Title
Diffusion of Breakthrough Technologies in the United States (1975-2005)

Author
Rietmann, Carsten Philipp

Publication Date
2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Diffusion of Breakthrough Technologies

in the United States (1975-2005)

A thesis submitted in partial satisfaction

of the requirements for the degree Master of Arts

in Geography

by

Carsten Philipp Rietmann

2017

ABSTRACT OF THE THESIS

Diffusion of Breakthrough Technologies

in the United States (1975-2005)

by

Carsten Philipp Rietmann

Master of Arts in Geography

University of California, Los Angeles, 2017

Professor David L. Rigby, Chair

This thesis examines the determinants of the spatial diffusion and adoption of breakthrough technologies, across industries and over time. It sets its focus on the United States between 1975 and 2005. Using patent data, this study uses survival analysis methods to test how geographical, social, and cognitive proximity, as well as additional covariates influence technological diffusion. In particular, an Extended Cox Hazard model is estimated and adapted to different subsets of the data. In total, 406 narrow technological fields within the United States Patent Classification are analyzed. These are all major technologies that were introduced after 1975. The thesis engages with breakthrough invention and novelty literature, as well as classic literature on spatial (innovation) diffusion as well as more recent proximity literature and technology-centered case studies. The results affirm that the expectations derived from theory regarding the role of proximities hold empirically. However, it also emphasizes the partial heterogeneity in these effects.

The thesis of Carsten Philipp Rietmann is approved.

Ka Yuet Liu

Eric Stewart Sheppard

David L. Rigby, Committee Chair

University of California, Los Angeles

2017

Table of Contents

List of Figures

List of Tables

1. Introduction

What is technological novelty? How do breakthrough technologies travel across space? Where are they adopted after their initial invention, and when? What influences these processes of diffusion? What roles do various dimensions of proximity play? Is it possible to detect explanatory factors for related technologies within certain technological fields? Do the dynamics of diffusion differ between diverse technologies, or are there some universal mechanisms at play?

These are the guiding questions that this study explores. The goal is to examine determinants of the spatial diffusion and adoption of novel and important technological fields, across various industries and technological domains, and over time. The focus is on the United States between 1975 and 2005. In this sense, the thesis is not concerned with the conditions that led to the invention of a breakthrough technology, but rather takes an ex-post perspective and examines the spatio-temporal diffusion and adoption of new ideas. Survival analysis is used to analyze and model the process of diffusion using patent data. In particular, a Cox Extended Hazard Model with both time-dependent and time-fixed covariates is developed and implemented for different levels of aggregation across technologies and time. Though geographical proximity has frequently been shown to determine the spillover or transmission of innovation, other forms of proximity – social and cognitive – are incorporated to provide a more comprehensive picture of the conditioning factors.

From a theoretical standpoint, the thesis engages with different branches of the literature: A brief review of work on the notion of novelty, the origin of inventions, and breakthrough technologies is provided. Further, more general work on spatial diffusion, different mechanisms of expansion and relocation, as well as more specific research on innovation diffusion is revisited.

Here, the classic literature (e.g., Hagerstrand, 1953; Griliches, 1957; Brown, 1981; Rogers, 1983) is used as much as more recent studies on the role of proximities. Lastly, the thesis surveys existing case studies of the spatial diffusion and adoption of technologies.

This study is structured as follows. Section 2 provides a brief overview regarding the origins of inventions, breakthrough technologies and novelty, and their use in research. Section 3 reviews relevant literature concerning the spatial diffusion of innovations, and covers various fields of research from general diffusion, through innovation, to newer concepts of proximity. Section 4 outlines contextual information on the use of patents in geographical innovation research and describes the data sources used for this study. Section 5 describes how novelty and breakthrough innovations are defined and identified. Section 6 introduces the methodology and describes the general and particular set-up of the Cox model in survival analysis. Section 7 shows empirical results from model estimation, and Section 8 offers a short conclusive discussion.

2. Novelty, breakthrough technologies, and the origins of inventions

This study takes an ex-post perspective and focuses on the adoption and diffusion of novelty: new breakthrough technologies. The conditions that led to the conception and invention of those technologies are not discussed. Yet, it is important to provide a fundamental understanding of novelty, its origins, and the characteristics of breakthrough technologies below.

2.1 Defining novelty and breakthrough technologies

Based on Arthur's (2009) seminal work on the nature of technology (2009), Verhoeven et al. (2016) distinguish two natures of inventions: Novelty in Recombination and Novelty in (Technological or Scientific) Knowledge Origins. First, Novelty in Recombination refers to the degree to which a new technology is based on the recombination of previously existing knowledge and components. Second, Novelty in (Technological or Scientific) Knowledge Origins is related to the extent to which a new technology is rooted in areas of knowledge which have not yet been exploited.

Since this study is framed around breakthrough innovations, it is important to revisit the various definitions of "breakthrough" that have been developed and employed in research. Usually, the term is used to distinguish discontinuous and radical inventions from incremental inventions (Dahlin and Behrens, 2005). While the first can be called 'structural deepening' (Arthur, 2009), the latter infers that a new category of technology for an invention has to be generated, for instance by issuing a new class in technological classification systems, or that 'redomaining' (Arthur, 2009) takes place. Incremental innovations are understood as small changes in existing technologies or simple extensions that do not alter the purpose of a technology, while breakthroughs refer to radical, unique, and novel advances in technology (Zhou et al., 2005) that also have an impact on technologies in the future (Dahlin and Behrens, 2005). This view reverberates with Dosi's (1982) distinction between 'normal procedures' and 'extraordinary breakthroughs', that also emphasizes the role of breakthroughs in shaping and initiating technological trajectories. This view can be extended by analyzing how an invention shapes later technologies or inventions in the future, for instance through implementing its technological elements into other later patents or by initiating novel technological fields, and is hence also used as a common definition for a breakthrough (e.g., Trajtenberg, 1990; Rosenberg, 1994; Fleming, 2001) or a radical invention (e.g., Schoenmakers and Duysters, 2010). Section 4 will also outline how these definitions are related to operationalizing breakthrough inventions in the context of this study.

Breakthrough inventions are also assumed to be more valuable and to have greater economic impact (Griliches, 1990). For instance, Kerr (2010) uses citations as a proxy for a patent's value and defines the patents within the top percent of forward citations of each technology as breakthrough inventions. Singh and Fleming (2010) use a similar approach for certain subpopulations of all patents. Equally analyzing at the subpopulation level, Castaldi et al. (2015) define thresholds to identify 'superstar patents' endogenously, rather than exogenously derived criteria like Kerr (2010).

2.2 Ex-ante characteristics of inventions

Conceptually, scholars have been analyzing radical inventions ex ante and ex post (Verhoeven et al., 2016). From an ex-ante perspective, particular (mainly technological) characteristics of inventions that have led to their creation are scrutinized. From an ex-post perspective, followed in this study, the impact on the subsequent technological development or the diffusion and adoption of the technology across space and over time is examined.

In general, the ex-ante standpoint often examines the characteristics of new technology and understands breakthroughs as disproportionately different from previous practices and procedures (Nelson and Winter, 1982; Shane, 2001), as adding new knowledge (Verhoeven et al., 2016) and as combining and re-assembling knowledge unexpectedly (Nooteboom, 2000). Note that some scholars argue breakthrough inventions do not necessarily employ previously existing technologies (Arthur, 2009; Banerjee and Cole, 2011) and it is this novelty that makes them valuable. Novelty is a *sine qua non* criterion to assess an invention's patentability, and hence it can be assumed that all patents introduce novelty. However, some patents are more radical and novel than others, as will be shown below. Some studies ignore the patent/invention-level and analyze the degree of novelty of a firm's capabilities and procedures as part of the breakthrough invention (Chandy and Tellis, 1998). In addition to these internal competences, the crucial knowledge inducing a radical discovery is believed to exist outside the scope of a particular firm's knowledge base (Hill and Rothaermel, 2003). Hence, firms need to adapt their internal processes to build absorptive capacity (Nahm et al., 2003).

Over recent decades, the large-scale analysis of patents has been identified as an alternative to qualitative methods, which are described below for the ex post perspective. The major advantage of patent data are their availability and historical coverage. There are various characteristics and components of patents that can be used for research. Patent information is used to examine ex-ante technological characteristics that can be linked to radical inventions. For instance, some studies rely on technological classifications or backward citation information to analyze the knowledge base and technological tradition that a patent relies upon. For example, Trajtenberg et al. (1997) look at the diffusion and location of backward citations across technology classes and argue that the more diverse the covered field of classes is, the more "original" the invention is. Further, a relatively large number of backward citations can also be understood as an index of reliance on previous knowledge and hence as reduced novelty (Sung et al., 2015; Wang et al., 2016). Related to this argument, Shane (2001) also counts the number of technological classes cited by a patent outside of its own technology classes to gauge radicalness. Additionally, the references and citations to scientific publications (non-patent references, the age of the patents cited, and the variance in these ages have been argued to reflect the novelty of knowledge embedded in inventions (Gittelman and Kogut, 2003; Verhoeven et al., 2016).

In addition to information on citations, inventions can be examined with other patent components as well. One possibility here are technological classifications. These are used in this study to define technological fields, but analyze them from an ex-post perspective regarding diffusion and adoption. In particular, the works by Fleming (2001, 2007), Strumsky and Lobo (2015), Akcigit et al. (2013), Youn et al. (2015), and Verhoeven (2016) on recombination and novelty being identified from technological class co-occurrences on patents need to be mentioned. The authors use a variety of definitions, such as the pairwise combination of technological classes being co-listed on a patent, to trace and identify novel recombinations as well as completely new subclasses that did not previously exist.

2.3 Ex-post characteristics of inventions

An ex-post perspective, adopted in this study, primarily examines the impact of inventions on the subsequent technological development or the diffusion and adoption of technology across space and over time. In addition to analyzing the technological and cognitive characteristics of an invention, the extent to which an invention shapes future technological development has generated considerable research. Scholars that relate to Dosi's (1982) concept of technology trajectories understand radical inventions through their impact on the subsequent evolution of knowledge. Connected to Kuhn (1962), they look for inventions with such a strong impact that they establish new knowledge frameworks or paradigms that channel future discovery. Further, from an economic standpoint, radical inventions can be examined through their impact on enterprises, industries, and economies (Anderson and Tushman, 1990; Henderson, 1993).

Qualitative methods have primarily been used to analyze the technological trajectories associated with core technologies (Dosi, 1982) and the impact of particular inventions on them and their position in them. Systematically analyzing and outlining technological trajectories has mainly been conducted on a case-level basis. This work relies on expert judgements on the radical components of a technology or invention produced in surveys of industry experts or managers (e.g. Dewar and Dutton, 1986; Pavitt et al., 1987; Acs and Audretsch, 1990). These methods are exposed to a higher risk of subjectivity and might produce a bias towards technologies that are popularly perceived as important, or towards those with which the particular expert is most familiar (Verhoeven et al, 2016). As it is often difficult to continuously gather data to create a panel dataset that is large enough to trace these temporal trajectories, these studies are often limited to case studies for specific industries. Also, patents have been used to measure the commercial value of inventions (Hall et al., 2005) and the broader technological and economic impact (Griliches, 1990; Fleming, 2001; Jaffe and Trajtenberg, 2002) and from an ex post perspective. Specific literature on the diffusion and adoption of patented technologies from an ex-post perspective will be reviewed in Section 3.

3. Literature review: Spatial diffusion of innovations

This section presents a literature review on the spatial diffusion of innovations. Starting from theory on spatial diffusion in general, it narrows its focus subsequently to cover the particular characteristics of the adoption of innovations and technologies across space, covering both classic and more recent research. After that, the concept of proximity and its different dimensions, crucial for this study, is introduced.

3.1 Spatial diffusion

Gould (1969) broadly distinguishes between two mechanisms of spatial diffusion. First, expansion diffusion is related to a spread of adoption across space. With expansion diffusion, the number of adopters increases over time, while earlier adopters in the center do not cease to use or be aware of the particular phenomenon of diffusion and continue to display high rates of incidences and activity. The diffusion of news or rumors are examples of expansion diffusion. Second, relocation diffusion refers to a variant of diffusion where a phenomenon moves across space. With relocation diffusion, the size of the phenomenon being studied does not necessarily increase. Migration or wildfires are examples of relocation diffusion.

In general, there are two mechanisms that drive the process of diffusion (Cohen and Tita, 1999). Contagious diffusion involves spread by direct contact and a non-hierarchical, primarily bilateral, transmission of a certain phenomenon. Taking an example from epidemiology, one person may be individually responsible for transmitting a disease to non-infected members of a population at risk. Hierarchical diffusion refers to the transmission of some phenomenon "through an ordered sequence of classes and places" (Cliff et al., 1981, p. 9). This can either be related to dissemination through centralized institutions such as the media, or a process that moves from larger metropolitan centers to the rural periphery. In many cases, diffusion cannot be uniquely assigned to one of these two mechanisms, but is often a combination of both. These two mechanisms could either be at work simultaneously, or be spatially selective, or may vary over time as the diffusion process unfolds.

