Exploiting Industry 4.0: Mapping Firm Level Technology Development and Adoption across the U.S. Using Patent and Trademark Data

Justine Miller

Student number: 5891256

Under the supervision of Prof. Dr. Carolina Castaldi

Second reader: Prof. Dr. Beate Volker

Word count: 14,924

Utrecht University

Utrecht, June 2020

Abstract

The Fourth Industrial Revolution, Industry 4.0 (I4), will impact economies and labour markets in pervasive and substantial ways. However, studies are still establishing how it will do so, and which technologies will account for these changes. In order to monitor I4, it is crucial to measure invention and innovation, as they are key to understanding how labour markets could be impacted by the industrial revolution. Therefore, this thesis studies how firms have been developing and adopting I4 technologies in the United States (U.S.), across space and time. I use both trademark and patent data for the period 2008-2017, addressing recent calls in the literature for combining both intellectual property rights data. Using a keyword filter, I am able to establish that patents indicate that the geography of invention is rather concentrated in certain U.S. states and cities, and that specialised firms dominate filings in core technologies. Similarly, trademark filings show that innovation diffusion is in general more spread across space even though spatial concentration in cities appears to be increasing. Additionally, the leading trademarking firms display lower shares of intellectual property rights ownership between 2008 and 2017, which suggests more spread filing activities across industries such as electronic gaming and gambling. The findings indicate that developments of I4 technologies are more concentrated than their applications both spatially and at the firm level, and further suggest mixed implications for labour markets.

Acknowledgments

I would first like to express my gratitude to my supervisor Prof. Dr. Carolina Castaldi who throughout this thesis gave me support and clear guidance. Her door was always open for advice, and I really enjoyed being pushed to go above and beyond to improve this thesis.

I would also like to thank family and especially my parents for their unfailing support. I'm sure that for them, reading and processing this thesis must not have been the lightest read. I would also like to thank my "quaranteam", Anaya, Anna, and Rea who made this lockdown much more enjoyable. Elsa, thank you for those daily sports Skype sessions and for helping recreate a semblance of sports form home. Haydn, our regular Skype sessions were always a great time to chat over the most random range of topics. You all helped me not go crazy in my small apartment for these past three months.

Table of Contents

List of Abbreviations

Abbreviation	Definition
AI	Artificial intelligence
EU	European Union
EPO	European Patent Office
EUIPO	European Union Intellectual Property Office
GPT	General purpose technology
I4	Industry 4.0
I4T	Industry 4.0 Technologies
IBM	International Business Machines
ICT	Information and Communication Technology
IT	Information Technology
IoT	Internet of Things
IPR	Intellectual Property Rights
JPO	Japan Patent Office
KIPO	Korean Intellectual Property Office
OECD	Organisation for Economic Co-operation and Development
R&D	Research & Development
CNIPA	State Intellectual Property Office of the People's Republic of China
U.S.	United States
USPTO	United States Patent and Trademark Office
WIPO	World Intellectual Property Organisation

List of Tables and Figures

1. Introduction

Industry 4.0 – the fourth industrial revolution – is set to radically impact industrial production and socio-economic systems. Industry 4.0 (I4) is characterised by the increased level of combinations between the virtual and real worlds, and consists of technologies such as the Internet of Things, cyber-physical systems, and artificial intelligence (AI) (Kumar et al., 2019; Ménière et al., 2017). Estimates indicate that the economic relevance of I4 will be significant as it could contribute to a $7trillion market in hardware and software development by the end of 2020 (Baldassari & Roux, 2017). Moreover, it also has the potential of universally impacting society by fostering medical advancements for instance, or by allowing firms and individuals to use resources more efficiently and pollute less. Essentially, I4 will assist in the digitalisation of manufacturing processes and in the automation of work. Several ongoing debates concern the effects these changes will have on labour markets: some jobs might disappear, and the sets of skills required by firms will most likely evolve (Pereira & Romero, 2017). In any case, studying the geographical unfolding of I4 is therefore key to guarantee labour markets are spared. How firms across countries embrace I4 and its technologies will define the extent to which these processes will impact local and national labour markets (Reischauer, 2018; Baldassari & Roux, 2017).

Industrial revolutions are either defined based on the innovation principles and paradigms they introduced under so-called technological revolutions (Perez, 2010; Grinin & Korotayev, 2015), or based on a few key technologies, also known as general purpose technologies, that enabled them (Bekar et al., 2018). Accordingly, I4 can be described as a new industrial revolution. It consists of the development of innovation principles through the convergence of connected technologies (Chiarello et al., 2018; Ménière et al., 2017), and it is often associated with certain crucial technologies whose characteristics are capable of radically impacting growth paths (Klinger et al., 2018; Bekar et al., 2018). Measuring innovation in I4 will therefore prove crucial in monitoring its diffusion and impact across economies. Consequently, the literature has turned to delimiting the scope of the term "Industry 4.0" and has yet failed to reach a consensus, although considerable progress has been made (Pereira & Romero, 2017; Chiarello et al., 2018; Ciffolilli & Muscio, 2018). Some studies have identified the main I4 technologies based on discourse analyses, taxonomy and literature reviews (Oztemel & Gursev, 2020; Ciffolilli & Muscio, 2018; Reischauer, 2018), while others have followed quantitative approaches and used bibliometric (Janik & Ryszko, 2018) or keyword searches (Balland & Bochma, 2019; Chiarello et al., 2018).

This thesis aims to address this lack of consensus on how I4 technologies are being exploited. More specifically, it contributes an original empirical study, by adding to both innovation and regional studies. Firstly, it provides a methodological contribution using both patent and trademark data to provide a more complete picture of the engagement of the United States (U.S.) with the ongoing industrial revolution and its technologies. Using both patent and trademark data allows me to benefit from the advantages of both intellectual property rights (IPR) types, while also addressing some of the shortcomings of patents with trademark data. I select the data applying a keyword filter, something new in the literature on trademarks (Castaldi, 2019). Additionally, I contribute to the literature on innovation diffusion by associating patents with invention in I4, and trademarks with innovation in I4. Secondly, this research is the first that combines a geographical and industry-level analysis to monitor I4. Previous studies have shown that some places display differing capacities to embrace I4 (Balland & Bochma, 2019). Other studies have also demonstrated that some regions may perform well in terms of invention but poorly in innovation levels (Filippetti et al., 2019). These two branches of literature highlight the urgency of studying how I4 and its technologies will unfold across the U.S. Accordingly, this research adds to the literature on I4 by considering the following question: "*How are firms developing and adopting I4 technologies across the U.S.?*"

Firstly, building on the literature that aims to identify the scope of Industry 4.0, I determine which technologies are being exploited by firms to answer the following sub question: "*How can patent and trademark filings reveal patterns of development and adoption of I4 technologies?*" In order to differentiate between the development and application of I4 technologies, I build on strands of literature that have highlighted the potential offered by trademark data to complement innovation studies (Castaldi, 2019; deGrazia et al., 2020; Gotsch & Hipp, 2012; Filippetti et al., 2019). For instance, trademarks offer firms attractive alternatives to capture non-technological and/or service-oriented innovations while being more easily obtainable than patents (Castaldi, 2019). In addition, they are capable to a certain degree of capturing firm strategies (Castaldi, 2019; Gao & Hitt, 2012), hence providing opportunities to further study the breadth of Industry 4.0. Therefore, I compile a list of keywords for different technology categories according to the literature on I4 and apply it to both patent and trademark data. Across all technology categories, I find increasing filing levels in I4 only until 2015 as the end of the sample may be affected by truncation. Moreover, I conclude that patents can be associated with hardware-related technologies, while trademarks can be more closely identified with varied services.

Secondly, borrowing tools from spatial economics, the research considers the following question: "*To what extent are I4 technologies being developed and applied spatially across the U.S.?*" Therefore, it sheds light on the geographical spread of I4 across the U.S. by comparing IPR filings throughout U.S. states and cities. Furthermore, before doing so, I study how firms exploit I4 technologies as they will be the actors who determine how the industrial revolution spreads across space. Results at the firm-level indicate that large and specialised firms in semiconductor and computer hardware industries dominate patenting activity, and slightly more diversified and international firms lead trademark rankings in activities such as electronic gambling and gaming. In addition, firms with the highest levels of patenting in I4 owned larger shares of IPR over the period studied compared to the firms with the most trademarks in I4. The geographical analysis shows that IPR filings are clustered within certain states and cities, although diffusion patterns are slightly more noticeable in trademark-related innovative activity. Moreover, both invention and innovation in I4 seem to have become more concentrated over the years. Overall, the geography of innovation appears to fluctuate more over time.

The conclusions reached in this research allow for a better understanding of the development and application of I4 technologies and provide insights as to how U.S. states and cities may be exploiting said technologies. They also open avenues for future studies to use trademark data, despite a few clear limitations: truncation of the data especially present in patent filings, inventor addresses being restricted to the U.S., and drawbacks of applying word filters on IPR data. Moreover, this study paves the way for further research to specifically examine the spatial implications for firms in U.S. states and cities that exploit I4 technologies or fail to do so.

2. Technological Change and Revolutions

2.1. Theories of Industrial Revolutions

History is often described as being shaped by the impact that radical technological breakthroughs, such as the steam engine, have had over time on societies. As such, economists have set to establish the mechanisms by which technologies were able to affect the growth paths of countries. While it has been widely agreed that technological change can lead to economic growth (Perez, 2010; Klinger et al., 2018; Petralia et al., 2017; Bresnahan & Trajtenberg, 1995), different theories have been introduced to explain the underlying mechanisms. The linear approaches which assume that technological change led to homogenous development have largely been discarded in favour of theories such as "technological revolutions" and "general purpose technologies". Technological revolutions are defined by combining insights from neo-Schumpeterian and evolutionary economics to argue that socio-technical systems evolve under paradigms which define and shape common perceptions, social rules, and future trends (Perez, 2010; Grinin & Korotayev, 2015). Consequently, this indicates that technological change and innovation are intrinsically linked, as one drives the other and vice versa.

However, the discourse on innovation has highlighted the ambiguity of this relationship between innovation and technological development (Fagerberg, 2005). The main argument resides with the fact that innovation implies novelty, although classifying ideas under the label 'new' can be somewhat ambiguous (Smith, 2005). For instance, the introduction of a new type of phone tends to be collectively considered as innovation, but a one-hour increase in the battery life of a traditional phone may be a more questionable novelty. The differences between the two are subtle: the former example involves a radical rethinking of capabilities and a challenge of established standards, while the latter builds upon previous technological advancements and only partially refines standards (Smith, 2005). In his 1990 paper, Romer develops a model of endogenous growth that partially accounts for this uniqueness aspect: he introduces a parameter for the relatedness of the "innovation". It establishes that an inventor creating a completely novel invention can reap more benefits, using patents, because of the unrelatedness of his creation compared to prior inventions.

The link between innovation and technological change is further emphasized by the theory of technological revolutions considering that the socio-technical systems also affect so-called "innovation principles" according to Perez (2010). Accordingly, she describes the first revolution, more famously known as the Industrial Revolution (ca. 1760), as a time that was characterised by innovation principles such as the predominance of mechanisation of production, productivity goals. She then argues that the learning trajectories, determined by the paradigms, lead to path-dependent and inter-related innovations. Those can thus bring about a technological revolution that has the potential to radically affect the current paradigm while opening avenues for wealth creation (Perez, 2010; Grinin & Korotayev, 2015; Dosi & Nelson, 2010). Ultimately, this theory addresses the ambiguity of the innovation definition: technological change introduces the "invention" side of innovation, while the paradigms enable the diffusion of those technologies in applications across all economic sectors thus highlighting the "innovation" aspect of the term innovation.

The theory of general purpose technologies (GPTs) is an alternative approach to linear models that also aims to understand the mechanisms linking technological change and economic growth. This theory, like Perez's, also builds upon evolutionary notions but is more closely related to growth models (Bekar et al., 2018). GPTs can be defined as crucial technologies, such as the steam engine, electricity, and microprocessors, that radically affected growth trends (Bresnahan & Trajtenberg, 1995; Perez, 2010; Petralia, 2019). Three main features allowed these key technologies to essentially change socio-economic landscapes and shape new socio-technical regimes: pervasiveness, their capacity to ensure increasing returns-to-scale, and to complement innovation (Petralia, 2019; Bresnahan & Trajtenberg, 1995). GPTs are dynamic and evolve over time: once they are introduced to markets, they continue their transformations and can become more efficient, which further contributes to the increased productivity levels, and improves their potential for commercialisation (Helpman & Trajtenberg, 1996; Bresnahan & Trajtenberg, 1995).

However, criticisms of the term deplore the difficulty of categorising technologies into GPT or non-GPT categories (Bekar et al., 2018). Ultimately, it seems that while most technologies bear some characteristics of GPTs, they have not managed to have such wide impacts on regimes throughout the world as technologies considered to be GPTs have achieved. Nevertheless, studying GPTs is highly relevant considering the study of Petralia (2019) who looked at the economic impact of places that had adopted GPTs. More specifically, he investigated the case of Electrical and Electronic technologies in the U.S. and concluded that GPTs were "engines of growth". The author found that places that were able to produce patents related to those technologies benefited from a higher economic growth and paid higher wages compared to places that did not manage to do so.

To sum up, an industrial revolution can be essentially boiled down to: (1) the degree of pervasiveness of its GPTs, and (2) the underlying innovation principles of its technological paradigm – or the main driver of the technological developments. Table 1 links the theoretical concepts introduced so far to the previous industrial revolutions societies have undergone. As shown in Table 1, in the mid-18th century, the first – and possibly the most famous – industrial revolution began and was mainly enabled by the steam engine. The pervasiveness of the technology was mostly notable due to how it impacted industries. For instance, where some factories were once restricted to locating near water streams, steam liberated them from these restrictions. Alternatively, industries were finally capable of increasing their productivity by replacing the old less powerful machines. Later on, with the development of steam locomotives, they became able to transport and distribute their goods faster and across wider market pools. The main innovation principle of this revolution was the aim to mechanise production lines, increasing the productivity of factories in the process.

About a century later – around 1870 –, advances in electrical engineering sparked the production of conveyor belts and assembly lines, which triggered the second industrial revolution. It was marked by the innovation principle of mass production and thus built upon the foundations of the previous revolution (Broadberry & O'Rourke, 2010). Moreover, the pervasive aspect of these technological developments also led to the adoption of infrastructure systems such as water and sewage systems. In this sense, this revolution was the first one for which the new inventions and innovations permeated not only across the work environment, but also across the whole society (Broadberry & O'Rourke, 2010).

The third industrial revolution saw the arrival of electronics and information and communication technologies (ICT) (Perez, 2010; Petralia, 2019). There exists some debate in regard to the starting date of this revolution based on which specific technological advancements authors consider most important. Essentially, the revolution can be said to have begun in earnest with the invention and proliferation of transistors and digital circuits, around the 1950s-60s (Cortada 2012). Nevertheless, the technologies were pervasive because they impacted workplaces and private lives, possibly even more so than the previous industrial revolution (David & Wright, 2003). In particular, phones, personal computers, and the rise of the Internet largely contributed to the increased (digital) connectivity of the world. The increased connectivity and the development of automated production lines were the key innovation principles which defined the revolution.

More recently, the fourth industrial revolution, I4, started and began emphasising the connectivity and digitalisation of both work and personal lives. The development of GPTs such as artificial intelligence (van Beuzekom et al., 2020) or deep learning and the creation of systems such as the Internet of Things (IoT)[1] accentuated the pervasive aspect of I4. Furthermore, the high profitability prospects make these technologies capable of radically impacting growth paths (Klinger et al., 2018; Bekar et al., 2018). I4 focuses on connected technologies enabling the link between virtual and real worlds which foster the main innovation principles of this revolution and have spurred the developments of "intelligent homes" and "intelligent workplaces" (Chiarello et al., 2018; Ménière et al., 2017).

