Title
Exploring the Changing Structures of Inventor Collaboration in U.S. Cities between 1836 and 1975

Author
van der Wouden, Frank

Publication Date
2018

UNIVERSITY OF CALIFORNIA

Los Angeles

Exploring the Changing Structures of Inventor Collaboration in U.S. Cities between 1836 and 1975

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Geography

by

Frank van der Wouden

2018

ABSTRACT OF THE DISSERTATION

Exploring the Changing Structures of

Inventor Collaboration in U.S. Cities

between 1836 and 1975

by

Frank van der Wouden

Doctor of Philosophy in Geography

University of California, Los Angeles, 2018

Professor David L Rigby, Chair

ABSTRACT

The production of novel knowledge is seen as a key driver of economic development. However, knowledge production is unevenly distributed across space, giving rise to patterns of regional disparities in economic fortunes. Recent empirical evidence has shown that the production of knowledge is increasingly being dominated by collaboration and teamwork. This is explained by the rising complexity of knowledge, the production of which demands resources that exceed the capacity of individuals. The interactions of collaboration can provide a platform over which resources can flow between regions, boosting opportunities for knowledge sharing and learning. Yet, to date, there is little long-run, systematic, empirical evidence on the relationship between collaboration, geography and knowledge production before 1975.

This dissertation contributes to this gap by examining the structures of inventor collaboration in U.S. cities between 1836 and 1975. A new inventor-patent database is constructed that identifies all (co-)inventors and their geographical location(s) on more than 3 million patents granted by the USPTO during this time-period. The results of the analyses provide a number of new insights that aid the understanding of the role of inventor collaboration in knowledge production and regional economic development. This dissertation presents evidence on a significant positive relationship between the complexity of a patent and collaboration. Increasing complexity is associated with local collaboration. Geographical distance negatively impacts the odds of collaboration, while having a first- or second-order relationship boosts these odds. Inventors are more likely to collaborate with individuals with similar knowledge portfolios, especially during times of crises. Inventors who moved across space or between firms are found to have greater future productivity. These findings can help policy makers and corporate executive design policy that fosters interactions that are firm and place-specific.

The dissertation of Frank van der Wouden is approved.

Eric Stewart Sheppard

Michael C. Storper

Jacob Gates Foster

David L. Rigby, Committee Chair

University of California, Los Angeles

2018

Contents

List of figures

Chapter 1. Co-inventors on historical US patents: Changing Patterns in Collaboration, Complexity and Geography

Chapter 2. What Mechanisms Structure Tie-Formation among U.S. Inventors: Empirical Evidence from U.S. Patents between 1836-1975

Chapter 3. Inventor Mobility and Productivity: A Long-Run Perspective

List of tables

Chapter 1. Co-inventors on historical US patents: Changing Patterns in Collaboration, Complexity and Geography

Chapter 2. What Mechanisms Structure Tie-Formation among U.S. Inventors: Empirical Evidence from U.S. Patents between 1836-1975

Chapter 3. Inventor Mobility and Productivity: A Long-Run Perspective

Acknowledgements

During my education I've been lucky to study with a number of very talented and enthusiastic educators and scholars. All of them have supported me throughout high school, undergraduate studies and graduate education. I'm very grateful for all their efforts and help. I'm most thankful for the unceasing support, insights and feedback (and espressos) of David Rigby. David brought me out to Los Angeles to work on my Master's thesis and introduced me to the wonders of quantitative analysis and rigorous analytical thinking. He is also the reason why I decided to come back to UCLA for the PhD and work on collaboration, geography and knowledge production.

I've been very fortunate to have been supported by an excellent dissertation committee. I'm grateful to Michael Storper, Eric Sheppard and Jacob Foster for their time, insights, feedback, listening ear and motivating words during the various stages of researching, constructing, coding, visualizing and writing about patenting, geography and inventor collaboration. Without their input and ideas this dissertation would have never been written. Thank you!

Next to this committee I've grateful for the advice and feedback from Dora Costa, Cesar Hidalgo, Ron Boschma, Andrea Morrison, Sergio Petralia, Pierre-Alexander Balland, Koen Frenken, Mathieu Steijn, Matté Hartog, Dario Diodato, Nicola Cortinovis and Ian Fellows.

During my time at UCLA I've worked with a group of outstanding graduate students. I am thankful for the contributions of Austin Madson, Matthew Cooper, Chris Esposito, Melissa

Haller, Christoph Rösler, Anthony Frigon, Dylan Connor, Chelsea Robinson, Mia Bennett, Colin Gleason, Tyler Harlan, Sara Hughes, Luis Alvarez León, Tom Narins, Scott Stephenson, Cameran Ashraf and many others at various stages of the dissertation.

Throughout my graduate studies I've received financial support from private and public institutions. Without this support it would have been considerably more difficult to do research in Los Angeles. I want to thank *UCLA Graduate Division* for employing me as a teaching fellow, allowing me to learn and grow as an educator, and for granting me the Graduate Summer Research Mentorship Award, the University Fellowship, and Dissertation Year Fellowship. I'm also grateful for the generous support of the *John Randolph Haynes and Dora Haynes Foundation* for granting me the Haynes Lindley Doctoral Dissertation Fellowship, the *Prins Bernhard Cultuurfonds* for naming me a Bernard Fellow, the *Fundatie van de Vrijvrouwe van Renswoude* for naming me a Fundatie van Renswoude Fellow and the *UCLA Center for Economic History* for providing a graduate student grant to me. Finally, I'm grateful to the *Institute for New Economic Thinking*, the *Summer Institute for Economic Geography*, the *AAG Economic Geography Specialty Group*, the *Young Economic Geographers Network* and the *Regional Studies Association* for travel funding.

I want to thank my family for listening to my long rants about patents, inventors and their interactions, but also for providing for me. Their continuous and generous support has allowed me to explore my interests, see the world and get to learn who I am. I'm particularly grateful to my father Klaas Gerard van der Wouden for taking hours and hours to help me design and code text-mining algorithms that have been the cornerstone of this dissertation. Since building data is an iterative process that takes multiple tries, my father had to sacrifice numerous days, evenings and weekends to help me out. Without him there would be no inventor-patent data.

Finally, I need to express my gratitude to my partner Mirjam for standing by my side, often from the other side of the world and in different time-zones. Thank you for listening to me, motivating me when things looked somber and discussing ideas when my head was flowing over with ideas. Your amazing smile and empowering words energize me.

Vita

Frank van der Wouden
fvanderwouden@ucla.edu

University of California, Los Angeles
1255 Bunche Hall
Los Angeles, CA 90095, USA

Education

University of California, Los Angeles

Ph.D. Candidate (2013-Present), Expected Graduation: October 2018

C.Phil - March 18th 2016

Dissertation: "Exploring the Changing Structures of Inventor Collaboration in U.S. cities between 1836 and 1975"

Committee: David Rigby, Michael Storper, Eric Sheppard and Jacob Foster

Utrecht University - Research M.Sc. in Economic Geography (Honors, 2012)

Thesis: "The role of Technological Relatedness in Economic Variety on Knowledge Production in U.S. Cities between 1975 - 2005" (A+)

Committee: Ron Boschma, David Rigby, Frank van Oort

University of California, Berkeley - Haas School of Business

Visiting student – Organizational Behavior (2010)

Utrecht University - B.Sc. in Human Geography and Urban Planning (Honors, 2010)

Thesis: "U.S. Companies in the polder: the effect of U.S. KIBS on spatial concentration and innovation in The Netherlands" (A+)

Teaching Experience

Teaching Fellow (2013-present) at University of California, Los Angeles

Globalization - prof. dr. Eric Sheppard

Regional development - prof. dr. David Rigby (x3)

Economic Geography - prof. dr. David Rigby

People & Earth Ecosystems - prof. dr. Kyle Cavanaugh

Global Economy - prof. dr. John Agnew (x2)

Junior Assistant Professor (September 2012 - September 2013) - Utrecht University

Lecturer: (1) Spatial Economics
(2) Firm and Region
(3) Economic Geography

Supervision Bachelor Theses: 15+ students in Economics & Geosciences faculties

Research Interests

Innovation	Social Network Analysis	Data Science
Inventor Networks	Big Data	Machine Learning
Economic History	Complex Systems	Geography of Innovation
Spatial Economics	Evolutionary Economics	Entrepreneurship
Regional Development	Collaboration	Dynamic Learning Processes

Publications & Working papers

Under Review:

Van der Wouden, F., *Co-inventors on US patents: Changing patterns in Collaboration, Complexity and Geography*

Van der Wouden, F. & D. Rigby, C*o-inventor Networks and Knowledge Production in Specialized and Diversified Cities*

Van Niftrik, C.H.B. & F. van der Wouden, *Pre-operative Screening Model to Predict Post-intracranial Tumor Surgery Complications using Machine Learning Techniques*

Van der Wouden, F. & D. Rigby, *Inventor Mobility and Productivity: A Long-Run Perspective*

Working Papers:

Van der Wouden, F., *What Mechanisms Structure Tie-Formation among U.S. Inventors: Empirical Evidence from U.S. Patents between 1836-1975*

Van der Wouden, F. & D. Rigby, C. Esposito and C. Rosler, *The Geography of Novel Knowledge*

Van der Wouden, F., *Using Machine Learning Techniques to Mine and Process Data in Social Sciences*

Policy Recommendations:

van Oort, F. & T. de Graaff, H. Koster, H. Olden & F. van der Wouden (2013), *The city as the Engine of Growth: Theoretical exploration, empirical analysis and policy recommendations for the Randstad*

Awards and Grants

2018 Travel Bursary - Institute for New Economic Thinking ($300)
2018 University Fellowship - UCLA Geography Department ($500)
2017 Graduate Student Grant - UCLA Center for Economic History ($3,000)
2017 Dissertation Year Fellowship - UCLA Graduate Division ($36,325)
2017 Outstanding Teaching Certificate - UCLA Geography Department
2017 Travel Bursary - Regional Studies Association (£215)
2016 Graduate Student Grant - UCLA Center for Economic History ($5,000)
2016 Travel Bursary - Young Economic Geographers Network (CHF 430)
2016 Travel Bursary - AAG Economic Geography Specialty Group ($250)
2016 Travel Bursary - Summer Institute for Economic Geography ($1,100)
2015 Haynes Lindley Doctoral Dissertation Fellowship - Haynes Foundation ($20,000)
2015 Graduate Summer Research Mentorship Award - UCLA Graduate Division ($6,000)
2013 PhD funding offer - UCLA Graduate Division ($170,000)
2013 University Fellowship - UCLA Graduate Division ($20,000)
2013 Bernhard Fellow - Prins Bernhard Cultuurfonds (€5,000)
2013 Fundatie van Renswoude Fellow - Fundatie van de Vrijvrouwe van Renswoude (€6,000)
2013 Marc de Smidt Prize - Best Master Thesis - Faculty GeoSciences (€500)
2013 Marc de Smidt Prize - Best Presentation - Faculty GeoSciences

Other

Reviewer: Journal of Economic Geography, Scientometrics, Research Policy
Affiliations: American Association of Geographers, Economic History Association, Regional Studies Association, California Center for Population Research
Languages: Dutch (Native), English (Fluent), German (Good), French (Intermediate)

Introduction

Understanding innovation is key to our understanding of economic development. Individuals provide valuable inputs and resources that are required to produce innovations. As the technological frontier is being pushed further into technologies of increased complexity, some have argued that it becomes increasingly difficult for these individuals to produce novelty and thus slows down rates of economic growth (Jones, 2009). This ‘burden of knowledge’ drives individuals to deepen their knowledge in specialized areas of expertise and the need to collaborate with others (Jones et al., 2008; Wuchty et al., 2007). This changing nature of innovation thus has the potential of negatively affecting long-run economic growth.

In an increasingly globalizing economy, the role of geography in economic processes is changing. While consumption and production have traditionally been strongly localized together, developments in telecommunication and transportation technologies have allowed for the spatial separation of production and consumption. Firms no longer have to be close to their customers to serve them. The spatial scales on which economic interactions take place have rapidly grown the last centuries, giving rise to global markets, global production chains and international competition. Some scholars argue that these trends signify the ‘death of distance’ (Ohmae, 1990) and that the world is becoming flat (Friedman, 2005).

On the other side, individuals and firms are found to be clustering in space. Urbanization rates are increasing in virtual all parts of the world (Ravallion et al., 2007; Wang et al., 2012). People move to cities for reasons that range from protection against violence to reuniting with family and access to services, but also for economic opportunities. Densely populated areas facilitate interactions between economic agents and the trading of services and goods, positively affecting the productivity of those located there. The spatial concentration of firms has long been recognized (Glaeser, 1999; Malmberg & Maskell, 2002; Marshall, 1920; Porter, 1990; Storper, 1992). The local presence of economic activity attracts economic agents that benefit from close geographical proximity to other agents, setting in motion forces of agglomeration (Duranton & Puga, 2004). The strength of these forces fluctuates over space due to the specific quality, quantity and relatedness of localized economic activities, as well as the institutionalized practices that help shape the economic fortunes of regions. Thus the role of geography in economic processes is perhaps best described as a paradox (Porter, 2000).

Economic production that is knowledge intensive tends to be more spatially concentrated than other types of economic processes (Audretsch & Feldman, 1996; Balland et al., 2018). This is especially relevant because the production of knowledge might lead to innovations that give rise to novel technological trajectories and new economic geographies. From a Schumpeterian perspective, innovations are at the forefront of the emergence of new industries that provide significant employment opportunities. While processes of globalization have spread employment to different locations around the world, knowledge-based production remains highly clustered in large part because of its dependence on agglomerations of highly-skilled workers (Cowan, David, & Foray, 2000; McInerney, 2002; Zollo & Winter, 2002). These high technology clusters have attracted considerable attention for their role in job generation and regional growth (Michael E Porter, 1990; Saxenian, 1994).

While the literature around agglomeration is extensive, it is only relatively recently that researchers have begun to focus on the micro-connections between economic agents within clusters (Gertler, 1995; Giuliani and Bell, 2005; Boschma and Ter Wal, 2007; Giuliani, 2007). Most of this work reports on snapshots of linkages that cover only a relatively short time-period. Hence, we have little information about the longer run development of collaboration within and between urban and regional concentrations of economic activity. In general, we know that economic actors rarely act in isolation. Rather, they interact with numerous partners – customers, suppliers, collaborators and competitors. These interactions connect agents building teams within firms and building inter-firm networks, both local and non-local, across which critical resources flow (Powell et al., 1996; Bathelt et al., 2004; Storper and Venables, 2004; Wuchty et al., 2007).

Recent work on spatial differences in knowledge production has explored how networks linking inventors influence patent productivity. Fleming et al. (2007) and Lobo and Strumsky (2008) explore how the social networks of inventors impact the rate of innovation. Whittington et al. (2009) and Breschi and Lenzi (2016) examined interactions between social and spatial relationships influencing invention. Cantner and Graf (2004) developed the connection between the technological specialization of regions and the structure of cooperation networks. Van der Wouden and Rigby (2018) extend this work by examining the link between the degree of specialization of U.S. cities and the structure of inventor collaboration networks.

Analysis of regional economic performance and of the interactions between economic agents within urban-regional clusters is typically restricted to relatively short time intervals because of the lack of long-run data. Identification of the emergence of regional clusters, of the linkages between actors, and the formation of the institutions in which they operate is difficult without detailed historical data, at least in part, because the legacy of past rounds of growth and development and inter-regional relationships often go unobserved. This is not to say that investigation of geographies of economic activity over shorter time scales are not useful, but merely to note that histories of growth and decline and path dependent development exert a long reach and shape future economic geographies in complex ways. One attraction of long-run historical data is that they offer researchers greater control over the influence of past events, even if only through larger numbers of observations.

Evolutionary economic geography is concerned with understanding the dynamics of geographical uneven development. Research in this sub-field would also benefit significantly from the availability of historical data. While Census records provide glimpses of the state of the U.S. space economy back for more than one hundred years, these data are relatively thin in terms of the number of variables available. Economists, geographers, historians and others seeking high quality historical data have looked increasingly to patent and inventor records for detailed information that might shed light on patterns and processes of regional uneven development.

Patent related research in the U.S. really took off in the 1980s with the development of the NBER digital patent database (Jaffe & Trajtenberg, 2002). Individual patent records provide detailed information on the nature of new technologies developed, about the timing of such developments, the inventors involved, ownership of the patents, and linkages between technologies traced through citations. In aggregate the patent data inform researchers about the volume of invention over time and through space, about the linkages between patents and inventors, the diffusion of knowledge and so much more (Feldman et al., 2014; Griliches, 1984; Hall et al. ,2001; Jaffe et al., 1993; Singh, 2005; Van der Wouden & Rigby, 2018). At the same time, patent data usually come with a series of "health warnings" to the researcher that they fail to capture all processes of knowledge discovery, that patents vary widely in terms of their importance, and that patent linkages, at least in the form of citations, are compromised (Griliches, 1990; Hall et al., 2001; Pavitt, 1985).

The widely available NBER digital patent data for the United States only dates back to 1975. Thus, most recent research on the geography and history of U.S. invention only dates back 40 years or so. However, newer work has constructed patent data back to 1836, and in some cases 1790 (Petralia, Balland, & Rigby, 2016; Ackcigit et al., 2017). While these efforts are important, they remain incomplete because they offer only a glimpse of the information that can be extracted from historical patent records. As a result, very little is known about the characteristics of U.S. knowledge production before 1975, apart from a small number of important surveys (see Lamoreaux & Sokoloff, 1996), thus obscuring our long-run understanding of the processes of invention and their role in structuring regional development.

This dissertation contributes to this gap by examining the relationships between regional knowledge production, collaboration, geography and mobility among metropolitan inventors on U.S. patents between 1836 and 1975. This period is of special interest to economic historians because it covers the 1870-1970 "Golden Age" of U.S. knowledge production and invention (Akcigit et al., 2017; Gordon, 2016). This period also covers much of the early history of U.S.

industrialization and so provides an important perspective on the emergence of an early economic geography.

This dissertation comprises three chapters and a new inventor-patent data-base. The data generated for this data-base draws on and extends the publicly available HistPat (Petralia et al., 2016) database. The HistPat database contains geographical information for historical patents provided by the USPTO between 1790 and 1975. Scholars from Utrecht University and UCLA scraped the text from digitized historical patent files available on Google Patents and EspaceNet, and recorded the first inventor and a geographical location. Unfortunately, HistPat provides no information for any possible additional co-inventors. I contribute to these data by identifying all inventors and their geographical locations for each USPTO patent between 1836 and 1975[1]. Using the raw scraped text files for the 4,125,734 patents, I examine whether each word in these text files is part of a first or family name, or a geographical location in the U.S. The data for the lists of names comes from inventor names in HistPat, the digital USPTO patents from 1975 up to 2005 (Li et al. 2011) and the U.S. Census. The data for the geographical locations come from the same sources as well as the U.S. Bureau of Economic Analysis.

After a (fuzzy) match between a word on the patent text with one of the lists with names and/or location occurs, a series of complex algorithms is run. These algorithms can broadly be placed in two groups. The first set of algorithms examines the words before and after the matched word to determine whether the matched word is a name or geographical location. Examining the text before and after the matched word helps to distinguish between name and

[1] I limit my analysis to 1836 because before 1836 the USPTO did not make use of patent examiners. It is generally accepted by economic historians that this institutional change has significantly impacted US patenting activity (Lamoreaux, Sokoloff, & Sutthiphisal, 2011; Sokoloff, 1988).

location. The second set of algorithms record a series of more than 30 statistics for each matched word. These statistics are used for the machine learning exercise, discussed below, to determine whether an observed name is truly an inventor and not the name of a witness, examiner, corporation or reference. Similar operations are undertaken for an observed geographical location that is linked to a name, but those statistics are only used to generate a likelihood measure of a correct name-location link.

The next step is to distinguish between inventors and non-inventors in the patent records. Non-inventor names can correspond to witnesses, attorneys, companies, references or other entities that have name-like characteristics. I use supervised machine learning techniques to classify each of the 8 million observed names as either an inventor (1) or a non-inventor (0). The details of this approach are outlined in the first chapter of this dissertation. The resulting inventor-patent database identifies 1,922,754 inventors, with 4,437,960 observations on 3,365,253 unique US patents between 1836 and 1975. For all but three years more than 80% of the patents are in our database. For a lot of years more than 90% of the patents are included. The wide coverage of the annual number of patents and lack of theoretical or empirical motivations to expect systematic bias in the unobserved patents suggests that this database is a representative sample of the historical USPTO patents granted between 1836 and 1975. This data-base is used as the main source of data in this dissertation, and it is put to use to examine three core questions:

- What are the relationships between complexity, collaboration and geography in knowledge production?
- What mechanisms structure tie-formation between U.S. inventors?

- What is the form of long-run patterns of firm and spatial mobility of U.S. inventors, and does mobility boost productivity?

The first chapter explores the relationships between collaboration, geography and complexity. Since 1950 the production of knowledge has become increasingly the product of collaboration (Wuchty et al., 2007). The reasons for scientific collaboration are widespread, ranging from resource optimization (Eaton, 1951), increased productivity (de Solla Price, 1986), access to ideas and resources (Wray, 2002), and to intellectual or social linkages (Thorsteinsdottir, 2000). The underlying argument in this field of study is that knowledge production is becoming increasingly complicated and requires inputs that exceed that of the individual. Large, complex projects cannot be undertaken by a single individual, especially when scientists and engineers are becoming increasingly specialized.

Knowledge is increasingly perceived as sets of embodied capabilities and routines constructed through processes of learning (Lundvall & Johnson, 1994; Zollo & Winter, 2002). The embodied nature of knowledge restricts its movement across space and between entities. This is especially the case for complex ideas resting heavily on non-ubiquitous forms of tacit knowledge (Kogut & Zander, 1992; Von Hippel, 1994). To diffuse complex, tacit knowledge repeated face-to-face interaction are required, emphasizing the need for co-location of economic agents (Singh, 2005; Sorenson, Rivkin, & Fleming, 2006; Storper & Venables, 2004). Complex knowledge tends to be limited to larger cities because its production rests heavily on face-to-face interaction (Balland et al., 2018).

Scholarly interests in collaboration and complexity rests upon the idea that both have a positive association with the value, quality or quantity of production. Collaborative work is found to boost creativity (Uzzi & Spiro, 2005), have higher acceptance rates for academic publication (Presser, 1980), receive more citations (Glanzel, 2002), raise productivity (Lee & Bozeman, 2005) and enables researchers to engage with large research questions (Thagard, 1997). Similarly, complex knowledge is seen as a crucial source of competitive advantage (Kogut & Zander, 1992; Maskell & Malmberg, 1999; Sorenson et al., 2006). Access to complex knowledge allows inventors to engage with more cutting-edge activities and capture the benefits that might arise from such engagement. These benefits may be monetary within firms and/or manifest as increased levels of economic development within society.

However, empirical evidence on the relationship between complexity and collaboration is limited. Long-run, large scale, systematic and historical evidence linking collaboration and complexity is lacking, primarily because empirical data on collaboration is absent. The research on collaboration tends to focus on relative short time-frames, specific fields or projects. The same is true for measures of knowledge complexity. Measures of complexity are also difficult to design, construct and operationalize over time and space.

The primary aim of this chapter is to examine patterns of co-invention on US patents along the axes of collaboration, complexity and geography. Using the new inventor-patent data-base to merge with information on the technology classes of U.S. patents, available from the USPTO, measures of complexity for each patent are constructed. The resulting data reveals the characteristics of U.S. inventor collaboration on patents of varying complexity, across geographies and at different moments in time. The key results of this study show that (1) collaboration has increased sharply since the 1930s; (2) there is a positive and significant relationship between complexity and collaboration on US patents; and (3) increasing complexity on patents promotes local, within-city collaboration.

The second chapter investigates the mechanisms that structure collaboration among US inventors. Most scholars agree that the change towards science-based knowledge production resulted in increased complexity of knowledge production. To produce novel knowledge, the increasing complexity requires inputs that exceed the resources of the individual. Some have labeled this the 'burden of knowledge' and argue it is increasingly becoming more difficult to produce innovations – especially for individuals (B F Jones, 2009). Empirical results show that knowledge production by scientist and inventors has become gradually more dominated by teams since the 1960s (Wuchty, Jones, & Uzzi, 2007).

There are numerous reasons for knowledge producers to collaborate. Knowledge producers might collaborate to optimize resources (Eaton, 1951), increase productivity (Lee & Bozeman, 2005; Melin, 2000; Price, 1986) or creativity (Uzzi & Spiro, 2005), access complementary ideas (de Solla Price, 1970) or resources (Katz & Martin, 1997; Wray, 2002), or for intellectual and social reasons (Edge, 1979; Stokes & Hartley, 1989; Thorsteinsdottir, 2000). Collaborative work tends to outperform work produced by single individuals. Collaborative work is more frequently cited citations than individual work (Fox, 1991; Frenken, Hölzl, & Vor, 2005; Glanzel, 2002; Katz & Martin, 1997; Lindsey, 1978), has higher acceptance rates for academic publication (Presser, 1980), and enables researchers to engage with large research questions

(Thagard, 1997). Faems et al. (2005) finds that collaborating firms are more likely to produce commercially successful products than non-collaborating firms. Katz & Martin (1997) provide a critical review of the literature on R&D collaboration.

Although these reasons to collaborate are well documented and understood, the understanding of tie-formation processes in collaborative knowledge production remains incomplete. To date, research on tie-formation among knowledge producers has largely focused on the firm-, region- or nation-state level, ignoring the level of the individual (a notable exception is Crescenzi, Nathan, & Rodríguez-Pose, 2016). In addition, the limited research on collaboration between individual knowledge producers has concentrated on the impact of individuals' attributes on tie-formation and ignoring relational, dyadic and triadic processes. As a result, little structural empirical evidence on the dyad-level mechanisms governing actual tie-formation among individual knowledge producers is available. This is surprising, because there is a vast social sciences literature that stresses the importance of dyad-level processes on tie-formation (Borgatti, Mehra, Brass, & Labianca, 2009; Oswald, Clark, & Kelly, 2004).

This chapter uses the new long-run inventor-patent data to examine the impact of three popular dyad-level mechanisms – geographical, social and technological distance – on tie-formation between individual inventors. Exponential random graph models (ERGMs) are used to estimate the impact of these three measures of proximity on the probability of a tie-formation between inventors and how this impact has evolved over time. This method has a main advantage over conventional statistical methods because ERGMs are designed to deal with the interdependence of observations in relational data and thus produces more accurate estimates on the dyad-level mechanisms modelled than conventional statistical models.

The results are the first long-run systematic empirical evidence on the mechanisms that foster and hinder tie-formation between inventors on US patents between 1836 and 1975. The key findings are that (1) geographical distance negatively impacts collaboration, that this effect has been decreasing over time but has levelled off in the 20th century; (2) technological proximity between inventors' patent portfolio positively impact collaboration, especially in times of uncertainty; and (3) social proximity promotes collaboration, but only to a certain extend. These results have important implications for the understanding of knowledge production because they stress the complexity of innovation dynamics

The third chapter examines whether mobility raises the productivity of inventors. The mobility of skilled workers within the economy has generated significant interest for decades. The main concern with mobility focuses on the impacts of the movement of skilled-workers across the economy. Scholars have captured the efficiency gains of worker sorting and matching, and separated these from the returns to agglomeration (Behrens, Duranton, & Robert-Nicoud, 2014; Jackson, 2012; Topel & Ward, 1992). In the innovation literature these gains are reflected as higher rates of inventor and firm productivity following movement (Hoisl, 2007; Kaiser et al., 2015).

The mobility of inventors has attracted attention for its role as drivers of knowledge flows and spillovers between firms and over space (Breschi & Lissoni, 2009; Zucker, Darby, &

Armstrong, 1998). At the regional level, the flow of knowledge through the movement of technology embodied in skilled workers is associated with high performing regions (Almeida & Kogut, 1999; Audretsch & Feldman, 1996; Saxenian, 1994). On the national scale, the movement of skilled workers is linked to inventive activities in emerging economies (Nathan, 2014; Saxenian, 2005, 2007).

The primary goals of this chapter are to explore long-run shifts in inventor mobility between firms and over space and to examine whether that mobility raises the inventor productivity. Most studies concerning the mobility of skilled-workers and inventors focus on specific growth sectors, regions or limited time-periods. There are no data comparing the rates of movement of inventors over the long-run, nor whether mobility raises the productivity of inventors. It is important to know whether rates of inventor mobility have significantly changed over time and thus whether the flow of knowledge may play as important a role in regional economic performance as many suggest.

To date, answering the above questions has been difficult because individual inventors in the United States have not been systematically identified and tracked. The new inventor-patent data-base constructed for this dissertation allows doing so. In addition, using search, match and machine learning techniques the assignee data for these historical patents is extended. This allows tracking individual inventors in time and across space *and* firms. Matching algorithms are used to match mobile to non-mobile inventors on a series of covariates, such that there is no connection or bias between the treatment variable (mobility) and the control variables. These matched samples are used in statistical models examining the effect of firm and spatial mobility on the future number of patents produced by inventors.

The key findings are that firm mobility and spatial mobility raise the future productivity of inventors. In general, inventors moving in space tend to have greater patent production over the following five years than their non-mobile counterparts. This effect is strongest for the last 30 years of the sample. Similarly, inventors moving between firms produced more patents than inventors who stayed with the same firm. This positive effect on productivity is observed for all time-periods in the sample, except for the aftermath of the Second World War. Inventors who moved both between firms and locations have significantly higher patent productivity than their counterparts who moved only between firms or locations. Inventors who moved between firms have greater productivity gains compared to their control group than the inventors who moved geographically compared to their control group. This suggests that firm mobility impacts inventor productivity more than spatial mobility.

The remainder of this dissertation is organized as follows. The next chapter presents the research on the relationships between complexity, collaboration and geography. The third chapter examines the mechanisms that structure patterns of collaboration among US inventors. The fourth chapter focuses on the relationship between inventor mobility and productivity. The last chapter presents a series of conclusions and discussion that points to future research possibilities.

Literature

Akcigit, U., Grigsby, J., & Nicholas, T. (2017). The Rise of American Ingenuity: Innovation and Inventors of the Golden Age.

Almeida, P., & Kogut, B. (1999). Localization of knowledge and the mobility of engineers in regional networks. *Management Science*, *45*(7), 905–917.

Audretsch, D. B., & Feldman, M. P. (1996). R & D Spillovers and the Geography of Innovation and Production. *The American Economic Review*, *86*(3), 630–640.

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

Behrens, K., Duranton, G., & Robert-Nicoud, F. (2014). Productive cities: Sorting, selection, and agglomeration. *Journal of Political Economy*, *122*(3), 507–553.

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network Analysis in the Social Sciences. *Science*, *323*(5916), 892 LP-895.

Breschi, S., & Lenzi, C. (2016). Co-invention networks and inventive productivity in US cities. *Journal of Urban Economics*, *92*, 66–75. http://doi.org/10.1016/j.jue.2015.12.003

Breschi, S., & Lissoni, F. (2009). Mobility of skilled workers and co-invention networks: An anatomy of localized knowledge flows. *Journal of Economic Geography*, *9*(4), 439–468.

