UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE Augustin Cournot

ED221

THÈSE présentée par :

Fenja SCHEU

soutenue le : 26 Juillet 2023

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : « Sciences économiques »

Trademarks and Textual Data: A Broader Perspective on Innovation

(Marques et données textuelles : Une perspective élargie sur l'innovation)

THÈSE dirigée par :

M. COWAN Robin	Professeur, Université de Strasbourg
Mme. OTT Ingrid	Professeur, Karlsruhe Institute of Technology, Allemagne

RAPPORTEURS :

M. VANNUCCINI Simone	Professeur Junior, Université Côte d'Azur
M. SURE-VETTER York	Professeur, Karlsruhe Institute of Technology, Allemagne

AUTRES MEMBRES DU JURY :

M. LLERENA Patrick	Professeur, Université de Strasbourg
Mme. NIEKEN Petra	Professeur, Karlsruhe Institute of Technology, Allemagne

Trademarks and Textual Data:
A Broader Perspective on Innovation

(Marques et données textuelles : Une perspective élargie sur l'innovation)

Zur Erlangung des
akademischen Grades eines
Doktors der Wirtschaftswissenschaften
(Dr. rer. pol.)

von der KIT-Fakultät für Wirtschaftswissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte DISSERTATION

von

Pour obtenir le grade de

Docteur

dans la Discipline/ Spécialité «Sciences économiques»
de l'Université de Strasbourg (UniStra)

THÈSE DE DOCTORAT *approuvée*

de

MSc Fenja Scheu

Referentin/ Directrice de Thèse :	Prof. Dr. Ingrid Ott	Karlsruher Institut für Technologie, Deutschland
Referent/ Directeur de Thèse :	Prof. Robin Cowan, PhD	Université de Strasbourg, France
Tag der mündlichen Prüfung/ Jour de l'examen oral :	26.07.2023	

Acknowledgement

I would like to thank my supervisors, Ingrid Ott and Robin Cowan, for their advice and academic guidance during my Cotutelle programme. Their unwavering support has been crucial to the progression of my research. Special thanks to Ingrid Ott for providing the opportunity to pursue my PhD under her supervision, and to Robin Cowan for supporting the refinement of my research. Equally, thanks to my monitoring committee, Simone Vannuccini and Patrick Llerena, for their continuous support in my Cotutelle journey. My appreciation extends to my colleagues, especially Ramona Bodemer and Petra Kuchem-Braner for their invaluable organisational support and Nikas Scheidt for our fruitful collaboration. Thanks to the Stifterverband für die Deutsche Wissenschaft e.V. for their financial support of my research through the INNcentive funding programme. This thesis is part of the BMBF-funded research project "Comprehensive Assessment of Innovation Ecosystems (GEI-OE)."

I am truly fortunate that this PhD journey introduced me to invaluable friends. Heartfelt thanks to Anna Nguyen, who started as a colleague and became a dear friend. Our writing sessions together have been a cornerstone of my doctoral journey, and I am grateful for your companionship. My thanks also to my Bronnbacher friends who enriched my PhD journey with cultural and artistic joy. I hold a special place in my heart for Almuth Schwarz, with whom I have shared numerous moments of hardship and joy. Your constant presence was an important source of strength.

Finally, immense thanks to my partner Maximilian Löffel, your continuous motivation and insights were invaluable. Despite the challenges posed by my dissertation, your unwavering support helped me navigate this path. I also want to thank my beloved siblings, Sven, Arne, and Neele Scheu, and my parents, Hagen Scheu and Sylvia Würtele, being a constant source of motivation and encouragement. Your steadfast belief in me and your support have played a crucial role in reaching this milestone.

This achievement is the result of a long journey and I am deeply grateful to all the remarkable people who contributed to this journey.

Abstract

Innovation measurement as a driver of welfare is often based on patents. Patents are suitable for capturing technical innovations, but service innovations or innovations in low-technology areas are hardly covered. Trademarks are considered a complement for measuring innovation. Designs might also gain further importance as a linking element between intellectual property rights (IPRs). However, challenges are the combination of these IPRs, the need for more detail in the trademark classification system and the innovation linkage of trademarks to enable sophisticated analyses. Textual data of trademarks might solve these challenges and simultaneously combine trademarks with patents and designs to improve innovation measurement. Thus, this thesis considers textual data of trademarks and examines innovation and the use of intellectual property rights, focusing on combining trademarks with patents and designs. The thesis first addresses the use of textual data in innovation research to establish a general understanding before the use of textual data of trademarks for innovation measurement is dealt with intensively in the following chapters: A literature review first shows the possibilities of textual data analysis, such as the analysis of large data amounts, the combination of different data sources, the achievement of data-driven insights, the discovery of hidden information, and the mapping of technology developments. Patents and publications are still the most common data source used for textual data. Clustering approaches are commonly applied, while only some studies apply classifications. The overview contributes to a better understanding of textual data analysis in general. Trademarks are then combined textually with patents and compared with findings from the literature regarding the coverage of innovations in patents and trademarks to assess if a combination is possible and if the combination adds additional insights into innovation. The analysis is performed on Robotics, representing a high-technology sector, and Footwear, a low-technology sector. A combination is possible, depending, however, on the subject covered. Trademarks display different text structures depending on the topic covered, leading to different results. The textual data shows more detail in the high-technology subjects than in the low-technology subjects. In terms of insights added, the two data sources cover innovations in general, as well as service, product and technical innovations. Trademarks are still more in services and products, and patents are more in technical innovation. Overall, the textual combination of patents and trademarks provides broader coverage and more detail on innovation, depending on the area under consideration. Finally, trademarks, patents and designs are textually combined to measure the technological transformation and firm involvement in the Musical Instrument sector, which is transforming from classical to electronic to digital musical instruments. The application of IPRs is further examined. The analysis reveals that data and digitalisation are becoming more relevant, which are covered in patents. Firms from high-technology sectors with a background in software and hardware are involved in the sector's transformation. These firms predominantly use patents to protect their technologies. In addition, gaming firms cover aspects such as signal processing and video games and protect these with trademarks. Designs present themselves as a linking element between the trademark and patent topics. Overall, it is revealed that firms with different backgrounds contribute to the transformation, that the IPR usage depends on the firm background and that the textual data of different IPRs cover different aspects of the transformation. The methodology allows the differentiation between firms and their areas of involvement, providing additional information for future research. The combination thus for capturing transformation in the sector and firm involvement to address economic questions. To conclude, the thesis contributes to a better understanding of textual data in general and of trademarks in combination with patents and designs in innovation research. The integration of textual data of trademarks in the measurement of innovation broadens the coverage of innovation. Further research is necessary to improve the selection of trademarks, the focus on innovative trademarks and the integration of the insights in economic modelling.

Deutsche Zusammenfassung

Die Messung von Innovationen als Wohlfahrttreiber basiert häufig auf Patenten. Diese sind geeignet, technische Innovationen zu erfassen. Dienstleistungsinnovationen oder technologiearme Innovationen werden jedoch kaum erfasst. Marken werden als Ergänzung zur Innovationsmessung angesehen und auch Designs könnten als Bindeglied zwischen Schutzrechten weiter an Bedeutung gewinnen. Herausforderungen sind jedoch die Kombination dieser Rechte, die Detaillierung des Markenklassifizierungssystems und die Innovationsverknüpfung von Marken, um anspruchsvolle Analysen zu ermöglichen. Textdaten von Marken könnten diese Herausforderungen lösen. Diese Arbeit befasst sich daher mit diesen, untersucht Innovation und die Nutzung von geistigen Eigentumsrechten, wobei der Schwerpunkt auf der Kombination von Marken mit Patenten und Designs liegt. Ein Literaturüberblick zeigt zunächst die Möglichkeiten der textuellen Datenanalyse auf, wie z.B. die Analyse großer Datenmengen, die Kombination verschiedener Datenquellen, die Erzielung datengetriebener Ergebnisse, die Entdeckung verborgener Informationen und die Abbildung von Technologieentwicklungen. Patente und Publikationen sind nach wie vor die am häufigsten genutzte Textdatenquelle. Clustering-Ansätze werden häufiger als Klassifizierungen verwenden. Der Überblick trägt zu einem besseren Verständnis der Analyse textueller Daten bei. Marken werden dann textuell mit Patenten kombiniert und mit Erkenntnissen über die Erfassung von Innovationen in Patenten und Marken verglichen, um zu beurteilen, ob eine Kombination möglich ist und ob diese zusätzliche Erkenntnisse über Innovationen liefert. Die Analyse wird für den Hochtechnologiesektor Robotik und den Niedrigtechnologiesektor Schuhe durchgeführt. Es wird festgestellt, dass eine Kombination möglich, jedoch vom behandelten Thema abhängig ist. Die Marken weisen je nach Thema unterschiedliche Textstrukturen auf, was zu unterschiedlichen Ergebnissen führt. Die Textdaten sind in den Hochtechnologiethemen detaillierter als in den Niedrigtechnologiethemen. Die beiden Datenquellen decken Innovationen im Allgemeinen sowie Dienstleistungs-, Produkt- und technische Innovationen ab. Marken sind nach wie vor eher bei Dienstleistungen und Produkten zu finden, Patente eher bei technischen Innovationen. Insgesamt bietet die textliche Kombination von Patenten und Marken eine breitere Abdeckung und je nach betrachtetem Bereich mehr Details über Innovationen. Schließlich werden Marken, Patente und Designs textlich kombiniert, um den technologischen Wandel von klassischen zu elektronischen Musikinstrumenten und die Beteiligung von Firmen im Musikinstrumentensektor zu messen. Die Anwendung von Rechten des geistigen Eigentums wird weiter untersucht. Die Analyse zeigt, dass Daten und Digitalisierung an Bedeutung gewinnen und über Patente erfasst sind. Firmen aus Hochtechnologiesektoren mit einem Hintergrund in Software und Hardware sind an der Transformation des Sektors beteiligt, welche überwiegend Patente zum Schutz ihrer Technologien nutzen. Gaming-Firmen decken Aspekte wie Signalverarbeitung und Videospiele ab und schützen diese mit Marken. Designs stellen sich als verbindendes Element zwischen den Marken- und Patentthemen dar. Es zeigt sich, dass verschiedene Firmen zur Transformation beitragen, dass die Nutzung von Schutzrechten vom Firmenhintergrund abhängt und dass die textlichen Daten verschiedener Schutzrechte unterschiedliche Aspekte der Transformation abdecken. Die Methodik ermöglicht eine Differenzierung zwischen den Firmen und deren Tätigkeitsbereichen, und liefert so zusätzliche Informationen. Die Kombination ermöglicht es also, den Wandel im Sektor und die Beteiligung der Unternehmen zu erfassen, um wirtschaftliche Fragen zu beantworten. Diese Arbeit trägt folglich zu einem besseren Verständnis von Textdaten im Allgemeinen und von Marken in Kombination mit Patenten und Designs in der Innovationsforschung bei. Die Einbeziehung von Textdaten von Marken in die Innovationsmessung erweitert den Erfassungsbereich von Innovation. Weitere Forschung ist notwendig, um die Auswahl der Marken, den Fokus auf innovative Marken und die Integration der Erkenntnisse in die ökonomische Modellierung zu verbessern.

Résumé Français

La mesure de l'innovation en tant que moteur de la prospérité se base souvent sur les brevets, qui sont appropriés pour saisir les innovations techniques. Les marques sont considérées comme un complément permettant d'appréhender les innovations de services ou les innovations à faible contenu technologique, et les designs pourraient également gagner en importance en tant que lien entre les droits de propriété intellectuelle. Cependant, les défis sont la combinaison de ces droits, le détail du système de classification des marques et le lien d'innovation entre les marques. Les données textuelles des marques pourraient résoudre ces défis. Ce travail se penche donc sur ces derniers, examine l'innovation et l'utilisation des droits de propriété intellectuelle, en se concentrant sur la combinaison des marques avec les brevets et les designs. Une revue de la littérature montre tout d'abord les possibilités offertes par l'analyse des données textuelles, telles que l'analyse de grandes quantités de données, la combinaison de différentes sources de données, l'obtention de résultats axés sur les données, la découverte d'informations cachées et la cartographie des évolutions technologiques. Les brevets et les publications restent la source de données textuelles la plus utilisée. Les approches de clustering sont plus souvent utilisées que les classifications. Cet aperçu contribue à une meilleure compréhension de l'analyse des données textuelles. Les marques sont ensuite combinées textuellement avec les brevets et comparées avec les connaissances sur la couverture des innovations dans les brevets et les marques afin d'évaluer si une combinaison est possible et si elle fournit des connaissances supplémentaires sur les innovations. L'analyse est effectuée pour le secteur de haute technologie de la robotique et le secteur de basse technologie des chaussures. On constate qu'une combinaison est possible, mais qu'elle dépend du thème traité. Les marques présentent des structures textuelles différentes selon le thème, ce qui donne des résultats différents. Les données textuelles sont plus détaillées dans les thèmes de haute technologie que dans les thèmes de basse technologie. Les deux sources de données couvrent les innovations en général ainsi que les innovations en matière de services, de produits et de technologie. Les marques continuent à être plus présentes dans les services et les produits, tandis que les brevets sont plus présents dans les innovations techniques. Dans l'ensemble, la combinaison textuelle des brevets et des marques offre une couverture plus large et, selon le domaine considéré, davantage de détails sur les innovations. Enfin, les marques, les brevets et les dessins et modèles sont combinés textuellement pour mesurer l'évolution technologique des instruments de musique classiques vers les instruments de musique électroniques et la participation des entreprises dans le secteur des instruments de musique. L'application des droits de propriété intellectuelle fait l'objet d'une analyse plus approfondie. L'analyse montre que les données et la numérisation gagnent en importance et sont couvertes par des brevets. Les entreprises des secteurs de haute technologie avec un arrière-plan de logiciels et de matériel sont impliquées dans la transformation du secteur, qui utilisent principalement des brevets pour protéger leurs technologies. Les entreprises de jeux couvrent des aspects tels que le traitement des signaux et les jeux vidéo et les protègent par des marques. Les designs se présentent comme un élément de liaison entre les thèmes des marques et des brevets. Il s'avère que différentes entreprises contribuent à la transformation, que l'utilisation des droits de propriété intellectuelle dépend du contexte de l'entreprise et que les données textuelles de différents droits de propriété intellectuelle couvrent différents aspects de la transformation. La méthodologie permet de différencier les entreprises et leurs domaines d'activité, et fournit ainsi des informations supplémentaires. Cette combinaison permet donc d'appréhender la transformation du secteur et l'implication des entreprises afin de répondre à des questions économiques. Ce travail contribue à une meilleure compréhension des données textuelles et des marques en combinaison avec les brevets et les designs dans la recherche sur l'innovation. Les données textuelles des marques élargissent la portée de l'innovation. Des recherches supplémentaires sont nécessaires pour améliorer la sélection des marques, l'accent sur les marques innovantes et l'intégration dans la modélisation économique.

Contents

List of Figures

List of Tables

List of Queries

1 Introduction

Innovation is a fundamental driver of economic growth and social development. Firms are eager to innovate and be innovative to stay competitive (Schumpeter 2010). Major innovations like the steam engine or semiconductors are only some of the innovations that had major impacts on the development of society (Kuznets 1973). Not only large but also small innovations shape our lives. Recent advances can be observed in the provision of services where digitalisation is enabling music or video streaming, car-sharing, online banking or e-commerce. Innovation impacts economic growth, technological progress, investments, productivity, and resource allocation. It can take various forms, be it technological innovations, new goods or new services being introduced, an improvement in processes or organisational change. The market introduction is an essential aspect of an innovation (OECD and Eurostat 2005). As innovation is important for the welfare of a country and the competitiveness of firms, policymakers worldwide are trying to make their regions innovative, hoping to spur economic growth and social development. The policy should therefore be based on evidence. Innovation research intends to provide the evidence needed to guide policymakers in good decision-making.

Most of the evidence on innovation and innovation policy has been based on patent data. Patents are a means for firms to protect their inventions from imitation and gain a monopoly right on their invention (Neuhäusler 2009). Patents are considered a formal protection mechanism and are part of intellectual property rights (IPR) (WIPO 2020d). For patent protection, an inventive technological step is required, and the patent needs to publicly describe the invention (WIPO 2021h). Patent protection is globally available. Patents are systematically registered and cover long-time spans (Basberg 1987). The patent system further provides a sophisticated classification system that simplifies patent searches and allows for detailed research. This makes patents interesting for innovation policy, as they are data sources that reveal information on the current technological state-of-the-art on a global scale (Kleinknecht et al. 2002). In innovation research, patents tend to be used as an indicator for product innovation (Dziallas and Blind 2019).

Even though patents have their advantages as an indicator, they also have some shortcomings:

- Not every patent covers a commercialised invention. This means that the invention is not introduced into the marketplace. Therefore, these patents only cover inventions, not innovations (Basberg 1987). It is thereby not distinguishable which patents are used in the marketplace. For innovation policy, it is therefore of interest to cover the diffusion aspect to ensure a market introduction, as this is necessary to bring progress and affect economic growth.

- Patent application differs across sectors and firms. As patents require an inventive step in a technological area, sectors with lower technological content are underrepresented (Kleinknecht et al. 2002; Neuhäusler and Frietsch 2015). This does, however, not mean that there is no innovation in these sectors. It is just not captured in patent-based innovation indicators (Hirsch-Kreinsen 2008). Research showed that low-technology sectors contribute to economic growth (Mendonça 2009).

- Patents have difficulties covering services, tacit knowledge, or software inventions (Millot 2009; Mendonça 2009; USPTO 2017; USPTO 2015). These, however, are increasingly important. First, industrialised countries have major service sectors that depend on services being provided. Second, digitalisation is impacting not only technological sectors but all sectors. Therefore, a shift

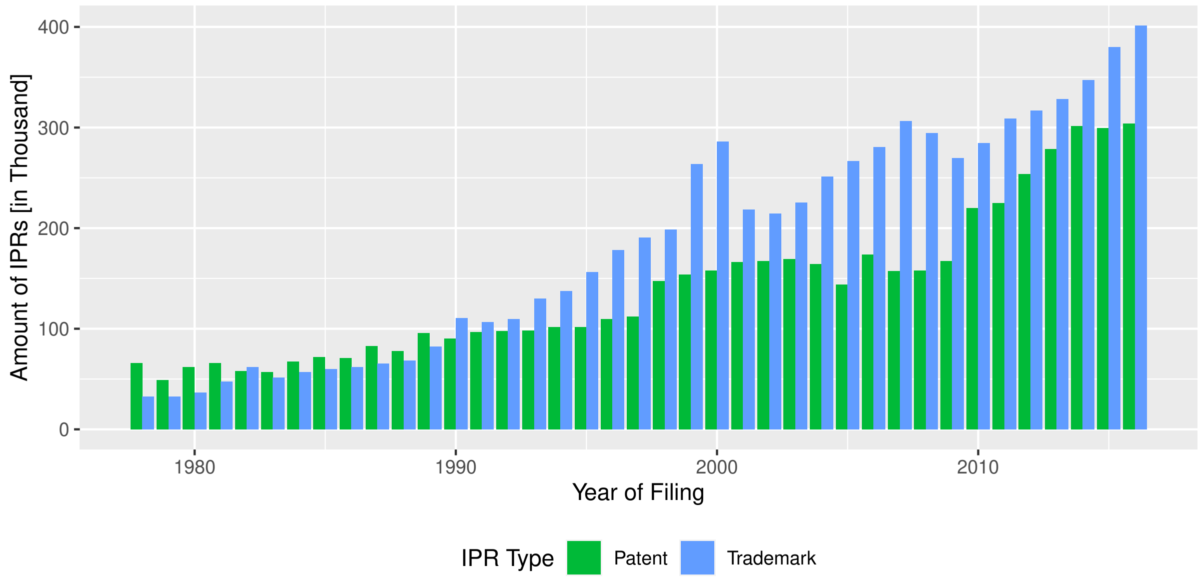


Figure 1.1: Total Yearly Development of Granted USPTO Patents and Registered USPTO Trademarks from 1978 to 2016.
The figure depicts the patents granted and trademarks registered by the USPTO from 1978 to 2016, with years referenced to the application filing year.
Source: Own representation based on USPTO data.

in innovation activities should be observed. Relying only on patent data might convey a limited perspective. Patent-based indicators could have overlooked the services mentioned above.

One possibility to include these perspectives and, in principle, overcome those shortcomings are trademarks. Trademarks can take various forms, mostly known are names representing a brand. For example, a trademark like “Apple Music” is foremost a sign or brand used to protect the name but can simultaneously stand for innovation in music provision. The name mostly requires registration for specific areas that indicate the usage of the trademark to attain trademark protection. Thus, the trademark prevents confusion and ensures distinctiveness between different goods or service providers (WIPO 2019b). Like patents, trademarks are also registered, widely available, and cover long-time spans (Flikkema et al. 2015), but are also applicable in all sectors and for product and services (S. J. H. Graham et al. 2013; Millot 2009). They are also formal protection mechanisms and intellectual property rights (WIPO 2020d). Trademark registrations increased in recent years and, for example, surpass patent registration in the United States Patent and Trademark Office (USPTO) (see Figure 1.1). Unlike patents, trademarks often require a market introduction, enabling a diffusion perspective. They are applicable in services and low-technology sectors. Overall, trademarks are of interest to innovation research. Millot (2009), Mendonça et al. (2004), and S. J. H. Graham and Hancock (2014) suggest already their use as an innovation indicator, pointing out their ability to cover not only technical product innovation but also service, marketing or business process innovation.

Nevertheless, trademarks also have shortcomings when it comes to innovation research:

- The link of trademarks to innovation is not ensured. In contrast to patents, an inventive step is not a requirement during trademark registrations. That makes trademarks more complex as an innovation indicator. Several authors, therefore, research the linkage of trademarks and innovation: Seip et al. (2018) find that trademarks are registered at different stages of the innovation process, depending on the innovation to be protected. Mendonça et al. (2004) argue that innovative firms are more active in trademarking than in patenting based on results of the Third Community Innovation Survey (CIS 3). Flikkema et al. (2014) could show in their survey that above 50% of the trademark applications in their sample link to innovation, of which service innovation is captured mostly only in trademarks and not in patents. So an innovation linkage is present. It is, however, not ensured for trademarks. Nevertheless, in certain areas with limited patentability, trademarks are one of the only formal protection mechanisms covering innovation activities in these areas.

- Another shortcoming of trademarks is their limited granularity. In patents, the patent classification system is very sophisticated and elaborates on various innovation areas. The system classifies each patent according to its technological content, allowing for detailed analyses of specific fields and relations between these fields (WIPO 2020c). The trademark classification system, however, provides fewer opportunities for innovation analysis, as the categories are broader and less specific. For several million trademarks, only 45 NICE classes are commonly available (WIPO 2022a), compared to over 75,000 categories in the patenting system (EPO 2022). This becomes a challenge when using trademarks for specific innovation areas as this implies that the level of analysis remains general, limiting the level of granularity that can be attained.

Thus, two challenges need to be addressed: First, trademarks do not always cover innovation, and second, the granularity of trademark analysis based on current classifications remains limited. Bringing trademarks and patents together could address the innovation challenges of trademarks and address the shortcoming of patents. In contrast to trademarks, patents always cover inventions but not always innovation. Together, these protection mechanisms could allow us to cover inventions and ensure their market application. Further, the insights gained on technological innovations based on patents could be enhanced with innovations in services or low-technology sectors based on trademarks. Malmberg (2005) combine trademarks and patents in the pharmaceutical industry to capture innovation activities. They find that both IPRs link to innovation activities in the field, with trademarks covering short-term activities. The authors suggest using both IPRs for valuable insights but also point out that industry-specific differences exist in the applicability of trademarks as an innovation indicator. Ribeiro et al. (2022) assess the correlation of trademarks and patents for different sectors. The authors find that trademarks are especially of interest in sectors with low patentability or where patents are not applied, services, and less developed countries. Their results indicate that the combination of trademarks and patents improves the understanding of technological and non-technological sectors. Even though joint analyses of patents and trademarks might be useful, integrating trademarks with patents is not straightforward. A possibility is the use of concordance tables to combine trademarks with patents. Zolas et al. (2017) link trademarks on economic data based on keyword matching to derive insights on firms' export behaviour. The use of keywords enables already a more detailed perspective on trademarks and enables a class-level perspective. However, the authors point out a need for additional approaches if further granularity is intended. Other authors like Flikkema et al. (2015) integrate patents and trademarks via the same legal entities that registered the different IPRs. However, this approach always limits itself to direct links between patents and trademarks. Firms that are only active in trademarks or only active in patents but contribute to an innovation field and its development and diffusion are overlooked. Further, this makes it hard to combine the advantages of the different data sources, as much information is lost.

The challenge of limited granularity of trademark analysis still needs to be solved. Trademarks provide more granular information, containing descriptions of the trademark application for goods and services. These descriptions are textual data that can be exploited. The data can provide detailed information on the application areas of trademarks and innovations that are protected with trademarks. Text analysis covers techniques to identify patterns from unstructured, textual data to generate data-driven insights (Aggarwal and Zhai 2012). Antons et al. (2020) argue for the application of text analysis to improve existing and develop new measurements in innovation research.

Authors value or apply textual data analysis for various reasons:

- Text analysis enables the analysis of large data sets: This is handy when the data sets to be analysed surpass the amount of data that humans can process in a feasible time. Ozcan et al. (2021) base their analysis on 22,891 tweets from Twitter, Larsen and Thorsrud (2019) use 459,745 newspaper articles, or Feng et al. (2020) apply their analysis on 41,994 patents. Applying text analysis to textual data enables the coverage of larger amounts. Several authors name this argument as one of the main reasons to apply textual data analysis (Loshin 2013; Antons and Breidbach 2018; Bakhtin et al. 2020; Kayser 2017; J. Kim and C. Lee 2017; N. Kim et al. 2015; Kohler et al. 2014; Shen et al. 2020; B. Wang and Z. Wang 2018; Zhu and Porter 2002).

- Text analysis allows for the analysis of new data sources: As textual data are contained in various sources, it is possible to include data sources into the analysis that could not be analysed with standard approaches. For example, Dahlke et al. (2021) integrate fundamental human needs and innovations to assess how crises shape innovation. The authors use web-scraped innovation project descriptions for the innovation state and compare them to nine fundamental human needs described in the literature. J. Kim and C. Lee (2017) extract futuristic data from websites like Siemens, MIT technology review, or the World Future Society to extract scientifically relevant future topics and compare them to the state-of-the-art revealed in patent data. They intend to identify weak signals and extract novel topics by doing so. Mi et al. (2021) take advantage of survey data of older people to determine the demand for gerontechnology and compare the insights gained to reports of emerging technologies and patents to forecast future trends. A final example is Fiordelisi et al. (2019) that use 10-K reports of firms to extract information on corporate creativity to measure the impact on firm innovativeness and value. The authors find that creative culture fosters innovation. These are only some examples of how new data sources reveal interesting insights to be included in economic models. Although text analysis opens up more possibilities for data use and other data sources are available, patents are still the most commonly used data source. However, the amount of possible textual data available for innovation research is large and increases every year: In 2020 alone, over 2.6 million new research articles were published on Web of Science (WoS 2021). Regarding intellectual property rights such as patents, trademarks and designs, 304,126 patents, 234,444 trademarks and 28,886 designs were registered in 2016 alone in the U.S., with a total of 5,439,151 registered patents, 4,246,859 registered trademarks and 527,902 registered designs from 1978 until 2016.[1]

- Text analysis allows for the combination of different data sources, or use of heterogeneous data: Barriers imposed by the difference in the data structure of different data sources can be overcome with the usage of language and shared words. Several authors take advantage of this: Bakhtin et al. (2020) combine various data sources from scientific articles, patents, news, and awards to analytical reports with natural language processing to identify emerging patterns in agriculture and food production to provide insights for governments and firms for strategic planning activities. Kayser (2017) integrate media with the abstracts of publications via matched terms for an innovation diffusion perspective, while Shen et al. (2020) compare the textual content of patent and publication clusters to discovering opportunities of scientific advancement and technological innovation in smart health monitoring technologies. Combining several data sources can thus reveal new insights, and text analysis helps with the integration.

Another important argument in this context is the time-efficient analysis options (Bakhtin et al. 2020; Chiarello et al. 2021; Dahlke et al. 2021; Kayser 2017; J. Kim and C. Lee 2017; Zhou et al. 2020; Zhu and Porter 2002), which enable quick results and therefore quick statements. This is particularly relevant in rapidly changing fields where new findings are constantly being published. In addition, data-based knowledge extraction is relevant, so that knowledge can be gained independently of experts (Bakhtin et al. 2020; Basole et al. 2019; Larsen and Thorsrud 2019; Mi et al. 2021; Shen et al. 2020; Song et al. 2017; Zhou et al. 2020). This reduces expert bias and allows more people to generate knowledge. Overall, the application of textual analysis in innovation research can reveal interesting insights. Studies have been done, for example, on the current development of technologies like biofuels Curci and Mongeau Ospina (2016) or blockchain (Chiarello et al. 2021), the diffusion of technologies like big data in scientific research (Y. Zhang et al. 2019) or upcoming topics in innovation like batteries or charging connectors in

[1] Own data based on Query A.1, Query A.1 and Query A.1

electric vehicles (Feng et al. 2020). The textual analysis can further be used to generate new ideas for innovation areas by extracting ideas from public sources like Twitter (Ozcan et al. 2021) or identifying promising areas by comparing consumer needs with available technologies (Mi et al. 2021). In the context of firms, text analysis can further reveal insights on the convergence of firms (N. Kim et al. 2015) or make the creativeness of firms traceable (Fiordelisi et al. 2019).

The various arguments for text analysis make it also interesting for the analysis of trademarks: Large amounts of trademarks are available for analysis. The textual data of trademarks can be used to extract detailed insights into different innovation areas and aspects of diffusion. Moreover, the textual data of trademarks allows us to bridge trademarks to other data sources. The combination of trademarks with different data sources can be of interest to capture, e.g., services but also the development of products more broadly. The combination could be achieved textually between trademarks and patents. This implies that shared words among trademarks and patents link the different intellectual property rights. This would also overcome the large differences between the available classification systems in trademarks and patents. The main assumption is that the description of an invention is similar across different intellectual property rights, but the area of protection and application is changing, with a slight shift of focus when it comes to the information conveyed. This means that patents cover more of the technological aspects, while trademarks shed light on the market application and diffusion of an invention. In this thesis, the combination of trademarks with patents is, thus, primarily of interest. However, designs are also considered an interesting data source for innovation areas with high importance on other shapes and patterns. Designs are applied to protect novel ornamental elements (WIPO 2021c). Like trademarks, designs are currently not included as a standard data source in analyses of innovation. In this work, they are included in Chapter 5, as they are of interest in the field under consideration. Like trademarks, they might contribute to a more accurate measurement of innovation and could provide interesting insights.

First attempts to use textual data of trademarks for analysis have been made: Flikkema et al. (2019) derive brand strategies of creation, extension or modernisation based on mark names and NICE coverage. Semadeni (2006) compares the wording of different firms' trademark applications to determine the firms' relative position to each other. The wording of trademarks can further reveal the likelihood of service innovations to be imitated (Semadeni and Anderson 2010). Even though some examples of the use of textual data of trademarks exist, it has yet to be common to apply textual data analysis. Castaldi (2020) points out in her overview that the focus is still primarily on structured, non-textual data of trademarks that are used for analysis and that the trademark information is not used yet to their full potential.

In terms of combination, textual data of trademarks and patents have been combined: Authors like M. Lee and S. Lee (2017) use the patents of firms to display technological knowledge and obtain details on the market positioning of competitors based on the trademark descriptions. H. Kim et al. (2017) match patents and trademarks broadly with keywords shared among patents and trademarks to uncover how technological knowledge can be applied in the market and identify potential areas for diversification. Patents thereby provide the invention perspective, while trademarks reveal areas of market introduction of these inventions. However, the approaches to combine the data sources still mostly focus on keywords. Another methodology to extract information from textual data that considers more than just keywords from the textual data is topic modelling. Topic modelling is an approach for extracting the topics occurring in a set of documents based on the words in these documents (Blei et al. 2003). In contrast to keyword-based matching, topic modelling can also consider synonyms or the context of words. Example applications in the context of innovation research are, for example, Basole et al. (2019), who identify topics and clusters in the entrepreneurial ecosystem via topic modelling and reveal that clusters of entrepreneurs are not only industry driven but also related to the similarity of their topics. Another example is Larsen and Thorsrud (2019), who apply topic modelling to identify valuable topics in news articles that are then integrated into economic models to improve the prediction of economic booms.

Overall, textual data of trademarks is the focus of this thesis. The information in the textual description of the application area of trademarks in goods and services could provide a broader perspective on innovation as they cover technical and non-technical areas and applications in goods and services. The textual data further allows for the combination of trademarks with patents that primarily consist of textual data, such as abstracts that summarise the content or patent claims that cover the protected areas. Analysing the textual data of trademarks is one solution to extract more information from trademarks and simultaneously provide a possibility to combine trademarks with other data sources. By doing so, this thesis extends innovation coverage with the textual inclusion of trademarks to provide a broader perspective on innovation. The broad innovation coverage supports empirically-based policy recommendations. However, to take advantage of the potential of textual data of trademarks, the data needs to be better comprehensively assessed, and the characteristics of the data source must be elaborated.

Several aspects need to be considered in the context of this thesis to extract the advantages of textual data of trademarks for innovation research:

Trademarks are a data source that provides, e.g. a perspective on services or low-technology areas but also covers high-technology areas. Trademarks, however, have the drawback of unclear innovation linkage and limited levels of detail for sophisticated analyses.

Textual Data can provide detailed information and, in that sense, provides an additional perspective in combination with structured data. Further, as textual data is available in various data sources, it holds the potential to allow for new combinations of data and, thus, new insights gained from these combinations. In the context of textual data, however, it remains to be seen how the possibility of their analysis is used in innovation research. It is of interest which methods and data are used to address different research questions.

Textual Data of Trademarks allows for the combination of trademarks with other data sources, thereby providing a broader perspective on innovation in general through the inclusion of services and low-technology innovation in the analyses. However, it needs to be made clear where trademark texts differ from, for example, patent texts and, thus, where they add value. Patents could provide the inventive step, while trademarks ensure market application. By that, the coverage of innovation could be ensured.

Innovation in Textual Data of Trademarks is not ensured. Some trademarks are linked to innovation, but not every trademark is. Therefore, it is still being determined how this affects the analysis of trademark text data and whether trademark textual data can provide interesting insights into activities in innovation areas, even if only to shed light on diffusion aspects.

Overall, the thesis addresses these aspects and intends to keep the strength of trademarks while simultaneously addressing the limitations of this data source using textual data. To build an understanding of trademarks and, in that context, textual data analysis, the thesis is structured as follows: In Chapter 2, the general background on innovation, patents, trademarks and designs as part of intellectual property rights and innovation measurement is given. Further, textual data analysis in general and with a focus on the main method used in this thesis is presented. This knowledge serves as a general foundation for the thesis. Following the Foundations, Chapter 3 focuses on the current use of textual data for innovation research in general. This literature review provides a general understanding of the state of the art of textual data, its analysis, reasons to apply textual data analysis and potential application areas in innovation research. Chapter 4 and Chapter 5 then perform textual data analyses of trademarks in combination with other data sources. Chapter 4 focuses on how trademarks and patents can be effectively combined to give a broader picture of innovation. It assesses the textual combination and how aspects of innovation are covered in each data source but with a focus, especially on trademarks. The underlying assumptions of this chapter are that insights on trademarks of structured and text-based data differ and that textual data of trademarks generally contribute to a broader understanding of innovation. The analysis is performed on Robotics and Footwear as representatives of a high-technology and a low-technology, respectively. The insights gained serve as a foundation for further analyses of innovation based on textual data of trademarks. By contrast, Chapter 5 combines patents, trademarks and designs to cover the transformation of the musical instrument sector from low technology to high technology and introduces a focus on firms. Even though one would expect an increase in patent applications in the sector due to the increasing relevance of electronic instruments, trademarks are the relevant intellectual property right in the sector, not only for classic but also for electronic musical instruments. The analysis contributes to a better understanding of this phenomenon by providing insights into the intellectual property rights application in the context of technological transformation and firm background. The thesis concludes in Chapter 6, where general insights are reflected and discussed, especially concerning textual data of trademarks, and the main conclusion of the thesis is drawn. In general, the work contributes to a better understanding of textual data of trademarks. It sheds light on textual data in innovation research as a whole, the contribution of trademarks combined with other data sources and uses the insights gained to answer economic questions.

For the general structure of the thesis, it can thus be said that the chapters start with a general perspective on textual data and become more specific over the course of the thesis: The thesis focuses first on textual data in general, then looks at textual data of trademarks in comparison to patents before finally applying the knowledge gained to assess technological transformation concerning intellectual property right use, covering patents, trademarks and designs. The structure of the thesis is graphically represented in Figure 1.2. Here, the commonly coloured shape contours of the boxes indicated shared aspects between the different chapters of the thesis. The legend can be found on the left.

Innovation Areas/ Sectors: All chapters cover different aspects of innovation (yellow). The foundation's section provides the general aspects of innovation and its measurement. The review in Chapter 3 is not restricted to an innovation area or sector. It is only relevant that the articles in the sample are published in a business or an economic area, cover innovation and use textual data. Chapter 4 compares Robotics as a high-technology sector with Footwear as a low-technology sector to gain generalisable insights. Lastly, Chapter 5 looks at the transformation of musical instruments, which covers aspects of low- and high-technology applications due to classical and electronic musical instruments.

Intellectual Property Rights: Intellectual property rights (violet) with a focus on trademarks are present across all parts of the thesis. The foundation's chapter presents the background of intellectual property rights and especially patents, trademarks and designs and the measurement of innovation, as well as an introduction to the textual data analysis method applied in this thesis. In Chapter 3, intellectual properties are only indirectly considered as part of the textual data sources being used in some of the analysed publications. Trademarks and patents are then closely assessed in Chapter 4 and Chapter 5 with the addition of designs in the latter. However, the main focus of the thesis is on trademarks.

Textual Data: Textual data (green) and its analysis are relevant in all chapters. The analysis of textual data is described in the foundation's chapter. This thesis uses Structural Topic Modelling developed by Roberts et al. (2019). Topic Modelling is a text analysis technique to derive topics out of a large number of text documents. Structural Topic Modelling is a specific topic modelling approach that identifies topics in existing documents under the consideration of additional information (Roberts et al. 2019) (see further details in Section 2.4). Chapter 3 analyses published articles using textual data in innovation research. The articles in the sample are not restricted to a specific data source. Chapter 4 then analyses patents and trademarks, while Chapter 5 additionally includes designs into the analysis.

Figure 1.3 provides an integrated perspective on the chapters. In the course of the thesis, the innovation areas covered are narrowed down to increase the focus, level of detail, and the integration of high- and low technology (triangle in yellow). The areas are broadest in the literature review of Chapter 3, where all kinds of areas in relation to innovation and textual data are assessed. In Chapter 4, high-technology Robotics and low-technology Footwear are analysed as innovation areas. In contrast, high- and low-technology aspects in Musical Instruments are jointly considered in Chapter 5. The triangle in green

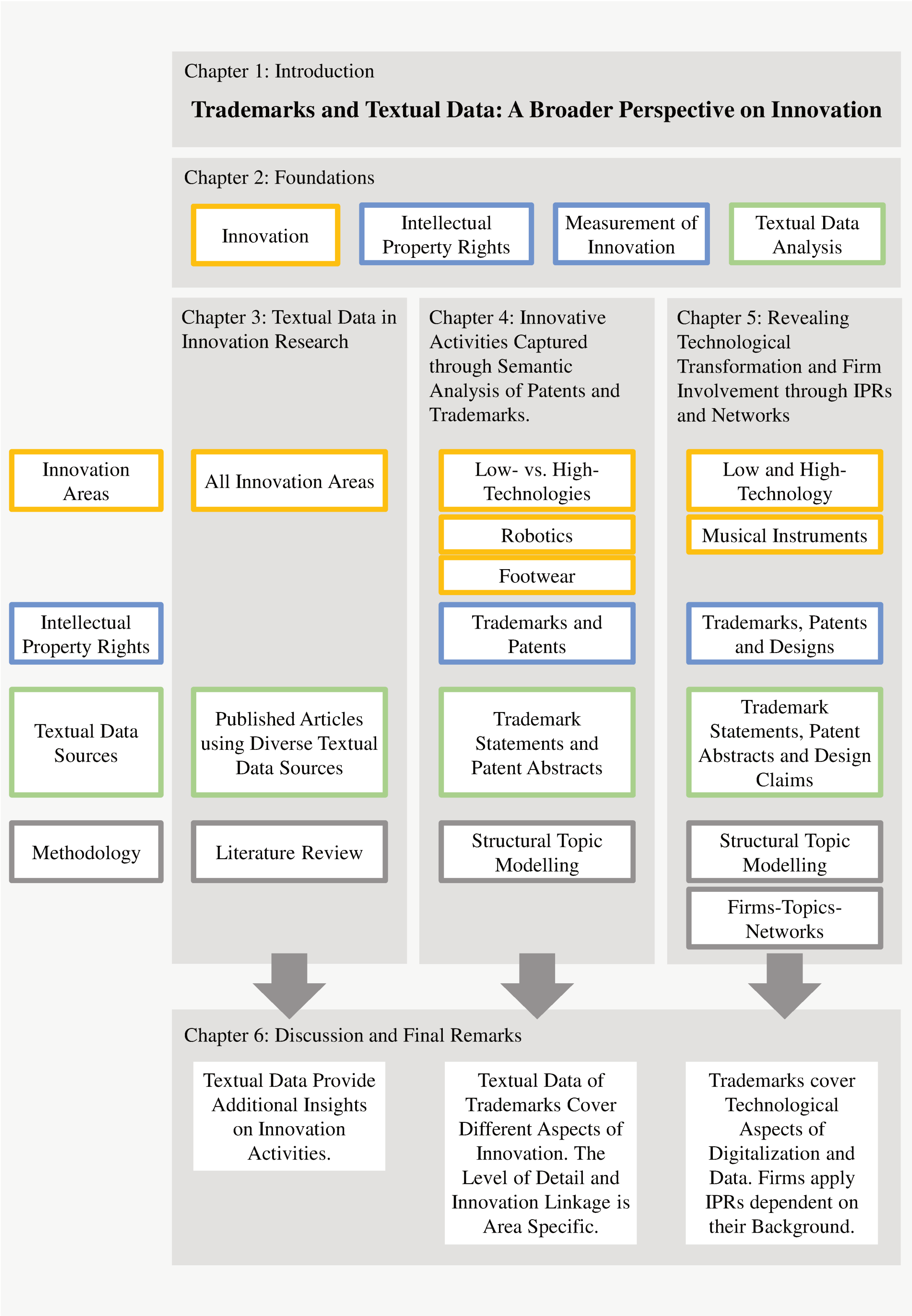


Figure 1.2: Structure of the Thesis.
The figure illustrates the thesis structure. Chapter 1 introduces the thesis, while the foundation chapter explores key themes such as innovation, intellectual property rights, innovation measurement, and textual data analysis, which are relevant throughout the entire thesis. Chapters 3 to 5 delve into the application of textual data in innovation research, with a specific focus on trademarks, patents, and designs. In these chapters, the left-hand side boxes explain the meaning of shape and colour: blue signifies aspects related to intellectual property rights, green represents textual data sources for analysis, and orange denotes the covered areas of innovation. Grey is associated with the methodology. Chapter 6 offers a summary of the conclusions drawn from individual chapters and the overarching thesis.

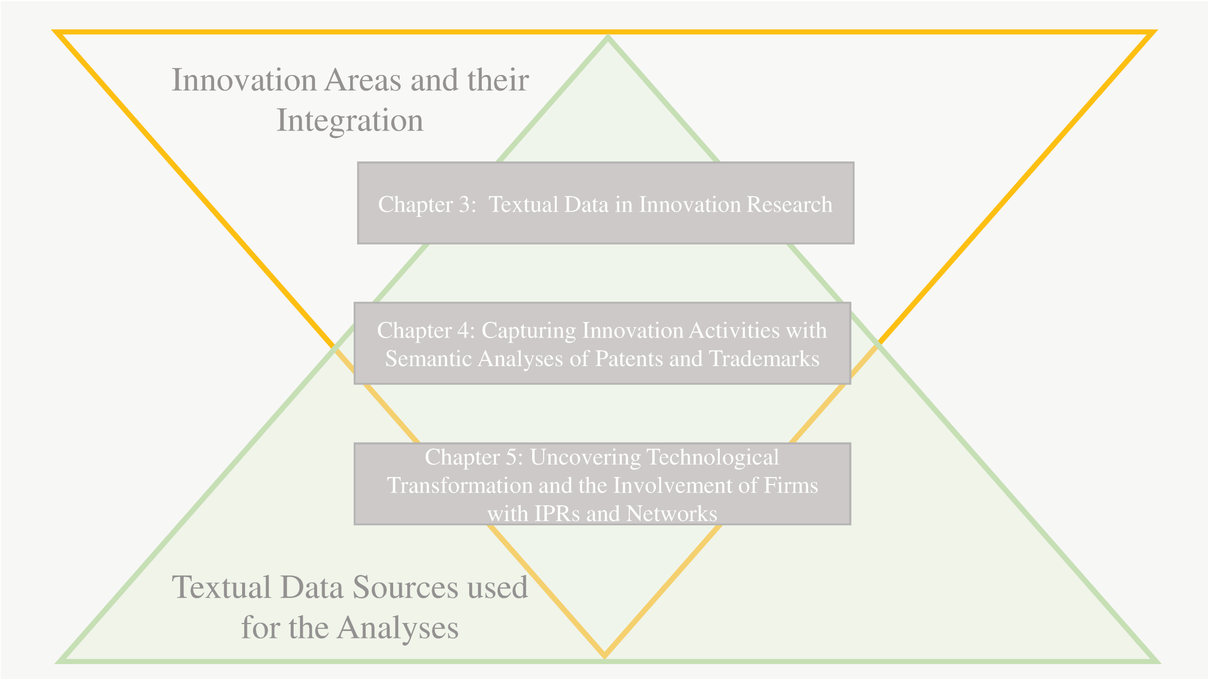


Figure 1.3: Integrated Perspective on the Research Chapters.
The triangle in yellow represents the breadth of innovation areas covered. The triangle in green stands for the breadth of data sources used.

represents the breadth of data sources used for the analyses in each chapter, which increases during the thesis. In the beginning, publications are considered, followed by a joint analysis of trademarks and patents, and concluding with a joint analysis of patents, trademarks and designs.

The remainder of the thesis goes as follows: First, further information on innovation, intellectual property rights, innovation measurements and textual data analysis is provided in Chapter 2. In Chapter 3, insights on textual data in innovation research are provided before performing different analyses on textual data of trademarks in Chapters 4 and 5. Finally, the findings of the thesis are summarised, and general topics are discussed before closing with the main contributions of the thesis in Chapter 6.

2 Foundations

In this chapter, the foundations of this thesis are presented. First, Section 2.1 addresses the meaning of innovation, its definition and diffusion. Section 2.2 deals with patents, trademarks and designs as intellectual property rights and their use for the protection of innovation. Section 2.3 then discusses the basis for measuring innovation. Finally, Section 2.4 lays the foundation of the analysis of textual data.

2.1 Innovation and Technology

Schumpeter (2010) identifies innovation as a driving force of economic growth. Through innovation, higher productivity can be reached, and new firms and industries emerge. Firms, therefore, are incentivised to keep being innovative to stay competitive. If innovation performance declines, they perish (Schumpeter 2010). Innovation impacts, e.g. technological progress, investments, productivity or resource allocation (OECD 2015). Kuznets (1973) argues that major innovations are the basis for structural transformation and modern economic growth enabled through scientific breakthroughs. A major innovation forms the basis for widespread applications and improvements, has an impact over a long period, and leads to economic transformation. The mass adoption of technological innovations interacts with scientific advances, driving further technological advances. Innovation is an important factor to be considered from a policy perspective.

Definition What is understood with "innovation" is not easily defined: According to Schumpeter (2010), innovation includes, but is not limited to, the introduction of new goods, development of new markets, technological changes, new business organisations and the recombination of existing factors in a new way. Henderson and Clark (1990) distinguish innovations based on their impact on linkages or the core concept of components. Incremental and modular innovations leave the connections untouched, while architectural and radical innovations change them. The core concept is discarded by modular and radical innovations and reinforced by incremental and architectural innovations. While Henderson and Clark (1990) look more at the types of innovation, the OECD and Eurostat (2005)'s definition sheds more light on the application areas of innovation. According to OECD and Eurostat (2005), product, process, marketing, and organisational innovation are the four distinguished types of innovation. The Oslo Manual is used as the basis for numerous innovation studies and policy analyses. For example, until 2016, the Community Innovation Survey (CIS), which collects information on innovation and enterprises in the European Union, assesses products against process innovation (Eurostat 2021a). In the *Oslo Manual 2018*, a revised definition with a broader innovation concept is developed. Therein, product innovation and business process innovation are distinguished. The definition of product innovation covers services, goods and knowledge-capturing products. The latter are offerings with characteristics of goods and services. Business process innovations are differentiated according to their business function reference. Functional areas include production, marketing and sales, information and communication systems, administration and management, or product and business process development. Common to both types of innovation is that a significant difference must be present compared to the firm's existing products or processes. New products must be introduced to the market, while processes require use within the firm (OECD and Eurostat 2018). The Community Innovation Survey in 2018 (Eurostat 2021a) adopts the differentiation to product and business process innovation. As can be seen, innovation is relevant in several areas, with varying definitions being applied. Baregheh et al. (2009) address this issue and develop the following overarching definition based on the understanding of innovation across multiple disciplines and over time:

> "Innovation is the multi-stage process whereby organisations transform ideas into new/improved products, services, or processes, to advance, compete and differentiate themselves successfully in their marketplace." Baregheh et al. (2009, p.1334)

Innovation Diffusion The successful introduction of innovation in the marketplace and, thus, the diffusion of innovation is highly relevant for considering innovation. Already Ryan and Gross (1943) analysed diffusion based on hybrid corn in Iowa, U.S. It is a method of breeding that replaced former methods within several years. Ryan and Gross (1943) find that the dimensions of time, acceptance, information channels and individual characteristics are relevant. The diffusion pattern is slow initially, rises steeply, and declines at the end. These elements are also part of the definition of Rogers (1983), which defines diffusion as:

> "The process by which (1) an innovation (2) is communicated through certain channels (3) over time (4) among the members of the social system." (Rogers 1983, p. 10)

To summarise, innovation has a major impact on the welfare of societies. It occurs in various areas and, per definition, requires a market application. Its diffusion can replace existing, e.g., products, processes, or services. Due to its importance, it is necessary to capture it broadly. Intellectual property rights such as patents, trademarks and designs are formal mechanisms to do so. While patents cover the state-of-the-art, diffusion considerations and the market introduction of inventions are key arguments for using trademarks for innovation measurement and, thus, of interest in this thesis.

2.2 Intellectual Property Rights Protection

As innovations are essential for firms and often require significant investments, firms apply formal and informal protection mechanisms to protect their innovations from imitation and ensure returns from their investments (Rammer 2002; Neuhäusler 2009). Formal protection mechanisms include, among other things, patents, trademarks, and copyrights, while informal protection mechanisms are, e.g., secrecy, complexity, or time to market (Rammer 2002). The mechanisms serve for protection and are also used to measure innovation activities. Especially patents are often considered (Dziallas and Blind 2019). However, they are not applied in the same way across all industries and sectors: Neuhäusler (2009) find that R&D-intensive industries apply more patent protection than lower R&D-intensity industries . Rammer (2002) compare trademarking and patenting and show that patents are used in higher numbers in the manufacturing sector than trademarks. In contrast, trademarks are used more than patents in the service sector. Thomä and Bizer (2013) find that most firms do not apply formal protection mechanisms in the small enterprise sector. Their innovations are either not protected or only protected with informal mechanisms such as secrecy, complexity, or design. This is in line with studies from Levin et al. (1987) and Cohen et al. (2000). According to Cohen et al. (2000), firms protect their product inventions with several mechanisms like patents, secrecy, and lead time advantage. The manufacturing industry prefers secrecy and lead time advantages over patents. Larger firms prefer patents. Firms use their patents to profit from licencing and commercialisation, to push for negotiation or prevent competition through blocking. The survey of Levin et al. (1987) reveals that most firms consider patent protection not appropriate for the protection of new processes. Patents did score better in the protection of products. However, lead time, learning curve, sales or services were still regarded as more effective in protecting product inventions. Patents displayed limited effectiveness in most industries, except for the pharmaceutical industry with drug development. Hall et al. (2014) also find that patents are important in the chemical, pharmaceutical, and medical industries. Yet, across all industries, lead time is often preferred to protect innovation, and patents are considered ineffective. The authors further find that the focus on patents neglects copyright and trademarks, which are widely used. The use of the protection mechanism is generally determined by the invention's characteristics like product or process, discreteness, and complexity. The authors point out that a better understanding on an invention level instead of a firm level is needed to understand better the different mechanisms to protect inventions.

Trademarks, patents and designs are considered intellectual property rights (IPRs) and protect the "creations of mind" (WIPO 2020d, p. 1). While patents, for example, promote and protect innovation, trademarks identify the origin of a product or service (Gorman 2017). Intellectual property protection is regulated under the *Paris Convention for the Protection of Industrial Property*, which was established in 1883 and last amended in 1979. It applies to patents, industrial designs, trade and service marks, among other things. The convention ensures that common rules apply, that foreign applications are treated the same as national ones and that a priority application exists[1] (WIPO 1979). The convention applies to all members of the Union. In 2022, 179 countries took part (WIPO 2022b).

In the following, further information on patents, trademarks and designs is provided, with a major focus on trademarks. Examples are provided to explain the textual data in each IPR. This information serves as a basis for the textual work in Chapters 4 and 5.

2.2.1 Patents

> "A patent is an exclusive right granted for an invention, which is a product or a process that provides, in general, a new way of doing something, or offers a new technical solution to a problem." (WIPO 2021h)

Patents protect inventions (EPO 2018). Besides protection, motives to patent are blockage, reputation, exchange, commercialisation, licensing, and incentives. All motives are applied, but sectoral differences exist (Blind et al. 2006; Cohen et al. 2000). The patent application requires a new invention that improves the status quo, is applicable in the industry and concerns a "physical activity of 'technical character' " (EPO 2018, Part G, Ch. III-1).[2] Patents thus require the disclosure of something new but do not require a market introduction. With patent protection, a monopoly is granted to the owner of the patent for a limited period, generally 20 years, in exchange for the public disclosure of the invention (WIPO 2021h). For the duration of the patent protection, the patent owner has the exclusive right to use the invention and can decide on its use by others. Patent protection is limited to a geographical scope determined by the granting authority (WIPO 2021h). The "United States Patent and Trademark Office (USPTO)" grants patents that are valid for the U.S. market (USPTO 2015), while European patents are granted by the "European Patent Office (EPO)" (EPO 2018). In addition, there are other patent authorities such as the "Deutsches Patent und Markenamt (DPMA)" in Germany (DPMA 2021b) or the "Institut National de la Propriété Industrielle (INPI)" in France, each of which is responsible for national protection. The geographical protection sought determines the responsible patent authority. Patents are arranged globally hierarchically according to the "International Patent Classification (IPC)" System. This system structures the patents according to their technical content into sections, classes, subclasses, and groups (WIPO 2020b). About 75,000 groups are in use. It facilitates finding information and working with patents (WIPO 2020c). An extension of the system is the "Cooperative Patent Classification (CPC)" system with around 250,000 entries provided by the EPO and USPTO that also covers new technical fields (EPO 2022). An example in the context of the thesis is CPC subclass "G10H" that covers electrophonic musical instruments with section G "Physics" and G10 "Musical Instruments" (USPTO 2022a).

[1] Priority implies that the applicant for protection has twelve months in case of patents or six months in case of industrial designs or marks to seek protection in other countries. If another applicant seeks protection, the applicant of the first regular application can refer to its priority (WIPO 2021i). This means that if the criteria for protection are met, only the time of application is decisive for deciding who receives the protection rights.

[2] An invention of technical character concerns a technical problem in a technical field with a technical feature (EPO 2018).

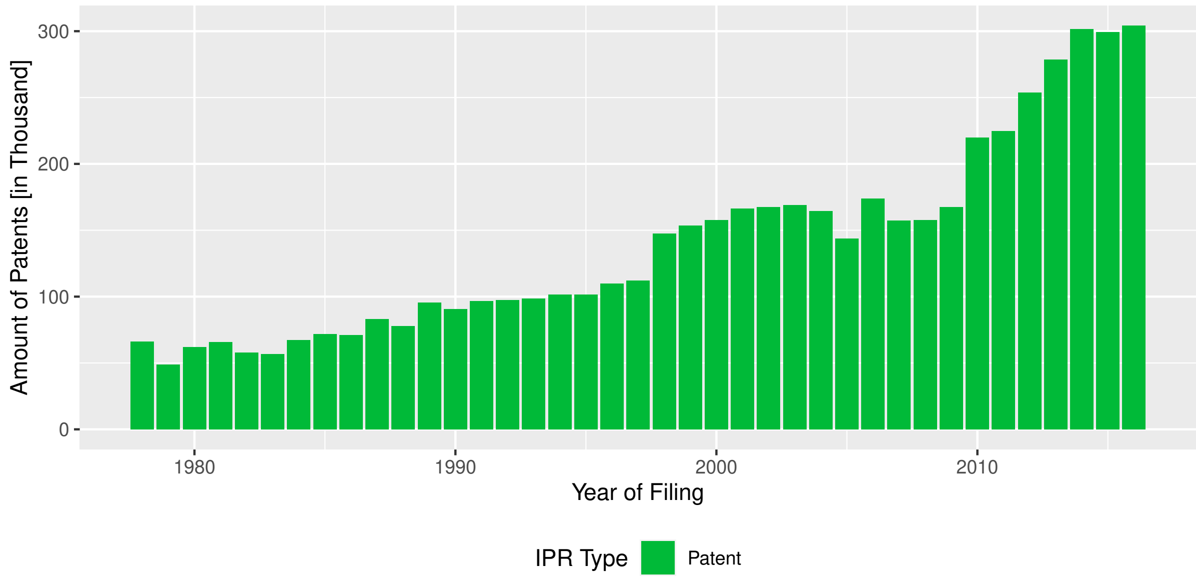


Figure 2.1: Total Yearly Development of Granted USPTO Patents between 1978 and 2016.
The figure depicts the patents granted by the USPTO from 1978 to 2016, with years referenced to the application filing year.
Source: Own representation based on USPTO data (see Query A.1).

For the analyses in this thesis, granted USPTO utility patents are considered. In the U.S. system, a distinction is made between design, utility, and plant patents. Utility patents are the most common and cover, e.g., processes or machines (USPTO 2022c). Utility patents are referred to in this thesis as patents. In total, 5,439,151 patents were registered from 1978 until 2016. The yearly development can be seen in Figure 2.1.

As textual data are used in this thesis, it is further necessary to understand the structure of a patent document. An example of musical instruments is U.S. patent 4,788,896. The patent's front page is displayed in Figure 2.2. In total, the patent document consists of 37 pages. On the front page on the top left, the inventors of the invention described in the patent, here Uchiyama and Suzuki (1988), are named. The patent number and registration date are found on the top right. On the left, further information is provided, like the title, assignee or filing date. The patent was filed in June 1986 and registered in December 1988 by "Nippon Gakki Seizo Kabushiki Kaisha", a predecessor of Yamaha Corporation. The CPC class G10H is also found on the left. On the right, the patent abstract is displayed. This thesis uses the abstracts of patents for textual data analysis. The other pages of the patent document contain further illustrations and a detailed description of the invention and the scope of protection.

Patents overall provide information on inventions. The introduction of the protected invention is thereby not a requirement. An innovation, however, does require introduction in the marketplace. Trademarks could provide this market use.

2.2.2 Trademarks

Unlike patents, trademarks often require introducing protected goods and services to the market. However, an inventive step is not necessary for a trademark to be valid, only an application in a market. An example of a musical instrument trademark document can be found in Figure 2.3. Trademarks protect a name, here "Parallel Digital Imaging". The textual description of "Sampled sound tone generator for electronic organs" introduces a trademark application. The trademark is registered in the context of Nice Class 15 "Musical instruments" which implies that the trademark protection is foremost for the area of musical

United States Patent [19]
Uchiyama et al.

[11] **Patent Number: 4,788,896**
[45] **Date of Patent: Dec. 6, 1988**

[54] **TONE GENERATOR HAVING A VARIABLE NUMBER OF CHANNELS WITH A VARIABLE NUMBER OF OPERATING UNITS**

[75] Inventors: **Yasuji Uchiyama; Shigeru Suzuki,** both of Hamamatsu, Japan

[73] Assignee: **Nippon Gakki Seizo Kabushiki Kaisha,** Hamamatsu, Japan

[21] Appl. No.: **875,479**

[22] Filed: **Jun. 18, 1986**

[30] **Foreign Application Priority Data**

Jun. 21, 1985 [JP] Japan 60-135793

[51] **Int. Cl.⁴** **G10H 1/06;** G10H 7/00
[52] **U.S. Cl.** .... **84/1.01;** 84/1.03; 84/1.19; 84/1.24; 340/365 S
[58] **Field of Search** .... 84/1.01, 1.03, 1.19, 84/1.24, DIG. 12, DIG. 22; 340/365 S

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,882,751	5/1975	Tomisawa et al.	84/1.01
4,018,121	4/1977	Chowning	84/1.01
4,543,869	10/1985	Kawashima et al.	84/1.01

FOREIGN PATENT DOCUMENTS

29519 6/1983 Japan .

Primary Examiner—Stanley J. Witkowski
Attorney, Agent, or Firm—Spensley Horn Jubas & Lubitz

[57] **ABSTRACT**

A tone generator includes a plurality of tone generation channels each comprising one or more operation channels. One operation channel performs basic operation of a tone signal computing operation (e.g., FM or AM) and a tone signal is generated by using one or more operation channels for one tone generation channel. Tone generation channels in the tone generator are established according to a selected performance mode. In a first mode, the operation channels are divided into N groups and N tone generation channels are established in correspondence to the respective operation channel groups. In a second mode, the operation channels are divided into M groups and M tone generation channels are established in correspondence to the respective operation channel groups wherein N is a different number from M. Tone generation channels of the number required according to the selected performance mode can be established by efficiently utilizing the operation channels of a limited number.

10 Claims, 9 Drawing Sheets

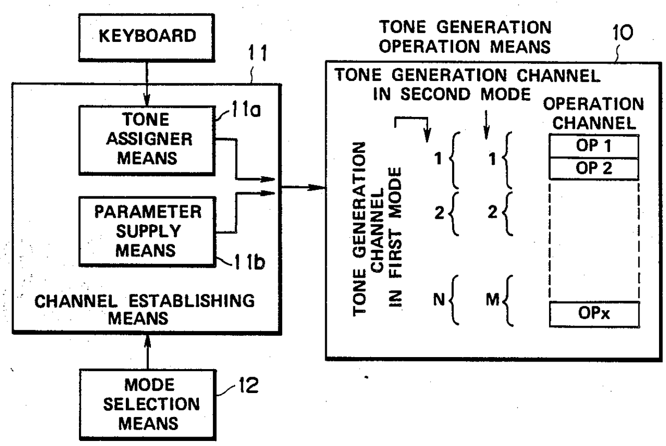


Figure 2.2: Example of U.S. (Utility) Patent No. 4,788,896.

The figure illustrates the front page of a U.S. patent for tone generators. The inventors, Uchiyama et al., are listed at the top left. The patent number and date of patent registration can be found at the top right. The patent filing date is situated next to the number 22. *Source: Front Page of U.S. Patent No. 4,788,896 (Uchiyama and Suzuki 1988).*

Int. Cl.: 15

Prior U.S. Cls.: 21 and 36

Reg. No. 1,719,447

United States Patent and Trademark Office **Registered Sep. 22, 1992**

TRADEMARK
PRINCIPAL REGISTER

PARALLEL DIGITAL IMAGING

RODGERS INSTRUMENT CORPORATION (OREGON CORPORATION)
1300 N.E. 25TH AVENUE
HILLSBORO, OR 97124

FOR: SAMPLED SOUND TONE GENERATOR FOR ELECTRONIC ORGANS, IN CLASS 15 (U.S. CLS. 21 AND 36).

FIRST USE 12-20-1990; IN COMMERCE 12-20-1990.

NO CLAIM IS MADE TO THE EXCLUSIVE RIGHT TO USE "DIGITAL" AND "IMAGING", APART FROM THE MARK AS SHOWN.

SN 74-024,399, FILED 1-31-1990.

J. TINGLEY, EXAMINING ATTORNEY

Figure 2.3: Example of U.S. Trademark Registration No. 1,719,447 (Serial No. 74,024,399).
The figure presents the U.S. Trademark bearing Serial No. 74,024,399. However, the trademark registration number, which is assigned post-registration, is 1,719,447. The trademark is identified as "parallel digital imaging" and falls under Nice class 15, registered for the purpose of a sampled sound tone generator for electronic organs.
Source: Registration Certificate of U.S. Trademark Serial No. 74,024,399 (USPTO 1990).

instruments (see also Table 2.1). Textual data of trademark applications are used in this thesis as a basis for textual data analysis.

Trademark registrations are viewed as a protection of marketing investments and are valued by financial markets. Established trademarks are thereby valued more highly than new trademarks (Sandner and J. Block 2011) and can account for a significant share of a firm's market capitalisation (WIPO 2013). However, the effect of trademark families [3] on the market valuation of large firms differs: brand development, extension and modernisation are valued. In contrast, brand creation and hedging have no effect (J. H. Block et al. 2014). Further, Castaldi and Dosso (2018) find a positive direct effect on growth premium for high-tech sector firms with more than average trademarks. Firms with more trademark filings than the median signal their capabilities to perform in the market. The demand for and use of trademarks correlates with economic cycles: For example, the financial crisis led to a falling demand for trademarks (Myers 2013) and deGrazia et al. (2019) show that declining first uses of trademarks predict business cycle downturns. Seip et al. (2018) argue that the timing of trademark application differs across sectors, product or service innovation and firm types and relates to different stages within the innovation process. In their empirical analyses of trademark applications, the authors find that the timing of trademarking depends on the firm and innovation to be protected. Earlier-stage trademarking is related to start-ups, especially in the case of product innovation, and to firms with trademarking experience. The high-technology manufacturing industry applies late-stage trademarking. Service innovations cannot be related to a specific stage but are applied in all stages. Further, according to De Vries et al. (2017), factors like high market concentration, business-to-consumer markets and venture capitalists increase the likelihood of start-ups first filing trademarks instead of patents. However, existing trademarks, other

[3] Trademark families are groups of trademarks with common characteristics like "Pampers" being protected by more than 40 different trademark applications (J. H. Block et al. 2014).

protection means or joint innovations with others decrease the likelihood of innovators trademarking their innovation (Athreye and Fassio 2019). Likewise, Bei (2019) finds that firms without valuable trademarks prefer internal innovation. The likelihood to commercialise innovations from external sources when entering new industries increases with the applicability of valuable trademarks in another industry. Overall, trademarks are applied across different companies and sectors for different reasons and at different stages of the innovation process.

Definition Formally seen, a trademark is "used to identify products" (EUIPO 2019) or to "identify and distinguish the goods/ services of one seller or provider from those of others, and to indicate the source of the goods/ services" (USPTO 2019). According to the definition of WIPO (2019b), "a trademark is a sign capable of distinguishing the goods or services of one enterprise from those of other enterprises." Trademarks are used for product identification and differentiation (EUIPO 2019; Millot 2009). Besides, for example, words, names, symbols or signs, trademarks can take the form of sounds, motions or multimedia marks (USPTO 2019; EUIPO 2019). Both services and goods can be covered under a "trademark" (analogous to the definition of (WIPO 2019b)). In some cases, these are also referenced separately as "service marks" or "trademarks" ("trade marks"). This thesis uses the term "trademark" to refer to service and goods applications.

According to Berlit (2019), trademarks need to fulfil five different legal functions:

1. The *origin function* makes the origin of the goods or services clear to the customer.
2. The *differentiation function* enables to distinguish firms or products from each other.
3. The *quality function* indicates the quality to be expected from the brand.
4. The *investment function* builds a brand value and generates additional value from licensing.
5. The *advertising function* eases the advertisement of goods and services.

In the registration process, the focus is particularly on the functions of origin and differentiation (Berlit 2019). The origin and differentiation functions ensure that the link between a product or service offering and its firm becomes evident in the market. In the service sector, the differentiation effect of trademarks can support firms in protecting their services that supplement "new to the market" product innovations (Crass and Schwiebacher 2017). The characteristics of trademarks further favour new product innovations or innovations in knowledge-intensive services, as their use reduces the likelihood of substitution through confusion between these innovations and competing offers (Crass and Schwiebacher 2013). Apart from the mentioned functions, trademarks can enhance products or become functional parts of them. Examples are fake Rolex watches, bought because of their name and not because of the product (Kozinski 1993).

For innovation research, the relationship between trademarks and innovation is particularly relevant. The legal functions checked during the registration process do not ensure the innovation reference. While trademark applications require the brand's use, they are not checked for their inventive step. Innovation is not a necessity in trademarks. Despite this, researchers find a link between trademarks and innovation. Flikkema et al. (2015) find in their study on Benelux SMEs that close to 60% trademark applications related to new innovations and that especially service innovations are mostly only covered by trademarks. The authors conclude that trademarks can capture non-technological inventions, intangibles and services, commercialised inventions, and new inventions to the firm or sector. Mendonça et al. (2004) find that trademark use in innovative firms is higher than patent use. The application of trademarks or patents in non-innovative firms is lower in both cases. Product-related trademarks still make up the largest part of trademark applications, but the share of service marks is rising. The authors further advocate measuring innovations with new trademark applications rather than firms' trademark portfolios to capture innovation activities instead of established products. Research thus implies that trademarks can trace innovations. So although the innovation content of trademarks is not checked in the registration process, they are often related to innovation or filed by innovative firms. This is an important prerequisite to serve as an innovation indicator.

Protection Trademark protection can be attained worldwide, but the protection scope depends on the applied trademark legislation. The German Patent and Trademark Office (DPMA) is responsible for trademark registrations in Germany. International registrations are filed with the World Intellectual Property Organization (WIPO), while EU trademarks are filed with the European Intellectual Property Organization (EUIPO) (Berlit 2019). The USPTO is responsible for registering trademarks in the U.S. (S. J. H. Graham et al. 2013). This thesis will first refer to German trademark law before looking at the differences between the European and U.S. trademarking system to illustrate the legal situation. Although the focus will be on U.S. trademarks and U.S. law in the remainder of the thesis, looking at German and EU regulations allows for a better understanding of the differences and similarities in worldwide trademark protection.

The German trademark law dates to the "Reichswarenschutzgesetz" of 1874. In 1967, compulsory use was introduced. Service marks can be registered since 1979. A comprehensive overall reform of the trademark law took place with the "Markenrechtsreformgesetz" (Trademark Law Reform Act), which was promulgated in the "Bundesgesetzblatt" (Federal Law Gazette) on October 25, 1994, (Berlit 2019). In the U.S., the state of California first introduced trademark registration in 1863. Federal registration was not introduced until 1870; before that, trademarks were only protected by common law (Duguid 2018). The introduction of a federal U.S. registration became necessary during globalisation and increased international trade, as European countries required prior registration of trademarks in the country of origin to be accepted in Europe (Gorman 2017). The 1870 legislation was overturned by the Supreme Court in 1879, making further legislation necessary (Duguid 2018). The Paris Convention of 1883 created international intellectual property law to ensure the same principles of design, patent and trademark protection applied internationally (Higgins 2012). In 1946, the Laham Act brought U.S. jurisdiction up to international standards and introduced service mark protection (Duguid 2018). As can be seen, trademarks have been applied for a long time. They have been in use longer than patents or designs Gorman (2017).

According to §4 MarkenG[4], trademark protection can be obtained by registration or by reputation, which is obtained through use[5]. However, registration offers the advantage that the time of protection and the geographical scope of protection is determined. The trademark must be sufficiently distinguishable from existing trademarks to obtain trademark protection (Berlit 2019). The trademark protection and authority determine the geographical scope of protection: The German Federal territory is considered the area of protection for trademarks registered in Germany. The protection of EU trademarks extends to the territories of the member states of the EU. International trademark protection is determined by the countries to which the protection extends and thus differs for each international trademark. The protection period of the trademarks is initially ten years and can be extended indefinitely if the prerequisites for extension are still fulfilled (Berlit 2019). Therefore, there exist trademarks that have had their protection for a long time. Well-known examples are Nivea (U.S. Serial Number: 71069150, Registered in 1913) or Coca-Cola (U.S. Serial Number: 71254696, Registered in 1928). Use of the trademark must be made under §26 MarkenG. If a trademark is not used for five years without interruption, a cancellation of the trademark can be requested (§49 MarkenG). Similarly, for EU trademarks, use must take place within five years of registration; otherwise, the registration may not be renewed (Section 3 Article 18 Clause 1 EUTMR).[6] In the U.S., proof of mark use and the specified areas of protection must be provided at the time of application according to §1(a) of the Trademark Act, 15 U.S.C. §1051(a). Since 1989, intent-to-use (ITU.) registrations have been possible (Myers 2013). If use is only intended but has not yet occurred,

[4] **markeng_1994_GesetzUberSchutz** as last amended on December 11, 2018

[5] Protection through use is not applicable for European trademarks (Berlit 2019)

[6] *Regulation (EU) 2017/1001 of the European Parliament and of the Council of 14 June 2017 on the European Union Trade Mark (Text with EEA Relevance.)*

an ITU. may be made (Section 1(b) of the Trademark Act, 15 U.S.C. §1051(b)). The applicant then has six months from approval to demonstrate use. The time may be extended upon request in 6-month increments to a maximum of 36 months (Section 1(d) of the Trademark Act, 15 U.S.C. §1051(d)).[7] Due to the different use requirements upon application in the U.S. compared to the European system, the Nice class coverage differs in the systems. Von Graevenitz et al. (2021) compare the scope of registration of Community Trademarks (CTM)[8] and U.K. trademark to their U.S. equivalents based on the word list provided in trademark descriptions. They find that the coverage of U.K. and CTM trademarks is more extensive than in the U.S. As the U.S. has a stricter use requirement than the EU or U.K., this implies that European or U.K. trademarks protect products or services not used in the market, preventing the use of these trademarks for others. The use requirement is important for this thesis and the main reason to focus on U.S. trademarks for the textual data analyses. The requirement ensures that trademarks cover actual market introduction.

U.S. Trademarks for Analyses Even though trademarks have been filed in the U.S. at the federal level since 1870, not all trademarks are available for analysis at the United States Patent and Trademark Office (USPTO). Trademarks registered before electronic recording and without renewal are not included. Trademarks registered as of 1978 are predominantly present. As of 1982, non-registered trademarks are also included in full (S. J. H. Graham et al. 2013). From 1978 to 2016, 7.22 Million trademarks have been applied for in the U.S., of which 4.24 Million trademarks have been registered. The trademark applications and registrations increased over time, with around 400 thousand applications in 2016 compared to below 100 thousand before 1990 ((see Query A.3)). The development can be derived from Figure A.1. Since the mid-1990s, registrations have increased. However, most marks are not renewed and are cancelled before the sixth year.

Nice Classification For the analyses, the trademark classification system is of interest. Trademarks are classified according to the Nice Classification, officially called the "International Classification of Goods and Services for the Purposes of the Registration of Marks", established in 1957 by the Nice Agreement and updated on a regular basis (WIPO 2021b). 89 countries contract the Nice Agreement. The agreement entered into force in the U.S. in 1972, while in France, the agreement had already been in force since 1961 or in Germany since 1962. (WIPO 2021g). The Nice Classification has been actively used in the U.S. since September 1, 1973. Before that, the U.S. applied its own classification system, which is still used in parallel (S. J. H. Graham et al. 2013). The latest version of the classification, which came into force on January 1, 2022, contains 45 classes, of which classes 1 to 34 are for goods, and 35 to 45 are for services. The classes organise the trademark applications and determine the main area of protection (see Table 2.1)) (WIPO 2022a).

Figure 2.4 displays the trademark registrations in the U.S., which are available for analysis. As can be seen, Nice class 9 "Science or research instruments" is the most important in the U.S., with 621 thousand registered trademarks, followed by class 35 "Advertising or business management" with 481 thousand registrations, class 41 "Education and training" with 438 thousand registrations and class 42 with 417 thousand registrations.[9] Trademark registration is possible in several Nice classes at the same time. In the U.S., however, the trademarks are mostly U.S. owned and mainly registered in only one Nice class. In case of an extension of use, the extension cannot be amended to the already registered trademark, but a new filing becomes necessary (Myers 2013). This is essential for innovation research because old trademark registrations then cover the at that time relevant status quo, so the temporal reference is ensured. Therefore, new registrations for an existing trademark also mean that the scope of this trademark application has expanded at a new point in time. As trademarks cover services in contrast to patents,

[7] *Chapter 15 - Trademark Act of 1946, as Amended*
[8] replaced by EU trademarks
[9] Class descriptions can be found in Table 2.1

Table 2.1: Overview of the Nice Classification, Version 2021-11

Class No.	Description	Class No.	Description
	Goods		
1	Chemical products	18	Leather and imitations of leather
2	Paints, colourants and preparations	19	Materials for building and construction
3	Non-medicated toiletry preparations	20	Furniture and parts
4	Industrial oils and greases, fuels and illuminants	21	Small, hand-operated utensils and apparatus for household and kitchen use
5	Pharmaceuticals and other preparations for medical or veterinary purposes	22	Canvas and other materials, raw fibrous textile materials
6	Common metals	23	Natural or synthetic yarns and threads for textile use
7	Machines and machine tools, motors and engines	24	Fabrics and fabric covers for household use
8	Hand-operated tools and implements	25	Clothing, footwear and headwear
9	Science or research apparatus and instruments, audiovisual and information technology equipment, safety and life-saving equipment	26	Dressmakers' articles, natural or synthetic hair for wear, small decorative items
10	Surgical, medical, dental and veterinary apparatus	27	Coverings for floors and walls
11	Environmental control apparatus and installations	28	Toys, apparatus for playing games, sports equipment, amusement and novelty items, certain articles for Christmas trees.
12	Vehicles and apparatus for transport	29	Foodstuffs of animal origin, vegetables and other horticultural comestible products
13	Firearms and pyrotechnic products	30	Foodstuffs of plant origin, auxiliaries intended for the improvement of food flavour
14	Precious metals, jewellery, clocks and watches	31	Land and sea products, live animals and plants, foodstuffs for animals
15	Musical instruments, parts and accessories	32	Non-alcoholic beverages, beer
16	Paper, cardboard, office requisites	33	Alcoholic beverages, essences and extracts
17	Electrical, thermal and acoustic insulating materials and plastics for manufacturing	34	Tobacco and articles used for smoking
	Services		
35	Advertising, business management, organisation and administration, office functions	41	Education, providing of training, entertainment, sporting and cultural activities
36	Financial, monetary and banking services, insurance services, real estate affairs	42	Scientific and technological services, research and design, industrial analysis, industrial research and industrial design services, quality control and authentication services, design and development of computer hardware and software
37	Construction services, installation and repair services, mining extraction, oil and gas drilling	43	Services for providing food and drink, temporary accommodation
38	Telecommunications services	44	Medical services, veterinary services, hygienic and beauty care, agriculture, aquaculture, horticulture and forestry services
39	Transport, packaging and storage of goods, travel arrangement	45	Legal services, security services for the physical protection, personal and social services
40	Treatment of materials, recycling of waste and trash, air purification and treatment of water, printing services, food and drink preservation		

The table offers an overview of the Nice Classification, Version 2021-11, used for trademark classification. There are 45 classes in total, comprising 34 for goods and 11 for services. The descriptions provided for each class offer a concise summary of their respective classifications.
Source: Own representation based on WIPO (2022a).

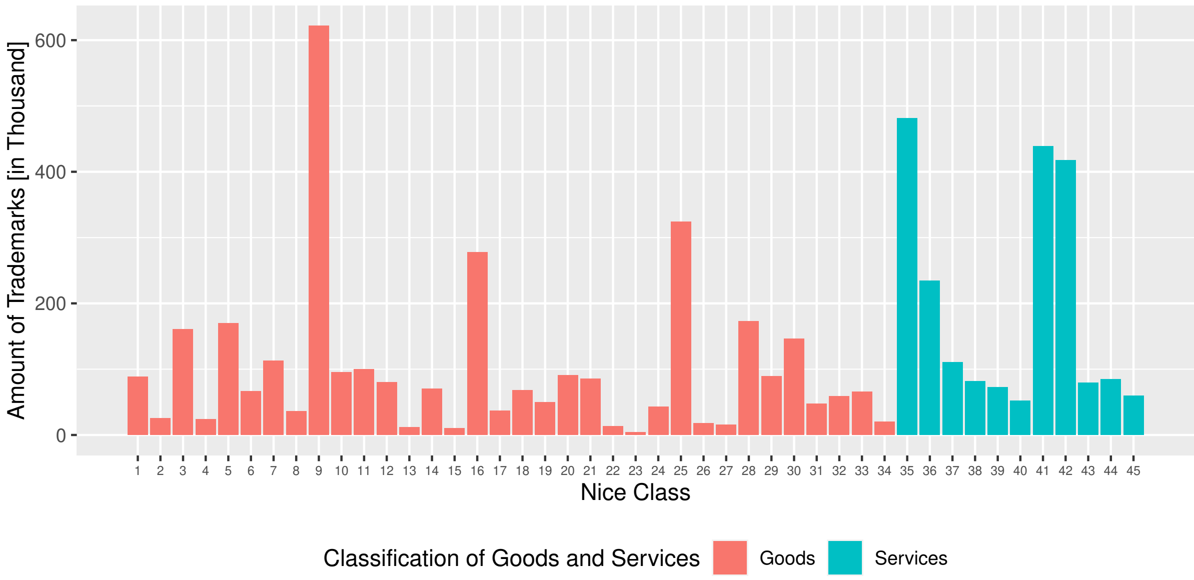


Figure 2.4: Nice Classes of Registered Trademarks between 1978 and 2016.
The figure illustrates the distribution of registered trademarks across various Nice classes at the USPTO, considering only trademarks registered between 1978 and 2016. Notably, classes 1 to 34 represent categories for goods, whereas classes 35 to 45 encompass services.
Source: Own representation based on USPTO data (see Query A.2).

Figure 2.5 displays the development of the goods and service registrations of trademarks from 1978 until 2016. As can be seen, the share of service trademarks and trademarks in absolute numbers increases over time.

2.2.3 Designs

Apart from patents and trademarks, this thesis considers designs as a protection mechanism. The reference of designs depends on the countries: most offer protection to registered designs under the industrial design law while others, like the USPTO (USPTO 2022b), apply patent law and refer to the designs as design patents (WIPO 2021c). During this thesis, industrial designs or design patents are referred to as “designs”. A “design constitutes the ornamental aspects of an article” (WIPO 2021c), consisting of two- or three-dimensional features of the article. They are applied to various products, from textiles to electronic devices. An example of a U.S. design is given in Figure 2.6. The figure displays design protection in musical instruments. Of interest for the textual data analysis is the design claim of an ornamental design for a musical tone controller. In total, 527,902 design patents are registered in the U.S. The registrations increased from around five thousand per year in the 1980s to above 20 thousand since 2006 (see Figure 2.7).

In general, designs are assigned to 32 classes according to the 13th version of the Locarno Classification, valid since January 1, 2021, (WIPO 2021e; WIPO 2021f). The Locarno International Classification was first introduced with the Locarno Agreement in 1968 (WIPO 2021a). The classification is valid in 59 countries, including Germany and France (WIPO 2021d). The European Union Intellectual Property Office (EUIPO) has compiled a European word list based on the Locarno Classification for registering designs (EUIPO 2021). The United States Patent and Trademark Office (USPTO) applies a design classification according to the U.S. Design Classification. The classifications are similar but not necessarily the same, and the U.S. system is more refined when it comes to the number of subclasses (USPTO 2020a; USPTO 2021c) (see Table 2.2). In the example mentioned above (see Figure 2.6), the design relates to U.S. Class D17 “Musical instruments”. In the U.S. system, class D14, “Recording,

Figure 2.5: Total Yearly Development in Goods and Services of Registered Trademark between 1978 and 2016.
The figure illustrates the yearly filings of registered trademarks within goods and services Nice classes at the USPTO from 1978 to 2016. It's important to note that if a trademark appears in multiple service or goods classes, it is counted only once.
Source: Own representation based on USPTO data (see Query A.2).

Communication, or Information Retrieval Equipment", is the most important class, followed by class D06, "Furnishings", (see Figure 2.8).

2.2.4 Combination of Protection Mechanisms

Each intellectual property right brings its mechanisms and makes it possible to protect different aspects. Firms combine different protection strategies to protect their innovations optimally. Amara et al. (2008) analyse eight different protection methods of knowledge-intensive business services. The mechanisms used by the firms are confidentiality agreements at 77%, lead time advantage at 61%, secrecy at 53.8%, copyright at 41%, complexity at 37.6%, trademarking at 34.5%, patenting at 15.7% and design protection in 11.8% of the cases. The authors find that design, patents, and trademarks are complementary legal methods. Firms use legal and informal methods jointly, whereby some protection mechanisms are used as complements, some as substitutes and some independent of each other that protect different aspects. Thus, managers "should pay attention not only to patents but also to the other protection mechanisms that complement the use of patents to protect innovations from imitation." In the case of Flikkema et al. (2014), 25% of the product innovations and 3% of service innovations are protected by trademarks and patents in SMEs. S. Graham and Somaya (2004) model the complementarity or substitution effect of copyrights, patents and trademarks as IPR protection strategies in the software industry. In the 1990s, patent and especially software patent protection grew in the U.S., while the scope of trademark protection remained limited. The authors find a complementary use of different types of intellectual property enforcement in the firm's legal process, particularly applying for copyrights and trademarks as plaintiffs and patents as defendants. J. H. Block et al. (2015) survey 600 mostly U.S. SMEs and find that while manufacturing firms use trademarks for marketing purposes to differentiate themselves from their competitors, they protect their inventions by other means. Service firms, in contrast, use trademarks in combination for marketing and exchange intentions, especially since trademarks are the most important property right in the service sector.Since innovations are protected by different means, separately and in combination, US00D347853S

United States Patent [19]

Masubuchi et al.

[11] **Patent Number: Des. 347,853**

[45] **Date of Patent: ** Jun. 14, 1994**

[54] **MUSICAL TONE CONTROLLER**

[75] Inventors: **Takamichi Masubuchi; Hisanori Kato,** both of Hamamatsu, Japan

[73] Assignee: **Yamaha Corporation,** Shizuoka, Japan

[**] Term: **14 Years**

[21] Appl. No.: **830,267**

[22] Filed: **Feb. 4, 1992**

[30] **Foreign Application Priority Data**

Aug. 6, 1991 [JP] Japan 3-23641

[52] **U.S. Cl.** **D17/99**

[58] **Field of Search** .... 84/600, 644, 670, 687, 84/718, DIG. 7; D17/99

[56] **References Cited**

U.S. PATENT DOCUMENTS

3,704,339	11/1972	Niinomi	84/687
4,542,291	9/1985	Zimmerman	250/231 R
4,635,516	1/1987	Giannini	84/600
5,005,460	4/1991	Suzuki et al.	84/600
5,022,303	6/1991	Suzuki et al.	84/600
5,125,313	6/1992	Hiyoshi et al.	84/600
5,127,301	7/1992	Suzuki et al.	84/600

FOREIGN PATENT DOCUMENTS

0264782	4/1988	European Pat. Off. .	
2029070	3/1980	United Kingdom .	
2221557	2/1990	United Kingdom	84/718

Primary Examiner—Melvin B. Feifer
Assistant Examiner—Adir Aronovich
Attorney, Agent, or Firm—Weingarten, Schurgin, Gagnebin & Hayes

[57] **CLAIM**

The ornamental design for a musical tone controller, as shown and described.

DESCRIPTION

FIG. **1** is a front elevational view of musical tone controller showing our new design;
FIG. **2** is a right-side view thereof;
FIG. **3** is a left-side view thereof;
FIG. **4** is a rear view thereof;
FIG. **5** is a top plan view thereof;
FIG. **6** is a bottom plan view thereof; and,
FIG. **7** is a front perspective view thereof.
The broken lines are shown for illustrative purposes only and form no part of the claimed design. The cable is fragmentarily shown to indicate indeterminate length.

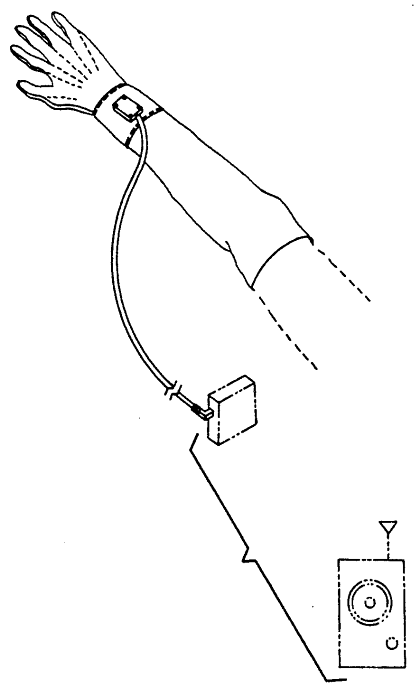


Figure 2.6: Example of U.S. (Design) Patent No. Des 347,853.
This is an example of a design registered with the USPTO. The claim offers insights into the design protection. In the United States, design protection is governed by patent law, and thus, this registered design is referred to as a "USPTO patent". The inclusion of "Des." in the USPTO patent number at the top right signifies that it is a design protected by the USPTO.
Source: Excerpt from U.S. Design Patent No. D347853 (Masubuchi and Kato 1994).

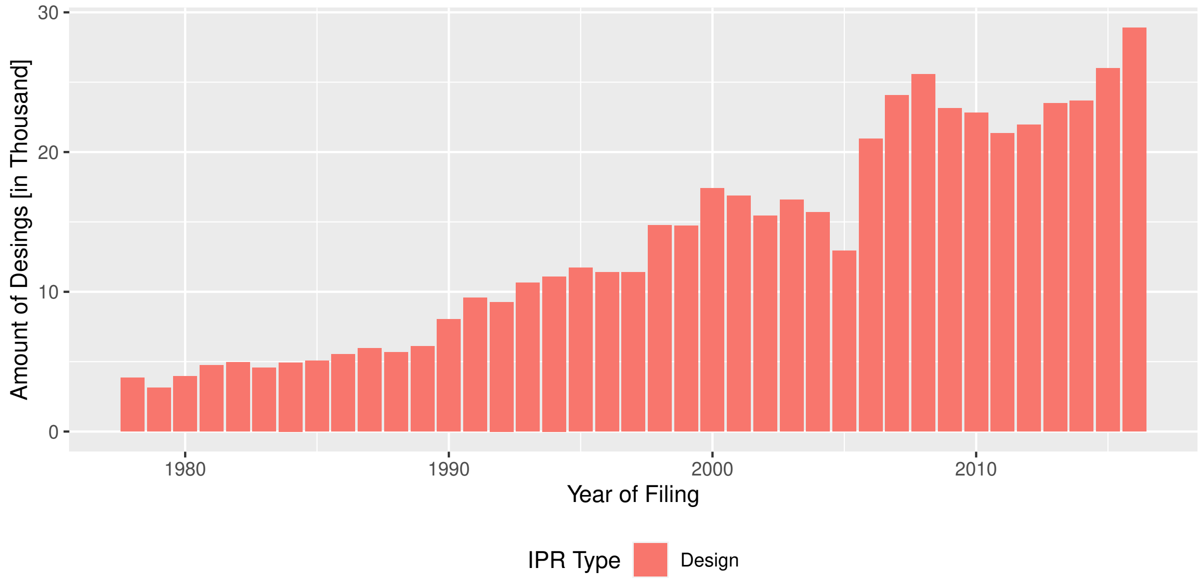


Figure 2.7: Total Yearly Development of Granted Designs between 1978 and 2016.
The figure illustrates the yearly filings of granted designs at the USPTO from 1978 to 2016.
Source: Own representation based on USPTO data (see Query A.4).

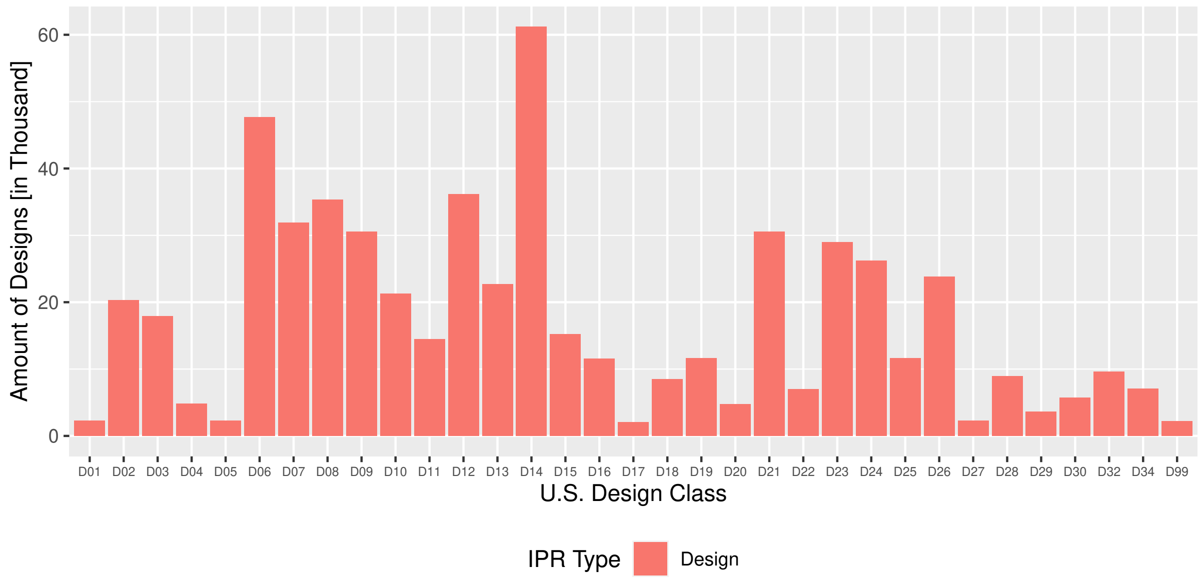


Figure 2.8: U.S. Design Classes of Granted Designs between 1978 and 2016.
The figure illustrates the distribution of granted designs across various U.S. design classes at the USPTO. It takes into account only designs filed between 1978 and 2016.
Source: Own representation based on USPTO data (see Query A.5).

Table 2.2: Overview of Design Classifications

Locarno Classification (13th Version)		**U.S. Design Classes**	
Class No.	**Description**	**Class No.**	**Description**
1	Foodstuffs	D1	Edible Products
2	Articles of clothing and haberdashery	D2	Apparel and Haberdashery
3	Travel goods, cases, parasols and personal belongings, not elsewhere specified	D3	Travel Goods, Personal Belongings, and Storage or Carrying Articles
4	Brushware	D4	Brushware
5	Textile piece goods, artificial and natural sheet material	D5	Textile or Paper Yard Goods; Sheet Material
6	Furnishing	D6	Furnishings
7	Household goods, not elsewhere specified	D7	Equipment for Preparing or Serving Food or Drink Not Elsewhere Specified
8	Tools and hardware	D8	Tools and Hardware
9	Packaging and containers for the transport or handling of goods	D9	Packages and Containers for Goods
10	Clocks and watches and other measuring instruments, checking and signalling instruments	D10	Measuring, Testing or Signaling Instruments
11	Articles of adornment	D11	Jewelry, Symbolic Insignia, and Ornaments
12	Means of transport or hoisting	D12	Transportation
13	Equipment for production, distribution or transformation of electricity	D13	Equipment for Production, Distribution, or Transformation of Energy
14	Recording, telecommunication or data processing equipment	D14	Recording, Communication, or Information Retrieval Equipment
15	Machines, not elsewhere specified	D15	Machines Not Elsewhere Specified
16	Photographic, cinematographic and optical apparatus	D16	Photography and Optical Equipment
17	Musical instruments	D17	Musical Instruments
18	Printing and office machinery	D18	Printing and Office Machinery
19	Stationery and office equipment, artistic and teaching materials	D19	Office Supplies; Artists' and Teachers' Materials
20	Sales and advertising equipment, signs	D20	Sales and Advertising Equipment
21	Games, toys, tents and sports goods	D21	Games, Toys and Sports Goods
22	Arms, pyrotechnic articles, articles for hunting, fishing and pest killing	D22	Arms, Pyrotechnics, Hunting and Fishing Equipment
23	Fluid distribution equipment, sanitary, heating, ventilation and air-conditioning equipment, solid fuel	D23	Environmental Heating and Cooling, Fluid Handling and Sanitary Equipment
24	Medical and laboratory equipment	D24	Medical and Laboratory Equipment
25	Building units and construction elements	D25	Building Units and Construction Elements
26	Lighting apparatus	D26	Lighting
27	Tobacco and smokers' supplies	D27	Tobacco and Smokers' Supplies
28	Pharmaceutical and cosmetic products, toilet articles and apparatus	D28	Cosmetic Products and Toilet Articles
29	Devices and equipment against fire hazards, for accident prevention and rescue	D29	Equipment for Safety, Protection and Rescue
30	Articles for the care and handling of animals	D30	Animal Husbandry
31	Machines and appliances for preparing food or drink, not elsewhere specified	D32	Washing, Cleaning or Drying Machines
32	Graphic symbols and logos, surface patterns, ornamentation	D34	Material or Article Handling Equipment
		D99	Miscellaneous

The table provides an overview of design classifications, comparing the Locarno classification (13th Version) on the left with the U.S. design Classes on the right. Most classes in the Locarno and U.S. systems contain similar content under corresponding numbering. For instance, Locarno Class 7 corresponds to U.S. Class D32, and Locarno Class 31 aligns with U.S. Class D7. However, there is no equivalent in the U.S. system for Locarno Classes 34 and 99, and conversely, there is no equivalent in Locarno for U.S. Class 32.
Source: Own representation based on WIPO (2021f) and USPTO (2021c).

combining several intellectual property rights to measure innovation makes sense. The combination of patents, trademarks and designs is considered in this thesis.

2.3 Measurement of Innovation

Innovation measurement and data serve policymakers as a basis for understanding economic and social change, the importance of innovation for goal achievement, and determining the efficiency and effectiveness of policy decisions (OECD and Eurostat 2018). Innovation indicators exist for innovation culture, strategy, knowledge and competence, organisational structure, research and development activities and input, and financial innovation performance, among others (Dziallas and Blind 2019). Dziallas and Blind (2019) differentiate between product and process innovation indicators. Process indicators are based on, e.g., time for idea generation, ongoing innovations, rate of suggestions implementation or implemented process improvements. Product indicators are based on, e.g., intellectual property rights such as patents, the novelty of the product, technological significance, or new product announcements. Besides, a shift to qualitative indicators related to technical and non-technical innovation is observable among the indicators used, which is explained by an increasing focus on service industries. Qualitative indicators are of interest in the early stages of innovation to assess innovative ideas where product and quantitative indicators may not be available and to cover sectors such as services(Dziallas and Blind 2019). As part of qualitative indicators, especially textual data could provide valid information. Authors like Ozcan et al. (2021) or Shen et al. (2020) apply textual data analysis to uncover ideas for further innovation. Current textual data for innovation research is more closely assessed in Chapter 3.

2.3.1 R&D-Expenditure or Patent-based Innovation Indicators

Patents and R&D expenditures are considered traditional innovation indicators, whereby patents are output, and R&D expenditures are input indicators. Sales, total innovation expenditures or new product announcements are among the new indicators used (Kleinknecht et al. 2002). With data on R&D expenditures, long time series can be created, or countries, industries or firms can be compared. Although expenditures cannot be broken down into technological fields, they can be separated into product and process innovations. However, R&D expenditures as a basis for innovation measurements only reflect one input factor of several for innovation and based on this, the innovation output cannot be directly inferred. Also, the input does not necessarily imply an equivalent output. Regional distinctions are difficult, and innovations by small firms or in the service sector are underestimated (Kleinknecht et al. 2002).

Pakes and Griliches (1980) first find a positive correlation between R&D spending and patent counts, while considering sectoral differences. Patents are mostly associated with the development stage of the innovation process (Basberg 1987) and are often used as an indicator for product innovations (Dziallas and Blind 2019). Patents tend to be registered in the concept phase and are therefore categorised as ex-ante product innovation indicators (Dziallas and Blind 2019). Patent data allow the analysis of long periods. They are publicly accessible and are mostly available in electronic form. A sophisticated classification system can be used for the analysis, and an overview of the technical knowledge is made possible. They can be disaggregated geographically (Kleinknecht et al. 2002). Patent citations can be used to measure the value of innovations, with high patent citations indicating major innovations (Trajtenberg 1990). Considering Patent citations tackles the problem of patents having different economic value (Pakes and Griliches 1980) and that the patenting system does not distinguish between small and large improvements (Basberg 1987). However, a drawback of patent-based indicators is that not all innovations are patented (Pakes and Griliches 1980) and that not every patenting leads to commercialisation so that often only inventions and not innovations are covered (Basberg 1987). Moreover, the propensity to patent varies across sectors and firms. Underestimating innovation in low-technology sectors or by small firms and overestimating firms with R&D collaborations or the innovation intensity of small patent holders are underlying systematic errors in the use of patent indicators (Kleinknecht et al. 2002). Blind et al. (2003) point out problems in the patenting system, such as difficulties in patenting services, tacit knowledge, software or even products and services in sectors such as the financial or business service sector. Becheikh et al. (2006) report that out of 108 studies interested in technological innovation in manufacturing, indirect indicators such as patents (18%) or R&D expenditures (6%) are used less frequently than direct innovation indicators like innovation count (25%) based on, e.g. product or process announcements, journals or databases or firm based surveys (24%).

To conclude, patents especially provide valuable information for, e.g., product innovation and the technological state. However, their coverage of areas like services or low-technology areas is weak.

2.3.2 Trademark-based Innovation Indicators

To overcome limitations of patent or R&D expenditure-based indicators, Millot (2009) proposes the use of trademark-based indicators. Trademarks reflect the commercial application of products and services. They can cover marketing and service sector innovation, thus mostly non-technological innovations, and are positively correlated with innovation on the firm-level. As seen in Subsection 2.2.2, trademarks are systematically collected, classified and disaggregated for different levels (Flikkema et al. 2015). However, their relation to innovation is not guaranteed. Different authors use trademarks to measure innovation in products and services:

Product Innovations: Malmberg (2005) explores trademarks as a product indicator for innovations new to the firm in the automotive, electromechanical and pharmaceutical industries in Sweden. For automotive and electromechanical products, trademarks are not considered reliable as new product introductions are not adequately registered in trademarks. In contrast, trademark registrations in the pharmaceutical industry align in the long-term perspective with new product introductions. Faurel et al. (2022) also measure new product innovation based on product trademarks. A significant positive relationship is found between new trademark applications, sales and profitability. Gao and Hitt (2012) use trademarks to measure IT-related product variety, with increased IT capital leading to more trademarks and thus indicating greater product variety. Overall, an increased rate of innovation can be observed with greater product diversity and product novelties caused by IT capital.

Service Innovations: Hipp and Grupp (2005) analyse the service industry in which product and process innovations occur. They find that, in contrast to the manufacturing sector, technology-based innovations are less important, and patents play only a minor role. To measure service innovation, they hence propose trademarks as an alternative measure. This is in line with a case study of Blind et al. (2003) among 65 European service firms that revealed that across all firm sizes, trademarks are used to a large extent as a protection mechanism and reflect the market introduction of a new service. Patenting is nevertheless relevant in services but covers ICT-, technology- or hardware-related services. The results are confirmed among German service firms using patents only when technical aspects are related to their service offerings (Blind et al. 2003).

Comparing services and product innovation, an online survey among European Community or Benelux trademark applicants reveals the following: First, new trademarks protect product innovations more frequently than service innovations. Second, the first trademark of a start-up or for brand creation is more likely to stand for product innovations than the first trademark of mature firms. Third, trademarks with a broader geographical coverage are less likely to be related to service innovations. Finally, trademarks with a narrower scope of protection are more likely to be related to product innovations (Flikkema et al. 2019).

Trademarks, therefore, open up broad possibilities for analysis due to their wide applicability and the possibility to measure non-technical, service or marketing innovations through trademarks. However, in the selection and analysis of trademarks, further aspects such as differences in the legal system, the home bias of trademark applications, differences between firms, sectors and countries (Millot 2009) and, above all, the unverified link with innovations in the application of trademarks as an innovation indicator must be taken into account. Further, not all products are covered by trademarks, and the importance of a trademark application is hard to measure (Gao and Hitt 2012). In summary, trademarks, like patents, have both their strengths and weaknesses: In the context of innovation, it is important to note that while they are not primarily intended to protect innovation, they can be associated with it and particularly capture service innovations or non-technical innovations. In comparison to and interaction with other indicators, they may make a valuable contribution to understanding innovation. The textual contribution of trademarks in the context of innovation remains to be assessed in this thesis.

2.3.3 Combination of Intellectual Property Rights

As discussed, indicators such as patents and trademarks have their strengths in capturing innovation but also their shortcomings. Combined, trademarks and patents could maintain these strengths while addressing their shortcomings. Therefore, this thesis addresses the shortcomings of trademarks through the combination of trademarks with other data sources, especially with patents and designs. According to Becheikh et al. (2006), this is already performed in other contexts: Direct measures like a firm-based survey or innovation count as well as indirect measures like R&D expenditure or patents come with shortcomings and advantages. Some studies combine different approaches to reduce shortcomings and maintain strengths. Related to the combination of trademarks and patents, Malmberg (2005) finds that combined analyses of patents and trademarks in the pharmaceutical industry may reveal insights on innovation, as trademark registrations go in the long run in line with product innovation. Also, Flikkema et al. (2015) find that the combined analysis of trademarks and patents brings large benefits. Trademarks relate to patented and unpatented innovation. As such, they find that trademarks filed related to brand creation or extension capture innovation, of which patents do not cover 80%. The combined use changes depending on the service or manufacturing background of the applicant. Li (2016) combines patents and trademarks to measure technological and non-technological types of innovation in firms as a basis for future earnings. New trademarks predict earnings one quarter as well as one year ahead. Patents and trademarks serve as a predictor for five-year earnings growth. The authors conclude that analysts, on the one hand, use the information provided by new trademarks to forecast short-term earnings but underestimate their effect. On the other hand, they jointly use patents and trademarks for long-term earnings, but they are overly optimistic. Overall, the combined consideration of trademarks and patents improves existing innovation measurements. In addition, qualitative indicators provide additional information on early-stage innovation. The use of trademarks, combined with other data sources and from a qualitative, textual perspective, therefore provides the potential to cover innovation activities from a broader perspective.

2.4 Textual Data Analysis

Textual data of trademarks are the main focus of this thesis. For its analysis, textual data analysis is applied. Textual data analysis is especially of interest in large data sets where an individual assessment of each document is not feasible or intended and where data are not structured into specific, identifiable fields. Further, the textual data can reveal granular information on the data provided and, as such, enable further insights. Insights are data-driven and less dependent on experts, so the expert bias is reduced. In the context of this thesis, text data opens up the possibility of directly linking various data sources. By analysing textual data, authors intend to, e.g., gain additional insights on the development of a field, extract or generate new ideas or identify new areas of future research (for further details, see Chapter 3). The challenge of textual data analysis is that the methodologies do not have a human understanding of the content of text documents. Textual data analysis, therefore, extracts information from documents based on patterns and statistical occurrences (Aggarwal 2015). For the analysis of textual data, different methods exist. The selection of the method depends on the problem that should be addressed. Most problems can be classified to be either clustering or classification problems:

Clustering problems identify groups of words (clusters) in the data set based on similar characteristics which tend to co-occur. Clustering can be used, e.g. for data summarisation, group identification, community detection or as a basis for classification models. It is based on similarity measures (Aggarwal 2015). Different clustering approaches exist, such as K-means clustering or Topic Modelling.

K-means clustering groups documents in a space of words. Consider an N-dimensional space where N is the number of unique words occurring in a set of documents. A document can be represented by a vector showing the number of times each word occurs in that document. Now, each document exists in the space, and the goal is to find groups of documents that cluster together. The algorithm begins by defining K randomly selected centres of clusters (centroids) — arbitrary locations in the space. Each document is assigned to the closest centre (using Euclidean distance). Iteratively, the centres are then re-determined to represent better the group of documents assigned to them. The assignment of documents to the centres is then reviewed and optimised for a new centre if this improves the distance between the document and its centre. The process continues until no more improvement is achieved, and thus, the cluster assignment remains stable (Aggarwal 2015). A drawback of K-means is that it partitions the documents in the word space. That is, each document belongs to exactly one cluster.

Topic modelling relaxes that position and assumes that a document can be composed of several underlying topics. However, topics are not explicit but only latent. Topic modelling makes them explicit through a probabilistic process which estimates the probability that a topic exists in a document. Topic modelling is a frequently used approach for dimensionality reduction. Several approaches for making this inference exist, among which a common one is Latent Dirichlet Allocation (LDA) of Blei et al. (2003). The results gained from applying LDA are topics in the sample documents provided to the model. These can then be used to interpret the documents or the area of study. A modification of the LDA is Structural Topic Modelling (STM) of Roberts et al. (2019), which makes it possible to determine and analyse the influence of additional information on the sample. In this thesis, the textual data analysis is performed with Structural Topic Modelling (STM) of Roberts et al. (2019). LDA and STM are explained in further detail below.

Classification problems address specific features of the data set. They are considered supervised problems because the models learn the classification feature based on a given training data set. The model then applies the classification to test data and assigns labels to the data (Aggarwal 2015). For example, Ozcan et al. (2021) use the approach to distinguish between tweets with and without ideas. The authors, therefore, extracted a set of tweets from Twitter. Of these, they manually labelled a subset and provided the positive feature “idea” or negative feature “not an idea”. This was then used as a training set for the model. Another example is the study of Kreuchauff and Korzinov (2017), which separates a Robotic patent data set into service robotics and industrial robotics. Compared to clustering approaches, classification problems are considered more potent as the classification is defined externally by the user (Aggarwal 2015).

Other methods of interest for analysing textual data are, e.g. sentiment analyses or word embedding. *Sentiment analyses* intend to extract the attitude or feeling of the writers towards a subject or topic out of the wordings used (Liu and L. Zhang 2012). In *Word embeddings* the context of words is analysed to identify words occurring in a similar context and revealing similar relations. “Paris”, for example, has the same relationship to “France” as “Berlin” has to “Germany”. Both are the capitals of the respective countries. The model can be used to uncover such relationships in textual data (Mikolov et al. 2013).

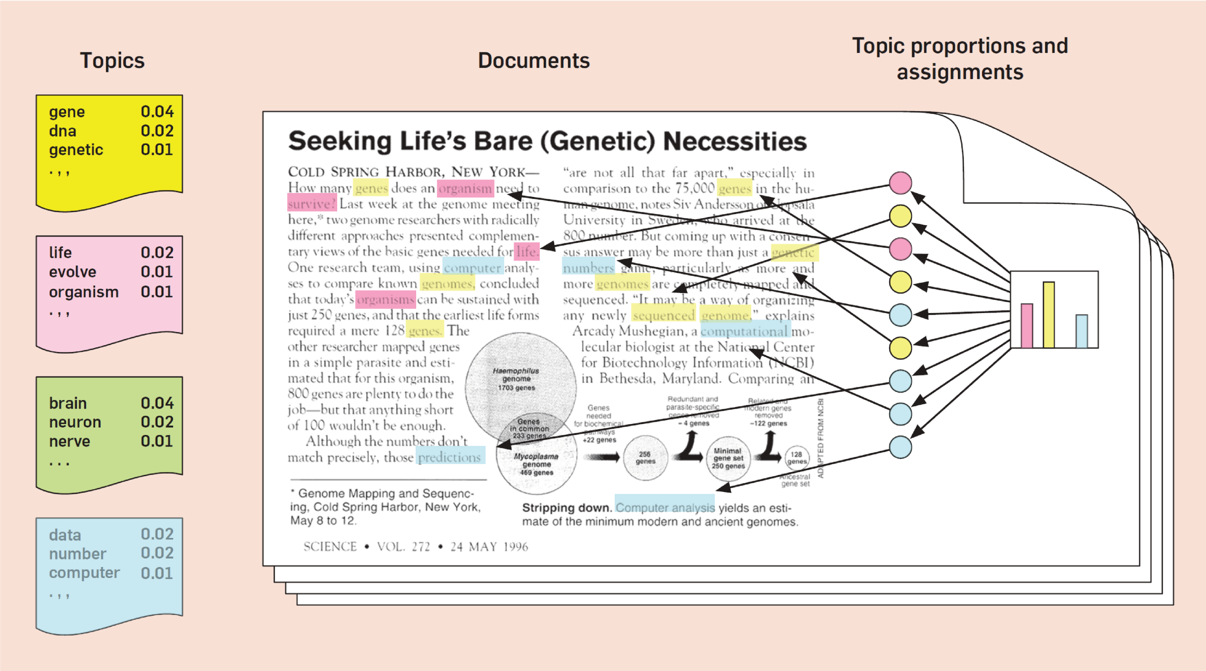


Figure 2.9: Example Representation of Topics in Documents.
The figure illustrates the intuition behind topic modelling. The basic components are topics, words and documents. Topics are presented on the left, along with associated words and their respective weights. The middle section displays the documents.On the right, the figure depicts topic proportions and assignments within the example documents. Small bars indicate the topic proportion in each document. Each topic is linked to words through coloured dots, and each dot corresponds to a word in the document. Arrows represent the direction of the computational estimation. A word distribution is derived based on an estimated topic distribution, which leads to the observed occurrence in the document.
Source: Extracted from Blei (2012, p. 78).

2.4.1 Latent Dirichlet Allocation

A topic modelling and clustering approach to extract information from textual data is Latent Dirichlet Allocation (LDA) introduced by Blei et al. (2003). LDA is applied to a sample of documents to extract the topics occurring in these documents. Like a document analysis performed by a human, the approach provides an overview of the topics in a sample of documents. It is, however, scalable to a large data sample, which would otherwise not be feasible to analyse. Yet, to reach that understanding, the model needs to estimate the topics. Figure 2.9 illustrates an example of the approach (Blei 2012) and is used to underline the explanations. The figure displays a document titled "Seeking Life's Bare (Genetic) Necessities". This is one of several documents that constitute the text corpus to be analysed.

The text data would first be extracted, preprocessed and transformed to be in an analysable form. These steps cover the selection of the data to be analysed, the dimensionality reduction to reduce the variation in the data and increase focus, and the structuring of the textual data. The steps are described in detail in Subsection 2.4.3.

Before estimating the topics, the analyst decides how many topics the process will discover, as the model requires the topics to be assumed as input. As the optimal number is unknown in advance, the model is calculated for several topic numbers, and the optimal number is then chosen based on several metrics, which are explained in Subsection 2.4.2). In the example shown in Figure 2.9, the analyst has chosen four as the number of topics to be estimated.

The next step is then the estimation of these four topics. To get these, the model works as follows. First, two multivariate distributions are determined, representing the probability of topics in documents and the probability of words in topics. These distributions together define the words occurring in each document. The estimation procedure is one of incremental Bayesian inference: Start with a random assignment of each word to one of the pre-determined numbers of topics, which distributes the topics across documents. Then iteratively reallocate each word in each document, one at a time, to a new topic. The assumption is that the topic allocation of the word to be reallocated is wrong, while the word-topic allocation for all other words in the document is correct. The allocation of a word to a new topic depends on two aspects: a) the extent to which that topic is present in the word's document, and b) the extent to which the word itself is allocated to the topic across all documents. If both probabilities are high, the word is likely to be reallocated to that topic. This reallocation changes the two probability distributions and is iterated until the distributions become stable (Blei et al. 2003; Blei 2012).

In Figure 2.9, the resulting four estimated topics are shown on the left. These can then be analysed. The topics evolve around genetics and DNA (yellow), life and organisms (pink), brain functions (green), and computation (blue). The topic-word distribution determines the allocation of the words in the topics. For example, the yellow topic of genetics and DNA displays three words and their occurrence probability in the topic. These are "gene (0.04)", "dna (0.02)" and "genetic (0.01)". On the right, the topic distribution in the document can be observed: of the four topics, only three are represented. Genetics and DNA have the highest proportion in the document, followed by life and organism. The two distributions determine the word assignment in the document. The red topic, for example, contributes "organism" and "survive", while the blue topic focuses on "predictions", "computers" and "numbers". As the green topic of brain function is not part of the topic distribution for this document, no words are assigned to the document. LDA is considered a reversed generative process because it first estimates the distributions representing the hidden structure before generating the documents with observable words based on these distributions (Blei 2012).

2.4.2 Structural Topic Modelling

Structural Topic Modelling (STM) of Roberts et al. (2019) is a refined version of the LDA. STM allows additional information about the documents to be provided to the model to improve model estimation. It enables the analysis of this additional information. The additional information influences two parameters: the word occurrence in the topics or the topic occurrence in the documents. The former is called the "topical content" parameter, while the latter is referred to as the "topic prevalence" parameter. The topic prevalence parameter μ changes the topic-document-probability while the topical content parameter β affects the word-topic-probability. These parameters are included in the model calculation and cause, for example, documents with the same features to have similar topic or word probabilities, while the probabilities of documents with other features differ. For the model analysis, γ provides the topic distribution in a document and is used as a Normal prior distribution that influences the topic prevalence (Roberts et al. 2019). A main difference to LDA is the topic dependence assumption. In LDA, topic independence is assumed, while in STM, this is rejected and replaced by the assumption that topics dealt with in a document are dependent. In LDA, the topic distribution is determined by a Dirichlet distribution. In STM, this is replaced by a logistic normal distribution (Roberts et al. 2019; Blei and Lafferty 2007). This means, for example, that the joint occurrence of a music topic and genetics topic is less likely than the joint occurrence of a genetics topic and brain function topic. The dependence assumption shows better performance on real text documents and is introduced in the Correlated Topic Modelling (CTM) of Blei and Lafferty (2007). Structural Topic Modelling builds on this CTM. Therefore, if no additional information is transferred to the model, the model corresponds to CTM (Roberts et al. 2019).

Structural Topic modelling can be applied to answer different research questions: Roberts et al. (2013) apply STM in the context of news articles related to China. The publisher and year are used to influence the topic prevalence, and the news outlet to influence the topical content. The publisher and year influence the topic prevalence assuming that both pieces of information are crucial in determining which topics occur in a document. For example, in 2004, the Taiwan election was relevant, while other topics may be more important at other times. This can then be investigated. The assumption behind the publisher's influence on topical content is that it makes a difference whether the global or the Chinese press reports on a topic. The words used for the same topic are different, e.g. "elect, democrat, vote" versus "reunif, one-china, province". The analysis helps to understand how China sees itself compared to the world. In the context of innovation research, Cripps et al. (2020) use STM to differentiate Tweets from firms using Twitter for innovation and crowdsourcing purposes and Tweets from firms without that relation. This binary information is contributed to the model as a topic prevalence. With the help of STM, the authors assess if Twitter can serve as a source for innovation-related information.

Like in LDA, STM requires the topic number as input. Residuals, held-out likelihood, semantic coherence and exclusivity are metrics to evaluate the models and select a topic number in STM. *Residuals* provide insights into the deviation between topic estimation and the resulting, estimated words in a document and the observed words in a document. A low residual value is desired and indicates that the topics provide a good approximation of the documents. *Held-out likelihood* shows how well the model predicts words that were previously removed from documents. *Exclusivity* describes the uniqueness of terms, while *semantic coherence* indicates the consistency of a topic. These metrics must be balanced against each other by the analyst (Roberts et al. 2019). This means that if a topic has a high number of words that are exclusive to this topic, the exclusivity of the topic is high. The more common words a topic has, the less exclusive a topic is. Common words are shared among several topics.

During this thesis, assessing the differences between IPRs and covering the dynamics over time is of interest. STM enables these analyses and is the main method to analyse textual documents in this thesis. It is applied in Chapter 4 and Chapter 5. The analyses are based on trademark, patent, and design documents to cover innovation and diffusion in different innovation areas. Additional information is provided to the STM to assess the differences between the IPR types, combine the different data sources and cover the dynamics of innovation. Topic prevalence is the main feature of STM exploited in this thesis. For the analyses, the following aspects are considered in the STM application:

Dynamics of innovation: Innovation is dynamic, and new innovation topics emerge continuously. This is, e.g. reflected in regular updates of the patent classifications system to reflect the latest developments. To reflect this changing dynamic over time, the year of application of the IPRs is considered in this thesis. This allows for changing topic occurrence in different years. Further, the topic occurrence is considered non-linear. A topic can thus often occur in the early 1990s, be less prominent in the 2000s and have a revival in the 2010s. This information is included as a spline calculation in topic prevalence ($s(year)$).

Topic differences between IPRs: Patents, trademarks, and designs have different functions in protecting innovation and diffusion. Patents cover inventions, while designs focus more on design aspects. Trademarks have a focus on market introduction. This implies that the topics covered by the different IPRs vary. To assess this and analyse if there is a difference in the topic coverage, the IPR type of the document is provided to the model (*IPR type*).

Time-wise differences between IPRs: Patents, trademarks, and designs are filed at different points in the innovation process. Patents and trademarks can refer to different stages of the innovation process (Seip et al. 2018). These stages can deviate. This means an invention protected by trademarks and patents can have different application years depending on the IP right. To capture this aspect, an interaction between the years and the type of IP right is enabled ($s(year) \cdot$ *IPR type*).

2.4.3 Data Preparation

LDA and STM require the textual data to be in an analysable form. The approaches use a document-term matrix for topic estimation. This requires the extraction, preparation and transformation of textual data. These steps are done the following way:

Data Extraction: To start with, a sample of textual documents or textual data is chosen. This data is considered unstructured. A document w is hereby seen as a sequence of N words, also called terms. A word is a discrete unit from a vocabulary. The information contained in the words of the document varies: *Stop words*, which are among others, "and", "or", and "we", or *common words* that occur very often in the sample are considered as words without meaning. Other words, like "guitar" or "electronic", can provide valuable information. It depends on the context and the intention of the analysis, which words are considered valuable.

Preprocessing: Taking all unique words from all text documents together constitutes the text corpus for the data analysis. The information is saved in a *document-term-matrix*, with documents in the rows, terms in the columns and the absolute occurrence as the values in the matrix. Since this matrix can become very large due to many terms and documents being involved, dimensionality reduction, also called preprocessing, is necessary. Dimensionality reduction reduces the computational power required by the topic estimation as the size of the matrix is reduced. The matrix size is reduced by removing words from the text corpus, which reduces the number of columns and by removing documents, which reduces the number of rows. Removing words might also lead to fewer documents if these are only focused on some words and are thus empty after cleaning. Preprocessing further reduces the amount of variation in the corpus.

Different approaches exist to remove words from the matrix. The intention is to keep only "meaningful" words, whereby "meaningful" is determined based on the context. *Term Frequency-Inverse Document Frequency (TF-IDF)* is often applied, which is a technique for normalisation of the documents. In a text document, a term can occur very often, giving it a high weight. However, this term does not necessarily contain valuable information, as it might be a widespread term in the text corpus, thus being contained in most documents. TF-IDF intends to reduce the bias introduced by words with repeated frequency and, by doing so, should introduce greater stability to the text and increase the similarity between words.

TF-IDF may be defined as follows:

$$\text{TF-IDF} = \text{TF}(t, d) \cdot \text{IDF}_t \tag{2.1}$$

t denotes the term or word, d the documents. $\text{TF}(t, d)$ is then a function of the *Term Frequency*, revealing how often a term t occurs in a document d. IDF is the *Inverse Document Frequency*. It determines in how many documents a specific term occurs in and can be calculated with $\text{IDF}_t = log\frac{N}{n_t}$, where index t denotes the term, N the number of documents in the corpus, and n_t the number of documents with term t. IDF is used to remove noise by reducing the word weight of higher-frequency words. Applied in TF-IDF, this thus means that a word which often occurs in many documents will receive a lower overall weight than words that often occur in only a small amount of documents (Aggarwal 2015; Aggarwal and Zhai 2012).

Apart from the removal of common words, preprocessing steps include, e.g. the removal of stop words like "a", "the" or "you" or stemming that reduces the words to their stem, like "going" to "go" (Aggarwal 2015). Further, tokenisation or lemmatising that tries to map words to their basic form like "went" to "go" can be applied (Allahyari et al. 2017). The unique terms remaining after preprocessing constitute the final dictionary. The entirety of the documents M that remain in the preprocessed sample is referred to as text corpus D.

Transformation: To analyse textual data, the data needs to be transformed into a multidimensional representation. LDA requires a *document-term-matrix* as an input for the model estimation. The documents in the matrix are listed in rows, and the terms are in columns. The values of the individual cells then correspond to the frequency of each term in the respective document. The unique words across all documents and the words' frequencies in each document are determined to get the values of the document-term-matrix. The representation of a text corpus in this way is also called a representation in a *Vector Space Model (VSM)*. Each document can therefore be represented as a vector consisting of all corpus' terms, indicating how often each word appears in that document (Aggarwal 2015; Allahyari et al. 2017). The vector representation can be used to compare documents. Cosine similarity is a commonly used method for that (Aggarwal 2015), which compares the direction of two vectors and determines their similarity (Allahyari et al. 2017).

The representation of the text documents in a document-term-matrix in LDA is possible because LDA does not consider the order of the words, i.e. only the frequency counts. It is therefore also called a *"bag-of-words" concept*, which neglects the order of the word sequence (Blei 2012). The concept can be considered a bag full of words, as the term suggests. Each document has its bag of words. This bag contains all the document's words, but the content is chaotic as the words are in a bag. The original order of the words is not kept, and the words mingle. The assumption behind this concept is that the word order of texts often does not provide additional information, and by neglecting it, the analyses of larger data sets become possible. Consequently, a text can be represented as a list of unique words and their occurrences.

2.5 Conclusion

Innovation is important for economic growth and the welfare of society. Measuring innovation is necessary for policymakers to understand the impact of innovation and policies. Innovation covers the introduction of, e.g. new services and products to the market. Without an introduction, only an invention is taking place. Traditional indicators are based, e.g. patents or R&D expenditures. Patents cover the inventive step of a technical invention but have shortcomings such as the missing market introduction, the non-coverage of services or low applicability in low-technology areas. Trademarks are considered to overcome these limitations. They cover services and require a market introduction. However, they do not require an inventive step, even though research has shown that they are often introduced with an innovation. Trademarks also have limitations, such as the missing granularity of the trademark classification system, making it challenging to perform sophisticated analyses. To provide further detail on innovation and to be able to combine trademarks with other data sources to cover innovation broadly, the textual data of trademarks are of main interest in the course of this thesis. This thesis applies Structural Topic Modelling to combine trademarks with patents and designs and to derive insights into these intellectual property rights.

3 Textual Data in Innovation Research

Chapter Abstract

The chapter assesses the current application of textual data in innovation research. 23 articles applying text analysis in innovation in Business and Economics are selected from Web of Science and Scopus. A detailed review of these articles follows, addressing the motivation for text analysis, the data sources used, the methodologies applied, and the research questions answered with textual data and its analysis. It becomes evident that authors apply text analyses as these enable them to analyse text data and large data sets and gain objectivity through data-driven results. The methodology can be used for rapid analyses, uncovering hidden information, and gaining data-based insights. The textual data sources are still mostly publications and patents, but other data sources are also relevant. The combination of different data sources is also applied. Most studies focus on clustering techniques for dimensionality reduction and uncovering topics of interest. Classification is applied to identify, e.g., relevant data sets or determine the innovativeness of firms. Research questions answered mostly focus on technology mapping, innovation trajectory and the identification of novel topics. Overall, textual data provide additional information on the development of innovation. The methodologies and variety of potential textual data sources provide a large potential for innovation research and can help to gain further insights and improve existing innovation measures.

KEYWORDS: Textual Data, Text Analysis, Innovation Research, Literature Review

JEL Code: O30

3.1 Motivation

The amount of data potentially available for innovation research is enormous: Currently, more than 5.4 million utility patents, 4.2 million trademarks and 500 thousand design patents are available for analysis at the United States Patent and Trademark Office (USPTO) (see Section 2.2). Further, in 2020, more than 2.6 million published articles were available via Web of Science (WoS 2021). Patents and publications are a major source for innovation research, while trademarks are considered a potential new data source to cover developments in services or market application (Millot 2009). However, the information available in these data sources is only partially stored in a way that is easy to analyse. The information is predominantly stored in texts. To extract the information from them, it is first necessary to transform it into a form that can be analysed and evaluated. Furthermore, a large amount of data available presents a challenge. Text analysis is dedicated to these subjects. With new methodologies for the analysis of textual data and increased computational power, mining this data and gaining insights from textual data becomes more accessible and interesting for various fields: Within ten years, the number of final publications outside of computer sciences, engineering and mathematics[1] in text mining increased from 3,550 (from 2001 to 2010) to 17,739 (from 2011 to 2020) final publications. Thereby, the Social Sciences (18%), Decision Sciences (14%), Medicine (13%) and Business, Management and Accounting (10%) are the main areas of text mining publications.[2] For example, Tseng et al. (2007) analyse text mining techniques for patent analyses and show that large data sets can be analysed via automatic procedures that preserve relevant content-based information in comparison to existing classifications. The techniques can help in several steps of patent analysis. Also, Antons et al. (2020) enforce that data mining, especially text mining, can help leverage the potential of hidden knowledge for innovation economics and extend the existing portfolio of analytical techniques.

Text mining or text analysis techniques are increasingly used. In innovation research, they offer new insights from traditional data sources like patents or publications, as they might extract further details besides the existing classifications. The analysis of textual data sources besides the traditional ones might also be facilitated, as textual data are genuinely available. Different data sources might be combined textually, which could surface new insights thanks to the combination. As text analysis offers several opportunities for analysis, this chapter analyses the current state of textual data and its analysis in innovation research to guide readers upon the application of textual data analysis. Therefore, a literature search in Web of Science and Scopus in business and economics is conducted to extract a sample of 23 published articles that are analysed in further detail.

The review of the articles focuses on why authors apply textual data analysis, the textual data sources used, the methodologies applied, and the research questions answered. The review contributes to the existing literature by providing insights into textual data for innovation research and detailed information on the articles in the sample. These serve as examples of current text analysis in innovation research. By doing so, it extends the review of Antons et al. (2020) that provide a general overview of text analysis in innovation research and aspects that need to be considered but do not provide further detail. In their assessment of 124 research articles, the authors focus on the major aspects to be observed, such as the relevant journal outlets and underlying data sources, the underlying topics, general methodological aspects, and the reporting quality. The authors provide a general overview of the articles authors' common keywords and generic topic clusters. The discussion of individual articles remains high due to the number of underlying research articles. The research questions that were answered and the authors' reasoning remain unclear. However, an overview of different articles is necessary to decide upon applying textual data analysis for own research questions in innovation research. This overview should highlight the potential areas of textual data, its analysis, and its value-added. The review in this chapter addresses this need, contributing to an understanding of the current applications of text data analysis in innovation research.

[1] These fields are since 1991 the main three fields of text mining publications.
[2] Data based on Query B.1).

The chapter discloses the methodology and descriptives on the sample in Section 3.2. The review results are described in Section 3.3. The review closes with the conclusion and discussion in Section 3.4

3.2 Methodology

As a basis for the literature review, the relevant literature on innovation and the use of textual data is identified first. The selected sample is then examined regarding the authors' motivation, data source, applied methodology, and answered research questions.

3.2.1 Data Selection

The data extraction is carried out in the subject areas of business and economics. Business and economics are considered relevant fields to capture developments in innovation research. The attention of the overview is not restricted to any specific journals to capture various applications of textual data and its analysis. This contrasts with Antons et al. (2020) that restrict their data selection to the top innovation, management and strategy journals. To finally extract and restrict the sample, the following procedure is chosen:

1. English research articles have been identified based on "text*" and "innov*" in the title or abstract in the areas of business or economics in the databases of Web of Science (WoS) (Query B.2) and Scopus (Query B.3). Abstracts with "textile" are excluded, as these introduce too much noise. 250 documents in Web of Science and 789 documents in Scopus are identified. However, an overlap between the data sources can occur.[3] Duplicates are later manually removed.
2. The "author keywords" of all research articles are then extracted and analysed. This leads to a list of 2,438 author keywords. Each keyword is evaluated individually in terms of using text data and related methods to produce a list of relevant keywords based on the author's assessment. Among others are "text mining", "text analytics", "big data", "sentiment analysis" or "unstructured data". The keywords relate to the data structure or means of analysis. The final list comprises 100 keywords (see Table B.1). Keywords that are related to a data source but not explicitly to textual data, like "Patent", "Analysis" or very broad words that are also used in other contexts, like "Patent Analysis", "Classification" or "Mining", are not included. A list is provided in Table B.2. The list of relevant keywords is then used to select the remaining articles for the next step. If the article contains non-relevant words, it will still be considered as long as it contains at least one relevant keyword.
3. The abstracts of 171 unique research articles are then assessed to identify the final sample. For the consideration of an article, a focus on text analysis in combination with innovation must be observable in the abstract. For this purpose, we reviewed all abstracts and evaluated them concerning their focus on innovation and text data. In case of ambiguity, the articles' content was also considered. The final sample consists of 23 articles (see Table 3.1).

3.2.2 Data Description

The 23 articles were published in 2002 and between 2014 to 2021.[4] Most articles have been published since 2014. Only the article of Zhu and Porter (2002) is published in the early 2000s. It is one of the early attempts to employ keyword-based analysis in innovation research. In absolute numbers, most

[3] The data extraction was conducted on June 21, 2021.
[4] Note that 2021 is only partially included in the data extraction.

Table 3.1: Overview of the Sample

Source	Title	Author Keywords	Journals
Antons and Breidbach (2018)	"Big Data, Big Insights?"	big data; research agenda; service innovation; service design; topic; modeling	Journal of Service Research
Bakhtin et al. (2020)	"The Future of Food Production – a Text-Mining Approach"	food innovation; agriculture; text mining; genetically modified; organisms	Technology Analysis and Strategic Management
Basole et al. (2019)	"Visual Analysis of Venture Similarity in Entrepreneurial Ecosystems"	ecosystems; entrepreneurship; technological innovation; industries; visualization; economics; cluster analysis; entrepreneurial ecosystems; strategic positioning; text analytics; visualization	IEEE Transactions on Engineering Management
Chiarello et al. (2021)	"Value Creation in Emerging Technologies through Text Mining"	blockchain; emerging technology; sentiment analysis; value creation	Technology Analysis and Strategic Management
Cripps et al. (2020)	"The Use of Twitter for Innovation in Business Markets"	social media; innovation; twitter; crowdsourcing; topic modelling; business to business marketing	Marketing Intelligence and Planning
Curci and Mongeau Ospina (2016)	"Investigating Biofuels through Network Analysis"	biofuel; innovation; patent data; topic model; text mining; network; analysis	Energy Policy
Dahlke et al. (2021)	"Crisis-Driven Innovation and Fundamental Human Needs"	content analysis; covid-19; human needs; innovation systems; topic modeling	Technological Forecasting and Social Change
Feng et al. (2020)	"The Technology Convergence of Electric Vehicles"	electric vehicle; link prediction; network analysis; technology convergence; text mining	Journal of Cleaner Production
Fiordelisi et al. (2019)	"Creative Corporate Culture and Innovation"	corporate culture; creative companies; creativity; innovation; r&d; textual analysis	Journal of International Financial Markets Institutions and Money
Kayser (2017)	"Comparing Public and Scientific Discourse in the Context of Innovation Systems"	innovation system; text mining; foresight; media analysis; publications; analysis; future technology analysis	Technological Forecasting and Social Change
N. Kim et al. (2015)	"Dynamic Patterns of Industry Convergence"	co-occurrence-based analysis; industry convergence; industry convergence; index; industry convergence map; unstructured data	Research Policy
J. Kim and C. Lee (2017)	"Novelty-Focused Weak Signal Detection in Futuristic Data"	weak signal; novelty detection; futuristic data; text mining; local; outlier factor; signal-portfolio map	Technological Forecasting and Social Change
Kohler et al. (2014)	"Service Innovation Analytics"	innovativeness of firms; machine learning; service innovation; service innovation capability; service innovation framework; text documents; text mining; unstructured data	International Journal of Information System Modeling and Design
Larsen and Thorsrud (2019)	"The Value of News for Economic Developments"	news; latent dirichlet allocation (lda); business cycles	Journal of Econometrics
Mi et al. (2021)	"Forecasting and Evaluating Emerging Technologies Based on Supply and Demand Matching – a Case Study of China's Gerontechnology"	emerging technologies; gerontechnology; semantic analysis; supply and demand matching; technology forecasting	Technology Analysis and Strategic Management
Ozcan et al. (2021)	"Social Media Mining for Ideation"	crowdsourcing; decision-making; semi-supervised learning; support vector machines; sustainability; text mining	Technovation
Shen et al. (2020)	"Discovering the Potential Opportunities of Scientific Advancement and Technological Innovation"	text mining; orclus; cosine similarity of tf-idf vectors; smart health; monitoring	Technological Forecasting and Social Change
Song et al. (2017)	"Discovering New Technology Opportunities Based on Patents"	new technology ideas; patent analysis; f-term; text-mining; technology; opportunity	Technovation
B. Wang and Z. Wang (2018)	"Heterogeneity Evaluation of China's Provincial Energy Technology Based on Large-Scale Technical Text Data Mining"	data mining; energy technology; heterogeneity evaluation; large-scale text data; lda topic model	Journal of Cleaner Production
Wu et al. (2020)	"Screening Patents of ICT in Construction Using Deep Learning and NLP Techniques"	ict in construction; nlp; deep learning; information management	Engeineering Construction and Architectural Management
Y. Zhang et al. (2019)	"Discovering and Forecasting Interactions in Big Data Research"	technological evolution; text mining; bibliometrics; big data	Technological Forecasting and Social Change
Zhou et al. (2020)	"Identifying and Assessing Innovation Pathways for Emerging Technologies"	altmetrics; bibliometrics; clustering algorithms; gold nanoparticles; machine learning; principal component analysis; sentiment analysis; sentiment analysis; technological innovation; technological innovation pathways; text mining	IEEE Transactions on Engineering Management
Zhu and Porter (2002)	"Automated Extraction and Visualization of Information for Technological Intelligence and Forecasting"	competitive technological intelligence; technology forecasting; text; mining; innovation indicators; technology maps	Technological Forecasting and Social Change

The table offers an overview of the research articles in the final sample, arranged in alphabetical order by authors.