The technology diffusion process that is modeled in this study with the Cox Extended Hazard Model can be identified as a non-hierarchical process of expansion diffusion since only the conditions leading to the first incidence of adoption (i.e., patenting in a certain technological field for the first time) in a city are examined. The non-hierarchical character is due to the way cities and their relationality are modeled in this thesis: through spatial proximity and not between different types or categories of cities. This will be further outlined below in Section 6.

3.2 Innovation diffusion

From a not exclusively spatial standpoint, Attewell (1992) focuses on two areas that are relevant for the analysis of innovation diffusion. First, adopter studies consider the particular characteristics of the adopters of the innovation, within firms often focusing on firm size or profitability as well as other organizational and environmental attributes. Second, macro-diffusion departs from the scale of the individual firm and investigates diffusion across the population of potential adopters. Rogers' (1983) uses the gravity model to examine diffusion across regions in space. These models are driven by the size of regions and by the distance between them. However, as Mansfield (1968) has noted, the gravity model proved less successful at explaining the diffusion of technological inventions than for social phenomena.

The macroscopic approach also identified the logistic or S-shaped curve that depicts the rate of adoption over time: Initially, few agents adopt the innovation, thereafter rates of adoption increase strongly before declining again as the innovation has saturated a population of potential adopters. In economic terms, the increase is often explained by a sizable decrease of a new technology's price which leads to a rise in demand (Hippel, 1988), or by increased awareness of the benefits from adopting a new technology From a sociological perspective, Burt (1987) argued that different social configurations lead to differently shaped S-adoption-curves. While cohesion is related to the concept of contagion outlined below and refers to direct interpersonal contact between an adopter and a potential user, structural equivalence is related to a similar position in relevant social networks of simultaneous adopters. Following Burt's research, structural equivalence has been found to be more influential than cohesion. Geroski (2000) offers a comprehensive critique of the S-curve that has been described above and compares this epidemic model with the probit model of diffusion that focuses on the different characteristics and interests of actors when considering the adoption of a novel phenomenon.

Yet, it is necessary to distinguish between the adoption of commercialized innovations by consumers or practitioners and the adoption of a new technology by researchers to advance research and development. As Rogers (1983) points out, the so-called hardware and software dimensions of a new technology differ in their diffusion dynamics. While "hardware is the tool, machine or physical object that embodies the technology" (Geroski, 2000, p. 605), software is seen as the necessary information base or knowledge to be acquired to actually use and operate the technology. Particularly with regard to patentable research, the latter aspect is obviously essential. Literature regarding this dimension of technology diffusion and adoption is outlined below.

Moreover, the common sequence of invention, innovation, and diffusion is only valid to a limited extent in the case of the research question in this study. As Silverberg (1991) points out, particularly the last two segments – innovation and diffusion – are not fully separable, since the original invention is always incrementally developed and altered when diffused. This is particularly the case for patents which are listed under the same technology code or in the same technological field. Still, however, the criteria of novelty and non-triviality have to be passed by every patent.

3.3 Classic literature

Foundations for the analysis of spatial diffusion of innovations were developed by Hägerstrand (1953) in the mid-20th century. Researching the adoption of agricultural practices by farmers, he postulated that adoption is heavily dependent on information about them. This information could either be transmitted through centralized media or interpersonal communication. Since the latter is spatially selective and constrained, Hägerstrand assumed that the geography of diffusion is dependent on the geography of personal contacts. In empirical analysis of tuberculosis controls in dairy herds, Hägerstand (1968) was able to show how spatial proximity between adopters shaped the diffusion process, and particularly positively influenced the time of adoption: Using Monte Carlo simulations and empirical analyses people living closer to earlier adopters were found to adopt the practice first, emphasizing the essential role of geographical propinquity. In addition, the size of settlements' populations also played a role and interacted with this mechanism of spatial proximity.

Brown (1981) reviews the innovation diffusion and adoption process and develops perspectives of communication, development, economic history, and market infrastructure. While the first two focus on adopters, the third is concerned with characteristics of the old (incumbent) and new technology, which is to be adopted. Particularly emphasizing the fourth viewpoint of market infrastructure, Brown argued that Hägerstrand's diffusion process focused only on the demand for information. He extended these early models by giving agency to the providers of innovation through encouraging non-propagators to adopt, since called diffusion agency or a supply-side perspective. One focus was related to the organizational structure of such propagating agencies (Deshpande et al., 1983). Brown compartmentalized the diffusion and adoption process into three steps: the establishment of diffusion agencies precedes the establishment of the innovation, which refers to actions towards triggering adoption in the particular region's population. Subsequently, the actual adoption takes place. Brown integrated these three dimensions in his market infrastructure framework.

In his examination of the creation and dissemination of hybrid corn, Griliches (1957) considered economic factors such as the profitability of the technology, the role of public and private institutions, as well as geographical spillovers and the share of potential adopters already having adopted the technology. Related to Brown (1981), he noted that "it does not make sense to blame the Southern farmers for being slow in acceptance, unless one takes into account the fact that no satisfactory hybrids were available to them" (Griliches, 1957, p. 507). He measured diffusion through the heterogeneity in the context of potential adopters and through the availability of hybrid corn, hence also giving agency to the suppliers, based on economic motivations: Specifically, the existence of differential financial incentives and, as mentioned, the innovation's profitability regulate the deviations in the rate of diffusion.

Mansfield (1968) provided support for Griliches' economic approach to explain diffusion, and developed a micro-level economic model to examine diffusion of industrial technology. While he focused on the general dissemination of information concerning a new technology in earlier work, he identified seven factors determining the intra- and interfirm diffusion rate, which can be categorized as structural dimensions, as well as adopter- and technology-related aspects. For instance, analyzing the diffusion of diesel locomotives in the railroad sector, he finds the investment's profitability, interfirm variations in liquidity and size, as well as differences in the date of first adoption to be significant variables. Further, this research emphasized the accumulation of knowledge, and how it relates to logistic diffusion and its accelerating pace.

3.4 Review of proximity concepts

The role of geographical proximity in the flow or the diffusion of technological knowledge has been emphasized not only in geography but across other disciplines (Hagerstrand 1953; Griliches, 1957). Such arguments have been confirmed by more recent evidence. For instance, in research on agglomeration economies, Arzaghi and Henderson (2008) show the importance of close spatial proximity for the creation of social networks in the advertising industry on New York City's Madison Avenue. With regard to innovation, Jaffe et al. (1993) and Sonn and Storper (2008) examine patent citations as evidence of the localization of knowledge flows. In these classical diffusion studies, geographical proximity has played a central role. Other forms of proximity bring this importance of spatial geography with regard to structuring knowledge flows into question, both theoretically (Boschma, 2005; Lagendijk and Oinas, 2005) and empirically (e.g., Breschi and Lissoni, 2001, regarding the importance of social networks).

The classic focus on geographical proximity in the analysis of diffusion and related areas has been widened to incorporate other kinds of proximity, to demonstrate how spatial proximity is conditioned by other forces, and to point at their interdependencies. For instance, Nooteboom (2000) shows that the governance and design of organizational structures, and in particular the proximity within these structures, is inherently connected to cognitive proximity. Aggregating different findings, including the French School of Proximity Dynamics (among others, Torre and Gilly, 2000), Boschma (2005) distinguishes five dimensions of proximity: cognitive, organizational, social, institutional, and geographical. The underlying assumption for all types of proximity is that more similar or proximate actors are more likely to behave in a similar or related way, such as adopting a new technology. The justification for these additional dimensions of proximity is the critique that spatial proximity is often driven through the underlying influence of the other types of proximity (Boschma, 2005). In addition to geographical proximity, this study incorporates measures for cognitive and social proximity.

Cognitive proximity postulates that agents are more proximate if they share a greater degree of intersection and correspondence of their knowledge and its structures. These refer to individual as well as organizational routines, skills, and procedures (Nelson and Winter, 1982). If these are overlapping, the absorptive capacity (Cohen and Levinthal, 1990) of the agents for a certain set of related knowledge is assumed to increase. Further, higher cognitive proximity is assumed to lead to higher cognitive homophily, which can be realized through exchange of knowledge and collaboration between actors with related sets of knowledge (Feldman et al., 2015).

Social proximity is related to the closeness of individual actors and the strength of the social ties between them. As Boschma (2005) points out, trust is the foundation of the strength of such ties, and is the outcome of repeated interaction. The relevance of social proximity has been increasing in research with scholars suggesting that the localized knowledge transfers of Jaffe et al. (1993) are driven as much by social proximity as spatial proximity. Breschi and Lissoni (2001) show that prior social ties between inventors exert significant influence on the structure of patent citations examined by Jaffe et al. (1993). They further their argument by the conjecture that studies have not isolated social from geographical proximity, since these two dimensions condition each other.

Additionally, Bathelt et al. (2004) show that interactive learning processes do not necessarily have to take place in permanent co-location (‘local buzz’), but can also be realized through pipelines that can both be maintained over distance or are created or reinforced periodically in person, for instance at trade fairs or conferences. In their view of the learning economy, Lundvall and Johnson (1994) add that these interpersonal connections go beyond a strict market- and exchange-based logic and that sharing information is part of the process of building trust, as well as its outcome. While Granovetter (1973) has found that simple – but useful – information can flow between weak social ties (e.g., between distant acquaintances), Breschi and Lissoni (2009) stress that more complex information, such as patentable knowledge, is being transmitted through strong social linkages. These are formed through repeated interaction. Examples are co-employment or long-term collaborations.

3.5 Case studies of diffusion and adoption of technologies

This section outlines a few key case studies in the research literature that explore the spatial diffusion and adoption of technologies. It is remarkable that there exist only a limited number of such studies with a spatial focus, while most examples are rather related to the analysis of the evolution of technological trajectories, as described above with relation to Dosi (1982). The latter can be categorized as aspatial and are organized separately in this section.

As a first example for case studies of explicitly spatial diffusion, Feldman et al. (2015) develop a case study in the framework of evolutionary economic geography of the spatial and temporal diffusion and adoption of the recombinant DNA technology after its invention in 1980. rDNA technology was introduced with U.S. patent US4237224 ("Process for Producing Biologically Functional Chimeras") by Stanley Cohen and Herbert Boyer. This technology was of a breakthrough nature in the sense that it introduced knowledge that could not be placed into an existing patent class and thus a new sub-class was added, United States Patent Classification (USPC) subclass *435 69/1* ("Recombinant DNA technique included in method of making a protein or polypeptide"). Tracing the emergence of new patents in this subclass allows the authors to examine the diffusion of rDNA technology away from the Bay Area, the site of its introduction.

When examining the relative influence and importance of geographical, cognitive, and social proximity in the diffusion of the rDNA technology, the authors find that diffusion to other regions was primarily conditioned by social proximity, which was identified from a co-inventor network in this subclass. Still, the channels of diffusion prove to be more complex, and also influenced by cognitive and – to a limited extent – by spatial proximity. The latter is shown to be influential in the later spread of the technology from the mid-1990s onward. Further, incongruent with theoretical assumptions, the role of university R&D is found to be negative.

Most studies only examine the distribution and spread of innovative activity as a whole rather than focusing on individual technologies. This is mainly conducted in a rather descriptive manner. For example, Moreno et al. (2005) analyze the diffusion of innovative activity, operationalized through patenting, in 17 European countries between 1978 and 1997. A knowledge production function is used to model patenting behavior, and the authors report that economic activity, agglomeration economies, and R&D expenditure are significantly related to patenting. With regard to the third of these, the authors find strong decay affects as the R&D activity of geographical proximate regions is found to be positive, too.

In the following, main examples of aspatial studies on technology diffusion are reviewed. As this thesis is aggregating different related technologies within the same ‘branch’ of the United States Patent Classification (USPC) technological structure, as outlined in the subsequent chapters, it is useful to provide a few examples of case studies on the technological trajectories of specific industries. So far, this has been mainly conducted through citation networks and bibliometric analysis.

Verspagen (2007) analyzes the trajectory of fuel cell research by mapping a patent citation network. Understanding different types of patents within such a trajectory as sources, sinks, and intermediate points, the study uses various metrics such as search path link counts and search path node pairs to identify the exact route of these trajectories. However, the search is not completely exploratory since relevant subclasses within the area of fuel cells are defined a priori. Verspagen builds on the earlier work of Hummon and Doreain (1989) who constructed a network of citations between scientific publications on DNA discovery and took an approach to build a ‘main path’ within this setup which is postulated to coincide with the primary transmission of knowledge and ideas in the field of research. Further examples include Su and Lee (2009) whose study in the field of electrical conducting polymer nanocomposites examines its technological evolution through a citation network with centrality measures. Here, relevant patents are selected through a search for United States Patent and Trademark Office (USPTO) patents with the keywords “nano” and “composite” in their title or abstracts. Lee and Wu (2010) examine the same industry in a closely related study, using centrality metrics in a patent citation network, while Barbieri (2015) follows technological trajectories of electric and hybrid vehicles through a citation analysis, employing evolutionary theory and defining the field through subclasses within the International Patent

Classification (IPC). Ruffaldi et al. (2010) investigate the rehabilitation and surgical robotics industry through citations.