Table 1. Technological Change and Revolutions: key characteristics based on two schools of thought.

Industrial Revolutions	Time period	GPT/Technological Revolutions	Innovation principles	Leading country/ies
1st Industrial Revolution	Ca. 1760	Steam engine	Mechanisation of production lines	Britain
2nd Industrial Revolution	Ca. 1870	Electrical engineering	Mass production via efficient assembly lines	Britain, Western Europe (especially Germany) and U.S.
3rd Industrial Revolution	1950s – 1960s	ICT, electronics	Automation and connectivity	U.S., EU, East Asia
4th Industrial Revolution	2010's	Artificial intelligence, deep learning	Digitalisation of work and personal lives	U.S., East Asia, EU

2.2. The Geography of Industrial Revolutions

The first three industrial revolutions have shown that the shifts in techno-economic paradigms and the development of GPTs can affect the geography of innovation (Balland & Boschma, 2019; Klinger et al., 2018). Different theories aim to explain how these geographical patterns came into existence. On one hand, geographical proximity has been identified as one of the ways through which firms and

[1] The Internet of Things (IoT) is a digital environment which contains different manufacturing resources that are embedded and connected by digital tools and sensors (Kumar et al., 2019)

various actors can exchange knowledge (Boschma, 2005). For instance, studies have found that technological innovation is concentrated in large cities and slow to diffuse across space (Petralia et al., 2017; Jaffe et al., 1993; Balland et al., 2020). Similarly, technological change has also been found to be negatively associated with distance: larger geographic distances hamper the diffusion of innovation (von Graevenitz et al., 2019).

On the other hand, geographical proximity fails to account for all the ways through which innovation can take place and spread across space (Boschma, 2005). For instance, Britain was the country that initiated many technological developments during the first industrial revolution, which guaranteed its position as a global leading economy. Path dependency played a key role in enabling the rise of the country via geographic (large coal supply), institutional (little guild restrictions), and historical (recent wars' impact) factors (Grinin & Korotayev, 2015). During the second industrial revolution, the combination of advances in steel manufacturing and access to an abundance of raw materials enabled the U.S. and countries of Western Europe to become heavily industrialised. Britain remained a world leader during the two first industrial revolutions but was slowly overtaken by the U.S. and East Asia when in the late 20th century, globalisation as well as digitalisation technologies helped integrate manufacturing processes throughout the world. These shifts in the geography of innovation also took place at lower spatial levels. For example, the Rust Belt states who used to be manufacturing leaders during the end of the second industrial revolution experienced an industrial decline from the 1960s onwards. The Sun Belt[2] states rose at their expense as new leaders in ICT and the renowned Silicon Valley established itself as a high-tech cluster (Balland & Boschma, 2019).

The previous examples illustrate that cognitive and institutional proximities played a key role in the unfolding of industrial revolutions (Boschma, 2005). In particular, cognitive proximity can help understand the geography of industrial revolutions through the literature on relatedness. The concept studies, among others, how regions could either participate in these "invention" or "innovation" processes by defining the concepts of related and unrelated variety. Related variety states that regions are more likely to diversify into new activities that are more closely linked to the existing capabilities (Hidalgo et al., 2018; Balland & Boschma, 2019; Boschma, 2005). Conversely, unrelated variety is more often associated with radical shifts from the existing capabilities which would allow places to exit cases of lock-in that are often detrimental to innovation levels (Boschma, 2005). Furthermore, factors such as the complexity of knowledge should be considered (Balland & Rigby, 2017). The idea is that knowledge that is harder to replicate or imitate will render the underlying invention more valuable and will also limit its capacity to be diffused across space (Balland & Rigby, 2017). Consequently, industrial revolutions, through the arrival of new and more advanced technologies have the capacity to affect the diffusion of innovation.

Clusters of industries are especially vulnerable because new technologies have the potential to either completely erase any competitive advantage companies in the cluster may have had over other locations, or they may on the contrary reinforce the cluster's dominance (Klinger et al., 2018). Ultimately, the first scenario can lead to job losses whereas the second case can lead to new employment opportunities. Moreover, Petralia (2019) has demonstrated how local economies that adopted certain GPTs benefited economically. To sum up, understanding the geography of industrial revolutions and innovation diffusion requires examining how firms are adopting the technologies and enabling invention and innovation processes. In this spirit, many countries are trying to seize the economic opportunities offered by I4 while safeguarding their labour markets from risks related to the automation of work (Liao et al., 2018; Ciffolilli & Muscio, 2018). Prior studies have accordingly studied how countries or firms were embracing I4 and some of its technologies (Balland & Boschma, 2019; van Beuzekom et al., 2020; Dernis et al. 2019). To my knowledge, no research has yet conducted a geographic study enriched with a firm-level analysis of the unfolding of I4 across the U.S.

[2] The Sun Belt states refer to an area that stretches from the Pacific to the Atlantic and which covers states such as California, Nevada, Arizona, New Mexico, Texas, Louisiana, Mississippi, Alabama and Georgia

3. Industry 4.0

3.1. Definition

Just like the steam engine triggered the first technological revolution, electrical engineering advances prompted the second Revolution, and ICT launched the third one (Perez, 2010; Petralia, 2019; David & Wright, 2003), the recent advent of digital technologies sparked the fourth industrial revolution (Kumar et al., 2019; Klinger et al., 2018). This period is characterised by the prominence of digital technologies, IoT, Big Data, AI and other similar technologies which are slowly and steadily gaining importance in our daily lives (Kumar et al., 2019; Ménière et al., 2017). Subsequently, researchers have pinpointed the term Industry 4.0 to designate "the full integration of information and communication technologies in the context of manufacturing and application areas such as personal, home, vehicle, enterprise and infrastructure" (Ménière et al., 2017, p. 17). In other terms, these new technologies will make our living and working environments more intelligent and capable of decision making to varying degrees (Kumar et al., 2019). Alternatively, Klinger et al. (2018) specifically study deep learning, a subset technology of AI, and conclude that it has the characteristics of a GPT: its importance is rapidly increasing which highlights its pervasiveness and it has a worldwide reach, although knowledge creation in the field is rather geographically concentrated.

Beyond altering manufacturing processes, the pervasive potential of these digital technologies will also affect the service industry through the phenomenon of "servitisation". The term designates the combination of the development of products and services with ensuing unique experiences tailored to each customer (Lee et al., 2014). It originally helped the manufacturing firms to differentiate themselves from the competition by providing integrated after-sale services (Lee et al., 2014). This practice came to be when information technologies (IT)[3] from the third industrial revolution enabled manufacturers to quickly adapt to changing consumer demands (Kelley, 1994). Indeed, these technologies allowed for increased efficiency in the workplace and for increased flexibility of manufacturing processes (Kelley, 1994). The trend of "servitisation" has since gained momentum, and accordingly, companies often seek to collect large volumes of data to help them mass customise their products and services (Kumar et al., 2019; Baldassari & Roux, 2017).

As mentioned before, I4 will largely affect economies and markets (Lee et al., 2014; Oztemel & Gursev, 2020; Baldassari & Roux, 2017; Reischauer, 2018). The intensity of this economic impact has been further studied from various approaches. For instance, a study of 2014 has evaluated that by 2025, the gained efficiencies in operational processes resulting from I4 technological advances will contribute an estimated €78bln to Germany's GDP (Hermann et al., 2015). Another analysis of 2017 claimed that by 2020, breakthroughs in the fields of hardware and software development would help develop a $7trillion market worldwide (Baldassari & Roux, 2017). Aside from opening new lucrative commercialisation opportunities (Dosi & Nelson, 2010), the pervasiveness of I4 will permeate over other industries aside from high-tech sectors. For instance, advances in 3D (bio)printing could lead to radical breakthroughs in medicine and healthcare by allowing mass production of organs.

3.2. Capturing I4 Technologies: Current Efforts

While most can agree on the broad definition of I4, there is still an apparent lack of consensus as to which technologies the term covers and as to how these technologies should be classified (Pereira & Romero, 2017; Chiarello et al., 2018; Ciffolilli & Muscio, 2018). Kumar et al. (2019) identify the following key technologies as central to I4: cyber-physical systems, IoT, AI, ICT, big data, and cloud computing. They argue that certain technologies, such as AI, enable industrial processes to develop human-like cognitive skills, which makes them "intelligent". Similarly, the authors claim that IoT has facilitated the development of intelligent manufacturing by first facilitating digitalisation and enabling cloud computing – which interconnects states of manufacturing processes in digital environments. Ménière et al. (2017) develop the idea of enabling technologies even further. The authors classify I4 technologies (I4T) under three categories: core, enabling, and application domains. In this manner,

[3] ICT can be considered an extension of IT that helps link systems established through IT

they classify core technologies as the ones that "make it possible to transform any object into a smart device connected via the internet" (p .10); enabling technologies build upon the core ones by being combined with the smart objects, and application domains capture the utilisation of the technologies and connected objects.

Other studies have also attempted to capture the full spread of I4T by conducting inductive keyword analyses based on patent data and bibliometric searches. A selection of those has been included in Table 2. Janik & Ryszko (2018) conduct a bibliometric search in order to both refine the definition of the term Industry 4.0, and to assess its importance. Since the first official inclusion of the term "Industry 4.0" in 2011 in a strategic document of the German Government, the authors find that the term has been increasingly included in publications. Their estimations indicate that the number of mentions rose from 6 publications in 2001 to 1186 by the end of 2017. In a similar way, Balland & Boschma (2019) establish two criteria to study patent data – the association with I4 and direct association with the Cooperative Patent Classification – which allows them to identify nine I4T. They find most of these nine categories to be increasingly at the centre of patent publications. This rise of interest in the potential of I4 is not only observable in patent data. For instance, consultancy firms and governments often explicitly include I4 and its related technologies in development strategies (Reischauer, 2018). Despite these advances in the literature on I4, further progress is required to study the industry's technological scope.

Table 2. An overview of selected studies that have captured the depth of I4 and its technologies (in chronological order).

Study	Method	Data Sources	Key Findings Related to Defining I4
Ménière et al. (2017)	Analysis of innovation within I4	Patents from the European Patent Office (EPO) until 2016	Core, enabling, and application domains categories of technologies established (as well as the technologies these terms cover)
Janik & Ryszko (2018)	Bibliometric analysis	Publications from the Web of Science database (1990 – 2018)	Network of the most frequent keywords in I4, identification of key publications and general publication trends
Chiarello et al. (2018)	Analysis of keywords based on semantic entries	Wikipedia free online encyclopaedia, Scopus database	Mapping a dictionary of I4 and technology clusters
Ciffolilli & Muscio, (2018)	Taxonomy analysis of I4 related projects	RED database of Ismeri Europa[4]	Classification of I4T
Balland & Boschma (2019)	Inductive keyword search of pre-selected patents (2002-2016)	Patents from the OECD[5]-REGPAT database, geocoded at the regional level (2002 – 2016)	Identification of nine main I4T and how they conceptually build on each other
Dernis et al. (2019)	Worldwide analysis of inventive activities of top corporate R&D investors	Patents: EPO, JPO, KIPO, CNIPA, USPTO; Trademarks: EUIPO, JPO, USPTO; Publications: Scopus database (2014 – 2016)[6]	Identification of AI and ICT key technologies
Oztemel & Gursev (2020)	Literature review	Eight publication databases[7]	Identification of key I4 technologies and their implementation in economies
van Beuzekom et al. (2020)	Bibliometric analysis, keyword and text mining analyses to identify AI developments	Publications: Scopus database; Open-source software: GitHub; Patents: EPO, JPO, KIPO, CNIPA, USPTO (1996 – 2016)	Improved insights into the scope of AI-related technologies

Note : The papers and reports cited in this table have much broader findings. The aim of this table was simply to compile a brief summary of the key findings directly related to I4T.

[4] The RED database consists of regionalised information of projects financed by the EU (budgets, organisation information, key research areas and technologies) (Ciffolilli & Muscio, 2018)
[5] Organisation for Economic Co-operation and Development (OECD)
[6] Japan Patent Office (JPO), Korean Intellectual Property Office (KIPO), State Intellectual Property Office of the People's Republic of China (CNIPA), United States Patent and Trademark Office (USPTO), European Union Intellectual Property Office (EUIPO)
[7] CiteSeerX, ACM, AISeL, EBSCOhost, Emerald Insight, Taylor Francis, Science Direct and Google Academic

The attention dedicated to I4 is justified as its societal and economic impacts will most likely affect labour markets through restructuration processes and re-alignment of skills (Reischauer, 2018; Ménière et al., 2017; Ciffolilli & Muscio, 2018; Boschma & Balland, 2019). Firms aiming to increase their productivity will most likely automate their production lines by adopting new technologies, which in turn will require different sets of skills from employees (Reischauer, 2018). Due to the lack of consensus on the formal definition of I4, most firms have not yet been able to completely exploit I4 and its technologies (Janik & Ryszko, 2018).

In addition, few publications on the macro and meso scales have directly looked at how firms have engaged with I4 and its technologies. For instance, studies have considered the impact of I4 on global value chains (Strange & Zucchella, 2017), they analysed the potential impacts of the industry on manufacturing processes (Tonelli et al., 2016), or also researched the interactions between service sectors and I4 (Frank et al., 2019), but to my knowledge, none directly scrutinised the involvement of firms with I4. However, this aspect has been partially explored by international organisations such as the OECD (Dernis et al., 2019) who, using different data sources focused on ICT, studied the engagement of firms with AI technologies. Therefore, this research contributes to the overall literature by adding to the studies that refine the scope of I4T by considering the firm-level dimension.

3.3. Capturing I4: The Opportunities from Trademark Data

The link between innovation and technological development calls for the use of innovation measures when estimating the scope of a industrial revolution (Fagerberg, 2005; Smith, 2005). Accordingly, the academic community has explored how different types of IPR data could help capture innovation. The key argument of the discourse relates to the idea that the term "innovation" is actually composed of both "invention" and "innovation" stages (Fagerberg, 2005). However, to this day, the most predominant approach has been to use Research & Development (R&D) and patent data in order to measure innovation in its broad sense (Mendonça, 2014). According to the United States Patent and Trademark Office (USPTO), patents are:

A property right granted […] to an inventor "to exclude others from making, using, offering for sale, or selling the invention throughout the U.S. or importing the invention into the US" for a limited time in exchange for public disclosure of the invention when the patent is granted.

In other words, patents grant their inventor, also titled "assignee", exclusive rights on their invention. Once the patent is granted, it becomes publicly available, which prompts the design of new inventions through knowledge spillovers (Petralia et al., 2016). Patents provide the clear benefit of being easily accessible and measurable partially owing to their worldwide availability at institutions that record their registration (Filippetti et al., 2019). Not only are patents very specific – they provide detailed information on the classification and the patent's owner –, but they have also been collected for rather long periods of time. This, combined with their capacity to estimate technological innovation, has made them the most prevalent innovation indicator (Mendonça et al., 2004; Filippetti et al., 2019).

On the other hand, patents' disadvantages have been increasingly highlighted in recent years. Smith (2005), building on Fagerberg's claim, draws attention to the fact that patents trace inventions rather than innovation, although they are still intrinsically linked to innovation (Fontana et al., 2013; Mendonça et al., 2004). As such, patents are biased in favour of industries whose inventions are more patentable such as high-tech industries. However, they are less able to grasp innovation in, for instance, more service-oriented sectors (Filippetti et al., 2019). Furthermore, not all inventions are patentable, which implies that a discovery, as novel as it may be, which does not become patented would not be considered as an invention. However, the thought-process that led to the discovery has most likely spurred alternative ideas and given rise to other knowledge creation mechanisms. Additionally, it has been shown that a large share of innovative outputs never becomes patented, and that the effectiveness of patents as an innovation indicator varies according to industry types (Fontana et al., 2013). Other drawbacks of the measurement include the tendency to underestimate the contribution of small and medium enterprises, and, specific to the case of patents, the difficulty to capture heterogeneity across industries and sectors (Mendonça, 2014; Mendonça et al., 2004).