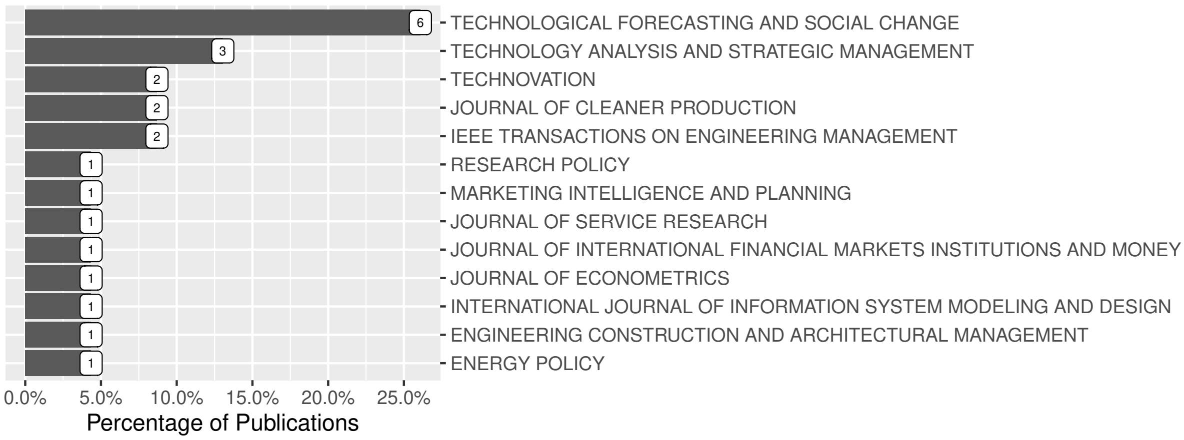


Figure 3.1: Publishing Journals in the Sample.
The journals in which the articles are published are listed, along with the absolute and relative numbers of articles in the sample. The relative numbers are expressed as a percentage of the total sample, for instance, six out of 23 articles, which represents 26%.
Source: Own representation based on the articles in the sample.

articles are published in the year 2020. The articles are published in 13 different journals, which can be derived from Figure 3.1. “Technological Forecasting and Social Change” is the most prevalent journal, with six out of 23 articles. It is a journal related to technological forecasting and future studies and a known innovation policy journal (Elsevier 2023b). In addition, “Technology Analysis and Strategic Management”, “Technovation”, “IEEE Transaction on Engineering Management” and “Research Policy” are also known innovation policy journals related to the included articles. All these journals have a focus on innovation. In total, the innovation-related journals contribute 14 out of 23 articles in the sample. The “Journal of Econometrics” with one article in the sample, is a methodological journal for theoretical and applied econometrics (Elsevier 2023a). Apart from these journals, the remaining eight articles are field-specific and relate to service research, marketing, energy policy, finance, construction, sustainability, or informatics. In the paper of Antons et al. (2020), the most important journal is also “Technology Forecasting and Social Change”. However, more applied journals that contain innovation research and contribute to the field are neglected.

3.3 Literature Review

The review is structured according to the four main fields of interest into four Subsections: First, the authors’ motivations for using text analysis to measure innovation are discussed (Subsection 3.3.1). Second, the data sources applied in the sample articles, and the use of these sources is considered (Subsection 3.3.2). Third, in Subsection 3.3.3, the authors’ principal methodologies are presented. Finally, the research questions the authors answer with textual data analyses are revealed. For better understanding, the articles are structured according to their content. Main topics considered are “Identification of Innovation”, “Technology Discovery and Exploration”, “Industry Patterns and Regions”, “Innovation Trajectory and Diffusion”, “Emerging or Novel Topics”, “Idea Generation and Extraction”, “Demand Assessment” and “Prediction Measure”. Further details on the articles and insights described in this chapter are disclosed in Appendix B.

3.3.1 Reasoning for Textual Data Analysis

This subsection provides an overview of why the sample articles’ authors use textual data and its analysis. Arguments that the authors explicitly state are mentioned. In addition, the authors’ advantages of text analysis are elaborated, particularly concerning the concrete application of text-based analyses to answer various research interests. The reasons mentioned are categorised into Data, Methodology, Objectives and Visualisation according to the main field of the reason. For example, if the reasoning relates to the data, then the reason is assigned to the data category. Objectives cover the main authors' intentions of the application of textual data analysis to answer their research questions: For example, eight articles intend to uncover hidden information through the application of textual data analysis. All reasons are listed in Table 3.2.

Data: In the data category, the possibilities of analysing large data sets or textual data, as well as the objectivity of analyses, i.e. data-driven analysis, are stated by a third of the articles as reasons for using text-based analyses. Furthermore, the possibility of combining different data sources is listed. In addition, one article each mentions the cost-efficient use of public data and the processing of all information, which means that a restriction to, for example, only the structured data does not have to take place.

Methodology: Methodologically, the most important argument is the fast and efficient analysis of textual data. In this respect, previous methods, such as surveys or manually reading the respective data sources, take much work. The advantages of direct analysability, repeatability, and reproducibility are related to this. The data can be evaluated directly without expert involvement. The direct evaluation and reduction of manual factors also increase the reproducibility of the results.

Objectives: The most important objectives are to uncover hidden information from the data and perform time-based analyses to discover trends over time. Furthermore, new opportunities for innovations or emerging themes are to be identified, i.e., themes that will become more relevant in the future are recognised. Four articles each examine relationships between, for example, topics or firms, use the text-based methods to divide the information into classes or use the possibility of obtaining a greater degree of detail and thus a better understanding of the object of investigation through the text-based methods. Also, the possibilities are used to identify the important information, distinguish it from the unimportant data, and verify findings obtained elsewhere.

Visualisation: Finally, several articles in the sample use the possibility of visualising the textual data and the analysis findings.

To summarise, authors value the possibility of efficiently analysing large data sets of textual data. The focus is mostly on discovering information, trends, opportunities or topics. The textual analysis provides mostly an overview that allows for further elaborations.

3.3.2 Data Sources Used

Both patents and publications are the most frequently analysed data sources, with 15 out of 23 articles either using patents, publications or both (Table 3.3).

Patents: In most cases, patents represent the current technological state of development or the technologically available supply. These are usually provided by the United States Patents and Trademarks Office (USPTO). In addition, patents are obtained from the JPO (Song et al. 2017), which includes patents from Japan, the global databases Incopat (Mi et al. 2021) and the Derwent Innovation Index (Feng et al. 2020). Finally, BioPat is used for technology-specific research in the field of biotechnology (Curci and Mongeau Ospina 2016). Where indicated, 245 to over 41,000 patents are used as a basis for the analyses.

Publications: Publications are used as a data source to represent progress in the field and the frontier of knowledge in basic research and emerging topics. Five out of eight articles obtain their publication data from Web of Science. Scopus is only used once. Furthermore, Cross Ref is used for scientific

Table 3.2: Overview of Reasons for Textual Data Analysis Application in the Sample

Category	Reason	No. Articles	Relative No. Articles
Data	Large Data Set	9	39%
	Objectivity	7	30%
	Enabling Unstructured/ Textual Data use	6	26%
	Multiple Data Sources	4	17%
	Cost-efficient Public Data	1	4%
	Full Information use	1	4%
Methodology	Efficient and Rapid Analyses	7	30%
	Direct Analysability	4	17%
	Reproducibility	3	13%
	Repeatability	2	9%
Objectives	Uncovering Latent/ Hidden Information	8	35%
	Time-based/ Trend Analyses	8	35%
	Opportunity Detection	6	26%
	Identification of Emerging Topics	6	26%
	Assessment of Relationships	4	17%
	Classification	3	13%
	Gain Detailed Insights	4	17%
	Identification of Relevant Information/ Noise Reduction	3	13%
	Result Verification	1	4%
Visualisation	Visualisation	5	22%

The table offers an overview of the reasons for applying textual data analysis within the sample. These reasons are categorized into four groups. The absolute number represents the count of articles in the sample that mention each reason. The relative number is expressed as a percentage of the total sample of articles. For example, nine out of 23, which is 39%, of the articles emphasize the importance of analyzing large data sets.
Source: The sample for this evaluation is provided in Table B.3.

references (Bakhtin et al. 2020) and INSPEC is used (Zhu and Porter 2002) with a focus on engineering research.

News Articles: When working with newspapers, no main data source can be identified. The data sources used range from existing databases to self-crawled data from newspaper pages. The rationale for using newspaper articles also varies greatly depending on the research focus. The diffusion of themes (J. Kim and C. Lee 2017) or convergence in industries (N. Kim et al. 2015) can be captured.

Social Media : Chiarello et al. (2021) use Twitter as a knowledge source for innovation, sharing and crowdsourcing, while Ozcan et al. (2021) gain ideas from tweets related to consumer and product informations. Zhou et al. (2020) apply Altmetrics as a diffusion measure that provides information on the impact of research.

Firm documents: Firm documents entail venture descriptions, website information or annual reports like 10-Ks. These are used to reveal information on the firm like ecosystem information (Basole et al. 2019), cultural insights (Fiordelisi et al. 2019) or innovation capabilities (Kohler et al. 2014).

Others: Other data sources are, e.g., documents on specific topics to answer the research questions.

Combination: A total of six articles combine several data sources. This often involves comparing two fields to identify gaps or patterns. The combination of these data sources mainly occurs based on keywords in each data source and mostly includes patents or publications: Kayser (2017) compares the word occurrence of publications and news article keywords with pie charts. Mi et al. (2021) identify areas of insufficient supply by comparing publications and reports on the need of older people. Semantic correlation is calculated based on the similarity between demand and supply and words' frequency. Zhou et al. (2020) construct topics based on publications and evaluates them with Altmetrics data. The combination of the data occurs based on joined words. J. Kim and C. Lee (2017) identify novelties by comparison of patents with futuristic data, which contain information about future trends expected from experts. They calculate the occurrence of words or documents from the futuristic data set in the patent data set to determine the unrelatedness of the upcoming topics to the current technological state. Shen et al. (2020) point out where further technical research or scientific investigation is needed based on patent data and reports on emerging technologies. Therefore, clusters in each data set are determined before matching the clusters based on common terms. Apart from a keyword combination, only Bakhtin et al. (2020) apply textual data analysis to all their data sources jointly, here patents, publications, news articles and others, before performing further analysis. The authors do not disclose if data source structure differences are considered during the data preparation.

Patents and publications are still the main data source when it comes to innovation analysis. Other data sources mainly contribute to public opinion like news articles or provide inside perspectives of firms like Tweets from firms or firm-specific documents like venture articles or 10-k-filings. When several data sources are combined, patents or publications are mostly included. Trademarks are missing from the data sources used in the sample articles and are also not covered by the analysis of Antons et al. (2020). This might indicate that it is still a small field of research. In terms of innovation data sources, the data sources have two things in common: they are publicly available and accessible, and the link to innovation is ensured as the data sources are either related to innovation like patents or publications, venture articles or innovation project descriptions, their connection to innovation is assessed like in the case of the Tweets of firms (Chiarello et al. 2021) or focusing more on effects of innovation like diffusion aspects (J. Kim and C. Lee 2017) or convergence (N. Kim et al. 2015).

3.3.3 Methodological Approaches

The analysed articles apply different methods of text analysis. The methodologies used are provided in Table B.6. The methodologies are summarised, and the most important ones are presented in this

Table 3.3: Overview of Data Sources in the Sample

Data Source	No. Articles	Relative No. Articles
Patents	9	39%
Publications	8	34%
News Articles	4	17%
Social Media	3	13%
Firm Documents	3	13%
Others	4	17%
More than one Data Source	6	26%

The table presents an overview of the data sources utilized for textual data within the sample. The relative shares are expressed as percentages relative to the total number of articles in the sample.
Source: The sample for this evaluation is provided in Table B.4 and in Table B.5.

subsection. As this chapter focuses on analysing textual data in innovation research, all articles in the sample use a textual data source and analyse it in some form: Basic analyses that are used contain word counts or simple word co-occurrences. More refined approaches to analyse texts are also applied: Eleven articles in the sample use clustering algorithms, and three apply classification algorithms to address their research problem (Table B.6). The results align with the results of Antons et al. (2020), where most articles apply clustering algorithms rather than classifications.

Data Preparation: Textual data need to be prepared to apply refined text analysis approaches (Allahyari et al. 2017). A method to do so is the Term Frequency-Inverse Document Frequency (TF-IDF), which five authors apply as one of their main methods for normalisation (see Basole et al. (2019), Mi et al. (2021), Ozcan et al. (2021), Shen et al. (2020), and Wu et al. (2020)). TF-IDF should introduce greater stability to the text and increase word similarity. It determines the frequency of a word in a document (term frequency) and the word across the documents of a data set (document frequency). The term frequency is multiplied by the inverse document frequency. As such, the frequency of words is normalised, reducing the bias introduced by words with repeated frequency across documents (Aggarwal and Zhai 2012; Aggarwal 2015).

The textual corpus can be represented in a vector space model (VSM) based on word frequency. Cosine similarity can be used to compare these vectors and determine their similarity (Allahyari et al. 2017). Cosine similarity uses the angle between two vectors to determine their similarity. Smaller angles indicate higher similarity between the vectors. The measure is independent of the length of the vectors (Aggarwal 2015). Authors use this measure to assess the similarity of entrepreneurial ventures (Basole et al. 2019) and document clusters (Shen et al. 2020), to determine a connection between documents, here patents, (Curci and Mongeau Ospina 2016), or to assess the degree of technological overlap (Song et al. 2017).

Clustering: 11 out of 23 articles apply clustering methods. These, for example, summarise data or identify groups or communities in the data sets (Aggarwal 2015). Y. Zhang et al. (2019) apply K-means clustering to identify topics and their evolution. K-means clustering use the distance between data points to create clusters. It starts with randomly selected centres, assigns each data point to a centre and then iteratively optimises the centres of the cluster until no improvement is achieved (Aggarwal 2015). The authors apply clustering to articles and measure their similarity to identify topics. If new articles emerge, their similarity is measured against the existing topics. If it is similar to existing topics, it is considered state-of-the-art, while the novelty increases with increasing distance from existing topics. As such, the approach discovers the evolution of topics in the context of big data over time (Y. Zhang et al. 2019). Dahlke et al. (2021) apply hierarchical clustering to group 16 domains into six clusters, while J. Kim and C. Lee (2017) use it to determine the number of clusters in their data, which is used as input to assign keywords to the clusters. Hierarchical clustering creates a hierarchical order of the clusters, which can then be exploited (Allahyari et al. 2017). In the context of innovation research, both articles applied hierarchical clustering to minimise the number of clusters to be analysed.

The most frequently used approach in the sample for dimensionality reduction is topic modelling: Eight articles in the sample apply topic modelling. Seven use Latent Dirichlet Allocation (LDA) to identify underlying topics in the textual data because of its data-driven and time-efficient approach. LDA was introduced by Blei et al. (2003). The model assumes that each document consists of a topic mixture that needs to be estimated. It, therefore, uses two distributions: the distribution of topics over words and the distribution of topics over documents. These distributions are randomly drawn in the beginning. The distributions are optimised by comparing the estimated word distribution in the document with the actual occurrence of words in the document. The authors mainly use LDA to identify topics in their area of research out of large data sets: Antons and Breidbach (2018) and Chiarello et al. (2021) use LDA on publications. Antons and Breidbach (2018) create a topic network of service innovation and service design research to highlight linkages and identify knowledge gaps, while Chiarello et al. (2021) identify nine blockchain issues. Three authors use patents for topic discovery: Curci and Mongeau Ospina (2016) create a network structure based on the patents. They then calculate the topics and assign each patent to an identified biofuel topic to discover shared characteristics among the patents. Feng et al. (2020) identify promising converging topics in electric vehicle patents based on promising links and B. Wang and Z. Wang (2018) identify 14 energy technology topics in energy patents. Other textual data used are innovation projects around Covid-19 to identify innovation domains (Dahlke et al. 2021) and news articles to assess the explanatory power of their topics on economic fluctuations (Larsen and Thorsrud 2019). In summary, LDA is used to summarise and extract topics from large data sets in innovation research. The results are then often analysed qualitatively.

Two articles apply refined versions of the LDA: Cripps et al. (2020) use Structural Topic Modelling (STM) to identify topics of innovative firms. In STM by Roberts et al. (2013), the estimation of the distributions can be modified with two different covariates that either affect the topic-word distribution or the topic-document distribution. In the case of Cripps et al. (2020), the authors divided their data, Twitter Tweets, into Tweets from firms with and without innovation and crowdsourcing purposes. This binary information is contributed to the model as a covariate that influences the topic-document distribution. Including the covariate enables the authors to analyse the Tweet topics concerning their purpose and identify the topics related to innovation. As the model enables the analysis of these covariates after topic estimation, the STM is interesting for studies of these covariates. Another approach is a provincial topic modelling introduced by B. Wang and Z. Wang (2018) introduce a provincial topic modelling. It is a modified model with an institution parameter to attain the probability of a topic occurring in a region. It allows the authors to explore the relationship between energy topics and regional contribution, which helps to formulate region-specific policy implications.

Classification: Three out of 23 articles address a classification problem. Classification problems address specific features of the data set and can be used to label data. They are supervised problems because the models learn the classification feature based on a training data set. The model then applies the classification to test data and estimate labels to the data. Classification approaches are considered more powerful than clustering as the classification is defined externally by the user (Aggarwal 2015). In the context of innovation research, the classification approaches are used to identify innovative firms (Kohler et al. 2014), ideas (Ozcan et al. 2021) or relevant patents of a technological field (Wu et al. 2020).

Other: Other methods used are e.g. sentiment analyses (Chiarello et al. 2021; Zhou et al. 2020), word embedding (Bakhtin et al. 2020; Mi et al. 2021) or network graph analyses (Antons and Breidbach 2018; Curci and Mongeau Ospina 2016; Feng et al. 2020).

In the context of innovation measurement, sentiment analysis is used to identify promising topics which have the best net effect in terms of expert expressions in relation to the evaluated topics (Zhou et al. 2020) or to identify problems that should be addressed by focusing on sentences with negative connotation (Chiarello et al. 2021). Sentiment analysis categorises statements into positive or negative based on the words used (Aggarwal and Zhai 2012). It intends to extract the attitude or feeling of the writers towards a subject or topic (Liu and L. Zhang 2012). Positive emotions are associated with words like "well" or "like" while negative emotions are expressed with words like "bad" or "hate" (Medhat et al. 2014). The two articles mentioned used the extracted opinion for evaluation.

Mi et al. (2021) use word embeddings to generate word vectors of the patent documents, while Bakhtin et al. (2020) build the topic-term relation with the approach. Word embeddings consider the context of words analysed to identify words occurring in a similar context and revealing similar relations. An example is "Paris" is to "France" like "Berlin" is to "Germany" where the model can be used to uncover such relationships (Mikolov et al. 2013).

Network graphs are used to identify important topics and measure network density (Antons and Breidbach 2018), to identify clusters and evolutions over time (Curci and Mongeau Ospina 2016; Y. Zhang et al. 2019), and to identify convergence (Feng et al. 2020). They are further used for visualisation of affiliations (Zhu and Porter 2002), country collaborations or citations (Y. Zhang et al. 2019), and term co-occurrences (Kayser 2017).

Several approaches enable the discovery of subjects in an extensive data set; here, the topic modelling approaches, word embeddings or network graphs of, e.g. words. Highlighting linkages or identifying linkages was further of interest, with, for example, linkages of topics, words, documents or patent classes. The classification approaches help in categorising the data. Most authors combine multiple approaches: Chiarello et al. (2021) combine sentiment analysis and LDA to identify interesting issues and the topics among these. Curci and Mongeau Ospina (2016) create a patent network with patents as the nodes that then are shaped and coloured according to their LDA topic to identify underlying characteristics. The suitable approach depends on the data analysed and the intention of the analysis. The overview shows that various methods are used depending on the main intention of the research. The approaches are mainly qualitative, focusing on discovering topics and linkages.

3.3.4 Analysed Research Questions

Before the content of the work is evaluated in detail, the relevant keywords are identified first. The most common author keyword in the sample is "text-mining" occurring in ten out of 23 papers (Figure 3.2). Related to text mining are keywords such as unstructured data or methodological information such as sentiment analysis, machine learning, big data, bibliometrics or analysis. Innovation and innovation-related keywords like technology forecasting, technological innovation or service innovation are further important in the sample.

Subsequently, the articles were structured according to their content focus and categorised accordingly. The main focus, purpose and contribution, and the articles' categories, can be seen in Table B.7 and Table B.8. The largest category is "Technology Discovery and Exploration" with six related articles. All the other categories contain two to three articles.

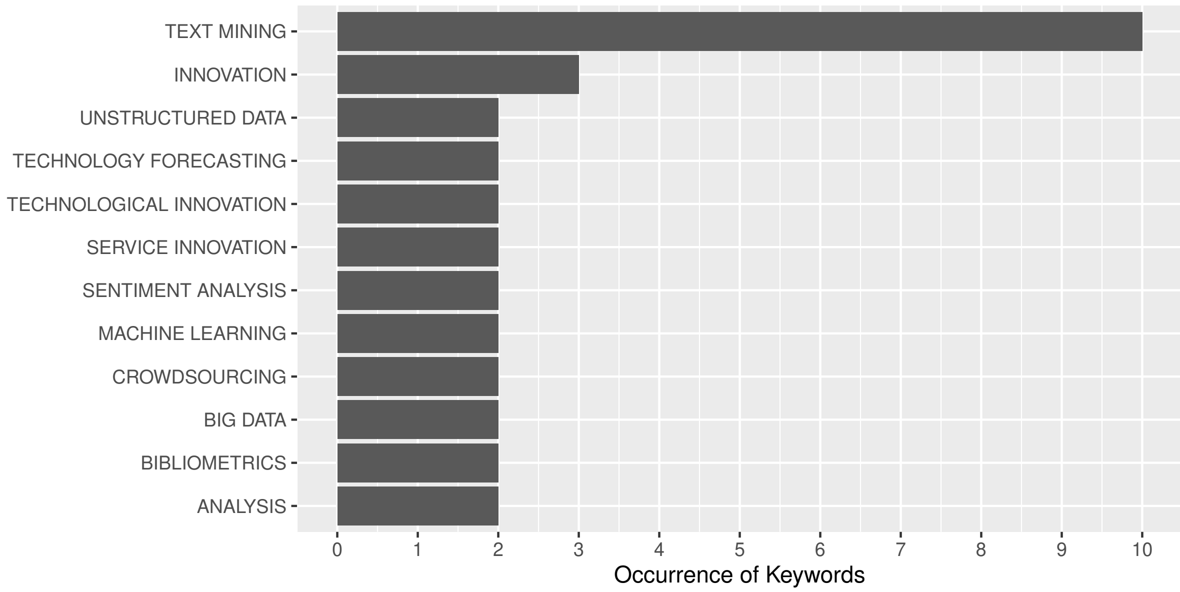


Figure 3.2: Most Frequent Author Keywords in the Sample.
The figure depicts the most frequently occurring keywords within the sample of articles. Only keywords that appear more than once are shown.
Source: Own representation based on author keywords of the sample.

Identification of Innovation In this category, the authors use different approaches to categorise or identify data for their innovation measurement: Wu et al. (2020) intend to identify USPTO patents of information and communication technologies in construction. Because the authors experience a lack of suitable patent classification and find the keyword-based patent identification insufficient, the authors train a deep learning model for identification. A relevant set of patent texts are used to train the model and identify other related patents. The authors show the model's advantages over other identification strategies and provide an approach for patent identification in fields without suitable patent classification. Kohler et al. (2014) classify firms' innovativeness based on firms' textual document. The model's prediction of the classification is compared to external classifications and provides good results. As such, the authors distinguish innovative from non-innovative firms. Cripps et al. (2020) apply Structural Topic Modelling to distinguish between Tweets from innovative firms from non-innovative firms. The authors assess the behaviour of innovative versus other SMEs on Twitter compared to interview results. They find that the interview results are verified by textual analysis and confirm that Twitter is used for open innovation and crowdsourcing. Overall, the authors in this category thus address situations in which the data sources do not provide an ensured innovation reference, or no delineated data set is available for the research question.

Technology Discovery and Exploration This category covers articles that intend to discover and explore the subjects relevant to technological areas. Already in 2002, Zhu and Porter provide a framework for technology assessment and knowledge generation based on textual data. This automated analysis generates technology maps and indicators that provide insights on keyword-based hot domains in the technology, potential growth opportunities, authors, affiliations, and countries involved (Zhu and Porter 2002). Curci and Mongeau Ospina (2016), Antons and Breidbach (2018), Cripps et al. (2020), and Chiarello et al. (2021) perform technology mapping with the help of topic modelling. Topics in, e.g. biofuels (Curci and Mongeau Ospina 2016), service innovation research (Antons and Breidbach 2018) and blockchain (Chiarello et al. 2021) are identified with the application of Latent Dirichlet Allocation (LDA) by Blei et al. (2003). Curci and Mongeau Ospina (2016) investigate the relevance of different production paradigms in biofuels. LDA is used for knowledge discovery, while cosine similarity compares the document vectors. This approach produces a network projection with edges based on similarity and nodes based on documents and topics. The projection reveals two main clusters of biofuels, alcohol and fat based. Chiarello et al. (2021) refined the topic modelling analysis of blockchain to focus on problems in the field. They included sentiment analysis in the data preparation to keep only sentences with negative scores. Negative scores stand for negative emotions. The sentiment analysis enables the authors to focus on problems in blockchain, as problems are associated with negative emotions. Bakhtin et al. (2020) apply word embedding and Cosine Similarity calculations to identify emerging topics within agriculture and food production. Relevant terms and their close neighbours are identified and clustered. Possible future trends can be predicted based on the occurrence frequency and annual term growth rate. The terms are categorised according to growth leadership, weak signalling, stable and niche areas. The different approaches enable the discovery of technological trends and developments and allow for the monitoring of technologies.

Emerging or Novel Topics Two articles focus on the detection of emerging or novel topics: Feng et al. (2020) explore topics and future trends in electric vehicles. To identify future trends, the authors perform a link prediction and identify promising converging relationships of IPCs. Therefore, IPC co-occurrences in patents are analysed, and the ones with the largest growth are identified. Patents with these promising co-occurrences are then extracted and explored via LDA. J. Kim and C. Lee (2017) also use patent data but consider it to represent the current state. Their goal is the identification of weak signals, as these are considered early indicators of future change. The potential signals are extracted from futuristic data, which are data about the future of technologies. Information novelty is measured with rarity in the data and unrelatedness from the existing knowledge base. The resulting signals are categorised into strong, stagnant weak and feeling signals, depending on their rarity and unrelatedness measure. In the case of augmented reality, weak signals relate to the application in gaming or maintenance but also to identity thefts or increased transparency and reputation impact due to real-time information on, e.g., restaurants.

Innovation Trajectory and Diffusion Zhou et al. (2020) and Y. Zhang et al. (2019) choose different ways to analyse the innovation trajectory. Zhou et al. (2020) identify 21 promising topics in gold nanoparticles and use sentiment analysis to evaluate their future potential. The relationship between topics, their ancestors and predecessors are then determined based on the main subject and object of the topics. Combined, the evolutionary path of the technology can be mapped. The market potential is finally judged based on scholarly activities, commentaries, and related social activities to reflect the need to attract society's attention for technological success. Y. Zhang et al. (2019) first visualise Big Data research with the journal citation maps, term-correlation maps, organisation collaboration and countries involved. They then determine the scientific evolutionary pathway. K-means is used to determine the initial topics in the first period. In subsequent periods, the similarity of the new articles to the identified topics is determined and, depending on the deviation from existing topics, considered state-of-the-art, evolutionary or novel. With the popularity of Big Data, already existing topics, such as "unstructured data", "predictive models" or "machine learning", have been revisited. Kayser (2017)'s analyses focus on the diffusion of innovation. Comparably to Zhou et al. (2020), public perception is used to measure diffusion based on the media discourses. The authors compare publication abstracts and news article terms in three case studies. Pie bubble analyses illustrate the term use per data set (Kayser 2017). Overall, the authors identify a starting point of an innovation or topic area and project or identify the trajectory or diffusion.

Idea Generation and Extraction Link detection can further be used to identify innovation opportunities. Textual data of technology, measured via patents, and state of scientific research, measured via publications, is therefore represented in a Vector Space Model. The evolving topical clusters in technology and science can be compared via the cosine similarity of the vectors. This reveals connections between the fields, and missing links become obvious. Addressing these links can be an opportunity for further research or innovation (Shen et al. 2020). Song et al. (2017) generate new ideas by transferring technological components from one technology to another. The authors start by defining a target technology. Based on the target technology keywords, possible reference technologies are identified using cosine similarity and selected based on their similarity or growth potential. The components of the reference technology are then extracted and compared to the actual components of the target technology. This comparison can reveal gaps and provide opportunities for new ideas by transferring components already used in the reference technology to the target technologies. The field of automotive braking systems in combination with ICT implies, e.g. the detection of signal phases or the application of time-division transmission for noise reduction. Twitter also serves as a data source for new ideas: Ozcan et al. (2021) extract Tweets related to sustainability and idea. To only focus on Tweets with new ideas, the data is then labelled and classified as an idea or non-idea with the help of classification approaches. The final ideas are grouped into sustainable production, education, packaging, and energy.

Demand Assessment In this category, the authors analyse if existing demands are met by new innovations: Dahlke et al. (2021) examine how innovations that have emerged in the context of the Covid-19 pandemic address people's needs. 707 crowdsourced innovations are assigned to 16 innovation domains using LDA for topic modelling. The authors find that new innovations are meeting emerging human needs: For example, the high need for medical and protective equipment at the beginning of the pandemic situation decreased, indicating that the need was met by innovation. Further, the increasing need for virtual space for interaction is met by firms preparing long-term solutions. Mi et al. (2021) match technological demand and supply in China's gerontechnology field. Technological demand is extracted based on surveys and reports on the ageing population. Patents about technology demand and old age provide the supply perspective. A semantic correlation of demand and supply is used to extract the saturation intensity. Undersaturated technologies show a higher demand level than the provided supply level, implying that further development is needed.

Industry Patterns and Regions Different approaches are possible to capture industry patterns and regional aspects. N. Kim et al. (2015) measure within and between industry convergence based on firms' co-occurrence in news articles. The link of firms to industries is performed via their industry classification. Out of firm co-occurrences, industry pairs are gained, which can be mapped and positioned in a 2-dimensional space. Finally, the authors find that within-industry convergence is greater than between-industry convergence. Overall, an increase in convergence is observable. However, the pattern is dynamic as some industries converge while others do not. Basole et al. (2019) map the ecosystem of entrepreneurs to provide a granular view of the ecosystem. The position within the ecosystem is determined based on the cosine similarity of firms' position statements. After that, communities are automatically detected and described with their TF-IDF keywords. The proposed approach enables an endogenous, data-driven clustering that reveals patterns besides the traditional industry systematics.

Finally, B. Wang and Z. Wang (2018) analyse China's energy technology development with LDA and a provincial topic modelling: 14 topics cover the main categories of energy savings, emission reduction and household energy in different provinces of China. The similarity in topic coverage is compared to provide regional policy recommendations.

Prediction Measure Fiordelisi et al. (2019) use textual analysis to determine the creativeness of company cultures based on words used in firm documents. This indicator is then used in regression analyses to assess the impact of company culture on innovation output. Based on the regression results, the authors conclude that creativity-fostering culture has a positive impact on successful innovation outcomes. Larsen and Thorsrud (2019) build an indicator based on LDA-extracted news articles topics and assess the predictive power of these news topics in regression analyses. Even though some topics are noise, many are considered relevant for predicting output growth and consumption. The authors thus conclude that news articles have predictive power for the economic outcome.

To conclude, the research questions answered often generate an overview of the current state of technologies, emerging areas or innovation trajectories. Some authors try to generate or extract new ideas or identify unmet demands. Different authors tackle the positioning of industries, firms or regions. Most of these approaches are exploratory. Only two authors combine the insights from textual data analysis in regression models to generate predictive measures.

3.4 Discussion and Conclusion

In the chapter, a detailed analysis of 23 articles is performed in the context of textual data and innovation research. "Technological Forecasting and Social Change" or other relevant innovation policy journals published most of these articles. However, a third of the articles are published in a specific field in which the innovation analyses are performed. The articles apply the analysis of textual data as this enables them to analyse large data sets data-driven. Text analysis methods enable rapid, time-efficient analysis. The authors perform time-based analyses, detect opportunities, or uncover hidden information. For their analyses, primarily patent data and publications are used. However, other data sources are also relevant, especially news articles, social media or firm documents. Six articles combine several data sources to address their research questions, including insights about, e.g., firm culture, customer feedback, and public reaction in innovation analyses. Many authors address their research questions with advanced methods of textual data analysis. Eleven articles use clustering algorithms, and, especially LDA. Three apply classification methodologies. While authors use classifications to identify ideas or innovative firms, clustering is more exploratory and used to identify topics and opportunities in the text corpus. The research questions answered by the articles in the sample are various. One of the main research areas addressed is discovering and exploring technologies, mostly based on topic modelling. Further areas of interest are industry convergence and regions, innovation trajectory and diffusion, and the emergence of novel topics or idea generation. Text analysis is also used to assess technologies' supply and demand and extract features for predictive models.

The reasons for applying textual data analysis are still primarily based on the large data to be analysed, the rapid analysis and the data-driven approach. Repeatability and reproducibility are also highlighted. The argumentation is often on cost and time. Only some authors compare their results from textual data to insights from interviews (Cripps et al. 2020) or external classifications (Kohler et al. 2014) for validation and, as such, show that their approach works and surpass traditional approaches. The discussion on when textual data add value to the discussions is an ongoing subject to address. For us, textual data is interesting in areas where common structured data reach their limits. Limits are, for example, no or insufficient classifications for data identification, the missing structure of the data to be analysed or insufficient data coverage in common data sources.

Even though textual data analysis allows the exploration of data sources besides patents and publications, most articles in the sample still focus on these. Trademarks, however, are not covered by this review and are also not part of the 124 articles analysed by Antons et al. (2020). Based on this, it can be said that the textual data of trademarks is not yet among the standard data sources of text analysis in innovation research, even though they hold a potential for measuring services or market-related innovation (Millot 2009). Here, two things can be considered: New opportunities in standard data sources and new opportunities in new data sources. Patents and publications are established data sources for innovation studies. Studies focusing on these data sources tried to extract new insights from these already well-studied data sources. They combined structured data like classification classes and unstructured data like keywords (see (Song et al. 2017)) or compared them on a more detailed level than formerly possible (see (Shen et al. 2020)). Studies outside of these data sources need to ensure the linkage of the data to innovation, which can be challenging. Identifying innovation and classification approaches were addressed in this overview. Authors tried different approaches to address this challenge, which allowed them to distinguish and analyse different data points. The innovation linkage still also remains a challenge in trademarks. Further, the authors solved this by combining new data sources with patents or publications. Patents or publications then still provided the state-of-the-art innovation perspective while the new data sources amended additional information. The additional data source should add something to the current discussion that is not otherwise covered and unique for the data source. For example, Fiordelisi et al. (2019) extracted the firm culture, or news articles are used to extract the public opinion. Interesting are combined approaches, where expert knowledge is used to generate further insights: For example, Zhou et al. (2020) use expert expressions for topic evaluation. Further, the classification approaches also require expert knowledge to generate the training data.

Method-wise, clustering and LDA were often applied, especially for discovering and exploring technologies. These are data-driven approaches and do not require expert knowledge for topic extraction. Besides LDA, a large variety of different methods was applied. Authors further combined different approaches to answering their research questions and account for the problems at hand. This resembles the complexity of textual data and the wide variety of questions to be addressed. However, a standard approach that should be used for specific problems has yet to be established. This would improve the comparability of results.

The studies were still mainly explorative and qualitative, which aligns with the focus on more explorative clustering methods. The studies provided, for example, a general overview of the current state of technology. This is quick and repeatable with the approaches. The insights gained can then be used in further studies. Some authors tried to, for example, predict links to transfer the current state of one technology to another (see Song et al. (2017)) or use insights from textual data to predict the economic situation (see Larsen and Thorsrud (2019)). However, these are still the minority of the applications of textual data analysis. In particular, text analyses often stand on their own and are not incorporated into existing modelling concepts as in Song et al. (2017) or Fiordelisi et al. (2019). However, a common use should be aimed at in order to use the capabilities of all approaches optimally.

The chapter contributes to understanding the current state of textual data analysis in innovation research. It provides the authors' reasons for the analysis of textual data and the application of text analysis, displays the data sources available and considered, the methodologies applied and, most importantly, highlights the areas of innovation research in which text analysis is currently used. The methodologies available for text analysis will improve further. Data sources formerly not possible will become interesting research outlets. The research in the sample articles was still primarily exploratory and descriptive. Further research needs to be done to include the insights gained from text analysis into economic indicators and models. Further research is also necessary on integrating textual data of trademarks in innovation research. Also, combining different textual data sources and approaches needs further refinement. So far, existing approaches and text analysis still stand separately next to each other.

Text analysis has its strengths, mainly because it opens up textual data for analysis. However, text analysis is not better than existing approaches. It mainly addresses other questions and has other fields of application. Ideally, text analysis methods expand the standard repertoire of analyses and, thus, of questions that can be answered and data that can be analysed. A case-by-case consideration is necessary, considering the advantages and disadvantages of the approaches. The different methods are then combined and used depending on the question. Overall, the overview helps to understand the current use of textual data and the possibilities its analyses bring to innovation research. As such, it can serve as an inspiration for further text-based analyses. We encourage the consideration of textual data to answer innovation-related questions.

4 Innovative Activities Captured through Semantic Analysis of Patents and Trademarks.

Chapter Abstract

The chapter analyses patent and trademark textual data for innovation research. It thereby focuses on two main questions: Can patents and trademarks be combined based on their textual data, and does the combination add new insights to the innovation discussion? Therefore, the textual data of patents and trademarks are combined via Structural Topic Modelling in Robotics and Footwear. The analysis first focuses on the consistency of the text-based combination before analysing the results regarding innovation in general and specifically for the product, market, service, and technical innovation to assess additional insights from the combination. The analysis reveals that the textual combination is possible and that trademarks contribute a service and application perspective. However, trademarks' textual structure differs depending on the innovation field covered, influencing the possibility of textually combining patents and trademarks.

KEYWORDS:	Innovation, Intellectual Property Rights (IPR), Patents, Trademarks, Unsupervised Machine Learning, Structural Topic Model (STM), Text Mining
JEL Code:	O34

4.1 Motivation

Dziallas and Blind (2019) review the indicators used in the literature to measure innovation. They find patents are a standard measure for product innovations or innovation in the product concept phase. Patents protect inventions, not innovations, meaning they have an inventive step but not necessarily a market introduction. Moreover, not all inventions are protected by patents or protectable by patents (Basberg 1987). According to the United Patent and Trademark Office (USPTO), a patentable invention requires a technical or industrial process, a machine, a manufactured article or a composition of matter such as chemical compositions (USPTO 2017; USPTO 2015). However, innovation entails more than just technical innovations. According to the definition of OECD and Eurostat (2018), innovation requires a significant improvement in the firm's status quo of a good, a service, a knowledge-intensive offering, or a business process in areas such as production, marketing and sales, or information and communication systems. Innovation is, therefore, not limited to technical areas but is more broadly defined. Nevertheless, services are only patentable in the case of technical components of the invented product.

Millot (2009), Mendonça et al. (2004), and S. J. H. Graham and Hancock (2014) suggest the use of trademarks as an alternative intellectual property-based innovation measure to capture service, business process or marketing innovation in addition to technical product innovation. Trademarks can capture non-technological inventions, intangibles and services, commercialised inventions, and new inventions to the firm or sector. They are systematically collected, classified, and disaggregated for different levels (Flikkema et al. 2015). Trademarks are brands used for product or service identification and differentiation (S. J. H. Graham et al. 2013; Millot 2009). Their application has increased over the years (Myers 2013) and so has their importance as a basis for innovation measurements. Compared to patents, a trademark application requires its use but is not inspected for its inventive step, so the relation to innovation is not ensured. Therefore, researchers like Flikkema et al. (2015) and Castaldi and Dosso (2018) conduct surveys of firms to capture the innovative value of trademarks or select innovative firms to measure their use of trademarks or patents. In joint analyses of trademarks and patents, authors often consider whether the firms use trademarks and patents separately or together (Llerena and Millot 2020; Grazzi et al. 2019; Thomä and Bizer 2013). The possibilities to analyse trademarks are, however, limited. The trademark system does not provide a similarly sophisticated classification system with a high level of detail as the patent system does. The trademark classification system is a limitation to be considered for analyses of innovation.

Trademarks and patents thus have their strength and also their shortcomings. Trademarks can capture non-technical and other types of innovation and ensure market introduction, while patents provide an invention perspective. Combining these data sources could bring interesting new insights into innovation. Malmberg (2005) suggests using patents and trademarks to capture innovation activities. Ribeiro et al. (2022) find that this joint analysis contributes across sectors, but it is especially of interest in sectors with low patentability, like non-technological sectors. Different approaches exist to combine the data sources: Flikkema et al. (2015) use the same legal entity to identify joint trademark and patent applications. Flikkema et al. (2014) approach trademarking firms to analyse their innovation protection strategies and identify protection with other intellectual property rights (IPRs). Combining the data sources via industry-level concordance tables is also possible. A less explored option is the use of textual data to extract detailed insights from the data sources. Textual data can provide detailed information, but the analysis of this data is not yet part of the standard analyses (Antons et al. 2020). Using textual data also enables the combination of different data sources via their texts: Bakhtin et al. (2020) combine data sources such as patents, publications or news articles via their sentences to perform joint analyses. J. Kim and C. Lee (2017) and Shen et al. (2020) identify future areas of development based on the comparison of patents with other data sources. Thereby, each data source contributes different information to address the research problem. In Chapter 3, many authors value the possibilities of text analyses to detect opportunities and uncover hidden information. In the context of trademarks and patents, for example, H. Kim et al. (2017) combine patents and trademarks based on shared keywords to identify potential areas for diversification, while M. Lee and S. Lee (2017) compare the technological knowledge of firms extracted from patents with the market positioning of competitors from trademarks. However, these text-based approaches mainly focus on keywords. More sophisticated methods to combine patents and trademarks on a detailed level could thus provide further information on innovation, including inventions and their diffusion. Yet, the review of Castaldi (2020) reveals that the available trademark information is only partially used. Analyses still focus on the structured information of a trademark, such as the year of filing, NICE class application or ownership. The potential of unstructured, textual information remains largely untapped.

In this chapter, patents and trademarks are therefore combined based on their textual data for various reasons:

- The combination of patents and trademarks might help overcome the limitations of each data source, e.g., the missing link to services, low technologies and diffusion in patents or the uncertain link to inventions in trademarks. The textual data of each data source might further enable the combination of patents and trademarks on a detailed level, independent of existing classifications and concordances. The chapter investigates if this combination on a detailed level is possible.

- The combined textual data set might cover the innovation area under study more comprehensively, allowing trends or insights to be captured more holistically. This chapter, therefore, investigates if the textual data of the data sources add to the innovation discussion. It therefore analysis if insights from previous studies on trademark and patent use can equally be reproduced based on the textual information provided in trademark and patent documents. The literature provides conclusions concerning the coverage of innovation in general and specifically for product, market, service, and technical innovations for patents and trademarks. These assumptions are expected to be transferable to textual data of patents and trademarks.

Structural Topic Modelling (STM) of Roberts et al. (2013) provides a convenient means to combine these two data sources. STM identifies the underlying topics in a set of documents. It further allows to differentiate between different features, here aspects of patents and trademarks, which can then be used for investigation. In this chapter, these features are then used to ask whether combining trademarks and patents is plausible and comprehensive, whether trademarks do in fact relate to innovation and whether the combination of patents and trademarks provide deeper information useful in considering market, service and technological aspects of innovation. The analysis is performed for the innovation areas of Robotics and Footwear. The combined results are synthesised to reject or support the assumptions.

The chapter offers the first in-depth description of the textual combination of patents and trademarks via their textual data. The combination of the data sources is mainly plausible. Further insights are gained on the textual differences between trademarks and patents and trademarks themselves. The topic-wise analysis of patents and trademarks highlights the market introduction of the patent-protected inventions. In conclusion, textual data of patents and trademarks provide the opportunity to combine the data sources on a detailed level and to provide further information on the innovation area and an overall broader perspective on innovation through better coverage of technical and service aspects. The combination can be used especially for studies where the market or service aspect of innovations is of interest or patents have known limitations. Overall, the chapter contributes to a better understanding of textual data of trademarks in the context of innovation.

The chapter is organised as follows: The theoretical background of trademark-innovation studies is given in Section 4.2. Section 4.3 describes the data sets and the methodology. The results are presented in Section 4.4 and Section 4.5 and synthesised in Section 4.6. Limitations of the research are revealed. The chapter concludes in Section 4.7.

4.2 Trademarking Literature

According to the USPTO (2020b, p.2) a trademark is "a word, phrase, symbol, or design, or a combination thereof, that identifies and distinguishes the source of the goods of one party from those of others." The same condition applies to the definition of service marks. Trade- and service marks enable customers to attribute goods or services to their origin. In contrast to patents, an inventive step is not required, but the use of the trademark in the marketplace is required. The market application and diffusion aspects make trademarks interesting for innovation research.

Trademark-Innovation Linkage As trademarks are not assessed for their inventive steps, several researchers try to establish whether there is a link between innovation and trademarking: Potepa and Welch (2018) show that patent announcements or trademark counts provide consistent and robust results to serve as proxies for a firm's innovation value. Flikkema et al. (2014) analyse 660 surveys of Benelux SMEs to explore the reference of trademarks to innovations. They find that 58% of trademarks link to a product, process, or service innovation. 33% of the overall trademarks refer to product innovation, 28% to process innovation and 35% to service innovation. In the case of new products, 25% are protected by patents and trademarks, while patents and trademarks simultaneously protect only 3% of the service innovations. Combined with additional intellectual property rights (IPRs), it becomes evident that IPRs other than trademarks cover 44% of product innovations but only 27% of service innovations. The survey of J. H. Block et al. (2015) among 600 mainly U.S. SMEs reveals mixed results regarding the objectives to trademark: According to the study, manufacturing firms use trademarking to distinguish from competitors but protect their inventions with other mechanisms. Service firms combine the protection, marketing and exchange objectives, primarily since trademarks serve as the most crucial IP right in services (J. H. Block et al. 2015). Young, small firms with high R&D spending combine formal (e.g., patent, trademarks, designs) and informal (e.g., secrecy) IP mechanisms to protect their innovations (Veugelers and Schneider 2018). In the knowledge-intensive business service sector, the likelihood of trademark registrations is significant for product-related services. Trademark registrations further link to successful innovation, especially in the case of product innovations (Gotsch and Hipp 2012). Gotsch and Hipp (2012) conclude that trademarks can serve as an innovation indicator that provides a market-based perspective due to their market application. Flikkema et al. (2019) conducted an online survey among applicants of the European Community or Benelux trademarks to find innovation and non-innovation-related trademark properties. Based on their observations, they conclude that 1) new trademarks protect goods innovations more frequently than service innovations, 2) the first trademark for start-ups or brand creation is more likely to stand for goods innovations than the first trademark of mature firms, 3) trademarks with a broader geographical coverage are less likely to be related to service innovations, and 4) trademarks with a narrower scope of protection are more likely to be related to goods innovations. In conclusion, there is strong evidence that trademarks are linked to innovation even though no checking of the inventive step occurs during the registration process. This fact makes this data source especially interesting for innovation research. As trademark applications link to the introduction of innovation, the trademark textual data could further reveal information on the current state of the diffusion.

Trademarks and Patents Trademarks and patents are used differently by firms and for research studies: According to Flikkema et al. (2015), combining trademarks with other data like patent data can enhance the insights gained on trademarking. The authors survey to investigate the complementarity of trademark and patent information and the trademark-innovation linkage. In the survey, trademarks relate to patented and unpatented innovation. For example, the authors find that trademarks filed to create or extend brands capture innovations that are 80% not covered by patents. The combined use changes depending on the service or manufacturing background of the applicant. Grazzi et al. (2019) consider the impact of Italian manufacturing firms' trademark and patent portfolios on their performance. The analyses reveal a positive relationship between IP registration to firm performance measures. In the model of Llerena and Millot

(2020), the incentives of firms to hold all combinations of patents or trademarks are analysed. The model assumes a duopoly situation in which firms can use trademarks or patents individually, jointly, or not at all as a protection mechanism. The model reveals that high advertisement depreciation rates lead to patent or trademark substitution, while high potential advertising spillovers imply complementary use of the IPRs.

Overall it can be said that in joint analyses of patents and trademarks, patents often stand for innovation (Flikkema et al. 2015), technology (Bei 2019) or technological capabilities (Nam and Barnett 2011; Castaldi and Dosso 2018). In contrast, trademarks stand for complementary assets (Bei 2019), market capabilities (Castaldi and Dosso 2018), or commercialisation (Nam and Barnett 2011). While patents are applied more by high-tech manufacturing or large, mature firms, trademarks are applied in low-tech manufacturing and services (Castaldi 2020).

Textual Data of Trademarks In trademark studies, textual data of trademarks are mainly not utilised: Castaldi (2020) compares trademarks to patents in detail with regard to organisational assets, market strategies or capabilities. The author finds that regarding trademark information used, most studies focus on the number of trademarks, stock information, and first filings in NICE classes or countries. Most studies still focus on whether firms use trademarks and patents separately or together (Grazzi et al. 2019; Llerena and Millot 2020; Thomä and Bizer 2013). One criticism that can be made is that the simultaneous filing of a patent or trademark application by the same firm only sometimes means that the protection covers the same subject matter. Further, the level of detail available with conventional combination methods, such as through concordance statistics, needs to be improved. These problems could be avoided through the use of textual data.

First attempts of textual data use in trademarks are performed by Semadeni (2006), Semadeni and Anderson (2010), and Flikkema et al. (2019). The mark name, together with NICE coverage, is used by Flikkema et al. (2019) to derive brand strategies of creation, extension or modernisation based on the similarity between different marks. Semadeni (2006) and Semadeni and Anderson (2010) use words of trademark descriptions. Semadeni (2006) analyses the positioning of firms relative to other firms based on their wording in trademark applications. The results suggest that firms need to balance their positioning: Firms might locate their services close to each other with a large textual overlap in their marks to increase legitimacy, decrease uncertainty or seek spillover benefits. Firms position themselves at a distance to reduce competition overlap and differentiate themselves. Hereby, small firms try to benefit from a near position of their services to others. As firms grow larger, a higher distance is intended to differentiate from competitors. In Semadeni and Anderson (2010), the authors assess the likelihood of imitation in service offerings. The imitation variable is calculated based on trademark descriptions. The trademark description-based indicator reveals that the more radical an innovation is, the less likely it is to be imitated. This is due to the inherent uncertainty of innovation. However, with an increasing reputation of the firm introducing the new service offering, the imitation likelihood increases.

M. Lee and S. Lee (2017) and H. Kim et al. (2017) display first attempts to combine patents and trademarks based on their textual data: M. Lee and S. Lee (2017) analyse how competitors with similar technological knowledge position themselves in the market. Patents display technological knowledge, while trademark NICE classes are used for business area identification. After the market opportunities have been identified, text mining is applied to the trademark descriptions to obtain further details on the competitor's market positioning. H. Kim et al. (2017) further enhance the analyses. They measure concentric diversification to define how firms diversify with new products or services. The authors use USPTO patents and trademarks to build a technology-product matrix. For semiconductors, they extract potential products from the trademark textual information and match these with patent titles. Their approach reveals potential diversification paths based on the firm's current state of technological knowledge.

Overall, there are first approaches to analysing textual data of trademarks and their combination with patents. However, there needs to be a more detailed analysis of trademark texts in the context of innovation measurement to understand the contribution of trademarks. In the joint analysis with patents, the mentioned studies' focus is on the market application of innovations covered in patents and not on which innovations trademarks contribute. In addition, the combinations are currently only based on individual words. Using larger text parts could lead to the capture of new subjects. To better understand the development of technologies and innovation, a higher level of detail is of interest.

Based on the literature review, in total, four main underlying conclusions concerning trademarks and patents can be drawn from the literature:

1. Trademarks capture innovation.
2. Trademarks have, in comparison to patents, their strengths in market and product-related areas.
3. Services are covered more in trademarks than in patents.
4. Patents cover technical innovations and do this more than trademarks.

These conclusions are mostly based on non-textual data. However, textual data may not contain any further information on these conclusions, and knowledge cannot be gained on this based on text. Therefore, this chapter discusses whether the textual combination of patents and trademarks contributes to innovation research. Prior to this question, however, is whether patents and trademarks can be combined at all on a detailed level. Therefore, the following two research questions arise:

RQ4.1: Can trademarks and patents be combined via their textual data on a detailed level?

RQ4.2: Does the textual combination of trademarks and patents add new insights to the discussion of innovation?

The insights contribute to a better understanding of textual data of trademarks and patents. They are mainly interesting for analysis where trademarks could bring additional insights due to low patent coverage.

4.3 Methodology and Data

The research questions are examined in this chapter based on the textual data of patents and trademarks. The textual data of patents and trademarks are combined via Structural Topic Modelling (STM) of Roberts et al. (2019), which is an approach to extract topics out of textual data. The topics are analysed regarding patents and trademarks, plausibility and their coverage of the assumptions described in the literature overview (see Section 4.2). The textual data is extracted for two research areas, Robotics and Footwear, to derive differences between different application areas. The methodology applied in this chapter is described in Subsection 4.3.1. The application areas of Robotics and Footwear are then shortly described in Subsection 4.3.2, before explaining in detail the data extraction in Subsection 4.3.3 and model selection Subsection 4.3.4.

4.3.1 Methodology

To answer the research questions, it is necessary first to combine the data of the two sources on a detailed level and then analyse the results in terms of plausibility and consistency, and with regard to the assumptions drawn from the literature. Thus, topics that occur in trademarks and patents should be derived. Structural Topic Modelling (STM) of Roberts et al. (2019) combines the data sources and derives these topics. It combines the different data sources on a topic level, meaning the model estimates topics in the sample and assigns the documents to these topics. The topics are used to assess the plausibility of the combination of patents and trademarks. They further serve as a basis for analysing innovation in trademarks and the differences between patents and trademarks in services, market, technology and product aspects.

Structural Topic Modelling STM of Roberts et al. (2019) is a machine learning model used to estimate topics in documents under the consideration of additional information. The topics are not directly observable in the textual documents. They are considered hidden and therefore need to be estimated. Each document is a collection of words that belong to several topics. The semantic structure is not taken into account. The words are observable and can be used for the estimation. The general assumption is that every document consists of various topics. For example, a Robotics document could be about surgery robotics, covering robotic components (topic 1) and surgery (topic 2). For the model to identify these two topics in the document, it is necessary to estimate them.

Each topic is represented by a collection of words that occurs in the documents and is observable. In the robotics example, the document could contain, for example, the words “axis”, “arm”, “rotation”, “tissue”, “skin”, “blood”, and “endoscopy”, among others. The model would then start with estimating the topics and then assign the words to the topics. Here, topic 1 robotic component would cover “axis”, “arm”, and “rotation”, and topic 2 surgery “tissue”, “skin”, “blood”, and “endoscopy”. For the topic estimation, it is important to note that in STM, topics are considered correlated and not independent. This implies that if three topics exist, like robotic components, surgery and shoes, the model would consider the co-occurrence of robotic components and surgery more likely than the co-occurrence of robotic components and shoes.

Out of the topic estimations and the word assignment, the model assigns words and topics to the documents and compares the assigned words with the observable words in the documents provided as input. In STM, the documents are thus considered a result of a generative process with estimated topics and assigned words (Blei 2012). A difficulty in topic modelling is that the model requires the total number of topics K occurring in the data set as input. The model cannot estimate that number by itself. It uses that information to estimate the topics, the word assignment and the document distribution. However, the number of topics is dependent on the data set. Therefore, different approaches exist to determine the number of topics to be considered. As a good number of K is not known in advance, several models with varying K need to be calculated (Roberts et al. 2014; Roberts et al. 2019). Standard metrics to evaluate the models are:

Residuals provide insights into how well the topics represent the texts based on the distance between the estimated document-word occurrence and the observed document-word occurrence. A low value indicates a good approximation of the documents (Roberts et al. 2019).

Exclusivity indicates whether the words describing the topic are unique to the topic or are commonly found in other topics (Roberts et al. 2019).

Semantic Coherence addresses the internal consistency of a topic. It is high when the words in a topic frequently co-occur (Roberts et al. 2019).

Held-out Likelihood provides information on the model’s predictive ability, meaning how well the model predicts previously removed words from the documents (Roberts et al. 2019).

A low value of residuals is desirable while maximising the exclusivity and coherence of the resulting topics. Exclusivity and coherence are opposing metrics that must be balanced against each other (Roberts et al. 2019). As input, STM not only uses the document information and the number of topics but can also handle additional information provided to the model. This additional information can have two effects:

Topic prevalence affects the topic occurrence in the document and allows for differences between different document types (Roberts et al. 2019). This means that, for example, the information about the IPR type of the document, here patents or trademarks, allows the model to assign a higher probability of occurrence of a robotic component topic to patents than to trademarks.

Topical content affects the word assignment in the topics per document type. It allows for different words to be used for the same topic depending on the document type (Roberts et al. 2019). This means that patents could use different words than trademarks to describe the topic of robotic components.

Modification of Structural Topic Modelling The study intends to assess the textual combination of trademarks and patents (see RQ4.1) and to make a topic-wise comparison of patents and trademarks concerning their market, product, service and technical coverage (see RQ4.2). STM makes it possible to address both research questions. With the use of topic prevalence, STM enables the combination of patents and trademarks while considering their differences. Therefore, the additional information of IPR type (*type*, "patent" or "trademark") and application year (*year*, e.g. 1980, 2000, 2010) is handed over to the model as $type \cdot s(year)$. The intuition behind this is the following:

IPR Type: The document type, as a topic prevalence covariate, allows for the difference between patents and trademarks. This is important as the text structure between patents and trademarks is different, and the topics covered might occur to varying shares depending on the document type. Patent documents and thus the textual description centre around one specific invention. A patent requires an inventive step (EPO 2018). Different knowledge fields might come together to create the invention, but the focus is still on one invention. Trademarks enhance differentiation and identification and are applicable in services (Millot 2009). Trademark documents, thus, centre around the application areas of the trademark. A trademark document can centre around one specific application area or can list several areas of application, depending on the scope of the trademark use. These application areas can be widely dispersed and not focused.

Year: The application year is also considered as a topic prevalence as the timing of patent and trademark applications differs. This is in line with Blei and Lafferty (2006) who demonstrated in their dynamic topic modelling that time influences the topics of "Science" publications. In this chapter, the application year is provided to the model as a spline ($s(year)$), meaning that the topic can occur in varying proportions for different years. Linearity is rejected, as, over a long period, a topic proportion will not have the same growing or falling tendencies. Considering the application year in the model allows for changing topic probabilities over time. This means that a topic with a high probability in the 1980s can have a different probability in the 2000s. In the context of Robotics, for example, this allows surgery robotics to evolve.

Finally, an interaction between the application year and the document type is assumed ($type \cdot s(year)$) as topics represented in trademarks are not necessarily simultaneously present in patents and vice versa. By allowing for the application year and document type interaction, changes in the topics covered in the different intellectual property rights can be studied.

Application of Structural Topic Modelling The approach used in this chapter from the data extraction until the model selection is displayed in Figure 4.1. The data extraction and model selection are performed separately for Robotics and Footwear application areas.

1. Data Extraction: As a first step, the data needs to be extracted from the patent and trademark data sources. The data is identified based on relevant classifications and keywords in patents and trademarks. The data extraction is described in Subsection 4.3.3.

2. Pre-processing: In the next step, the documents are cleaned and prepared for the structural topic modelling. The structural topic modelling requires a document-term matrix as input, with the documents on the rows, the unique words on the columns and the occurrence of words in the

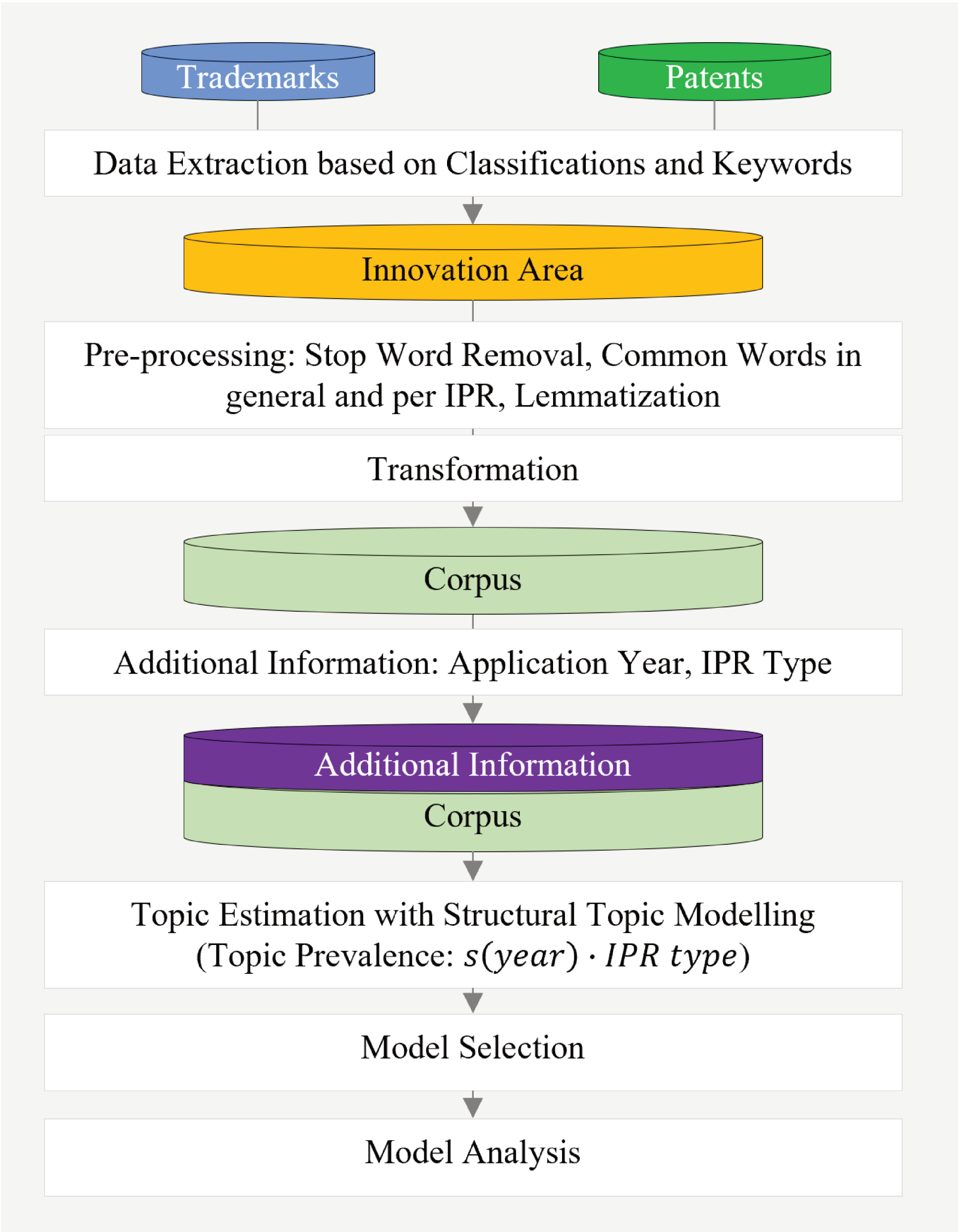


Figure 4.1: Approach Overview in Robotics and Footwear.
The figure illustrates the approach used in this chapter from data extraction to model selection. Data extraction relies on trademarks and patents related to the innovation areas, here Robotics and Footwear. This data is subsequently pre-processed and transformed to create the corpus for textual data analysis. Additional information is added, forming the basis for topic estimation, which is carried out using structural topic modelling. Among various topic estimations, a model with a specific number of topics is selected and employed for the analysis in this chapter.

document as values. The dimensions of the document-term matrix influence the complexity of the model, meaning more words and documents increase the matrix. To reduce the complexity, non-meaningful words are excluded. First, lemmatising is applied, which returns words to their original form (e.g., “was” to “be”). Afterwards, English stopwords[1], numbers, single-digit letters, words that occur only in one document, words that occur less than five times, and words that occur in more than 50% of patents or trademark documents are removed. Only unigrams are considered. If a document does not contain any words after the cleansing process, this document is removed from the data set. This may occur when a document consists only of common words. From the reduced dictionary, a document-term matrix is created. This means that each word in the cleaned data set is counted for each considered document. The textual data and the additional information per document (IPR type, application year) are then provided to the model.

3. STM is then calculated for various number of topics to perform the model selection (further details are described in Subsection 4.3.4).

The results are shown in Figure 4.2. For each innovation area, here Robotics and Footwear, a topic overview is derived. The topic overview displays all estimated topics that occur in the data set. The topics are then analysed in further detail to be able to answer the research question. The topic occurrence displays the differences in the occurrence probability of a topic per IPR type and over time. This is used to understand if there is a difference concerning the topic for different IPRs. The topic classification then displays the main classifications of patent and trademark documents. Patent classes are expected to be more specific than trademark classes, as the trademark system is less refined.

The documents on the topic are used to compare examples of patents and trademarks to the topic to understand if the topic is consistent, meaning if patent and trademark documents align, and to see what information can be gained from the documents. This analysis is repeated for several topics to draw general conclusions. The topics are selected based on their IPR occurrence probability in patents, trademarks or neither of those. The analysis is performed for Robotics and Footwear separately in Section 4.4 for Robotics and in Section 4.5 for Footwear. The results are then jointly discussed in Section 4.6 to draw general conclusions and answer the research questions.

4.3.2 Innovation Areas

Millot (2009) points out a high correlation between trademarks and innovation, especially for knowledge-intensive services and in the high-tech sector. Millot (2009) finds further that industries with low patenting activities are high on trademarks such as clothing and footwear or printing and publishing. Therefore in this thesis, a high-technology area and a low-technology area are chosen for the analysis of innovation areas: Robotics and Footwear. Representatives from low- and high-technology areas enable the discovery of general application patterns in patents and trademarks, irrespective of the underlying degree of high- or low-technology.

High-Technology Area of Robotics Robotics is an innovation area where multiple fields and disciplines are involved in its development. It is considered a technology of great future importance. Major regions such as the United States of America, the European Union, Japan, and China have robotic development and application plans. Application areas are, e.g., manufacturing, medicine, logistics, military, entertainment, or education (T.-M. Wang et al. 2018). According to the International Federation of Robotics (IFR), robots are used for industrial automation, delivering services to humans, or performing tasks on equipment. Especially the application of service robotics like, e.g., lawn mowing or vacuum cleaning robots has increased in recent years (IFR 2019a; IFR 2019b). Robotics has evolved to enable

[1] The stopword list is taken from stop_words from the R package tidytext. The list combines the sources SMART, snowball and onix.

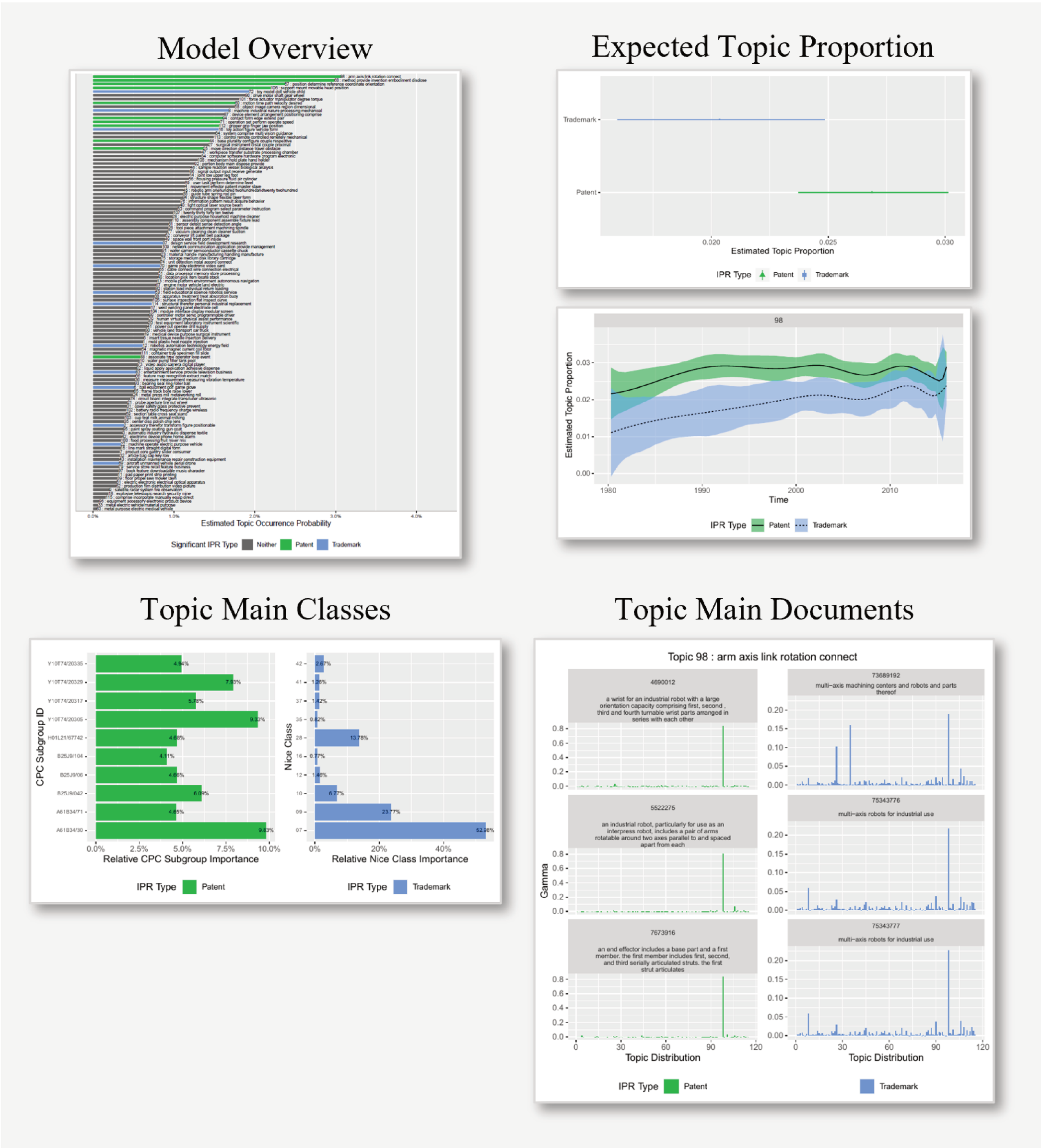


Figure 4.2: Illustrative Analyses in Robotics and Footwear.
The figure offers an overview of the various analyses in this chapter. The "Model Overview" presents all the topics of the model and indicates whether there is a significant difference between patents and trademarks (for more details, see also Figure 4.11) The "Expected Topic Proportion" is analysed for each topic, both in general and over time (see also Figure 4.8). "Topic Main Classes" provides a reference for the topic to classifications of patents and trademarks (see also Figure 4.9). Finally, "Topic Main Documents" examines documents where the topic is highly present to assess the consistency of the topic and the focus of the documents on that particular topic (see also Figure 4.10).

applications in unstructured, unknown environments with human-machine interaction compared to previously structured, known environments with explicit separation from humans. This advancement has been achieved by integrating artificial intelligence, which allows the technology to be adapted to environmental requirements and thus enables new applications (Sorbe et al. 2018). Robotic technology is one of several key technologies for industry 4.0. Robots will not only be used for repetitive work but will increasingly interact with humans. Manufacturing will change, enabling new outcomes, mass customisation and new business models (Kamarul Bahrin et al. 2016). The economic activities in the fields of Robotics can be classified according to NACE Rev. 2 [2]. The main activities are taking place in medium-high technology areas like equipment manufacturing. Service activities also exist but focus on wholesale, retail and repair (see Table 4.1). Overall, Robotics serves as a representative of a high-technology area.

Low-Technology Area of Footwear The activities in Footwear mainly focus on medium-low to low technology applications or less knowledge-intensive services (see Table 4.1). The activities in Footwear that relate to machinery and more complex elements of Footwear, such as heels or soles, are considered medium-low to medium-high technology activities. However, most activities related to producing and repairing Footwear are considered low in technology or knowledge-intensity. The high product variety of shapes and materials makes the production process complex and challenging to automate (Ronthaler et al. 2011; Raffaeli and Germani 2011). According to (OECD 2011), the footwear industry belongs to low-technology industries as their R&D intensity is generally low.

4.3.3 Data Extraction

For the analysis of Robotics and Footwear, the textual descriptions of trademarks' "goods and service" field and patents' abstracts are used. The trademarks and patent applications are taken from the United States Intellectual Property Office (USPTO). The data sets are limited to the years 1978 to 2016.[3]

The criteria for data extraction in the case of **Robotics** are the following:

- *For Patents:* The registered USPTO patents are identified and extracted based on (1.) robotic CPC classes[4], supplemented with (2.) a keyword-based search (%robot%) in the patents' title and abstract (see Query C.1 and Query C.2).

    1. A CPC-based search is applied, as experts assign patents to these classes based on their affiliation to, for example, Robotics. For the data extraction, the CPC classes are identified based on the keyword (%robot%) and each class's description. Only CPC classes that focus primarily on robotics are included, while classes that use robotics only as an example are not explicitly included. In total, 883 different subgroups are specified (see Table C.1 and Table C.2). Relevant subgroups are, e.g., "A47L2201/00" (Robotic cleaning machines, i.e. with automatic control of the travelling movement or the cleaning operation), "B25J19/0029" (Accessories fitted to manipulators, e.g., for monitoring, for viewing; Safety devices combined with or specially adapted for use in connection with manipulators -Means for supplying energy to the end effector arranged within the different robot elements) or "G05B2219/2661" (Program-control systems-Pc systems-Pc applications-Milking robot).

    2. The search is supplemented by a keyword-based search so that robotics patents not included in the explicitly listed classes are also covered. The keyword-based search also increases the comparability of patents and trademarks since, in the case of robotics trademarks, only

[2] "Nomenclature statistique des activités économiques dans la Communauté européenne" (Eurostat 2008)

[3] This is due to data availability. Trademark data coverage is limited before 1978 according to S. J. H. Graham et al. (2013). After 2016, the USPTO patent data coverage was not complete at the point of extraction.

[4] CPC classes are used for the classification of patents (see also Subsection 2.2.1).

Table 4.1: Economic Activities in the Innovation Areas of Robotics and Footwear

Technology classification	Robotics		Footwear	
	NACE Class	Description	NACE Class	Description
Manufacturing				
Medium-high technology	28.22	Manufacture of lifting and handling equipment; including mechanical manipulators and industrial robots specifically designed for lifting, handling, loading or unloading	28.94	Manufacture of machinery for textile, apparel and leather production; including machinery for making or repairing footwear or other articles of hides, skins, leather or fur skins
	28.99	Manufacture of other special-purpose machinery n.e.c.; including Manufacture of industrial robots for multiple tasks for special purposes		
Medium-low technology			22.19	Manufacture of rubber boot and shoe heels and soles and other rubber footwear parts
			22.29	Manufacture of plastic footwear parts
Low technology			14.19	Manufacture of other wearing apparel and accessories; including manufacture of footwear of textile material without applied soles
			15.20	Manufacture of footwear
			16.29	Manufacture of wooden shoe parts
			32.30	Manufacture of ski-boots
			32.50	Manufacture of orthopaedic shoes
Services				
Less knowledge-intensive (market) services	46.69	Wholesale of other machinery and equipment; including wholesale of production-line robots	46.16	Agents involved in the sale of textiles, clothing, fur, footwear and leather goods
			46.42	Wholesale of clothing and footwear
			46.49	Wholesale of other household goods; including wholesale of sport goods, including special sports footwear such as ski boots
			47.72	Retail sale of footwear and leather goods in specialised stores
			47.82	Retail sale via stalls and markets of textiles, clothing and footwear
			77.29	Renting and leasing of other personal and household goods; including textiles, wearing apparel and footwear
			95.23	Repair of footwear and leather goods

The table offers an overview of economic activities in Robotics and Footwear. It includes technology classifications, NACE classes, and descriptions of the related activities. On the left, you can find the technology classifications, which are then described in the context of Robotics and Footwear. The technology classifications are further categorized into manufacturing and services. *Source: Own representation based on Eurostat (2008) and Eurostat (2021b)*

Table 4.2: Overview of the Data Sets of Robotics and Footwear

Data Set	Robotics		Footwear	
	Original	**Cleaned**	**Original**	**Cleaned**
Patent Documents	26,650	26,650	14,253	14,253
Trademark Documents	10,352	10,349	87,391	79,811
Total Documents	37,002	36,999	101,644	94,064
Terms	33,467	12,104	31,499	12,164
Selected Number of Topics		**115**		**89**

The table offers an overview of the documents and terms used as the foundation for structural topic modelling, along with the selected number of topics based on various metrics. It displays the initial number of documents extracted from various data sources, as well as the document numbers provided to the model after data preprocessing and cleaning. The time-frame is limited to the years 1978 through 2016.

this search is available. Nevertheless, the CPC-based search is used in patents to include the expert perspective on robotics patenting.

- *For Trademarks:* The registered USPTO trademarks are identified based on a keyword-based search (%robot%) in (1.) the trademark name (or pseudo name[5]) or in (2.) the "goods and service" field description (see Query C.3, Query C.4 and Query C.5).

  1. Trademarks cover specific goods or services. The names associated should not lead to confusion or wrong assumptions. Trademarks with "robot" in the trademark's name create the association that the trademark is applied in robotics and should thus cover robots in a broader sense. Subsequently, all "goods and service" fields of the trademark are related to robotics, even though "robot" might not be part of the descriptions. Therefore, if a trademark's name contains "robot", all of its "goods and service" fields are included in the analysis.

     An example is "ROBOTOL" which stands for robot-related ear surgery (Serial no: 79207597). The trademark directly creates the association that this brand is for robotic products. Therefore, all available descriptions are used for the analyses.

  2. The search is supplemented by a search in the "goods and services" field of the trademarks. If the name of a trademark does not designate robotics, there may still be coverage of a robotics application. The robotic application must be explicitly described. If a mark covers multiple NICE classes, only the classes related to "robot" are considered in the analysis. For example: "TIM" is a trademark registered for an underwater robot (serial number: 73329434). The name does not indicate robot affiliation, but the NICE class description explicitly mentions robotics.

In total, 37,002 Robotics documents are extracted from the data sources. Thereof, 36,999 documents remain after the cleaning process (as described in Subsection 4.3.1). The number of terms, here unique words in the data set, is reduced from 33,467 to 12,104 (see Table 4.2).

The data extraction for Footwear is analogous. For patents, CPC classes concerning shoes and footwear are assessed (see Query C.6 and Query C.7). 619 different CPC subgroups are included (see also Table C.3), among which are "A43B1/00" (Footwear characterised by the material), "A43B19/00" (Shoe-shaped inserts; Inserts covering the instep) or "A43D69/00" (Shoe-nailing machines). Footwear for animals is excluded. Further, patents with combination of "%foot%" and "%shoe%" or with "%footwear%" in abstract or title are included. The combination of foot and shoe is necessary as the word "shoe" is not

[5] The USPTO gives pseudo names to improve the searchability of the data sets for specific keywords. For example, a name with "bot" might be amended to "robot" to include the trademark in the results of "robot" searches.

precise. It also occurs in a context such as brake shoes. For Footwear trademarks, trademarks are selected based on their (1.) trademark name, (2.) pseudo name or (3.) "goods and service" field description (see Query C.8, Query C.9 and Query C.10).

The final data extraction consists of 101,644 documents, of which 94,064 remain after the cleaning process and serve as a basis for the model selection.

4.3.4 Model Selection

Based on the data extraction, the models are selected for the analysis for each data set separately. The final selected number of topics K for Robotics and Footwear can be seen in Table 4.2.

Model Selection in Robotics Out of the 36,999 robotic documents and 12,104 terms, a document-term matrix is created as input for the structural topic model. Various models are calculated as the number of topics needs to be decided upon before the model calculation. The best-fitting number of topics is then chosen based on Figure 4.3. The figure provides an overview of model diagnostics related to exclusivity, held-out likelihood, residuals, and semantic coherence. These metrics are calculated for every model and allow for the comparison of the models.

The residuals form an inverted U-shape as shown in Figure 4.3a. A model at the bottom of the U-shape is preferred. In the case of Robotics, residuals are lowest between the model of 100 and 130 topics. In Figure 4.3b, the close-up of the residuals reveals that the models with 114 topics and 122 topics have the lowest residuals among all models. Therefore these are included in further analysis. Besides topics 114 and 122, other topics are considered for the final selection. These are chosen based on their semantic coherence and exclusivity. Semantic coherence determines the relatedness of the terms within a topic. With an increase in number of topics, the coherence declines. In parallel, exclusivity increases, which indicates how unique the terms are for the topic. A trade-off between semantic coherence and exclusivity exists. In general, models with fewer topics have a higher semantic coherence but lower exclusivity (Roberts et al. 2014). Figure 4.4a shows the mean exclusivity-coherence trade-off for all models between 100 and 130. The mean exclusivity-coherence trade-off for a model is calculated based on the mean of each topic's exclusivity-coherence trade-off in the model. Models 106, 115, 121 and 128 provide an interesting mean exclusivity-coherence trade-off to consider further. Overall, six models are looked at in closer detail. Figure 4.4b displays exclusivity-semantic coherence-trade-off per topic for each model. The model with 121 topics has, for example, many topics with a good trade-off (top right corner) but also many low exclusivity topics. Based on the comparison, the model with 115 topics is chosen.[6]

Model Selection in Footwear In the case of Footwear, the 94,064 cleaned documents resulted in a model with 89 topics. The selected model displays a good trade-off between exclusivity and semantic coherence. The metrics to approximate the number of topics K in Footwear are provided in the Appendix, Subsection C.2.2.

4.4 Results of the Robotic Analysis

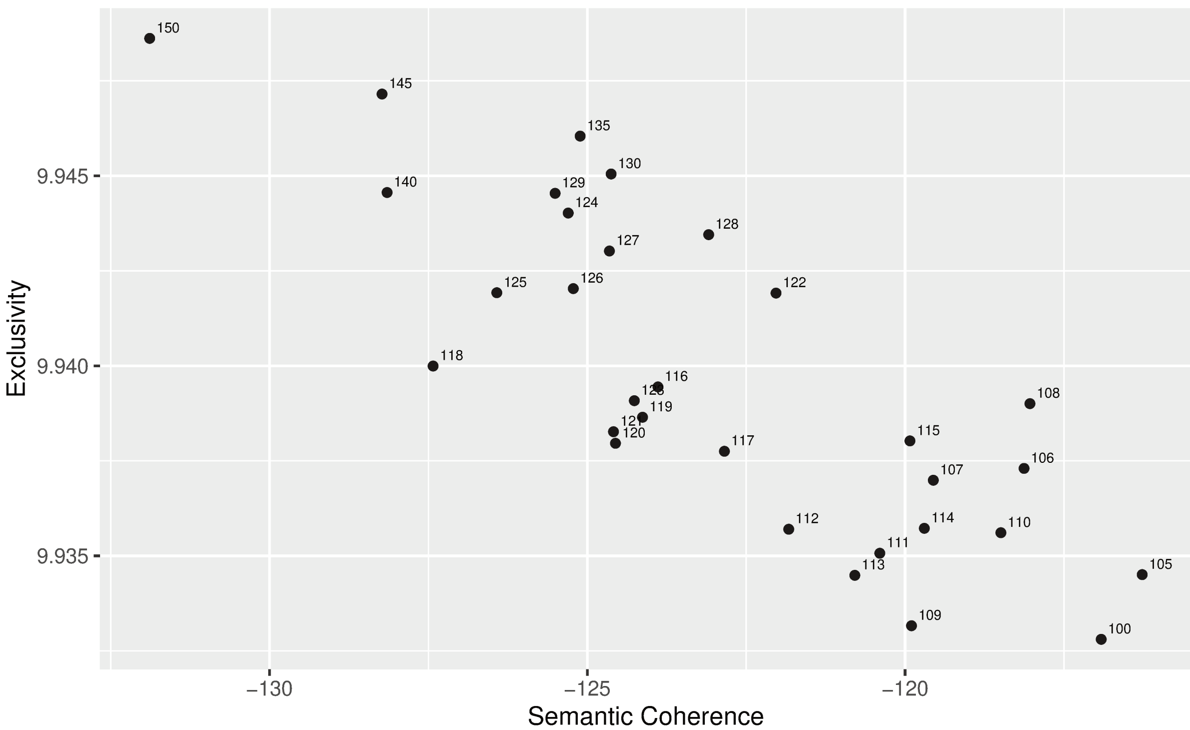
The results are presented for Robotics first in Section 4.4, before briefly discussing Footwear in Section 4.5. The results of the two areas are then compared and discussed in Section 4.6.

Before analysing the estimated topics from the field of robotics, the data set is described in terms of the annual applications and the associated classifications to provide a general understanding of the data

[6] The exclusivity-coherence trade-off of model 115 is displayed in the Appendix, Figure C.1.

Figure 4.3: Metrics to Evaluate the Number of Topics K in Robotics.
The general diagnostics offer an initial overview to determine the appropriate number of topics (K). For each calculated model, metrics such as exclusivity, held-out likelihood, residuals, and semantic coherence are assessed. Each point on the graph represents a distinct model and its corresponding diagnostics. These metrics are evaluated in relation to one another to decide on the optimal K. Specifically, in the case of residuals, models located at the lowest point of the slope are of particular interest. The figure presents both (a) an overall view and (b) a closer, zoomed-in perspective to facilitate the selection of models that warrant closer examination.

(a) Mean Exclusivity-Semantic Coherence Trade-off of Different Models.



(b) Exclusivity-Semantic Coherence Trade-off of the Topics of Selected Models.

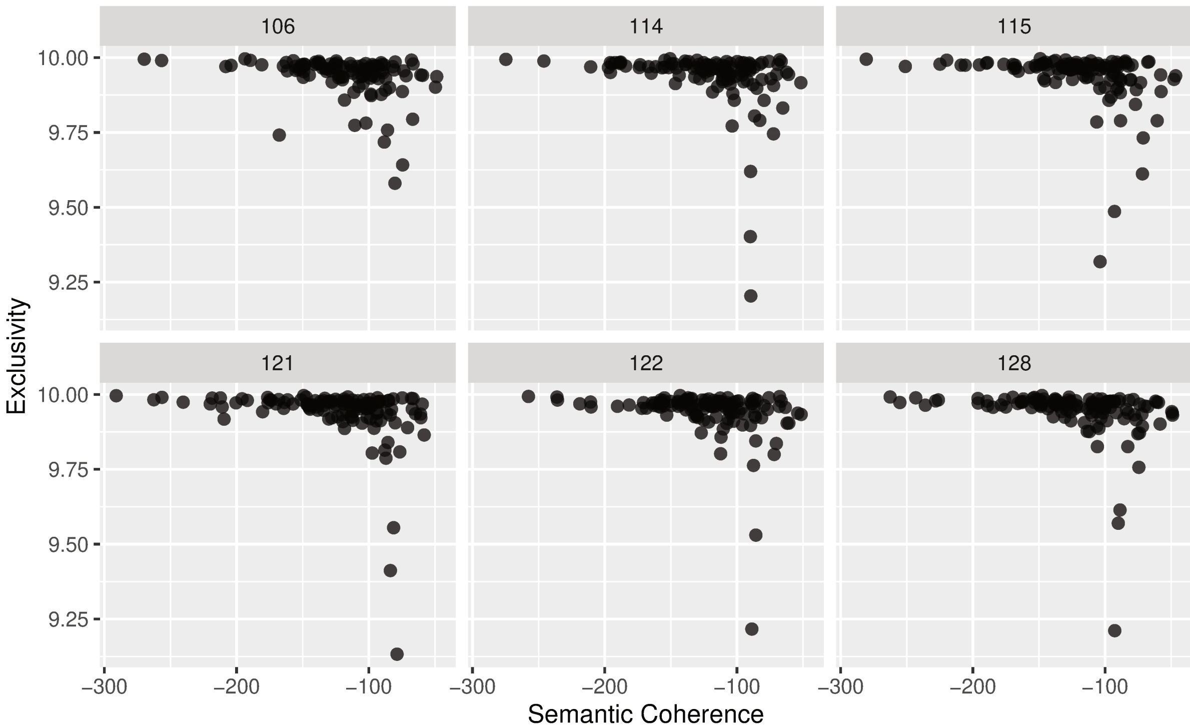


Figure 4.4: Exclusivity and Semantic Coherence of Robotics Models in Comparison.
A trade-off between exclusivity and semantic coherence needs to be determined. Models with fewer topics exhibit higher semantic coherence but lower exclusivity compared to models with more topics. Subfigure (a) presents the mean exclusivity against the mean semantic coherence for each model, with each point representing a model and its respective trade-off. Potential models are selected for further examination. Subfigure (b) illustrates the exclusivity-semantic coherence trade-off for these chosen models on a per-topic basis. In Subfigure (b), each frame corresponds to a model, and each point represents the trade-off for a specific topic within the model.

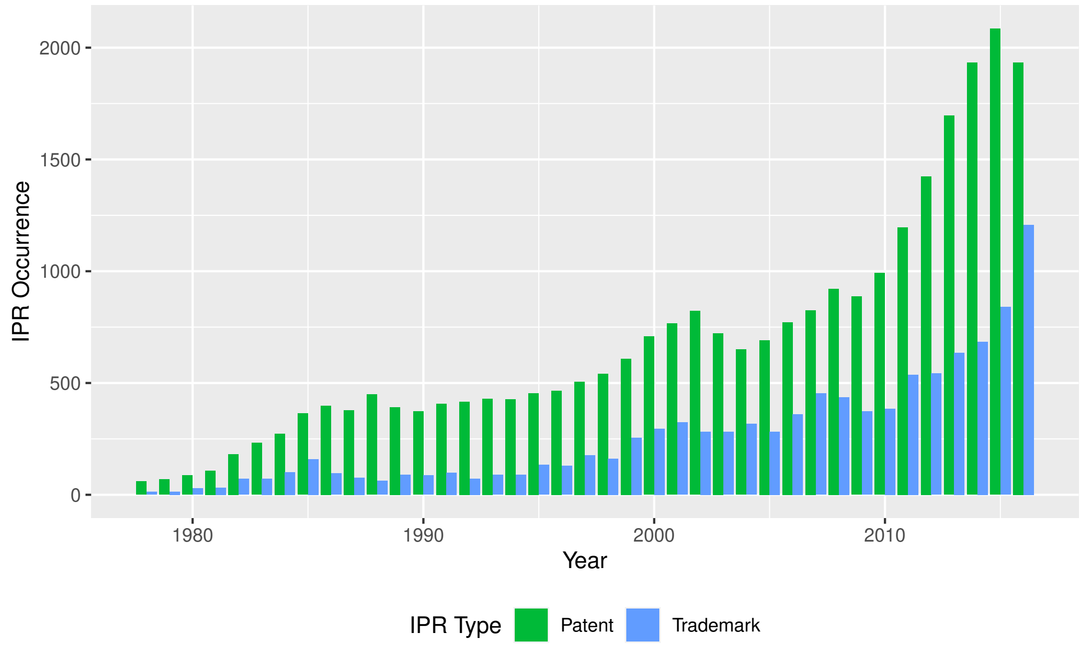


Figure 4.5: Robotic Patent and Trademark Registrations per Year.
The figure illustrates the yearly filings of registered trademarks and patents in the Robotics data set, with years referenced to the application filing year in trademark or patent application.
Source: Own representation based on the Robotic data set.

set. The data description refers to the final data set of 36,999 cleaned documents. In total and per year, more patent documents are related to Robotics than trademarks. As Robotics is a high-technology area, patents can be widely applied. The general trends in patents and trademarks are similar: overall, the number of documents is low in the 1980s but has increased since. Noticeable is a peak around 2002 and a decline in 2009 (see Figure 4.5), which might relate to general economic cycles. According to S. J. H. Graham et al. (2013), a general peak of trademark applications occurred during the dot-com boom around 1999-2000 and a decline is visible in 2009 related to the global financial crisis. The increase in trademarks is surprising, as Robotics is still a high-technology area. The analysis should provide further insights into the application areas for trademark protection in Robotics.

Patents and trademarks in Robotics are categorised in existing classification schemes:

- Figure 4.6 displays the classification of patent documents into CPC groups. A patent can be classified into one or more CPC groups. The patent occurrences in a CPC group are summarised to determine the importance of each CPC group within the data set. Each patent occurrence in a group is counted as “one” in this process. If a patent is contained in multiple CPC groups, then the patent is not split between multiple CPC groups but is counted fully in each group. The sum is divided by the total number of unique patent documents in the data set to determine the relative share of the CPC group. The CPC groups represented in Figure 4.6 are all groups with a share of more than 1%. In Robotics, the CPC groups B25J, G05B, Y10S, Y10T and A61B have the highest share of patents. CPC group B25J, with a share of 48.75%, inherits all patents about “Manipulators”. CPC group G05B (19.00%) stands for “Control or regulating systems in general”, Y10S (17.50%) and Y10T (17.27%) cover technical subjects of the former U.S. classification system like apparel, baths, beds or compound tools and A61B (13.44%) for “Diagnosis” and “Surgery” (CPC 2020).

- Figure 4.7 displays the trademark documents in the 45 NICE Classes in relative shares. Trademarks can be registered in multiple NICE classes simultaneously. Every NICE class occurrence is counted separately. Figure 4.7 represents the total share of robotic trademarks in all 45 NICE classes. NICE class 7 “Machines, machine tools, engines and vehicles” (35.22%), class 9 “Scientific, research,

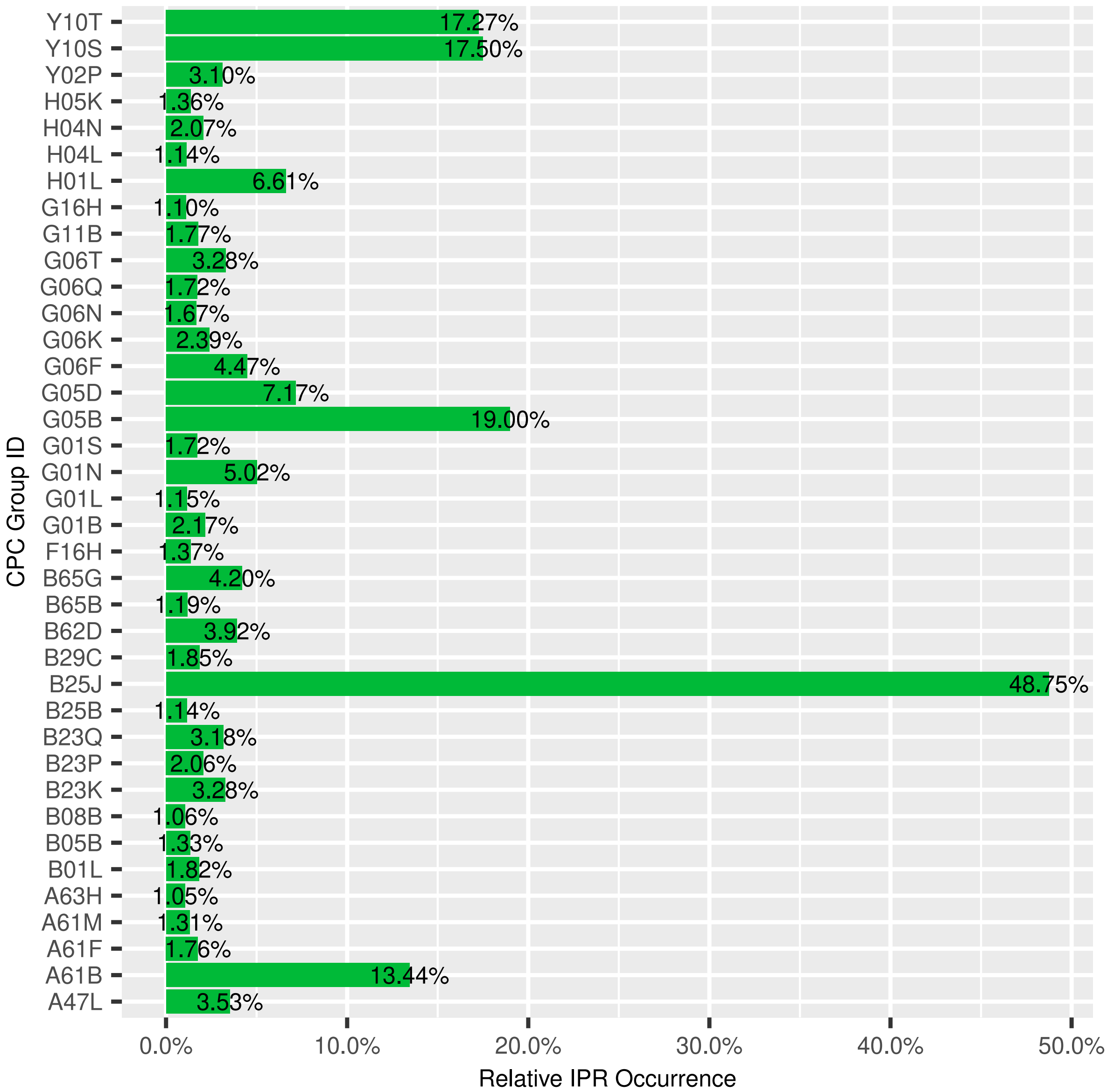


Figure 4.6: Robotic Patent Classification.
The figure presents the CPC groups of patent documents with a share exceeding 1%. The CPC groups are arranged in descending alphabetical order. The length of each bar reflects the relative occurrence, thus the occurrence of the CPC group in comparison to the total number of patent documents. It's important to note that since a patent can be assigned to more than one CPC group, the percentages of the groups do not add up to 100%.

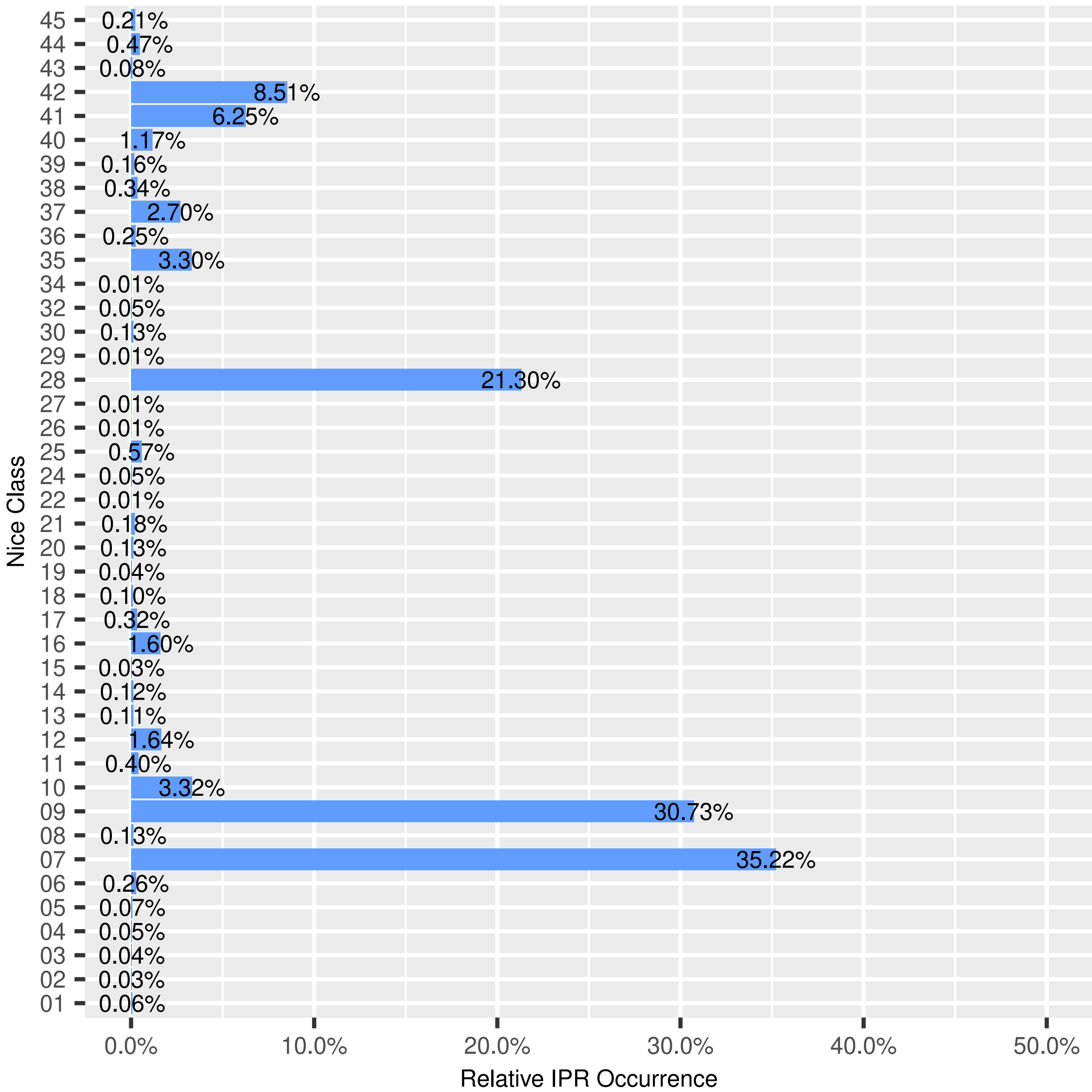


Figure 4.7: Robotic Trademark Classification.
The figure displays the NICE classes of the trademark documents. The NICE classes are ordered in descending order. The length of the bar indicates the relative occurrence, thus the occurrence of the NICE class compared to the total amount of trademark documents. As a document can be registered in more than one NICE class, the NICE classes do not sum up to 100%.

processing of sounds, images or data, computer software" (30.73%) and class "Games, toys, video game apparatus, sport articles" 28 (21.30%) are the main trademark classes. Also of relevance are the trademark service classes 42 (8.41%) with "Scientific and technological services" and 41 (6.25%) with "Education" (WIPO 2019a).

Patents and trademark classifications thus both concern machinery. Patents additionally cover the subject of surgery, which belongs to the area of service robotics. The trademark classes also cover games, software and services. These subjects should be reflected in the topic estimation and result in topics around, e.g., games, services, and medicine.

4.4.1 Exemplified Analysis of Topic 98

In the following, topic 98 is described to exemplify the analysis of different topics of the structural topic modelling before comparing topics in Robotics to the research questions. Topic 98 has the highest occurrence probability in the data set, with 3.06%. The topic's share is calculated by summarising the topic's occurrence probability in every document in the data set. The main words in topic 98 are "arm", "axis", "link", "rotation" and "connection", which are essential functions of robots. The topic has a high semantic coherence and exclusivity, thus a good exclusivity-semantic coherence trade-off. The topic should thus be consistent, and the words should provide a reasonable explanation for the topic (see also Appendix, Figure C.1). As described in Paragraph 4.3.1, the Topic Occurrence, Topic Classifications, and the Documents on the Topic are analysed for each topic. The results of the analysis are described in the following:

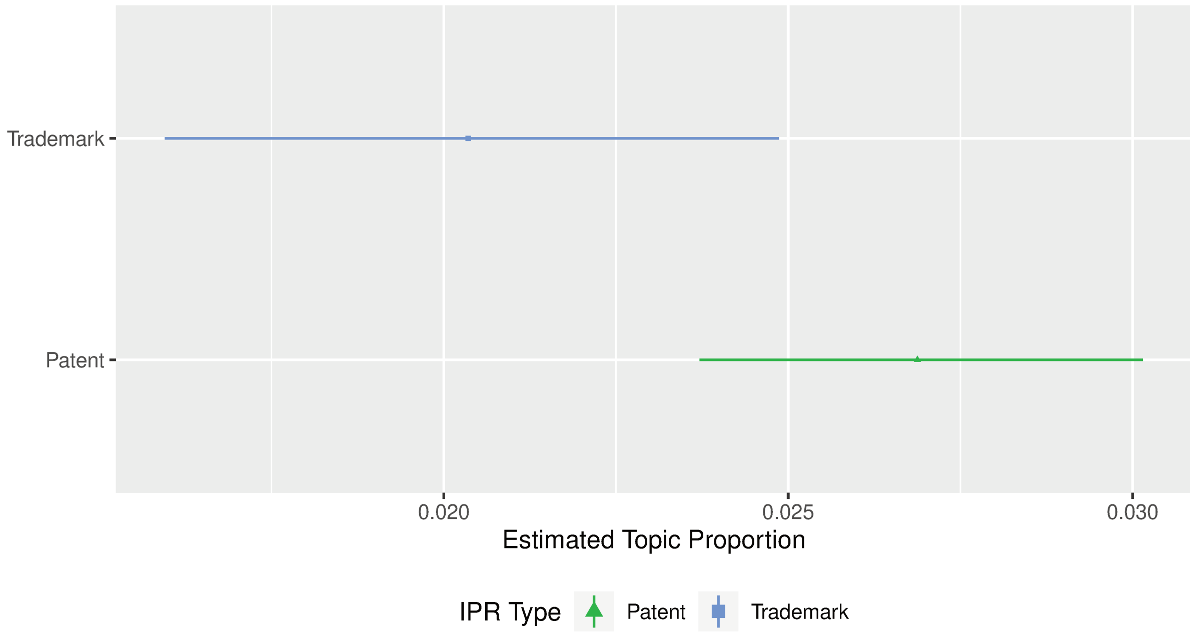
Topic Occurrence: Figure 4.8a shows the occurrence probability of topic 98 separately for trademarks and patents. The topic has a higher estimated topic proportion in patents than in trademarks. Even though the ranges of trademarks and patents overlap, the estimated mean is higher in patents. A t-test determines if the difference is significant. At the 95% confidence interval, the topic occurrence probability is significant in patents (see Appendix, Figure C.2). The topic is thus considered a patent topic.

Figure 4.8b displays the development of the topic over time, separated for patents and trademarks. As can be seen, the estimated topic proportion for patents and trademarks is converging towards each other. The topic's estimation and confidence intervals per IPR were separated in the 1980s but started to overlap around 2008. Interestingly, the topic share increases in trademarks from around 1% to 2.5%, while the share remains relatively stable in patents at around 3%. The increasing topic share in trademarks implies that the subject of robotic components has become more relevant in trademarks, which is reflected in the trademark descriptions and words. The reason could be that especially service robots are increasingly introduced to the market and thus are more present in trademarks than before.

Topic Classifications: The relation to existing classification systems is then analysed, as these provide additional information about the technologies behind them. Figure 4.9 displays the ten main patent CPC Subgroups and the trademark NICE classes in the topic. The main classes are calculated based on the topic occurrence in each document: Every topic occurs in every document, and every document is in different classes. The document's topic occurrence weights the classes of each document. These weights are summarised for each class to determine the ten main classes for patents and trademarks.

In trademarks, NICE classes are determined by the application area of the trademark. The NICE classes are generally broad, as there are only 45 classes. The trademark descriptions, however, can cover more than just the application area. Topic modelling focuses on this content. Therefore, it is expected that the class assignment of a topic and the topic description only partly align, depending on the topic. For topic 98 with robotic components, a high relation to NICE class 7 (machines and

(a) For All Documents considered in the model.



(b) For All Documents in the Respective Year.

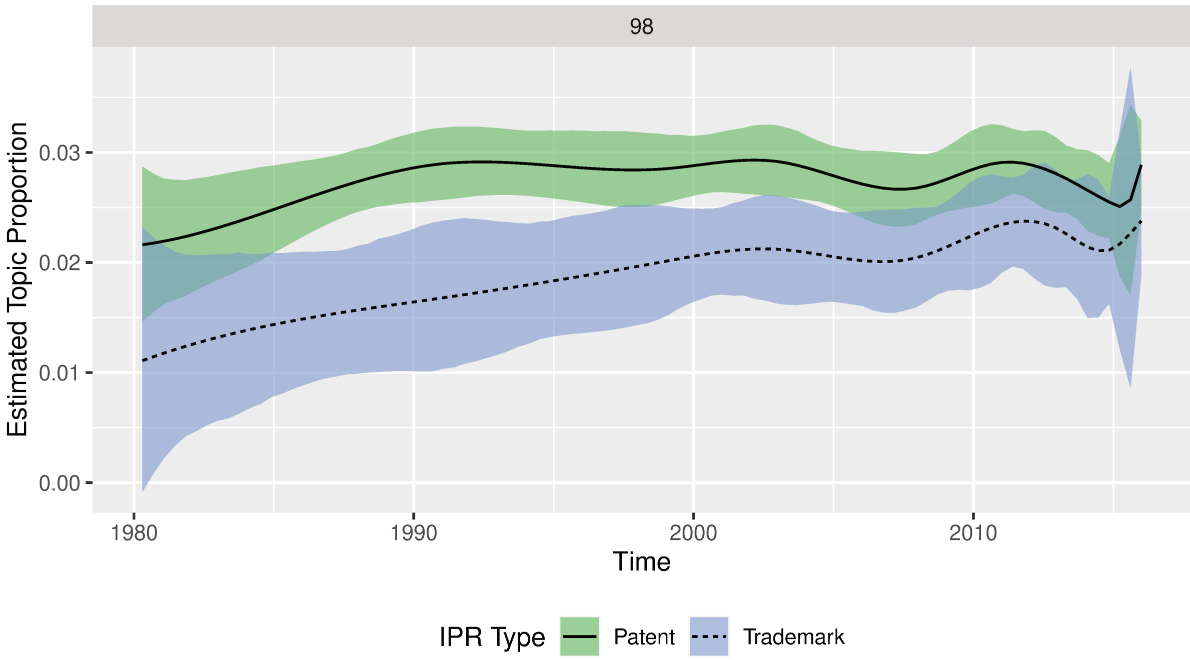


Figure 4.8: Expected Topic Proportion per IPR in Topic 98.
Subfigure (a) presents the occurrence probability of topic 98 across different document types. For topic 98, there is an overlap in the variances of trademarks and patents, but the estimated mean is higher in patents. A significance test, specifically a t-test, is conducted based on both variance and mean. The results are presented in Figure C.2. In the case of topic 98, the topic occurrence probability is found to be significant in patents. Subfigure (b) shows the evolution of the occurrence probability over time. The line represents the mean occurrence, while the coloured area signifies the variance.

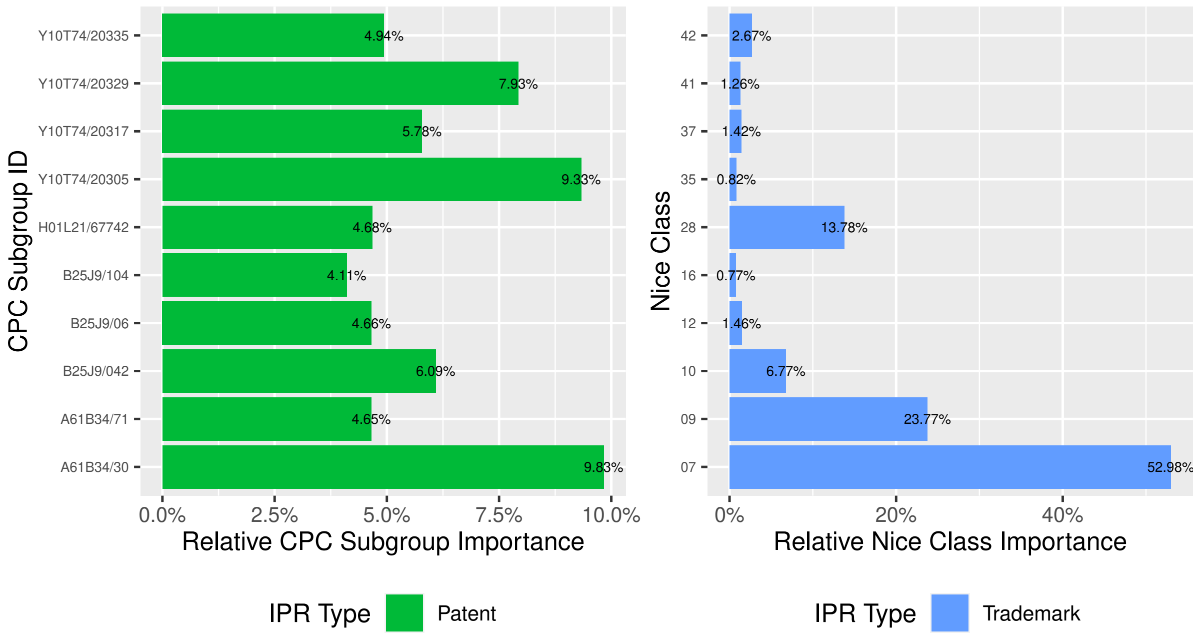


Figure 4.9: Ten Main Classes of Topic 98 in Patents and Trademarks.
The figures display the ten main CPC subgroups and NICE classes of patents and trademarks that are related to topic 98. The class names and shares are provided.

engines) is expected. However, other classes that focus more on application aspects of robotics are also likely. In trademarks, NICE class 7 (machines and engines) with 52.98%, NICE class 9 (23.77%, data and computer software) and 28 (13.78%, games and toys) are the three main trademark classes. NICE class 7 is an expected result. NICE class 9 displays the importance of data and software in the field, while games are a robotics application area. In contrast to the patent classifications, the trademark classes do not allow for further analysis. The level of detail provided by the class description is limited.

In comparison, the patent classification is more refined with related classes such as Y10T74 "Machine element or mechanisms" with Y10T74/20335 "Wrist", Y10T74/20329 "Joint between elements", Y10T74/20317 "Including electric motors", Y10T74/20305 "Robotic arm", and B25J9 "Programme-controlled manipulators" with B25J9/104 "Cables, chains or ribbons", B25J9/06 "Characterised by multi-articulated arms", B25J9/042 "Comprising an articulated arm". A61B34/30 includes surgical robots and A61B34/71 drive cable-operated manipulators. These are application areas of multi-axis robots or robotic arms. The application areas in patents differ from those in trademarks. The machine elements are covered by both, but the patent perspective does not cover the data or gaming aspects.

Regarding the topic, the classifications are in line with the topic description. Several aspects are important to notice: The comparison of the topic with the classifications underlines the result of the modelling. It also supports the interpretation of topics by merging them with existing metrics. Concerning trademarks, the combination enables the application of the patent classifications to trademarks and to capture overarching subjects of trademarks that the broad trademark classification cannot identify. In terms of content, the core aspects such as machines are covered by both trademarks and patents. Differences exist in the fields of application with surgery in patents and data and games in trademarks.

Documents on the Topic: Finally, the topic content is analysed in closer detail. The purpose is to analyse if the topic estimation and the combination of patents and trademarks are consistent and more detailed than the classification. Consistency is considered when the topic words, and the patent and trademark documents align. Further, the documents should be focused on the topic. Consistency implies that in the respective topic, the detailed combination of trademarks and patents based on their textual data is possible. This is necessary to answer the first research question. It is further assessed if the combination brings new insights. Therefore, the main documents of the topic are considered. As each document in the data set relates to the topic but with a varying share, the documents with the highest share differentiated per IPR type are extracted. The main trademark and patent documents in topic 98 are represented in Figure 4.10. On the top of Figure 4.10, the topic name and the five words with the highest frequency in the topic are displayed. The patent documents are displayed on the left, and the trademark documents are on the right. Every facet provides the document ID (patent ID for patents or serial number for trademarks) to relate the documents to the original document. [7] Beneath each document number, a snippet from the document is shown. This is an excerpt of the original textual description in the document. The document snippets combined with the highest probability words of the topic help to interpret the topic.

Beneath the document snippets, the gamma value for each topic in the document is shown. Gamma provides the association of the document to each topic of the model. The gamma values of all topics add up to 1. High gamma values in a few topics indicate that the textual description is focused, while high gamma values in many topics indicate that the textual description is broad. The document's gamma distribution of all topics helps to interpret the text snippets, especially when the text snippet captures multiple topics. The distribution then reveals if the model detected that dispersion. In patents, the gamma values are only high for topic 98 (around 80%). The patent documents are thus focused on and present in topic 98. The trademark documents are related to topic 98 with above 15%. Even though the documents also cover other topics, the text snippets and the topic's words are coherent.

From the observation, patent and trademark documents are about the same topic. The clustering of the model thus seems consistent and plausible. Overall, the textual descriptions reveal information on a subject relevant in Robotics. Compared to the classifications, the topic results are more refined than the trademark classifications. The patent classification is very detailed, so the topics do not add additional information in this case. The combination, however, reveals that the topic is covered in patents as well as in trademarks. Hereby, the patent text contributes to the invention aspects, while trademarks reveal the market introduction. Combined, an area of innovation is thus covered.

To summarise, topic 98 is significant in patents, but trademarks' relevance increases over time. The classifications and the textual description of the topic are aligned. The textual descriptions provide, however, more details than the classification, especially in the case of trademarks. The textual combination of trademarks and patents is further consistent. Patents still provide more details in their textual description than the trademark descriptions. Nevertheless, trademarks also cover information about the robots and their application area.

4.4.2 Robotic Analysis

In the following, different robotic topics are summarised and compared to answer the research questions. The Robotic model with 115 topics is selected. The topics are overviewed in Figure 4.11. The figure shows each topic's expected probability in the data set. Every bar represents a topic and its topic occurrence probability. The topic occurrence probabilities of the topics add up to 100%. Behind every bar, the topic's number and the five words with the highest occurrence probability within the topic are shown. For every topic, it can be determined if an IPR type is occurring with a significantly higher probability than the other. This is determined based on a t-test with a 95% confidence interval that tests

[7] The document ID can be used to search the document in the USPTO database.

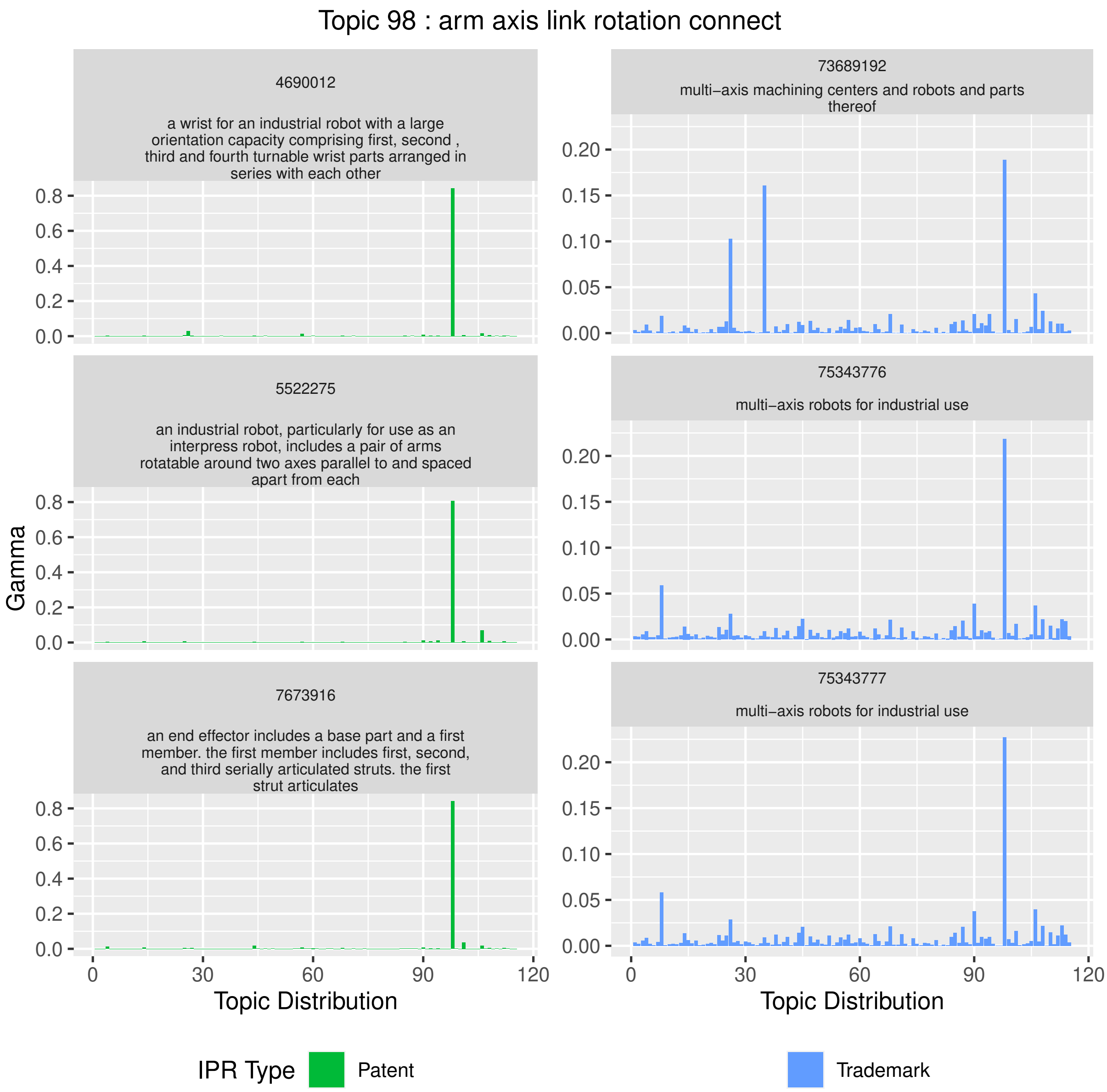


Figure 4.10: Main Patent and Trademark Documents in Topic 98 of Robotic Model 115. The figure displays different information about topic 98 of model 115. At the top, the five words with the highest occurrence probability in the topic are displayed. Further, a selection of the three main documents per document type is displayed. The registration number is given as well as a short snippet of the textual description. The bar graph underneath the header of each document displays the overall topic distribution of the document. The gamma value for each topic is shown, which provides the association of a topic to the document. Each value is determined by gamma. This helps in the interpretation of the data as it provides information on the level of focus on specific topics of the document.

whether the occurrence difference between patents or trademark documents in that topic is significant. The occurrence difference is calculated based on the topic occurrence probability separated for each data source as seen in Figure 4.8. The colouring of the bar in Figure 4.11 indicates the significant IPR type in the topic. The bar is coloured blue for significance in trademarks, green for patents, and grey if there is no significant difference.[8] As seen in Figure 4.11, most topics cannot be attributed to a specific IPR type (displayed in grey). Only some have significance in either patents or trademarks (coloured accordingly). Especially topics with a high frequency have a clear document significance. Of the 20 main topics, ten are significant in patents, while only three are significant in trademarks. Topics with a lower occurrence probability are, however, either significant in trademarks or occur in trademarks and patents without a significant difference.

In the following, different topics are analysed in closer detail and compared. Therefore, the main patent and trademark topics and the main topic without a significant difference in the occurrence probability between trademarks and patents are considered to understand which topics are covered by the different data sources. Further of interest are topics with a clear trademark association but low occurrence probability to understand that phenomenon. The general conclusion for each topic is described here.[9] The topics are organised according to their importance and significant IPR type.

Patent Topics: Out of 115 topics, 11 topics are significantly associated with patents. The first four topics are in patents. Further, ten topics out of the 20 main topics are in patents. The main associated words of the topics are e.g. “arm”, “axis”, “method”, “position”, “operation”, “speed”, “gripper”, “move” and “direction”. These words can be brought into the context of robotic components or core functionalities of robots like movement, gripping, and operation. Most of the patent topics are represented in patent class A61B “diagnosis” and “surgery” with 13.44% of the total patent documents. In trademarks, these topics are mostly in NICE class 7 or 9, covering machines or data and software aspects in general. Topic 98 is the most important topic in general, with an occurrence probability of 3.06% and is significant in patents. It is summarised in the following.[10]

Topic 98: Topic 98 is about arm rotation and the axis of robots. It is mainly associated with patent documents, which are focused on this topic and not highly present in other topics. The patent descriptions provide details about the development of these axes, while the trademarks only reveal that robots with multi-axis exist in the industrial context. The trademarks are less focused on the topic than the patents. The topic is present in patents and trademarks but significantly more common among patent documents.

In summary, the core of the patent topics is very technical. While the classification of trademarks only reveals a relation to machines in general, the patent classification provides detailed information on the technical background of the invention. This also helps to understand the trademark documents better.

Trademark Topics: 13 out of 115 topics are trademark topics. Only three trademark topics are among the 20 main topics. Trademark documents provide around one-fourth of the data to the data set. This might explain why the main topics are mostly significant in patents. The trademark documents are diverse in their related words: “toy”, “machine”, “research”, “game”, “educational”, “personal”, “automation”, “energy”, “entertainment”, “service”, “electric”, “vehicle”, and “unmanned” are among the main words. The topics can be related to the areas of toys and games (topics 72, 16 and 70; NICE cl. 28), education and entertainment (topics 53, 83; NICE cl. 41), research (topic 37; NICE cl. 42), and machines (topics 8, 114, 22, 69; NICE cl. 7). Topic 69 covers the area of unmanned aerial vehicles, which is an important application area of service robotics. Besides NICE class 9, it is also covered in Cl. 12 (vehicles) and 28 (e.g. sport equipment, amusement).

[8]The results of the t-test for each topic can be seen in detail in the Appendix, Figure C.2.
[9]Further details for each topic are provided in the Appendix, Subsection C.1.2.
[10]Topic 98 is used to exemplify the analysis. The details are provided in Subsection 4.4.1.

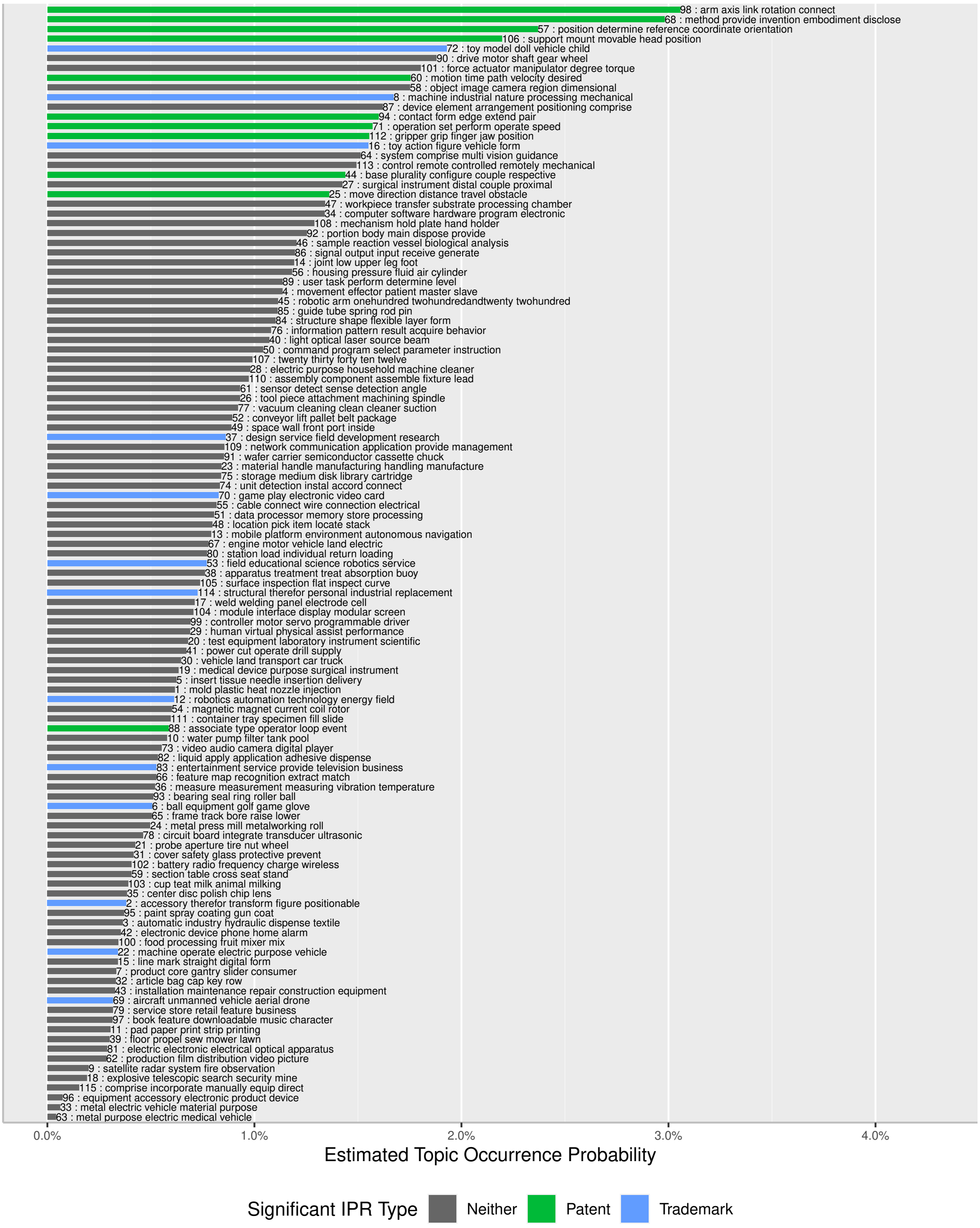


Figure 4.11: Overview of the Model with 115 Topics in Robotics.
The figure provides an overview of the 115 topics in Robotics. The topics are displayed according to their relative share in the overall data set and ordered decreasingly. The colouring indicates the document type, which is significant for the topic with a 95% confidence interval (Figure C.2). In case of no significant difference, the colouring of the bar is kept in grey. The description to the right of each bar displays the topic's number and the five words with the highest occurrence probability in the topic.

Toys and games are often also represented in the patent class A63H, which also relates to toys. "Service" occurs among the main words in topics 37, 53, 83, and 79. All of these topics are significant in trademarks except for topic 79, where neither trademarks nor patents are significant. Topic 72, as the most important trademark topic, topic 37, as a service topic, and topic 69, covering unmanned aerial vehicles, are described in closer detail.

Topic 72: Topic 72 (Figure C.3) deals with game-related or toy-related aspects. The patent documents are not focused on this topic but also cover other topics. In comparison, the trademark documents are very focused on the topic. The patents provide more insights into toy inventions, while trademarks highlight robotic toys and their application. The highest frequency words in the topic align with the most important NICE Class 28, which contains trademarks for games and toys, but also to the patent class A63H that relates to toys. Overall, the general area of invention and application matches.

Topic 37: Topic 37 (Figure C.4) relates to designs, research, and technical engineering. The patents assigned to the topic are not focused on topic 37 but also relate to other topics. By contrast, the trademark documents linked to this topic are more focused on it (with a weak presence in other topics). Within topic 37, the trademark occurrence probability is generally higher than for patents. Trademark NICE class 42 relates to scientific and technological services and design, which aligns with the associated trademark documents and the topic description. The patent classes associated with topic 37 cover diverse subjects, like A61B "surgical and diagnosis", B25J "robotics", and G01N " analysis of materials". The service relation from the document snippets is present in trademarks and patent documents. In detail, however, it is noticeable that although the text snippets generally cover research, a common subject within research is missing.

Topic 69: Topic 69 (Figure C.4) represents aircraft, unmanned, aerial, vehicles, and drones. It is more associated with trademarks than with patents. Looking temporally, interesting peaks are visible in the 1985s and around 2010, indicating the market introduction of unmanned aerial vehicles (UAVs) around that time. The trademark classes 7 and 9 are related to machines and instruments for controlling and monitoring aircraft, while the patent classes relate to aerodynamic aspects (B64C), Master-slave robots (A61B34), and controls (G05D). Further of interest is that the trademark, as well as the patent documents, are very focused on the topic of unmanned aerial vehicles. This is in contrast to topic 72 and topic 37, where only trademarks tend to be focused, whereas the patents are also present elsewhere in the topic space.

In summary, the trademark topics generally cover aspects similar to the main NICE classes. The topics provide further details on the underlying subjects like unmanned aerial vehicles. In trademark topics, patents are less focused, while trademarks are generally focused. This pattern is also visible in patent topics where the trademarks are less focused than the patents. An exception is topic 69, where the patent and the trademark documents are similarly focused on the topic. Here, the document descriptions of patents and trademarks are very aligned. Nevertheless, the topic occurs in trademarks with a higher probability than in patents and represent essential developments in the field.

Neither: Most topics are not identifiable as trademark or patent topics. This implies that patents and trademarks occur in the topics. However, no significant difference in the occurrence probability is observable. The topics represented by this group are very diverse, with many main topics focusing on core robotic aspects like "motor", "drive", "manipulator", "system", "control", "computer", and "mechanism". A group of topics focus on aspects around communication like "network", "communication", "data", "wire", and "connection". Others focus on application areas like "welding", "medical", or "milking". Topic 90, as the topic with the highest occurrence probability in this group, which is also related to core robotics, topic 109, as a communication topic, topic 27, for medical applications, and topic 103, for milking applications, are chosen for closer analysis.

Topic 90: In topic 90 (Figure C.6) no IPR type is significant. The topic nevertheless has a clear technical relation with “drive”, “motor”, “shaft”, “gear” and “wheel” as the main words. The main patent and trademark documents are focused to a high degree on the topic. Also, the shares overlap over time, with a lower mean of occurrence in trademarks than in patents. The most important trademark class is NICE class 7, which relates to machines and engines. This is again in line with the topic estimation.

Topic 109: Topic 109 (Figure C.7) describes “network”, “communication”, and “application” aspects. The main documents describe, e.g. voice communication, internet methods, and telecommunication. The topic strongly overlaps patents and trademarks in its estimated occurrence and in the time overview. The patent classes relate to robots and positioning, while the NICE classes are related to, among others, machines (cl. 7), data transmission (cl. 9) and the development of computers and software (cl. 42). The main documents in the topic are focused on the topic. Both IPRs display a high level of detail in their descriptions and are in line with each other.

Topic 27: The medical topic 27 (Figure C.8) is occurring in both trademark and patent topics, with higher shares in the latter. Over time, the estimated probability is similar but slightly higher for patents. The related classes also pick up the topic of medicine, with medical robotics in CPC classes and medical instruments and services in the NICE classes (like NICE class 10 with surgical and medical apparatus).

Topic 103: Milking of animals is relevant in topic 103 (Figure C.9). The topic is equally represented in trademark and patent documents. The topic development over time is very similar, with, generally, very small shares. The related NICE classes are, among others, about machines and tools (cl. 7), scientific and research instruments as well as data processing (cl. 9), apparatus for cooking and feeding (cl. 11), plastics, and packaging and pipes (cl. 17). CPC class A10J5 and A01J7 include accessories for milking machines and devices, and A01K1 the housing of animals.

The combination of patents and trademarks is very consistent for topics that are to a similar proportion in patents and trademarks. This means that the topic words, patent and trademark descriptions align. Patents and trademarks combined give perspectives on the same topic and provide different perspectives. Remarkably, the documents have a similar focus on the topics. The topics cover multiple aspects of robotics, from core functions to more specific applications.

Before synthesising the results to answer the research questions of robotics, firms’ involvement is considered. As the involvement of firms is of interest for the studies on innovation, the involved firms are exemplarily derived for topics 27 and 103 (see Figure C.10). In line with the topics, the highly involved firms in terms of IPR application have a background in either surgical robotics or milking. For example, the firm Intuitive Surgical Operations develops surgical robots and is responsible for one-quarter of the topic.[11] Ethicon Inc., Ethicon Endo Surgery Inc., Hansen Medical and Covidien are also from the medical sphere. It is interesting that Ethicon Inc. and Ethicon Endo Surgery Inc. are both associated with topic 27 as these firms merged in 2013 (Arnold 2013). In topic 103, DeLaval (2021) and Lely (2021) are firms active in the dairy and farming industry and, thus, apparently also active in the development of milking robots. The combination of trademarks and patents thus can detect firms that are active in similar areas and even provide insights on mergers from a content perspective. In general, the model displays a high firm-topic-relation.