With regard to bibliometric analyses, Kaplan and Vakili (2015) focus on fullerenes in nanotechnology and use the technique of topic modeling as a variant of text mining. Liu et al. (2011) explore the photovoltaics industry with a keyword co-occurrence analysis that identifies five sub-sectors within the photovoltaic industry which are subsequently used to analyze patent growth trajectories for these clusters.

However, many of these studies lack an explicit spatial focus and rather focus on the evolution of particular technological fields. As shown above, substantial research has been conducted on the diffusion of technologies by adopters and consumers (see also Baptista, 1999) as well as the importance of regional effects (Alderman and Davies, 1990). Also, the geographical diffusion of innovative activity in general without considering technological classes and forms of proximity has often been examined. Such studies are supplemented by work researching spatial diffusion through patent citations (e.g. Jaffe et al, 1993; Almeida, 1996; MacGarvie, 2005). Further, other fields of research look at the impact of R&D knowledge spillovers on productivity on proximate regions (Keller, 2002; Aldieri and Cincera, 2009). Feldman et al. (2015) serve as a notable example of studying a particular technology and subclass in the USPC. This thesis contributes to the literature by examining the spatial diffusion of a large group of technological fields that have recently emerged and have not been studied yet. The analysis is run at different levels of aggregation and clustering of technologies to compare and contrast the influence of the dimensions of proximity and other covariates.

4. Patents in geographical innovation research and data sources

This section briefly reviews the use of patents in geographical research on innovation, and describes the data sources used in this thesis as well as important definitions. It particularly emphasizes the connection between the identification of subclasses and their first occurrence, their aggregation to technological fields, and how their diffusion is examined in this study.

In the last few decades, patents have been increasingly employed to explore a range of technology-related issues in various disciplines, such as innovation studies, economics, economic history, and economic geography (Lamoreaux and Sokoloff, 1996; Jaffe and Trajtenberg, 2002; Acemoglu et al., 2013). Different elements of patents such as citations, claims, classification codes, as well as information on inventors, assignees, and their geography have been used – as described in the previous sections – to examine questions related to patterns of innovative activities, flows of knowledge and their conditions, the direction and pace of technological development, as well as the commercialization of scientific knowledge, among many others.

The use of patent data has proliferated due to the range of variables that may be extracted from patent documents, but also because of the historical and geographical coverage of patent information, and its availability. At the same time, it is important to note that the use of patent data has also been questioned, with some asking whether patents provide a representative and unbiased picture of innovative and economic activity. For instance, Pavitt (1985) stresses the fact that firms patent strategically and that there are other means of protecting inventions, such as speed of commercialization or secrecy. Still, Feldman et al. (2015) defend the use of patents by pointing to the informational character of the inventive operations of different organizations, particularly in industries where protecting intellectual property is crucial.

This work uses patent data to examine the diffusion of novelty and technological breakthroughs in the United States since 1975. I examine novel technologies, which are identified through patents that introduce a new subclass. The precise procedure to identify these patents will be described in the next section. I focus on new subclasses because I can locate their emergence in one Metropolitan Statistical Area (MSA) and then trace their diffusion from this location over time.

For these purpose, the NBER patent dataset is used as the primary data source to cover all patents that have been issued after 1975. These data are linked to further datasets that include the various covariates used in this study such as employment, population, National Science Foundation (NSF) grant information, and geographical coordinates. The latter have been used to compute Euclidian distances between all pairs of cities.

The United States Patent and Trademark Office (USPTO) is the governmental agency that inter alia examines patent applications and grants patents, which are a form of property right for the exclusive commercial use of a technology by the patent assignee. The USPTO differentiates between three main groups of patents: utility, design, and plant. Utility patents deal with the usefulness and function of technologies and can be grouped in the categories of (a) an improvement of an existing idea, (b) a composition of matter, (c) a manufacture, (d) a machine, and (e) a process. Design patents are issued for a new, original and ornamental design for a product and hence protects its aesthetic appearance. Plant patents are the most uncommon patent category and are issued for the discovery or invention of plants that are reproduced asexually.

This study is limited to utility patents which account for 90% of all granted patents (Strumsky and Lobo, 2015) and are designated for the invention of new and useful processes, machines, artifacts, or compositions of matter. The USPTO categorizes the technologies that account for the novelty of an invention with a scheme of technology codes, the United States Patent Classification (USPC). The USPTO has identified 438 primary technology classes, subdivided into approximately 167,000 sub-classes, hierarchically organized to provide increasingly detailed definitions of technology types. For instance, the primary Class 435 ("Chemistry: Molecular Biology and Microbiology") consists of patents within 823 subclasses. An example for a subclass is class "435 69/1 - Recombinant DNA technique included in method of making a protein or polypeptide", which had been discussed above.

At the most aggregate level, Hall et al. (2001) define six technological categories (Chemical, Computers & Communications, Drugs & Medicals, Electrical & Electronic, Mechanical, and Others) and 36 nested technological subcategories. These classes are aggregations of the 438 primary classes of the USPC.

The USPC classification is updated bi-annually. This includes the introduction of new subclasses, the focus of this analysis as a way to define novelty, as well as the subsequent reclassification of previous classes and the patents related to them (Harris et al., 2010; Wang et al., 2016). So-called classification orders are published by the Technology Centers of the USPTO. These display changes of the USPC to the previous month and state whether patents have been reclassified, whether new classes have been created and whether classes have been deleted (Strumsky et al., 2012). It is important to note that when the USPTO identifies a new class or eliminates an existing class this sets in motion an examination of all patents and their membership within a particularly class category. Hence, the current classification system provides a reliable structure for all patents issued by the USPTO since 1836.

Rather than discard patent data for subclasses below the first tier, all patents in the USPTO database are rolled up to the first-tier level. An example of the hierarchical structure of the USPC is provided in Figure 1 below. For the primary class 984, a series of 9 first tier subclasses are identified. One of these, 984 25/0, may be further disaggregated into the subclasses shown. All patents within the bounded area in Figure 1 are aggregated up into class 984 25/0.

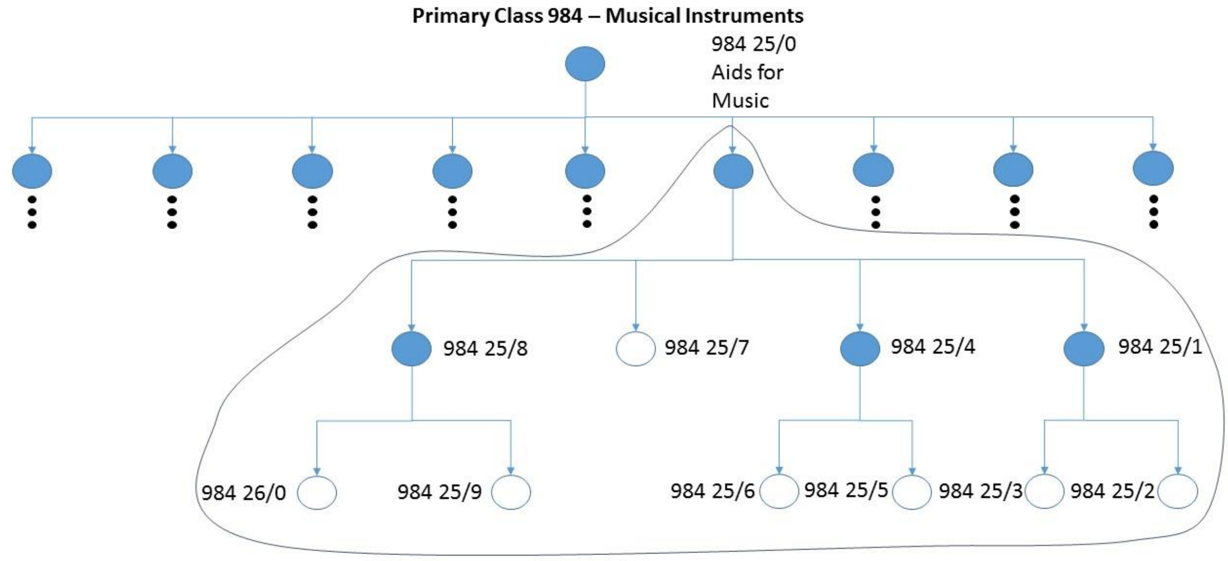

Fig. 1: Example of the structure of the USPC hierarchy, and of a technological field

5. The identification of novelty

This study uses patent data to identify technological novelty and to analyze the historical development of new ideas and their movement over space. Technological novelty is defined as the first occurrence of a USPC subclass on a patent. Patents have numbers that run consecutively, thus it is possible to identify the first instance of a new subclass in the USPTO system.

Before the procedure to identify these first instances of USPC subclasses is described in detail, some descriptive statistics on novelty and its geography are provided. Table 1 and Figure 2 display the development of novelty – operationalized through the introduction of new subclasses within the USPC – and patenting overall since 1975. Within a larger historical frame since 1836, two trends are apparent. First, the rate of the introduction of newly introduced subclasses steadily increased after 1836, reached a climax in the 1860s, and has been decreasing steadily since. Second, overall patenting activity in the United States rose monotonically until 1975, and experienced a particularly strong increase since 1980, reaching nearly 900,000 patents between 2001 and 2010. Figure 3 shows the overall spatial distribution of birthplaces of 10,195 subclasses that were introduced by the USPTO between 1975 and 2005. The five metropolitan areas with the highest counts are New York-Newark-Jersey City (1,139 subclasses), San Jose-Sunnyvale-Santa Clara (643), Boston-Cambridge-Newton (574), Los Angeles-Long Beach-Anaheim (551), and San Francisco-Oakland-Hayward (535). Figure 4 then depicts the distribution of the birth places of the novel 406 technological fields examined in this study among metropolitan areas since 1975. As described above in the previous section, these are 406 first-tier subclasses which have initiated the development of related 'lower-tier' subclasses. Significantly, only 99 metropolitan areas are among these birthplaces, with New York-Newark-Jersey City (48 subclasses), San Francisco-Oakland-Hayward (31), Los Angeles-Long Beach-Anaheim (25), and San Jose-Sunnyvale-Santa Clara (24) at the top.

Time period	Novelty	Patents
1975-1984	4,912 (491)	452,742 (45,274)
1985-1994	3,706 (371)	469,592 (46,959)
1995-2005	1,577 (143)	857,556 (77,960)
∑	10,195	1,779,890

Table 1: Development of novelty (newly introduced subclasses) and patenting in the US. Annual averages in parentheses.