Cantner, U., & Graf, H. (2004). Cooperation and specialization in German technology regions. *Journal of Evolutionary Economics*, *14*(5), 543–562.

Cowan, R., David, P. A., & Foray, D. (2000). The explicit economics of knowledge codification and tacitness. *Industrial and Corporate Change*, *9*(2), 211–253.

Crescenzi, R., Nathan, M., & Rodríguez-Pose, A. (2016). Do inventors talk to strangers? on proximity and collaborative knowledge creation. *Research Policy*, *45*(1), 177–194.

de Solla Price, D. J. (1970). Citation measures of hard science, soft science, technology, and nonscience. In *Communication Among Scientists and Engineers* (pp. 3–22). Lexington Books, D.C. Heath and Company.

Duranton, G., & Puga, D. (2004). Micro-foundations of urban agglomeration economies. In *Handbook of regional and urban economics* (Vol. 4, pp. 2063–2117). Elsevier.

Eaton, J. W. (1951). Social Processes of Professional Teamwork. *American Sociological Review*, *16*(5), 707–713.

Edge, D. (1979). Quantitative measures of communication in science: A critical review. *History of Science*, *17*(2), 102–134.

Faems, D., Van Looy, B., & Debackere, K. (2005). Interorganizational Collaboration and Innovation: Toward a Portfolio Approach. *Journal of Product Innovation Management*, *22*, 238–250.

Feldman, M. P., Kogler, D. F., & Rigby, D. L. (2014). rKnowledge : The Spatial Diffusion and Adoption of rDNA Methods. *Regional Studies*, 1–21

Fleming, L., King, C., & Juda, A. (2007). Small Worlds and Regional Innovation. *Organization Science*, *18*(6), 938–954.

Fox, M. F. (1991). Gender, environmental milieu, and productivity in science. In H. Zuckerman, J. Cole, & J. Bruer (Eds.), *The outer circle* (pp. 188–204). WW Norton & Co, New York.

Frenken, K., Hölzl, W., & Vor, F. de. (2005). The citation impact of research collaborations: the case of European biotechnology and applied microbiology (1988–2002). *Journal of Engineering and Technology Management*, *22*(1–2), 9–30.

Friedman, T. (2005). The world is flat. *New York: Farrar, Straus and Giroux*, 488.

Glaeser, E. L. (1999). Learning in Cities. *Journal of Urban Economics*, *46*(2), 254–277.

Glanzel, W. (2002). Coauthorship patterns and trends in the sciences (1980-1998): A bibliometric study with implications for database indexing and search strategies. *Library Trends*, *50*(3), 461–473.

Gordon, R. J. (2016). *The rise and fall of American growth: The US standard of living since the civil war*. Princeton University Press.

Griliches, Z. (1984). *R&D, Patents and Productivity*. Chicago: The University of Chicago Press.

Griliches, Z. (1990). Patent Statistics as Economic Indicators: A Survey. *Journal of Economic Literature*, (28), 1661–1707.

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001). The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools.

Jackson, C. K. (2012). Match Quality, Worker Productivity, and Worker Mobility: Direct Evidence from Teachers. *The Review of Economics and Statistics*, *95*(4), 1096–1116.

Jaffe, A. B., & Trajtenberg, M. (2002). *Patents, citations, and innovations: A window on the knowledge economy*. MIT press.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics*, *108*(3), 577–598.

Jones, B. F. (2009). The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder? *Review of Economic Studies*, *76*(1), 283–317.

Jones, B. F. (2009). The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder? *The Review of Economic Studies*, *76*(1), 283–317.

Jones, B. F., Wuchty, S., & Uzzi, B. (2008). Multi-University Research Teams: Shifting Impact, Geography, and Stratification in Science. *Science*, *322*(5905), 1259–1262.

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, *26*(1), 1–

18.

Kogut, B., & Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science*, *3*(3), 383–397.

Lamoreaux, N. R., & Sokoloff, K. L. (1996). Long-term change in the organization of inventive activity. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(November), 12686–12692.

Lamoreaux, N. R., Sokoloff, K. L., & Sutthiphisal, D. (2011). The Reorganization of Inventive Activity in the United States during the Early Twentieth Century. In *Understanding Long-Run Economic Growth: Geography, Institutions, and the Knowledge Economy* (pp. 235–274).

Lee, S., & Bozeman, B. (2005). The Impact of Research Collaboration on Scientific Productivity. *Social Studies of Science*, *35*(5), 673–702.

Lindsey, D. (1978). *The scientific publication system in social science*. Jossey-Bass Inc Pub.

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

Malmberg, A., & Maskell, P. (2002). The elusive concept of localization economies: Towards a knowledge-based theory of spatial clustering. *Environment and Planning A*, *34*(3), 429–449.

Marshall, A. (1920). *Principles of Economics*. London: Macmillan and Co., Ltd.

Maskell, P., & Malmberg, a. (1999). The Competitiveness of Firms and Regions: “Ubiquitification” and the Importance of Localized Learning. *European Urban and*

Regional Studies, *6*(199901), 9–25.

McInerney, C. (2002). Knowledge management and the dynamic nature of knowledge. *Journal of the American Society for Information Science and Technology*, *53*(12), 1009–1018.

Melin, G. (2000). Pragmatism and self-organization: Research collaboration on the individual level. *Research Policy*, *29*(1), 31–40.

Nathan, M. (2014). The wider economic impacts of high-skilled migrants: a survey of the literature for receiving countries. *IZA Journal of Migration*, *3*(1), 4.

Ohmae, K. (1990). The borderless world. *McKinsey Quarterly*, (3), 3–19.

Oswald, D. L., Clark, E. M., & Kelly, C. M. (2004). Friendship Maintenance: An Analysis of Individual and Dyad Behaviors. *Journal of Social and Clinical Psychology*, *23*(3), 413–441.

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

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*.

Porter, M. E. (1990). The Competitive Advantage of Nations. *Harvard Business Review*, *68*(2), 73–93.

Porter, M. E. (2000). Location, Competition, and Economic Development: Local Clusters in a Global Economy. *Economic Development Quarterly*, *14*(1), 15–34.

Presser, S. (1980). Collaboration and the quality of research. *Social Studies of Science*, *10*(1), 95–101.

Price, D. J. de S. (1986). Little Science, Big Science and Beyond. Columbia University Press.

Ravallion, M., Chen, S., & Sangraula, P. (2007). New evidence on the urbanization of global poverty. *Population and Development Review*, *33*(4), 667–701.

Saxenian, A. (1994). *Regional advantage: culture and competition in Silicon Valley and route 128*. Cambridge, MA: Harvard University Press.

Saxenian, A. (2005). From brain drain to brain circulation: Transnational communities and regional upgrading in India and China. *Studies in Comparative International Development*, *40*(2), 35–61.

Saxenian, A. (2007). *The new argonauts: Regional advantage in a global economy*. Harvard University Press.

Singh, J. (2005). Collaborative Networks as Determinants of Knowledge Diffusion Patterns. *Management Science*, *51*(5), 756–770.

Sokoloff, K. L. (1988). Inventive Activity in Early Industrial America: Evidence From Patent Records, 1790–1846. *The Journal of Economic History*, *48*(04), 813.

Sorenson, O., Rivkin, J. W., & Fleming, L. (2006). Complexity, networks and knowledge flow. *Research Policy*, *35*(7), 994–1017.

Stokes, T. D., & Hartley, J. A. (1989). Coauthorship, social structure and influence within specialties. *Social Studies of Science*, *19*(1), 101–125.

Storper, M. (1992). The Limits to Globalization : Technology Districts and International Trade. *Economic Geography*, *68*(1), 60–93.

Storper, M., & Venables, A. J. (2004). Buzz: Face-to-face contact and the urban economy. *Journal of Economic Geography*, *4*(4), 351–370.

Thagard, P. (1997). Collaborative knowledge. *Noûs*, *31*(2), 242–261.

Thorsteinsdottir, O. (2000). External research collaboration in two small science systems. *Scientometrics*, *49*(1), 145–160.

Topel, R. H., & Ward, M. P. (1992). Job mobility and the careers of young men. *The Quarterly*

Journal of Economics, *107*(2), 439–479.

Uzzi, B., & Spiro, J. (2005). Collaboration and Creativity: The Small World Problem. *American Journal of Sociology*, *111*(2), 447–504.

Van der Wouden, F., & Rigby, D. L. (2018). *Inventor collaboration networks and different levels of specialization in U.S. cities*. Working Paper

Von Hippel, E. (1994). “Sticky information” and the locus of problem solving: implications for innovation. *Management Science*, *40*(4), 429–439.

Wang, H., He, Q., Liu, X., Zhuang, Y., & Hong, S. (2012). Global urbanization research from 1991 to 2009: A systematic research review. *Landscape and Urban Planning*, *104*(3–4), 299–309.

Whittington, K. B., Owen-smith, J., & Powell, W. W. (2009). Kjersten Bunker Whittington. *Administrative Science Quarterly*, *54*(1), 90–122.

Wray, K. B. (2002). The Epistemic Significance of Collaborative Research. *Philosophy of Science*, *69*(1), 150–168.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036–9.

Zollo, M., & Winter, S. G. (2002). Deliberate learning and the evolution of dynamic capabilities. *Organization Science*, *13*(3), 339–351.

Zucker, L. G., Darby, M. R., & Armstrong, J. (1998). Geographically Localized Knowledge: Spillovers or Markets?, *XXXVI*(January).

Chapter 1: Co-inventors on historical US patents: Changing Patterns in Collaboration, Complexity and Geography

Abstract

This paper examines co-invention on U.S. patents between 1836 and 1975. The patent data show that the production of knowledge is increasingly the result of collaboration. This is explained by the rising complexity of knowledge, the production of which demands resources that exceed the capacity of individual specialized inventors. To date, there is little long-run, systematic, empirical evidence on the relationship between complexity and collaboration. This paper contributes to the literature by reporting the growth of inventor collaboration, highlighting the timing of that growth and its geography. The paper then explores the statistical relationship between inventor collaboration and knowledge complexity. A new inventor-patent database is built that identifies all (co-)inventors along with their geographical location for more than 3 million US patents. The key findings are that (1) collaboration has increased sharply since the 1930s; (2) there is a positive relationship between complexity and collaboration on US patents; and (3) increasing complexity on patents is associated with within-city collaboration.

Introduction

Recent work shows that since 1950 the production of knowledge has become increasingly the output of collaboration (Wuchty et al., 2007). Theories explaining scientific collaboration are extensive, ranging from resource optimization (Eaton, 1951), increased productivity (de Solla Price, 1986), access to ideas and resources (Wray, 2002), and to intellectual or social linkages (Thorsteinsdottir, 2000). However, the key underlying argument is that knowledge production is becoming increasingly complicated and requires inputs that exceed that of the individual. For example, to generate new insights and knowledge on particle physics The Large Hadron Collider project pools resources from over 100 countries, linking thousands of scientists and engineers. Such complex projects cannot be undertaken by a single individual, especially when scientists and engineers are becoming more specialized.

The production of knowledge is unevenly distributed across space (Balland & Rigby, 2017). Increasingly, knowledge is perceived as sets of embodied capabilities and routines constructed through processes of learning (Lundvall & Johnson, 1994; Zollo & Winter, 2002). The embodied nature of knowledge restricts its movement between entities and across space. This is especially the case for complex ideas that rest heavily on non-ubiquitous forms of tacit knowledge (Kogut & Zander, 1992; Von Hippel, 1994). The diffusion of complex, tacit knowledge requires repeated face-to-face interaction, emphasizing the need for co-location of economic agents (Singh, 2005; Sorenson et al., 2006). Because complex knowledge production rests heavily on face-to-face interaction, it tends to be limited to larger cities (Balland et al., 2018)

The political and scholarly interests in collaboration and complexity rest upon the notion that both have a positive relationship with the value, quality or quantity of output. Collaborative work is found to spur creativity (Uzzi & Spiro, 2005), receive more citations (Glanzel, 2002), have higher acceptance rates for academic publication (Presser, 1980), raise productivity (Lee & Bozeman, 2005) and enables researchers to engage with large research questions (Thagard, 1997). In similar vein, complex knowledge is seen as a key source of competitive advantage (Kogut & Zander, 1992; Maskell & Malmberg, 1999; Sorenson et al., 2006). Access to complex knowledge allows inventors to engage with more advanced activities and capture the benefits that might arise from such engagement. These benefits may be pecuniary within firms or manifest as increased levels of economic development within society.

Surprisingly, empirical evidence on the relationship between complexity and collaboration is limited. There is no long-run, large scale, systematic and historical evidence linking collaboration and complexity, primarily because empirical data on collaboration is lacking. Most research on collaboration focuses on relative short time-frames, specific fields or projects. The same is true for measures of knowledge complexity. Measures of complexity are also difficult to design, construct and operationalize over time and space.

The primary aim of this paper is to examine patterns of co-invention on US patents along the axes of collaboration, complexity and geography. Using state-of-the-art machine learning and text mining algorithms, I construct a unique inventor-patent database that identifies all (co-)inventors and their geographical location(s) on more than 3 million patents between 1836 and 1975. The raw data originate in the United States Patent and Trademark Office (USPTO). I merge the inventor collaboration data with information on the technology classes of US patents, available from the USPTO, to construct measures of complexity for each patent. The resulting data reveals the characteristics of US inventor collaboration on patents of varying complexity, across geographies and at different moments in time. The key results of this study show that (1) collaboration has increased sharply since the 1930s; (2) there is a positive and significant relationship between complexity and collaboration on US patents; and (3) increasing complexity on patents is associated with within-city collaboration.

The rest of this paper is structured as follows. In the second section I motivate the research and discuss recent work concerning collaboration, knowledge complexity, geography of collaboration and economic history. The third section describes the construction of the historical long-run U.S. patent-inventor database. Section four presents the empirical analyses that examine co-invention on US patents with respect to collaboration, complexity and geography. The final section presents and discusses the key findings of this paper.

2. Literature

This section reviews relevant literature on collaboration, knowledge complexity, geography of collaboration and economic history. Four hypotheses are posited.

2.1 Collaboration

Scholars across different disciplines have recognized the increasing presence of collaboration in the production of knowledge. Since the 1950s there is empirical evidence that knowledge production in general is no longer the product of the 'lone-wolf' (Patel, 1973) or individual genius (Merton, 1968), but increasingly the outcome of cooperation. This trend is observed in almost every field and across the globe (Crescenzi et al.,2016; Hoekman et al., 2010; Jones et al., 2008; Merton, 1973; Wuchty et al., 2007).

The increasing complexity of contemporary knowledge production requires inputs that exceed that of the individual. De Solla Price (1963) refers to this trend as the shift to 'big science', emphasizing the increasing importance of complex science-based technologies (Noble, 1979; Pavitt, 1984). Within such system, knowledge producers might collaborate to optimize resources (Eaton, 1951), increase productivity (de Solla Price, 1986), access complementary ideas (de Solla Price, 1970), access to resources (Wray, 2002), or for intellectual or social reasons (Thorsteinsdottir, 2000). Jones (2009) argues that the cumulative nature of knowledge produces a 'knowledge burden' that makes subsequent innovation all the harder. This knowledge burden is dampened to some degree by collaboration that spurs creativity (Uzzi & Spiro, 2005).

Collaboration is positively associated with the value, quality or quantity of output. Collaborative work receives more citations than individual work (Frenken et al., 2005; Glanzel, 2002), it has higher acceptance rates in academic publication (Presser, 1980), it raises productivity (Lee & Bozeman, 2005) and enables researchers to engage with larger research questions (Thagard, 1997). Faems et al. (2005) find that collaborating firms are more likely to produce commercially successful products than non-collaborating firms. Katz & Martin (1997) provide a critical review on research collaboration.

Analysis of collaboration has been restricted to relatively recent data. Using USPTO patent records from 1975 to 1995, Wuchty et al. (2007) report that the average number of inventors on patents has been increasing over time. Others report similar findings (Crescenzi et al., 2016; Fleming & Frenken, 2007; Lobo & Strumsky, 2008). While this research shows the rise of inventor collaboration over the last 30-40 years, we have little information about the longer-run history of co-invention. In particular, we do not know if the recent data exhibit a significant break from the past. A first research question thus focuses on the long run history of collaboration on U.S. patents. In relation to this question, I pose a first hypothesis to examine:

H1: *Collaboration on US patents increased between 1836 and1975*

2.2 Knowledge complexity

The production of knowledge has widely been identified as the key input to innovation. It is therefore seen as the source of competitive advantage and the driver of long-run economic growth (Lucas, 1988; Romer, 1986; Solow, 1957). Combined with a growing awareness of the heterogeneity of firms (Nelson & Winter, 1982), this led to the development of knowledge-based views of the firm, in which coordination, recombination and integration of the (specialized) knowledge of individuals is seen as central to firm performance (Grant, 1996). These views rapidly became extended to incorporate the coordination of inter-firm processes, as inter-firm collaboration and mobility of employees produced knowledge spillover and knowledge networks (Almeida & Kogut, 1999; Kogut, 2000).

Not all knowledge is equal and some forms of knowledge impact the competitiveness of firms and regions more than others (Maskell & Malmberg, 1999). Knowledge can be differentiated across multiple dimensions (Winter, 1998). The classic work by Polanyi (1958) distinguishes knowledge based on the degree to which it is codifiable. Unlike codified knowledge, tacit knowledge is difficult to codify, it is embodied within and across economic agents and structures the routines that they employ, often unconsciously. The non-ubiquitous and relatively immobile nature of tacit knowledge is widely seen as a critical agent of regional competitive advantage in our global, interconnected age (Asheim & Gertler, 2005).

Knowledge also varies in complexity, though Kogut & Zander (1992) argue that complexity is a critical dimension of what makes knowledge tacit. Different perceptions on knowledge complexity exist. For Fleming & Sorenson (2001) knowledge complexity can be derived from the interaction between the components it combines and how easily these components can be combined. Others argue that knowledge is complex when it can surprise the observer and its characteristics cannot simply be linked to its components (Axelrod & Cohen, 2000; Tsoukas, 2005). From both perspectives, complex knowledge is emergent and more likely to rely on tacit, non-ubiquitous knowledge than less complex knowledge. Like tacit knowledge, complex ideas tend not to flow readily over space (Balland & Rigby, 2017; Sorenson et al., 2006).

Many scholars have argued that knowledge is becoming increasingly complex over time, witnessed in the shift to big science and the importance of science-based technologies (Pavitt, 1984; de Solla Price, 1963). Some stress this is troublesome as producing new knowledge becomes increasingly difficult and costly, slowing down economic growth (Jones, 2009). Surprisingly, there is very limited empirical evidence that shows whether knowledge is becoming more complex. This paper aims to provide such evidence using historical USPTO patents.

These remarks lead to the following hypotheses:

H2: *The complexity of patents increased between 1836 and1975*

and,

H3: *There is a positive relationship between knowledge complexity and collaboration on US patents*

2.3 Geography of collaboration

Relational ties can act as links between agents at different geographical locations over which knowledge is shared. Knowledge is distributed unevenly across space. This is witnessed by localized pockets of specializations across which the quality and value of knowledge varies (Balland & Rigby, 2017). Many scholars have emphasized the role of relational linkages and the network structures they create in the transfer of knowledge (Broekel et al., 2014; Singh, 2005; van Oort & Lambooy, 2014), the diffusion of technological change (Feldman, Kogler, & Rigby, 2014) and regional knowledge production (Van der Wouden & Rigby, 2018). Indeed, collaborations facilitate the flow of knowledge over space and allow for external knowledge sourcing.

Although complex technologies might be more prone to collaboration than less complex technologies, the geographical patterns of collaboration on complex and less complex knowledge remains unexplored. On the one hand, it is argued that the production of complex knowledge relies heavily on face-to-face interactions provided by *local buzz* – the dynamics and interactions that arise from the co-location of a set of related economic agents (Storper & Venables, 2004). From this perspective, collaboration on complex knowledge, in comparison to less complex knowledge, is more likely to take place by inventors in the same location. On the other hand, scholars have argued that the inputs to complex knowledge are highly specialized and produced in specific but not necessarily the same locations. Thus, collaboration on complex knowledge might span *pipelines* that connect different knowledge sets at multiple locations (Bathelt et al., 2004) Following these claims, I construct the following hypothesis:

H4: *Complexity promotes within-city collaboration rather than between-city collaboration*

2.4 Economic and institutional context

The time-window of this study covers events that have significantly impacted the U.S. economy[2]. From a technological perspective, the period 1870-1970 is often referred to as the "Golden Age" of U.S. invention (Akcigit et al., 2017; Gordon, 2016). Researchers of American innovation have long stressed the importance of the institutional environment constructed in the US since the 19th century. In general, institutional, political and legal frameworks secure property and enforce contracts (Landes, 1969). Intellectual property rights provide incentives to explore new technological endeavors and encourage risk-taking behavior because it makes economic agents immune from political turmoil, and allows the appropriation of new technologies (North & Thomas, 1973; Rosenberg, 1982).

In contrast to popular perception, recent research provides evidence of how the rapid industrialization of the US in the late 19th century co-occurred with the creation of robust democratic governmental institutions (Bensel, 2000; Novak, 2008). Two specific institutions are particularly important for technological progress. First, the construction and enforcement of a patent system created a market for technologies which allowed inventors to appropriate, sell or license their inventions, providing incentives for inventors to specialize in technology (Lamoreaux et al., 2013; Lamoreaux & Sokoloff, 1996). Although the United States was not the first country to develop such a system, it is generally perceived as constructing the first *modern*

[2] For example, the American Civil War, First and Second World War and the Great Depression

patent system in 1836[3]. The US established a patent evaluation process conducted by certified patent examiners that required detailed information on the specifications of the technology. This system created a modern interference and administrative appeal practices (Lamoreaux & Sokoloff, 2005). This system also promoted the diffusion of technological knowledge because the patent records were publicly available (Lamoreaux & Sokoloff, 1999). In the 1840s networks of interconnected inventors, patent agents, lawyers and journals became organized around the publicly available patent records and actively diffused the technical knowledge across the country (Lamoreaux & Sokoloff, 2002). There are no reports of similar networks emerging in 19th century England, The Netherlands, France, Germany or other industrializing nations.

Importantly, under the original 1836 US patent system only individuals were able to apply for patents. Firms were not allowed to patent – even if the invention was created in their shop (Coriat & Weinstein, 2012). Fisk (1998) argues that part of this legislation rested upon the commonly hold notion that an invention is the result a one person's 'genius'. To maintain their competitive position, firms readily acquired new technological ideas. Patents as tradeable and profitable goods in a market for technologies gave rise to specialized 'entrepreneurial' inventors. These inventors became a great source of technical knowledge, because they made a career producing specialized patents and assigning the rights to firms (Lamoreaux & Sokoloff, 2005).

Second, the construction of the US postal system proved to have a significant impact on technological progress. During the late 18th and early 19th century, the postal system rapidly extended communications throughout vast country, providing access to information and long-distance communication for rural and urban populations (John, 2009). Moreover, Khan (2005)

[3] US Congress, "An Act to Promote the Progress of Useful Arts", July 1836

notes that inventors could send their patent application to the USPTO at the expense of the postal system. Acemoglu et al. (2016) provide empirical evidence that U.S. counties with a post office had significantly greater patenting rates than counties without a post office.

Aside from the construction of the US patent and postal system, transformations of the US capitalist system significantly impacted the production of knowledge in the U.S. during the19th and 20th centuries. The US was transforming from entrepreneurial capitalism to corporate capitalism dominated by large firms (Coriat & Weinstein, 2012). With the anti-trust Sherman Act of 1890, large corporations faced increased competition and relying on acquiring new technologies produced by independent inventors was an increasingly risky strategy. As an alternative they began to construct in-house R&D facilities to internalize knowledge production. By the early 1900s highly productive inventors increasingly developed long-term relationships with firms (Khan & Sokoloff, 2001).

The construction of in-house R&D labs resulted in questions about ownership of the technologies invented. The 1836 Patent Act allowed only individuals to apply for patents. This began to change with the emergence of the ‘shop right’ doctrine in the 1880s. Shop rights’ granted employees ownership of patents they invented, but employers were granted licenses to the patented technology as compensation for their funding of R&D.

By the early 1900s it was clear that the romantic vision of the lone genius was essentially dead. New technologies became increasingly complicated, demanding multiple and different skills and resources. Fisk (1998, p.1141) argues that “collective research and development had become the source of most inventions long before courts and the public finally realized it”. This new reality challenged the individualistic paradigm of US patent law and set in motion the move towards the second major change in US patent legislation. In 1933 the US Supreme Court ruled that 'the respective rights and obligations of employer and employee, touching an invention conceived by the latter, spring from the contract of employment' (in Coriat & Weinstein, 2012). Thus, if employees were 'hired-to-invent', the product of their labor belongs to the firm (Coriat & Weinstein, 2012). Indeed, inventors recruited by corporates became colleagues and possible collaborators instead of competitors.

Finally, developments in transportation and telecommunication technologies rapidly changed in the 19^{th} and 20^{th} century. By the end of the 19^{th} century railroads had extended across most of the US, fostering economic growth by connecting markets (Atack et al., 2010) and inventive activities (Perlman, 2016). The extension of the railroad was accompanied by the telegraph, making use of the railroads' right-of-way. According to Nonnenmacher (2001) the travel time from New York to Chicago was 2 days in 1860, but the telegraph allowed for almost instant messaging. Around 1920, US commercial air travel took off, significantly cutting down the travel time between cities. Perhaps even more importantly, the early 1900s are characterized by the rapid introduction of the telephone and automobile. By 1930 around 40% of the US homes had a telephone connection and 20% of American families had an automobile (Fischer & Carroll, 1988). These developments make it significantly easier to collaborate in general and across distance.

3. Data

In this section I discuss the methods used to construct the collaboration data.

3.1 Searching, matching and recording

The data generated for this work extends the publicly available HistPat database (Petralia, Balland, & Rigby, 2016). HistPat contains geographical information for historical patents provided by the USPTO between 1790 and 1975. The authors of HistPat scraped the text from digitized historical patent files available on Google Patents and EspaceNet, and recorded the first inventor and a geographical location. Unfortunately, HistPat provides no information for any possible additional co-inventors.

I contribute to these data by identifying all inventors and their geographical locations for each USPTO patent between 1836 and 1975[4]. I use the raw scraped text files for the 4,125,734 patents and examine whether each word in these text files is part of a first or family name, or a geographical location in the US. Figure 1 shows an example of the data for USPTO patent 1. The upper part of the image is the scanned copy of the original document. The lower part is the digitized version of this document, published by Google Patents. The data for the lists of first and family names comes from the digital USPTO patents from 1975 up to 2005 (Lai et al., 2012), inventor names in HistPat, and the U.S. Census. The data for the geographical locations come from the same sources as well as the U.S. Bureau of Economic Analysis.

Once a (fuzzy) match between a word on the patent text with one of the lists with names and/or location occurs, a series of complex algorithms is run. These algorithms can broadly be placed in two groups. The first set examines the words before and after the matched word to determine whether the matched word is a name or geographical location. For example, the word

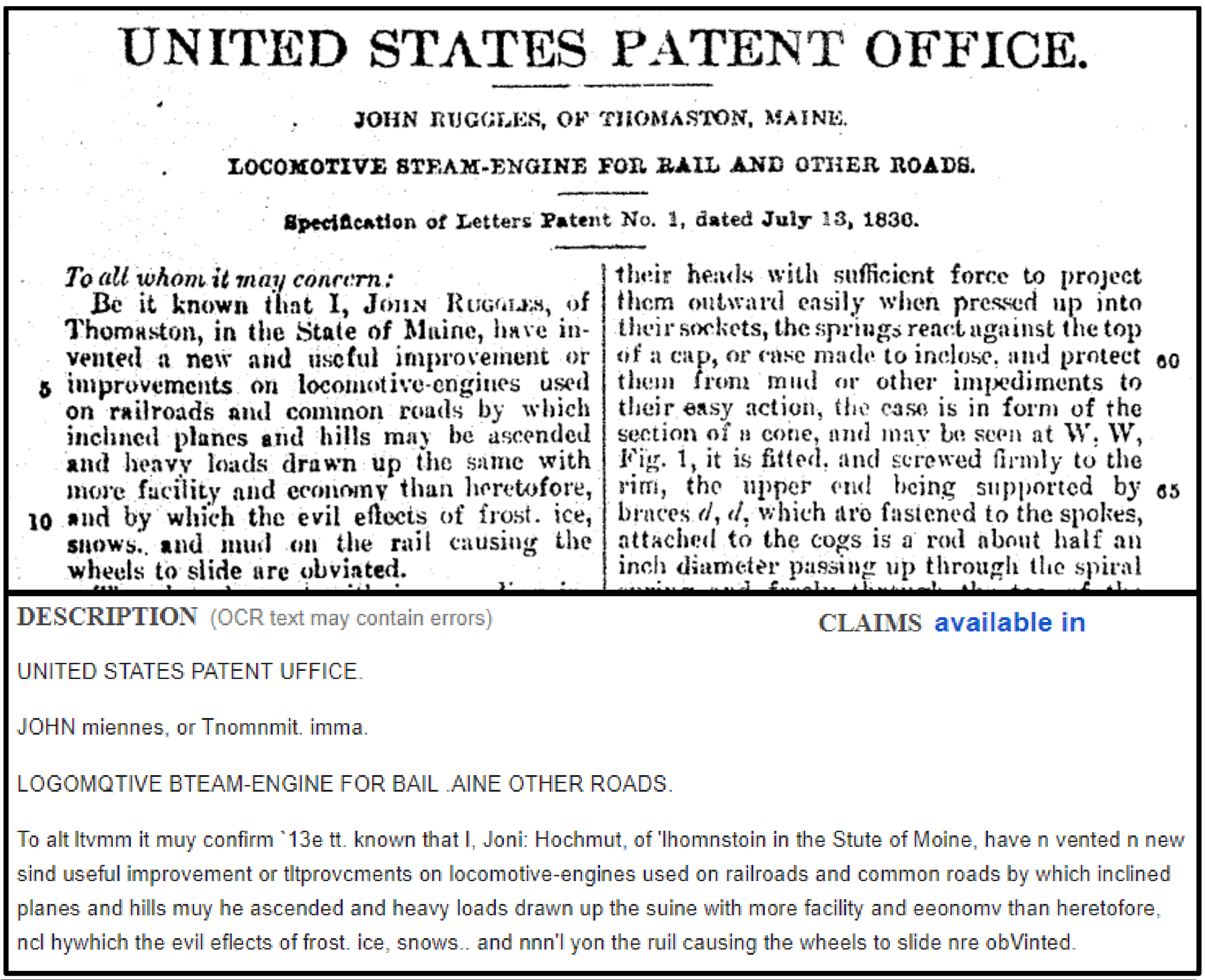
[4] The analysis is limited to 1836 because before 1836 the USPTO did not make use of patent examiners. It is generally accepted that this institutional change has significantly impacted US patenting activity (Lamoreaux et al., 2011; Sokoloff, 1988).