In recent years, the general discourse on measuring innovation has slightly shifted focus and has successfully begun advocating for the inclusion of data on trademarks (deGrazia et al., 2020; Gotsch & Hipp, 2012; Filippetti et al., 2019; Castaldi, 2019). This was primarily led by the rise of the service industry, and most specifically of ICTs (Amara et al., 2008), which sparked the need to find alternatives to patents due to their bias for manufacturing innovation (Gallouj & Savona, 2008; Gotsch & Hipp, 2012; Filippetti et al., 2019). In this digital era that is often service oriented, trademarks present themselves as fundamental tools to capture innovation (Castaldi, 2019; Gao & Hitt, 2012; von Graevenitz et al., 2019), and are defined by the USPTO as follows:

[Trademarks] protect words, names, symbols, sounds, or colors that distinguish goods and services from those manufactured or sold by others and to indicate the source of the goods. Trademarks, unlike patents, can be renewed forever as long as they are being used in commerce.

The official definition highlights one of the crucial advantages offered by trademarks over patents: as long as the renewal fees are paid and that the mark is being exploited as it was intended, trademarks have no expiration date (Castaldi, 2019). In addition, the filing dates of trademarks tend to be a more accurate representation of the time of innovation than patents (von Graevenitz et al., 2019). The low fees required for filing trademarks most likely also play a part in their popularity among firms. Moreover, trademarks also offer smaller companies the possibility to protect goodwill, tacit knowledge and codified output (Amara et al., 2008; Graham et al., 2013). On one hand, trademarks offer a direct link to markets by providing both a signalling function and a reflection of product differentiation, which allows them at the same time to grasp to some extent firm strategies (Castaldi, 2019; Gao & Hitt, 2012). On the other hand, trademarks are not capable of fully capturing the extent of product variety of firms (Gao & Hitt, 2012). Alternatively, before 2013, the accessibility of data was rather limited, until the Case Files Dataset was released by the USPTO. It provided detailed data retroactively, which has since enabled empirical studies to start using trademarks (Castaldi, 2019; Graham et al., 2013). Since then, various opportunities arising from working with trademark data have been highlighted. Specifically, in her study, Castaldi (2019) suggests research avenues such as using keyword analyses over traditional classification conventions, or constructing metrics capable of capturing the value and novelty of trademarks.

Conversely, using trademark data to capture innovation also presents drawbacks. The first one being that the industrial data remains available at aggregated classification levels, which limits comparison opportunities with patent classifications (Filippetti et al., 2019). Similar to the case of patents, the tendency to trademark across industries significantly varies and can lead to possible biases (Mendonça, 2014). Additionally, it has been proven that the competitive value of trademarks not only varies over time but is also likely to differ across industries (Gao & Hitt, 2012), which further complicates comparison analyses. The traditional approach to this matter when using patent data is to evaluate patent citations, but that is not possible for trademark data (Gao & Hitt, 2012). Alternatively, there is also no guarantee that a trademark will be linked with a single product or service, nor that it will not be applied to different sectors (Filippetti et al., 2019). Moreover, the geographical information of trademarks is recorded at the location of the headquarters (Graham et al., 2013). This narrows the possibilities of studying how trademarks differ in their exploitation across space: it is possible to do so across countries but not within countries.

Nonetheless, recent research has highlighted the potential of solely exploiting trademark data in innovation studies. von Graevenitz et al. (2019) use trademark data to study the diffusion of innovation across the U.S. The authors find that the diffusion of ideas and innovation is negatively associated with distance, a conclusion that is usually reached using patent citations data. Gao & Hitt (2012) investigate the link between IT and trademark data. They conclude that the relationship is significantly positive, implying that higher IT stock can lead to higher rates of trademarks held by a company. Alternatively, the authors demonstrate that trademarks in IT have higher turnover rates: IT firms revise their products frequently which shortens the lifespan of trademarks and ultimately reduces their levels.

The consensus is that trademark data is most insightful when applied to knowledge-intensive services and small and medium enterprises (Gotsch & Hipp, 2012; Filippetti et al., 2019; Flikkema et al., 2014). Mendonça et al. (2004) argue, based on the Community Innovation Survey, that firms considered to be innovative frequently favoured trademarks over patents. However, patents retain their key advantage of being able to capture technological inventions, hence raising the possibility of using different types of IPR at the same time to get a more complete picture. Llerena & Millot (2013) contribute to this idea by assessing whether patents and trademarks should be considered as complementary or substitutes when it comes to protecting innovation. The authors conclude that the answer is not straightforward and is based upon factors such as advertising spillovers and depreciation rates. In addition, trademarks present interesting complementary characteristics to patents as they tend to cover the 'innovation' aspect of the definition of Fagerberg (2005). Indeed, they are more adequate to reflect the products and services which exploit the technologies developed through patent data (Castaldi, 2019). Moreover, in the words of Graham et al. (2013, p.3), "trademark data may capture innovations that are not patented, either because they are not patentable or because their inventors choose not to seek patent protection". Overall, these findings strongly suggest that these two types of IPR data should be combined to cover a broader range of innovation capabilities (Flikkema et al., 2014).

In this spirit, studies have begun capturing innovation with different types of IPR data. For instance, Dernis et al. (2019) use patents, trademarks and publications centred around ICT to explore trends in innovation in AI throughout the world. Similarly, Filippetti et al. (2019) analyse the innovative profile of European Union (EU) regions. They base their analysis on three types of IPR data: patents that cover technological innovation – or effectively invention, trademarks to capture innovation in knowledge-intensive services, and design rights for low- and medium-tech enterprises. It allows them to draw more comprehensive pictures of the EU and findings include the fact that while eastern and southern European regions maybe be weaker when looking at patent data, they tend to perform rather well in the other two types of IPR data. Seip et al. (2019) also consider different IPR types and develop categories of IPR applicants based on filing intensity and variety. They find that there is a cluster of a few firms specialised in high-tech industries which display high levels of IPR applications across both spectrum of intensity (variety and intensity).

4. Methodology

4.1. Research Questions and Hypotheses

The literature review on industrial revolutions and innovation of the two previous sections paved the way to answer the main research question: "*How are firms developing and adopting I4 technologies across the U.S.?"* It established that because I4 is an industrial revolution, it is important to study the interactions between invention and technological development with innovation and application of technologies. This can be done by building on the literature on capturing innovation and by combining two types of IPR data: patents and trademarks. Moreover, considering the potential social and economic impacts that I4T may have (Petralia, 2019; Baldassari & Roux, 2017), it is also crucial to determine which technologies firms are exploiting both conceptually and spatially. Ultimately, this will also contribute to the literature on innovation diffusion by associating patents with the geography of invention, and trademarks with the geography of market application.

The first sub-question to be addressed is as follows: "*How can patent and trademark filings reveal patterns of development and adoption of I4 technologies?"* In other words, I build on the literature on capturing innovation and use patents to determine the areas of invention in I4, and trademarks to identify the key areas of innovation within I4. Based on the literature, I would expect patents to capture I4 inventions better, and trademarks to capture I4 innovation and more specifically innovation in service-oriented sectors (Filippetti et al., 2019).

Through geographical and cognitive proximities, the development and adoption of I4T will play a key role in determining whether labour markets are positively or negatively affected by the industrial revolution. As previously mentioned, firms may embrace or fail to seize the economic opportunities of I4T and in turn, local and national labour markets may suffer. To understand how firms are exploiting I4T and how the unfolding of I4 is taking place across space, the second sub-question addresses the following: "*To what extent are I4T being developed and applied spatially across the U.S.?"* Given the importance of mass individual data collection in I4 (Oztemel & Gursev, 2020; Kumar et al., 2019), I except the firms analysis – which focuses on cognitive proximity – to confirm the importance of major tech firms such as Google or Amazon who are able to both gather the data and use it to tailor their products and services. As to the geographical analysis, I anticipate finding that the exploitation of technologies in I4 will be rather concentrated in a few key places in the U.S. Moreover, it is likely that invention will prove to be more spatially concentrated than innovation for two main reasons: (1) studies on GPT diffusion have shown technological inventions to be more concentrated than their application (Petralia et al., 2017; von Graevenitz et al., 2019), and (2) the main technologies of I4 can be considered to be "complex" and therefore hardly replicable and stickier (Balland & Rigby, 2017; Balland et al., 2020).

4.2. Data Preparation

In order to capture the extent to which firms invent and innovate across U.S. states and cities, I collected data from the USPTO website. I used the Trademark Case Files Dataset for trademarks, and bulk data from PatentsView for (utility) patents. I chose to focus on the time period 2008-2017 to capture I4 IPRs even before the label "Industry 4.0" began being used. It is important to consider years prior to 2011 as the term "Industry 4.0" builds upon "concepts and perspectives that evolved over the years" (Pereira & Romero, 2017, p. 1208). By excluding the years 2008 to 2010, I would have failed to capture these early developments that led to the development of the term.

Afterwards, I filtered both datasets so that only registered marks and patents remained. Therefore, IPRs that have expired, whose application has never been completed, or that have been cancelled were not included in the analysis. The registration process acts in a way as a quality check, and this procedure also ensures that the dates used will match the actual invention date as closely as possible. The average total pendency, or the average time elapsed between IPR filing and registration, can be especially high for patent data: it reached almost two years in 2019 (USPTO, 2020). Based on this estimation, I did not include later years (2018 and 2019) in the analysis to take into account data truncation. Indeed, a large amount of IPRs from 2018 and 2019 would have been excluded from the analysis as their application has not yet been finalised.

I collected the datasets for U.S. filers only, based on the country of the inventor for patents and the country of the owner for trademarks. If there were multiple inventors based in different countries, the data was excluded because of limited downloading capacities[8]. This is a limitation to the quality of the data as it excludes IPRs resulting from international collaboration projects. I collected a small sample of patent data from the UPSTO and calculated how many patents I have approximately excluded by only keeping U.S.-based inventors. According to my own estimations, I exclude about 3.5% of patents. Furthermore, for patent data, the assignee information was gathered based on the 'assignee' and 'patent assignee' datasets made available on the USPTO website. Regarding the trademark data, I used the publicly available 'owner' data file to obtain information on the owners that have filed and/or registered marks.

I gathered geographic information for both datasets at the state and city levels. Given computing capacity constraints, I only kept city names at the urban level. For trademarks, the owner information for non-individuals (firms, corporations or agencies) is located based on the position of the headquarters (Graham et al., 2013). Unfortunately, this limits the potential to study the geography of the application of technologies by examining how trademarks are being used across space. To improve the spatial comparability of both IPR types, I chose to use the location of the patent assignee rather than that of the patent inventor. This prevents me from accurately determining the geography of invention in I4, but in this manner, spatial analyses for patent and trademarks datasets are more comparable and both focus on the assignee level.

[8] This thesis was written from home between March and June 2020, in the middle of the corona crisis. I was not able to conduct certain analyses because of the limited computing power of my personal laptop.

Despite the absence of an exhaustive classification of I4T (Oztemel & Gursev, 2020), many attempts have been made to define the scope of I4 (as compiled in Table 2). Therefore, I followed a similar method to Balland & Boschma (2019) and used keywords to select data related to Industry 4.0. More specifically, I mined the data based on pre-selected keywords, using R software packages[9]. Those keywords were chosen based on the studies listed in Table 2.

In the case of trademarks, each mark contains a statement text that describes the good/service covered by the trademark and helps identifying it (Graham et al., 2013). As to patents, that information is contained within the patent abstract which also aims to summarise its contents. Filtering those statements and abstracts based on a certain number of keywords offers benefits in two different ways: (1) it allows to compare findings between patents and trademarks and overcome the differences in aggregation and classifications between the two (Filippetti et al., 2019) and (2) it contributes to the existing literature that has already largely implemented such keywords filter on patent data (Joung & Kim, 2017) but has not apparently done it as often using trademark data (Castaldi, 2019).

To enhance the depth of the analysis, the keywords are divided into three categories (see Table 3) based on Ménière et al. (2017): core technologies, enabling technologies, and application domains. Overall, these keywords should be able to capture the innovation principle (Perez, 2010) of I4: the digitalisation of the work and private lives. Consequently, one would expect core technologies to be more predominant in patent data by capturing invention, whereas trademarks and application domains would be more closely linked as they capture innovation (Filippetti et al., 2019). Enabling technologies are somehow "stuck in the middle", because they still combine fundamental elements of I4T, but could also be rather easily developed into final products or services. I expect those technologies to be slightly more associated with trademarks than with patents for that reason. The "General" category consists of generic terms of I4 and of broad-reaching terms such as IoT, cyber-physical systems and Big Data. These terms are the ones that best capture the pervasive aspect of I4, and therefore placing them in one of the other three types of technologies would have overshadowed their reach.

[9] A list of all the R packages used throughout this research can be found under Appendix 1

Table 3. Main keywords used to filter patent and trademark data.

Technology Category	Main Classes	Examples of Keywords
Core Technologies	Hardware Software Connectivity	Graphical processing units, quantum computers, RAM, adaptive display, ICT, random access memory, (micro) processors and controllers, cloud computing …
Enabling Technologies	Analytics User interfaces Three-dimensional support systems Artificial intelligence Position determination Power supply Security	Additive manufacturing, 3D printing, augmented and virtual reality, system integration, cybersecurity, deep learning, blockchain, programmable logic controller, GPS systems, Machine-to-Machine (M2M) communication …
Application Domains	Home Personal Enterprise	Drones, Smart industries & offices, autonomous robots and vehicles, robotics, advanced manufacturing solutions, intelligent lighting & heating systems, …
General	Internet of Things, Big Data, Industry 4.0, Cyber-Physical Systems	

Note: The actual keyword filter based on the technologies in column 3 was more specific and was written so that it considers possible variations in spelling or abbreviation.

Using a keyword filter-based method is not without flaws. Firstly, the selection of the keywords is subjective and static (Chiarello et al., 2018), while technological revolutions and GPTs are dynamic (Perez, 2010; Helpman & Trajtenberg, 1996). Secondly, in the case of trademarks, the statements used as a basis for the filter can also be problematic, granting that filers are not required to describe the underlying technologies of products and services (Graham et al., 2013). Because of this, many potential applications of the IPR are automatically discarded from the filter. However, it is worth mentioning that business opportunities would lead most filers to elaborate as much as possible on the trademarks they are applying for, as it represents a direct link to their customers. Especially in the case of business-to-business firms, filers could be incentivised to display their product/service in more precise ways to increase their chances of engaging with transactions with other firms. On the other hand, for marks dedicated towards exchanges between businesses and individual consumers, firms might be less inclined to fully disclose the technologies behind their product/service. This is especially true considering that customers will most likely not read the underlying statements: they will only read the name of the trademark. However, I only applied the keyword filters to the statements – and patent abstracts – and did not include the title of the IPR.

Once the data is filtered for these keywords, I conduct a manual check in order to avoid including irrelevant data. In order to match the information on inventors and owners with the selected data, both trademark and patent datasets are merged with owner and assignee information based on the serial numbers. To facilitate later analyses, both final patent and trademark datasets are further cleaned by removing punctuation signs[10] from the abstract and statement entries.