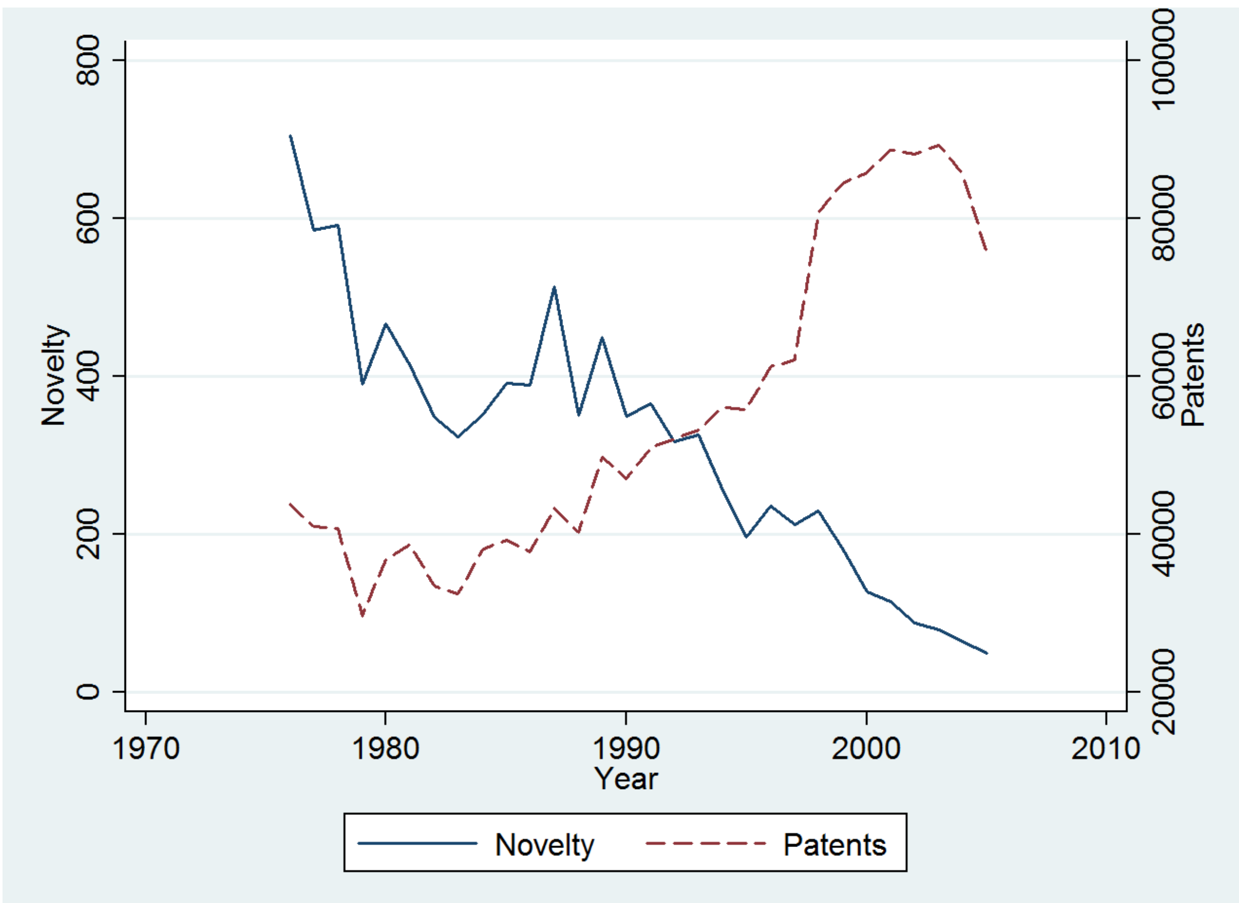

Fig. 2: Development of novelty and patenting in the US, 1975-2005

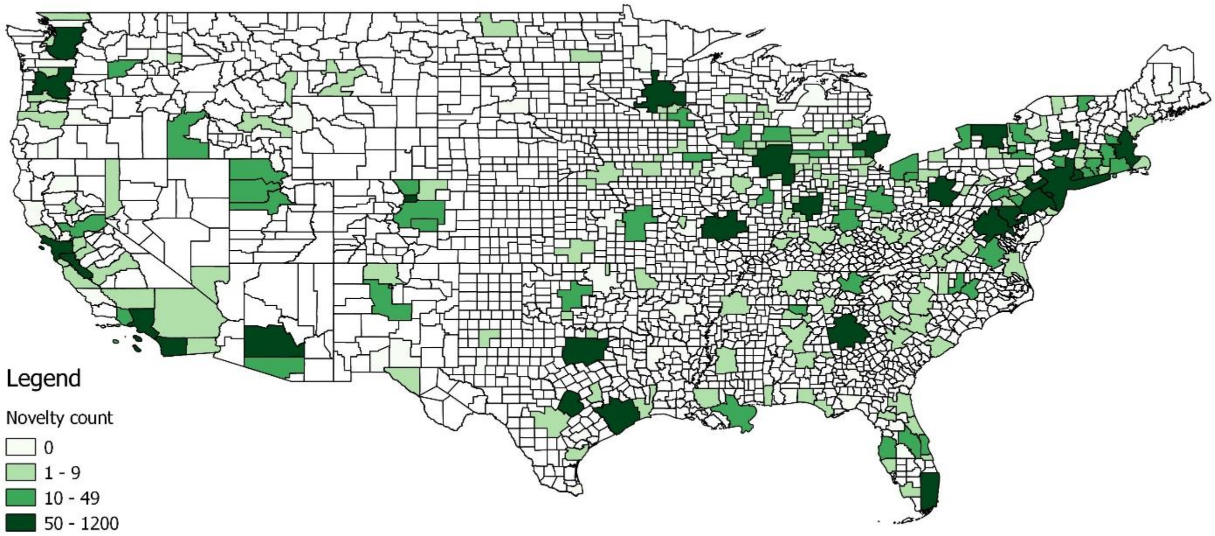

Fig. 3: Birthplace distribution of all new subclasses issued between 1975 and 2005

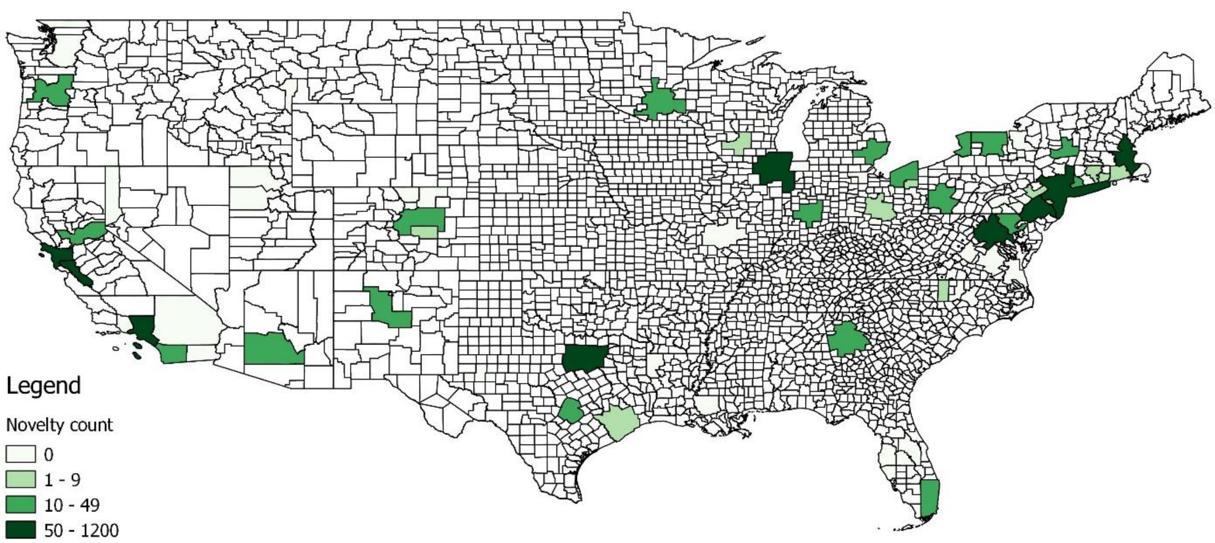

Fig. 4: Birthplace distribution of the novel 406 technological fields, 1975-2005

However, identification of the patent that introduces a new subclass solely based on the grant date or patent number is problematic. The problem can be exemplified by the case of recombinant DNA technology that introduced the subclass 435/69.1 as noted above by Feldman et al. (2015).

The authors determine the origin of this technology to be the famous Cohen-Boyer patent US4237224 ("Process for Producing Biologically Functional Chimeras"). With the introduction of this new subclass, the USPTO then identified two prior patents, generated before US4237224, that belonged in the subclass. Hence USPTO records now show the patent (US4237224) as the third patent in subclass 435/69.1.

As a remedy for this general problem, I introduce an additional criterion to identify the first patent that is associated with the emergence of a new USPC subclass. Since the introduction of a new subclass involves the production of a new type of technological knowledge, this should be reflected by a larger number of forward citations for the patent that introduces that new knowledge. Thus, forward citations are counted for every patent within a particular new subclass, up to five years after their first appearance. I use this five-year window to calculate the mean and standard deviation of citations for each subclass. I then define the true first instance of a subclass to be the historically first patent that has a citation count higher than the mean citation plus one standard deviation within this five-year window.

This study only considers first instances of subclasses at the 'first-tier' level (see Figure 3). The rationale for this decision is based on three reasons. First, the 'depth' of the hierarchy is very heterogeneous across primary classes, ranging from one to sixteen tiers. Hence, this aggregation to the first tier is designed to create standards for the broadest possible comparison. Second, many of the subclasses at very low levels of the USPC hierarchy do not contain enough patents granted between 1975 and 2005 to examine spatial diffusion and to compute the proximity metrics. Third, the computational power needed to compute the measures and separate models for all 10,000 subclasses created after 1975 would be beyond the scope of that available.

6. Model and Construction of Covariates

In this section of the thesis, I present a model of technology diffusion across metropolitan regions of the U.S. economy. The dependent variable in this analysis is binary: whether or not a city adopted a particular technological class in a given year. The probability of technology diffusion is modeled as a function of a series of time-fixed and time-varying covariates. The framework employed in the analysis is the Cox extended hazard model. The form of this model is explored below before presenting of the results of the analysis.

6.1 Extended Cox Hazard Model

This study employs event history and survival analysis to examine the spatial diffusion and adoption of breakthrough technologies between 1975 and 2005. This work uses the Extended Cox Hazard Model. Despite the fact that it does not explicitly control for unobserved heterogeneity, it has multiple advantages (cf. Mills, 2011) over the frequently used fixed effects panel models.

First, in contrast to other models in survival analysis, it is a semi-parametric model that is flexible in form, since a specific probability distribution does not need to be chosen a priori. Second, its “robust nature” (Mills, 2011: 90) implies that it generally fits data well. Third, although the baseline hazard probability (h_0) is unspecified, the effects of multiple independent variables can still be examined through the analysis of parameter estimates. Fourth, the exponent that is the essential part of the model also guarantees that estimated hazards are zero or positive. This is crucial since the hazard function inherently needs to take on values between zero and infinity. Fifth, different from logistic regressions, censored observations that experience no event in the examined time frame are included in the model’s computations and not dropped. This is important, for model parameters are derived from a much larger information set that includes many null events. The extended form of the Cox hazard model allows time-varying independent variables to be incorporated into the modeling framework, and is employed in the analysis below.

In its general form, the Cox Proportional Hazards Model is given by

$$h(t, X) = h_0(t)^{\sum_{i=1}^{p} \beta_i X_i}$$

where $h_0(t)$, the baseline hazard, is time-dependent but not dependent on covariates. The exponential term $\sum_{i=1}^{p} \beta_i X_i$ includes all covariates, but is not time-dependent. Hence, all independent variables are fixed over time. Thus, at a particular time t, the hazard is the result of the interaction of two terms: the time-dependent baseline hazard and the exponent which is the sum of all $\beta_i X_i$. The model is semi-parametric as the baseline hazard is flexible in its form, yet the independent variables are incorporated linearly.

The extended Cox model allows time-dependent covariates:

$$h(t, X(t)) = h_0(t)^{\sum_{i=1}^{p} \beta_i X_i + \sum_{j=1}^{p2} \delta_j X_j(t)}$$

where the exponential term is extended by adding $\sum_{j=1}^{p2} \delta_j X_j(t)$, which includes the time-varying covariates. Hence, contrary to the general form of the model, the exponential term now includes both time-independent and –dependent predictors. As in the general model, the coefficients of the regression are estimated by maximizing a partial likelihood function. Notably, the extended Cox model assumes that the impact of a time-varying independent variable on the probability of survival/failure at a certain time t is not contingent on the variable's values at earlier or later times, but solely on its value at time t. Further, although the independent variable may change over time for time-dependent covariates, there is only a single hazard ratio or coefficient (δ_j) computed for each time-dependent variable in the model. Box-Steffensmeier and Jones (2004) suggest that time-varying covariates in survival models should be incorporated in a temporally lagged way, to remedy the simultaneity problem of cause and outcome. In innovation diffusion research, Feldman et al. (2015) lag their time-varying independent variables by one year, an approach followed in this work. The reader is referred to Kleinbaum and Klein (2006) for an in-depth discussion of the extended Cox model.

The hazard ratio can be derived from a model coefficient and is provided in the empirical model output below. In this work, the hazard is the adoption of a technology in a city in a certain year. It can be interpreted as follows. The particular ratio can be conveniently converted into percentages, (ratio – 1)*100, representing the percent change of the hazard for each additional unit of a particular covariate for time-fixed independent variables, ceteris paribus. The parameter estimates for the time-varying independent variables for time-dependent continuous covariates can be also understood as the change in the probability of the hazard occurring for a unit change of the time-varying covariate.

As described above, this study uses not only time-fixed covariates such as the average distance to other cities, but also – and mainly – time-dependent independent variables such as two measures for social proximity or the sum of NSF grants. These will be introduced further below in this section. Incorporating time-dependent covariates into a Cox hazard model implies that the hazards in the model are no longer proportional. Hence, this adjusted model that used these time-varying measures as predictors is called the 'Extended Cox Hazard Model' (Mills, 2011).

6.2 Data structure

The data for the survival analysis through the extended Cox model used in this study are set up as a counting process (Therneau and Grambsch, 2000), also known as episode splitting (Blossfeld et al., 2012), to be explained below. This structure is advantageous to take both time-varying and time-fixed covariates into account. As Therneau and Grambsch (2000, p. 68, emphasis in the original) point out, "the basic viewpoint [of the method of a counting process] is to think of each subject [i.e., a city, in which a particular subclass is to be adopted] as the realization of a *very slow* Poisson process. Censoring is not "incomplete data", rather, the Geiger counter just hasn't clicked yet." Hence, splitting the survival analysis for a certain city over the time period from 1975 to 2005 into different episodes (i.e., years in this thesis) – regardless of the question whether the city eventually adopts the technology or not – enables the incorporation of time-dependent covariates and the estimation of the model through statistical software.