"EDISON" can be matched to inventor "THOMAS EDISON", but also to the location "EDISON, NEW JERSEY". Examining the text before and after the matched word helps to distinguish between name and location. At the end of this stage, more than 8 million names on the 4.1 million patents are recorded. For about 60% of these observed names, the algorithms also associate a location. All the words that are categorized as geographical locations but can't be linked to a name on the patent are ignored.

The second set of algorithms record a series of more than 30 statistics for each matched word. These statistics are used for the machine learning exercise, discussed below, to determine whether an observed name is truly an inventor and not a witness, examiner, corporation or reference. For instance, once a name is observed I count how often it occurs in the text, how *far* the name is from the word "inventor", how many words are between the observed name and the top and bottom of the patent, if it is adjacent to corporate identifications, and so on. Similar operations are undertaken for an observed geographical location that is linked to a name, but those statistics are only used to generate a likelihood measure of a correct name-location link. Table 1 shows a truncated snapshot of the resulting database. Note that the algorithms have been able to match words from the digitized text in Figure 1 to actual names and locations. At this stage it is still unclear whether the observed name corresponds to an inventor.

A great challenge for the data construction is to deal with the poor quality of the scanned documents and the corresponding digital text. As is clear from Figure 1, for some patents the text is difficult to read and match against names and locations. To overcome this issue I employ fuzzy match algorithms if an exact match is not found. Fuzzy matching allows for deviation between two text strings to be matched. This means that if a word is misspelled, a positive match may still occur if it meets a set of criteria. However, it also allows for false-positive matches when words are matched that are not the same. Given the poor quality of the digital text for some years, the benefits of fuzzy matching outweigh the drawbacks. Moreover, the machine learning exercises described below limit the potential impact of false-positive matches because they are less likely to correspond to an inventor and thus ignored.

Figure 1. Photo-copied and digitized records of USPTO patent US1A



UNITED STATES PATENT OFFICE.

JOHN RUGGLES, OF THOMASTON, MAINE.

LOCOMOTIVE STEAM-ENGINE FOR RAIL AND OTHER ROADS.

Specification of Letters Patent No. 1, dated July 13, 1836.

To all whom it may concern:

Be it known that I, JOHN RUGGLES, of Thomaston, in the State of Maine, have invented a new and useful improvement or improvements on locomotive-engines used on railroads and common roads by which inclined planes and hills may be ascended and heavy loads drawn up the same with more facility and economy than heretofore, and by which the evil effects of frost, ice, snows, and mud on the rail causing the wheels to slide are obviated.

their heads with sufficient force to project them outward easily when pressed up into their sockets, the springs react against the top of a cap, or case made to inclose, and protect them from mud or other impediments to their easy action, the case is in form of the section of a cone, and may be seen at W, W, Fig. 1, it is fitted, and screwed firmly to the rim, the upper end being supported by braces *d*, *d*, which are fastened to the spokes, attached to the cogs is a rod about half an inch diameter passing up through the spiral

DESCRIPTION (OCR text may contain errors)

CLAIMS available in

UNITED STATES PATENT UFFICE.

JOHN miennes, or Tnomnmit. imma.

LOGOMQTIVE BTEAM-ENGINE FOR BAIL .AINE OTHER ROADS.

To alt ltvmm it muy confirm `13e tt. known that I, Joni: Hochmut, of 'lhomnstoin in the Stute of Moine, have n vented n new sind useful improvement or tltprovcments on locomotive-engines used on railroads and common roads by which inclined planes and hills muy he ascended and heavy loads drawn up the suine with more facility and eeonomv than heretofore, ncl hywhich the evil eflects of frost. ice, snows.. and nnn'l yon the ruil causing the wheels to slide nre obVinted.

Source: Google Patents

Table 1. Observed names and geographical locations for USPTO patents (truncated)

Patent	Year	Inventor	City	County	State	A1	A_x
1	1836	JOHN RUGGLES	THOMASTON	KNOX	MAINE	3	…
2	1836	GEORGE SULLIVAN				0	…
2	1836	FRANKLIN BROWN				0	…
3	1838	FOWLER M RAY	CATSKILL		NEW YORK	-48	…
4	1853	BENJAMIN IRVING	GREENPOINT	KINGS	NEW YORK	0	…

3.2 Supervised machine learning to identify inventors

The next step is to distinguish between inventors and non-inventors in the data. Non-inventor names can correspond to other entities that have name-like characteristics. I use supervised machine learning techniques to classify each of the 8 million observed names as either an inventor (1) or a non-inventor (0). Each observed name on a patent is called an event.

Supervised binary classification machine learning techniques typically make use of three different datasets. The training set contains verified or *known* information on events and the corresponding characteristics. Machine learning algorithms can be deployed to this dataset to model and *learn* which characteristics and structure among the characteristics can best classify events. A validation dataset is used to tune the parameters of the candidate models and prevent the models from over-fitting to the training data. A test set is used to evaluate the performance of the candidate models and select the final model. The final model is used to predict for the events in the mined data whether each event is an inventor.

To construct the co-inventor data I follow this approach, but have to overcome a key issue. The only training data that can be built for this project is the first inventor data recorded in

HistPat[5]. If the observed names generated by my algorithms match with the inventor listed on the patent recorded in HistPat, I have a *known* event for an inventor with corresponding characteristics[6]. However, I have no information on *known* events for non-inventors. Such situations call for a one-class classifier approach in which only characteristics for one class can be learned. Events are classified into that class if their characteristics fit into the class given certain thresholds derived from distributions in the training data. The quality and accuracy of one-class classifiers often underperform multi-class classifiers because they can only learn from one class. Based on this underperformance, I follow a multi-class classifying approach[7].

To be able to continue with the multi-class classifier approach, I generate artificial *known* event data for non-inventors. From the work of Wuchty et al (2007) it is reasonable to assume that the average number of inventors on a patent is well below two before 1975. Given the fact that there are more than 8 million events for 4 million patents a fair share of events are non-inventors. I take a random sample of 30% of all the events that are not *known* inventors and artificially classify them in my training data as *known* non-inventors[8]. The remaining 70% of not *known* inventors are excluded from the training data. The training data now consists of over 3 million events with *known* inventors and more than 1 million artificial *known* non-inventors.

---

[5] Inventor names are scraped from Google Patents and EspaceNet, cleaned and formatted by Petralia et al. (2016).

[6] This assumes that the scraped inventor names recorded in the HistPat data are correct.

[7] Several extensive one-class classifier algorithms are ran. None came close to the performance of multi-class classifiers.

[8] In larger samples the quality of my trained models decreases, which indicates that I falsely artificially assign inventors to the non-inventor class, obscuring the learning process. In smaller samples there are not enough correct non-inventors artificially assigned as non-inventor in the training data to learn from.

Although there might be false-negative events in the latter group (non-inventors that are actually inventors), the majority of events provide data to learn from.

There is an expansive battery of machine learning algorithms available that can be used to train a supervised binary classification model. It is uncertain which algorithm will generate the best performance. Therefore, it is best practice to explore a variety of algorithms and examine fitness criteria. Some of the algorithms have multiple parameters that can be optimized to increase the performance of the models. This leads to extensive searches through parameter space and requires considerable computing power. For the exercise described below, I've experimented with support vector machine, generalized linear models, random forest, gradient boosting machine and deep learning algorithms, as well as stacking and ensemble techniques. The final, best performing trained model is fit using gradient boosting machine algorithms. This model performed the best on the test data with an average accuracy of 83%. Other algorithms that produced interesting results were deep learning and random forest algorithms, with respectively 79% and 80% accuracy.

It is important to note that after the trained model is applied on the mined data, roughly 200 results have been randomly selected and manually compared to the original photo-copied patents available online[9]. Although a handful of patents were incorrectly classified, the analysis of these patents didn't show any sign of a systematic bias.

The next step is to disambiguate the inventors that are assigned to patents and assign each unique inventor with a classifying ID. This means that we want to know if Josephine Bruin

[9] Available at http://google.patents.com

living in Los Angeles in 1919 is the same inventor as Josephine Bruin living in Los Angeles in 1921. To deal with such issues, many disambiguation approaches have been developed. I loosely follow Ventura et al. (2015) and construct a supervised machine learning approach. This approach involves a training, validation and test dataset. My training, validation and test datasets come from the Lai et al. (2012) database that holds information on disambiguated inventors on US patents between 1975 and 2012. All databases hold event-level information with a series of similar characteristics. Importantly, the training database also holds information about which inventors are the same.

The disambiguation approach involves the following steps:

1. Select characteristics on which pairs of inventors are to be compared. These characteristics need to occur in all databases. I select: first name, middle name, last name, year, city, county, state, technology class (1:5).
2. Generate for each of the characteristics a method to compare similarity between two strings (i.e. how similar are 'Josephine' and 'Josphine')
3. Pair-wise compare all inventors in the training database and find the similarity score for each of the characteristics. This results in a vector of 12 similarity scores.
4. For each pair-wise comparison in the training database we know whether the pair of inventors are the same unique inventor. We can classify the vector of similarity scores with a 1 if the compared inventors are the same and 0 if they are not.
5. Train a series of supervised machine learning algorithms to 'learn' which combination of similarity characteristics correspond to pair-wise comparison of identical (1) and different inventors (0). The trained model can be used to predict whether a vector of similarity characteristics correspond to the same inventor or not.

6. Repeat step 3 and 4 for the newly mined event data. This results in hundreds of millions of vectors with similarity characteristics.
7. Apply the model generated in step 5 to the data generated in step 6.
8. Assign the inventors that are identified as the same individuals with the same ID.

The resulting inventor-patent database identifies 1,922,754 inventors, with 4,437,960 observations on 3,365,253 unique US patents between 1836 and 1975. Figure 2 shows the annual number of patents granted by the USPTO (left axis) and the percentage of the patents for which at least one inventor is identified. For all but three years more than 80% of the patents are in the database. For a lot of years more than 90% of the patents are included. The wide coverage of the annual number of patents and lack of theoretical or empirical motivations to expect systematic bias in the unobserved patents suggests that this database is a representative sample of the historical USPTO patents granted between 1836 and 1975.

Table 2 shows the top 10 inventors identified by these exercises. All names in this table are well known inventors. Surprisingly, Thomas A. Edison is not the most observed inventor in our dataset, although he is generally regarded as the most inventive individual in the late 19th and early 20th century. Times Magazine assigns 1,093 patents to his name, while I only find 536 patents on which he is an inventor. In our dataset, Francis H. Richard is found on 737 patents, while Times Magazine connects him to 894 patents. There are multiple reasons why Edison might not be identified on the other 500 or so patents: the quality of the OCR was too poor to read or due to usages of different spellings of his name (i.e. Thomas *Alva* Edison). More likely, it is because his name occurs on a disproportionally large numbers of patents as a reference (thus not an inventor). The machine learning algorithms have *learned* this and require a large number of positive scores to identify “Thomas Edison” as an inventor. If the algorithms can’t construct these scores in the patent document, the probability that “Thomas Edison” is truly an inventor of the patent is assessed as low by the learning algorithm and is not regarded as such.

Figure 2. Annual number of patents granted and percentage of patents in database

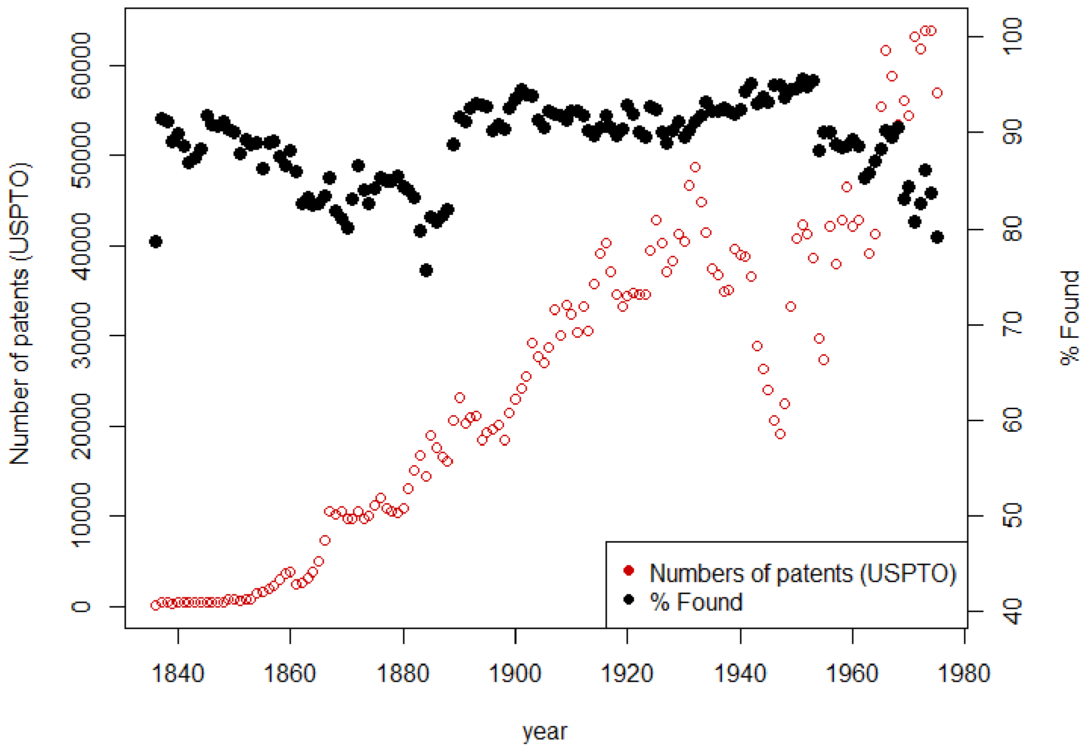


Table 2. Top 10 inventors in database

Number	Name	Patents
1	Francis H. Richard	738
2	Thomas A. Edison	536
3	Elihu Thomson	516
4	John F. O’Connor	511
5	Edwin H. Land	465
6	Clyde C. Farmer	458
7	George Albert Lyon	453
8	Carleton Ellis	481
9	Louis H. Morin	406
10	Thomas E. Murray	388

4. Empirical results

In this section, the patterns of collaboration on U.S. patents since 1836 are explored.

4.1 Collaboration on US patents

Collaboration on US patents has increased over time. Figure 3 indicates that up until the 1920s, annual shares of collaborated patents remained below 20 percent. After the 1920s a rapid increase occurs and by the 1960s about 30 percent of patents result from collaboration. By the 1970s, inventor collaborations generate about 40 percent of patents.

The average number of inventors on patents has also expanded over time. Figure 4 shows the annual average number of inventors on patents. The average number of inventors remained rather stable up until the 1940s. After 1940, a constant increase in the number of inventors on patents is observed. By 1970 the average team size is 1.6 inventors per patent compared with a value of about 1.2 inventors at the end of the 19th century. The same trend is observed if only collaborated patents are considered. The in-set in Figure 4 shows that in the 1940s the average team-size increases rapidly. This indicates that the increase in the average number of inventors on a patent is not produced by fluctuations in the annual number of single inventors, but can be attributed to growing average team-sizes. Note that the average numbers in the late 1970s presented here align with the average numbers documented for 1975 in Wuchty et al. (2007).

There is variation in the average number of inventors on patents in different technological categories. Patent examiners classify each patent into one or more classes depending on the knowledge claims that they make. There are 438 different primary level technology classes in the USPTO into which all utility patents are placed. Hall et al. (2001) produced a classification that aggregates these 438 classes into 6 broad categories. Figure 5 plots the annual average number of inventors on patents for each of these broad categories. In early years there is substantial variation because of the low number of counts. By the late 19th century the ‘Computer & Communication’ category has the highest average number of inventors on patents – the era of great advances in telecommunication technologies. Throughout the entire 20th century, patents in the ‘Chemical’ category have the highest average number of inventors. Together these findings confirm hypothesis 1.

Figure 3. Share of collaborated patents between 1836 and 1975.

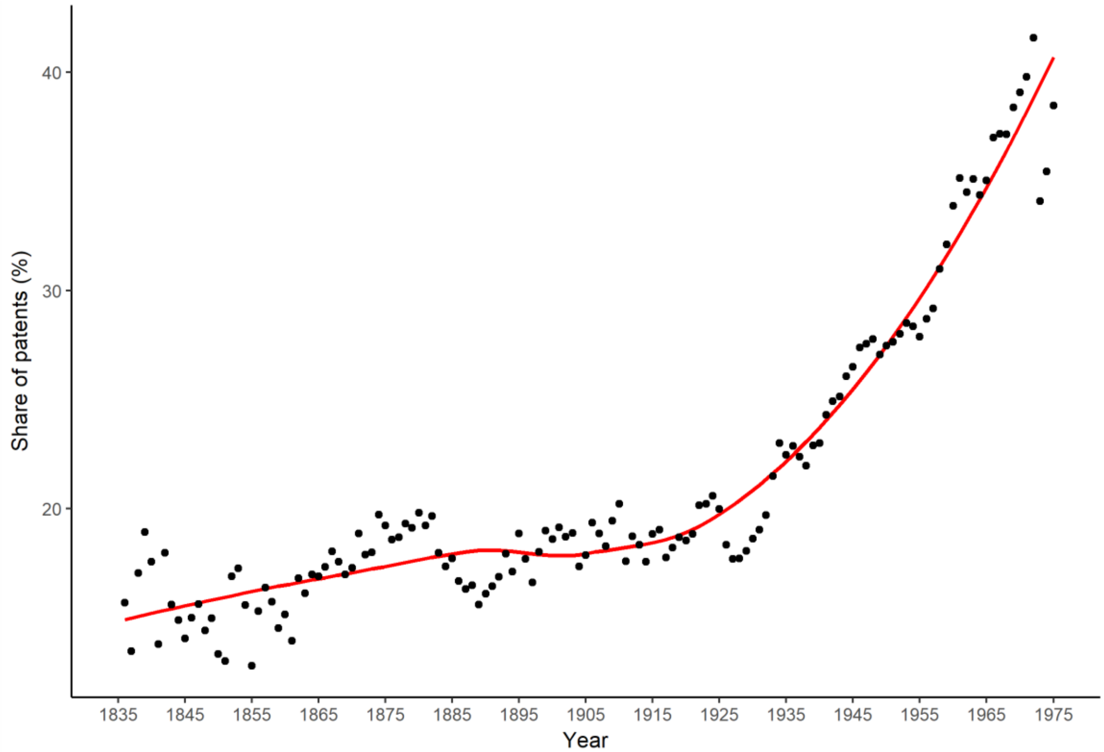


Figure 4. Average number of inventors on a patent between 1836 and 1975

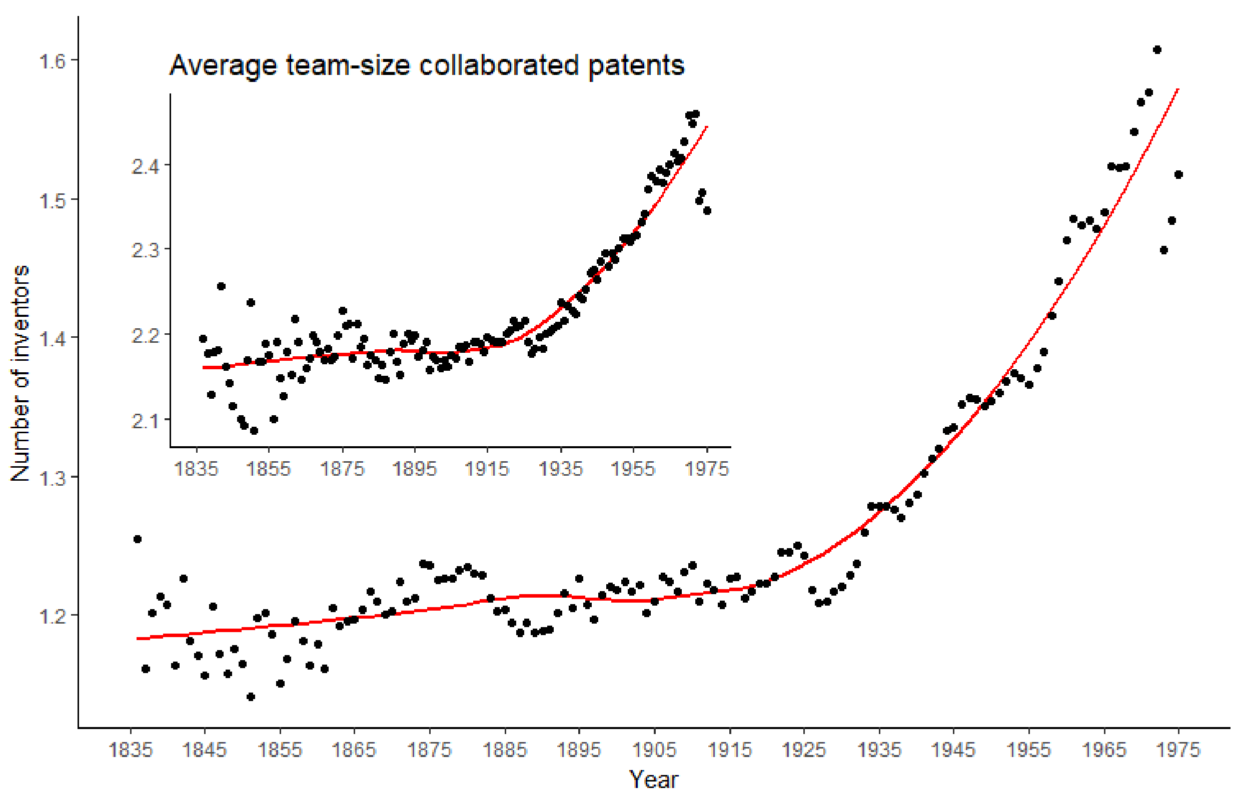
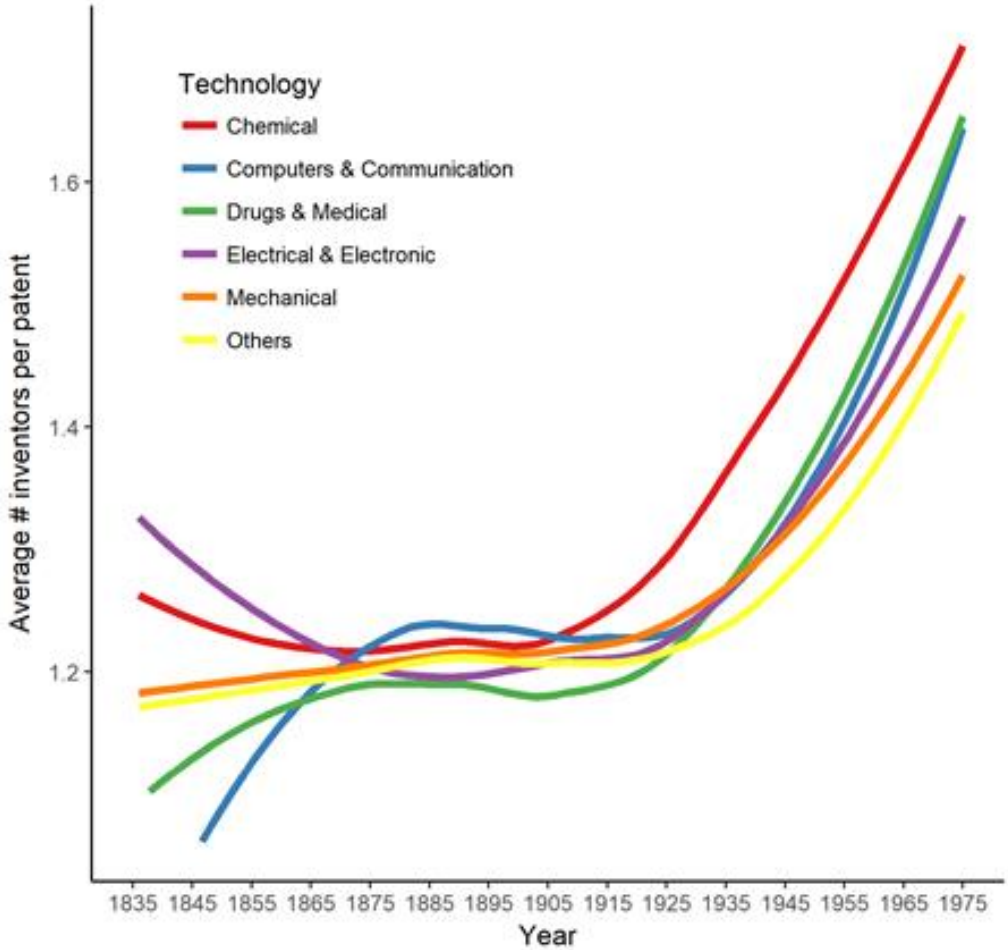


Figure 5. Average number of inventors per patent by technological category



4.2 Complexity and collaboration

Did the complexity of patents increase over time, and is there a positive relationship between complexity and collaboration? The *complexity* measure of a patent originates in Fleming & Sorenson (2001) and is based on the *NK* model of Kauffman (1993). Patent examiners classify each patent into technological classes. The patents examined here are classified in 438 primary classes that can be disaggregated to 10,562 mainline classes and 153,305 sub-classes. The proposed complexity measure is based on the ease with which each mainline class can be recombined with other mainline classes on a single patent. That is, if a mainline class co-occurs relatively often with other mainline classes on a patent, the ease of recombination of the technology class is relatively high. Formally, the ease of recombination of mainline class *i* is defined as

$$E_i = \frac{Count\ of\ Mainline\ Classes\ Previously\ Combined\ with\ Mainline\ Class\ i}{Count\ of\ Previous\ Patents\ in\ mainline\ class\ i} \quad (1)$$

However, the number of mainline classes increases rapidly over time and has an influence on the ease of recombination measure. Patents in 1840 are classified in 372 mainline classes, while the 1970 patents are classified across 7,722 mainline classes. This means that the ease of recombination of technology classes in early year patents is relatively low compared to later year patents because there are fewer classes in which a patent can be located. As a control I construct a standardizing coefficient *S* defined as

$$S = \frac{Count\ of\ unique\ mainline\ classes\ in\ 1975}{Count\ of\ unique\ mainline\ classes\ in\ year\ t} \quad (2)$$

This standardizing coefficient S is used to get an adjusted ease of recombination score

$$EA_i = E_i * S \tag{3}$$

The complexity of patent l is defined as the number of mainline classes divided by the sum of the ease of recombination of these classes. Formally,

$$Complexity_l = \frac{Count\ of\ Mainline\ Classes\ on\ Patent\ l}{\sum_{l \in i} EA_i} \tag{4}$$

The ease of recombination for a patent is calculated using a ten-year window of patent recombinations recorded up to the grant year date of the focal patents. Thus, the ease of recombination for technology classes shifts over time.

Figure 6 shows that the average complexity of patents has been increasing over time. From the inset in Figure 6, average complexity increases through the nineteenth century, it falls from the 1920s to the mid-1940s and then increases once more. The period of initial increase matches the Second Industrial Revolution (Gordon, 2016) and the 'golden age' of US invention (Akcigit et al., 2017). The drop in complexity after the 1920s aligns with the Great Depression. The increase in complexity after the 1940s supports the literature on the shift to 'big science' in knowledge production.

The average complexity of patents varies by technological category. Up until the 1900s the 'Drugs & Medical' patents were most complex. Throughout the first half of the 20th century, patents in the 'Computers & Communication' category are most complex, on average. This period marks the development of the telephone. The decrease after the 1920s might be the result of the 'exhaustion' of this generation of communication technologies. Indeed, complexity rises again in the 1950s when computer related technologies begin to emerge. In the 1960s 'Chemicals' patents are on average the most complex. Patents in the 'Drugs & Medical' category increased in complexity when the production of drugs became integrated with synthetic organic chemistry post World War II (Drews, 2000). Remarkably, the 'Mechanical' category has long been among the most complex categories, but they fell to the bottom of the complexity ranking by 1975.

The average complexity of the patent stock across US cities increased over time. Complex knowledge is unevenly distributed across space (Balland & Rigby, 2017). In Figure 7, the annual distribution of US cities in terms of average complexity of their patent portfolio is shown[10]. The figure shows clear variation amongst the complexity of the patent portfolio of US metropolitan region. Over time the complexity of patents produced in US cities is increasing as the distributions of cities shift to the right. Moreover, the spread in the distribution is increasing over time. At the end of the 19th century the coefficient of variance was below 0.5, while in the 1970s it reached to above 0.60. These findings confirm hypothesis 2: Complexity on patents increased on average between 1836 and 1975.

[10] Outliers are not plotted because these are generated by extremely low patent counts.