In order to create maps, I geocoded the data on a state basis, after having excluded data on Alaska and Hawaii. The other analyses and visualisations that did not require mapping still include the two states. I was not able to generate similar maps at the city levels because geocoding the data at that spatial level was too much to process for my personal laptop. For both datasets, the assignee and owner organisations as well as city names contained multiple entries that were equivalent to the

[10] The punctuation signs removed were : ( ) , ; - * ” ’ [ ]

same name. For instance, “GOOGLE INC”, “GOOGLE, INC” and “new york” and “NEW YORK” were all recoded into one unique name. This cleaning was not done systematically, but many checks were done throughout the process to ensure that companies would not have been affected by an oversight.

Additionally, some cities were recorded under erroneous states (New York into Delaware for instance). Large cities which only exist in one unique state were easily recoded, but cities that could be in one or multiple states were not recoded in order to avoid wrong guesses. For instance, if New York was incorrectly based in Delaware, I recoded it to the state of New York. However, cities such as Auburn could be from either of four states (New York, Washington, Alabama, Maine), it was not possible to determine the correct state without the complete address – which was excluded because of computing constraints. For text analyses based on word frequency, the code was designed in order to automatically select the most frequent words, while prepositions, pronouns, adverbs as well as verbs and nouns that did not carry a meaningful value were deleted. The complete list of those deleted words can be found in Appendix 2.

5. Results

5.1. Focus of IPR Filings

Answering the overarching research question involves initially establishing whether I4 patents can reflect technology development in I4 and whether trademarks can reveal patterns of technology adoption in I4. In other words, it is necessary to assess whether patents are a more adequate indicator of invention in I4 and conversely whether trademarks are more insightful when capturing innovation in I4. Using text mining tools, I conduct a word cloud analysis on both datasets across all technology categories. In this manner, this allows me to evaluate the hypothesis about using the two IPR types in combination and in turn to unpack the areas of technology development and application for I4.

For patents (Figure 1, left side), the terms appear rather technical and hardware related. More specifically, a majority of the most mentioned words appear – either directly or indirectly – in the core and enabling technology categories. Regarding trademarks (Figure 1, right side), the words are, in a way, closer to what one might associate with services, although some core-related words are still noticeable (“hardware”, “device”, “systems”). Words like “game”, “downloadable”, “video” could be an illustration of technologies that are directly linked to the usage consumers might make of them, whereas in the case of patents, it seems harder to directly relate those words to market applications. Therefore, patents seem more suited to study invention and technological development in I4, and trademarks appear more suited for studying the exploitation of these technologies in their applications, as was to be expected based on the literature (Filippetti et al., 2019; Flikkema et al., 2014; Mendonça et al., 2004).

Figure 1. Word cloud of the patent and trademark data (2008-2017).

Patents | Trademarks

Note: The size of the words is proportional to the number of times the words were mentioned in the patent abstracts or trademark statements.

For a more thorough analysis, I constructed specific word clouds for both patent and trademark data, for each of the technology categories defined in Table 3. The results in Figures 2 and 3 show that while most words can be traced across both general and specific clouds, some only appear in distinct technology categories. For example, application domains of both patent and trademark data are not as predominant in the main word clouds, which insinuates that most I4 technologies are not mainly focused on exploiting the use of connected objects. As before, application domains for patents seem more concentrated around physical objects and components ("system", "device", "vehicle", …), whereas for trademarks, they appear more directed at commercial opportunities, as illustrated by words including "development", "design", "marketing", "business", or "services". Another interesting feature in the trademark word cloud of core technologies (Figure 3) is the presence of the term "medical". This could point towards the idea of pervasiveness for I4 by showing potential beyond ICT-related fields.

As mentioned above, the core and enabling categories appear to be central to both patents and trademarks. Once again, patents focused more on 'technical' aspects of I4, whereas trademarks were targeted around application potentials of the technologies. This links back to Perez's theory of technological revolutions (2010): the interconnectedness between virtual and real world paradigms of I4 has allowed I4 technologies to bring about both technological change through patent and inventions, while also enabling the development of application of the technologies through the innovation aspect of trademarks. With regards to the "General" category, few differences are observable for trademarks, although for patents, in accordance with the keywords table (Table 3), terms such as "Internet of Things", and "Cyber-Physical Systems" as well as "Data" (also present is "big", probably combining into the term "Big Data") are present. This might exhibit a more explicit attention borne by IPR applicants to the central themes and technologies to I4. A majority of the words highlighted in Figures 2 and 3 are also present in Ménière et al.'s report from 2017, in the respective I4 technological categories, which in a way attests to the quality of the keyword filter.

It is worth noting that some words across all clouds may appear rather generic when considered on their own. The list below includes – for a small sample of words – a few examples of how those words appeared in context in patent abstracts and trademark statements. To improve the quality of this analysis, some words should be left together based on context (eg: global positioning *system*) while others should be kept out (eg: business *consulting* in the fields of…) as they bring little insight into this analysis. This method could also possibly allow technologies such as AI to appear on the clouds of the enabling category. Given how time-consuming the task of analysing each word in the context of an expression would have been, I did not make such distinctions in this research.

- System: global positioning system, computer operating system
- Device: mobile device, IoT enabled device, wireless device
- Method(s): teaching methods for AI, methods to cloud
- Application: mobile application, application integration, software application
- Management : computer database management, management of computer software
- Design: software design, cloud-based computer aided design
- Service: software as a service, platform as a service, educational service
- Consulting: computer software consulting, business consulting in the fields of …
- Output: output signal, colour output
- Information: processing information, bits of information
- Communication: communication device, communication network
- Time : in real-time, for a certain time-period

Figure 2. Word cloud of the technology categories of patent data (2008-2017).

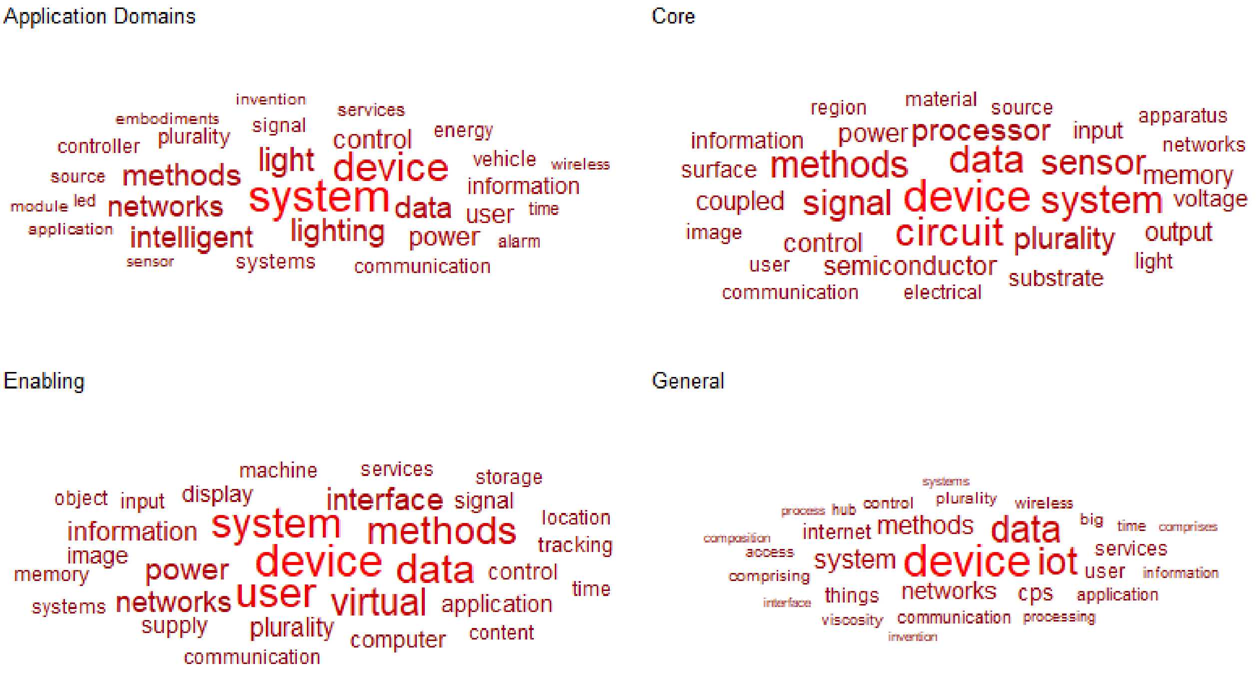
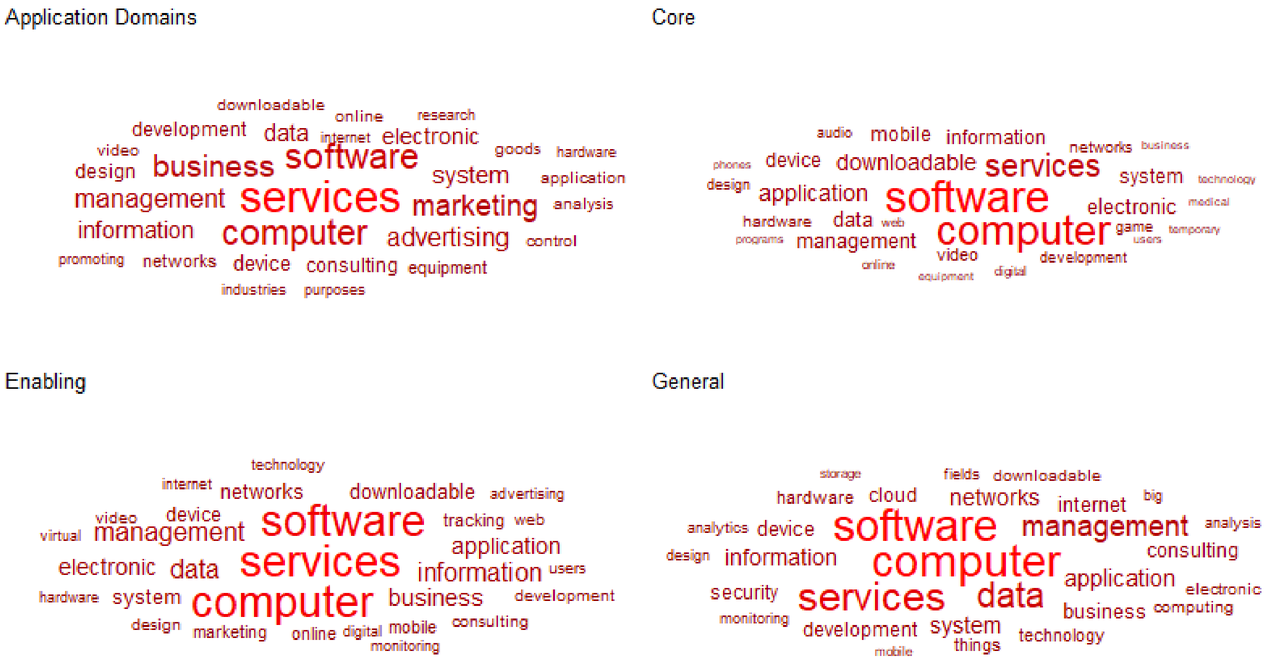


Figure 3. Word cloud of the technology categories of trademark data (2008-2017).



5.2. Technology Categories of I4

The second step in understanding patterns of development and adoption of I4 technologies requires a better understanding of the data. More specifically, it calls for the need to expand on how I4 filings are divided not only across patents and trademarks, but also across time and technology categories (as defined in Table 3).

While the label "Industry 4.0" only appeared in 2011 in Germany (Pereira & Romero, 2017), some of the technologies were already being developed prior to that date (Balland & Boschma, 2019). This is indeed shown in Table 4 (and Figure 4) as before 2011, I4T were already being patented and trademarked. While there is no large increase of registered patents in I4 after 2011, there is one for trademarks whose numbers are then stable for a few years. Whether this difference between 2011 and 2012 can be directly attributed to the first official label of the term I4 remains unsure and should be studied more thoroughly.

However, the general trends for patents are rather puzzling. Overall, we observe an increase from 2008 to 2013, except for a dip in 2009 – probably linked to the financial crisis (WIPO, 2019), which seems to confirm the increase in interest in the field. From 2014 onwards, the numbers start decreasing. While the numbers for 2016 and 2017 could be partially explained by the truncation of the data and by lower application levels towards the end of 2017 (WIPO, 2019), the results of 2014 and 2015 fall outside the general observations in I4. Those tendencies are reflected in the subcategories of application, core, and enabling technologies. Only the category capturing general terms associated with the industry displays increasing numbers from 2012 onwards. Overall, for the years 2008-2015[11] patenting activity in Industry 4.0 increased ranging from 12% to 19%, with the "General" category seeing a growth of almost 400%, thus illustrating a more explicit interest in the central themes of I4.

In the case of trademarks, the findings are much less ambiguous, Table 4 depicts large increases in trademark filing activity across all the technology categories of I4. This could reflect a shift in commercialisation strategies towards I4-related opportunities (Castaldi, 2019, Gao & Hitt, 2012). This is in line with the observations of Oztemel & Gursev (2020) who claimed that the term of Industry 4.0 has now been accepted and is increasingly being embraced by firms. Additionally, increases in IPR levels throughout the country can be expected given the increased national interest in I4. For instance, in 2011 the Obama administration launched the Advanced Manufacturing Partnership that was dedicated towards enabling innovation and facilitating developments in I4T throughout the country (Liao et al., 2018).

Overall, the numbers across categories corroborate one conclusion of the word clouds: I4 patents are more associated with core technologies. More specifically, Table 5 confirms that for both IPR types, core and enabling technologies are the key focus of innovation in I4. Nonetheless, I4 technologies aggregate into growing proportions out of all registered IPR in the U.S. over the time period studied which reflects the idea of pervasiveness of industrial revolutions: I4 technologies could be permeating across a broader range of industries. However, it is worth noting that patenting and trademarking activities have been on the rise over the time period studied across most sectors and activities (WIPO, 2019). Therefore, future research could aim to disentangle the effects between overall increases in protecting intellectual property and increased interest in I4.

[11] This time period limits the impact of data truncation

Figure 4. Total number of patents and trademarks selected from the keyword analysis.

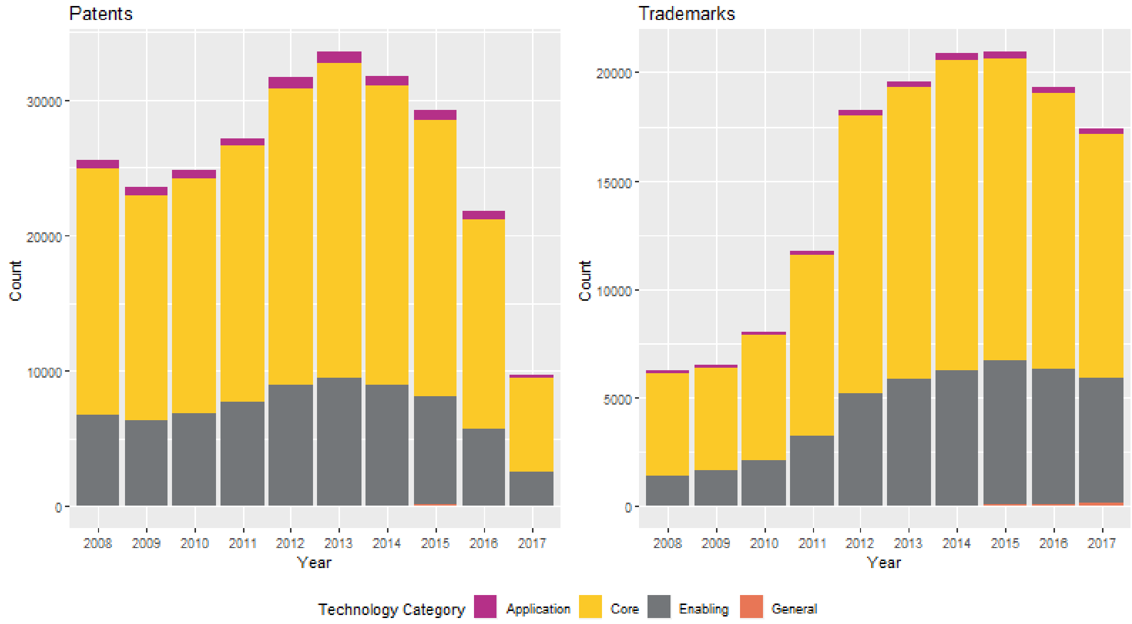


Table 4. Total number of patents and trademarks selected from the keyword analysis.