In total, there are observations for 406 technology classes for each of 366 metropolitan areas spanning 31 years from 1975 to 2005. Contrary to the conventional Cox Hazards model, each city consists of multiple sub-episodes (also referred to as splits). In this study, the data are structured in an isochronous way, with each observation representing a single year between the first global occurrence of the new subclass that represents a technological field (1975 at the earliest) and its first adoption in the specific city (2005 at the latest). Hence, the one-year window is not varied and flexible.

The event (or failure, as commonly called in survival analysis) status is coded as a binary variable that reflects whether a Metropolitan Statistical Area (MSA) *c* patents in a certain subclass j in year t (1), or not (0). All observations for cities that have not adopted a particular technology by 2005 are considered right-censored, but remain in the model, as explained above as one of the advantages of the Cox model in comparison with logistic regressions. Each city can only record a failure status (i.e., adoption) once for each technology. The model stops for each subclass and city at the first occurrence of such an event. All previous observations before the year of adoption in a city are recorded as non-failures, but contain the corresponding values of the time-varying independent variables for the particular year. An observation under the method of episode splitting hence contains (a) values of the covariates – both time-fixed and time-varying – during the subperiod, (b) the start and end year of the subperiod (e.g. 1975 and 1976), and (c) information on the status (failure or survival) of the entity (i.e., the city) during the subperiod (Blossfeld et al., 2012).

It is important to note that the concept of a technological field has been realized in the data set through recoding all subclasses in a particular field (review Figure 1 for a visualization of the concept) to their respective first-tier subclass. The construction of covariates will be described below. Overall, the hazard model explores how the independent variables influence the probability of technology adoption, occurring (diffusion) over space through time.

6.3 Construction of covariates

This sub-section briefly defines the major independent variables that are used in the Cox hazard model. These variables measure the geographical proximity of cities to one another, the geographical proximity of cities to neighbors that have already adopted a particular technology, the cognitive proximity of cities to a particular technology class and the social proximity of cities to one another. A series of additional, more general, covariates are also included in the model. These are discussed later.

Geographical proximity

From the theoretical arguments outlined above, geographical proximity is assumed to exert a positive influence on the diffusion of technology. Thus, cities that are closer together might be expected to transfer knowledge more readily than cities that are further apart. Geographical proximity between all pairs of cities is computed in Euclidean form based on latitude and longitude coordinates for the centroids of each metropolitan area.

This proximity dimension is computed in the following way. A matrix of geographical distances between all MSAs, and hence between all metropolitan regions in the United States, is being calculated. The Euclidean distances are based on the latitudes and longitudes of each MSA's centroid. Conventionally, it is hypothesized that cities spatially closer to other cities that have already developed the particular technology in previous time periods are more likely to develop patents in the same technological field, too.

There are two primary possibilities to incorporate geographical proximity into the Cox model. In a first, rather crude, form, the proximity of a city to all other cities in the U.S. city-system is employed. This measure is based on the average Euclidean distance from one city to all other MSAs, regardless of whether they have developed a particular technology or not. This is a time-fixed measure since the city's absolute locations do not change over time. The unit of measurement is kilometers. In a second form, the minimum distance from a potential adopting city to another scity that has already developed patents in a particular technology class is used. This study uses the Euclidean distance in kilometers to the nearest neighboring MSA with a previously developed patent in the relevant field – and hence the minimum distance – as the independent variable. Clearly, this decision is ambiguous, since the literature provides no clear answer. However, early computations confirmed the consistency of this choice. Further, this research follows Feldman et al.'s (2015) robust results in choosing this measure.

Year	*1980*	*1990*	*2000*
Minimum	0	0	0
Maximum	6169	5597	5476
Mean	1170	882	583
Standard Deviation	1008	870	659

Table 2: Descriptive statistics for minimum-distance-based geographical proximity (Euclidean distances, measures in kilometers)

Cognitive proximity

The "distance" between technology fields, what is sometimes referred to as cognitive proximity, is generally based upon the frequency with which technology classes are grouped together or co-classified on individual patents. Such calculations of cognitive proximity are usually performed between the 438 primary classes found in USPTO data. The technological fields examined in this thesis are more disaggregate than the primary class level, occupying the next tier down in the USPTO class hierarchy, what is referred to here as the first-tier subclass level. The probability of a city developing a patent in one of these first-tier technological fields might reasonably be thought to depend upon the distribution of existing inventive efforts within the city and how far those efforts are in knowledge space from the first-tier subclass.

The cognitive proximity, or relatedness of a city's knowledge base to a particular technological field, is measured in the following way. First, for each technological field with its first-tier subclass, a frequency vector is computed, which contains all USPC primary classes that have been grouped together on patents with this subclass between 1975 and 2005. In detail, C_{psu} =1 if patent p is classified under both subclass s and primary class u, and 0 otherwise. Summing over all patents in the time period results in this frequency vector.

While the simple co-class counts represent the technological relatedness of all primary classes to a certain subclass, they are of course also impacted by the absolute number of patents in these classes. Hence, these co-class counts need to be standardized by the square root of the product of the overall number of patents in this primary class and the number of patents in the particular subclass examined:

$$S_{us} = \frac{\sum_p C_{psu}}{\sqrt{\sum p_u * \sum p_s}}$$

Where $\sum p_u$ is the number of patents in primary class u and $\sum p_s$ is the number of patents in subclass s.

Subsequently, this global measure needs to be projected to a city-based perspective to compute the technological proximity of a city's knowledge stock to the particular subclass. For this purpose, the average relatedness of a city's patents to this subclass is calculated. High values of relatedness are assigned to patents that frequently co-list the subclass. In general, this average relatedness score AR for city m in year t is computed as follows.

$$AR^{mt} = \frac{\sum_j S_{us}^t * D_s^{ct}}{N^{ct}}$$

where S_{us}^t stands for the technological relatedness between patents in the 438 primary classes and patents in the analyzed subclass. This has been described above as the standardized co-occurrence.

Moreover, D_s^{ct} represents the patent count in technology s in year t in city c. N^{ct} is the total number of patents in year t in city c.

Computationally, this procedure is executed as follows. First, a matrix with primary class counts per city between 1975 and 2005 is computed. Second, counts for all 406 technological fields/subclasses per city over the same time period are appended. Third, the co-occurrences are standardized taking the count of the co-occurring primary classes of a particular subclass and dividing it by the square root of the product of the overall count of the particular primary class and the respective subclass' count. Both figures are not city-specific. Then, the average relatedness AR is computed as described above. Lastly, all average relatedness values are scaled by 100 to fit the extended Cox Hazard model calibration.

This measure is not a perfect choice since it is based on averages and not on the distribution of distances between technologies. For instance, as Feldman et al. (2015) note, it does not distinguish between small cities that produce patents that are an average distance away from the particular subclass and larger cities with higher innovative activity and patents that are distant and patents that are close to the relevant subclass. These two scenarios would yield similar results when calculating the average relatedness as presented above.

Year	*1980*	*1990*	*2000*
Minimum	0	0	0
Maximum	211.2	211.2	211.2
Mean	1.3	1.1	1
Standard Deviation	5.2	4.6	4.1

Table 3: Descriptive statistics for cognitive proximity

Social Proximity

The third proximity dimension that is part of the model in this study is social proximity. In the literature review above, it has been particularly contrasted with geographical proximity (cf. Breschi and Lissoni, 2001). The social interactions of relevance are the interactions between inventors working on particular technologies and endowed with particular knowledge sets related to the relevant technological field, or subclass. If these inventors collaborate on patents, the probable flow and transmission of knowledge is of interest and is assumed to connect cities to each other in these knowledge domains. To operationalize this aspect of proximity for this study, a metropolitan area social proximity matrix is being set up with 366 x 366 cells, the dimensions being the number of metropolitan areas (MSA) in the United States. Initially, all cells are filled with zeros. Then, for every subclass, the patents listed under the primary class of this subclass in a certain time period are identified. Co-inventors for patents in this primary class residing in two cities c_1 and c_2 increments the matrix cells (c_1, c_2) and (c_2, c_1) by 1. For co-inventors residing in the same city, the diagonal cell value is increased by 1. Once this has been re-iterated for all relevant patents, the row or column average is computed.

This value represents the average social proximity for each metropolitan area to all other metropolitan areas for the primary class of the analyzed subclass. The values for each city and primary class were computed for a moving three-year window preceding the year of a particular observation. After testing different time frames, three years were found to be the most appropriate window. For example, for subclass 436 80/2 in CBSA 10180 (Abilene, TX) in 1977, all inventor data on patents in primary class 436 (Chemistry: Analytical and Immunological Testing) in this CBSA between 1975 and 1977 were identified and used to compute the statistic for social proximity. The primary class has been chosen as a broader category than the subclass itself to ensure that enough observations are available and to take closely related subclasses into account. It needs to be reflected that this measure can be critiqued for its generality. Also, all patents with sole inventors are excluded from this analysis.

It can be argued that particularly complex technologies are spatially sticky and difficult to transmit them over distances. Thus, another measure of social proximity is introduced that takes this consideration into account, using the same methodology described above but focusing on inventor collaborations within the same city, the main diagonal of the matrix described above. As with the first variant of social proximity, it is based on a moving three-year time frame and inventions within the primary class of the particular technology. Here, the share of intra-city collaborations is computed by dividing the number of co-inventorships within a certain city through the sum of all of a city's co-inventorships. The values are scaled by 10 to fit the Extended Cox Hazards model calibration properly.

Year	*1980*		*1990*		*2000*	
	Conventional SP	*Intra-city SP*	*Conventional SP*	*Intra-city SP*	*Conventional SP*	*Intra-city SP*
Minimum	0	0	0	0	0	0
Maximum	7.2	10	18.75	10	53.45	10
Mean	0.02	0.65	0.04	0.87	0.15	1.07
Standard Deviation	0.16	2.25	0.3	2.46	1.03	2.5

Table 4: Descriptive statistics for social proximity (SP)

Additional covariates

In addition to the four proximity measures described above, other covariates have been used in this analysis. First, the number of employees in every city by year, between 1975 and 2005 (*Employment*). Second, the number of inhabitants of every city, by year between 1975 and 2005 is used as well (*Population*). Third, the cumulative age of backward citations of patents issued in a particular city in a certain year is included (*Age of Citations*) as a proxy for the average novelty and radicalness of inventions, assuming that a higher average age of backward citations makes inventions less radical and valuable. Fourth, the sum of NSF grants per city per year is used as a proxy for R&D resources (*NSF Grants*). All covariates are lagged one year. It is important to note that all these covariates as well as the majority of the proximity measures are time-dependent, which has important implications for the structure and set-up of the model for survival analysis.

7. Empirics

In total, the spatial diffusion and adoption of 406 technological fields (i.e., subclasses within utility primary classes) is being examined in this study. Digest as well as design classes are excluded. Table 5 shows the distribution of these technologies with respect to Hall's six aggregate classes *Chemical*, *Computers & Communication*, *Drugs & Medical*, *Electrical & Electronic*, *Mechanical*, and *Others*. The 406 subclasses examined here are distributed across 117 of the 438 primary classes of the USPTO. Analysis focuses on these 406 patent first-tier subclasses because they have come into existence since 1975 and thus the development of patents within these classes can be traced year-on-year across the metropolitan areas of the United States.

Chemical	*Computers & Communication*	*Drugs & Medical*	*Electrical & Electronic*	*Mechanical*	*Others*	$\sum$
64	109	72	77	45	39	406

Table 5: Distribution of the 406 subclasses in Hall's (2001) six classes

A number of models are estimated in this study. First, I explore aggregate results from a model that spans all years and technological fields. Thereafter, I run the diffusion model separately for a series of technology classes at different levels of aggregation and over different time-periods. The disaggregate analysis is performed to identify sector- or time-specific results that may be of interest.

Table 6 shows the results of the Extended Cox Hazard Model for the complete sample of all 406 subclasses, 366 cities, and 30 years. The model includes 12,420 'failures', i.e. events of adopting a subclass in a city, amongst a total of 2.5 million observations.

	Hazard ratios
	Model 1: Full sample
Lag Geographic Proximity	0.99900***
	(0.00003)
Three-year Social Proximity	1.04435***
	(0.00507)
Intra-city Social Proximity	1.20630***
	(0.00257)
Cognitive Proximity	1.00788***
	(0.00059)
Average Geographic Distance to Other Cities	1.00032***
	(0.00002)
Lag NSF Grants	1.00000***
	(0.00000)
Lag Employment	1.00000***
	(0.0000001)
Lag Population	1.00000***
	(0.0000000)
Lag Age of Citations	1.00000***
	(0.0000001)
Failures	12420
Observations	2,481,496
Log Likelihood	-62,057.97000
Note:	*p<0.1; **p<0.05; ***p<0.01
	Ties are being handled with the Breslow method. Except for cognitive proximity and the average geographic distance, all covariates are lagged one year. The p-values are computed by logging the hazard ratios (which yields the regression coefficient) and by dividing through the standard error.