Figure 6. Average complexity per technological category between 1836 and 1975

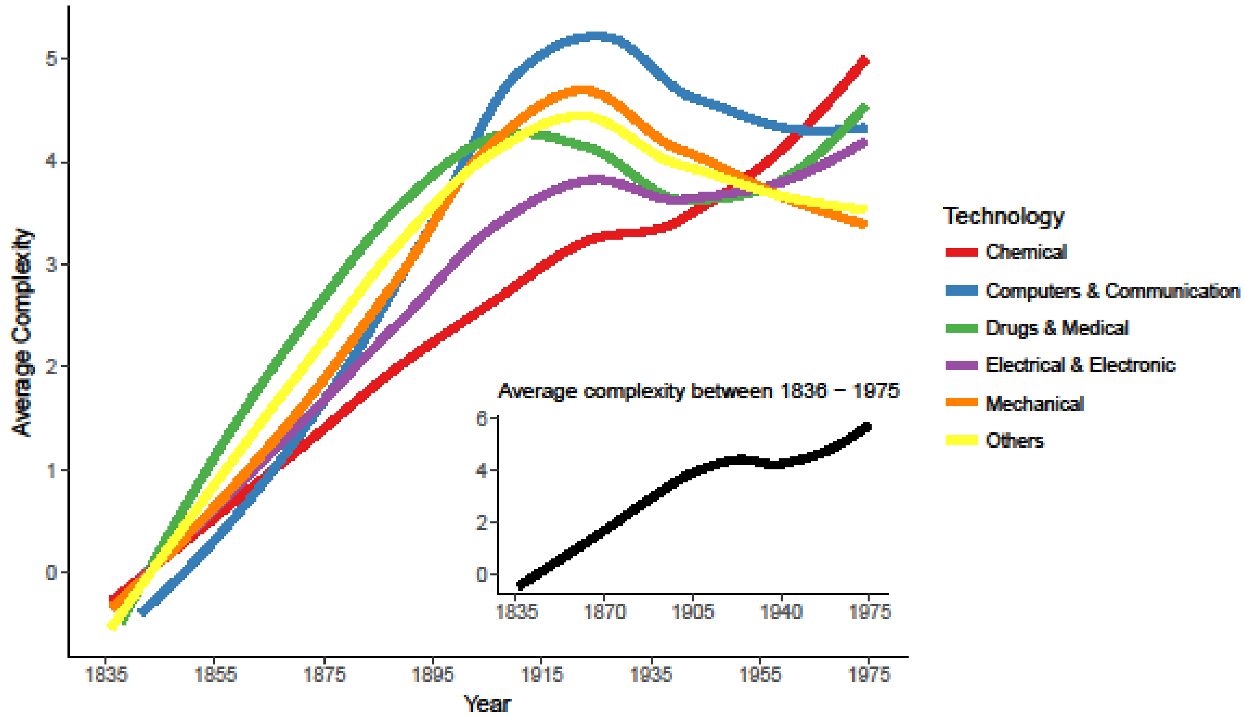
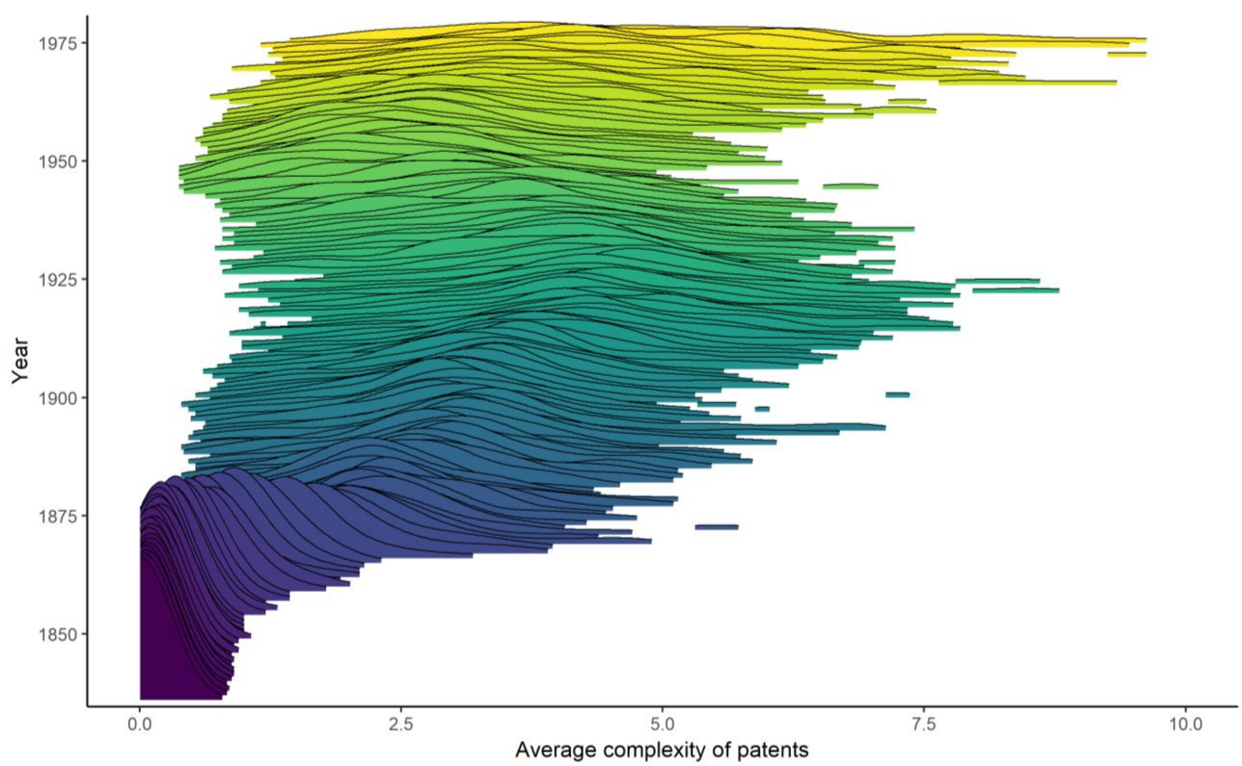
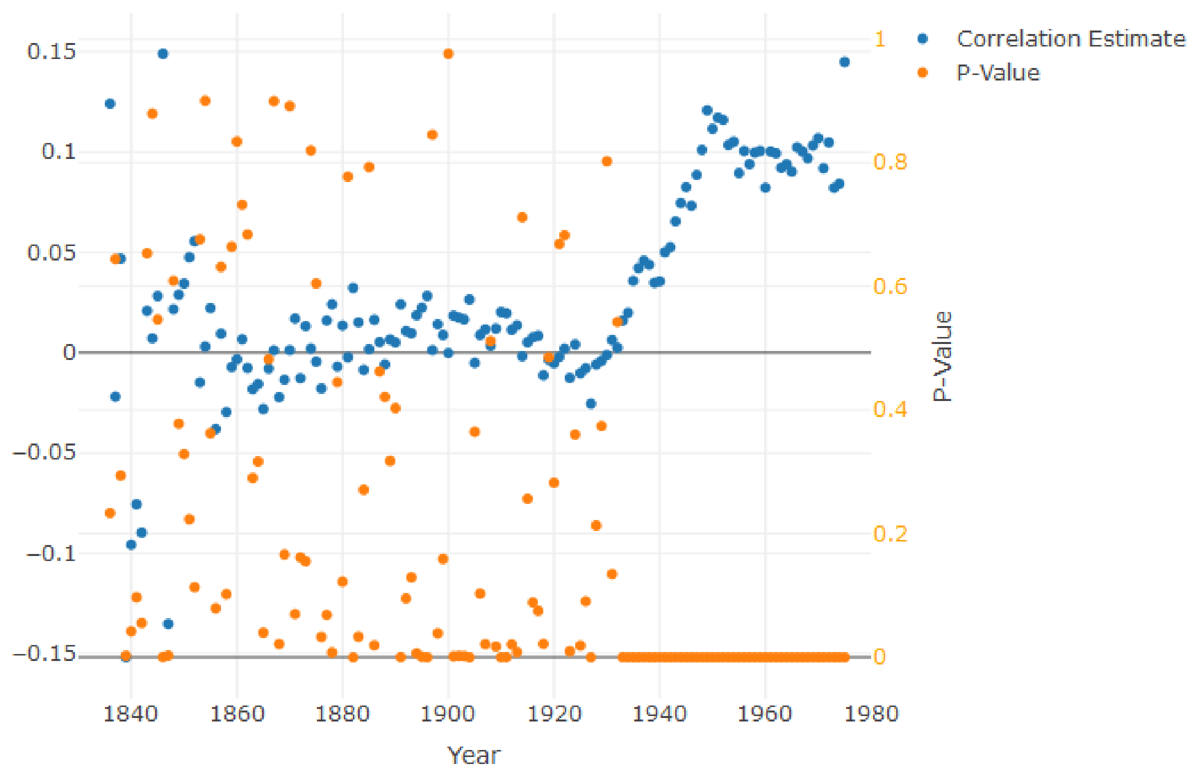


Figure 7. Distribution of US cities based on average complexity of patents between 1836 and 1975



A key question in this paper is whether there is a positive relationship between complexity and collaboration at the patent level. Figure 8 plots the annual correlation between complexity and collaboration on all patents by year. There is no consistent and systematic relationship between complexity and collaboration before the 1920s. During the 1920s there is a significant but weak negative relationship between complexity and collaboration. Interestingly, from 1940s onwards there is an increasing, significant and positive relationship, suggesting that increasing complexity is associated with a higher probability of a patent being produced through collaboration.

Figure 8. Correlation between complexity and collaboration on a patent between 1836 and 1975



These ideas are extended in models examining the relationship between complexity and collaboration in more detail. The dependent variable is a binary variable indicating whether the patent is the result of a collaboration (1) or not (0). The key independent variable of interest is the complexity of the patent. A positive and significant coefficient for this variable would confirm hypothesis 3. The second independent variable is a time dummy that indicates whether the patent is produced after 1940 (1) or not (0). This cut-off point is chosen because Figure 6 indicates there is a structural relationship between complexity and collaboration that after 1940s. Given the evidence on increasing collaboration and complexity over time presented above, a positive coefficient is expected. The third independent variable is the interaction between complexity and the time dummy. A positive coefficient for this interaction term means that for patents produced after 1940 the effect of complexity on collaboration increases with the estimated coefficient for the interaction term. In addition, city, year and technology fixed effects are estimated. It is plausible that specific city-level characteristics promote or hinder collaboration. Year fixed effects control for time-specific shocks in collaboration across the sample. Technology fixed effects control for the possible heterogeneity in the propensity to collaborate across the technology classes. The sample size is restricted to patents for which the primary inventor (first listed on patent) lives in a MSA to satisfy the requirements for the city fixed effects variable. The unit of analysis is the patent. Each patent is assigned to the MSA of the first inventor.

The results presented in Table 3 indicate that there is indeed a positive and significant relationship between complexity and collaboration on a patent, regardless of the model specifications. Model 1 is the baseline model and reports a positive and significant relationship between complexity and collaboration. A coefficient of β_1 = 0.15 means that a doubling of complexity is associated with a change in the odds of a patent being a collaboration by a factor of 1.11. In models 2-5, technology, city and year fixed effects are introduced. The estimated coefficient for complexity remains roughly of the same value and is positive and significant. The estimated coefficient for complexity on collaboration remains positive and significant. In model 7 the time dummy and its interaction with complexity are introduced. The coefficient for complexity now corresponds to patents produced before 1940 (dummy variable is zero). For these patents, controlling for city and technology fixed effects, a coefficient of $\beta_1 = 0.06$ means that a doubling of complexity is associated with a change in the odds that a patent is the result of a collaboration by a factor of 1.04. The positive and significant coefficient for the time dummy (0.54) means that the odds of a patent being collaborated changes by 1.72 if a patent is produced after 1940. The significant coefficient for the interaction term between the time dummy and complexity indicates that when a patent is produced after 1940 (dummy variable is one) the effect of complexity on collaboration increases with 0.03 to 0.09 in terms of log-odds. A doubling of complexity for patents produced after 1940 is associated with a change in the odds that a patent is a collaboration by the factor of 1.06. This indicates that the positive relationship between complexity and collaboration is significantly stronger for patents produced after 1940, than patents produced before 1940. The results presented here confirm hypothesis 3.

Table 3. Results of Logistic Regression with fixed effects

Dependent variable	Is patent collaborated (0/1)?						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Complexity (log)	0.15***	0.10***	0.15***	0.10***	0.10***	0.03***	0.06***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Dummy > 1940							0.54***
							(0.01)
Complexity (log) * Dummy > 1940							0.03***
							(0.004)
Fixed Effects:							
- Technology		✓		✓		✓	✓
- City			✓	✓		✓	✓
- Year					✓	✓	
Constant	-1.23***	-1.78***	-1.38***	-1.78***	-1.09***	-1.88***	-1.95***
	(0.003)	(0.02)	(0.12)	(0.13)	(0.42)	(0.44)	(0.13)
N	1,764,733	1,764,733	1,764,733	1,764,733	1,764,733	1,764,733	1,764,733
Log Likelihood	-997,311	-976,428	-993,265	-974,236	-972,388	-960,603	-964,249
AIC	1,994,627	1,953,710	1,987,254	1,950,044	1,945,059	1,923,055	1,930,074

*p < .1; **p < .05; ***p < .01

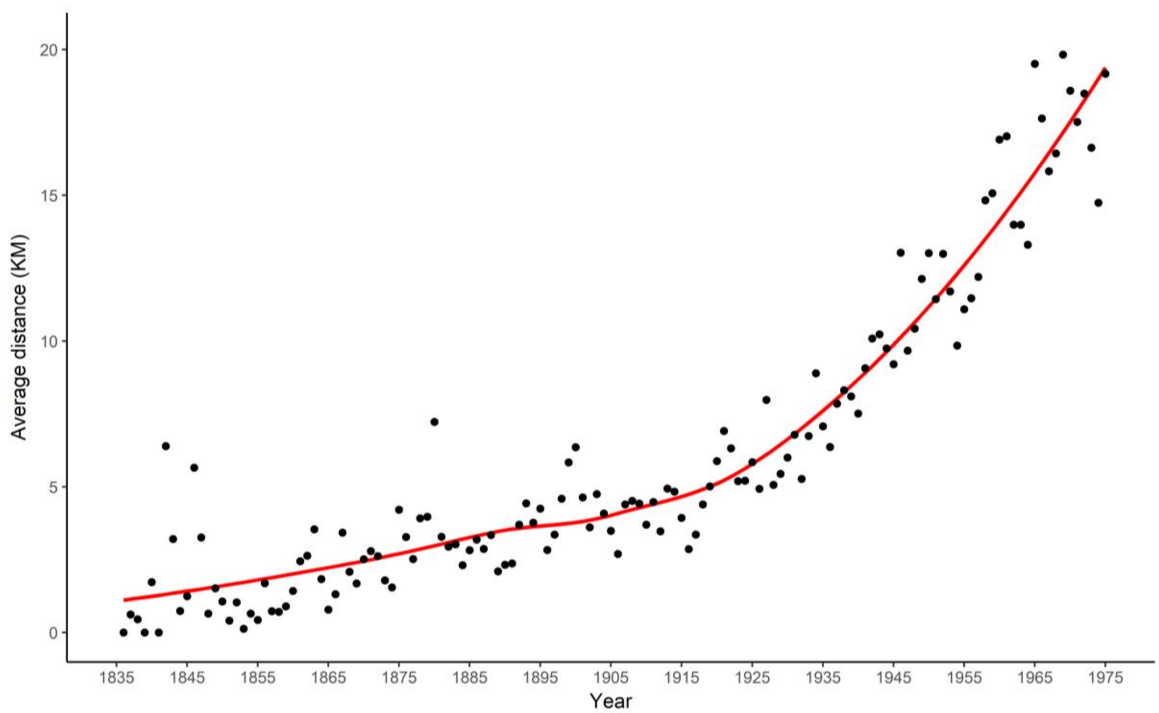
4.3 Geography of co-invention

In this section the distinct patterns to the geography of co-invention on historical US patents is described. The data used consist only of patents for which all inventors reside in an MSA.

Geographical distance of co-invention

Figure 9 reveals that the geographic distance between co-inventors on US patents has been steadily growing over time. The distance of between-city collaboration is measured as the straight-line distance in kilometers between the centers of U.S. metropolitan regions. In the 1850s, the average distance between co-inventors was below 10 kilometers. This increased to around 20 kilometers in the 1930s and to 40 kilometers in the 1960s. This increase in the distance between collaborators can result from two processes: a greater share of patents resulting from collaborations between inventors in different cities, or inventors collaborate over greater distance. I find evidence for both processes. Between-city collaboration increased from roughly 1 percent through the 19th century up to 5 percent in 1975. In addition, the average distance of collaboration involving multiple cities has doubled from about 200 kilometers in the mid-19th century to about 400 kilometers in 1975. Interestingly, the average distance of between-city co-inventors decreased slightly after the 1940s. This might have to do with the dynamics in the spatial distribution of the US population and inventors as the US expanded to the West. Such expansion might have spurred coast-to-coast collaboration in the initial years of settlement in the West, but leveled off as the population on the west-coast started to grow and provided an increasing local pool of possible (co-)inventors, limiting the need for coast-to-coast collaboration.

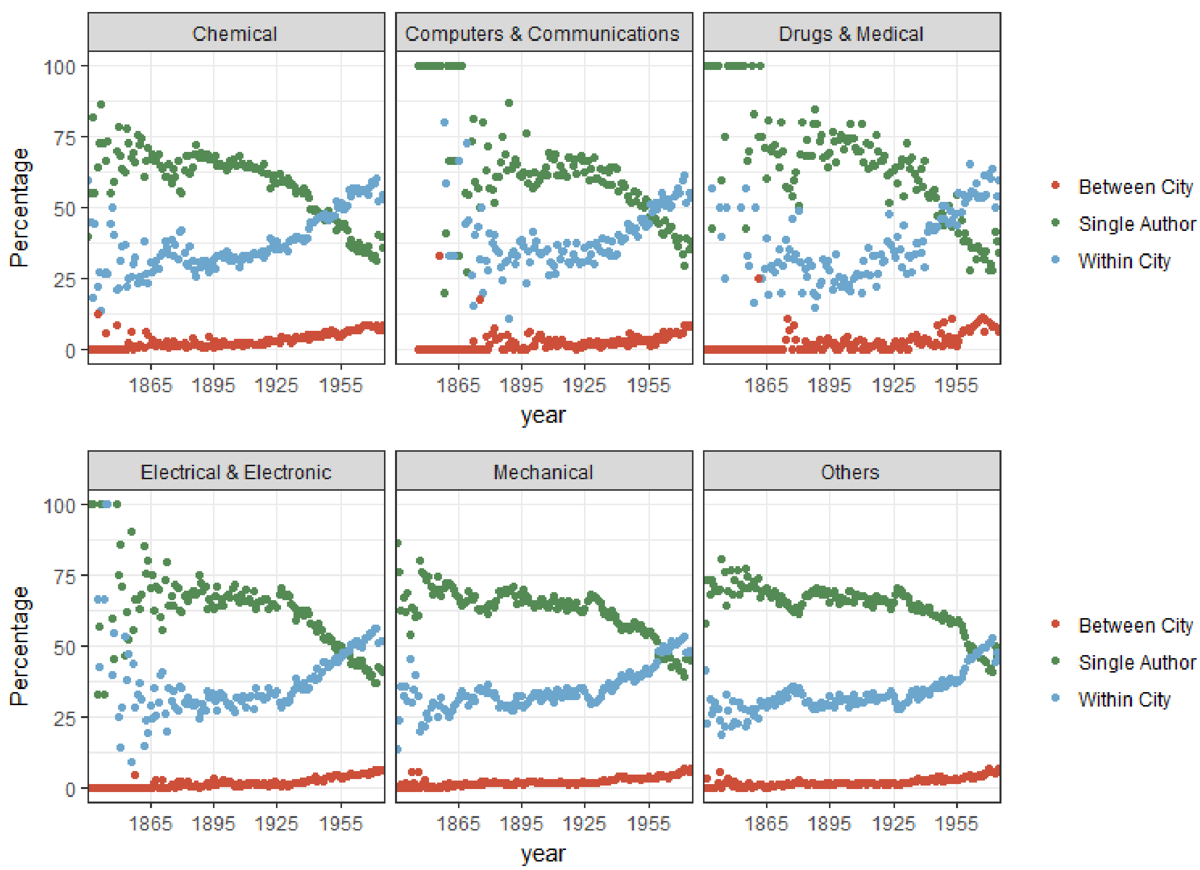
Figure 9. Average distance (KM) on patents in 1836-1975



Internal and external collaboration:

Collaboration occurs primarily by inventors from the same city. Figure 10 shows that within-city collaboration (blue dots) is also growing at a faster rate than between-city collaboration (red dots). Single authorship of patents is dropping rapidly across all categories. Although these general trends are observed across the six aggregate technology categories, patterns of collaboration differ across categories. For instance, since the 1940s the majority of patents in the 'Chemicals' category result from collaboration, whereas it takes until the 1960s for patents in the 'Mechanicals' and 'Others' category to reach this point.[11]

Figure 10. Collaboration across technology categories



[11] Note that the data in Figure 10 slightly deviates from Figure 1. The data used to construct Figure 1 includes patents with inventors who reside in non-MSA locations. These are excluded from Figure 10 because these are not metropolitan areas.

Geography, collaboration and complexity

Does complexity promote within-city collaboration rather than between-city collaboration? Although complex patents are found to result from collaboration more frequently than less complex patents, the geographical patterns of these collaborations are unknown. Collaboration on complex patents might be local, because its construction relies more heavily on repeated face-to-face interactions and trust than less complex patents do. On the other hand, inputs to complex patents might be heavily localized in pockets of specialization. Inventors collaborating on complex patents might span between-city *pipelines*, connecting different pockets of specialization.

To examine this issue, logistic regression models are estimated. The unit of analysis is the patent. Only collaborated patents are selected for which all inventors resided in an MSA. The dependent variable is whether a patent is the result of a within-city (0) or between-city (1) collaboration. The key independent variable is the *Complexity* of a patent. The second independent variable is the number of inventors on a patent. Larger teams might be more likely to be non-local. If a patent becomes complex to the point that inputs from a larger number of inventors is needed, the probability this input is not found locally might increase too. That is, increasing number of inventors on patents might raise the probability of between-city collaboration. Finally, city, year and technology fixed effects are estimated to control for specific unobserved characteristics that promote or hinder between or within-city collaboration.

To isolate the effect of complexity on collaboration, a matching procedure has been undertaken in which between-city collaborated patents are matched to within-city collaborated patents with identical scores on the number of inventors, city, year and primary technology class of the patent. Unmatched observations are dropped. The outcome is a sample with balanced covariates for within and between-city collaborated patents, except for the *Complexity* score. *Complexity* is now closer to being independent of the covariates. This makes estimates based on the parametric analyses less reliant on model specifications and modeling choices (Ho et al., 2007).

The results presented in Table 4 demonstrate a negative and significant relationship between complexity and between-city collaboration. In model 1 the estimated coefficient for *Complexity* (β_1 = -0.03) indicates that a doubling of the complexity score is associated with an expected change in the odds of between-city collaboration by a factor of 0.98. Similar results are found for models with different specifications. In model 6, controlling for *Number of Inventors* and year, city and technology fixed effects, the estimated coefficient for *Complexity* (β_1 = -0.16) suggest that a doubling of the complexity score is associated with an expected change in the odds of between-city collaboration by a factor of 0.89. The *Number of Inventors* on a patent has a positive and significant relationship with between-city collaboration in all models. In model 6, controlling for *Complexity* and year, city and technology fixed effects, the estimated coefficient for *Number of Inventors* (β_2 = 0.43) indicates that for every additional inventor on the patent, the odds of the patent being collaborated between cities increases by 1.54. These findings confirm hypothesis 4: Increasing complexity on patents is associated with within-city collaboration.

Table 4. Statistical models on within- and between-city collaboration

Dependent variable:	Patent is collaborated within-city (0) or between-city (1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Complexity (log)	-0.08***	-0.07***	-0.32***	-0.18***	-0.15***	-0.16***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Number of Inventors	0.34***	0.38***	0.36***	0.38***	0.40***	0.43***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Fixed effects:						
- City		✓				✓
- Year			✓		✓	✓
- Technology				✓	✓	✓
Constant	-2.14***	-1.80***	-1.78***	-2.29***	-2.15***	-1.69***
	(0.03)	(0.07)	(0.68)	(0.14)	(0.75)	(0.75)
N	160,484	160,484	160,484	160,484	160,484	160,484
Log Likelihood	-76,919	-75,809	-76,768	-76,177	-76,066	-75,098
AIC	153,844	151,983	153,775	153,131	153,143	151,566

$^{*}p < .1$; $^{**}p < .05$; $^{***}p < .01$

5. Conclusion

This paper examined historical U.S. inventor collaboration for the years 1836 to 1975. Recent research has shown the increasing importance of collaboration in the production of knowledge. The key underlying assumption is that the rising complexity of knowledge requires resources that exceed the inputs of individuals. Yet, little empirical evidence on the relationship between collaboration and complexity is available. The key findings are that (1) collaboration has increased sharply since the 1930s; (2) there is a positive relationship between complexity and collaboration on US patents; and (3) increasing complexity on patents is associated with within-city collaboration.

Using the raw text files of the HistPat database I utilize search, match and machine learning techniques to identify and disambiguate all (co-)inventors including their geographical location as recorded on US patents between 1836 and 1975. The resulting inventor-patent database contains about 80-90% of the patents granted by the USPTO and identifies 1,922,754 inventors, from 4,437,960 observations on 3,365,253 unique patents. These data are used to examine co-invention on U.S. patents.

The research finds evidence for a growing tendency to collaborate on US patents between 1836 and 1975. This finding is in line with other research on inventor collaboration on US patents (Wuchty et al., 2007) and scientific collaboration on papers (Jones et al., 2008) that both extend back to 1975. The observed increase in the percentage of patents collaborated, along with the growth in average team-size after the 1930s might be linked to the 1933 US Supreme Court ruling to assign the rights of technologies developed by inventors hired-to-invent directly to the firm. This provided an incentive for firms to recruit and hire inventors, instead of buying licenses to patented technologies by inventors operating privately. In doing so, the firm becomes a platform in which inventors can collaborate instead of being direct competitors. More general, infrastructure and transportation developments throughout the 19^{th} and 20^{th} century have made it easier for inventors to collaborate across increasing distance. By the 1920s almost all major US cities were connected by the railroads; in 1930 about 40% of US homes had a telephone connection and 20% of Americans had access to an automobile.

The observed positive relationship between complexity and collaboration on US patents confirms the theoretical claim that complexity and collaboration are significantly correlated. The evidence presented in this paper suggests that the impact of complexity on the odds of a patent being collaborated becomes markedly stronger after the 1940s. This supports a shift to ‘big science’ after World War II, in which the complexity of knowledge production accelerated. However, the direction of a possible causal link between complexity and collaboration remains unclear.

Co-invention on US historical patents has a distinct geography. Metropolitan collaboration is mostly between inventors from the same city. Moreover, as complexity increases, the odds of between-city collaboration decreases. This suggests that the production of complex knowledge relies more strongly on the geographical co-location of inventors than does less complex knowledge. Local collaboration allows repeated face-to-face meetings, spontaneous encounters and other interactions facilitated by the *local buzz,* than non-local collaboration. These interactions might be particularly important for the production of complex knowledge, because such knowledge relies difficult to diffuse tacit knowledge.

The paper presented here has at least two major short-comings. First, this paper has focused on inventors located in US cities. In doing so, the geography of collaboration as presented in this work is biased because it doesn’t include rural-urban and foreign collaboration. Second, the poor quality of the text files makes it difficult to retrieve all (co-)inventors and geographical locations on patents. As a consequence, about 5-20% of patents per year are missing in the database.

Future research could focus on our understanding of patterns of rural-urban and foreign collaboration on historical US patents and examine the direction of causality between complexity and collaboration. Moreover, little is known on the mechanisms that structure tie formation between co-inventors and to what extent these structures have evolved over time. In addition, much remains unclear on how the movement of inventors in space affects patterns of collaboration and impact urban networks of invention. Finally, the role of firms has been neglected and remains unclear in collaboration on historical US patents.

Literature

Acemoglu, D., Moscona, J., & Robinson, J. A. (2016). State capacity and American technology: evidence from the nineteenth century. *The American Economic Review*, *106*(5), 61–67.

Akcigit, U., Grigsby, J., & Nicholas, T. (2017). The Rise of American Ingenuity: Innovation and Inventors of the Golden Age.

Almeida, P., & Kogut, B. (1999). Localization of knowledge and the mobility of engineers in regional networks. *Management Science*, *45*(7), 905–917.

Asheim, B., & Gertler, M. (2005). The geography of innovation. In *The Oxford handbook of innovation* (pp. 291–317). Oxford: Oxford University Press.

Atack, J., Bateman, F., Haines, M., & Margo, R. A. (2010). Did railroads induce or follow economic growth? *Social Science History*, *34*(2), 171–197.

Axelrod, R. M., & Cohen, M. D. (2000). *Harnessing complexity*. Basic Books.

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

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.

Bensel, R. F. (2000). *The political economy of American industrialization, 1877-1900*. Cambridge University Press.

Broekel, T., Balland, P.-A., Burger, M., & van Oort, F. (2014). Modeling knowledge networks in economic geography: a discussion of four methods. *The Annals of Regional Science*.

Coriat, B., & Weinstein, O. (2012). Patent Regimes, Firms and the Commodification of knowledge. *Socio-Economic Review*, *10*, 267–292.

Crescenzi, R., Nathan, M., & Rodríguez-Pose, A. (2016). Do inventors talk to strangers? on proximity and collaborative knowledge creation. *Research Policy*, *45*(1), 177–194.

de Solla Price, D. J. (1970). Citation measures of hard science, soft science, technology, and nonscience. In *Communication Among Scientists and Engineers* (pp. 3–22). Lexington Books, D.C. Heath and Company.

Drews, J. (2000). Drug discovery: a historical perspective. *Science*, *287*(5460), 1960–1964.
Eaton, J. W. (1951). Social Processes of Professional Teamwork. *American Sociological Review*, *16*(5), 707–713.
Faems, D., Van Looy, B., & Debackere, K. (2005). Interorganizational Collaboration and Innovation: Toward a Portfolio Approach. *Journal of Product Innovation Management*, *22*, 238–250.
Feldman, M. P., Kogler, D. F., & Rigby, D. L. (2014). rKnowledge : The Spatial Diffusion and Adoption of rDNA Methods. *Regional Studies*, 1–21.
Fischer, C. S., & Carroll, G. R. (1988). Telephone and automobile diffusion in the United States, 1902-1937. *American Journal of Sociology*, *93*(5), 1153–1178.
Fisk, C. L. (1998). Removing the “Fuel of Interest” from the “Fire of Genius”: Law and the Employee Inventor, 1830-1930. *University of Chicago Law Review*.
Fleming, L., & Frenken, K. (2007). the Evolution of Inventor Networks in the Silicon Valley and Boston Regions. *Advances in Complex Systems*, *10*(01), 53–71.
Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: Evidence from patent data. *Research Policy*, *30*(7), 1019–1039.
Freeman, L. C. (1979). Centrality in Social Networks Conceptual Clarification. *Social Networks*, *1*(1968), 215–239.
Frenken, K., Hölzl, W., & Vor, F. de. (2005). The citation impact of research collaborations: the case of European biotechnology and applied microbiology (1988–2002). *Journal of Engineering and Technology Management*, *22*(1–2), 9–30.
Glanzel, W. (2002). Coauthorship patterns and trends in the sciences (1980-1998): A bibliometric study with implications for database indexing and search strategies. *Library Trends*, *50*(3), 461–473.
Gordon, R. J. (2016). *The rise and fall of American growth: The US standard of living since the civil war*. Princeton University Press.
Grant, R. M. (1996). Toward a Knowledge-Based Theory of the Firm. *Strategic Management Journal*, *17*(Winter Special Issue), 109–122.
Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001). The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools.
Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis*, *15*(3), 199–236.
Hoekman, J., Frenken, K., & Tijssen, R. J. W. (2010). Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Research Policy*, *39*(5), 662–673.
John, R. R. (2009). *Spreading the news: The American postal system from Franklin to Morse*. Harvard University Press.
Jones, B. F. (2009). The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder? *Review of Economic Studies*, *76*(1), 283–317.
Jones, B. F., Wuchty, S., & Uzzi, B. (2008). Multi-University Research Teams: Shifting Impact, Geography, and Stratification in Science. *Science*, *322*(5905), 1259–1262.
Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, *26*(1), 1–18.
Kauffman, S. A. (1993). *The origins of order: Self organization and selection in evolution*. Oxford university press.

Khan, B. Z. (2005). *The Democratization of Invention: patents and copyrights in American economic development, 1790-1920*. Cambridge University Press.

Khan, B. Z., & Sokoloff, K. L. (2001). History Lessons : The Early Development of Intellectual Property Institutions in the United States. *The Journal of Economic Perspectives*, *15*(3), 233–246.

Kogut, B. (2000). The Network as Knowledge: Generative Rules and the Emergence of Structure. *Strategic Management Journal*, *21*(3), 405–425.

Kogut, B., & Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science*, *3*(3), 383–397.

Lai, R., D'Amour, A., Yu, A., Sun, Y., & Fleming, L. (2012). Disambiguation and Co-authorship Networks of the U.S. Patent Inventor Database (1975 - 2010). Harvard Dataverse.