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Total
Patents	**25,530**	**23,530**	**24,844**	**27,180**	**31,661**	**33,514**	**31,730**	**29,246**	**21,793**	**9,737**	**258,765**
Application	595	532	589	591	801	786	736	689	579	278	**6,176**
Core	18,176	16,679	17,421	18,907	21,891	23,230	22,019	20,435	15,543	6,884	**181,185**
Enabling	6,730	6,304	6,802	7,645	8,936	9,448	8,899	7,988	5,587	2,538	**70,877**
General	29	15	32	37	33	50	76	134	84	37	**527**
Trademarks	**6,246**	**6,521**	**8,046**	**11,777**	**18,265**	**19,562**	**20,872**	**20,923**	**19,278**	**17,423**	**148,913**
Application	93	101	123	186	250	279	326	331	252	288	**2,228**
Core	4,736	4,768	5,803	8,372	12,789	13,423	14,284	13,885	12,669	11,206	**101,935**
Enabling	1,417	1,652	2,120	3,219	5,211	5,827	6,199	6,605	6,228	5,794	**44,272**
General					15	33	63	102	129	135	**477**
Total	**31,776**	**30,051**	**32,890**	**38,957**	**49,926**	**53,076**	**52,602**	**50,169**	**41,071**	**27,160**	**407,678**

Table 5. Shares of I4 technology categories out of the total number of registered IPR (with U.S. inventors/owners).

In %	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Patents										
Application	0.56	0.53	0.55	0.51	0.62	0.59	0.58	0.63	0.74	0.82
Core	17.22	16.67	16.28	16.33	16.88	17.30	17.38	18.57	19.99	20.34
Enabling	6.38	6.30	6.36	6.60	6.89	7.03	7.02	7.26	7.19	7.50
General	0.03	0.01	0.03	0.03	0.03	0.04	0.6	0.12	0.11	0.11
Trademarks										
Application	0.16	0.17	0.18	0.21	0.19	0.20	0.22	0.22	0.18	0.19
Core	8.06	8.05	8.68	9.39	9.79	9.64	9.86	9.28	8.86	7.30
Enabling	2.41	2.79	3.17	3.61	3.99	4.18	4.28	4.41	4.36	3.77
General	-	-	-	-	0.01	0.02	0.04	0.07	0.09	0.09

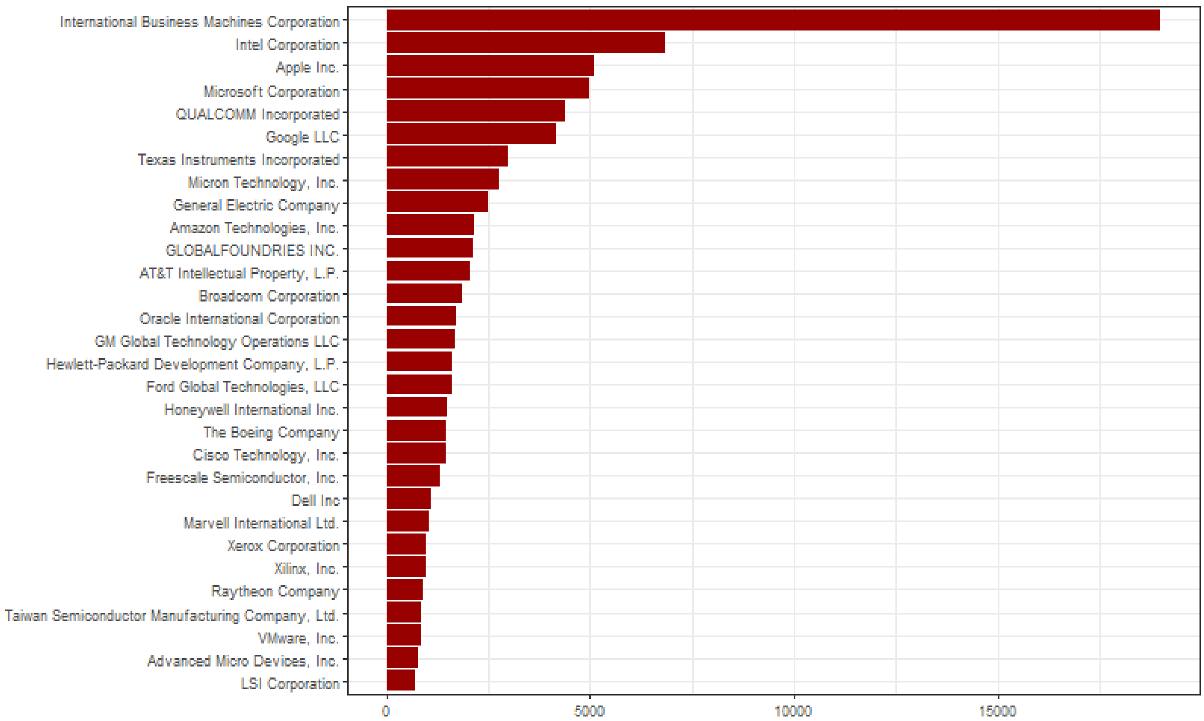
5.3. Firms and Innovation Activity in I4

Proceeding to the second sub-research question, I start by considering the firm level. Building on the literature and findings of the word clouds, I determine which organisations are behind the registered IPR. After that, I examine the extent to which invention and innovation in I4 are concentrated at the firm level. It is worth noting that some corporations use specific holdings to carry out their innovation activities. For instance, Oath Inc. appears in the figures but is actually a subsidiary of Verizon. In this research I limit myself to identifying those key firms in I4, rather than studying how those corporations conduct their research practices.

5.3.1. IPR Filings at the Firm Level

Firstly, by looking at the x axes of both Figures 5 and 6, and comparing it with the absolute numbers of Table 4, we see that patent applications appear to be more concentrated in a few firms compared to trademark filings. This is in line with stylised facts, whereby patenting is more complex and requires more conditions to be met (Castaldi, 2019). The clear leader in patenting activity in I4 over the time period is the company International Business Machines (IBM). This observation seems plausible given that it is a company that focuses on producing hardware and software for computers. It is therefore more likely to capture different technology categories (core and enabling mainly) based on the keywords selected.

Figure 5. Main organisations (based on assignee data) behind patent applications (2008-2017).



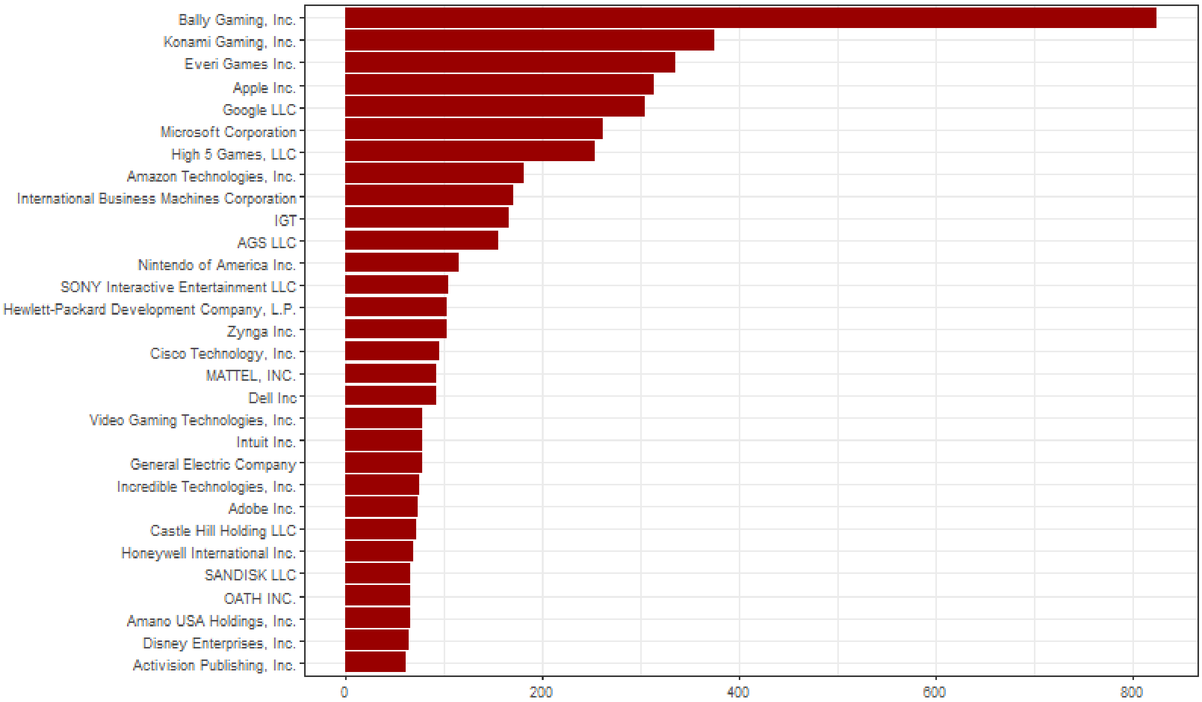
Regarding trademarks (Figure 6), a completely different pattern emerges, in line with the word cloud analysis of the previous section. Bally Gaming, a company that specialises in slot machines manufacturing, is ahead of the ranking, but by smaller margins than IBM had in the patents ranking. To give an example of which technologies a slot machine manufacturer like Bally Gaming exploits to apply I4 technologies, I randomly selected one of their trademarks after it had been selected by the keyword filter:

Video game software, electronic gaming and non-gaming systems which are auxiliary to live casino gaming tables and bingo, namely, computer hardware, LED and LCD display units; computer software used to randomly choose values of jackpots and randomly choose cards, numbers and other symbols associated with live casino table games play, as well as programmable options for playing along with the live casino table games and bingo.

Based on this trademark description, it seems that I4 technologies such as specialised hardware and software are being used to enhance the experience of gambling and gaming services of companies like Bally Gaming. However, this trademark statement illustrates that I4 technologies, as selected in Table 3 may include technologies and concepts from the previous industrial revolution. This is rather a rather logic consequence of the keyword filter considering that I4 partially builds on concepts and techniques from the end of the third industrial revolution (Pereira & Romero, 2017). However, in the context of I4, new technologies such as three-dimensional support systems enable improvements in fields such as gambling and gaming, and consequently improve old technologies (Ménière et al., 2017).

The "Big Four" tech companies, Apple, Google, Microsoft, and Amazon all appear on the trademark ranking, as was expected. Indeed, firms often rely on collecting large amounts of data to better ensure heavy customisation of services (Kumar et al., 2019), and companies such as these four are well placed to compile that data. Moreover, a few companies[12] are present on both rankings which implies that they are involved in both technology development and technology application in I4. Aside from these ten firms, the graphs differ significantly in terms of companies.

Figure 6. Main organisations behind trademark filings (2008-2017).



Note: Google Inc and Google LLC data has been aggregated into Google LLC as it is now the official denomination under Alphabet holdings.

Out of the fifty unique firms present in Figures 5 and 6, the large majority is U.S.-based or has filed IPR through a U.S. holding, which is to be expected considering how the data was selected in the first place. Regarding the parent companies, one is based in Taiwan (the eponymous Taiwan Semiconductor Manufacturing Company), four are Japanese (Nintendo, Amano, Sony, Konami), one is British (IGT) and the rest are from the U.S.

Figures 7 and 8 confirm the higher concentration of patents compared to trademarks; a majority of trademarks in I4 is filed by small firms. For patents, IBM is the only company that owns more than 3% of all I4 patents filed between 2008 and 2017: its share of all I4 patents reaches 7.3%. Intel also owns comparatively more I4 patents than the other leading companies, but afterwards, the spread between firms starts decreasing. However, in the case of trademarks, all companies own less

[12] IBM, Apple Inc, Google, General Electric Company, Amazon, Microsoft, Hewlett-Packard, Honeywell, Cisco Technology, Dell

than 1% of I4 trademarks, and the differences in percentage shares are rather small. This is again logical given stylised facts: it is comparatively easier to file a trademark than a patent (Castaldi, 2019), so firms are less dependent on their sizes to file trademarks and there is little concentration in filing activity. The thirty firms present in trademark filing rankings cumulatively own about 3.30% of all I4 trademarks filed between 2008 and 2017.

Figure 7. Share of patents owned by main organisations behind I4 patent applications (2008-2017).

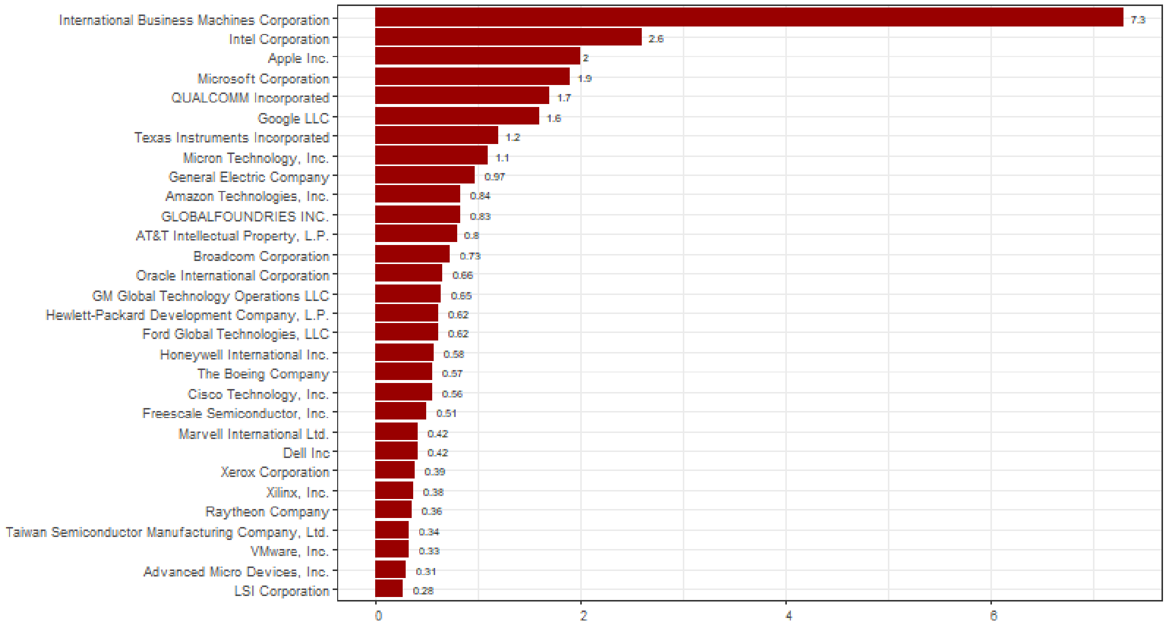
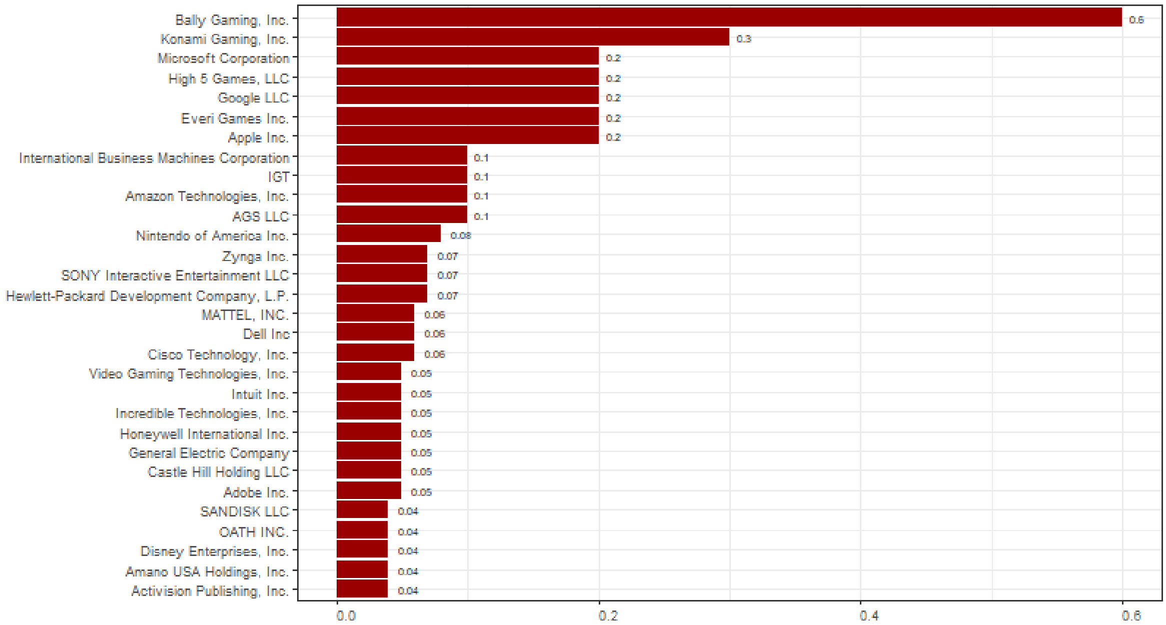


Figure 8. Share of trademarks owned by main organisations behind I4 trademark filings (2008-2017).