Table 6: Model results of complete sample

All hazard ratios are significant, not unexpected given the large number of observations in the model. A hazard ratio significantly different from unity suggests that the independent variable in question exerts a significant influence on the probability of a city adopting a technology. Here I briefly run through the influence of all the independent variables for the general model reported above. First, geographical proximity is smaller than 1 and reflects that a greater geographical distance to the nearest neighboring city decreases the likelihood for the adoption of a technology to occur.

Second, as the only counter-intuitive exception among the hazard ratios, the average geographic distance enters the model with a marginally greater-than-one hazard ratio, reflecting that a greater mean distance to other cities actually increases the probability for a technology's adoption. While the reasons for this result remain unclear, it can be speculated that this time-fixed measure is very coarse, and hence presents no clear mechanism for the transmission of knowledge. Also, this could be an indication of a 'big-city-effect' determining the diffusion of technologies, with large cities with greater spatial distances between them disproportionately often adopting the technologies. A number of models have been estimated without this variable. Since results for estimated coefficients qualitatively stayed the same, the variable has not been dropped to guarantee consistency.

Third, the hazard ratios for both metrics of social proximity are greater than one, implying that cities with greater social proximity in the particular technological primary classes are more likely to develop patents in a certain technology.

Fourth, cognitive proximity also enters the model with a coefficient significantly greater than one. Hence, greater cognitive proximity of a city's knowledge stock to the particular technological field positively influences the probability for adoption.

Fifth, the hazard ratios for all four additional city-level covariates of NSF grants, employment, population, and the age of citations are significant, but deviate only very slightly from 1. Overall, these are affirming results that underline the importance of the theoretically derived impacts on knowledge transmission – geographical, social, and cognitive proximity – even when estimating the Extended Cox Hazards Model across all cities, years, and technologies.

Separate models are now explored for technology sub-classes that are part of the six broad technology classes identified by Hall et al. (2001). The results are presented in Table 7. As in the previous model, the hazard ratios are significant and consistently in the expected direction for geographical proximity, cognitive proximity, and intra-city social proximity.

	Hazard ratios					
	Model 2: Chemical	Model 3: Computers and Communication	Model 4: Drugs and Medical	Model 5: Electrical and Electronic	Model 6: Mechanical	Model 7: Others
Lag Geographic Proximity	0.99922***	0.99904***	0.99921***	0.99889***	0.99885***	0.99856***
	(0.00007)	(0.00006)	(0.00007)	(0.00008)	(0.00011)	(0.00013)
Three-year Social Proximity	1.20404***	0.98694***	1.08192***	1.05790***	1.16482***	2.13474***
	(0.02655)	(0.00847)	(0.01319)	(0.01053)	(0.02151)	(0.16872)
Intra-city Social Proximity	1.22136***	1.19237***	1.19484***	1.23305***	1.16177***	1.16136***
	(0.00732)	(0.00424)	(0.00603)	(0.00641)	(0.00877)	(0.01442)
Cognitive Proximity	1.06454***	1.09165***	1.02355***	0.99868***	1.05118***	1.37189***
	(0.00453)	(0.00650)	(0.00119)	(0.00128)	(0.00954)	(0.02452)
Average Geographic Distance to Other Cities	1.00032***	1.00038***	1.00029***	1.00035***	1.00031***	1.00039***
	(0.00006)	(0.00003)	(0.00004)	(0.00005)	(0.00007)	(0.00007)
Lag NSF Grants	1.00000***	1.00000***	1.00000***	1.00000***	1.00000***	1.00000***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Lag Employment	1.00000***	1.00000***	1.00000***	1.00000***	1.00000***	1.00000***
	(0.0000002)	(0.0000001)	(0.0000001)	(0.0000002)	(0.0000003)	(0.0000003)
Lag Population	1.00000***	1.00000***	1.00000***	1.00000***	1.00000***	1.00000***
	(0.0000001)	(0.0000001)	(0.0000001)	(0.0000001)	(0.0000002)	(0.0000001)
Lag Age of Citations	1.00000***	1.00000***	1.00000***	1.00000***	1.00000***	1.00000***
	(0.0000002)	(0.0000001)	(0.0000002)	(0.0000002)	(0.0000002)	(0.0000003)
Failures	1404	4772	2362	1919	1057	906
Observations	435,500	568,218	472,202	463,695	282,074	259,807
Log Likelihood	-6,676.00400	-23,347.85000	-11,704.43000	-9,496.04900	-5,666.35400	- 4,610.83300

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Ties are being handled with the Breslow method. Except for cognitive proximity and the average geographic distance, all covariates are lagged one year. The p-values are computed by logging the hazard ratios (which yields the regression coefficient) and by dividing through the standard error. The model is clustered on cities.

Table 7: Model results for Hall's (2001) six classes

The results in Table 7 are largely in line with expectations, though the coefficients in two cases are significant with the wrong sign. In terms of the three-year measure of social proximity, Model 3 for *Computers and Communication* has a coefficient that is significantly below 1 (or negative). Thus, as social proximity between inventors in different cities increases, so the probability of technology adoption declines.

It could be argued that patents in this technological class, which consists of complex knowledge and is hence spatially sticky, is rather developed through intra-city collaborations between inventors. This conjecture is supported with the coefficient on intra-city social proximity that is greater than one (positive). The coefficient for cognitive proximity is less than one (negative) for *Model 5: Electrical and Electronic*, implying that cities with a knowledge base close to technologies in electrical and electronic classes are less likely to adopt a new related technology. While the reasons remain unclear, it could be argued that cities with a strong base in these areas of research have outdated knowledge and are hence not capable to participate in the development of novel technologies. Additionally, it may be the case that this aggregate class consists of various highly heterogeneous technologies which are not necessarily closely related. As Table 5 indicated above, *Electrical and Electronic* represents the class with the second highest number of subclasses in the model (77 of 406). This conjecture is explored further below.

Table 8 continues the path of estimating the Cox model for more disaggregate technological areas and provides a more detailed analysis and examines a selection of Hall's (2001) 36 aggregated technological classes. The results for *11 – Agriculture, Food, Textiles* (3), *22 – Computer Hardware & Software* (66), *32 – Surgery & Medical Instruments* (13), as well as *33 – Biotechnology* (22) (number of subclasses in estimation)are displayed. Still, the tendency of hazard ratios is similar to the models run for the complete data set as well as Hall's six classes. The only exception is *11 – Agriculture, Food, Textiles*, which produces a counter-intuitive hazard ratio for the time-fixed average geographic distance, as well as disproportionately high values for cognitive and, particularly, the conventionally measured three-year social proximity. It needs to be mentioned that this cluster consists of only three subclasses. Hence, extreme results are not averaged out as could be the case in bigger clusters.

	Hazard ratios			
	Model 8: Agriculture, Food, Textiles	Model 9: Computer Hardware and Software	Model 10: Surgery and Medical Instruments	Model 11: Biotechnology
Lag Geographic Proximity	0.99505***	0.99891***	0.99929***	0.99907***
	(0.00203)	(0.00007)	(0.00017)	(0.00011)
Three-year Social Proximity	27.60224***	0.97987***	1.09236***	1.21083***
	(2.07842)	(0.00958)	(0.06120)	(0.03332)
Intra-city Social Proximity	1.49080***	1.18291***	1.15925***	1.18799***
	(0.12926)	(0.00497)	(0.01986)	(0.00774)
Cognitive Proximity	2.92296***	1.07599***	1.76190***	1.02249***
	(1.01715)	(0.00684)	(0.13264)	(0.00123)
Average Geographic Distance to Other Cities	0.99749***	1.00033***	1.00037***	1.00021***
	(0.00324)	(0.00004)	(0.00016)	(0.00006)
Lag NSF Grants	1.00000***	1.00000***	1.00000***	1.00000***
	(0.0000001)	(0.00000)	(0.00000)	(0.00000)
Lag Employment	0.99999***	1.00000***	1.00000***	1.00000***
	(0.00001)	(0.0000001)	(0.000001)	(0.0000002)
Lag Population	1.00000***	1.00000***	1.00000***	1.00000***
	(0.000004)	(0.0000001)	(0.0000003)	(0.0000001)
Lag Age of Citations	1.00000***	1.00000***	1.00000***	1.00000***
	(0.000002)	(0.0000001)	(0.000001)	(0.0000002)
Failures	6	3528	200	1504
Observations	15,014	290,502	73,968	140,946
Log Likelihood	-16.88079	-17,499.28000	-1,014.18800	-7,534.78900

Note: *p<0.1; **p<0.05; ***p<0.01

Ties are being handled with the Breslow method. Except for cognitive proximity and the average geographic distance, all covariates are lagged one year. The p-values are computed by logging the hazard ratios (which yields the regression coefficient) and by dividing through the standard error. The model is clustered on cities.

Table 8: Model results for technologies in Hall’s (2001) 36 aggregated technological fields

Table 9 depicts a differentiated analysis for the conditions of spatial diffusion and adoption before and after 1990 for all technological fields. It could be possible that technological changes and other developments have led to different mechanisms for technology diffusion being determinant between these time periods. However, the fact that the hazard ratios deviate from one in the same direction opposes this hypothesis. An additional model with dummy and interaction effects for the proximity variables is estimated to test whether there are significant differences between the pre- and post-1990 coefficients. The p-values are 0.011 (geographical proximity), 0.00003 (social proximity), 0.11 (intra-city social proximity), and 0.007 (cognitive proximity), reflecting that these differences are indeed significant at a 95% confidence level.

	Hazard ratios	
	Model 12: Years before 1990	Model 13: Years after 1989
Lag Geographic Proximity	0.99887***	0.99913***
	(0.00005)	(0.00004)
Three-year Social Proximity	1.26200***	1.04800***
	(0.03124)	(0.00512)
Intra-city Social Proximity	1.22804***	1.19383***
	(0.00508)	(0.00302)
Cognitive Proximity	1.02079***	1.00633***
	(0.00153)	(0.00068)
Average Geographic Distance to Other Cities	1.00053***	1.00026***
	(0.00004)	(0.00002)
Lag NSF Grants	1.00000***	1.00000***
	(0.00000)	(0.00000)
Lag Employment	1.00000***	1.00000***
	(0.0000001)	(0.0000001)
Lag Population	1.00000***	1.00000***
	(0.0000001)	(0.0000000)
Lag Age of Citations	1.00000***	1.00000***
	(0.0000002)	(0.0000001)
Failures	2990	9430
Observations	752,922	1,728,574
Log Likelihood	-14,424.33000	-47,445.18000
Note:	*p<0.1; **p<0.05; ***p<0.01	
	Ties are being handled with the Breslow method. Except for cognitive proximity and the average geographic distance, all covariates are lagged one year. The p-values are computed by logging the hazard ratios (which yields the regression coefficient) and by dividing through the standard error. The model is clustered on cities.	

Table 9: Model results for different time periods

Table 10 compares the model results for cities with high and low patenting activity. Following Feldman et al.'s (2015) distinction between cities with either less than 10 or more than 90 patents per subclass, the model reveals that the determinants for the adoption of a novel subclass in a particular city almost universally have the same direction and significance. Cognitive proximity represents the only exception. While a higher cognitive proximity of a city towards the respective new subclass' field of technology increases the probability for low-patent city to adopt the technology, the opposite is found for high-patent cities.

	Hazard ratios	
	Model 14: Low-patent cities (< 10 patents)	Model 15: High-patent cities (> 90 patents)
Lag Geographic Proximity	0.99943***	0.99905***
	(0.00003)	(0.00024)
Three-year Social Proximity	1.03167***	0.99895***
	(0.00875)	(0.02097)
Intra-city Social Proximity	1.04441***	1.03803***
	(0.00342)	(0.01618)
Cognitive Proximity	1.00431***	0.99844***
	(0.00183)	(0.00185)
Average Geographic Distance to Other Cities	1.00009***	1.00010***
	(0.00002)	(0.00011)
Lag NSF Grants	1.00000***	1.00000***
	(0.00000)	(0.00000)
Lag Employment	1.00000***	1.00000***
	(0.0000001)	(0.0000003)
Lag Population	1.00000***	1.00000***
	(0.0000000)	(0.0000001)
Lag Age of Citations	1.00000***	1.00000***
	(0.0000001)	(0.000001)
Failures	9513	503
Observations	100,706	1,298
Log Likelihood	-27,861.31000	-1,170.16200
Note:	*p<0.1; **p<0.05; ***p<0.01	
	Ties are being handled with the Breslow method. Except for cognitive proximity and the average geographic distance, all covariates are lagged one year. The p-values are computed by logging the hazard ratios (which yields the regression coefficient) and by dividing through the standard error. The model is clustered on cities.	