Lamoreaux, N. R., & Sokoloff, K. L. (1996). Long-term change in the organization of inventive activity. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(November), 12686–12692.

Lamoreaux, N. R., & Sokoloff, K. L. (1999). Inventors, firms, and the Market for Technology in the Late Nineteenth and Early Twentieth Centuries. In N. R. Lamoreaux, D. M. G. Raff, & P. Temin (Eds.), *Learning by Doing in Markets, Firms, and Countries* (pp. 19–60). University of Chicago Press.

Lamoreaux, N. R., & Sokoloff, K. L. (2002). *Intermediaries in the US Market for Technology, 1870-1920*. National bureau of economic research.

Lamoreaux, N. R., & Sokoloff, K. L. (2005). *The decline of the independent inventor: a schumpterian story?* National Bureau of Economic Research.

Lamoreaux, N. R., Sokoloff, K. L., & Sutthiphisal, D. (2011). The Reorganization of Inventive Activity in the United States during the Early Twentieth Century. In *Understanding Long-Run Economic Growth: Geography, Institutions, and the Knowledge Economy* (pp. 235–274).

Lamoreaux, N. R., Sokoloff, K. L., & Sutthiphisal, D. (2013). Patent Alchemy: The Market for Technology in US History. *Business History Review*, *87*(01), 3–38.

Landes, D. S. (1969). *The Unbound Promethous*. Cambridge, New York: Press Syndicate of the University of Cambridge.

Lee, S., & Bozeman, B. (2005). The Impact of Research Collaboration on Scientific Productivity. *Social Studies of Science*, *35*(5), 673–702.

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.

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, *22*(1), 3–42.

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

Maskell, P., & Malmberg, a. (1999). The Competitiveness of Firms and Regions: "Ubiquitification" and the Importance of Localized Learning. *European Urban and Regional Studies*, *6*(199901), 9–25.

Merton, R. K. (1968). The Matthew effect in science. *Science*, *159*(3810), 56–63.

Merton, R. K. (1973). The sociology of science: Theoretical and empirical investigations. University of Chicago press.

Nelson, R. R., & Winter, S. G. (1982). *An Evolutionary Theory of Economic Change*. Harvard

University Press.
Noble, D. F. (1979). *America by design: Science, technology, and the rise of corporate capitalism*. Oxford University Press, USA.
Nonnenmacher, T. (2001). History of the US Telegraph Industry. *EH. Net Encyclopedia, Edited by Robert Whaples*.
North, D. C., & Thomas, R. P. (1973). *The rise of the western world: A new economic history*. Cambridge University Press.
Novak, W. J. (2008). The Myth of the "Weak" American State. *The American Historical Review*, *113*(3), 752–772.
Patel, N. (1973). Collaboration in the Professional Growth of American Sociology. *Social Science Information*, *12*(6), 77–92.
Pavitt, K. (1984). Sectoral patterns of technical change: Towards a taxonomy and a theory. *Research Policy*, *13*(6), 343–373.
Perlman, E. R. (2016). Connecting the Periphery: Three Papers on the Developments Caused By Spreading Transportation and Information Networks in the Nineteenth Century United States.
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*.
Polanyi, M. (1958). *Personal Knowledge: Towards A Post-Critical Philosophy*. London: Routledge.
Presser, S. (1980). Collaboration and the quality of research. *Social Studies of Science*, *10*(1), 95–101.
Price, D. de S. (1963). Big science, little science. *Columbia University, New York*, 119.
Price, D. J. de S. (1986). Little Science, Big Science and Beyond. Columbia University Press.
Romer, P. M. (1986). Increasing Returns and Long-Run Growth. *Journal of Political Economy*, *94*(5), 1002–1037.
Rosenberg, N. (1982). The historiography of technical progress. In *Inside the Black Box: Technology and Economics* (pp. 3–33). Cambridge: Cambridge University Press.
Singh, J. (2005). Collaborative Networks as Determinants of Knowledge Diffusion Patterns. *Management Science*, *51*(5), 756–770.
Sokoloff, K. L. (1988). Inventive Activity in Early Industrial America: Evidence From Patent Records, 1790–1846. *The Journal of Economic History*, *48*(04), 813
Solow, R. (1957). Technical Change and the Aggregate Production Function. *The Review of Economics and Statistics*, *39*(3), 312–320.
Sorenson, O., Rivkin, J. W., & Fleming, L. (2006). Complexity, networks and knowledge flow. *Research Policy*, *35*(7), 994–1017.
Storper, M., & Venables, A. J. (2004). Buzz: Face-to-face contact and the urban economy. *Journal of Economic Geography*, *4*(4), 351–370.
Thagard, P. (1997). Collaborative knowledge. *Noûs*, *31*(2), 242–261.
Thorsteinsdottir, O. (2000). External research collaboration in two small science systems. *Scientometrics*, *49*(1), 145–160.
Tsoukas, H. (2005). *Complex knowledge: Studies in organizational epistemology*. Oxford University Press.
Uzzi, B., & Spiro, J. (2005). Collaboration and Creativity: The Small World Problem. *American Journal of Sociology*, *111*(2), 447–504.
van Oort, F. G., & Lambooy, J. G. (2014). Cities, knowledge, and innovation. In *Handbook of*

Regional Science (pp. 475–488). Springer.
Ventura, S. L., Nugent, R., & Fuchs, E. R. H. (2015). Seeing the non-stars: (Some) sources of bias in past disambiguation approaches and a new public tool leveraging labeled records. *Research Policy*, *44*(9), 1672–1701.
Von Hippel, E. (1994). “Sticky information” and the locus of problem solving: implications for innovation. *Management Science*, *40*(4), 429–439.
Winter, S. (1998). Knowledge and competence as strategic assets. In *The strategic Management of Intellectual Capital*. Butterworth-Heinemann.
Wray, K. B. (2002). The Epistemic Significance of Collaborative Research. *Philosophy of Science*, *69*(1), 150–168.
Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036–9.
Zollo, M., & Winter, S. G. (2002). Deliberate learning and the evolution of dynamic capabilities. *Organization Science*, *13*(3), 339–351.

Chapter 2: What Mechanisms Structure Tie-Formation among U.S. Inventors: Empirical Evidence from U.S. Patents between 1836-1975

Abstract

What mechanisms structure US inventor collaboration? Scholars have emphasized the role of collaboration and increasing team-sizes in the production of knowledge. Although the reasons to collaborate are many and well understood, long-run structural empirical evidence on the mechanisms that structure tie-formation amongst inventors is lacking. This paper investigates this process by examining collaborative tie formation between inventors in US metropolitan regions between 1836 and 1975. This long-run era covers the 'golden age of US invention' and captures the development and introduction of numerous transport and tele-communication technologies that facilitate collaboration and at greater distance. Within this light of rapidly changing technologies, this paper explores to what extent geographical, social and technological distance influence tie-formation between US inventors and how these have evolved over time. Exponential random graph models are used to statistically estimate the impact of these forms of distance on tie-formation. The three key findings are that (1) geographical distance negatively impacts collaboration, that this effect has been decreasing over time but has levelled off in the 20th century; (2) technological proximity between inventors' patent portfolio positively impact collaboration, especially in times of uncertainty; (3) social proximity promotes collaboration, but only to a certain extend. Together these findings contribute to the existing literature by identifying a set of key mechanisms that structure collaborative knowledge production. These results are of relevance for policymakers constructing innovation policies and corporate executives concerned with innovation because they highlight the complexity of collaborative knowledge production.

Introduction

Scholars across different disciplines have reported the increase in collaboration in the production of knowledge. Most scholars agree that the shift towards science-based knowledge production increased the complexity of knowledge production. To produce new knowledge, the increasing complexity requires inputs that exceed the resources of the individual. Some have labeled this the 'burden of knowledge' and argue it is increasingly becoming more difficult to produce innovations – especially for individuals (Jones, 2009). Empirical results show that knowledge production by scientist and inventors has become gradually more dominated by teams since the 1960s (Wuchty, Jones, & Uzzi, 2007). Using long-run US inventor-patent data, Van der Wouden (2018) finds that the average team-size and the percentage of collaborated US patents has increased over time, especially since the 1940s.

There are numerous reasons for knowledge producers to collaborate. They might collaborate to optimize resources (Eaton, 1951), increase productivity (Lee & Bozeman, 2005; Melin, 2000; D. J. de S. Price, 1986) or creativity (Uzzi & Spiro, 2005), access complementary ideas (Price, 1970) or resources (Katz & Martin, 1997; Wray, 2002), or for intellectual and social reasons (Edge, 1979; Stokes & Hartley, 1989; Thorsteinsdottir, 2000). Collaborative work pays off: it is more frequently cited citations than individual work (Fox, 1991; Frenken, Hölzl, & Vor, 2005; Glanzel, 2002; Katz & Martin, 1997; Lindsey, 1978), has higher acceptance rates for academic publication (Presser, 1980), and enables researchers to engage with large research questions (Thagard, 1997). On the firm level, Faems et al. (2005) finds that collaborating firms are more likely to produce commercially successful products than non-collaborating firms. A critical review of R&D collaboration is provided by Katz & Martin (1997).

Although these reasons to collaborate are relatively well documented and understood, our understanding of tie-formation processes in collaborative knowledge production remains incomplete. Thus far, research on tie-formation among knowledge producers has largely focused on the firm-, region- or nation-state level, ignoring the level of the individual (a notable exception is Crescenzi, Nathan, & Rodríguez-Pose, 2016). Moreover, the limited research on collaboration between individual knowledge producers has concentrated on the impact of individuals' attributes on tie-formation, ignoring relational, dyadic and triadic processes. Consequently, little structural empirical evidence is available on the dyad-level mechanisms governing actual tie-formation among individual knowledge producers. This is surprising, because there is a vast social sciences literature that stresses the importance of dyad-level processes on tie-formation (Borgatti, Mehra, Brass, & Labianca, 2009; Oswald, Clark, & Kelly, 2004).

This paper uses long-run US inventor data between 1836 and 1975 to examine the impact of three popular dyad-level mechanisms – geographical, social and technological distance – on tie-formation between individual inventors. Exponential random graph models (ERGMs) are used to estimate the influence of these three measures of proximity on the probability of a tie-formation between inventors and how this influence has changed over time. This method has a main advantage over conventional statistical methods because ERGMs are designed to deal with the interdependence of observations in relational data and thus produces more accurate estimates on the dyad-level mechanisms modelled than conventional statistical models.

The results of this paper are the first long-run systematical empirical evidence on the mechanisms that foster and hinder tie-formation between inventors on US patents between 1836 and 1975. The key findings are that (1) geographical distance negatively impacts collaboration, that this effect has been decreasing over time but has levelled off in the 20th century; (2) technological proximity between inventors' patent portfolio positively impact collaboration, especially in times of uncertainty; and (3) social proximity promotes collaboration, but only to a certain extent. These results have important implications for understanding knowledge production, because they stress the complexity of innovation dynamics.

The remainder of this paper is structured as follows. The next section provides the theoretical motivation of this research and identifies three hypotheses derived from the existing literature on collaboration. Section three discusses the data and introduces the exponential random graph models. The descriptive and statistical results are presented in section four. The last section provides the discussion of the findings.

2. Tie formation amongst inventors on US patents

Scholars have reported a recent rise in collaboration among knowledge producers (Crescenzi et al., 2016; Hoekman, Frenken, & Tijssen, 2010; Jones, Wuchty, & Uzzi, 2008; Wuchty, Jones, & Uzzi, 2007). Van der Wouden (2018) shows that collaboration among US inventors has been increasing rapidly since the 1930s. These works relate to earlier research on the diminishing role of the 'individual genius' (Merton, 1968) or 'lone-wolf' (Patel, 1973), that has been connected to the general shift in knowledge production towards complex science-based knowledge (Noble, 1979; Pavitt, 1998; D. de S. Price, 1963). The complex nature of science-based knowledge requires inputs and resources that often exceed that of the individual and hence promotes collaboration. Jones (2009) argues that the increasing complexity weighs knowledge production down, because increasingly more effort and resources are needed to produce novelty.

Knowledge producers have numerous reasons and benefits to collaborate. Collaborating individuals have access to additional resources (Katz & Martin, 1997; Wray, 2002), (complementary) ideas (de Solla Price, 1970), and allows individuals to take on more elaborate research questions (Thagard, 1997). Others have argued that individuals collaborate to optimize resources (Eaton, 1951), for intellectual or social reasons (Edge, 1979; Stokes & Hartley, 1989; Thorsteinsdottir, 2000), or to increase their creativity and productivity (Lee & Bozeman, 2005; Melin, 2000; D. J. de S. Price, 1986; Uzzi & Spiro, 2005). The benefits from collaboration are tangible for individuals and firms. Collaborated academic articles tend to have higher acceptance rates (Presser, 1980) and are cited more frequently (Fox, 1991; Frenken, Hölzl, & Vor, 2005; Katz & Martin, 1997; Lindsey, 1978). Firms engaged in collaboration benefit from greater learning performance and increased innovativeness (Ahuja, 2000), are more likely to produce commercial successful products (Faems, Van Looy, & Debackere, 2005), tend to survive longer (Mitchell & Singh, 1996; Singh & Mitchell, 1996) and grow faster (Powell, Koput, & Smith-Doerr, 1996), and have privileged access to potential collaborators (Stuart, 1998). A critical review is provided by Katz & Martin (1997).

While the reasons and benefits of collaboration received much research attention, the structures that promote collaborative tie-formation among knowledge producers remain a black box. Over the last two decades a set of literature has emerged that stresses the impact of different forms of proximity between knowledge producers on collaboration (Bathelt &

Gluckler, 2003; Bathelt, Malmberg, & Maskell, 2004; Bercovitz & Feldman, 2011; Boschma, 2005; Breschi & Lissoni, 2009; Cassi & Plunket, 2014, 2015; Crescenzi et al., 2016; Hoekman et al., 2010; Maggioni, Nosvelli, & Uberti, 2007; Storper & Venables, 2004). In this literature, it is argued that geographical proximity facilitates collaboration because it lowers the costs of interactions. These costly repeated interactions are required to build the trust that, in turn, is fundamental for knowledge sharing. In addition, when knowledge-sharing occurs, being close in geographical space eases the processes of coordination.

Hypothesis 1: Geographical distance hinders tie-formation between inventors

Social proximity also promotes collaboration. Agents with a social relationship are more likely to trust each other than agents without a relationship (Coleman, 1990; M. Granovetter, 1985; Uzzi, 1996). Agents might also trust others with which they are indirectly connected. For instance, two unconnected agents with a mutual collaborator are more likely to trust and connect with each other than two unconnected agents without such common connection. This behavior is described as triadic closure and is found to influence tie-formation across different contexts (Burt, 1992; Granovetter, 1973; Harary, 1955; Martin, Ball, Karrer, & Newman, 2013; Rapoport, 1953; Shi, Foster, & Evans, 2015; Ter Wal, 2013). In such case, agents infer the trust from the relationship their collaborator has with the agent. From this perspective, well-connected agents in a social network are proximate to a greater number of agents than less well-connected agents and have access to a greater number of potential collaborators. However, it is important to note here that geographical and social proximity are often entangled. The social interactions of agents tend to be geographically localized, making it difficult to unravel and allocate the effects of both proximities on collaboration and knowledge production (see Boschma, 2005 and Malmberg & Maskell, 2006)

Hypothesis 2: Social proximity promotes tie-formation between inventors

Likewise, technological or cognitive proximity is argued to foster collaboration. Agents are more likely to collaborate if they have similar sets of knowledge because “they speak the same language”. This overlap promotes communication and thus facilitates knowledge sharing and learning (Brown & Duguid, 2002; Jaffe, 1989). However, too much overlaps leaves little opportunities for learning. For effective collaboration agents require some degree of similarity and variation in their knowledge sets (Broekel & Boschma, 2012; Nooteboom, 2000).

Hypothesis 3: Technological proximity promotes tie-formation between inventors

During the time-window of this study major developments in telecommunication and transportation technologies have made it significantly easier to collaborate and at greater distance. A clear example is the development of the telegraph. From 1851 onwards telegraph wires were installed along railroad tracks. Whereas it still took roughly two days to travel between New York and Chicago in 1860, sending a telegraph between the cities was nearly simultaneously (Nonnenmacher, 2001). Railroad tracks connected most of the US cities by the end of the 19th century, connecting markets and promoting economic activities (Atack, Bateman, Haines, & Margo, 2010) and stimulate inventive activities (Perlman, 2016). After the railroad and telegraph, new technologies were introduced in the early 19th century. By the 1930s roughly

20% of US citizens had access to an automobile and 40% of the US homes had a telephone connection (Fischer & Carroll, 1988). Finally, commercial air-travel took off in the 1920s, connecting US cities at great geographical distance. All these developments allow US citizens to connect, communicate and collaborate with more ease than ever before.

Changes in the US patent laws have arguably impacted trends in US inventor collaboration. The original 1836 US patent system only granted patents to individual men or women and excluded firms to patent (Coriat & Weinstein, 2012). Individual inventors exploited their patents by selling or licensing the rights to the patent to firms. This market for technologies made inventors competitors and provided incentives for inventors to specialize in certain technologies. Firms acquired these new technologies to enhance their competitive position (N R Lamoreaux & Sokoloff, 1999; Naomi R Lamoreaux & Sokoloff, 2005). In 19th and 20th century the US economy was transforming from entrepreneurial capitalism to corporate capitalism (Coriat & Weinstein, 2012; Freeman, 1979). The production and sole appropriation of technological knowledge became increasingly important for the competitiveness of firms (Fisk, 1998; Noble, 1979; Pavitt, 1984), but was hindered by inventors who could license their patents to multiple firms, undermining the competitive advantage of firms. Although firms increasingly started hiring inventors to incorporate the development of technologies in-house, to the dismay of firms the legal rights to these technologies still belonged to individual inventors. In 1933 the US Supreme Court responded with the ruling that 'the respective rights and obligations of employer and employee, touching an invention conceived by the latter, spring from the contract of employment' (in Coriat & Weinstein, 2012). Thus, the rights to the technologies developed by employees 'hired-to-invent' belong to the firm (Coriat & Weinstein, 2012; Fisk, 1998).

Effectively, this ruling made an increasing number of corporate inventor collaborators instead of competitors.

It is important to note that geographical distance is, unlike social and technological proximity, posited here as an exogenous variable. That is, geographical distance is seen as given. While this perspective has several computational advantages (i.e. a fixed distance), geographical distance is produced by socio-economic processes in society. Of course, cities connected by a direct highway or railway are more easily accessible than when such transportation infrastructure didn't exist. Socio-economic processes, put forward in political and economic power, ultimately decide which places become connected. The *relative* geographical distance between connected places decreases with respect to other places. This note is especially of relevance in early time-periods of this study when the connectivity between U.S. cities was rapidly developing over time[12].

3. Data and methods

3.1 Data

The empirical analysis of this paper uses the historical US inventor-patent collaboration data introduced and described in Van der Wouden (2018). This data-set is an extension to the HistPat database (Petralia et al. 2016). This latter database contains technical and geographical information for historical patents provided by the USPTO between 1790 and 1975. The authors of HistPat scraped the text from digitized historical patent files made available by Google Patents and EspaceNet, and recorded the first inventor and a geographical location. However, the HistPat database holds no information for any possible additional co-inventor(s) and their geographical location(s). Van der Wouden (2018) used a series of complex searching, matching and machine-

[12] In the analysis presented below, I've used railroad data (Atack, 2016) to proxy for changing connectivity over time, but did not change the observed results.

learning algorithms to identify and disambiguate all US inventors and their geographical locations on historical US patents between 1836 and 1975. Merging the new inventor data (and their geographical location) with the technical information available in HistPat, a rich inventor-patent data becomes available.

Table 3.1 provides a simplified and semi-hypothetical example on how the raw information is structured. The unit of observation is the inventor-patent pair. The first patent has only one inventor – John Ruggles from Thomaston, Maine. This patent is classified in technology class 12. Up to five technology classes are recorded for each patent. The third patent, granted in 1853, is invented by two inventors. Benjamin Irving lived in Greenpoint, New York, and Josephine Bruin lived in Los Angeles, California. Patent three is thus a collaboration by two inventors living in different cities. The unique inventor ID allows us to follow inventors through time and space. For instance, two years later Josephine Bruin patents again but this time from Greenpoint, New York. In 1856 Joe and Josephine Bruin collaborate on patent five. The variables marked with an asterisk originate from the HistPat database (Petralia et al. 2016).

Table 3.1 Simplified and semi-hypothetical example of structure inventor-patent data.

Patent*	Year*	ID	Inventor	City	County	State	$Techn_n$*	…
1	1836	1	JOHN RUGGLES	THOMASTON	KNOX	MAINE	12	…
2	1838	2	FOWLER M RAY	CATSKILL		NEW YORK	33	…
3	1853	3	BENJAMIN IRVING	GREENPOINT	KINGS	NEW YORK	19	…
3	1853	4	JOSEPHINE BRUIN	LOS ANGELES		CALIFORNIA	19	…
4	1855	4	JOSEPHINE BRUIN	GREENPOINT		NEW YORK	19	…
5	1856	4	JOSEPHINE BRUIN	LOS ANGELES		CALIFORNIA	20	…
5	1856	5	JOE BRUIN	LOS ANGELES		CALIFORNIA	20	…

The inventor-patent data introduced by Van der Wouden (2018) is used to construct an inventor collaboration network in which the *nodes* are individual inventors. If two or more inventors collaborate on the same patent they are linked with an *edge*. Both nodes and edges have a number of attributes – that is, associated information. Node attributes are *inventor ID*, *inventor name*, *number of previously granted patents* and *geographical location*. Edge attributes are *patent number, technology classes* and *year granted*[13]. Note that if an inventor occurs on multiple patents (edges), this inventor can have different geographical locations due to movement in geographical space. Figure 3.1 represents the network of the inventor-data in table 3.1. The red dots indicate the inventors as nodes. The grey curved edges between the inventors indicate the patent these inventors collaborated on. If a patent is not collaborated, the corresponding edge is a self-loop. These patents are not plotted in graphs and are not incorporated in the statistical analysis outline below because they are not collaborations.

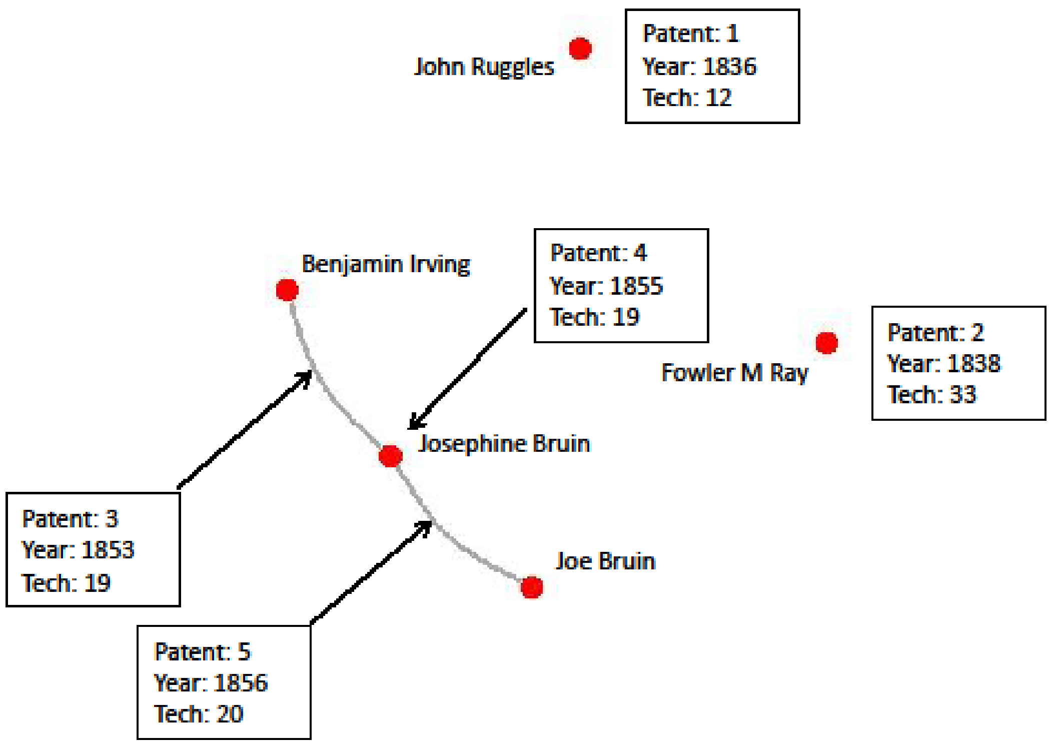
The total number of inventor-patent observations in the database is 4,446,008. A little more than 2 million observations are dropped because (1) no correct geographical location is available because the patent text was unreadable; (2) the inventor was not living in the USA; or (3) the inventor was not living in a Metropolitan Statistical Area[14]. The number of inventor-patent observations used in the analysis is 2,350,248 and consists of 1,057,805 unique inventors and 1,780,118 patents. These patents are granted to inventors living in 360 different cities and across 140 years. The distribution of inventors across cities is highly uneven: most cities only host a handful of inventors, whereas only a handful of cities host thousands of inventors. Similarly, the distribution of inventors is uneven across time – there are 60 unique inventors in 1836, but 34,065 in 1970. For a detailed description of the data see Van der Wouden (2018).

---

[13] Note that these edge attributes are directly derived from HistPat.

[14] Van der Wouden (2018) describes that the combination of city, county and state is used to assign a 2008 CBSA code to the inventor-patent observation.

Figure 3.1 Network representation of example data in table 3.1



3.2 Method, variables and statistics

3.2.1 Method

In this article, I examine the processes that give rise to collaborative ties between inventors in US metropolitan regions between 1836 and 1975. The structure of the ties in the population of inventors originates from both the distribution of individual attributes and the dynamics of interactions. For example, the tendency of inventor collaboration relies on the number of possible inventors and the aptitude to collaborate. The ability of inventors to collaborate with multiple inventors gives rise to complex social network structures in which collaborative ties may be interdependent. A classic example tells that friends of friends are much more likely to become friends than one would expect at random. This process is often referred to as *triadic closure* (Rapoport, 1953) and is observed across a wide range of social networks (for instance, Ter Wal, 2013). Importantly, this dependency between observations renders the use of most statistical approaches inappropriate because it violates the assumption of independent observation. Moreover, population-level structures can act as a platform over which knowledge might flow that can shape and/or reshape individual-level interactions that, in turn, might influence population-level structures and characteristics (Cartwright & Harary, 1956; Davis, 1970; Goodreau, Kitts, & Morris, 2009; Kohler, Behrman, & Watkins, 2001).

Thus, social networks are complex structures that arise from mechanisms that operate at multiple levels simultaneously and in which observations are interdependent. The recently developed exponential random graph models (ERGM) can handle these types of data, allowing researchers to make statistical inference from such data. This group of models has become popular in social sciences for at least three reasons. First, ERGMs allows for interdependency among observations. Tie-formation is seen as an endogenous process that involves dyadic dependence. Second, ERG models estimate effects across individual, dyadic, triadic and network levels simultaneously and can help researchers distinguish the effects of multiple micro-level dynamics. Third, ERG models allow researchers to test network theories and model the structural properties of networks. To do so, these models rely heavily on probability theory and stochastic procedures for their statistical inference. The observed network is generally regarded as an outcome from a set of possible networks with similar characteristics - the stochastic process. Because the information on this stochastic process is unknown, the goal in constructing the model is to formulate and test hypotheses that structure this process. Indeed, the objective of ERG modelling is to find a model for the structural characteristics of a network that maximizes the likelihood of an observed network being constructed (Robins, Pattison, Kalish, & Lusher, 2007; Snijders, 2011).

Exponential random graph models specify the probability of a random network ***Y*** with a set of *n* nodes and their attributes as

$$P(Y = y|n \text{ inventors}) = \left(\frac{1}{k}\right) \exp\left\{\sum_{S} \eta_s g_s(y)\right\}$$

in which P(Y = y) represents the probability that the structural characteristics of network Y constructed in the exponential random graph process are identical to the structural characteristics of the observed network (y). Y_{ij} represents the tie between inventors *i* and *j*, and takes the value of 1 if a tie is present and 0 otherwise. Thus, network Y can be represented as a matrix of *n* x *n* in which each element indicates whether or not a tie between inventors is present. The network statistic(s) that represents the structural characteristic(s) of the network is denoted as $g_s(y)$. η_s refers to the parameter corresponding to network configuration *S*. If the score on this parameter of the network statistic is the same as in the observed network, then $g_s(y) = 1$, otherwise it is 0. Finally, *k* is a normalizing constant to generate a proper probability distribution (Robins et al, 2007).

Following Goodreau et al. (2009), the equation above can be rewritten as the conditional log-odds of a single tie between inventor *i* and *j* as:

$$logit\ \left(P\left(Y_{ij} = 1 \mid n \text{ inventors}, Y_{ij}^{k}\right)\right) = \sum_{s=1}^{S} \eta_s\ \delta g_s(Y)$$

in which Y_{ij}^{k} indicates all dyads other than Y_{ij}. The δ term represents the extent to which the network statistic(s) changes when a tie is present or not between inventor i and j. The η vector can now be interpreted as the increase in log-odds of a tie forming if the forming of that tie increases g_s by one unit.

In a proposed model with a set of statistics (g), the η vector maximizing model likelihood can be approximated using Markov chain Monte Carlo simulation techniques (Hunter et al. 2008). These techniques use stochastic simulation processes to produce a distribution of random graphs using an initial set of parameter values. The generated random graphs are then compared against the real-world observed graph. The values of the set of parameter are continuously being refined to come as close to the characteristics of the observed graph. The parameter estimates stabilize when additional refinement of the estimates under the proposed models (with a set of g statistics) is impossible (Snijders, 2001, 2002). If the specified model is unlikely to proximate the observed graph there is model degeneracy and the model fails to converge or produce very poor fits to the data (Handcock, 2003). To assess a model's goodness-of-fit, the structure of the simulated graphs can be compared to the structure of the observed graph (Hunter et al., 2008a).

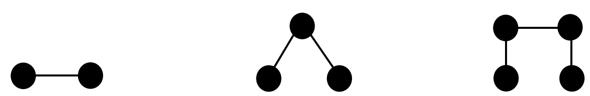
3.2.2 Proximity variables

In this paper three key mechanisms often proposed to affect tie-formation and are examined in the context of US inventor collaboration. These mechanisms (geographical, social and technological proximity) are exogenous dyad level covariates and are constructed for each of the 140 years in the analysis and for any possible pair of inventors, regardless of whether they collaborate. For example, if there are 100 inventors in year t, matrix m of [100 x 100] is constructed in which the 10,000 elements are filled with the value of the measure that operationalizes the proposed mechanism of tie-formation.

The three mechanisms are operationalized as follows. First, the *geographical proximity* between any possible pair of inventors in a given year is calculated by taking the absolute distance in kilometers between the centers of the US metropolitan regions each inventor lives in. If inventors live in the same city, the geographical distance between them is coded as zero.