5.3.2. Industry Activity

Zooming out of the firm level to the industry/sectoral level can help understand where patenting and trademarking activities are the highest. So, for each firm publicly listed company in both Figures 5 and 6 (and Figures 7 and 8), I compiled industry information from websites such as Bloomberg and Yahoo Finance[13]. For private firms, I consulted the official website and identified the one or two key industry segments they belonged to. In both cases (public and private firms), the results are not exhaustive but were rather aimed at completing the picture on I4 applications provided by the results so far.

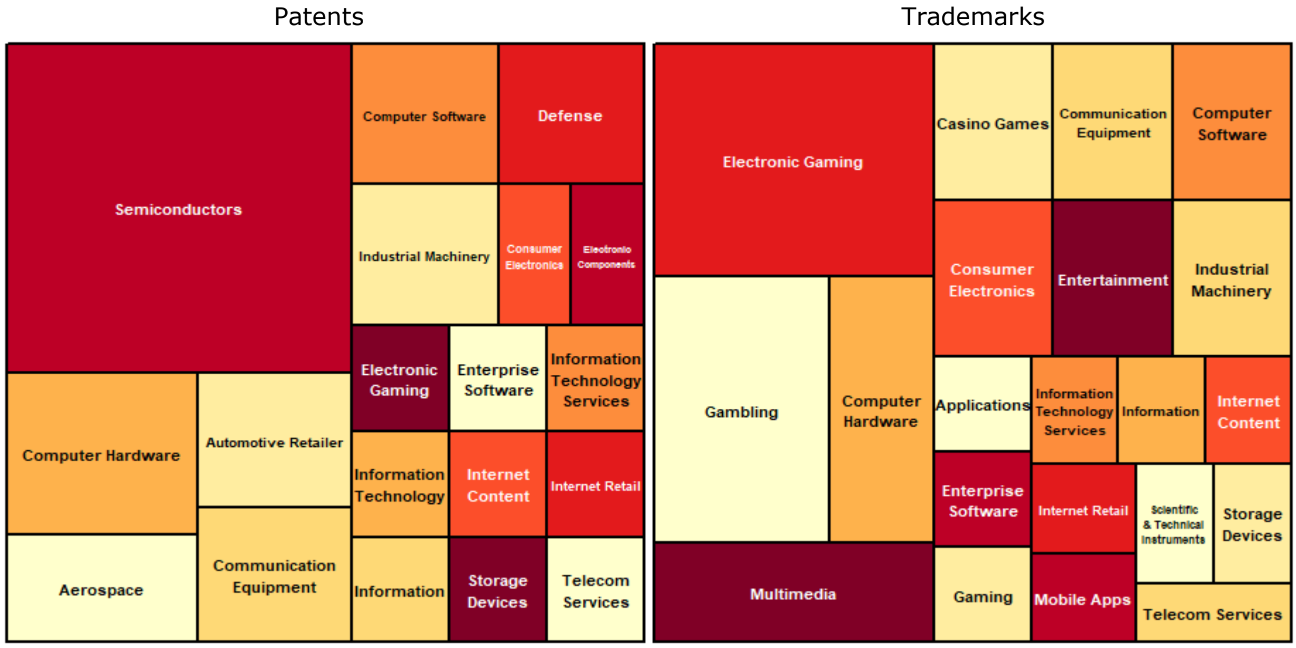
Figure 9 shows the primary industries within which the leading firms in I4 are involved. For patents, semiconductors are the main activity, followed by computer software and hardware, automotive retail, and communication equipment. This outcome echoes the word clouds results for patents. Likewise, computer hardware is also an important industry for trademark filings in I4. Also corroborating the results of the word clouds (cf. Figure 3), gaming, gambling and consumer electronics are largely involved in I4 innovative activity, although their share of total trademarks remains very small. Moreover, such results also substantiate the claim that trademarks seem more connected to services, whereas patenting firms focus around core technologies. Additionally, it is interesting to note that firms displaying high levels of trademark filings also tend to have more differentiated industry focus (from gaming to telecom services, via mobile apps and industrial machinery). On the contrary, the majority of patenting firms evolve in a rather specific range of industries (semiconductors, hardware, machinery) and tend to be less diversified.

The notable presence of the defence industry can be associated with Boeing who benefits from contracts with the U.S. government. Aside from this, there are a few possible explanations behind the prominence of gaming and gambling industries in the U.S. Firstly, the U.S. has placed itself as one of the world leaders in the video game industry, alongside Japan and Europe (Tschang & Vang, 2008). It was able to do so by building its activities around previously established industries such as arcade games and personal computers (Izushi & Aoyama, 2006). The four largest U.S. clusters are located in Los Angeles, New York, San Francisco and Seattle, based on high co-location of video game development studios (Tschang & Vang, 2008). Another plausible reason behind the importance of gaming and gambling industries is that state legislations have over time authorised gambling and have enabled the industry to significantly grow across the country (Morse et al., 2007). Because of the difficulty to regulate it, the advent of the Internet and of modern technologies considerably helped the market for online gambling to develop. For instance, it is estimated that between 2002 and 2004, the market for online poker grew from $100million to $1billion (Morse et al., 2007). In this example, the capacity of certain I4 technologies to be "engines of growth" (Petralia, 2019) is clearly highlighted. The high profitability of technologies has radically impacted the growth path of online gambling (Klinger et al., 2018; Bekar et al., 2018). Moreover, the involvement of Native Indian Tribes in the US with gambling (and casino games) could also help explain the high importance of the industry within I4 (Schaap, 2010).

In addition, gaming and gambling are business to consumer activities which often rely significantly of branding and product differentiation. In the case of electronic gaming, each new game gets a new title but often builds on old software and hardware technologies which enables video game publishers to exploit their IPRs (Tschang & Vang, 2008). For this reason, the presence of gaming and gambling might be less related to innovation activities within I4, and more about product differentiation.

[13] The industry information on these websites is easily accessible and tends to be well considered by investors

Figure 9. Industry/sector information of the leading firms.



5.4. Geographical Diffusion of I4T

This section aims to bring in the geographical aspect to answer the second sub-research question. I first consider how I4 filings are spread across U.S. states, and then I conduct a similar analysis at the city level. In this manner, I can assess whether invention in I4 is indeed more concentrated than innovation in I4, as hypothesised.

5.4.1. At the State Level

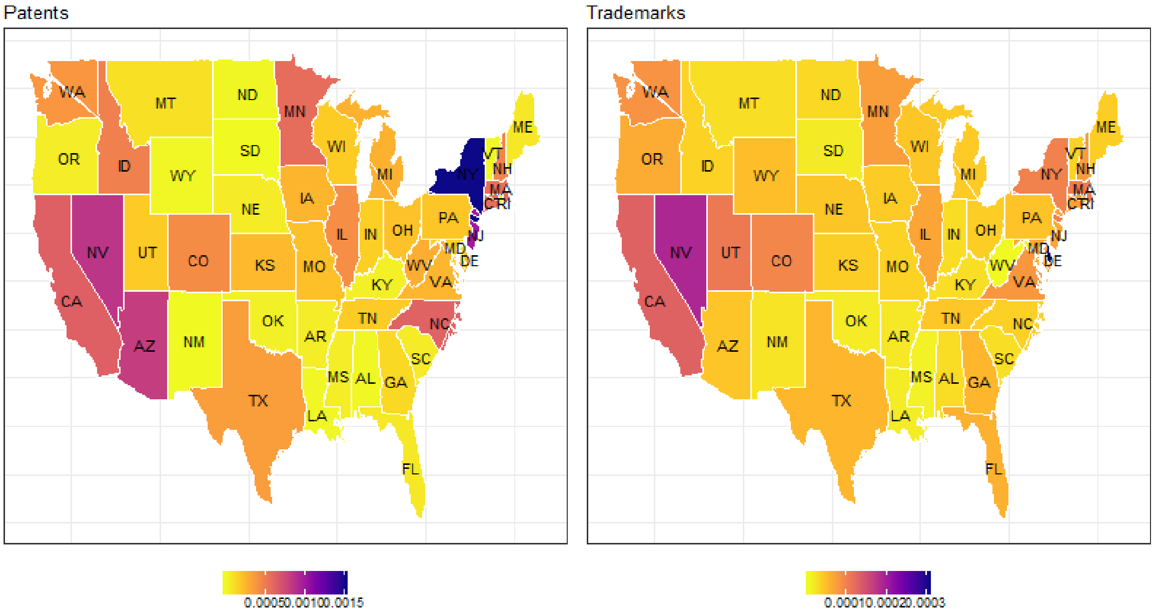
The maps in Figure 10 show the spatial concentration of registered IRP data per capita in 2015 across all technological categories of I4. I focus on the year 2015 instead of 2017 – the end of the data sample. In this manner, I am able to work with IPRs that run less risk of being affected by truncation – or being excluded because they have not yet been registered. Delaware and Nevada are the states with the highest concentration in trademarks, closely followed by California, Utah and New York. For patents, a few of the leading states are located in the Sun Belt states (California, Nevada, Arizona, North Carolina). The others include the states of New York, New Jersey, and Minnesota. The leading position of New York may be explained by its lower population levels compared to states such as California or Texas.

An interesting feature is that the differences in concentration are greater for patents than for trademarks. This is a similar observation to the one made for total firm IPR filings in I4 but in this case the conclusion is slightly different: it appears that innovation in I4 is more spatially spread compared to invention. Furthermore, some states seem to be scoring low on both patents and trademarks. This is the case for Ohio, Louisiana, Mississippi and both North and South Dakota. It is important to keep in mind that because the geographical data is based on assignee and owner data, states with higher concentrations will tend to also be ones that are attractive to firms in terms of taxes or general legislation.

At this spatial level, maps showing the evolution of total registered IPR over the time period did not show any consistent patterns. This is logical considering the time period I am studying, which only covers the advent of I4. If any patterns were to be observable, they would have most likely happened when the Fourth Industrial Revolution established itself as the new technological paradigm and took over from the previous Industrial Revolution. For instance, if my dataset had covered data from the 1960s until 2017, then I would have most likely been able to observe geographical patterns such as the rise of the Sun Belt states over time (Balland & Boschma, 2019). In addition, the shift between 'old' and 'new' innovation principles would have made apparent the readiness of states to adopt those new principles based on their existing capabilities, as advocated by the literature on relatedness (Hidalgo et al., 2018; Balland & Boschma, 2019; Boschma, 2005). However, 2008 being the first year studied in this research, I could not study whether or how states embraced the new technological paradigm.

Another feature visible on the maps is that some of the states that show higher concentrations of IPR applications are also states that have either eased gambling laws or that are still home to tribal lands. For instance, Nevada – the state displaying higher concentration levels in both maps – was the only state that allowed gambling between the 1930s and mid-1970s, which gave it time to establish itself as a lucrative centre for gambling in the U.S. (Morse et al., 2007). Likewise, there are a few tribal lands located in Nevada which have also been encouraged to invest in the industry so that they could become more financially independent (Morse et al., 2007).

Figure 10. Registered IPR per capita across all technology types within I4 (2015).



Note: The population information is collected from the U.S. Census Bureau and contains total population per state in 2015. The North-Eastern States abbreviations are as follows (from North to South): Maine, Vermont (left of New Hampshire), New Hampshire, New York, Massachusetts, Connecticut (left of Rhode Island), Rhode Island, Pennsylvania, New Jersey, Maryland, Delaware.

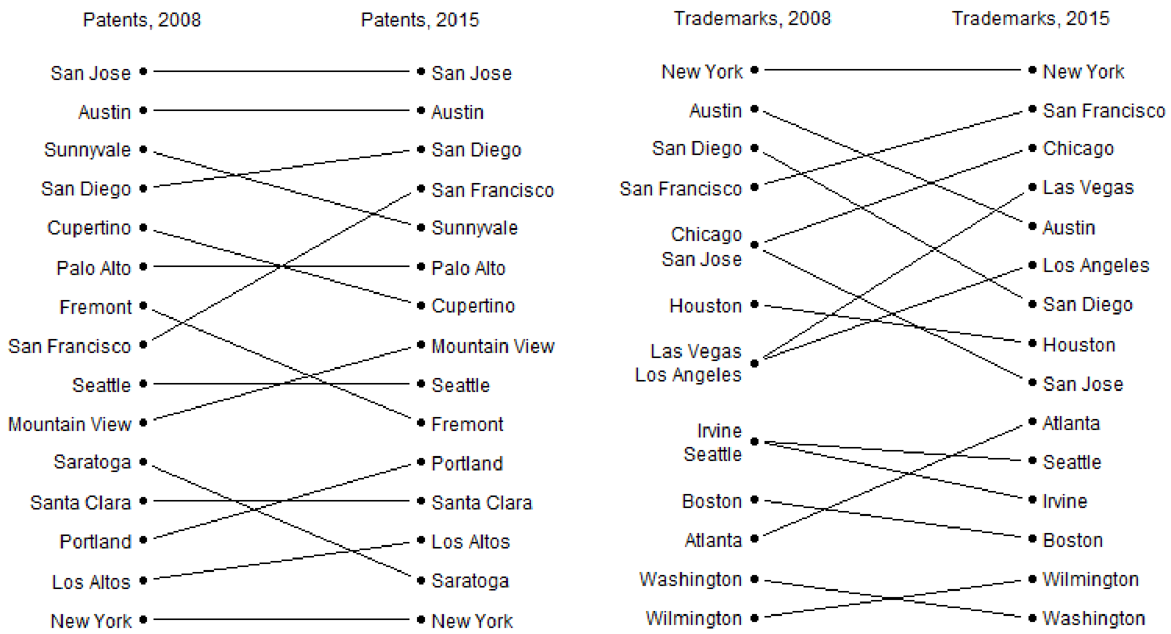
5.4.2. At the City Level

By zooming in at the city level and analysing the trends over time, it is possible to discern clear winner and losers. As shown in Figure 11, San Jose, which is part of Silicon Valley was and remains a leader in patenting levels. However, Sunnyvale, Cupertino, and Saratoga, which are also located in what is defined as Silicon Valley have slightly declined in their patenting levels. Nonetheless, out of fifteen cities present in the ranking, eight are located in the high-tech cluster of Silicon Valley. San Francisco is the city that displays the highest rise having overtaken four cities. These results echo the findings of Balland et al. (2020) who found patents in complex activities to be spatially concentrated around large cities such as San Jose, San Francisco, Los Angeles or New York. A number of cities maintained their position, thus contributing to the overall stability of the ranking. Once again, since I am considering a very limited time period, it is logical to observe rather few fluctuations in the graphs between 2008 and 2015.