The analysis implies the following for the research questions:

[11] This means that across all documents, the share of documents that relate to the firm and the topic sum up to 24.25%.

RQ4.1: *Can trademarks be combined via their textual data on a detailed level?*
The analysis of the different topics revealed that the combination, in general, is possible. The different topics are consistent regarding the matched documents, the words with the highest frequency and the related classification. This means that, for example, a topic about milking in robotics is contained in patent and trademark documents that both focus on milking. Thematically, this means that the topics present in the patents are also present in the trademarks. Trademarks and patents are most aligned in the topics where no significant difference between trademark and patent occurrence probability is observable. Compared to the classification systems of patents and trademarks, the topics display a higher level of detail. Further, most topics are covered by several classes. The topics provide more details than, especially the NICE classes.

RQ4.2: *Does the textual combination of trademarks and patents add new insights to the discussion of innovation?*
Different aspects need to be considered to answer whether the textual combination of trademarks and patents adds new insights to the discussion of innovation:

- Trademarks and Innovation: The matching of patents and trademarks is generally very consistent. This implies that topic-wise, the inventions in patents are also introduced to the market, at least thematically. Further, topic 69 of unmanned aerial vehicles, for example, is an innovation area in robotics. It is, however, significantly present in trademarks. The inclusion of trademarks into the analysis increases the importance of topics with a higher occurrence probability in trademarks than in patents. Overall, the analysis thus indicates that diffusion of inventions takes place and that trademarks cover topics that are related to innovation. The analysis of trademarks could therefore be sufficient in cases where an overview of the market introduction of inventions is sought. The textual analysis does, however, not ensure that the same firm performs the market introduction. It could, nevertheless, provide important insights to understand the evolution of topics over time and across multiple firms. The firm-perspective on milking and medicine gave first insights into that direction: firms were linked via joint topics that are also firm-wise connected (like Ethicon Inc. and Ethicon Endo Surgery Inc.).

- Market and Products: The trademark descriptions ensure the market introduction of the topics. The topics of trademarks often display an application area like toys, education or energy. Further, some topics are focused solely on end products.

- Services: Services are represented in trademark topics or topics with no significant IPR difference. The services in the trademark topics displayed a list of service areas without a clear focus. Services were less present in patents. In areas of service robotics, like milking or farming, patents and trademarks are equally represented. This can mean that in general, services are more in trademarks. However, with an increasing link to technology, as in milking and farming, the patentability increases and as such the occurrence probability in patents increases.

- Technical innovation: The patent topics are mostly covering core functions of robotics. Also, many topics that are equally in patents and trademarks focus on core components or the communication aspects of modern robots. Trademarks display less detail in technical aspects and rather represent the final product than individual components. However, trademarks nevertheless cover technical aspects, and their share in Robotics is also increasing. Topic-wise, it is noticeable that topic 69 about unmanned aerial vehicles is significant in trademarks and that topics like topic 90 with general aspects or topic 27 and topic 103 display no signifiant difference in occurrence probability between patents and trademarks. Example descriptions of trademarks are "drones in the nature of unmanned aerial vehicles for the purposes of aerial photography" (see Figure C.5d), "tightening rings for connecting gear drive shafts sold as a component part of gear" (see Figure C.6d), "surgical, medical, dental and veterinary apparatus and instruments, in particular, medical devices, namely, access devices in the nature of trocars for use in manual and robotically" (see Figure C.8d) and "equipment for stable and pasture and for dairy cattle, namely, machines for supplying fodder concentrate in fodder troughs" (see Figure C.9d). These example texts reveal technical aspects present in trademarks.

In terms of new insights added thanks to the combination, it can thus be said that on one hand, patents still focus more on inventions than trademarks. On the other hand, trademarks provide insights especially on the market introduction or on the application area of technologies. Additionally, some topics like unmanned aerial vehicles gained increasing importance thanks to the combination with trademarks. Services are also still mainly in trademarks. When it comes to technical aspects, trademarks increasingly cover insights into the technologies.

Overall, for the innovation area of Robotics, the combination of trademarks and patents based on textual data is consistent and provides a high level of detail and insights into robotics development. Trademarks, as well as patents, are increasing in Robotics. Interestingly, technical aspects are not only covered by patents but also by trademarks, even though the share is still higher in patents. Each IPR thus can contribute to the understanding of innovation in Robotics. The combination can be especially of interest in areas where patent coverage is low and further information on the market introduction is of interest.

4.5 Results of the Footwear Analysis

In the case of Footwear, the 94,064 cleaned documents resulted in a model with 89 topics. Trademark documents are the dominant document type, providing 84.85% of the total documents to the model (see Table 4.2). The number of trademarks registered shows a large increase over time, while patent registrations do not grow very much (Figure 4.12). The three most important CPC classes are A43B with 70.34% ("Characteristic features of footwear; parts of footwear"), A43C with 15.65% ("Fastenings or attachments for footwear; laces in general") and Y10T with 13.39% ("Technical subjects covered by former U.S. classification"). A43, in general, is associated with "Footwear" (CPC 2020). In the case of NICE classes, most trademarks are registered in class 25 with 86.44% (used "mainly for clothing, footwear and headwear for human beings") and class 35 with 11.28% ("Advertising; business management, organisation and administration; office functions") (WIPO 2019a) (see Figure C.13). From a firm perspective, "Nike Inc." has the most IPR documents in the data set, with 2,048 associated total IPR documents. The firm has the most patent applications (1,861) and is one of the largest applicants for trademark protection, with 187 trademarks. "Asics Corporation" has the highest number of trademark applications (307) (see Figure C.14).

Figure 4.13 displays the 89 topics in Footwear. Eight topics are clearly either patent or trademark topics, but the majority of topics displays no significant difference in occurrence probability in trademarks or patents. Compared to Robotics, the highest frequency words are more focused on clothing items and do not express a connection to innovation activities. Closer analyses are performed for patent and trademark topics as well as for topics without a significant difference between patents and trademarks:

Patent Topics: The words with the highest frequency in patent topics are e.g., "sleeved", "portion", "sole", "boot", "binding", and "strap". While topic 23 covers diverse aspects of clothing and is more focused in trademarks than patents (Figure C.16), topic 25 reveals information on sole mounting (Figure C.17) or topic 14 on ski bindings (Figure C.18). The topics are mostly consistent, with the topics being present in patents and trademarks. The descriptions are generally less technical in patents than in trademarks. In Topic 14 (Figure C.18), for example, aspects of skis and boots are present in patents and trademarks. Patents are very detailed like "a safety ski binding mounted on member movable with respect to the ski but fixedly secured thereto.". The trademark descriptions, however, do not disclose further information on the boot mechanisms and rather describe the

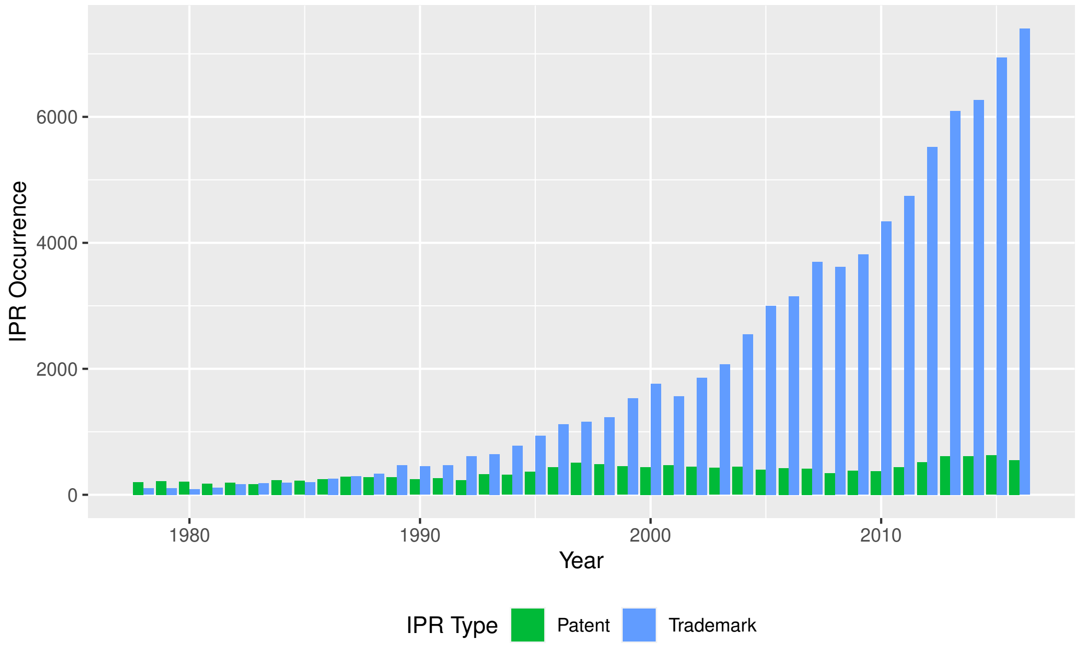


Figure 4.12: Footwear Patent and Trademark Registrations per Year.
The figure illustrates the yearly filings of registered trademarks and patents in the Footwear data set, with years referenced to the application filing year in trademark or patent application.
Source: Own representation based on the footwear data set.

application with “downhill ski equipment, namely, a flexible boot and rigid support combination for use in skiing”.

Trademark Topics: The trademark topics cover “coat”, “top”, “sweater”, “headgear”, and “sportswear”. These are, in general, clothing items without a technical relation. Topic 12 (above 15%) and topic 41 (above 5%) are the topics with the highest occurrence probability in the data set. Topic 12 “coat, sweater, belt, sock, scarf” (Figure C.19) is not the focus of trademarks or patent documents. The trademark descriptions are a list of clothing items that relate to the main words (“clothing, namely, pants, shirts, sweaters, footwear, underwear, coats, ties, [...]”). The patent documents are present also in other topics and cover aspects like rotatable shoe racks usable for belts. The clothing items are thereby not the focus of inventions but cover products in trademarks or side-topics of patent descriptions. The same applies to topic 41 (Figure C.20), where the patent documents are only distantly related to the topic. The combination of patents and trademarks is, in general, more consistent in the patent topics. While the patent documents contribute different inventions that can be derived from the descriptions, the trademark documents only provide a list of clothing items and no further detail for an improved understanding of the innovations in Footwear.

Neither: Most topics display no significant difference in occurrence probability in patents or trademarks. Examples are topic 21, topic 29, topic 56 and topic 70. Topics 21 and 29 are the most important in that group. To cover services, topics 81 and 26 are of interest. Topics 56 and 70 are examples of topics with consistent patent and trademark matching. Topics 21 (Figure C.21) and 29 (Figure C.22) are the focus of trademarks and less of patents. The former is on shoes, the latter on dresses. The main words in the topics are dominated by the trademarks associated with the topics. The patents relate to, e.g. topic 21 and describe aspects of shoe and boot inventions. The trademarks only contribute to final products without further details on the technologies behind them.

In topics 53 and 70, the patent and trademark documents are focused on each topic. Topic 53 (Figure C.23) focuses on “computer”, “electronic” and “devices”. Patent and trademark documents provide information on the same areas and the topic is very consistent. Here again, patents describe

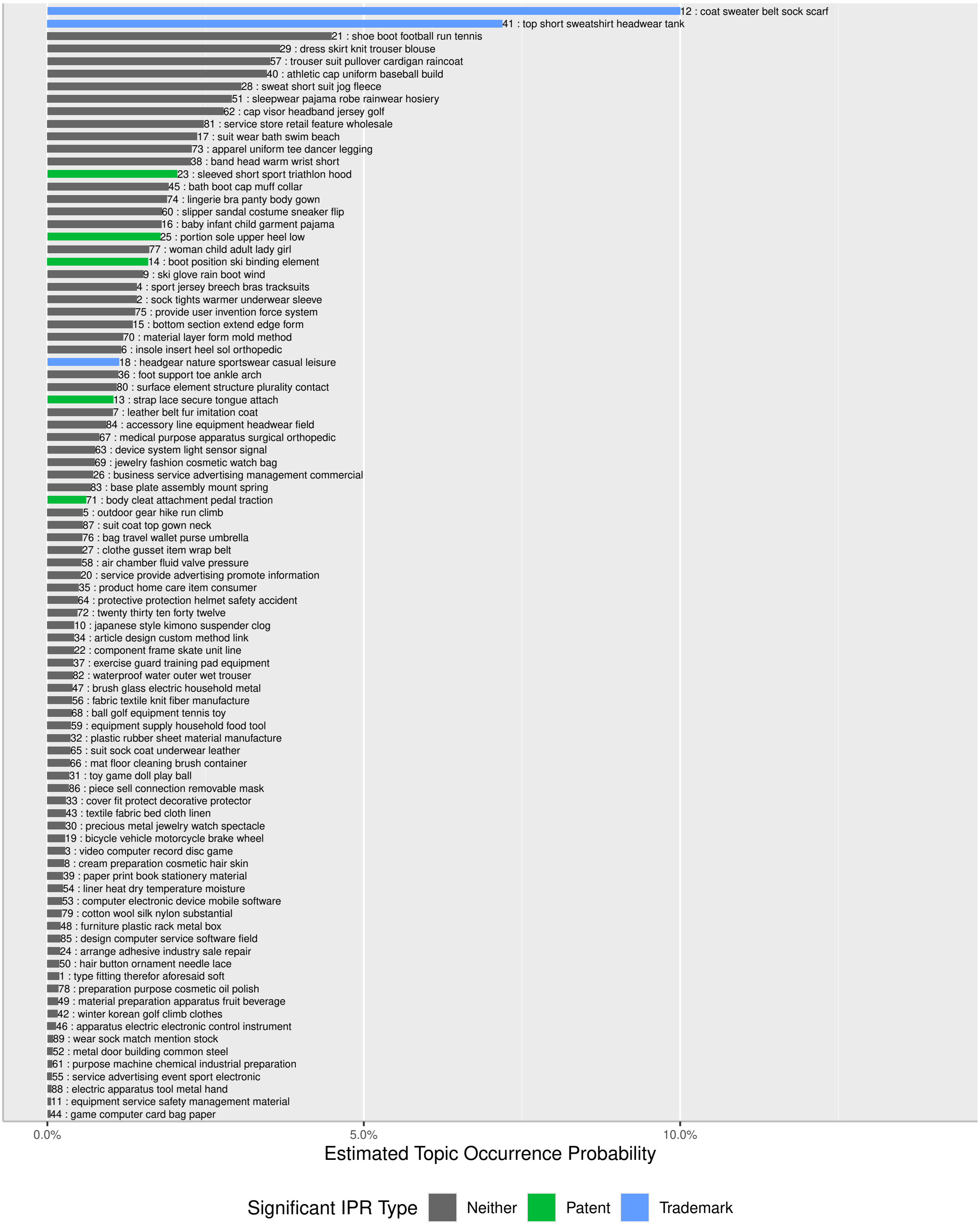


Figure 4.13: Overview of the 89 Topics in Footwear.
The figure provides an overview of the 89 topics in Footwear. The topics are displayed according to their relative share in the overall data set and ordered decreasingly. The colouring indicates the significant document type, with a 95% confidence interval (Figure C.15). In case of no significant difference, the colouring of the bar is kept grey. The legend on the right displays the topic's number and the five main words with the highest occurrence probability in the topic.

in greater detail the invention, while trademarks focus on products like computer hardware, monitors or drivers. According to the documents in which the topic is present (see also Figure C.23), this covers for example a locket containing medical information that are incorporated in shoes (USPTO patent No. 6419158) or stomp detecting sensors in shoes for use in playing computer games (USPTO trademark No. 78683123). Another example is Topic 70 (Figure 4.14), that gives insights on materials, forming and moulding. It provides further details also in trademark documents. The different materials are described in detail. Here also, trademarks and patent documents are focused on the topic.

Services are covered in topics 81 and 26. Topic 81 (Figure C.24) is a service topic. Both patents and trademark documents are about service-related aspects, with aspects around the storing of Footwear. Topic 26 (Figure C.25) covers more advertising and subscription aspects of Footwear. Both service topics are generally consistent and reveal how services are used in the context of Footwear.

Concerning the research questions, this implies the following:

RQ4.1: *Can trademarks and patents be combined via their textual data on a detailed level?*
The combination of trademarks and patents based on their textual data is possible, but the quality of the combination varies. The topics provide insights into, for example, technological aspects or materials when patent and trademark documents are consistently on the topic. Then, insights on Footwear besides the final product can be gained. However, this is often not the case, with only lists of clothing dominating the topics.

RQ4.2: *Does the textual combination of trademarks and patents add new insights to the discussion of innovation?*
The insights gained for innovation research are less obvious than in Robotics. Foremost, the relation of trademarks to innovation must be questioned. While some trademarks still provide similar or further details, most do not provide additional information on the inventions or their application areas. In consistent topics, a broad connection between patents and trademarks can be made. The level of detail is however, less granular than in Robotics.

Overall, the Footwear topics are less consistent than the Robotics topics. The combination is thus less reliable. The insights from the combinations are further limited: in topics with consistent results, trademarks and patents broadly contribute to the same topic. However, the link to innovation is less clear than in Robotics. The trademark documents also reveal a difference: The trademark documents consist more of a list of application areas than precise descriptions of the inventions of shoes. This is particularly noticeable when comparing the texts with robotics.

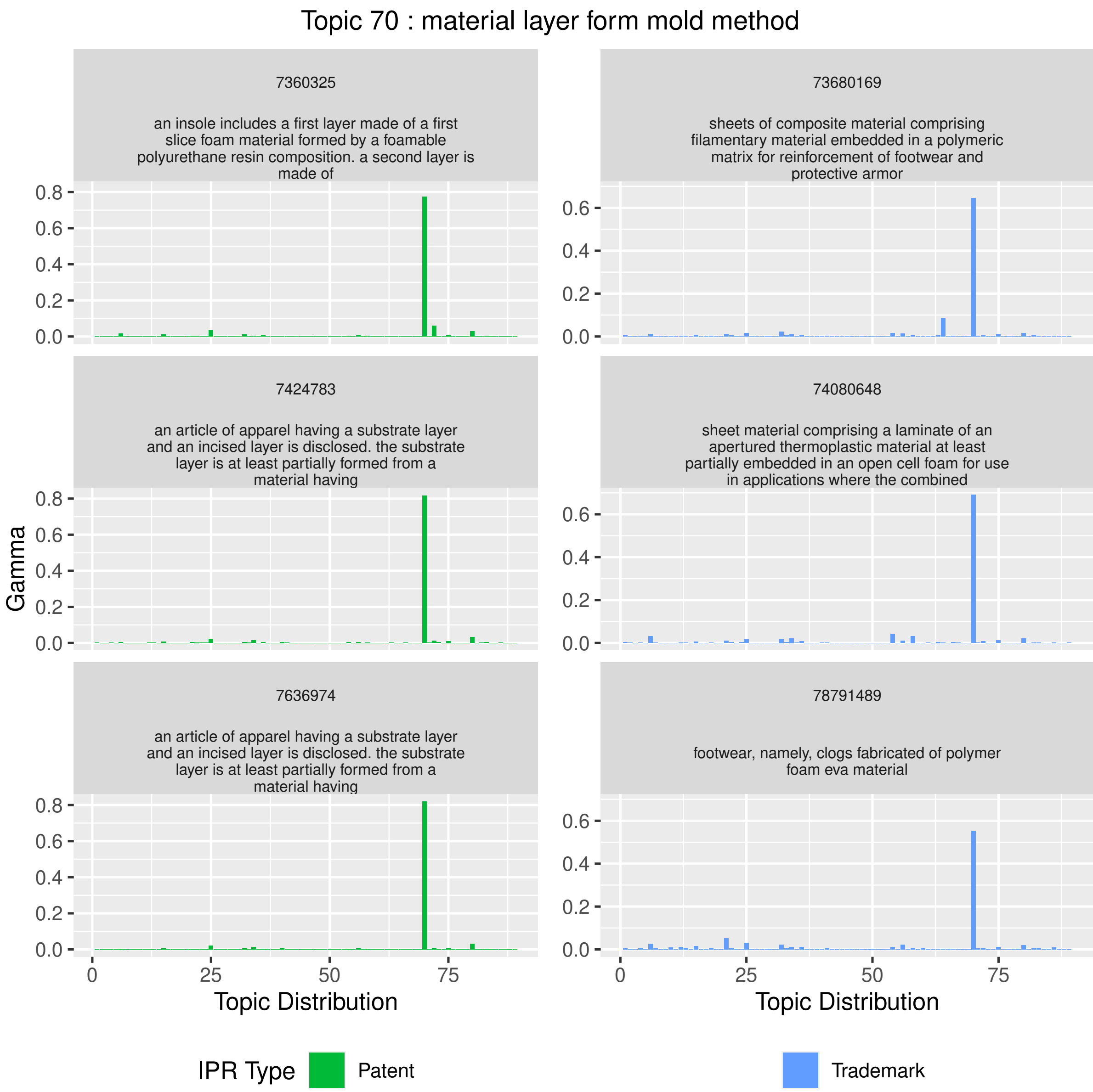


Figure 4.14: Main Patent and Trademark Documents in Topic 70 of Footwear Model 89. The figure displays information on topic 70 of model 89. At the top, the five words with the highest occurrence probability in the topic are displayed. Further, the three main documents are selected per document type. The official registration number and a short snippet of the textual description are given. The bar graph underneath each document's header displays the document's overall topic distribution. This helps interpret the data as it provides information on the level of focus on specific topics of the document.

4.6 Discussion

Comparing Footwear and Robotics, the Robotics data set contains more patent documents than trademark documents. In Footwear, the data set contains around five times as many trademarks as patents. For both areas, for some topics a significant occurrence probability is observable for patents or trademarks. Yet, for most topics no difference in occurrence probability between trademarks and patents is observable. Most topics in Robotics and Footwear occur similarly in both IPRs. Within Robotics, the topic occurrence probabilities of document types converge over time, with trademark probabilities converging towards those of patents. In Footwear, the probabilities of patents and trademarks are clearly separated before 2000. Afterwards, they start to converge. However, the convergence is less pronounced than in Robotics. For patent topics, trademark topics and topics without a significant difference in IPR occurrence, the following can be said:

Patent Topics: Patent topics were, e.g., about signalling or positioning in Robotics or bindings, mounting, or soles in Footwear. These are functions, mechanisms or parts found in technical products but not final products themselves. The patent documents in Robotics were generally focused on the resepctive topics, while the trademark documents were less focused on one topic.

Trademark Topics: Trademark topics covered clothing in Footwear and machines in Robotics, which are broad categories of final products. In Robotics, the trademark documents were focused on the topic, while the patents were less focused on the topic. The topics were, in general, in line with the classification. Trademarks contributed further details on Robotics like unmanned aerial vehicles or technical aspects that are described in the trademark applications like medical apparatus or milking equipment. In Footwear, however, the topics displayed a listing of clothes in the trademark documents.

Neither: Topics with no significant difference between the IPRs were, e.g., surgery or milking robots in Robotics or clothing items in Footwear. In Robotics, the documents of trademarks and patents were both focused on the topics and aligned. In Footwear, only some topics were consistent. Patents provided more details on the invention than trademarks.

For reasons of robustness, models with different numbers of topics per area were also examined. The results remained similar. Applied to the research questions, the following results:

RQ4.1: *Can trademarks and patents be combined via their textual data on a detailed level?*
The analysis displayed mixed results. In Robotics, the combination of trademarks and patents via their textual data displayed consistent results. The topics occurrence probability in trademarks and in patents converged over time, highlighting that the two data sources are increasingly aligned. Thereby, the level of details in trademark descriptions increased over time: While *"Industrial Robots"* or other short trademark descriptions were common in the 1980s in Robotics (like in USPTO trademark serial no.'s: 73325186 or 73422044), the detail of the description increases to, e.g., *"Robots for personal, educational and hobby use and structural parts therefor; robots for educational use in schools and components and sub-assemblies for use with the robots; robot kits comprised of servos, wheels, threaded nuts, screws, and structural parts; and software for controlling robots"* (USPTO trademark serial no.: 86871821). The textual combinations were in line with the classifications of patents and trademarks but provided a more detailed perspective.

In Footwear, the combination of patents and trademarks was only partially consistent. The data set consisted mainly of trademarks, only a small number of patents was available. The information contained in the textual data of trademark and patent was very different. Trademarks revealed little information on inventions to generate overall consistent topics. The Footwear trademarks mostly focused on final products without any or only some further detail. An example is USPTO trademark serial no. 75456406: ‘*‘Clothing, namely, pants, shirts, sweaters, footwear, underwear, coats, ties, sweatshirts, scarves, belts, gloves, hats, socks, jackets and blazers."* The trademark descriptions of topics with consistent results displayed a different textual structure. Here the trademarks provided more information than just the final product, with higher similarity to patent abstracts, like USPTO trademark serial no. 73680169: *"Sheets of composite material comprising filamentary material embedded in a polymeric matrix for reinforcement of footwear and protective armor."*

In general, the trademark descriptions in Footwear are less specific regarding the production of Footwear than the trademark descriptions in Robotics. They contain rather different application areas of trademark themselves, like other forms of clothing, than further information on the background or related innovations. In analysis of the topics, it became obvious that compared to Robotics, Footwear is a low-technology area with fewer inventions and technological applications. Firms are more focused on the marketing of their products than in the invention of new technologies.

Generally, the combination of trademarks and patents is possible. However, the level of detail available in trademark descriptions differs depending on the innovation area looked at and the time covered. This influences the consistency of the topic modelling results. Consistent topics are necessary to cover additional insights with trademarks. This needs to be considered when trademarks' textual data are used to analyse innovation.

RQ4.2: *Does the textual combination of trademarks and patents add new insights to the discussion of innovation?*
Different aspects are again considered to answer the question of additional insights on innovation through the combination of trademarks and patents via their textual data:

Trademarks and Innovation: The first aspect concerns the innovative content of trademark descriptions. This is relevant, as the combination should add information to the discussion. Therefore the trademarks need to relate to innovation and add additional insights to the discussion on innovation.

In Robotics, the combination of trademarks and patents in topics was generally consistent. The topic words, the trademarks and patents, as well as the classes, were thus aligned. The trademark documents covered the same area of inventions as patents. The representation of each topic in patents or trademarks varied depending on the subject covered. Some topics, like unmanned aerial vehicles, were present more in trademarks than patents. In Footwear, the topic consistency was given only for some topics. In these cases, trademarks did provide additional information on innovation, like materials. Often trademarks provide only information on the products without further details on the related inventions or application areas.

Overall, the relation of trademarks to innovation is partly visible in the text data, partly the textual data rather suggests a marketing focus. The degree to which information on innovation can be extracted from trademarks highly depends on the innovation area. A certain degree of technological relatedness is required to combine trademarks and patents and to generate topics in trademarks where additional information on innovation is available.

Market and Products: A disadvantage of patents is the unclear diffusion of inventions into the marketplace. Trademarks cover the market introduction; the combination could help ensure that an innovation and not just an invention is covered. In addition, trademarks are known for their market application. Therefore, trademark topics should be mainly about products.

Most topics in Robotics and Footwear did not display a significant difference between trademark or patent occurrence probability. In Robotics, the topics' combination of patents and trademarks was very consistent, and the documents were well aligned. In that aspect, trademarks ensured the market perspective of the invention topic covered in patents. Even though this does not ensure the introduction of an individual invention, this nevertheless shows that the subject area of several inventions is used in the market. This takes into account that the invention and the market introduction might be performed by different firms. Market and product applications were also described in patents but not as the main focus. At the same time, patents helped to interpret and bundle trademarks, as the patent classification system provided more structured details than the trademark classification system.

For the topics that were significant in patents or trademarks in Robotics, it became obvious that the trademark topics covered more different application areas of Robotics like energy or education or covered the final product. Patent topics were stronger in the technical aspects. In Footwear, the topics of patents and trademarks were often not consistent. Therefore, the trademarks could not provide clear details about the market introduction of inventions. The trademarks provided the market introduction perspective for the topic in the case of consistent topics. Then, the inventions described in patents were equally present in the description of trademarks. In Footwear, the trademarks were mostly focused on final products without further details on their invention. Here the advantage of the combination and textual data of trademarks for innovation research needs to be questioned.

In general, trademarks cover more final products and the market perspective. They can add the market and product perspective in the analysis of innovation, while patents focus more in detail on the inventions behind them. However, the quality of the results depends highly on the application area analysed and the trademarks covered in the data set.

Services: One of the main intentions of the combination of trademarks and patents was to cover services in the context of innovation better. The question is whether the textual data of trademarks did provide additional insights.

The service topics were present either in trademark topics or in topics with a similar occurrence probability in patents and trademarks. In Robotics, a tendency of these topics to occur in trademark was observable. The services are related to education, development, and entertainment. Further, service robotics areas like milking or surgery were covered by topics with similar occurrence probabilities in trademarks and patents. Trademarks and patents thus equally captured these areas. In Footwear, service or retail topics were equally represented in trademarks and patents. Insights on services could thus be gained from both data sources, with a higher representation in trademarks.

Overall, services without a straightforward technical application are more in trademarks like entertainment or education. However, services with a technical component are also observable in patents like service robotics or footwear subscriptions. This is in line with Blind et al. (2003) that find that services with technical aspects are also present in patents. Thus, the inclusion of trademarks provided additional insights on services and highlighted the service areas in Robotics or Footwear.

Technical innovation: Inventions protected by patents are required to cover a technical aspect. Trademarks do not have an invention or technology requirement. Information on the technologies themselves may not be displayed in trademarks.

The patent topics in Robotics and Footwear mostly covered technical aspects and focused on functions or components. In Robotics, trademarks also contribute detailed technical information on various topics independent of the significant IPR. An example is trademark serial no. 87255341: *"Surgical, medical, dental and veterinary apparatus and instruments, in particular, medical devices, namely, access devices in the nature of trocars for use in manual and robotically-assisted minimally invasive surgical procedures in the nature of laparoscopic, endoscopic,gynecological, urological, thoracic, colo-rectal, gastrointestinal and bariatric and general surgery; [...]"*.

In Footwear, however, the technical details are provided mainly by patents and rarely by trademarks. Regarding contribution, patent descriptions still provide more technical information than trademarks. Yet, the level of detail in trademarks increases, as described above.

In summary, trademarks do contribute further details to the discussion of innovation. Trademarks shed light on the market application of inventions and include further details on, e.g., services. Technical aspects are also covered, but are still more important in patents. The insights gained from trademarks still depend on the consistency of the topics and the innovation areas covered. Yet, the share of trademarks in technical topics and also the level of detail in trademark description increases. Thereby, the potential additional information that can be extracted from trademark descriptions increases, which makes them increasingly interesting for further textual data analysis.

The textual data analysis revealed differences in trademark descriptions. It is interesting to note the different details of the documents depending on the area covered. In Robotics, trademarks do not provide the same, but similar information as patents. The share of each IPR in each topic thereby varies. This makes the combination interesting as, for example, trademarks can cover service-related topics, or the diffusion of inventions can be ensured. In the low-technology area of Footwear, patents and trademarks and associated textual descriptions diverge largely, with trademarks providing, in general, less information on technological and innovation aspects. The structure of the trademark descriptions also differs, with trademarks in Robotics providing further details on the products, while in Footwear, a list structure is observable. This affects the consistency of the topics.

In general, it seemed like the trademarks of Robotics did cover largely innovations, while the trademarks of Footwear mostly covered marketing aspects. Only the trademarks in topics with high consistency between patents and trademarks seemed to relate to innovation. The possibilities to combine and the level of detail are thus sector or technology-area specific and require the coverage of innovation in trademarks.

4.7 Conclusion and Further Research

This chapter addresses the potential of textual data of trademarks in combination with patents for innovation research. Textual data thereby allow for an increased level of detail to overcome current limitations of trademarks, such as limited classification systems, difficult combinations with other data sources or missing checks for innovativeness. At the same time, limitations of patents, like the missing market introduction, could be overcome. The research questions, therefore, address whether trademarks and patents can be textually combined and to what extent this combination contributes to the understanding of innovation.

The combination of patents and trademarks is performed via Structural Topic Modelling on the innovation areas of Robotics and Footwear. Robotics serves as a representative for a high-technology area, while Footwear represents a low-technology area. The structural topic modelling approach uses textual data of the USPTO trademark goods and service descriptions and USPTO patent abstracts as input. The model extracts topics from the data sets that are then analysed.

For the first research questions, the resulting topics are assessed in terms of consistency to answer the question of the possibility of combining patents and trademarks on their textual data. In Robotics, the topics displayed a high consistency and trademarks and patents described similar subjects. The topics generally corresponded to the classification of patents and trademarks but were even more detailed than these classifications. Here, the trademark classifications were the most abstract; the patent classification already provided more detailed information, and the topics focused on detailed subjects with corresponding granularity. In Footwear, this was only observed in some topics where trademarks and patents were considered consistent. Then, the trademarks showed more detail than those in non-consistent topics.

The general impression was that a difference between the innovational content in textual data of trademarks exists that is also observable in the structure of the trademark descriptions. The possibility of the combination is thus very dependent on the area covered.

To study the additional insights gained from the combination of textual data of patents and trademarks for innovation research, common conclusions concerning the data sources were drawn from the literature. These covered the innovativeness of trademarks and the coverage of the product, market, service and technical innovations in patent and trademark data. In Robotics, the trademarks were aligned with the patents and provide detailed information. Trademarks here contributed information on services and included the market perspective. Trademarks and patents covered technological aspects, with patents still providing more details. The topic presence in patents and trademarks varied depending on the subject. The combination could thus cover additional insights as both data sources highlighted different areas and contributed different information. In comparison, the Footwear analysis and especially the trademarks did often not cover central aspects of the innovations. Only some consistent topics revealed some insights.

To conclude, the combination of trademarks and patents can bring interesting insights, with trademarks being strong in services and market application and patents covering technological aspects. The text-based combination is possible and overcomes existing limitations of the classification systems. However, it is highly dependent on the area analysed. High topic consistency is required in order to gain additional information on innovation. The chapter contributes to a better understanding of textual data of trademarks in different innovation areas, helps to overcome the existing limitation of the data sources, and serves as a basis for further textual analysis of trademarks and patents.

The analysis revealed several areas for further research: The trademark descriptions revealed consistency with patents and also a link to innovation. However, further research is necessary to understand, if the textual data of trademarks in these consistent topics only cover the market introduction or also new aspects themselves. In addition, it was observed that the trademarks of different innovation areas and over time revealed a different structure. This structures of trademarks could be used to separate innovative from non-innovative trademarks. Even if the innovation content is difficult to assess, focusing only on trademarks with a high level of information to be extracted might be of interest. The reasoning behind the varying level of detail in the descriptions of trademarks also remained unclear. A reason might be the more technical background of Robotics where the advancement of the technology is an important aspect versus the more marketing focus of Footwear, where the brand and style is more important than a new invention. Here, further research is necessary. The analysis in Robotics and Footwear covered a high- and a low-technology area, with varying results. Analysing additional innovation areas with varying levels of technology may offer further insights into the differences within the textual data of patents and trademarks. On a content level, trademarks and patents were combined based on joint topics and not based on individual firms. The idea was that multiple firms might contribute to a topic with inventions and product introductions. Further research is necessary to disentangle the innovation contribution of several firms to the same topic and the patent and trademark coverage. Lastly, it is interesting to use the textual combination of patents and trademarks to extract insights into various fields.

5 Revealing Technological Transformation and Firm Involvement through IPRs and Networks

The transformation of both low-technology and high-technology sectors is being driven by advancements in information and communication technology. However, low-technology sectors tend to be under-represented in discussions about innovation, especially when relying on patent-based indicators. Consequently, capturing the ongoing changes in these sectors poses a significant challenge. This chapter seeks to evaluate this transformation by leveraging a combination of patents, trademarks, and designs to achieve a more comprehensive understanding of sector activities. Its objective is to assess the capacity of these intellectual property rights (IPRs) to capture the shift from low-technology to high-technology sectors and to provide deeper insights into the sector's development. Moreover, this chapter adopts a firm-centric perspective to comprehend the role played by firms in driving this transformation. Additionally, it explores whether there are any changes in IPR usage patterns over time. To explore this, the chapter focuses on the transformation within the musical instruments sector. To achieve this goal, the chapter extracts trademarks, patents, and designs relevant to musical instruments. It combines their textual data using Structural Topic Modelling to estimate the topics most pertinent to musical instruments. Based on these topic estimates, networks are constructed to evaluate the sector's transformation. These networks illustrate differences between electronic and classical musical instruments. Electronic topics predominantly feature in patents, with trademarks gaining significance over time. Designs serve as a bridge between patents and trademarks, connecting the topics associated with both. Topics linked to digital sound creation and mobile aspects are more closely associated with trademarks, while classical instruments are more prominently covered by trademarks. To provide a more comprehensive understanding, these networks are further enriched with firm-specific information. These networks unveil a thematic positioning of firms, aligning with their backgrounds. For example, hardware and software companies are more active in patenting, while retailers and gaming firms tend to focus on trademarks. In conclusion, the combined use of different IPRs effectively mirrors the technological transformation within the sector and enriches the discussion with context-specific information about the transformation process itself.

KEYWORDS: Sectoral Transformation, Intellectual Property Rights, Patents, Designs, Trademarks, Structural Topic Modelling, Text Mining, Textual Data

JEL Code: O33, O34

Musical instruments are an important cultural asset, and the production is a craft rich in tradition. For example, German organ building was recognised as an intangible cultural heritage in 2018. Organ building can look back on 2000 years of history (DUK 2018). Germany is considered a traditional manufacturing country for musical instruments, dating back to the 13th century (Böcher 2008). Playing a musical instrument remains important for a large part of the German population: A survey in 2021 revealed that around 13,8% of the German population above age 14 play a musical instrument and around 1.97 million people do this several times a week (VuMA 2021). The musical instrument sector is further of economic relevance: In Germany, the sector revenue reached 756,38 Million Euro in 2019 and is expected to grow above 800 Million Euro by 2025 (Statista 2021). Worldwide, revenue reached 34,15 Billion Euro in 2022 and is expected to grow by 8.28% per year to above 50 Billion Euro until 2027 (Statista 2022). Musical instruments and their parts or accessories are additionally traded internationally and shipped around the world (OECD 2021). Interestingly, a technological transformation is taking place in the musical instruments sector: Enders (2017) notes a change in musical instruments towards electrification, modularisation and digitalisation of musical instruments. This change implies a transformation from a traditional low-technology sector with activities in the production, retail, and repair of these instruments (Eurostat 2008) to a sector with activities in high-technology areas like the production of amplifiers for musical instruments. This is reflected in the global revenue, which covers not only classical music instruments but also electronic and electromechanical musical instruments and the components and parts thereof (Statista 2022).

A change towards electrification and digitalisation is not only taking place in the musical instrument sector but is observable in all sectors, independent of being considered a low-technology or high-technology sector: Authors find evidence for transformation in the footwear sector (Raffaeli and Germani 2011), in the financial industry (Cuesta et al. 2015; Brandl and Hornuf 2020), in the energy sector (Erlinghagen and Markard 2012), in life sciences (Gbadegeshin 2019) or the automotive sector (Llopis-Albert et al. 2021). The technological diversification in the sectors is not only driven by the information and communication (ICT) industry, but also non-specialised industries are contributing to the development (Mendonça 2006). Likewise, Mendonça (2009) state that the boundaries between low-technology and high-technology sectors are blurring in the wake of digitalisation.

Overall, therefore, not only are high-tech sectors changing, but low-tech sectors are also changing. In the case of musical instruments, this means that electrification and digitisation are reinventing traditional, classical instruments with electronic possibilities. For example, an electronic piano still uses keys to strike a note, like a classical piano but generates sound not by strings but by electronic components and amplifiers. Even though innovation accordingly also takes place in the traditional low-technology sectors, these are often overlooked in the discussion of innovation (Hirsch-Kreinsen 2008). However, a better understanding of these sectors is of interest, as they absorb technologies and contribute to technological progress: Mendonça (2009) find that traditional, low-technology sectors are taking in knowledge and adapting it to their field of activity. Further, Also Neuhäusler and Frietsch (2015) find that low-technology sectors are increasingly active in high-technology areas and vice versa. This chapter, therefore, intends to assess musical instruments to contribute to the discussion of technological transformation. Erlinghagen and Markard (2012) further note that a closer look at firms and their contribution could add explanatory power to the overall discussion of transformation. It is thus of interest to understand the transformation in musical instruments, how the aspects of high-technologies change the sectors and how this relates to the firms involved.

Intellectual property rights are of interest to study the transformation, as patents, trademarks, and designs are all applied protection rights in musical instruments. Patents do, in general, cover innovation activities. However, sectors with low-R&D intensity are significantly underestimated, with patents as the common data source for innovation. Most patenting activities still relate to R&D-intensive technologies, even though patenting activities in non-R&D-intensive technologies are increasing (Neuhäusler and Frietsch 2015). In terms of transformation due to digitalisation, patent-based indicators further underestimate software-based innovations or organisational changes (Mendonça 2009). In low-technology sectors

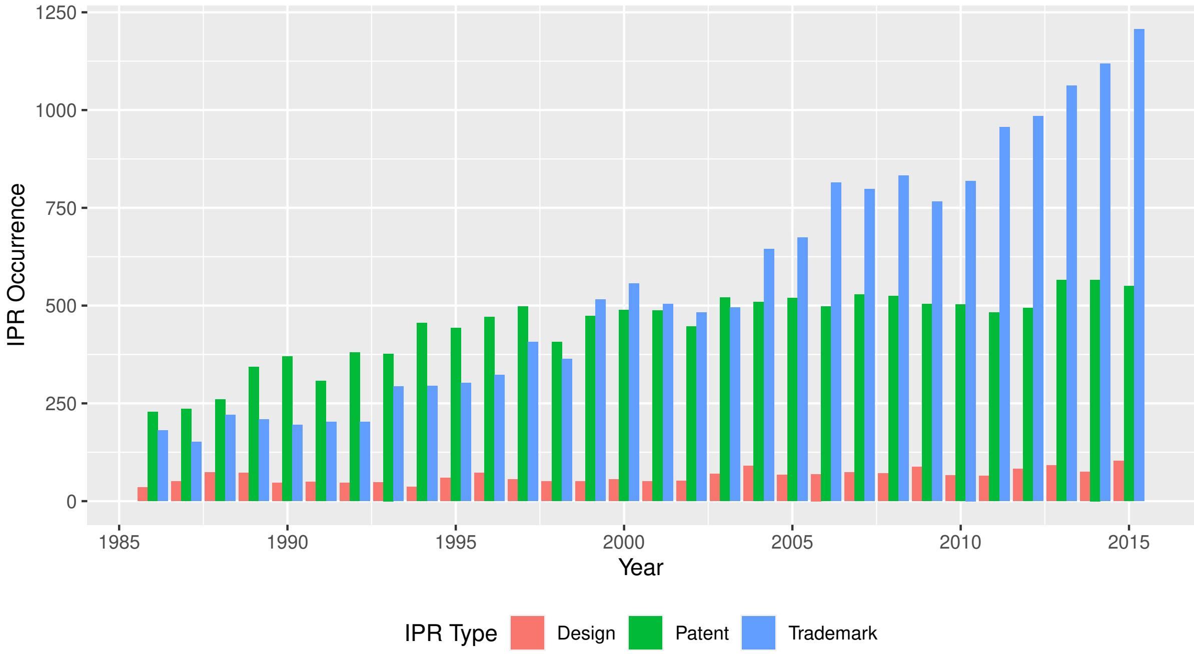


Figure 5.1: Musical Instrument IPR Registrations per Year and per IPR Type.
The figure illustrates the yearly filings of registered designs, trademarks and patents in Musical instruments, with years referenced to the application filing year in design, trademark and patent application.
Source: Own representation based on the musical instrument data set.

(Millot 2009) and in small and medium-sized firms (Flikkema et al. 2014), trademarks are used to a greater extent. In musical instruments, trademarks, in general, can be considered an indicator of musical instruments due to the importance of the origin of an instrument. E.g., a Stradivarius violin is immediately considered high quality and value. As the sound body and, thus, the product design plays a unique role in the sound generation of musical instruments, design protection is further relevant in musical instruments. Design protection is further a complement to trademarks in case of non-applicability of patents (Crouch 2010). Moreover, each intellectual property classification system considers musical instruments as a separate category, which underlines the frequent use of IP rights for musical instruments.

The combination of patents, trademarks and designs is considered to assess the change in musical instruments and to gain further insights into firms' involvement. In addition, technological transformation in the sector could lead to a change in the application of intellectual property rights: Musical instrument activities are considered to change over time from classical musical instruments to electronic musical instruments. Therefore, innovations related to the electrification and digitisation of musical instruments will become more relevant over time. This change should change the focus of firms, or the participation of firms in general, potentially involving more high-technology firms. Technological change should lead to a shift in the use of intellectual property rights toward more patenting. However, this is not the case, as seen in Figure 5.1. Contrary to the assumption that patents cover technology and are applied more frequently as electronic activities increase, trademarks began to overtake patents in 2004. A detailed analysis of IPRs in terms of topic coverage and transformation in musical instruments will be carried out to understand this pattern.

The combination of patents, trademarks, and designs in this chapter is achieved with the textual data of each data source. Textual data analysis furthermore provides content-wise information on transformation and the firms involved. Topics are extracted from the textual data to understand the transformation and IPR coverage. This is done with Structural Topic Modelling based on Roberts et al. (2019). The approach enables the combination of different data sources and maps the transformation in musical instruments across topics. The topics are related to each other, and their linkages are used to create a topic network with topics as the nodes and their relation as the links. The networks are used to identify patterns and changes related to the transformation, firms and intellectual property rights. The analysis reveals that trademark registrations surpass patents in general and electronic musical instruments. The trademarks thereby cover, among other things, aspects of digitalisation and data and remain strong in classical musical instruments. Networks are also used in the analysis to assess the connection between topics and their relationship to firms. The networks reveal that over time, firms from high-technology areas like information and communication technology become active and contribute their technical knowledge to the field. But also known musical instrument producers contribute to technological transformation. Furthermore, a firm-specific application of IPR can be observed. Firms with a background in hardware and software are more involved in patent topics, while gaming firms are more likely to protect digital aspects through trademarks. Overall, the analysis contributes to a better understanding of the transformation, the related use of intellectual property rights and firms' involvement. The chapter contributes to a better understanding of technological transformations in low-technology sectors in general.

The remainder of the chapter is organised as follows: Section 5.1 discusses the transformation in musical instruments and Section 5.2 different sectors and their relation to intellectual property rights. Section 5.3 presents the methodology and data to answer the research questions. The results are presented in two parts: a sectoral perspective in Section 5.4 and a firm perspective in Section 5.5. Before concluding the chapter in Section 5.7, the results of both parts are jointly discussed in Section 5.6.

5.1 Technological Transformation of Musical Instruments

Despite having a long history and tradition, musical instruments are subject to constant change. Electrification has been taking place since 1900, later modularisation, and since the 1980s, all the components for the digitisation of musical instruments have been available. Modularisation allows the free combination of further components, such as synthesisers, while digitisation includes digital sound generation and processing. This development continues with software-based applications, the simulation of traditional instruments, or the application of artificial intelligence (Enders 2017). New technological opportunities change the production and consumption of music. New music genres like rock music evolved, enabled by improved or new musical instruments. E.g., improved mechanics in pianos or wind instruments increased volume and improved intonation, while electronic and digital technologies enabled electric guitars or synthesisers (Lerch 2018). However, unlike an acoustic guitar, an electric guitar cannot stand alone. It needs other components, such as amplifiers or pedals, influencing the sound (Théberge 2017). In terms of transformation, this could indicate that different firms are involved in covering the different requirements of the instrument.

The definition of musical instruments is broad: von Hornbostel (1933, p. 129) defines musical instruments as everything "with which sound can be produced intentionally" and thus proposes the term "sound-producing instruments". Hardjowirogo (2017) considers the definition of musical instruments in a modern context, where electronic devices such as cell phones can produce sound. She names seven dimensions relevant for musical instruments: (1) Sound Production analogous to von Hornbostel (1933), (2) Intention, (3) Learnability through exercises, (4) Playability or controllability where actions result in sounds. (5) expressivity of play, (6) cultural embeddedness and (7) audience perception. As a sub-sector of the music ecosystem, the musical instrument sector includes manufacturers of classical and electronic musical instruments and retailers. Overlaps with Audio Equipment & Loudspeakers distributors and manufacturers are possible via e.g. the development of amplifiers for musical instruments (SoundDiplomacy 2021).

5.2 Technological Transformation of Sectors

Digitalisation and electrification are not only relevant in the musical instruments sector. According to Mendonça (2006), ICT is one of the main components to drive technological diversification. The development of ICT is driven by specialised and non-specialist industries, having their main technological capabilities elsewhere, contributing to the development. After looking at a sector perspective, Mendonça (2009) concludes that separating sectors into low- and high-technology might be questioned in the wake of renewal due to ICT. Several authors find evidence for an ICT-induced transformation: In the footwear sector, there is a need for ICT solutions to reduce labour costs. Automation can support shoe design through virtual prototyping and accelerate processes (Raffaeli and Germani 2011). In the financial industry, FinTechs provide novel offerings, which forces banks to change their offerings and pursue digital strategies (Cuesta et al. 2015). Banks are still reluctant to embrace the change, but new technologies such as blockchain can potentially disrupt the financial industry (Brandl and Hornuf 2020). In the energy sector, ICT firms introduce new business models and technologies. For example, ICT firms promote mobile communication, while incumbents support power line communication. ICT firms and startups are often enterprise software providers (Erlinghagen and Markard 2012). In retail, digitalisation caused more data-driven offerings. Especially multi-sided digital platforms like Amazon.com or Alibaba Group generate value through their digital ecosystem (Hänninen et al. 2017). In life sciences, digitalisation changes the commercialisation process from simultaneous to parallel, from product development to marketing. With AI or big data analytics, e.g., methods, trends or upcoming technologies are identified early on (Gbadegeshin 2019). In the automotive sector, digitalised processes create a competitive advantage for service and product development. Environmental requirements and demand in e.g. digital services, or autonomous vehicles fuel the transformation (Llopis-Albert et al. 2021). However, according to Mendonça (2009) patent-based indicators do not cover the breadth of the transformation in mature industries as they underestimate software-based innovations or organisational changes. Transformation due to technological development is taking place across sectors: Mendonça (2009) analyses the performance of all sectors in new technologies such as information and communication technology (ICT) or biotechnology-based on their patenting activities. The author finds that low-technology and high-technology sectors perform similarly, even though the measurement of patenting underestimates the innovation activities in areas like software. The sectors do not only adapt technologies but also contribute to their development.

Yet, low-technology or traditional sectors are often overlooked in the analyses of their innovation activity in the consideration of the knowledge society (Hirsch-Kreinsen 2008). However, these sectors not only apply technologies but also stimulate innovations in high-tech sectors through their demand. For the generation of new knowledge in the low-technology sectors, the absorption of external knowledge and the relationship to high-tech sectors are particularly relevant (Hirsch-Kreinsen 2008). In Germany, for example, low-technology or non-R&D-intensive sectors are of considerable economic relevance. Thereby, the group of low-technology sectors is heterogeneous, so each sector has its particularities (Wydra and Nusser 2015). Due to the importance of these sectors, several authors advocate the comprehensive inclusion of the sectors in the economic analysis (Hirsch-Kreinsen 2008; Wydra and Nusser 2015; Som and Kirner 2015). The sector classification is based on the intensity of R&D spending: High-technology industries have a higher average intensity by value-added and by production than low-technology industries. However, this does not mean that a high-technology industry cannot be involved in low-technology products (OECD 2011). Thornhill (2006) analyses firms and innovation in Canada and finds that high-technology firms introduce more new products than low-technology firms, whereas Hauknes and Knell (2009) observe knowledge transfers from high-technology to low-technology sectors. Still, they find only a little evidence of knowledge transfers in the other direction.

In musical instruments, the sector's economic activities vary from low to high technology activities, which also relates to low to high R&D intensity. According to the NACE Rev. 2 [1], economic activities in the field of musical instruments belong to classes shown in Table 5.1. Most activities are related to

[1] "*N*omenclature statistique des *A*ctivités économiques dans la *C*ommunauté *E*uropéenne" (Eurostat 2008)

low- or medium-low technologies in manufacturing or less knowledge-intensive services. The related R&D-intensity is on a medium or low level. Only the production of amplifiers as part of electronic musical instruments is classified as high-technology activity.The low- to high-tech activities align with the expectation that the knowledge required to produce and create these instrument types thus differs. Electronic musical instruments require knowledge of technological aspects. Classical instruments are often handcrafted based on experience and feeling. Unfortunately, the ratio of each activity for the sector of musical instruments remains unclear. As the change toward the increasing application of digitisation and electrification of musical instruments takes place, the importance of high-technology activities is expected to increase over time. However, low-technology activities such as musical instrument manufacturing or repair will remain important for classical musical instruments.

Table 5.1: Economic Activities and R&D-Intensity in Musical Instruments

Musical Instruments				
Technology Classification	**NACE Rev. 2**	**Description**	**R&D Intensity (%)**	**R&D Intensity Classification**
Manufacturing				
High technology	26.40	Manufacture of consumer electronics; including manufacture of amplifiers for musical instruments and public address systems	24.05	**High R&D intensity**
Medium-low technology	33.19	Restoring of organs and other historic musical instruments	1.93	**Medium R&D intensity**
Low technology	32.2 / 32.20	Manufacture of musical instruments	3.52	
	32.40	Manufacture of toy musical instruments	3.52	
Services				
Less knowledge-intensive (market) services	46.49	Wholesale of other household goods; including wholesale of musical instruments	0.28	**Low R&D intensity**
	47.59	Retail sale of furniture, lighting equipment and other household articles in specialised stores; including retail sale of musical instruments and scores	0.28	
	77.29	Renting and leasing of other personal and household goods; including jewellery, musical instruments, scenery and customers	0.18	
	95.29	Repair of other personal and household goods; including repair of musical instruments	0.11	

The table offers an overview of economic activities in the field of musical instruments. On the left, technology classification and NACE Rev. 2 codes are displayed. On the right, R&D-Intensity and classification information is provided. The descriptions are in the center. The relationship between NACE Rev. 2 and R&D intensity is established through the use of 2-digit ISIC codes. The classifications are categorized into manufacturing and services.
Source: Own representation based on Eurostat (2008), Eurostat (2021b), and Galindo-Rueda and Verger (2016).

Concerning intellectual property rights, Neuhäusler and Frietsch (2015) find that, on the one hand, firms from the non-R&D-intensive sectors are increasingly patenting and applying high-technologies, and on the other hand, firms from the high-technology, R&D-intensive sectors are increasingly active in non-R&D-intensive sectors. Of all EPO patents from 2000 to 2010, only 40% related to non-R&D-intensive technologies. However, non-R&D-intensive firms increasingly enter the high-technology area: Patents of German firms belonging to non-R&D-intensive sectors contribute 7% to advanced technologies or 28% to high technologies. Likewise, the shares of high-technology firms in non-R&D-intensive areas have increased. In comparison, Millot (2009) finds that trademarks play a major role in the low-technology sector and are used to a greater extent than patents. Apart from patents and trademarks, industrial designs are of interest in traditional sectors: In the furniture industry, for example, a traditional low-tech sector with high importance on design, design is a question of innovation (Lindman et al. 2008). Crouch (2010) further argues that design protection complements trademarks and can be a fallback if patent protection fails. Brem et al. (2017) find that industrial design applications and designs in combination with open innovation in SMEs and trademark application in small firms relate positively to firm performance. Overall, sectors with low technology intensity are often overlooked in the analysis. Patent-based analyses do not well cover these sectors. Especially trademarks, but also designs, could be alternative data sources for the coverage of innovation activities.

Overall, the boundaries between low-technology and high-technology, between R&D-intensive and non-R&D-intensive sectors, are blurring. Across all sectors, transformation is taking place, driven by e.g., digitalisation and ICT. Patents capture some of the innovational activities and the transformation in the sectors but do not provide a comprehensive picture. A combination of designs, trademarks and patents could capture the sectoral transformation. Erlinghagen and Markard (2012) note that in the discussion of transformation, a closer look at the firm types involved and their contribution would help provide explanatory insights.

As becomes obvious, musical instruments cover low and high technology. Patents, trademarks and designs protect innovations in that sector (see also Subsection 5.3.2). To better understand transformations with IPR use, this chapter analyses the transformation of the musical instrument sector in relation to the use of patents, trademarks, and designs. First, it is interesting to understand the sector's technological transformation and then relate this transformation to the use of intellectual property rights (IPRs). Second, as firms are involved in the development, it is of further interest which contributes to the transformation and how these firms apply IPRs. Overall, this leads to two perspectives: A sectoral perspective of the transformation with a focus on technological transformation from low-technology to high-technology and a firm perspective to understand firms' contribution to the transformation. The related research questions (RQs), therefore, are:

RQ5.1: What is the nature of the sector transformation from a low-technology sector to a sector with low- and high-technology?

RQ5.2: How does the transformation relate to IPR use?

The related research questions from the firm perspective are:

RQ5.3: Who contributes to the transformation?

RQ5.4: How does this relate to IPR use?

5.3 Methodology and Data

To analyse the transformation of the musical instrument sector and innovation activities, textual information based on patents, trademarks and designs is used as input. The use of these three intellectual property rights ensures that low-technology aspects, as well as high-technology aspects of musical instruments, are included in the analysis. The textual data of the intellectual property rights should provide information on the transformation and combine the different data sources. To answer the research questions, it is necessary to understand what is happening in the sector. It is, therefore, of interest to understand the thematic development of the sector. The thematic development can be extracted from the IPRs, since the IPRs depict innovation activities and diffusion aspects. They can therefore provide a representation of the activities in the sector. Furthermore, firms and topics are linked, so an analysis of the firms and their involvement in the thematic development of the sector is possible.

5.3.1 Methodological Approach

A structural topic modelling approach following Roberts et al. (2019) is applied to generate these topics.[2] Topic modelling enables the discovery of unknown topics in textual data and the exploration of these topics. To do so, the model first estimates the distribution of words in topics and second the distribution of topics in documents. Like that, the model assigns words to documents based on the topic-document-distribution. The model compares the resulting estimated documents to those provided to the model. The model estimation is iteratively optimised so that the estimation and the provided documents converge. This approach leads to the final analysable topics. The model requires the topic number as input, which is defined externally. The model only estimates the topics' content based on the textual data itself (Blei et al. 2003). Furthermore, Structural Topic Modelling provides additional information to the model that can influence the topic-word or document-topic distributions. It can then distinguish this additional information and allows for its examination.

In the case of musical instruments, this results in the following approach, visualised in Figure 5.2. First, musical instrument textual data from USPTO trademarks, patents and designs are extracted, preprocessed and transformed.[3] This results in the text corpus necessary for the topic modelling. The corpus consists of the documents, all words occurring in the documents and a word-document count. Then additional information is provided to the model as topic prevalence, application year and IPR type. The topic prevalence influences the topic-document distribution.

Application year refers to the year for which the trademark, patent or design is applied. This information allows for different years to cover different topics. Application year as a topical content prevalence enables an analysis of the topics over time and how the sector shifts in topics. As the topics can have different relevance over time, a spline function accounts for the changes (*s(year)*).

IPR type refers to the type of IPR from which the textual data is extracted, patent, trademark or design. This information allows for the different IPRs to cover different topics. IPR type as a topical content prevalence enables an analysis of the topic coverage in different IPRs and how the IPRs capture the transformation.

An interaction between the additional information is allowed ($s(year)) \cdot IPR\ type$), meaning different IPRs can cover different topics at different times. Structural Topic Modelling then estimates the topics. Model selection then takes place, meaning that a suitable model is selected based on different metrics. This is described in Subsection 5.3.3. The selected model is then analysed to answer the research questions. Before providing the results in detail, the general approach is described in Paragraph 5.3.1.

Model Analysis Overview After the model selection, the model results are analysed. The analysis is divided into two parts that address the sector and the firm perspective to answer the research questions:

Part I: Sector Analysis The sector analysis covers the first two research questions. The analysis considers the technological transformation of the sector from classical to electronic instruments and covers the different IPR types that lead to the topics. First, the perspectives of IPRs per topic and the Hornborstel-Sachs Classification[4] are taken to address the transformation. These are two dimensions used for data extraction. As seen in Figure 5.3, the yearly development is considered (upper left corner). The development is divided into the dimensions under consideration. In the example, these are electronic (blue bar) and classical musical instruments (red bar). In addition,

[2] For further detail on the approach also see Section 2.4 and Subsection 4.3.1.

[3] The process is described in detail in Subsection 5.3.2.

[4] The Hornborstel-Sachs Classification is a common classification system in musical instruments used since the early 1900s (von Hornbostel and Sachs 1914). For the analysis, the instruments separate into string, percussion, and wind instruments. Percussion instruments cover idiophones (vibration of the instrument itself) and membranophones (vibrating membrane). Documents from musical instruments, which are not part of one of the three categories, are grouped under "Generic".

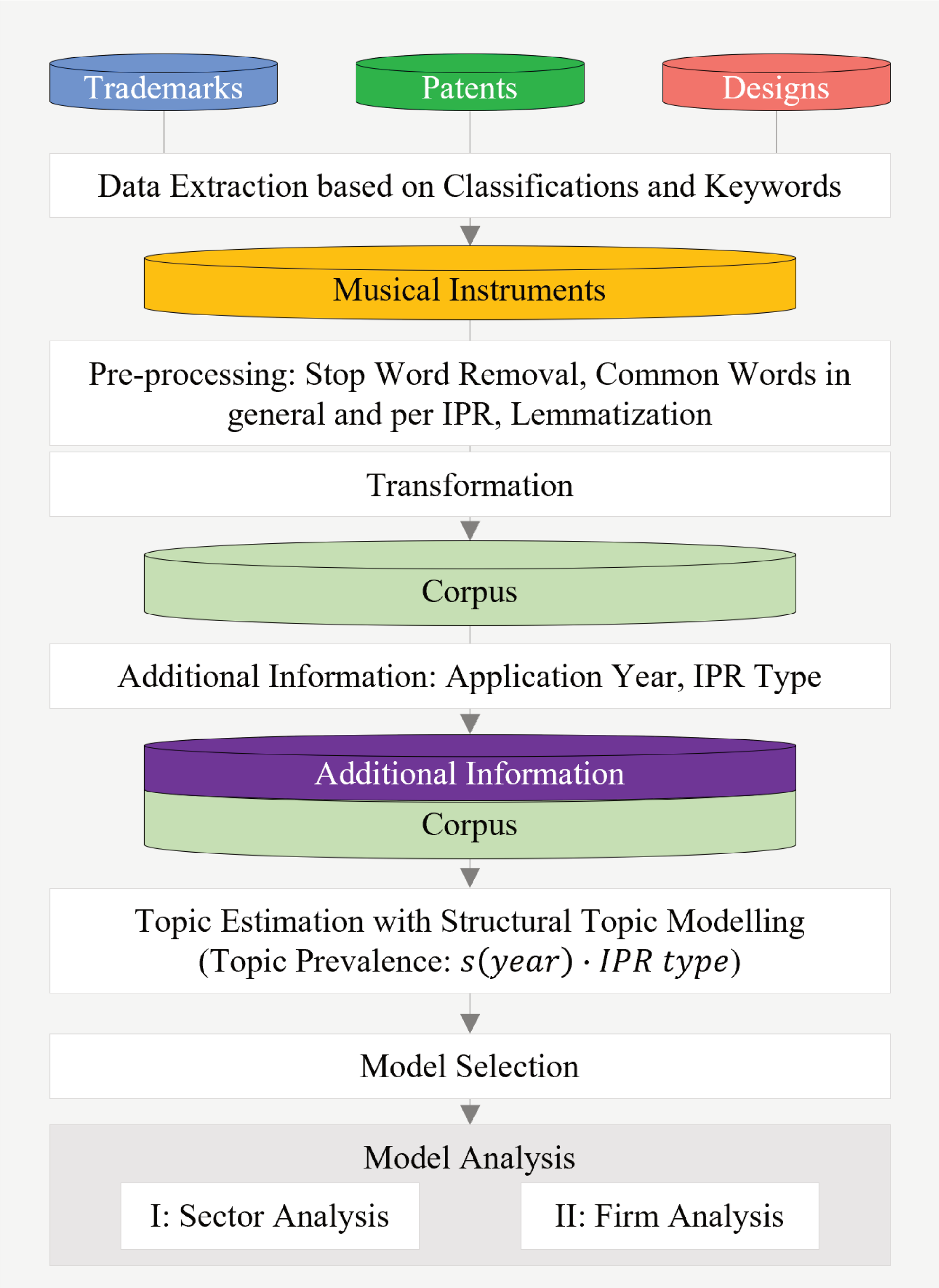


Figure 5.2: Approach Overview in Musical Instruments.
The figure illustrates the approach used in this chapter from data extraction to model selection. Data extraction relies on trademarks and patents related to Musical Instruments. This data is subsequently pre-processed and transformed to create the corpus for textual data analysis. Additional information is added, forming the basis for topic estimation, which is carried out using structural topic modelling. Among various topic estimations, a model with a specific number of topics is selected and employed for the analysis in this chapter.

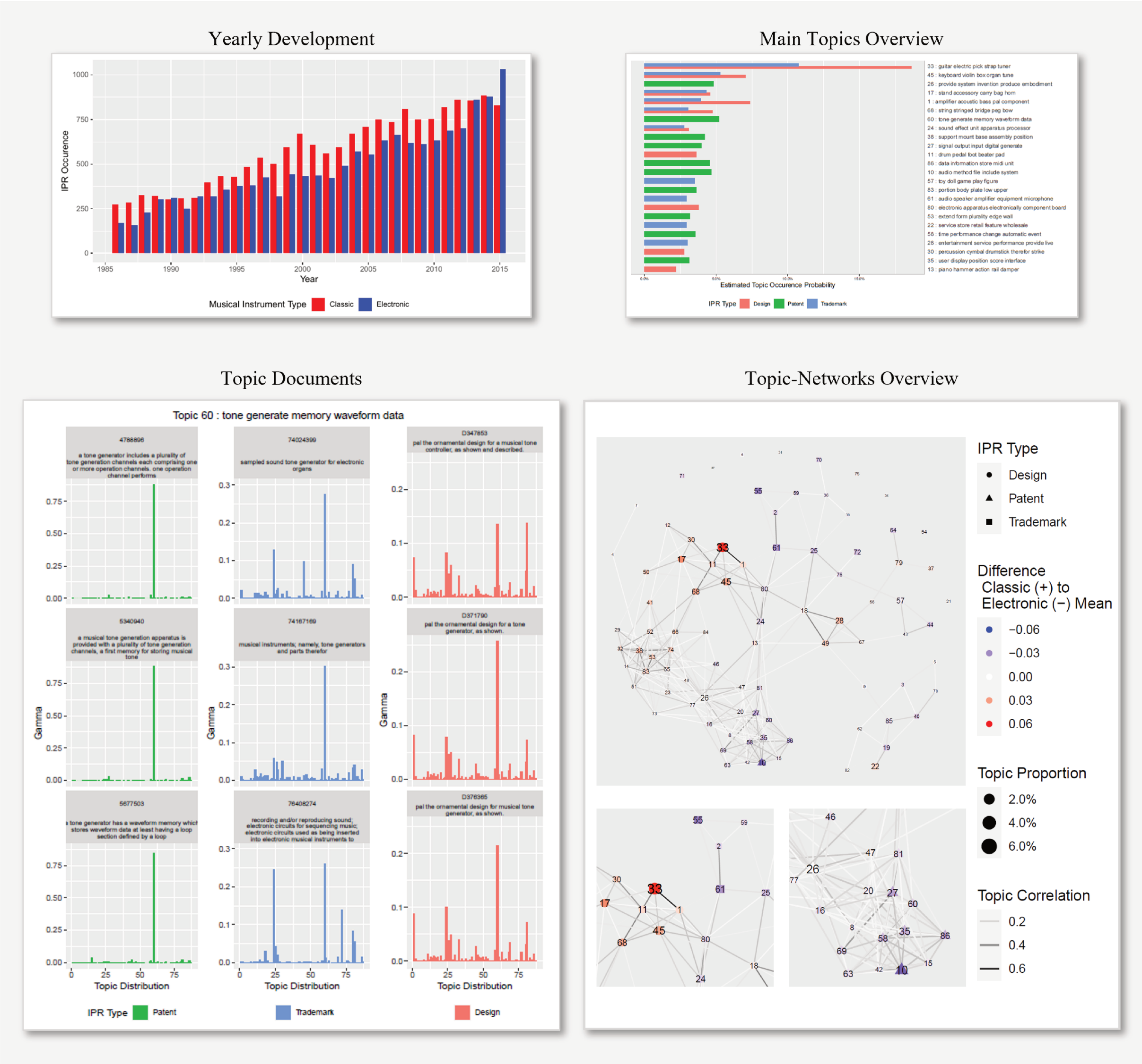


Figure 5.3: Part I: Illustrative Sector Analysis Overview.
The figure provides an overview of the types of results provided in Section 5.4. The yearly development is provided per e.g. IPR type, the Hornborstel-Sachs Classification, and musical instruments (see Figure 5.12). The IPR occurrence is determined based on the number of associated IPR documents. The main topics overview provides an overview of the main topics in the sample, here separated per IPR type (see Figure 5.7). It is used to compare, for example, topics in different IPRs and understand the topic occurrence. Topic documents provide then further details on the main documents of a topic separated for different IPRs (see Figure 5.8). Finally, different networks are analysed. An example of the topic-networks is provided (see Figure 5.19). Topics are the nodes. The links between topics are determined based on the co-occurrence of the topics. The nodes contain further information: the node shape displays the IPR with the highest relative share in the topic. The node colour indicates the node's relation, here, to electronic or classical instruments. The node size displays the topic proportion in the data set.

the main topics are considered (upper right corner). The x-axis indicates the probability that the topic occurs in the model. In the example, the probabilities are further differentiated by IPR type. This means that the probability of the topic occurring in the respective IPR is shown. These may vary from IPR to IPR, as the topics covered in each IPR type are not necessarily the same. In general, the consistency of the topics and thus the topic words and the main topic documents are considered (lower left corner).

The analysis of the sector transformation is mainly performed on networks of the topics. The network consists of nodes and links. The nodes represent topics amended with additional information like topic number (label), topic proportion (size), IPR type (shape), and tendency to classical or electronic musical instruments (colour). Depending on the network, the node's colour can also relate to, e.g. the IPR type or Hornborstel-Sachs-Type. This information is based on the documents that are related to each topic. The links represent the correlation of topics. The correlation of the topics is calculated with Pearson correlation[5]. The Pearson correlation returns the correlation of two variables between -1 (strong negative correlation) and 1 (strong positive correlation). If the correlation of two topics is larger than 0.08, it is displayed in the network. The higher the correlation, the more likely it is that two topics co-occur. In the lower right corner, an example of a topics-network is displayed. These networks are constructed from 1986 to 2015 and have five-year intervals to identify patterns that explain the change. Clusters are identified via visual inspection and are verified with short random walks to detect communities (Pons and Csardi 2021; Pons and Latapy 2005).[6] The analysis can be found in Section 5.4.

Part II: Firm Analysis The interaction between topics and firms is the focus of the second part. As the firms are owners of patents, trademarks, and designs, they can be related to topics. The analysis is limited to the major firms in the data set. These are identified based on the number of applications per firm over all intellectual properties. The firm's names are cleaned before the firm identification: The names are put in lowercase, and punctuation and spaces before or after the names are removed. Further, the names are manually unified for the major firms in the data set. The unification implies that, e.g., "nippon gakki seizo kabushiki kaisha", "yamaha corporation", "yamaha corporation of america", "yamaha hatsudoki kabushiki kaisha" and "yamaha international corporation" is unified to "Yamaha Corporation" and displayed as Yamaha. No further aggregation is performed to assess firms on an individual level. The IPRs are then grouped per firm to determine the major firms.

Based on the data, firms-topics-networks are created. The nodes reflect either topics or firms. The topic nodes are labelled with the topic number. The node shape represents the IPR type with the highest mean like design (round), patent (triangular), and trademark (rectangular). The topics nodes are coloured blue (electronic instruments) or red (classical instruments), according to the electronic or classical documents in the topic. Two topics co-occur if they both occur in a document. Based on the joint occurrence of the topics in the documents, the Pearson correlation between the topics is determined. A link between topics is displayed if the Pearson correlation between the topics is larger than 0.08. The firm nodes are labelled with the firm name and have a rhombus shape. The firm node colour represents the firm background of classical producers, game developers, hardware, producers, other, retail, and software. The node size reflects the importance of the firm or topic. Firms can be linked to topics via the occurrence of a firm's document in a topic. The firms are then considered to be active in a topic. Therefore, only the documents of the specific firm are selected, and the mean occurrence probability of all documents for each topic is calculated. The link is displayed if the mean occurrence probability of the firm's documents is larger than 0.02. Firms are not directly linked to each other, and no direct connection is displayed between firms. Nevertheless, firms can be indirectly related through their shared activities in a topic. For example, firm a and firm b are active in topic 1, so firms a and b are indirectly connected.

[5]The calculation is based on the "cor()" function of package "stats" version 4.0.3

[6]The community detection function "cluster_walktrap" of igraph is applied (Pons and Csardi 2021), where sub-graphs or communities are identified via short random walks (Pons and Latapy 2005).

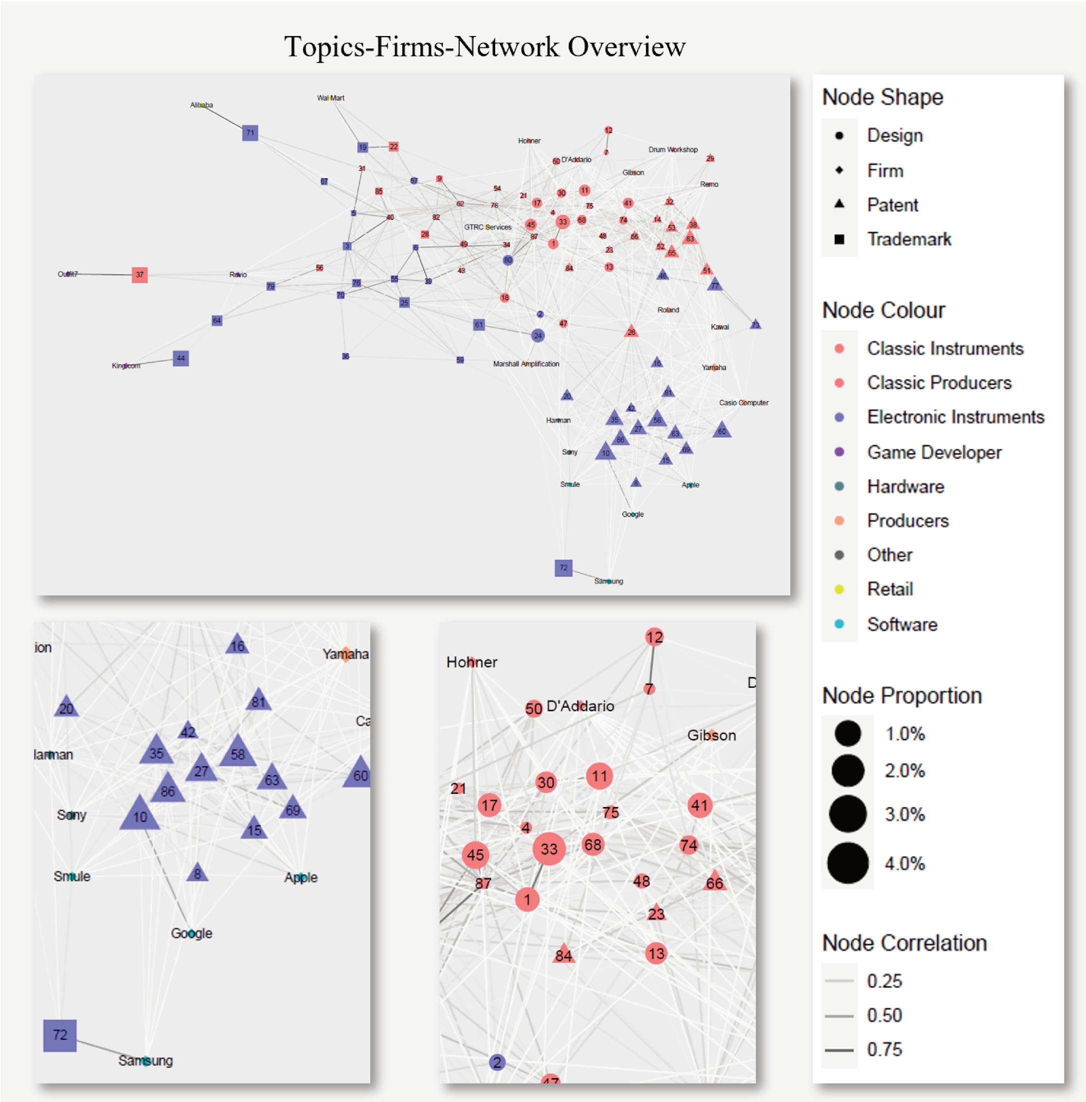


Figure 5.4: Part II: Illustrative Firm Analysis Overview.
The figure provides an example of the networks in Section 5.5. The network can be seen in closer detail in Figure 5.23. The topics-firms-network contains several elements. The nodes relate to topics and to firms. As several dimensions are shown in the network, the node shapes, colours and labels are used to separate the different relations. The firm nodes are labelled with the firm name. Firm nodes have a rhombus shape. Their colouring depends on the firm background. The background is separated into classical producers, game developers, hardware, producers, other, retail and software. Classical producers and producers are separated, as the former only produce classical musical instruments, while the latter also produce electronic instruments. The topic nodes are labelled with the topic number. The node shape of the topic is round for designs, triangular for patents or rectangular for trademarks. The topics have been coloured either blue for electronic instruments or red for classical instruments. The size of the nodes in both cases indicates the proportion of the topic or firm in the model. Direct links exist between topics or between topics and firms. A link between two topics is determined based on the co-occurrence of the topics and a positive correlation between these topics. A link between a topic and a firm is based on the occurrence of IPR documents of the firm in a topic and the positive correlation between these.

The network is extended with firm information to answer research questions on firm involvement (see Figure 5.4). [7] To cover the sector's transformation, the different networks are analysed from 1986 to 2015 and 5-year intervals. The analysis can be found in Section 5.5.

5.3.2 Data Extraction

For the analysis, textual data of patents, designs and trademarks are extracted from the USPTO. The data extraction is based on a combination of official classifications of musical instruments and keyword-based searches. Various queries identify the data related to musical instruments. For the official classifications, classes of musical instruments are of interest. For the keyword-based searches, general terms of musical instruments, as well as specific terms of musical instruments like organs, guitars or drums, are combined.

Classification of Musical Instruments In the context of musical instruments, the Hornborstel-Sachs classification is of relevance (Dräger 1948): In 1914, von Hornbostel and Sachs proposed the classification of musical instruments along four dimensions of tone generation: idiophones (vibration of the instrument itself), membranophones (vibrating membrane), chordophones (vibrating strings) and aerophones (vibrating air) (von Hornbostel and Sachs 1914). Idiophones and membranophones are occasionally grouped together under percussion instruments (Marcuse and Bowles 2015). Sachs (1940) later mentions the fifth dimension of electrophones with either electric sound creation or transmission. The Musical Instruments Museums Online (MIMO) project updated the classification in 2011: Especially electrophones are further divided into electro-acoustic instruments, electromechanical instruments, analogue electronic instruments, digital instruments, hybrid configurations and software (MIMO-Consortium 2011). The Hornborstel-Sachs classification continues to be used to organise musical instruments in museums and, as such, remains a critical reference system to this day (Koch and Kopal 2014). Patents, trademarks, and designs have their classification systems. However, the DPMA aligns the patent and the Hornborstel-Sachs classification and assigns the existing patent classes to chordophones, aerophones and percussion instruments. In this, percussion instruments cover idiophones and membranophones together (DPMA 2021a). The different available classifications are considered during the data extraction.

General Data Extraction The data extraction is performed in several steps. First, a general search is performed for all three intellectual property rights based on the IPR's classification systems, supplemented by a general search on "musical instruments" or "music instruments".

Patents: CPC[8] class G10 contains patents that are classified to relate to musical instruments. The G10 class is subdivided into the groups G10B to G10L mainly along the dimension of tone generation (USPTO 2021a). The classification of DPMA (2021a) provided for chordophones, aerophones and percussion instruments, as well as electrophonic and automatic musical instruments, serve as a basis for the patent data extraction (see also Appendix, Query D.1). The extraction is enhanced with all patents included in CPC classes "G10B", "G10C", "G10D", "G10F", "G10G" and "G10H" and supplemented with keyword-based searches for "%musical instrument%" or "%music instrument%" (see Appendix, Query D.2 and Query D.3). All patents are granted.

Trademarks: The trademark data consists of trademarks extracted from the musical instrument Nice class 15 and a keyword-based search of "%music%" and "%instrument%" in the trademark name or pseudo mark or "musical instrument" or "music instrument" in the trademark description (see Appendix, Query D.16). The keyword-based search in the trademark description is more strict,

[7] The analysis of S. Lee and Rha (2018) inspired the approach, as the authors display keyword and article nodes in the same network to find network structures.

[8] The "Cooperative Patent Classification (CPC)" system of the USPTO and EPO classifies patents into more than 250,000 categories (EPO 2022).

as the chance of including trademarks belonging to music and instruments but with no reference to musical instruments is higher than in patents. Nice class 15 contains trademarks concerning electronic, electric, stringed and mechanical musical instruments and musical boxes or robotic drums. It does not contain an amplification apparatus or downloadable music (WIPO 2020a). All trademarks considered are registered trademarks.

Designs: Designs related to musical instruments are also included in the analyses. U.S. designs are organised in 33 classes under which design class "D17" relates to Musical instruments (USPTO 2005). Further, class "84" of the United States Patent Classification (USPC) with the title of music focuses on musical instruments according to its description (USPTO 2012).[9] The query used to derive the designs can be seen in Appendix, Query D.9.

Musical Instrument Specific Data Extraction The data extraction is refined with musical instrument-specific words. The reasons are twofold: first, the different intellectual property rights classification systems are not aligned. By searching for a specific list of keywords, the differences are addressed. Second, the specific word list allows for the assessment of differences in the instrument related to the Hornborstel-Sachs classification as the established classification system in the field. Thus, it is interesting to assess whether the separation of the instruments in wind, string and percussion instruments helps explain the sector's transformation. The list of instruments used as a basis for the specific keyword search is based on the trademark class 15 description (WIPO 2020a). The class description contains specific instruments like accordions, harps, kettledrums, or saxophones. The instruments mentioned are manually assigned to the classes of wind, stringed or percussion instruments and used as a basis for identifying patents, trademarks, and designs in the specific classes. The instruments associated with each class can be found in Table 5.2. In the case of patents, the wind instruments identified by terms are supplemented with patents in CPC class G10B. Analogue for stringed instruments in CPC class G10C (Query D.1).

The data extraction is done prudently. All included documents fulfil the criteria. Nevertheless, the data extraction approach includes producers of classical or electronic musical instruments and other firms, for example, gaming firms. Although these firms are not immediately associated with musical instruments, the text-based analysis reveals that they also contribute to transforming musical instruments (seen in Section 5.5). It is in line with the report of SoundDiplomacy (2021), which points out that apart from classical and electronic musical instrument manufacturers, e.g. audio equipment and loudspeaker manufacturers are active in the field of musical instruments and develop amplifiers for musical instruments. As seen in Table 4.1, amplifiers are part of the high technology activities of the sector. Thus, firms with diverse backgrounds likely contribute to the transformation of musical instruments.

Electronic and classical musical instruments The data set is further separated into classical and electronic musical instruments to cover the technical transformation in musical instruments, especially electronic instruments. The electronic musical instruments are identified based on the data sets created for musical instruments and their subgroups of wind, stringed and percussion instruments. Therefore, '%electr%' is searched in the title, claim, abstract or statement depending on the related IPR (see also Query D.7, Query D.14, Query D.21). Utility patents in class "G10H" are also associated with electronic musical instruments (Query D.1). All musical instruments not included in electronic musical instruments are assigned to classical ones. Therefore, each document is either assigned to the electronic or the classical data set (Subsubsection D.1.1, Subsubsection D.1.2, Subsubsection D.1.3).

[9] The USPC classification is used in the U.S. for design and plant classifications (USPTO 2021b).

Table 5.2: Musical Instruments Terms in the Hornborstel-Sachs Classification

Instrument Classification	**Hornborstel-Sachs**	**Related Instrument Terms**
Wind instruments (Appendix, Query D.4, Query D.17, Query D.10)	Aerophones	aerophones, wind instruments, bagpipes, bamboo flutes, bandonions, barrel organs, buccins [trumpets], clarionets, clarions, concertinas, cornets [musical instruments], flutes, harmonicas, harmoniums, horns [musical instruments], melodicas, oboes, ocarinas, organs [musical instruments], saxophones, sheng [Chinese musical wind instruments], suona [Chinese trumpets], trombones, trumpets
String instruments (Appendix, Query D.5, Query D.19, Query D.11)	Chordophones	chordophone, string/ stringed instruments, balalaikas [stringed musical instruments], banjos, basses [musical instruments], double basses, guitars, harps, huqin [Chinese violins], Jews' harps [musical instruments], lyres, mandolins, pianos, pipa [Chinese guitars], stringed musical instruments, violas, violins, zithers
Percussion instruments (Appendix, Query D.6, Query D.20, Query D.12)	Idiophones and Membranophones	percussion instruments, idiophone , membranphone, accordions, carillons [musical instruments], castanets, cymbals, gongs, handbells [musical instruments], triangles [musical instruments], xylophones, kettledrums, drums, robotic drums, tambourines, tom-toms

The table offers an overview of common musical instrument classifications, the Hornbostel-Sachs classification, and related instrument terms used for data extraction. These terms are obtained from the Nice class description, which lists related musical instrument terms. The terms are then categorized according to the Hornbostel-Sachs classification, forming the basis for differentiating between wind, string, and percussion instruments.
Source: Own instrument extraction based on WIPO (2020a). Since not all terms are unambiguous for musical instruments, the corresponding ambiguous terms were supplemented in the queries with a reference of the text to music.

Table 5.3: Overview of the Data Set of Musical Instruments for Structural Topic Modelling

Data Set	Musical Instruments	
	Original	**Cleaned**
Patent Documents	15,969	15,969
Trademark Documents	19,891	19,208
Design Documents	2,368	2,350
Total Documents	**38,228**	**37,527**
Terms	30,908	11,576
Selected Topic Number		**87**

The table offers an overview of the documents and terms used as the foundation for structural topic modelling, along with the selected number of topics based on various metrics. It displays the initial number of documents extracted from various data sources, as well as the document numbers provided to the model after data preprocessing and cleaning. The time-frame is limited to the years 1978 through 2016.

Musical Instruments The final data set of musical instruments consists of all individual class and keyword-based searches combined. The data is limited to the years 1978 to 2016.[10] For the topic modelling, as much data as possible are considered. The resulting data can be derived from Table 5.3. To prepare the data sets for Structural Topic Modelling, English stopwords[11], words that occur only in one document, and words that occur in equal or more than 50% of all documents and per IPR type are removed. Only unigrams are considered. During preprocessing, common words such as “music” and “instrument” are removed, which can lead to empty documents if only, e.g., “musical instrument” is given as the textual description. In this case, the document as such is removed from the analysis. The total document number after preprocessing is, therefore, smaller than before. Primarily trademark and design documents are removed. After preprocessing, 37,527 documents remain with 11,576 related terms. Further information on the data set is provided in Section 5.4.

5.3.3 Model Selection

The topic number is an input for the model estimation. To select a fitting topic number and select a model, the model is run several times for different topic numbers. Different metrics support the model selection. These are semantic coherence, exclusivity, held-out likelihood and residuals. *Residuals* reveal the deviation between prediction and observation. *Held-out likelihood* shows the model’s performance on previously held-out documents. *Exclusivity* provides information on the common use of the words of a topic compared to other topics. *Semantic coherence* provides information on the relatedness of the words of the topic to each other (Roberts et al. 2014). Figure 5.5 provides the metrics to decide the topic number. The model is estimated for topic number k between 10 and 150. The residuals are lowest between 70 and 90. A trade-off must be made between semantic coherence and exclusivity, as these two metrics oppose each other. Higher exclusivity leads to lower coherence and vice versa. Yet, the trade-off is still different between models with different topic numbers. Based on the overview of trade-offs (see Figure 5.6a) and models’ metrics, some models are selected for a closer view. For musical instruments, these are model 81, 83, 84, 86, and 87. 83, 84, and 87 have low residuals, while 83 and 86 display a good exclusivity-semantic coherence trade-off. For these models, the semantic coherence and exclusivity

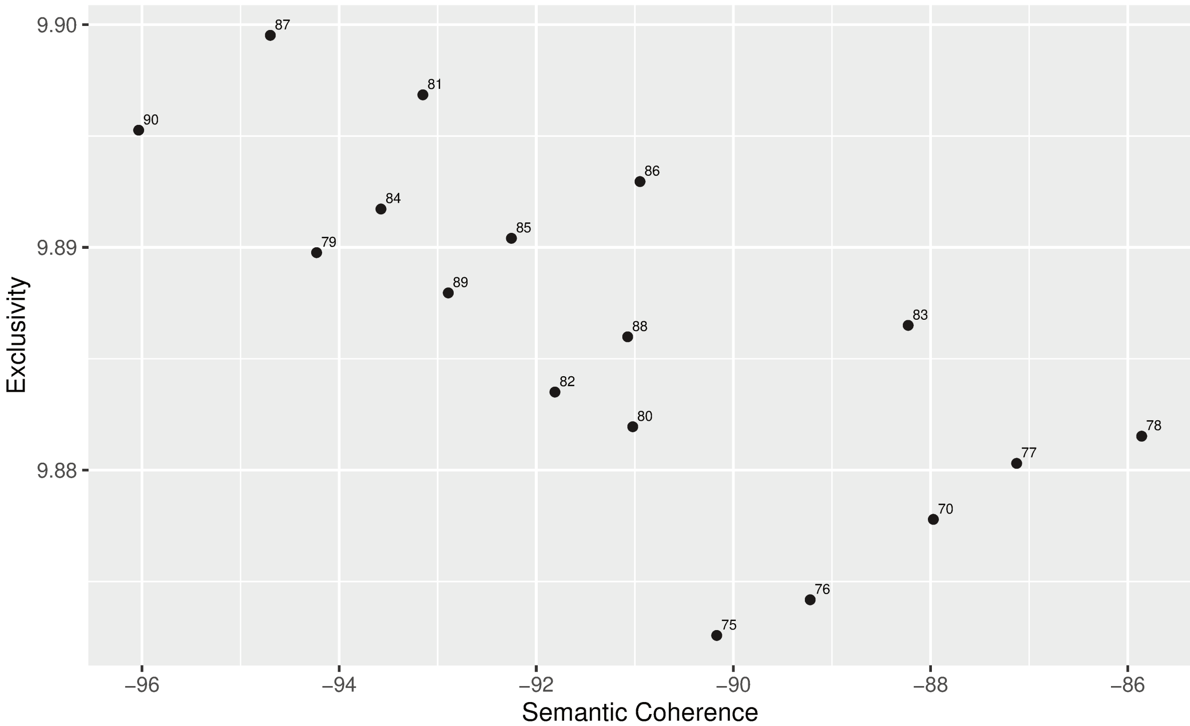
[10]This is due to data availability. Trademark data coverage is limited before 1978 according to S. J. H. Graham et al. (2013). After 2016, the USPTO patent and design data coverage are not complete.

[11]The stopword list is taken from stop_words from the R package *tidytext*. The list combines the sources “SMART”, “snowball” and “onix”.

trade-off on a per-topic level is considered in Figure 5.6b. The model with 87 topics is chosen for further analyses due to the models' mean and a good exclusivity-coherence trade-off with acceptable levels of exclusivity while semantic coherence of various topics remains high. The residuals are lowest in model 83. However, model 87 has a better trade-off of coherence and exclusivity. The topic proportions of model 87 can be derived from Appendix, Figure D.1.

Figure 5.5: Metrics to Evaluate the Number of Topics K in Musical Instruments. The general diagnostics offer an initial overview to determine the appropriate number of topics (K). For each calculated model, metrics such as exclusivity, held-out likelihood, residuals, and semantic coherence are assessed. Each point on the graph represents a distinct model and its corresponding diagnostics. These metrics are evaluated in relation to one another to decide on the optimal K. Specifically, in the case of residuals, models located at the lowest point of the slope are of particular interest. The figure presents both (a) an overall view and (b) a closer, zoomed-in perspective to facilitate the selection of models that warrant closer examination.

(a) Mean Exclusivity-Semantic Coherence Trade-off of different Models



(b) Exclusivity-Semantic Coherence Trade-off of the Topics of Selected Models

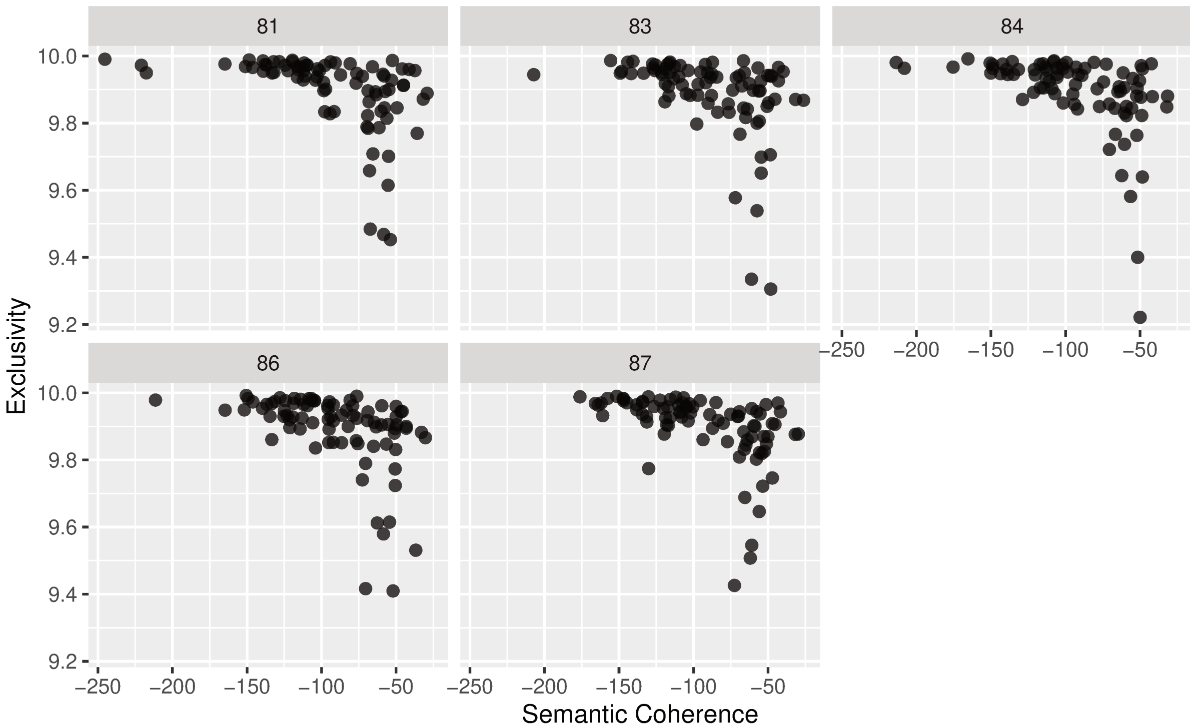


Figure 5.6: Exclusivity and Semantic Coherence of Musical Instrument Models in Comparison. A trade-off between exclusivity and semantic coherence needs to be determined. Models with fewer topics exhibit higher semantic coherence but lower exclusivity compared to models with more topics. Subfigure (a) presents the mean exclusivity against the mean semantic coherence for each model, with each point representing a model and its respective trade-off. Potential models are selected for further examination. Subfigure (b) illustrates the exclusivity-semantic coherence trade-off for these chosen models on a per-topic basis. In Subfigure (b), each frame corresponds to a model, and each point represents the trade-off for a specific topic within the model.

5.4 Part I: Transformation of the Musical Instrument Sector

The results of the analyses are initially presented regarding IPRs (Subsection 5.4.1), followed by the Hornbostel-Sachs classification (Subsection 5.4.2), and finally, in the context of electronic and classical instruments (Subsection 5.4.3). The analyses build on each other and are supplemented by the information identified as relevant. The analysis period covers the years from 1986 to 2015.

5.4.1 Intellectual Property Rights

This subsection provides a general overview of the data set before presenting the topic modelling results. Thereby, the topics in general and their consistency is of interest. Then a network perspective is taken.

IPR Overview From 1986 to 2015, in total, 31,928 documents are covered with 13,434 patents, 1,920 designs and 16,574 trademarks. The development over time is displayed in Figure 5.1. Overall, the yearly registrations have increased from 1986 to 2015. Patents and designs are relatively stable: patent registrations reach around 500 registrations per year, while design registrations are constant at below 100 registrations per year. Trademark registrations display an interesting trend: the registrations were below patent registrations before 2000 but surpassed patent registrations since 2004 and have since increased. This is in contrast to expectations. The patents mainly relate to CPC class G10 “Musical instruments, Acoustics” and especially to G10H “Electrophonic Musical Instruments” and G10D with, among others, “Stringed Musical Instruments” and “Percussion Musical Instruments” (see Appendix, Figure D.2b). Most trademarks are related to Nice class 15 that contains, e.g., “musical instruments” or “music stands” (see Appendix, Figure D.2c). D17 as the “musical instrument class” is the most important design class (see Appendix, Figure D.2a).

IPR Topic Modelling Results The 87 Structural Topic Modelling topics can be associated with different intellectual property rights (see Appendix, Figure D.3). For better readability, Figure 5.7 displays only the ten main topics per intellectual property right. As the topic estimation is based on the underlying documents, each topic can be related to each document. Every topic is represented in every document, but the degree of the relation varies. At the same time, every document is a particular intellectual property right, either a patent, a trademark or a design. The ratio of each IPR, in general, can be determined and related to the topic. If a topic is not among the main ten topics of that intellectual property right, the proportion is not displayed even though it exists. The three main topics are topic 33 (guitar, electric, pick, strap, tuner), topic 45 (keyboard, violin, box, organ, tune) and topic 26 (provide, system, invention, produce, embodiment).

Topic 33 has the highest occurrence probability in the data set. This topic is highly relevant in designs and trademarks and among the main ten topics of these data sources. Patent documents only have a very low associated topic occurrence probability, and the topic is, therefore, not among the ten main topics of patents. Content-wise, the topic relates to (electric) guitars.

Topic 45 is the second most important topic, and relates to violins and includes synthesisers. This topic is also among the main ten topics of trademarks and designs but not of patents. Both topics relate to the final products of musical instruments.

Topic 26 is the third most important topic and is the first topic with a high patent occurrence probability. It seems to relate to patent-specific wordings in general, like ‘provide’, ‘method’ or ‘disclose’, which are not specific to musical instruments. The occurrence probability of designs and trademarks is low; therefore, these are not among the main ten topics of these data sources. Only the patent occurrence probability is displayed in the overview of the main topics of each IPR.

This pattern is generally observable: The main patent topics are disjunct from the main trademark and design topics. Trademarks and designs share six out of ten topics that are among the most important topics for these IPRs, while patent topics are disjunct (see Figure 5.7).

Patent topics are associated with tone generation and data (topic 60), support mounts (topic 38), digital tone generation and signal processing (topic 27), data information storage (topic 86), audio and media content (topic 10) or frequency and wave amplitude (topic 69). Designs and trademark topics are related to guitars and accessories (topic 33), keyboards, violins, and synthesisers (topic 45), instrument boxes for transportation and stands (topic 17), acoustic amplification (topic 1), strings and bows (topic 68) and sound effects (topic 24). Designs additionally relate to drums and pedals (topic 11), electronic apparatus (topic 80), percussion, cymbals, drumsticks (topic 30) and piano hammers (topic 13). At the same time, trademarks include toy and gameplay (topic 57), audio speakers (topic 61), retail and sale stores (topic 22) and entertainment and performance (topic 28). The patent topics relate to digital or electronic music generation, music storage or data processing. Design topics are more about the instruments and include electronic or acoustic elements. Trademarks provide an additional perspective on the use of musical instruments in child-play and entertainment or the offerings of musical instruments in retail. They also include instruments, their components and acoustic or audio aspects.

Each topic is related to each intellectual property right. However, the consistency differs: E.g., patent topic 60 has a high consistency among the main documents of patents, trademarks and designs with tone generation and data as the main topic (see Figure 5.8). The figure shows the topic number and the topic's main words on the top. Underneath, three main documents for each intellectual property right are displayed, with the document number, a snippet of the document text as well as the topic distribution of the document. Topic 19 and Topic 76 are both mostly in trademarks. However, their consistency differs: Topic 19, which relates to accessories, equipment, and supply, includes patents on music generation, storage and placement, while the trademarks are focused on electronic retail and designs cover the cases for guitars (see Appendix, Figure D.4). The trademark documents are very focused on that topic and contribute the most to the supply aspects, while the other IPR documents are only partly focused on this topic. Topic 76 also highly relates to trademarks. It contains gaming aspects in the main trademarks, patent and design documents and can be considered consistent (see Appendix, Figure D.5. The electronic aspects are more in the latter than in the former topic.

IPR-Topic-Network The topics are connected via their common document occurrence. Figure 5.9 displays the topic connection. The node size shows the topic proportion per topic. To ensure readability, the edges displayed are larger than a threshold of 0.08 points. The pattern remains the same with a lower threshold. The colouring indicates the strength of the correlation, with darker colours symbolising a higher correlation. The node colour further symbolises the IPR type with the highest topic occurrence probability in the data set. As can be seen, the topics with the same major IPR form groups. Designs thereby separate patent and trademark topics. The relation of topics to intellectual property rights is thus considered an important aspect to consider to understand the sector transformation.

To summarise, patents, trademarks and designs cover separate aspects of innovation in musical instruments. The main trademark and design topics overlap, while patents are separated. Design topics occur in the networks mainly in between patents and trademarks topics, bridging the information of these two data sources.

5.4.2 Hornborstel-Sachs Classification

In this subsection, first, an overview of the Hornborstel-Sachs classification is provided. The Topic Modelling results are then presented before analysing the topics-network in relation to the Hornborstel-Sachs classification.

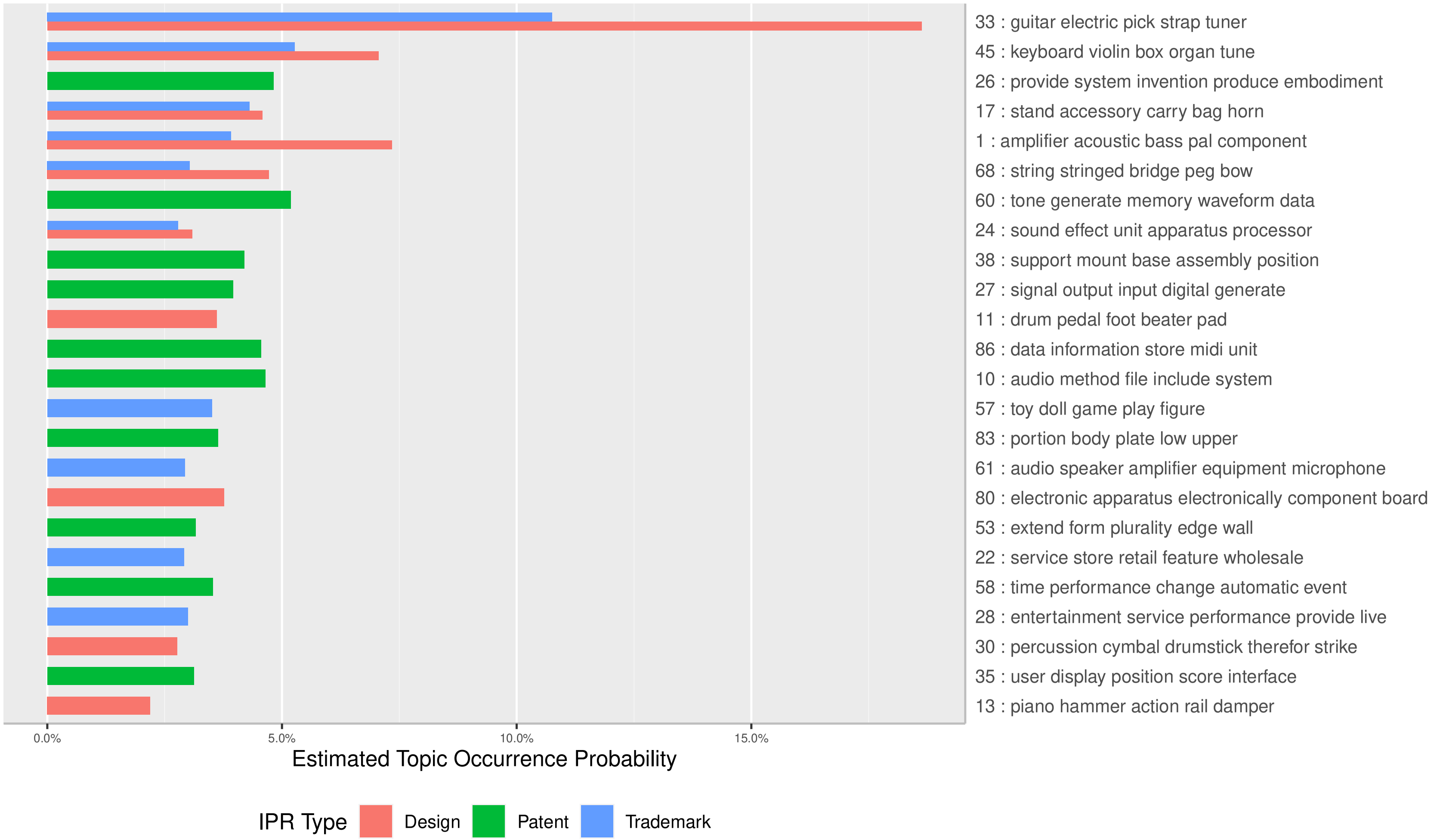


Figure 5.7: Overview of the Ten Main Topics per IPR Type in Musical Instruments.
The figure presents the ten main topics per intellectual property right (IPR type), including trademarks, designs, and patents. The right-hand y-axis displays the topic number and the main words associated with each topic. It also illustrates the topic proportion within each IPR type. If a topic is not within the main ten topics for a specific IPR type, it is denoted by a missing bar, indicating that the proportion for that topic is not displayed.

Table 5.4: Overview of the Data Set according to the Hornborstel-Sachs Classification

Hornborstel-Sachs Classification	IPR Type			Total Documents
	Design	Patent	Trademark	
Generic	919	8,323	6,681	15,923
Percussion	101	972	2,268	3,341
String	825	3,680	7,920	12,452
Wind	48	604	1,427	2,079

The table offers an overview of the number of documents related to various IPR types and the Hornbostel-Sachs Classification from 1986 to 2015. A document can be associated with multiple Hornbostel-Sachs Classes simultaneously, depending on the referenced instruments. Documents lacking a classification are categorized under “Generic”.
Source: Own representation based on the data extraction. The categorisation is determined based on the search queries provided in Appendix, Query D.6, Query D.5, Query D.4, Query D.12, Query D.11, Query D.10, Query D.20, Query D.19 and Query D.17.

Hornborstel-Sachs Classification Overview For the analysis of the Hornborstel-Sachs classification, each document was classified according to the terms provided in Table 5.2. If a document includes terms in more than one class, the document is assigned to both. If no specific Hornborstel-Sachs term is available in the document, the document is considered a generic document in the data set. An overview of the data set is provided in Table 5.4. Most documents belong to the generic category, followed by string instruments with 12,452 documents. Per year, the instruments in the categories “generic” and “string” contribute the most to the data set. The documents of every category increase over time (see Appendix, Figure D.6). String instruments especially relate to trademark documents, while patents and trademarks mainly cover the generic category. Over time, string instruments mainly increase due to trademark registrations, while patents and trademarks increase simultaneously in Generics. Most patent documents are in the generic category, followed by the category of string instruments. The same applies to design protection. Design protection is overall low in numbers (Figure 5.10).

Hornborstel-Sachs Classification Topic Modelling Results The network relation of the topics to the Hornbostel-Sachs classes is displayed in (see Appendix, Figure D.7). The topic descriptions are partly in line with the mainly associated Hornbostel-Sachs class. E.g., topic 33 relates to guitars and is highly associated with string instruments; topic 17 relates to flutes, horns, saxophones and trumpets and is highly in wind instruments. The words describing topic 45 mainly consist of violins, violas and bass instruments which belong to string instruments. Document-wise, however, wind instruments are the strongest, followed by percussion and string instruments.

Hornborstel-Sachs-Topic-Network The topic-network is displayed in Figure 5.11. Each topic is coloured according to the category with the highest occurrence probability. The colours seem randomly assigned, and no pattern becomes directly observable. The only remarkable aspect is the cluster of generic topics centred around topic 58. These topics relate mostly to patent documents, as seen in Figure 5.9. Content wise, these topics are concerned with e.g., “time”, “performance” (topic 58), “control”, “circuit”, “switch” (topic 81), “tone”, “generate”, “data” (topic 60), “signal”, “output” or “digital” (topic 27). All these terms relate to the digital processing of tone and data. This, however, seems to be a general aspect of the development within musical instruments that is not specifically related to any Hornborstel-Sachs class.

The Hornborstel-Sachs classification into wind, string and percussion instruments does not contribute to a better understanding of the transformation of the musical instrument sector. The generic topics are

Table 5.5: Overview of the Data Set, Differentiated for Electronic or Classical Musical Instruments

Musical Instrument Type	IPR Type			Total Documents
	Design	Patent	Trademark	
Classical	1,548	5,412	10,394	17,354
Electronic	372	8,022	6,180	14,574
Total Documents	1,920	13,434	16,574	31,928

The table offers an overview of the number of documents associated with different IPR types and either electronic or classical musical instruments from 1986 to 2015. A document is categorized as either related to the former or the latter but not both simultaneously.
Source: Own representation based on the data extraction. The categorisation is determined based on the search queries provided in Appendix, Query D.7, Query D.14 and Query D.21.

important in terms of data and signal processing. This is, however, not related to a specific instrument type. As Hornborstel-Sachs Classification is not decisive for the topic relations, the classification is thus not the focus of further analyses.

5.4.3 Electronic and Classical Musical Instruments

In the subsection, electronic and classical musical instruments are analysed. The technological transformation in musical instruments is leading to a change from classical to electronic to digital musical instruments. Consequently, this should also be reflected in using IPRs with patents being applied more. The subsection first provides an overview of electronic and classical musical instruments. Then the topics are analysed before taking a network perspective.

Electronic and Classical Musical Instruments Overview In recent years, electronic instruments have surpassed classical instruments (Figure 5.12a). This trend is, however, not due to patents but due to an increase in trademarks in electronic instruments. Until 2000, primarily patent documents are related to electronic musical instruments. This has changed since 2000 with a rapid increase in trademark registrations. Since 2009, trademark registrations in electronic instruments have surpassed patents. Electronic instrument trademarks even surpass classical instruments' trademarks in recent years (Figure 5.12b). Overall, most documents still relate to classical instruments, with 17,354 documents compared to 14,574 documents on electronic musical instruments (Table 5.5). In total numbers, patent applications still surpass trademarking in electronic musical instruments. However, if the trend of increased trademarking related to electronic instruments continues, trademarks will surpass patents. In classical instruments, trademarks and design registrations are strong. Subdivided according to the official classification of the respective IPR types, classical instruments in trademarks are predominantly in class 15, "Musical instruments" (see Appendix, Figure D.2c). Electronic instruments are mainly in class 15 and class 9 "Electrical and Scientific Apparatus". Class G10H is inherently the main class in electronic instruments and is exclusively assigned to electronic instruments in patents. G10D is strongly represented in classical instruments and refers mainly to string instruments (see Appendix, Figure D.2b). Interesting is also the electronic instrument class G06F with the topic of digital data processing. Designs of classical and electronic musical instruments are represented primarily in class D17 "Musical Instruments" (see Appendix, Figure D.2a). Here, already further details can be gained. Apparently, data and digital plays a role in musical instruments and electronic aspects in general. Trademarks and patents capture this phenomenon. Further details are expected from the topic analysis.

Electronic and Classical Musical Instruments Topic Modelling Results The ten main topics in electronic instruments, displayed in Figure 5.13, cover hardware aspects like tuners (topic 33) and speakers (topic 61), but also data processing aspects such as memory (topic 60), sound effect (topic 24), signal (topic 27) or data (topic 86). The data processing aspect is in line with the patent classification G06F, which also covers digital data processing. Classical topics are more directly related to music instrument aspects themselves, like guitars (topic 33), keyboards (topic 45), strings (topic 68), support mounts (topic 38) or pedals (topic 11). Only topics 33 and 26 are among the main topics of electric and classical musical instruments. Topic 33 relates to guitars which are often also electric, so these terms occur together in the topic. Topic 26 is a more general topic with terms related to patent documents.

Electronic and Classical Musical Instruments Networks A network perspective is taken to understand the relation of topics and the influence of classical and electronic musical instruments (see Figure 5.14). The nodes of the network display the IPR type that is strongest in the underlying topic. The links represent the correlation of the topics. The node's colouring represents the topic's relation to classical or electronic musical instruments. The stronger the topic is in electronic compared to classical instruments, the more intense the blue. The reverse is valid for the classical instruments with the red colouring. To calculate the colouring, the topic's relation to electronic or classical instrument documents is decisive. Based on the documents, the topic's share in electronic and classical documents is determined. The difference between these shares then determines the intensity of the colouring. The blue is particularly intense around the topics 60, 10, 86, 27 and 58, which appear together and are all represented among the ten main electronic musical instrument topics. The same applies among others to classical topics 33, 45 and 17. Topic 26 is equally represented in electronic and classical instruments (as already seen in Figure 5.13). It is correspondingly neutral in colour. In terms of content, it is more concerned with methodological aspects and combines classical with electronic topics. In total, 46 topics tend to be classical, and 41 topics tend to be electronic in the overall time period. Visual inspection of the graph reveals that topics related to electronic instruments and topics related to classical instruments occur together. The electronic instruments around topic 60 are, at the same time, dominantly in patent documents. Clusters are determined via random walk calculation to verify the visual inspection. The results are visible in Figure 5.15. The calculation confirms the grouping of electronic topics around topics 58 and 69 (see cluster 5). Cluster 10 evolves around the topics 52 (e.g., tune, block, adjustment) and 53 (e.g., extend, form and plurality). They belong to classical instruments but are less pronounced than, e.g., topic 33. Topic 33 and topic 1, however, belong to Cluster 2, wherein also some electronic topics belong, like topic 61. Content-wise, this makes sense as topic 33 relates to guitars and electric guitars, topic 1 relates to amplifiers and acoustic basses, and topic 61 revolves around amplifications and audio speakers. The topics are thus content-wise connected, and these instruments and components are used together. The networks are additionally enriched with the most important IPR type (related intellectual property right) per topic. The network perspective displays that in the overall period, the most important electronic topics are, at the same time, predominantly located in patents. Electronic topics are also represented in trademarks but to a smaller extent. Clusters 5 and 10 are primarily in patents, while cluster 1 relates to trademarks and designs. The separated perspective on classical and electronic instruments documents reveals that the separation of patents, trademarks and designs is maintained (see Appendix, Figure D.10 and Figure D.11). Design topics are more prevalent in electronic than in classical instruments. In addition, trademarks are more dominant in classical instruments. Overall, the cluster reveal groups of topics in the network.

In summary, the topics are separated into classical and electronic musical instruments. The topics cover hardware for sound generation like speakers and tuners, software aspects like data processing and signals and components of instruments like keyboards, strings, pedals and guitars. The main topics align with the classification of the IPRs, here, for example, G06F data processing. The network revealed a relationship between patents and electronic instruments and trademarks and classical instruments. To conclude, electronic or classical musical instruments build clusters in the topic network that explain the network from a content perspective. The clusters further relate to specific IPRs, like the topics of the electronic cluster are dominant in patents. The cluster overview highlights topics that are strongly connected to each other.

5.4.4 Transformation within the Sector

Until now, the complete period from 1986 to 2016 has been covered. Intellectual properties contribute different information to the analyses, as the topics are represented to a different degree by the different IPRs. The network perspective of topics and IPRs revealed a clear separation between the different IPRs and related topics to each IPR. The perspective on the Hornborstel-Sachs classification revealed that a generic cluster is observable that does not link to a specific instrument category. However, the string, wind and percussion classifications did not further contribute to understanding the network structure. The separation of classical and electronic musical instruments revealed a clustering pattern related to the intensity of technological components. The clusters overlap with intellectual property use. From the transformation perspective, these are already interesting insights, as the technology intensity is decisive for clustering and the combination of different IPRs contributes different aspects to the development.

To further assess the aspect of transformation, a time perspective is needed. The data set is therefore separated into five-year intervals from 1986 to 2015:

1986–1990: Three clusters are visible in this interval: A cluster around topic 33, one around topic 60 and one around topic 83 (see Figure 5.16). Topic 33 is a major topic in classical musical instruments. It covers guitars, electric, picks and straps. It is covered in designs and surrounded by topics 45 (keyboard, violin), 68 (strings, bow) and 17 (stand, accessory). Topic 45 and 17 are trademark topics. The cluster is around instruments with a focus on string instruments. The topics are all related to designs or trademarks. The patent documents only have small proportions in these topics. Another cluster evolves around topic 60[12], tone generation and data. Topic 60 is the major topic in electronic instruments. It is surrounded by other electronic topics like 81 (control, circuit), 27 (signal, output) and 69 (frequency, components). These are all patent topics. The cluster covers aspects of sound generation, like components of amplifiers and sound interface circuits. A third cluster is visible around topic 83 (portion body plate), 53 (extend form plurality), and 38 (support mount base) with a light relation to classical instruments. These topics cover aspects of instrument assembly. The remaining topics are mainly small and without a clear relation. In terms of innovation, it can be said that the three clusters around sound generation, instrument assembly and classical instruments are still quite separate from each other. Topic 68 around strings connects the assembly and string instrument cluster, while topic 26 with provide, system, embodiment connects the assembly and the sound generation cluster in patents. This could give an indication that innovation is taking place but not yet very integrated. Further, regarding other topics covered, it becomes obvious that the main focus is on these clusters. Topics outside of these clusters remain rather small.

1991–1995: This interval covers the same clusters as before with similar insights on innovation (see Appendix, Figure D.12).

1996–2000: Here, topic 86 is becoming equally important in electronic instruments as topic 60 (see Appendix, Figure D.13). It covers data and information. Here, it is visible that a change of the subjects occurs within the technological cluster. While amplifier components are rather hardware-focused, the innovation focus changes towards data.

[12] The cluster around electronic instrument topic 60 is verified via random walk (e.g. cluster 5 in Appendix, Figure D.15, cluster 3 in Appendix, Figure D.16, and cluster 11 in Appendix, Figure D.17.

2001–2005: In the electronic patent cluster, topic 60 is still relevant, but topics 86 and 10 are now more decisive with data, information, and audio files (see Figure 5.17). Apart from the already identified clusters, the electronic instruments around topics 61, 25, 76, and 24 become more intense. The cluster evolves around audio, speakers, amplifiers, sound, effects, video, games and computers. The topics overall relate to digital sound design, which is relevant in the context of videos and gaming. These topics contrast the cluster around topic 60 in trademarks and not patents. The patents cover more the hardware components of amplifiers than the topics in trademarks, which are more related to sound design. This also connects to topics 57 and 44, linked to topic 76. These all cover games and toys, while topic 76 also covers the video and computer aspects of gaming. Noteworthy are further the classical topics 49 and 28 connected to topic 18, which cover the aspects of education, entertainment and music. The clusters are verified via random walk and displayed in Figure 5.18. In terms of innovation, it is interesting that the amount of topics associated with electronic subjects is increasing.

2006–2010: The relation of the trademark topics and electronic musical instruments increases (see Appendix, Figure D.14). As such, the topics around digital sound design and gaming increase in importance, as well as the classical instrument topics around education and entertainment. The electronic cluster around topic 60 is still important. The linkages to other electronic topics start to increase. Innovation-wise new topics become important that reveal the diversity of the development. Amplifiers are still at the core of electronic musical instruments, while digital sound design and especially gaming are more distant subjects but relevant in terms of digital music creation.

2011–2015: In the final network, the trademark topics are now more pronounced. The cluster around topics 61 (audio and speaker) and 25 (record, video, computer) are now even further related to electronic instruments with their sound design (Figure 5.19). Strongly visible is now also topic 55 (device phone battery computer camera), which is connected to topic 59 (computer software system hardware network) and topic 2 (cable electrical power electric connector). This topic apparently links to mobile phone development. Education and entertainment (topics 28 and 49) are still in classical instruments. Here again, further electronic topics increase in importance. By now, most topics have a similar occurrence probability compared to the first period. This indicates that innovation is now taking place in diverse areas related to musical instruments.

Compared to the late 1980s, the clusters around sound generation, string instruments and instrument assembly are still present from 2011 to 2015. However, over time, the topics that are strong in trademarks increase in relevance and are increasingly connected to electronic musical instruments. The trademark clusters cover digital sound design, mobile phones, and gaming elements. These are all areas of more recent technological advancement. Further, while only some topics were strong in the late 1980s, the occurrence probability of all topics in the last period is similar. Regarding innovation, the topics evolve over time. While in the beginning, mainly patent topics were strongly occurring with a focus on electronic components and assembly; the focus shifted towards trademark topics with gaming and data processing. This could indicate that the patent topics remain relevant. However, the focus on innovation in musical instruments has shifted: Patent topics still occur as a potential part of an instrument or as innovation is still ongoing. Still, from a sector perspective, other topics emerged that also shape the sector and are present more in trademarks.

Electronic Musical Instruments In terms of electronic musical instruments, it can be said that over time, the importance of electronic instruments increases: While from 1986 to 1990, 28 out of 87 topics are in electronic instruments (Figure 5.16)), 41 topics (47%) are related to electronics from 2011 to 2015 (Figure 5.19). In addition, the electronic topics are mostly represented by patent documents and centred to 1996 mainly around the topic 60 “tone generation” (Figure 5.16). From 1996 onwards, topic 86, “data storage processing”, becomes more important (see Appendix, Figure D.13). Innovation is mainly taking place concerning electronic topics. Even though electronic subjects change over time, electronic aspects become more relevant overall in the topics.

Classical Musical Instruments Correspondingly, the proportion of topics with predominantly classical instruments has decreased from 59 in 1986-1990 (Figure 5.16)) to 46 in 2011-2015 (Figure 5.19). The main clusters in classical instruments are instrument assembly in patents and (string) instruments in designs and trademarks. Later, the topics of education and entertainment become more important. Some of the classical topics change over time to electronic topics: This implies that there is a shift to more electronic or digital-related aspects within the topics. This again enforces the assumption that innovation is driven from the electronic development.

There is no thematic shift in patent-related topics. However, the patent-related cluster becomes less important as the proportion of the topics decreases overall. The cluster around trademarks covers more digital sound creation, gaming, and mobile aspects, besides entertainment and education. Over time, the related topics gained larger importance and are increasingly associated with electronic instruments. So overall, a thematic shift occurred towards trademarks: not only did their numbers increase, but they also covered different aspects than patents. Trademarks revealed that gaming and data contribute to the transformation of musical instruments, while patents covered the amplification and sound generation aspects. Patents and trademarks thus jointly contributed to the analysis and led to a broader coverage of the transformation in musical instruments.

The analysis implies the following for the research question related to the technological transformation of the sector:

RQ5.1: *What is the nature of the sector transformation from a low-technology sector to a sector with low- and high-technology?*
In the 1980s, most topics were still assigned to classical instruments. By the 2010s, this had levelled out so that electronic and classical topics contributed about half of the topics each. Thematically, aspects of digital sound creation, gaming and mobile have been added over time that were not so strongly represented in the initial phase. Sound generation, however, had already been included since the 1980s. Change in innovation is thus taking place especially in electronic and technological subjects. Classical instruments are nevertheless relevant, however the newer and the changes of topics occur in the technology sphere.

RQ5.2: *How does the transformation relate to IPR use?*
The combination of different intellectual property rights, here of patents, designs, and trademarks, captures the technological transformation of the sector. Thematically, patents cover different aspects than trademarks and designs. Trademark and design topics are often aligned, yet designs are positioned between patents and trademarks. As musical instruments became more technical and electronic, one would expect an increase in patenting (see Neuhäusler (2009)). In our case, however, trademarks become more important. In electronic instruments, trademarks even surpass patents in recent years (Figure 5.12b). This relates to the topics patents and trademarks cover: While the patent clusters remain quite stable over time with sound generation and instrument assembly, the trademark topics change and evolve towards sound creation, gaming and mobile. These aspects are not captured from the patent perspective. In the 1980s, most trademark topics were not strongly related to electronic topics and mostly to classical musical instruments. This thus changed over time. A reason for the strength of trademarks in these digital aspects might be that these are easier to protected in trademarks than in patents. This would be in line with Millot (2009) pointing out the strength of trademarks in software. Further, trademarks are important in the gaming industry in general, so that trademark protection might be more in line with the general protection strategy of these firms.

The analysis revealed that digitalisation and mobile aspects have become more relevant in the sector's transformation. These influence the technologies of sound creation and, thus, music production. Trademarks mainly cover the transformation. While the patent clusters remain stable over time, the aspects that increase in importance are introduced by the trademark analysis. The combination of trademarks, designs and patents thereby contributed to covering the broad transformation of the sector. Each IPR contributes its aspects, whereby the patent topic coverage deviates from trademarks and designs. Focusing solely on patent analysis to cover musical instruments would have overlooked the transformation. The inclusion of trademarks thus contributes to a broader understanding of innovation. The firms relating to the thematic shift remain unclear. This will be covered in part II.

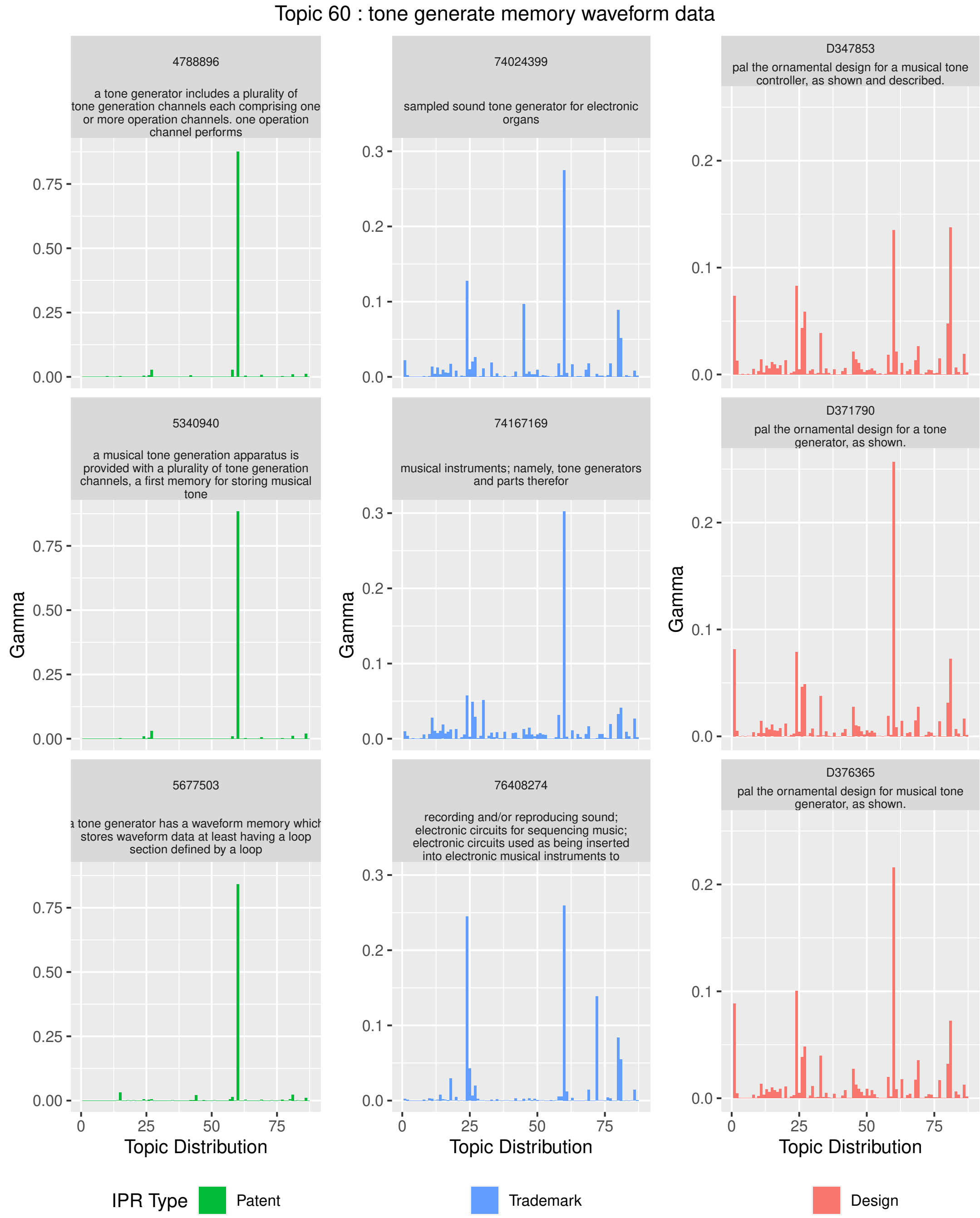


Figure 5.8: Main Documents in Musical Instruments Topic 60.
The figure displays different information on topic 60 of model 87. At the top, the five words with the highest occurrence probability in the topic are displayed. Further, a selection of three main documents per document type is displayed. The official registration number is given, and a short snippet of the textual description. The bar graph underneath each document's header displays the document's overall topic distribution. The gamma value for each topic is shown, which provides the association of a topic to the document. The gamma distribution helps interpret the data as it provides information on the level of focus on specific topics of the document.

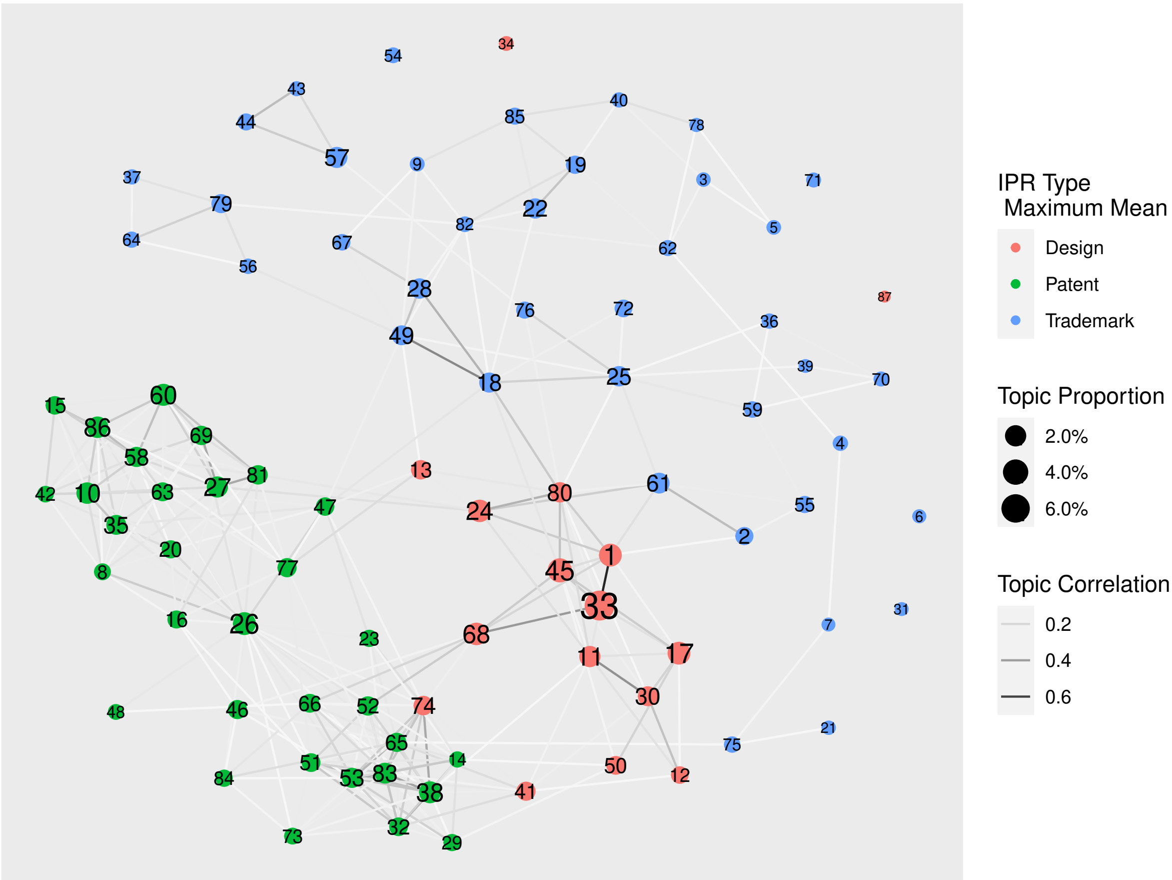


Figure 5.9: Network of Musical Instrument Topics in Patent, Trademark and Design Documents. The network displays the topic relation differentiated for the IPR types of patents, trademarks and designs. Nodes represent the topics, while links represent the likelihood of these topics to co-occurrence. The node size represents the topic proportion. Here, the node colour further represents the IPR type with the highest mean occurrence in the topic. All documents from 1986 to 2015 are considered. The colouring of the nodes represent the IPR type with the highest estimated mean occurrence probability (for further details, see Appendix, Figure D.3).

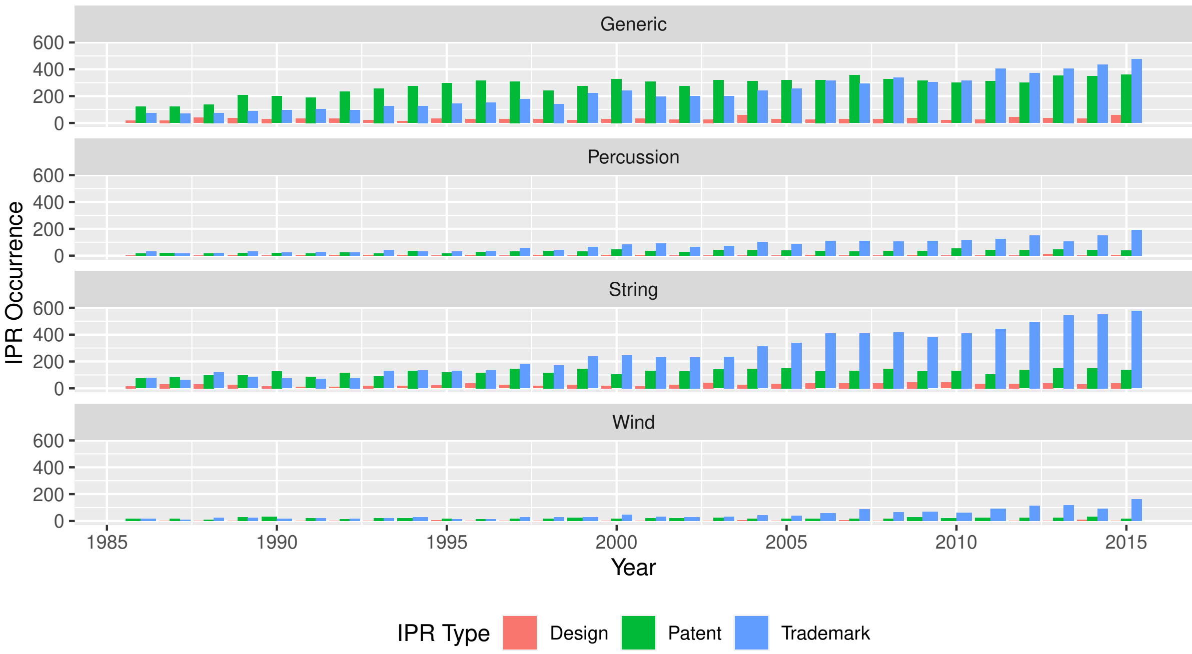


Figure 5.10: Musical Instrument IPR Registrations per Year, per IPR Type and according to the Hornborstel-Sachs Classification.
The figure displays the yearly development of IPR registrations, according to the Hornborstel-Sachs Classification, between 1986 and 2015. The generic category contains all documents that are not explicitly related to another category.