Second, *social proximity* between any possible pair of inventors in a given year is calculated by retrieving for each inventor his/her previous collaborators (1-order), the collaborators of these collaborators (2-order) and their collaborators (3-order) from the previous ten years (that is, *t-1 : t-10*). This measure indicates which inventors were in the social network of an individual inventor in the ten years prior to the patent(s) in the focal year. In this case, only the first-, second- and third-order *neighborhood* of an inventor is recorded, but one could extend this to any *n*-order up to the maximum path-length of a network. Here the arbitrary cut-off point is the third neighborhood order because neighborhoods of growing order are increasingly less likely to impact tie-formation. Figure 3.2 shows a network representation of the three orders of neighborhoods. For each of the three orders a square matrix is built that holds information whether any pair of inventors are connected (0/1) by a minimum of that order. This means that the inventor-pairs that can be reached with an order of one or two are not recorded in the third-order matrix. Making this distinction allows to disentangle the effects of different orders of neighborhood connections on tie-formation. One could also extend the time-period further back in time. In this context, it seems unlikely that social connections from more than 10 years ago promote tie-formation in the focal year. Note that the social proximity measure is operationalized in three separate variables.

Figure 3.2 Example of 1-order, 2-order and 3-order connections



Third, *technological proximity* between any pair of inventors is based on the similarity of the technology classes in their patent portfolios of the previous 10 years. The technological proximity of a pair of technology classes is measured using co-classification on patents. Pairs of technologies that often co-occur on patents are assumed to be more proximate than technology classes that do not co-occur on patents. These pair-wise technological proximity scores are calculated annually by examining the patents produced in the focal year. These annual proximity scores are used to calculate the average technological proximity of a pair of inventors' patent portfolio. Note that the proximity of the portfolios is calculated using the pair-wise technological proximity scores in the focal year and not the years in which the patents in the portfolio are produced. For a detailed description of the calculations see Van der Wouden & Rigby (2018).

Table 3.2 and 3.3 illustrate the calculation. In year *t* Josephine and Joe are patenting (either collaborative or autonomous). In the previous ten years, Josephine has patented in class 1 and 3, whereas Joe patented in class 2 and 4. Technology classes 1 and 2 seldom co-occurred on a patent in year *t*, so the technology proximity between these classes is relative low. Technology classes 3 and 2 co-occurred frequently, resulting in a higher proximity score. The average technological proximity between Josephine's and Joe's patent portfolio is 2.01.

Table 3.2 Technological portfolio of inventors Josephine and Joe

	Josephine	Joe
Technology	1	2
Technology	3	4

Table 3.3 Pair-wise technological proximity between technologies of inventor portfolios

Technologies Josephine	Technologies Joe	Proximity
1	2	0.01
1	4	1.52
3	2	4.3
3	4	2.2
Average proximity		2.01

3.2.3 Additional model statistics

Next to the three key variables introduced above, three other statistics are used to specify the ERGM for the annual inventor collaboration networks between 1836 and 1975.

First, a network statistic for tie-density is calculated. This term, labelled '*edges*', is the total number of collaborations between inventors in the network. It represents the density of the network and the estimated coefficient can be interpreted as the log-odds of a tie between any random two inventors in the network, ceteris paribus.

Second, a term is added to account for differences in the number of patents inventors have produced in previous years. It is likely that inventors with large number of patents produced are more preferable collaborators than inventors with smaller number of patents produced. Thus, the probability of a collaboration might be a function of the inventor's number of patents. This term, labelled '*Node covariate – Patents*', is equal to the sum of the number of patents of inventors that have collaborated together. The estimated coefficient should be interpreted as the change in log-odds of a tie between any two inventors caused by the increase of one patent in the patent portfolio of the two inventors.

Third, a statistic called the *geometrically weighted degree distribution* (GWDEGREE) is added (Hunter, 2007). This term estimates the change in tie-formation given the number of connected inventors and thus helps to model the degree distribution in the network. *Geometrically weighted* refers to the fact that the effect of any additionally connected inventor is discounted by scale parameter α. This discount is necessary to prevent model degeneracy in which the model produces unrealistic networks (Handcock, 2003). For instance, if collaborating with one inventor renders an additional tie 30% more likely, collaborating with four inventors makes an additional tie 120% more likely. These values become quickly unrealistic, so marginally decreasing the weight of an additional collaborator is preferable (see Goodreau et al. (2009) for an intuitive explanation).

The final model that is used to model the 140 networks is as follows:

Collaboration (0/1) = f(*edges, geographical distance, technological proximity,*
social proximity (order 1), social proximity (order 2),
social proximity (order 3), number of patents in previous years,
geometrically weighted degree distribution)

4. Empirical results

4.1 Descriptive statistics of networks

In Table 4.1 the descriptive statistics show how the inventor networks have changed on ten characteristics between 1840 and 1970.

In general terms the networks have strongly increased in size. The number of inventors increased from 251 to 34,060 between 1840 and 1970. The growth in the number of collaboration is even stronger. It has grown from just 58 collaborations in 1840 to 28,709 collaborations in 1970. This means that the inventor-to-collaboration ratio changed from roughly 4 in early years to close to 1 in last years. This indicates a trend towards more collaboration on US patents. Note that a ratio of close to one does not mean that each inventor collaborates, because the count of collaboration is inflated by increasing team-sizes over time. For instance, a patent with 5 inventors results in 10 collaborations. For a more detail description on the evolution on collaboration on US patents between 1836 and 1975, see Van der Wouden (2018).

The density of the networks dropped over time. This makes sense because the density is measured as the observed collaborations divided by the possible number of collaborations. Adding an additional inventor raises the number of possible collaboration faster than the number of collaboration an average inventor adds to the network. Networks with extremely low density indicate that the overall population of inventors is very poorly connected. Indeed, it is very unlikely that the inventor of the "rocking bath-tub" (patent US643094A, 1900) collaborates with the inventor of the "chicken eye protectors" (patent US730918A, 1903[15]).

Over time the average geographical distance between all pairs of inventors, regardless of whether they collaborate or not, has increased over time. The average geographical distance between all inventors was about 370 kilometers in 1840, but reached roughly 1300 kilometers in 1970. This trend reflects the changing spatial distribution of innovative activity in the US. An increasing number of cities is producing inventors. In the 19th century most of the US economic and innovative activities took place in the northeastern part of the country. But throughout the 20th century the expansion and rapid settlement of the West was a fact. For instance, in 1880 the estimated population of California was around 800,000 people. In 1910 there were already close to 2.4 million Californians. This means that the West now also started to increasingly produce inventors, raising the average geographical distance.

[15] Both examples originate from http://www.wipo.int/patents/en/historical_patents.html

The average technological proximity between the patent portfolios of inventors has been trending upwards until the 1900s. In the 20th century the average technological proximity between inventors remains rather stable. An increase in proximity means the patent portfolio of US inventors are filled with patents that hold technology classes more proximate to the technology classes on patents of other inventors. Three developments might explain part of the trend. First, Lamoreaux & Sokoloff (1996) report that by the late 19th century the ability to produce patents and license the rights of the patents to firms promoted specialization in invention and facilitated the growth of 'career' inventors. This increasing group of inventors specialized in the production of certain technologies and thus created portfolios with rather low technological proximity to the portfolio of other specialized inventors. Second, by the 1930s inventor collaboration starts to take-off (Van der Wouden, 2018). If collaboration brings together inventors specialized in different technologies and produce novel cross-technological patents, the patent portfolios of both inventors will become more technologically diversified and proximate. Third, the introduction and spread of general purpose technologies (GPTs) throughout patent portfolios in the late 19th century might have made the portfolios of inventors more technologically proximate[16]. That is, when an increasing number of patents rely, make use or build upon these technologies of great applicability, the technology classes of the GPTs will be listed on many patents and reflected in an inventor's portfolio.

The total social proximity between inventors in collaboration networks also has been steadily growing over time. Up to 1870 inventors didn't have any first, second or third order connections made in the previous ten years with the other inventors that year. The inventors in 1970 together had 3,204 first order connections with other 1970 inventors. This increase results from three developments: (1) an increasing proportion of the inventors are repeating inventors

[16] Examples of GPTs in the late 19th century are electricity and the internal combustion engine.

(as witnessed by the increasing average number of previous patents); (2) inventors collaborate more over time; and (3) average team-size of patents is increasing over time (Van der Wouden, 2018). The connections made on earlier patents are counted in the social networks of subsequent years if the collaborating inventor(s) of these earlier patents also patent in the focal year.

Table 4.1 Descriptive statistics of annual networks in periods of 10 years

					(1)	(2)	(3)	(4)	(5)	(6)
Year	Inventors	Collaborations	Density	Cities	Average Geographical Distance	Average Techn. Proximity	Sum Previous Collaborators in Network	Sum Collaborators of Previous Collaborators in Network	Sum Collaborators of Previous Collaborators' Collaborators in Network	Average Patents Previous Years
1840	251	58	0.0018	46	370.5	0.01	0	0	0	0.08
1850	461	85	0.0008	55	485.2	0.35	0	0	0	0.28
1860	2,181	439	0.0002	126	635.3	0.19	0	0	0	0.25
1870	5,024	1,283	0.0001	172	757.1	0.06	12	8	6	0.34
1880	6,168	1,876	0.00009	193	862.9	0.82	176	152	122	1.11
1890	11,457	3,052	0.00005	242	1009.5	0.70	334	257	154	1.47
1900	11,652	3,415	0.00005	262	1085.4	1.86	355	396	372	2.53
1910	15,772	4,732	0.00004	278	1231.3	1.43	892	1,213	1,592	3.45
1920	19,135	6,157	0.00004	282	1281.5	1.53	933	1,179	1,588	2.16
1930	16,136	5,464	0.00004	260	1272.1	1.78	1,527	1,838	2,223	4.11
1940	17,354	7,741	0.00005	262	1154.9	1.60	1,869	2,609	3,830	4.59
1950	21,854	13,382	0.00006	270	1231.5	0.92	2,416	3,364	4,940	3.79
1960	27,153	19,367	0.00005	288	1272.7	1.79	2,707	2,327	2,815	2.28
1970	34,060	28,709	0.00005	308	1329.7	1.58	3,234	4,331	8,057	1.84

4.2 Statistical results

This section presents the empirical findings of the estimation of annual exponential random graph models. The results of the models between 1836 and 1840 are excluded from the figures because the impact of geographical distance, technological and social proximity couldn't be properly estimated due to very limited number of non-zero values. The black dots in the figures correspond to the estimated coefficient in log-odds. The blue line represents the Theil-Sen slope indicating the overall trend in the time-series. This estimator fits a slope to a series of points in the plane based on the median of the slopes through all pairs of points, making it robust to outliers. The red dotted vertical line indicates there is a significant Pettitt trend-break in the time-series. The Pettitt trend-break uses the Mann and Whitney statistical test to compare segmentations of the complete time-series. The solid red lines represent the Theil-Sen slopes for each separate time-series.

4.2.1 Probability of connection

The log-odds of tie-formation between any random pair of inventors – independent of the other variables in the model – has dropped between 1840 and 1975. Figure 4.1 shows the overall trend is slightly negative, indicating that the tie-formation between any random pair of inventor is decreasing over time. This is to be expected because the networks have grown over time, making it less likely a random pair of inventors to collaborate. This is probably also causing the significant negative slope in the period before the 1865 trend-break. However, after the trend-break a significant positive trend is observed, indicating that regardless of the growing network size, the log-odds that any two random inventors collaborate is increasing over time. This can be explained by the increasing tendency to collaborate and the growing team-sizes on US patents over time and especially since the 1940s (Van der Wouden, 2018).

4.2.2 Geographical distance

Geographical distance has a negative impact on collaboration. The coefficients plotted in Figure 4.2 are to be interpreted as the change in log-odds when geographical distance between pair of inventors increases by 1 kilometer. For instance, in 1850 the log-odds of a collaboration between two random inventors living 100 kilometers apart is -1.80. However, in 1975 these log-odds are -0.086. The negative influence of geographical distance on tie-formation between inventors has decreased over time. The biggest decrease occurs between the 1840 and 1870. During this period of time the US railroads and telegraph were rapidly expanded, connecting cities in the north-eastern states. This allowed people to travel and communicate more easily – easing the friction of geographical distance. Interestingly, after the 1870 trend-break the positive slope of the trend is very small. This indicates that the negative impact of geographical distance on tie-formation did not change much over time. For example, the negative impact of one kilometer distance between two inventors on tie-forming in 1910 is -0.0030 and -0.0028 in 1960. These data thus suggest that after the initial reduction in the friction of geographical distance on tie-formation in the late 19th century, the impact of geographical distance remained rather stable throughout the 20th century.

Figure 4.1 The log-odds of tie-formation between any pair of inventors between 1840 and 1975

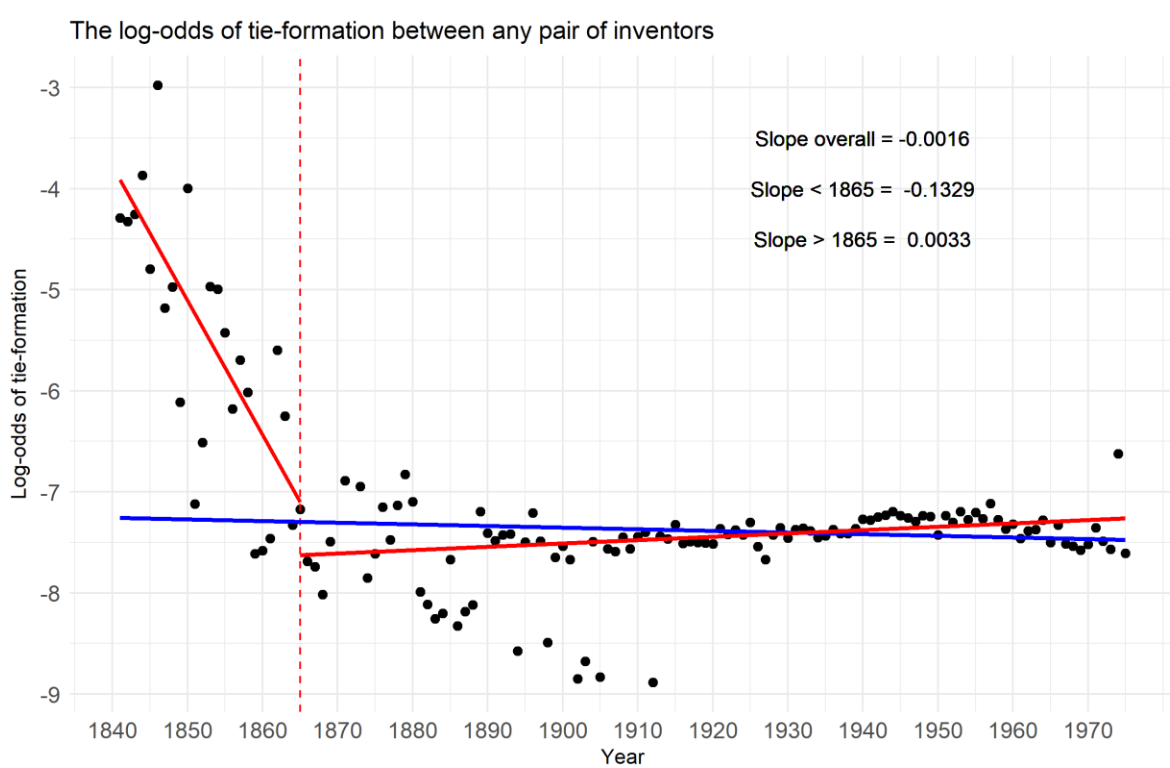
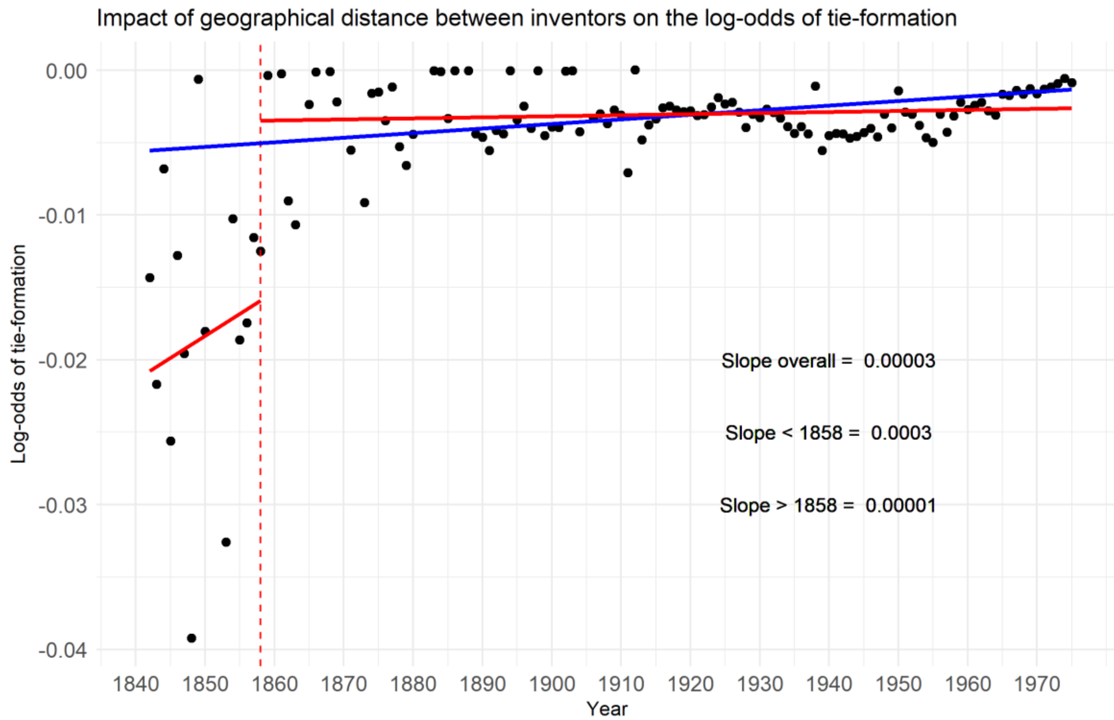


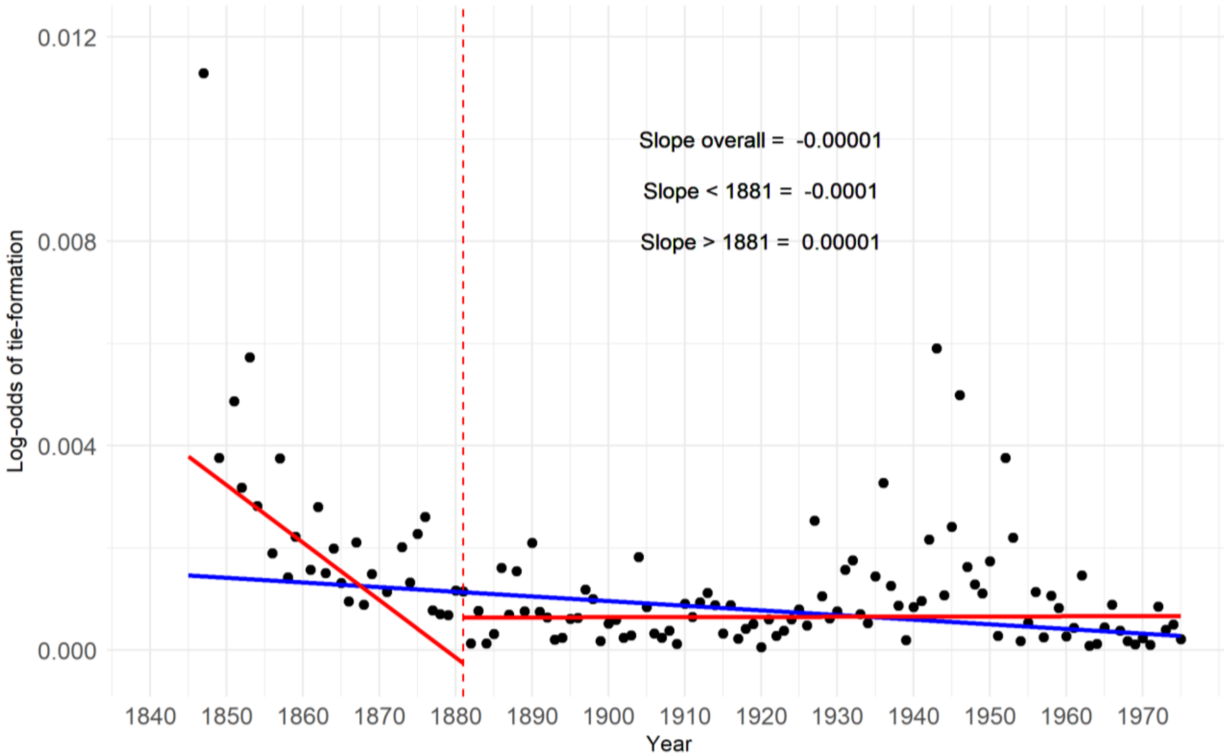
Figure 4.2 Impact of geographical distance on the log-odds of tie-formation



4.2.3 Technological proximity

Technological proximity between the patent portfolios of two inventors has a positive impact on tie-formation. This means that the log-odds of a tie-formation are greater when two inventors have produced patents in similar technology classes. However, Figure 4.3 shows an overall negative trend over time. In the mid-19th century technological proximity had a relative strong positive impact on tie-formation. This impact dropped until the wake of Great Depression in the mid-1920s. During the Great Depression, throughout the Second World War and its aftermath the impact of technological proximity on tie-formation increased again. From the mid-1950s onwards the impact of technological proximity was returned to its pre-war impact on tie-formation. Still, inventors are more likely to collaborate if they have similar technological expertise.

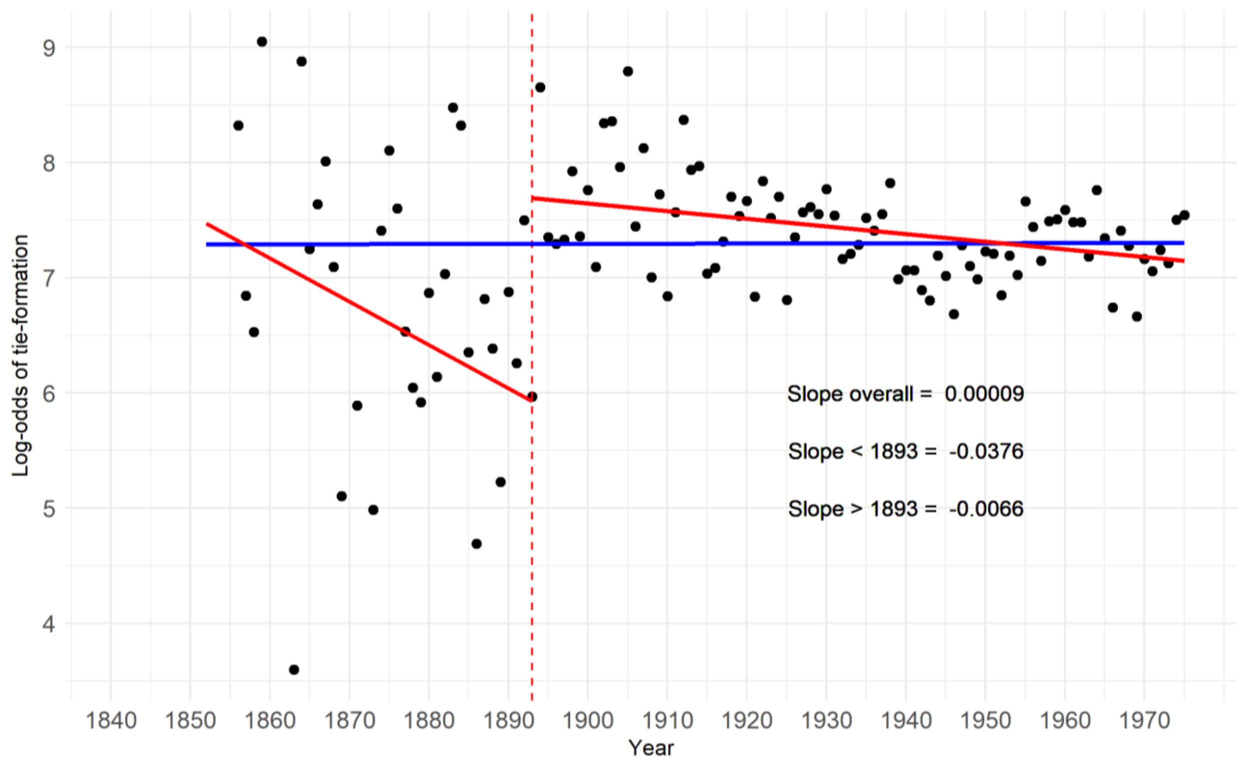
Figure 4.3 Impact of technological proximity on the log-odds of tie-formation between inventors



4.2.4 Social proximity

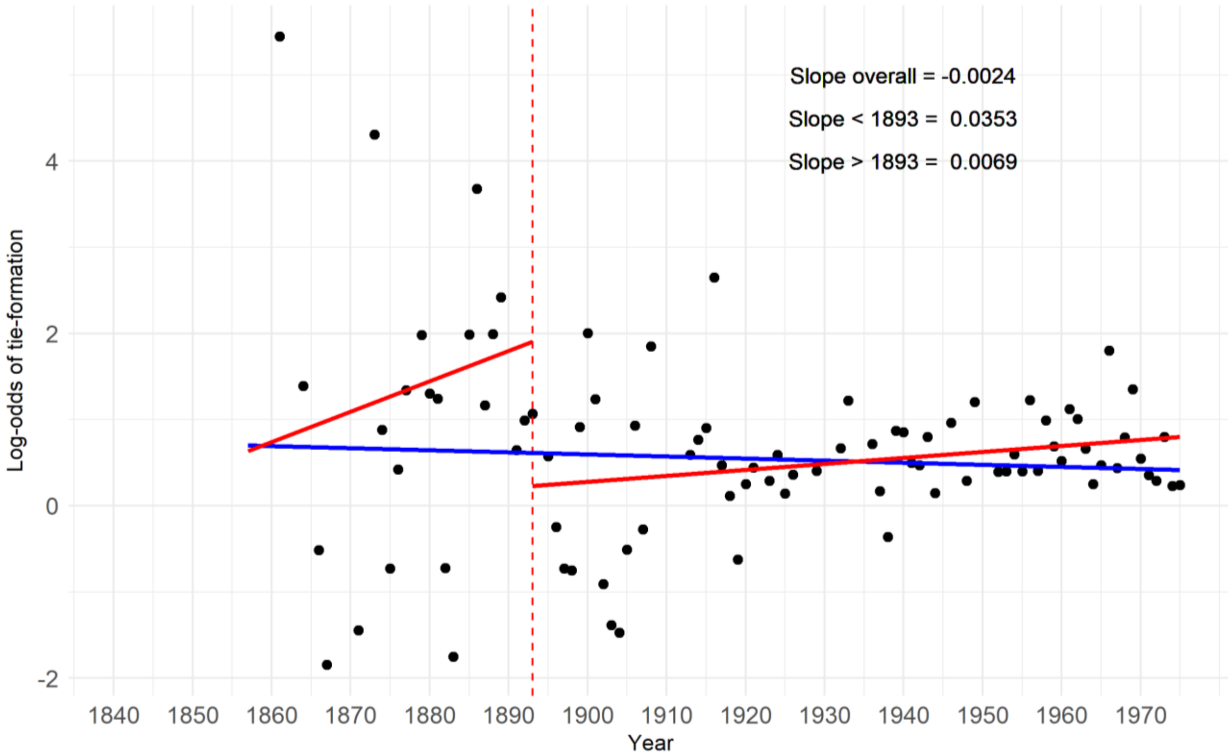
Social proximity between inventors has, in general, a positive effect on tie-formation between inventors. Figure 4.4 show how the log-odds of a tie-formation between two inventors changes when they have collaborated on a patent in the previous ten years. The impact is positive and relatively strong. The overall trend shows an (insignificant) upward slope, suggesting that the impact is increasing over time. However, the trends before and after the trend-break are both negative, suggesting that the log-odds of tie-formation decreases over time. Yet, the impact of having a previous collaboration with an inventor changes the odds of a tie tremendously. In 1950 the odds of a collaboration change by a factor of 1376 if two inventors have a previous collaboration. This suggests that pairs of inventors are rather robust and connections endure over time. Moreover, these results indicate that first-order connections made in previous years are a strong predictor of tie-formation between inventors, but that the impact is decreasing in the 20^{th} century.

Figure 4.4 Impact of previous collaboration on the log-odds of tie-formation



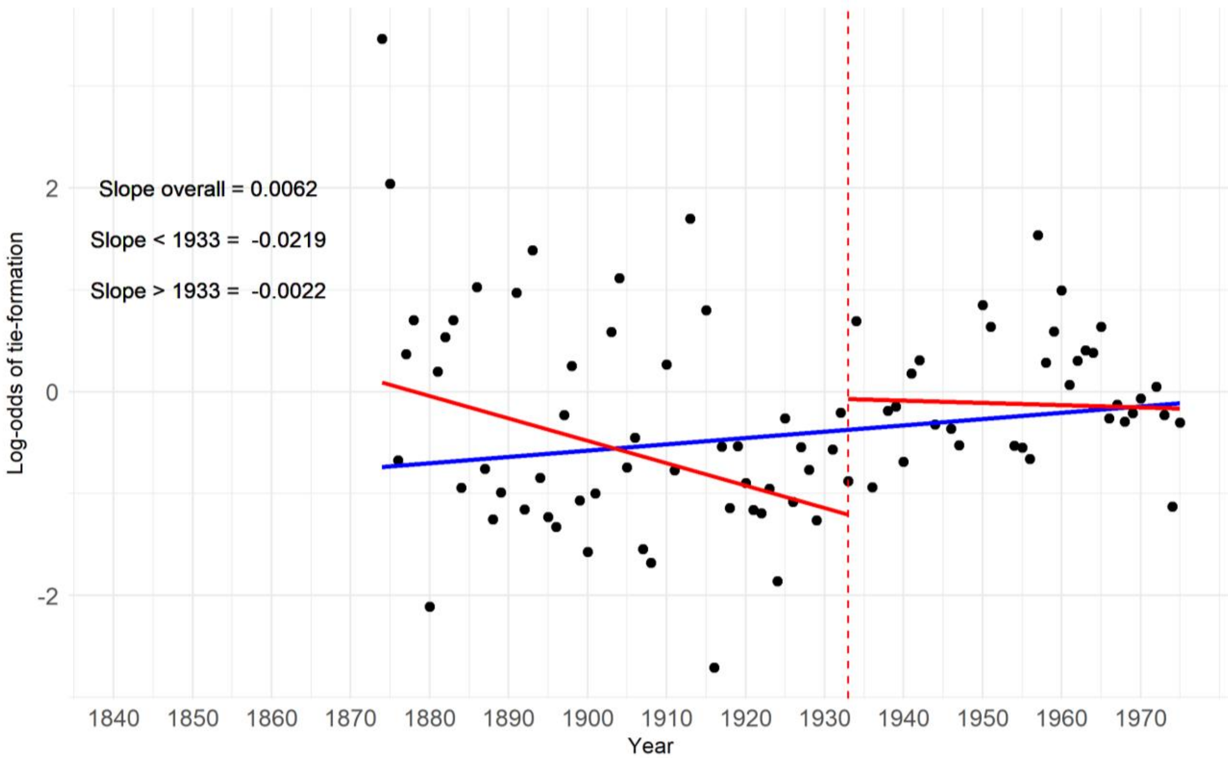
Having a second-order connection in the previous ten years has a precarious effect on tie formation up to the 1940s. Figure 4.5 indicates that there is an overall negative trend in impact of having a second-order connection on the log-odds of tie-formation between a pair of inventors. However, this trend and the trends before and after the 1892 trend-break are insignificant. Interestingly, having a second-order connection has a relatively strong and consistent positive impact on tie-formation since the 1940s.