Table 10: Model results for cities with differing levels of patenting activity

While these results affirm the general mechanisms – particularly regarding geographical, social, and cognitive proximity – that influence and condition the spatial diffusion and adoption of technologies, a question remains: Does aggregating data across individual technological fields drive some of the results reported above? Thus, running the Cox model for individual technology subclasses might be a useful extension. Some evidence of the variability in coefficients across individual technology subclasses is presented in Figure 5. Discussion of estimating the Cox model across individual USPTO subclasses continues below. Note that repeated hypothesis testing of this kind raises questions about the levels of significance at which the null hypothesis should be rejected. However, because this work is largely exploratory, I do not focus on these concerns here.

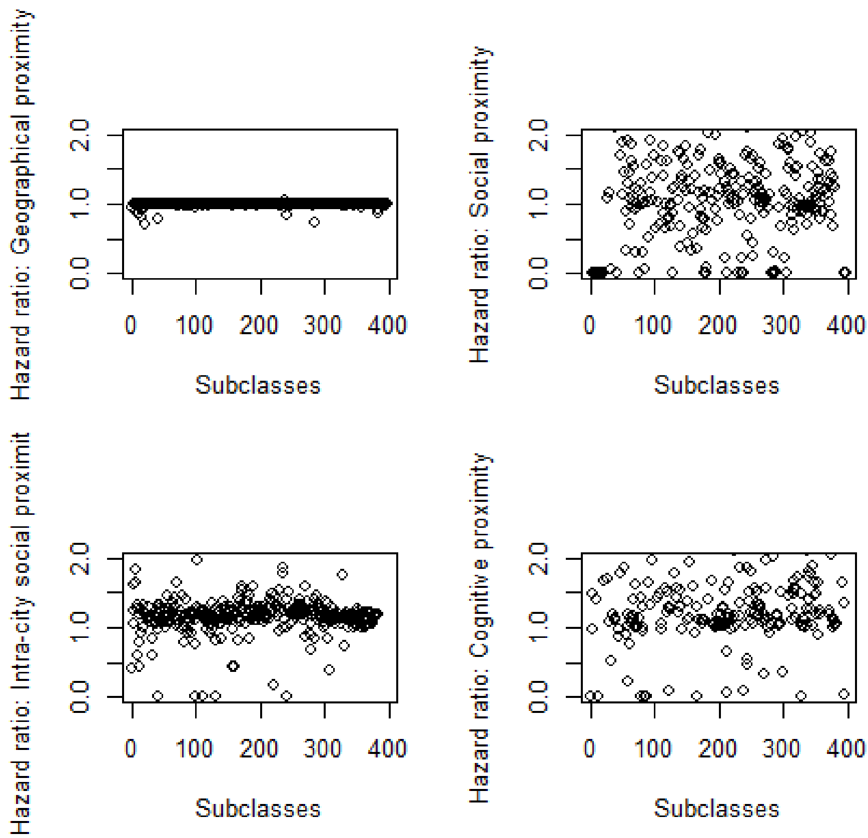

Fig. 5: Distribution of proximity hazard ratios across all 406 subclasses

It is informative to compare the results for the hazard ratios and related model coefficients for individual technological fields. Figure 5 plots the distribution of hazard ratios for all four employed metrics for the three proximity dimensions. A visual and brief statistical analysis reveals that the values of the conventionally measured social proximity are most dispersed, followed by cognitive proximity (mean = 4.89, standard deviation = 8.7), intra-city social proximity (mean = 9.39, standard deviation = 1.84), and geographical proximity (mean = 0.99, standard deviation = 0.03). Particularly the hazard ratios of the latter seem to be very homogenous. Figure 6 shows a significantly more homogeneous picture for the hazard ratios of the additional covariates for the average geographical distance, NSF grants, employment, population, and the age of citations, with the time-fixed geographical proximity measure of average spatial distance being the only exception. The reader may refer to the 'big-city-effect' described above for the interpretation of this independent variable's hazard ratios, with large cities with greater spatial distances between them disproportionately often adopting the technologies.

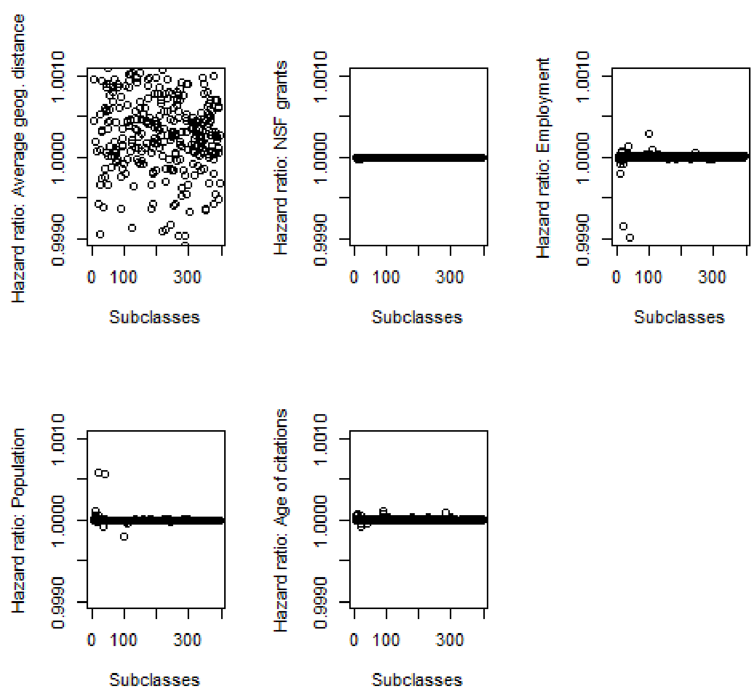

Fig. 6: Distribution of additional covariate hazard ratios across all 406 subclasses

In the following, individual technological fields with different configurations of their proximity hazard ratios of their spatiotemporal diffusion and adoption between 1975 and 2005 will be presented. Focusing on the crucial split between greater-than-one and less-than-one hazard ratios, Table 11 shows the distribution of individual subclasses above and below this threshold. To remind the reader, both social proximity metrics as well as the cognitive proximity statistics are expected have a hazard ratio greater than one, while geographical proximity is expected to have values less than one. As the preceding estimated models already suggested, these separate results reflect that the majority of subclasses have proximity hazard ratios with the respective expected sign. Still, the shares of hazard ratios with an unexpected hazard ratio range between 8.6% for geographical proximity and 29% for the conventionally measured social proximity.

	Geographical proximity	*Social proximity*	*Intra-city social proximity*	*Cognitive proximity*
≥1	34 (8.6%)	272 (71%)	341 (89%)	283 (89.3%)
<1	362 (91.4%)	111 (29%)	42 (11%)	34 (10.7%)
∑	396	383	383	317

Table 11: Distribution of subclass-based hazard ratios

To deepen this analysis of the Cox model estimations for single technological fields, Tables 12 and 13 provide an overview of possible hazard ratio combinations of the three key proximity metrics nearest-neighbor geographical proximity, intra-city social proximity, and cognitive proximity. The hazard ratios are categorized as either less than (*0*) or greater than (*1*) one. Both average-distance spatial proximity and the conventionally measured social proximity are not considered here due to their counter-intuitive results that could be attributed to their generality as well as to the significant variation of values of social proximity due to varying patent data availability in different technology classes, respectively. Eight different combinations of these three hazard ratios are possible as displayed in the table. The signs that theory suggests are highlighted. For example, the combination *011* means that the hazard ratios for a particular technological fields are less than one for geographical proximity, greater than one for intra-city social proximity and also greater than one for cognitive proximity. For 304 of all 406 subclasses, complete results have been reported for the model's estimation. The remaining 102 subclasses have partially missing values for their hazard ratios that are due to the low number of patents issued in certain technologies and cities, which were insufficient to estimate the Cox model. Interestingly, 225 of these 304 technologies (74%) with complete results report a combination of the expected signs for the three proximities. The frequency of hazard ratio combinations for these outliers is given below. A group of these outliers are due to limited numbers of patents in these fields and cities that adopted these technologies.

Hazard Ratio Combination	Geographical Proximity	Intra-City Social Proximity	Cognitive Proximity	Count (N = 304)
000	<1	<1	<1	6
001	<1	<1	>1	22
010	<1	>1	<1	23
011	<1	>1	>1	225
100	>1	<1	<1	1
101	>1	<1	>1	3

110	>1	>1	<1	4
111	>1	>1	>1	20

Table 12: Distribution of subclass-based hazard ratios. 0 in the hazard ratio combination stands for a less-than-one and 1 for a greater-than-one hazard ratio. Highlighted cells: Hazard ratios as expected from theory.

Hazard Ratio Combination	1 - Chemical	2 – Computers & Communication	3 – Drugs & Medical	4 – Electrical & Electronic	5 - Mechanical	6 - Others	%
000	1	2	0	0	2	1	2.0
001	2	4	4	4	7	1	7.2
010	2	3	4	5	4	5	7.6
011[1]	36	79	42	43	18	7	74.0
100	0	0	0	0	0	1	0.3
101	0	2	1	0	0	0	1.0
110	1	1	0	1	1	0	1.3
111	5	6	5	3	1	0	6.6
%	15.5	31.9	18.4	18.4	10.9	4.9	**N = 304**

Table 13: Distribution of subclass-based hazard ratios across Hall's (2001) six classes. 0 in the hazard ratio combination stands for a less-than-one and 1 for a greater-than-one hazard ratio.[1]: Combination of expected signs for the three proximity hazard ratios (Code 011).

8. Discussion

This study of novelty and the spatial diffusion and adoption of breakthrough inventions in the United States between 1975 and 2005 has produced three main findings. First, the results show that certain factors, theoretically and previously empirically found to condition the spatial diffusion and adoption of certain technologies, are equally relevant across different industries and breakthrough technologies, different time frames, and cities with differing patenting activity. This particularly emphasizes the importance of the three examined dimensions of proximity – geographical, social, and cognitive – in influencing the likelihood of technology adoption in a particular city.

Second, through estimating the Extended Cox Hazard Model for incrementally smaller subsets of the patent data set and peaking at subclass-based model runs, more complex dynamics are found. Although the vast majority of technological areas exhibit similar influences through the proximity measures and additional covariates, a few outliers are identified. Particularly the latter ones were simply averaged out when executing the model at more aggregate data levels. A group of these outliers are due to limited numbers of patents in these fields and cities that adopted these technologies. It may be an interesting additional research project to provide case studies of these outliers to find out what the reasons for these (partially) unexpected proximity values are.

Third, this study affirms the accuracy of the institutionally defined United States Patent Classification by examining the diffusion of groups of related technologies. The definition for relatedness is solely derived from its proximity in the classification's structure and hierarchy, as demonstrated in Chapter 4. Still, this institutionally defined relatedness produces consistent results in the Extended Cox Hazard Model, which are largely consistent with theory. Here, all subclasses have been aggregated to their respective 'parent' subclass at the first tier of the hierarchy. As mentioned above, only first-tier subclasses introduced between 1975 and 2005 were considered in this thesis.

This thesis contributes to the literature by examining the spatial diffusion of a large group of technological fields that have recently emerged and have not been studied yet. The analysis is run at different levels of aggregation and clustering of technologies to compare and contrast the influence of the dimensions of proximity and other covariates.

There are different potential extensions to the study presented here. It would certainly be interesting to link these technology diffusion models to citation data in order to compare different notions of technological trajectories and relatedness. Further, the existing model could be extended by incorporating other covariates that focus on certain dimensions of the technologies. For instance, different measures for the ‘growth’ or ‘branching’ of a first-tier subclass such as the number of ‘downstream’ subclasses could be used. In addition, more computational power would allow more precise calculation of proximities with more fine-grained resolutions than at the primary class level, which was partially used. Also, with available data in the future, the analysis could be extended to years prior to 1975 and after 2005. Another related research question could be how the conditions for spatial diffusion differ between breakthrough and non-breakthrough technologies.

There are different dimensions in which this study and its results have to be reflected critically. First, there exists a potential right-censoring issue in the identification of first instances which has been based on grant data and the additional criterion of relatively high citation counts. This is due to the fact that earlier patents had more time to being cited. Second, Kerr’s (2010) critique on the risk of upward bias of the model coefficients through unmodeled factors needs to be reinstated here. These could lead to higher shares of breakthroughs and generally higher numbers of patents. Further, if citations are localized (as demonstrated, for instance, in Jaffe et al., 1993 and Sonn and Storper, 2008), breakthroughs could disproportionately be identified in cities with unusually high patent growth. Another critique that extends an argument made above is the generality of the cognitive and social proximity measures. It can be argued that their broad focus on primary classes is not necessarily representative for proximity characteristics at the subclass level. Particularly regarding social proximity, an individual-based co-inventor network could have been constructed and used for the calculation of proximity values (cf. Breschi and Lissoni, 2009; Feldman et al., 2015) but proved to be too complex for the 406 first-tier subclasses examined in this study.

9. References

Acemoglu, D., Akcigit, U., Bloom, N., & Kerr, W. R. (2013). Innovation, reallocation and growth (No. w18993). *National Bureau of Economic Research*.

Acs, Z. J., & Audretsch, D. B. (1990). The determinants of small-firm growth in US manufacturing. *Applied Economics*, 22(2), 143-153.