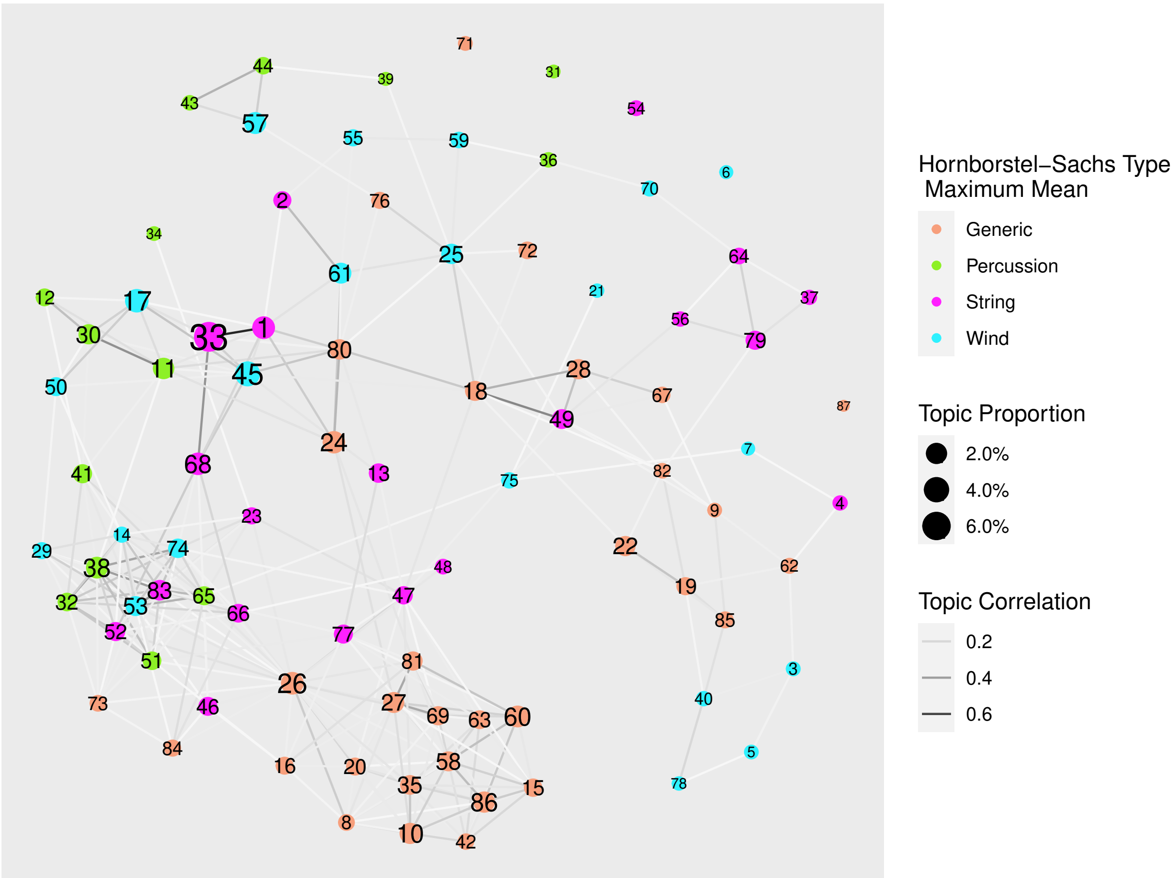
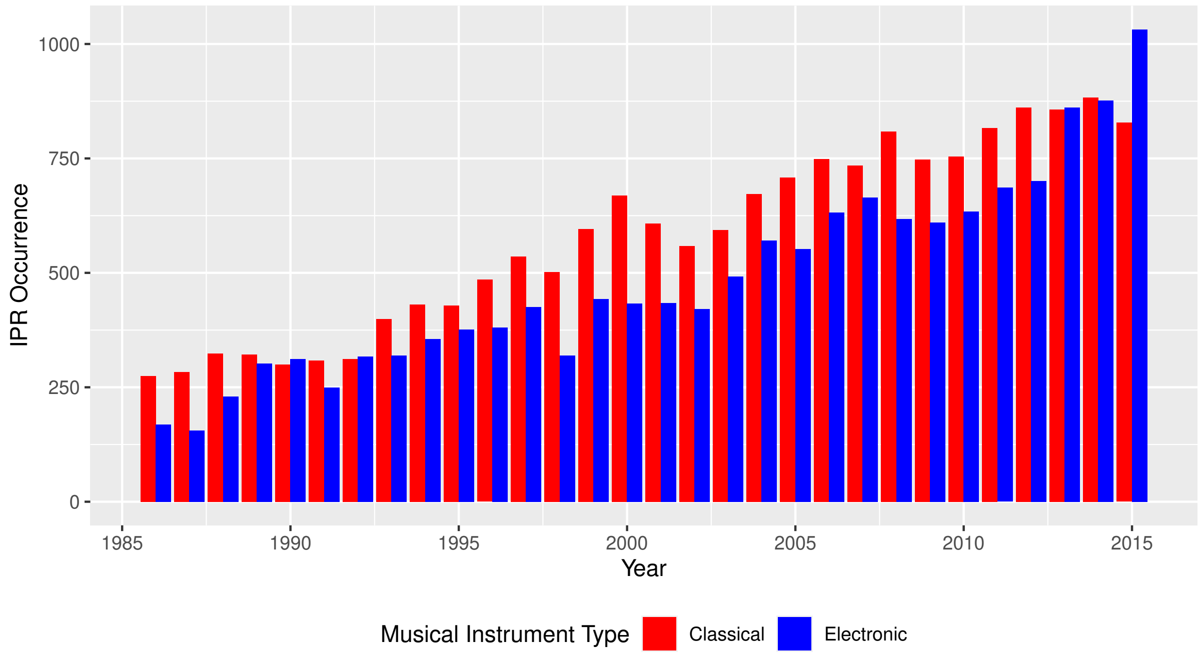


Figure 5.11: Network of Musical Instrument Topics according to the Hornborstel-Sachs Classification. The network illustrates the topic relationships categorized by the Hornborstel-Sachs Classification. Nodes represent the topics, and links depict the likelihood of co-occurrence between topics. Node size indicates the topic proportion, and node color indicates the Hornbostel-Sachs Classification category with the highest mean occurrence in the topic. The analysis considers documents from 1986 to 2015. Node color represents the category with the highest estimated mean occurrence probability (for further detail, see Appendix, Figure D.7).

(a) Document Registration per Year and Musical Instrument Type



(b) Differentiated for IPR Types

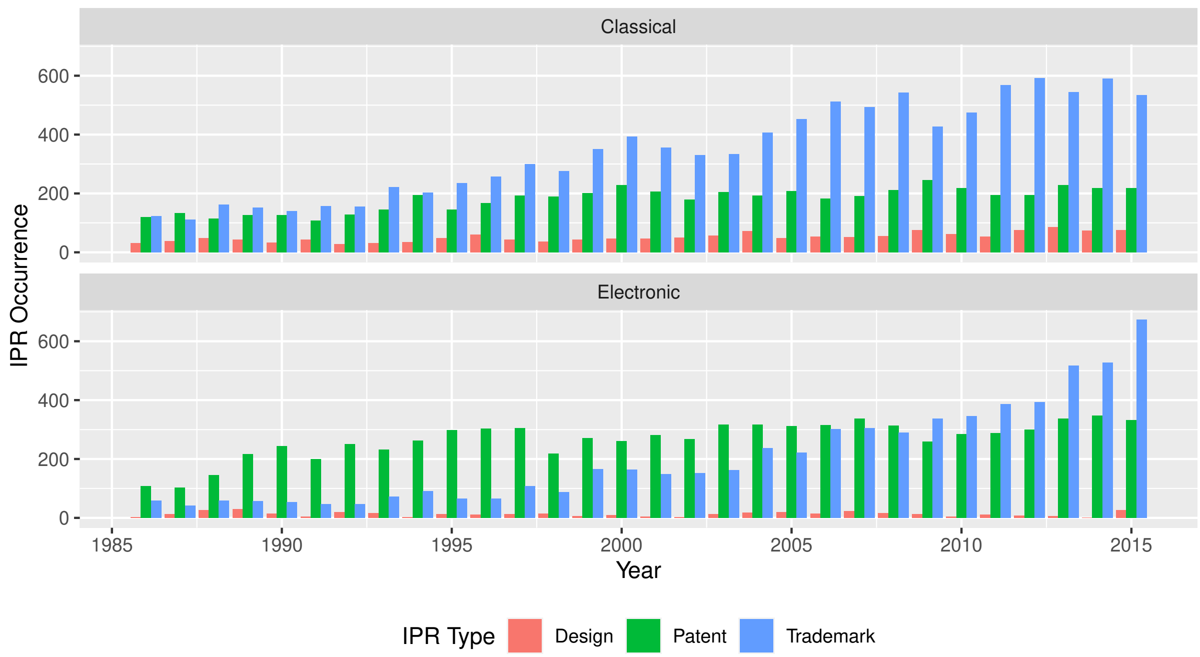


Figure 5.12: Musical Instrument Document Registration per Year and Musical Instrument Type. The figures display the yearly development of IPR registrations, differentiated for classical and electronic musical instruments, with Subfigure (a) representing the overall development, and Subfigure (b) illustrating the IPR development within Electronic and Classical musical instruments.

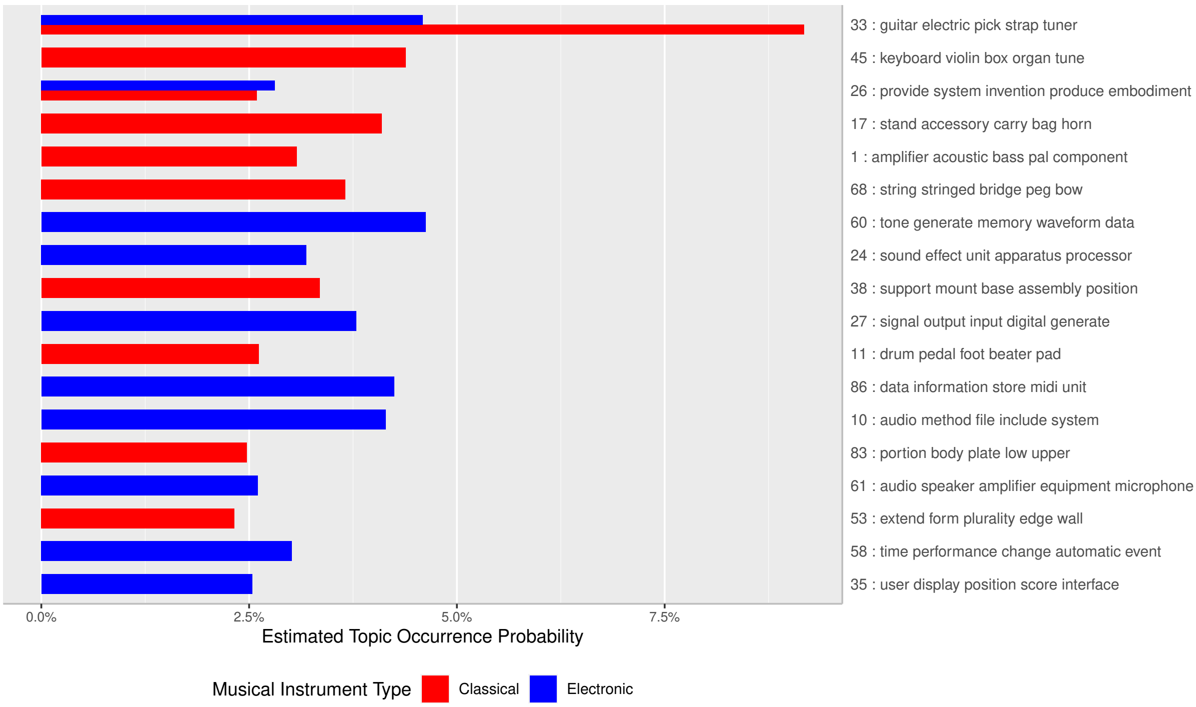


Figure 5.13: Overview of the Ten Main Topics in Electronic and Classical Musical Instruments. The figure presents the ten main topics per musical instrument type, including electronic and classical instruments. The right-hand y-axis displays the topic number and the main words associated with each topic. It also illustrates the topic proportion within each musical instrument type. If a topic is not within the main ten topics for a specific IPR type, it is denoted by a missing bar, indicating that the proportion for that topic is not displayed.

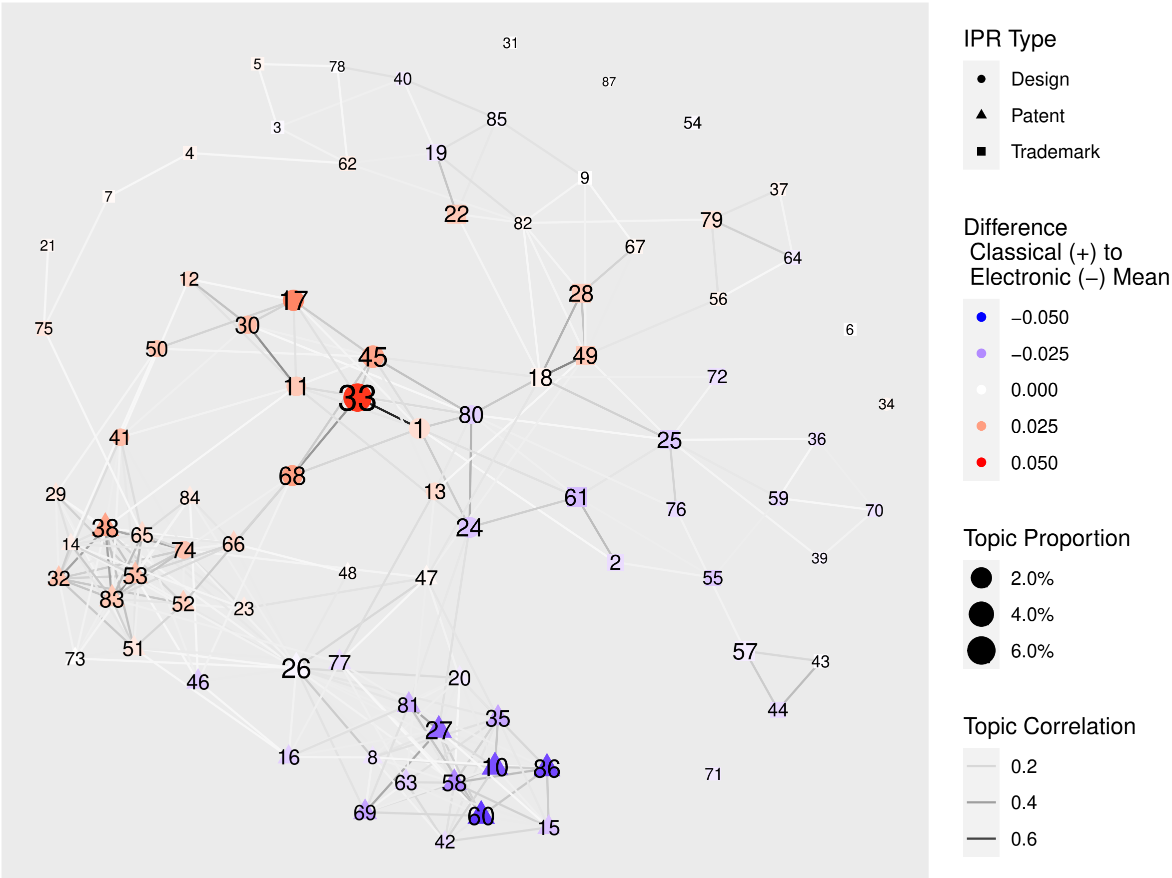


Figure 5.14: Network of Musical Instrument Topics, Differentiated for Electronic and Classical Instrument.
The network illustrates the topic relationships from 1986 to 2015, categorized by IPR types and electronic or classical musical instruments. Nodes represent topics, while links represent the likelihood of these topics to co-occur. The node size corresponds to the topic's proportion. IPR types are indicated by node shapes, while node colours represent the musical instrument type. Intensive blue signifies strong relevance to electronic musical instruments, and red represents a focus on classical musical instruments. In this network, there are 46 nodes related to classical topics and 41 related to electronic topics. The edges display the strength of correlation, with lighter colours indicating lower correlation and darker colours indicating stronger correlation.

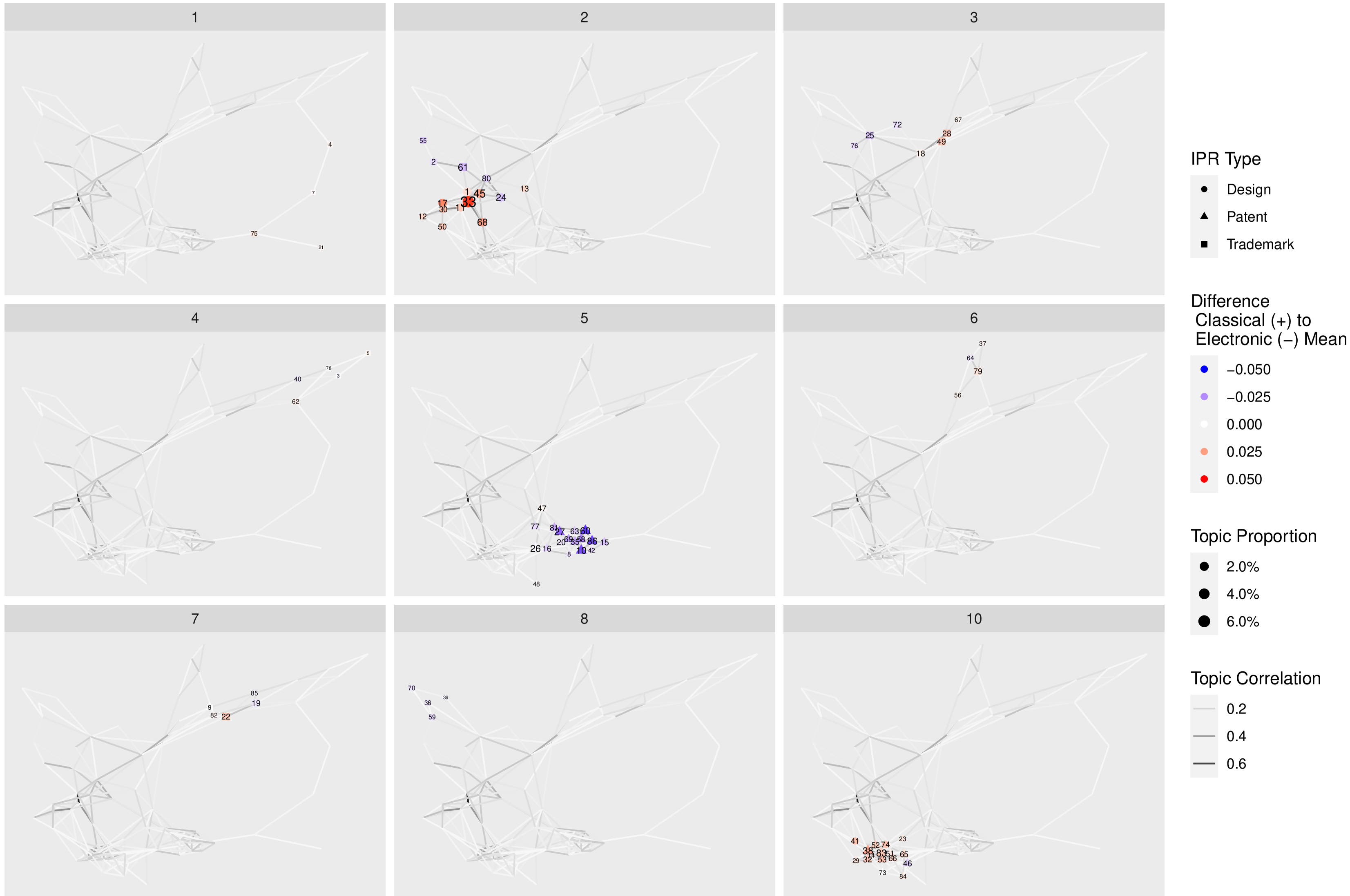


Figure 5.15: Cluster Identification in Topic Networks of Electronic and Classical Musical Instrument from 1986 to 2015.
The figure displays clusters with more than three topics in electronic and classical musical instruments from 1986 to 2015, determined via random walk.

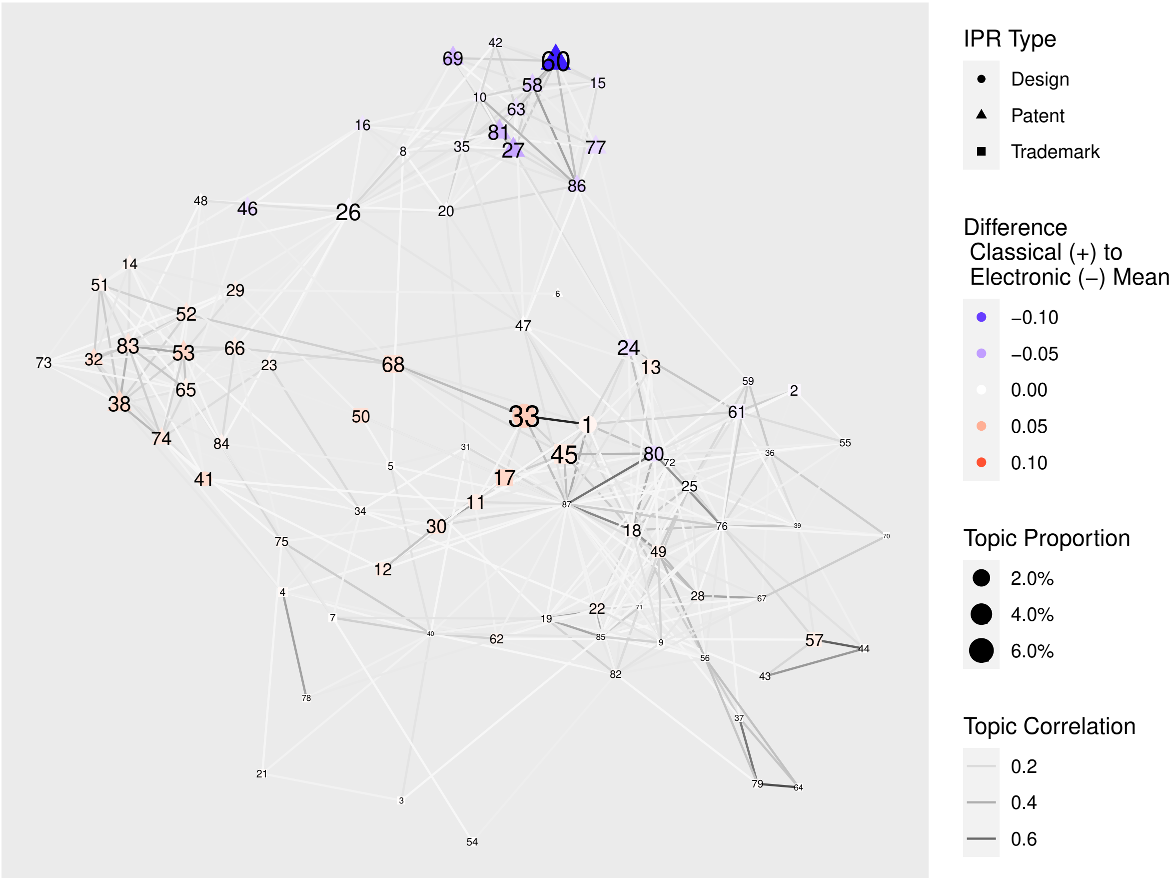


Figure 5.16: Network of Musical Instrument Topics, Differentiated for Electronic and Classical Instruments from 1986 to 1990.
The network illustrates the topic relationships from 1986 to 1990, categorized by IPR types and electronic or classical musical instruments. Nodes represent topics, while links represent the likelihood of these topics to co-occur. The node size corresponds to the topic's proportion. IPR types are indicated by node shapes, while node colours represent the musical instrument type. Intensive blue signifies strong relevance to electronic musical instruments, and red represents a focus on classical musical instruments. In this network, there are 59 nodes related to classical topics and 28 related to electronic topics. The edges display the strength of correlation, with lighter colours indicating lower correlation and darker colours indicating stronger correlation.

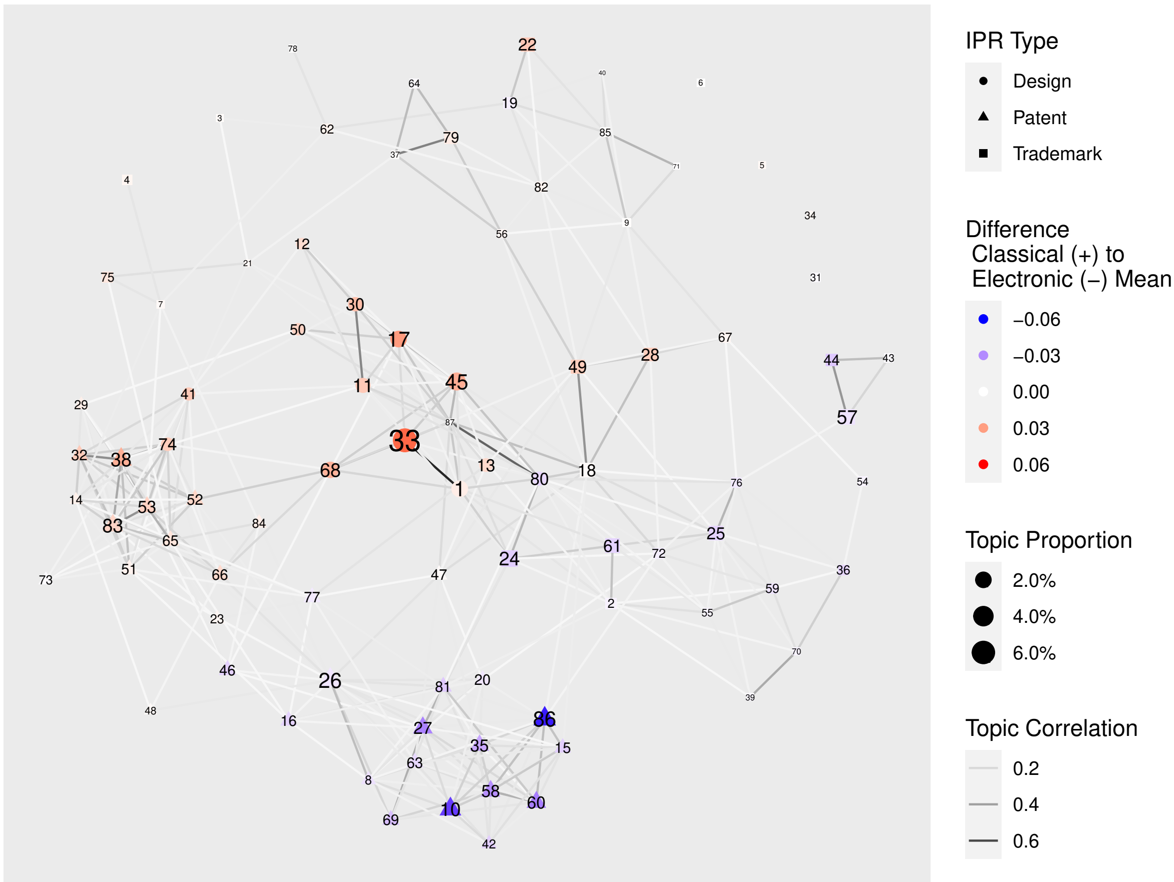


Figure 5.17: Network of Musical Instrument Topics, Differentiated for Electronic and Classical Instruments from 2001 to 2005.
The network illustrates the topic relationships from 2001 to 2005, categorized by IPR types and electronic or classical musical instruments. Nodes represent topics, while links represent the likelihood of these topics to co-occur. The node size corresponds to the topic's proportion. IPR types are indicated by node shapes, while node colours represent the musical instrument type. Intensive blue signifies strong relevance to electronic musical instruments, and red represents a focus on classical musical instruments. In this network, there are 48 nodes related to classical topics and 39 related to electronic topics. The edges display the strength of correlation, with lighter colours indicating lower correlation and darker colours indicating stronger correlation.

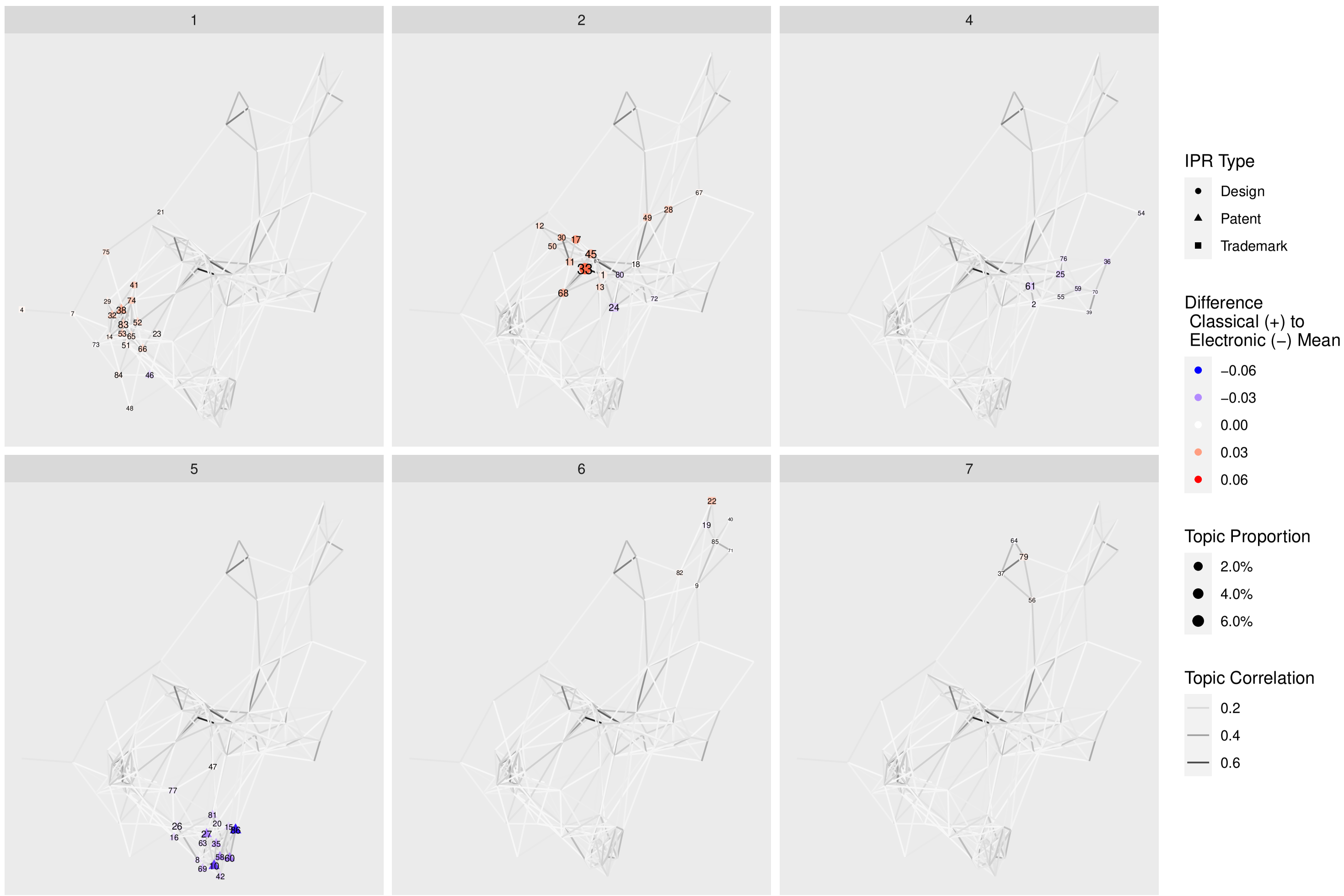


Figure 5.18: Cluster Identification in Topic Networks of Electronic and Classical Musical Instrument from 2001 to 2005.
The figure displays clusters with more than three topics in electronic and classical musical instruments from 2001 to 2005, determined via random walk.

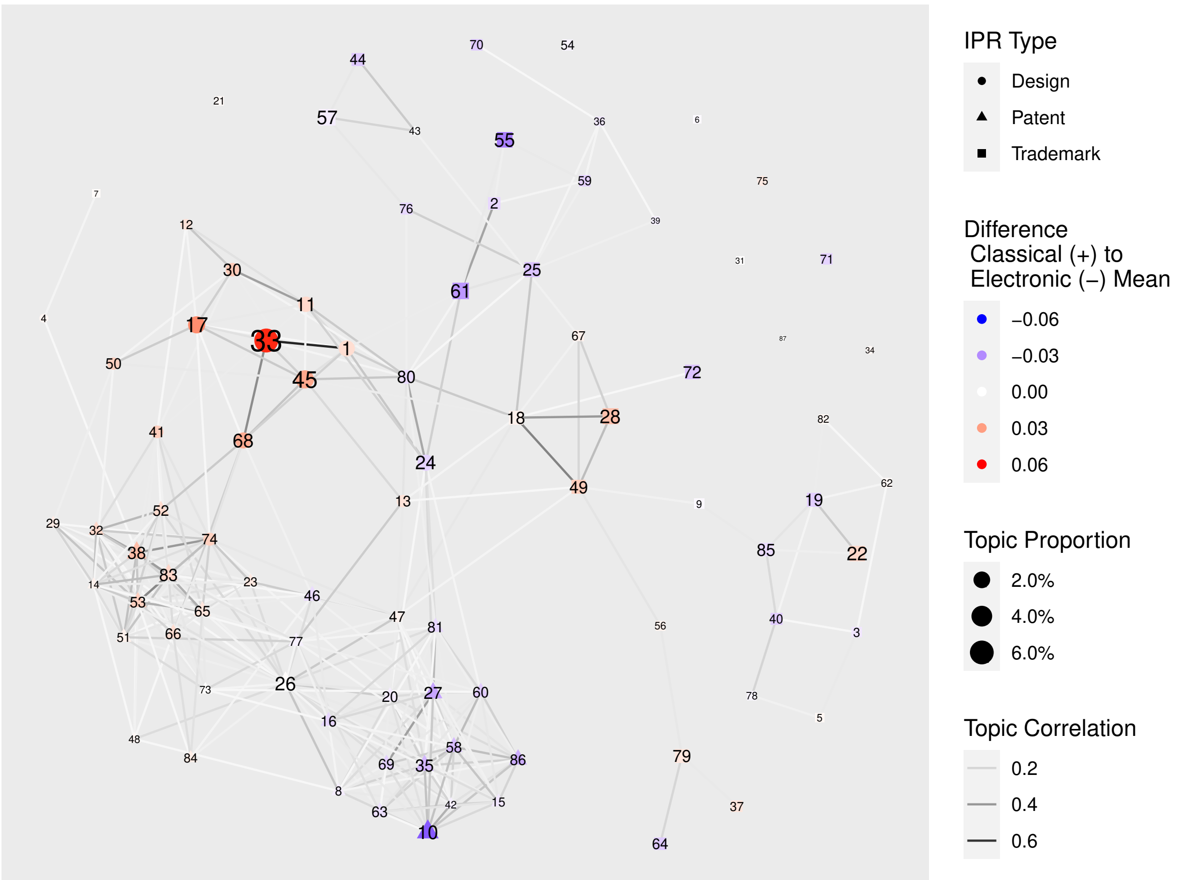


Figure 5.19: Network of Musical Instrument Topics, Differentiated for Electronic and Classical Instruments from 2011 to 2015.
The network illustrates the topic relationships from 2011 to 2015, categorized by IPR types and electronic or classical musical instruments. Nodes represent topics, while links represent the likelihood of these topics to co-occur. The node size corresponds to the topic's proportion. IPR types are indicated by node shapes, while node colours represent the musical instrument type. Intensive blue signifies strong relevance to electronic musical instruments, and red represents a focus on classical musical instruments. In this network, there are 46 nodes related to classical topics and 41 related to electronic topics. The edges display the strength of correlation, with lighter colours indicating lower correlation and darker colours indicating stronger correlation.

5.5 Part II: Major Firms in the Musical Instrument Sector

The analysis in Section 5.4 revealed that the technological transformation occurs from classical to electronic musical instruments. This technological transformation is not only covered by patents, but trademarks contribute information on the transformation with digitalisation, mobile and gaming. This section considers major firms' involvement in the sector's transformation. The intention is to identify firms behind the transformation, relate the topic development with the firm background and assess the related registrations of intellectual property rights. The major firms were identified based on their intellectual property rights registrations between 1986 and 2015. The main assumption is that the firm's backgrounds align with their activities in electronic and classical instruments, digitalisation, mobile and gaming. In terms of low-technology and high-technology, it is expected that high-technology firms are contributing technological aspects and components to the sector's transformation. This means that firms outside the classical musical instrument sector become involved over time. In contrast, low-technology firms adapt technical knowledge and apply it to musical instruments. Before the firm analysis is performed, an overview of the firms and the data set is provided. The networks are then analysed for different intervals and discussed regarding the research questions.

5.5.1 Firm Overview

The firm analysis is performed in five-year intervals from 1986 to 2015. The firms involved in musical instruments are displayed in a network graph together with topics. This creates a firms-topics-network with 87 topics and many potential firms. However, restriction on the major firms is necessary to inspect the firms-topics-network visually. Around 20 firms per firms-topics-network are found to be suitable in terms of visual analysability. This implies that the joint visualisation of topics and firms in networks is still readable and that data cleaning is feasible while the general insights remain similar. A robustness check of the results was performed on 100 and 300 firms. The results presented in this section remain similar for more firms (see also Appendix, Figure D.22 and Figure D.23).

The respective 20 major firms for the analyses are determined for each five-year interval. These are then combined to reach the overall firm data set. The major firms have applied and registered the most intellectual property rights in five-year intervals. The combination of all firms from every interval results in 62 unique firms. Not all firms are active simultaneously in, e.g., electronic and classical musical instruments. Further, only some firms apply every IPR type. The relation of these 62 firms to patents, trademarks and designs, as well as classical and electronic musical instruments, can be derived from Table 5.6. The table provides an overview of the number of firms. It is observable, for example, that 39 firms patent classical instruments. 47 firms apply for patents, and 58 firms are active in classical instruments. In the firm data set, the share of firms in classical and electronic instruments is nearly the same. Looking specifically at electronic instruments, interestingly, most firms (48) apply trademarks to protect their technological developments. This is in line with the findings from the sector perspective.

In the data set of 62 firms, 4,288 patents, 2,429 trademarks and 430 designs are registered in musical instruments (see Table 5.7). Patents are highly represented and contribute 60% of all documents in the data set of Part II, which is a higher patent share than the 42% of all documents in the data set of Part I. Most IPRs relate to electronic musical instruments, with 4,760 registrations. There are about twice as many documents for electronic instruments as for classical Instruments. The number of documents in the total data set is similar to classical and electronic instruments. It is thus expected that in the firm analysis, mostly patenting activities and aspects of electronic musical instruments will be observable. Overall, the 62 firms registered 7,147 documents. They thus represent approximately 22% of the total data set. Further information on the 62 firms is provided in the Appendix, Section D.3.

Table 5.6: Major Firms in Electronic and Classical Musical Instrument IPRs from 1986 to 2015

Firms	IPR Type			Total Firms
	Patent	Trademark	Design	
Classical	39	53	31	58
Electronic	37	48	9	54
Total Firms	47	59	34	62

The table provides an overview of major firms involved in registering intellectual property rights for musical instruments, distinguishing between classical and electronic instruments. To qualify as a major firm, the firm must rank among the top 20 firms in patent, trademark, and design registrations during five-year intervals from 1986 to 2015. The sum of classical and electronic firms may not equal the total number of firms because a single firm can be involved in either classical, electronic, or both. The same principle applies to the IPR types. Each number represents the subset of firms meeting the criteria out of a total of 62 firms.

Table 5.7: Overview of IPRs in relation to Major Firms from 1986 to 2015

	Mean	Median	Min	Max	Total IPRs	Firms
Design	12.65	4	1	175	430	34
Patent	91.23	28	1	1,903	4,288	47
Trademark	41.17	27	1	275	2,429	59
Classical	41.16	18.5	1	461	2,387	58
Electronic	88.15	27	1	1,892	4,760	54
Total	115.27	59.5	9	2,353	7,147	62

The table provides an overview of the IPRs in relation to the major firms in Musical Instruments from 1986 to 2015.

5.5.2 Firm Involvement and Firms-Topics-Networks

Based on the documents of the identified firms, networks are generated. The 62 firms are displayed together only in the general overview, covering the entire period from 1986 to 2015. In the five-year intervals, the major 20 firms of that interval are displayed. In the networks, a node can be either a topic or a firm. A network has two types of links: Topic-topic and Topic-firm links. Links are defined differently depending on the nodes involved:

Topic-topic link: A vector can characterise a topic, one entry for each document, showing the strength of the topic in that document. For each pair of topics, we calculate the correlation between their vectors. We define a link between topics if the correlation is greater than 0.08.

Topic-firm link: The topic-firm links are defined based on the documents a firm "owns". The firm's mean occurrence probability of its documents for each topic is calculated. If the mean occurrence of the firm's documents in a topic is larger than 0.02, we say there is a link between these.

Between two firms, no direct link is calculated nor displayed in the network. Two firms can only be indirectly related through their joint contribution to a topic. This implies that two firms own documents highly related to a topic in the network, and therefore these firms contribute to the same topic. If two firms contribute to the same topics, their positioning within the topic-network will be similar. This suggests that the firms have similar competencies or interests due to their similar topic activities. Figure 5.20 displays the network for these firms from 1986 to 2015. Two different node types need to be differentiated: The colour expresses the topic's relation to classical or electronic musical instruments: dark blue for electronic instruments and red nodes for classical instruments. The shapes of the nodes are determined by the IPR type that is most decisive for the topic: circular for designs, triangular for patents, and rectangular for trademarks. The nodes are always labelled according to their topic number, distinguishing them from firms. However, the colours are solids, which means that the strength of the relationship cannot be displayed. This restriction had to be made to add additional information of firms. Firms have a rhombus shape and are coloured according to their main field of activity in musical instruments. This is determined based on background information on the firms, enriched by the main words found in the documents of the respective firm (see Appendix, Table D.1, D.2, D.3, D.4, and D.5)). For all firms, firm types are determined. This leads to seven categories: Classical Producers (Red), Game Developers (Purple), Hardware (Dark green), Producers (Orange), Other (Grey), Retail (Yellow) and Software (Turquoise) "Classical Producers" cover firms that focus solely on classical musical instruments, while "Producers" offer electronic and classical musical instruments. These firms are directly involved in producing musical instruments or related components. No further distinction is being made between electronic and classical instruments. However, some firms focus on, for example, classical guitars or pianos, while others develop electronic guitars or pianos, and still others provide special components, especially for electronic musical instruments. Game developers include firms that mostly develop game apps or video games but also classical board games. Software and hardware firms are mostly IT-related and either contribute to data processing and related applications or develop processors and computer components. Firms in the retail category are involved in distributing musical instruments or merchandise in general. One important thing to note is that the same colouring for firms and topics is used in cases where topics and firms are related to similar backgrounds and should help with the interpretation. This is the case for red nodes that relate to "classical instruments" (topics) and "classical producers" (firms) and for orange nodes, which relate to generic instruments (topics) and producers (firms), where both classical and electronic musical instruments can be associated.

In the network, it becomes evident that the same types of firms are similarly positioned (Figure 5.20): Hardware firms are related to topic 60 (tone, generate, data, store, read), 69 (filter, wave, frequency, amplitude), 81 (control, circuit, switch), 63 (note, pitch, chord), 26 (system, produce, method), 27 (signal, output), 15 (voice, piece, pattern, rhythm, song), 86 (data, store, apparatus) and 58 (time, performance, event, detect). These topics are primarily in patents and related to electronics. Software firms are also

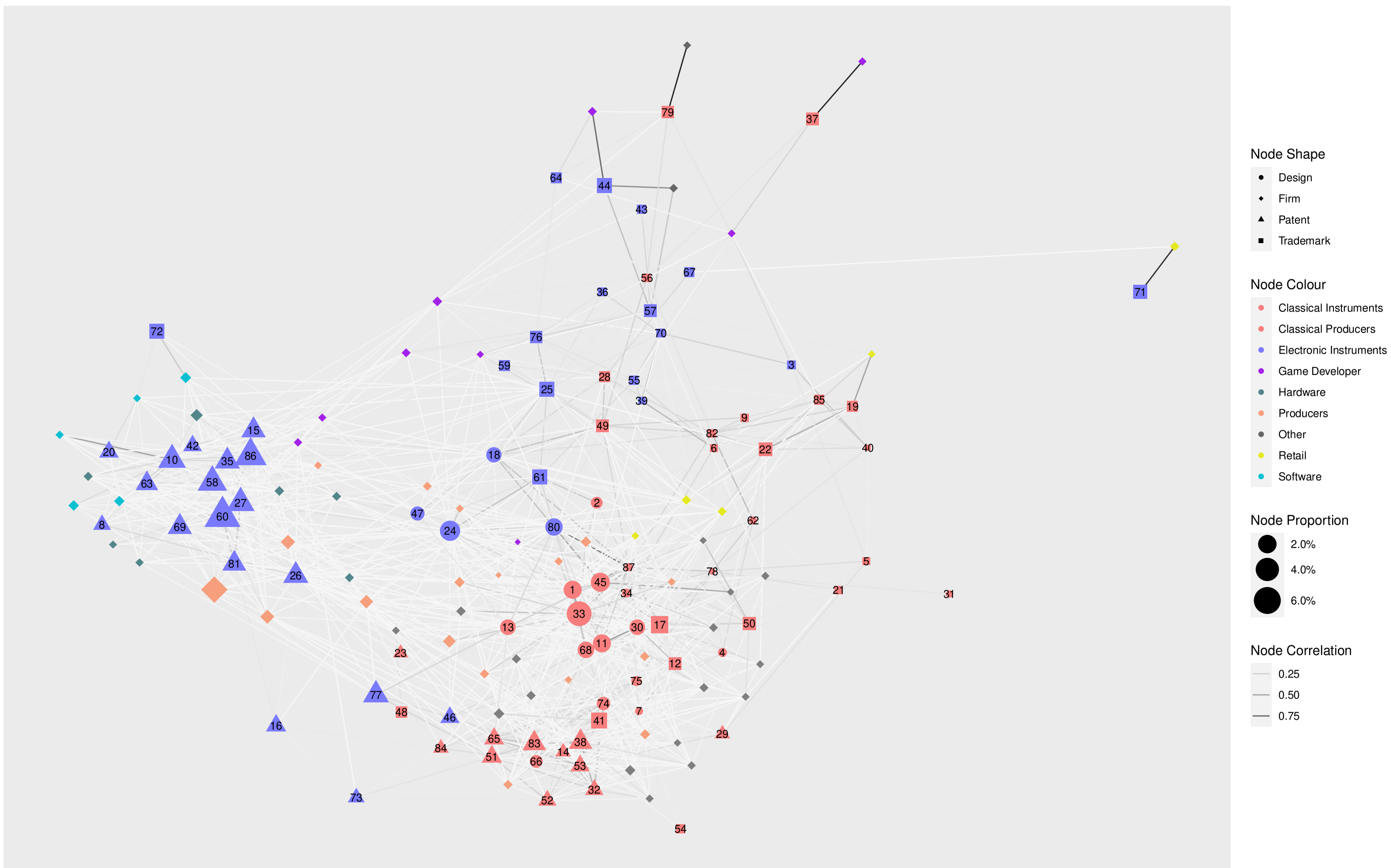


Figure 5.20: Major Topics-Firms-Network from 1986 to 2015.
The network displays major firms in the musical instruments field from 1986 to 2015. Nodes represent firms and different intellectual property rights, including design, patent, and trademark. Topic nodes are numbered, indicating the topic they represent. Topic nodes are coloured either red or blue, depending on their association with electronic or classical musical instruments. The shape and colour of topic nodes are determined by the predominant IPR type or musical instrument type in the dataset. Firm nodes are rhombus-shaped and coloured based on their firm type. For clarity, the overview only provides node colourings and does not include the firm names. Edges represent topic correlations from firms to topics and between topics. Lighter colours indicate lower correlations.

in proximity to these topics. The firms are involved in topics 10 (audio, file, system), 20 (device), 8 (model), 42 (melody, composition, create), and 72 (computer, software, mobile, digital, phone, player). Computers and data processing are naturally relevant in the hardware and software area. Most of the topics of firms in retail or game development are represented by trademarks. Game-developing topics are assigned to the electronic sphere. The firms are positioned among others close to the topics 57 (toy, game, play), 76 (game, video, computer), 37 (paper, book, write), 44 (game, toy, glove), and 35 (user, display, position). In retail, topics 71 (service, business, computer, rental), 22 (service, store, retail, wholesale), 19 (accessory, product, equipment, supply, home), and 85 (service, advertising, information, commercial, business) are of relevance. Some are more focused on musical instruments and then positioned more centrally in the network, while retail generally is more positioned apart. Finally, in musical instruments firms, the breadth of topics is only covered by the interaction of all property rights, here patents, designs, and trademarks, whereby both classical and electronic instruments are mapped. The classical producers are thereby positioned close to classical instrument topics represented in trademarks and designs. Firms involved in classical and electronic musical instruments are generally positioned between the classical and the electronic topics, which focus more on data and signals.

In the following, the networks are examined in further detail in five-year intervals. This helps to understand the contribution of different firms to the transformation and their use of IPRs. In general, the number of producers of musical instruments is highest in the late 1980s, with 16 out of 20 firms being producers. This decreases to 8 firms out of 22 until the late 2010s (see Appendix, Table D.7).

1986–1990: Of the 20 firms, 18 produce musical instruments, of which eight focus on classical musical instruments (see Figure 5.21). Casio Computers is in between, as it is a traditional producer of electronic hardware but also produces electronic musical instruments. In comparison to other musical instrument producers, however, the firm does not produce classical musical instruments. E-mu Systems also produced electronic hardware for musical instruments, focusing on synthesizers and digital sound creation. Both are classified as musical instrument producers. Sam Ash distributes services and guitars. Quake Oats[13] also developed games and provides electronic guitar keyboards. Interestingly, the firms are positioned accordingly in the network. E-mu System links to circuits, controllers and switches (topic 81), frequency, components and waves (topic 69), sensors and modules (topic 16), and electronic apparatus (topic 80). Casio Computers is related to keyboards and touch (topic 77), sound effects and processor (topic 24), electronic components (topic 80) and keyboards (topic 45). The classical instrument producers mainly link to classical topics. In general, patent topic 60 is large with terms like “tone”, “generate” and “data”.

In the network, the topics and firms are placed in such a way that they are best related to the associated nodes while at the same time ensuring readability. This means, for example, that firms such as Latin Percussion, which are only strongly associated with classical themes, are positioned outside the classical themes, while firms such as E-mu Systems, in particular, are more strongly represented among the electronic themes. This implies that, e.g. Yamaha also links to topic 83 (plate portion body) in classical instruments but is nevertheless positioned outside of the cluster around topic 60 (tone, generate, memory), as most connections address these topics. Therefore, although the proximity and positioning of firms and topics in the network is not statistically calculated, firms with similar backgrounds are related to similar topics and are positioned similarly accordingly. Overall, all firms are strongly related to the production of musical instruments, related aspects or their distribution.

1991–1995: Seven out of 20 firms are related to hardware, software and gaming (see Appendix, Figure D.24). Firms like IBM, Sony, Panasonic, Samsung, and Pioneer Electronics relate to studio hardware, consumer electronics, audio equipment or data. The firms relate to topic 86 (data). The focus of the innovations taking place is likewise changing. Patents cover mostly electronic topics, but classical musical instrument topics are also covered. Firms that focused on classical instruments

[13]The firm mainly has a background in food production, but also had a video game division in the past.

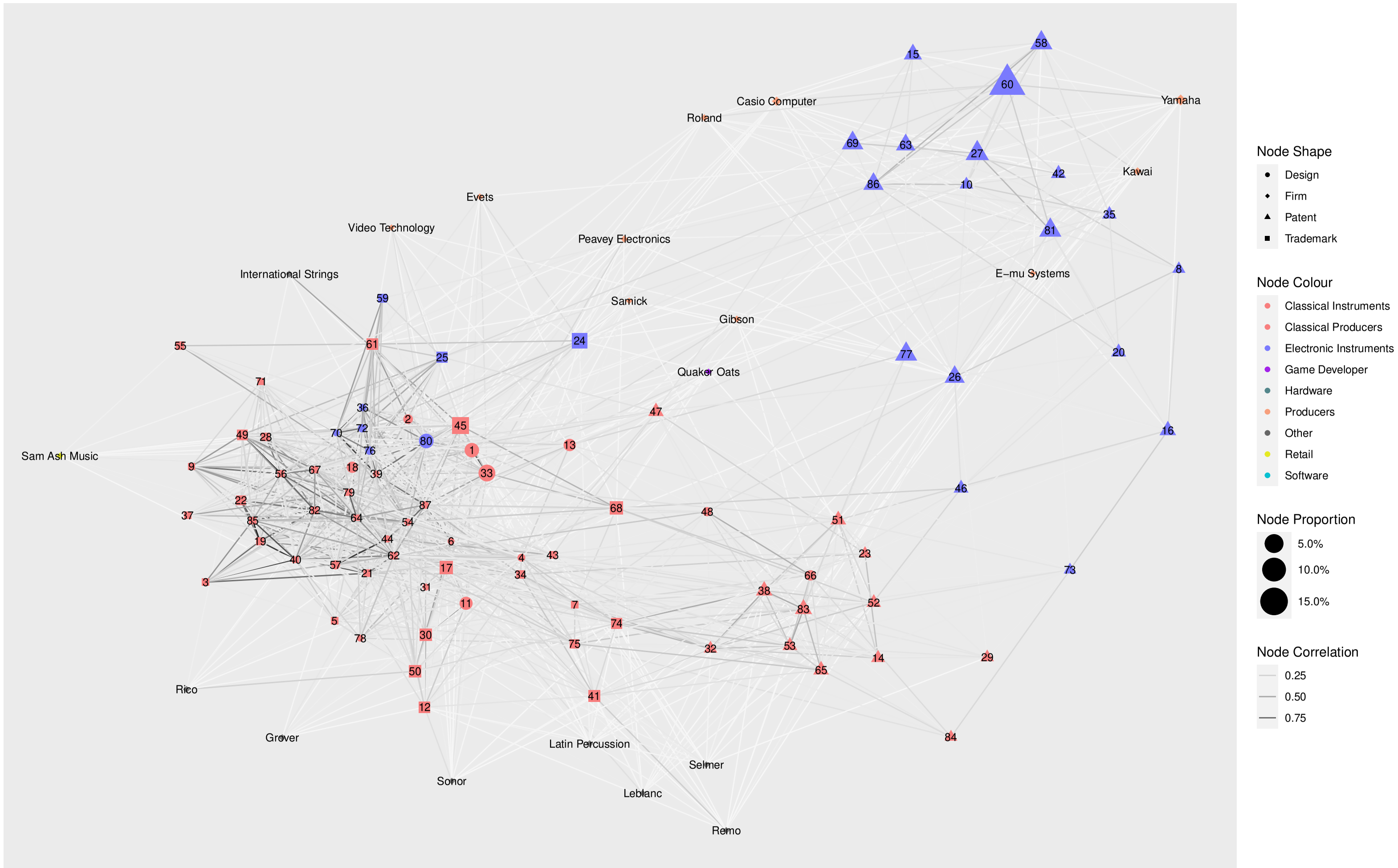


Figure 5.21: Major Topics-Firms-Network from 1986 to 1990.
The network displays the 20 major firms in the field of musical instruments from 1986 to 1990. Nodes represent firms and topics. Topic nodes are labelled with numbers, and their shapes indicate the IPR document with the highest mean occurrence: design (round), patent (triangular), and trademark (rectangular). Topic nodes are coloured either red or blue, depending on their association with electronic or classical musical instruments. The shape and colour are determined by the strongest document type or musical instrument type in the dataset. Firm nodes are rhombus-shaped and coloured according to their firm type, with firm names displayed. Links represent the topic correlation from firms to topics and between topics, with lighter colours indicating lower correlations.

were scarce in this period. Most firms are producers that have a connection to electronic musical instruments or have a strong electronic background. Compared to the first period, more firms from other sectors, like the information and technology sector and with a stronger technological background, are involved and contributing to the transformation.

1996–2000: Trademarks become increasingly important in electronic musical instruments (see Appendix, Figure D.25). Nevertheless, classical musical instrument producers are still present among the major firms. Further, firms from other sectors are relevant, which could already be seen in 1991-1995 and can now be observed again: Sony Interactive Entertainment was founded to cover the aspects of video games. The firm is positioned differently from Sony. This aligns with the firm background, where Sony produces hardware, and Sony Interactive Entertainment combines the knowledge of music, video games and consumer goods. Microsoft is now also among the major firms and provides possibilities for music consumption. In the sector, the innovation areas covered are thus further increasing.

2001–2005: Over time, the representation of electronic topics in trademarks increases further and relates to rental services, computers and games (see Appendix, Figure D.26). Apple joins the transformation and contributes to the development of electronic patent topics.

2006–2010: Samsung, Apple, Microsoft, and Sony are positioned next to each other and highly interlinked with patent topics (see Figure 5.22). The involvement of these high-technology firms indicates that the innovation taking place is also related to high-technology requirements. Due to their linkages, video game developers like Nintendo and Sony Interactive Entertainment are also involved and positioned between electronic patent and trademark topics. In addition, Alibaba and Guitar Center are present with retail services and connected to trademark topics. Overall, the background of firms contributing to musical instruments becomes more diverse, with hardware, software, gaming and retail firms being involved, apart from classical and electronic musical instrument producers.

2011–2015: Until 2015, four different software firms were involved (see Figure 5.23). Samsung, Apple and Google are three major players. They all provide software for music consumption and possibilities to create music. Their documents relate to audio systems (Google, Apple) or sound processing (Samsung). Closer to music performance is the firm “Smule”, which provides apps for users to sing together (see Appendix, Table D.4). The associated topics reflect the firms’ focus: computer software (topic 72), audio files (topic 10), song, voice and rhythm (topic 15). By now, 19 electronic topics are represented in trademarks. Retail firms like Wal-Mart or Alibaba are involved. Related topics are service, business (topic 71), accessories, products, and equipment (topic 19). Trademarks dominate the most relevant topics. Also, gaming firms relate mostly to trademark topics. King.com, Rovio and Outfit7 are involved in gaming app development. For these, they need sound design and animation. The highly associated firms thus changed from initially mainly instrument producers towards firms with a strong technology background or retail relation. This impacts the innovations taking place in the sector as these firms drive their topics. Several musical instrument producers are also involved: Yamaha and Kawai produce classical and electronic instruments like acoustic guitars or pianos. Casio Computers is an example of an electronic consumer goods producer that moved into musical instruments. The firm does not offer classical instruments but focuses solely on electronic musical instruments. The same applies to Marshall Amplification, with a background in speakers and amplifiers that naturally links to sound generation topics.

In general, the topic of classical musical instruments and electronic musical instruments are grouped separately in the late 1980s. Over time, this clear separation disappears, and electronic and classical topics jointly occur. Further, topic 60 and related topics increase in proportion and become more intertwined. This is similar to the effects seen in the analyses of the technological transformation in the sector (Subsection 5.4.4). The design topics are overall low in numbers. However, they are connecting patent and trademark topics.

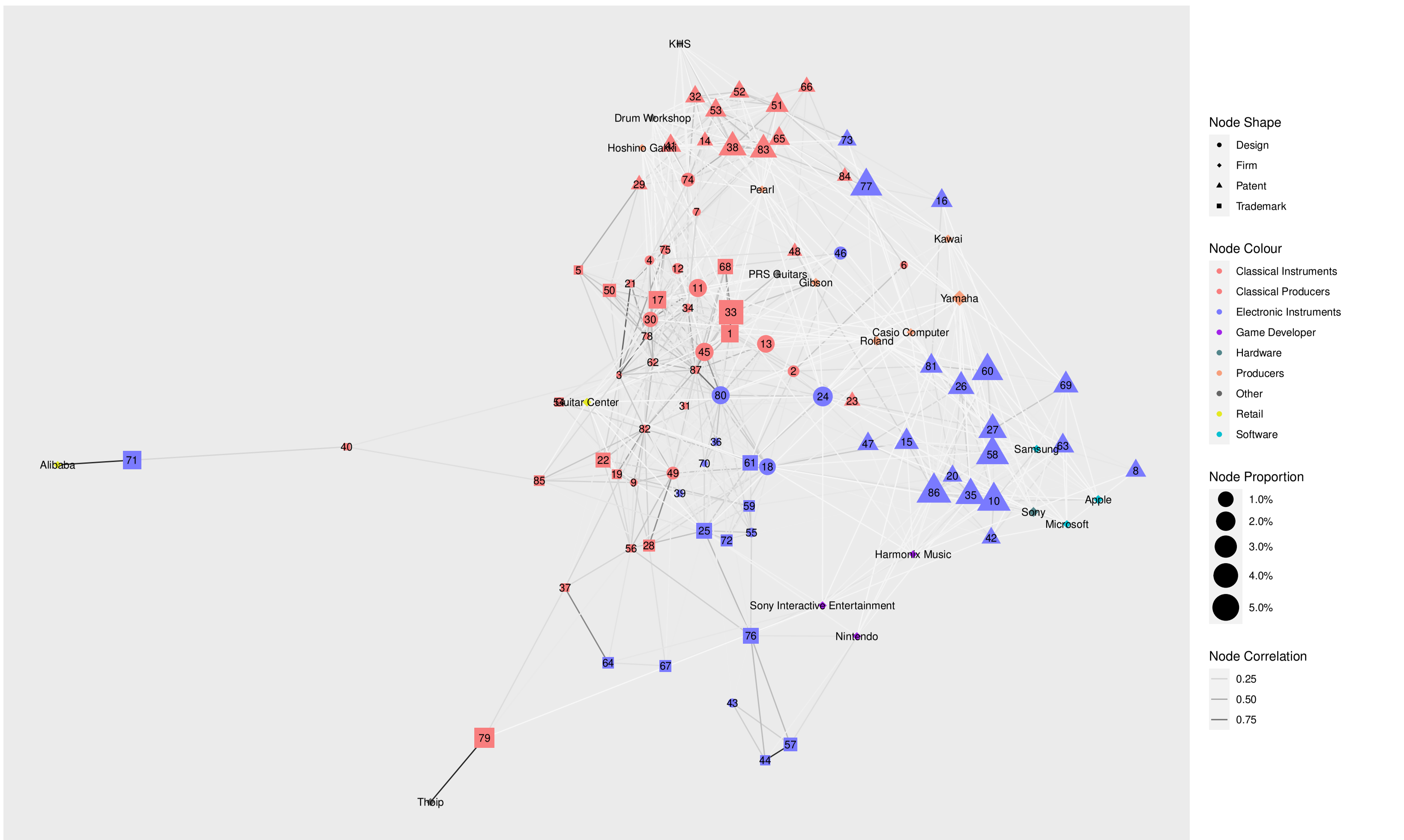


Figure 5.22: Major Topics-Firms-Network from 2006 to 2010.
The network displays the 20 major firms in the field of musical instruments from 2006 to 2010. Nodes represent firms and topics. Topic nodes are labelled with numbers, and their shapes indicate the IPR document with the highest mean occurrence: design (round), patent (triangular), and trademark (rectangular). Topic nodes are coloured either red or blue, depending on their association with electronic or classical musical instruments. The shape and colour are determined by the strongest document type or musical instrument type in the dataset. Firm nodes are rhombus-shaped and coloured according to their firm type, with firm names displayed. Links represent the topic correlation from firms to topics and between topics, with lighter colours indicating lower correlations.

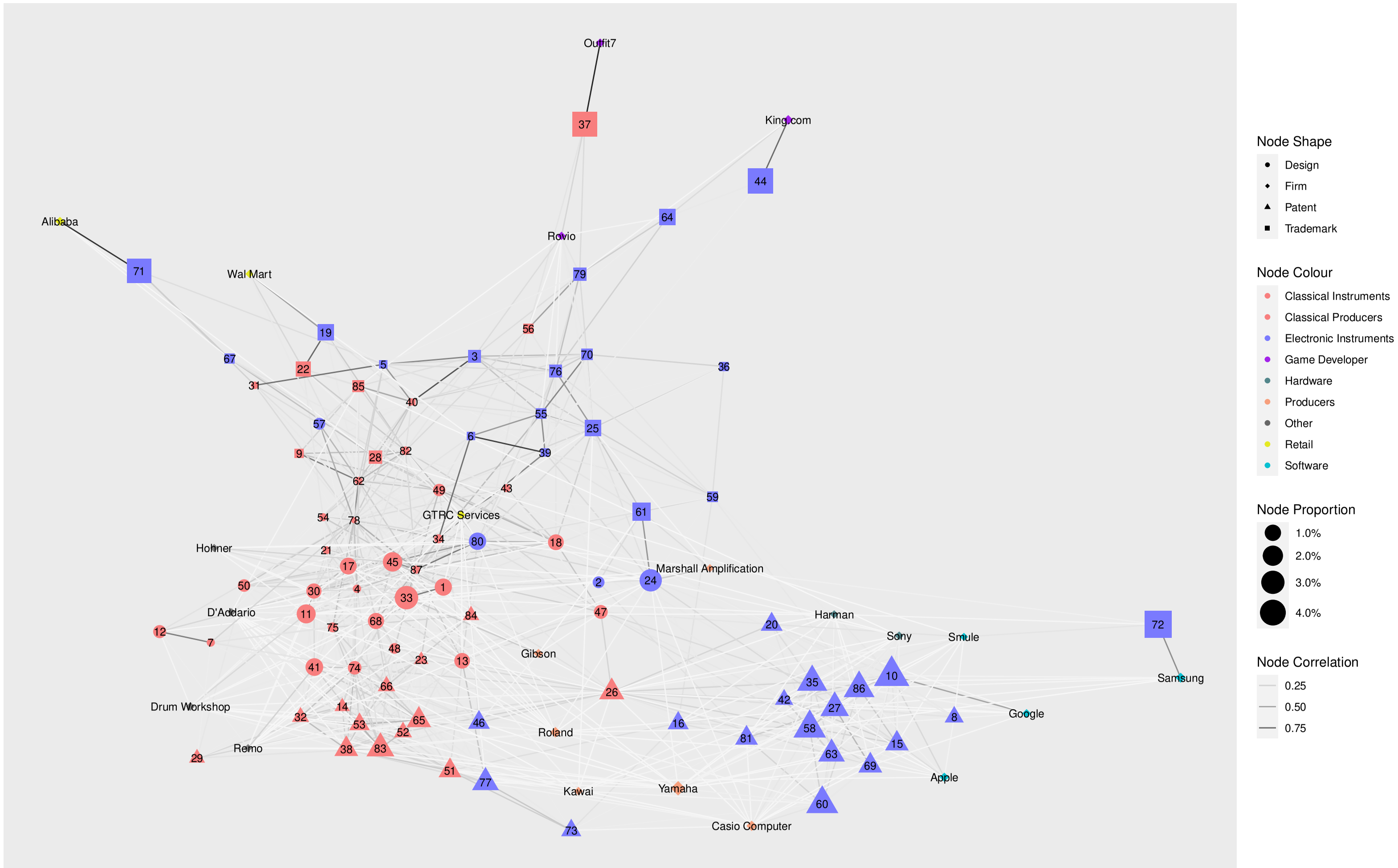


Figure 5.23: Major Topics-Firms-Network from 2011 to 2015.
The network displays the 22 major firms in the field of musical instruments from 2011 to 2015. Nodes represent firms and topics. Topic nodes are labelled with numbers, and their shapes indicate the IPR document with the highest mean occurrence: design (round), patent (triangular), and trademark (rectangular). Topic nodes are coloured either red or blue, depending on their association with electronic or classical musical instruments. The shape and colour are determined by the strongest document type or musical instrument type in the dataset. Firm nodes are rhombus-shaped and coloured according to their firm type, with firm names displayed. Links represent the topic correlation from firms to topics and between topics, with lighter colours indicating lower correlations.

We further saw that firms that started outside the sector (video game developers and computer or electronic hardware) are moving increasingly into the musical instruments sector. The background of the firms mainly involved in the transformation changes: in 1986-1990, mostly instrument producers were active, while over time, more firms with a high-technology background strongly contributed. Interestingly, firms with similar backgrounds still occur in proximity to each other. Producers focusing on electronic and classical instruments bridge the technology firms and the classical producers. Yamaha is the most important firm in all years. Its positioning changes over time but remains closely connected to electronic patenting topics. Like other producers of electronic musical instruments, it is positioned between electronic and classical topics in more recent periods. Regarding innovation, this implies that the innovation subjects covered shifted. While the sector's development in 1986-1990 was mainly on core musical instrument subjects, more technological aspects became relevant over time, which also changed musical instruments and music creation. The firms' background further relates to their linkages in patent, design or trademark topics: Software and hardware firms linked to patent topics, musical instrument producers to patent, design and trademark topics, while retailers and gaming firms were more active in trademarking topics. The joint analysis of all three intellectual property rights thus allowed the inclusion of firms with varying backgrounds that nevertheless contribute to the sector's transformation.

Finally, it is worth taking a brief look at the relationships of firms and the assignment of topics to the Hornborstel-Sachs classification. It is striking that hardware and software firms are not assigned to any specific instrument type (Figure 5.24). Some musical instrument firms are strongly connected through their respective focus (Figure 5.23). For example, drum-related firms, such as Remo or Drum Workshop, are linked to percussion topics (e.g., topics 41, 53, 65, 32). These are usually separated from the firms that are concerned with the production of guitars and pianos. Here it can also be observed that, e.g., piano producers tend to be close to other piano producers, but the separation is less clear (see Yamaha or Kawai). In some cases, proximity to electronics is revealed, or the firms are heavily involved in components such as amplifiers (e.g., Marshall Amplification). Here again, it is shown that firms with similar background cluster together and link to similar topics. Sharing the same topics implies that these firms also contribute similar innovations.

Overall, the firms-topics-networks are consistent with what we know about firm activities. It further displays the context in which firms are active and links different firms together through their shared topic contribution. The networks display that over time, firms like hardware and software producers are becoming more involved, and also game developers contribute to the transformation of musical instruments. Electronic aspects were mostly covered by patents in the 1980s and are increasingly covered by trademarks. However, trademarks are used by firms with different backgrounds, like game developers or retailers, and cover more data and digital-related aspects. Regarding innovation, it becomes obvious that the focus shifts in line with the firms being mainly involved in the sector's transformation.

Concerning the research questions, the following can be said about the involvement of firms and IPR use:

RQ5.3: *Who contributes to the transformation?*

Over time, hardware firms, then software, gaming and retail firms, become more relevant. Of the major firms, the hardware and software firms are known players like Microsoft, Apple, Google, and Samsung. Thus, it is confirmed that high-technology firms from other sectors are becoming active in the field of musical instruments. Apart from firms with their main focus outside the musical instrument sector, musical instrument producers also contribute to the transformation, e.g., Yamaha being the most important player. In the case of Yamaha, they actively intake new technologies and develop their competencies in digital technologies like the internet of things, artificial intelligence, signal processing or sound sources (Yamaha 2020). The firms themselves, here musical instrument producers, are advancing the instruments technologically. This is in line with Neuhäusler and Frietsch (2015). Further, in retail, there are known multi-sided digital platform contributors like Alibaba (see also Hänninen et al. (2017)). Their areas of contribution to

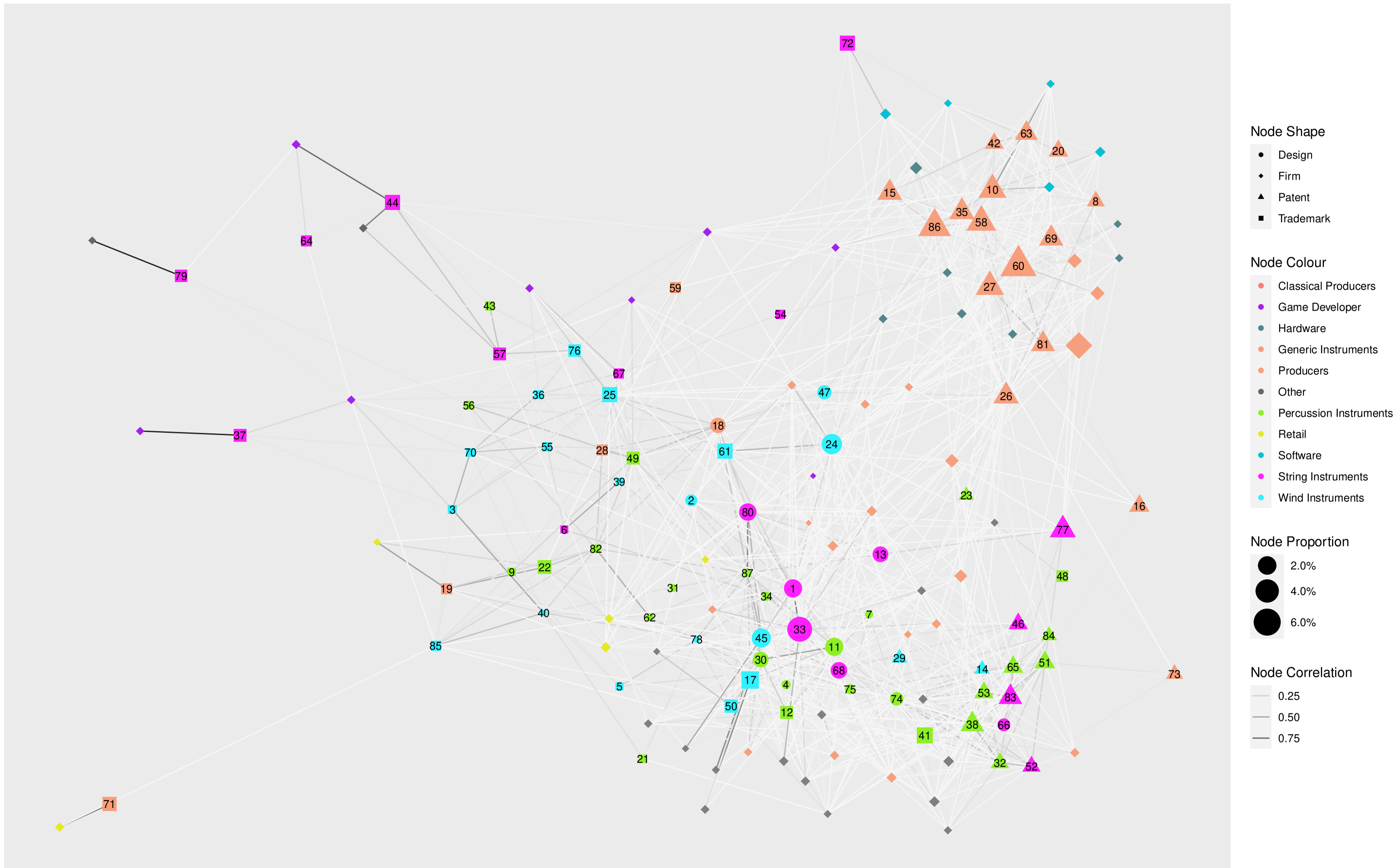


Figure 5.24: Major Topics-Firms-Network from 1986 to 2015 differentiated according to the Hornborstel-Sachs Classification. The network illustrates the major firms in the field of musical instruments from 1986 to 2015. Nodes represent firms and topics. Topic nodes are labelled with numbers. The shape of the topic nodes is categorised into the IPR document with the highest mean occurrence, here: design (round), patent (triangular), and trademark (rectangular). The colouring of topic nodes is either orange, light blue, green, or violet, depending on the connection to music, wind, percussion, or string instruments. The shape and colouring are determined by the dominant document type or musical instrument category in the dataset. The firm nodes have a rhombus shape and are coloured according to their firm type. The firm names are displayed. The links illustrate the topic correlation from firms to topics and from topics to topics, with lighter colouring indicating a weaker correlation.

the sector differ. The hardware and software-related topics are more data or electronically-driven, like signal processing or voice recording. Gaming topics relate to offline games but also online gaming. These firms contribute to the development but also use musical instruments or related innovations. Their involvement in the musical instrument sector comes with a surprise. These firms are included thanks to the text-based data selection. The analysis reveals that gaming firms contribute to the sector due to their development in sound processing. This highlights a strength of the text-based analysis method that allows for uncovering new results based on the textual data provided. Retail firms sell musical instruments and offer themselves digital platforms but are rather not involved in developing innovations in the sector. The musical instrument producers display a large diversity: some are entirely focused on classical instruments, others provide solely electronic musical instruments, and some combine both. Further, some firms in the field of electronic musical instruments focus on parts like amplifiers, strings, or guitar pedals. So the degree of electrification and related knowledge within the producers vary.

RQ5.4: *How does this relate to IPR use?*
The use of intellectual property rights also changes: While electronic topics are in patents initially, they are also increasingly represented in trademarks. This is in line with the findings of the sector perspective. Striking is the interplay of firm focus and intellectual property rights use. Hardware and software firms are predominantly positioned in patent-related topics, while games firms and retailers are found in trademarks. The cluster around topics 25, 55, and 76, identified in the sectoral transformation, is centrally positioned in the firms-topics-networks and in proximity to hardware, electronic musical instrument producers and gaming firms. The topics revolved around video recording, gaming and computers, which aligns with the firms in proximity. Classical musical instrument producers are more closely positioned to topics with a strong trademark or design relation than to patenting topics.

The transformation is driven by firms that produce musical instruments and integrate technologies into them. In addition, firms from other sectors, such as ICT or gaming, contribute to the technological developments. The use of different IPRs for the analysis reflects this development. The use of IPRs differs depending on the background of the firm.

5.6 Discussion of the Transformation in Musical Instruments

The musical instrument sector exemplified the analyses of technological transformation from a low- to a high-technology sector. The sector was chosen due to several reasons: In musical instruments, patents, trademarks, and designs are of relevance, and the sector has undergone a technical transformation from classical musical instruments to the electrification and digitalisation of musical instruments. Textual data were used to combine the different intellectual property rights and to provide information on the transformation. Structural Topic Modelling was applied to discover latent topics and to gain context information on the transformation and the interrelation with IPRs. The analysis was divided into two parts: first, the focus was on the sector before focusing on major firms.

Part I: Sector Analysis The sector was first assessed based on trademarks, patents and designs. Network analyses showed that the topics of musical instruments are divided along the lines of intellectual property rights. This suggests that intellectual property rights represent different content. Patents covered more electronic and digital music generation as well as aspects of data processing. Designs were about musical instruments and related patents and trademarks, while trademarks covered services such as retail and entertainment and components of musical instruments. The design topics remained of strong importance in the firm analysis.

In the case of musical instruments, an influence of the musical instrument type (e.g., wind, string or percussion) according to the Hornborstel-Sachs classification was assumed. However, the type of musical instrument played a minor role. Only an overlap of generic topics and electronic topics could be observed.

Electronic and classical instruments were separated to assess the impact of digitalisation and electrification on the sector. It was found that electronic and classical topics cluster. Furthermore, over time, electronic topics are increasingly located in trademarks with a thematic background in digital sound design, mobile and gaming. This is an interesting result, as one would expect trademarks to be related only to the low-technology aspects of musical instruments. However, they also cover high-technology topics. Patents, trademarks and designs overall covered separate aspects, whereby the patent topics deviated from trademarks and designs.

Part II: Firm Analysis The firm perspective was assessed with a focus on the major firms between 1986 and 2015 in five-year intervals. It was found that firms with similar backgrounds are linked to similar topics and therefore position similarly in the network. For example, hardware or software firms, but also producers of guitars or drums, are similarly located in the network.

It was further observed that the intensity of technological relatedness drives the positioning of firms in relation to topics and, therefore, in the network. Firms focused on classical instruments were mainly linked to classical topics and positioned accordingly. Firms that also produce electronic instruments are likewise more connected to electronic topics and are also positioned accordingly. Here, it can be noted that the degree of electronic instrument production varies. Over the years, more and more hardware firms have been among the major firms, presumably responsible for the provision of electronic components and further development. Technologically, the processing of signals is relevant, both in amplifiers and in the recording and reproducing of sounds. Accordingly, hardware firms are concerned with sound recording and reproduction, such as headphones.

A surprising development is the addition of software developers and game manufacturers over time. The occurrence of software firms highlights the development towards digital musical instruments: These firms eased the consumption of music via, e.g. streaming services and enabled the digital creation of music via, e.g., app-simulated musical instruments. Game manufacturers are often video game or app-game developers. Sound design is important for these games, leading to these firms being thematically represented in musical instruments. Retailing of musical instruments occurs directly through the producers, retailers specialising in musical instruments, or retailers offering diverse product groups. The firms related to gaming and retailing were mostly connected to trademark topics, while software and hardware firms were linked to patent-related topics and positioned accordingly.

Overall, the firm composition changes over time towards firms from other sectors and with other thematic - especially more digital or technical - backgrounds (like Google, Microsoft, Apple or Samsung) entering the musical instrument sector. The thematic and technical background played a major role in the application of intellectual property rights.

While the sector analysis herein focused on the topics, the firm analysis revealed the actors behind them. Even though the firm analysis covered a smaller data set than the sector analysis, the same trends could be identified: over time, gaming and related topics, thus digital sound design and videos became more important. Further, mobile developers entered the sector. Relating the results to the low-technology and high-technology activities of the sector (see Table 5.1), low-technology topics still exist, like the manufacturing of musical instruments. However, high-technology topics play an increasingly important role in the sector, covering activities like the development of amplification, processing of data and signals and the creation of digital sound. Patents and trademarks jointly cover this technological transformation, developed by musical instrument firms and firms active in high-technology sectors. Thereby, the musical instrument producers still cover the core of musical instruments and their production. However, firms with ICT, gaming or retail background still contribute to the sector's development. The analysis thus provided more context and detail on the transformation. In terms of innovation, it remains unclear how the gaming aspects directly improve musical instruments. Here, a more indirect contribution through the improvement of the sound is assumed.

5.7 Conclusion and Further Research

Concerning the research questions presented in Section 5.2, the following can thus be said:

RQ5.1: *What is the nature of the sector transformation from a low-technology sector to a sector with low- and high-technology?*
The musical instrument sector has transformed from the provision of classical musical instruments to electronic instruments and digitalisation. The related topics revealed that digital sound creation, mobile and gaming have been of increasing interest over time.

RQ5.2: *How does the transformation relate to IPR use?*
In line with technological transformation, the use of intellectual property rights has changed. Even though one would expect an increase in patenting activities due to the increase in technological aspects, this is not observable. Trademarking covers the aspects of digital sound creation, mobile and gaming.

RQ5.3: *Who contributes to the transformation?*
In the late 1980s, mostly musical instrument producers were active. Over time, hardware and software firms, retailers and game developers entered the sector. The transformation and technological change is thus driven also by firms that are not directly involved in musical instrument production. Firms with a software or hardware background contribute knowledge of data processing and computational aspects. Gamings firms are involved in digital sound creation. Nevertheless, also musical instrument producers contributed to the transformation and developed musical instruments further.

RQ5.4: *How does this relate to IPR use?*
The hardware and software firms were closely related to electronic topics and patenting, while firms covering game-developing aspects and sound creation were more related to trademarks. The inclusion of trademarks made the entry of the latter firms more obvious. Most musical instrument firms applied trademarks or design protection furthermore. The use of IPR thus is highly firm background dependent, as patents and trademarks cover technical aspects, but the technical subject covered differs.

To summarise, the analyses contained a sectoral and a major firm perspective. From the sectoral perspective, aspects like sound design, mobile and gaming, which have been increasingly present in the musical industry, were captured mainly by trademarks. Designs displayed an important connection between trademarks and patents, while patents mainly contributed to electronic or technical aspects of musical instruments. The firm perspective provided additional insights: Firms from various backgrounds contribute to the sector's transformation. As a result, the application of intellectual property relates to the background of the firms. Trademarks especially highlight the involvement of gaming firms, which aligns with the sector perspective results. Also, from the firm perspective, designs remained important, topic-wise positioned in between patents and trademarks.

To conclude, combining the different IP rights captures the technological transformation in musical instruments. The use of text-based analysis enabled the discovery of topics that are related to different firm backgrounds. With the combination of different IPRs and the usage of textual data, some of the dynamics were traced. Firms like Yamaha have an integrated IPR strategy so that the firms' innovation activities are only covered with the combination of different IPRs. Textual data enables the combination, highlighting different nuances of the innovations. The joint consideration of patents, designs and trademarks depicts the transformation of the musical instrument sector more comprehensively - and thus presumably better -

than a single property right would have been able to. If, for example, the focus had only been on patents, game developers and classical musical instrument manufacturers would have been less present. Similarly, if only trademarks had been considered, the technological transformation driven by software and hardware firms would not have been represented. Especially in areas with low patent applicability, like software and services, the combination of different IPRs is particularly suitable for mapping the transformation comprehensively. The combination allows for events to be depicted more broadly and provides a comprehensive perspective on innovation. Nevertheless, the innovation-relatedness of trademark documents remains an open question: Trademarks mostly represent the topics of gaming firms and retailers. Game developers are involved in developing gaming and sound innovations, while the innovation reference in retail must be questioned.

Method-wise, the networks with additional data contributed to understanding the sector. The shift to new technologies becomes observable. Further, the topics provide contextual information on the change and changing importance. The combination of topics and firms further allowed for a more content-wise perspective. It reveals not only which firms are involved but also in which aspects. For the analysis, the firm names were cleaned to a single writing pattern before aggregation. Sub-firms were, however, not aggregated. This was done assuming that close thematic integration within the firm would result in sub-firms being positioned similarly and appearing together while a different focus could be mapped accordingly. This assumption was successfully demonstrated. In the case of Sony, this led Sony and Sony Interactive Entertainment to be analysed separately. Accordingly, the firms are thematically set up and positioned differently, with the former having a closer connection to hardware topics while the latter covers aspects of video games. Ultimately, this approach can be used to analyse firms in closer detail and to break down the positioning of sub-firms more precisely. The analysis can also provide more background information, in our case, about the technological transformation of the sector. Even though the analysis of the chapter focuses on musical instruments, the analyses are also of interest to understanding the transformation occurring in other sectors. Applying the methodology to other sectors could reveal additional insights, mainly covering firms that are solely active in trademarks. Design protection could also be of interest in other sectors, bridging trademarks and patents or providing information on less technical or design-heavy innovations.

For future research, several points are to be mentioned: The technological transformation and firms' involvement were measured based on textual data analysis. The joint analysis of several intellectual property rights and especially the inclusion of trademarks covered the transformation activities in the sector. They enabled a broad analysis to improve the understanding of innovation. However, not every trademark is related to innovation. The reference to innovation was ensured via the combination of patents and designs. For future research, it is interesting to analyse if trademark textual data of innovations can be distinguished from those that do not represent an innovation but, for example, mark diffusion. Combining topics with firm positioning in a firms-topics-network can provide further insights into technological transformation and firm strategies. Further research could include that information or analyse the positioning of firms with multiple sub-firms in closer detail. Further, the patenting and trademarking shares in the total data set versus the firm data set deviated. The share of patents was higher in the firm data set than in the overall data set. This might indicate different firm types are active in the different intellectual property rights: Patents might be more of interest for large firms, while trademarks might be registered more by small firms and to a smaller degree. Comparing the positioning of large and small firms in a firms-topics-network might lead to further interesting results. In general, the analysis provides context information and is descriptive. Further research might be of interest to include insights, e.g., on firm types or the relation of firms to specific topics to improved models of intellectual property rights use.

6 Discussion and Final Remarks

The thesis covered various aspects of textual data of trademarks in the context of innovation research. Before concluding this thesis, a summary of each chapter is provided in Section 6.1 followed by a discussion. The discussion focuses on insights gained on textual data of trademarks from the joint consideration of the different chapters. The thesis concludes in Section 6.3.

6.1 Summary and Contribution

The thesis evolves around the use of trademark textual data in innovation research. To capture innovation broadly and especially cover service innovation and innovation in low-technology areas, this thesis examined trademarks as an innovation indicator. To combine trademarks with other data sources and to increase the level of detail of the analyses, textual data of trademarks were assessed.

After presenting background information on innovation, intellectual property rights, the measurement of innovation, and the analysis of textual data in Chapter 2, Chapter 3 provides a general perspective on the use of textual data in innovation research. Chapter 4 then compares patents and trademarks from a textual data perspective in Robotics and Footwear. The data sources are evaluated against their coverage of products, services and technical innovation. Chapter 5 applies the insights gained in the previous chapters and uses them to measure technological transformation based on the textual data of patents, trademarks and designs in Musical Instruments. It further provides information on the involvement of different firms in the transformation.

Each chapter of the thesis thus focused on different aspects. However, the chapters are connected by various elements. All chapters cover innovation and different innovation areas (see Figure 6.1). The breadth of innovation areas (see yellow triangle) covered is broadest in Chapter 3, as here, all areas of innovation are considered under the condition that textual data are used for the analysis of innovation. Chapter 4 covers Robotics and Footwear: the former represents a high-technology area and the latter represents a low-technology area. In Chapter 5 the Musical Instruments sector is assessed as a sector in which a technological transformation takes place and where aspects of low- and high-technology are relevant. In parallel to a decreasing breadth of innovation area, the integration of high-technology and low-technology increases. In Chapter 4, the high- and low-technologies are analysed individually and then compared for Robotics and Footwear, while low- and high-technology aspects are jointly analysed in musical instruments. The data sources used (see green triangle) during the thesis become more diverse. At first, the research is based on a sample of publications related to textual data in innovation research. In the following Chapter 4, patent and trademark textual data is used as the basis for analysis and the textual content of the different data sources are compared to each other. Finally, in Chapter 5, patents, trademarks and designs are jointly used to provide broad coverage of the innovations in musical instruments and the topics derived from the analysis are used to identify general patterns. Figure 6.1 provides an integrated perspective on the chapters. The joint aspects are coloured accordingly.

Further detail on each chapter is provided in the following:

Chapter 3 provides a general overview of the current use of textual data in innovation research. Of special interest are the motivation to apply textual data analysis, the data sources used, the methodology applied, and the research questions answered based on textual data. The chapter contributes to a

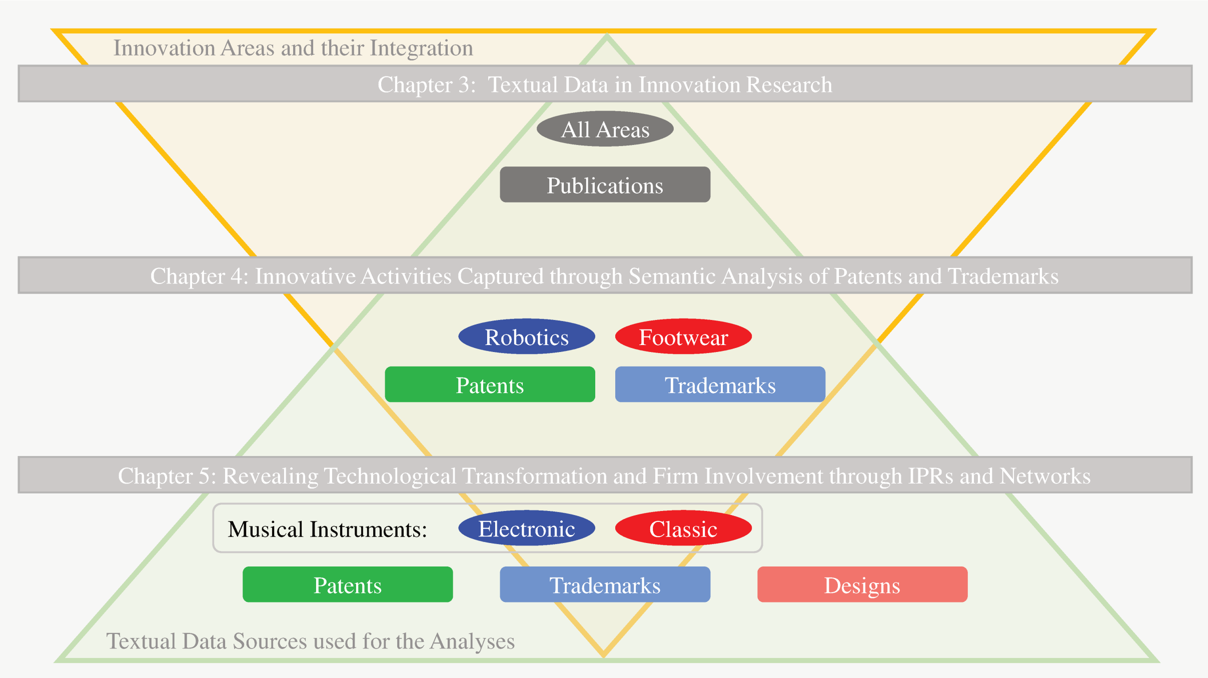


Figure 6.1: Summary of the Integrated Perspective on the Research Chapters
The figure provides an integrated perspective on the research chapters of the thesis. The yellow triangle represents the breadth of covered innovation areas, while the green triangle signifies the range of used textual data sources. Further detail of the innovation areas covered is provided in oval boxes: these are either grey, dark blue, or red, indicating no specific technological focus, a high-technology area, or a low-technology area. The textual data sources are presented in rectangular boxes, with grey boxes for publications that cover various data sources for their analysis, green boxes for patents, light blue for trademarks, and light red for designs. This consistent colour scheme is maintained throughout the thesis for patents, trademarks, designs, high-technology, and low-technology, ensuring clarity across the chapters. Overall, it can be observed that the scope of intellectual property rights covered expands throughout the thesis, while there is a more integrated consideration of low- and high-technology. For instance, musical instruments encompass aspects of both low- and high-technology.

better understanding of textual data in innovation research and the different methodologies applied. In the context of the thesis, an overview of text-analysis methods are given, which help to build an understanding on text-analysis. For this purpose, articles from the fields of economics and business that contain text and innovation in the title or abstract are considered. The data set is successively narrowed down based on the author-provided keywords and an intensive analysis of the abstracts. The final sample consisted of 23 articles that are related to innovation and textual data.

The analysis revealed that most authors appreciate the possibility of analysing large textual data sets in an objective, data-driven and time-efficient way. Primarily patents and publications are used as data sources. Clustering approaches were mainly applied. The research questions covered the discovery and exploration of technologies, industry convergence, innovation trajectory and diffusion, and the emergence of novel topics or idea generation. The chapter thus provides an understanding of the current state of textual data analysis. This contributes to the application of textual data analysis in innovation research. It highlights potential research areas that benefit from the use of textual data: Text-based analyses enable the analyses of text data sources and the combination of different data sources via language. Due to their reduced expert involvement, objectivity increases. Analysing text-based data becomes faster and easier with the use of text-based methods, allowing for regular repetition. Methodologies can be used for a variety of objectives. For the most part, they are used to identify topics or emerging trends and make them visible. Overall, the chapter reveals that research on textual data in innovation can provide an additional perspective on innovation. So far, textual data analysis is still mostly descriptive, but some concepts to use the information gained in further analysis exist.

Chapter 4 combines textual data of patents and trademarks and assesses how aspects of innovation are covered in each data source, focusing especially on trademarks. The chapter contributes to a better understanding of the application of textual data of trademarks and the combination of the data source with patents. This enables further analyses of innovation based on trademark textual data. The chapter addresses two research questions:

RQ4.1: Can trademarks and patents be combined via their textual data on a detailed level?

RQ4.2: Does the textual combination of trademarks and patents add new insights to the discussion of innovation?

To answer the research questions, trademarks and patents are combined textually with Structural Topic Modelling of Roberts et al. (2019). Robotics and Footwear as representatives of high-technology and low-technology are analysed. The results are analysed in terms of textual combination and insights gained. Therefore, general assumptions are drawn from the literature on innovation in trademarks, markets, services, and technology.

The first question covers the possibility of combining patents and trademarks textually: In terms of combination and consistent topics, the analysis displayed mixed results: In Robotics, the textual combination of trademarks and patents displayed consistent results. They were in line with the classifications of patents and trademarks but provided a more detailed perspective. Patents highlighted the invention and technical aspects, while trademarks added details and mainly covered the application areas. For trademark analysis, the approach further connects the trademark topics to the patent classification system, allowing for a better interpretation of the trademarks.

Footwear is a low-technology area with fewer inventions and technological applications. This could be observed in the textual analysis: Only a small number of patents compared to trademarks was available. Further, the trademarks mainly focused on several clothing items instead of innovative technologies. The combination of patents and trademarks was only partially consistent. This means that the topic words, patent and trademark documents do not match. In these topics, trademarks only revealed little information on inventions but focused on end products instead. Some consistent topics in Footwear existed. In these cases, trademark documents were aligned with the content of the patent documents and the topics and provided further details on the subject.

The results in Robotics and Footwear thus provided different results. It became apparent that the textual structure of the trademarks differed between Robotics and Footwear: In Robotics, the trademark descriptions are generally more detailed with information on the product, the purpose of the robot and aspects of protection. The trademark descriptions in Footwear are less specific and detailed than those in Robotics. The Footwear trademark documents displayed a similar textual structure to those in Robotics in topics where the topic words, patent and trademark documents aligned and the topic was thus considered consistent. Further, in Robotics, an increasing level of detail is observable in the trademark descriptions over time. Generally, however, the level of detail of the trademark descriptions differs: Trademarks with a high level of technical information contain similar information to patents. Trademarks with few technological and innovative aspects deviate from patents. Trademark descriptions thus can capture innovation activities and provide a detailed perspective, but this depends on the technological aspects considered. The level of detail is sector or innovation-area-specific as well as time-dependent. This should be considered when trademarks' textual data are used to analyse innovation.

The second question focused on the added insights due to the combination of patents and trademarks. General conclusions about the data sources are extracted from the literature to assess the added insights. These are that trademarks cover more service or market-related innovations, and patents provide information on technical innovation. The chapter challenges these conclusions to assess if the results are reproducible based on textual data. This might not be the case, as the information contained in textual data is different. For example, service-related information might be textually covered in patents and trademarks, even though patents do not protect services. The reason can be that service applications might be described in a patent even though it protects a technological aspect. Therefore, it is necessary to understand the contribution of textual data of trademarks in comparison to patents. The different assumptions are evaluated against the results of the textual data analysis and reveal the following:

Trademarks and Innovation: The link between trademarks and innovation is observable in the textual data. However, a certain technological degree is required to combine trademarks and patents and to generate topics in trademarks where additional information on innovation is available.

Market and Products: In Robotics, trademarks ensure the market perspective of the inventions covered in patents. The market and product application were also covered in patents but not as the primary focus. Patents helped to interpret and bundle trademarks, as the patent classification system provided more structured details than the trademark classification system. In Footwear, trademarks only provided the market introduction perspective in the case of consistent topics. The trademarks were mostly focused on final products without further details on their invention. Here the advantage of the combination and textual data of trademarks for innovation research may not be very strong. In general, trademarks cover more final products and the market perspective. They can add the market and product perspective in the analysis of innovation, while patents focus more in detail on the inventions behind them.

Services: Insights on services could be gained from both data sources, with a higher representation in trademarks. Services like entertainment or education are more present in trademarks than they are in patents. Services with a technical component, however, are also observable in patents like service robotics or footwear subscriptions, which is in line with the findings of Blind et al. (2003). The inclusion of trademarks provided additional insights into services and highlighted the service areas in both Robotics or Footwear.

Technical innovation: The patent topics in both Robotics and Footwear covered technical aspects and focused on functions or components. In Robotics, trademarks contribute detailed technical information on various topics independent of the significant IPR. In Footwear, the technical details are provided mainly by patents and rarely by trademarks. Generally, patent descriptions still provide more technical information than trademarks. Trademark descriptions, nevertheless, cover technical aspects, the extent is just highly dependent on the field of innovation looked at, and the level of detail is lower than patents.

The general conclusions extracted from the literature, thus, in general, also apply to textual data of trademarks and patents. In terms of additional insights, trademarks can thus contribute to innovation coverage. The results' quality depends highly on the application area analysed and the trademarks covered in the data set.

Overall, the combination of trademarks and patents revealed that the data sources can be combined on a detailed level and that the combination adds insights into the innovation discussion. As the field of analysis highly impacts the quality of the results, the field of analysis needs to be considered. The chapter contributes to a better understanding of textual data of trademarks as an innovation indicator – in general and for specific innovation areas of low and high technology. Further, the chapter highlights the potential of the combination of trademarks and patents and the differences between these two textual data sources in innovation. The insights gained serve as a foundation for further analyses of innovation based on textual data of trademarks, especially in joint analysis with patents, which can be used to improve existing innovation measurements.

Chapter 5 assesses the technological transformation in musical instruments and the involvement of major firms in the transformation with joint textual analysis of patents, trademarks and designs. The musical instrument sector combines low-technology and high-technology aspects. It is a sector with, on the one hand, a long tradition and low-technology products such as classical musical instruments. On the other hand, it is a sector undergoing a technological transformation through the electrification and digitalisation of musical instruments. Even though musical instruments are becoming increasingly technical, trademark registrations surpass patent and design registrations.

In musical instruments, patent, trademark and design protection is applied. To understand the transformation, the involved firms and the use of intellectual property rights, the chapter combines textual data of patents, trademarks and designs in musical instruments to attain a broad coverage of technological transformation and to assess a change of use of the IPRs concerning this transformation and involved firms.

Thus, the analysis provides a broader perspective on the transformation, firms and the application of different intellectual property rights (IPRs). The analysis is separated into two parts: Part I focuses on the sector in general, and part II includes major firms in the discussion. The related questions are:

RQ5.1: What is the nature of the sector transformation from a low-technology sector to a sector with low- and high-technology?

RQ5.2: How does the transformation relate to IPR use?

The related research questions from the firm perspective are:

RQ5.3: Who contributes to the transformation?

RQ5.4: How does this relate to IPR use?

The different intellectual property rights serve as a basis for the analysis. Their textual data provide content-based insights into the transformation. For this purpose, the data sources are combined textually, and the topics of the data set are extracted with Structural Topic Modelling of Roberts et al. (2019). Based on the topic estimation, networks can be formed with the topics as nodes and the topic co-occurrences as linkages. The networks are analysed for patterns that are related to the change taking place. In the sector analysis, the relevance of IPR use, different musical instrument types and classical versus electronic musical instruments are assessed, and changes over time are considered. From the major firm perspective, the topic network is then enhanced with firms, building a firms-topics-network.

The sector perspective reveals that the musical instrument sector has transformed from the provision of classical musical instruments to electronic instruments and digitalisation, as expected. The topics covered change towards digital aspects like sound creation, mobile and gaming. With technological transformation, the use of intellectual property rights has changed. Each IPR type provides different information: Electronic topics or data-related topics are relevant in patents but are increasingly present also in trademarks. Designs are an intermediate between patents and trademarks and often cover similar topics as trademarks. The IPR use has changed over time, with trademarking becoming more important in electronic musical instruments and covering the digital aspects of digital sound creation, mobile and gaming.

The firm perspective focuses on the major firms in five-year intervals from 1986 until 2015 and categorises them according to their background. These firms are the major applicants of intellectual property rights across all IPRs. The analysis reveals that the background of firms contributing to the transformation changes over time, with firms from various backgrounds contributing to the sector's transformation. Mainly musical instrument producers were active in the beginning. Over time, hardware and software firms, retailers and game developers entered the sector. Their topics were predominantly related to computation, signalling, voice capturing and data processing. The transformation and technological change is thus driven also by firms that are not directly involved in musical instrument production. The IPR use overlaps with the firm background. Software and hardware firms are more related to patent topics, while gaming and retail firms are close to trademark topics. The positioning of musical instrument producers depends on the degree of electrification of their instruments. Some are positioned closer to patent topics, while most are in the trademark or design topic domains. The results are in line with the sector perspective.

Overall, the electrification and digitalisation of musical instruments is a combined effort of hardware, software and gaming firms leading to new developments in the sector and musical instrument producers that introduce electrification to their instruments. The chapter shows that technological transformation is captured via the combined use of different intellectual property rights. Especially the inclusion of trademarks adds additional aspects like gaming to the analysis. The use of text-based analysis enabled the discovery of topics that are related to the transformation and different firms. This enabled a differentiated analysis of the transformation. Method-wise, the firms-topics-network provides a better understanding of firm positioning in general. The chapter contributes to a better understanding of intellectual property rights application in the context of technological transformation and involved firms. A technological transformation is taking place in musical instruments and across other industries, especially fuelled by digitalisation. The methodology of the chapter can be applied to understand the transformation of different sectors better, as not only high-technology aspects but also low-technology aspects are covered, and to assess firms on a more detailed level.

6.2 Discussion

Considering the chapters together, some general aspects remain to be discussed. The thesis started with the need for a broader innovation perspective to cover low-technology areas and services not covered in patents. Trademarks were treated as an alternative data source to provide this perspective. The combination of trademarks with patents and designs was proposed based on textual data, assuming that these provide the potential to combine several data sources and a broader perspective on low-technology innovation and services. The textual data were preprocessed and combined throughout the thesis with Structural Topic Modelling. The analyses revealed several strong points of the approach but also challenges that still need to be addressed:

Trademarks: Trademark data provide not only information on services but also technological areas. This became obvious in Robotics, where trademarks also contained detailed information on innovations. In Musical Instruments, trademarks introduced aspects of digitalisation like mobile, digital sound design and gaming. Interestingly, the firm's background is more decisive for the IPR application than the technological aspects. One would have expected all high-technology aspects to be found more in patents. However, software and hardware firms were found to be active in patents, while firms with gaming backgrounds were active in trademarks. Here, further research is needed on the differences that explain this behaviour. A reason could be the speed of development in the gaming sector and the importance of marketing and customer closeness.

Through the preprocessing and analyses, it became evident that the textual data of different intellectual property rights differ, especially in the dimensions of legal and common words used, sentence structure and level of detail. These affect the ability to combine and compare the different data sources. The documents' IPR type was therefore considered during preprocessing to account for these differences: Highly common words specific to each IPR were considered and removed. Care was taken in the cleaning process to preserve the differences between the data sources. However, this resulted in words without meaning being included in the final data sets. In order to further improve the analysis of multiple data sources, a stop word list specific to each IP right could therefore be of interest, containing, in particular, the common legal terms of the respective IP right. As such, further focus on meaningful words could be ensured while considering content-related differences in the data sources.

An additional aspect to consider in the textual data analysis of the IPR documents is that the documents were not equally represented in the different innovation areas. In Robotics, most documents were patents, while in Footwear, most documents were trademarks. In the data set of musical instruments, patents and trademarks were equally represented, while designs were rarely included. The share of each IPR type in the data set impacts the topics estimated, as the topics are estimated based on the documents provided and thus lean towards the IPR type with the highest share. This means that the smaller the proportion of an IPR, the greater the uncertainties of the estimate concerning this IPR. We have seen in the different chapters that as long as consistency and relatedness of the documents to the topics are ensured, the difference in the share of IPRs is, however, neglectable. In Robotics, the topics were very consistent. In Footwear, the topics with a high relation to patents were also consistent. In the case of trademark topics, the consistency depended on the structure of the trademarks. In cases where trademarks provided only items of clothing, the consistency was lower, and the topics provided less detail on the production or innovation contained. In musical instruments, topics related to hardware and software firms and mostly patents display a high consistency among the main documents of patents, trademarks, and designs. However, the consistency of trademark topics depends on the electronic aspects. Overall, the consistency and, thus, quality differs depending on the textual structure. A high technological relatedness often leads to a high quality of the topics. Additionally, the quality and use of the IPRs in the area covered need to be considered.

Finally, the text length must be taken into account. Trademarks contain predominantly short texts or concise word lists. Designs are also short in their textual description. Patents provide extensive descriptions or short abstracts. This thesis used the abstracts of patents to align with the textual length of trademarks and designs. This approach worked for the combination. For patents, it could be interesting to extend the analysis concerning the patent description to achieve an even higher level of detail.

Textual Data of Trademarks: The thesis focused on textual data of trademarks in combination with other data sources to gain a broader perspective on innovation. Among the articles in the sample of Chapter 3, no article used trademarks as a data source for textual data. This is in line with the review of Antons et al. (2020), where also no trademarks were covered in the articles applying text analysis in innovation research. However, throughout this thesis, it is shown that textual data of trademarks do contain information on innovation and provide an opportunity to combine trademarks with other data sources. This provides additional insights into firm positioning or innovations in different innovation areas.

The analyses of textual data of trademarks, patents and designs made it possible to discover more detail about Robotics, Footwear and the transformation of Musical Instruments, especially in comparison to the available classifications. Nevertheless, the amount of text and kind of text structure is also decisive for the level of detail: Even though trademark descriptions are relatively short, differences in their structure exist. The trademark descriptions of Robotics or Musical Instrument trademarks covering high-technology aspects displayed a more extensive description and detail than trademarks in, e.g., the low-technology area of Footwear, where the trademark description resembled a list of goods without further detail. Here further research is needed to understand the reasons for this difference and potentially focus on trademark descriptions with a high level of detail.

The analysis of textual data of trademarks in this thesis was used to explore trends. The topic modelling approach supports this intention. However, trend exploration is only one aspect of textual data used for innovation research. More research needs to be done to include the insights gained thanks to textual data in existing economic models.

Innovation in Textual Data of Trademarks This thesis deals with the assumption that trademark texts represent interesting insights about innovations. This is despite the fact that the link between trademarks and innovation is not guaranteed. The analyses of Robotics, Footwear and Musical Instruments were able to show that trademarks provide detailed information about innovation comparable to patents. They could ensure the perspective of market introduction and diffusion. However, differences in the trademark texts were found to influence how well the texts of trademarks contribute to the understanding of innovation. This different structure could be related to the extent to which trademarks represent innovation and requires further investigation.

Overall, textual data of intellectual property rights provide interesting insights into the development of innovation. The thesis addressed textual data of trademarks and revealed their contribution.

6.3 Closing Remarks and Future Research

To conclude, the thesis adds to the discussion of textual data of trademarks in innovation research. It, therefore, analysed the current state of textual data analysis in innovation research, compared trademark and patent textual data and applied textual data from patents, trademarks, and designs to measure technological transformation. Overall, the textual data of trademarks enabled the combination of several data sources. The inclusion of trademarks captured service and low-technology aspects. The textual data of trademarks added aspects like digitalisation in musical instruments. The thesis contributes to a better understanding of textual data use, unlocks textual data of trademarks and applies textual data of patents, trademarks, and designs to assess technological transformation. Combining different intellectual property rights through textual data provides a broader perspective on innovation while overcoming existing limitations. The combination of trademarks, patents, and designs enhances the strengths of each data source for innovation research. The textual data further provides detailed information to understand technological development or the behaviour of firms. The thesis contributes to the area of textual data use for innovation research and especially the application of trademark textual data analysis. In a broader context, the thesis contributes to evidence-based policymaking by enhancing our understanding and measurement capabilities of innovation. This improvement aids in deriving more comprehensive recommendations for political initiatives, especially in low-technology innovation areas.

Further research is necessary. The challenges mentioned above remain to be addressed. In this thesis, the analysed patent, trademark, and design documents were identified using class-based and keyword-based searches. This is necessary as, across different data sources, no common classification system exists. At the same time, it must be ensured that the criteria for the documents to be included are similar across the data sources. In the case of more complex areas of interest, like green technologies in developing countries, a more sophisticated approach to identifying the relevant data across different data sources is of interest as the results are only as good as the data used. Further, the innovation linkage of textual data of trademarks could be improved: In areas with a high technological relation, the trademarks in this thesis displayed a high level of detail. However, in areas with low-technology intensity, like clothing, the trademark descriptions covered mostly a list of clothing items. Here, an approach to differentiate between relevant and non-relevant trademark descriptions for innovation research might be of interest to increase the focus on innovation. Lastly, insights gained by the approach, like the background of firms being decisive for IPR use, remain to be included in further economic models to improve their results.

References

Aggarwal, C. C. (2015): *Data Mining*. Cham: Springer International Publishing. DOI: 10/ghngr6.

Aggarwal, C. C.; Zhai, C., eds. (2012): *Mining Text Data*. Ed. by C. C. Aggarwal; C. Zhai. Boston, MA: Springer US. DOI: 10/c38g.

Allahyari, M.; Pouriyeh, S.; Assefi, M.; Safaei, S.; Trippe, E. D.; Gutierrez, J. B.; Kochut, K. (2017): "A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques". arXiv: 1707.02919.

Amara, N.; Landry, R.; Traoré, N. (2008): Managing the Protection of Innovations in Knowledge-Intensive Business Services. In: *Research Policy* 37(9), pp. 1530–1547. DOI: 10/fg7krp. (Visited on 03/22/2019).

Antons, D.; Breidbach, C. F. (2018): Big Data, Big Insights? Advancing Service Innovation and Design With Machine Learning. In: *Journal of Service Research* 21(1), pp. 17–39. DOI: 10/gcs6qq.

Antons, D.; Grünwald, E.; Cichy, P.; Salge, T. O. (2020): The Application of Text Mining Methods in Innovation Research: Current State, Evolution Patterns, and Development Priorities. In: *R&D Management* 50(3), pp. 329–351. DOI: 10/ggxk3m.

Arnold, M. (2013): J&J Consolidates, Rebrands Ethicon Units. In: *MM+M Medical Marketing and Media*. URL: https://www.mmm-online.com/home/channel/jj-consolidates-rebrands-ethicon-units/ (visited on 10/19/2021).

Athreye, S.; Fassio, C. (2019): Why Do Innovators Not Apply for Trademarks? The Role of Information Asymmetries and Collaborative Innovation. In: *Industry and Innovation*, pp. 1–21. DOI: 10/gjdkd4.

Bakhtin, P.; Khabirova, E.; Kuzminov, I.; Thurner, T. (2020): The Future of Food Production – a Text-Mining Approach. In: *Technology Analysis & Strategic Management* 32(5), pp. 516–528. DOI: 10/gkr9nw.

Baregheh, A.; Rowley, J.; Sambrook, S. (2009): Towards a Multidisciplinary Definition of Innovation. In: *Management Decision* 47(8), pp. 1323–1339. DOI: 10/d8cqgw.

Basberg, B. L. (1987): Patents and the Measurement of Technological Change: A Survey of the Literature. In: *Research Policy* 16, pp. 131–141. DOI: 10/fbtftf.

Basole, R. C.; Park, H.; Chao, R. O. (2019): Visual Analysis of Venture Similarity in Entrepreneurial Ecosystems. In: *IEEE Transactions on Engineering Management* 66(4), pp. 568–582. DOI: 10/ghmpcb.

Becheikh, N.; Landry, R.; Amara, N. (2006): Lessons from Innovation Empirical Studies in the Manufacturing Sector: A Systematic Review of the Literature from 1993–2003. In: *Technovation* 26(5-6), pp. 644–664. DOI: 10/bv9q78.

Bei, X. (2019): Trademarks, Specialized Complementary Assets, and the External Sourcing of Innovation. In: *Research Policy* 48(9), p. 103709. DOI: 10/gjnzmj.

Berlit, W. (2019): *Markenrecht*. 11th ed. München: C.H. Beck oHG. ISBN: 978-3-406-71996-7.

Blei, D. M. (2012): Probabilistic Topic Models. In: *Communications of the ACM* 55(4), pp. 77–84. DOI: 10/b39c.

Blei, D. M.; Lafferty, J. D. (2006): “Dynamic Topic Models”. In: *Proceedings of the 23rd International Conference on Machine Learning - ICML ’06*. The 23rd International Conference. Pittsburgh, Pennsylvania: ACM Press, pp. 113–120. DOI: 10/bjkmq9.

Blei, D. M.; Lafferty, J. D. (2007): A Correlated Topic Model of Science. In: *The Annals of Applied Statistics* 1(1), pp. 17–35. DOI: 10/bzkc9z.

Blei, D. M.; Ng, A. Y.; Jordan, M. I. (2003): Latent Dirichlet Allocation. In: *Journal of Machine Learning Research 3*, pp. 993–1022. DOI: 10/gjng2q.

Blind, K.; Edler, J.; Frietsch, R.; Schmoch, U. (2006): Motives to Patent: Empirical Evidence from Germany. In: *Research Policy* 35(5), pp. 655–672. DOI: 10/c4583d.

Blind, K.; Edler, J.; Schmoch, U.; Anderson, B.; Howells, J.; Miles, I.; Roberts, J.; Green, L.; Evangelista, R.; Hipp, C.; Herstatt, C. (2003): *Patents in the Service Industries - Final Report*. Fraunhofer Institut für Systemtechnik und Innovationsforschung. URL: http://publica.fraunhofer.de/dokumente/N-21425.html (visited on 08/01/2023).

Block, J. H.; Fisch, C.; Sandner, P. G. (2014): Trademark Families: Characteristics and Market Values. In: *Journal of Brand Management* 21(2), pp. 150–170. DOI: 10/gjk3gr.

Block, J. H.; Fisch, C. O.; Hahn, A.; Sandner, P. G. (2015): Why Do SMEs File Trademarks? Insights from Firms in Innovative Industries. In: *Research Policy* 44(10), pp. 1915–1930. DOI: 10/f7x6jp.

Böcher, B. (2008): *Musikinstrumentenbau*. Bonn: Deutscher Musikrat/ Deutsches Musikinformationszentrum (MIZ). URL: http://www.miz.org/static_de/themenportale/einfuehrungstexte_pdf/07_Musikwirtschaft/boecher_musikinstrumentenbau.pdf (visited on 02/05/2020).

Brandl, B.; Hornuf, L. (2020): Where Did FinTechs Come From, and Where Do They Go? The Transformation of the Financial Industry in Germany After Digitalization. In: *Frontiers in Artificial Intelligence* 3, p. 8. DOI: 10/gnhpqz.

Brem, A.; Nylund, P. A.; Hitchen, E. L. (2017): Open Innovation and Intellectual Property Rights: How Do SMEs Benefit from Patents, Industrial Designs, Trademarks and Copyrights? In: *Management Decision* 55(6), pp. 1285–1306. DOI: 10/gkfg75.

Castaldi, C. (2020): All the Great Things You Can Do with Trademark Data: Taking Stock and Looking Ahead. In: *Strategic Organization* 18(3), pp. 472–484. DOI: 10/gndgvh.

Castaldi, C.; Dosso, M. (2018): *From R&D to Market: Using Trademarks to Capture the Market Capability of Top R&D Investors*. 2018-01. Seville, Spain: European Commission, Joint Research Centre (JRC), pp. 1–32. URL: http://hdl.handle.net/10419/202177.

Chiarello, F.; Belingheri, P.; Bonaccorsi, A.; Fantoni, G.; Martini, A. (2021): Value Creation in Emerging Technologies through Text Mining: The Case of Blockchain. In: *Technology Analysis & Strategic Management*, pp. 1–17. DOI: 10/gkr9nd.

Cohen, W.; Nelson, R.; Walsh, J. (2000): *Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)*. w7552. Cambridge, MA: National Bureau of Economic Research, pp. 1–31. DOI: 10/b3jr6d.

CPC (2020): *Cooperative Patent Classification (CPC) - Table on CPC Scheme and CPC Definitions*. URL: https://www.cooperativepatentclassification.org/cpcSchemeAndDefinitions/table (visited on 10/06/2020).

Crass, D.; Schwiebacher, F. (2013): *Do Trademarks Diminish the Substitutability of Products in Innovative Knowledge-Intensive Services?* 13-061. DOI: 10/gh3mrz.

Crass, D.; Schwiebacher, F. (2017): The Importance of Trademark Protection for Product Differentiation and Innovation. In: *Economia e Politica Industriale* 44(2), pp. 199–220. DOI: 10/ghb6zd.

Cripps, H.; Singh, A.; Mejtoft, T.; Salo, J. (2020): The Use of Twitter for Innovation in Business Markets. In: *Marketing Intelligence & Planning* 38(5), pp. 587–601. DOI: 10/gjd5b5.

Crouch, D. D. (2010): A Trademark Justification for Design Patent Rights. In: 24. URL: http://ssrn.com/abstract=1656590.

Cuesta, C.; Ruesta, M.; Tuesta, D.; Urbiola, P. (2015): The Digital Transformation of the Banking Industry. In: *BBVA Research – Digital Economy Watch*. URL: https://www.researchgate.net/publication/291357544_The_digital_transformation_of_the_banking_industry (visited on 12/08/2021).

Curci, Y.; Mongeau Ospina, C. A. (2016): Investigating Biofuels through Network Analysis. In: *Energy Policy* 97, pp. 60–72. DOI: 10/f843j9.

Dahlke, J.; Bogner, K.; Becker, M.; Schlaile, M. P.; Pyka, A.; Ebersberger, B. (2021): Crisis-Driven Innovation and Fundamental Human Needs: A Typological Framework of Rapid-Response COVID-19 Innovations. In: *Technological Forecasting and Social Change* 169(120799). DOI: 10/gjt4sp.

De Vries, G.; Pennings, E.; Block, J. H.; Fisch, C. (2017): Trademark or Patent? The Effects of Market Concentration, Customer Type and Venture Capital Financing on Start-Ups' Initial IP Applications. In: *Industry and Innovation* 24(4), pp. 325–345. DOI: 10/ghf6ch.

deGrazia, C. A. W.; Myers, A.; Toole, A. A. (2019): Innovation Activities and Business Cycles: Are Trademarks a Leading Indicator? In: *Industry and Innovation*, pp. 1–20. DOI: 10/gsgr4j.

DeLaval (2021): *Digital Innovations for a Sustainable Dairy Industry*. URL: https://corporate.delaval.com/ (visited on 10/19/2021).

DPMA (2021a): *IPC-Verzeichnis - G - Physics*. URL: https://depatisnet.dpma.de/ipc/ipc.do (visited on 04/07/2021).

DPMA (2021b): *Patente - Prüfung und Erteilung*. Deutsches Patent- und Markenamt. URL: https://www.dpma.de/patente/pruefung_erteilung/index.html (visited on 11/10/2021).

Dräger, H. H. (1948): *Prinzip einer Systematik der Musikinstrumente*. Kassel, Basel: Bärenreiter-Verlag. URL: http://www.musikwissenschaft.uni-wuerzburg.de/fileadmin/04070000/Instrumentensammlung/Materialien_Instrumente/Dra__ger_Systematik_Musikinstrumente_1948_korr.pdf (visited on 11/17/2019).

Duguid, P. (2018): Early Marks: American Trademarks before US Trademark Law. In: *Business History* 60(8), pp. 1145–1168. DOI: 10/gjzvwt.

DUK (2018): *„Orgelbau und Orgelmusik" als Immaterielles Kulturerbe der Menschheit ausgezeichnet | Deutsche UNESCO-Kommission*. Bonn: Deutsche UNESCO-Kommission (DUK), Zentralverband des Deutschen Handwerks (ZDH). URL: https://www.unesco.de/kultur-und-natur/immaterielles-kulturerbe/immaterielles-kulturerbe-weltweit/orgelbau-und-0 (visited on 10/12/2021).

Dziallas, M.; Blind, K. (2019): Innovation Indicators throughout the Innovation Process: An Extensive Literature Analysis. In: *Technovation* 80–81, pp. 3–29. DOI: 10/gfc38p.

Elsevier (2023a): *Aims and Scope - Journal of Econometrics*. URL: https://www.sciencedirect.com/journal/journal-of-econometrics/about/aims-and-scope (visited on 01/06/2023).

Elsevier (2023b): *Aims and Scope - Technological Forecasting and Social Change*. URL: https://www.sciencedirect.com/journal/technological-forecasting-and-social-change/about/aims-and-scope (visited on 01/06/2023).

Enders, B. (2017): "From Idiophone to Touchpad. The Technological Development to the Virtual Musical Instrument". In: *Musical Instruments in the 21st Century: Identities, Configurations, Practices*. Ed. by

T. Bovermann; A. de Campo; H. Egermann; S.-I. Hardjowirogo; S. Weinzierl. Singapore: Springer, pp. 45–58. DOI: 10/kjxk.

EPO (2018): *Guidelines for Examination in the European Patent Office*. Munich, Germany: European Patent Office (EPO). ISBN: 978-3-89605-214-8. URL: https://www.epo.org/law-practice/legal-texts/guidelines.html (visited on 04/03/2019).

EPO (2022): *Cooperative Patent Classification (CPC)*. URL: https://www.epo.org/searching-for-patents/helpful-resources/first-time-here/classification/cpc.html (visited on 03/21/2022).

Erlinghagen, S.; Markard, J. (2012): Smart Grids and the Transformation of the Electricity Sector: ICT Firms as Potential Catalysts for Sectoral Change. In: *Energy Policy* 51, pp. 895–906. DOI: 10/f4n45z.

EUIPO (2019): *Trade Mark Definition*. URL: https://euipo.europa.eu/ohimportal/en/trade-mark-definition (visited on 03/12/2019).

EUIPO (2021): *Locarno Classification (Designs)*. URL: https://euipo.europa.eu/ohimportal/en/locarno-classification (visited on 05/10/2021).

Eurostat (2008): *NACE Rev. 2*. Luxembourg: Office for Official Publications of the European Communities. ISBN: 978-92-79-04741-1. URL: https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF (visited on 06/01/2023).

Eurostat (2021a): *Database - Science, Technology and Innovation*. URL: https://ec.europa.eu/eurostat/web/science-technology-innovation/data/database (visited on 03/18/2021).

Eurostat (2021b): *Eurostat Indicators on High-tech Industry and Knowledge - Intensive Services - Annex 3 - High-tech Aggregation by NACE Rev. 2*. URL: https://ec.europa.eu/eurostat/cache/metadata/Annexes/htec_esms_an3.pdf (visited on 02/17/2021).

EUTMR (2017): *Regulation (EU) 2017/1001 of the European Parliament and of the Council of 14 June 2017 on the European Union Trade Mark (Text with EEA Relevance.)* URL: http://data.europa.eu/eli/reg/2017/1001/oj/eng (visited on 05/18/2021).

Faurel, L.; Li, Q.; Shanthikumar, D.; Teoh, S. H. (2022): Bringing Innovation to Fruition: Insights from New Trademarks. In: *SSRN*. DOI: 10/gsgr4m.

Feng, S.; An, H.; Li, H.; Qi, Y.; Wang, Z.; Guan, Q.; Li, Y.; Qi, Y. (2020): The Technology Convergence of Electric Vehicles: Exploring Promising and Potential Technology Convergence Relationships and Topics. In: *Journal of Cleaner Production* 260(120992), pp. 1–13. DOI: 10/ghhwwn.

Fiordelisi, F.; Renneboog, L.; Ricci, O.; Lopes, S. S. (2019): Creative Corporate Culture and Innovation. In: *Journal of International Financial Markets, Institutions and Money* 63(101137), pp. 1–14. DOI: 10/gh3bzd.

Flikkema, M.; Castaldi, C.; de Man, A.-P.; Seip, M. (2015): Explaining the Trademark-Innovation Linkage: The Role of Patents and Trademark Filing Strategies.t. In: *Academy of Management Proceedings* 1(16624). DOI: 10/gktncj.

Flikkema, M.; Castaldi, C.; de Man, A.-P.; Seip, M. (2019): Trademarks' Relatedness to Product and Service Innovation: A Branding Strategy Approach. In: *Research Policy* 48(6), pp. 1340–1353. DOI: 10/ghf3hr.

Flikkema, M.; de Man, A.-P.; Castaldi, C. (2014): Are Trademark Counts a Valid Indicator of Innovation? Results of an in-Depth Study of New Benelux Trademarks Filed by SMEs. In: *Industry and Innovation* 21(4), pp. 310–331. DOI: 10/gh3mrx.

Galindo-Rueda, F.; Verger, F. (2016): OECD Taxonomy of Economic Activities Based on R&D Intensity. In: *OECD Science, Technology and Industry Working Papers*. OECD Science, Technology and Industry Working Papers 04. DOI: 10/gktncj.

Gao, G. G.; Hitt, L. M. (2012): Information Technology and Trademarks: Implications for Product Variety. In: *Management Science* 58(6), pp. 1211–1226. DOI: 10/ggfxpj.

Gbadegeshin, S. (2019): The Effect of Digitalization on the Commercialization Process of High-Technology Companies in the Life Sciences Industry. In: *Technology Innovation Management Review* 9(1), pp. 49–63. DOI: 10/gnhpq3.

Gorman, C. (2017): The Role of Trademark Law in the History of US Visual Identity Design, c.1860–1960. In: *Journal of Design History* 30(4), pp. 371–388. DOI: 10/gjzvqj.

Gotsch, M.; Hipp, C. (2012): Measurement of Innovation Activities in the Knowledge-Intensive Services Industry: A Trademark Approach. In: *The Service Industries Journal* 32(13), pp. 2167–2184. DOI: 10/ctvn49.

Graham, S.; Somaya, D. (2004): “Complementary Uses of Patents, Copyrights and Trademarks by Software Firms: Evidence from Litigation”. In: DRUID Conference Paper. Copenhagen: Danish Research Unit for Industrial Dynamics. URL: zinc.zew.de/pub/zew-docs/div/IKT04/Paper_Graham.pdf (visited on 06/01/2023).

Graham, S. J. H.; Hancock, G. (2014): The USPTO Economics Research Agenda. In: *The Journal of Technology Transfer* 39(3), pp. 335–344. DOI: 10/gsjnxr.

Graham, S. J. H.; Hancock, G.; Marco, A. C.; Myers, A. F. (2013): The USPTO Trademark Case Files Dataset: Descriptions, Lessons, and Insights. In: *Journal of Economics & Management Strategy* 22(4), pp. 669–705. DOI: 10/gjk3gg.

Grazzi, M.; Piccardo, C.; Vergari, C. (2019): Concordance and Complementarity in Intellectual Property Instruments. In: *SSRN Electronic Journal*. DOI: 10/gsjnxs.

Hall, B.; Helmers, C.; Rogers, M.; Sena, V. (2014): The Choice between Formal and Informal Intellectual Property: A Review. In: *Journal of Economic Literature* 52(2), pp. 375–423. DOI: 10/gjmph7.

Hänninen, M.; Smedlund, A.; Mitronen, L. (2017): Digitalization in Retailing: Multi-Sided Platforms as Drivers of Industry Transformation. In: *Baltic Journal of Management* 13(2), pp. 152–168. DOI: 10/ghb7q9.

Hardjowirogo, S.-I. (2017): “Instrumentality. On the Construction of Instrumental Identity”. In: *Musical Instruments in the 21st Century: Identities, Configurations, Practices*. Ed. by T. Bovermann; A. de Campo; H. Egermann; S.-I. Hardjowirogo; S. Weinzierl. Singapore: Springer, pp. 9–24. DOI: 10/kms7.

Hauknes, J.; Knell, M. (2009): Embodied Knowledge and Sectoral Linkages: An Input–Output Approach to the Interaction of High- and Low-Tech Industries. In: *Research Policy* 38(3), pp. 459–469. DOI: 10/bfnx4b.

Henderson, R. M.; Clark, K. B. (1990): Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms. In: *Administrative Science Quarterly* 35(1), p. 9. DOI: 10/dzhftg.

Higgins, D. M. (2012): "Forgotten Heroes and Forgotten Issues": Business and Trademark History during the Nineteenth Century. In: *The Business History Review* 86(2), pp. 261–285. DOI: 10/ggcssc.

Hipp, C.; Grupp, H. (2005): Innovation in the Service Sector: The Demand for Service-Specific Innovation Measurement Concepts and Typologies. In: *Research Policy* 34(4), pp. 517–535. DOI: 10/dcznfr.

Hirsch-Kreinsen, H. (2008): "Low-Technology": A Forgotten Sector in Innovation Policy. In: *Journal of Technology Management & Innovation* 3(3). DOI: 10/fwbnws.

IFR (2019a): *Industrial Robots*. International Federation of Robotics (IFR). URL: https://ifr.org/industrial-robots (visited on 12/20/2019).

IFR (2019b): *Service Robots*. International Federation of Robotics (IFR). URL: https://ifr.org/service-robots/ (visited on 12/20/2019).

Kamarul Bahrin, M. A.; Othman, M. F.; Nor Azli, N. H.; Talib, M. F. (2016): Industry 4.0: A Review on Industrial Automation and Robotic. In: *Jurnal Teknologi* 78(6-13), pp. 137–143. DOI: 10/gf9vxp.

Kayser, V. (2017): Comparing Public and Scientific Discourse in the Context of Innovation Systems. In: *Technological Forecasting and Social Change* 115, pp. 348–357. DOI: 10/gf6w7x.

Kim, H.; Hong, S.; Kwon, O.; Lee, C. (2017): Concentric Diversification Based on Technological Capabilities: Link Analysis of Products and Technologies. In: *Technological Forecasting and Social Change* 118, pp. 246–257. DOI: 10/ghnkww.

Kim, J.; Lee, C. (2017): Novelty-Focused Weak Signal Detection in Futuristic Data: Assessing the Rarity and Paradigm Unrelatedness of Signals. In: *Technological Forecasting and Social Change* 120, pp. 59–76. DOI: 10/gkr9kg.

Kim, N.; Lee, H.; Kim, W.; Lee, H.; Suh, J. H. (2015): Dynamic Patterns of Industry Convergence: Evidence from a Large Amount of Unstructured Data. In: *Research Policy* 44(9), pp. 1734–1748. DOI: 10/f7q9pw.

Kleinknecht, A.; Van Montfort, K.; Brouwer, E. (2002): The Non-Trivial Choice between Innovation Indicators. In: *Economics of Innovation and New Technology* 11(2), pp. 109–121. DOI: 10/frx4j8.

Koch, L.-C.; Kopal, R. (2014): Klassifikation von Musikinstrumenten – Zum 100-jährigen Bestehen der Hornbostel-Sachs-Systematik. In: *Zeitschrift für Ethnologie* 139(2), pp. 281–302. JSTOR: 24365030.

Kohler, M.; Feldmann, N.; Kimbrough, S. O.; Fromm, H. (2014): Service Innovation Analytics: Leveraging Existing Unstructured Data to Assess Service Innovation Capability. In: *International Journal of Information System Modeling and Design* 5(2), pp. 1–21. DOI: 10/gkr9r4.

Kozinski, A. (1993): Trademarks Unplugged. In: *New York University Law Review* 68(4), pp. 960–978. URL: https://heinonline.org/HOL/LandingPage?handle=hein.journals/nylr68&div=37&id=&page= (visited on 06/01/2023).

Kreuchauff, F.; Korzinov, V. (2017): A Patent Search Strategy Based on Machine Learning for the Emerging Field of Service Robotics. In: *Scientometrics* 111(2), pp. 743–772. DOI: 10/f959t8.

Kuznets, S. (1973): Modern Economic Growth: Findings and Reflections. In: *The American Economic Review* 63(3), pp. 247–258. JSTOR: 1914358.

Larsen, V. H.; Thorsrud, L. A. (2019): The Value of News for Economic Developments. In: *Journal of Econometrics* 210(1), pp. 203–218. DOI: 10/gf5nd5.

Lee, M.; Lee, S. (2017): Identifying New Business Opportunities from Competitor Intelligence: An Integrated Use of Patent and Trademark Databases. In: *Technological Forecasting and Social Change* 119, pp. 170–183. DOI: 10/gd4whf.

Lee, S.; Rha, J. (2018): A Network Text Analysis of Published Papers in Service Business, 2007–2017: Research Trends in the Service Sector. In: *Service Business* 12(4), pp. 809–831. DOI: 10/gfm3d5.

Lely (2021): *Farm of the Future*. URL: https://www.lely.com/ (visited on 10/19/2021).

Lerch, A. (2018): "The Relation Between Music Technology and Music Industry". In: *Springer Handbook of Systematic Musicology*. Ed. by R. Bader. Springer Handbooks. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 899–909. DOI: `10/kms8`.

Levin, R. C.; Klevorick, A. K.; Nelson, R. R.; Winter, S. G.; Gilbert, R.; Griliches, Z. (1987): Appropriating the Returns from Industrial Research and Development. In: *Brookings Papers on Economic Activity* 1987(3), pp. 783–831. DOI: `10/cmrhjk`.

Li, Q. (2016): "Do Analysts Understand Innovation? Evidence from Patents and Trademarks". Irvine: University of California. URL: `https://escholarship.org/uc/item/4rg4q80x` (visited on 11/19/2018).

Lindman, M.; Scozzi, B.; Otero-Neira, C. (2008): Low-tech, Small- and Medium-sized Enterprises and the Practice of New Product Development: An International Comparison. In: *European Business Review* 20(1), pp. 51–72. DOI: `10/b9fb8p`.

Liu, B.; Zhang, L. (2012): "A Survey of Opinion Mining and Sentiment Analysis". In: *Mining Text Data*. Ed. by C. C. Aggarwal; C. Zhai. Boston, MA: Springer US, pp. 415–464. DOI: `10/km5w`.

Llerena, P.; Millot, V. (2020): Are Two Better than One? Modelling the Complementarity between Patents and Trademarks across Industries. In: *Industry and Innovation* 27(1-2), pp. 52–79. DOI: `10/gsjnxt`.

Llopis-Albert, C.; Rubio, F.; Valero, F. (2021): Impact of Digital Transformation on the Automotive Industry. In: *Technological Forecasting and Social Change* 162, p. 120343. DOI: `10/ghjg3q`.

Loshin, D. (2013): "Market and Business Drivers for Big Data Analytics". In: *Big Data Analytics*. Elsevier, pp. 1–9. DOI: `10/kms9`.