The slope graph for trademarks (Figure 11) reflects the findings of the state level: the data is less concentrated and is spread across nine states, including California and Texas. Additionally, the positions of cities in this ranking vary more over time than the ones for patents. This can be associated with the fact that I4 patents possibly build upon more complex knowledge, which makes invention stickier across space (Balland & Rigby, 2017) compared to trademarks. This would in turn allow leading cities in I4 patenting to maintain their positions more easily over time. The trademark rankings consist of some of the country's biggest cities and centres of consumption. Given that trademarks are associated with knowledge intensive services (Filippetti et al., 2019), it seems rather logical that they appear to be concentrated in cities that are amongst the most connected partially owing to their capacity of providing global services (Taylor & Derudder, 2004). Furthermore, the four cities that have in past literature (Tschang & Vang, 2008) been identified as video games clusters (Los Angeles, New York, San Francisco, Seattle) are all present in this ranking, in line with the observations of Figure 10.

It is worth mentioning again that as some cities were recorded under incorrect states and could not all be re-classified with certainty; it is possible that the ranking is not an entirely accurate depiction of reality. For instance, the presence of Wilmington (Delaware) can come across as a surprise. Although, Delaware did appear in Figure 10 – possibly because of the same data issue –; and it is the sixth state with the highest density throughout the country which could in turn facilitate the development of clusters.

Figure 11. Evolution of registered IPR across all technology types in U.S. cities.

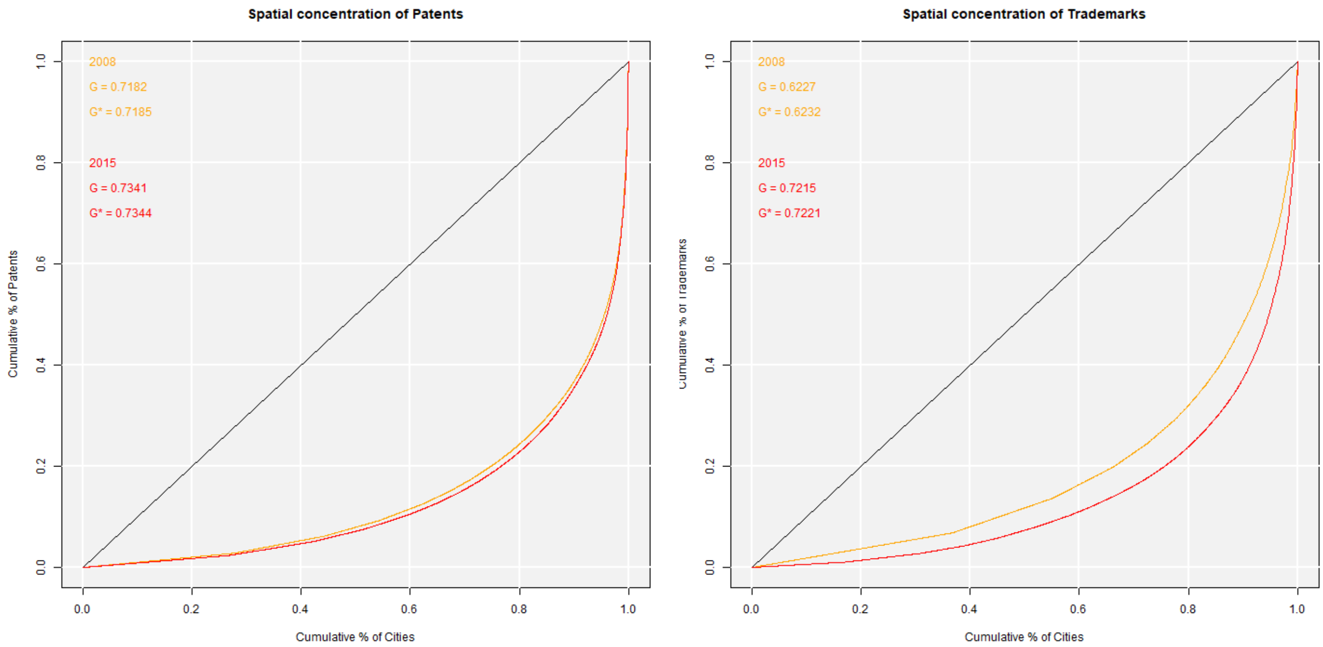


To evaluate the extent to which IPR filings in I4 are spatially concentrated across cities, the literature on regional inequalities has established that the Gini coefficient could be used as a measure of concentration (Wieland, 2019)[14]. A Gini coefficient of 0 implies that there is perfect equality, or in this case, no concentration, whereas a value of 1 indicates no diffusion whatsoever. Combined with the Lorenz curve for graphical purposes, Figure 12 shows how Gini coefficients evolved between 2008 and 2015 for patents and trademarks. In both cases, we observe relatively strong concentrations of innovation activity in cities, with 0.73 for patents and 0.72 for trademarks in 2015. While patents display higher spatial concentration, both IPR show increases in spatial concentration. However, that increase is larger in the case of trademarks and reaches 16%. Both Figures 11 and 12 indicate that the diffusion of I4 technologies from development to market application is limited: knowledge is rather sticky. This is especially true in the case of patents, as was shown in previous studies (Balland

[14] The Gini coefficient allows to measure concentration by estimating the discrepancy between the distribution of a certain variable and a reference – or equal – distribution, in this case, the Lorenz curve (Wieland, 2019)

& Rigby, 2017). It is worth noting that if the geographic information of trademarks had not only been assigned to headquarters, then the concentration would most likely be even lower.

Figure 12. Spatial concentration of IPR filings at the city level.



6. Discussion

In this research, I present how firms are developing and adopting I4 technologies using both patent and trademark data. To answer the first research question, I started by studying the vocabulary range of patents and trademark, and then I examined which I4 technologies were being developed and exploited in their applications across three different technology categories (core, enabling, application). In this manner, it was possible to inspect two key elements to industrial revolutions: technological development and adoption. Patent data, as expected, was insightful when investigating the geography of invention and determining the I4 technologies firms are developing (Mendonça et al., 2004; Filippetti et al., 2019). On the other hand, trademarks were used to estimate the innovation side of I4 or how firms applied I4 technologies. However, I was unable to discern whether technology adoption did indeed illustrate patterns of innovation or whether it showed patterns of diffusion and product differentiation. Nevertheless, the word clouds showed that trademarks could be more closely associated with services, as the literature had argued (Castaldi, 2019). Across both IPR types, there seemed to be an increased interest surrounding I4T, as made evident by the growth trends between 2008 and 2015. In addition, the variety of words present in the trademark word clouds hinted towards the pervasive potential of I4 by showing a glimpse of how I4T were being applied across different sectors.

Furthermore, general trends across all industries beyond the scope of I4 showed larger trademarks filings compared to patent applications (WIPO, 2019). This is a logical observation given the lesser restrictions associated with trademark filing and the lower costs required (Castaldi, 2019). Yet, the numbers in Table 4 showed a contrasting pattern: there were more patents than trademarks (except in 2017). Consequently, this could imply that there is still a large amount of unexploited potential for firms to invest and trademark in I4. Alternatively, this could also be a result of the trademarks' statements being non-exhaustive. Because of this, each trademark would have less chances of being selected by the keyword filter. One way to partially address the limitations of keyword filters would be to inductively select the keywords following more advanced algorithms, in the spirit of Joung & Kim (2017). Perhaps a better way to tackle this would be to use companies with I4T patents as a starting point and then study all these firms' trademarks to determine how exactly the application of technologies is happening. This would also help circle around the issue mentioned before: when filing for trademarks, companies are not required to specify which technologies they are using.

The second research question was divided into two parts. Firstly, it focused on identifying the type of firms that were involved in the development and application of I4T. Secondly, it determined the extent to which the technologies were being exploited across the U.S. Results of the industry-level analysis showed that the semiconductor business within I4 proved to be one of the key industries within which firms evolved and patented heavily. This also helped explain the prominence of IBM in I4 technology development. The other firms presented in the patent rankings were also what could be considered 'large' firms and evolved in high-tech industries, also confirming that patent data can be biased towards certain industries (Filippetti et al., 2019). In the context of this study, this bias for high-tech industries was reinforced by two aspects: the first simply being that patents are often more associated with high-tech manufacturing given their filing requirements (Filippetti et al., 2019; Castaldi, 2019), and the second reason being that I4 largely consists of high-tech industries. Considering this bias, replicating this study by also including IPR types, such as design rights, that tend to be better at capturing innovation of low- and medium-tech firms would possibly allow to counter this effect (Filippetti et al., 2019). The leading firms of the patent ranking reflected the results of Ménière et al. (2017). The authors had looked at EPO applicants and had also found that patent applications were dominated by large firms who tended to focus on ICT. Moreover, their findings also indicated relatively large concentration levels for patenting activity, and in this research, the thirty leading firms in terms of I4 patent filings owned a cumulative share of 31.6% of all I4 patents filed by U.S. inventors between 2008-2017.

As expected, trademarks were more disseminated across firms compared to patents, as displayed by the lower concentration levels of IPR filings by the leading firms. Because of the manner in which the data was filtered, all firm rankings included a majority of U.S. firms. Although across trademarks, there was a slightly larger presence of international firms. As in the case of patents, the presence of large conglomerates was noticeable, which could indicate that I4 is susceptible to resource accumulation (such as large amounts of data for instance), thus limiting greater filing levels among smaller firms. However, the most recurring industries in I4 trademark filings were concentrated around gaming, electronic and gambling-, casino- related industries. This reflected the closer association of trademark data with services and commercial application of technologies (Castaldi, 2019). This does appear to be a logical finding considering that gaming industries often rely on advancements in technologies such as cloud-computing, microchips (for improved graphics), microprocessors, or virtual reality which are all central technologies to I4. Moreover, these findings can be explained considering the involvement of native Americans with casino and gambling-related activities (Schaap, 2010), or by taking into account the welcoming legislation for gambling throughout the U.S. (Morse et al., 2007). Besides, the U.S. has positioned itself as a world leader in the video game industry, which has allowed several clusters to flourish throughout the country (Tschang & Vang, 2008).

The implications of the presence of firms[15] on both patent and trademark application rankings are twofold. On the one hand, it implies a certain dominance over the industry. These firms could be "IPR strategists" – firms that heavily file across different IPR types – in I4 based on the description by Seip et al. (2019). However, this can only be partially assessed by this data as it only covers two IPR types and focuses on the specific case of I4. On the other hand, this could reflect patterns of 'servitisation' of the economy (Lee et al., 2014): companies would not only be developing technologies (high patenting levels), but they would also be selling the resulting products or services that build on the applications of these technologies (high trademarking levels). Nevertheless, to test this assumption, future studies should also separate products from services as this thesis does not distinguish one from the other, even though trademark data can to some extent capture both (Castaldi, 2019).

[15] IBM, Apple Inc, Google, General Electric Company, Amazon, Microsoft, Hewlett-Packard, Honeywell, Cisco Technology, Dell

The geographical analysis aimed to study diffusion patterns of I4T across space, at different spatial levels. Patent applications were concentrated spatially: about a dozen states displayed higher levels of I4 patents per capita compared to the other states. Trademarks depicted lower concentration levels across U.S. states, and displayed larger diffusion patterns, although these appeared to be decreasing towards the end of the sample. Nevertheless, the observations implied that technology development is stickier than the application of technologies, as expected based on the literature (Petralia et al., 2017; von Graevenitz et al., 2019; Balland & Rigby, 2017). Combining this conclusion with the apparent overlap between states showing higher levels of concentration of both trademarks and patents confirms the argument of von Graevenitz et al. (2019) who argued that a larger distance limits the diffusion of innovation. Alternatively, these findings could also point towards an idea of "spatial servitisation" (Lee et al., 2014): some states could be centres of invention and innovation. In other words, these states could be aiming to secure both technology development and adoption. However, to assess this claim, one would need to look at whether it is indeed the same technologies who are fully being exploited in-state. However, as mentioned before, further research disentangling products and services would be necessary to exactly determine how firms are exploiting the application of technologies into either products or services.

At lower spatial levels, cities belonging to Silicon Valley illustrated their leadership as technology hubs. Conversely, the ranking of cities in terms of trademark filings showed much more dispersed patterns but also displayed greater fluctuations between positions, a finding also observable in Gini coefficients. In both IPR data instances, there are relatively large concentration levels across cities. Besides, it seems rather logical that the most inventive cities also happened to be among the largest cities in the U.S. considering the research of Balland et al. (2020) who found that a few large cities were accountable for greater shares of invention. Consequently, it would be interesting to establish the possible link between the involvement of large firms and the high spatial concentration in I4.

Despite the contributions made in this study on the potentials of combining IPR types to study technological development and innovation trends, there are a few other limitations to the ones mentioned before. According to the World Intellectual Property Organisation (WIPO), IPRs in the U.S. have been increasingly filed by international inventors and firms (WIPO, 2019). However, given the data used in this project was cleaned keeping only U.S. inventors owners, it considerably limited possibilities to observe how foreign firms trademarked and patented in the country if not done through U.S. holdings. For instance, the Korean firm Samsung Electronics was notably absent from this research, although previous reports had noted its world predominance in AI patenting (Dernis et al., 2019). Therefore, enlarging the data to world investors could improve the quality of the results presented in this research, and would also allow to consider how the dominant firms in both patent and trademark rankings are competing for market shares in I4.

Further limitations to this research include the already mentioned truncation of the data. By reproducing this study either over different time periods, or either waiting until all 2017 applications are finalised, results should become more consistent in later years. Also, identical trademarks/patents were sometimes selected more than once by the filter given the presence of several keywords in the statements and abstracts. For all analyses, I excluded duplicates in an impartial manner: if two datapoints were classified as "Core" and "General" respectively, only the first one was selected. Better ways to tackle this would have involved either studying the overlap between categories (as in Ménière et al., 2017), or to keep the IPR in the technology category where there were more keywords appearing. Due to time constraints, this was unfortunately not done. Moreover, a crucial aspect that was left out of this study was the value of IPR. While it would be possible to estimate patents' valuation using patent citation for instance, this would still prove to be a difficult feat for trademarks, which represents a profitable avenue for future research (Gao & Hitt, 2012; Castaldi, 2019). Furthermore, running the same analyses as done throughout this research on a better computer could also provide more insightful and detailed results.

Additionally, future research could further extend the approach of this project by including design rights to also capture innovation in low and medium tech enterprises, as previously done by Filippetti (2019). Alternatively, there is still room to conduct studies focused on disentangling the effects between overall increases in protecting intellectual property and increased interest in I4. One course of action would be to study trends in I4 patenting and trademarking over time, by also comparing innovation clusters between the previous industrial revolution and this one. Alternatively, larger datasets could be compiled by also including data from non-US inventors. In this scenario, more foreign firms might appear on the IPR filing rankings. In a similar manner, to directly assess the diffusion of innovation with trademark data, it would be possible to follow von Graevenitz et al. (2019) and to consider the frequency and location of the use trademarks' terms.