Figure 4.5 Impact of having a previous collaborator in common on the log-odds of tie-formation



The impact of a third-order connection in the previous ten years on tie formation is less clear. Figure 4.6 shows that in the period before 1874 having such connection did not have a significant impact on tie-formation. After 1874 a third-order connection has a precarious effect on tie-formation between a pair of inventors. The volatile nature of the data in early years can be explained by the very low annual counts of such connections. But even when counts are relatively high after the 1930 trend-break, third-order connections still have an unclear effect on tie-formation. As a result, no significant trends are detected. What is clear is that there is no strong, consistent, positive effect of a third-order connection on tie-formation between inventors.

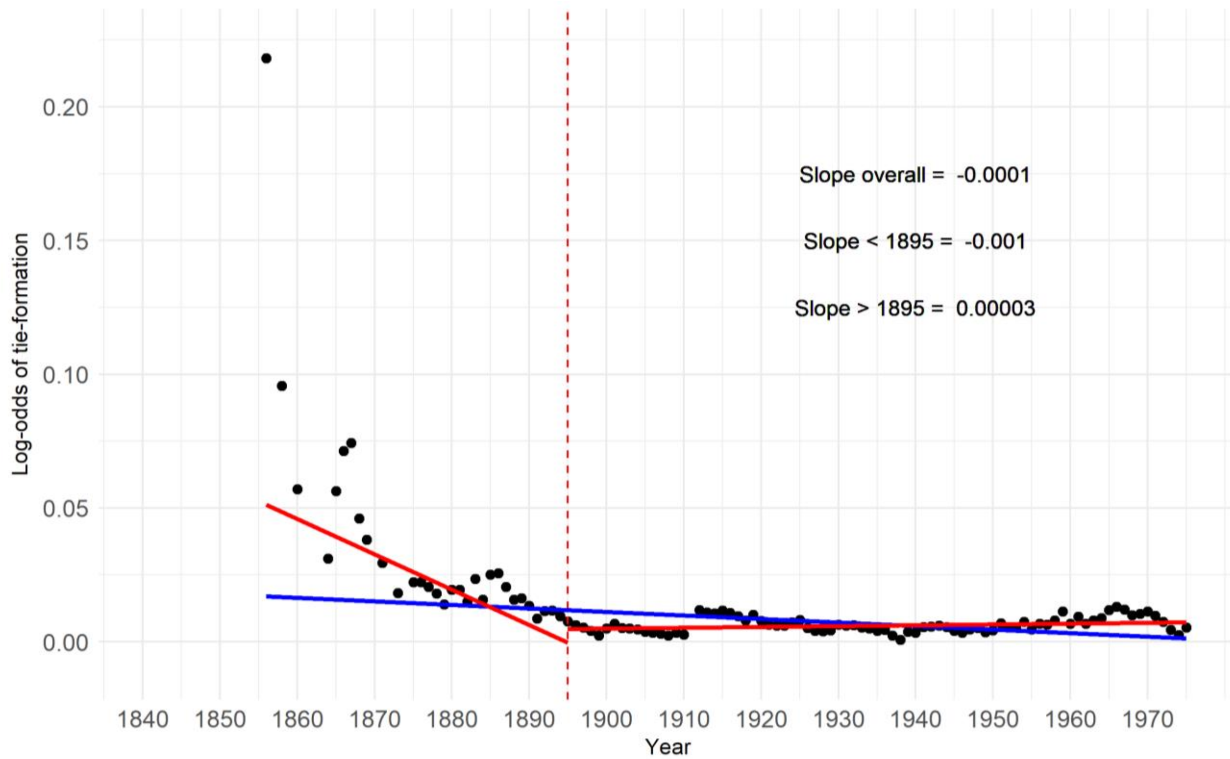
Figure 4.6 Impact of having a third-order connection on the log-odds of tie-formation



Number of previous patents

Having previous patents significantly increases the likelihood of tie-formation between two inventors, but this impact is decreasing over time. Figure 4.7 indicates that having one previous patent increases the log-odds of tie-formation in 1858 with 0.096. This number drops until the 1895 trend-break. After 1895 there is a rather weak, insignificant positive trend. In the 19th century, experienced inventors are more likely to form ties with other inventors than do experienced inventors in the 20th century.

Figure 4.7 Impact of the number of previous patents on the log-odds of tie-formation



Changing impacts over time

The probability of tie-formation has changed over time. To provide a hands-on understanding of how these probabilities have changed, Table 4.2 lists three examples of situations for tie-formation for inventor *i* and *j*. These inventors have 1 and 2 patents in previous years, respectively.

To estimate the probability of tie-formation, the log-odds parameter model estimates need to be transformed logistically. While the impact of a unit change of a certain variable on the log-odds of tie-formation can be examined independently of the other variables in the model, to properly find the probability of tie-formation, all model variables need to take a certain value (Luke, 2015). In the *Geographical* situation the inventors are close in geographical space. The technological and social proximity variables take a (arbitrary) constant value. In the

Technological example, inventors *i* and *j* have a rather great technological proximity in their knowledge portfolio. The geographical distance and social proximity take a constant value. Finally, in the *Social* example, the inventors have a lot of connections in common. Again, the geographical distance and technological proximity take a constant value.

Table 4.2 Three example situations for tie-formation among two inventors

Situation	Geog.	Techn.	Social 1	Social 2	Social 3	$\textbf{Patents}_i$	$\textbf{Patents}_j$
Geography	0	1	1	0	0	1	2
Technology	1000	800	1	0	0	1	2
Social	1000	1	2	1	1	1	2

Table 4.3 shows how tie-formation across these three situations has changed over time. The baseline estimates (identical to the 'edges' term discussed above) represents the overall probability that any two inventors collaborate in the focal year. This baseline is used to evaluate the change in probability of collaboration across the three situations. As expected, inventors close in geographical space are much more likely to collaborate. The probability that two inventors in the Geography situation collaborate is 28.5% in 1880 and drops to 10.8% in 1970. While the benefits of being close in geographical space are decreasing over time, the impact geographical proximity is still significant. An opposite effect is observed in the Technology situation – the impact of technological proximity is increasing over time. In 1880, the probability that two inventors with a relatively great technological proximity collaborate is roughly 1.2% and increases up to 2.8% in 1970. Having first- and second-order prior social ties has the largest impact on tie-formation, increasing over time. The probability that two inventors in the Social

situation (see Tables 4.2 and 4.2) are collaborating increases from 67.4% up to a staggering 98.1%.

Table 4.3 Change in probability of tie-formation across three situations and time

	Probability of tie-formation		
Year	**1880**	**1925**	**1970**
Baseline	0.0008	0.0007	0.0005
Geography	0.285	0.121	0.108
Technology	0.012	0.025	0.028
Social	0.674	0.915	0.981

Goodness-of-fit

From the 135 fitted models all but 3 models converged. The models for years 1850, 1922 and 1971 didn't converge. It is unclear why the models for these years fail to converge, but the spread of these models across the time-window of this study does not suggest a systematical issue. However, while model convergence indicates that the algorithms were able to settle on stable coefficient estimates, additional information is needed to evaluate the goodness-of-fit of the models. Given the generative nature of ERGM, Hunter et al. (2008) suggest to use the estimated model coefficients and same MCMC algorithms to simulate a number of networks, and compare the simulated networks to the observed network on a series of graph-level structural properties. If the properties of the simulated networks differ significantly from the properties of the observed network, the goodness-of-fit of the estimated model is poor and caution is needed to interpret the results of this model. The goodness-of-fit of the models presented here is generally good, except for the models in early years (pre-1850) and later years (post 1961).

5. Conclusion

This research has examined the mechanisms that structure collaboration among US inventors between 1836 and 1975. Although the main reasons to collaborate are well defined and understood, long-run structural empirical evidence on the mechanisms that structure tie-formation amongst inventors is missing. To address this gap, three frequently posited mechanisms that impact collaboration are examined: geographical distance; technological proximity; and social proximity. The three key findings are that (1) geographical distance negatively impacts collaboration, that this effect has been decreasing over time but has levelled off in the 20th century; (2) technological proximity between inventors' patent portfolio positively impact collaboration, especially in times of uncertainty; (3) social proximity promotes collaboration, but only to a certain extend.

Using new inventor-patent data (Van der Wouden, 2018), 140 annual networks of inventor collaboration are constructed and examined. For each year, matrices are constructed that hold information on the geographical distance and technical and social proximity for each possible inventor pair. The inventor collaboration network and the matrices are used as inputs for 140 exponential random graph models (ERGMs). These ERGMs estimate how the three proposed mechanisms, together with three control variables, impact the likelihood of observing a tie between two inventors. ERGMs have a number of advantages over traditional statistical techniques. These models allows for interdependency among observations, which is essential for network data. Tie-formation is seen as an endogenous process that involves dyadic dependence. In addition, ERGMs model effects across individual, dyadic, triadic and network levels simultaneously. This allows researchers to distinguish the effects of multiple micro-level dynamics. Finally, using probability theory and stochastic procedures for statistical inference, ERGMs can test network theories and model the structural properties of networks. Together, these characteristics enable researchers using ERGMs to make proper and accurate model estimates for mechanisms structuring tie-formation in networks.

This research finds evidence that geographical distance has a negative effect on tie-formation between inventors. This finding has already been well documented in the literature (Cassi & Plunket, 2015; Crescenzi et al., 2016; Ter Wal, 2013). What is novel is that this research shows there is a large decrease in the negative impact in the 19th century. Geographical distance matters less. However, the negative impact of geographical distance on tie-formation only decreased incrementally throughout the 20th century. This finding is surprising when keeping in mind the incredible developments in telecommunication and transport technologies during this era. Geographical distance still negatively affects tie-formation. The advances in communication and transportation developments seem not to have been able to match the benefits of geographical proximity (see Audretsch & Feldman, 1996; Feldman & Kogler, 2010; Jaffe et al. 1993; Storper & Venables, 2004).

Technological proximity has a positive impact on tie-formation between two inventors. The evidence presented in this paper shows that the impact of technological proximity on tie-formation is largest in two specific time-periods: during the mid-1800s and between 1930 and 1950. Both periods are characterized by high levels of uncertainty. In the first time-period, the 1836 US patent system was still relative new. It is likely that the introduction of this novel system confronted inventors with unknown rules and processes that determined their chances of successful application. In the second period, the Great Depression, the Second World War and its aftermath raised uncertainty on a different level. During crises individuals might be less likely to experiment because it is unclear if society will appreciate their experiments. Instead, in both time-periods inventors are more likely to make traditional, less risky decisions and collaborate with inventors with similar technological skills. That is, in times of uncertainty, inventors might choose the exploitation of familiarity over the experimentation with new technologies (Granovetter, 2005; March, 1991; Schumpeter, 1934). In this case, the exploitation of familiarity is the collaboration with inventors with similar technological portfolios.

Social proximity has a strong positive effect on the likelihood of collaboration. However, this only holds for first-order connections and for second order connections after the 1930s. First-order connections are established between two inventors when they have collaborated on a patent in the previous 10 years. A positive impact of a first-order connection thus means that inventors are more likely to re-collaborate with a previous collaborator. Inventors know the expertise of previous collaborators and can more readily evaluate their contributions on future patents than that of unfamiliar inventors, increasing the likelihood of collaboration. Second-order collaborations have a consistent positive effect on collaboration after the 1930s. This could be explained by the rapid increase in US inventor collaboration after the 1930s, reported by Van der Wouden (2018). This development generated a large number of connections throughout an increasing pool of inventors that has experience with collaboration. This is likely to increases the overall likelihood of new collaborations to occur and especially promotes collaboration with an inventor with whom previous collaborators already collaborated. The impact of a third-order connection on tie formation is less distinct. Throughout the entire time-frame of this study the impact has been volatile, suggesting that these type of connections have limited impact on future tie-formation.

The number of previous patents has a positive effect on tie-formation, but this impact is decreasing over time. Its strongest effect on collaboration was in the 19^{th} century. Through the 20^{th} century, its effect on collaboration was rather stable. There are a number of possible explanations for these findings. First, the processes of applying for and being granted a patent might not have been as transparent and obvious in the 19^{th} century as it was throughout the 20^{th} century. The 1836 patent system was only just introduced and only a relatively small number of Americans had experience with the novel system. New inventors might decide to team-up with inventors with successful patenting experience to overcome this ignorance. Second, patenting costs money. Although patenting in the US was relatively inexpensive compared to other countries, it still was costly (Khan & Sokoloff, 2001; Lerner, 2000) . Financial assistance might not have been as readily available as in later years, because inventors mostly worked independently of firms in the 19^{th} century (Lamoreaux & Sokoloff, 2005). Collaborating and jointly applying for a patent allows cost sharing. In the 20^{th} century, firms would increasingly finance patents. Third, experienced inventors are more attractive partners if they had a good reputation. In the 19^{th} century most inventors operated independently and made a living by licensing specialized patents to firms. Inventors who did this successfully would have built good reputations and become attractive to other (novel) inventors to collaborate with. This effect drops over time, because inventors increasingly become recruited by firms. In 1933 the US Supreme Court ruled that the patent and its rights that result from inventors hired by firms belong to the firm (Coriat & Weinstein, 2012). This took away the direct monetary incentive for an increasing pool of corporate inventors to collaborate with experienced inventors.

This paper has four contributions to research on knowledge production. First, this research has shown that the negative effect of geographical distance on collaboration changes over time, but seems to be rather resilient to the technological advances in tele-communication and transportation introduced in the 20th century. Second, not all social connections have a positive effect on collaboration. Only after the 1930s having a second-order connections consistently affects collaboration positively. Third-order connections don't seem to positively impact collaboration in a systematic fashion. Third, in times of uncertainty US inventive collaborative behavior is different. Inventors are more likely to collaborate with inventors with similar knowledge stocks than in times with less uncertainty. Finally, this paper has shed light on collaboration in a period US knowledge production that has been undocumented before. The results presented here are the first systematical long-run evidence identifying mechanism that structure tie-formation among US inventors.

The results of this analysis are of relevance for policymakers constructing innovation policies and corporate executives concerned with innovation. As collaborative knowledge production has numerous benefits on firm and regional level, understanding the mechanisms structuring collaboration is important. First, geographical proximity is still an important driver of collaboration and suggests the significance of clustering economic activities in space – even with telecommunication and transportation technologies in place and controlling for social connections and technological proximity. Second, simply clustering activities might not be enough. Social connections between inventors, either first- or second-order, significantly promotes collaboration. Facilitating environments that stimulate the creation of dense social networks will benefit collaborative knowledge production. Promoting individuals to interact and engage with previous collaborators and their close connections is most likely to result in successful collaborative output, ceteris paribus. Third, the direction of knowledge production is impacted by the degree of uncertainty. In times of high uncertainty, US inventors are found to avoid risk and collaborate with technologically similar inventors. Policies aimed at reducing the inherent uncertainties of knowledge production could result in less path dependent and more diversified outcomes.

This research has several limitations that should be taken into account. First, the role of the firm in tie-formation is unclear. Currently, there is no reliable data that links firms to the historical US patents. This becomes an issue after 1933, because then firms become legally able to apply for patents. Second, this paper has focused on historical US patents. Knowledge production is much broader than the production of patents. Clearly, not all knowledge is patented. Therefore, caution is advised generalizing these findings. Third, the data used for this research is incomplete. Only patents that are assigned to inventors living in MSAs are incorporated in the study. This means that patents containing rural or foreign inventors are dropped from the analysis. Finally, the analysis in this paper is limited from 1836 to 1975. It is unclear how these mechanisms structure tie-formation after 1975. What is especially of interest is how the introduction of the Internet will impact the effect of geographical, social and technological proximity on tie-formation among US inventors.

Literature

Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative Science Quarterly*, *45*(3), 425–455.

Atack, J., Bateman, F., Haines, M., & Margo, R. A. (2010). Did railroads induce or follow economic growth? *Social Science History*, *34*(2), 171–197.

Audretsch, D. B., & Feldman, M. P. (1996). R & D Spillovers and the Geography of Innovation and Production. *The American Economic Review*, *86*(3), 630–640.

Bathelt, H., & Gluckler, J. (2003). Toward a relational economic geography. *Journal of Economic Geography*, *3*(2), 117–144.

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.

Bercovitz, J., & Feldman, M. (2011). The mechanisms of collaboration in inventive teams: Composition, social networks, and geography. *Research Policy*, *40*(1), 81–93.

Boschma, R. (2005). Proximity and Innovation: A Critical Assessment. *Regional Studies*, *39*(1), 61–74.

Breschi, S., & Lissoni, F. (2009). Mobility of skilled workers and co-invention networks: An anatomy of localized knowledge flows. *Journal of Economic Geography*, *9*(4), 439–468.

Broekel, T., & Boschma, R. (2012). Knowledge networks in the Dutch aviation industry: the proximity paradox. *Journal of Economic Geography*, *12*(2), 409–433.

Brown, J. S., & Duguid, P. (2002). Local Knowledge: Innovation in the Networked Age. *Management Learning*, *33*(4), 427–437.

Burt, R. S. (1992). *Structural Holes*. MA: Harvard Business Press.

Cartwright, D., & Harary, F. (1956). Structural balance: a generalization of Heider's theory. *Psychological Review*, *63*(5), 277.

Cassi, L., & Plunket, A. (2014). Proximity, network formation and inventive performance: in search of the proximity paradox. *Annals of Regional Science*, *53*(2), 395–422.

Cassi, L., & Plunket, A. (2015). Research Collaboration in Co-inventor Networks: Combining Closure, Bridging and Proximities. *Regional Studies*, *49*(6), 936–954.

Coleman, J. S. (1990). *Foundations of Social Theory*. Belknap Press of Harvard University Press.

Coriat, B., & Weinstein, O. (2012). Patent Regimes, Firms and the Commodification of knowledge. *Socio-Economic Review*, *10*, 267–292.

Crescenzi, R., Nathan, M., & Rodríguez-Pose, A. (2016). Do inventors talk to strangers? on proximity and collaborative knowledge creation. *Research Policy*, *45*(1), 177–194.

Davis, J. a. (1970). Clustering and Hierarchy in Interpersonal Relations: Testing Two Graph Theoretical Models on 742 Sociomatrices. *American Sociological Review*, *35*(5), 843.

de Solla Price, D. J. (1970). Citation measures of hard science, soft science, technology, and nonscience. In *Communication Among Scientists and Engineers* (pp. 3–22). Lexington Books, D.C. Heath and

Company.

Eaton, J. W. (1951). Social Processes of Professional Teamwork. *American Sociological Review*, *16*(5), 707–713.

Edge, D. (1979). Quantitative measures of communication in science: A critical review. *History of Science*, *17*(2), 102–134.

Faems, D., Van Looy, B., & Debackere, K. (2005). Interorganizational Collaboration and Innovation: Toward a Portfolio Approach. *Journal of Product Innovation Management*, *22*, 238–250.

Feldman, M. P., & Kogler, D. F. (2010). *Stylized facts in the geography of innovation*. *Handbook of the Economics of Innovation* (1st ed., Vol. 1). Elsevier BV.

Fischer, C. S., & Carroll, G. R. (1988). Telephone and automobile diffusion in the United States, 1902-1937. *American Journal of Sociology*, *93*(5), 1153–1178.

Fisk, C. L. (1998). Removing the "Fuel of Interest" from the "Fire of Genius": Law and the Employee Inventor, 1830-1930. *University of Chicago Law Review*.

Fox, M. F. (1991). Gender, environmental milieu, and productivity in science. In H. Zuckerman, J. Cole, & J. Bruer (Eds.), *The outer circle* (pp. 188–204). WW Norton & Co, New York.

Freeman, L. C. (1979). Centrality in Social Networks Conceptual Clarification. *Social Networks*, *1*(1968), 215–239.

Frenken, K., Hölzl, W., & Vor, F. de. (2005). The citation impact of research collaborations: the case of European biotechnology and applied microbiology (1988–2002). *Journal of Engineering and Technology Management*, *22*(1–2), 9–30.

Goodreau, S. M., Kitts, J. A., & Morris, M. (2009). Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks. *Demography*, *46*(1), 103–125.

Granovetter, M. (1985). Economic action and social structure: the problem of embeddedness. *American Journal of Sociology*, *91*(3), 481–510.

Granovetter, M. (2005). The impact of social structure on economic outcomes. *Journal of Economic Perspectives*, 33–50.

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

Handcock, M. S. (2003). *Assessing degeneracy in statistical models of social networks* (Center for Statistics and the Social Science, University of Washington).

Harary, F. (1955). The Number of Linear, Directed, Rooted, and Connected Graphs. *Transactions of the American Mathematical Society*, *78*(2), 445–463.

Hoekman, J., Frenken, K., & Tijssen, R. J. W. (2010). Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Research Policy*, *39*(5), 662–673.

Hunter, D. R. (2007). Curved Exponential Family Models for Social Networks. *Social Networks*, *29*(2), 216–230.

Hunter, D. R., Goodreau, S. M., & Handcock, M. S. (2008a). Goodness of fit of social network models. *Journal of the American Statistical Association*, *103*(481), 248–258.

Hunter, D. R., Goodreau, S. M., & Handcock, M. S. (2008b). Goodness of Fit of Social Network Models. *Journal of the American Statistical Association*, *103*(481), 248–258.

Jaffe, A. B. (1989). Real effects of academic research. *The American Economic Review*, 957–970.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics*, *108*(3), 577–598.

Jones, B. F. (2009). The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder? *The Review of Economic Studies*, *76*(1), 283–317.

Jones, B. F., Wuchty, S., & Uzzi, B. (2008). Multi-University Research Teams: Shifting Impact, Geography, and Stratification in Science. *Science*, *322*(5905), 1259–1262.

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, *26*(1), 1–18.

Khan, B. Z., & Sokoloff, K. L. (2001). History Lessons : The Early Development of Intellectual Property Institutions in the United States. *The Journal of Economic Perspectives*, *15*(3), 233–246.

Kohler, H.-P., Behrman, J. R., & Watkins, S. C. (2001). The density of social networks and fertility decisions: Evidence from South Nyanza District, Kenya. *Demography*, *38*(1), 43–58.

Lamoreaux, N. R., & Sokoloff, K. L. (1996). Long-term change in the organization of inventive activity. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(November), 12686–12692.

Lamoreaux, N. R., & Sokoloff, K. L. (1999). Inventors, firms, and the Market for Technology in the Late Nineteenth and Early Twentieth Centuries. In N. R. Lamoreaux, D. M. G. Raff, & P. Temin (Eds.), *Learning by Doing in Markets, Firms, and Countries* (pp. 19–60). University of Chicago Press.

Lamoreaux, N. R., & Sokoloff, K. L. (2005). *The decline of the independent inventor: a schumpterian story?* National Bureau of Economic Research.

Lee, S., & Bozeman, B. (2005). The Impact of Research Collaboration on Scientific Productivity. *Social Studies of Science*, *35*(5), 673–702. http://doi.org/10.1177/0306312705052359

Lerner, J. (2000). 150 Years of Patent Protection. *NBER Working Paper 7478*.

Lindsey, D. (1978). *The scientific publication system in social science*. Jossey-Bass Inc Pub.

Luke, D. A. (2015). Statistical Network Models. In D. Luke (Ed.), *A User's Guide to Network Analysis in R* (pp. 163–187). Cham: Springer International Publishing.

Maggioni, M. a., Nosvelli, M., & Uberti, T. E. (2007). Space versus networks in the geography of innovation: A European analysis. *Papers in Regional Science*, *86*(3), 471–493.

Malmberg, A., & Maskell, P. (2006). Localized learning revisited. *Growth and Change*, *37*(1), 1–18.

March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, *2*(1), 71–87.

Martin, T., Ball, B., Karrer, B., & Newman, M. E. J. (2013). Coauthorship and citation patterns in the Physical Review. *Physical Review E*, *88*(1), 12814.

Melin, G. (2000). Pragmatism and self-organization: Research collaboration on the individual level. *Research Policy*, *29*(1), 31–40.

Merton, R. K. (1968). The Matthew effect in science. *Science*, *159*(3810), 56–63.

Mitchell, W., & Singh, K. (1996). Survival of Businesses Using Collaborative Relationships to Commercialize Complex Goods. *Strategic Management Journal*, *17*(3), 169–195.

Noble, D. F. (1979). *America by design: Science, technology, and the rise of corporate capitalism*. Oxford University Press, USA.

Nonnenmacher, T. (2001). History of the US Telegraph Industry. *EH. Net Encyclopedia, Edited by Robert Whaples*.

Nooteboom, B. (2000). Learning by interaction: Absorptive capacity, cognitive distance and governance. *Journal of Management and Governance*, *4*(1–2), 69–92.

Patel, N. (1973). Collaboration in the Professional Growth of American Sociology. *Social Science Information*, *12*(6), 77–92.

Pavitt, K. (1984). Sectoral patterns of technical change: Towards a taxonomy and a theory. *Research Policy*, *13*(6), 343–373.

Pavitt, K. (1998). Technologies, products and organization in the innovating firm: What Adam Smith tells us and Joseph Schumpeter doesn't. *Industrial and Corporate Change*, *7*(3), 433–452.

Perlman, E. R. (2016). Connecting the Periphery: Three Papers on the Developments Caused By Spreading Transportation and Information Networks in the Nineteenth Century United States.

Powell, W. W., Koput, K. W., & Smith-Doerr, L. (1996). Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative Science Quarterly*, 116–145.

Presser, S. (1980). Collaboration and the quality of research. *Social Studies of Science*, *10*(1), 95–101.

Price, D. de S. (1963). Big science, little science. *Columbia University, New York*, 119.

Price, D. J. de S. (1986). Little Science, Big Science and Beyond. Columbia University Press.

Rapoport, A. (1953). Spread of information through a population with socio-structural bias: I. Assumption of transitivity. *The Bulletin of Mathematical Biophysics*, *15*(4), 523–533.

Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social Networks*, *29*(2), 173–191.

Schumpeter, J. A. (1934). *The Theory of Economic Development: An Inquiry Into Profits, Capital, Credit, Interest, and the Business Cycle*. Harvard University Press.

Shi, F., Foster, J. G., & Evans, J. A. (2015). Weaving the fabric of science: Dynamic network models of science's unfolding structure. *Social Networks*, *43*, 73–85.

Singh, K., & Mitchell, W. (1996). Precarious Collaboration: Business Survival After Partners Shut Down or Form New Partnerships. *Strategic Management Journal*, *17*, 99–115.

Snijders, T. A. B. (2001). The statistical evaluation of social network dynamics. *Sociological Methodology*, *31*(1), 361–395.

Snijders, T. A. B. (2002). Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, *3*(2), 1–40.

Snijders, T. A. B. (2011). Statistical Models for Social Networks. *Annual Review of Sociology*, *37*(1), 131–153.

Stokes, T. D., & Hartley, J. A. (1989). Coauthorship, social structure and influence within specialties. *Social Studies of Science*, *19*(1), 101–125.

Storper, M., & Venables, A. J. (2004). Buzz: Face-to-face contact and the urban economy. *Journal of Economic Geography*, *4*(4), 351–370.

Stuart, T. E. (1998). Network Positions and Propensities to Collaborate: An Investigation of Strategic Alliance Formation in a High-Technology Industry. *Administrative Science Quarterly*, *43*(3), 668–698.

Ter Wal, A. L. J. (2013). The dynamics of the inventor network in german biotechnology: Geographic proximity versus triadic closure. *Journal of Economic Geography*, *14*(3), 589–620.

Thagard, P. (1997). Collaborative knowledge. *Noûs*, *31*(2), 242–261.

Thorsteinsdottir, O. (2000). External research collaboration in two small science systems. *Scientometrics*, *49*(1), 145–160.

Uzzi, B. (1996). The sources and consequences of embeddedness for the economic performance of organizations: The network effect. *American Sociological Review*, *61*(4), 674–698.

Uzzi, B., & Spiro, J. (2005). Collaboration and Creativity: The Small World Problem. *American Journal of Sociology*, *111*(2), 447–504.

Wray, K. B. (2002). The Epistemic Significance of Collaborative Research. *Philosophy of Science*, *69*(1), 150–168.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036–9.

Chapter 3: Inventor Mobility and Productivity: A Long-Run Perspective

Abstract

U.S. inventor mobility is traced between firms and over space from 1836 to 1975. Historical patent data are used to identify all U.S. inventors on patents, the locations of those inventors and whether the patents were assigned on issue. Mobility is identified for the set of repeat inventors who change assignee or location over time. Rates of inventor mobility are generated by decade over the period examined. Firm and geographical mobility increase over time, with only temporary reversals around the Great Depression and Second World War. Comparisons of patent productivity among matched samples of mobile and immobile knowledge producers reveals that firm mobility and spatial mobility raise the future number of patents produced by inventors. Firm mobility has a greater impact than geographical mobility on inventor productivity. However, the impact of both forms of mobility change over time.

Introduction

The mobility of skilled workers, researchers and inventors within the economy has generated considerable interest for some time. The concern with mobility hinges primarily on the impacts of the movement of skilled-workers across the economy. From Topel and Ward (1992) through to Jackson (2013) and Behrens et al. (2014), capturing the efficiency gains of worker sorting and matching, and separating these from the returns to agglomeration has long been a research target. These gains play out in the innovation literature as higher rates of inventor and firm productivity following movement (Hoisl, 2007; Kaiser et al., 2015). At the same time, inventor mobility has attracted more recent attention for its role in driving knowledge flows or spillovers between firms and over space (Zucker et al., 1998; Breschi and Lissoni, 2009). Such flows are a key component in models of economic growth resting on increasing returns, after Romer (1986, 1990). At the regional level, knowledge sharing through the movement of technology embodied in skilled workers is thought to characterize high performing regions (Saxenian 1994; Audretsch and Feldman, 1996; Almeida and Kogut, 1999; Miguelez and Moreno, 2013). At broader spatial scales inventor and skilled-worker mobility is also linked to innovation in emerging economies (Saxenian, 2005, 2007; Nathan, 2014) prompting calls for a mobility-driven model of economic growth and international development (Clemens, 2010).