Malmberg, C. (2005): *Trademarks Statistics as Innovation Indicator? - A Micro Study*. 17. Lund University, CIRCLE - Center for Innovation, Research and Competences in the Learning Economy.

Marcuse, S.; Bowles, E. A. (2015): *Percussion Instrument*. Encyclopedia Britannica. URL: `https://www.britannica.com/art/percussion-instrument` (visited on 04/09/2021).

Masubuchi, T.; Kato, H. (1994): "Musical Tone Controller". U.S. pat. D345853. URL: `https://pdfpiw.uspto.gov/.piw?Docid=D0347853` (visited on 03/21/2022).

Medhat, W.; Hassan, A.; Korashy, H. (2014): Sentiment Analysis Algorithms and Applications: A Survey. In: *Ain Shams Engineering Journal* 5(4), pp. 1093–1113. DOI: `10/gfkmhz`.

Mendonça, S. (2006): The Revolution within: ICT and the Shifting Knowledge Base of the World's Largest Companies. In: *Economics of Innovation and New Technology* 15(8), pp. 777–799. DOI: `10/dmb6f6`.

Mendonça, S. (2009): Brave Old World: Accounting for 'High-Tech' Knowledge in 'Low-Tech' Industries. In: *Research Policy* 38(3), pp. 470–482. DOI: `10/fb63c8`.

Mendonça, S.; Pereira, T. S.; Godinho, M. M. (2004): Trademarks as an Indicator of Innovation and Industrial Change. In: *Research Policy* 33(9), pp. 1385–1404. DOI: `10/dr4z85`.

Mi, L.; Huang, L.-c.; Han, Z.-x.; Miao, H.; Wu, F. (2021): Forecasting and Evaluating Emerging Technologies Based on Supply and Demand Matching – a Case Study of China's Gerontechnology. In: *Technology Analysis & Strategic Management*, pp. 1–17. DOI: `10/gkr9m9`.

Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. (2013): "Efficient Estimation of Word Representations in Vector Space". arXiv: `1301.3781`.

Millot, V. (2009): *Trademarks as an Indicator of Product and Marketing Innovations*. 2009/06. Paris: OECD Publishing. DOI: `10/dwnnwn`.

MIMO-Consortium (2011): *Revision of the Hornbostel-Sachs Classification of Musical Instruments by the MIMO Consortium*. Musical Instrument Museums Online (MIMO) Consortium. URL: http://mimo-international.com/documents/Hornbostel%20Sachs.pdf (visited on 04/06/2021).

Myers, A. (2013): *What Is behind the Growth in Trademark Filings? An Analysis of United States Data Background Research for the World IP Report 2013*. 2013-1. Alexandria, Virgina: U.S. Patent and Trademark Office. URL: http://ssrn.com/abstract=2849614.

Nam, Y.; Barnett, G. A. (2011): Globalization of Technology: Network Analysis of Global Patents and Trademarks. In: *Technological Forecasting and Social Change* 78(8), pp. 1471–1485. DOI: 10/bs62sx.

Neuhäusler, P. (2009): *Formal vs. Informal Protection Instruments and the Strategic Use of Patents in an Expected-Utility Framework*. 20. Karlsruhe: Fraunhofer ISI, p. 32. DOI: 10/kmtb.

Neuhäusler, P.; Frietsch, R. (2015): "Patent Activities in Non-R&D-Intensive Technology Areas". In: *Low-Tech Innovation*. Ed. by O. Som; E. Kirner. Cham: Springer International Publishing, pp. 51–66. DOI: 10/km5z.

OECD (2011): *ISIC Rev. 3 Technology Intensity Definition*. OECD Directorate for Science, Technology and Industry, Economic Analysis and Statistics Division. URL: https://www.oecd.org/sti/ind/48350231.pdf (visited on 11/18/2021).

OECD (2015): *OECD Innovation Strategy 2015 - An Agenda for Policy Action*. Paris, France: OECD. URL: https://ec.europa.eu/futurium/en/system/files/ged/27-oecdandcouncilmnisterailmeeting-innovation-strategy-2015.pdf (visited on 12/08/2021).

OECD (2021): *International Trade by Commodity Statistics Volume 2021 Issue 3: Belgium, Canada, Germany, Italy, Japan, Korea, United States*. Paris, France. DOI: 10/kmtc.

OECD; Eurostat (2005): *Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data*. 3rd ed. 54261 2005. Paris, France: OECD Publishing. DOI: 10/bg2nzq.

OECD; Eurostat (2018): *Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation, 4th Edition*. 4th ed. The Measurement of Scientific, Technological and Innovation Activities. Paris/ Eurostat, Luxembourg: OECD Publishing. DOI: 10/fb2r.

Ozcan, S.; Suloglu, M.; Sakar, C. O.; Chatufale, S. (2021): Social Media Mining for Ideation: Identification of Sustainable Solutions and Opinions. In: *Technovation* 107(102322), pp. 1–12. DOI: 10/gkhq35.

Pakes, A.; Griliches, Z. (1980): Patents and R&D at the Firm Level: A First Report. In: *Economics Letters* 5(4), pp. 377–381. DOI: 10/c8hcqh.

Pons, P.; Csardi, G. (2021): "Cluster_walktrap - Community Structure via Short Random Walks". In: Nepusz, T. *Package 'igraph' - Network Analysis and Visualization*. 1.2.7. CRAN, pp. 81–82. URL: https://igraph.org/r/#docs (visited on 11/22/2021).

Pons, P.; Latapy, M. (2005): "Computing Communities in Large Networks Using Random Walks (Long Version)". arXiv: physics/0512106.

Potepa, J.; Welch, K. T. (2018): *Calibrating Innovation Proxies with an Accurate Tobin's Q and Appraised Innovation Value*. SSRN Scholarly Paper ID 2990481. Rochester, NY: Social Science Research Network. DOI: 10/kmtd.

Raffaeli, R.; Germani, M. (2011): Advanced Computer Aided Technologies for Design Automation in Footwear Industry. In: *International Journal on Interactive Design and Manufacturing (IJIDeM)* 5(3), pp. 137–149. DOI: 10/dfxdmb.

Rammer, C. (2002): *Patente und Marken als Schutzmechanismen für Innovationen*. 11-2003. Mannheim: Zentrum für Europäische Wirtschaftsforschung (ZEW), p. 16. URL: http://ftp.zew.de/pub/zew-docs/gutachten/SDI_11_03.pdf (visited on 11/12/2021).

Ribeiro, L. C.; dos Santos, U. P.; Muzaka, V. (2022): Trademarks as an Indicator of Innovation: Towards a Fuller Picture. In: *Scientometrics* 127(1), pp. 481–508. DOI: 10/gqwcts.

Roberts, M. E.; Stewart, B. M.; Tingley, D. (2019): Stm: R Package for Structural Topic Models. In: *Journal of Statistical Software* 91(2). DOI: 10/ggc8cz.

Roberts, M. E.; Tingley, D.; Stewart, B. M.; Airoldi, E. M. (2013): "The Structural Topic Model and Applied Social Science". In: Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluatoin, pp. 1–20. URL: https://api.semanticscholar.org/CorpusID:59893873.

Roberts, M. E.; Stewart, B. M.; Tingley, D.; Lucas, C.; Leder-Luis, J.; Gadarian, S. K.; Albertson, B.; Rand, D. G. (2014): Structural Topic Models for Open-Ended Survey Responses. In: *American Journal of Political Science* 58(4), pp. 1064–1082. DOI: 10/f6mv8k.

Rogers, E. M. (1983): *Diffusion of Innovations*. 3rd ed. New York, London: Free Press; Collier Macmillan. 453 pp. ISBN: 978-0-02-926650-2.

Ronthaler, M.; Fernández, J. d. G.; Vögele, T. (2011): Automatisierung in der Schuhindustrie. In: *Industrie Management* 27, pp. 59–61.

Ryan, B.; Gross, N. (1943): The Diffusion of Hybrid Seed Corn in Two Iowa Communities. In: *Rural Sociology* 8(1), pp. 15–23.

Sachs, C. (1940): *The History of Musical Instruments*. New York: W. W. Norton & Company, Inc. ISBN: 0-393-02068-1. URL: https://archive.org/details/the_history_of_musical_instruments_curt_sachs/page/n15/mode/2up (visited on 04/06/2021).

Sandner, P. G.; Block, J. (2011): The Market Value of R&D, Patents, and Trademarks. In: *Research Policy* 40(7), pp. 969–985. DOI: 10/bzpmcv.

Schumpeter, J. A. (2010): *Konjunkturzyklen: Eine Theoretische, Historische Und Statistische Analyse Des Kapitalistischen Prozesses*. Vol. 8453. UTB / UTB. Vandenhoeck & Ruprecht. 1132 pp. ISBN: 978-3-8252-8453-4.

Seip, M.; Castaldi, C.; Flikkema, M.; de Man, A.-P. (2018): The Timing of Trademark Application in Innovation Processes. In: *Technovation* 72–73, pp. 34–45. DOI: 10/gsjnxv.

Semadeni, M. (2006): Minding Your Distance: How Management Consulting Firms Use Service Marks to Position Competitively. In: *Strategic Management Journal* 27(2), pp. 169–187. DOI: 10/c3pd8k.

Semadeni, M.; Anderson, B. S. (2010): The Follower's Dilemma: Innovation and Imitation in the Professional Services Industry. In: *Academy of Management Journal* 53(5), pp. 1175–1193. DOI: 10/cw6gpb.

Shen, Y.-C.; Wang, M.-Y.; Yang, Y.-C. (2020): Discovering the Potential Opportunities of Scientific Advancement and Technological Innovation: A Case Study of Smart Health Monitoring Technology. In: *Technological Forecasting and Social Change* 160, p. 120225. DOI: 10/gkr9jp.

Som, O.; Kirner, E., eds. (2015): *Low-Tech Innovation*. Ed. by O. Som; E. Kirner. Cham: Springer International Publishing. DOI: 10/dvff.

Song, K.; Kim, K. S.; Lee, S. (2017): Discovering New Technology Opportunities Based on Patents: Text-mining and F-term Analysis. In: *Technovation* 60–61, pp. 1–14. DOI: 10/gjthwm.

Sorbe, S.; Gal, P.; Millot, V. (2018): *Can Productivity Still Grow in Service-Based Economies? Literature Overview and Preliminary Evidence from OECD Countries Economics Department*. 1531. OECD.

URL: https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ECO/WKP%282018%2979&docLanguage=En (visited on 12/08/2021).

SoundDiplomacy (2021): *Der Wert von Musik - Ökonomische Qirkungsanalyse Des Musikökosystems*. Berlin: Sound Diplomacy. URL: https://static1.squarespace.com/static/593eb2a837c58172ed556cbb/t/6101412a53472e6bd2f8ae3e/1627472184940/2021_Der_Wert_von_Musik_O%CC%88konomische_Wirkungsanalyse.pdf (visited on 07/29/2021).

Statista (2021): *Prognose: Umsatz der Branche Herstellung von Musikinstrumenten in Deutschland von 2012 bis 2019 und Prognose bis zum Jahr 2025 (in Millionen Euro)*. Statista. URL: https://de.statista.com/prognosen/313870/herstellung-von-musikinstrumenten-umsatz-in-deutschland (visited on 10/08/2021).

Statista (2022): *Musical Instruments - Worldwide | Statista Market Forecast*. Statista. URL: https://www.statista.com/outlook/cmo/toys-hobby/musical-instruments/worldwide?currency=EUR (visited on 12/04/2022).

TFSR (1946): *Chapter 15 - Trademark Act of 1946, as Amended*. URL: https://tmep.uspto.gov/RDMS/TFSR/current#/current/TFSR-15USCd1e1.html (visited on 05/18/2021).

Théberge, P. (2017): "Musical Instruments as Assemblage". In: *Musical Instruments in the 21st Century: Identities, Configurations, Practices*. Ed. by T. Bovermann; A. de Campo; H. Egermann; S.-I. Hardjowirogo; S. Weinzierl. Singapore: Springer, pp. 59–66. DOI: 10/km5v.

Thomä, J.; Bizer, K. (2013): To Protect or Not to Protect? Modes of Appropriability in the Small Enterprise Sector. In: *Research Policy* 42(1), pp. 35–49. DOI: 10/f4kv3t.

Thornhill, S. (2006): Knowledge, Innovation and Firm Performance in High- and Low-Technology Regimes. In: *Journal of Business Venturing* 21(5), pp. 687–703. DOI: 10/d882k9.

Trajtenberg, M. (1990): A Penny for Your Quotes: Patent Citations and the Value of Innovations. In: *The RAND Journal of Economics* 21(1), pp. 172–187. JSTOR: 2555502.

Tseng, Y.-H.; Lin, C.-J.; Lin, Y.-I. (2007): Text Mining Techniques for Patent Analysis. In: *Information Processing & Management*. Patent Processing 43(5), pp. 1216–1247. DOI: 10/cb4vqv.

Uchiyama, Y.; Suzuki, S. (1988): "Tone Generator Having a Variable Number of Channels with a Variable Number of Operating Units". U.S. pat. 4788896A. Nippon Gakki Co. Ltd. URL: https://pdfpiw.uspto.gov/.piw?Docid=04788896 (visited on 03/21/2022).

USPTO (1990): *Trademark Serial No. 74024399: Parallel Digital Imaging*. USPTO and Rogers Instrument Corporation. URL: https://tsdr.uspto.gov/#caseNumber=74024399&caseType=SERIAL_NO&searchType=statusSearch (visited on 03/21/2022).

USPTO (2005): *Handbook of Classification*. United States Patent and Trademark Office (USPTO). URL: https://www.uspto.gov/sites/default/files/web/offices/opc/documents/handbook.pdf (visited on 05/07/2021).

USPTO (2012): *Classes within the U.S. Classification System*. United States Patent and Trademark Office (USPTO). URL: https://www.uspto.gov/sites/default/files/patents/resources/classification/overview.pdf (visited on 05/07/2021).

USPTO (2015): *Patents: Getting Started - Patent Basics - General Information Concerning Patents*. Alexandria, Virgina: United States Patent and Trademark Office (USPTO). URL: https://www.uspto.gov/patents-getting-started/general-information-concerning-patents#heading-2 (visited on 03/06/2019).

USPTO (2017): "2100 Patentability". In: *Manual of Patent Examining Procedure (MPEP)*. 9. United States Patent and Trademark Office (USPTO). URL: https://www.uspto.gov/web/offices/pac/mpep/old/mpep_E9R08-2017.htm.

USPTO (2019): *Trademarks: Getting Started - Trademark Basics*. United States Patent and Trademark Office USPTO. URL: https://www.uspto.gov/trademarks-getting-started/trademark-basics (visited on 03/06/2019).

USPTO (2020a): *Manual of Patent Examining Procedure - Chapter 0900 - Section 907 Locarno Classification Designations [R-07.2015]*. United States Patent and Trademark Office (USPTO). URL: https://www.uspto.gov/web/offices/pac/mpep/s907.html (visited on 05/10/2021).

USPTO (2020b): *Protecting Your Trademark - Enhancing Your Rights through Federal Registration*. United States Patent and Trademark Office (USPTO). URL: https://www.uspto.gov/sites/default/files/documents/Basic-Facts-Booklet.pdf.

USPTO (2021a): *CPC Cooperative Patent Classification - G PHYSICS*. URL: https://www.uspto.gov/web/patents/classification/cpc/html/cpc-G.html#G10 (visited on 04/07/2021).

USPTO (2021b): *Patent Classification*. URL: https://www.uspto.gov/patents/search/classification-standards-and-development (visited on 05/07/2021).

USPTO (2021c): *Seven - Classification of Design Patents*. URL: https://www.uspto.gov/patents/laws/examination-policy/seven-classification-design-patents (visited on 11/15/2021).

USPTO (2022a): *CPC Definition - G10H Electrophonic Musical Instruments*. URL: https://www.uspto.gov/web/patents/classification/cpc/html/cpc-G10H.html#G10H (visited on 03/21/2022).

USPTO (2022b): *Design Patent Application Guide*. URL: https://www.uspto.gov/patents/basics/types-patent-applications/design-patent-application-guide (visited on 04/24/2022).

USPTO (2022c): *Nonprovisional (Utility) Patent Application Filing Guide*. URL: https://www.uspto.gov/patents/basics/types-patent-applications/nonprovisional-utility-patent (visited on 03/21/2022).

Veugelers, R.; Schneider, C. (2018): Which IP Strategies Do Young Highly Innovative Firms Choose? In: *Small Business Economics* 50(1), pp. 113–129. DOI: 10/gsjnxw.

Von Graevenitz, G.; Ashmead, R.; Greenhalgh, C. (2021): *Cluttering and Non-Use of Trade Marks in Europe*. Newport: The Intellectual Property Office. ISBN: 978-1-910790-22-9. URL: https://www.gov.uk/government/publications/cluttering-and-non-use-of-trade-mark-in-europe (visited on 05/18/2021).

Von Hornbostel, E. M. (1933): The Ethnology of African Sound-Instruments. Comments on *Geist Und Werden Der Musikinstrumente* by C. Sachs. In: *Africa* 6(2), pp. 129–157. DOI: 10/cntfhv.

Von Hornbostel, E. M.; Sachs, C. (1914): Systematik der Musikinstrumente. In: *Zeitschrift für Ethnologie* 64, pp. 553–590. URL: https://www.archive.org/stream/zeitschriftfre46berluoft#page/553/mode/1up (visited on 11/17/2019).

VuMA (2021): *Konsumenten Im Fokus - Basisinformationen Für Fundierte Mediaentscheidungen*. Arbeitsgemeinschaft Verbrauchs- und Medienanalyse. URL: https://www.vuma.de/fileadmin/user_upload/PDF/berichtsbaende/VuMA_Berichtsband_2021.pdf (visited on 10/08/2021).

Wang, B.; Wang, Z. (2018): Heterogeneity Evaluation of China's Provincial Energy Technology Based on Large-Scale Technical Text Data Mining. In: *Journal of Cleaner Production* 202, pp. 946–958. DOI: 10/gfjm3n.

Wang, T.-M.; Tao, Y.; Liu, H. (2018): Current Researches and Future Development Trend of Intelligent Robot: A Review. In: *International Journal of Automation and Computing* 15(5), pp. 525–546. DOI: 10/ggb2m3.

WIPO (1979): *Paris Convention for the Protection of Industrial Property*. World Intellectual Property Organization (WIPO). URL: wipolex.wipo.int/en/text/287556 (visited on 05/10/2021).

WIPO (2013): *World Intellectual Property Report Brands – Reputation and Image in the Global Marketplace*. URL: https://www.wipo.int/publications/en/details.jsp?id=384&plang=EN (visited on 05/11/2019).

WIPO (2019a): *Nice Classification*. URL: https://www.wipo.int/classifications/nice/en/index.html (visited on 03/07/2019).

WIPO (2019b): *Trademarks*. URL: https://www.wipo.int/trademarks/en/index.html (visited on 03/12/2019).

WIPO (2020a): *Class 15*. URL: https://www.wipo.int/classifications/nice/nclpub/en/fr/?basic_numbers=show&class_number=15&explanatory_notes=show&lang=en&menulang=en&mode=flat&notion=&pagination=no&version=20200101 (visited on 04/09/2021).

WIPO (2020b): *Guide to the International Patent Classification*. Geneva, Switzerland: World Intellectual Property Organization (WIPO). URL: https://www.wipo.int/edocs/pubdocs/en/wipo_guide_ipc_2020.pdf (visited on 11/10/2021).

WIPO (2020c): *International Patent Classification (IPC)*. Geneva, Switzerland: World Intellectual Property Organization (WIPO). URL: https://tind.wipo.int/record/43106/files/wipo_brochure_ipc.pdf (visited on 11/10/2021).

WIPO (2020d): *What Is Intellectual Property?* Geneva, Switzerland: World Intellectual Property Organization (WIPO), p. 28. URL: https://www.wipo.int/edocs/pubdocs/en/wipo_pub_450_2020.pdf (visited on 11/12/2021).

WIPO (2021a): *About the Locarno Classification*. URL: https://www.wipo.int/classifications/locarno/en/preface.html (visited on 05/10/2021).

WIPO (2021b): *About the Nice Classification*. URL: https://www.wipo.int/classifications/nice/en/preface.html (visited on 05/10/2021).

WIPO (2021c): *Industrial Designs*. URL: https://www.wipo.int/designs/en/index.html (visited on 05/10/2021).

WIPO (2021d): *Locarno Agreement (Total Contracting Parties: 59)*. URL: https://wipolex.wipo.int/en/treaties/ShowResults?search_what=C&treaty_id=14 (visited on 05/10/2021).

WIPO (2021e): *Locarno Classification*. URL: https://www.wipo.int/classifications/locarno/en/index.html (visited on 05/10/2021).

WIPO (2021f): *Locarno Classification - List of Classes and Subclasses with Explanatory Notes*. URL: https://www.wipo.int/classifications/locarno/locpub/en/fr/?class_number=32&explanatory_notes=show&lang=en&menulang=en&notion=class_headings&subclasses=show&version=20210101 (visited on 05/10/2021).

WIPO (2021g): *Nice Agreement (Total Contracting Parties: 89)*. URL: https://wipolex.wipo.int/en/treaties/ShowResults?search_what=C&treaty_id=12 (visited on 05/10/2021).

WIPO (2021h): *Patents*. URL: https://www.wipo.int/patents/en/index.html (visited on 05/18/2021).

WIPO (2021i): *Summary of the Paris Convention for the Protection of Industrial Property (1883)*. World Intellectual Property Organization (WIPO). URL: https://www.wipo.int/treaties/en/ip/paris/summary_paris.html (visited on 05/10/2021).

WIPO (2022a): *Nice Classification - List of Classes with Explanatory Notes*. URL: https://www.wipo.int/classifications/nice/nclpub/en/fr/?explanatory_notes=show&lang=en&menulang=en&notion=class_headings&version=20220101 (visited on 03/22/2022).

WIPO (2022b): *Paris Convention (Total Contracting Parties: 177)*. URL: https://wipolex.wipo.int/en/treaties/ShowResults?search_what=C&treaty_id=2 (visited on 10/28/2022).

WoS (2021): *Publication Year 2020 from Web of Science Core Collection*. URL: https://www.webofscience.com/wos/woscc/summary/5847144f-02c6-4d9c-934c-28f4ea67ff28-11441487/relevance/1 (visited on 11/05/2021).

Wu, H.; Shen, G.; Lin, X.; Li, M.; Zhang, B.; Li, C. Z. (2020): Screening Patents of ICT in Construction Using Deep Learning and NLP Techniques. In: *Engineering, Construction and Architectural Management* 27(8), pp. 1891–1912. DOI: 10/gkr9nr.

Wydra, S.; Nusser, M. (2015): "Economic Relevance and the Future Potential of Non-R&D-Intensive Industries". In: *Low-Tech Innovation*. Ed. by O. Som; E. Kirner. Cham: Springer International Publishing, pp. 33–50. DOI: 10/km52.

Yamaha (2020): *Yamaha Group Annual Report 2020*. Japan: Yamaha Corporation. URL: https://www.yamaha.com/en/ir/publications/pdf/an-2020e.pdf (visited on 11/16/2021).

Zhang, Y.; Huang, Y.; Porter, A. L.; Zhang, G.; Lu, J. (2019): Discovering and Forecasting Interactions in Big Data Research: A Learning-Enhanced Bibliometric Study. In: *Technological Forecasting and Social Change* 146, pp. 795–807. DOI: 10/gd4whz.

Zhou, X.; Guo, Y.; Li, F.; Wang, J.; Wei, H.; Yu, M.; Chen, S. (2020): Identifying and Assessing Innovation Pathways for Emerging Technologies: A Hybrid Approach Based on Text Mining and Altmetrics. In: *IEEE Transactions on Engineering Management*, pp. 1–12. DOI: 10/gkr9pc.

Zhu, D.; Porter, A. L. (2002): Automated Extraction and Visualization of Information for Technological Intelligence and Forecasting. In: *Technological Forecasting and Social Change* 69(5), pp. 495–506. DOI: 10/crf849.

Zolas, N.; Lybbert, T. J.; Bhattacharyya, P. (2017): An 'Algorithmic Links with Probabilities' Concordance for Trademarks with an Application Towards Bilateral IP Flows. In: *The World Economy* 40(6), pp. 1184–1213. DOI: 10/gsjnxx.

A Appendix to Chapter 2: Foundations

This appendix relates to Chapter 2: Foundations.

A.1 Patents

```sql
select distinct pa.id, year(date) as year
from uspto.patent pa
where type in ('utility') AND
year(date) > 1977 AND year(date) <= 2016
```

Query A.1: Granted Patents per Year

A.2 Trademarks

```sql
select distinct cf.serial_no, cl.intl_class_cd as class,  year(filing_dt) as year
from trademarks.case_file as cf
join trademarks.intl_class cl on cl.serial_no = cf.serial_no
where year(filing_dt) > 1977 and year(filing_dt) <= 2016 and registration_no not in ('0000000')
and cl.intl_class_cd not in ('200', 'A', 'B')
```

Query A.2: Registered Trademark Filings in Nice Classes per Year

```sql
select distinct cf.serial_no as serial_no, year(filing_dt) as year, registration_no
from trademarks.case_file as cf
where year(filing_dt) > 1977 and year(filing_dt) <= 2016
```

Query A.3: Trademark Filings per Year

A.3 Designs

```sql
select distinct pa.id, year(date) as year
from uspto.patent pa
where type in ('design') AND
year(date) > 1977 AND year(date) <= 2016
```

Query A.4: Granted Designs per Year

```sql
select count(distinct cf.serial_no), cl.intl_class_cd
from trademarks.case_file as cf
join trademarks.intl_class cl on cl.serial_no = cf.serial_no
where year(filing_dt) > 1977 and year(filing_dt) <= 2016 and registration_no not in ('0000000')
group by cl.intl_class_cd
order by count(distinct cf.serial_no)
```

Query A.5: Granted Designs at the USPTO between 1978 and 2016 per U.S. Design Class

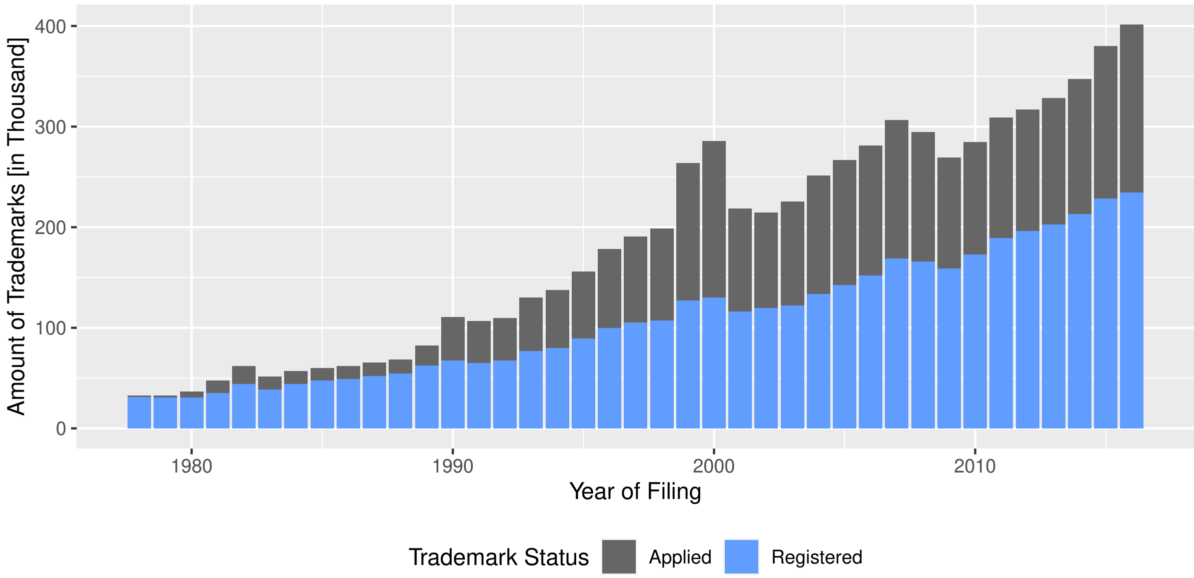


Figure A.1: Total Yearly Development of Trademark Filings between 1978 and 2016.
The figure displays the yearly filings of trademarks at the USPTO differentiated for applied and registered trademarks from 1978 until 2016.
Source: Own representation based on USPTO data (see Query A.3).

B Appendix to Chapter 3: Textual Data in Innovation Research

B.1 Data Selection

```
(TITLE-ABS-KEY (text AND mining) AND (LIMIT-TO ( PUBSTAGE, "final"))
```

Query B.1: Scopus Search (by January 01, 2022)

```
(TI = (text*) OR AB = (text*)) AND (TI = (innov*) OR AB = (innov*))
AND SU = (Business & Economics)
NOT TS = (textile)
AND DOCUMENT TYPES: (Articles)
AND Language: English
```

Query B.2: Web of Science Search Resulting in 250 Documents (by June 21, 2021)

```
((TITLE(text*) OR ABS(text*)) AND (TITLE(innov*) OR ABS(innov*)))
AND (SUBJAREA (econ) OR SUBJAREA (busi))
AND NOT TITLE-ABS-KEY(textile)
AND (LIMIT-TO(DOCTYPE, "ar"))
AND (LIMIT-TO(LANGUAGE, "English"))
```

Query B.3: SCOPUS Search Resulting in 789 Documents (by June 21, 2021)

Table B.1: Keywords Selected for the Sample Identification

text mining (63), big data (13), sentiment analysis (10), machine learning (8), natural language processing (8), textual analysis (7), content analysis (7), data mining (6), text (6), text analysis (6), text analytics (6), bibliometrics (6), leximancer (5), deep learning (4), topic modeling (4), cluster analysis (4), text classification (3), text messaging (3), unstructured data (3), bibliometric analysis (3), clustering (3), tech mining (3), big data analytics (2), artificial intelligence (2), data analytics (2), latent dirichlet allocation (lda) (2), lda topic model (2), network text analysis (2), semantic analysis (2), supervised learning (2), support vector machine (2), support vector machines (2), text-mining (2), text clustering (2), topic modelling (2), artificial intelligence algorithms (1), artificial neural networks (1), automated content analysis (1), automated sentiment analysis (1), big' data (1), big data analytics techniques (1), computer-aided text analysis (1), convolutional neural network (1), cosine similarity of tf-idf vectors (1), latent dirichlet allocation (1), lda (1), lexical semantics (1), lexicon-based approach (1), machine learning approach (1), natural language processing (nlp) (1), opining mining (1), opinion analysis (1), patent mining (1), patent text mining (1), qualitative and quantitative methods (1), qualitative meta-analysis (1), semantic brand score (1), semantic topological space (1), semantic web (1), semantic web mining (1), semi-supervised learning (1), sentimental analysis (1), social media data mining (1), term clumping (1), terms mining approach (1), text-as-data analysis (1), text -mining (1), text and sentimental analysis (1), text as data (1), text documents (1), text innovation (1), textual (1), tfidf (1), tokenization (1), topic model (1), vector semantics (1), web scraping (1), webscraping (1), word embedding (1), word frequency (1), wordformation (1), clustering algorithms (1), co-citation (1), co-occurrence-based analysis (1), co-occurrence analysis (1), co-occurrence networks (1), computer based learning (1), context-based concepts (1), contextual framing (1), latent aspect rating analysis (lara) (1), latent aspect rating regression (1), meaning extraction (1), scientometric (1), scientometric evaluation (1), scientometrics (1), sequential pattern mining (1), technology maps (1), technology mining and trend analysis (1), large-scale text data (1), user-oriented context-to-text recognition for learning (1)

The table provides an overview of all keywords selected for the sample identification. The keywords are sorted in descending order according to their occurrence in the data set. The occurrence is provided in brackets.
Source: Own overview based on the data extraction of Query B.2 and Query B.3.

Table B.2: Keywords Not-Selected for the Sample Identification

innovation (71), entrepreneurship (13), open innovation (12), education (10), management (10), technology (10), creativity (9), discourse (9), patents (9), sustainability (9), knowledge management (8), analysis (7), patent analysis (7), sustainable development (7), critical discourse analysis (6), governance (6), information technology (6), marketing (6), service innovation (6), social media (6), classification (5), crowdsourcing (5), history (5), learning (5), performance (5), qualitative research (5), technological innovation (5), women (5), art (4), china (4), communication (4), consumer behaviour (4), consumers (4), culture (4), design (4), emerging technologies (4), empowerment (4), engagement (4), ethics (4), experiential learning (4), family business (4), foresight (4), innovation management (4), knowledge (4), marketing strategy (4), mining (4), netnography (4), rhetoric (4), teaching (4), technology forecasting (4), user-generated content (4), discovery (3), agency (3), australia (3), blended learning (3), blockchain (3), blogs (3), brand personality (3), business value (3), case studies (3), case study (3), climate change (3), creative industries (3), curriculum design (3), customer satisfaction (3), delphi method (3), discourse analysis (3), e-commerce (3), eco-innovation (3), efficiency (3), entrepreneur (3), entrepreneurialism (3), ethnography (3), fashion (3), fintech (3), flexibility (3), food innovation (3), healthcare (3), internet (3), literature review (3), management accounting (3), media (3), methodology (3), methods (3), motivation (3), network (3), networks (3), online communities (3), online community (3), online reviews (3), ontology (3), product development (3), productivity (3), qualitative (3), quality (3), radical innovation (3), recommendation system (3), research (3), retailing (3), risk management (3), schumpeter (3), security (3), sharing economy (3), social enterprise (3), social network analysis (3), tourism (3), university (3), web (3), language (2), triz (2), crowdfunding (2), visualization (2), academic entrepreneurship (2), accommodation (2), accounting research (2), accounting system (2), administrative reform (2), advertising concerts (2), advertising history (2), aesthetics (2), agriculture (2), airbnb (2), algorithmic management (2), analytics (2), annual reports (2), architecture (2), as-if (2), assessment (2), attributes (2), augmented reality (2), authentication (2), benchmarking (2), binary opposition (2), biotechnology (2), brand associations (2), business (2), business-to-business marketing (2), business model (2), business process (2), capital market (2), change (2), change management (2), choice (2), citation analysis (2), cloud computing (2), co-creation (2), collaboration (2), collective housing (2), commercialisation of leisure (2), companies (2), competencies (2), competitive advantage (2), competitiveness (2), conjoint analysis (2), consumer (2), consumer marketing (2), consumer preferences (2), contingency theory (2), corporate culture (2), corporate social responsibility (2), covid-19 (2), czech republic (2), data security (2), decision-making (2), demand volatility (2), design science (2), development (2), digital ecosystems (2), digital transformation (2), digitalization of management (2), disruptive innovation (2), disruptive technologies (2), double-entry bookkeeping (2), dynamic capabilities (2), eco-design (2), economics (2), eighteenth century (2), electronic commerce (2), emerging technology (2), employer branding (2), engineers (2), entrepreneurs (2), experience (2), f53 (2), factor analysis (2), fashion industry (2), feature selection (2), financial crisis (2), financial risk (2), financial service systems (2), forecasting (2), framework (2), friedman (2), gerontechnology (2), globalization (2), grand duchy of tuscany (2), green-oriented design (2), green technologies (2), grounded theory (2), hospitality (2), human resource management (2), hybrid (2), ict (2), idea generation (2), ideology (2), impulse purchasing (2), inclusive innovation (2), industrial marketing (2), industries (2), industry 4.0 (2), information asymmetry (2), information management (2), information retrieval (2), information systems (2), innovation indicators (2), innovation projects (2), innovative teaching (2), innovativeness (2), institutional theory (2), institutions (2), instrumentalism (2), insurance (2), intellectual capital (2), interdisciplinary (2), interdisciplinary training (2), internet of things (2), knowledge creation (2), knowledge structure (2), leadership (2), lean production (2), legislation (2), link prediction (2), literature-based discovery (2), literature-related discovery (2), machine (2), management history (2), management volume (2), measurement (2), memory (2), microstructure (2), mobile commerce (2), modeling (2), multidisciplinary (2), municipality (2), myth (2), narratives (2), new engineering education (2), new product development (2), new technology (2), news (2), newspaper advertising (2), online review analysis (2), online vacation rentals (2), opportunity (2), organizational change (2), organizational design (2), organizational innovation (2), organizational restructuring (2), perception (2), personalization (2), persuasion in advertising (2), philosophy (2), pitch (2), pluralism (2), power (2), practice theory (2), product life cycle (2), project management (2), project owner (2), qualitative-quantitative approach (2), qualitative approach (2), qualitative content analysis (2), qualitative methods (2), quality-related managerial insights (2), r&d (2), realism (2), reform (2), regulation (2), religion (2), renewable energy (2), representation (2), research methods (2), research trends (2), research trends in the service sector (2), responsibility (2), risk (2), risk assessment (2), risk factor (2), roadmaps (2), scenario (2), science and technology (2), science and technology policy (2), science mapping (2), scientific research (2), search engines (2), sense of place (2), sensemaking (2), simulation (2), smes (2), sms (2), social (2), social housing (2), social network (2), spatial dimension (2), startup (2), strategic management (2), strategic planning (2), students (2), supply and demand matching (2), survey (2), sweden (2), systematic literature review (2), talent cultivation (2), teaching and learning (2), teaching of economics (2), technology intelligence (2), technology roadmapping (2), tension (2), textbook writer (2), textbooks (2), themes (2), theory (2), translation (2), twitter (2), uncertainty (2), universities (2), university theses (2), unrealistic assumptions (2), urbanism (2), user innovation (2), value (2), value creation (2), virtual worlds (2), voice of customers (2), voice of employees (2), volatility (2), website (2), willingness to pay (2), workshops (2), www (2)

The table provides an overview of keywords not selected for the sample identification. These are keywords with more than one occurrence in the data set. The keywords are sorted in descending order according to their occurrence in the data set. The occurrence is provided in brackets.
Source: Own overview based on the data extraction of Query B.2 and Query B.3.

B.2 Literature Review

In the following, detailed information on the articles in the sample are provided that lead to the analyses results expressed in Section 3.3. First, in Table B.3, the reasons are shown in relation to the articles of the sample in which these occurred. Table B.4 and Table B.5 then provide an overview of the textual data sources used in the articles. The tables further provide information on the data sources, the amount of data and the reasoning of to use these data sources. The applied methodologies are then disclosed in Table B.6, with a textual description of the methodology and a categorization of the methodologies according to selected areas of textual data analyses. Before providing detailed information on the content of the different articles in Table B.7 and Table B.8, Figure 3.2 and Figure B.2 provide first information on the keywords and main words in the sample and articles.

Table B.3: Overview of Reasons for Textual Analysis Application Stated in the Sample

Category	Reason	Absolute Summary	Relative Summary	Antons and Breidbach (2018)	Bakhtin et al. (2020)	Basole et al. (2019)	Chiarello et al. (2021)	Cripps et al. (2020)	Curci and Mongeau Ospina (2016)	Dahlke et al. (2021)	Feng et al. (2020)	Fiordelisi et al. (2019)	Kayser (2017)	J. Kim and C. Lee (2017)	N. Kim et al. (2015)	Kohler et al. (2014)	Larsen and Thorsrud (2019)	Mi et al. (2021)	Ozcan et al. (2021)	Shen et al. (2020)	Song et al. (2017)	B. Wang and Z. Wang (2018)	Wu et al. (2020)	Y. Zhang et al. (2019)	Zhou et al. (2020)	Zhu and Porter (2002)
Data	Large Data Set	9	39%	1	1								1	1	1	1				1		1				1
	Objectivity	7	30%		1	1											1	1		1	1				1	
	Enabling Unstructured/ Textual Data Use	6	26%						1					1	1	1							1			1
	Multiple Data Sources	4	17%	1							1									1						1
	Cost-efficient Public Data	1	4%																1							
	Full Information Use	1	4%															1								
Methodology	Efficient and Rapid Analyses	7	30%		1		1			1			1	1											1	1
	Direct Analysability	4	17%					1				1				1		1								
	Reproducibility	3	13%		1	1	1																			
	Repeatability	2	9%		1											1										
Objectives	Uncovering Latent/ Hidden Information	8	35%						1	1	1	1								1				1	1	1
	Time-based/ Trend Analyses	8	35%		1				1	1			1	1	1	1								1		
	Opportunity Detection	6	26%	1														1	1		1	1			1	
	Identification of Emerging Topics	6	26%		1		1				1						1	1						1		
	Assessment of Relationships	4	17%						1		1		1											1		
	Classification	3	13%													1			1				1			
	Gain Detailed Insights	4	17%			1					1			1									1			
	Identification of Relevant Information/ Noise Reduction	3	13%			1							1		1											
	Result Verification	1	4%					1																		
Visualisation	Visualisation	5	22%						1				1					1	1							1

The table provides an overview of reasons that authors of the sample state in their articles for the application of text analysis on textual data. The reasons are categorized into the data, methodology, objectives or visualization.

Table B.4: Textual Data Use of the Articles in the Sample

Reference	Database/ Data source	Text Data	Data Source Meaning	Patents	Publications	News Articles	Social Media	Firm Documents	Others	More than One Data Source
Antons and Breidbach (2018)	Web of Science	641 business and economic journal articles in service innovation and service design	Content of journal articles for advancements in service innovation and service design research		1					
Bakhtin et al. (2020)	Open Source Data: e.g. CrossRef database, USPTO, National Science Foundation, global news portals, Analytical Organisations	30 million documents: scientific articles, USPTO patents, news, granted awards, analytical reports	Content of different data sources for emerging patterns in agriculture and food production	1	1	1			1	1
Basole et al. (2019)	Crunchbase	24,068 operating venture descriptions	Content of venture description for the identification of entrepreneurial ecosystems					1		
Chiarello et al. (2021)	Scopus	6,601 journal articles	Content of journal articles for emerging market topics		1					
Cripps et al. (2020)	Twitter	15,054 multilingual tweets from 38 companies	Twitter as potential knowledge source for innovation, sharing and crowdsourcing				1			
Curci and Mongeau Ospina (2016)	Green Inventory (GI) database, BioPat	245 Patents (Title, Abstract, Claim, Description)	Patents as technological state-of-the-art	1						
Dahlke et al. (2021)	COVID Innovations web-platform	707 web-scraped innovation project descriptions	Innovation assessment based on innovation project descriptions, human needs from the fundamental humans needs by Max-Neef et al (1989)						1	
Feng et al. (2020)	Derwent Innovation Index	41,994 patents	Patents as technological state-of-the-art	1						
Fiordelisi et al. (2019)	Compustat, CRSP, SEC Edgar Database, Kogan et al (2017)	S&P500 firms: 10-Ks for corporate culture index	10-Ks for the corporate culture, patents included as innovation indicator					1		
Kayser (2017)	Web of Science, LexisNexis	Publication abstracts, News Articles	Journal articles for the innovation development, news articles to represent diffusion		1	1				1
J. Kim and C. Lee (2017)	Futuristic data websites (e.g. Siemens, MIT Technology Review, World Future Society), USPTO	USPTO patent data, futuristic data	Patents as technological state-of-the-art, futuristic data to identify future relevant topics	1					1	1
N. Kim et al. (2015)	Web-crawled newspaper articles	1,878,019 daily newspaper articles in the US from 1989 to 2012	Newspaper articles for industry convergence			1				

The table discloses the data sources used by the authors in the sample. The information is broken down into the database used, the textual data used and the rationale for using this textual data. If explicitly stated, the amount of data on which the research is based is also disclosed. Lastly, it is listed whether more than one data sources are used.

Table B.5: Textual Data Use of the Articles in the Sample (continued)

Reference	Database/ Data source	Text Data	Data Source Meaning	Patents	Publications	News Articles	Social Media	Firm Documents	Others	More than One Data Source
Kohler et al. (2014)	Web-crawled firm websites or SEC	455 annual reports from 132 firm websites or 10-K filings	Annual reports to reveal service innovation capabilities					1		
Larsen and Thorsrud (2019)	Daegens Naeringsliv newspaper articles from the Atekst Database	459,745 Norwegian newspaper articles	newspaper articles as a description of the economy, where highly mentioned topics address the economic future and needs			1				
Mi et al. (2021)	Survey data, report, Incopat database	Survey data, Emerging Technologies and Support an Ageing Population Report,	Survey data of elderly people for demand, Report on emerging technologies as future needs, Patent data for the technological supply	1					1	1
Ozcan et al. (2021)	Twitter	22,891 tweets from Twitter	Social media for consumer and product information				1			
Shen et al. (2020)	Web of science Science Citation Index publications, USPTO patents	930 patents and 4,543 scientific papers	Publications as basic research and scientific frontier, patents as technological innovation	1	1					1
Song et al. (2017)	JPO patents	patents	patents provide technological solutions	1						
B. Wang and Z. Wang (2018)	IncoPat patent database	57,158 patents from 1985 to 2017	Patents as scientific and technological information	1						
Wu et al. (2020)	USPTO patent database	in total 348 ICTC and non-ICTC patents	patents for technical and scientific information	1						
Y. Zhang et al. (2019)	Web of Science (WoS)	5,840 articles	Scientific articles as representatives of the research frontier		1					
Zhou et al. (2020)	Web of Science, Altmetrics	22,901 web of science documents from 2013 to 2018 in gold nanoparticles, altmetrics for 6,102 records	Altmetrics to measure impact of research based on social media data		1		1			1
Zhu and Porter (2002)	INSPEC	3552 nanotechnology abstracts of journal articles and conference paper	Publications for innovation and technology domain information		1					
			No. Articles	9	8	4	3	3	4	6

The table discloses the data sources used by the authors in the sample. The information is broken down into the database used, the textual data used and the rationale for using this textual data. If explicitly stated, the amount of data on which the research is based is also disclosed. Lastly, it is listed whether more than one data sources are used. The information is finally aggregated for all articles in the sample.

Table B.6: Methodologies Used for the Analysis of Textual Data in the Sample

Reference	Methodological Text-mining Approach	Clustering	Classification	Topic Modelling	LDA	Word Embedding	Cosine Similarity	Co-Occurrence Analyses
Antons and Breidbach (2018)	LDA Topic Modelling (69 topics), Topic network graphs	1		1	1			
Bakhtin et al. (2020)	Word2Vec, Semantic map					1		
Basole et al. (2019)	TF-IDF, Cosine similarity, visualization map with clusters						1	
Chiarello et al. (2021)	Sentiment Analysis, LDA Topic Modelling (9 topics), Development over time	1		1	1			
Cripps et al. (2020)	Structural Topic Modelling (20 topics)	2		1				
Curci and Mongeau Ospina (2016)	Keyword topic co-occurrence network, Cosine distance for similarity, Krackhardt efficiency measure for threshold value, LDA Topic Modelling (6 topics)	1		1	1		1	1
Dahlke et al. (2021)	LDA Topic Modelling (16 topics), Herfindahl-Hirschman index (HHI) for need concentration measure, Hierarchical Clustering of Domains based on thematic dissimilarity	1		1	1			
Feng et al. (2020)	Network analyses on IPC classes, Resource Allocation index for link prediction, LTA topic modelling (10 topics per converging relationship)	1		1	1			
Fiordelisi et al. (2019)	Creativity index based on word frequency, adjusted word frequency, Create Bag of Words; Regression model							
Kayser (2017)	Term network for each dataset, Pie bubble chart of matched terms based on summed relative frequency per dataset of the term							
N. Kim et al. (2015)	Industry classification based on firm SIC code, Industry convergence based on firm co-occurrence, Industry Co-occurrence (IC) Index for industry convergence measure, industry convergence map							1
J. Kim and C. Lee (2017)	keyword-document-matrix, novelty detection technique for rarity and paradigm unrelatedness assessment based on local outlier factor (LOF), hierarchical clustering	1						
Kohler et al. (2014)	Vocabulary list based on expert interviews for service innovation capabilities, classification tree to discriminate innovative and less innovative firms		1					
Larsen and Thorsrud (2019)	LDA topic modelling, network graphs, news index based on yearly topic occurrence and topic predictive power on economic output	2		1	1			
Mi et al. (2021)	word segmentation, Word2Vec, TF-IDF, semantic correlation, S&D matching model based on intensity					1		1
Ozcan et al. (2021)	TF-IDF, BERT (word embedding), Synthetic minority oversampling technique (SMOTE), Transducive SVM (TSVM) and logistic regression for idea classification and clustering		1			1		
Shen et al. (2020)	Vector space model (VSM), TFIDF, ORCLUS for clustering, cosine similarity between patent and scientific clusters,						1	
Song et al. (2017)	F-term use and solution transfer within and across industries, cosine similarity, technology growth and technology applicability						1	
B. Wang and Z. Wang (2018)	Latent Dirichlet Allocation (LDA), Provincial Topic Model (inclusion of institution information in LDA), Vector Space Model for provincial similarity calculation	1		1	1			
Wu et al. (2020)	Part-of-speech (POS), TF-IDF weighting, Deep learning MLP model with neuronal networks		1					
Y. Zhang et al. (2019)	Term correlation map, citation map, country collaboration map, K-means algorithm, Scientific evolutionary pathways (SEP)	1						1
Zhou et al. (2020)	Principal Component Analyses (PCA) for unsupervised feature extraction, Argumentative Zooming (AZ) for topic recognition, term clumping; Sentiment analyses for attitudes, SAO analysis for meaning and relationship extraction/ technology paths	1						
Zhu and Porter (2002)	Technology opportunity analysis (TOA), co-occurrence of terms, PCA for main relationship extraction, visualization in maps							1
	No. Articles	11	3	8	7	3	4	5

The table discloses the methodologies applied in relation to textual data in the different articles of the sample. The methodologies are further categorized and the common methodologies are displayed.

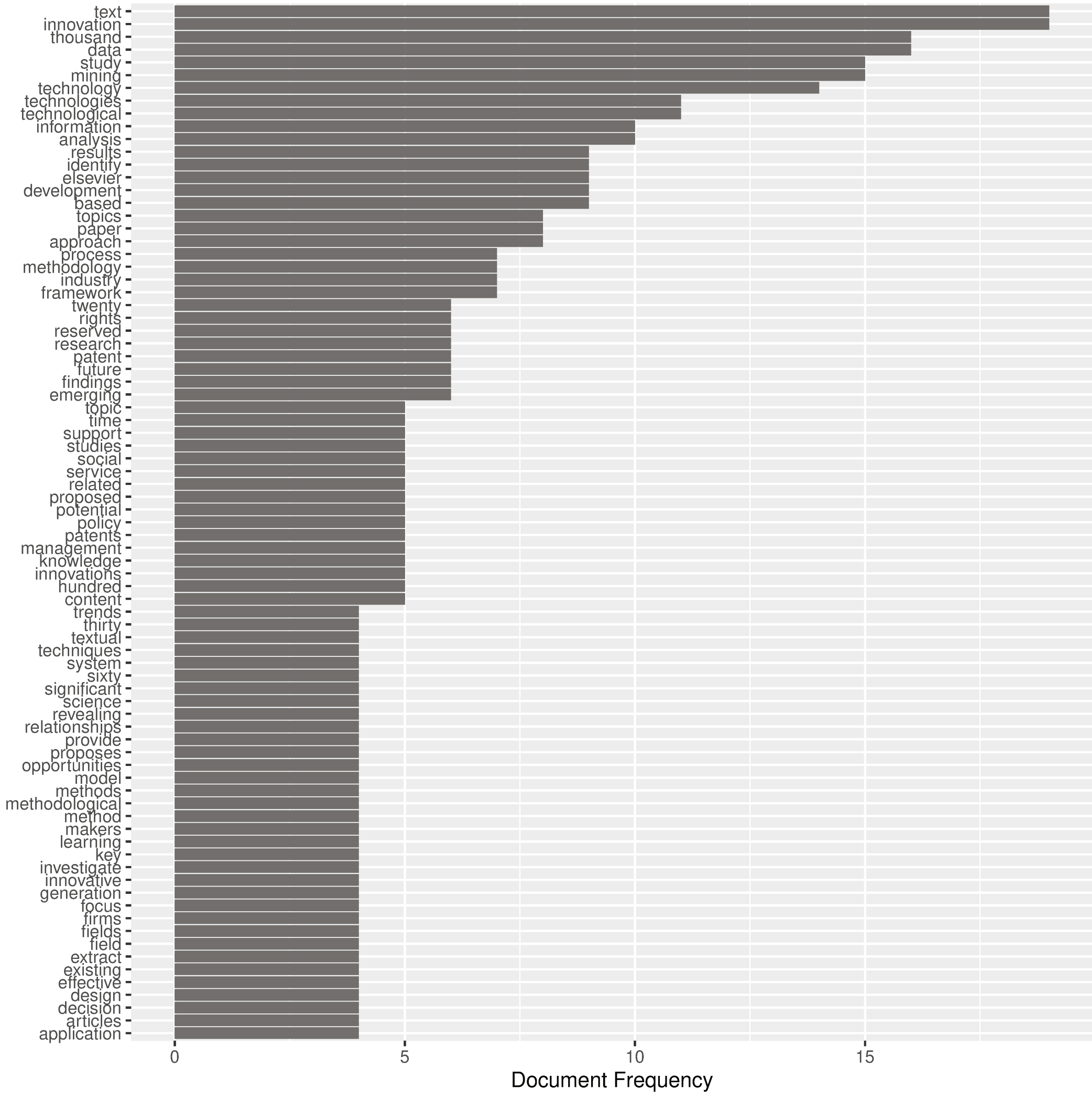


Figure B.1: Highest Document Frequency Words in the Sample.
The figure displays the most frequent words in the corpus, which is based on the abstracts of the documents in the sample.

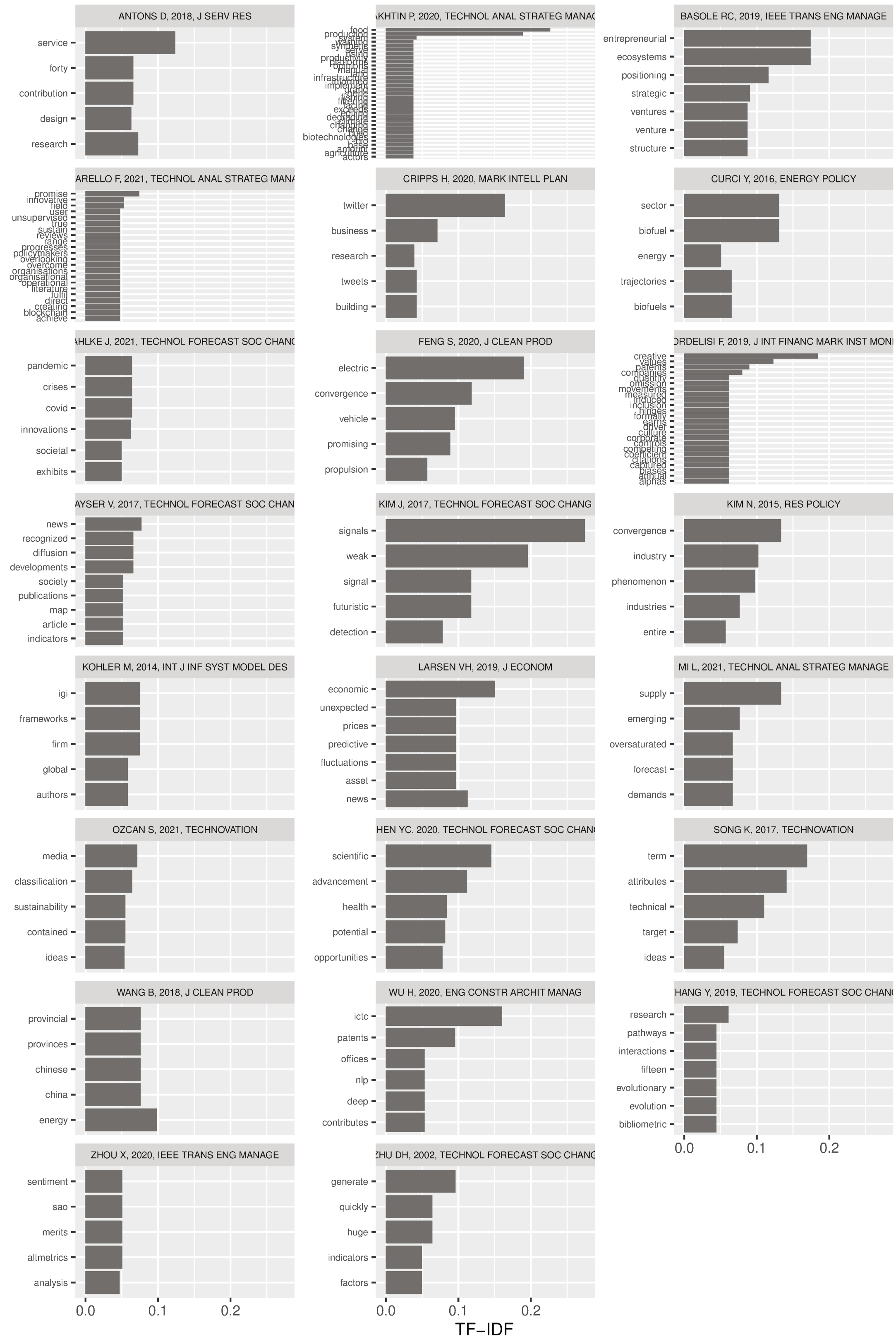


Figure B.2: Most Expressive Words for Each Abstract in the Sample.
The words are identified based on the TF-IDF calculation. TF (**T**erm **F**requency) measures the number of terms in the document and is multiplied with IDF (**I**nverted **D**ocument **F**requency), that determines the uniqueness of the word in relation to full corpus based on all documents.

Table B.7: Detailed Literature Overview of the Main Content of the Articles in the Sample

Reference	Main Focus	Purpose	Contribution	Category
Antons and Breidbach (2018)	Service Innovation, Service Design Research	Identification of knowledge gaps in service innovation and service design research, illustration of the current state	Integrate service design into service innovation, provides guidance for additional research opportunities	Technology Discovery and Exploration
Bakhtin et al. (2020)	Agriculture and Food production, Sectoral Innovation	Identifying emerging agriculture and food production topics	Provide insights in the agriculture and food sector for strategic planning activities of governments and firms	Technology Discovery and Exploration
Basole et al. (2019)	Entrepreneurial Ecosystem	Provide a granular entrepreneurial ecosystem view	Entrepreneurial clusters not only industry driven but related to similar topics; tool for competitive market intelligence analyses	Industry Patterns and Regions
Chiarello et al. (2021)	Blockchain, Emerging Markets	Extract technological problems from scientific literature to identify and map challenges of emerging markets	Reveal valuable technological information in scientific literature, Provide early mapping of technological problems in emerging markets, indicate potential of value creation, method of innovation extraction	Technology Discovery and Exploration
Cripps et al. (2020)	Knowledge Exchange, Innovation in B2B	Assess of Twitter for B2B marketing, Twitter for crowdsourcing, knowledge sharing and innovation, Relation of interviewee response to Twitter use via text-mining	Crowdsourcing, knowledge sharing and innovation creation in B2B markets occur on Twitter.	Identification of Innovation
Curci and Mongeau Ospina (2016)	Biofuel Sector, Technological Trajectory	Assess the relevance of different production paradigms and intra-sectoral technology closeness	Proliferation of biofuels patents due to sustainability concerns, adaption of the alternative energy source to the transport/ industrial sector	Technology Discovery and Exploration
Dahlke et al. (2021)	Crisis-driven Innovation	Integration of human needs and innovation	Clustering of emerging domains of innovation in crises, classification among human needs, identify technological, frugal and social innovations, Matching of human needs and innovations a complex crisis phenomenon	Demand Assessment
Feng et al. (2020)	Technology convergence, Electric vehicles	Explore technology convergence and topics in electric vehicles	Detect promising converging links in electrical vehicles, framework for trend analyses	Emerging or Novel Topics
Fiordelisi et al. (2019)	Company creativity culture, Innovation output	Impact of creativity on company output and firm value	Creative-oriented corporate culture fosters innovation	Prediction Measure
Kayser (2017)	Innovation Diffusion	integrate media in the innovation discourse for innovation diffusion analyses	Integration of innovation diffusion based on media, quantitative examination of news articles	
J. Kim and C. Lee (2017)	Weak Signals, Novelty Detection	Extraction of weak signals out of futuristic and patent data	Identification of weak signals based on public and expert knowledge, novelty consideration with futuristic data based on rarity and paradigm unrelatedness	Emerging or Novel Topics
N. Kim et al. (2015)	Industry Convergence	Analyse within and in between industry convergence depending on the industry scope	Industry convergence increase, higher intensity of within industry convergence than between industry convergence, dynamic process of convergence where some industries converge and others not	Industry Patterns and Regions

Table B.8: Detailed Literature Overview of the Main Content of the Articles in the Sample (continued)

Reference	Main Focus	Purpose	Contribution	Category
Kohler et al. (2014)	Service Innovation	Representation of service innovation capability and assessment based on text mining	Firm capability validation framework, assessment of innovation capability level	Identification of Innovation
Larsen and Thorsrud (2019)	Business Cycle	Distinguish between news and noise in news articles and assess types of news leading to a news shock	News topics can predict economic outcome. Economic boom identified after unexpected innovation news	Prediction Measure
Mi et al. (2021)	Emerging Technologies, Supply and Demand Matching, Gerontechnology	Forecasting gerontechnologies and future trends considering demand and supply	Identify emerging technologies with high demand intensity and ineffective supply	Demand Assessment
Ozcan et al. (2021)	Sustainable Innovation	Trend exploration and idea identification from social media data, identification of sustainable innovation	Identification of ideas based on twitter data classification algorithm, identification of idea clustering	Idea Generation and Extraction
Shen et al. (2020)	Smart Health Monitoring Technologies	Discovering opportunities of scientific advancement and technological innovation in smart health monitoring technologies	Scientific and technological fields are connected to each other, Retaining information with text mining, comparison of science and technology	Idea Generation and Extraction
Song et al. (2017)	Idea Generation	Idea generation and opportunity analyses on firm level based on F-term patent mining	Framework to generate new rapid ideas for technical solutions based on solutions within or across industries that are similar to the initial knowledge base. At best, these provide growth opportunities and are applicable.	Idea Generation and Extraction
B. Wang and Z. Wang (2018)	Energy Technology, Regional Development	Understand the development and direction of innovation in energy technology in China and different regions thereof from a content perspective	Identification of energy technology trends: no energy technology flow from high technology to high emission areas, low mobility of energy technologies; cooperation and technology transfer between provinces	Industry Patterns and Regions
Wu et al. (2020)	Information and Communication Technology in Construction (ICTC)	Identification of ICTC patents via patent deep learning classification	Automatic binary classification model for the identification of ICTC patents that enable the assessment of ICTC data	Identification of Innovation
Y. Zhang et al. (2019)	Big Data Research, Science Community	Discovery of the evolution of big data research and forecasting for recommendations	Machine learning and cloud computing as hot topics in big data research, large application in bioinformatics, developments in artificial intelligence reason for new directions in big data analytics	Innovation Trajectory and Diffusion
Zhou et al. (2020)	Innovation Trajectory, Gold Nanoparticles	Predict innovation trajectory, identify hot and valuable topic, include sentiment analyses combined with quantitative analyses	Framework to identify pathway of technological innovation and future potential for commercialization based on a combined text mining, machine learning, sentiment analyses and altmetrics approach	Innovation Trajectory and Diffusion
Zhu and Porter (2002)	Automatic Technology Forecasting, Nanotechnology	Improve technology forecasting and respond to challenges such as large amount of textual data or generation of rapid findings	Automated process framework for knowledge generation from bibliographic text mining via maps of e.g. keywords, affiliations, authors	Technology Discovery and Exploration

C Appendix to Chapter 4: Innovative Activities Captured through Semantic Analysis of Patents and Trademarks

As in Chapter 4, the appendix first provides further information on the analyses of Robotics and its data followed by information on Footwear.

C.1 Robotics

Subsection C.1.1 discloses the queries used to identify the Robotics data set. Additionally, background information on the considered patent CPC subgroups are provided. Subsection C.1.2 then provides further detail of the model selection and several topic examples.

C.1.1 Data Selection

Patent Selection

```
select distinct cp.patent_id
into fs.robot_uspto_pat
from uspto.cpc_current cp
inner join fs.robot_uspto_cpc_subgroup_selected se on se.id = cp.subgroup_id

union


select distinct cp.patent_id
from uspto.cpc_current cp
where cp.group_id in ('B25J')

union

select distinct pa.id as 'patent_id'
from uspto.patent pa
where (abstract like '%robot%' or title like '%robot%')
```

Query C.1: All Patents of Robotics

```
select distinct patent_id
into fs.robot_uspto_pat_util
from fs.robot_uspto_pat mu
join uspto.patent pa on mu.patent_id = pa.id
where pa.type in ('utility')
```

Query C.2: Only Utility Patents of Robotics

Table C.1: CPC Subgroup IDs used of the Selection of Robotic Patents

CPC Subgroup ID	Description
A47L2201/00	Robotic cleaning machines, i.e. with automatic control of the travelling movement or the cleaning operation
A47L2201/02	Robotic cleaning machines, i.e. with automatic control of the travelling movement or the cleaning operation-Docking stations; Docking operations
A47L2201/022	Robotic cleaning machines, i.e. with automatic control of the travelling movement or the cleaning operation-Docking stations; Docking operations-Recharging of batteries
A47L2201/024	Robotic cleaning machines, i.e. with automatic control of the travelling movement or the cleaning operation-Docking stations; Docking operations-Emptying dust or waste liquid containers
A47L2201/026	Robotic cleaning machines, i.e. with automatic control of the travelling movement or the cleaning operation-Docking stations; Docking operations-Refilling cleaning liquid containers
A47L2201/028	Robotic cleaning machines, i.e. with automatic control of the travelling movement or the cleaning operation-Docking stations; Docking operations-Refurbishing floor engaging tools, e.g. cleaning of beating brushes
A47L2201/04	Robotic cleaning machines, i.e. with automatic control of the travelling movement or the cleaning operation-Automatic control of the travelling movement; Automatic obstacle detection
A47L2201/06	Robotic cleaning machines, i.e. with automatic control of the travelling movement or the cleaning operation-Control of the cleaning action for autonomous devices; Automatic detection of the surface condition before, during or after cleaning
A61B2034/305	Computer-aided surgery; Manipulators or robots specially adapted for use in surgery-Surgical robots-Details of wrist mechanisms at distal ends of robotic arms
A61B2034/306	Computer-aided surgery; Manipulators or robots specially adapted for use in surgery-Surgical robots-Details of wrist mechanisms at distal ends of robotic arms-Wrists with multiple vertebrae
A61B34/30	Computer-aided surgery; Manipulators or robots specially adapted for use in surgery-Surgical robots
A61B34/32	Computer-aided surgery; Manipulators or robots specially adapted for use in surgery-Surgical robots-operating autonomously
A61B34/35	Computer-aided surgery; Manipulators or robots specially adapted for use in surgery-Surgical robots-for telesurgery
A61B34/37	Computer-aided surgery; Manipulators or robots specially adapted for use in surgery-Surgical robots-Master-slave robots
A61B6/4458	Apparatus for radiation diagnosis, e.g. combined with radiation therapy equipment -Constructional features of the device for radiation diagnosis-related to the mounting of source units and detector units -the source unit or the detector unit being attached to robotic arms
B07C2501/0063	Sorting according to a characteristic or feature of the articles or material to be sorted-Using robots
B25J11/0015	Manipulators not otherwise provided for-Manipulators having means for high-level communication with users, e.g. speech generator, face recognition means-Face robots, animated artificial faces for imitating human expressions
B25J13/089	Controls for manipulators -by means of sensing devices, e.g. viewing or touching devices-with position, velocity or acceleration sensors-Determining the position of the robot with reference to its environment
B25J19/0029	Accessories fitted to manipulators, e.g. for monitoring, for viewing; Safety devices combined with or specially adapted for use in connection with manipulators -Means for supplying energy to the end effector-arranged within the different robot elements
B25J19/0033	Accessories fitted to manipulators, e.g. for monitoring, for viewing; Safety devices combined with or specially adapted for use in connection with manipulators -Means for supplying energy to the end effector-arranged within the different robot elements-with axial connectors in end effector flange
B25J19/0037	Accessories fitted to manipulators, e.g. for monitoring, for viewing; Safety devices combined with or specially adapted for use in connection with manipulators -Means for supplying energy to the end effector-arranged within the different robot elements-comprising a light beam pathway, e.g. laser
B25J19/0041	Accessories fitted to manipulators, e.g. for monitoring, for viewing; Safety devices combined with or specially adapted for use in connection with manipulators -Means for supplying energy to the end effector-arranged within the different robot elements-having rotary connection means
B25J9/0003	Programme-controlled manipulators-Home robots, i.e. small robots for domestic use
B25J9/065	Programme-controlled manipulators-characterised by multi-articulated arms-Snake robots
B25J9/1035	Programme-controlled manipulators-characterised by positioning means for manipulator elements-Gears specially adapted therefor, e.g. reduction gears -Pinion and fixed rack drivers, e.g. for rotating an upper arm support on the robot base
B29C2945/76317	Indexing scheme relating to injection moulding, i.e. forcing the required volume of moulding material through a nozzle into a closed mould-Measuring, controlling or regulating-Location of measurement -Auxiliary devices-robots, grippers
B29C2945/76795	Indexing scheme relating to injection moulding, i.e. forcing the required volume of moulding material through a nozzle into a closed mould-Measuring, controlling or regulating-Location of control -Auxiliary devices-robots, grippers
B32B2038/1891	Ancillary operations in connection with laminating processes-Handling of layers or the laminate-Using a robot for handling the layers
B64G2001/247	Cosmonautic vehicles-Parts of, or equipment specially adapted for fitting in or to, cosmonautic vehicles-Guiding or controlling apparatus, e.g. for attitude control -Advanced control concepts for autonomous, robotic spacecraft, e.g. by using artificial intelligence, neural networks or autonomous agents
B64G2004/005	Tools specially adapted for use in space-Robotic manipulator systems for use in space
B65F2230/14	Shapes of refuse receptacles-Robot
B65H2555/31	Actuating means-Multi-axis-Robots
B67D2007/0403	Apparatus or devices for transferring liquids from bulk storage containers or reservoirs into vehicles or into portable containers, e.g. for retail sale purposes -for transferring fuels, lubricants or mixed fuels and lubricants-arrangements for automatically fuelling vehicles, i.e. without human intervention-Fuelling robots
B67D2007/0405	Apparatus or devices for transferring liquids from bulk storage containers or reservoirs into vehicles or into portable containers, e.g. for retail sale purposes -for transferring fuels, lubricants or mixed fuels and lubricants-arrangements for automatically fuelling vehicles, i.e. without human intervention-Fuelling robots-Opening devices
B67D2007/0407	Apparatus or devices for transferring liquids from bulk storage containers or reservoirs into vehicles or into portable containers, e.g. for retail sale purposes -for transferring fuels, lubricants or mixed fuels and lubricants-arrangements for automatically fuelling vehicles, i.e. without human intervention-Fuelling robots-Opening devices-for fuel tank flaps
B67D2007/0409	Apparatus or devices for transferring liquids from bulk storage containers or reservoirs into vehicles or into portable containers, e.g. for retail sale purposes -for transferring fuels, lubricants or mixed fuels and lubricants-arrangements for automatically fuelling vehicles, i.e. without human intervention-Fuelling robots-Opening devices-for fuel tank flaps-using vacuum cups
B67D2007/0411	Apparatus or devices for transferring liquids from bulk storage containers or reservoirs into vehicles or into portable containers, e.g. for retail sale purposes -for transferring fuels, lubricants or mixed fuels and lubricants-arrangements for automatically fuelling vehicles, i.e. without human intervention-Fuelling robots-Opening devices-for fuel tank flaps-using grippers
B67D2007/0413	Apparatus or devices for transferring liquids from bulk storage containers or reservoirs into vehicles or into portable containers, e.g. for retail sale purposes -for transferring fuels, lubricants or mixed fuels and lubricants-arrangements for automatically fuelling vehicles, i.e. without human intervention-Fuelling robots-Opening devices-for fuel tank flaps-using air blast
B67D2007/0415	Apparatus or devices for transferring liquids from bulk storage containers or reservoirs into vehicles or into portable containers, e.g. for retail sale purposes -for transferring fuels, lubricants or mixed fuels and lubricants-arrangements for automatically fuelling vehicles, i.e. without human intervention-Fuelling robots-Opening devices-for filler caps
B67D2007/0417	Apparatus or devices for transferring liquids from bulk storage containers or reservoirs into vehicles or into portable containers, e.g. for retail sale purposes -for transferring fuels, lubricants or mixed fuels and lubricants-arrangements for automatically fuelling vehicles, i.e. without human intervention-Fuelling robots-Manipulator arms
B67D2007/0419	Apparatus or devices for transferring liquids from bulk storage containers or reservoirs into vehicles or into portable containers, e.g. for retail sale purposes -for transferring fuels, lubricants or mixed fuels and lubricants-arrangements for automatically fuelling vehicles, i.e. without human intervention-Fuelling robots-Fuelling nozzles
B67D2007/0421	Apparatus or devices for transferring liquids from bulk storage containers or reservoirs into vehicles or into portable containers, e.g. for retail sale purposes -for transferring fuels, lubricants or mixed fuels and lubricants-arrangements for automatically fuelling vehicles, i.e. without human intervention-Fuelling robots-Fuelling nozzles-with locking devices

The table provides an overview of the selected CPC Subgroups and their description.
Source: Own representation based on USPTO data.

Table C.2: CPC Subgroup IDs used of the Selection of Robotic Patents (continued)

CPC Subgroup ID	Description
B67D2007/0423	Apparatus or devices for transferring liquids from bulk storage containers or reservoirs into vehicles or into portable containers, e.g. for retail sale purposes -for transferring fuels, lubricants or mixed fuels and lubricants-arrangements for automatically fuelling vehicles, i.e. without human intervention-Fuelling robots-Fuelling hoses
B67D2007/0425	Apparatus or devices for transferring liquids from bulk storage containers or reservoirs into vehicles or into portable containers, e.g. for retail sale purposes -for transferring fuels, lubricants or mixed fuels and lubricants-arrangements for automatically fuelling vehicles, i.e. without human intervention-Fuelling robots-Fuelling hoses-comprising a single hose for several fuels
B67D2007/0426	Apparatus or devices for transferring liquids from bulk storage containers or reservoirs into vehicles or into portable containers, e.g. for retail sale purposes -for transferring fuels, lubricants or mixed fuels and lubricants-arrangements for automatically fuelling vehicles, i.e. without human intervention-Fuelling robots-Fuelling hoses-comprising several hoses for several fuels
B67D2007/0428	Apparatus or devices for transferring liquids from bulk storage containers or reservoirs into vehicles or into portable containers, e.g. for retail sale purposes -for transferring fuels, lubricants or mixed fuels and lubricants-arrangements for automatically fuelling vehicles, i.e. without human intervention-Fuelling robots-Fuelling hoses-having devices to avoid a mix up of different fuels
B67D2007/043	Apparatus or devices for transferring liquids from bulk storage containers or reservoirs into vehicles or into portable containers, e.g. for retail sale purposes -for transferring fuels, lubricants or mixed fuels and lubricants-arrangements for automatically fuelling vehicles, i.e. without human intervention-Fuelling robots-Moveable
B67D2007/0432	Apparatus or devices for transferring liquids from bulk storage containers or reservoirs into vehicles or into portable containers, e.g. for retail sale purposes -for transferring fuels, lubricants or mixed fuels and lubricants-arrangements for automatically fuelling vehicles, i.e. without human intervention-Fuelling robots-Moveable-according to a planar coordinate system
B67D2007/0434	Apparatus or devices for transferring liquids from bulk storage containers or reservoirs into vehicles or into portable containers, e.g. for retail sale purposes -for transferring fuels, lubricants or mixed fuels and lubricants-arrangements for automatically fuelling vehicles, i.e. without human intervention-Fuelling robots-Moveable-according to a planar coordinate system-with the ability to compensate movements of the car during filling
B67D2007/0436	Apparatus or devices for transferring liquids from bulk storage containers or reservoirs into vehicles or into portable containers, e.g. for retail sale purposes -for transferring fuels, lubricants or mixed fuels and lubricants-arrangements for automatically fuelling vehicles, i.e. without human intervention-Fuelling robots-Moveable-according to a spatial coordinate system
B67D2007/0438	Apparatus or devices for transferring liquids from bulk storage containers or reservoirs into vehicles or into portable containers, e.g. for retail sale purposes -for transferring fuels, lubricants or mixed fuels and lubricants-arrangements for automatically fuelling vehicles, i.e. without human intervention-Fuelling robots-Moveable-according to a spatial coordinate system-with the ability to compensate movements of the car during filling
F16H2057/0043	General details of gearing -Mounting or adjusting transmission parts by robots
F16H2061/0071	Control functions within ; control units of; change-speed- or reversing-gearings for conveying rotary motion ; ; Control of exclusively fluid gearing, friction gearing, gearings with endless flexible members or other particular types of gearing-Method or means for testing of transmission controls or parts thereof-Robots or simulators for testing control functions in automatic transmission
G01N35/0099	Automatic analysis not limited to methods or materials provided for in any single one of groups G01N1/00 - G01N33/00; Handling materials therefor-comprising robots or similar manipulators
G01S13/881	Systems using the reflection or reradiation of radio waves, e.g. radar systems; Analogous systems using reflection or reradiation of waves whose nature or wavelength is irrelevant or unspecified -Radar or analogous systems specially adapted for specific applications -for robotics
G05B2219	828 subgroups of G05B2219, covering aspects of Program-control systems

The table provides an overview of the selected CPC Subgroups and their description. Additionally, 828 subgroups belong to class G05B2219, that covers Program-control systems in various forms.
Source: Own representation based on USPTO data.

Trademark Selection

```
select *
into fs.robot_trademarks_statement
from trademarks.[statement]
where serial_no in (
select serial_no
from trademarks.case_file
where mark_id_char like '%robot%'
)
and left(statement_type_cd, 2) in ('GS')

union

select *
from trademarks.statement
where serial_no in (
select serial_no
from trademarks.[statement]
where statement_text like '%robot%'
and left(statement_type_cd, 2) in ('PM')
)
and left(statement_type_cd, 2) in ('GS')

union

select *
from trademarks.statement
where statement_text like '%robot%'
and left(statement_type_cd, 2) in ('GS')
```

Query C.3: All Trademarks of Robotics

```
select mu.statement_type_cd, mu.statement_text, mu.serial_no
into fs.robot_trademarks_statement_registered
from fs.robot_trademarks_statement mu
join trademarks.case_file cf on mu.serial_no = cf.serial_no
where cf.registration_no not in ('0000000')
```

Query C.4: Only Registered Trademarks of Robotics

```
select distinct mu.serial_no
into fs.robot_trademarks_registered
from fs.robot_trademarks_statement_registered mu
```

Query C.5: Distinct Registered Trademarks of Robotics

C.1.2 Topic Modelling Results

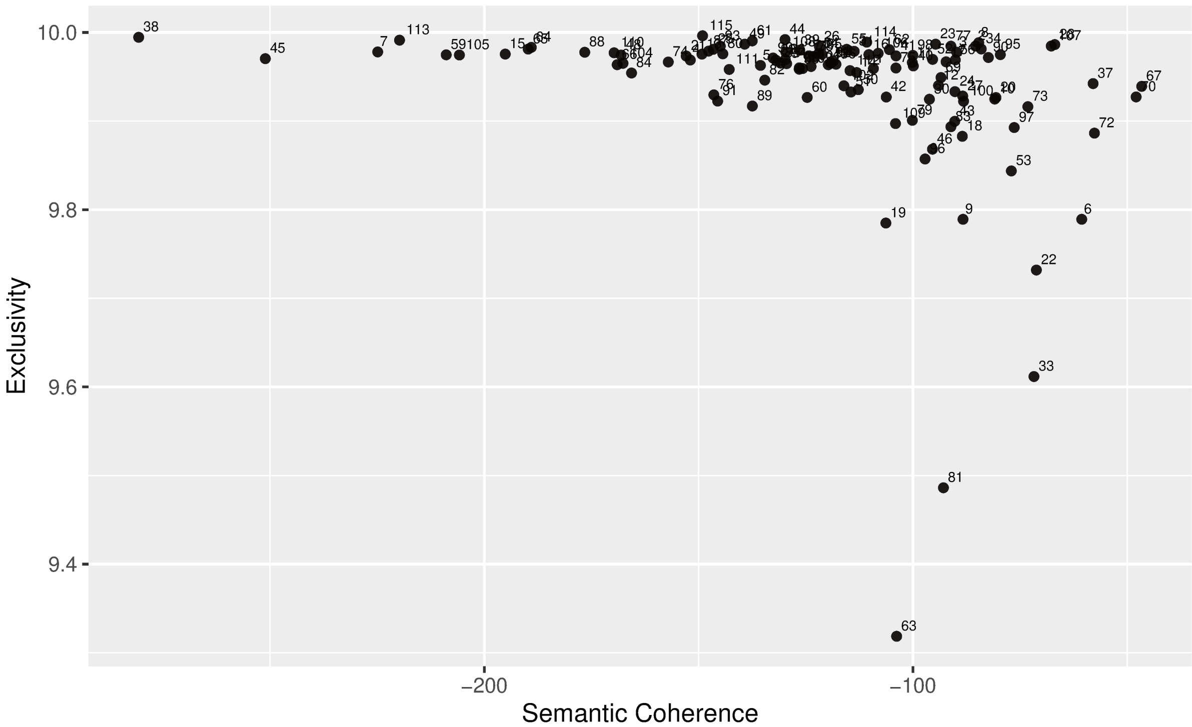


Figure C.1: Exclusivity-Semantic Coherence Trade-off of the Topics of Robotics Model 115. For the model with 115 topics, the trade-off between exclusivity and semantic coherence for each topic is displayed. This means that topic 38 for example is high on exclusive words but the words are less coherent, while topic 63 has a lot of common words with high coherence.

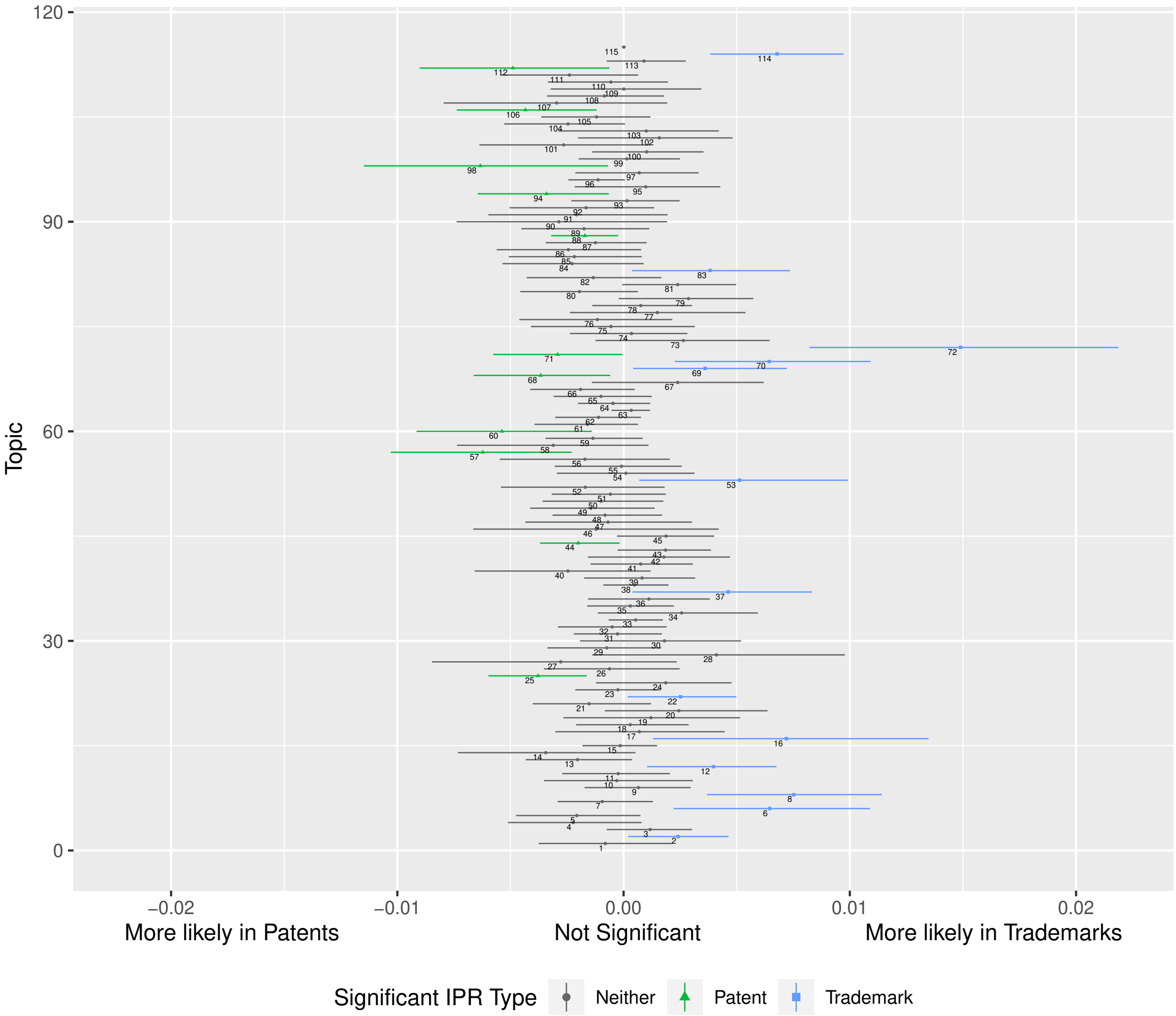
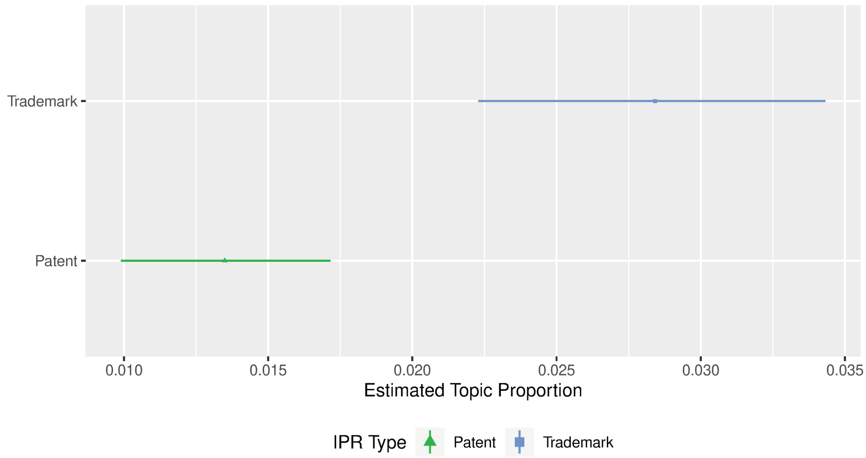


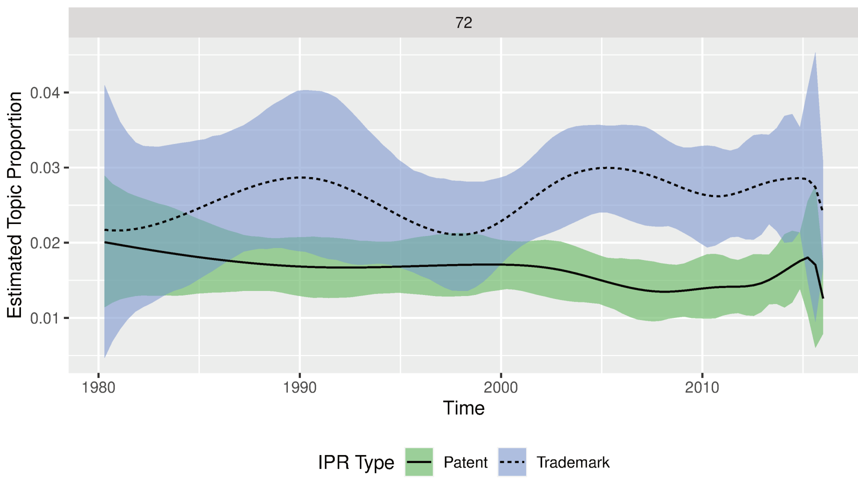
Figure C.2: Significance of Robotic Topics in Patents or Trademarks.
Each topics relates to a varying degree to patents or trademarks. The figure weights the occurrence probability of trademarks and patents against each other. If the mean and the variance in this figure are more likely in one document type, the occurrence is significance. In case of patents and trademarks, the related topics are coloured accordingly. The significance is calculated based on a 95% confidence interval.

Analysis of Robotic Topic 72

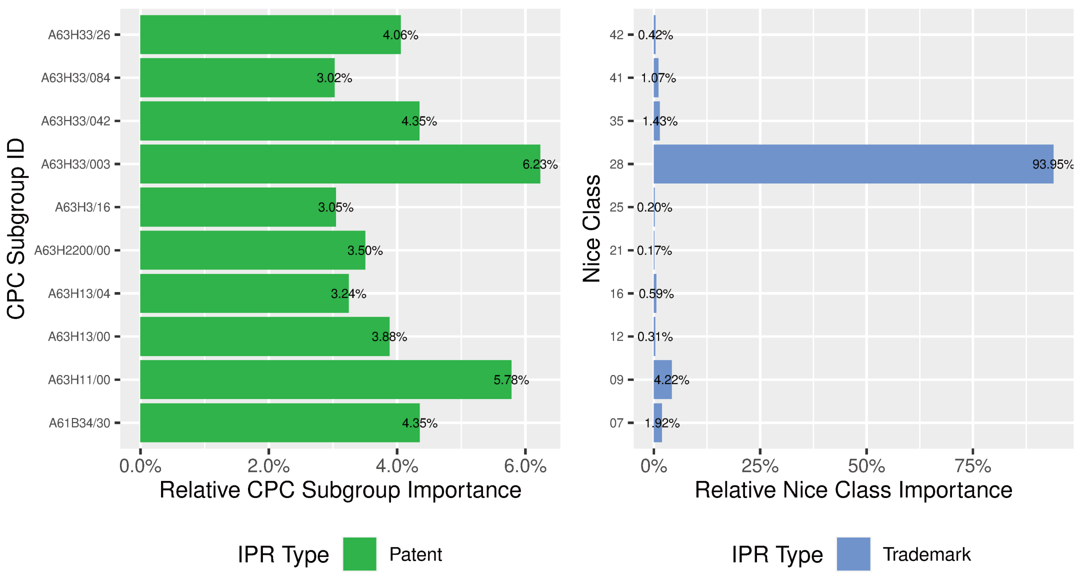
(a) Expected Topic Proportion per IPR in Topic 72



(b) Expected Topic Proportion per IPR in Topic 72 over Time



(c) Ten Main Classes of Topic 72 in Patents and Trademarks



(d) Main Documents in Topic 72

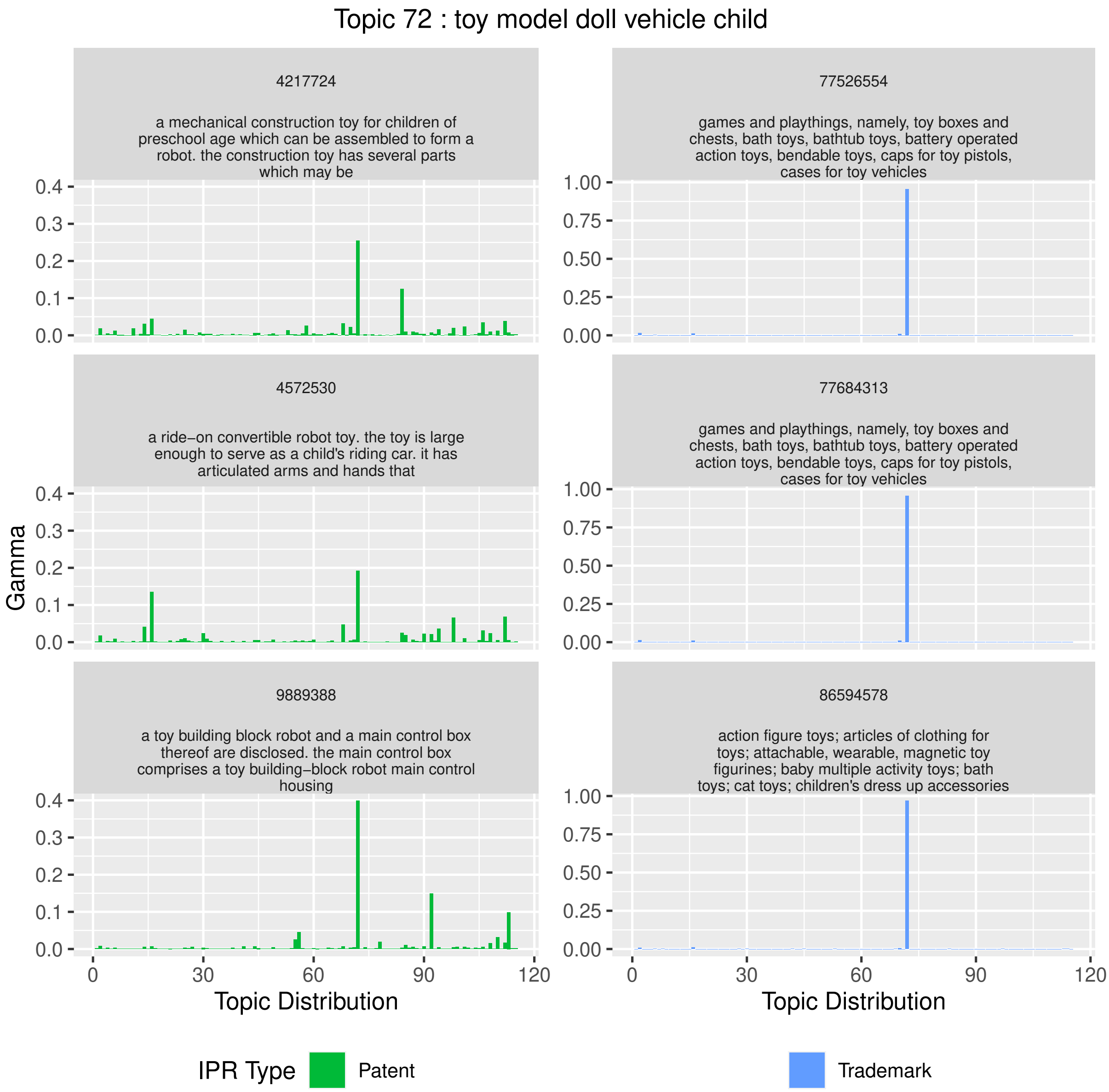
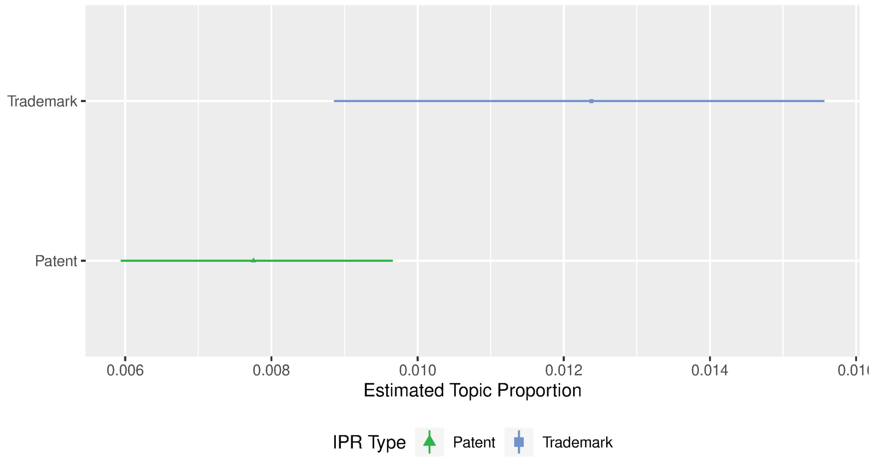


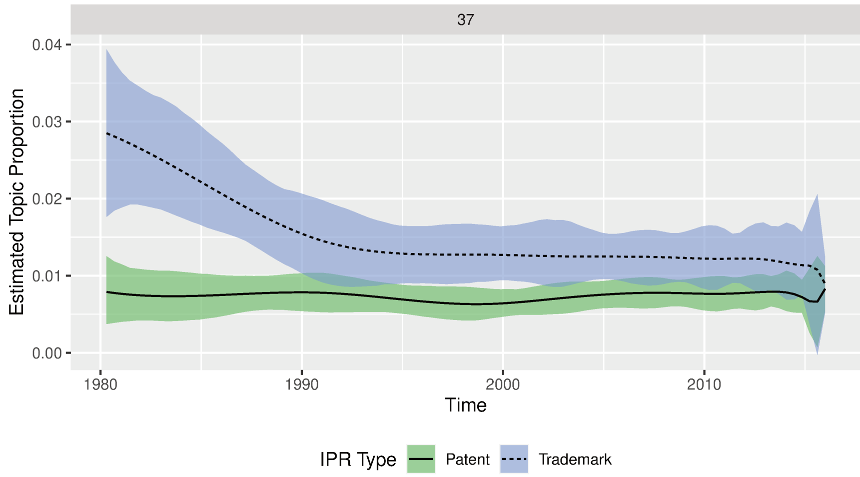
Figure C.3: Overview of Robotic Topic 72.
For topic 72, the expected topic proportion per IPR and over time are given. As can be seen, the topic is strong in trademarks and less relevant in patents, also in the time overview. The classifications related to games and toys. The documents in relation to the topics focus on toys in trademarks and there on toy products. The patent documents describe different kinds of robotic toys. The patent documents are less focused on topic 72 than the trademark documents, but also occur in other topics. Overall, the general area of invention and application matches. Patents provide more insights on the toy inventions themselves while trademarks highlight the occurrence of robotic toys and their application.

Analysis of Robotic Topic 37

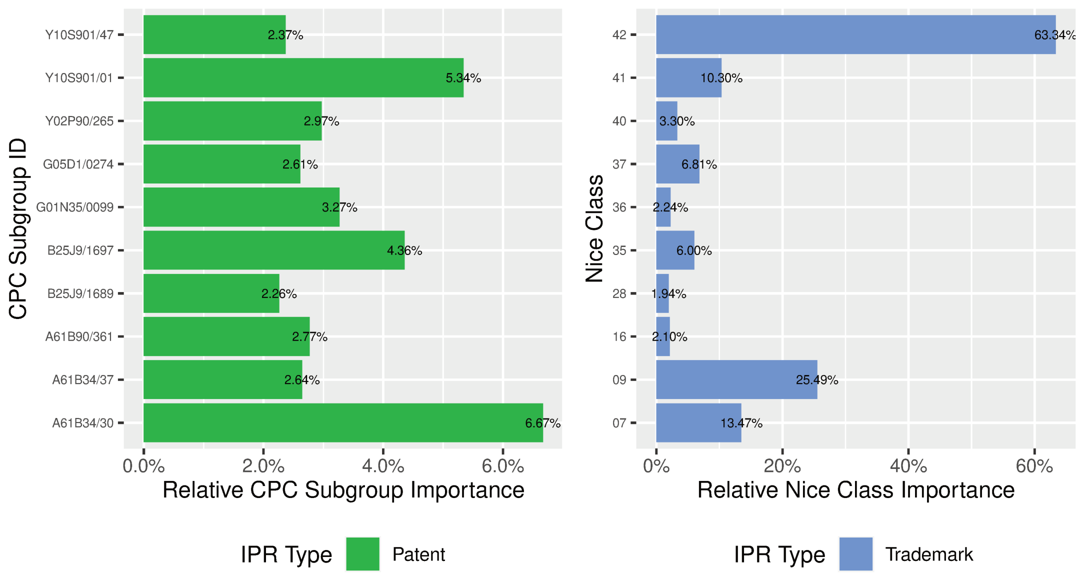
(a) Expected Topic Proportion per IPR in Topic 37



(b) Expected Topic Proportion per IPR in Topic 37 over Time



(c) Ten Main Classes of Topic 37 in Patents and Trademarks



(d) Main Documents in Topic 37

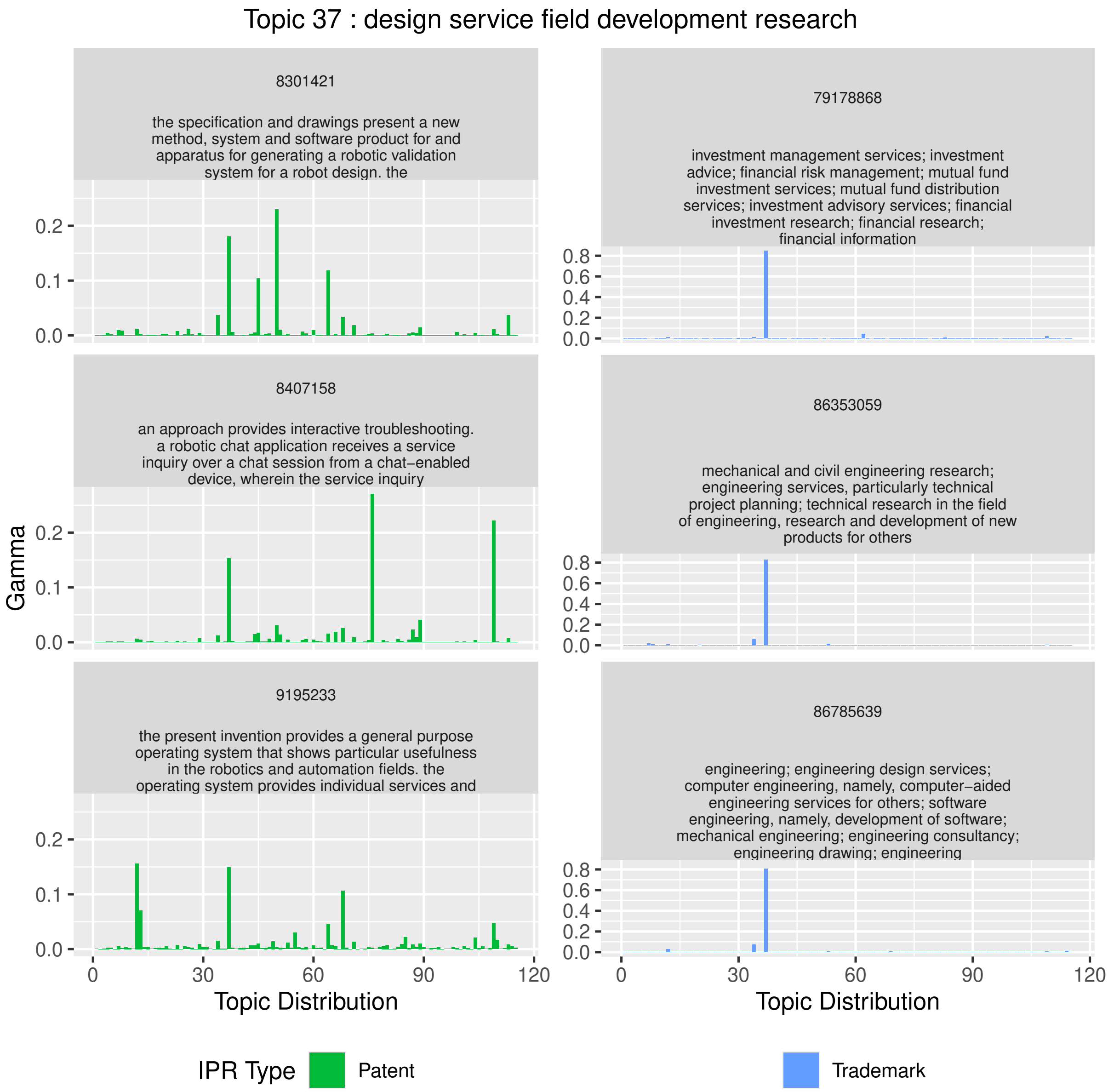
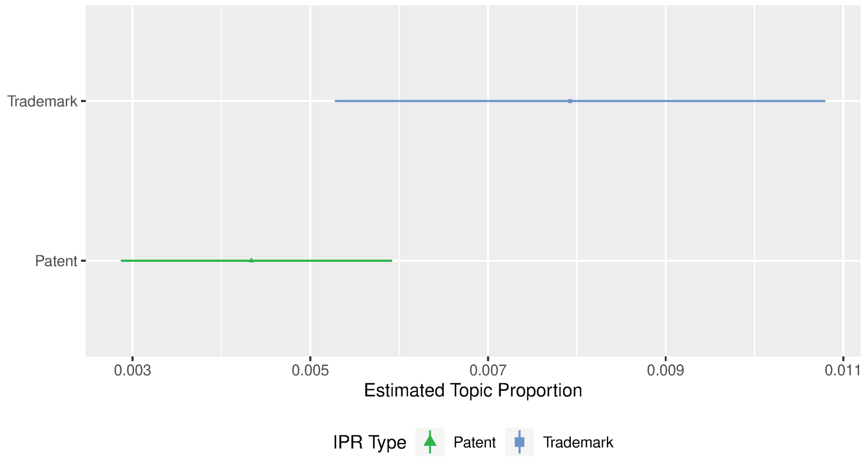


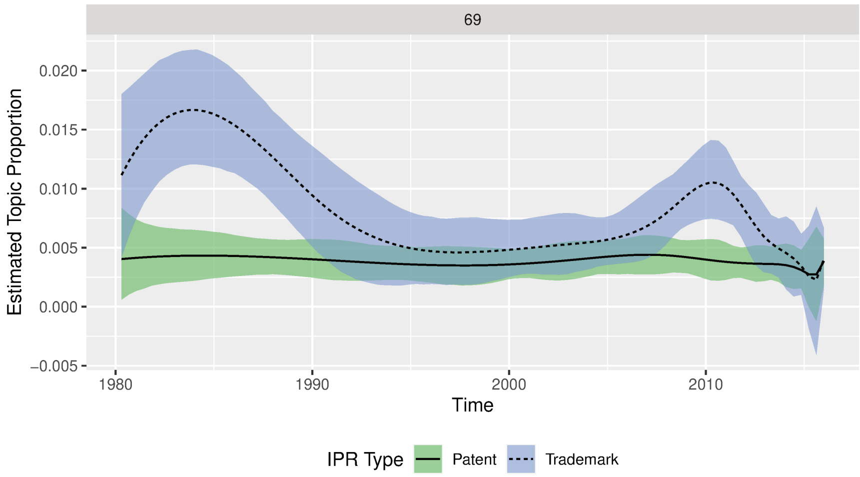
Figure C.4: Overview of Robotic Topic 37.
Trademarks are strong in topic 37 in comparison to patents. This is especially true in the beginning of 1980s. Since then, patents and trademarks started to converge, the topic becoming less important in trademarking. The trademark classes are especially related to service classes and here scientific research, while the patent classes relate to the investigation of different material. This is in line with the main words of the topic that also relate to services and research. In detail, however, it is noticeable that although the texts can generally be assigned to the field, no clear common theme can be identified within Research. Here, the trademarks are also very broadly positioned.

Analysis of Robotic Topic 69

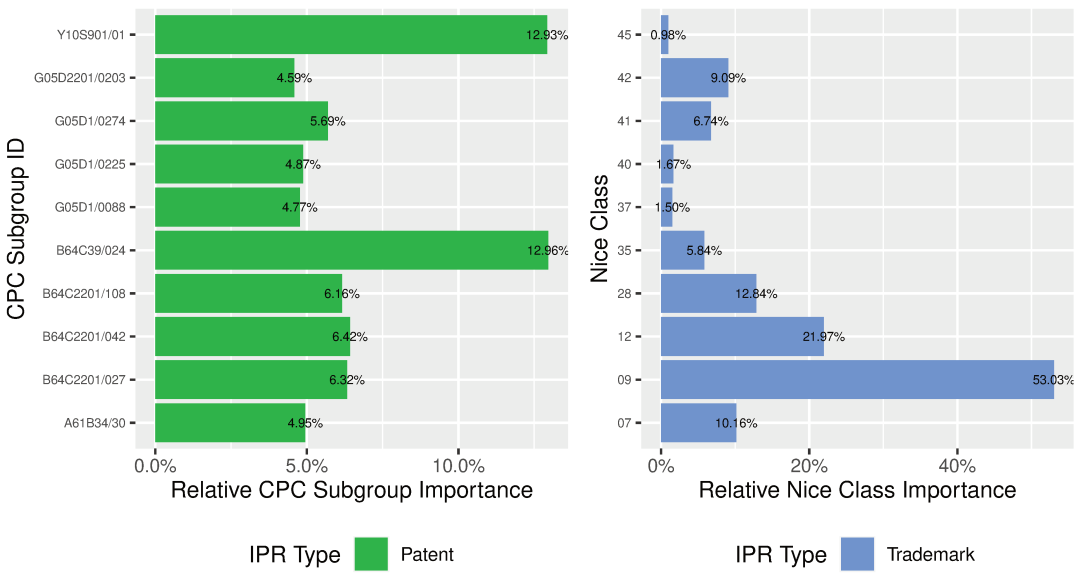
(a) Expected Topic Proportion per IPR in Topic 69



(b) Expected Topic Proportion per IPR in Topic 69 over Time



(c) Ten Main Classes of Topic 69 in Patents and Trademarks



(**d**) Main Documents in Topic 69

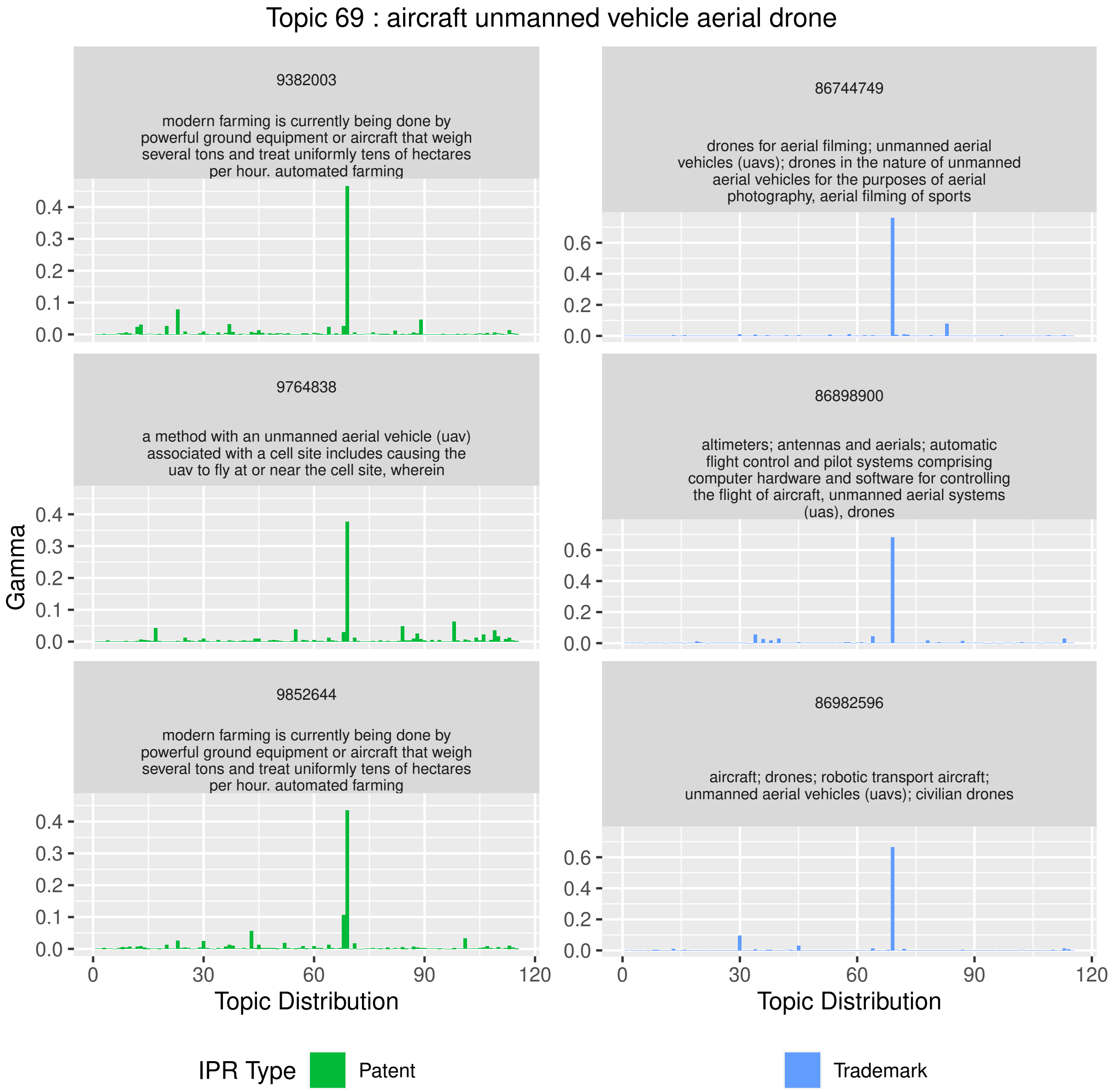
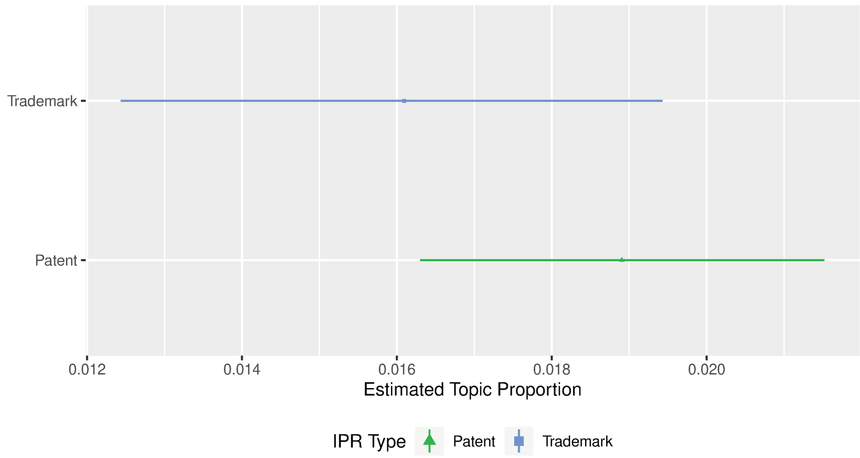


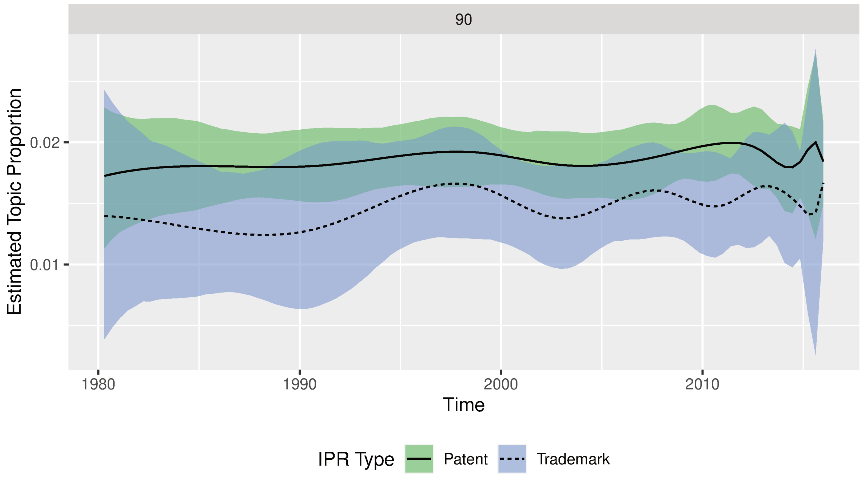
Figure C.5: Overview of Robotic Topic 69.
Trademarks are stronger in topic 69 in comparison to patents. There are interesting peaks visible in the 1985s and around 2010. The trademark classes 7 and 9 are especially related to machines and instruments for controlling and monitoring aircraft, while the patent classes relate to aerodynamic aspects (B64C), Master slave robots (A61B34), and controls (G05D). Further of interest is that the trademark as well as the patent documents are very focused on the topic of unmanned aerial vehicles. In topic 72 as well as in topic 37, only trademarks were focused while patents were present in diverse topics.

Analysis of Robotic Topic 90

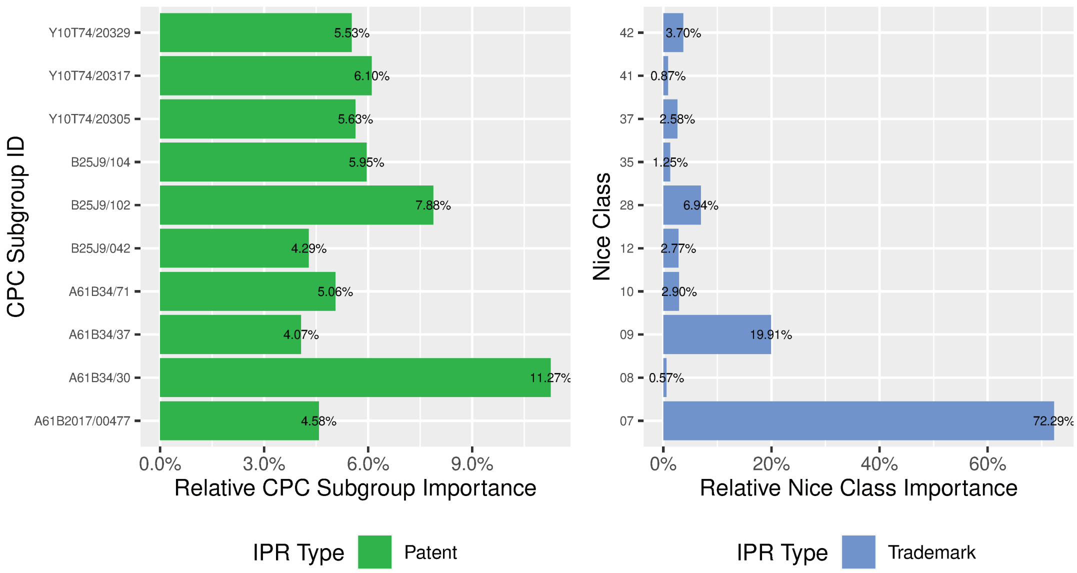
(a) Expected Topic Proportion per IPR in Topic 90



(b) Expected Topic Proportion per IPR in Topic 90 over Time



(c) Ten Main Classes of Topic 90 in Patents and Trademarks



(**d**) Main Documents in Topic 90

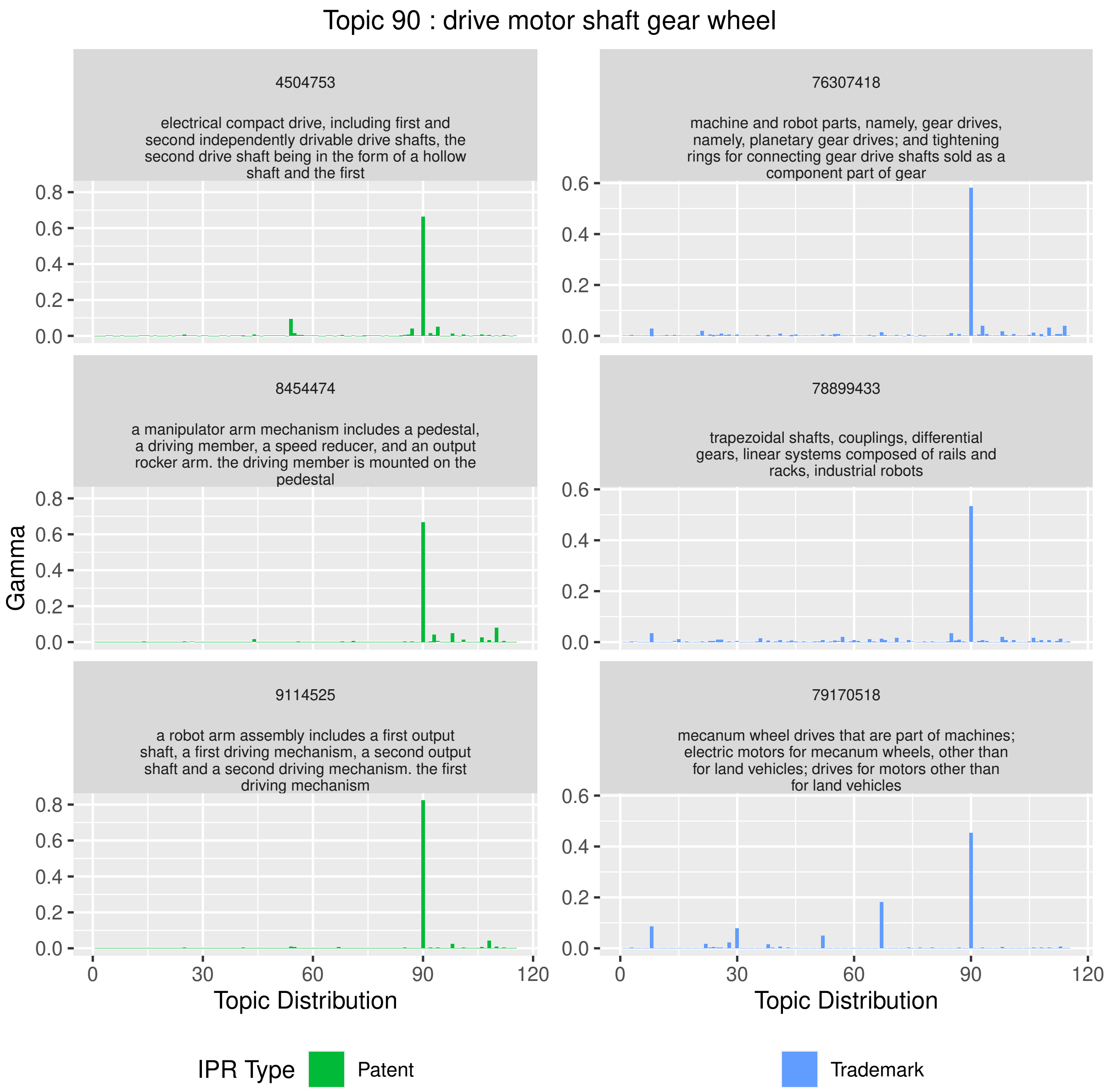
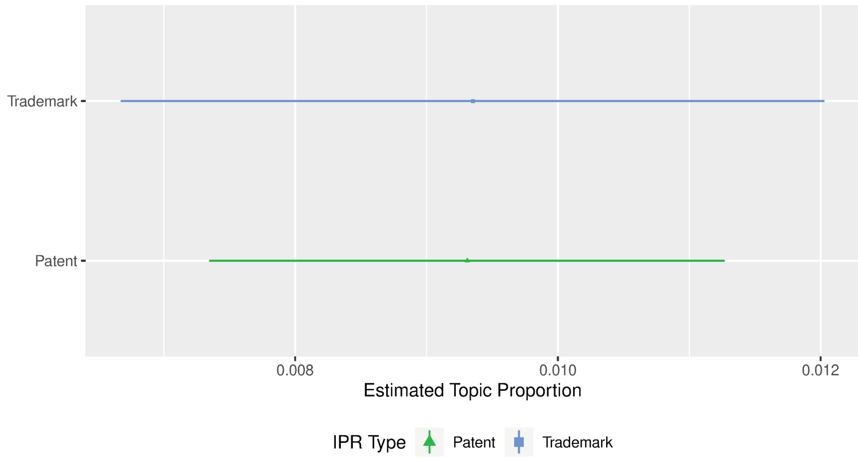


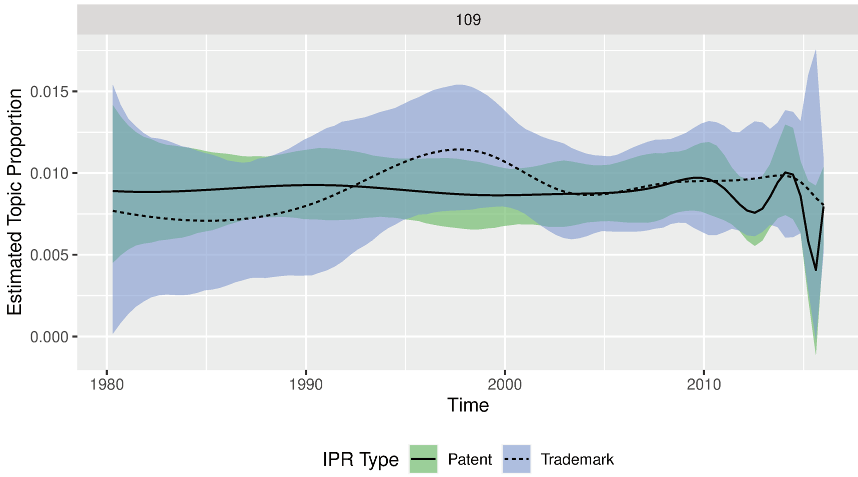
Figure C.6: Overview of Robotic Topic 90.
Topic 90 has a strong overlap of patents and trademarks in the estimated topic occurrence, also in the time overview. The classifications focus mainly on machinery and motor aspects. The documents in relation to the topics focus on machines and components in trademarks and patents. Both IPRs display a high level of detail in their descriptions and are in line with each other.

Analysis of Robotic Topic 109

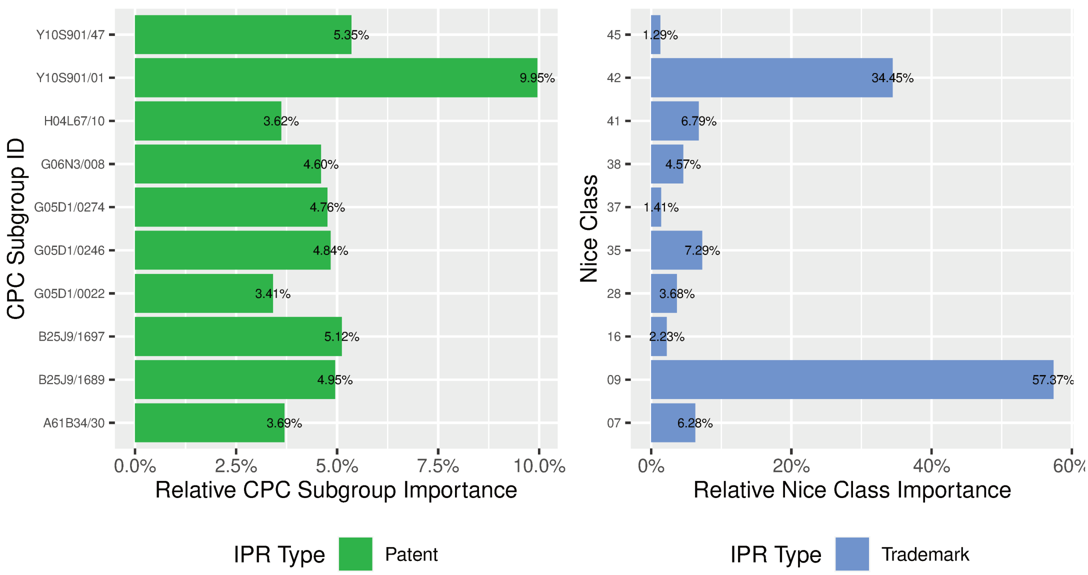
(a) Expected Topic Proportion per IPR in Topic 109



(b) Expected Topic Proportion per IPR in Topic 109 over Time



(c) Ten Main Classes of Topic 109 in Patents and Trademarks



(**d**) Main Documents in Topic 109

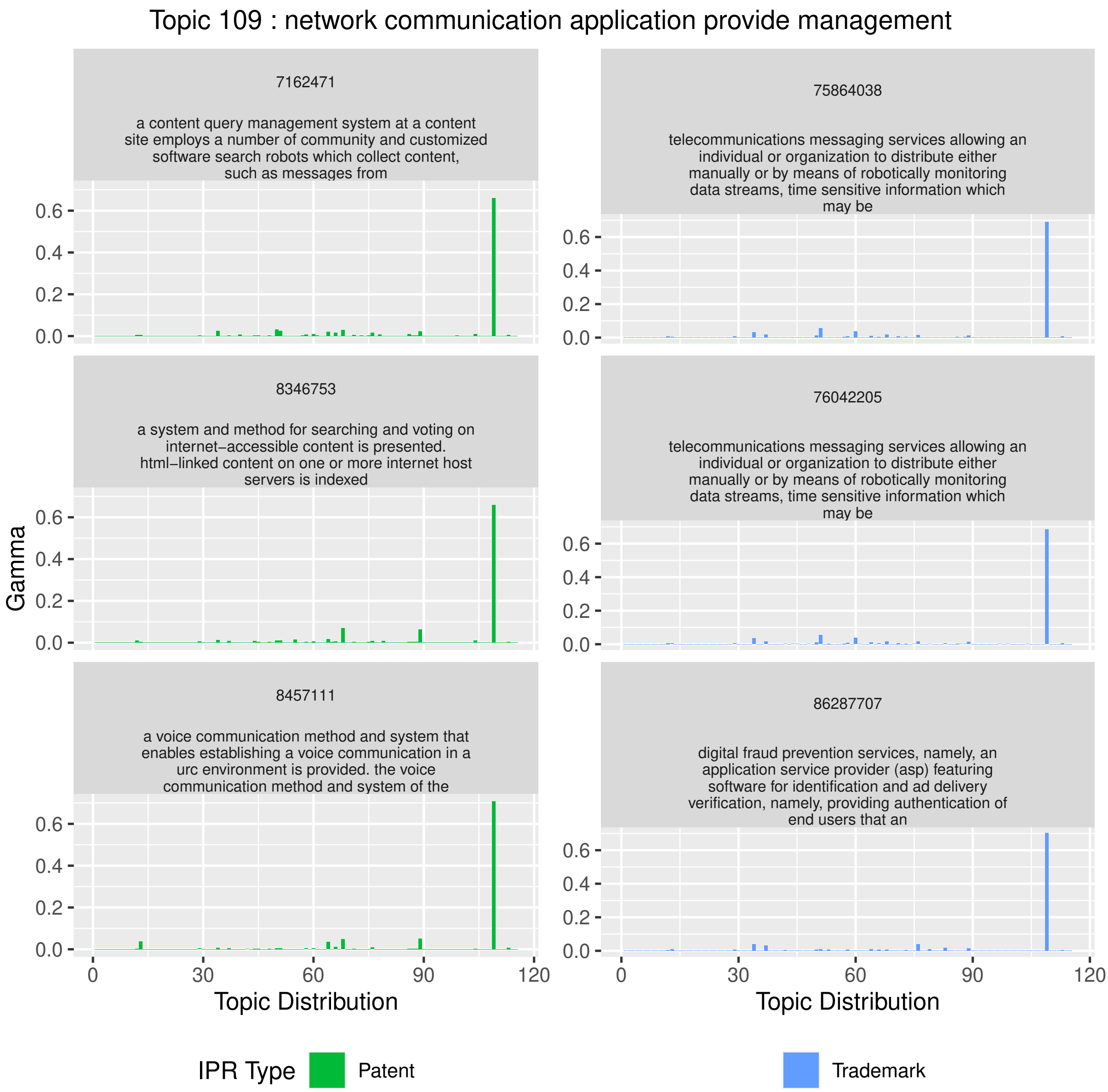
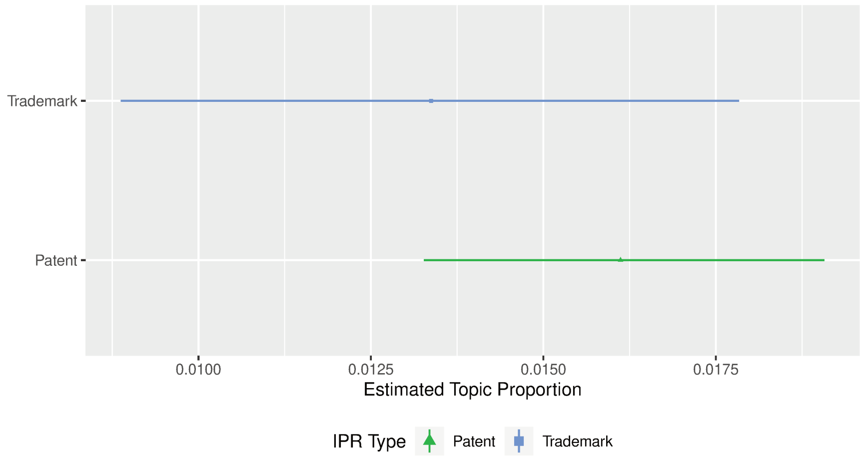


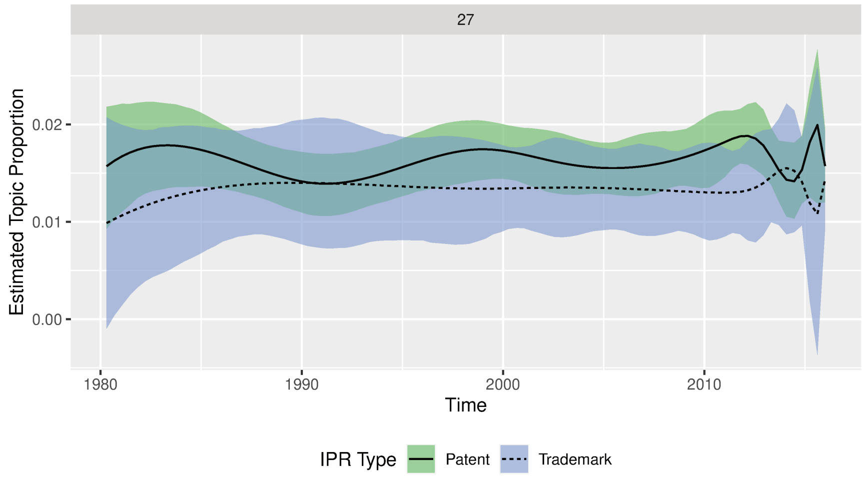
Figure C.7: Overview of Robotic Topic 109.
Topic 109 has a strong overlap of patents and trademarks in the estimated topic occurrence, also in the time overview. The patent classes relate to robots (B25J9, Y10S901) and positioning (G05D1), while the Nice classes are related to, among others, machines (cl. 7), data transmission (cl. 9) and the development of computer and software (cl. 42). The documents in relation to the topic are focused on the topic. Both IPRs display a high level of detail in their descriptions and are in line with each other.