Agrawal, A., Kapur, D., & McHale, J. (2008). How do spatial and social proximity influence knowledge flows? Evidence from patent data. *Journal of Urban Economics*, 64(2), 258-269.

Akcigit, U., Kerr, W. R., & Nicholas, T. (2013). The Mechanics of Endogenous Innovation and Growth: Evidence from Historical US Patents. Technical report, Working Paper.

Alderman, N., & Davies, S. (1990). Modelling regional patterns of innovation diffusion in the UK metalworking industries. *Regional Studies*, 24(6), 513-528.

Aldieri, L., & Cincera, M. (2009). Geographic and technological R&D spillovers within the triad: Micro evidence from US patents. *The Journal of Technology Transfer*, 34(2), 196-211.

Almeida, P. (1996). Knowledge sourcing by foreign multinationals: Patent citation analysis in the US semiconductor industry. *Strategic management journal*, 17(S2), 155-165.

Anderson, P., & Tushman, M. L. (1990). Technological discontinuities and dominant designs: A cyclical model of technological change. *Administrative science quarterly*, 604-633.

Arthur, W. B. (2009). The nature of technology: What it is and how it evolves.

Arzaghi, M., & Henderson, J. V. (2008). Networking off Madison Avenue. *The Review of Economic Studies*, 75(4), 1011-1038.

Attewell, P. (1992). Technology diffusion and organizational learning: The case of business computing. *Organization Science*, 3(1), 1-19.

Banerjee, P. M., & Cole, B. M. (2011). Globally radical technologies and locally radical technologies: the role of audiences in the construction of innovative impact in biotechnology. *IEEE Transactions on Engineering Management*, 58(2), 262-274.

Baptista, R. (1999). The diffusion of process innovations: A selective review. *International Journal of the Economics of Business*, 6(1), 107-129.

Barbieri, N. (2015). Technological trajectories and environmental policy: the transformation of the automobile (Doctoral dissertation).

Bathelt, H., Malmberg, A., & Maskell, P. (2004). Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography*, 28(1), 31-56.

Blossfeld, H. P., Golsch, K., & Rohwer, G. (2012). Event history analysis with Stata. Psychology Press.

Boschma, R. (2005). Proximity and innovation: a critical assessment. *Regional Studies*, 39(1), 61-74.

Box-Steffensmeier, J. M., & Jones, B. S. (2004). Event history modeling: A guide for social scientists. Cambridge University Press.

Breschi, S., & Lissoni, F. (2001). Knowledge spillovers and local innovation systems: a critical survey. *Industrial and Corporate Change* 10, 975–1005.

Breschi, S., & Lissoni, F. (2009). Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *Journal of Economic Geography*.

Brown, L. (1981). Innovation diffusion; a new perspective.

Burt, R. S. (1987). Social contagion and innovation: Cohesion versus structural equivalence. *American Journal of Sociology*, 92(6), 1287-1335.

Castaldi, C., Frenken, K., & Los, B. (2015). Related variety, unrelated variety and technological breakthroughs: an analysis of US state-level patenting. *Regional Studies*, 49(5), 767-781.

Chandy, R. K., & Tellis, G. J. (1998). Organizing for radical product innovation: The overlooked role of willingness to cannibalize. *Journal of Marketing Research*, 474-487.

Cliff, A. D., Haggett, P., Ord, J. K., and Versey, G. R. (1981). Spatial Diffusion: An Historical Geography of Epidemics in an Island Community, Cambridge University Press, Cmbridge.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. Administrative science quarterly, 128-152.

Cohen, J., & Tita, G. (1999). Diffusion in homicide: Exploring a general method for detecting spatial diffusion processes. *Journal of Quantitative Criminology*, 15(4), 451-493.

Dahlin, K. B., & Behrens, D. M. (2005). When is an invention really radical?: Defining and measuring technological radicalness. *Research Policy*, 34(5), 717-737.

Deshpande, R., Brown, L. A., Rogers, E. M., & Sahal, D. (1983). A comparative review of innovation diffusion books.

Dewar, R. D., & Dutton, J. E. (1986). The adoption of radical and incremental innovations: An empirical analysis. *Management Science*, 32(11), 1422-1433.

Dosi, G. (1982). Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Research Policy*, 11(3), 147-162.

Feldman, M. P., Kogler, D. F., & Rigby, D. L. (2015). rKnowledge: the spatial diffusion and adoption of rDNA methods. *Regional Studies*, 49(5), 798-817.

Fleming, L. (2001). Recombinant uncertainty in technological search. *Manag. Sci*. 47, 117–132.

Fleming, L. (2007). Breakthroughs and the" long tail" of innovation. *MIT Sloan Management Review*, 49(1), 69.

Geroski, P. A. (2000). Models of technology diffusion. *Research Policy*, 29(4), 603-625.

Gilly, J., & Torre, A. (2000). Proximity relations. Elements for an analytical framework. In: Green, M. and McNaughton, R. (Eds): *Industrial Networks and Proximity*, 1-16, Ashgate.

Gould, P. R. (1969). Spatial Diffusion, Resource Paper No. 4.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 1360-1380.

Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change. *Econometrica*, Journal of the Econometric Society, 501-522.

Griliches, Z. (1990). Patent statistics as economic indicators: a survey (No. w3301). *National Bureau of Economic Research.*

Hagerstrand, T. (1953). Spatial diffusion as an Innovation Process. Lund, University of Lund.

Hagerstrand, T. (1968). Innovation diffusion as a spatial process.

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001). The NBER patent citation data file: Lessons, insights and methodological tools (No. w8498). *National Bureau of Economic Research*.

Hall, B. H., Jaffe, A., & Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of Economics*, 16-38.

Harris, C. G., Arens, R., & Srinivasan, P. (2010). Comparison of IPC and USPC classification systems in patent prior art searches. In *Proceedings of the 3rd international workshop on Patent Information Retrieval* (pp. 27-32). ACM.

Henderson, R. (1993). Underinvestment and incompetence as responses to radical innovation: Evidence from the photolithographic alignment equipment industry. *The RAND Journal of Economics*, 248-270.

Hill, C. W., & Rothaermel, F. T. (2003). The performance of incumbent firms in the face of radical technological innovation. *Academy of Management Review*, 28(2), 257-274.

Hippel, E. von (1988). The Sources of Innovation. New York: Oxford University Press.

Hummon, N. P., & Dereian, P. (1989). Connectivity in a citation network: The development of DNA theory. *Social Networks*, 11(1), 39-63.

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

Jaffe, A., & Trajtenberg, M. (Eds) (2002). Patents, Citations, and Innovations: A Window on the Knowledge Economy. MIT Press, Cambridge, MA.

Kaplan, S., & Vakili, K. (2015). The double-edged sword of recombination in breakthrough innovation. *Strategic Management Journal*, 36(10), 1435-1457.

Keller, W. (2002). Geographic localization of international technology diffusion. *The American Economic Review*, 92(1), 120-142.

Kerr, W. R. (2010). Breakthrough inventions and migrating clusters of innovation. *Journal of Urban Economics*, 67(1), 46-60.

Kleinbaum, D. G., & Klein, M. (2006). Survival analysis: a self-learning text. Springer Science & Business Media.

Kuhn, T. S. (1962). The Structure of Scientific Revolutions. University of Chicago Press.

Lagendijk A., & Oinas P. (2005). Proximity, external relations, and local economic development. In: Lagendijk A, Oinas P, editors. *Proximity, Distance and Diversity, Issues on Economic Interaction and Local Development*. Burlington, VT: Ashgate; 2005. p. 3-22.

Lamoreaux, N., & Sokoloff, K. (1996). Long-term change in the organization of inventive activity, *Proceedings of the National Academy of Sciences*, 1993, 12686–12692.

Lee, P. C., Su, H. N., & Wu, F. S. (2010). Quantitative mapping of patented technology—The case of electrical conducting polymer nanocomposite. *Technological Forecasting and Social Change*, 77(3), 466-478.

Liu, J. S., Kuan, C. H., Cha, S. C., Chuang, W. L., Gau, G. J., & Jeng, J. Y. (2011). Photovoltaic technology development: A perspective from patent growth analysis. *Solar Energy Materials and Solar Cells*, 95(11), 3130-3136.

Lobo, J., & Strumsky, D. (2008). Metropolitan patenting, inventor agglomeration and social networks: A tale of two effects. *Journal of Urban Economics*, 63(3), 871-884.

Lundvall, B. Ä., & Johnson, B. (1994). The learning economy. *Journal of Industry Studies*, 1(2), 23-42.

MacGarvie, M. (2005). The determinants of international knowledge diffusion as measured by patent citations. *Economics Letters*, 87(1), 121-126.

Mansfield, E. (1968). Industrial research and technological innovation; an econometric analysis.

Mills, M. (2011). Introducing survival and event history analysis. Sage Publications.

Moreno, R., Paci, R., & Usai, S. (2005). Spatial spillovers and innovation activity in European regions. *Environment and Planning A*, 37(10), 1793-1812.

Nahm, A. Y., Vonderembse, M. A., & Koufteros, X. A. (2003). The impact of organizational structure on time-based manufacturing and plant performance. *Journal of Operations Management*, 21(3), 281-306.

Nelson, R. R., & Winter, S. G. (1982). An evolutionary theory of economic change.

Nooteboom, B. (2000). Learning and innovation in organizations and economies. OUP Oxford.

Pavitt, K. (1985). Patent statistics as indicators of innovative activities: possibilities and problems. *Scientometrics*, 7(1-2), 77-99.

Pavitt, K., Robson, M., & Townsend, J. (1987). The size distribution of innovating firms in the UK: 1945-1983. *The Journal of Industrial Economics*, 297-316.

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.

Rogers, E. (1983). The Diffusion of Innovation, (Third Ed.), New York: Free Press.

Rosenberg, N. (1994). Science, Invention, and Economic Growth. *Economic Journal* 84, 90-108.

Ruffaldi, E., Sani, E., & Bergamasco, M. (2010). Visualizing perspectives and trends in robotics based on patent mining. *In Robotics and Automation (ICRA), 2010 IEEE International Conference* (pp. 4340-4347).

Schoenmakers, W., & Duysters, G. (2010). The technological origins of radical inventions. *Research Policy*, 39, 1051–1059.

Shane, S. (2001). Technological opportunities and new firm creation. *Management Science*, 47(2), 205-220.

Silverberg, G. (1991). Adoption and diffusion of technology as a collective evolutionary process, in Nakienovic, N. and Grubler, A. (Eds): Diffusion of Technologies and Social Behavior, pp. 209–229. IIASA/Springer, Berlin.

Singh, J., & Fleming, L. (2010). Lone inventors as sources of breakthroughs: Myth or reality?. *Management Science*, 56(1), 41-56.

Sonn, J. W., & Storper, M. (2008). The increasing importance of geographical proximity in knowledge production: an analysis of US patent citations, 1975–1997. *Environment and Planning A*, 40(5), 1020-1039.

Sorenson, O. (2005). Social networks and industrial geography. In *Entrepreneurships, the New Economy and Public Policy* (55-69).

Su, H. N., & Lee, P. C. (2009). Dynamic and quantitative exploration on technology evolution mechanism: The case of electrical conducting polymer nanocomposite. *PICMET'09-2009 Portland International Conference on Management of Engineering & Technology* (pp. 2433-2440). IEEE.

Strumsky, D., Lobo, J., & Van der Leeuw, S. (2012). Using patent technology codes to study technological change. *Economics of Innovation and New Technology*, 21(3), 267-286.

Strumsky, D., & Lobo, J. (2015). Identifying the sources of technological novelty in the process of invention. *Research Policy*, 44(8), 1445-1461.

Therneau, T. M., & Grambsch, P. M. (2000). Modeling survival data: extending the Cox model. *Springer Science & Business Media*.

Trajtenberg, M. (1990). A penny for your quotes: patent citations and the value of innovations. *The Rand Journal of Economics*, 172-187.

Trajtenberg, M., Henderson, R., & Jaffe, A. (1997). University versus corporate patents: A window on the basicness of invention. *Economics of Innovation and New Technology*, 5(1), 19-50.

Verhoeven, D., Bakker, J., & Veugelers, R. (2016). Measuring technological novelty with patent-based indicators. *Research Policy*, 45(3), 707-723.

Verspagen, B. (2007). Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research. *Advances in Complex Systems*, 10(01), 93-115.

Wang, C., Sung, H. & Huang, M. (2016). Technological evolution seen from the USPC reclassifications. *Scientometrics* 107: 537.

Youn, H., Strumsky, D., Bettencourt, L. M., & Lobo, J. (2015). Invention as a combinatorial process: evidence from US patents. *Journal of The Royal Society Interface*, 12(106), 20150272.

Zhou, K. Z., Yim, C. K., & Tse, D. K. (2005). The effects of strategic orientations on technology- and market-based breakthrough innovations. *Journal of Marketing*, 69(2), 42-60.