7. Conclusion

Recent literature has highlighted the economic impact of the Fourth Industrial Revolution. Those studies have made important developments in establishing reliable classifications and topology of Industry 4.0 but have yet to determine the interactions between firms and I4 as a whole. The foundations of I4 allow it to be considered a revolution and some of its technologies to be associated with GPTs, hence placing technological change and innovation at the centre of its development. However, empirical studies on innovation have so far prioritised patent data over other IPR types despite the successful effort of many studies to demonstrate the promising future of trademark data. The main purpose of this research was to unveil the different ways in which firms are engaging with I4 and its technologies across the U.S., using both patent and trademark data.

This research illustrated that combining patent and trademark data to study invention and innovation in I4 was worthwhile. Patent data shed light on technological inventions and showed that I4 patents focused more around hardware-related industries, while trademark data was more apt to offer insights into companies' applications of these technologies. Trademarks were associated with gaming and gambling industries. The results also indicated that a few large firms whose main focus was the semiconductor industry dominated patent applications, while trademark filings were much more spread across the firms present in the sample.

To spatially monitor I4, both patent and trademark proved insightful to study the development and application of I4 technologies. More specifically, patents were exploited to study the geography of invention whereas trademarks were used to determine the geography of innovation and technology adoption. Invention in I4 across the U.S. proved to be rather concentrated as demonstrated by the presence of a few technological hubs. Conversely, innovation in I4 appeared to be more spread across the U.S., with the notable presence of service cities in trademark filing rankings. A majority of these findings were expected considering the existing literature on the subject, but still highlighted possible avenues for future research.

Despite the limitations of this research, I showed that large and rather specialised firms have increasingly dominated the technological developments of Industry 4.0 and that not all places seem to have equally embraced those opportunities. While the application of I4 technologies appeared less concentrated in the hands of a few firms, it showed similar trends of becoming more spatially concentrated. Places that display high levels of invention and innovation in I4 could either be better exploiting the economic benefits of I4, but they could also be more exposed to threats of automation. Conversely, places that exploit I4T to lesser extents are in the short term less likely to face losses of jobs but might also be failing to seize the potential of I4. In any case, this research has shown that in order to understand the possible economic consequences of I4, a geographical perspective is crucial.

8. References

Amara, N., Landry, R., & Traoré, N. (2008). Managing the protection of innovations in knowledge-intensive business services. *Research Policy*, *37*(9), 1530–1547. https://doi.org/10.1016/j.respol.2008.07.001

Baldassari, P., & Roux, J. D. (2017). Industry 4.0: Preparing for the future of work. *People & Strategy*, *40*(3), 20–24.

Balland, P.-A., & Boschma, R. (2019). *Mapping the potential of EU regions to contribute to Industry 4.0*. Utrecht University, Department of Human Geography and Spatial Planning, Group Economic Geography.

Balland, P.-A., Jara-Figueroa, C., Petralia, S. G., Steijn, M. P., Rigby, D. L., & Hidalgo, C. A. (2020). Complex economic activities concentrate in large cities. *Nature Human Behaviour*, 1–7.

Balland, P.-A., & Rigby, D. (2017). The geography of complex knowledge. *Economic Geography*, *93*(1), 1–23.

Bekar, C., Carlaw, K., & Lipsey, R. (2018). General purpose technologies in theory, application and controversy: A review. *Journal of Evolutionary Economics*, *28*(5), 1005–1033. https://doi.org/10.1007/s00191-017-0546-0

Boschma, R. (2005). Proximity and Innovation: A Critical Assessment. *Regional Studies*, *39*(1), 61–74. https://doi.org/10.1080/0034340052000320887

Bresnahan, T. F., & Trajtenberg, M. (1995). General purpose technologies 'Engines of growth'? *Journal of Econometrics*, *65*(1), 83–108. https://doi.org/10.1016/0304-4076(94)01598-T

Broadberry, S., & O'Rourke, K. H. (2010). *The Cambridge economic history of modern Europe: Volume 2, 1870 to the present*. Cambridge University Press.

Castaldi, C. (2019). All the great things you can do with trademark data: Taking stock and looking ahead. *Strategic Organization*, 1476127019847835. https://doi.org/10.1177/1476127019847835

Chiarello, F., Trivelli, L., Bonaccorsi, A., & Fantoni, G. (2018). Extracting and mapping industry 4.0 technologies using wikipedia. *Computers in Industry*, *100*, 244–257.

Ciffolilli, A., & Muscio, A. (2018). Industry 4.0: National and regional comparative advantages in key enabling technologies. *European Planning Studies*, *26*(12), 2323–2343. https://doi.org/10.1080/09654313.2018.1529145

Cortada, J. W. (2012). *The digital flood: The diffusion of information technology across the US, Europe, and Asia*. Oxford University Press.

David, P. A., & Wright, G. (2003). General purpose technologies and surges in productivity. *The Economic Future in Historical Perspective*.

deGrazia, C. A. W., Myers, A., & Toole, A. A. (2020). Innovation activities and business cycles: Are trademarks a leading indicator? *Industry and Innovation*, *27*(1–2), 184–203. https://doi.org/10.1080/13662716.2019.1650252

Dernis, Gkotsis, Grassano, Nakazato, Squicciarini, vanBeuzekom, & Vezzani. (2019). *World Corporate Top R&D Investors: Shaping the Future of Technologies and of AI*. https://www.oecd.org/sti/world-corporate-top-rd-investors-shaping-future-of-technology-and-of-ai.pdf

Dosi, G., & Nelson, R. R. (2010). Technical change and industrial dynamics as evolutionary processes. In *Handbook of the Economics of Innovation* (Vol. 1, pp. 51–127). Elsevier.

Fagerberg, J. (2005). Innovation—A Guide to the Literature. In D. C. Mowery, R. R. Nelson, & J. Fagerberg, *The Oxford handbook of innovation*. Oxford university press.

Filippetti, A., Gkotsis, P., Vezzani, A., & Zinilli, A. (2019). How to survive an economic crisis? Lessons from the innovation profiles of EU regions. 40.

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. *Industry and Innovation*, *21*(4), 310–331.

Fontana, R., Nuvolari, A., Shimizu, H., & Vezzulli, A. (2013). Reassessing patent propensity: Evidence from a dataset of R&D awards, 1977–2004. *Research Policy*, *42*(10), 1780–1792.

Frank, A. G., Mendes, G. H., Ayala, N. F., & Ghezzi, A. (2019). Servitization and Industry 4.0 convergence in the digital transformation of product firms: A business model innovation perspective. *Technological Forecasting and Social Change*, *141*, 341–351.

Gallouj, F., & Savona, M. (2008). Innovation in services: A review of the debate and a research agenda. *Journal of Evolutionary Economics*, *19*(2), 149. https://doi.org/10.1007/s00191-008-0126-4

Gao, G. (Gordon), & Hitt, L. M. (2012). Information Technology and Trademarks: Implications for Product Variety. *Management Science*, *58*(6), 1211–1226. https://doi.org/10.1287/mnsc.1110.1480

Gotsch, M., & Hipp, C. (2012). Measurement of innovation activities in the knowledge-intensive services industry: A trademark approach. *The Service Industries Journal*, *32*(13), 2167–2184. https://doi.org/10.1080/02642069.2011.574275

Graham, S. J. H., Hancock, G., Marco, A. C., & Myers, A. (2013). *The USPTO Trademark Case Files Dataset: Descriptions, Lessons, and Insights* (SSRN Scholarly Paper ID 2188621). Social Science Research Network. https://papers.ssrn.com/abstract=2188621

Grinin, L., & Korotayev, A. (2015). Great Divergence and Great Convergence. *A Global Perspective. N. p.: Springer International Publishing*.

Helpman, E., & Trajtenberg, M. (1996). *Diffusion of general purpose technologies*. National bureau of economic research.

Hermann, M., Pentek, T., & Otto, B. (2015). Design principles for Industrie 4.0 scenarios: A literature review. *Technische Universität Dortmund, Dortmund*.

Hidalgo, C. A., Balland, P.-A., Boschma, R., Delgado, M., Feldman, M., Frenken, K., Glaeser, E., He, C., Kogler, D. F., & Morrison, A. (2018). The principle of relatedness. *International Conference on Complex Systems*, 451–457.

Izushi, H., & Aoyama, Y. (2006). Industry evolution and cross-sectoral skill transfers: A comparative analysis of the video game industry in Japan, the United States, and the United Kingdom. *Environment and Planning A*, *38*(10), 1843–1861.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics*, *108*(3), 577–598.

Janik, A., & Ryszko, A. (2018). Mapping the field of Industry 4.0 based on bibliometric analysis. *Proceedings of 32nd IBIMA Conference*, 15–16.

Joung, J., & Kim, K. (2017). Monitoring emerging technologies for technology planning using technical keyword based analysis from patent data. *Technological Forecasting and Social Change*, *114*, 281–292. https://doi.org/10.1016/j.techfore.2016.08.020

Kelley, M. R. (1994). Productivity and information technology: The elusive connection. *Management Science*, *40*(11), 1406–1425.

Klinger, J., Mateos-Garcia, J., & Stathoulopoulos, K. (2018). Deep learning, deep change? Mapping the development of the Artificial Intelligence General Purpose Technology. *ArXiv:1808.06355 [Cs, Econ]*. http://arxiv.org/abs/1808.06355

Kumar, K., Zindani, D., & Davim, J. P. (2019). *Industry 4.0: Developments towards the fourth Industrial Revolution* (Vol. 1–1 online resource). Springer; WorldCat.org. https://doi.org/10.1007/978-981-13-8165-2

Lee, J., Kao, H.-A., & Yang, S. (2014). Service innovation and smart analytics for industry 4.0 and big data environment. *Procedia Cirp*, *16*(1), 3–8.

Liao, Y., Loures, E. R., Deschamps, F., Brezinski, G., & Venâncio, A. (2018). The impact of the fourth industrial revolution: A cross-country/region comparison. *Production*, *28*.

Llerena, P., & Millot, V. (2013). Are trade marks and patents complementary or substitute protections for innovation. *Bureau d'économie Théorique et Appliquée Document de Travail*, *1*.

Mendonça, S. (2014). National adaptive advantages: Soft innovation and marketing capabilities in periods of crisis and change. In *Structural Change, Competitiveness and Industrial Policy* (pp. 149–166). Routledge.

Mendonça, S., Pereira, T. S., & Godinho, M. M. (2004). Trademarks as an indicator of innovation and industrial change. *Research Policy*, *33*(9), 1385–1404. https://doi.org/10.1016/j.respol.2004.09.005

Ménière, Y., Rudyk, I., & Valdes, J. (2017). *Patents and the Fourth Industrial Revolution* (p. 100). European Patent Office.

Morse, E. A., Goss, E. P., & Morse, E. (2007). *Governing fortune: Casino gambling in America*. University of Michigan Press.

Oztemel, E., & Gursev, S. (2020). Literature review of Industry 4.0 and related technologies. *Journal of Intelligent Manufacturing*, *31*(1), 127–182. https://doi.org/10.1007/s10845-018-1433-8

Pereira, A. C., & Romero, F. (2017). A review of the meanings and the implications of the Industry 4.0 concept. *Procedia Manufacturing*, *13*, 1206–1214.

Perez, C. (2010). Technological revolutions and techno-economic paradigms. *Cambridge Journal of Economics*, *34*(1), 185–202.

Petralia, S. (2019). *Tracing the Impact of a General Purpose Technology: The Case of Electricity* (SSRN Scholarly Paper ID 3144071). Social Science Research Network. https://doi.org/10.2139/ssrn.3144071

Petralia, S., Balland, P.-A., & Morrison, A. (2017). Climbing the ladder of technological development. *Research Policy*, *46*(5), 956–969.

Petralia, S., Balland, P.-A., & Rigby, D. L. (2016). Unveiling the geography of historical patents in the United States from 1836 to 1975. *Scientific Data*, *3*(1), 1–14.

Reischauer, G. (2018). Industry 4.0 as policy-driven discourse to institutionalize innovation systems in manufacturing. *Technological Forecasting and Social Change*, *132*, 26–33. https://doi.org/10.1016/j.techfore.2018.02.012

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, *98*(5, Part 2), S71–S102.

Schaap, J. I. (2010). The growth of the Native American gaming industry: What has the past provided, and what does the future hold? *American Indian Quarterly*, *34*(3), 365–389.

Seip, M., Castaldi, C., Flikkema, M., & de Man, A.-P. (2019). A taxonomy of firm-level IPR application practices to inform policy debates. 41.

Smith, K. (2005). Measuring Innovation. In J. Fagerberg, D. C. Mowery, & R. R. Nelson, *The Oxford Handbook of Innovation*. Oxford university press.

Strange, R., & Zucchella, A. (2017). Industry 4.0, global value chains and international business. *Multinational Business Review*.

Taylor, P. J., & Derudder, B. (2004). *World city network: A global urban analysis*. Routledge.

Tonelli, F., Demartini, M., Loleo, A., & Testa, C. (2016). A novel methodology for manufacturing firms value modeling and mapping to improve operational performance in the industry 4.0 era. *Procedia CIRP*, *57*, 122–127.

Tschang, F. T., & Vang, J. (2008). Explaining the spatial organization of creative industries: The case of the US Videogames industry.

United States Patent and Trademark Office. (2020). *Fiscal Year 2019 Performance and Accountability Report*. https://www.uspto.gov/sites/default/files/documents/USPTOFY19PAR.pdf

van Beuzekom, B., Dernis, H., Harhoff, D., & Squicciarini, M. (2020). Identifying and measuring developments in artificial intelligence: Making the impossible possible.

von Graevenitz, G., Graham, S. J., & Myers, A. (2019). *Distance (Still) Hampers Diffusion of Innovations*.

Wieland, T. (2019). REAT: A Regional Economic Analysis Toolbox for R. *Region*, *6*(3), R1–R57.

WIPO (2019). *World Intellectual Property Indicators 2019*. Geneva: World Intellectual Property Organization.

9. Appendices

9.1. List of R Packages Used

R Package	Version	Author(s)
ggmap	2.6.1	D. Kahle and H. Wickham
ggplot2	3.0.0	H. Wickham
ggwordcloud	0.5.0	E. Le Pennec and K. Slowikowski
gridExtra	2.3	B. Auguie
maps	3.3.0	R.A. Becker, A.R. Wilks, R. Brownrigg
openxlsx	4.1.0	A. Walker
plotrix	3.7-7	J. Lemon
RColorBrewer	1.1-2	E. Neuwirth
REAT	3.0.2	T. Wieland
rlist	0.4.6.1	K. Ren
stringr	1.3.1	H. Wickham
tidytext	0.2.3	J. Silge, D. Robinson

tidyverse	1.2.1	H. Wickham
treemap	2.4-2	M. Tennekes
viridis	0.5.1	S. Garnier

9.2. Words Deleted from the Text Frequency Analyses

<table>
<tr><td>Trademark Statements</td><td rowspan="2">and, of, the, in, out, for, to, a, on, use, non, with, as, no, it, line, namely, featuring, other, others, or, an, by, all, their, that, providing, is, first, one, at, second, be, may, from, are, least, can, based, includes, configured, each, more, which, layer, between, having, include, provided, such, using, associated, also, within, including, portion, current, set, when, has, disclosed, field, wherein, provide, about, nature, used</td></tr>
<tr><td>Patent Abstracts</td></tr>
</table>