Analysis of Robotic Topic 27

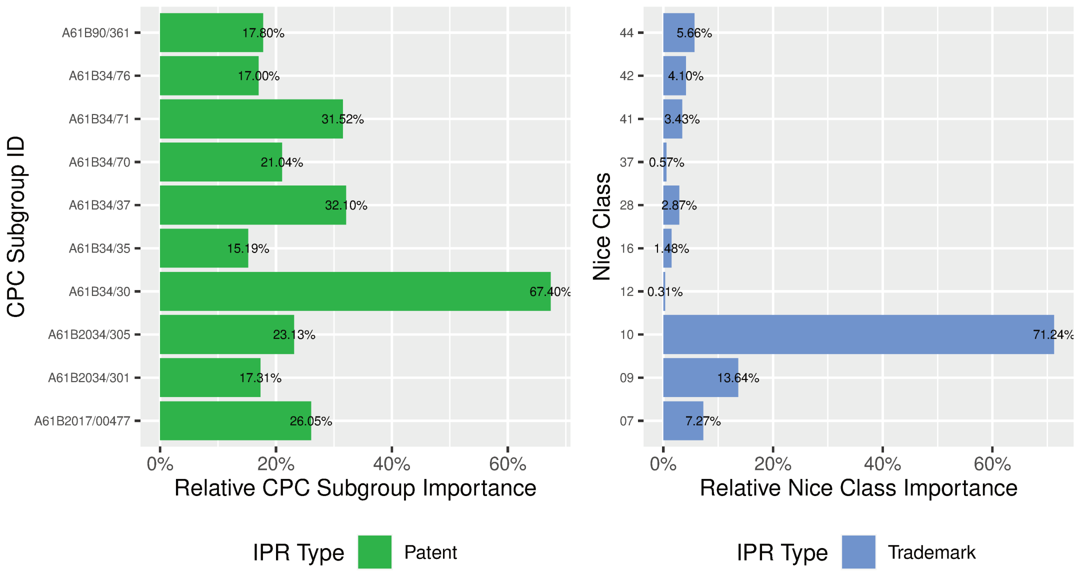
(a) Expected Topic Proportion per IPR in Topic 27



(b) Expected Topic Proportion per IPR in Topic 27 over Time



(c) Ten Main Classes of Topic 27 in Patents and Trademarks



(**d**) Main Documents in Topic 27

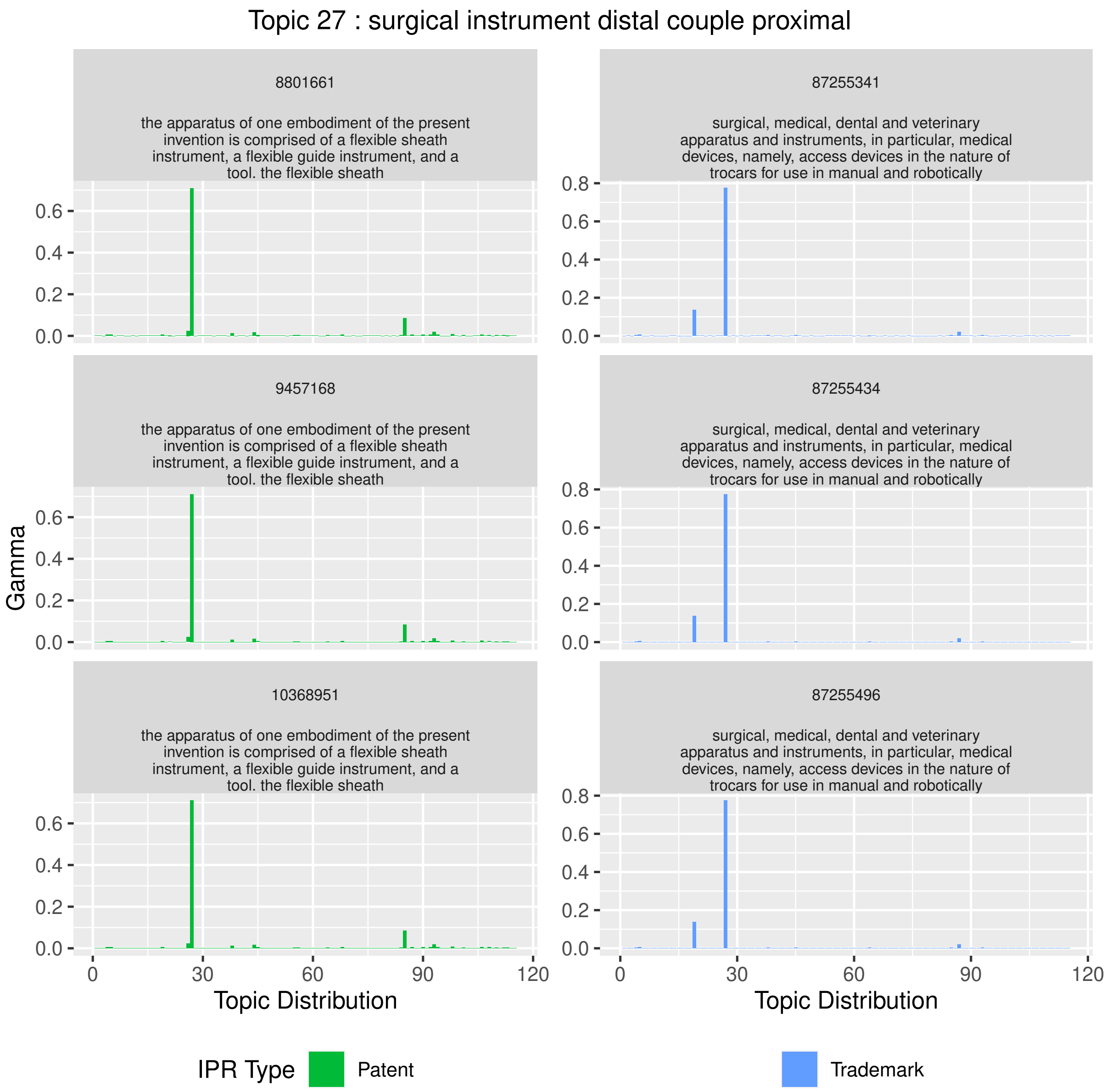
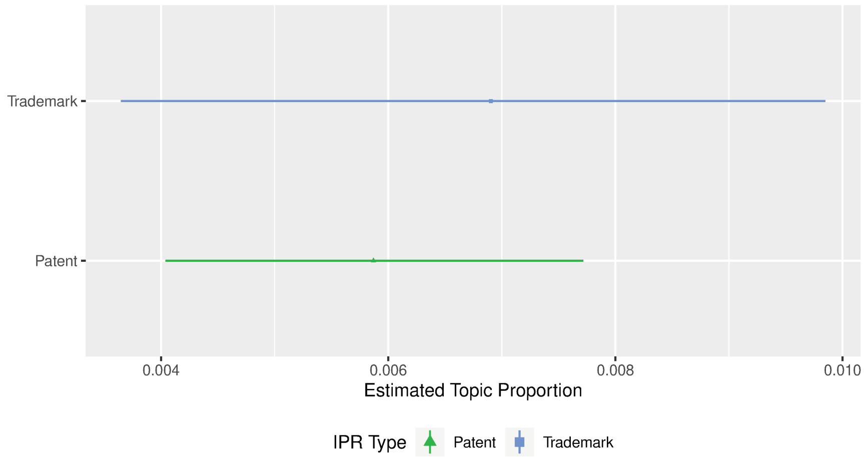


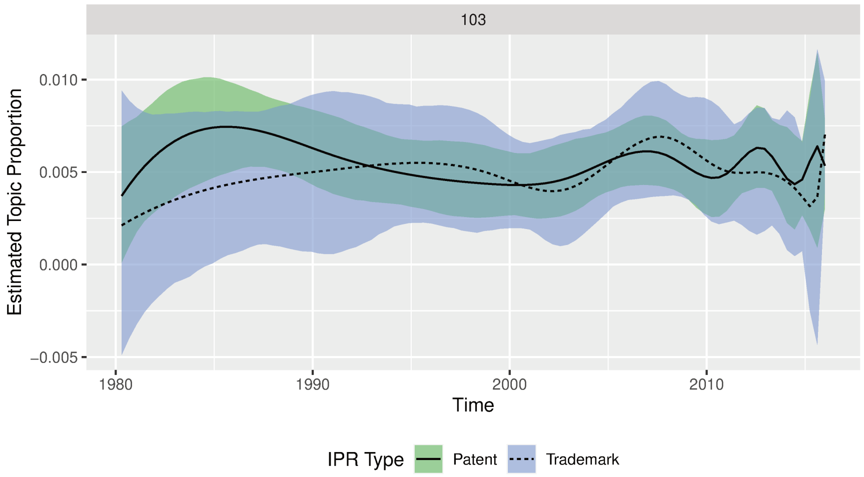
Figure C.8: Overview of Robotic Topic 27.
Trademarks and patents do strongly overlap in topic 27. Neither of the two IPRs is significant in the topic. Patent and trademark classification relate to surgery and medical topics. This is in line with the textual description. Here the patent description is more precise in the invention itself, while the trademarks describe the application area of the instruments in the surgical context.

Analysis of Robotic Topic 103

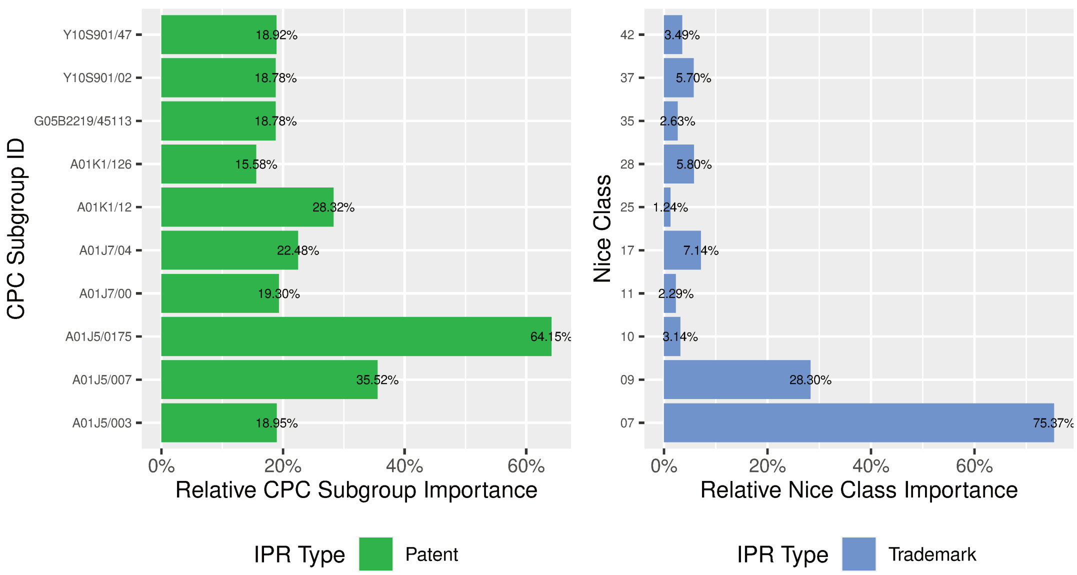
(a) Expected Topic Proportion per IPR in Topic 103



(b) Expected Topic Proportion per IPR in Topic 103 over Time



(c) Ten Main Classes of Topic 103 in Patents and Trademarks



(**d**) Main Documents in Topic 103

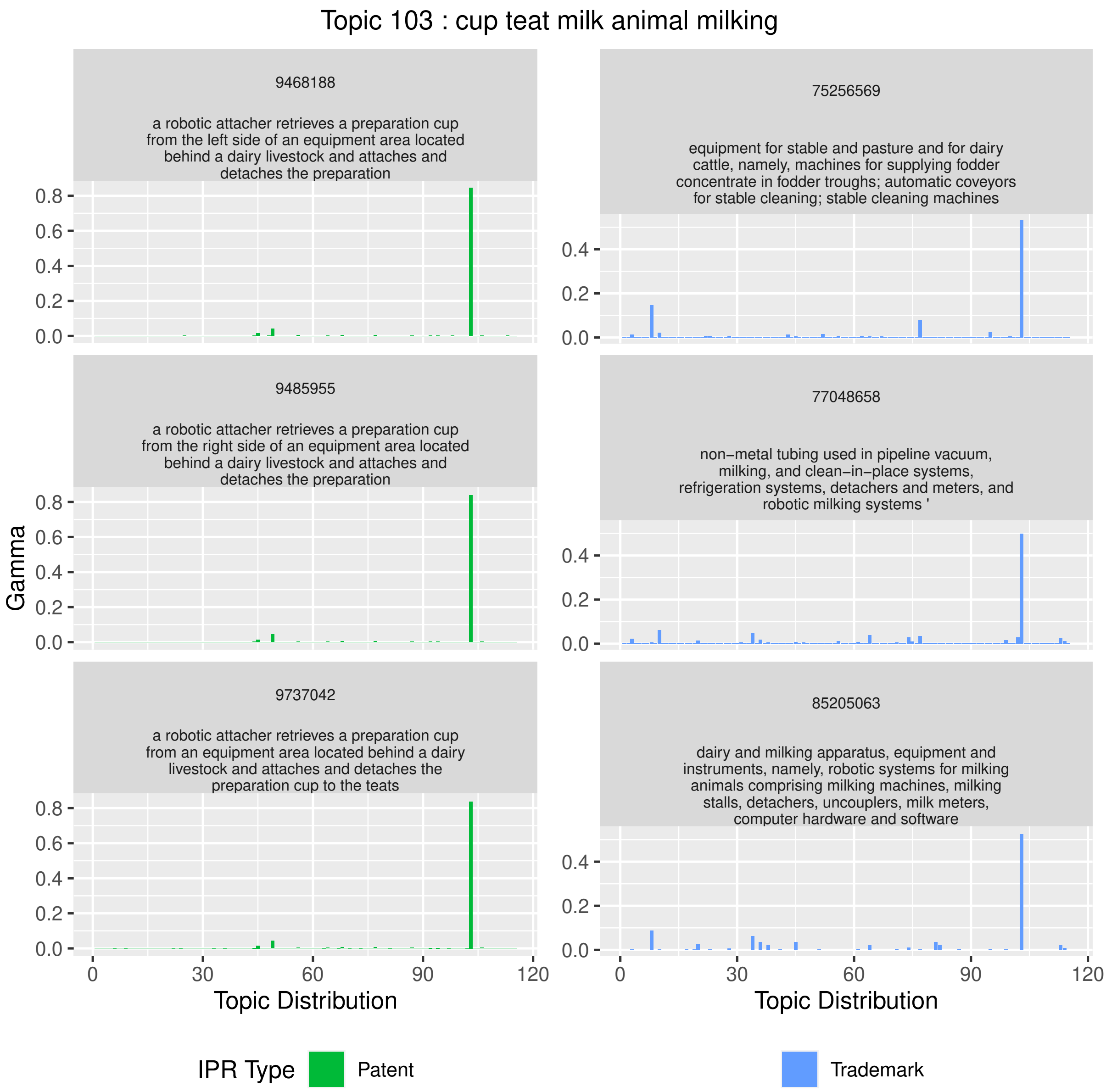
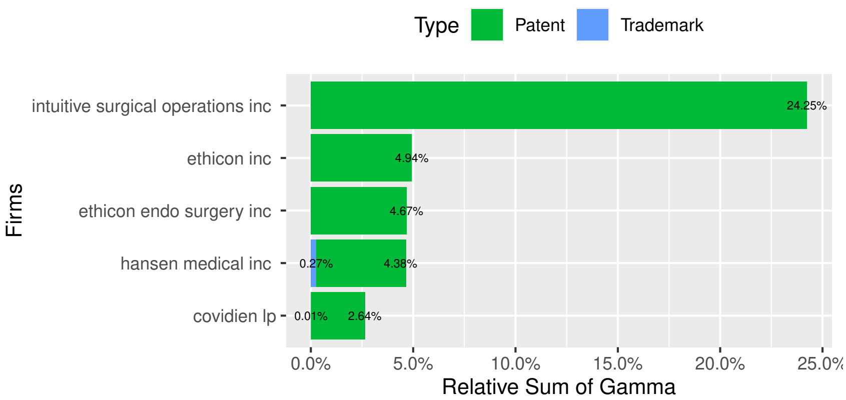


Figure C.9: Overview of Robotic Topic 103.
Topic 103 is equally likely to occur in trademarks and in patents. It covers diary and milking aspects in robotics. This is an area of service robotic application. The application areas described in trademarks are very precise and provide additional insights on the application of milking robots.

(a) Five Major Firms in Topic 27



(b) Five Major Firms in Topic 103

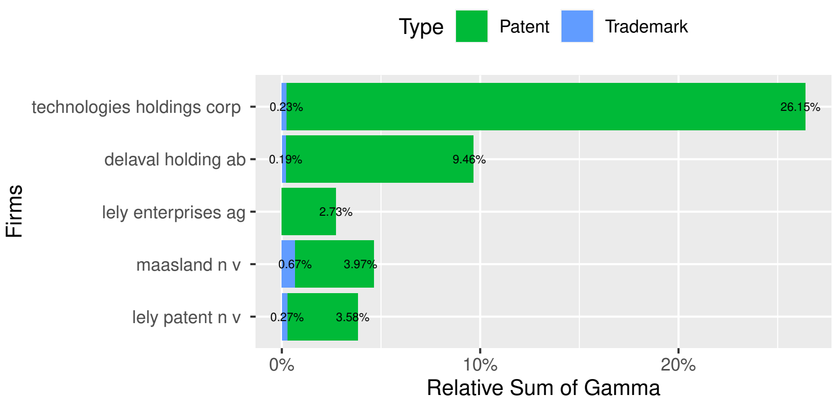


Figure C.10: Five Major Firms in Topics 27 and 103.
The figures display the major five firms active in trademarking and patenting in the topics 27 and 103. The relative sum of gamma is displayed which indicates the proportion of the firm documents in the topic. The relative sum of gamma over all firms add up to 100%.

C.2 Footwear

Subsection C.2.1 discloses the queries used to identify the Footwear data set. Additionally, background information on the considered patent CPC subgroups are provided. Subsection C.2.2 then provides further detail of the model selection, especially disclosing the metrics to evaluate the topic number K (see Figure C.11), the trade-off between exclusivity and semantic coherence in general (Figure C.12a), and for selected models (Figure C.12b). Subsection C.2.3 then discloses general information on the data set such as the development of registered documents over time (see Figure 4.12) or the distribution of patent (Figure C.13a) or trademark classes (Figure C.13b) in the data set, further information on the topics in general like the significance of the different topics in the IPRs (Figure C.15) or main documents for selected topics (Subsection C.2.3).

C.2.1 Data Selection

Patent Selection

```sql
select distinct cp.patent_id
into fs.foot_uspto_pat
from uspto.cpc_current cp
inner join fs.foot_uspto_cpc_subgroup_selected se on se.id = cp.subgroup_id

union

select distinct pa.id as 'patent_id'
from uspto.patent pa
where
(abstract like '%shoe%' or title like '%shoe%') and (abstract like '%foot%' or
title like '%foot%')
or (abstract like '%footwear' or title like '%footwear%')
```

Query C.6: All Patents of Footwear

```sql
select distinct patent_id
into fs.foot_uspto_pat_util
from fs.foot_uspto_pat mu
join uspto.patent pa on mu.patent_id = pa.id
where pa.type in ('utility')
```

Query C.7: Only Utility Patents of Footwear

Trademark Selection

Trademarks are selected based on keywords in the trademark name, pseudo name or in the goods and service field description (see Query C.8):

Case 1: Trademarks with “footwear” in the name, or with “shoe” in the name and “foot” in the name or description are selected.

Case 2: Trademarks with “shoe” and “foot”, or “footwear” as pseudo names are selected.

Case 3: Trademarks with “foot” and “shoe”, or with “footwear” in their goods and service description are selected.

Table C.3: CPC Subgroup IDs used of the Selection of Footwear Patents

CPC Subgroup ID	Description
A43B1/00	Footwear characterised by the material
A43B1/0009	Footwear characterised by the material -Footwear made at least partially of alveolar or honeycomb material
A43B1/0018	Footwear characterised by the material -Footwear made at least partially of flexible, bellow-like shaped material
A43B1/0027	Footwear characterised by the material -Footwear made at least partially from a material having special colours
A43B1/0036	Footwear characterised by the material -Footwear made at least partially from a material having special colours-with fluorescent or phosforescent parts
A43B1/0045	Footwear characterised by the material -Footwear made at least partially of deodorant means
A43B1/0054	Footwear characterised by the material -Footwear provided with magnets, magnetic parts or magnetic substances
A43B1/0063	Footwear characterised by the material -Footwear made at least partially of material that can be recycled
A43B1/0072	Footwear characterised by the material -Footwear made at least partially of transparent or translucent materials
A43B1/0081	Footwear characterised by the material -Footwear made at least partially of hook-and-loop type material
A43B1/009	Footwear characterised by the material -Footwear made at least partially of washable material
A43B1/02	Footwear characterised by the material -Footwear made of animal or plant fibres or fabrics made therefrom
A43B1/04	Footwear characterised by the material -Footwear made of animal or plant fibres or fabrics made therefrom-Braided, knotted, knitted, or crocheted footwear
A43B1/06	Footwear characterised by the material -Footwear made of wood, cork, card-board, paper or like fibrous material
A43B1/08	Footwear characterised by the material -Footwear made of metal
A43B1/10	Footwear characterised by the material -Footwear made of rubber
A43B1/12	Footwear characterised by the material -Footwear made of rubber-of rubber waste
A43B1/14	Footwear characterised by the material -Footwear made of gutta-percha, celluloid, or plastics
A43B11/00	Footwear with miscellaneous arrangements to facilitate putting-on or removing, e.g. with straps
A43B11/02	Footwear with miscellaneous arrangements to facilitate putting-on or removing, e.g. with straps-with built-in shoe-horns
A43B13/145	Soles ; Sole and heel units-characterised by the constructive form-provided with wedged, concave or convex end portions, e.g. for improving roll-off of the foot-Convex portions, e.g. with a bump or projection, e.g. 'Masai' type shoes
A43B13/38	Soles ; Sole and heel units-Built-in insoles joined to uppers during the manufacturing process, e.g. structural insoles; Insoles glued to shoes during the manufacturing process
A43B13/383	Soles ; Sole and heel units-Built-in insoles joined to uppers during the manufacturing process, e.g. structural insoles; Insoles glued to shoes during the manufacturing process-pieced
A43B13/386	Soles ; Sole and heel units-Built-in insoles joined to uppers during the manufacturing process, e.g. structural insoles; Insoles glued to shoes during the manufacturing process-multilayered
A43B13/39	Soles ; Sole and heel units-Built-in insoles joined to uppers during the manufacturing process, e.g. structural insoles; Insoles glued to shoes during the manufacturing process-with upset sewing ribs
A43B13/40	Soles ; Sole and heel units-Built-in insoles joined to uppers during the manufacturing process, e.g. structural insoles; Insoles glued to shoes during the manufacturing process-with cushions
A43B13/41	Soles ; Sole and heel units-Built-in insoles joined to uppers during the manufacturing process, e.g. structural insoles; Insoles glued to shoes during the manufacturing process-combined with heel stiffener, toe stiffener, or shank stiffener
A43B15/00	Welts for footwear
A43B17/00	Insoles for insertion, e.g. footbeds or inlays, for attachment to the shoe after the upper has been joined
A43B17/003	Insoles for insertion, e.g. footbeds or inlays, for attachment to the shoe after the upper has been joined -characterised by the material
A43B17/006	Insoles for insertion, e.g. footbeds or inlays, for attachment to the shoe after the upper has been joined -characterised by the material -multilayered
A43B17/02	Insoles for insertion, e.g. footbeds or inlays, for attachment to the shoe after the upper has been joined -wedge-like or resilient
A43B17/023	Insoles for insertion, e.g. footbeds or inlays, for attachment to the shoe after the upper has been joined -wedge-like or resilient-wedge-like
A43B17/026	Insoles for insertion, e.g. footbeds or inlays, for attachment to the shoe after the upper has been joined -wedge-like or resilient-filled with a non-compressible fluid, e.g. gel, water
A43B17/03	Insoles for insertion, e.g. footbeds or inlays, for attachment to the shoe after the upper has been joined -wedge-like or resilient-filled with a gas, e.g. air
A43B17/035	Insoles for insertion, e.g. footbeds or inlays, for attachment to the shoe after the upper has been joined -wedge-like or resilient-filled with a gas, e.g. air-provided with a pump or valve
A43B17/04	Insoles for insertion, e.g. footbeds or inlays, for attachment to the shoe after the upper has been joined -with metal insertions or coverings
A43B17/06	Insoles for insertion, e.g. footbeds or inlays, for attachment to the shoe after the upper has been joined -with metal springs
A43B17/08	Insoles for insertion, e.g. footbeds or inlays, for attachment to the shoe after the upper has been joined -ventilated
A43B17/10	Insoles for insertion, e.g. footbeds or inlays, for attachment to the shoe after the upper has been joined -specially adapted for sweaty feet; waterproof
A43B17/102	Insoles for insertion, e.g. footbeds or inlays, for attachment to the shoe after the upper has been joined -specially adapted for sweaty feet; waterproof-Moisture absorbing socks; Moisture dissipating socks
A43B17/105	Insoles for insertion, e.g. footbeds or inlays, for attachment to the shoe after the upper has been joined -specially adapted for sweaty feet; waterproof-Moisture absorbing socks; Moisture dissipating socks-Disposable
A43B17/107	Insoles for insertion, e.g. footbeds or inlays, for attachment to the shoe after the upper has been joined -specially adapted for sweaty feet; waterproof-waterproof
A43B17/12	Insoles for insertion, e.g. footbeds or inlays, for attachment to the shoe after the upper has been joined -made of wood
A43B17/14	Insoles for insertion, e.g. footbeds or inlays, for attachment to the shoe after the upper has been joined -made of sponge, rubber, or plastic materials
A43B17/16	Insoles for insertion, e.g. footbeds or inlays, for attachment to the shoe after the upper has been joined -with heel or toe caps
A43B17/18	Insoles for insertion, e.g. footbeds or inlays, for attachment to the shoe after the upper has been joined -Arrangements for attaching removable insoles to footwear
A43B19/00	Shoe-shaped inserts; Inserts covering the instep
A43B19/005	Shoe-shaped inserts; Inserts covering the instep -Weighted inserts for shoes, i.e. insert comprising an additional weight
A43B23/00	Uppers; Boot legs; Stiffeners; Other single parts of footwear
A43B23/02	Uppers; Boot legs; Stiffeners; Other single parts of footwear-Uppers; Boot legs
A43B23/0205	Uppers; Boot legs; Stiffeners; Other single parts of footwear-Uppers; Boot legs -characterised by the material
A43B23/021	Uppers; Boot legs; Stiffeners; Other single parts of footwear-Uppers; Boot legs -characterised by the material -Leather
A43B23/0215	Uppers; Boot legs; Stiffeners; Other single parts of footwear-Uppers; Boot legs -characterised by the material -Plastics or artificial leather
A43B23/022	Uppers; Boot legs; Stiffeners; Other single parts of footwear-Uppers; Boot legs -characterised by the material -Plastics or artificial leather-with waterproof breathable membranes

The table provides an extraction of the selected CPC Subgroups and their description. Full list available upon request.
Source: Own representation based on USPTO data.

```sql
select  st.statement_type_cd, st.statement_text, st.serial_no
into fs.foot_trademarks_statement_20210118
from trademarks.[statement] st
join trademarks.case_file cf on cf.serial_no = st.serial_no
where cf.serial_no in (
select cf.serial_no
from trademarks.case_file cf
join trademarks.statement st on cf.serial_no = st.serial_no
where
-- Case 1:
cf.mark_id_char like '%shoe%'  and  -- mark_id_char: Trademark Name
(cf.mark_id_char like '%foot%' or st.statement_text like '%foot%')
or cf.mark_id_char like '%footwear%'
)
and left(statement_type_cd, 2) in ('GS') -- GS: Goods and Service Field
union
select *
from trademarks.statement
where serial_no in (
select serial_no
from trademarks.[statement]
where
-- Case 2
((statement_text like '%shoe%' and statement_text like '%foot%') or statement_text
    like '%footwear%')
and left(statement_type_cd, 2) in ('PM') -- PM: Pseudo Mark
)
and left(statement_type_cd, 2) in ('GS')
union
select *
from trademarks.statement
-- Case 3
where ((statement_text like '%shoe%' and statement_text like '%foot%')
or statement_text like '%footwear%')
and left(statement_type_cd, 2) in ('GS')
```

Query C.8: All Trademarks of Footwear

```sql
select mu.statement_type_cd, mu.statement_text, mu.serial_no
into fs.foot_trademarks_statement_registered_20210118
from fs.foot_trademarks_statement_20210118 mu
join trademarks.case_file cf on mu.serial_no = cf.serial_no
where cf.registration_no not in ('0000000')
```

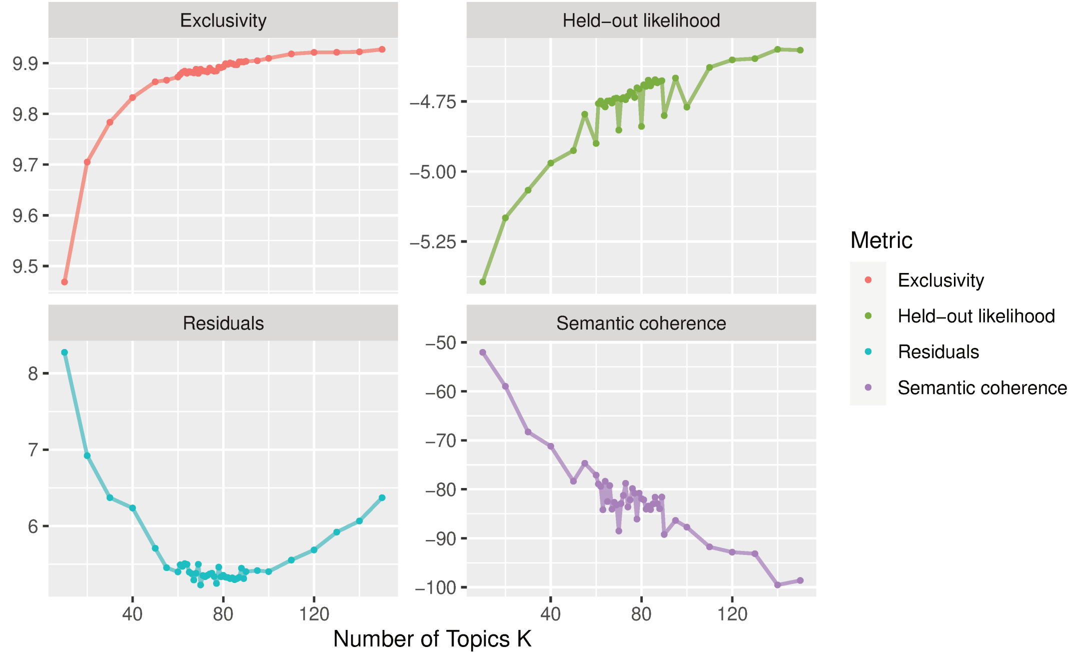
Query C.9: Only Registered Trademarks of Footwear

```sql
select distinct mu.serial_no
into fs.foot_trademarks_registered_20210118
from fs.foot_trademarks_statement_registered_20210118 mu
```

Query C.10: Distinct Registered Trademarks of Footwear

C.2.2 Model Selection in Footwear

(a) Overview of the Model Diagnostics in Footwear.



(b) Zoomed Model Diagnostics in Footwear.

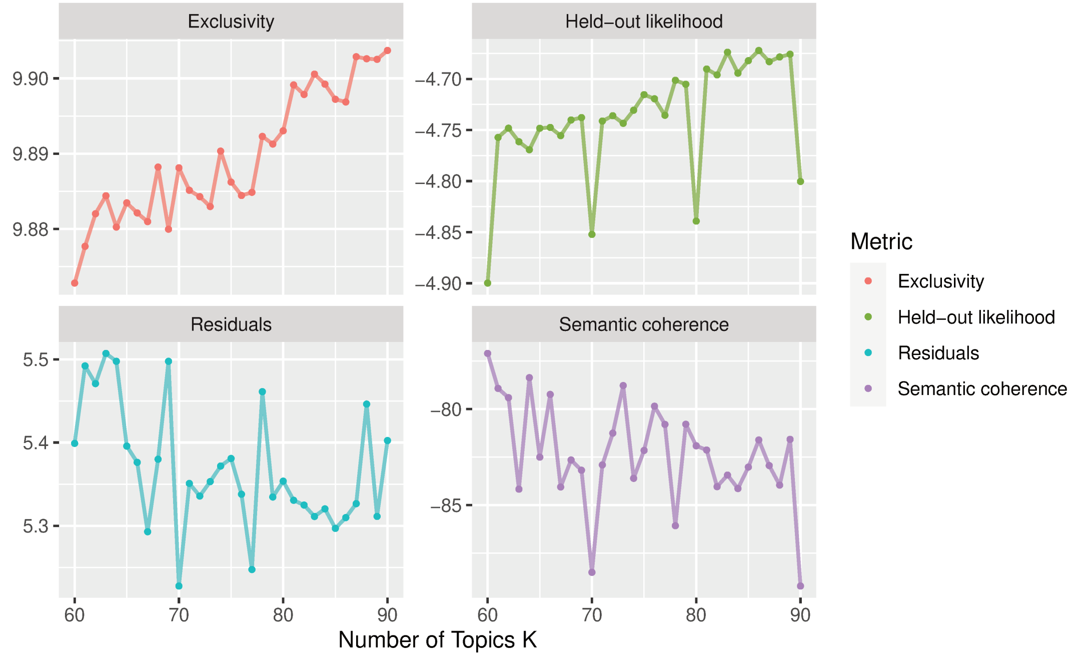
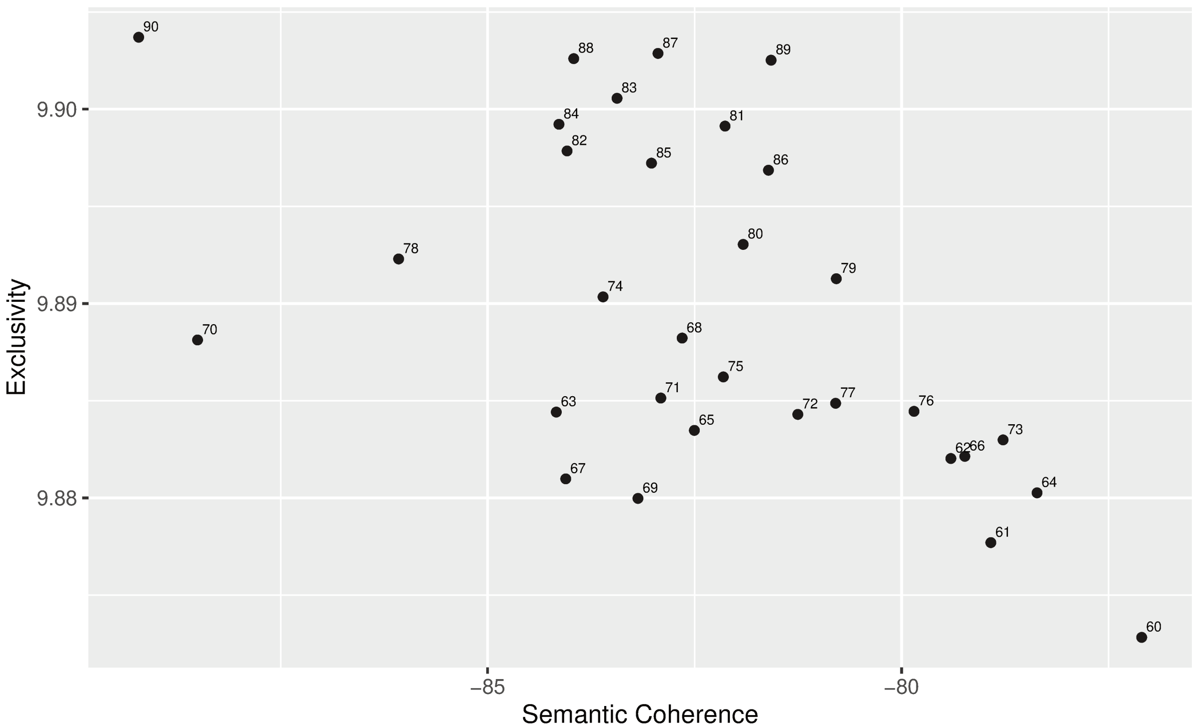
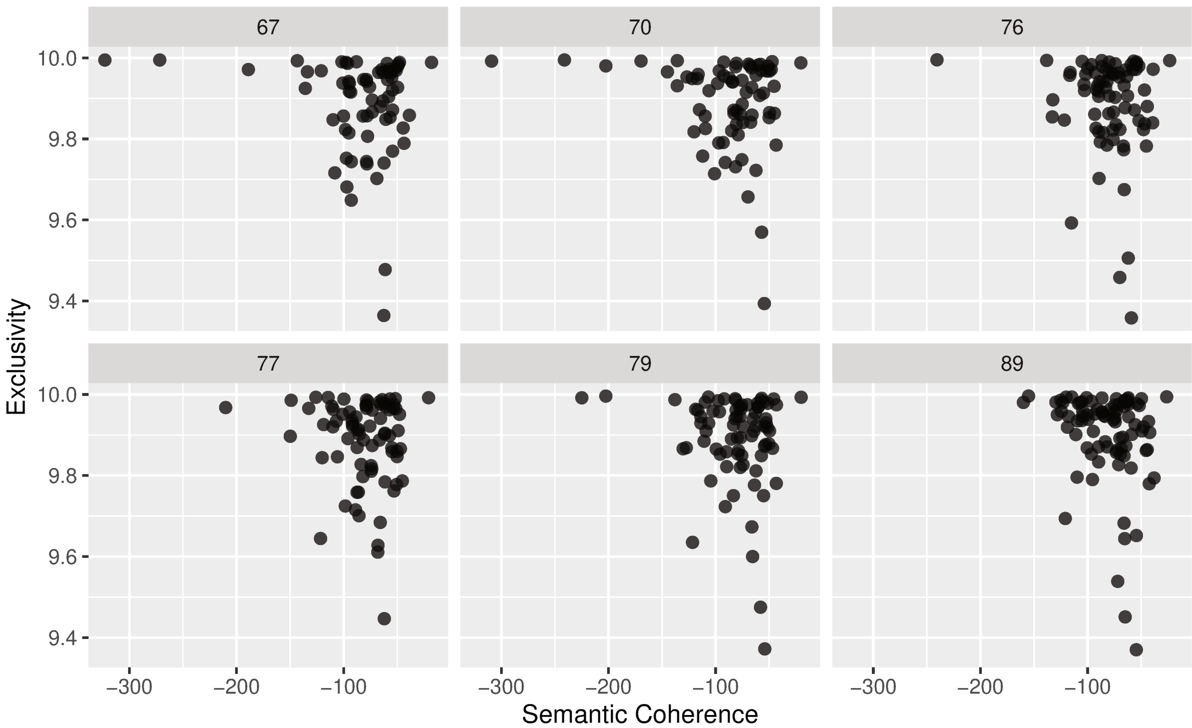


Figure C.11: Metrics to Evaluate the Topic Number K in Footwear.
The general diagnostics provide a first overview to determine the topic number K. For each model calculated, the exclusivity, held-out likelihood, residuals and semantic coherence are determined. Every point in the graph stands for a different model and its diagnostics. To determine K, the metrics are weighted against each other. In case of residuals, models at the lowest point of the slope are of interest. The figure provides an overview in Subfigure (a) and a zoomed perspective in Subfigure (b). The zoomed perspective makes it easier to select models where a closer look might be of interest.



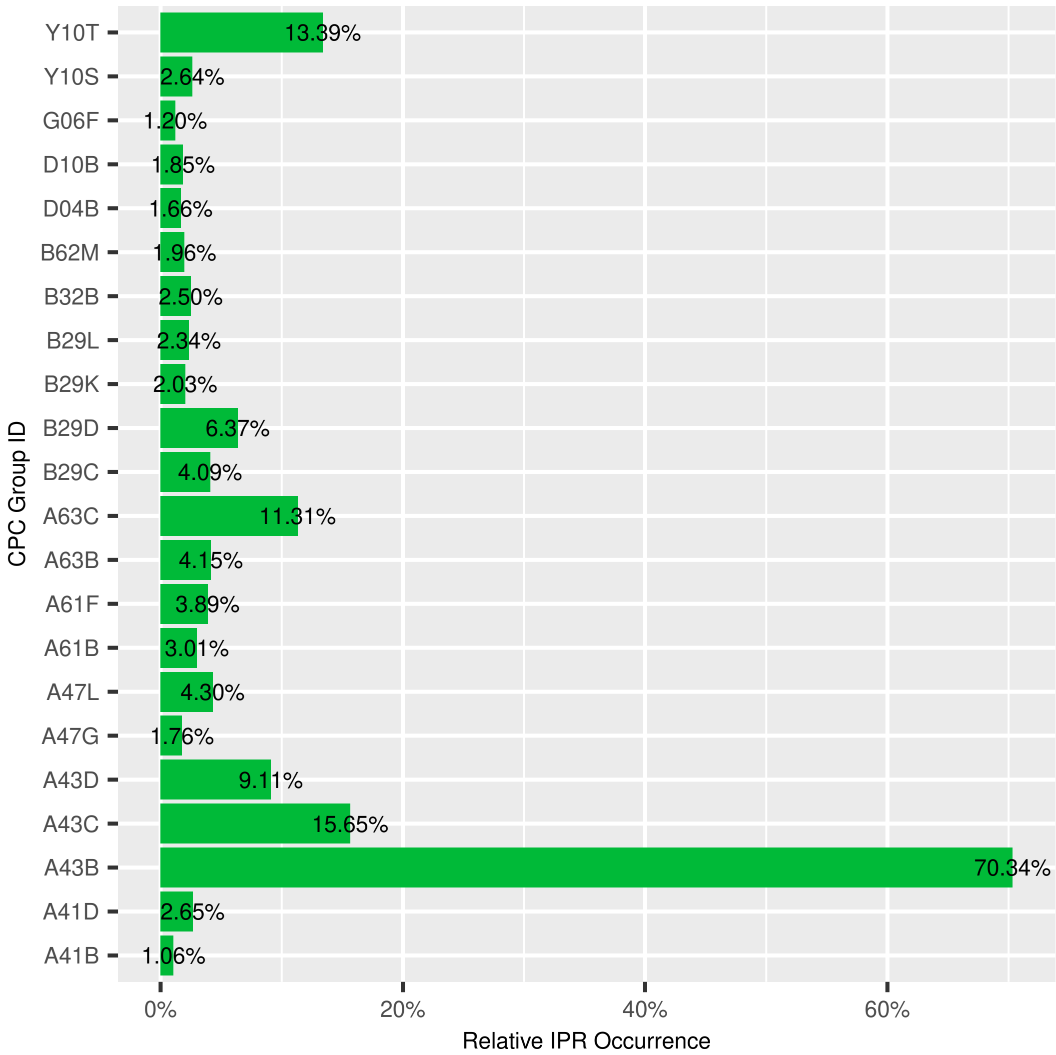
(a) Mean Exclusivity-Semantic Coherence Trade-off of different Models.



(b) Exclusivity-Semantic Coherence Trade-off of the Topics of Selected Models.

Figure C.12: Exclusivity and Semantic Coherence of Footwear Models in Comparison. Subfigure (a) displays the mean exclusivity against the mean semantic coherence for each model. A trade-off between exclusivity and semantic coherence needs to be found. After choosing potential models, the exact exclusivity-semantic coherence trade-off per topic in the relevant models are displayed in Subfigure (b). Here, each frame represents a model. Each point stands for the trade-off of a topic of the model. In general, models with fewer topics have a higher semantic coherence for more topics, but lower exclusivity.

(**a**) Patent Classification.



C.2.3 Topic Modelling Results

(b) Trademark Classification.

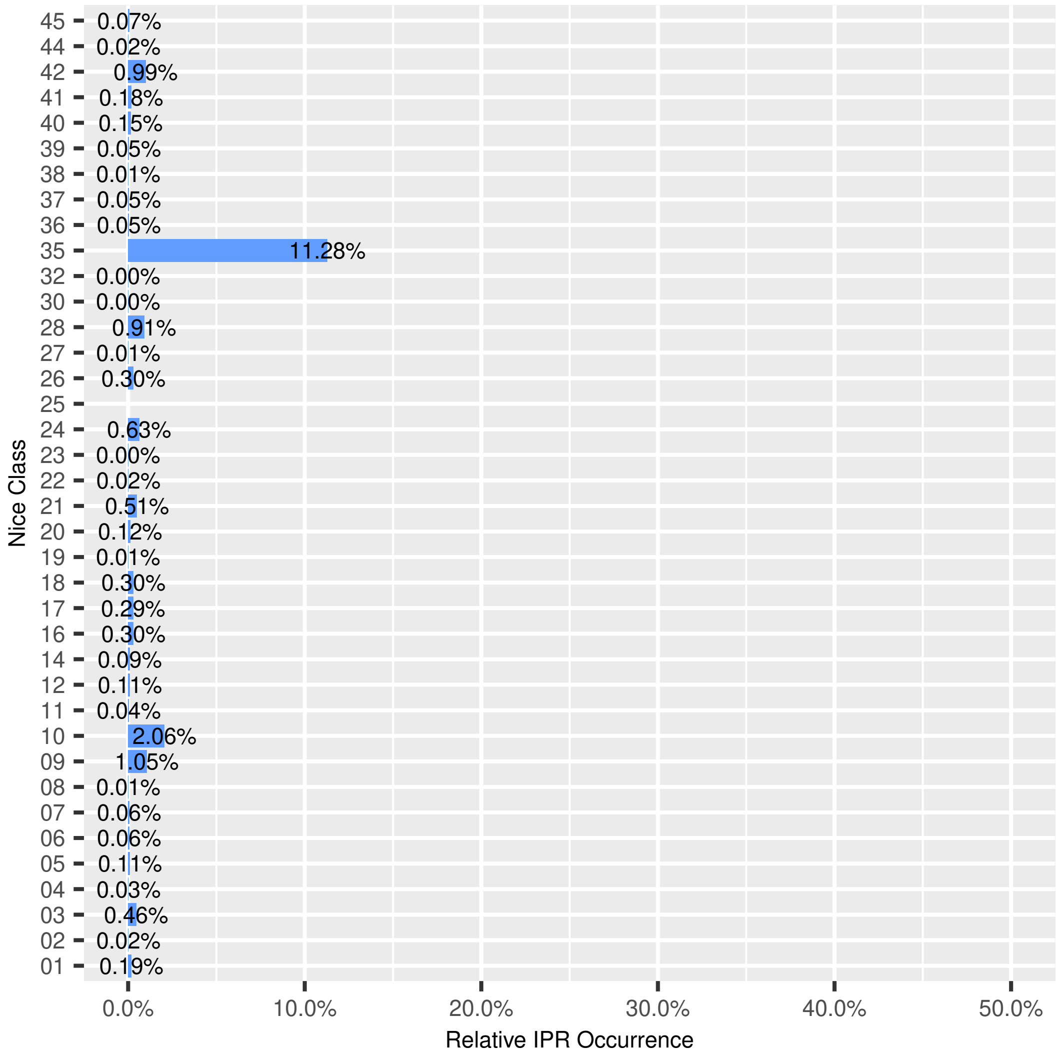


Figure C.13: Footwear Trademarks and Patents in Existing Classification Schemes. Subfigure (a) displays the CPC groups of the patent documents that have a total share of more than 1%. Subfigure (b) displays the Nice classes of the trademark documents. The respective classes are ordered alphabetically in a descending order. The length of the bar indicates the relative occurrence of the CPC group compared to the total amount of patent documents in case of patents or the relative occurrence of the Nice group compared to the total amount of trademark documents in case of trademarks. As a document can be assigned to more than one CPC group or more than on Nice class, the groups or classes do not sum up to 100%.

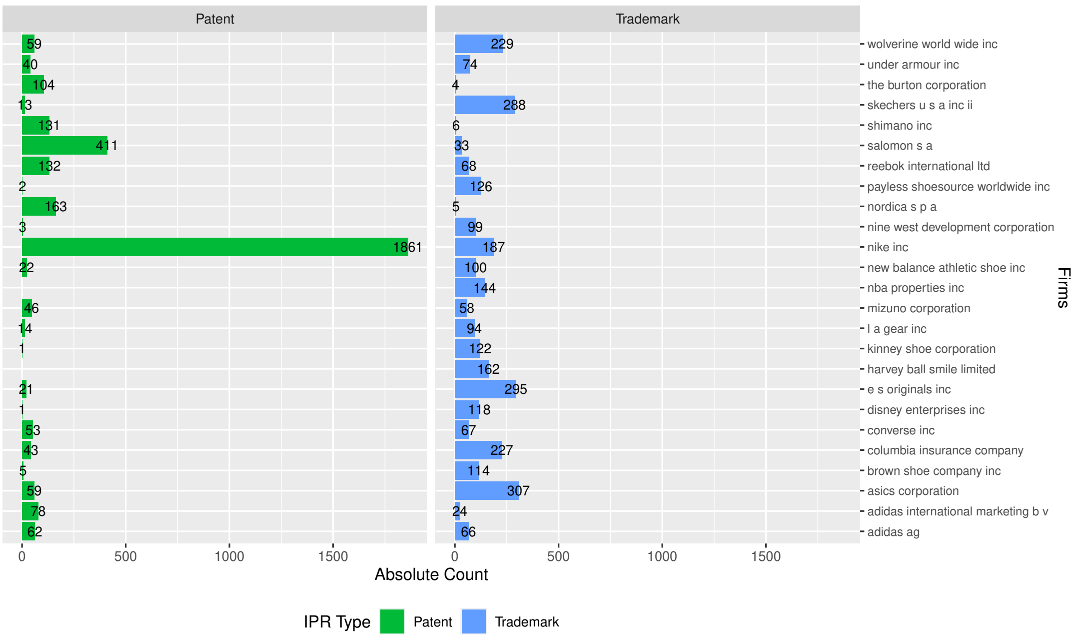


Figure C.14: Major 25 Firms in Footwear Patenting or Trademarking. The figure displays the major 25 firms with the most documents registered related to robotics. The numbers are determined based on the documents filed under the same owner name. For each firm, the entries are separated for patents and trademarks.

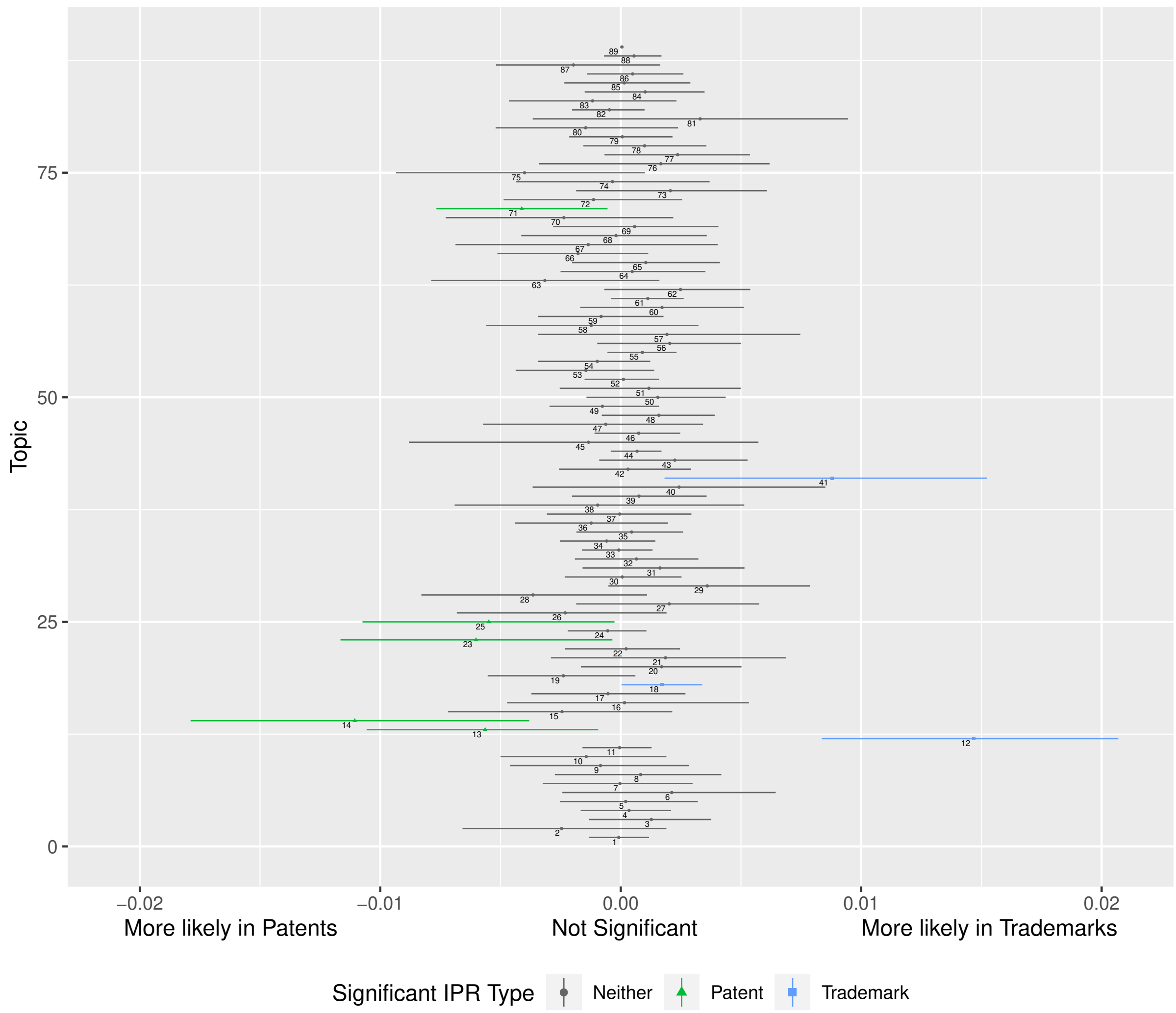


Figure C.15: Significance of Footwear Topics in Patents or Trademarks. Each topics relates to a varying degree to patents or trademarks. The figure weights the occurrence probability of trademarks and patents against each other. If the mean and the variance in this figure are more likely in one document type, the occurrence is significance. In case of patents and trademarks, the related topics are coloured accordingly.

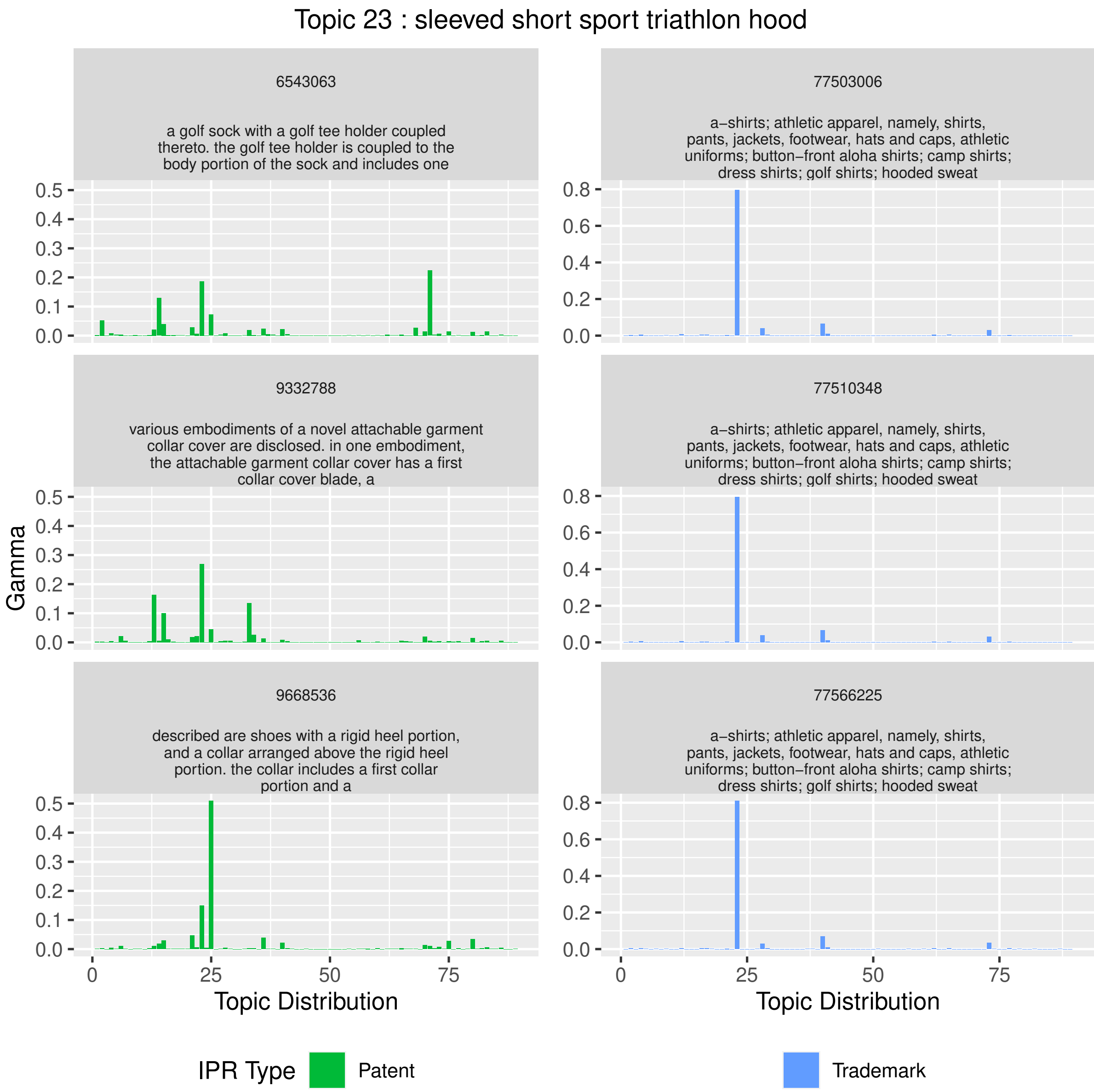


Figure C.16: Overview of Footwear Topic 23.

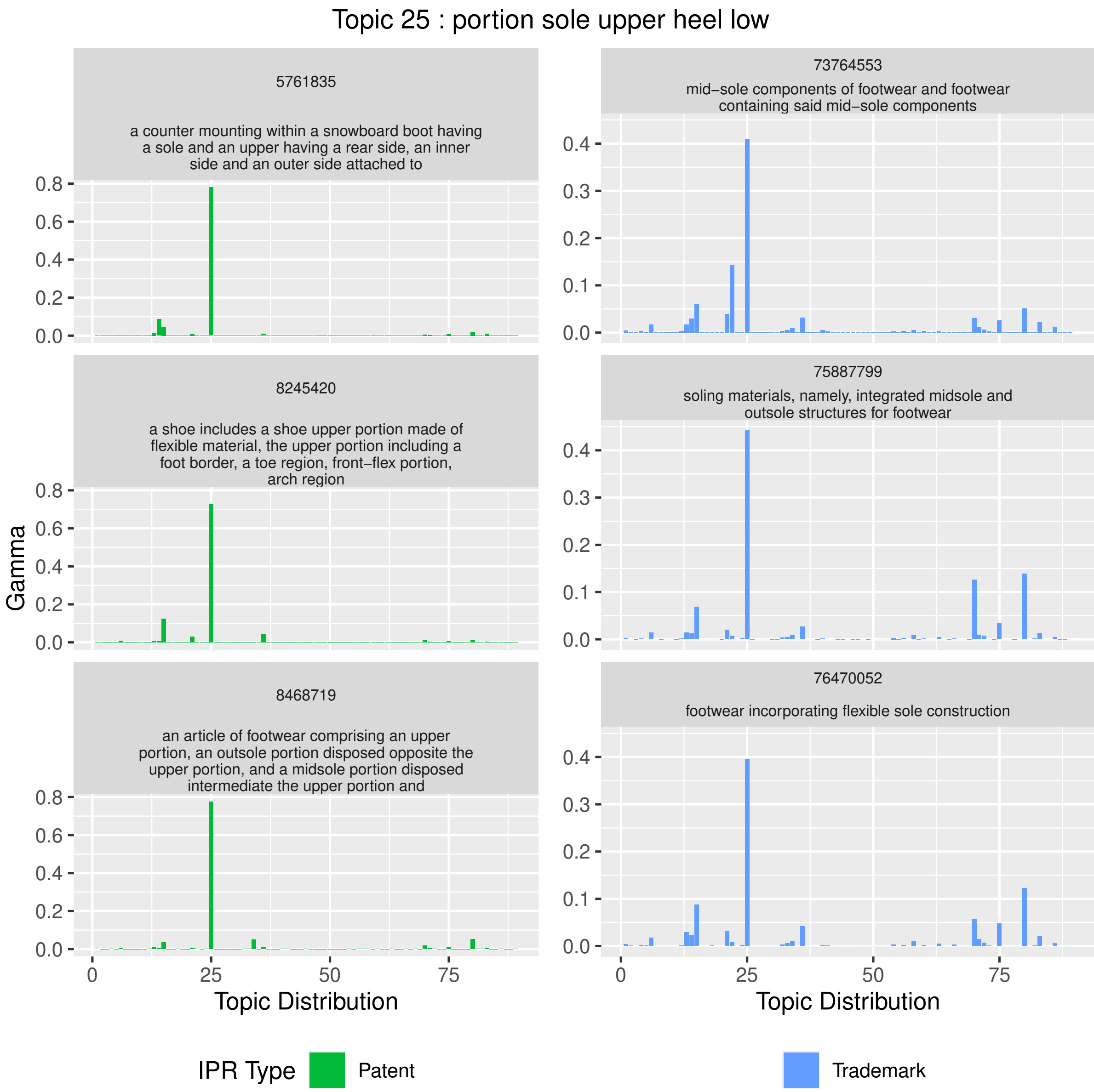


Figure C.17: Overview of Footwear Topic 25.

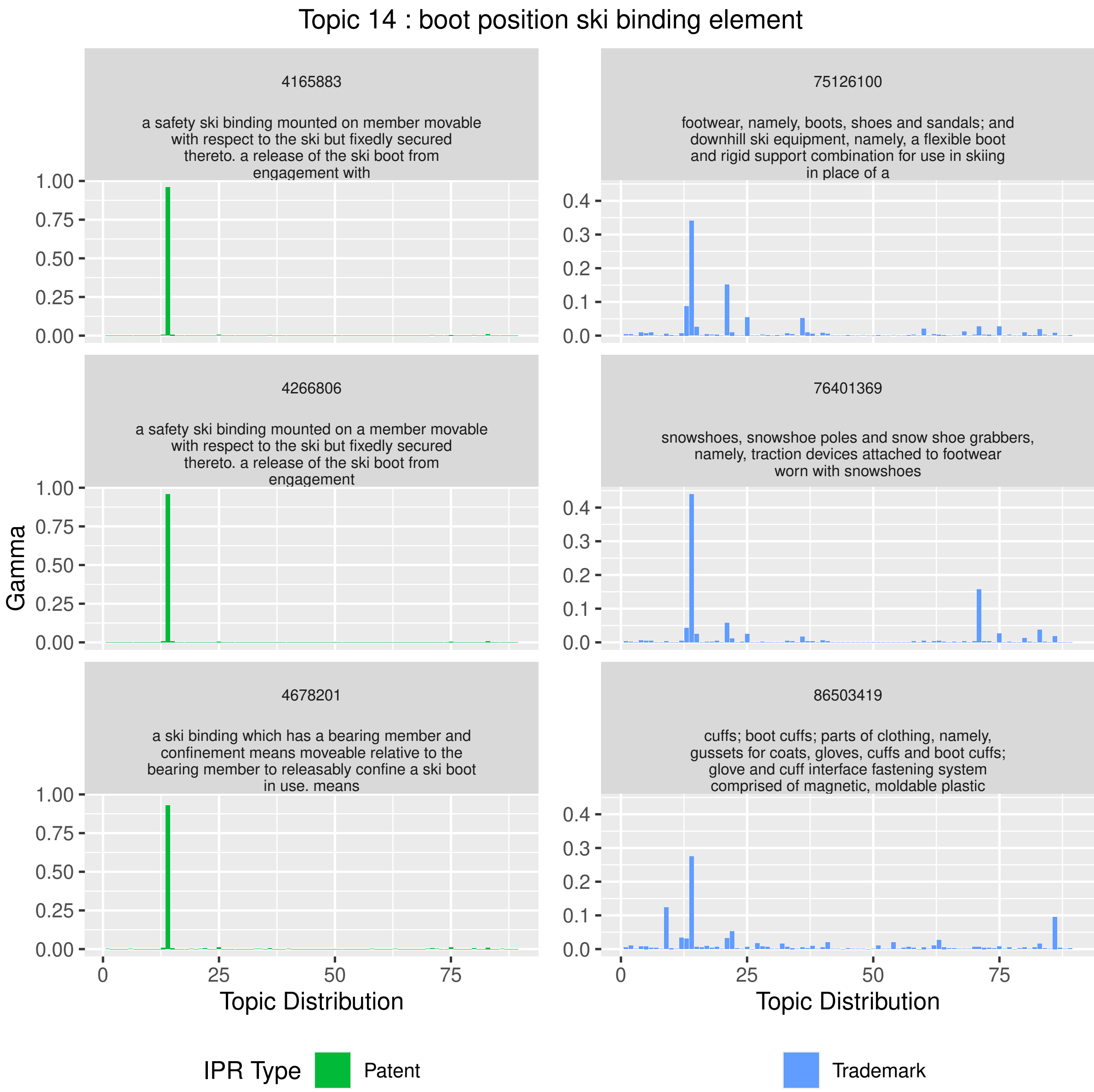


Figure C.18: Overview of Footwear Topic 14.

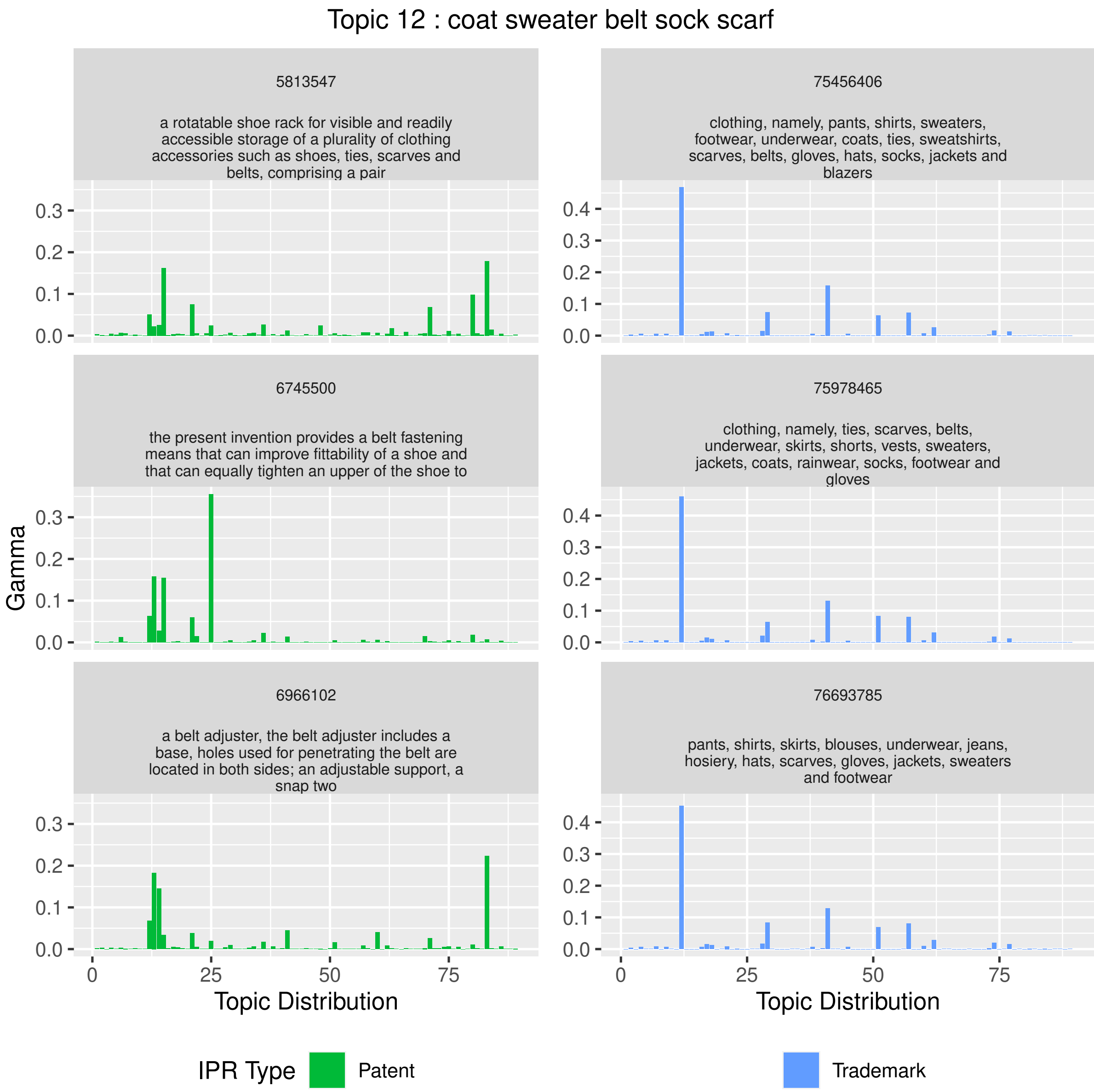


Figure C.19: Overview of Footwear Topic 12.

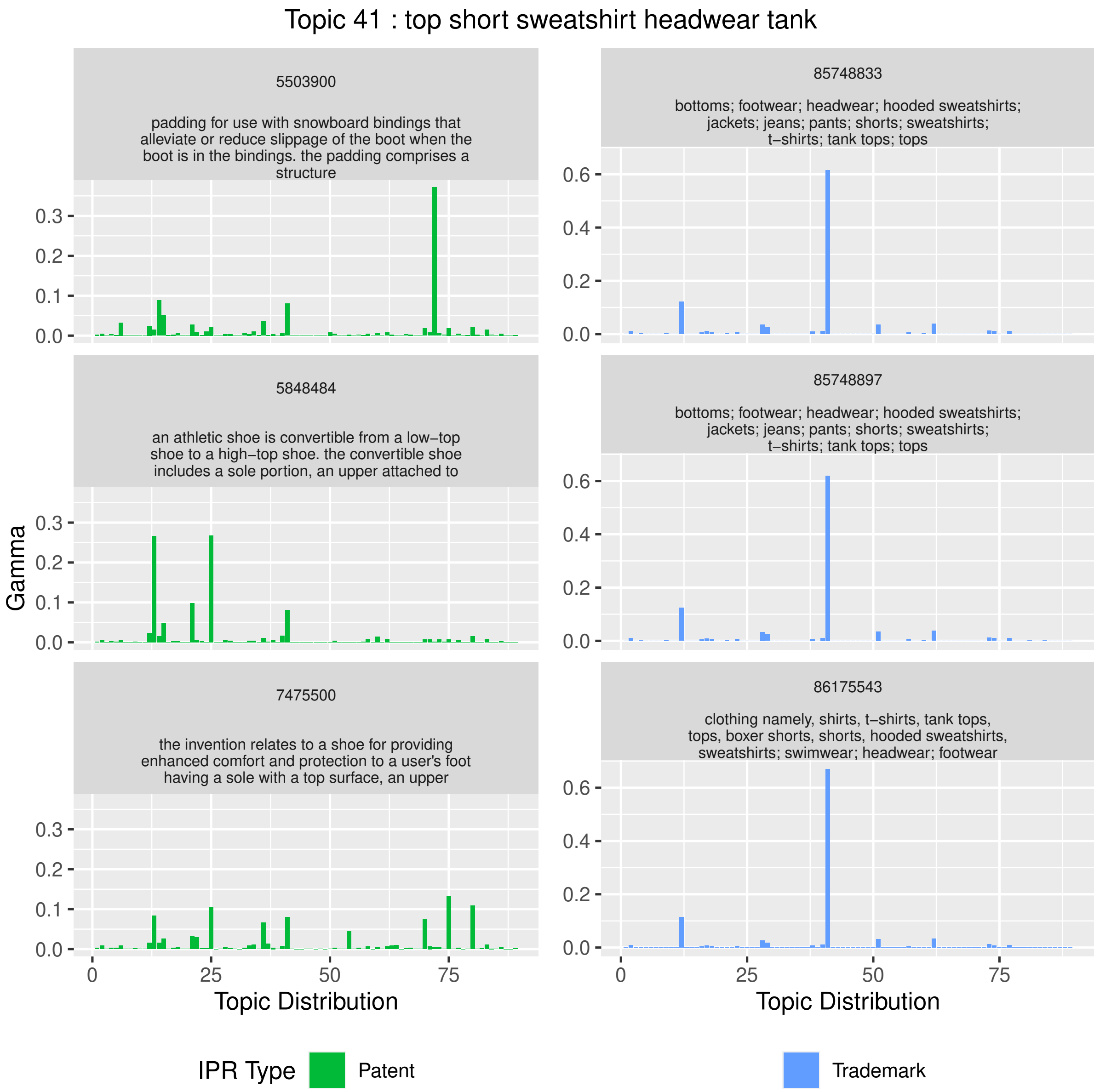


Figure C.20: Overview of Footwear Topic 41.

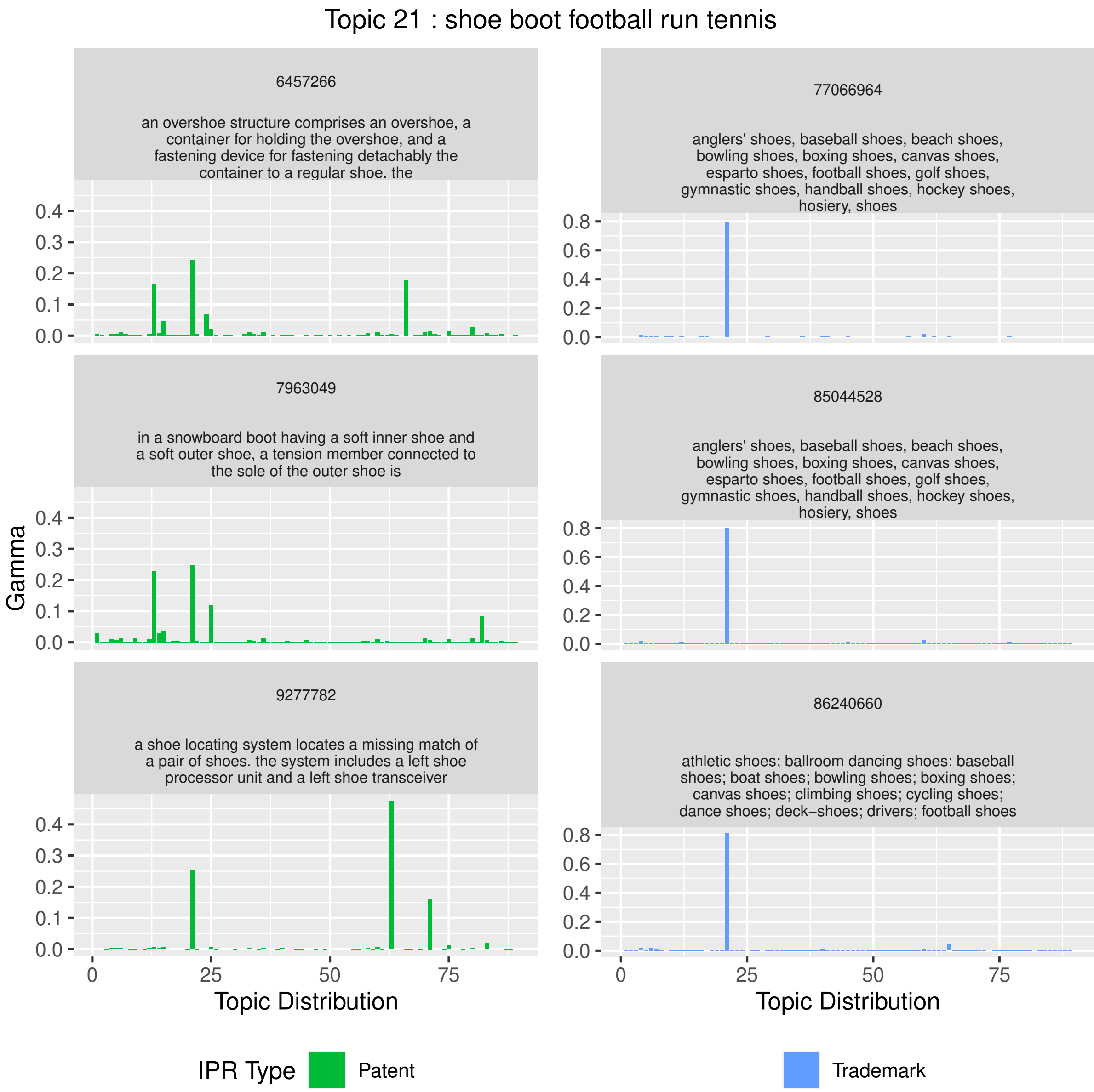


Figure C.21: Overview of Footwear Topic 21.

Figure C.22: Overview of Footwear Topic 29.

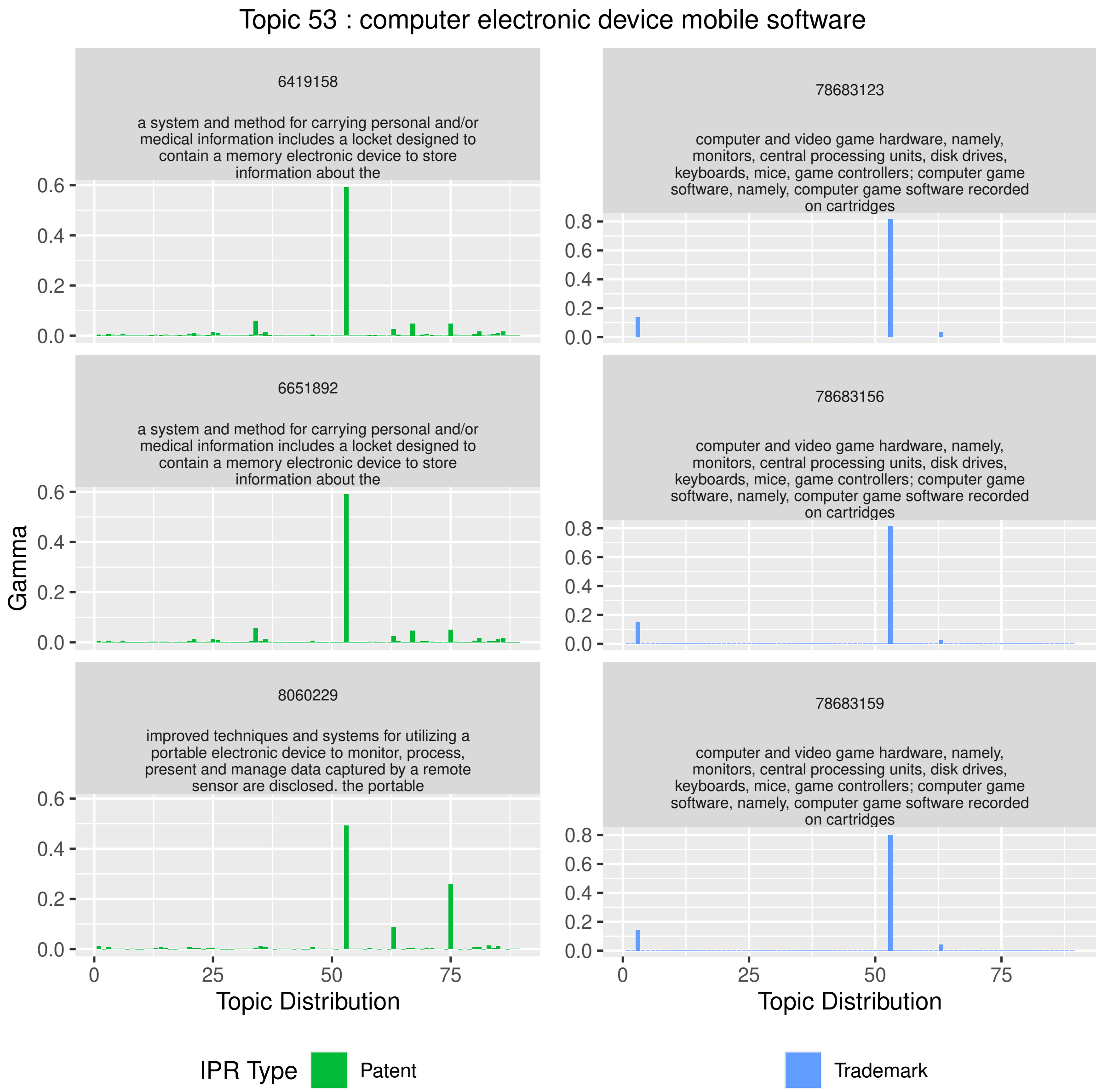


Figure C.23: Overview of Footwear Topic 53.

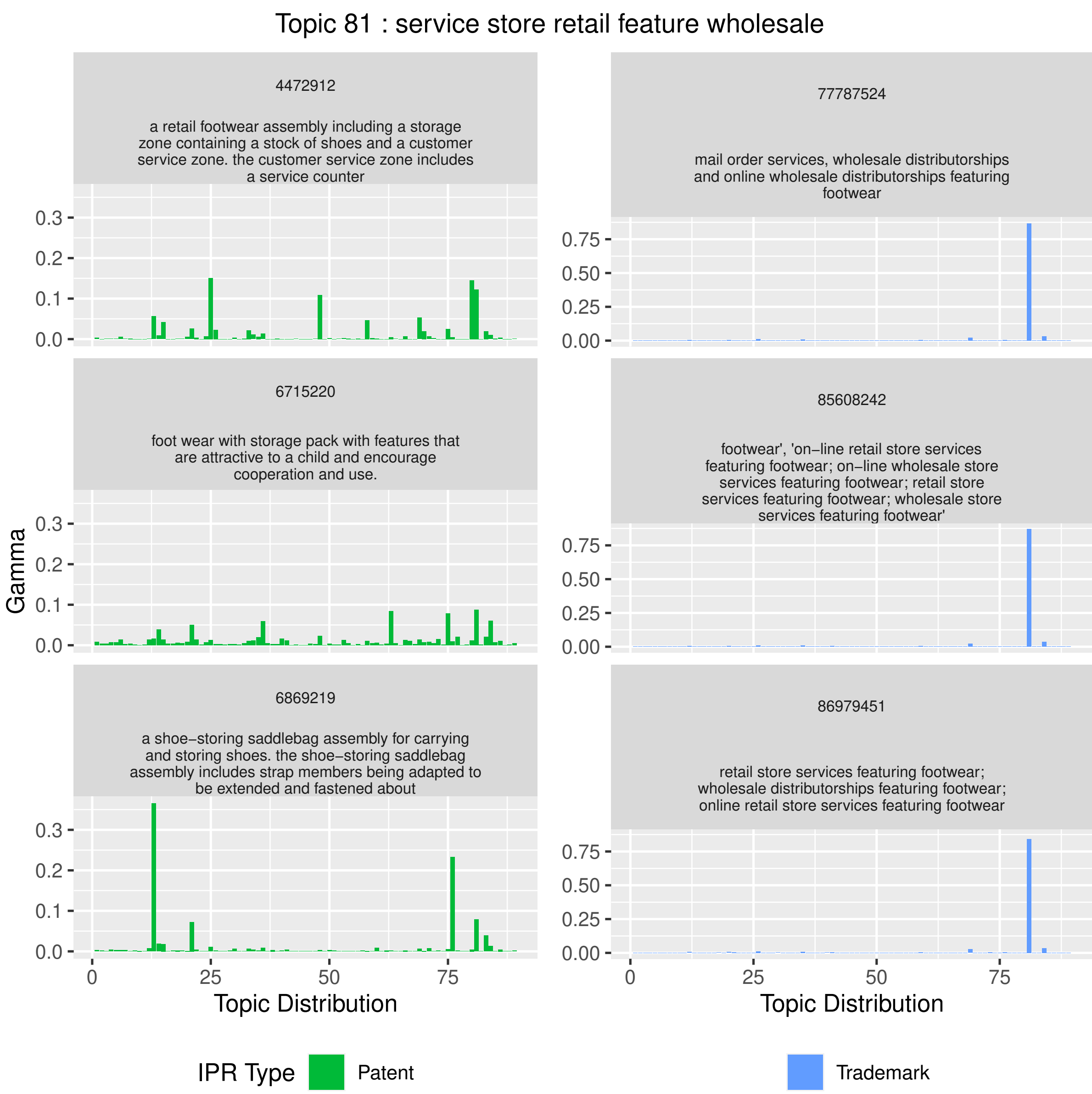


Figure C.24: Overview of Footwear Topic 81.

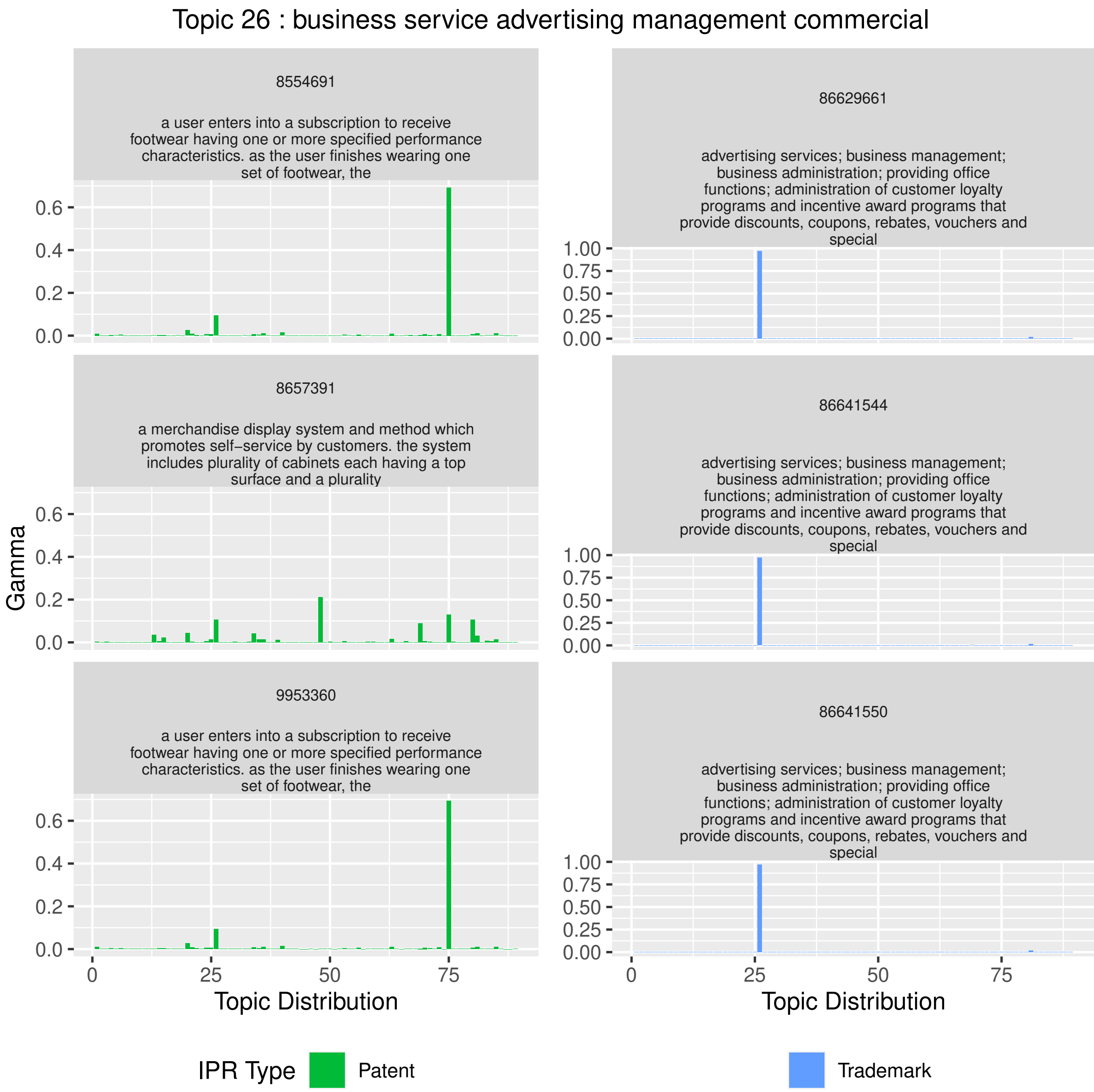


Figure C.25: Overview of Footwear Topic 26.

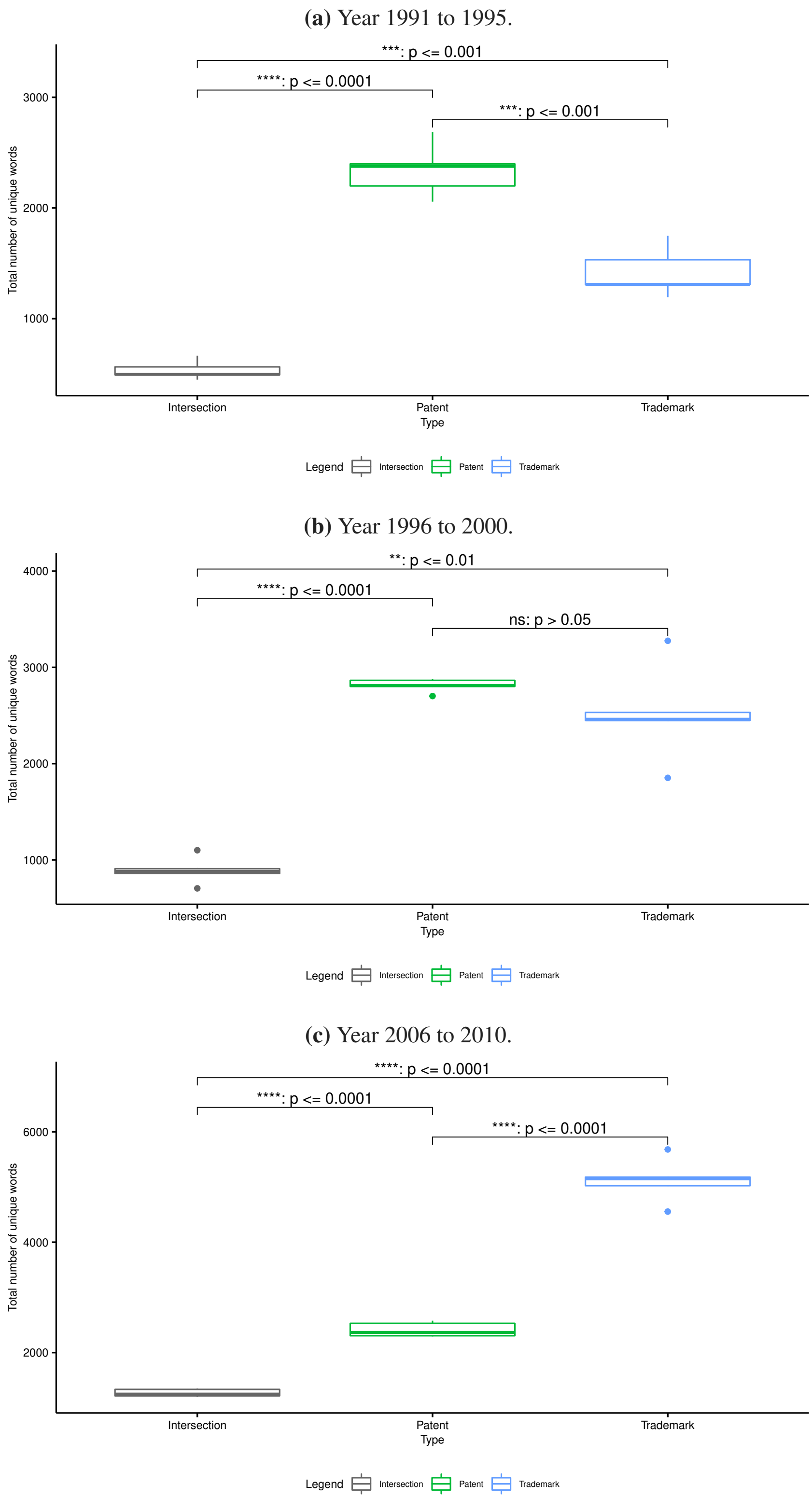


Figure C.26: Statistical t-Test of Unique Words per Document Type for different Five-Year Intervals in Footwear.

D Appendix to Chapter 5: Revealing Technological Transformation and Firm Involvement

D.1 Data Selection

D.1.1 Patents

```
--HORNBORSTEL-SACHS-CLASSIFICATION
select distinct cp.patent_id
into fs.music_uspto_pat_util_hsc
from uspto.cpc_current cp
join uspto.patent pa on cp.patent_id = pa.id
where (
cp.subgroup_id in
(-- Wind instruments (Aerophones)
-- Classical: fs.music_uspto_pat_hsc_wind_classic
'G10D7/00', 'G10D9/00', 'G10B1/00', 'G10B3/00', 'G10D11/00', 'G10D7/00',
    'G10K5/00', 'G10K9/00',
-- Automatic: fs.music_uspto_pat_hsc_wind_auto
'G10F1/12', 'G10F1/14',

-- String instruments (Chordophones)
-- Classical: fs.music_uspto_pat_hsc_string_classic
'G10D1/00', 'G10D3/00', 'G10C1/00', 'G10C3/00', 'G10C9/00', 'G10D1/00',
-- Automatic: fs.music_uspto_pat_hsc_string_auto
'G10F1/02', 'G10F1/04', 'G10F1/16', 'G10F1/18', 'G10F1/20',

-- Percussion instruments (Idiophones and Membranophones)
-- Classical: fs.music_uspto_pat_hsc_perc_classic
'G10K1/00', 'G10K3/00', 'G10D13/00',
-- Automatic: fs.music_uspto_pat_hsc_perc_auto
'G10F1/08', 'G10F1/10')

or cp.group_id in
(-- Automatic musical instruments
'G10F',
-- Electrophonic musical instruments
'G10H'))

and pa.type in ('utility')
```

Query D.1: Patents accoring to the Hornborstel-Sachs Classification

```
select distinct cp.patent_id
into fs.into fs.music_uspto_pat_util_cpc

from uspto.cpc_current cp
join uspto.patent pa on cp.patent_id = pa.id
where (cp.subgroup_id in (
'B25J11/004', -- Manipulators not otherwise provided for-Manipulators for
    entertainment-Playing a music instrument
'B62B2202/34', -- Indexing codes relating to type or characteristics of transported
    articles - Furniture - Music instruments, e.g. pianos
```

```
'H01H2231/018', -- Applications-Musical instrument
'Y10S224/91', -- Package and article carriers-Carrier for musical instrument
'Y10T29/49574') -- Metal working-Method of mechanical manufacture - Sound device making - Musical instrument or tuning fork making

or cp.group_id in (
'G10B', -- Organs, armoniums or similar wind musical instruments with associated blowing apparatus
'G10C', -- Pianos, harpischords, spinets or similar string musical instruments with one or more keyboards
'G10D', -- String musical instruments; wind musical instreumnts; accordions or concertinas; percussion musical instruments; aeolian harps; singing-flame musical instruments; musical instruments not otherwise provided for
'G10F', -- Automatic musical instruments
'G10G', -- Representation of music; recording music in notation form; accessories for music or musical instruments not otherwise provided for, e.g. supports
'G10H') --Electrophonic musical instruments; instruments in which the tones are generated by electromechanical means or electronic generators, or in which the tones are synthesised from a data store
)
and pa.type in ('utility')
```

Query D.2: Patents of Musical Instruments per CPC Classes

```
select distinct pa.id as 'patent_id'
into fs.music_uspto_pat_util_term
from uspto.patent pa
where ((abstract like '%music_instrument%' or abstract like '%musical_instrument%'
or title like '%music_instrument%' or title like '%musical_instrument%'))
and pa.type in ('utility')
```

Query D.3: Patents of Musical Instruments based on Terms

Wind Instruments (Aerophones)

```
select distinct pa.id as 'patent_id'
into fs.music_uspto_pat_util_term_wind
from uspto.patent pa
where (
(abstract like '%aerophone%' or
abstract like '%wind_instrument%' or
abstract like '%bagpipe%' or
abstract like '%bamboo_flute%'or
abstract like '%bandonion%' or
abstract like '%bandoneon%' or
abstract like '%barrel_organ%'or
abstract like '%buccin%'and (abstract like '%music%' or title like '%music%') or
abstract like '%clarionet%'or
abstract like '%clarion%'or
abstract like '%concertina%' and (abstract like '%music%' or title like '%music%') or
abstract like '%cornet%' and (abstract like '%music%' or title like '%music%') or
abstract like '%flute%' and (abstract like '%music%' or title like '%music%') or
abstract like '%harmonica%'and (abstract like '%music%' or title like '%music%') or
abstract like '%harmonium%' or
abstract like '%horn%' and (abstract like '%music%' or title like '%music%')or
abstract like '%melodica%'or
abstract like '%oboe%' and (abstract like '%music%' or title like '%music%')or
abstract like '%ocarina%'or
--organs [musical instruments] -- too many unrelated results
abstract like '%saxophone%'or
```

```
abstract like '%sheng%' and (abstract like '%music%' or title like '%music%') or --[Chinese musical wind instruments]
abstract like '%suona%' or --[Chinese trumpets]
abstract like '%trombone%'and (abstract like '%music%' or title like '%music%')  or
abstract like '%trumpet%' and (abstract like '%music%' or title like '%music%') )
or
(title like '%aerophone%' or
title like '%wind_instrument%' or
title like '%bagpipe%' or
title like '%bamboo_flute%'or
title like '%bandonion%' or
title like '%bandoneon%' or
title like '%barrel_organ%'or
title like '%buccin%'and (title like '%music%' or abstract like '%music%') or
title like '%clarionet%'or
title like '%clarion%'or
title like '%concertina%' and (title like '%music%' or abstract like '%music%') or
title like '%cornet%' and (title like '%music%' or abstract like '%music%') or
title like '%flute%' and (title like '%music%' or abstract like '%music%') or
title like '%harmonica%'and (title like '%music%' or abstract like '%music%') or
title like '%harmonium%' or
title like '%horn%' and (title like '%music%' or abstract like '%music%')or
title like '%melodica%'or
title like '%oboe%' and (title like '%music%' or abstract like '%music%')or
title like '%ocarina%'or
--organs [musical instruments] -- too many unrelated results
title like '%saxophone%'or
title like '%sheng%' and (title like '%music%' or abstract like '%music%') or --[Chinese musical wind instruments]
title like '%suona%' or --[Chinese trumpets]
title like '%trombone%'and (title like '%music%' or abstract like '%music%')  or
title like '%trumpet%' and (title like '%music%' or abstract like '%music%'))
)
and pa.type in ('utility')
```

Query D.4: Patents of Wind Instruments

String Instruments (Chordophones)

```
select distinct pa.id as 'patent_id'
into fs.music_uspto_pat_util_term_string
from uspto.patent pa
where (
(
abstract like '%chordophone%'or
abstract like '%string_instrument%' or
abstract like '%stringed_instrument%' or
abstract like '%balalaika%' and (title like '%music%' or abstract like '%music%')or
abstract like '%banjo%' or
abstract like '%bass%' and (title like '%music%' or abstract like '%music%')or
--double basses -- part of bass
abstract like '%guitar%'or
abstract like '%harp%'and (title like '%music%' or abstract like '%music%') or
abstract like '%huqin%' or -- [Chinese violins]
--Jews' harps [musical instruments] -- part of harp
abstract like '%lyre%' and (title like '%music%' or abstract like '%music%') or
abstract like '%yoke_lute%'or
abstract like '%mandolin%' and (title like '%music%' or abstract like '%music%') or
abstract like '%piano%'or
abstract like '%pipa%' and (title like '%music%' or abstract like '%music%') or --[Chinese guitars]
abstract like '%viola%'and (title like '%music%' or abstract like '%music%') or
```

```
abstract like '%violin%' or
abstract like '%zither%'
) or
(
title like '%chordophone%'or
title like '%string_instrument%' or
title like '%stringed_instrument%' or
title like '%balalaika%' and (title like '%music%' or abstract like '%music%')or
title like '%banjo%' or
title like '%bass%' and (title like '%music%' or abstract like '%music%')or
--double basses -- part of bass
title like '%guitar%'or
title like '%harp%'and (title like '%music%' or abstract like '%music%') or
title like '%huqin%' or -- [Chinese violins]
--Jews' harps [musical instruments] -- part of harp
title like '%lyre%' and (title like '%music%' or abstract like '%music%') or
title like '%yoke_lute%'or
title like '%mandolin%' and (title like '%music%' or abstract like '%music%') or
title like '%piano%'or
title like '%pipa%' and (title like '%music%' or abstract like '%music%') or --
    [Chinese guitars]
title like '%viola%'and (title like '%music%' or abstract like '%music%') or
title like '%violin%' or
title like '%zither%'
)
)
and pa.type in ('utility')
```

Query D.5: Patents of String Instruments

Percussion Instruments (Idiophones and Membranophones)

```
select distinct pa.id as 'patent_id'
into fs.music_uspto_pat_util_term_perc
from uspto.patent pa
where (
(
abstract like '%percussion%' and (abstract like '%music%' or abstract like
    '%instrument%' or title like '%music%' or title like '%instrument%') or
abstract like '%idiophone%' or
abstract like '%membranphone%' or
abstract like '%accordion%' and (title like '%music%' or abstract like '%music%') or
abstract like '%carillon%' and (title like '%music%' or abstract like '%music%') or
    -- [musical instruments]
abstract like '%castanet%' and (title like '%music%' or abstract like '%music%') or
abstract like '%cymbal%' and (title like '%music%' or abstract like '%music%') or
abstract like '%gong%' and (title like '%music%' or abstract like '%music%') or
abstract like '%handbell%' and (title like '%music%' or abstract like '%music%') or
    -- [musical instruments]
abstract like '%triangle%' and (title like '%music%' or abstract like '%music%') or
    -- [musical instruments]
abstract like '%xylophone%' or
--kettledrum' -- part of drum
abstract like '%timpani%' or
abstract like '%drum%' and (title like '%music%' or abstract like '%music%') or
--robotic drum -- part of drum
abstract like '%tambourine%' or
abstract like '%tom_tom%'
) or
(
title like '%percussion%' and (abstract like '%music%' or abstract like
    '%instrument%' or title like '%music%' or title like '%instrument%') or
```

```
title like '%idiophone%' or
title like '%membranphone%' or
title like '%accordion%' and (title like '%music%' or abstract like '%music%') or
title like '%carillon%' and (title like '%music%' or abstract like '%music%') or --
    [musical instruments]
title like '%castanet%' and (title like '%music%' or abstract like '%music%') or
title like '%cymbal%' and (title like '%music%' or abstract like '%music%') or
title like '%gong%' and (title like '%music%' or abstract like '%music%') or
title like '%handbell%' and (title like '%music%' or abstract like '%music%') or --
    [musical instruments]
title like '%triangle%' and (title like '%music%' or abstract like '%music%') or --
    [musical instruments]
title like '%xylophone%' or
--kettledrum' -- part of drum
title like '%timpani%' or
title like '%drum%' and (title like '%music%' or abstract like '%music%') or
--robotic drum -- part of drum
title like '%tambourine%' or
title like '%tom_tom%'
)
)
and pa.type in ('utility')
```

Query D.6: Patents of Percussion Instruments

Electronic or Classical Musical Instruments

```
select distinct te.patent_id
into fs.music_uspto_pat_util_string_elec
from fs.music_uspto_pat_util_string te
join uspto.patent pa on pa.id = te.patent_id
left join uspto.cpc_current cp on cp.patent_id = te.patent_id
where pa.abstract like '%electr%' or pa.title like '%electr%' or cp.group_id in
    ('G10H')
```

Query D.7: Patents of Electronic String Instruments

```
select *
into fs.music_uspto_pat_util_perc_classic
from fs.music_uspto_pat_util_perc mu
where mu.patent_id not in (select *
from fs.music_uspto_pat_util_perc_elec)
```

Query D.8: Patents of Classical Percussion Instruments

D.1.2 Designs

```
select distinct pa.id as patent_id
into fs.music_uspto_pat_design_cl
from uspto.patent pa
join uspto.uspc_current_april2021 pc on pc.patent_id = pa.id
join uspto.claim cl on cl.patent_id = pa.id
where type in ('design')
and
(mainclass_id in ('D17', '84') --D17 Musical instruments, 84 Music
or pa.title like '%music_instrument%'
or pa.title like '%musical_instrument%'
or cl.text like '%music_instrument%'
or cl.text like '%musical_instrument%')
```

Query D.9: Designs of Musical Instruments

Wind Instruments (Aerophones)

```
select distinct pa.id as 'patent_id'
into fs.music_uspto_pat_design_wind_term
from uspto.patent pa
join uspto.claim cl on cl.patent_id = pa.id
where type in ('design')
and (
(cl.text like '%aerophone%' or
cl.text like '%wind_instrument%' or
cl.text like '%bagpipe%' or
cl.text like '%bamboo_flute%'or
cl.text like '%bandonion%' or
cl.text like '%bandoneon%' or
cl.text like '%barrel_organ%'or
cl.text like '%buccin%'and (cl.text like '%music%' or title like '%music%') or
cl.text like '%clarionet%'or
cl.text like '%clarion%'or
cl.text like '%concertina%' and (cl.text like '%music%' or title like '%music%') or
cl.text like '%cornet%' and (cl.text like '%music%' or title like '%music%') or
cl.text like '%flute%' and (cl.text like '%music%' or title like '%music%') or
cl.text like '%harmonica%'and (cl.text like '%music%' or title like '%music%') or
cl.text like '%harmonium%' or
cl.text like '%horn%' and (cl.text like '%music%' or title like '%music%')or
cl.text like '%melodica%'or
cl.text like '%oboe%' and (cl.text like '%music%' or title like '%music%')or
cl.text like '%ocarina%'or
--organs [musical instruments] -- too many unrelated results
cl.text like '%saxophone%'or
cl.text like '%sheng%' and (cl.text like '%music%' or title like '%music%') or
    --[Chinese musical wind instruments]
cl.text like '%suona%' or --[Chinese trumpets]
cl.text like '%trombone%'and (cl.text like '%music%' or title like '%music%')  or
cl.text like '%trumpet%' and (cl.text like '%music%' or title like '%music%') )
or
(title like '%aerophone%' or
title like '%wind_instrument%' or
title like '%bagpipe%' or
title like '%bamboo_flute%'or
title like '%bandonion%' or
title like '%bandoneon%' or
title like '%barrel_organ%'or
title like '%buccin%'and (title like '%music%' or cl.text like '%music%') or
title like '%clarionet%'or
title like '%clarion%'or
title like '%concertina%' and (title like '%music%' or cl.text like '%music%') or
title like '%cornet%' and (title like '%music%' or cl.text like '%music%') or
title like '%flute%' and (title like '%music%' or cl.text like '%music%') or
title like '%harmonica%'and (title like '%music%' or cl.text like '%music%') or
title like '%harmonium%' or
title like '%horn%' and (title like '%music%' or cl.text like '%music%')or
title like '%melodica%'or
title like '%oboe%' and (title like '%music%' or cl.text like '%music%')or
title like '%ocarina%'or
--organs [musical instruments] -- too many unrelated results
title like '%saxophone%'or
title like '%sheng%' and (title like '%music%' or cl.text like '%music%') or
    --[Chinese musical wind instruments]
title like '%suona%' or --[Chinese trumpets]
title like '%trombone%'and (title like '%music%' or cl.text like '%music%')  or
title like '%trumpet%' and (title like '%music%' or cl.text like '%music%')
))
```

Query D.10: Designs of Wind Instruments

String Instruments (Chordophones)

```
select distinct pa.id as 'patent_id'
into fs.music_uspto_pat_design_string_term
from uspto.patent pa
join uspto.claim cl on cl.patent_id = pa.id
where type in ('design')
and ((
cl.text like '%chordophone%'or
cl.text like '%string_instrument%' or
cl.text like '%stringed_instrument%' or
cl.text like '%balalaika%' and (title like '%music%' or cl.text like '%music%')or
cl.text like '%banjo%' or
cl.text like '%bass%' and (title like '%music%' or cl.text like '%music%')or
--double basses -- part of bass
cl.text like '%guitar%'or
cl.text like '%harp%'and (title like '%music%' or cl.text like '%music%') or
cl.text like '%huqin%' or -- [Chinese violins]
--Jews' harps [musical instruments] -- part of harp
cl.text like '%lyre%' and (title like '%music%' or cl.text like '%music%') or
cl.text like '%yoke_lute%'or
cl.text like '%mandolin%' and (title like '%music%' or cl.text like '%music%') or
cl.text like '%piano%'or
cl.text like '%pipa%' and (title like '%music%' or cl.text like '%music%') or --
    [Chinese guitars]
cl.text like '%viola%'and (title like '%music%' or cl.text like '%music%') or
cl.text like '%violin%' or
cl.text like '%zither%'
) or
(
title like '%chordophone%'or
title like '%string_instrument%' or
title like '%stringed_instrument%' or
title like '%balalaika%' and (title like '%music%' or cl.text like '%music%')or
title like '%banjo%' or
title like '%bass%' and (title like '%music%' or cl.text like '%music%')or
--double basses -- part of bass
title like '%guitar%'or
title like '%harp%'and (title like '%music%' or cl.text like '%music%') or
title like '%huqin%' or -- [Chinese violins]
--Jews' harps [musical instruments] -- part of harp
title like '%lyre%' and (title like '%music%' or cl.text like '%music%') or
title like '%yoke_lute%'or
title like '%mandolin%' and (title like '%music%' or cl.text like '%music%') or
title like '%piano%'or
title like '%pipa%' and (title like '%music%' or cl.text like '%music%') or --
    [Chinese guitars]
title like '%viola%'and (title like '%music%' or cl.text like '%music%') or
title like '%violin%' or
title like '%zither%'
))
```

Query D.11: Designs of String Instruments

Percussion Instruments (Idiophones and Membranophones)

```
select distinct pa.id as 'patent_id'
into fs.music_uspto_pat_design_perc_term
from uspto.patent pa
--join uspto.uspc_current_april2021 pc on pc.patent_id = pa.id
join uspto.claim cl on cl.patent_id = pa.id
where type in ('design')
```

```
and (
(
cl.text like '%percussion%' and (cl.text like '%music%' or cl.text like '%instrument%' or title like '%music%' or title like '%instrument%') or
cl.text like '%idiophone%' or
cl.text like '%membranphone%' or
cl.text like '%accordion%' and (title like '%music%' or cl.text like '%music%') or
cl.text like '%carillon%' and (title like '%music%' or cl.text like '%music%') or -- [musical instruments]
cl.text like '%castanet%' and (title like '%music%' or cl.text like '%music%') or
cl.text like '%cymbal%' and (title like '%music%' or cl.text like '%music%') or
cl.text like '%gong%' and (title like '%music%' or cl.text like '%music%') or
cl.text like '%handbell%' and (title like '%music%' or cl.text like '%music%') or -- [musical instruments]
cl.text like '%triangle%' and (title like '%music%' or cl.text like '%music%') or -- [musical instruments]
cl.text like '%xylophone%' or
--kettledrum' -- part of drum
cl.text like '%timpani%' or
cl.text like '%drum%' and (title like '%music%' or cl.text like '%music%') or
--robotic drum -- part of drum
cl.text like '%tambourine%' or
cl.text like '%tom_tom%'
)
or
(
title like '%percussion%' and (cl.text like '%music%' or cl.text like '%instrument%' or title like '%music%' or title like '%instrument%') or
title like '%idiophone%' or
title like '%membranphone%' or
title like '%accordion%' and (title like '%music%' or cl.text like '%music%') or
title like '%carillon%' and (title like '%music%' or cl.text like '%music%') or -- [musical instruments]
title like '%castanet%' and (title like '%music%' or cl.text like '%music%') or
title like '%cymbal%' and (title like '%music%' or cl.text like '%music%') or
title like '%gong%' and (title like '%music%' or cl.text like '%music%') or
title like '%handbell%' and (title like '%music%' or cl.text like '%music%') or -- [musical instruments]
title like '%triangle%' and (title like '%music%' or cl.text like '%music%') or -- [musical instruments]
title like '%xylophone%' or
--kettledrum' -- part of drum
title like '%timpani%' or
title like '%drum%' and (title like '%music%' or cl.text like '%music%') or
--robotic drum -- part of drum
title like '%tambourine%' or
title like '%tom_tom%'
)
)
```

Query D.12: Designs of Percussion Instruments

All

```
select distinct *
into music.uspto_pat_design
from fs.music_uspto_pat_design_perc_term
UNION
select *
from fs.music_uspto_pat_design_cl
UNION
select *
from fs.music_uspto_pat_design_string_term
UNION
select *
from fs.music_uspto_pat_design_wind_term
```

Query D.13: All Designs of Musical Instruments

Electronic or Classical Musical Instruments

```
select distinct te.patent_id
into music_uspto_pat_design_elec
from fs.music_uspto_pat_design te
join uspto.patent pa on pa.id = te.patent_id
join uspto.claim cl on cl.patent_id = te.patent_id
where cl.text like '%electr%' or pa.title like '%electr%'
```

Query D.14: Designs of Electronic Musical Instruments

```
select *
into fs.music_uspto_pat_design_classic
from fs.music_uspto_pat_design mu
where mu.patent_id not in (select *
from fs.music_uspto_pat_design_elec)
```

Query D.15: Designs of Classical Musical Instruments

D.1.3 Trademarks

```
select *
into fs.music_uspto_tm_cl_statement

-- NICE class 15
from trademarks.statement
where serial_no in (select cf.serial_no
from trademarks.case_file cf
where Convert(date, cf.filing_dt, 120) >= '1973-09-01') and
left(statement_type_cd, 5) in ('GS015')

Union

-- Trademark name
select *
from trademarks.[statement]
where
serial_no in (
select serial_no
from trademarks.case_file
where
(mark_id_char like '%music%' and mark_id_char like '%instrument%')
)
```

```
and left(statement_type_cd, 2) in ('GS')

Union

-- Pseudomarks
select *
from trademarks.statement
where
serial_no in (
select serial_no
from trademarks.[statement]
where
(statement_text like '%music%' and statement_text like '%instrument%')
and left(statement_type_cd, 2) in ('PM')
)
and left(statement_type_cd, 2) in ('GS')

union

-- Trademark description
select *
from trademarks.statement
where (statement_text like '%music_instrument%' or statement_text like
    '%musical_instrument%')
and left(statement_type_cd, 2) in ('GS')
```

Query D.16: Trademarks of Musical Instruments

Wind Instruments (Aerophones)

```
select *
into fs.music_trademarks_wind_statement

-- Trademark name
from trademarks.[statement]
where
serial_no in (
select serial_no
from trademarks.case_file
where
mark_id_char like '%wind_instrument%'
)
and left(statement_type_cd, 2) in ('GS')

Union

-- Pseudomarks
select *
from trademarks.statement
where
serial_no in (
select serial_no
from trademarks.[statement]
where
statement_text like '%wind_instrument%'
and left(statement_type_cd, 2) in ('PM')
)
and left(statement_type_cd, 2) in ('GS')

union

-- Trademark description
```

```sql
select *
from trademarks.statement
where (
statement_text like '%aerophone%' or
statement_text like '%wind_instrument%' or
statement_text like '%bagpipe%' or
statement_text like '%bamboo_flute%'or
statement_text like '%bandonion%' or
statement_text like '%bandoneon%' or
statement_text like '%barrel_organ%'or
statement_text like '%buccin%'and statement_text like '%music%'  or
statement_text like '%clarionet%'or
statement_text like '%clarion%'or
statement_text like '%concertina%' and statement_text like '%music%'  or
statement_text like '%cornet%' and statement_text like '%music%'  or
statement_text like '%flute%' and statement_text like '%music%'  or
statement_text like '%harmonica%'and statement_text like '%music%'  or
statement_text like '%harmonium%' or
statement_text like '%horn%' and statement_text like '%music%' or
statement_text like '%melodica%'or
statement_text like '%oboe%' and statement_text like '%music%' or
statement_text like '%ocarina%'or
--organs [musical instruments] -- too many unrelated results
statement_text like '%saxophone%'or
statement_text like '%sheng%' and statement_text like '%music%'  or --[Chinese musical wind instruments]
statement_text like '%suona%' or --[Chinese trumpets]
statement_text like '%trombone%'and statement_text like '%music%'   or
statement_text like '%trumpet%' and statement_text like '%music%'
)
and left(statement_type_cd, 2) in ('GS')
```

Query D.17: Trademarks of Wind Instruments

```sql
select distinct mu.statement_type_cd, mu.statement_text, mu.serial_no
into fs.music_uspto_tm_wind_statement_reg
from fs.music_uspto_tm_wind_statement mu
join trademarks.case_file cf on mu.serial_no = cf.serial_no
where cf.registration_no not in ('0000000')
```

Query D.18: Only Registered Trademarks of Musical Instruments

String Instruments (Chordophones)

```sql
select *
into fs.music_uspto_tm_string_statement

-- Trademark name
from trademarks.[statement]
where
serial_no in (
select serial_no
from trademarks.case_file
where
mark_id_char like '%string_instrument%' or
mark_id_char like '%stringed_instrument%'
)
and left(statement_type_cd, 2) in ('GS')


union

```

```
-- Pseudomarks
select *
from trademarks.statement
where
serial_no in (
select serial_no
from trademarks.[statement]
where
(statement_text like '%string_instrument%' or
statement_text like '%stringed_instrument%')
and left(statement_type_cd, 2) in ('PM')
)
and left(statement_type_cd, 2) in ('GS')

union

-- String/ed instruments (Chordophones)
select *
from trademarks.statement
where (
statement_text like '%chordophone%'or
statement_text like '%string_instrument%' or
statement_text like '%stringed_instrument%' or
statement_text like '%balalaika%' and statement_text like '%music%'or
statement_text like '%banjo%' or
statement_text like '%bass%' and statement_text like '%music%'or
--double basses -- part of bass
statement_text like '%guitar%'or
statement_text like '%harp%'and statement_text like '%music%' or
statement_text like '%huqin%' or -- [Chinese violins]
--Jews' harps [musical instruments] -- part of harp
statement_text like '%lyre%' and statement_text like '%music%' or
statement_text like '%yoke_lute%'or
statement_text like '%mandolin%' and statement_text like '%music%' or
statement_text like '%piano%'or
statement_text like '%pipa%' and statement_text like '%music%' or -- [Chinese
    guitars]
statement_text like '%viola%'and statement_text like '%music%' or
statement_text like '%violin%' or
statement_text like '%zither%'
)
and left(statement_type_cd, 2) in ('GS')
```

Query D.19: Trademarks of String Instruments

Percussion Instruments (Idiophones or Membranophones)

```
select *
into fs.music_uspto_tm_perc_statement

-- Trademark name
from trademarks.[statement]
where
serial_no in (
select serial_no
from trademarks.case_file
where
mark_id_char like '%percussion_instrument%'
)
and left(statement_type_cd, 2) in ('GS')

union
```

```

-- Pseudomarks
select *
from trademarks.statement
where
serial_no in (
select serial_no
from trademarks.[statement]
where (
statement_text like '%percussion_instrument%'
and left(statement_type_cd, 2) in ('PM')
))
and left(statement_type_cd, 2) in ('GS')

union

-- Percussion instruments (Idiophones and Membranophones)
select *
from trademarks.statement
where (
statement_text like '%percussion%' and (statement_text like '%music%' or statement_text like '%instrument%') or
statement_text like '%idiophone%' or
statement_text like '%membranphone%' or
statement_text like '%accordion%' and statement_text like '%music%' or
statement_text like '%carillon%' and statement_text like '%music%' or -- [musical instruments]
statement_text like '%castanet%' and statement_text like '%music%' or
statement_text like '%cymbal%' and statement_text like '%music%' or
statement_text like '%gong%' and statement_text like '%music%' or
statement_text like '%handbell%' and statement_text like '%music%' or -- [musical instruments]
statement_text like '%triangle%' and statement_text like '%music%' or -- [musical instruments]
statement_text like '%xylophone%' or
--kettledrum' -- part of drum
statement_text like '%timpani%' or
statement_text like '%drum%' and statement_text like '%music%' or
--robotic drum -- part of drum
statement_text like '%tambourine%' or
statement_text like '%tom_tom%'
)
and left(statement_type_cd, 2) in ('GS')
```

Query D.20: Trademarks of Percussion Instruments

Electronic or Classical Musical Instruments

```
select distinct *
from fs.music_uspto_tm_statement_reg te
where te.statement_text like '%electr%'
```

Query D.21: Trademarks of Electronic Musical Instruments

```
select distinct mu.statement_type_cd, mu.statement_text, mu.serial_no
into fs.music_uspto_tm_wind_statement_reg_elec
from fs.music_uspto_tm_wind_statement_reg mu
join trademarks.case_file cf on mu.serial_no = cf.serial_no
where mu.statement_text like '%electr%'
```

Query D.22: Trademarks of Electronic Wind Instruments

```
select *
into fs.music_uspto_tm_perc_reg_classic
from fs.music_uspto_tm_perc_reg mu
where mu.serial_no not in (select *
from fs.music_uspto_tm_perc_reg_elec)
```

Query D.23: Trademarks of Classical Percussion Instruments

D.2 Sectoral Perspective

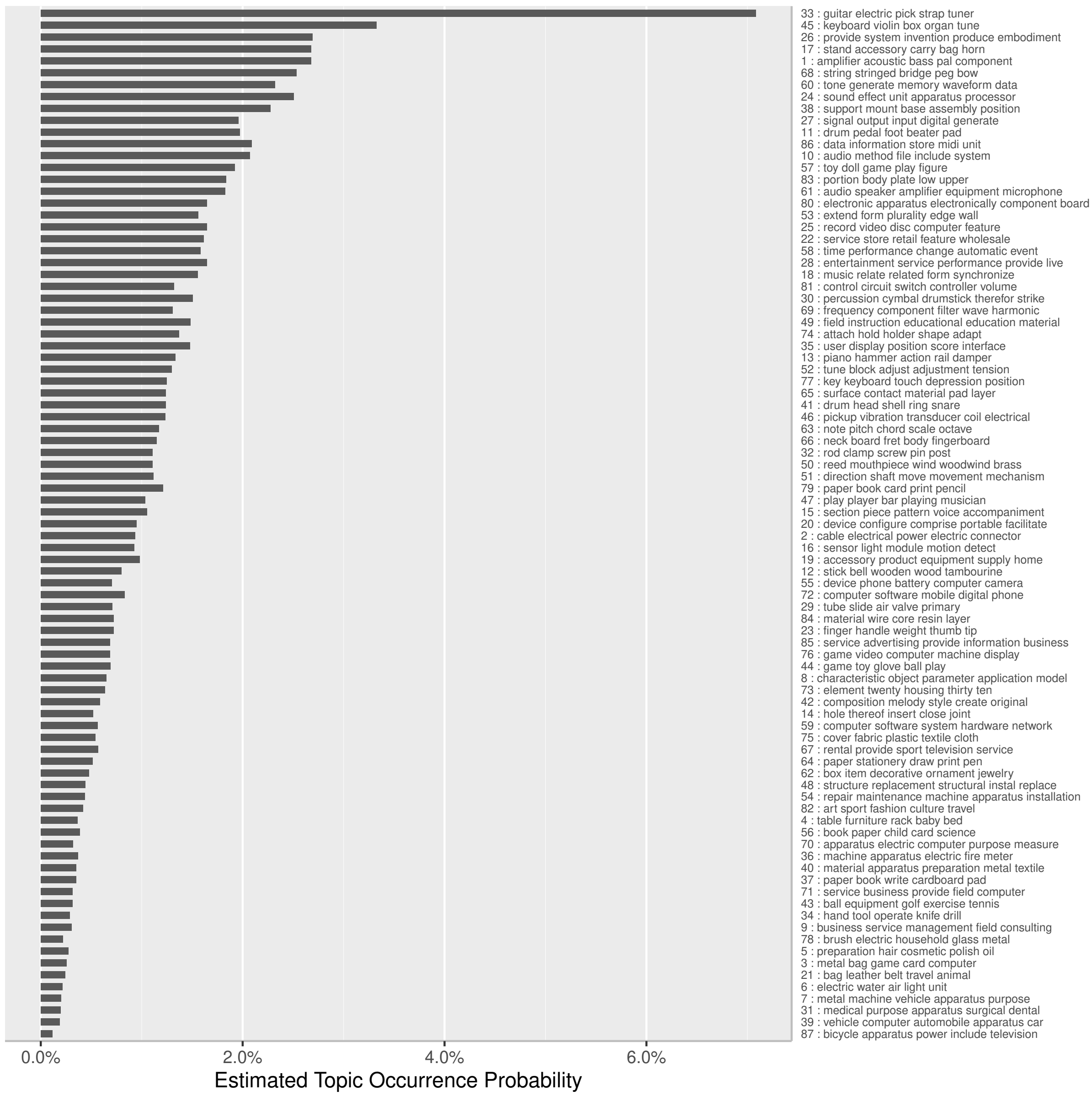
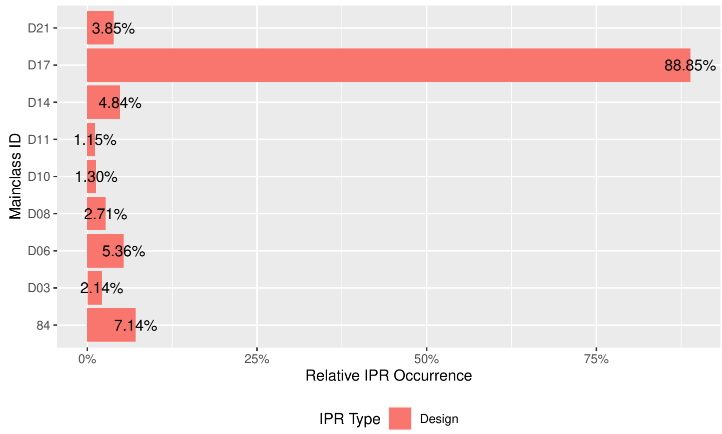


Figure D.1: Overview of the Model with 87 Topics in Musical Instruments.
The figure provides an overview of the 87 topics in Musical Instruments. The topics are displayed according to their relative share in the overall data set and ordered decreasingly.

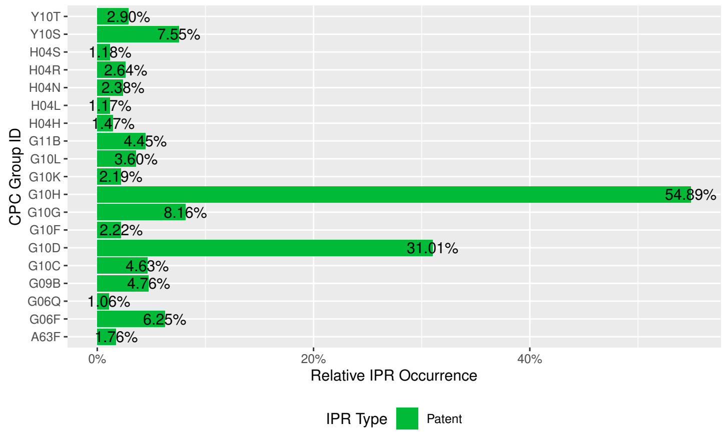
D.2.1 Intellectual Property Rights

Here, additional results differentiated for patents, designs and trademarks are provided.

(a) U.S. Design Classification



(b) CPC Classification for Patents



(c) Nice Classification for Trademarks

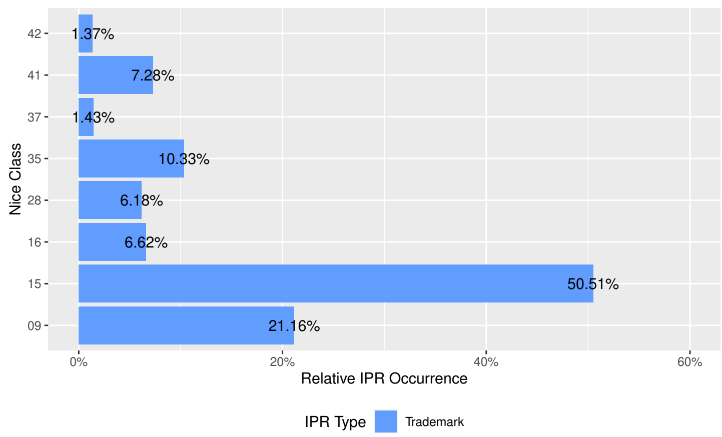


Figure D.2: Musical Instrument Patents, Designs and Trademarks in existing Classification Schemes. The Subfigures provide an overview of the most important classes in the different IPR types in the data set of musical instruments from 1986 to 2015.

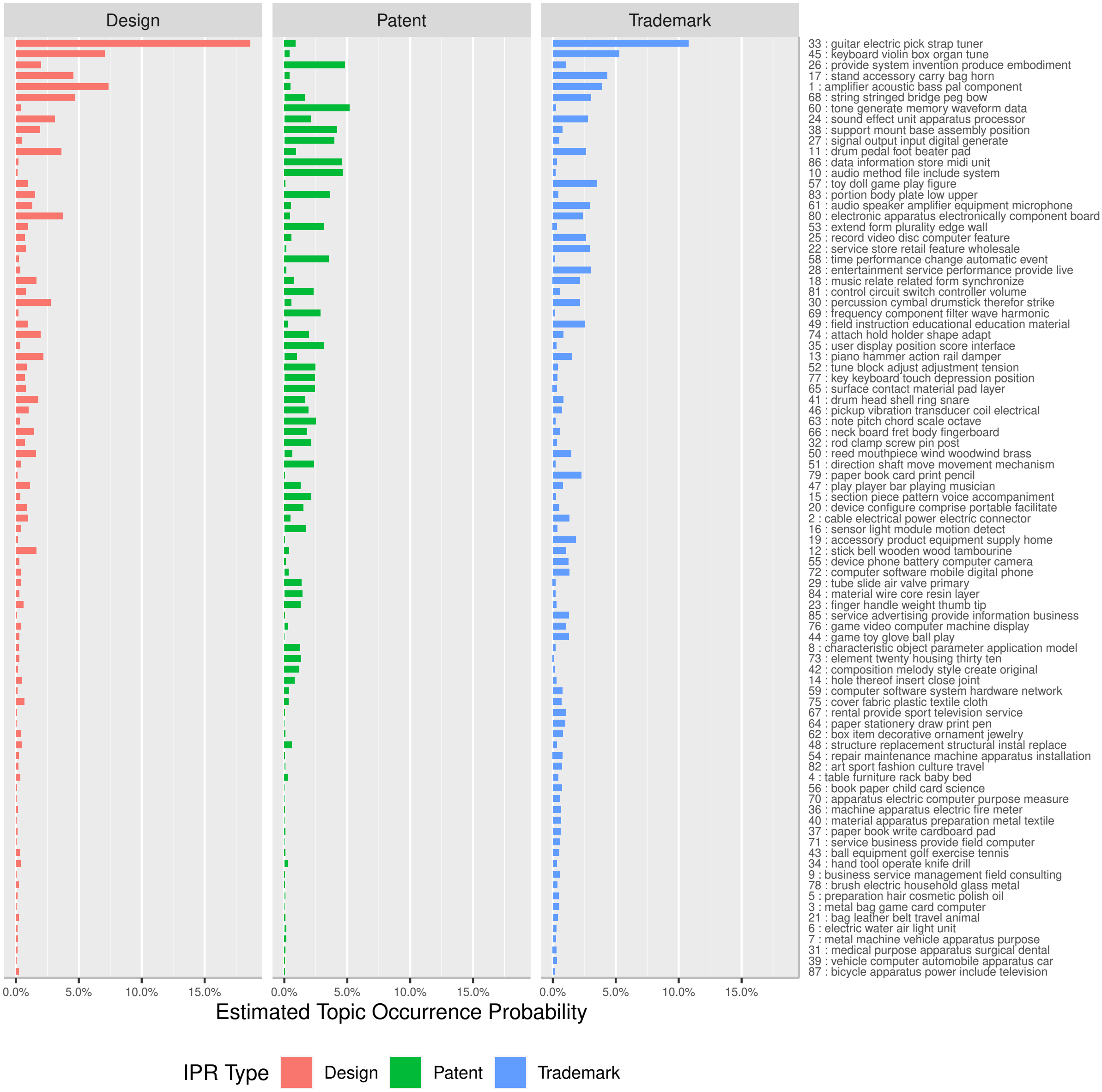


Figure D.3: Overview of the Model with 87 Topics in Musical Instruments, Differentiated by Document Types.
The figure provides an overview of the 87 topics in Musical Instruments. The topics are displayed according to their relative share in the overall data set and ordered decreasingly.

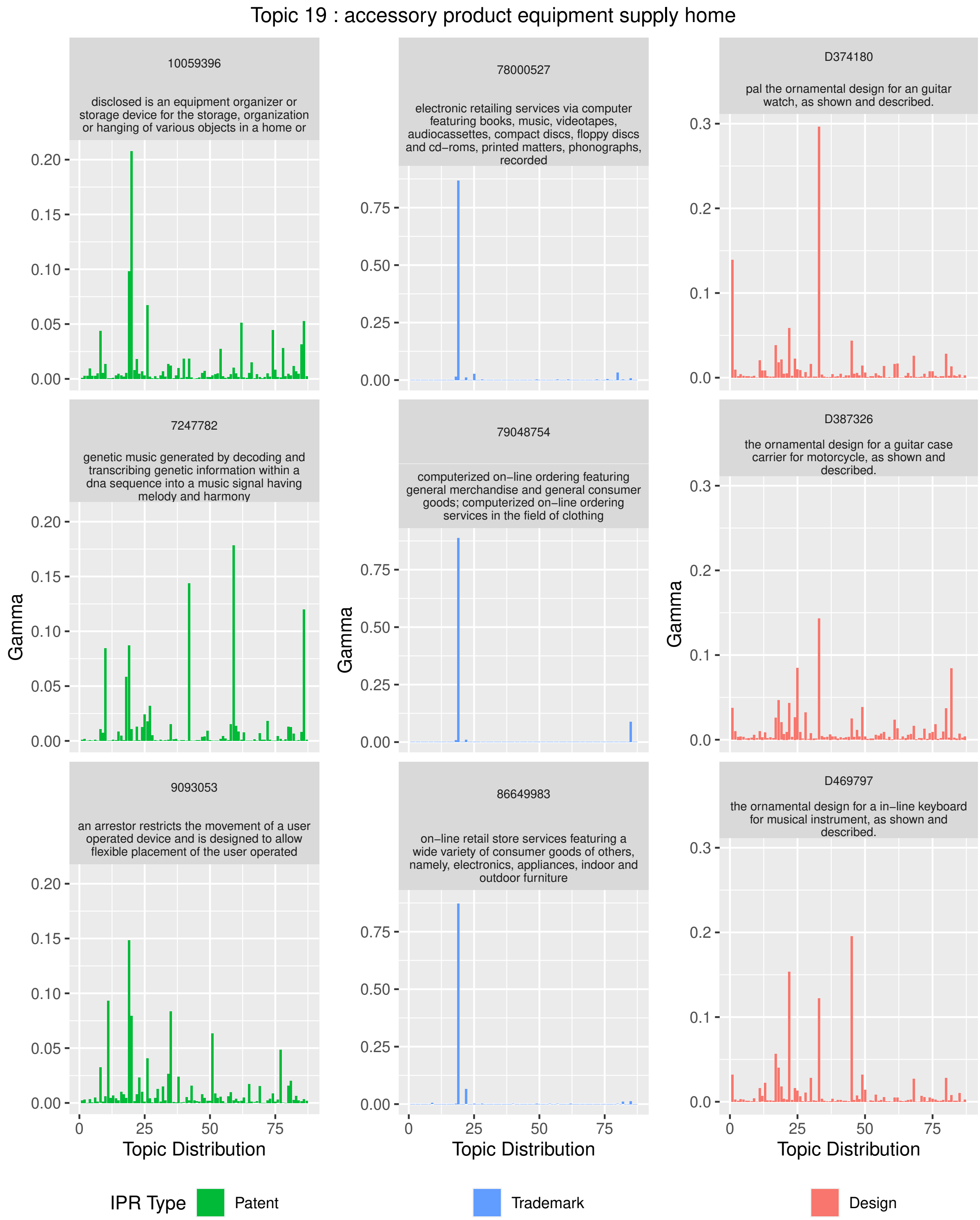


Figure D.4: Main Documents in Musical Instruments Topic 19.

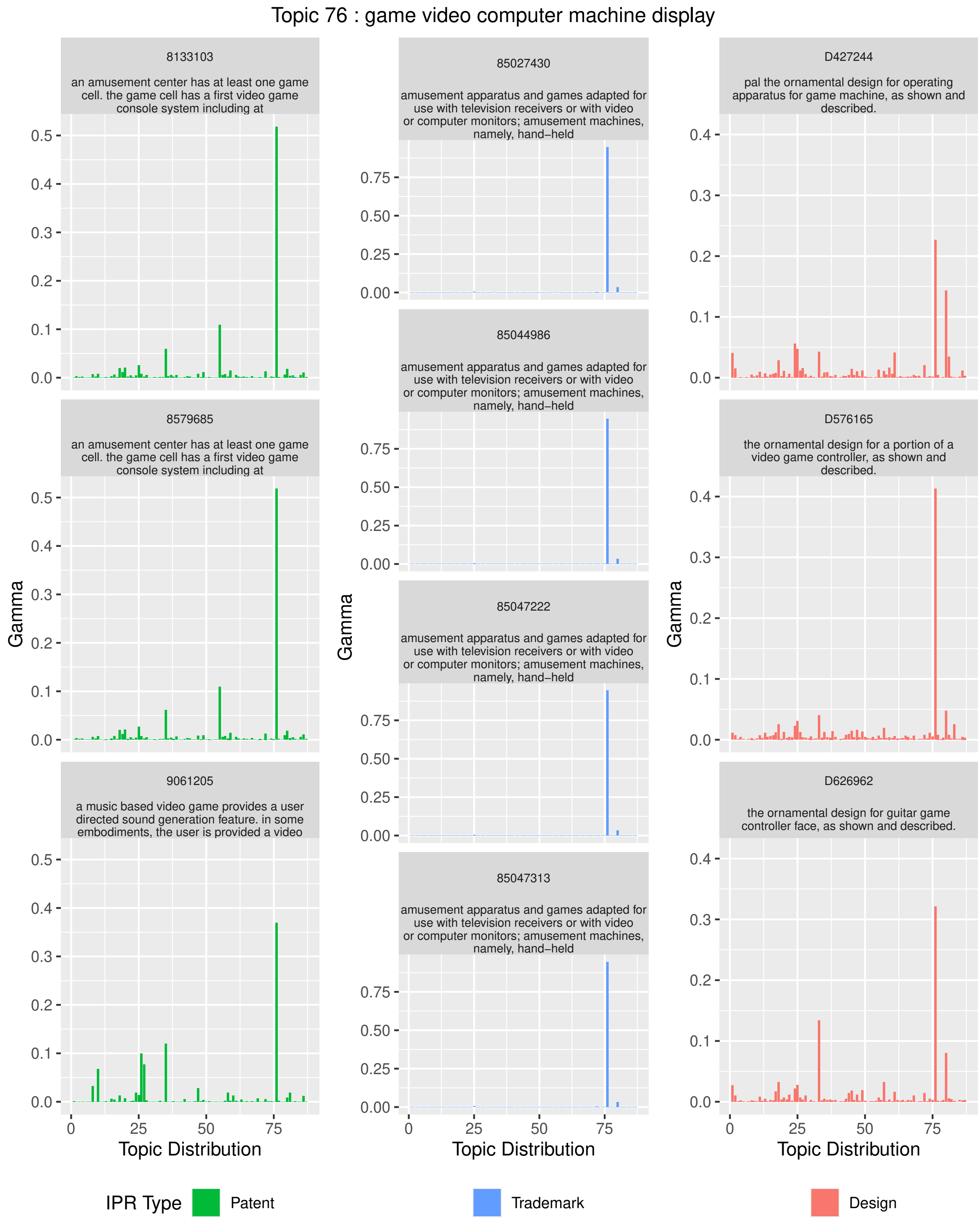


Figure D.5: Main Documents in Musical Instruments Topic 76.

D.2.2 Hornborstel-Sachs Classification

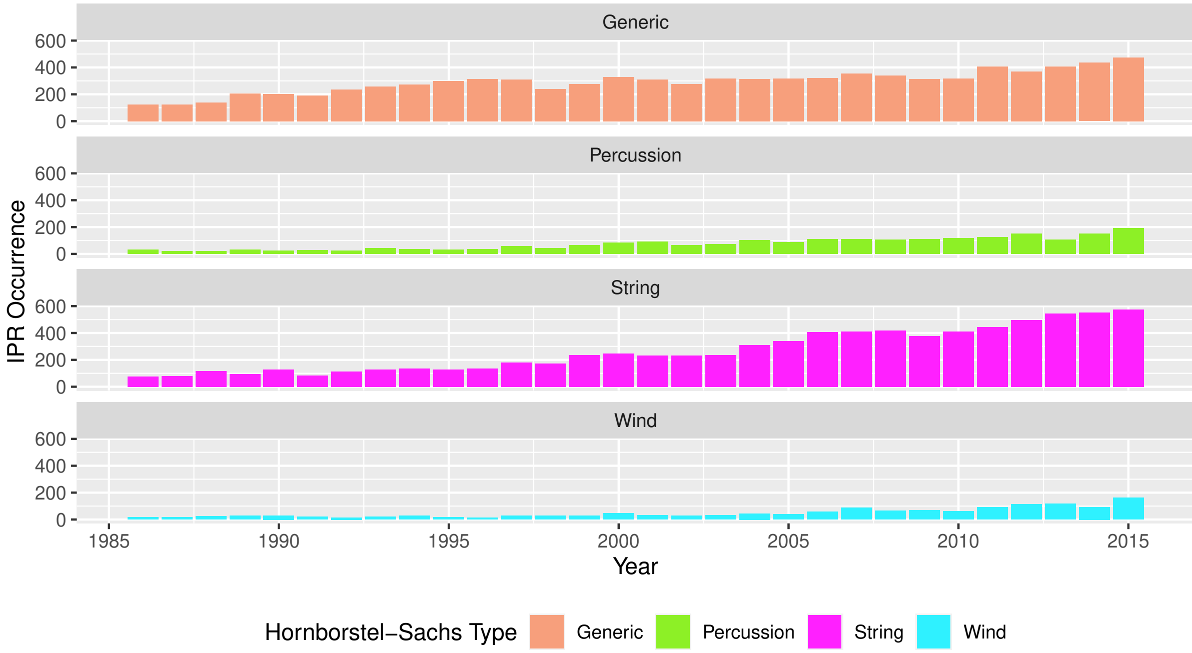


Figure D.6: Musical Instrument IPR Registration per Year and Hornborstel-Sachs Classes. The figures display the yearly development of the document registrations according to the Hornborstel-Sachs Classification.

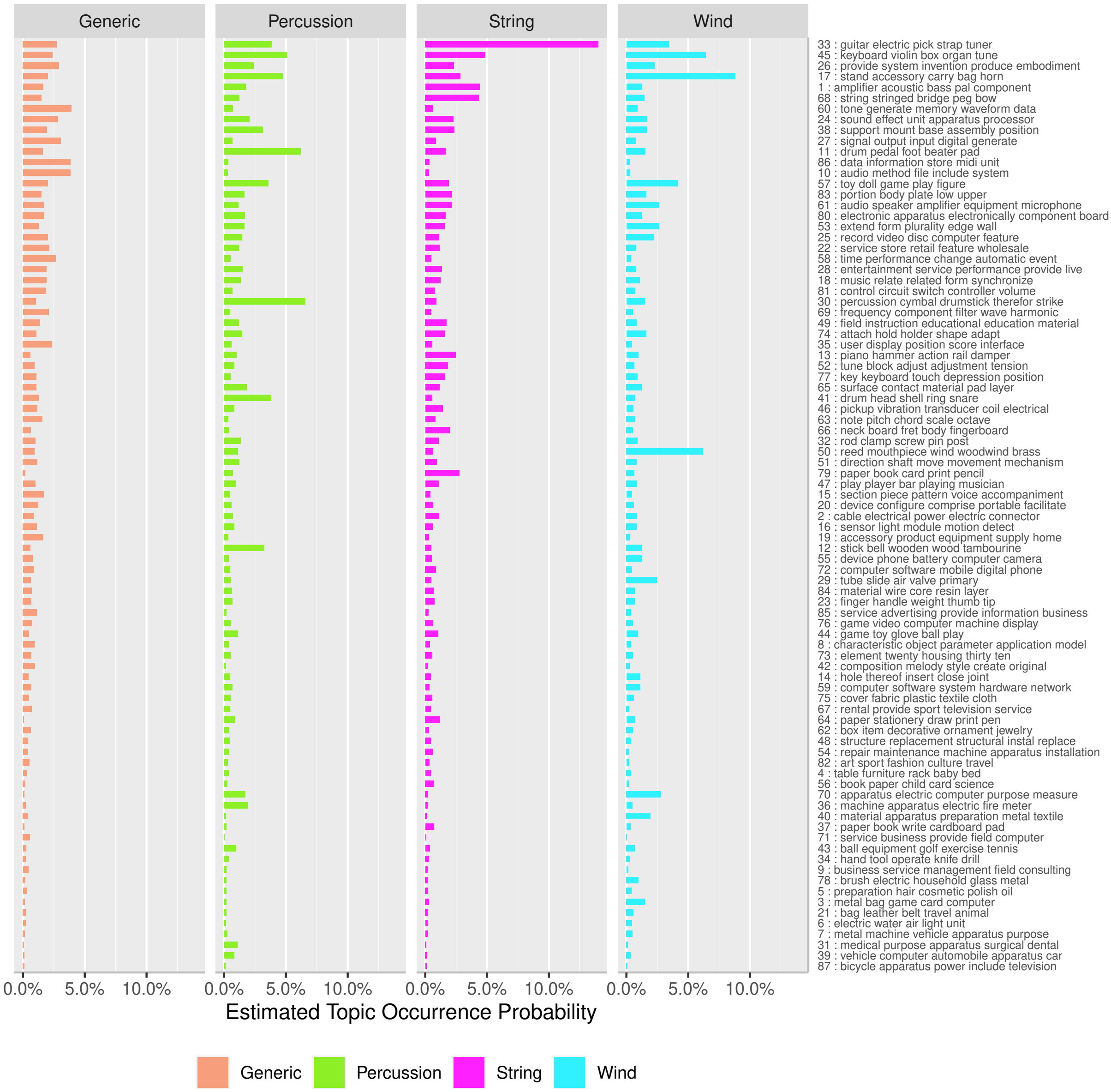


Figure D.7: Overview of the Model with 87 Topics in Musical Instruments according to the Hornborstel-Sachs Classification.
The figure provides an overview of the 87 topics in Musical Instruments. The topics are displayed according to their relative share in the overall data set and ordered decreasingly.

D.2.3 Electronic and Classical Musical Instruments

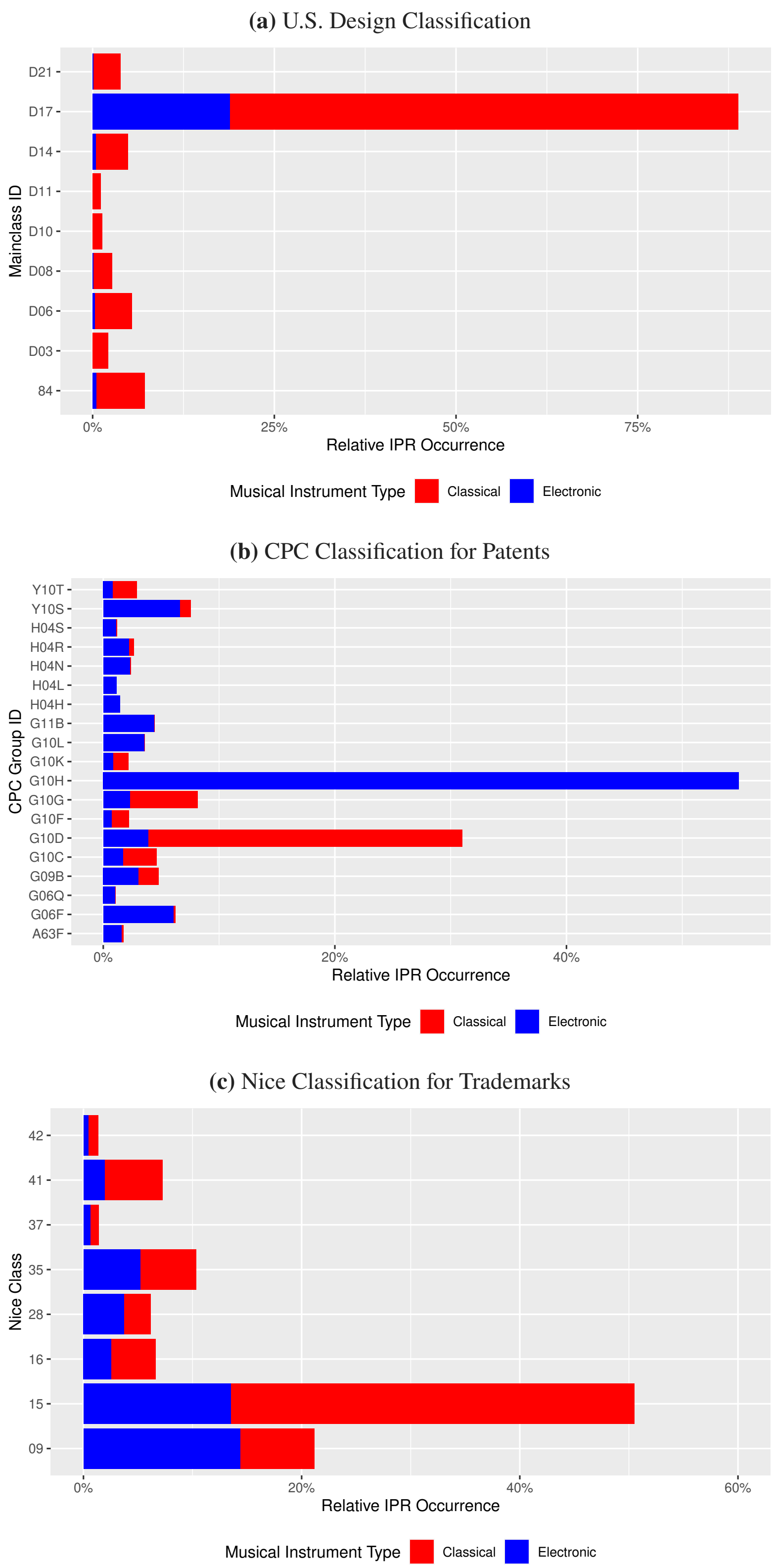


Figure D.8: Classical and Electronic Musical Instrument in existing Classification schemes. The figure offers an overview of the most significant classes in the various document types within the data set of musical instruments.

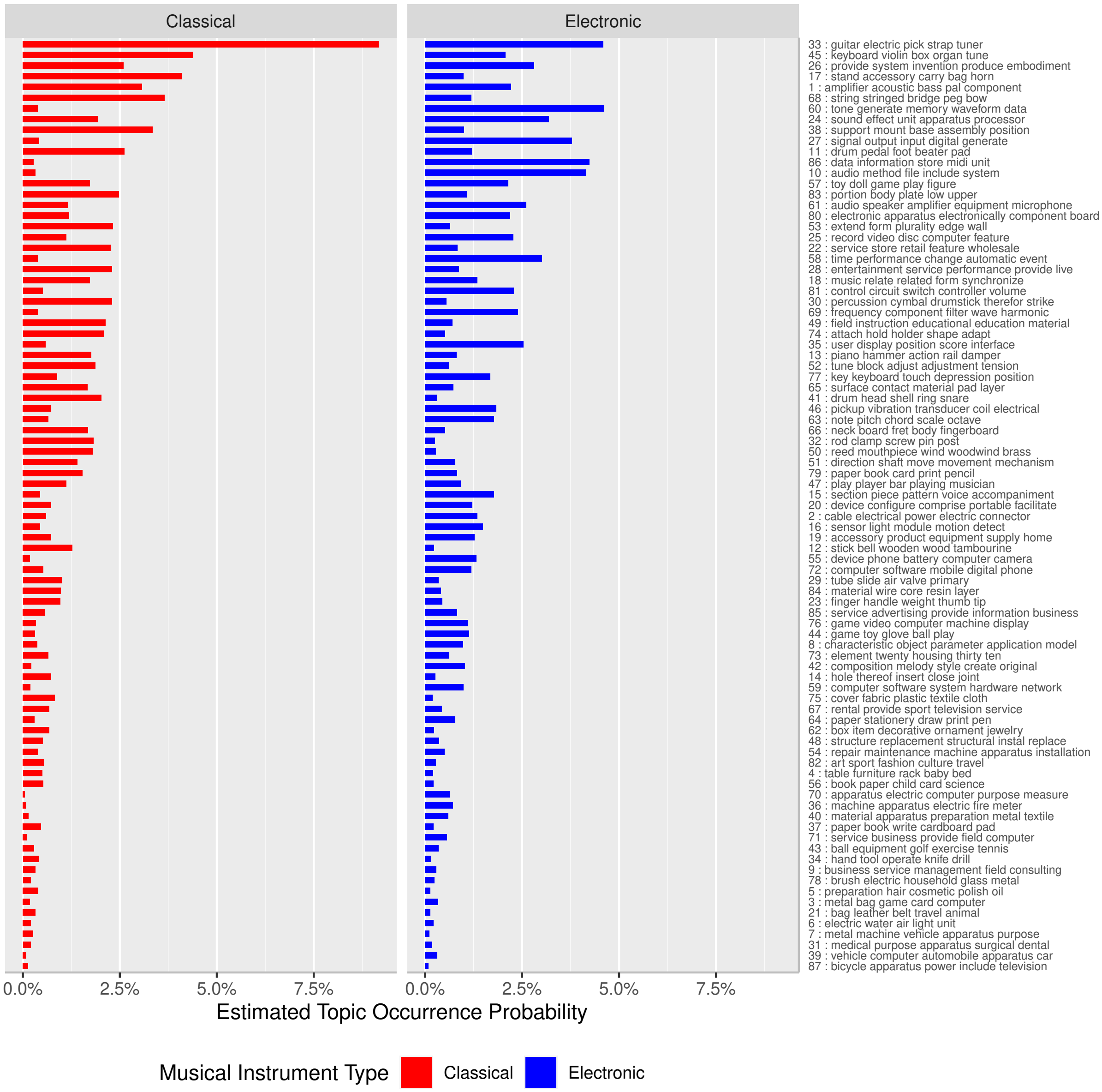


Figure D.9: Overview of the Model with 87 Topics in Musical Instruments, Differentiated for Electronic and Classical Instruments.
The figure presents an overview of the 87 topics in Musical Instruments. The topics are arranged based on their relative share in the overall dataset and are ordered in descending order.

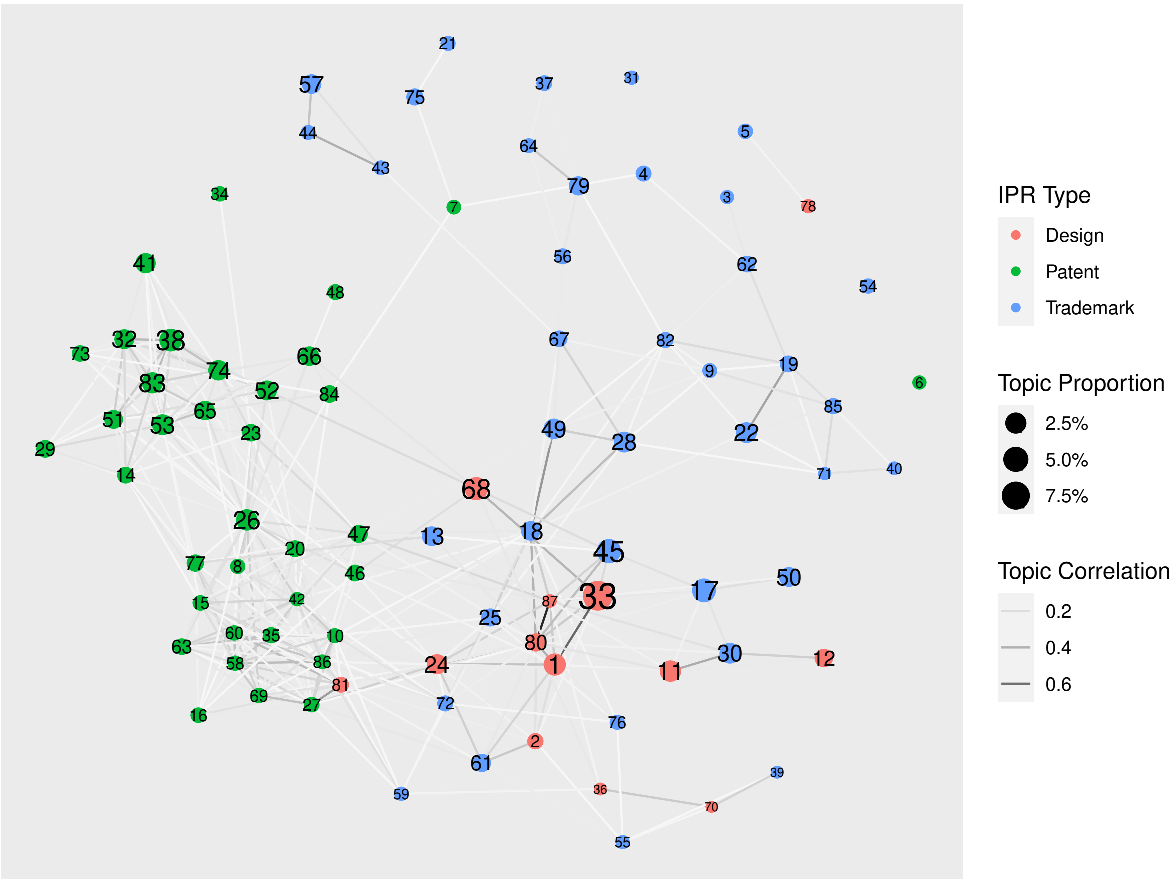


Figure D.10: Network of Classical Musical Instruments.
The network illustrates the topic relationships in classical musical instruments from 1985 to 2015. Nodes represent the topics, while links indicate the likelihood of these topics co-occurring. The node size corresponds to the topic's proportion. The nodes are coloured based on the document type with the highest mean topic occurrence.

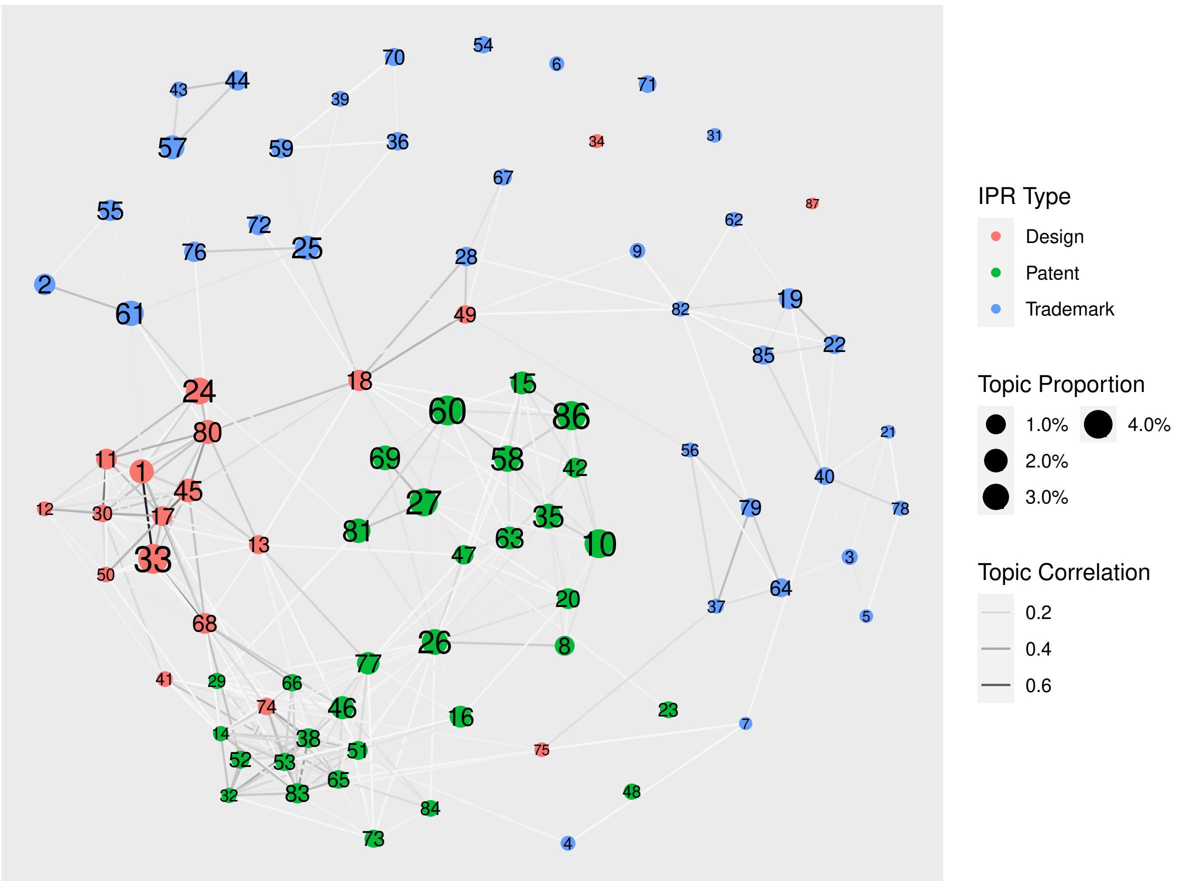


Figure D.11: Network of Electronic Musical Instruments.
The network illustrates the topic relationships in electronic musical instruments from 1985 to 2015. Nodes represent the topics, while links indicate the likelihood of these topics co-occurring. The node size corresponds to the topic's proportion. The nodes are coloured based on the document type with the highest mean topic occurrence.

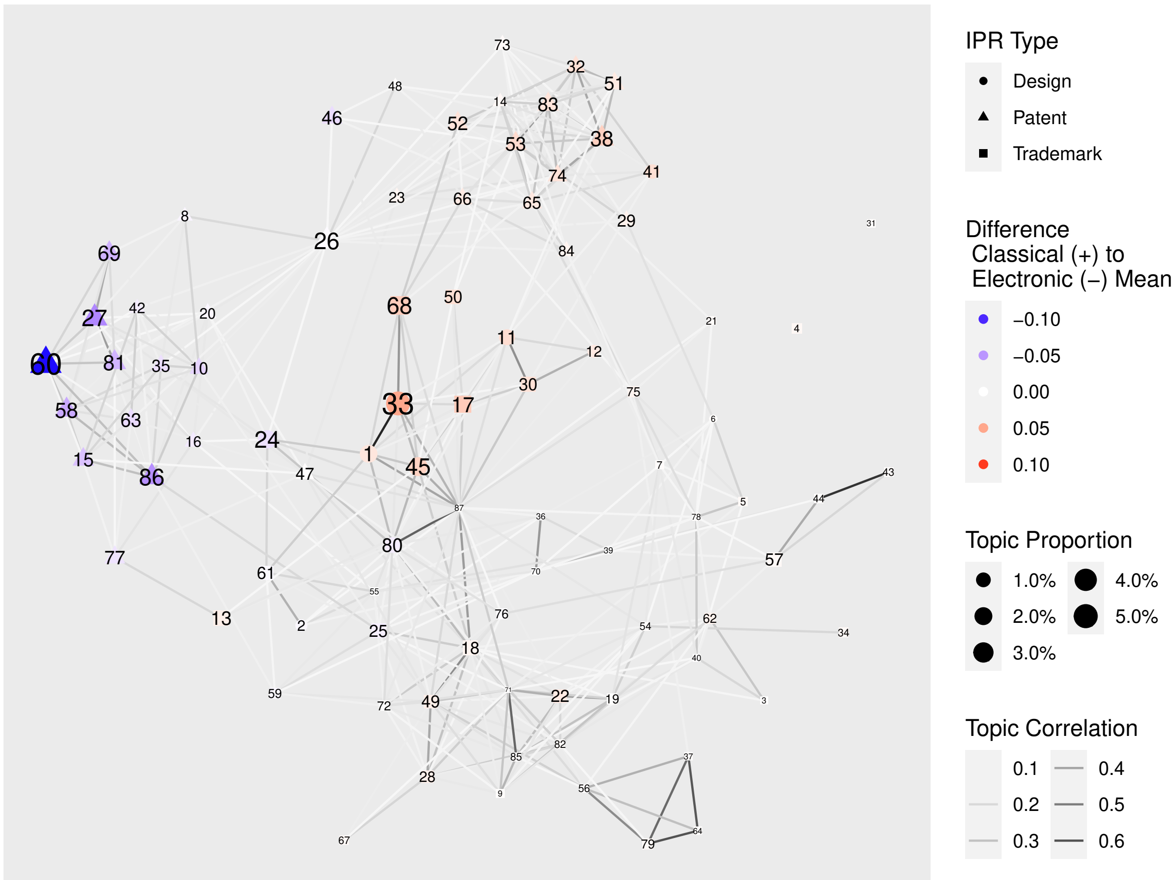


Figure D.12: Network of Musical Instrument Topics, Differentiated for Electronic and Classical Instruments from 1991 to 1995.
The network illustrates the topic relationships from 1991 to 1995, categorized by IPR types and electronic or classical musical instruments. Nodes represent topics, while links represent the likelihood of these topics to co-occur. The node size corresponds to the topic's proportion. IPR types are indicated by node shapes, while node colours represent the musical instrument type. Intensive blue signifies strong relevance to electronic musical instruments, and red represents a focus on classical musical instruments. In this network, there are 58 nodes related to classical topics and 29 related to electronic topics. The edges display the strength of correlation, with lighter colours indicating lower correlation and darker colours indicating stronger correlation.

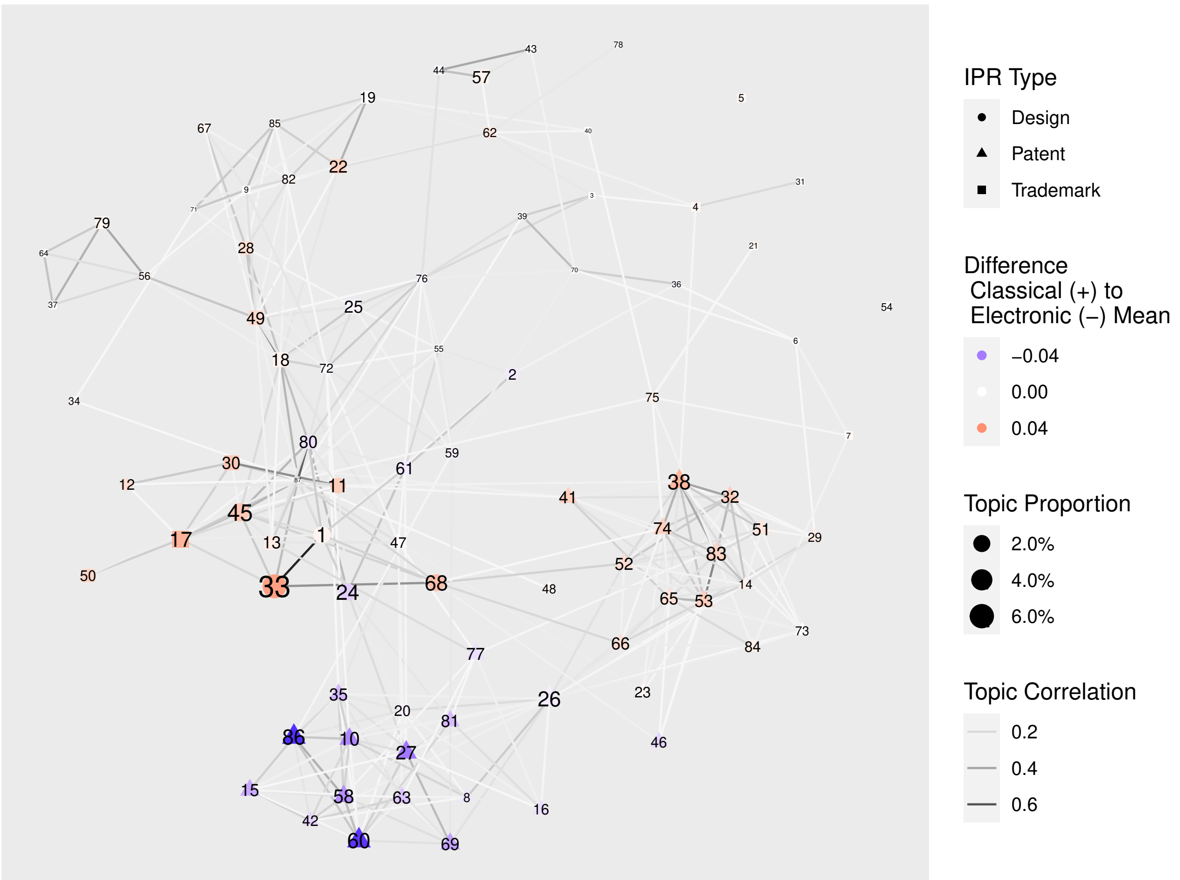


Figure D.13: Network of Musical Instrument Topics, Differentiated for Electronic and Classical Instruments from 1996 to 2000.
The network illustrates the topic relationships from 1996 to 2000, categorized by IPR types and electronic or classical musical instruments. Nodes represent topics, while links represent the likelihood of these topics to co-occur. The node size corresponds to the topic's proportion. IPR types are indicated by node shapes, while node colours represent the musical instrument type. Intensive blue signifies strong relevance to electronic musical instruments, and red represents a focus on classical musical instruments. In this network, there are 53 nodes related to classical topics and 34 related to electronic topics. The edges display the strength of correlation, with lighter colours indicating lower correlation and darker colours indicating stronger correlation.

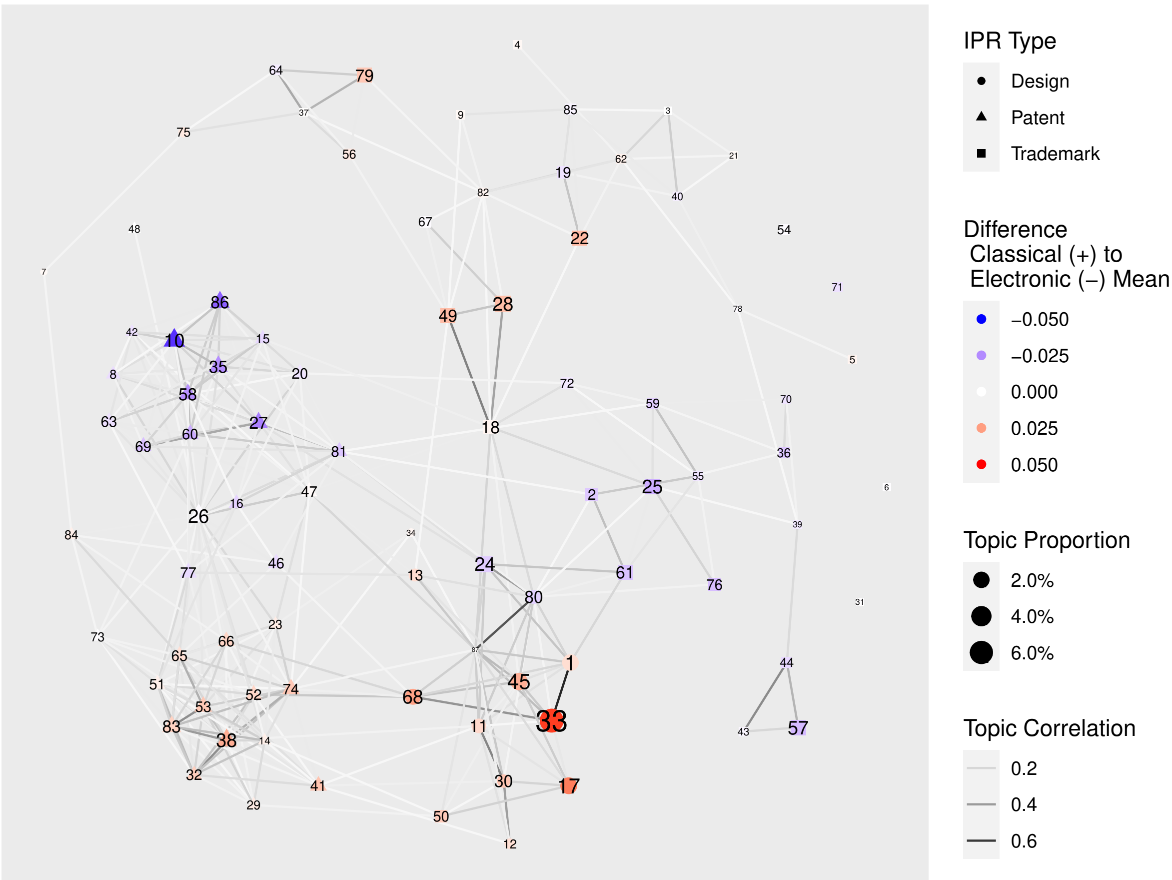


Figure D.14: Network of Musical Instrument Topics, Differentiated for Electronic and Classical Instruments from 2006 to 2010.

The network illustrates the topic relationships from 2006 to 2010, categorized by IPR types and electronic or classical musical instruments. Nodes represent topics, while links represent the likelihood of these topics to co-occur. The node size corresponds to the topic's proportion. IPR types are indicated by node shapes, while node colours represent the musical instrument type. Intensive blue signifies strong relevance to electronic musical instruments, and red represents a focus on classical musical instruments. In this network, there are 43 nodes related to classical topics and 44 related to electronic topics. The edges display the strength of correlation, with lighter colours indicating lower correlation and darker colours indicating stronger correlation.

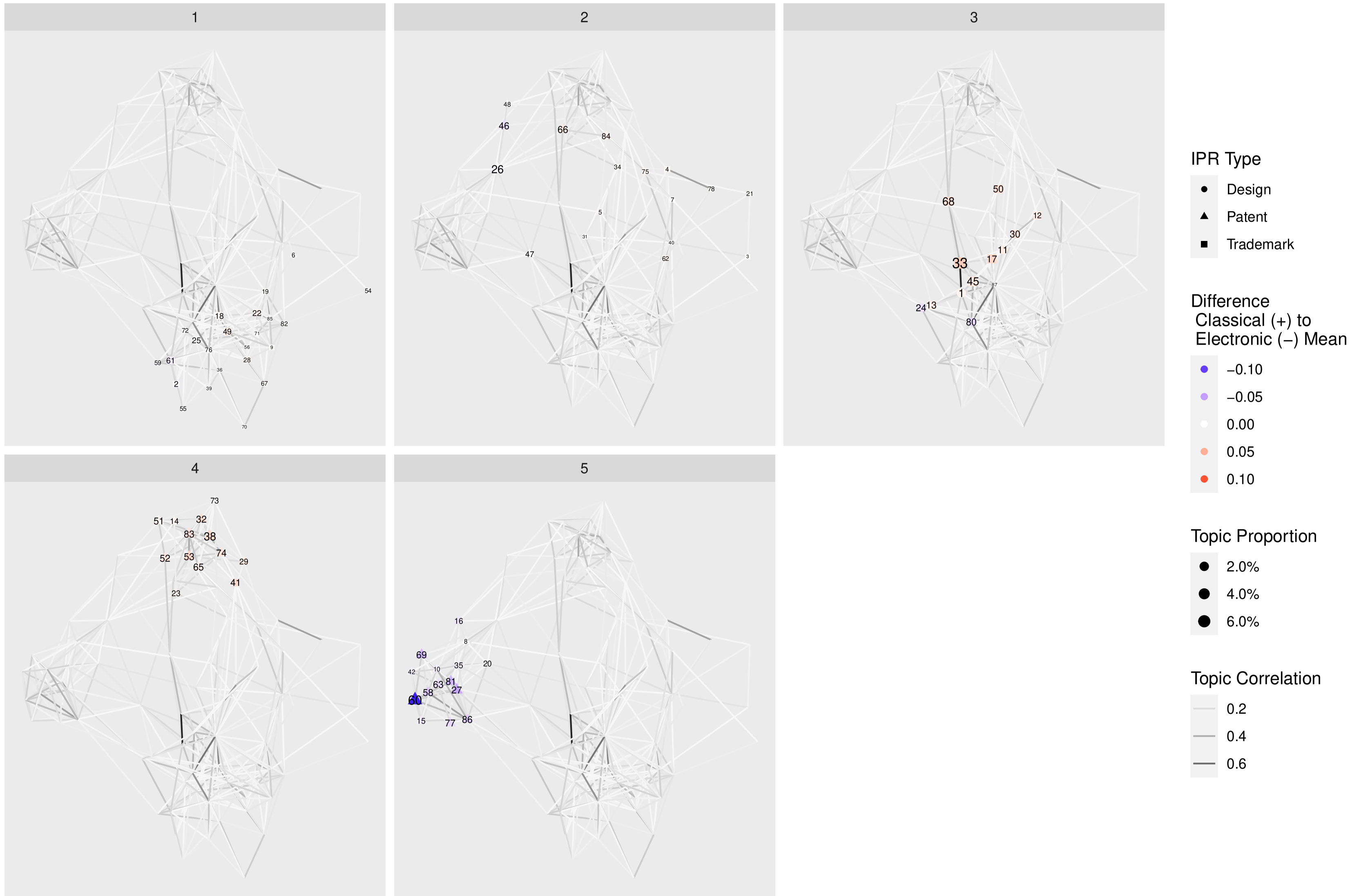


Figure D.15: Cluster Identification in Topic Networks of Electronic and Classical Musical Instrument from 1986 to 1990.
The figure displays clusters with more than three topics in electronic and classical musical instruments from 1986 to 1990, determined via random walk.

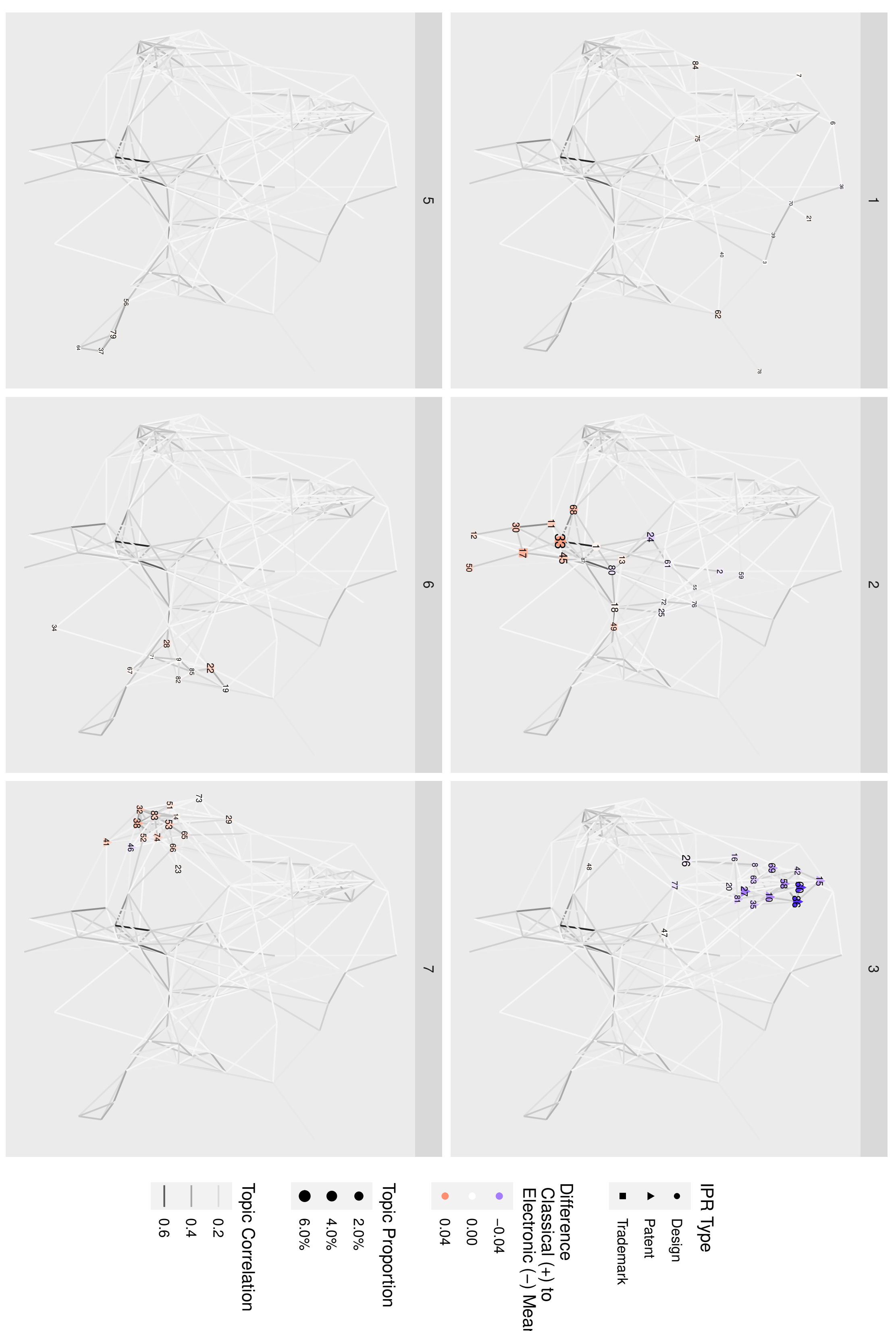


Figure D.16: Cluster Identification in Topic Networks of Electronic and Classical Musical Instrument from 1996 to 2000.
The figure displays clusters with more than three topics in electronic and classical musical instruments from 1996 to 2000, determined via random walk.

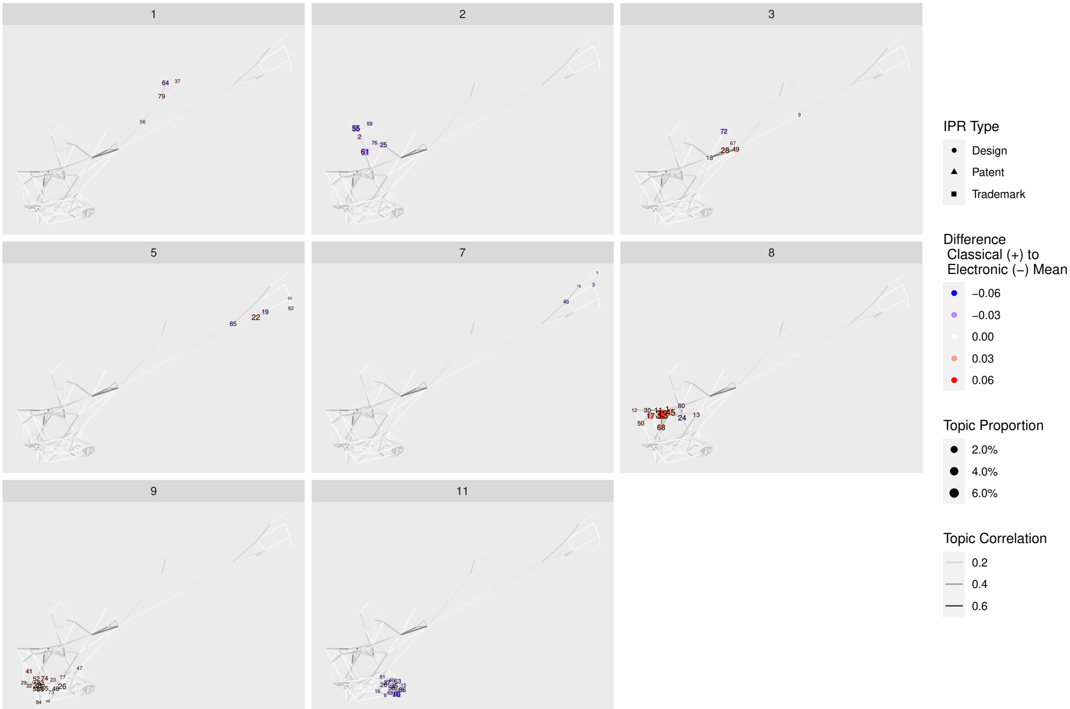


Figure D.17: Cluster Identification in Topic Networks of Electronic and Classical Musical Instrument from 2011 to 2015.
The figure displays clusters with more than three topics in electronic and classical musical instruments from 2011 to 2015, determined via random walk.

D.3 Firm Perspective

The appendix provides additional information on the firms involved as well as topics-firms-networks with more data.

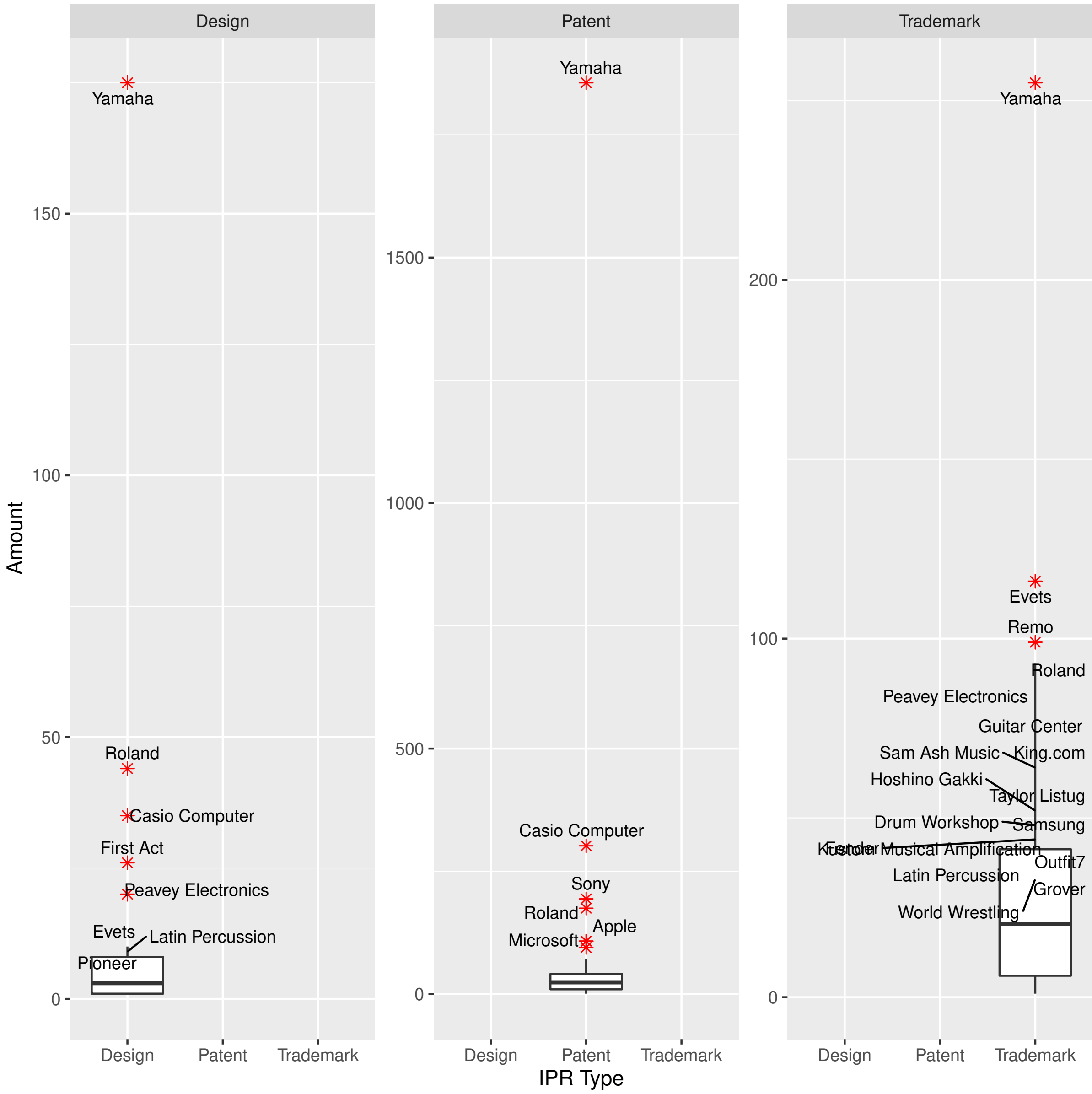


Figure D.18: Box Plot on IPRs per Major Firms.

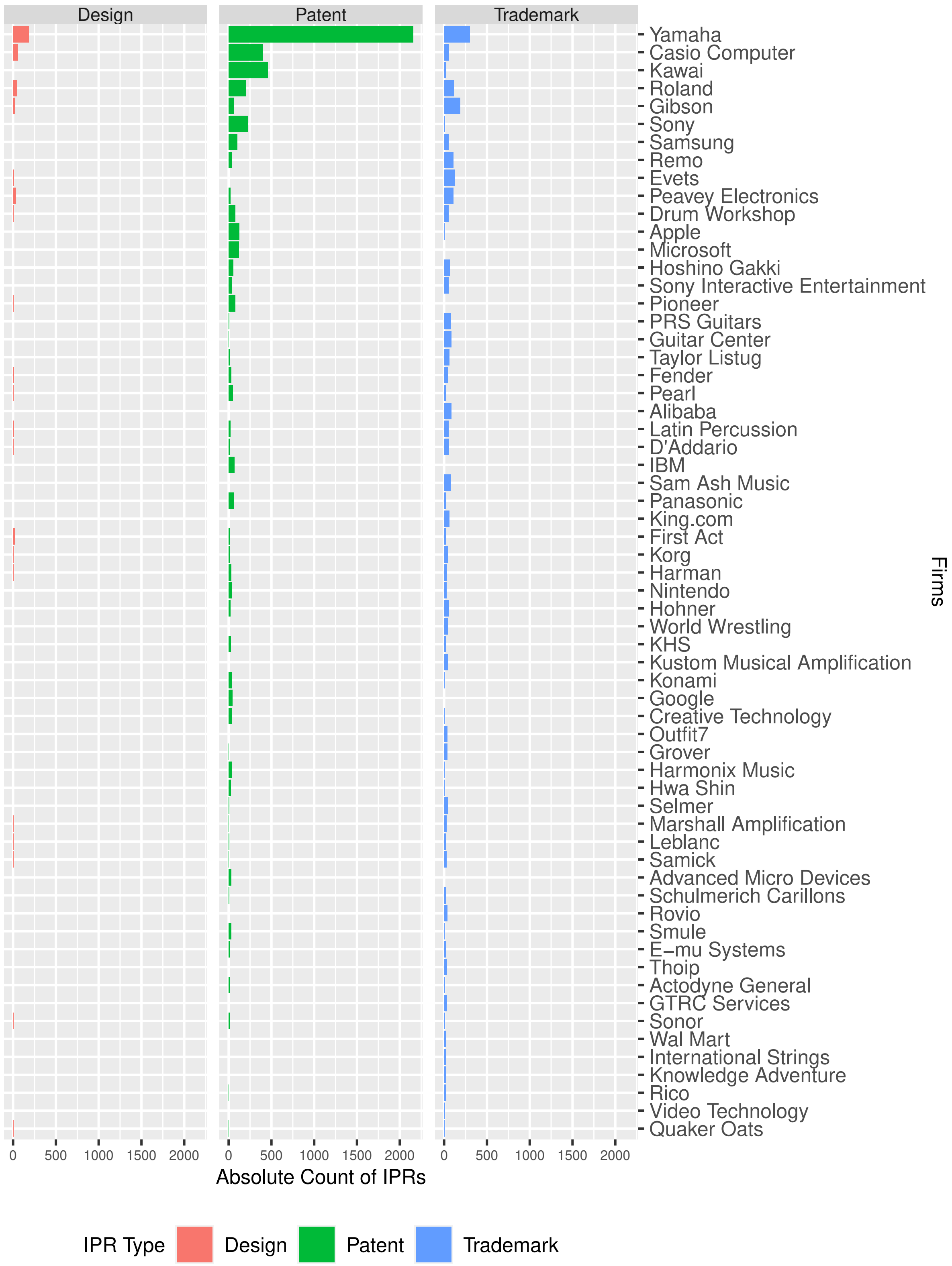


Figure D.19: Major Firms in Musical Instruments related to IPR Types.

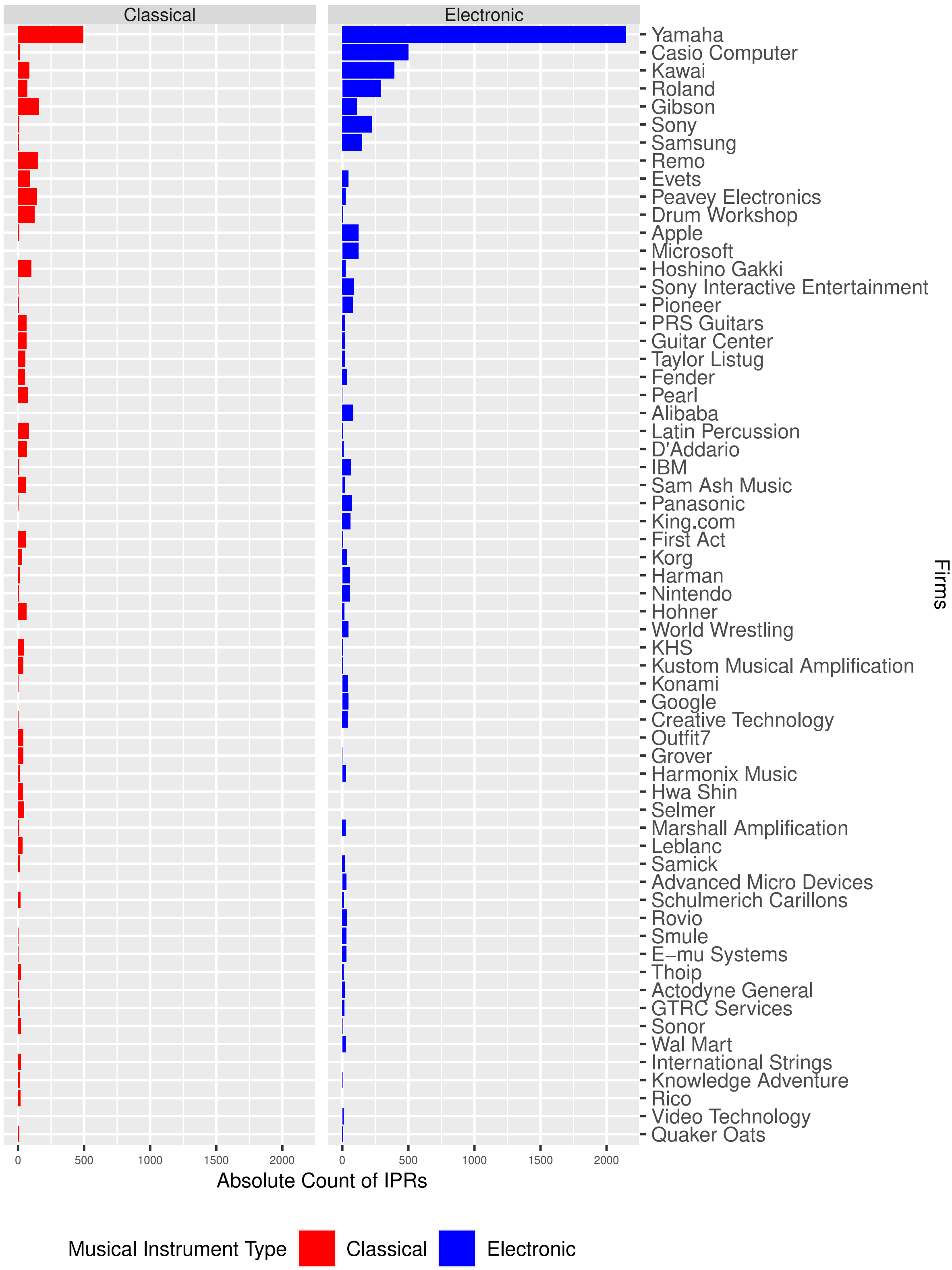


Figure D.20: Major Firms in Musical Instruments related to Electronic and Classical Musical Instruments.

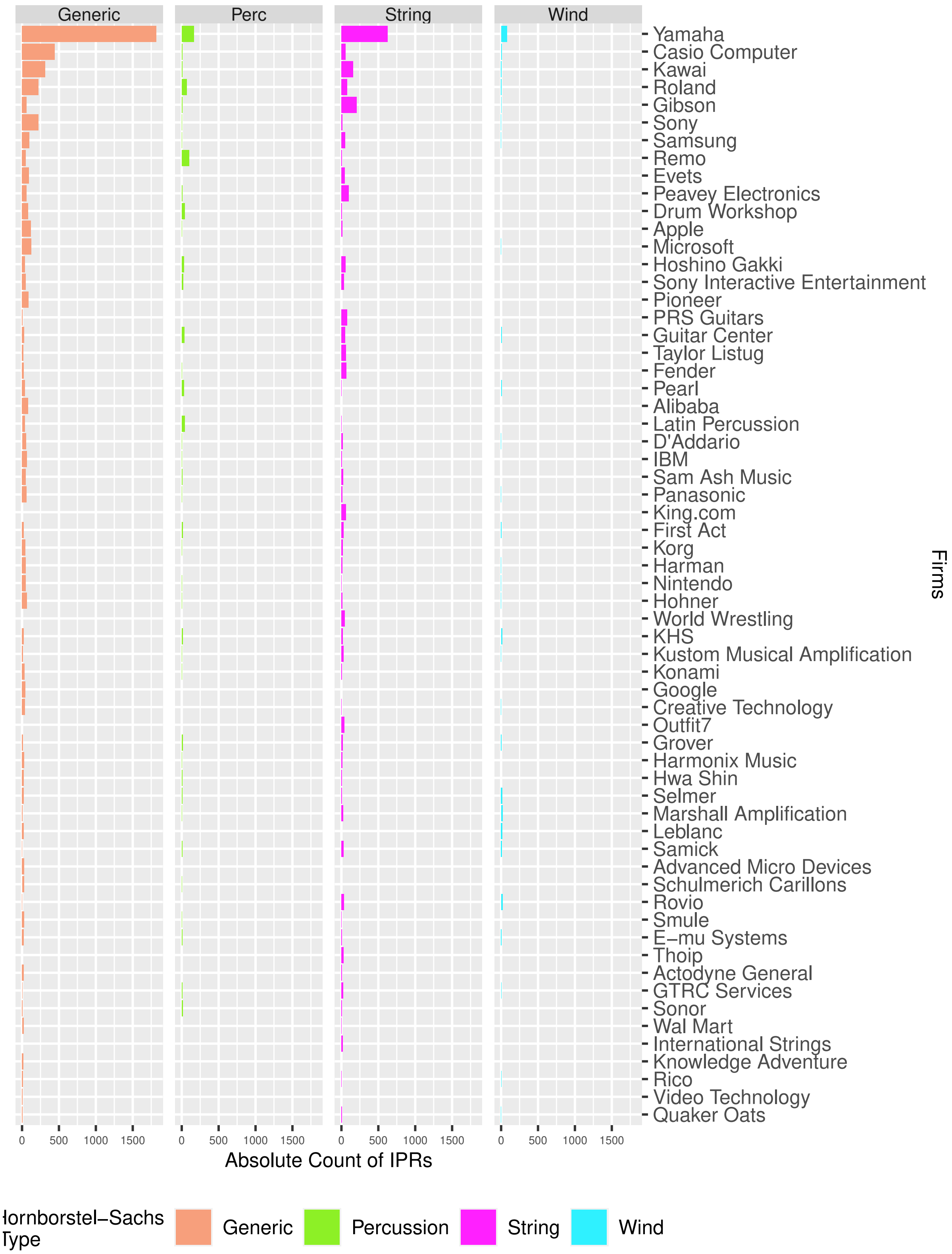


Figure D.21: Major Firms in Musical Instruments related to the Hornborstel-Sachs Classification.

Table D.1: Description of the Major Firms from A to F

Rank	Firm Name	Official Firm Name	Main Activities	Main Words	Firm Category	Source
54	Actodyne General	Actodyne General, Inc.	Electric Guitars, Accessories	stringed (18), include (15), string (15), assembly (13), dispose (12)	Producers	www.dnb.com /business-directory/ company-prof iles.actodyne_ general_inc. d567c49e3c402060d51705c8752832cb. html, www.lacemusic. com/
47	Advanced Micro Devices	Advanced Micro Devices, Inc.	Computer Hardware, Processors	audio (20), include (20), data (19), signal (19), memory (16)	Hardware	www.amd.com/de
22	Alibaba	Alibaba.com Limited	Retail	advertising (69), agency (69), business (69), commerce (69), communication (69)	Retail	www.alibaba.com/
11	Apple	Apple, Inc.	Apps (Digital Creation, Digital Consumption)	method (73), include (68), user (66), audio (58), system (52)	Software	www.apple.com/ios/ garageband/
2	Casio Computer	Casio Computer Co,. Ltd.	Keyboards, E-pianos, Consumer Electronics	electronic (178), data (135), tone (135), keyboard (131), generate (122)	Producer	www.casio-intl.com /np/en/emi/
38	Creative Technology	Creative Technology Limited	Amplifiers, Speakers, Accessories	provide (21), method (20), audio (19), data (17), digital (15)	Hardware	us.creative.com/
23	D'Addario	D'Addario Company, Inc.	Strings, Reeds, Accessories	string (32), guitar (13), accessory (11), drum (11), drumhead (10)	Classical Producers	www.daddario.com/
12	Drum Workshop	Drum Workshop, Inc.	Drums	drum (82), comprise (48), support (36), percussion (35), pedal (31)	Classical Producers	www.dwdrums.com/
51	E-mu Systems	E-mu Systems, Inc.	Studio Headphones, Electronic Musical Instruments	digital (22), sound (21), electronic (18), provide (18), multi (11)	Producers	www.emu.com/
9	Evets	Evets Corporation		sound (53), effect (50), pedal (47), guitar (42), electronic (34)	Producers	
20	Fender	Fender Musical Instruments Corporation	Acoustic and Electric Guitars	guitar (44), electric (19), amplifier (17), control (10), include (10)	Producer	www.fender.com /de-DE/start
27	First Act	First Act, Inc.	Classical Instruments For Education	guitar (24), claim (17), drum (16), include (13), string (12)	Classical Producers	www.jazwares.com /brands/first-act

The table provides an overview on the major Firms (A to F) in the data set. It provides the rank among the major Firms, the cleaned firm name, the main activity of the firm based on web searches. Further, the top words of documents of the firm are provided. As every firm is assigned to a category, this is also disclosed.

Table D.2: Description of the Major Firms from G to I

Rank	Firm Name	Official Firm Name	Main Activities	Main Words	Firm Category	Source
5	Gibson	Gibson Brands, Inc.	Guitar	guitar (192), stringed (66), string (58), electric (54), bass (41)	Producers	www.gibson.com/
39	Google	Google, Inc.	Digital App Platform	system (30), method (24), audio (23), base (23), determine (22)	Software	"research.google.com/audioset/ontology/musical_instrument_1.html, artsandculture.google.com/theme/-gJiyqdl_D_BIQ, musiclab.chromeexperiments.com/Experiments, music.youtube.com/googleplaymusic"
41	Grover	Grover Musical Products, Inc.	String Instruments	percussion (10), toy (9), accessory (7), guitar (7), drum (6)	Classical Producers	grotro.com/
55	GTRC Services	GTRC Services, Inc.	Musical Instrument Retail	electronic (12), keyboard (12), guitar (11), accessory (8), audio (8)	Retail	audiotools.com/en_mi_links_gt.html
17	Guitar Center	Guitar Center, Inc.	Musical Instrument Retail	accessory (44), guitar (44), feature (38), music (38), service (38)	Retail	www.guitarcenter.com/
30	Harman	Harman International Industries, Inc.	Audio Equipment: Speakers, Sound Production	signal (50), audio (41), sound (29), electronic (26), processor (26)	Hardware	www.harman.com/
42	Harmonix Music	Harmonix Music Systems, Inc.	Video Games, Soundtracks	game (27), rhythm (23), action (21), include (20), method (19)	Game Developer	www.harmonixmusic.com/
33	Hohner	Matth. Hohner AG	Harmonica, Accordions, Flutes	accordion (17), guitar (10), harmonica (10), ukulele (10), accessory (5)	Classical Producers	www.hohner.de/
14	Hoshino Gakki	Hoshino Gakki Co., Ltd	Drums, Guitars	drum (48), guitar (40), include (35), body (23), support (22)	Producers	www.hoshinogakki.co.jp/hoshino_e/products/index.html
43	Hwa Shin	Hwa Shin Musical Instrument Co., Ltd	Drums, Accessories	include (20), drum (17), screw (17), base (16), stand (16)	Classical Producers	matchory.com/supplier/hwa-shin-musical-instrument
25	IBM	IBM Corporation		method (41), system (35), include (27), audio (23), generate (22)	Hardware	www.ibm.com/de-de
58	International Strings	International Strings, Inc.	Violin, Cellos, Violas,	bass (18), bow (18), cello (18), string (18), viola (18)	Classical Producers	www.shopnemc.com/category/International_Strings

The table provides an overview on the major Firms (G to I) in the data set. It provides the rank among the major Firms, the cleaned firm name, the main activity of the firm based on web searches. Further, the top words of documents of the firm are provided. As every firm is assigned to a category, this is also disclosed.

Table D.3: Description of the Major Firms from K to P

Rank	Firm Name	Official Firm Name	Main Activities	Main Words	Firm Category	Source
3	Kawai	Kawai Musical Instruments Manufacturing Co., Ltd.	Piano, Keyboard	electronic (148), key (140), tone (135), keyboard (123), data (122)	Producers	www.kawai-global.com/
35	KHS	K.H.S. Musical Instruments Co., Ltd	Classical Musical Instruments	stand (18), mount (16), drum (15), pivotally (13), assembly (12)	Classical Producers	world.khsmusic.com/en
28	King.com	King.com Limited	Apps (Digital Consumption), Computer Games	bag (62), board (62), bow (62), box (62), card (62)	Game Developer	www.king.com/de
59	Knowledge Adventure	Knowledge Adventure, Inc.	Educational Games	accompaniment (19), exclusive (19), learn (19), play (19), provide (19)	Game Developer	www.knowledgeadventure.com/
37	Konami	Konami Corporation		game (31), player (26), music (23), device (20), provide (20)	Game Developer	www.konami.com/en/
31	Korg	Korg, Inc.	Production	sound (26), electronic (20), signal (18), synthesizer (16), tuner (16)	Producers	www.korg.com/de/products/
36	Kustom Musical Amplification	Kustom Musical Amplification, Inc.	Amplification	guitar (29), stringed (20), amplifier (15), control (11), speaker (11)	Producers	kustom.com/
24	Latin Percussion	Latin Percussion, Inc.	Drums	percussion (28), drum (20), mount (12), provide (12), block (8)	Classical Producers	www.lpmusic.com/
44	Leblanc	Leblanc, Inc.	Brass, Wind instruments	clarinet (15), woodwind (7), wind (5), brass (4), mouthpiece (4)	Classical Producers	www.company-histories.com/G-Leblanc-Corporation-Company-History.html
45	Marshall Amplification	Marshall Amplification plc	Guitar Amplification, Drums, Percussions	amplifier (24), cabinet (23), guitar (23), loudspeaker (23), sound (23)	Producers	marshall.com/
13	Microsoft	Microsoft Corporation	Apps (Digital Consumption)	provide (60), user (58), system (57), include (49), audio (48)	Software	www.microsoft.com/de-de/p/groove-musik/9wzdncrfj3pt?activetab=pivot:overviewtab
32	Nintendo	Nintendo Co., Ltd	Games, Equipment	game (40), sound (35), music (34), apparatus (32), display (30)	Game Developer	www.nintendo.com/
40	Outfit7	Out Fit 7 Limited	Gaming Apps	album (39), animate (39), bag (39), ball (39), bib (39)	Game Developer	outfit7.com/applications/
29	Panasonic	Panasonic Corporation		apparatus (25), signal (25), output (24), include (23), music (23)	Hardware	www.panasonic-electric-works.com/de/firmengeschichte.htm
21	Pearl	Pearl Musical Instrument Company	Percussion, Accessories	drum (38), mount (22), assembly (21), provide (21), comprise (20)	Producers	pearldrum.com/
10	Peavey Electronics	Peavey Electronics Corporation	Amplifiers, Speakers, Accessories	guitar (66), amplifier (48), bass (12), body (12), sound (12)	Producers	peavey.com/
16	Pioneer	Pioneer Corporation	Mischpult, DJ Equipment	apparatus (47), information (47), device (34), sound (34), include (32)	Hardware	www.pioneerelectronics.com/PUSA/Home
18	PRS Guitars	Paul Reed Smith Guitars	Guitars	guitar (72), bass (24), pick (11), pickup (10), tune (10)	Classical Producers	prsguitars.com/

The table provides an overview on the major Firms (K to P) in the data set. It provides the rank among the major Firms, the cleaned firm name, the main activity of the firm based on web searches. Further, the top words of documents of the firm are provided. As every firm is assigned to a category, this is also disclosed.

Table D.4: Description of the Major Firms from Q to S

Rank	Firm Name	Official Firm Name	Main Activities	Main Words	Firm Category	Source
61	Quaker Oats	The Quaker Oats Company		electronic (4), guitar (4), keyboard (4), pal (4), electric (3)	Game Developer	www.quakeroats.com/
8	Remo	Remo, Inc.	Drums	drum (91), drumhead (72), material (25), sound (25), surface (23)	Classical Producers	remo.com/products/
60	Rico	Rico Products	Clarinetts, Accessories	mouthpiece (8), reed (7), accessory (6), woodwind (6), material (4)	Classical Producers	
4	Roland	Roland Corporation	Piano, Synthesizers, Amplifiers	electronic (175), sound (78), provide (77), include (75), percussion (69)	Producers	www.roland.com/global/
49	Rovio	Rovio Entertainment, Ltd	Games, Sound Design	bag (30), book (30), box (30), card (30), computer (30)	Game Developer	investors.rovio.com/en
26	Sam Ash Music	Sam Ash Music Corporation	Musical Instrument Retail	accessory (30), service (24), retail (18), guitar (16), store (16)	Retail	www.samash.com/
48	Samick	Samick Musical Instruments Co., Ltd	Guitars, Pianos	guitar (17), piano (16), acoustic (15), electric (15), electronic (11)	Producers	samickpiano.com/
7	Samsung	Samsung Electronics Co., Ltd.	Headphones, Apps (Digital Creation, Digital Consumption)	data (85), audio (82), music (82), sound (81), information (79)	Software	news.samsung.com/global/samsung-devices-and-services-every-music-lover-should-know-about
50	Schulmerich Carillons	Schulmerich Carillons, Inc.	Bells	carillon (20), replacement (12), structural (12), thereof (12), electronic (11)	Classical Producers	schulmerichcarillons.com/
46	Selmer	The Selmer Company, Inc.	Clarinetts	accessory (8), drum (7), orchestral (7), band (6), horn (5)	Classical Producers	www.company-histories.com/The-Selmer-Company-Inc-Company-History.html
52	Smule	Smule, Inc.	Singing App, Collaboration Music	device (28), performance (28), user (24), capture (23), render (21)	Software	www.smule.com/
56	Sonor	Sonor GmbH	Drums	drum (14), percussion (10), include (8), support (8), device (7)	Classical Producers	www.sonor.com/
6	Sony	Sony Corporation	Audio Equipment, Headphones, Microphones	apparatus (105), data (100), information (93), include (92), signal (83)	Hardware	www.sony.net/Products/proaudio/en/
15	Sony Interactive Entertainment	Sony Interactive Entertainment, Inc.		game (59), video (56), music (55), sound (51), program (47)	Game Developer	www.sie.com/en/index.html

The table provides an overview on the major Firms (Q to S) in the data set. It provides the rank among the major Firms, the cleaned firm name, the main activity of the firm based on web searches. Further, the top words of documents of the firm are provided. As every firm is assigned to a category, this is also disclosed.

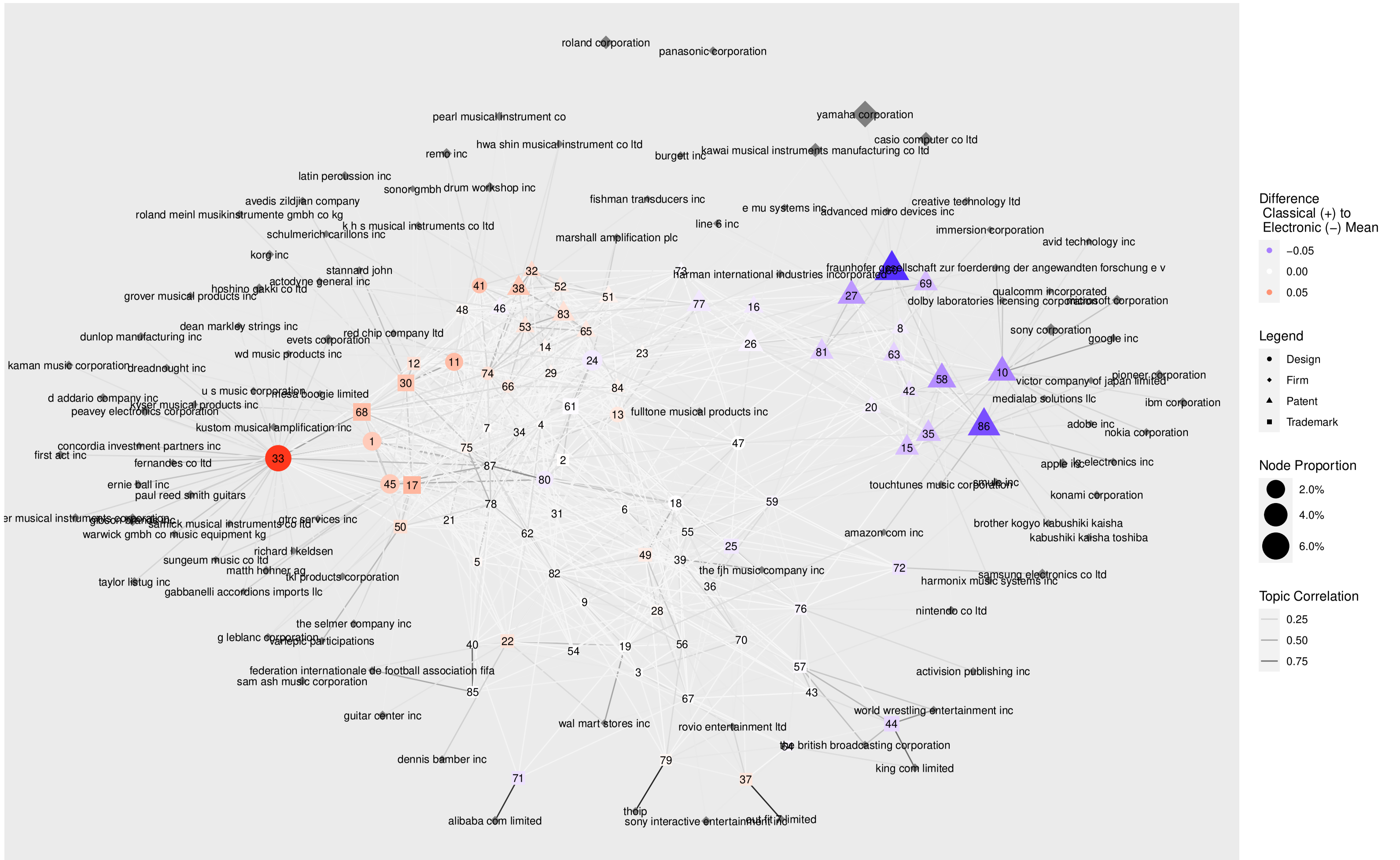


Figure D.22: 100 Major Firms in a Topics-Firms-Network from 1986 to 2015.
The network illustrates the 100 major firms in the field of musical instruments from 1986 to 2015. Nodes are linked to firms and various intellectual property rights, including design, patent, and trademark. Topic nodes are numbered to indicate the topic they represent. Topic nodes are coloured either in red or blue, depending on their association with electronic or classical musical instruments. The shape and colour of topic nodes are determined by the dominant IPR type or musical instrument category in the dataset. Firm nodes are rhombus-shaped and coloured based on their firm type. The connections represent the topic correlation from firms to topics and among topics. Lighter colours indicate a weaker correlation.

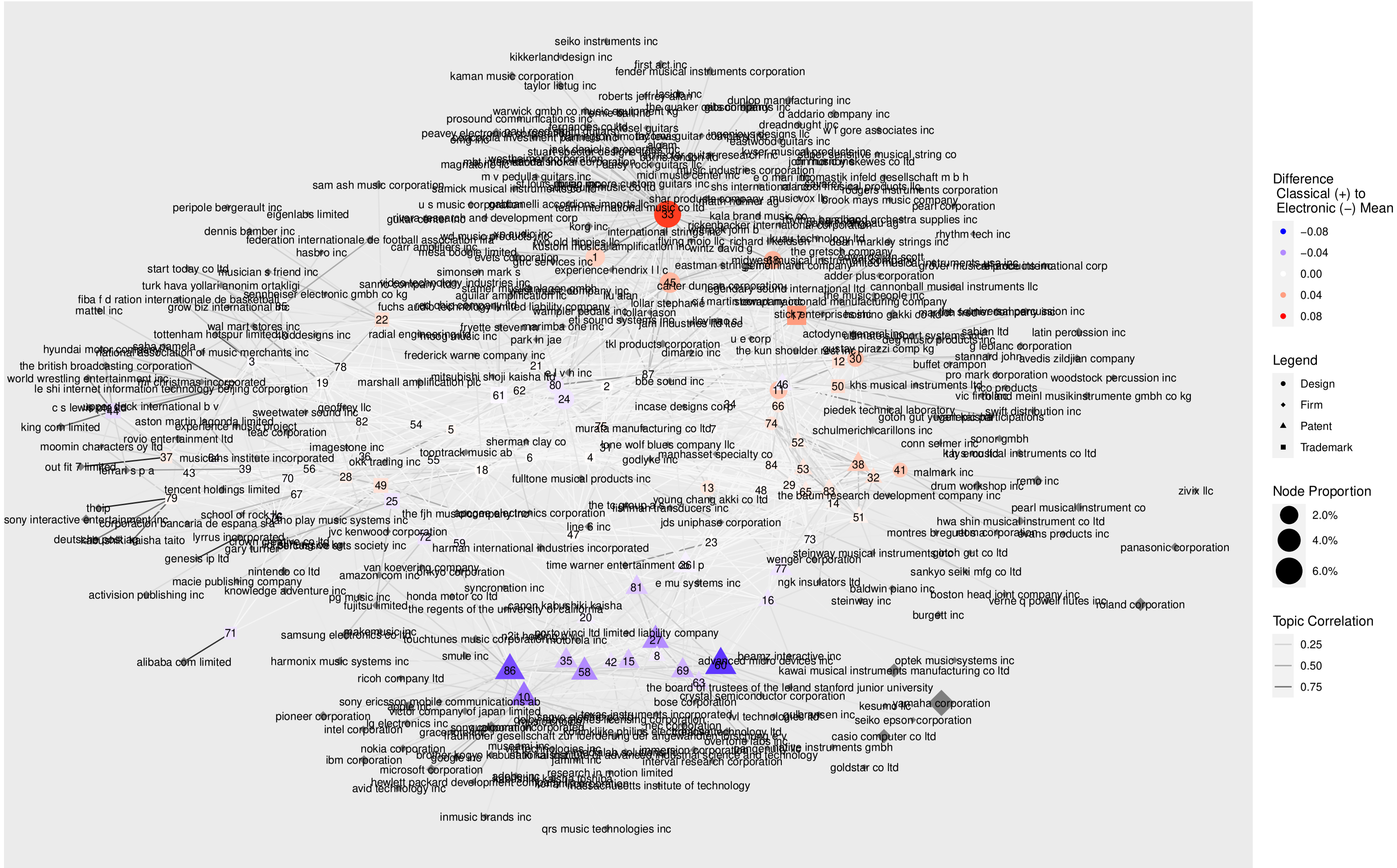


Figure D.23: 300 Major Firms in a Topics-Firms-Network from 1986 to 2015.
The network illustrates the 300 major firms in the field of musical instruments from 1986 to 2015. Nodes are linked to firms and various intellectual property rights, including design, patent, and trademark. Topic nodes are numbered to indicate the topic they represent. Topic nodes are coloured either in red or blue, depending on their association with electronic or classical musical instruments. The shape and colour of topic nodes are determined by the dominant IPR type or musical instrument category in the dataset. Firm nodes are rhombus-shaped and coloured based on their firm type. The connections represent the topic correlation from firms to topics and among topics. Lighter colours indicate a weaker correlation.

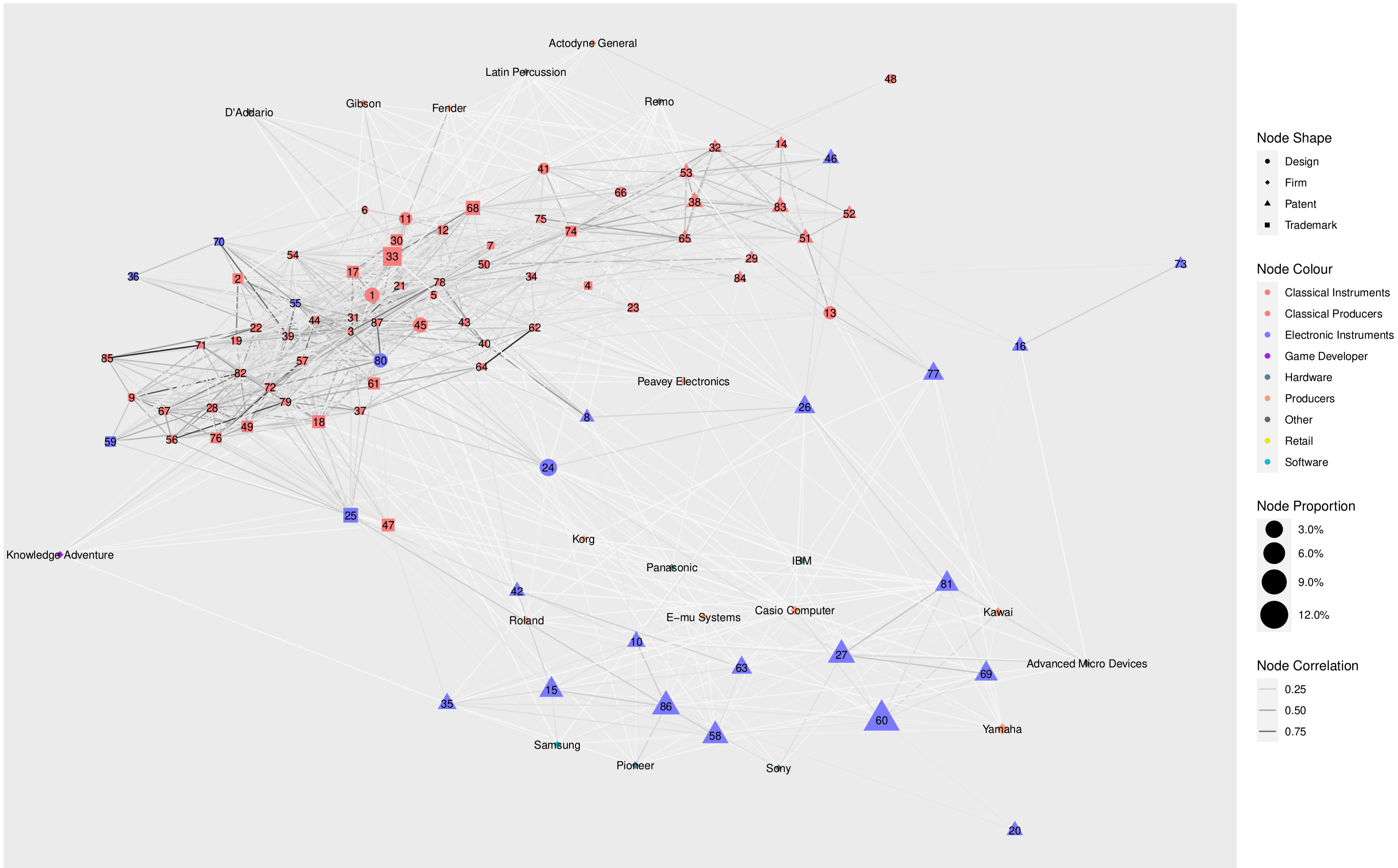


Figure D.24: Major Topics-Firms-Network from 1991 to 1995.
The network displays the 20 major firms in the field of musical instruments from 1991 to 1995. Nodes represent firms and topics. Topic nodes are labelled with numbers, and their shapes indicate the IPR document with the highest mean occurrence: design (round), patent (triangular), and trademark (rectangular). Topic nodes are coloured either red or blue, depending on their association with electronic or classical musical instruments. The shape and colour are determined by the strongest document type or musical instrument type in the dataset. Firm nodes are rhombus-shaped and coloured according to their firm type, with firm names displayed. Links represent the topic correlation from firms to topics and between topics, with lighter colours indicating lower correlations.

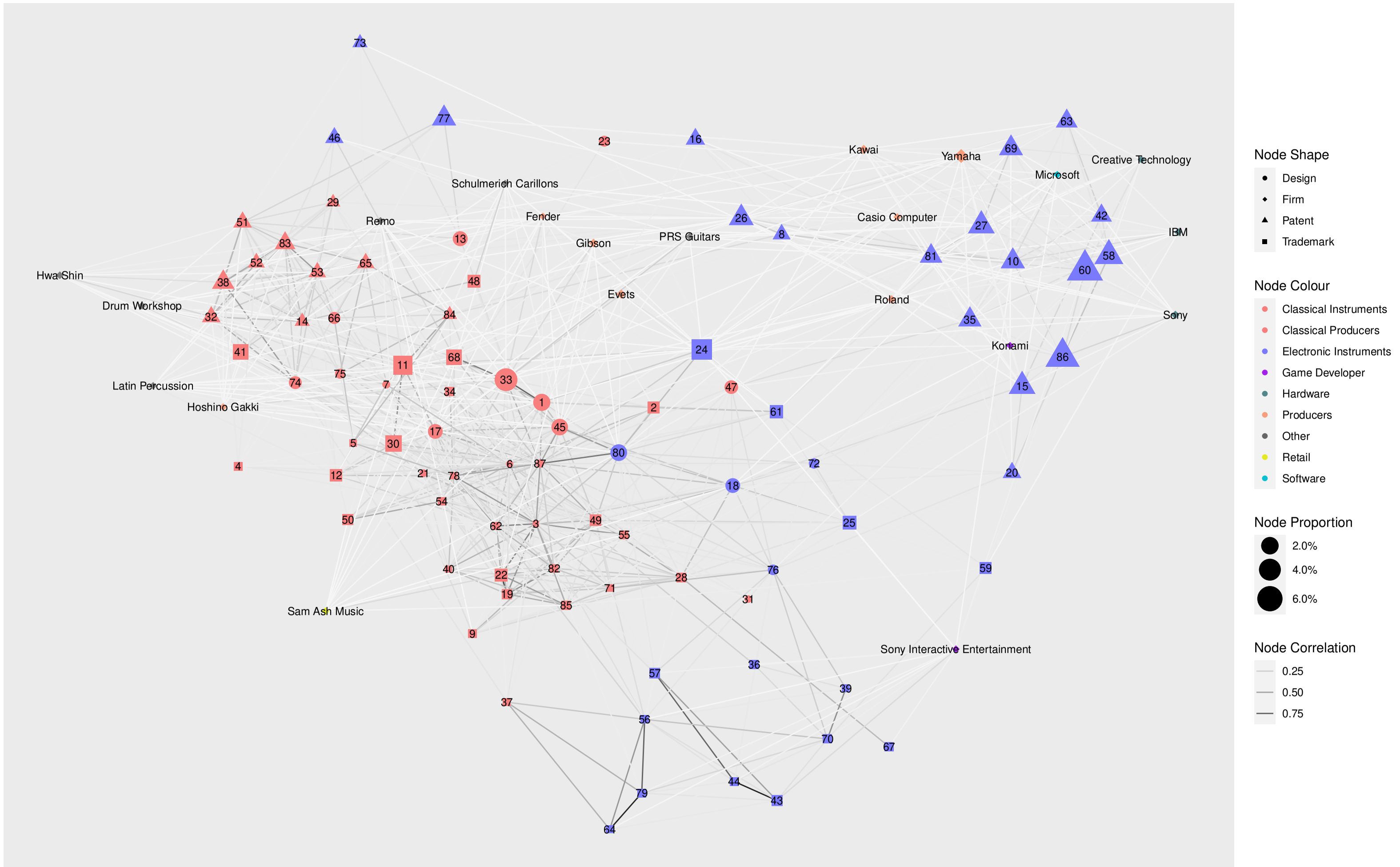


Figure D.25: Major Topics-Firms-Network from 1996 to 2000.
The network displays the 21 major firms in the field of musical instruments from 1996 to 2000. Nodes represent firms and topics. Topic nodes are labelled with numbers, and their shapes indicate the IPR document with the highest mean occurrence: design (round), patent (triangular), and trademark (rectangular). Topic nodes are coloured either red or blue, depending on their association with electronic or classical musical instruments. The shape and colour are determined by the strongest document type or musical instrument type in the dataset. Firm nodes are rhombus-shaped and coloured according to their firm type, with firm names displayed. Links represent the topic correlation from firms to topics and between topics, with lighter colours indicating lower correlations.

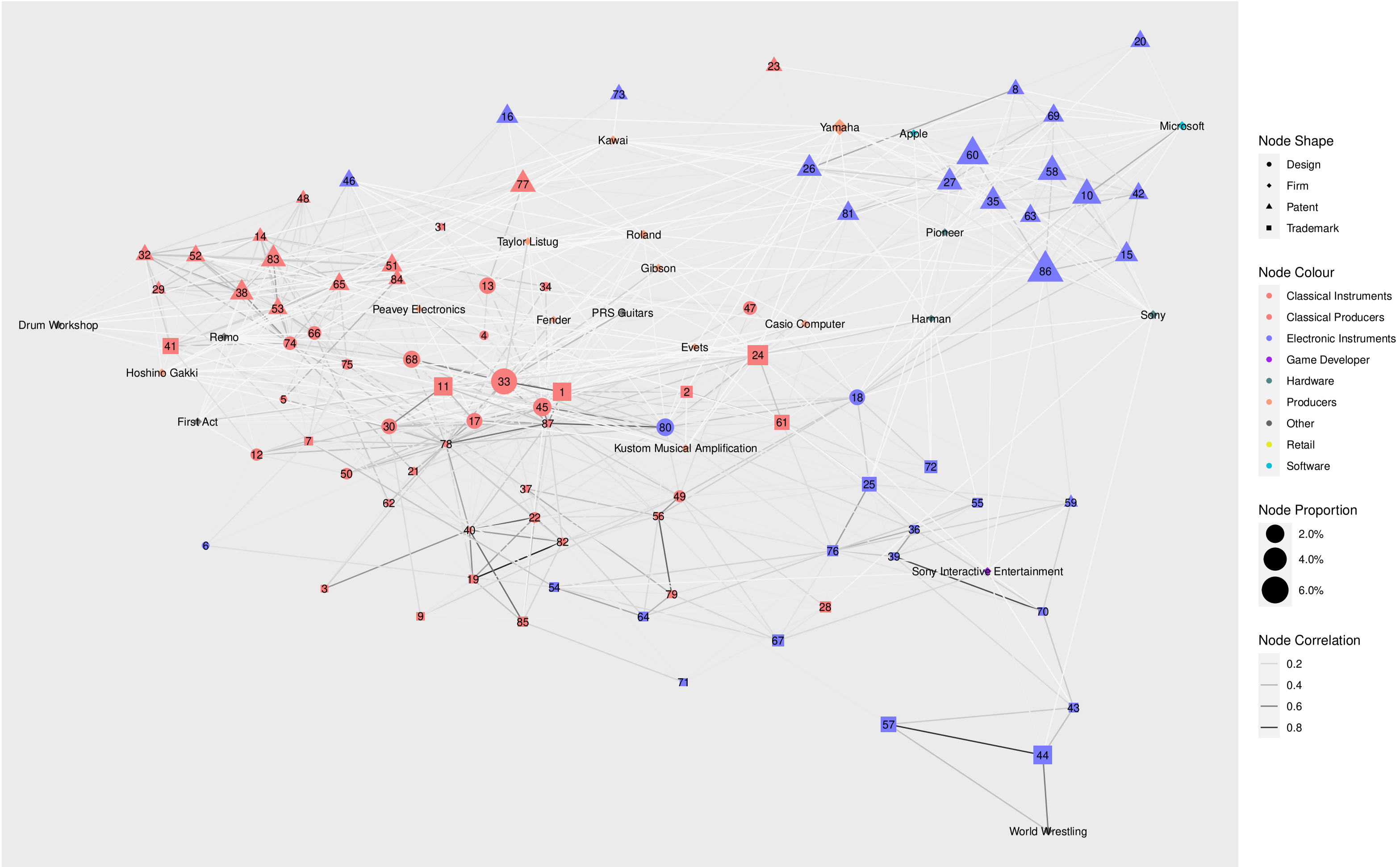


Figure D.26: Major Topic-Firm-Network from 2001 to 2005.
The network displays the 22 major firms in the field of musical instruments from 2001 to 2005. Nodes represent firms and topics. Topic nodes are labelled with numbers, and their shapes indicate the IPR document with the highest mean occurrence: design (round), patent (triangular), and trademark (rectangular). Topic nodes are coloured either red or blue, depending on their association with electronic or classical musical instruments. The shape and colour are determined by the strongest document type or musical instrument type in the dataset. Firm nodes are rhombus-shaped and coloured according to their firm type, with firm names displayed. Links represent the topic correlation from firms to topics and between topics, with lighter colours indicating lower correlations.

Table D.5: Description of the Major Firms from T to Y

Rank	Firm Name	Official Firm Name	Main Activities	Main Words	Firm Category	Source
19	Taylor Listug	Taylor Listug, Inc.	Piano, Organs, Others	guitar (54), carry (22), transducer (15), vibration (15), pickup (14)	Producers	`www.bloomberg.com/profile/company/7341418Z:US`
53	Thoip	Thoip		activity (28), address (28), adhesive (28), album (28), almanac (28)	Other	
62	Video Technology	Video Technology Industries, Inc.	Electronic Keyboards	electronic (9), keyboard (9)	Producers	
57	Wal Mart	Wal Mart Stores, Inc.	Retail	game (22), craft (21), furniture (20), appliance (19), beauty (19)	Retail	`www.walmart.com/`
34	World Wrestling	World Wrestling Entertainment, Inc.		accessory (46), action (46), animal (46), arcade (46), board (46)	Other	`www.wwe.com`
1	Yamaha	Yamaha Corporation	Electronic and Classical Musical Instruments	tone (834), data (800), generate (730), electronic (716), provide (662)	Producers	`de.yamaha.com/de/products/musical_instruments/index.html`

The table provides an overview on the major Firms (T to Y) in the data set. It provides the rank among the major Firms, the cleaned firm name, the main activity of the firm based on web searches. Further, the top words of documents of the firm are provided. As every firm is assigned to a category, this is also disclosed.

Table D.6: Documents per Major Firms per Five-Year Intervals for the Ranks 1 to 54

Rank	Firm Name	1986–1990	1991–1995	1996–2000	2001–2005	2006–2010	2011–2015	1986–2015
1	Yamaha	**315**	**292**	**421**	**588**	**435**	**302**	**2353**
2	Casio Computer	**145**	**97**	**24**	**25**	**33**	**86**	**410**
3	Kawai	**58**	**115**	**74**	**53**	**40**	**33**	**373**
4	Roland	**22**	**21**	**45**	**67**	**77**	**93**	**325**
5	Gibson	**17**	**47**	**53**	**44**	**60**	**47**	**268**
6	Sony	6	**16**	**38**	**55**	**82**	**32**	**229**
7	Samsung	3	**24**	5	12	**31**	**72**	**147**
8	Remo	**13**	**16**	**28**	**43**	8	**22**	**130**
9	Evets	**13**	7	**69**	**20**	4	13	**126**
10	Peavey Electronics	**23**	**38**	15	**23**	16	9	**124**
11	Apple	0	1	1	**20**	**53**	**47**	**122**
12	Drum Workshop	2	7	**26**	**25**	**28**	**34**	**122**
13	Microsoft	0	4	**19**	**58**	**29**	10	**120**
14	Hoshino Gakki	6	5	**19**	**34**	**23**	19	**106**
15	Sony Interactive Entertainment	0	0	**18**	**26**	**33**	11	**88**
16	Pioneer	7	**30**	14	**25**	9	1	**86**
17	Guitar Center	0	0	1	17	**45**	18	**81**
18	PRS Guitars	2	8	**17**	**26**	**22**	6	**81**
19	Taylor Listug	1	1	16	**26**	17	14	**75**
20	Fender	5	**13**	**18**	**20**	10	8	**74**
21	Pearl	5	10	4	16	**24**	14	**73**
22	Alibaba	0	0	0	0	**22**	**47**	**69**
23	D’Addario	2	**12**	14	10	8	**22**	**68**
24	Latin Percussion	**21**	**12**	**20**	15	0	0	**68**
25	IBM	5	**18**	**24**	4	6	9	**66**
26	Sam Ash Music	**8**	4	**20**	13	10	9	**64**
27	First Act	0	0	5	**38**	19	0	**62**
28	King.com	0	0	0	0	0	**62**	**62**
29	Panasonic	7	**14**	14	15	6	6	**62**
30	Harman	0	1	0	**22**	12	**26**	**61**
31	Korg	1	**18**	4	10	10	18	**61**
32	Nintendo	2	0	2	13	**32**	9	**58**
33	Hohner	3	5	7	2	9	**26**	**52**
34	World Wrestling	0	0	0	**42**	3	1	**46**
35	KHS	1	2	1	8	**24**	8	**44**
36	Kustom Musical Amplification	0	0	12	**25**	6	0	**43**
37	Konami	0	0	**19**	12	10	1	**42**
38	Creative Technology	0	2	**18**	15	4	1	**40**
39	Google	0	0	0	2	5	**33**	**40**
40	Outfit7	0	0	0	0	0	**39**	**39**
41	Grover	**13**	7	7	4	2	5	**38**
42	Harmonix Music	0	1	4	1	**24**	7	**37**
43	Hwa Shin	1	9	**24**	0	0	1	**35**
44	Leblanc	**8**	8	8	8	0	0	**32**
45	Marshall Amplification	0	1	0	1	5	**25**	**32**
46	Selmer	**9**	10	10	3	0	0	**32**
47	Advanced Micro Devices	0	**15**	15	1	0	0	**31**
48	Samick	**9**	1	8	12	0	1	**31**
49	Rovio	0	0	0	0	0	**30**	**30**
50	Schulmerich Carillons	1	1	**17**	6	0	5	**30**
51	E-mu Systems	**10**	**16**	3	0	0	0	**29**
52	Smule	0	0	0	0	4	**25**	**29**
53	Thoip	0	0	0	0	**27**	1	**28**
54	Actodyne General	0	**15**	8	3	1	0	**27**

The table provides an overview of the documents registered per Major Firms (rank 1 to 54) for the different five-year intervals from 1981 to 2015. The overview is ordered according to the rank of the firm among the Major Firms. If the firm is included in the networks related to that time interval, the document number is marked in bold.

Table D.7: Documents per Major Firms per Five-Year Intervals for the Ranks 55 to 62

Rank	Firm Name	1986–1990	1991–1995	1996–2000	2001–2005	2006–2010	2011–2015	1986–2015
55	GTRC Services	0	0	0	0	0	**25**	**25**
56	Sonor	**11**	6	3	3	0	1	**24**
57	Wal Mart	0	0	1	0	0	**22**	**23**
58	International Strings	**16**	0	0	2	0	2	**20**
59	Knowledge Adventure	0	**14**	5	0	0	0	**19**
60	Rico	**15**	1	1	0	0	0	**17**
61	Quaker Oats	**9**	0	0	0	0	0	**9**
62	Video Technology	**8**	1	0	0	0	0	**9**
	Classical Producers	8	3	6	5	3	4	16
	Game Developer	1	1	2	1	3	3	9
	Hardware	0	5	3	3	1	2	7
	Other	0	0	0	1	1	0	2
	Producers	10	10	8	11	7	6	18
	Retail	1	0	1	0	2	3	5
	Software	0	1	1	2	3	4	5
	Total Major Firms	20	20	21	22	20	22	62

The table provides an overview of the documents registered per major firms (rank 55 to 62) for the different 5-year intervals from 1981 to 2015. The overview is ordered according to the rank of the firm among the Major Firms. If the firm is included in the networks related to that time interval, the document number is marked in bold.

E English Summary

E.1 Introduction

Innovation is a fundamental driver of economic growth and social development. Firms are eager to innovate and be innovative to stay competitive (Schumpeter 2010). Major innovations like the steam engine or semiconductors are only some of the innovations that had major impacts on the development of society (Kuznets 1973). Not only large but also small innovations shape our lives. Recent advances can be observed in the provision of services where digitalisation is enabling music or video streaming, car-sharing, online banking or e-commerce. Innovation impacts economic growth, technological progress, investments, productivity, and resource allocation. It can take various forms, be it technological innovations, new goods or new services being introduced, an improvement in processes or organisational change. The market introduction is an essential aspect of an innovation (OECD and Eurostat 2005). As innovation is important for the welfare of a country and the competitiveness of firms, policymakers worldwide are trying to make their regions innovative, hoping to spur economic growth and social development. The policy should therefore be based on evidence. Innovation research intends to provide the evidence needed to guide policymakers in good decision-making.

Most of the evidence on innovation and innovation policy has been based on patent data. Patents are a means for firms to protect their inventions from imitation and gain a monopoly right on their invention (Neuhäusler 2009). Patents are considered a formal protection mechanism and are part of intellectual property rights (IPR) (WIPO 2020d). For patent protection, an inventive technological step is required, and the patent needs to publicly describe the invention (WIPO 2021h). Patent protection is globally available. Patents are systematically registered and cover long-time spans (Basberg 1987). The patent system further provides a sophisticated classification system that simplifies patent searches and allows for detailed research. This makes patents interesting for innovation policy, as they are data sources that reveal information on the current technological state-of-the-art on a global scale (Kleinknecht et al. 2002). In innovation research, patents tend to be used as an indicator for product innovation (Dziallas and Blind 2019).

Even though patents have their advantages as an indicator, they also have some shortcomings:

- Not every patent covers a commercialised invention. This means that the invention is not introduced into the marketplace. Therefore, these patents only cover inventions, not innovations (Basberg 1987). It is thereby not distinguishable which patents are used in the marketplace. For innovation policy, it is therefore of interest to cover the diffusion aspect to ensure a market introduction, as this is necessary to bring progress and affect economic growth.

- Patent application differs across sectors and firms. As patents require an inventive step in a technological area, sectors with lower technological content are underrepresented (Kleinknecht et al. 2002; Neuhäusler and Frietsch 2015). This does, however, not mean that there is no innovation in these sectors. It is just not captured in patent-based innovation indicators (Hirsch-Kreinsen 2008). Research showed that low-technology sectors contribute to economic growth (Mendonça 2009).

- Patents have difficulties covering services, tacit knowledge, or software inventions (Millot 2009; Mendonça 2009; USPTO 2017; USPTO 2015). These, however, are increasingly important.

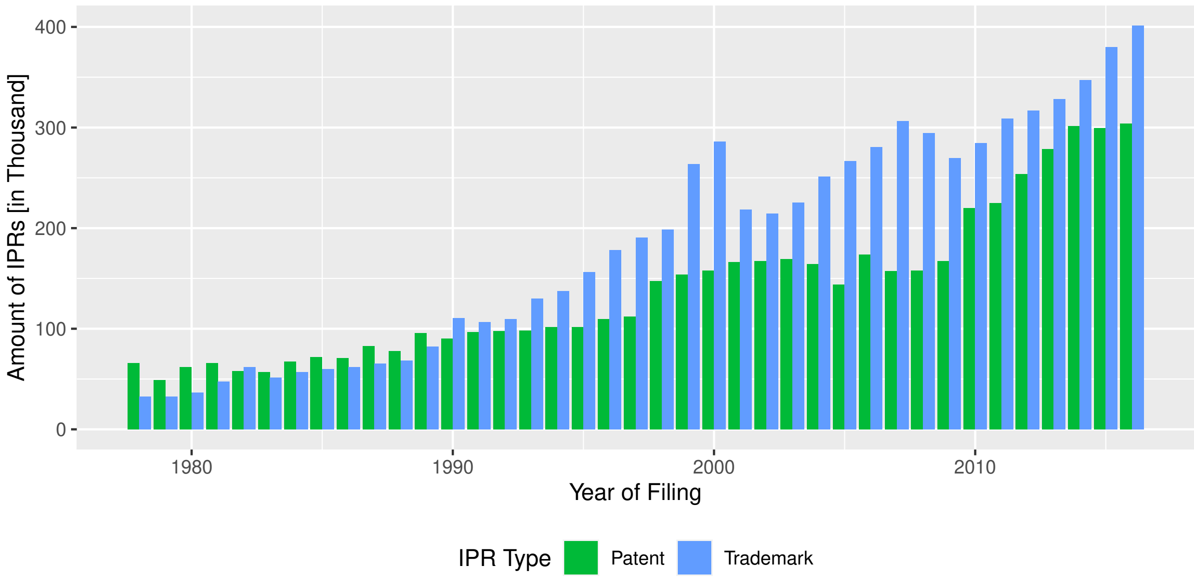


Figure E.1: Total Yearly Development of Granted USPTO Patents and Registered USPTO Trademarks between 1978 and 2016.
Source: Own representation based on USPTO data.

First, industrialised countries have major service sectors that depend on services being provided. Second, digitalisation is impacting not only technological sectors but all sectors. Therefore, a shift in innovation activities should be observed. Relying only on patent data might convey a limited perspective. Patent-based indicators could have overlooked the services mentioned above.

One possibility to include these perspectives and, in principle, overcome those shortcomings are trademarks. Trademarks can take various forms, mostly known are names representing a brand. For example, a trademark like "Apple Music" is foremost a sign or brand used to protect the name but can simultaneously stand for innovation in music provision. The name mostly requires registration for specific areas that indicate the usage of the trademark to attain trademark protection. Thus, the trademark prevents confusion and ensures distinctiveness between different goods or service providers (WIPO 2019b). Like patents, trademarks are also registered, widely available, and cover long-time spans (Flikkema et al. 2015), but are also applicable in all sectors and for product and services (S. J. H. Graham et al. 2013; Millot 2009). They are also formal protection mechanisms and intellectual property rights (WIPO 2020d). Trademark registrations increased in recent years and, for example, surpass patent registration in the United States Patent and Trademark Office (USPTO) (see Figure E.1). Unlike patents, trademarks often require a market introduction, enabling a diffusion perspective. They are applicable in services and low-technology sectors. Overall, trademarks are of interest to innovation research. Millot (2009), Mendonça et al. (2004), and S. J. H. Graham and Hancock (2014) suggest already their use as an innovation indicator, pointing out their ability to cover not only technical product innovation but also service, marketing or business process innovation.

Nevertheless, trademarks also have shortcomings when it comes to innovation research:

- The link of trademarks to innovation is not ensured. In contrast to patents, an inventive step is not a requirement during trademark registrations. That makes trademarks more complex as an innovation indicator. Several authors, therefore, research the linkage of trademarks and innovation: Seip et al. (2018) find that trademarks are registered at different stages of the innovation process, depending on the innovation to be protected. Mendonça et al. (2004) argue that innovative firms are more active in trademarking than in patenting based on results of the Third Community Innovation Survey (CIS 3). Flikkema et al. (2014) could show in their survey that above 50% of the trademark applications in their sample link to innovation, of which service innovation is captured mostly only in trademarks and not in patents. So an innovation linkage is present. It is, however, not ensured for trademarks. Nevertheless, in certain areas with limited patentability, trademarks are one of the only formal protection mechanisms covering innovation activities in these areas.

- Another shortcoming of trademarks is their limited granularity. In patents, the patent classification system is very sophisticated and elaborates on various innovation areas. The system classifies each patent according to its technological content, allowing for detailed analyses of specific fields and relations between these fields (WIPO 2020c). The trademark classification system, however, provides fewer opportunities for innovation analysis, as the categories are broader and less specific. For several million trademarks, only 45 NICE classes are commonly available (WIPO 2022a), compared to over 75,000 categories in the patenting system (EPO 2022). This becomes a challenge when using trademarks for specific innovation areas as this implies that the level of analysis remains general, limiting the level of granularity that can be attained.

Thus, two challenges need to be addressed: First, trademarks do not always cover innovation, and second, the granularity of trademark analysis based on current classifications remains limited. Bringing trademarks and patents together could address the innovation challenges of trademarks and address the shortcoming of patents. In contrast to trademarks, patents always cover inventions but not always innovation. Together, these protection mechanisms could allow us to cover inventions and ensure their market application. Further, the insights gained on technological innovations based on patents could be enhanced with innovations in services or low-technology sectors based on trademarks. Malmberg (2005) combine trademarks and patents in the pharmaceutical industry to capture innovation activities. They find that both IPRs link to innovation activities in the field, with trademarks covering short-term activities. The authors suggest using both IPRs for valuable insights but also point out that industry-specific differences exist in the applicability of trademarks as an innovation indicator. Ribeiro et al. (2022) assess the correlation of trademarks and patents for different sectors. The authors find that trademarks are especially of interest in sectors with low patentability or where patents are not applied, services, and less developed countries. Their results indicate that the combination of trademarks and patents improves the understanding of technological and non-technological sectors. Even though joint analyses of patents and trademarks might be useful, integrating trademarks with patents is not straightforward. A possibility is the use of concordance tables to combine trademarks with patents. Zolas et al. (2017) link trademarks on economic data based on keyword matching to derive insights on firms' export behaviour. The use of keywords enables already a more detailed perspective on trademarks and enables a class-level perspective. However, the authors point out a need for additional approaches if further granularity is intended. Other authors like Flikkema et al. (2015) integrate patents and trademarks via the same legal entities that registered the different IPRs. However, this approach always limits itself to direct links between patents and trademarks. Firms that are only active in trademarks or only active in patents but contribute to an innovation field and its development and diffusion are overlooked. Further, this makes it hard to combine the advantages of the different data sources, as much information is lost.

The challenge of limited granularity of trademark analysis still needs to be solved. Trademarks provide more granular information, containing descriptions of the trademark application for goods and services. These descriptions are textual data that can be exploited. The data can provide detailed information on the application areas of trademarks and innovations that are protected with trademarks. Text analysis covers techniques to identify patterns from unstructured, textual data to generate data-driven insights (Aggarwal and Zhai 2012). Antons et al. (2020) argue for the application of text analysis to improve existing and develop new measurements in innovation research.

Authors value or apply textual data analysis for various reasons:

- Text analysis enables the analysis of large data sets: This is handy when the data sets to be analysed surpass the amount of data that humans can process in a feasible time. Ozcan et al. (2021) base their analysis on 22,891 tweets from Twitter, Larsen and Thorsrud (2019) use 459,745 newspaper articles, or Feng et al. (2020) apply their analysis on 41,994 patents. Applying text analysis to textual data enables the coverage of larger amounts. Several authors name this argument as one of the main reasons to apply textual data analysis (Loshin 2013; Antons and Breidbach 2018; Bakhtin et al. 2020; Kayser 2017; J. Kim and C. Lee 2017; N. Kim et al. 2015; Kohler et al. 2014; Shen et al. 2020; B. Wang and Z. Wang 2018; Zhu and Porter 2002).

- Text analysis allows for the analysis of new data sources: As textual data are contained in various sources, it is possible to include data sources into the analysis that could not be analysed with standard approaches. For example, Dahlke et al. (2021) integrate fundamental human needs and innovations to assess how crises shape innovation. The authors use web-scraped innovation project descriptions for the innovation state and compare them to nine fundamental human needs described in the literature. J. Kim and C. Lee (2017) extract futuristic data from websites like Siemens, MIT technology review, or the World Future Society to extract scientifically relevant future topics and compare them to the state-of-the-art revealed in patent data. They intend to identify weak signals and extract novel topics by doing so. Mi et al. (2021) take advantage of survey data of older people to determine the demand for gerontechnology and compare the insights gained to reports of emerging technologies and patents to forecast future trends. A final example is Fiordelisi et al. (2019) that use 10-K reports of firms to extract information on corporate creativity to measure the impact on firm innovativeness and value. The authors find that creative culture fosters innovation. These are only some examples of how new data sources reveal interesting insights to be included in economic models. Although text analysis opens up more possibilities for data use and other data sources are available, patents are still the most commonly used data source. However, the amount of possible textual data available for innovation research is large and increases every year: In 2020 alone, over 2.6 million new research articles were published on Web of Science (WoS 2021). Regarding intellectual property rights such as patents, trademarks and designs, 304,126 patents, 234,444 trademarks and 28,886 designs were registered in 2016 alone in the U.S., with a total of 5,439,151 registered patents, 4,246,859 registered trademarks and 527,902 registered designs from 1978 until 2016.[1]
- Text analysis allows for the combination of different data sources, or use of heterogeneous data: Barriers imposed by the difference in the data structure of different data sources can be overcome with the usage of language and shared words. Several authors take advantage of this: Bakhtin et al. (2020) combine various data sources from scientific articles, patents, news, and awards to analytical reports with natural language processing to identify emerging patterns in agriculture and food production to provide insights for governments and firms for strategic planning activities. Kayser (2017) integrate media with the abstracts of publications via matched terms for an innovation diffusion perspective, while Shen et al. (2020) compare the textual content of patent and publication clusters to discovering opportunities of scientific advancement and technological innovation in smart health monitoring technologies. Combining several data sources can thus reveal new insights, and text analysis helps with the integration.

Another important argument in this context is the time-efficient analysis options (Bakhtin et al. 2020; Chiarello et al. 2021; Dahlke et al. 2021; Kayser 2017; J. Kim and C. Lee 2017; Zhou et al. 2020; Zhu and Porter 2002), which enable quick results and therefore quick statements. This is particularly relevant in rapidly changing fields where new findings are constantly being published. In addition, data-based knowledge extraction is relevant, so that knowledge can be gained independently of experts (Bakhtin et al. 2020; Basole et al. 2019; Larsen and Thorsrud 2019; Mi et al. 2021; Shen et al. 2020; Song et al. 2017; Zhou et al. 2020). This reduces expert bias and allows more people to generate knowledge. Overall, the application of textual analysis in innovation research can reveal interesting insights. Studies have been done, for example, on the current development of technologies like biofuels (Curci and Mongeau Ospina 2016) or blockchain (Chiarello et al. 2021), the diffusion of technologies like big data in scientific research

[1] Own data based on Query A.1, Query A.1 and Query A.1

(Y. Zhang et al. 2019) or upcoming topics in innovation like batteries or charging connectors in electric vehicles (Feng et al. 2020). The textual analysis can further be used to generate new ideas for innovation areas by extracting ideas from public sources like Twitter (Ozcan et al. 2021) or identifying promising areas by comparing consumer needs with available technologies (Mi et al. 2021). In the context of firms, text analysis can further reveal insights on the convergence of firms (N. Kim et al. 2015) or make the creativeness of firms traceable (Fiordelisi et al. 2019).

The various arguments for text analysis make it also interesting for the analysis of trademarks: Large amounts of trademarks are available for analysis. The textual data of trademarks can be used to extract detailed insights into different innovation areas and aspects of diffusion. Moreover, the textual data of trademarks allows us to bridge trademarks to other data sources. The combination of trademarks with different data sources can be of interest to capture, e.g., services but also the development of products more broadly. The combination could be achieved textually between trademarks and patents. This implies that shared words among trademarks and patents link the different intellectual property rights. This would also overcome the large differences between the available classification systems in trademarks and patents. The main assumption is that the description of an invention is similar across different intellectual property rights, but the area of protection and application is changing, with a slight shift of focus when it comes to the information conveyed. This means that patents cover more of the technological aspects, while trademarks shed light on the market application and diffusion of an invention. In this thesis, the combination of trademarks with patents is, thus, primarily of interest. However, designs are also considered an interesting data source for innovation areas with high importance on other shapes and patterns. Designs are applied to protect novel ornamental elements (WIPO 2021c). Like trademarks, designs are currently not included as a standard data source in analyses of innovation. In this work, they are included in Chapter 5, as they are of interest in the field under consideration. Like trademarks, they might contribute to a more accurate measurement of innovation and could provide interesting insights.

First attempts to use textual data of trademarks for analysis have been made: Flikkema et al. (2019) derive brand strategies of creation, extension or modernisation based on mark names and NICE coverage. Semadeni (2006) compares the wording of different firms' trademark applications to determine the firms' relative position to each other. The wording of trademarks can further reveal the likelihood of service innovations to be imitated (Semadeni and Anderson 2010). Even though some examples of the use of textual data of trademarks exist, it has yet to be common to apply textual data analysis. Castaldi (2020) points out in her overview that the focus is still primarily on structured, non-textual data of trademarks that are used for analysis and that the trademark information is not used yet to their full potential.

In terms of combination, textual data of trademarks and patents have been combined: Authors like M. Lee and S. Lee (2017) use the patents of firms to display technological knowledge and obtain details on the market positioning of competitors based on the trademark descriptions. H. Kim et al. (2017) match patents and trademarks broadly with keywords shared among patents and trademarks to uncover how technological knowledge can be applied in the market and identify potential areas for diversification. Patents thereby provide the invention perspective, while trademarks reveal areas of market introduction of these inventions. However, the approaches to combine the data sources still mostly focus on keywords. Another methodology to extract information from textual data that considers more than just keywords from the textual data is topic modelling. Topic modelling is an approach for extracting the topics occurring in a set of documents based on the words in these documents (Blei et al. 2003). In contrast to keyword-based matching, topic modelling can also consider synonyms or the context of words. Example applications in the context of innovation research are, for example, Basole et al. (2019), who identify topics and clusters in the entrepreneurial ecosystem via topic modelling and reveal that clusters of entrepreneurs are not only industry driven but also related to the similarity of their topics. Another example is Larsen and Thorsrud (2019), who apply topic modelling to identify valuable topics in news articles that are then integrated into economic models to improve the prediction of economic booms.

Overall, textual data of trademarks is the focus of this thesis. The information in the textual description of the application area of trademarks in goods and services could provide a broader perspective on innovation as they cover technical and non-technical areas and applications in goods and services. The textual data further allows for the combination of trademarks with patents that primarily consist of textual data, such as abstracts that summarise the content or patent claims that cover the protected areas. Analysing the textual data of trademarks is one solution to extract more information from trademarks and simultaneously provide a possibility to combine trademarks with other data sources. By doing so, this thesis extends innovation coverage with the textual inclusion of trademarks to provide a broader perspective on innovation. The broad innovation coverage supports empirically-based policy recommendations. However, to take advantage of the potential of textual data of trademarks, the data needs to be better comprehensively assessed, and the characteristics of the data source must be elaborated.

E.2 Research Questions

Several aspects need to be considered in the context of this thesis to extract the advantages of textual data of trademarks for innovation research:

Trademarks are a data source that provides, e.g. a perspective on services or low-technology areas but also covers high-technology areas. Trademarks, however, have the drawback of unclear innovation linkage and limited levels of detail for sophisticated analyses.

Textual Data can provide detailed information and, in that sense, provides an additional perspective in combination with structured data. Further, as textual data is available in various data sources, it holds the potential to allow for new combinations of data and, thus, new insights gained from these combinations. In the context of textual data, however, it remains to be seen how the possibility of their analysis is used in innovation research. It is of interest which methods and data are used to address different research questions.

Textual Data of Trademarks allows for the combination of trademarks with other data sources, thereby providing a broader perspective on innovation in general through the inclusion of services and low-technology innovation in the analyses. However, it needs to be made clear where trademark texts differ from, for example, patent texts and, thus, where they add value. Patents could provide the inventive step, while trademarks ensure market application. By that, the coverage of innovation could be ensured.

Innovation in Textual Data of Trademarks is not ensured. Some trademarks are linked to innovation, but not every trademark is. Therefore, it is still being determined how this affects the analysis of trademark text data and whether trademark textual data can provide interesting insights into activities in innovation areas, even if only to shed light on diffusion aspects.

Overall, the thesis addresses these aspects and intends to keep the strength of trademarks while simultaneously addressing the limitations of this data source using textual data. To build an understanding of trademarks and, in that context, textual data analysis, the thesis is structured as follows: In Chapter 2, the general background on innovation, patents, trademarks and designs as part of intellectual property rights and innovation measurement is given. Further, textual data analysis in general and with a focus on the main method used in this thesis is presented. This knowledge serves as a general foundation for the thesis. Following the Foundations, Chapter 3 focuses on the current use of textual data for innovation research in general. This literature review provides a general understanding of the state of the art of textual data, its analysis, reasons to apply textual data analysis and potential application areas in innovation research. Chapter 4 and Chapter 5 then perform textual data analyses of trademarks in combination with other data sources. Chapter 4 focuses on how trademarks and patents can be effectively combined to give a broader picture of innovation. It assesses the textual combination and how aspects of innovation are covered in each data source but with a focus, especially on trademarks. The underlying assumptions of this chapter are that insights on trademarks of structured and text-based data differ and that textual data of trademarks generally contribute to a broader understanding of innovation. The analysis is performed on Robotics and Footwear as representatives of a high-technology and a low-technology, respectively. The insights gained serve as a foundation for further analyses of innovation based on textual data of trademarks. By contrast, Chapter 5 combines patents, trademarks and designs to cover the transformation of the musical instrument sector from low technology to high technology and introduces a focus on firms. Even though one would expect an increase in patent applications in the sector due to the increasing relevance of electronic instruments, trademarks are the relevant intellectual property right in the sector, not only for classic but also for electronic musical instruments. The analysis contributes to a better understanding of this phenomenon by providing insights into the intellectual property rights application in the context of technological transformation and firm background. The thesis concludes in Chapter 6, where general insights are reflected and discussed, especially concerning textual data of trademarks, and the main conclusion of the thesis is drawn. In general, the work contributes to a better understanding of textual data of trademarks. It sheds light on textual data in innovation research as a whole, the contribution of trademarks combined with other data sources and uses the insights gained to answer economic questions.

For the general structure of the thesis, it can thus be said that the chapters start with a general perspective on textual data and become more specific over the course of the thesis: The thesis focuses first on textual data in general, then looks at textual data of trademarks in comparison to patents before finally applying the knowledge gained to assess technological transformation concerning intellectual property right use, covering patents, trademarks and designs. The structure of the thesis is graphically represented in Figure E.2. Here, the commonly coloured shape contours of the boxes indicated shared aspects between the different chapters of the thesis. The legend can be found on the left.

Innovation Areas/ Sectors: All chapters cover different aspects of innovation (yellow). The foundation's section provides the general aspects of innovation and its measurement. The review in Chapter 3 is not restricted to an innovation area or sector. It is only relevant that the articles in the sample are published in a business or an economic area, cover innovation and use textual data. Chapter 4 compares Robotics as a high-technology sector with Footwear as a low-technology sector to gain generalisable insights. Lastly, Chapter 5 looks at the transformation of musical instruments, which covers aspects of low- and high-technology applications due to classical and electronic musical instruments.

Intellectual Property Rights: Intellectual property rights (violet) with a focus on trademarks are present across all parts of the thesis. The foundation's chapter presents the background of intellectual property rights and especially patents, trademarks and designs and the measurement of innovation, as well as an introduction to the textual data analysis method applied in this thesis. In Chapter 3, intellectual properties are only indirectly considered as part of the textual data sources being used in some of the analysed publications. Trademarks and patents are then closely assessed in Chapter 4 and Chapter 5 with the addition of designs in the latter. However, the main focus of the thesis is on trademarks.

Textual Data: Textual data (green) and its analysis are relevant in all chapters. The analysis of textual data is described in the foundation's chapter. This thesis uses Structural Topic Modelling developed by Roberts et al. (2019). Topic Modelling is a text analysis technique to derive topics out of a large number of text documents. Structural Topic Modelling is a specific topic modelling approach that identifies topics in existing documents under the consideration of additional information (Roberts et al. 2019) (see further details in Section 2.4). Chapter 3 analyses published articles using textual data in innovation research. The articles in the sample are not restricted to a specific data source. Chapter 4 then analyses patents and trademarks, while Chapter 5 additionally includes designs into the analysis.

Figure E.3 provides an integrated perspective on the chapters. In the course of the thesis, the innovation areas covered are narrowed down to increase the focus, level of detail, and the integration of high- and

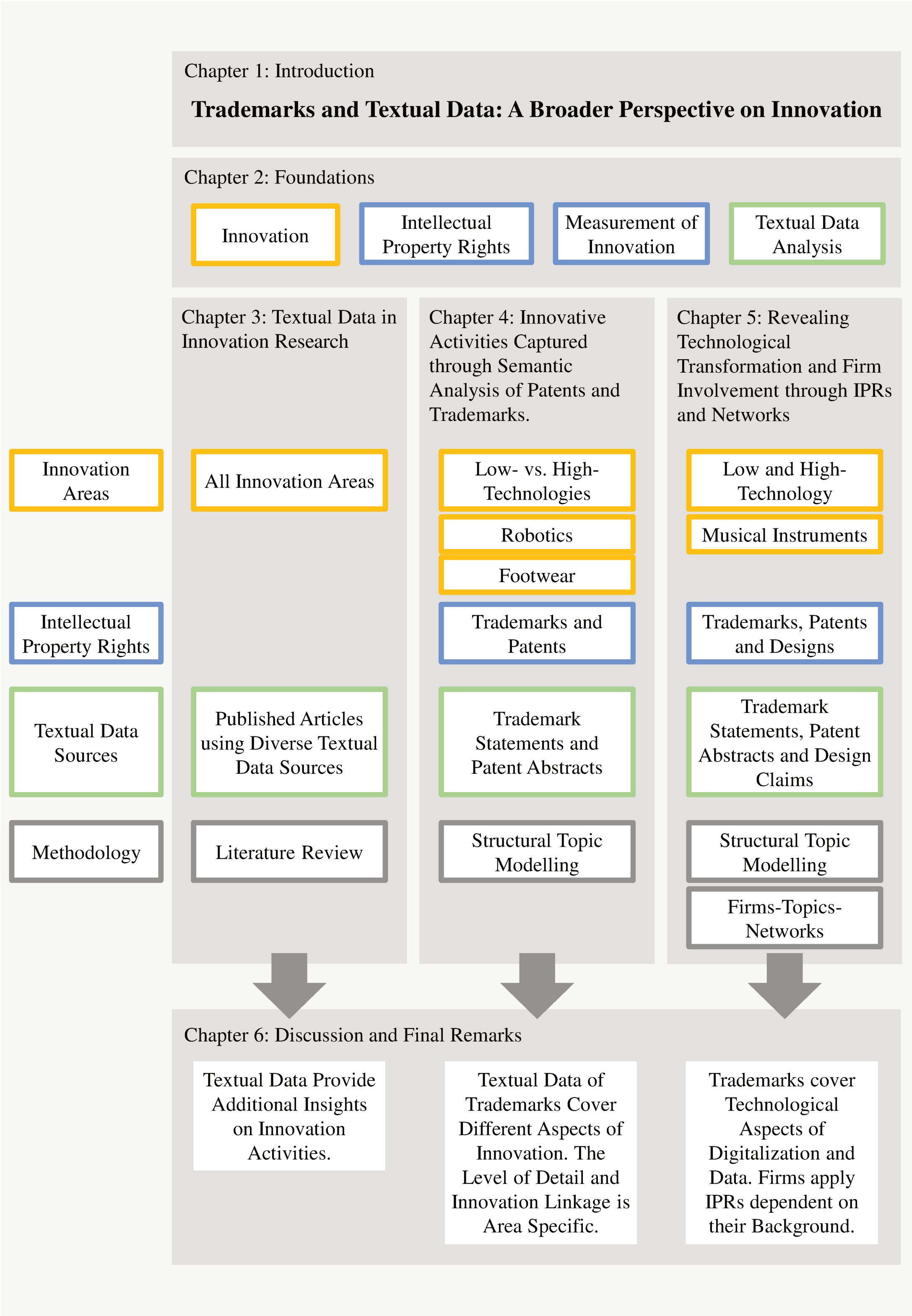


Figure E.2: Structure of the Thesis.
The boxes on the left indicate the meaning of the shape contour colouring. Grey relates to the methodology, violet to aspects of intellectual property rights, green to textual data sources used for the analyses and orange to innovation areas covered.

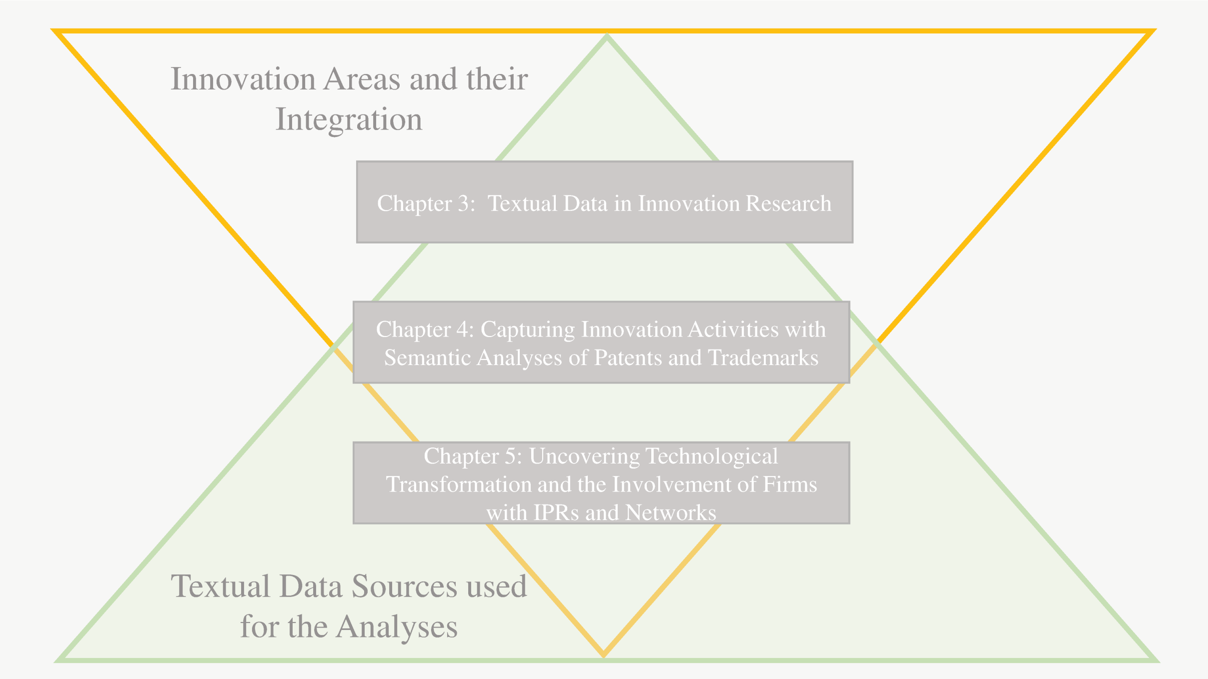


Figure E.3: Integrated Perspective on the Research Chapters.
The triangle in yellow represents the breadth of innovation areas covered. The triangle in green stands for the breadth of data sources used.

low technology (triangle in yellow). The areas are broadest in the literature review of Chapter 3, where all kinds of areas in relation to innovation and textual data are assessed. In Chapter 4, high-technology Robotics and low-technology Footwear are analysed as innovation areas. In contrast, high- and low-technology aspects in Musical Instruments are jointly considered in Chapter 5. The triangle in green represents the breadth of data sources used for the analyses in each chapter, which increases during the thesis. In the beginning, publications are considered, followed by a joint analysis of trademarks and patents, and concluding with a joint analysis of patents, trademarks and designs.

The remainder of the thesis goes as follows: First, further information on innovation, intellectual property rights, innovation measurements and textual data analysis is provided in Chapter 2. In Chapter 3, insights on textual data in innovation research are provided before performing different analyses on textual data of trademarks in Chapters 4 and 5. Finally, the findings of the thesis are summarised, and general topics are discussed before closing with the main contributions of the thesis in Chapter 6.

E.3 Summary and Contribution

The thesis covered various aspects of textual data of trademarks in the context of innovation research. Before concluding this thesis, a summary of each chapter is provided in Section 6.1 followed by a discussion. The discussion focuses on insights gained on textual data of trademarks from the joint consideration of the different chapters. The thesis concludes in Section 6.3.

The thesis evolves around the use of trademark textual data in innovation research. To capture innovation broadly and especially cover service innovation and innovation in low-technology areas, this thesis examined trademarks as an innovation indicator. To combine trademarks with other data sources and to increase the level of detail of the analyses, textual data of trademarks were assessed.

After presenting background information on innovation, intellectual property rights, the measurement of innovation, and the analysis of textual data in Chapter 2, Chapter 3 provides a general perspective on the use of textual data in innovation research. Chapter 4 then compares patents and trademarks from a textual data perspective in Robotics and Footwear. The data sources are evaluated against their coverage of products, services and technical innovation. Chapter 5 applies the insights gained in the previous chapters and uses them to measure technological transformation based on the textual data of patents, trademarks and designs in Musical Instruments. It further provides information on the involvement of different firms in the transformation.

Further detail on each chapter is provided in the following:

Chapter 3 provides a general overview of the current use of textual data in innovation research. Of special interest are the motivation to apply textual data analysis, the data sources used, the methodology applied, and the research questions answered based on textual data. The chapter contributes to a better understanding of textual data in innovation research and the different methodologies applied. In the context of the thesis, an overview of text-analysis methods are given, which help to build an understanding on text-analysis. For this purpose, articles from the fields of economics and business that contain text and innovation in the title or abstract are considered. The data set is successively narrowed down based on the authors' keywords provided and an intensive analysis of the abstracts. The final sample consisted of 23 articles that are related to innovation and textual data.

Most of the articles were published in known innovation policy journals, and a third was in a subject-specific field. Most authors appreciate the possibility of analysing large textual data sets in an objective, data-driven and time-efficient way. Textual data serves as a data source for time-based analyses and to detect opportunities or uncover hidden information. Primarily patents and publications are used as data sources. Other data sources are news articles, social media or firm documents. Most articles in the sample focus on an individual data source, while six articles combine several data sources to address their research questions, including insights about, e.g., firm culture, customer feedback, and public reaction in innovation analyses. Either patents or publications are always included in the combinations. Eleven articles apply clustering algorithms, especially for exploration and topic identification, while three apply classification methodologies to identify ideas or innovative firms. For clustering, especially Latent Dirichlet Allocation (LDA), which is a topic modelling approach (Blei et al. 2003) is used. Various research questions are covered: A main research area addressed is discovering and exploring technologies, mostly based on topic modelling. Further areas are industry convergence and regions, innovation trajectory and diffusion, and the emergence of novel topics or idea generation.

The chapter provides an understanding of the current state of textual data analysis. This contributes to the application of textual data analysis in innovation research. It highlights potential research areas that benefit from the use of textual data: Text-based analyses enable the analyses of text data sources and the combination of different data sources via language. Due to their reduced expert involvement, objectivity increases. Analysing text-based data becomes faster and easier with the use of text-based methods, allowing for regular repetition. Methodologies can be used for a variety of objectives. For the most part, they are used to identify topics or emerging trends and make them visible. Overall, the review reveals that research on textual data in innovation can provide an additional perspective. So far, textual data analysis is still mostly descriptive, but some concepts to use the information gained in further analysis exists.

Chapter 4 combines textual data of patents and trademarks and assesses how aspects of innovation are covered in each data source, focusing especially on trademarks. The chapter contributes to a better understanding of the application of textual data of trademarks and the combination of the data source with patents. This enables further analyses of innovation based on trademark textual data. The chapter addresses two research questions:

RQ4.1: Can trademarks and patents be combined via their textual data on a detailed level?

RQ4.2: Does the textual combination of trademarks and patents add new insights to the discussion of innovation?

To answer the research questions, trademarks and patents are combined textually with Structural Topic Modelling of Roberts et al. (2019). Robotics and Footwear as representatives of high-technology and low-technology are analysed, respectively. The results are analysed in terms of textual combination and insights gained. Therefore, general assumptions are drawn from the literature on innovation in trademarks, markets, services, and technology.

The model estimation is used to analyse Robotics and Footwear. In Robotics, for example, 115 topics are estimated based on the data set provided to the model. As can be seen in Figure E.4, the topics have different occurrence probabilities in each data source. The green bars display topics that have a significance for occurring in patents, while blue bar displays topics with a trademark significance. Gray bars do not have a clear occurrence in either data source. In the course of the analyses, the different topics in trademarks, patents and neither of both are assessed. The internal consistency of the topics is thereby considered, which implies that the general topic words and the documents that are related to the topic should be understandable for the human observer. The topic relation to official classification systems as well as the importance of the topic over time is also considered. The analysis is repeated for Footwear, before the results are combined and assessed in relation to the research questions.

The first question covers the possibility of combining patents and trademarks textually:

RQ4.1: *Can trademarks and patents be combined via their textual data on a detailed level?*
In terms of combination and consistent topics, the analysis displayed mixed results: In Robotics, the textual combination of trademarks and patents displayed consistent results. They were in line with the classifications of patents and trademarks but provided a more detailed perspective. Patents highlighted the invention and technical aspects, while trademarks added details and mainly covered the application areas. For trademark analysis, the approach further connects the trademark topics to the patent classification system, allowing for a better interpretation of the trademarks. Overall, the textual data of trademarks enhanced the analysis and provided information on additional areas. Footwear is a low-technology area with fewer inventions and technological applications. This could be observed in the textual analysis: Only a small number of patents compared to trademarks was available. Further, the trademarks mainly focused on several clothing items instead of innovative technologies. Trademarks covered only a little information on inventions to generate consistent topics; thus, the combination of patents and trademarks was only partially consistent. The results in Robotics and Footwear thus provided different results. It became apparent that the textual structure of the trademarks differed between Robotics and Footwear: In Robotics, the trademark descriptions are generally more detailed with information on the product, the purpose of the robot and aspects of protection. Further, an increasing level of detail is observable over time. The trademark descriptions in Footwear are less specific and detailed than those in Robotics. Trademark descriptions of topics with consistent results in Footwear displayed a similar textual structure to those in Robotics. The level of detail available supports the model in its calculations. Generally, the level of detail available in trademark descriptions differs depending on the innovation area looked at and the time covered. Trademarks with a higher degree of technical information contain similar information as patents, while trademarks with less technological and innovation aspects diverge from patents. Trademark texts thus capture innovation activities and provide a detailed perspective. However, the level of detail is sector or innovation-area-specific. When trademarks' textual data are used to analyse innovation, this should be considered.

RQ4.2: *Does the textual combination of trademarks and patents add new insights to the discussion of innovation?*

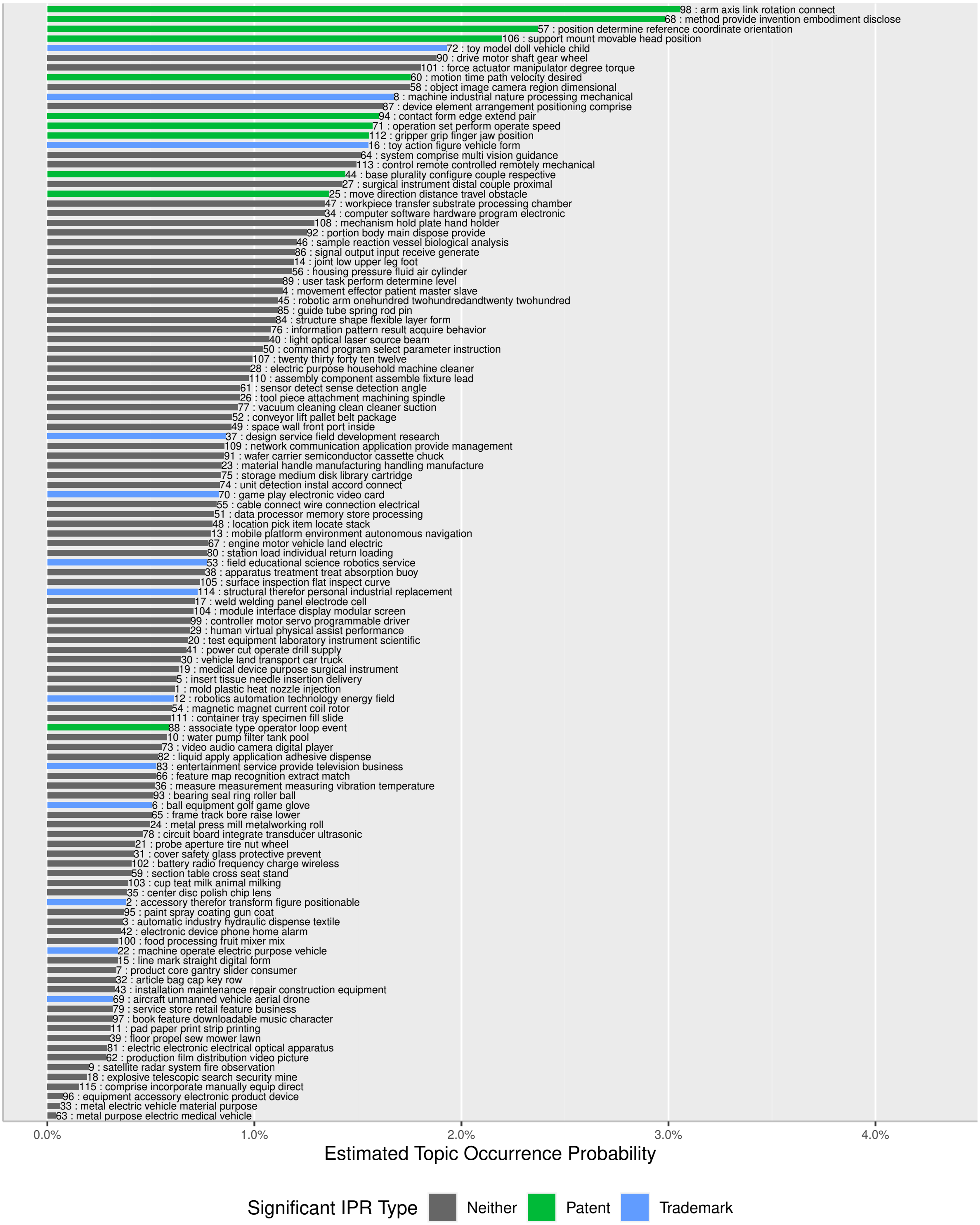


Figure E.4: Overview of the Model with 115 Topics in Robotics.
The figure provides an overview of the 115 topics in Robotics. The topics are displayed according to their relative share in the overall data set and ordered decreasingly. The colouring indicates the document type, which is significant for the topic, with a 95% confidence interval (Figure C.2). In case of no significance, the colouring of the bar is kept in grey. The legend on the right displays the topic number and the five words with the highest occurrence probability in the topic.

The second question focused on the added insights due to the combination. General assumptions about the data sources are extracted from the literature to assess the added insights. These are that trademarks cover innovation, more service or market-related innovations, and patents provide information on technical innovation. These assumptions are challenged as not all results might be reproducible based on textual data. For example, a difference in service coverage might not be observable textually. Service-related information might also be covered in patent textual data, or no further description of the services might be provided in the trademark texts. Therefore, it is necessary to understand the contribution of textual data of trademarks in comparison to patents. The different assumptions are evaluated against the results of the textual data analysis and reveal the following:

Trademarks and Innovation: The link between trademarks and innovation is observable in the textual data. However, a certain technological degree is required to combine trademarks and patents and to generate topics in trademarks where additional information on innovation is available.

Market and Products: In Robotics, trademarks ensure the market perspective of the inventions covered in patents. The market and product application were also covered in patents but not as the primary focus. Patents helped to interpret and bundle trademarks, as the patent classification system provided more structured details than the trademark classification system. In Footwear, trademarks only provided the market introduction perspective in the case of consistent topics. The trademarks were mostly focused on final products without further details on their invention. Here the advantage of the combination and textual data of trademarks for innovation research needs to be questioned. In general, trademarks cover more final products and the market perspective. They can add the market and product perspective in the analysis of innovation, while patents focus more in detail on the inventions behind them.

Services: Insights on services could be gained from both data sources, with a higher representation in trademarks. Services like entertainment or education are more in trademarks. Services with a technical component are also observable in patents like service robotics or footwear subscriptions, which is in line with the findings of Blind et al. (2003). The inclusion of trademarks provided additional insights into services and highlighted the service areas in Robotics or Footwear.

Technical innovation: The patent topics in Robotics and Footwear covered technical aspects and focused on functions or components. In Robotics, trademarks contribute detailed technical information on various topics independent of the significant IPR. In Footwear, the technical details are provided mainly by patents and rarely by trademarks. Generally, patent descriptions still provide more technical information than trademarks. Trademark descriptions, nevertheless, cover technical aspects, the extent is just highly dependent on the field of innovation looked at, and the level of detail is lower than patents.

The literature's assumptions, thus, in general, also apply to textual data of trademarks and patents. In terms of additional insights, trademarks can thus contribute to innovation coverage. The results' quality depends highly on the application area analysed and the trademarks covered in the data set.

Overall, the combination of trademarks and patents revealed that the data sources can be combined on a detailed level and that the combination adds insights into the innovation discussion. As the field of analysis highly impacts the quality of the results, the field of analysis needs to be considered. The chapter contributes to a better understanding of textual data of trademarks as an innovation indicator – in general and for specific innovation areas of low and high technology. Further, the chapter highlights the potential of the combination of trademarks and patents and the differences

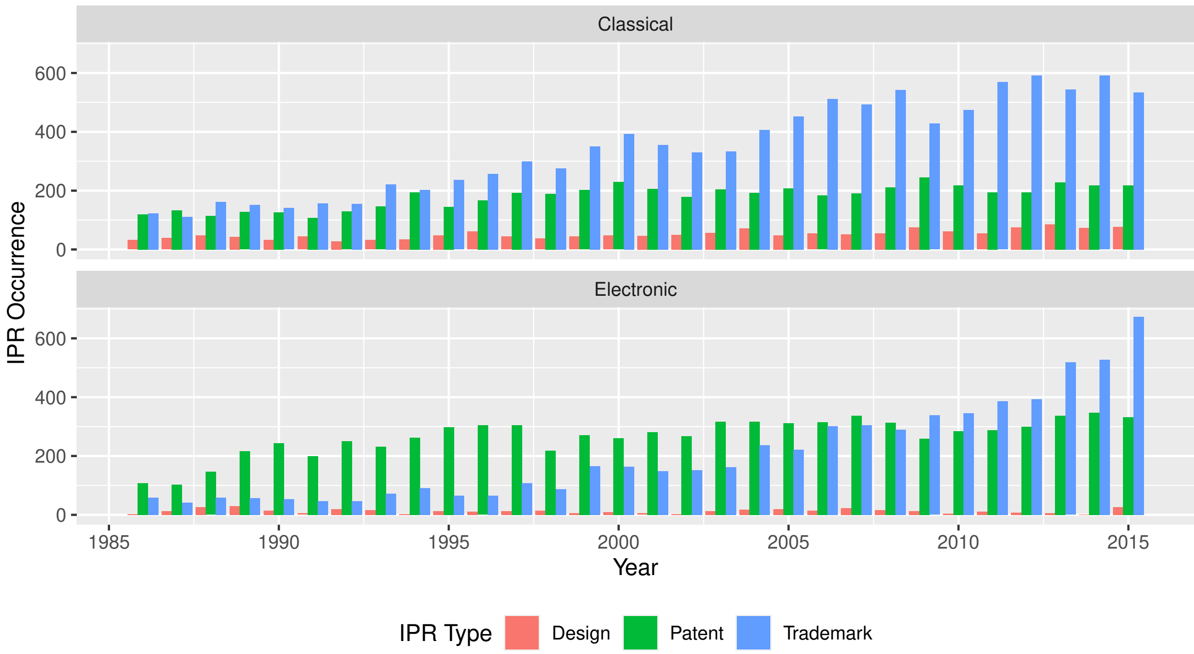


Figure E.5: Musical Instrument IPR Registration per Year and Musical Instrument Type. The figures display the yearly development of the document registrations, differentiated for classic or electronic musical instruments.

between these two textual data sources in innovation. The insights gained serve as a foundation for further analyses of innovation based on textual data of trademarks, especially in joint analysis with patents, which can be used to improve existing innovation measurements.

Chapter 5 assesses the technological transformation in musical instruments and the involvement of major firms in the transformation with joint textual analysis of patents, trademarks and designs. The musical instrument sector combines low-technology and high-technology aspects. It is a sector with, on the one hand, a long tradition and low-technology products such as classical musical instruments. On the other hand, it is a sector undergoing a technological transformation through the electrification and digitalisation of musical instruments. Even though musical instruments are becoming increasingly technical, trademark registrations surpass patent and design registrations (see Figure E.5).

In musical instruments, patent, trademark and design protection is applied. To understand the transformation, the involved firms and the use of intellectual property rights, the chapter combines textual data of patents, trademarks and designs in musical instruments to attain a broad coverage of technological transformation and to assess a change of use of the IPRs concerning this transformation and involved firms.

Thus, the analysis provides a broader perspective on the transformation, firms and the application of different intellectual property rights (IPRs). The analysis is separated into two parts: Part I focuses on the sector in general, and part II includes major firms in the discussion. The related questions are:

RQ5.1: What is the nature of the sector transformation from a low-technology sector to a sector with low- and high-technology?

RQ5.2: How does the transformation relate to IPR use?

The related research questions from the firm perspective are:

RQ5.3: Who contributes to the transformation?

RQ5.4: How does this relate to IPR use?

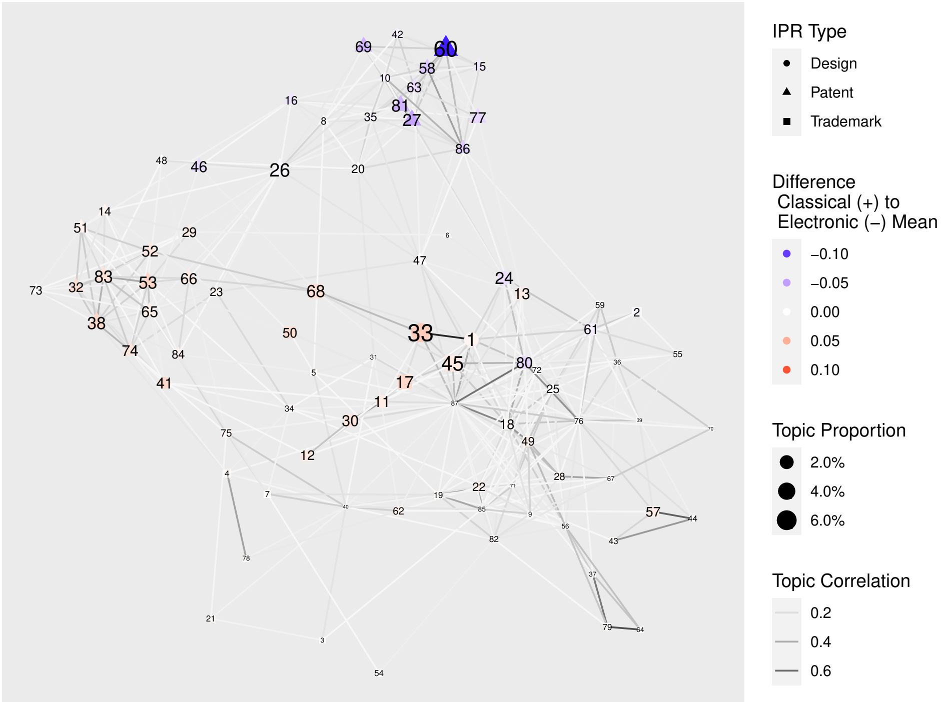
The different intellectual property rights serve as a basis for the analysis. Their textual data provide content-based insights into the transformation. For this purpose, the data sources are combined textually, and the topics of the data set are extracted with Structural Topic Modelling of Roberts et al. (2019). Based on the topic estimation, networks can be formed with the topics as nodes and the topic co-occurrences as linkages. The networks are analysed for patterns that are related to the change taking place. In the sector analysis, the relevance of IPR use, different musical instrument types and classical versus electronic musical instruments are assessed, and changes over time are considered. From the major firm perspective, the topic network is then enhanced with firms, building a firms-topics-network.

The sector perspective reveals that the musical instrument sector has transformed from the provision of classical musical instruments to electronic instruments and digitalisation. For example Figure E.6a displays the topic network from 1986 to 1990 from a sectoral perspective, while Figure E.6b displays the situation from 2011 until 2015. As can be seen, the topics in classical musical instruments decreased from 59 to 46 nodes, while the electronic topics increased over time. Electronic topics are covered by patents as well as trademarks. Designs connect the patent and trademark topics. The topics covered change towards digital aspects like sound creation, mobile and gaming. With technological transformation, the use of intellectual property rights has changed. Each IPR type provides different information: Electronic topics or data-related topics are relevant in patents but are increasingly also in trademarks. Designs are an intermediate between patents and trademarks and often cover similar topics as trademarks. The IPR use has changes over time, with trademarking becoming more important in electronic musical instruments and covering the digital aspects of digital sound creation, mobile and gaming.

The firm perspective focuses on the major firms in five-year intervals from 1986 until 2015 and categorises them according to their background. These firms are the major applicants of intellectual property rights across all IPRs. Figure E.7 displays the 87 topics in relation to the major firms from 1986 until 2015. The analysis reveals that the background of firms contributing to the transformation changes over time, with firms from various backgrounds contributing to the sector's transformation. Mainly musical instrument producers were active in the beginning. Over time, hardware and software firms, retailers and game developers entered the sector. Their topics were predominantly related to computation, signalling, voice capturing and data processing. The transformation and technological change is thus driven also by firms that are not directly involved in musical instrument production. The IPR use overlaps with the firm background. Software and hardware firms are more related to patent topics, while gaming and retail firms are close to trademark topics. The positioning of musical instrument producers depends on the degree of electrification of their instruments. Some are positioned closer to patent topics, while most are in the trademark or design topic domains. The results are in line with the sector perspective.

Overall, the electrification and digitalisation of musical instruments is a combined effort of hardware, software and gaming firms leading to new developments in the sector and musical instrument producers that introduce electrification to their instruments. The chapter shows that technological transformation is captured via the combined use of different intellectual property rights. Especially the inclusion of trademarks adds additional aspects like gaming to the analysis. The use of text-based analysis enabled the discovery of topics that are related to the transformation and different firms. This enabled a differentiated analysis of the transformation. Method-wise, the firms-topics-network provides a better understanding of firm positioning in general. The chapter contributes to a better understanding of intellectual property rights application in the context of technological transformation and involved firms. A technological transformation is taking place in

(a) Network of Musical Instrument Topics, Differentiated for Electronic and Classic Instruments from 1986 until 1990



(b) Network of Musical Instrument Topics, Differentiated for Electronic and Classic Instruments from 2011 until 2015

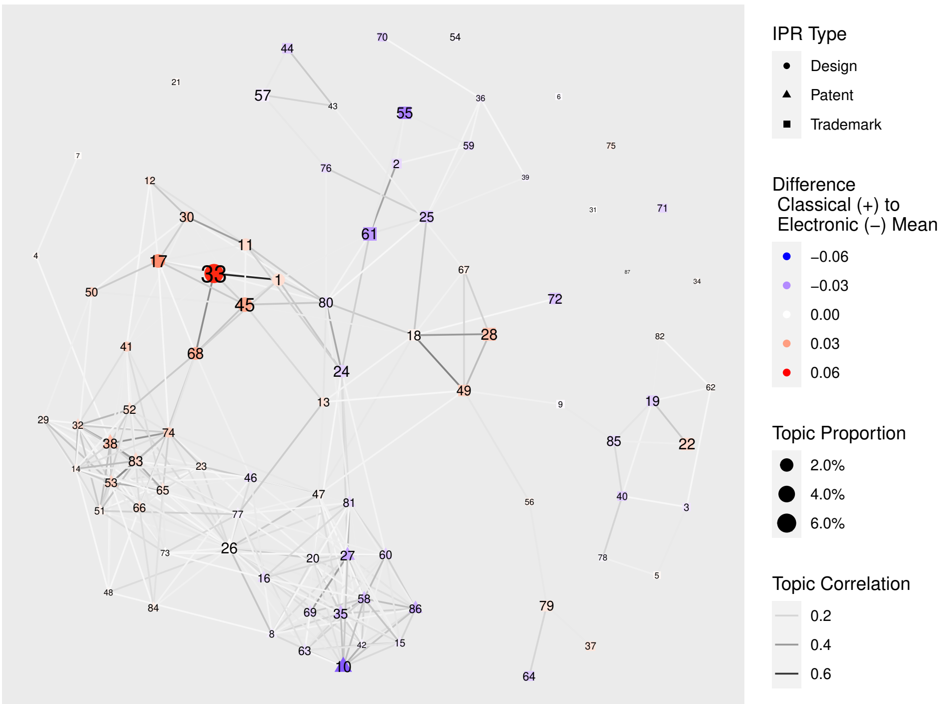


Figure E.6: Network of Musical Instrument Topics, Differentiated for Electronic and Classic Instruments.
The networks display the topic relation from 1986 until 1990 and 2011 until 2015, differentiated for IPR types and electronic versus classical musical instruments. Nodes represent the topics, while links represent the likelihood of these topics to co-occurrence. The node size represents the topic proportion. The IPR types are indicated via the shape of the nodes, while the musical instrument information is given by the colouring of the nodes. An intensive blue means that the topic is strong in electronic musical instruments and vice versa for classic musical instruments in red.The edges display the strength of correlation. Lighter colours have a lower correlation than darker colours. In Figure E.6a, 59 nodes are more in classical musical instruments compared to 28 nodes in electronic topics. In Figure E.6b 46 nodes are more in classical musical instruments compared to 41 nodes in electronic topics.

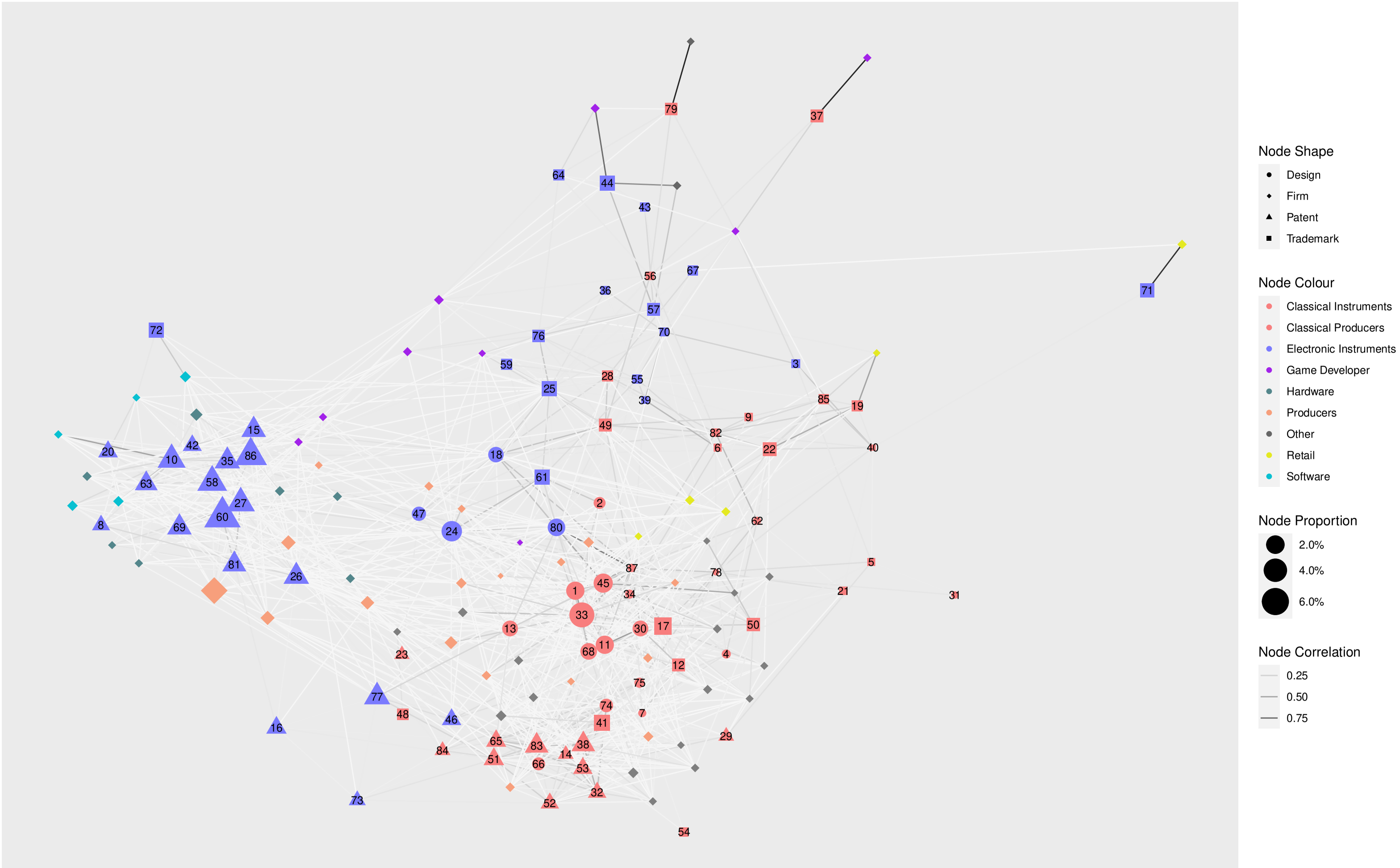


Figure E.7: Major Topics-Firms-Network from 1986 until 2015.
The network displays the major firms in the field of musical instruments from 1986 until 2015. Nodes relate to firms and the different intellectual property rights, here, design, patent, and trademark. Topic nodes always have numbers inside, indicating the topic number. The colouring of topic nodes is either red or blue, depending on the relation to electronic or classic musical instruments. The shape and the colouring are determined by the strongest IPR type or musical instrument type in the data set. The firm nodes have a rhombus shape and are coloured according to their firm type. The overview provides only the dot's colourings and not the naming of the firms for clarity. The edges show the topic correlation from firms to topics and from topics to topics, with a lighter colouring indicating a lower correlation.

musical instruments and across other industries, especially fuelled by digitalisation. The methodology of the chapter can be applied to understand the transformation of different sectors better, as not only high-technology aspects but also low-technology aspects are covered, and to assess firms on a more detailed level.

E.4 General Remarks on the Thesis

Each chapter of the thesis thus focused on different aspects. However, the chapters are connected by various elements. All chapters cover innovation and different innovation areas (see Figure E.8). The breadth of innovation areas (see yellow triangle) covered is broadest in Chapter 3, as here, all areas of innovation are considered under the condition that textual data are used for the analysis of innovation. Chapter 4 covers Robotics and Footwear, whereas the former represents a high-technology area and the latter represents a low-technology area. In Chapter 5 the Musical Instruments sector is assessed as a sector in which a technological transformation takes place and where aspects of low- and high-technology are relevant. In parallel to a decreasing breath of innovation area, the integration increases. In Chapter 4, the high- and low-technologies are analysed individually and then compared for Robotics and Footwear, while low- and high-technology aspects are jointly analysed in musical instruments. The data sources used (see green triangle) during the thesis become more diverse. At first, the research is based on a sample of publications related to textual data in innovation research. In the following Chapter 4, patent and trademark textual data is used as the basis for analysis and the textual content of the different data sources are compared to each other. Finally, in Chapter 5, patents, trademarks and designs are jointly used to provide broad coverage of the innovations in musical instruments and the topics derived from the analysis are used to identify general patterns. Figure E.8 provides an integrated perspective on the chapters. The joint aspects are coloured accordingly.

E.5 Discussion

Considering the chapters together, some general aspects remain to be discussed. The thesis started with the need for a broader innovation perspective to cover low-technology areas and services not covered in patents. Trademarks were treated as an alternative data source to provide this perspective. The combination of trademarks with patents and designs was proposed based on textual data, assuming that these provide the potential to combine several data sources and a broader perspective on low-technology innovation and services. The textual data were preprocessed and combined throughout the thesis with Structural Topic Modelling. The analyses revealed several strong points of the approach but also challenges that still need to be addressed:

Trademarks: Trademark data provide not only information on services but also technological areas. This became obvious in Robotics, where trademarks also contained detailed information on innovations. In Musical Instruments, trademarks introduced aspects of digitalisation like mobile, digital sound design and gaming. Interestingly, the firms' background is more decisive for the IPR application than the technological aspects. One would have expected all high-technology aspects to be found more in patents. However, software and hardware firms were found to be active in patents, while firms with gaming backgrounds were active in trademarks. Here, further research is needed on the differences that explain this behaviour. A reason could be the speed of development in the gaming sector and the importance of marketing and customer closeness.

Textual Data: The textual data enabled the analyses of various innovation areas in a data-driven way. Large data sets could be covered, and trends were detected without the involvement of experts. Trademarks, patents and designs could be combined and jointly analysed. The focus on the

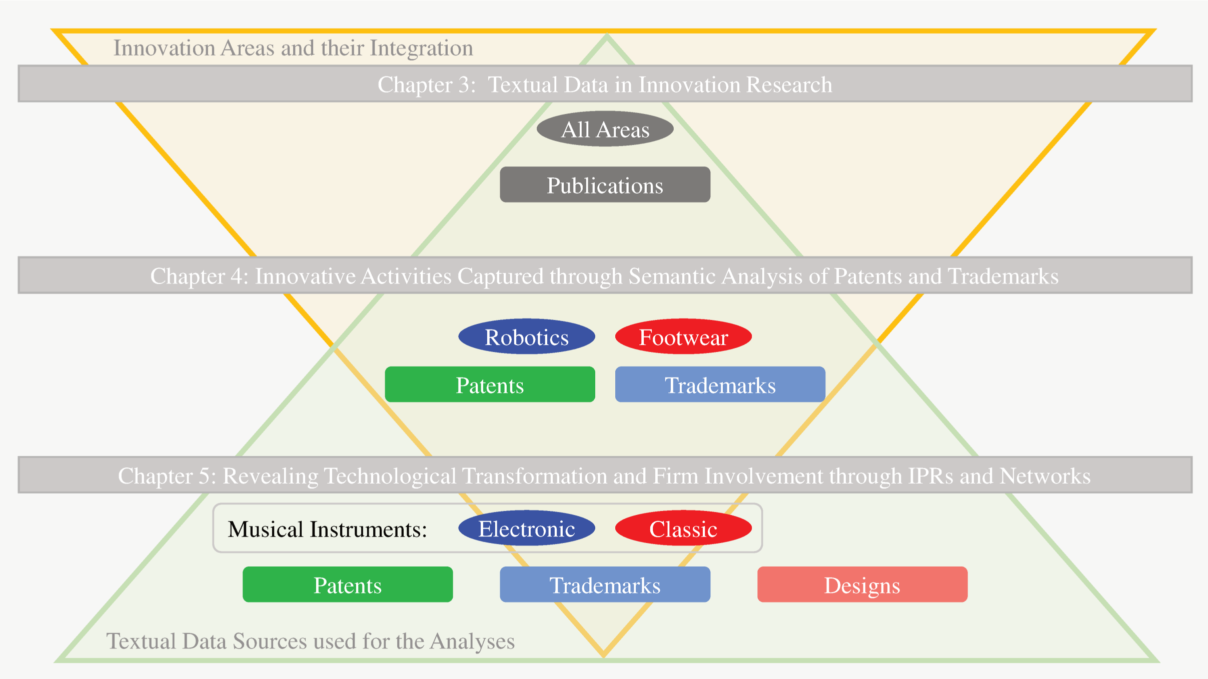


Figure E.8: Summary of the Integrated Perspective on the Research Chapters
The triangle in yellow represents the breadth of innovation areas covered. The triangle in green stands for the breadth of data sources used. The dark grey colour of the boxes in Chapter 3 indicates that no specific area of interest is covered. The literature review covers a sample of publications. However, the textual data applied in these publications use textual data from various data sources. The oval boxes are either coloured dark blue or red. Dark blue relates to high-technologies, while red relates to low-technology aspects. The intellectual property rights have a violet shape contour. Patents have a green filling, trademarks a light blue filling, and designs a light red filling. The colouring of the boxes was maintained throughout the thesis in terms of patents, trademarks, designs, high-technology, and low-technology to enable a common understanding throughout the chapters.

textual information thereby contributes a broader perspective, as seen in the analyses of Musical Instruments. The textual data provided information to explain the, e.g., use of trademarks in electronic musical instruments. Also, the approach enabled the thematic positioning of firms that displays their involvement in the transformation. Over time, it became evident that firms with different backgrounds became involved, which then introduced new topics to the transformation of musical instruments.

Through the preprocessing and analyses, it became evident that the textual data of different intellectual property rights differ, especially in the dimensions of legal and common words used, sentence structure and level of detail. These affect the ability to combine and compare the different data sources. The documents' IPR type was therefore considered during preprocessing to account for these differences: Highly common words specific to each IPR were considered and removed. Care was taken in the cleaning process to preserve the differences between the data sources. However, this resulted in words without meaning being included in the final data sets. In order to further improve the analysis of multiple data sources, a stop word list specific to each IP right could therefore be of interest, containing, in particular, the common legal terms of the respective IP right. As such, further focus on meaningful words could be ensured while considering content-related differences in the data sources.

An additional aspect to consider in the textual data analysis of the IPR documents is that the documents were not equally represented in the different innovation areas. In Robotics, most documents were patents, while in Footwear, most documents were trademarks. In the data set of musical instruments, patents and trademarks were equally represented, while designs were rarely included. The share of each IPR type in the data set impacts the topics estimated, as the topics are estimated based on the documents provided and thus lean towards the IPR type with the highest share. This means that the smaller the proportion of an IPR, the greater the uncertainties of the estimate concerning this IPR and, thus, the greater the variance. We have seen in the different chapters that as long as consistency and relatedness of the documents to the topics are ensured, the difference in the share of IPRs is, however, neglectable. In Robotics, the topics were very consistent. In Footwear, the topics with a high relation to patents were also consistent. In the case of trademark topics, the consistency depended on the structure of the trademarks. In cases where trademarks provided only items of clothing, the consistency was lower, and the topics provided less detail on the production or innovation contained. In musical instruments, topics related to hardware and software firms and mostly patents display a high consistency among the main documents of patents, trademarks, and designs. However, the consistency of trademark topics depends on the electronic aspects. Overall, the consistency and, thus, quality differs depending on the textual structure. A high technological relatedness often leads to a high quality of the topics. Additionally, the quality and use of the IPRs in the area covered need to be considered.

Finally, the text length must be taken into account. Trademarks contain predominantly short texts or concise word lists. Designs are also short in their textual description. Patents provide extensive descriptions or short abstracts. This thesis used the abstracts of patents to align with the textual length of trademarks and designs. This approach worked for the combination. For patents, it could be interesting to extend the analysis concerning the patent description to achieve an even higher level of detail.

Textual Data of Trademarks: The thesis focused on textual data of trademarks in combination with other data sources to gain a broader perspective on innovation. Among the articles in the sample of Chapter 3, no article used trademarks as a data source for textual data, which is in line with the results of others. However, throughout this thesis, it is shown that textual data of trademarks do contain information on innovation and provide an opportunity to combine trademarks with other data sources. This provides additional insights into firm positioning or innovations in different innovation areas.

The analyses of textual data of trademarks, patents and designs made it possible to discover more in detail about Robotics, Footwear and the transformation of Musical Instruments, especially in comparison to the available classifications. Nevertheless, the amount of text and kind of text structure is also decisive for the level of detail: Even though trademark descriptions are relatively short, differences in their structure exist. The trademark descriptions of Robotics or Musical Instrument trademarks covering high-technology aspects displayed a more extensive description and detail than trademarks in, e.g., the low-technology area of Footwear, where the trademark description resembled a list of goods without further detail. Here further research is needed to understand the reasons for this difference and potentially focus on trademark descriptions with a high level of detail.

The analysis of textual data of trademarks in this thesis was used to explore trends. The topic modelling approach supports this intention. However, trend exploration is only one aspect of textual data used for innovation research. More research needs to be done to include the insights gained thanks to textual data in existing economic models.

Innovation in Textual Data of Trademarks This thesis deals with the assumption that trademark texts represent interesting insights about innovations. This is despite the fact that the link between trademarks and innovation is not guaranteed. The analyses of Robotics, Footwear and Musical Instruments were able to show that trademarks provide detailed information about innovation comparable to patents. They could ensure the perspective of market introduction and diffusion. However, differences in the trademark texts were found to influence how well the texts of trademarks contribute to the understanding of innovation. This different structure could be related to the extent to which trademarks represent innovation and requires further investigation.

Overall, textual data of intellectual property rights provide interesting insights into the development of innovation. The thesis addressed textual data of trademarks and revealed their contribution.

E.6 Closing Remarks and Future Research

To conclude, the thesis adds to the discussion of textual data of trademarks in innovation research. It, therefore, analysed the current state of textual data analysis in innovation research, compared trademark and patent textual data and applied textual data from patents, trademarks, and designs to measure technological transformation. Overall, the textual data of trademarks enabled the combination of several data sources. The inclusion of trademarks captured service and low-technology aspects. The textual data of trademarks added aspects like digitalisation in musical instruments. The thesis contributes to a better understanding of textual data use, unlocks textual data of trademarks and applies textual data of patents, trademarks, and designs to assess technological transformation. Combining different intellectual property rights through textual data provides a broader perspective on innovation while overcoming existing limitations. The combination of trademarks, patents, and designs enhances the strengths of each data source for innovation research. The textual data further provides detailed information to understand technological development or the behaviour of firms. The thesis contributes to the area of textual data use for innovation research and especially the application of trademark textual data analysis. In a broader context, the thesis contributes to evidence-based policymaking by enhancing our understanding and measurement capabilities of innovation. This improvement aids in deriving more comprehensive recommendations for political initiatives, especially in low-technology innovation areas.

Further research is necessary. The challenges mentioned above remain to be addressed. In this thesis, the analysed patent, trademark, and design documents were identified using class-based and keyword-based searches. This is necessary as, across different data sources, no common classification system exists. At the same time, it must be ensured that the criteria for the documents to be included are similar across the data sources. In the case of more complex areas of interest, like green technologies in developing countries, a more sophisticated approach to identifying the relevant data across different data sources is of interest as the results are only as good as the data used. Further, the innovation linkage of textual data of trademarks could be improved: In areas with a high technological relation, the trademarks in this thesis displayed a high level of detail. However, in areas with low-technology intensity, like clothing, the trademark descriptions covered mostly a list of clothing items. Here, an approach to differentiate between relevant and non-relevant trademark descriptions for innovation research might be of interest to increase the focus on innovation. Lastly, insights gained by the approach, like the background of firms being decisive for IPR use, remain to be included in further economic models to improve their results.

F Résumé Français

F.1 Introduction

L'innovation est un moteur fondamental de la croissance économique et du développement social. Les entreprises sont désireuses d'innover et d'être innovantes pour rester compétitives (Schumpeter 2010). Les innovations majeures comme la machine à vapeur ou les semi-conducteurs ne sont que quelques-unes des innovations qui ont eu un impact majeur sur le développement de la société (Kuznets 1973). Ce ne sont pas seulement les grandes mais aussi les petites innovations qui façonnent nos vies. Des progrès récents peuvent être observés dans la fourniture de services où la numérisation permet la diffusion de musique ou de vidéos, le covoiturage, la banque en ligne ou le commerce électronique. L'innovation a un impact sur la croissance économique, le progrès technologique, les investissements, la productivité et l'allocation des ressources. Elle peut prendre diverses formes, qu'il s'agisse d'innovations technologiques, de l'introduction de nouveaux biens ou de nouveaux services, d'une amélioration des processus ou d'un changement organisationnel. L'introduction sur le marché est un aspect essentiel d'une innovation. L'innovation étant importante pour le bien-être d'un pays et la compétitivité des entreprises, les responsables politiques du monde entier s'efforcent de rendre leurs régions innovantes, dans l'espoir de stimuler la croissance économique et le développement social. La politique doit donc être fondée sur des preuves. La recherche sur l'innovation vise à fournir les preuves nécessaires pour guider les décideurs politiques dans la prise de bonnes décisions.

La plupart des données sur l'innovation et la politique d'innovation sont basées sur les données relatives aux brevets. Les brevets sont un moyen pour les entreprises de protéger leurs inventions de l'imitation et d'obtenir un droit de monopole sur leur invention (Neuhäusler 2009). Les brevets sont considérés comme un mécanisme de protection formel et font partie des droits de propriété intellectuelle (DPI) (WIPO 2020d). Pour obtenir une protection par brevet, il faut une activité technologique inventive et le brevet doit décrire publiquement l'invention (WIPO 2021h). La protection par brevet est disponible dans le monde entier. Les brevets sont systématiquement enregistrés et couvrent de longues périodes (Basberg 1987). Le système des brevets fournit en outre un système de classification sophistiqué qui simplifie les recherches de brevets et permet des recherches détaillées. Cela rend les brevets intéressants pour la politique d'innovation, car ils constituent des sources de données qui révèlent des informations sur l'état actuel de la technologie à l'échelle mondiale (Kleinknecht et al. 2002). Dans la recherche sur l'innovation, les brevets ont tendance à être utilisés comme un indicateur de l'innovation de produit (Dziallas and Blind 2019).

Même si les brevets présentent des avantages en tant qu'indicateur, ils ont aussi quelques défauts :

- Tous les brevets ne couvrent pas une invention commercialisée. Cela signifie que l'invention n'est pas introduite sur le marché. Par conséquent, ces brevets ne couvrent que les inventions et non les innovations (Basberg 1987). Il n'est donc pas possible de distinguer quels brevets sont utilisés sur le marché. Pour la politique d'innovation, il est donc intéressant de couvrir l'aspect diffusion pour assurer une introduction sur le marché, car cela est nécessaire pour apporter des progrès et affecter la croissance économique.
- L'application des brevets diffère selon les secteurs et les entreprises. Comme les brevets exigent une activité inventive dans un domaine technologique, les secteurs à faible contenu technologique sont sous-représentés (Kleinknecht et al. 2002; Neuhäusler and Frietsch 2015). Cela ne signifie

pas pour autant qu'il n'y a pas d'innovation dans ces secteurs. Elle n'est tout simplement pas prise en compte dans les indicateurs d'innovation basés sur les brevets (Hirsch-Kreinsen 2008). La recherche a montré que les secteurs de faible technologie contribuent à la croissance économique (Mendonça 2009).

- Les brevets ont des difficultés à couvrir les services, les connaissances tacites ou les inventions de logiciels (Millot 2009; Mendonça 2009; USPTO 2017; USPTO 2015). Or, ceux-ci sont de plus en plus importants. Premièrement, les pays industrialisés ont d'importants secteurs de services qui dépendent de la prestation de services. Deuxièmement, la numérisation a un impact non seulement sur les secteurs technologiques, mais sur tous les secteurs. Il convient donc d'observer une évolution des activités d'innovation. Le fait de s'appuyer uniquement sur les données relatives aux brevets pourrait donner une perspective limitée. Les indicateurs basés sur les brevets pourraient avoir négligé les services mentionnés ci-dessus.

Les marques constituent une possibilité d'inclure ces perspectives et, en principe, de surmonter ces lacunes. Les marques peuvent prendre diverses formes, les plus connues étant les noms représentant une marque. Par exemple, une marque comme "Apple Music" est avant tout un signe ou une marque utilisée pour protéger le nom, mais elle peut aussi représenter l'innovation dans la fourniture de musique. Le nom nécessite le plus souvent l'enregistrement de domaines spécifiques qui indiquent l'usage de la marque pour obtenir la protection de celle-ci. Ainsi, la marque empêche toute confusion et assure le caractère distinctif entre différents produits ou fournisseurs de services. Comme les brevets, les marques sont également déposées, largement disponibles et couvrent de longues périodes (Flikkema et al. 2015), mais elles sont également applicables dans tous les secteurs et pour les produits et services (S. J. H. Graham et al. 2013; Millot 2009). Elles constituent également des mécanismes de protection formels et des droits de propriété intellectuelle (WIPO 2020d). Les enregistrements de marques ont augmenté ces dernières années et dépassent, par exemple, les enregistrements de brevets à l'Office des brevets et des marques des États-Unis (USPTO) (voir Figure F.1). Contrairement aux brevets, les marques nécessitent souvent une introduction sur le marché, ce qui permet une perspective de diffusion. Elles sont applicables dans les services et les secteurs de basse technologie. Dans l'ensemble, les marques commerciales présentent un intérêt pour la recherche sur l'innovation. Millot (2009), Mendonça et al. (2004), and S. J. H. Graham and Hancock (2014) suggèrent déjà leur utilisation en tant qu'indicateur d'innovation, soulignant leur capacité à couvrir non seulement l'innovation technique de produit mais aussi l'innovation de service, de marketing ou de processus commercial.

Néanmoins, les marques présentent également des lacunes en matière de recherche sur l'innovation :

- Le lien entre les marques et l'innovation n'est pas garanti. Contrairement aux brevets, l'activité inventive n'est pas une exigence lors de l'enregistrement d'une marque. Cela rend les marques plus complexes en tant qu'indicateur d'innovation. Plusieurs auteurs étudient donc le lien entre les marques et l'innovation : Seip et al. (2018) constatent que les marques sont enregistrées à différentes étapes du processus d'innovation, en fonction de l'innovation à protéger. Mendonça et al. (2004) affirment que les entreprises innovantes sont plus actives dans le dépôt de marques que dans le dépôt de brevets sur la base des résultats de la troisième enquête communautaire sur l'innovation (CIS 3). Flikkema et al. (2014) ont pu montrer dans leur enquête que plus de 50

- Un autre défaut des marques est leur granularité limitée. En ce qui concerne les brevets, le système de classification des brevets est très sophistiqué et traite de divers domaines d'innovation. Le système classe chaque brevet en fonction de son contenu technologique, ce qui permet des analyses détaillées de domaines spécifiques et des relations entre ces domaines (WIPO 2020c). Le système de classification des marques, en revanche, offre moins de possibilités d'analyse de l'innovation, car les catégories sont plus larges et moins spécifiques. Pour plusieurs millions de marques, seules 45 classes NICE sont couramment disponibles (WIPO 2022a), contre plus de 75 000 catégories dans le système des brevets (EPO 2022). Cela devient un défi lorsque l'on utilise les marques pour

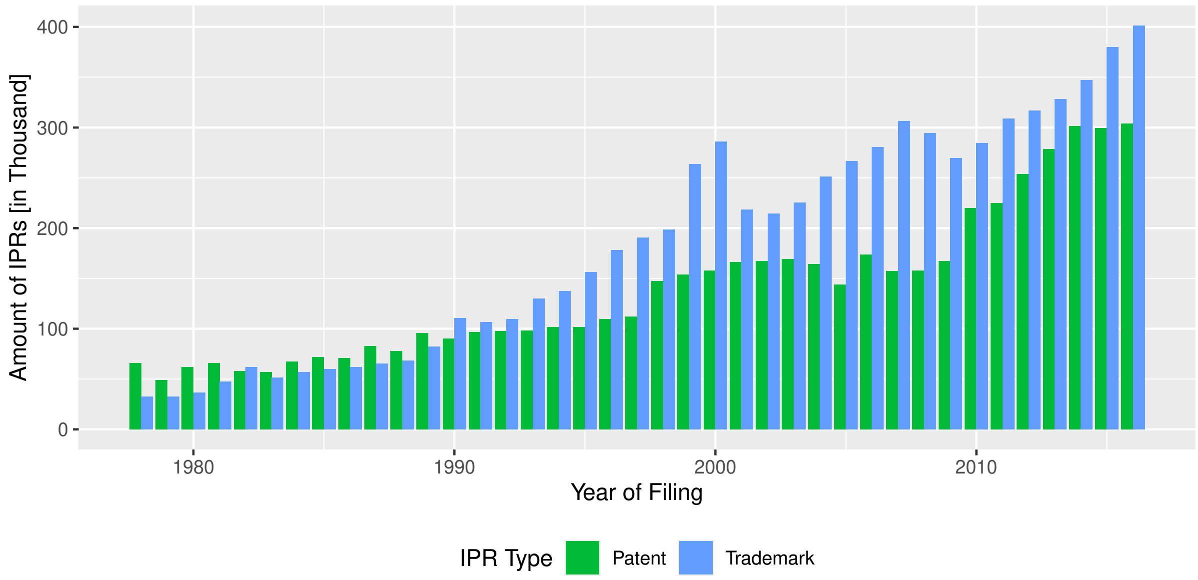


Figure F.1: Développement annuel total des brevets USPTO délivrés et des marques déposées à l'USPTO entre 1978 et 2016.
Source : représentation personnelle basée sur les données de l'USPTO.

des domaines d'innovation spécifiques, car cela implique que le niveau d'analyse reste général, ce qui limite le niveau de granularité qui peut être atteint.

Ainsi, deux défis doivent être relevés : Premièrement, les marques ne couvrent pas toujours l'innovation, et deuxièmement, la granularité de l'analyse des marques basée sur les classifications actuelles reste limitée. Premier défi Le rapprochement des marques et des brevets pourrait permettre de relever les défis des marques en matière d'innovation et de combler les lacunes des brevets. Contrairement aux marques, les brevets couvrent toujours les inventions mais pas toujours l'innovation. Ensemble, ces mécanismes de protection pourraient nous permettre de couvrir les inventions et de garantir leur application sur le marché. En outre, les connaissances acquises sur les innovations technologiques basées sur les brevets pourraient être enrichies par les innovations dans les services ou les secteurs de faible technologie basées sur les marques. Malmberg (2005) combinent les marques et les brevets dans l'industrie pharmaceutique pour saisir les activités d'innovation. Ils constatent que les deux DPI sont liés aux activités d'innovation dans le domaine, les marques couvrant les activités à court terme. Les auteurs suggèrent d'utiliser les deux DPI pour obtenir des informations précieuses, mais ils soulignent également qu'il existe des différences spécifiques à chaque secteur en ce qui concerne l'applicabilité des marques comme indicateur d'innovation. Ribeiro et al. (2022) évalue la corrélation entre les marques et les brevets pour différents secteurs. Les auteurs constatent que les marques commerciales sont particulièrement intéressantes dans les secteurs à faible brevetabilité ou dans lesquels les brevets ne sont pas appliqués, les services et les pays moins développés. Leurs résultats indiquent que la combinaison des marques et des brevets améliore la compréhension des secteurs technologiques et non technologiques. Même si des analyses conjointes des brevets et des marques peuvent être utiles, l'intégration des marques et des brevets n'est pas simple. Une possibilité est l'utilisation de tables de concordance pour combiner les marques et les brevets. Zolas et al. (2017) associe les marques déposées à des données économiques basées sur la concordance de mots-clés pour obtenir des informations sur le comportement des entreprises en matière d'exportation. L'utilisation de mots-clés permet déjà une perspective plus détaillée sur les marques et permet une perspective au niveau de la classe. Cependant, les auteurs soulignent la nécessité de recourir à des approches supplémentaires si l'on souhaite obtenir une plus grande granularité. D'autres auteurs comme Flikkema et al. (2015) intègrent les brevets et les marques via les mêmes entités juridiques qui ont enregistré les différents DPI. Toutefois, cette approche se limite toujours aux liens directs entre les brevets et les marques. Les entreprises qui ne sont actives que dans le domaine des marques ou des brevets, mais qui contribuent à un domaine d'innovation, à son développement et à sa diffusion, sont négligées. En outre, il est difficile de combiner les avantages des différentes sources de données, car de nombreuses informations sont perdues.

Le problème de la granularité limitée de l'analyse des marques doit encore être résolu. Les marques fournissent des informations plus granulaires, contenant des descriptions de la demande de marque pour les produits et les services. Ces descriptions sont des données textuelles qui peuvent être exploitées. Les données peuvent fournir des informations détaillées sur les domaines d'application des marques et les innovations qui sont protégées par des marques. L'analyse de texte couvre les techniques permettant d'identifier des modèles à partir de données textuelles non structurées afin de générer des informations axées sur les données (Aggarwal and Zhai 2012). Antons et al. (2020) plaident en faveur de l'application de l'analyse textuelle pour améliorer les mesures existantes et en développer de nouvelles dans la recherche sur l'innovation.

Les auteurs apprécient ou appliquent l'analyse des données textuelles pour diverses raisons :

- L'analyse textuelle permet l'analyse de grands ensembles de données : C'est pratique lorsque les ensembles de données à analyser dépassent la quantité de données que les humains peuvent traiter en un temps raisonnable. Ozcan et al. (2021) basent leur analyse sur 22 891 tweets de Twitter, Larsen and Thorsrud (2019) utilisent 459 745 articles de journaux, ou Feng et al. (2020) appliquent leur analyse sur 41 994 brevets. L'application de l'analyse textuelle à des données textuelles permet de couvrir des quantités plus importantes. Plusieurs auteurs citent cet argument comme l'une des principales raisons d'appliquer l'analyse des données textuelles. (Loshin 2013; Antons and Breidbach 2018; Bakhtin et al. 2020; Kayser 2017; J. Kim and C. Lee 2017; N. Kim et al. 2015; Kohler et al. 2014; Shen et al. 2020; B. Wang and Z. Wang 2018; Zhu and Porter 2002).

- L'analyse textuelle permet l'analyse de nouvelles sources de données : Les données textuelles étant contenues dans diverses sources, il est possible d'inclure dans l'analyse des sources de données qui ne pourraient pas être analysées avec des approches standard. Par exemple, Dahlke et al. (2021) intègre les besoins humains fondamentaux et les innovations pour évaluer comment les crises façonnent l'innovation. Les auteurs utilisent des descriptions de projets d'innovation extraites du Web pour définir l'état de l'innovation et les comparent à neuf besoins humains fondamentaux décrits dans la littérature. J. Kim and C. Lee (2017) extraient des données futuristes de sites web tels que Siemens, MIT technology review, ou la World Future Society pour extraire des sujets futurs scientifiquement pertinents et les comparer à l'état de l'art révélé dans les données de brevets. Ils ont l'intention d'identifier les signaux faibles et d'extraire ainsi de nouveaux sujets. Mi et al. (2021) tirent parti des données d'enquête sur les personnes âgées pour déterminer la demande de gérontechnologie et comparent les informations obtenues aux rapports sur les technologies émergentes et les brevets pour prévoir les tendances futures. Un dernier exemple est Fiordelisi et al. (2019) qui utilisent les rapports 10-K des entreprises pour extraire des informations sur la créativité des entreprises afin de mesurer l'impact sur la capacité d'innovation et la valeur des entreprises. Les auteurs constatent que la culture créative favorise l'innovation. Ce ne sont là que quelques exemples de la manière dont les nouvelles sources de données révèlent des informations intéressantes à inclure dans les modèles économiques. Bien que l'analyse de texte offre davantage de possibilités d'utilisation des données et que d'autres sources de données soient disponibles, les brevets restent la source de données la plus utilisée. Cependant, la quantité de données textuelles disponibles pour la recherche sur l'innovation est importante et augmente chaque année : Rien qu'en 2020, plus de 2,6 millions de nouveaux articles de recherche ont été publiés sur Web of Science (WoS 2021). Concernant les droits de propriété intellectuelle tels que les brevets, les marques et les dessins et modèles, 304 126 brevets, 234 444 marques et 28 886 dessins et modèles ont été enregistrés pour la seule année 2016 aux États-Unis, avec un total de 5 439 151 brevets enregistrés, 4 246 859 marques déposées et 527 902 dessins et modèles enregistrés de 1978 à 2016.[1].

- L'analyse textuelle permet de combiner différentes sources de données, ou d'utiliser des données hétérogènes : Les barrières imposées par la différence de structure des données de différentes sources de données peuvent être surmontées par l'utilisation du langage et des mots partagés. Plusieurs auteurs en tirent parti : Bakhtin et al. (2020) combinent diverses sources de données allant d'articles scientifiques, de brevets, d'actualités et de prix à des rapports analytiques avec le traitement du langage naturel pour identifier les modèles émergents dans l'agriculture et la production alimentaire afin de fournir des idées aux gouvernements et aux entreprises pour les activités de planification stratégique. Kayser (2017) intègrent les médias aux résumés des publications via des termes appariés pour une perspective de diffusion de l'innovation, tandis que Shen et al. (2020) comparent le contenu textuel des groupes de brevets et de publications pour découvrir les opportunités de progrès scientifique et d'innovation technologique dans les technologies de surveillance intelligente de la santé. La combinaison de plusieurs sources de données peut ainsi révéler de nouvelles perspectives, et l'analyse textuelle facilite cette intégration.

Un autre argument important dans ce contexte est le temps... d'analyse efficaces en termes de temps (Bakhtin et al. 2020; Chiarello et al. 2021; Dahlke et al. 2021; Kayser 2017; J. Kim and C. Lee 2017; Zhou et al. 2020; Zhu and Porter 2002), qui permettent d'obtenir des résultats rapides et donc des déclarations rapides. Cela est particulièrement important dans les domaines en évolution rapide où de nouvelles découvertes sont constamment publiées. En outre, l'extraction de connaissances basée sur des données est pertinente, de sorte que les connaissances peuvent être acquises indépendamment des experts. (Bakhtin et al. 2020; Basole et al. 2019; Larsen and Thorsrud 2019; Mi et al. 2021; Shen et al. 2020; Song et al. 2017; Zhou et al. 2020). Cela réduit les préjugés des experts et permet à un plus grand nombre de personnes de générer des connaissances. Dans l'ensemble, l'application de l'analyse textuelle à la recherche sur l'innovation peut révéler des informations intéressantes. Des études ont été réalisées, par exemple, sur le développement actuel de technologies comme les biocarburants (Curci and Mongeau Ospina 2016) ou la blockchain (Chiarello et al. 2021), la diffusion de technologies comme le big data dans la recherche scientifique (Y. Zhang et al. 2019) ou les sujets d'innovation à venir comme les batteries ou les connecteurs de charge dans les véhicules électriques (Feng et al. 2020). L'analyse textuelle peut également être utilisée pour générer de nouvelles idées de domaines d'innovation en extrayant des idées de sources publiques telles que Twitter (Ozcan et al. 2021) ou en identifiant des domaines prometteurs en comparant les besoins des consommateurs aux technologies disponibles (Mi et al. 2021). Dans le contexte des entreprises, l'analyse textuelle peut également révéler des informations sur la convergence des entreprises (N. Kim et al. 2015) ou rendre la créativité des entreprises traçable (Fiordelisi et al. 2019).

Les divers arguments en faveur de l'analyse de texte la rendent également intéressante pour l'analyse des marques : De grandes quantités de marques sont disponibles pour l'analyse. Les données textuelles des marques peuvent être utilisées pour extraire des informations détaillées sur différents domaines d'innovation et aspects de la diffusion. De plus, les données textuelles des marques nous permettent de relier les marques à d'autres sources de données. La combinaison des marques avec différentes sources de données peut être intéressante pour saisir, par exemple, les services mais aussi le développement des produits de manière plus large. La combinaison pourrait être réalisée textuellement entre les marques et les brevets. Cela implique que des mots partagés entre les marques et les brevets relient les différents droits de propriété intellectuelle. Cela permettrait également de surmonter les grandes différences entre les systèmes de classification disponibles dans les marques et les brevets. L'hypothèse principale est que la description d'une invention est similaire dans les différents droits de propriété intellectuelle, mais que le domaine de protection et d'application change, avec un léger changement d'orientation en ce qui concerne l'information transmise. Cela signifie que les brevets couvrent davantage les aspects

[1]Données propres basées sur Query A.1, Query A.1 et Query A.1

technologiques, tandis que les marques mettent en lumière l'application et la diffusion d'une invention sur le marché. Dans cette thèse, c'est donc la combinaison des marques et des brevets qui nous intéresse en premier lieu. Toutefois, les dessins et modèles sont également considérés comme une source de données intéressante pour les domaines d'innovation qui accordent une grande importance à d'autres formes et motifs. Les dessins et modèles sont appliqués pour protéger les nouveaux éléments ornementaux (WIPO 2021c). Comme les marques, les dessins et modèles ne sont pas actuellement inclus comme source de données standard dans les analyses de l'innovation. Dans ce travail, ils sont inclus dans Chapter 5, car ils présentent un intérêt dans le domaine étudié. Comme les marques, ils pourraient contribuer à une mesure plus précise de l'innovation et fournir des informations intéressantes.

Les premières tentatives d'utilisation des données textuelles des marques pour l'analyse ont été faites : Flikkema et al. (2019) déduire des stratégies de marque de création, d'extension ou de modernisation sur la base des noms de marque et de la couverture NICE. Positionnement de l'entreprise Semadeni (2006) compare le libellé des demandes de marques de différentes entreprises afin de déterminer la position relative des entreprises les unes par rapport aux autres. Le libellé des marques peut également révéler la probabilité d'imitation des innovations de services (Semadeni and Anderson 2010). Même s'il existe quelques exemples d'utilisation des données textuelles des marques, l'application de l'analyse des données textuelles n'est pas encore courante. Dans son aperçu, Castaldi (2020) souligne que l'accent est encore mis principalement sur les données structurées et non textuelles des marques qui sont utilisées pour l'analyse et que les informations sur les marques ne sont pas encore utilisées à leur plein potentiel.

En termes de combinaison, les données textuelles des marques et des brevets ont été combinées : Des auteurs comme M. Lee and S. Lee (2017) utilisent les brevets des entreprises pour afficher les connaissances technologiques et obtenir des détails sur le positionnement des concurrents sur le marché à partir des descriptions des marques. H. Kim et al. (2017) font correspondre les brevets et les marques de façon générale avec des mots-clés partagés entre les brevets et les marques pour découvrir comment les connaissances technologiques peuvent être appliquées sur le marché et identifier les domaines potentiels de diversification. Les brevets fournissent ainsi la perspective de l'invention, tandis que les marques commerciales révèlent les domaines d'introduction de ces inventions sur le marché. Cependant, les approches visant à combiner les sources de données se concentrent encore principalement sur les mots-clés. Approches autres que les mots-clés La modélisation thématique est une autre méthode d'extraction d'informations à partir de données textuelles qui prend en compte plus que les mots-clés des données textuelles. La modélisation thématique est une approche permettant d'extraire les thèmes apparaissant dans un ensemble de documents sur la base des mots contenus dans ces documents. Contrairement à la correspondance basée sur les mots clés, la modélisation des sujets peut également tenir compte des synonymes ou du contexte des mots. Des exemples d'applications dans le contexte de la recherche sur l'innovation sont, par exemple, Basole et al. (2019), qui identifie les thèmes et les groupes dans l'écosystème entrepreneurial via la modélisation thématique et révèle que les groupes d'entrepreneurs ne sont pas seulement liés à l'industrie mais aussi à la similarité de leurs thèmes. Un autre exemple est celui de Larsen and Thorsrud (2019), qui applique la modélisation thématique pour identifier les sujets intéressants dans les articles d'actualité qui sont ensuite intégrés dans des modèles économiques pour améliorer la prédiction des booms économiques.

Dans l'ensemble, les données textuelles des marques sont au centre de cette thèse. Les informations dans la description textuelle du domaine d'application des marques dans les biens et services pourraient fournir une perspective plus large sur l'innovation car elles couvrent des domaines et des applications techniques et non techniques dans les biens et services. Les données textuelles permettent en outre de combiner les marques avec les brevets qui se composent principalement de données textuelles, telles que les résumés qui résument le contenu ou les revendications de brevet qui couvrent les domaines protégés. L'analyse des données textuelles des marques est une solution pour extraire plus d'informations des marques et fournir simultanément une possibilité de combiner les marques avec d'autres sources de données. En faisant cela, cette thèse étend la couverture de l'innovation avec l'inclusion textuelle des marques pour fournir une perspective plus large sur l'innovation. La large couverture de l'innovation soutient les recommandations politiques basées sur des données empiriques. Cependant, pour tirer profit du potentiel des données textuelles des marques, les données doivent être évaluées de manière plus complète, et les caractéristiques de la source de données doivent être élaborées.

F.2 Questions de Recherche

Plusieurs aspects doivent être considérés dans le contexte de cette thèse pour extraire les avantages des données textuelles des marques pour la recherche sur l'innovation :

Les marques commerciales sont une source de données qui fournit, par exemple, une perspective sur les services ou les domaines de basse technologie mais qui couvre également les domaines de haute technologie. Les marques commerciales présentent toutefois l'inconvénient d'un lien peu clair avec l'innovation et de niveaux de détail limités pour des analyses sophistiquées.

Données textuelles peut fournir des informations détaillées et, en ce sens, offre une perspective supplémentaire en combinaison avec les données structurées. En outre, comme les données textuelles sont disponibles dans diverses sources de données, elles ont le potentiel de permettre de nouvelles combinaisons de données et, par conséquent, de nouvelles idées tirées de ces combinaisons. Dans le contexte des données textuelles, il reste toutefois à voir comment la possibilité de les analyser est utilisée dans la recherche sur l'innovation. Il est intéressant de savoir quelles méthodes et quelles données sont utilisées pour répondre aux différentes questions de recherche.

Données textuelles des marques permet de combiner les marques avec d'autres sources de données, offrant ainsi une perspective plus large sur l'innovation en général grâce à l'inclusion des services et de l'innovation de basse technologie dans les analyses. Toutefois, il convient de préciser en quoi les textes des marques diffèrent, par exemple, des textes des brevets et, par conséquent, en quoi ils apportent une valeur ajoutée. Les brevets pourraient fournir l'étape inventive, tandis que les marques garantissent l'application commerciale. Ainsi, la couverture de l'innovation pourrait être assurée.

Innovation dans les données textuelles des marques n'est pas garanti. Certaines marques sont liées à l'innovation, mais pas toutes. Par conséquent, il reste à déterminer comment cela affecte l'analyse des données textuelles des marques et si les données textuelles des marques peuvent fournir des indications intéressantes sur les activités dans les domaines de l'innovation, ne serait-ce que pour mettre en lumière les aspects de diffusion.

Dans l'ensemble, la thèse aborde ces aspects et entend conserver la force des marques tout en abordant simultanément les limites de cette source de données en utilisant des données textuelles. Afin d'acquérir une compréhension des marques et, dans ce contexte, de l'analyse des données textuelles, la thèse est structurée comme suit : Dans Chapter 2, le contexte général de l'innovation, des brevets, des marques et des dessins et modèles en tant que partie des droits de propriété intellectuelle et de la mesure de l'innovation est donné. En outre, l'analyse des données textuelles en général et avec un accent sur la principale méthode utilisée dans cette thèse est présentée. Ces connaissances servent de base générale à la thèse. Après le Foundations, le Chapter 3 se concentre sur l'utilisation actuelle des données textuelles pour la recherche sur l'innovation en général. Cette revue de la littérature fournit une compréhension générale de l'état de l'art des données textuelles, de leur analyse, des raisons d'appliquer l'analyse des données textuelles et des domaines d'application potentiels dans la recherche sur l'innovation. Chapter 4 et Chapter 5 effectuent ensuite des analyses de données textuelles de marques en combinaison avec d'autres sources de données. Chapter 4 se concentre sur la manière dont les marques et les brevets peuvent être combinés efficacement pour donner une image plus large de l'innovation. Il évalue la combinaison textuelle et la manière dont les aspects de l'innovation sont couverts dans chaque source de données, mais en se concentrant, en particulier, sur les marques. Les hypothèses sous-jacentes de ce chapitre sont que les perspectives sur les marques des données structurées et textuelles diffèrent et que les données textuelles des marques contribuent généralement à une compréhension plus large de l'innovation. L'analyse est effectuée sur la robotique et les chaussures en tant que représentants d'une haute technologie et d'une basse technologie, respectivement. Les résultats obtenus servent de base à d'autres analyses de l'innovation basées sur les données textuelles des marques. En revanche, Chapter 5 combine les brevets, les marques et les dessins et modèles pour couvrir la transformation du secteur des instruments de musique, qui passe d'une technologie faible à une technologie élevée, et met l'accent sur les entreprises. Même si l'on pourrait s'attendre à une augmentation des demandes de brevets dans le secteur en raison de la pertinence croissante des instruments électroniques, les marques déposées sont le droit de propriété intellectuelle pertinent dans le secteur, non seulement pour les instruments de musique classiques mais aussi pour les instruments électroniques. L'analyse contribue à une meilleure compréhension de ce phénomène en donnant un aperçu de l'application des droits de propriété intellectuelle dans le contexte de la transformation technologique et du contexte de l'entreprise. La thèse se termine dans Chapter 6, où des idées générales sont reflétées et discutées, en particulier concernant les données textuelles des marques, et la conclusion principale de la thèse est tirée. En général, le travail contribue à une meilleure compréhension des données textuelles des marques. Il met en lumière les données textuelles dans la recherche sur l'innovation dans son ensemble, la contribution des marques combinée à d'autres sources de données et utilise les connaissances acquises pour répondre à des questions économiques.

Pour la structure générale de la thèse, on peut donc dire que les chapitres commencent par une perspective générale sur les données textuelles et deviennent plus spécifiques au cours de la thèse : La thèse se concentre d'abord sur les données textuelles en général, puis examine les données textuelles des marques par rapport aux brevets avant de finalement appliquer les connaissances acquises pour évaluer la transformation technologique concernant l'utilisation des droits de propriété intellectuelle, couvrant les brevets, les marques et les modèles. La structure de la thèse est représentée graphiquement dans Figure F.2. Ici, les contours des boîtes, de couleur commune, indiquent les aspects partagés entre les différents chapitres de la thèse. La légende se trouve à gauche.

Domaines/ secteurs d'innovation : Tous les chapitres couvrent différents aspects de l'innovation (en jaune). Le chapitre de la fondation fournit les aspects généraux de l'innovation et de sa mesure. L'examen dans Chapter 3 n'est pas limité à un domaine ou à un secteur d'innovation. Il suffit que les articles de l'échantillon soient publiés dans un secteur d'activité ou un domaine économique, qu'ils traitent de l'innovation et qu'ils utilisent des données textuelles. Chapter 4 compare la robotique, secteur de haute technologie, et les chaussures, secteur de basse technologie, afin d'obtenir des informations généralisables. Enfin, Chapter 5 se penche sur la transformation des instruments de musique, qui couvre les aspects des applications de basse et haute technologie dues aux instruments de musique classiques et électroniques.

Droits de propriété intellectuelle : Les droits de propriété intellectuelle (violet) avec un accent sur les marques sont présents dans toutes les parties de la thèse. Le chapitre de la fondation présente le contexte des droits de propriété intellectuelle et notamment les brevets, les marques et les dessins et modèles et la mesure de l'innovation, ainsi qu'une introduction à la méthode d'analyse des données textuelles appliquée dans cette thèse. Dans Chapter 3, les propriétés intellectuelles ne sont qu'indirectement considérées comme faisant partie des sources de données textuelles utilisées dans certaines des publications analysées. Les marques et les brevets sont ensuite évalués de près dans Chapter 4 et Chapter 5 avec l'ajout des dessins dans ce dernier. Cependant, la thèse se concentre principalement sur les marques.

Données textuelles : Les données textuelles (en vert) et leur analyse sont pertinentes dans tous les chapitres. L'analyse des données textuelles est décrite dans le chapitre de la fondation. Cette thèse utilise la modélisation thématique structurelle développée par Roberts et al. (2019). La modélisation des sujets est une technique d'analyse de texte qui permet de déduire des sujets à partir d'un grand nombre de documents textuels. La modélisation thématique structurelle est

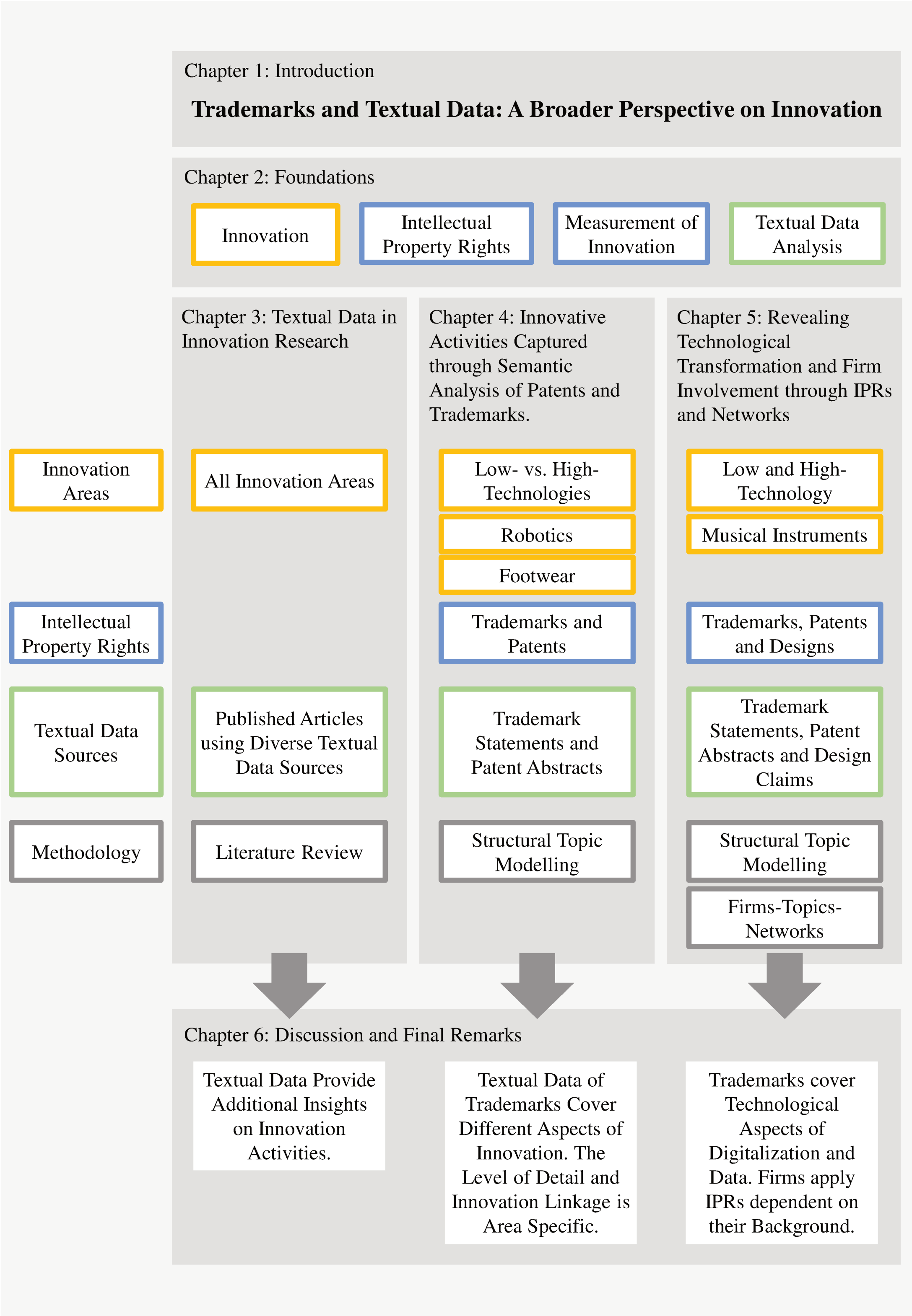


Figure F.2: Structure de la thèse.
Les cases à gauche indiquent la signification de la coloration des contours de la forme. Le gris se rapporte à la méthodologie, le violet aux aspects des droits de propriété intellectuelle, le vert aux sources de données textuelles utilisées pour les analyses et l'orange aux domaines d'innovation couverts.

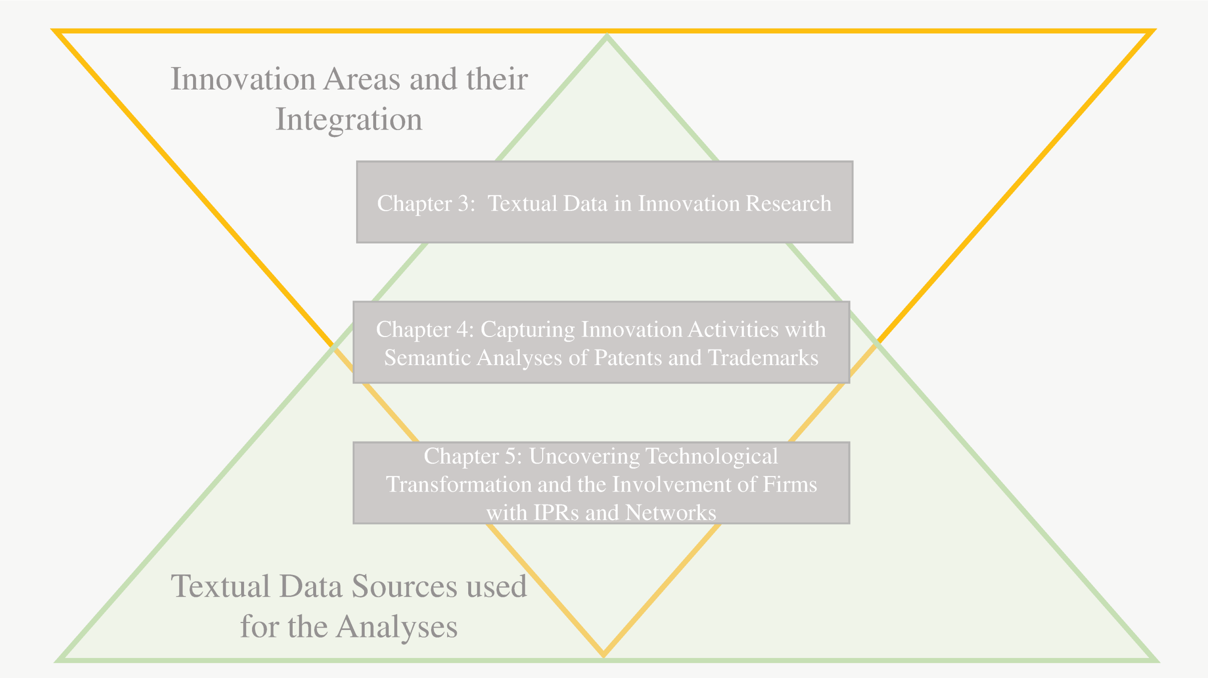


Figure F.3: Perspective intégrée sur les chapitres de recherche.
Le triangle en jaune représente l'étendue des domaines d'innovation couverts. Le triangle en vert représente l'étendue des sources de données utilisées.

une approche spécifique de modélisation thématique qui identifie les thèmes dans les documents existants en tenant compte d'informations supplémentaires (voir plus de détails dans Section 2.4). Chapter 3 analyse les articles publiés en utilisant des données textuelles dans la recherche sur l'innovation. Les articles de l'échantillon ne sont pas limités à une source de données spécifique. Chapter 4 analyse ensuite les brevets et les marques déposées, tandis que Chapter 5 inclut en plus les dessins et modèles dans l'analyse.

Figure F.3 offre une perspective intégrée des chapitres. Au cours de la thèse, les domaines d'innovation couverts sont réduits pour augmenter la concentration, le niveau de détail et l'intégration de la haute et de la basse technologie (triangle en jaune). Les domaines sont les plus larges dans l'analyse documentaire de Chapter 3, où toutes sortes de domaines en relation avec l'innovation et les données textuelles sont évalués. Dans Chapter 4, la robotique de haute technologie et les chaussures de basse technologie sont analysées comme des domaines d'innovation. En revanche, les aspects de haute et de basse technologie des instruments de musique sont considérés conjointement dans Chapter 5. Le triangle en vert représente l'ampleur des sources de données utilisées pour les analyses dans chaque chapitre, qui augmente au cours de la thèse. Au début, les publications sont considérées, suivies d'une analyse conjointe des marques et des brevets, et enfin d'une analyse conjointe des brevets, des marques et des dessins et modèles.

Le reste de la thèse se déroule comme suit : Tout d'abord, des informations supplémentaires sur l'innovation, les droits de propriété intellectuelle, les mesures de l'innovation et l'analyse des données textuelles sont fournies dans le chapitre Chapter 2. Dans le chapitre Chapter 3, des informations sur les données textuelles dans la recherche sur l'innovation sont fournies avant de réaliser différentes analyses sur les données textuelles des marques dans les chapitres 4 et 5. Enfin, les résultats de la thèse sont résumés, et les sujets généraux sont discutés avant de conclure avec les principales contributions de la thèse dans Chapter 6.

F.3 Résumé et Contribution

Cette thèse a couvert divers aspects des données textuelles des marques dans le contexte de la recherche sur l'innovation. Avant de conclure cette thèse, un résumé de chaque chapitre est fourni dans Section 6.1 suivi d'une discussion. La discussion se concentre sur les connaissances acquises sur les données textuelles des marques à partir de l'examen conjoint des différents chapitres. La thèse se conclut dans le chapitre Section 6.3.

La thèse évolue autour de l'utilisation des données textuelles des marques dans la recherche sur l'innovation. Afin de capturer l'innovation au sens large et de couvrir en particulier l'innovation de service et l'innovation dans les domaines de basse technologie, cette thèse a examiné les marques comme indicateur d'innovation. Afin de combiner les marques avec d'autres sources de données et d'augmenter le niveau de détail des analyses, les données textuelles des marques ont été évaluées.

Après avoir présenté des informations générales sur l'innovation, les droits de propriété intellectuelle, la mesure de l'innovation et l'analyse des données textuelles dans Chapter 2, Chapter 3 fournit une perspective générale sur l'utilisation des données textuelles dans la recherche sur l'innovation. Chapter 4 compare ensuite les brevets et les marques déposées du point de vue des données textuelles dans Robotics et Footwear. Les sources de données sont évaluées en fonction de leur couverture des produits, des services et de l'innovation technique. Chapter 5 applique les connaissances acquises dans les chapitres précédents et les utilise pour mesurer la transformation technologique sur la base des données textuelles des brevets, des marques et des dessins et modèles dans les instruments de musique. Il fournit également des informations sur l'implication des différentes entreprises dans la transformation.

Des détails supplémentaires sur chaque chapitre sont fournis dans ce qui suit :

Chapter 3 fournit un aperçu général de l'utilisation actuelle des données textuelles dans la recherche sur l'innovation. La motivation à appliquer l'analyse des données textuelles, les sources de données utilisées, la méthodologie appliquée et les réponses aux questions de recherche basées sur les données textuelles présentent un intérêt particulier. Ce chapitre contribue à une meilleure compréhension des données textuelles dans la recherche en innovation et des différentes méthodologies appliquées. Dans le contexte de la thèse, une vue d'ensemble des méthodes d'analyse de texte est donnée, ce qui permet de mieux comprendre l'analyse de texte. A cette fin, les articles des domaines de l'économie et des affaires qui contiennent le texte et l'innovation dans le titre ou le résumé sont considérés. L'ensemble de données est successivement réduit sur la base des mots-clés fournis par les auteurs et d'une analyse intensive des résumés. L'échantillon final se compose de 23 articles qui sont liés à l'innovation et aux données textuelles.

La plupart des articles ont été publiés dans des revues connues sur la politique de l'innovation, et un tiers était dans un domaine spécifique. La plupart des auteurs apprécient la possibilité d'analyser de grands ensembles de données textuelles d'une manière objective, axée sur les données et efficace en termes de temps. Les données textuelles servent de source de données pour des analyses temporelles et pour détecter des opportunités ou découvrir des informations cachées. Les sources de données utilisées sont principalement les brevets et les publications. Les autres sources de données sont les articles d'actualité, les médias sociaux ou les documents de l'entreprise. La plupart des articles de l'échantillon se concentrent sur une seule source de données, tandis que six articles combinent plusieurs sources de données pour répondre à leurs questions de recherche, y compris des informations sur, par exemple, la culture de l'entreprise, le retour d'information des clients et la réaction du public dans les analyses de l'innovation. Les combinaisons comprennent toujours soit des brevets, soit des publications. Onze articles appliquent des algorithmes de clustering, notamment pour l'exploration et l'identification des sujets, tandis que trois articles appliquent des méthodologies de classification pour identifier les idées ou les entreprises innovantes. Pour la mise en grappes, on utilise notamment l'allocation latente de Dirichlet (LDA), qui est une approche de modélisation des sujets : (Blei et al. 2003). Diverses questions de recherche sont abordées : L'un des principaux domaines de recherche abordés est la découverte et l'exploration des technologies, principalement sur la base de la modélisation des thèmes. D'autres domaines sont la convergence industrielle et les régions, la trajectoire et la diffusion de l'innovation, et l'émergence de nouveaux sujets ou la génération d'idées.

Ce chapitre permet de comprendre l'état actuel de l'analyse des données textuelles. Il contribue à l'application de l'analyse des données textuelles dans la recherche sur l'innovation. Il met en évidence les domaines de recherche potentiels qui bénéficient de l'utilisation des données textuelles : Les analyses textuelles permettent d'analyser des sources de données textuelles et de combiner différentes sources de données via le langage. En raison de leur implication réduite des experts, l'objectivité augmente. L'analyse des données textuelles devient plus rapide et plus facile avec l'utilisation de méthodes basées sur le texte, permettant une répétition régulière. Les méthodologies peuvent être utilisées pour une variété d'objectifs. La plupart du temps, elles sont utilisées pour identifier des sujets ou des tendances émergentes et les rendre visibles. Dans l'ensemble, l'examen révèle que la recherche sur les données textuelles dans l'innovation peut apporter une perspective supplémentaire. Jusqu'à présent, l'analyse des données textuelles reste essentiellement descriptive, mais il existe des concepts permettant d'utiliser les informations obtenues dans une analyse plus approfondie.

Chapter 4 combine les données textuelles des brevets et des marques et évalue comment les aspects de l'innovation sont couverts dans chaque source de données, en se concentrant particulièrement sur les marques. Contribution Ce chapitre contribue à une meilleure compréhension de l'application des données textuelles des marques et de la combinaison de cette source de données avec les brevets. Cela permet d'autres analyses de l'innovation basées sur les données textuelles des marques. Ce chapitre aborde deux questions de recherche :

RQ4.1 : Les marques et les brevets peuvent-ils être combinés via leurs données textuelles à un niveau détaillé ?

RQ4.2 : La combinaison textuelle des marques et des brevets apporte-t-elle de nouvelles perspectives à la discussion sur l'innovation ?

Pour répondre aux questions de recherche, les marques et les brevets sont combinés textuellement avec la modélisation thématique structurelle de Roberts et al. (2019). La robotique et les chaussures sont analysées en tant que représentants de la haute technologie et de la basse technologie, respectivement. Les résultats sont analysés en termes de combinaison textuelle et d'aperçus obtenus. Par conséquent, des hypothèses générales sont tirées de la littérature sur l'innovation dans les marques, les marchés, les services et la technologie.

L'estimation du modèle est utilisée pour analyser la robotique et les chaussures. Dans le cas de la robotique, par exemple, 115 sujets sont estimés sur la base de l'ensemble des données fournies au modèle. Comme on peut le voir dans Figure F.4, les sujets ont des probabilités d'occurrence différentes dans chaque source de données. Les barres vertes affichent les sujets dont l'occurrence dans les brevets est significative, tandis que les barres bleues affichent les sujets dont l'occurrence dans les marques est significative. Les barres grises n'ont pas d'occurrence claire dans les deux sources de données. Au cours des analyses, les différents thèmes dans les marques, les brevets et aucun des deux sont évalués. La cohérence interne des thèmes est ainsi considérée, ce qui implique que les mots généraux du thème et les documents qui y sont liés doivent être compréhensibles pour l'observateur humain. La relation entre le sujet et les systèmes de classification officiels ainsi que l'importance du sujet dans le temps sont également pris en compte. L'analyse est répétée pour les chaussures, avant que les résultats ne soient combinés et évalués par rapport aux questions de recherche.

La première question porte sur la possibilité de combiner textuellement les brevets et les marques :

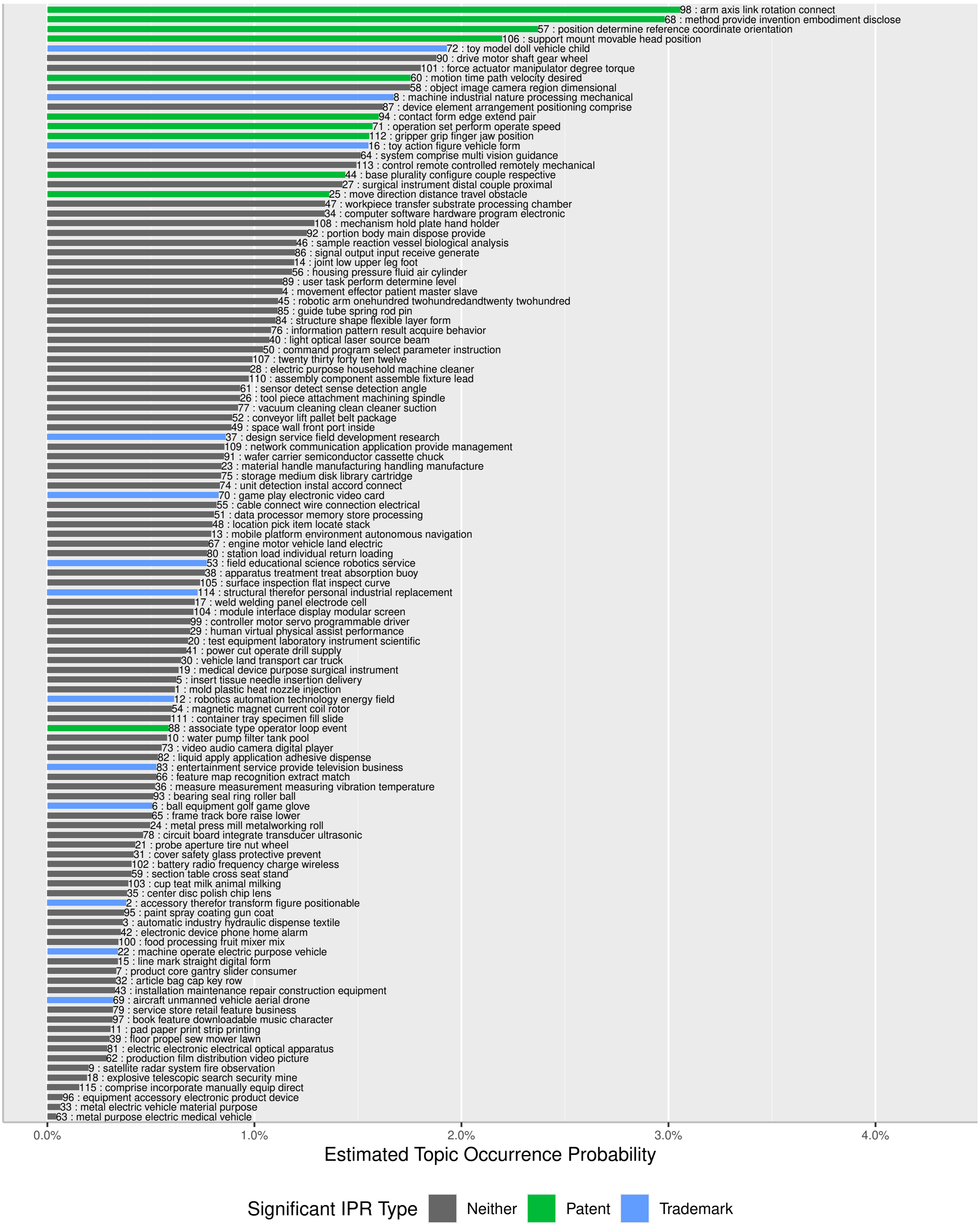


Figure F.4: Aperçu du modèle avec 115 sujets en robotique.
La figure présente une vue d'ensemble des 115 sujets de la robotique. Les sujets sont affichés en fonction de leur part relative dans l'ensemble des données et classés par ordre décroissant. La couleur indique le type de document qui est significatif pour le sujet, avec un intervalle de confiance de 95% (Figure C.2). En cas d'absence de signification, la coloration de la barre est maintenue en gris. La légende de droite indique le numéro du sujet et les cinq mots ayant la probabilité d'occurrence la plus élevée dans le sujet.

RQ4.1 : *Les marques et les brevets peuvent-ils être combinés via leurs données textuelles à un niveau détaillé ?*

En termes de combinaison et de thèmes cohérents, l'analyse a donné des résultats mitigés : En robotique, la combinaison textuelle des marques et des brevets a donné des résultats cohérents. Ils étaient en accord avec les classifications des brevets et des marques mais fournissaient une perspective plus détaillée. Les brevets soulignent l'invention et les aspects techniques, tandis que les marques ajoutent des détails et couvrent principalement les domaines d'application. Pour l'analyse des marques, l'approche relie davantage les thèmes des marques au système de classification des brevets, ce qui permet une meilleure interprétation des marques. Globalement, les données textuelles des marques ont amélioré l'analyse et fourni des informations sur des domaines supplémentaires. Les chaussures sont un domaine de faible technologie avec moins d'inventions et d'applications technologiques. Cela a pu être observé dans l'analyse textuelle : Seul un petit nombre de brevets par rapport aux marques était disponible. De plus, les marques se concentraient principalement sur plusieurs articles d'habillement plutôt que sur des technologies innovantes. Les marques commerciales ne couvraient que peu d'informations sur les inventions pour générer des sujets cohérents ; ainsi, la combinaison des brevets et des marques commerciales n'était que partiellement cohérente. Les résultats dans les domaines de la robotique et de la chaussure ont donc fourni des résultats différents. Il est apparu que la structure textuelle des marques différait entre la robotique et les chaussures : En Robotique, les descriptions des marques sont généralement plus détaillées avec des informations sur le produit, le but du robot et les aspects de la protection. De plus, un niveau de détail croissant est observable au fil du temps. Les descriptions de marques dans le domaine des chaussures sont moins spécifiques et détaillées que celles de la robotique. Les descriptions de marques des sujets avec des résultats cohérents dans les Chaussures ont montré une structure textuelle similaire à celles de la Robotique. Le niveau de détail disponible soutient le modèle dans ses calculs. En général, le niveau de détail disponible dans les descriptions de marques diffère selon le domaine d'innovation examiné et la période couverte. Les marques commerciales comportant un degré élevé d'informations techniques contiennent des informations similaires à celles des brevets, tandis que les marques commerciales comportant moins d'aspects technologiques et d'innovation s'écartent des brevets. Les textes de marque capturent donc les activités d'innovation et fournissent une perspective détaillée. Toutefois, le niveau de détail est spécifique au secteur ou au domaine d'innovation. Il convient d'en tenir compte lorsque les données textuelles des marques sont utilisées pour analyser l'innovation.

RQ4.2 : *La combinaison textuelle des marques et des brevets apporte-t-elle un nouvel éclairage à la discussion sur l'innovation ?*

La deuxième question se concentre sur les connaissances supplémentaires apportées par la combinaison. Des hypothèses générales sur les sources de données sont extraites de la littérature pour évaluer les informations supplémentaires. Celles-ci sont que les marques couvrent l'innovation, plus les innovations liées aux services ou au marché, et que les brevets fournissent des informations sur l'innovation technique. Ces hypothèses sont remises en question car tous les résultats ne sont pas forcément reproductibles sur la base de données textuelles. Par exemple, une différence dans la couverture des services pourrait ne pas être observable textuellement. Les informations relatives aux services peuvent également être couvertes par les données textuelles des brevets, ou aucune description supplémentaire des services peut être fournie dans les textes des marques. Par conséquent, il est nécessaire de comprendre la contribution des données textuelles des marques par rapport aux brevets. Les différentes hypothèses sont évaluées par rapport aux résultats de l'analyse des données textuelles et révèlent ce qui suit :

Marques et innovation : Le lien entre les marques et l'innovation est observable dans les données textuelles. Cependant, un certain degré technologique est nécessaire pour combiner les marques et les brevets et pour générer des sujets dans les marques où des informations supplémentaires sur l'innovation sont disponibles.

Marché et produits : En robotique, les marques assurent la perspective de marché des inventions couvertes par les brevets. Le marché et l'application du produit étaient également couverts par les brevets, mais pas en tant qu'objectif principal. Les brevets ont aidé à interpréter et à regrouper les marques, car le système de classification des brevets fournissait des détails plus structurés que le système de classification des marques. Dans le secteur de la chaussure, les marques n'ont fourni la perspective de l'introduction sur le marché que dans le cas de sujets cohérents. Les marques étaient principalement axées sur les produits finis sans autres détails sur leur invention. Ici, l'avantage de la combinaison et des données textuelles des marques pour la recherche sur l'innovation doit être remis en question. En général, les marques couvrent plus de produits finis et la perspective du marché. Elles peuvent ajouter la perspective du marché et du produit dans l'analyse de l'innovation, tandis que les brevets se concentrent plus en détail sur les inventions qui les sous-tendent.

Services : Des informations sur les services ont pu être obtenues à partir des deux sources de données, avec une plus grande représentation dans les marques. Les services comme le divertissement ou l'éducation sont plus présents dans les marques. Les services avec une composante technique sont également observables dans les brevets comme la robotique de service ou les abonnements de chaussures, ce qui est en accord avec les résultats de Blind et al. (2003). L'inclusion des marques a fourni des informations supplémentaires sur les services et a mis en évidence les domaines de services de la robotique ou des chaussures.

Innovation technique : Les thèmes des brevets dans les domaines de la robotique et des chaussures couvrent des aspects techniques et se concentrent sur les fonctions ou les composants. Dans le domaine de la robotique, les marques apportent des informations techniques détaillées sur divers sujets, indépendamment des DPI importants. Dans le secteur des chaussures, les détails techniques sont fournis principalement par les brevets et rarement par les marques. En général, les descriptions de brevets fournissent toujours plus d'informations techniques que les marques. Les descriptions de marques couvrent néanmoins des aspects techniques, mais leur étendue dépend fortement du domaine d'innovation considéré, et le niveau de détail est inférieur à celui des brevets.

Les hypothèses de la littérature s'appliquent donc, en général, également aux données textuelles des marques et des brevets. En termes d'informations supplémentaires, les marques peuvent donc contribuer à la couverture de l'innovation. La qualité des résultats dépend fortement du domaine d'application analysé et des marques couvertes dans l'ensemble de données.

Dans l'ensemble, la combinaison des marques et des brevets a révélé que les sources de données peuvent être combinées à un niveau détaillé et que cette combinaison permet de mieux comprendre le débat sur l'innovation. Étant donné que le champ d'analyse a un impact important sur la qualité des résultats, il faut tenir compte du champ d'analyse. Ce chapitre contribue à une meilleure compréhension des données textuelles des marques en tant qu'indicateur d'innovation - en général et pour des domaines d'innovation spécifiques de basse et haute technologie. En outre, le chapitre souligne le potentiel de la combinaison des marques et des brevets et les différences entre ces deux sources de données textuelles en matière d'innovation. Les connaissances acquises servent de base à d'autres analyses de l'innovation basées sur les données textuelles des marques, notamment dans le cadre d'une analyse conjointe avec les brevets, qui peuvent être utilisées pour améliorer les mesures existantes de l'innovation.

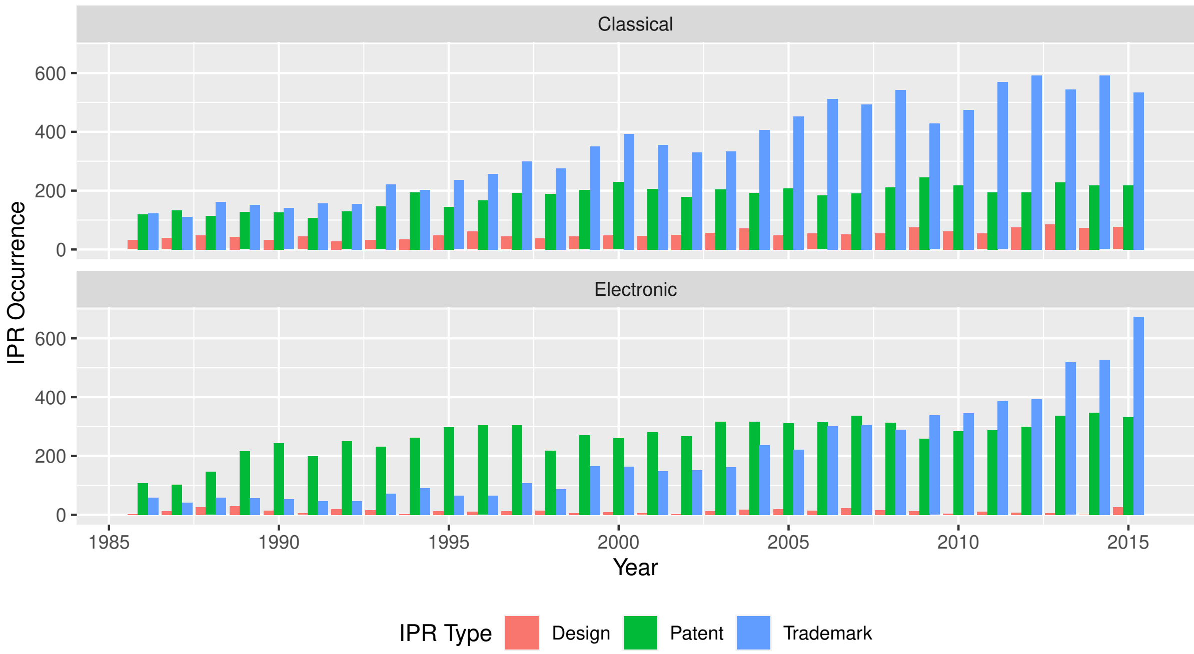


Figure F.5: Enregistrement des documents relatifs aux instruments de musique par année et type d'instrument de musique.
Les chiffres montrent l'évolution annuelle des enregistrements de documents, différenciés pour les instruments de musique classiques ou électroniques.

Chapter 5 évalue la transformation technologique des instruments de musique et l'implication des principales entreprises dans cette transformation grâce à une analyse textuelle conjointe des brevets, des marques et des dessins et modèles. Le secteur des instruments de musique combine des aspects de basse technologie et de haute technologie. Il s'agit d'un secteur avec, d'une part, une longue tradition et des produits de faible technologie tels que les instruments de musique classiques. D'autre part, il s'agit d'un secteur qui connaît une transformation technologique grâce à l'électrification et à la numérisation des instruments de musique. Même si les instruments de musique deviennent de plus en plus techniques, les enregistrements de marques dépassent les enregistrements de brevets et de modèles (voir Figure F.5).

Pour les instruments de musique, la protection des brevets, des marques et des modèles est appliquée. Pour comprendre la transformation, les entreprises concernées et l'utilisation des droits de propriété intellectuelle, le chapitre combine des données textuelles sur les brevets, les marques et les dessins et modèles dans les instruments de musique afin d'obtenir une large couverture de la transformation technologique et d'évaluer un changement d'utilisation des DPI concernant cette transformation et les entreprises concernées.

Ainsi, l'analyse fournit une perspective plus large sur la transformation, les entreprises et l'application de différents droits de propriété intellectuelle (DPI). L'analyse est séparée en deux parties : La partie I se concentre sur le secteur en général, et la partie II inclut les principales entreprises dans la discussion. Les questions connexes sont les suivantes :

RQ5.1 : Quelle est la nature de la transformation du secteur, qui est passé d'un secteur à faible technologie à un secteur à faible et haute technologie ?

RQ5.2 : Comment la transformation est-elle liée à l'utilisation des DPI ?

Les questions de recherche connexes du point de vue des entreprises sont les suivantes :

RQ5.3 : Qui contribue à la transformation ?

RQ5.4 : Comment cela se rapporte-t-il à l'utilisation des DPI ?

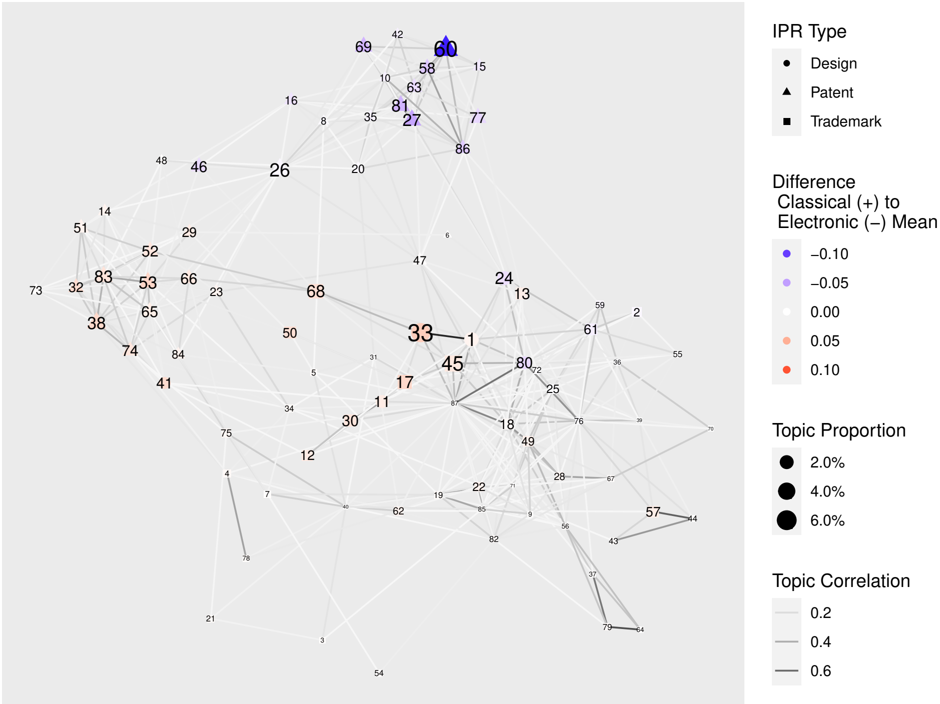
Les différents droits de propriété intellectuelle servent de base à l'analyse. Leurs données textuelles fournissent un aperçu de la transformation basé sur le contenu. À cette fin, les sources de données sont combinées textuellement et les thèmes de l'ensemble des données sont extraits à l'aide de la modélisation thématique structurelle de Roberts et al. (2019). Sur la base de l'estimation des thèmes, des réseaux peuvent être formés avec les thèmes comme nœuds et les cooccurrences de thèmes comme liens. Les réseaux sont analysés à la recherche de modèles liés au changement en cours. Dans l'analyse sectorielle, la pertinence de l'utilisation des DPI, les différents types d'instruments de musique et les instruments de musique classiques par rapport aux instruments électroniques sont évalués, et les changements dans le temps sont pris en compte. Du point de vue des grandes entreprises, le réseau thématique est ensuite enrichi d'entreprises, ce qui permet de construire un réseau entreprises-sujets.

La perspective sectorielle révèle que le secteur des instruments de musique s'est transformé, passant de la fourniture d'instruments de musique classiques à celle d'instruments électroniques et à la numérisation. Par exemple, Figure F.6a affiche le réseau thématique de 1986 à 1990 d'un point de vue sectoriel, tandis que Figure F.6b affiche la situation de 2011 à 2015. Comme on peut le constater, les thèmes relatifs aux instruments de musique classiques ont diminué, passant de 59 à 46 nœuds, tandis que les thèmes électroniques ont augmenté au fil du temps. Les thèmes électroniques sont couverts par des brevets ainsi que par des marques. Les dessins et modèles relient les thèmes relatifs aux brevets et aux marques. Les thèmes couverts évoluent vers des aspects numériques tels que la création sonore, le mobile et les jeux. Avec la transformation technologique, l'utilisation des droits de propriété intellectuelle a changé. Chaque type de DPI fournit des informations différentes : Les sujets électroniques ou les sujets liés aux données sont pertinents dans les brevets mais le sont aussi de plus en plus dans les marques. Les dessins et modèles sont un intermédiaire entre les brevets et les marques et couvrent souvent des sujets similaires à ceux des marques. L'utilisation des DPI a évolué au fil du temps, les marques devenant plus importantes pour les instruments de musique électroniques et couvrant les aspects numériques de la création de sons numériques, des mobiles et des jeux.

La perspective de l'entreprise se concentre sur les grandes entreprises à intervalles de cinq ans entre 1986 et 2015 et les classe en fonction de leurs antécédents. Ces entreprises sont les principaux demandeurs de droits de propriété intellectuelle, tous DPI confondus. Figure F.7 affiche les 87 thèmes en relation avec les principales entreprises de 1986 à 2015. L'analyse révèle que l'origine des entreprises contribuant à la transformation change au fil du temps, des entreprises d'origines diverses contribuant à la transformation du secteur. Au début, les producteurs d'instruments de musique étaient principalement actifs. Au fil du temps, des entreprises de matériel et de logiciels, des détaillants et des développeurs de jeux sont entrés dans le secteur. Leurs sujets étaient principalement liés au calcul, à la signalisation, à la capture de la voix et au traitement des données. La transformation et l'évolution technologique sont donc également le fait d'entreprises qui ne sont pas directement impliquées dans la production d'instruments de musique. L'utilisation des DPI se recoupe avec les antécédents des entreprises. Les entreprises de logiciels et de matériel informatique sont davantage liées aux thèmes des brevets, tandis que les entreprises de jeux et de vente au détail sont proches des thèmes des marques. Le positionnement des producteurs d'instruments de musique dépend du degré d'électrification de leurs instruments. Certains se positionnent plus près des thèmes liés aux brevets, tandis que la plupart se situent dans les domaines des marques ou du design. Les résultats sont conformes à la perspective sectorielle.

Dans l'ensemble, l'électrification et la numérisation des instruments de musique est un effort combiné des entreprises de matériel, de logiciels et de jeux qui conduisent à de nouveaux développements dans le secteur et des producteurs d'instruments de musique qui introduisent l'électrification dans leurs instruments. Ce chapitre montre que la transformation technologique est saisie par

(a) Réseau de rubriques d'instruments de musique différencié pour les instruments de musique électroniques et classiques de 1986 à 1990



(b) Réseau de sujets sur les instruments de musique différencié pour les instruments de musique électroniques et classiques de 2011 à 2015

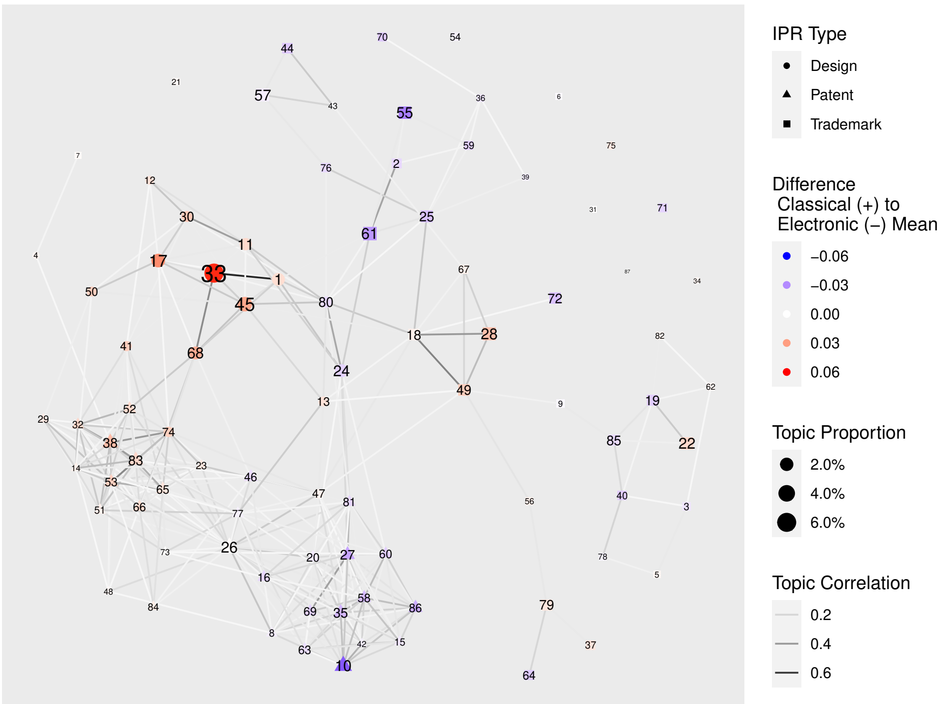


Figure F.6: Réseau de thèmes d'instruments de musique différenciés pour les instruments de musique électroniques et classiques.
Les réseaux présentent les relations thématiques entre 1986 et 1990 et entre 2011 et 2015, différenciées selon les types de DPI et les instruments de musique électroniques ou classiques. Les nœuds représentent les sujets, tandis que les liens représentent la probabilité de cooccurrence de ces sujets. La taille des nœuds représente la proportion de sujets. Les types de DPI sont indiqués par la forme des nœuds, tandis que l'information sur les instruments de musique est donnée par la couleur des nœuds. Un bleu intense signifie que le sujet est fort en instruments de musique électroniques et vice versa pour les instruments de musique classiques en rouge. Les arêtes indiquent la force de la corrélation. Les couleurs claires ont une corrélation plus faible que les couleurs foncées. Dans Figure F.6a, 59 nœuds sont plus liés aux instruments de musique classiques que 28 nœuds aux sujets électroniques. Dans Figure F.6b 46 nœuds concernent davantage les instruments de musique classiques, contre 41 nœuds pour les sujets électroniques.

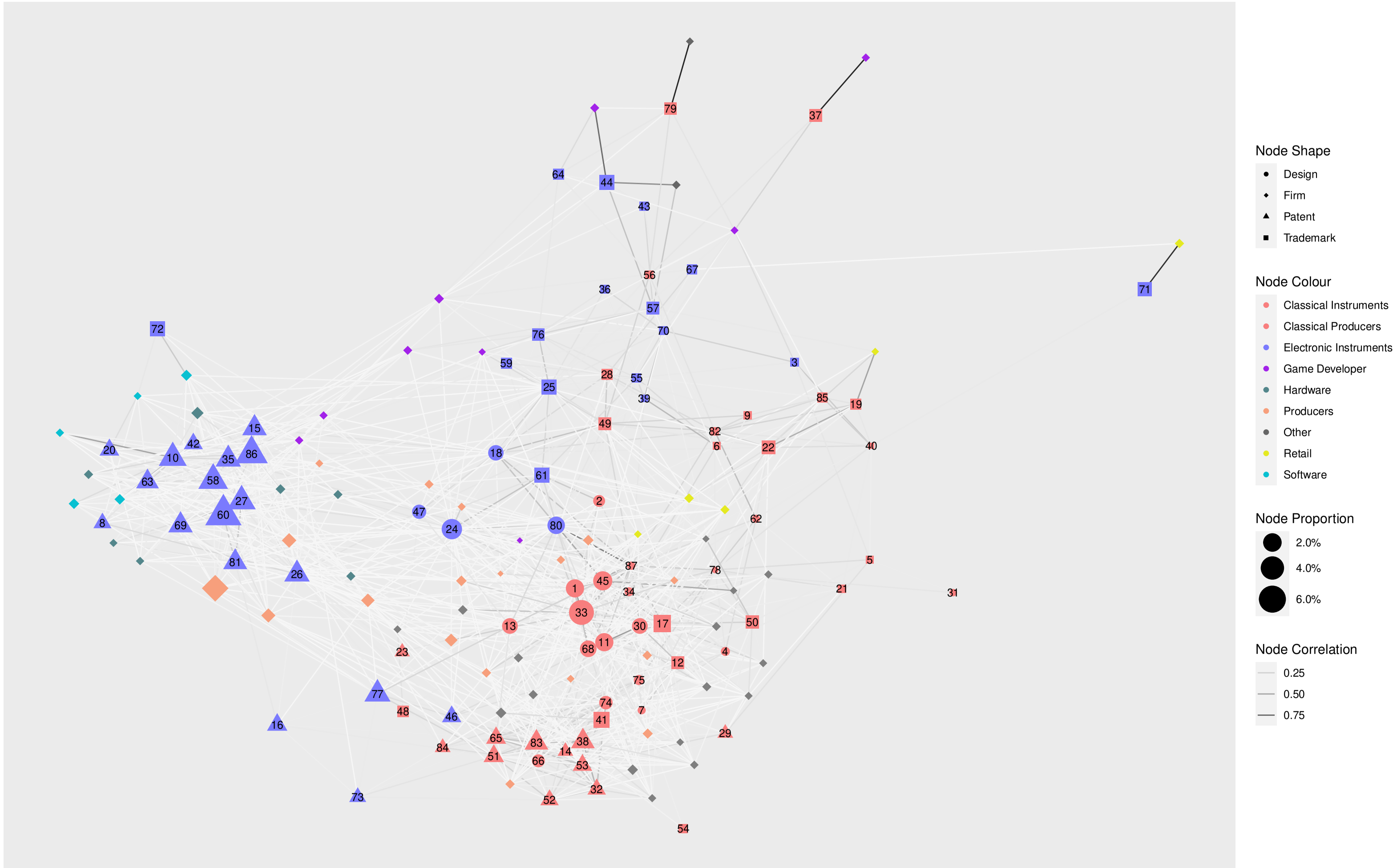


Figure F.7: Major Topics-Firms-Network from 1986 until 2015.
Le réseau présente les principales entreprises dans le domaine des instruments de musique de 1986 à 2015. Les nœuds se rapportent aux entreprises et aux différents droits de propriété intellectuelle, à savoir les dessins et modèles, les brevets et les marques. Les nœuds thématiques ont toujours des chiffres à l'intérieur, indiquant le numéro du thème. Les nœuds thématiques sont colorés en rouge ou en bleu, selon qu'il s'agit d'instruments de musique électroniques ou classiques. La forme et la couleur sont déterminées par le type de DPI ou d'instrument de musique le plus important dans l'ensemble de données. Les nœuds des entreprises ont une forme de losange et sont colorés en fonction de leur type d'entreprise. Par souci de clarté, la vue d'ensemble ne présente que la coloration des points et non la dénomination des entreprises. Les arêtes montrent la corrélation entre les entreprises et les thèmes et entre les thèmes, une couleur plus claire indiquant une corrélation plus faible.

l'utilisation combinée de différents droits de propriété intellectuelle. En particulier, l'inclusion des marques déposées ajoute à l'analyse des aspects supplémentaires tels que les jeux. L'utilisation de l'analyse textuelle a permis de découvrir des sujets liés à la transformation et à différentes entreprises. Cela a permis une analyse différenciée de la transformation. D'un point de vue méthodologique, le réseau firms-topics-network permet de mieux comprendre le positionnement des entreprises en général. Le chapitre contribue à une meilleure compréhension de l'application des droits de propriété intellectuelle dans le contexte de la transformation technologique et des entreprises concernées. Une transformation technologique est en cours dans le secteur des instruments de musique et dans d'autres secteurs, notamment grâce à la numérisation. La méthodologie du chapitre peut être appliquée pour mieux comprendre la transformation des différents secteurs, car elle couvre non seulement les aspects de haute technologie mais aussi ceux de basse technologie, et pour évaluer les entreprises à un niveau plus détaillé.

F.4 Remarques Générales sur la Thèse

Chaque chapitre de la thèse se concentre donc sur des aspects différents. Cependant, les chapitres sont reliés par divers éléments. Tous les chapitres traitent de l'innovation et de différents domaines d'innovation (voir Figure F.8). L'étendue des domaines d'innovation (voir triangle jaune) couverts est la plus large dans le chapitre Chapter 3, car ici, tous les domaines d'innovation sont pris en compte à condition que des données textuelles soient utilisées pour l'analyse de l'innovation. Chapter 4 couvre la robotique et les chaussures, alors que la première représente un domaine de haute technologie et la seconde un domaine de basse technologie. Dans le document Chapter 5, le secteur des instruments de musique est évalué comme un secteur dans lequel une transformation technologique a lieu et où les aspects de basse et de haute technologie sont pertinents. Parallèlement à une diminution de la part du domaine d'innovation, l'intégration augmente. Dans Chapter 4, les hautes et basses technologies sont analysées individuellement puis comparées pour la robotique et les chaussures, tandis que les aspects de basse et haute technologie sont analysés conjointement dans les instruments de musique. Les sources de données utilisées (voir triangle vert) au cours de la thèse se diversifient. Dans un premier temps, la recherche est basée sur un échantillon de publications relatives aux données textuelles dans la recherche sur l'innovation. Dans le Chapter 4 suivant, les données textuelles des brevets et des marques sont utilisées comme base d'analyse et les contenus textuels des différentes sources de données sont comparés les uns aux autres. Enfin, dans Chapter 5, les brevets, les marques et les dessins et modèles sont utilisés conjointement pour fournir une large couverture des innovations dans les instruments de musique et les thèmes dérivés de l'analyse sont utilisés pour identifier des modèles généraux. Figure F.8 fournit une perspective intégrée des chapitres. Les aspects communs sont colorés en conséquence.

F.5 Discussion

Si l'on considère l'ensemble des chapitres, certains aspects généraux restent à discuter. La thèse a commencé par la nécessité d'une perspective d'innovation plus large pour couvrir les domaines de basse technologie et les services non couverts par les brevets. Les marques déposées ont été traitées comme une source de données alternative pour fournir cette perspective. La combinaison des marques avec les brevets et les dessins et modèles a été proposée sur la base de données textuelles, en supposant que celles-ci offrent la possibilité de combiner plusieurs sources de données et une perspective plus large sur l'innovation et les services de basse technologie. Les données textuelles ont été prétraitées et combinées tout au long de la thèse à l'aide de la modélisation thématique structurelle. Les analyses ont révélé plusieurs points forts de l'approche mais aussi des défis qui doivent encore être relevés :

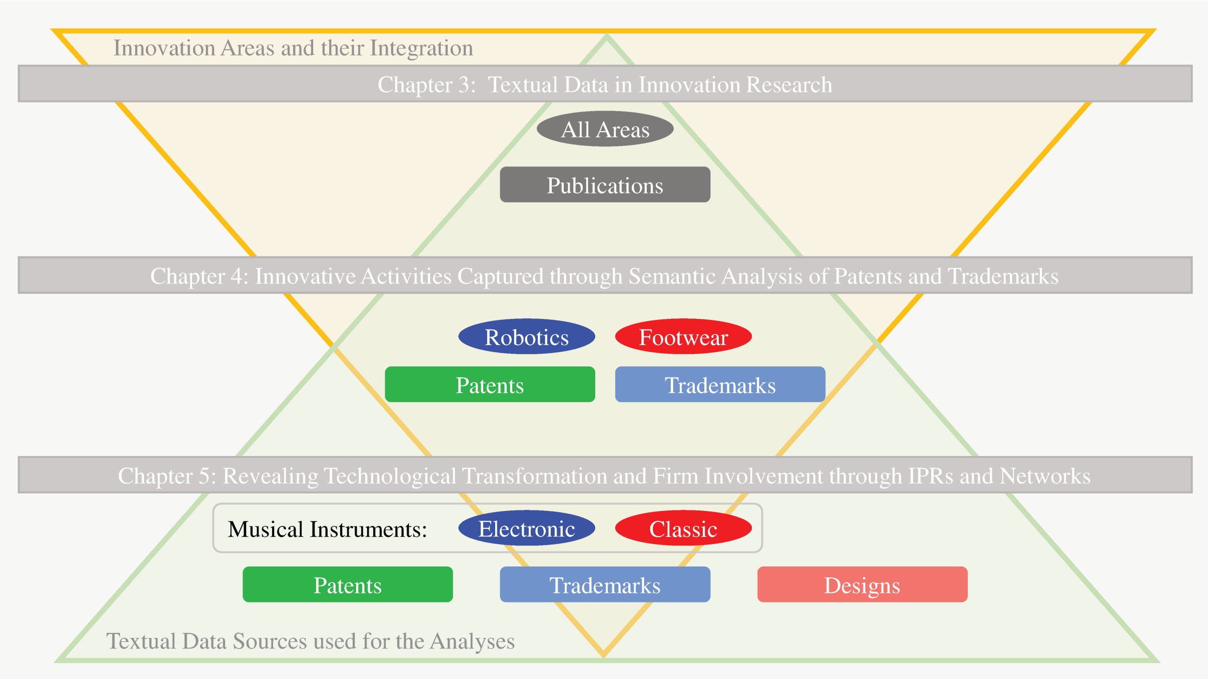


Figure F.8: Résumé de la perspective intégrée sur les chapitres de recherche
Le triangle en jaune représente l'étendue des domaines d'innovation couverts. Le triangle en vert représente l'étendue des sources de données utilisées. La couleur gris foncé des cases dans Chapter 3 indique qu'aucun domaine d'intérêt spécifique n'est couvert. L'analyse documentaire couvre un échantillon de publications. Cependant, les données textuelles appliquées dans ces publications utilisent des données textuelles provenant de diverses sources de données. Les cases ovales sont colorées en bleu foncé ou en rouge. Le bleu foncé concerne les hautes technologies, tandis que le rouge concerne les aspects de basse technologie. Les droits de propriété intellectuelle ont un contour de forme violette. Les brevets ont un contour vert, les marques un contour bleu clair et les dessins un contour rouge clair. La coloration des cases a été maintenue tout au long de la thèse en termes de brevets, marques, dessins et modèles, haute technologie et basse technologie afin de permettre une compréhension commune tout au long des chapitres.

Marques : Les données sur les marques fournissent non seulement des informations sur les services mais aussi sur les domaines technologiques. Cela est devenu évident dans le domaine de la robotique, où les marques contiennent également des informations détaillées sur les innovations. Dans le domaine des instruments de musique, les marques ont présenté des aspects de la numérisation tels que le mobile, la conception de sons numériques et les jeux. Il est intéressant de noter que le contexte des entreprises est plus déterminant pour l'application des DPI que les aspects technologiques. On aurait pu s'attendre à ce que tous les aspects de haute technologie se retrouvent davantage dans les brevets. Cependant, les entreprises spécialisées dans les logiciels et le matériel informatique sont actives dans le domaine des brevets, tandis que les entreprises spécialisées dans les jeux vidéo sont actives dans le domaine des marques. Dans ce cas, des recherches supplémentaires sont nécessaires pour déterminer les différences qui expliquent ce comportement. Une raison pourrait être la rapidité du développement dans le secteur des jeux et l'importance du marketing et de la proximité du client.

Données textuelles : Les données textuelles ont permis d'analyser divers domaines d'innovation d'une manière axée sur les données. De grands ensembles de données ont pu être couverts et des tendances ont été détectées sans l'intervention d'experts. Les marques, les brevets et les dessins et modèles ont pu être combinés et analysés conjointement. L'accent mis sur les informations textuelles apporte ainsi une perspective plus large, comme le montrent les analyses des instruments de musique. Les données textuelles ont fourni des informations pour expliquer, par exemple, l'utilisation des marques dans les instruments de musique électroniques. De plus, l'approche a permis le positionnement thématique des entreprises qui montre leur implication dans la transformation. Au fil du temps, il est devenu évident que des entreprises ayant des antécédents différents ont été impliquées, ce qui a introduit de nouveaux sujets dans la transformation des instruments de musique.

Lors du prétraitement et des analyses, il est apparu que les données textuelles des différents droits de propriété intellectuelle diffèrent, notamment en ce qui concerne les mots juridiques et courants utilisés, la structure des phrases et le niveau de détail. Ces différences affectent la capacité de combiner et de comparer les différentes sources de données. Le type de DPI des documents a donc été pris en compte lors du prétraitement afin de tenir compte de ces différences : Les mots très courants spécifiques à chaque DPI ont été pris en compte et supprimés. Lors du processus de nettoyage, nous avons pris soin de préserver les différences entre les sources de données. Cependant, cela a entraîné l'inclusion de mots sans signification dans les ensembles de données finaux. Afin d'améliorer encore l'analyse de sources de données multiples, il pourrait donc être intéressant de disposer d'une liste de mots vides spécifique à chaque droit de propriété intellectuelle, contenant notamment les termes juridiques courants du droit de propriété intellectuelle concerné. Ainsi, il serait possible de se concentrer davantage sur les mots significatifs tout en tenant compte des différences de contenu entre les sources de données.

Un autre aspect à prendre en compte dans l'analyse des données textuelles des documents DPI est que les documents n'étaient pas également représentés dans les différents domaines d'innovation. Dans le domaine de la robotique, la plupart des documents étaient des brevets, tandis que dans celui des chaussures, la plupart des documents étaient des marques. Dans l'ensemble de données sur les instruments de musique, les brevets et les marques étaient également représentés, tandis que les dessins et modèles étaient rarement inclus. La part de chaque type de DPI dans l'ensemble de données a un impact sur les thèmes estimés, car les thèmes sont estimés sur la base des documents fournis et penchent donc vers le type de DPI ayant la part la plus élevée. Cela signifie que plus la proportion d'un DPI est faible, plus les incertitudes de l'estimation concernant ce DPI sont grandes et, par conséquent, plus la variance est importante. Nous avons vu dans les différents chapitres que, tant que la cohérence et le lien entre les documents et les sujets sont assurés, la différence dans la part des DPI est négligeable. En robotique, les thèmes étaient très cohérents. Dans le secteur de la chaussure, les thèmes ayant un rapport élevé avec les brevets étaient également cohérents. Dans le cas des thèmes relatifs aux marques, la cohérence dépendait de la structure des marques. Dans les cas où les marques ne fournissaient que des articles d'habillement, la cohérence était plus faible, et les thèmes fournissaient moins de détails sur la production ou l'innovation contenues. Dans le domaine des instruments de musique, les thèmes relatifs aux entreprises de matériel et de logiciels, et principalement les brevets, présentent une cohérence élevée parmi les principaux documents de brevets, de marques et de modèles. Cependant, la cohérence des sujets relatifs aux marques dépend des aspects électroniques. Dans l'ensemble, la cohérence et, par conséquent, la qualité diffèrent en fonction de la structure textuelle. Une relation technologique élevée conduit souvent à une qualité élevée des sujets. En outre, il faut tenir compte de la qualité et de l'utilisation des DPI dans le domaine couvert.

Enfin, la longueur du texte doit être prise en compte. Les marques contiennent principalement des textes courts ou des listes de mots concis. Les dessins et modèles sont également courts dans leur description textuelle. Les brevets fournissent des descriptions détaillées ou des résumés courts. Cette thèse a utilisé les résumés des brevets pour s'aligner sur la longueur textuelle des marques et des dessins et modèles. Cette approche a fonctionné pour la combinaison. Pour les brevets, il pourrait être intéressant d'étendre l'analyse concernant la description du brevet pour atteindre un niveau de détail encore plus élevé.

Données textuelles des marques : La thèse s'est concentrée sur les données textuelles des marques en combinaison avec d'autres sources de données pour obtenir une perspective plus large sur l'innovation. Parmi les articles de l'échantillon de Chapter 3, aucun article n'a utilisé les marques de commerce comme source de données textuelles, ce qui est en accord avec les résultats des autres. Cependant, tout au long de cette thèse, il est démontré que les données textuelles des marques contiennent des informations sur l'innovation et offrent une opportunité de combiner les marques avec d'autres sources de données. Cela permet d'obtenir des informations supplémentaires sur le positionnement des entreprises ou sur les innovations dans différents domaines d'innovation.

Les analyses des données textuelles des marques, brevets et dessins ont permis de découvrir plus en détail la robotique, les chaussures et la transformation des instruments de musique, notamment par rapport aux classifications disponibles. Néanmoins, la quantité de texte et le type de structure du texte sont également déterminants pour le niveau de détail : Même si les descriptions de marques sont relativement courtes, il existe des différences dans leur structure. Les descriptions de marques de robotique ou d'instruments de musique couvrant des aspects de haute technologie présentaient une description et des détails plus étendus que les marques dans, par exemple, le domaine de basse technologie des chaussures, où la description de la marque ressemblait à une liste de produits sans plus de détails. Ici, des recherches supplémentaires sont nécessaires pour comprendre les raisons de cette différence et potentiellement se concentrer sur les descriptions de marques avec un niveau élevé de détails.

L'analyse des données textuelles des marques dans cette thèse a été utilisée pour explorer les tendances. Le sujet modellin

Innovation dans les données textuelles des marques Cette thèse traite de l'hypothèse selon laquelle les textes de marques représentent des aperçus intéressants sur les innovations. Ceci malgré le fait que le lien entre les marques et l'innovation n'est pas garanti. Les analyses de la robotique, des chaussures et des instruments de musique ont pu montrer que les marques fournissent des informations détaillées sur l'innovation, comparables aux brevets. Elles pourraient assurer la perspective de l'introduction sur le marché et de la diffusion. Cependant, des différences dans les textes des marques ont été trouvées pour influencer la façon dont les textes des marques contribuent à la compréhension de l'innovation. Cette structure différente pourrait être liée à la mesure dans laquelle les marques représentent l'innovation et nécessite une enquête plus approfondie.

Dans l'ensemble, les données textuelles des droits de propriété intellectuelle fournissent des informations intéressantes sur le développement de l'innovation. La thèse a abordé les données textuelles des marques et a révélé leur contribution.

F.6 Marques de Conclusion et Recherches Futures

Pour conclure, la thèse ajoute à la discussion sur les données textuelles des marques dans la recherche sur l'innovation. Elle a donc analysé l'état actuel de l'analyse des données textuelles dans la recherche sur l'innovation, comparé les données textuelles des marques et des brevets et appliqué les données textuelles des brevets, des marques et des dessins pour mesurer la transformation technologique. Globalement, les données textuelles des marques ont permis de combiner plusieurs sources de données. L'inclusion des marques a permis de saisir les aspects liés aux services et à la basse technologie. Les données textuelles des marques ont ajouté des aspects tels que la numérisation des instruments de musique. La thèse contribue à une meilleure compréhension de l'utilisation des données textuelles, débloque les données textuelles des marques et applique les données textuelles des brevets, des marques et des dessins pour évaluer la transformation technologique. La combinaison de différents droits de propriété intellectuelle par le biais de données textuelles offre une perspective plus large sur l'innovation tout en surmontant les limitations existantes. La combinaison des marques, des brevets et des dessins et modèles renforce les points forts de chaque source de données pour la recherche sur l'innovation. Les données textuelles fournissent en outre des informations détaillées permettant de comprendre le développement technologique ou le comportement des entreprises. Cette thèse contribue au domaine de l'utilisation des données textuelles pour la recherche sur l'innovation et en particulier à l'application de l'analyse des données textuelles des marques. Dans un contexte plus large, la thèse contribue à l'élaboration de politiques fondées sur des données probantes en améliorant notre compréhension et nos capacités de mesure de l'innovation. Cette amélioration permet de formuler des recommandations plus complètes pour les initiatives politiques, en particulier dans les domaines d'innovation à faible technologie.

Des recherches supplémentaires sont nécessaires. Les défis mentionnés ci-dessus restent à relever. Dans cette thèse, les documents de brevets, de marques et de dessins et modèles analysés ont été identifiés en utilisant des recherches basées sur des classes et des mots clés. Cela est nécessaire car il n'existe pas de système de classification commun aux différentes sources de données. En même temps, il faut s'assurer que les critères des documents à inclure sont similaires dans toutes les sources de données. Dans le cas de domaines d'intérêt plus complexes, comme les technologies vertes dans les pays en développement, il est intéressant d'adopter une approche plus sophistiquée pour identifier les données pertinentes dans différentes sources de données, car la qualité des résultats dépend de celle des données utilisées. En outre, le lien entre l'innovation et les données textuelles des marques pourrait être amélioré : Dans les zones ayant une relation technologique élevée, les marques déposées dans cette thèse ont affiché un niveau élevé de détails. Cependant, dans les domaines à faible intensité technologique, comme les vêtements, les descriptions des marques couvraient principalement une liste d'articles de vêtements. Ici, une approche visant à différencier les descriptions de marques pertinentes et non pertinentes pour la recherche sur l'innovation pourrait être intéressante pour augmenter la concentration sur l'innovation. Enfin, les connaissances acquises par l'approche, comme le contexte des entreprises étant décisif pour l'utilisation des DPI, restent à inclure dans d'autres modèles économiques pour améliorer leurs résultats.

Université de Strasbourg

Augustin Cournot Doctoral School

Fenja SCHEU

Trademarks and Textual Data : A Broader Perspective on Innovation

(Marques et données textuelles : Une perspective élargie sur l'innovation)

Résumé

Les brevets mesurent l'innovation technique, tandis que les marques couvrent l'innovation de basse technologie et de services. Cette thèse explore les données textuelles des marques pour combiner divers droits de propriété intellectuelle. Les données textuelles permettent l'analyse de données volumineuses, la combinaison de sources diverses et l'obtention d'informations fondées sur des données. La combinaison de la robotique (haute technologie) et des chaussures (basse technologie) offre une couverture et des détails accrus sur l'innovation, variant selon les secteurs. Dans le secteur des instruments de musique, les données textuelles sur les marques, les brevets et les dessins et modèles mettent en évidence les transformations technologiques en cours. Les brevets couvrent les données et la numérisation, tandis que les marques couvrent le traitement des signaux et les jeux vidéo. Les dessins et modèles servent d'éléments de liaison. Il est possible de différencier les entreprises et les domaines d'intervention. En conclusion, la thèse montre que l'intégration de données textuelles sur les marques élargit la couverture de l'innovation.

Mots clés :
Innovation, brevet, marque, dessin, droit de propriété intellectuelle, IPR, texte, transformation, STM, Structural Topic Modelling

Résumé en anglais

Patents often measure technical innovation, while trademarks cover low-tech and services. This thesis explores textual data of trademarks to combine different intellectual property rights. Textual data enable, for example, large data analysis, the combination of diverse sources, and data-driven insights. The combination of trademarks and patents in Robotics (high-tech) and Footwear (low-tech) provides broader coverage and detail on innovation, varying by sector. In the musical instrument sector, textual data of trademarks, patents, and designs highlight ongoing technological transformation. Patents cover data and digitalization subjects and are used by high-tech firms, while trademarks cover signal processing and video games of gaming firms. Designs act as a linking element. Differentiation between firms and areas of involvement is possible. In conclusion, the thesis reveals that integrating textual trademark data broadens innovation coverage.

Keywords :
Innovation, patent, trademark, design, intellectual property right, IPR, text, transformation, STM, Structural Topic Modelling