This paper explores inventor mobility in the U.S. economy from 1836 to 1975. The primary goals are to explore long-run shifts in patterns of inventor mobility between firms and over space and to examine whether that mobility raises the productivity of inventors. Most studies of skilled-worker and inventor mobility focus on specific growth sectors, regions or limited time-periods. There are no data comparing the rates of movement of inventors over the long-run, nor whether mobility raises the productivity of inventors. It is important to know whether rates of inventor mobility have significantly changed over time and thus whether the flow of their knowledge may play as important a role in regional economic performance as many suggest.

To date, answering the questions just posed has been difficult largely because individual inventors in the United States have not been systematically identified and tracked. Patent data available from the United States Patent and Trademark Office (USPTO) provides a potential source of inventor records, though this source is compromised as individual inventors are identified by name and are not disambiguated. Further complicating the task of examining long-run inventor mobility, patent records before 1975 have not been available in digital form until recently (Van der Wouden, 2018).

The data used in this paper draw heavily on the publicly available HistPat database (Petralia et al., 2016) and the disambiguated inventor data from Van der Wouden (2018). HistPat contains geographical and technological information for historical patents granted by the USPTO between 1836 and 1975 and can be linked to the disambiguated inventor data. Using search, match and machine learning techniques the assignee data for these historical patents is extended. This allows tracking individual inventors in time and across space and firms. Matching algorithms are used to match mobile to non-mobile inventors on a series of covariates, such that there is no connection or bias between the treatment variable (mobility) and the control variables. These matched samples are used in statistical models examining the effect of firm and spatial mobility on the future number of patents produced by inventors.

The key findings of this paper are that firm mobility and spatial mobility raise the future number of patents produced by inventors. In general, inventors who move in space tend to have greater patent production over the following five years than non-mobile inventors. This effect is strongest for the last 30 years of the sample. Similarly, inventors who moved between firms produce more patents than inventors who stayed with the same firm. This positive effect on productivity is observed for all time-periods in the sample, except for the aftermath of the Second World War. Inventors who moved both between firms and locations have significantly higher patent productivity than their counterparts who moved only between firms or locations. Inventors who moved between firms have greater productivity gains compared to their control group than the inventors who moved geographically compared to their control group. This suggests that firm mobility impacts inventor productivity more than spatial mobility.

The remainder of the paper is organized as follows. The next section presents an overview of the relevant literature in which the paper is embedded. Section three introduces the construction of the data and methods. In section four these data are described in more detail. The empirical results of the statistical models exploring the influence of mobility on inventor productivity are presented in section five. The last section presents a series of conclusions, a discussion of the findings and directions for future research.

2. Literature Review

Technology is increasingly important to the performance of firms and, in aggregate, regions. The management of technology within organizations, its production and its protection, is a critical component of the resource based view of the firm (Barney, 1991; Wernerfelt, 1984; Grant, 1996) and related models of the regional economy (Saxenian, 1994; Lawson and Lorenz, 1999; Asheim and Gertler, 2005). For most researchers, knowledge, especially that of a tacit nature, is considered locked inside individuals, groups of workers and firm routines (Kogut and Zander, 1992). The relative immobility of knowledge is then seen as a key determinant of the competitive advantage of firms and especially regions even in a world economy that is increasingly integrated (Maskell and Malmberg, 1999).

Differentials in firm and regional performance notwithstanding, some knowledge clearly does move within the capitalist space economy. Ideas flow between organizations as they increasingly share the burden of innovation through collaboration (Schilling and Phelps, 2007; Hoekman et al., 2010). Knowledge also diffuses over time across firms and regions via spillovers of various kinds (Jaffe et al. 1993; Owen-Smith and Powell, 2004; Acs and Varga, 2005; Miguelez and Moreno, 2013) and through the movement of skilled individuals (Almeida and Kogut, 1999; Zucker and Darby, 2006). The relative significance of these mechanisms is keenly debated (Zucker et al., 1998; Breschi and Lissoni, 2009).

The mobility of workers within the economy and the impact of that movement has generated considerable debate. From early theoretical discussion on learning by Arrow (1962), to work on sorting and matching in labor markets (Topel and Ward, 1992; Jackson, 2013) a large literature has emerged that attempts to explain wage differences across the economy (Abowd et al., 1999; Behrens et al., 2014; Dauth et al., 2016) and to separate the influence of worker and firm heterogeneity on productivity and wages from the characteristics of cities, or agglomeration.

Related research focuses on processes of firm learning through hiring skilled employees (Song et al., 2003; Rosenkopf and Almedida, 2003) and extensions of these claims to regional economic performance (Boschma et al., 2008; Faggian and McCann, 2008; Cappelli et al., 2018).

The global dimensions of worker mobility and linkages to technology transfer, economic growth and development are explored by Saxenian (2005; 2007) who traces the international mobility of high-skilled U.S. workers, the networks they form and their role in circulating ideas between industrialized and emerging economies. Kerr (2008), Edler et al, (2011) and Nathan (2014) extend this work, while Clemens (2010) goes further in advocating for a labor mobility agenda to fuel international development. Links to the literature on globalization emerge in research that examines the role of technology spillovers from foreign firms, often transnational subsidiaries, channeled through worker mobility (see Gorg and Strobl, 2005; Balsvik, 2011; Poole (2013). There is abundant evidence of the impacts of skilled foreign workers on economic performance and innovation in the U.S. and Europe (Niebuhr, 2010; Peri, 2012; Stuen et al., 2012; Bosetti et al., 2015)

An important, though small, body of literature has also focused more specifically on the impacts of inventor mobility. Inventors are a special class of skilled workers that have generally been tracked through information linked to patents. The core concern in this research is whether mobile inventors, those who move between firms and over space, are more productive than immobile inventors. Hoisl (2007), using a matched sample of German inventors and firms, finds that inventors who switch firms are more productive, but also that more productive inventors are less likely to move between firms. The latter result is echoed by Palomeras and Melero (2010). Thus, it is important to be aware of endogeneity that links inventor mobility and performance. While Agarwal and Ohyama (2013) also report that the mobility of scientists increases their productivity, Fallah et al. (2012) show that inventors in the U.S. telecom industry who move between firms exhibit a reduction in patenting performance compared with inventors who remained in the same company. Hoisl and de Rassenfosse (2014) push further to identify the mechanism through which mobility influences inventor productivity and suggest it is a matching effect driven by the knowledge fit of inventors and their employers. Related research reports that highly productive inventors in the U.S. are disproportionately foreign-born (Stephan and Levin, 2001) and that Russian mathematicians who left the Soviet Union were out-performing their U.S. counterparts (Borjas and Doran, 2012). Evidence of no migration effect on innovative output is reported by Stephan et al. (2007) and Hunter et al. (2009). Franzoni et al. (2014) review these studies and find a positive migration effect in their own studies of scientific output.

A somewhat broader literature reports on the impact of inventor moves on the innovative output of the firms that hire them. Kaiser et al. (2015) report that a new R&D worker coming from a firm that patents significantly raises the number of patent applications by the new employer. Rahko (2017) does not find significant evidence that inventor mobility raises the patenting profile of hiring firms, though it has a positive impact on future inventive efforts. Song et al. (2003) and Rosenkopf and Alameida (2003) suggest that learning through new hires is an effective way of boosting innovation in distant technology fields. Rahko (2017) reports similar results. Maliranta et al. (2009) find that hiring the R&D workers of competing firms into one's own R&D division does not boost firm productivity, perhaps because of a lack of absorptive capacity or worker fit. Palomeras and Melero (2010) and Zwick et al. (2017) discuss the many factors that shape the relationship between mobility and firm outcomes.

In general, the broad findings of a positive effect of mobility on firm innovation are used to support the claims of Saxenian (1994) and Samila and Sorenson (2011) that high performing regions are characterized by high rates of skilled worker turnover and the sharing of knowledge. Further evidence is provided by Almeida and Kogut (1999). Migeulez and Moreno (2013) also report the positive effects of spillovers on regional innovation and Cappelli et al. (2018) show that inventor mobility raises the productivity of Italian regions. Work by Marx et al. (2009) showing how non-compete laws limit labor turnover and patenting supports these general findings. A broad review of much of the work above is provided by Maedsley and Somaya (2016).

3. Data and Methods

The data used to examine inventor mobility and productivity in this paper draw, in large part, on the publicly available HistPat database (Petralia et al., 2016). HistPat contains geographical and technological information for historical patents granted by the USPTO between 1790 and 1975. However, HistPat only provides information on the first inventor of a patent and does not identify possible additional co-inventors. Co-inventor data are recovered using the procedures outlined in Van der Wouden (2018), employing searching, matching and machine learning algorithms to detect, identify and disambiguate inventors on the HistPat patents and to extend the HistPat data in different ways. Identifying all inventors on U.S. patents and disambiguating those individuals enables the tracking of inventors over time and space, as long as they generate additional patents. Further, identifying assignee information for HistPat patents adds useful information about knowledge ownership and the status of inventors.

Section 3.1 describes the methods used to extend HistPat, largely through adding disambiguated assignee data. Assignees are organizations, typically firms and universities, that own, or are given ownership of, patents generated by individual inventors or teams of co-inventors. In many cases the inventors are employees of these organizations. Identification of assignees is critical to examine the mobility of repeat inventors between firms. Inventor mobility over space is captured by examining the locational data for inventors identified by Van der Wouden (2018). Section 3.2 details construction of a sample of repeating inventors in which mobile inventors, those moving between firms, locations or both, are matched to a set of non-mobile inventors across a series of key variables. Following these inventor samples over time permits analysis of the causal effect of mobility on inventor productivity.

3.1 Building assignee data

Following the general approach outlined in Van der Wouden (2018), the raw scraped text files of 4,125,734 USPTO patents available in HistPat were examined to identify whether a text string in the patent document was part of a company name using fuzzy matching algorithms. Only firm assignees were identified in the historical data. Universities were not systematic assignees on patents much before passage of the Bayh-Dole Act in 1980. The data for the lists of company names comes from the digital USPTO patents spanning the period 1975 to 2005 (Lai et al. 2011), the assignee data that is explicitly recorded in HistPat files (originating from USPTO and EspaceNet), the Business Reference Services of the Library of Congress and several Wikipedia pages listing corporations by country[17]. Combined, these data comprise 217,850 unique company names and reflect a mix of historical, current, foreign and domestic firms.

The words that (fuzzy) match with company names in our historical lists of patents are recorded. For these matches four additional characteristics are used for the supervised machine learning exercise discussed below. Analysis records to what extent a full match is observed, how often the matched word(s) occur in the document, where in the document the matched word(s) first occur and if the matched word(s) are close to the terms 'assigned to' or a close variation of that phrase. These characteristics are used to predict whether the matched firm names are likely the assignees of the patent and not a reference to different firms.

The next step is to classify each of the observed matches as either an assignee (1) or a non-assignee (0). Each observed match on a patent is called an event. Since the outcome variable of the matching process takes only two values we use supervised binary classification machine learning techniques.

---

[17] For example, a list of companies in the United States can be found at https://en.wikipedia.org/wiki/List_of_companies_of_the_United_States

Supervised machine learning approaches make use of three different types of datasets. The base data with known outcomes is randomly split in chunks of 80%, 10% and 10%. The largest set is branded the training set and used to train a series of different machine learning algorithms. These algorithms *learn* which combinations of variables in the data can best predict the known outcome of events. The validation set (10% of events) is used to tweak the parameters of candidate models to improve the performance and prevent the over-fitting the training data. Over-fitted candidate models are too sensitive to the structures observed in the training data and generalize poorly to other data. The test set (also 10% of events) is used to assess the performance of the candidate models on a number of indicators selected by the researcher. Finally, the model of best performance is used to predict whether each event on patents with unknown assignees is an assignee or not.

An important step is to build *known* events. HistPat contains two vectors with information on the assignee that originates from Google Patents and EspaceNet. These vectors have missing values for more than 700,000 patents. These missing values are correct if the patent was not assigned or incorrect if the information on an assignee has not been retrieved from the document. Patents are typically assigned to companies. When patents are not assigned, they generally remain the property of the inventor. The primary objective of this step of the analysis is to identify those patents that are unassigned and belong to the inventor(s) and those that are assigned to firms. Assignee data for repeat inventors can then be used to infer whether or not an inventor moves between firms. Hoisl (2007), using sample data of German patents and the firms they work for, confirms that well over 90% of the assignees listed on patents are the organizations that employ inventors.

The assignee information in HistPat can be used to construct *known* events. We merge the two assignees vectors and only keep records where the Google and Espace assignee information is identical. In this case, both the Google and EspaceNet researchers have established the same assignee. We assume this is the true assignee of the patent[18]. A search algorithm is deployed to identify non-empty records holding words that are corporate identifiers such as ‘Inc.’, ‘Corporation’, ‘Company’, the Dutch ‘Maatschappij’ or German ‘Werke’. The events for which the observed words are the same as the vector with known assignees can then be used to train candidate models. After tuning the parameters of these models on the validation set, the final model is chosen based on performance using the test data set. This final model is used to predict which observed names for the patents without a matched assignee in the EspaceNet name vector is the assignee or not. In the final data more than 60,000 corporate assignees were matched to patents missing in the HistPat data.

3.2 Variables, matched samples and models

The core research question in this paper asks whether geographical mobility or firm mobility affect the productivity of inventors. To disentangle the effects of these different types of mobility on productivity, four samples are constructed to estimate a series of related models. Table 3.1 and Table 3.2 list these samples and the nature of the analysis for which they are used. Note that inventors in the treatment group have experienced at least one type of mobility (between firm or between location). In analyses 3 and 4, inventors in the control group did *not* have one specific type of treatment but received the other.

[18] We have manually removed clear mistakes, but some mistakes might exist. This is not problematic, because machine learning algorithms deal rather well with *noise* in the data.

Table 3.1 Types of samples based on mobility of repeating inventors

		Geographic Mobility	
		Yes	No
Firm Mobility	Yes	A	B
	No	C	D

The *initial sample* consists of all inventors who have applied for more than 1 patent over during the period 1836 to 1975. We limit the analysis to inventors who lived in a US metropolitan area (MSA) at the time of the patent application. This decision is made because rural locations are difficult to identify and disambiguate in the historical data. This sample of urban, repeating inventors is called the initial sample and is described in section 4 of the paper.

Table 3.2 Four analyses of the influence of different types of mobility on productivity

Analyses	Treatment Group		Control Group	
	Samples	Label	Samples	Label
1	A + C	Geographical mobility (ignoring firm mobility)	B + D	No geographical mobility (ignoring firm mobility)
2	A + B	Firm mobility (ignoring geog. mobility)	C + D	No firm mobility (ignoring geog. mobility)
3	A	Double mobility	C	Only geographical mobility
4	A	Double mobility	B	Only firm mobility

The initial sample of all repeating inventors over the period 1836 to 1975 is resampled over time. Starting in 1900 and for every ten years up to 1970, annual *focal year* samples of repeating inventors are taken. Repeat inventors through the nineteenth century are not very numerous and so are ignored here. A repeating inventor is included in the 1900 focal year sample if she/he applied for a patent in 1900 and at least one patent in the previous five years (back to 1896 in this case). For all the repeat inventors in the eight *focal year* samples, a series of variables are constructed (these are outlined below). These variables were subsequently employed in the matching procedures and statistical analysis reported below.

The first key variable *Geographical Mobility* takes the value of zero if the listed location of the inventor on the patent in the focal year is the same as the location listed on the inventor's previous patent. The mobility variable takes the value of one if inter-urban mobility occurred. If the inventor has multiple patents in the previous five years, the most recent pair of patents is used to define the mobility variable. Immobile inventors are removed from the focal year sample if they become mobile within five years after the focal year. This is done to remove the influence of mobility on productivity after the focal year.

The second key variable *Firm Mobility* is constructed in similar fashion as geographical mobility but indicates whether the corporate assignee listed on the focal patent is the same as the corporate assignee on the inventor's previous patent. If this is the case, the variable takes the value of zero and otherwise one. Repeating inventors who are associated with one unassigned and one assigned patent are also considered to have experienced firm mobility. The third key variable *Double Mobility* measures whether an inventor moved between both geographical locations and firms between the focal and prior patent.

Inventor productivity is the dependent variable of interest in this study. *Productivity* is a count of patents granted to the inventor in the five years after the focal year. We hypothesize that mobile inventors produce on average more patents than non-mobile inventors.

An important concern is that productive inventors might also be more mobile because they have a higher probability of being recruited by firms than less productive inventors. This makes it difficult to disentangle the causal effect of mobility on productivity (see Hoisl, 2007). To control for this, a measure of prior productivity is generated that is labelled *Previous Patents*. This variable is the count of patents granted previously to the inventor before a focal year. This variable is a critical matching variable linking inventors in treatment and control groups as reported below.

It is hypothesized that the social networks of inventors impacts their productivity. This is a second matching variable defined across treatment and control samples. The *Sum of Collaborators* specifies the number of collaborators an inventor has worked with throughout their career up to the date of the focal year patent. This variable measures the size of the inventor's collaborative network and proxies for access to important resources such as knowledge, practices and opportunities. Increased access to resources can boost an inventor's future productivity because these resources can be used to produce patents more readily (see Owen-Smith and Powell, 2004).

The variable *Assignor* indicates whether an inventor's focal year patent is assigned to a corporate assignee. It takes the value of zero if it is not assigned and one if assigned to a corporation. Inventors associated with corporations/firms at the time of patenting are likely to have greater financial support than inventors patenting privately. This support might boost the productivity of inventors and thus this is used as a third matching variable across samples.

In addition to these controls, three variables are recorded and included as fixed effects in the statistical models employed. For each inventor, fixed effects include the *MSA* of location at time of patenting, the focal *Year* of invention, and the *Primary Technology Class* in which they patent. These fixed effect variables are also used for building the matched samples.

To isolate the effect of mobility on productivity, observations in the focal sample years are pre-processed by matching mobile inventors to non-mobile inventors using coarsened exact matching (CEM) algorithms (Iacus et al., 2012). Mobility is regarded as the 'treatment'. Each inventor is placed in the treatment or control group depending on how they score on the two mobility questions. For each of the four types of analyses (see Table 3.2) a treatment and control group is selected. The CEM algorithm prunes observations that don't have close matches on a series of covariates in both the treated and control groups. The covariates used for matching are *Sum of Collaborators, Assignor, Previous Patents, MSA* and *Technology Class*. After matching, the resulting annual *analytic samples* hold observations on mobile and non-mobile inventors that have similar features and only differ on their mobility score. This means that the analytic samples are balanced, such that there is no connection or bias between the treatment variable (*Mobility*) and the control variables. Note that for each of the four analyses a new matching procedure is undertaken because the treatment and control groups are different.

Pre-processing the data using CEM has many advantages (see Iacus, King, & Porro, 2011; Iacus et al., 2012). One of these is that the results of the statistical models are less reliant on the model specifications and thus more robust. In addition, matching allows estimation of the causal effect of mobility on future productivity – the main aim of this paper. Important to note is that this approach has at least two limitations for this research. The sample size of our analysis is smaller, because unmatched observations are deleted. In addition, our analysis is representative only for inventors with rather common covariates, because inventors that are outliers on the set of matching criteria are not likely to be matched.

Four statistical models are utilized to estimate the effect of mobility on productivity. These models take the form of a Poisson regression because our dependent variable (*Productivity*) is a count variable. The *Productivity* of inventor *i* is defined as follows:

$$\begin{aligned} Productivity_i = {} & \beta_1 * Mobility_i + \beta_2 * Sum\ of\ Collaborators_i + \beta_3 \\ & * Previous\ Patents_i + \beta_4 * Assignor_i + \beta_5 * (Mobility_i \\ & * Sum\ of\ Collaborators_i) + \beta_6 * (Mobility_i * Previous\ Patents_i) + \beta_7 \\ & * (Mobility_i * Assignor_i) + MSA_i + Primary\ Technology\ Class_i + Year_i \\ & + \varepsilon_i \end{aligned}$$

in which *MSA*, *Primary Technology Class* and *Year* are fixed effects and ε_i represents the error term. *Mobility* can take the form of geographical, firm or double mobility, depending on the analysis. β_{5-7} are interaction terms and are interpreted as the change in the effect of the interacting variable (i.e. Sum of Collaborators) on Productivity when Mobility takes the value of one. This allows the researcher to examine how Mobility mediates the effect of the control variables on Productivity.

In Analyses 1 and 2 the effect of geographical and firm mobility (respectively) on productivity is examined (see Table 3.2). The observations in the treatment group are inventors who have been mobile. The control group consists of inventors who did not receive such mobility treatment. A positive regression coefficient for mobility indicates that mobile inventors are more productive than non-mobile inventors. From the evidence in Analyses 1 and 2 it is impossible to say whether firm or geographical mobility impacts productivity more strongly because the samples are not comparable. To engage with that issue Analyses 3 and 4 are designed (see Table 3.2). The observations in the treatment group in Analysis 3 and 4 are mobile inventors in terms of both geographical and firm movement. The control group for Analysis 3 only consists of inventors who were mobile in terms of geography but not in terms of firm movement. The control group for Analysis 4 consists of inventors who were mobile in terms of firm movement but not geography. Note that the observations in the treatment group for Analysis 3 and 4 are similar, but the observations in the control group are different. This is important because the difference in regression coefficients for Mobility in Analysis 3 and 4 indicates whether the effect of firm mobility or geographical mobility on productivity is greater.

Table 3.3 presents descriptive statistics of the variables in the model for the *initial sample*, encompassing all repeating US urban inventors. Note that productivity is not included in Table 3.3 because it is only measured for inventors in the matched *analytical samples*. Their productivity is measured using data from the initial sample.

Table 3.3 Descriptive statistics of variables in the model

	# Obs.	Average	SD	Min	Max	Range
Geog. Mobility	1,310,564	0.25	0.43	0	1	1
Firm Mobility	1,310,564	0.31	0.46	0	1	1
Double Mobility	1,310,564	0.11	0.32	0	1	1
Assignor	1,310,564	0.59	0.49	0	1	1
Collaboration	1,310,564	0.42	0.49	0	1	1
Prev. patents	1,310,564	9.39	21.73	1	695	694
Sum of Collaborators	1,310,564	5.48	14.35	0	503	503

	# Obs.	# Unique
MSA	1,310,564	366
Primary Tech. Class	1,310,564	430
Year	1,310,564	140

4. Description of data

This section of the paper provides descriptive statistics on trends in mobility, productivity, collaboration, corporate patenting and the geography of repeating US inventors in the initial sample data between 1836 and 1975.

The share of mobile inventors has changed over time. Figure 4.1 shows that throughout most of the 19th century the share of mobile inventors increased. It reached a peak in 1876 when more than 40 percent of the repeating inventors had moved locations in the last 5 years. After this peak the share of mobile inventors declined to around 15 percent during the Great Depression and Second World War. Inventor mobility seemed to be reduced by these crises and the uncertainties they bring about. The economic revival of the US economy after the Second World War witnessed a significant increase in the mobility of inventors. Within 25 years, the share of mobile inventors nearly doubled in the post-war period. This likely reflects improvements in modes of transportation and the decreased cost of movement. It also reflects the growth of cities in the south and west of the country and the geographical spread of centers of invention that marked the rapid economic expansion of the 1950s and 60s. The figure also shows a smoothed 5-year rate of inventor mobility. Note that Trajtenberg (2005, 2006) claims inter-state mobility of inventors in the original NBER post-1975 data of 13%. Combining this with intra-state mobility yields mobility rates that are not that dissimilar than those presented here.

There is considerable heterogeneity in the productivity of repeat inventors. The heat map of Figure 4.2 highlights the annual distribution of inventor productivity. Until the late-1870s more than 50 percent of repeating inventors only had one previous patent and none had more than 6 previous patents. Over time, this balance gradually shifted towards inventors with larger patent portfolios. At the turn of the 20th century repeating inventors with more than 10 patents represented almost 30 percent of the inventors of any focal year, whereas inventors with only 1 patent only accounted for about 20 percent of the overall inventor pool. This growth in the average number of patents per inventor is driven in part by the impact of well-known prolific inventors (i.e. Thomas Edison) who acquired great numbers of patents at this time, a period when Lamoreaux and Sokoloff (1996) report the emergence of a national market for technology in the United States and a deepening division of labor around the process of invention. This trend reversed during the 1910s and 1920s, when the size of inventor portfolios declined somewhat. Interestingly, during the Great Depression and Second World War, repeat patenting became dominated by prolific inventors once more. A relatively small share of inventors had only one or two patents in their patent portfolios, while more than 30 percent of inventors had more than 10 patents. This suggests that during crises, when uncertainty is high, the influx of new repeating inventors is relative low. Experienced inventors with multiple patents might be less influenced by this uncertainty and continue patenting. In the post-war period repeating inventors with lower number of previous patents become dominant once again.

Patterns of mobility are different for inventors producing new knowledge individually or in teams, and for inventors producing patents for firms (assigned patents) or for themselves. Figure 4.3 indicates that repeat inventors working on patents produced through collaboration are more likely to be mobile than their non-collaborative counterparts. Although the shares of mobility fluctuate over time, collaborative inventors are more mobile across the entire time-period. However, during the Great Depression and Second World War, when overall mobility is low, the gap between collaborative and non-collaborative inventors is relatively small. The mobility rate of collaborative inventors is impacted more strongly by these crises than that for solo inventors. Mobility rates for both collaborating and solo inventors increase after the 1950s.

Figure 4.1 Share of geographically mobile US inventors over time (5-year mobility between US counties)

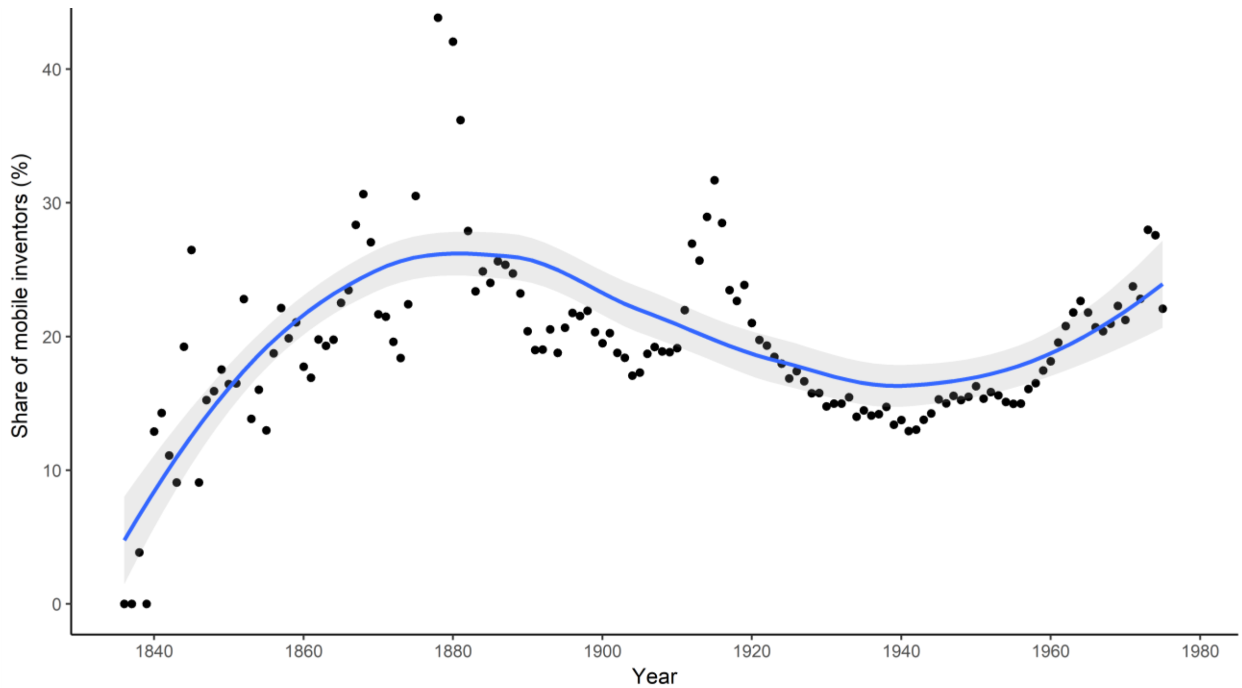


Figure 4.2 Annual distribution of the number of previous patents in repeating inventor's portfolios

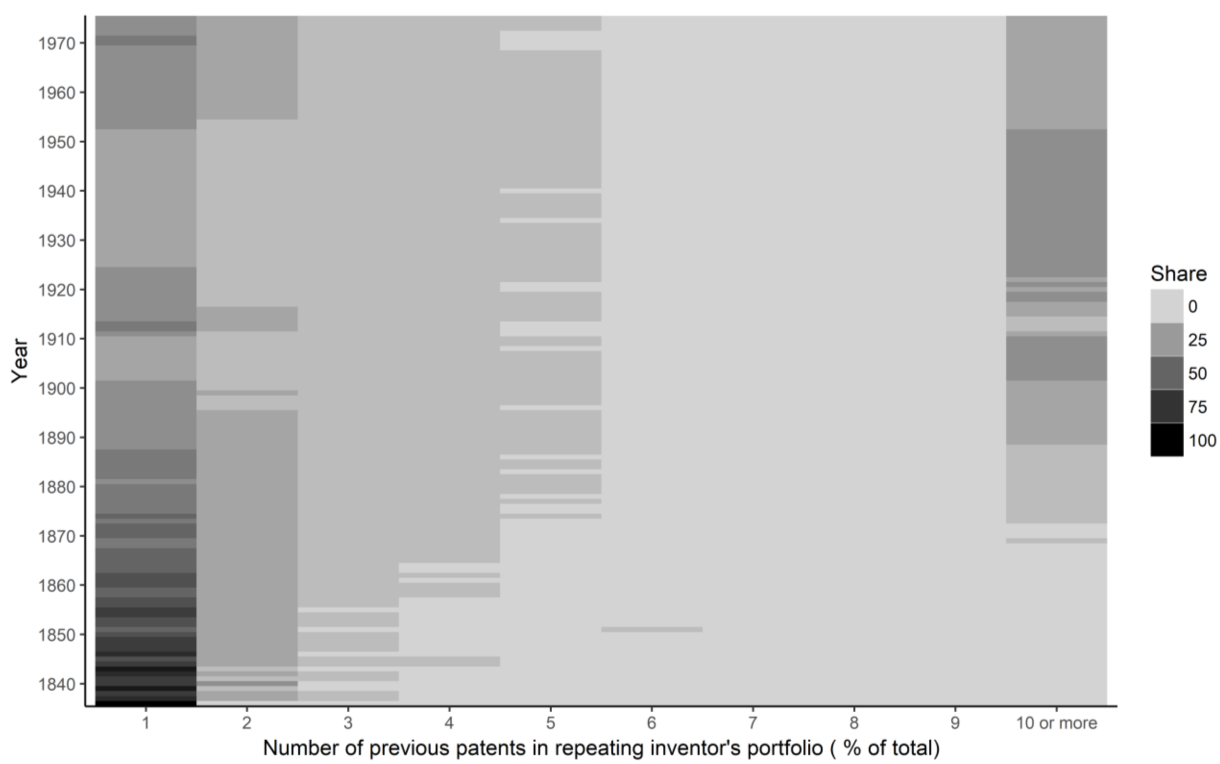
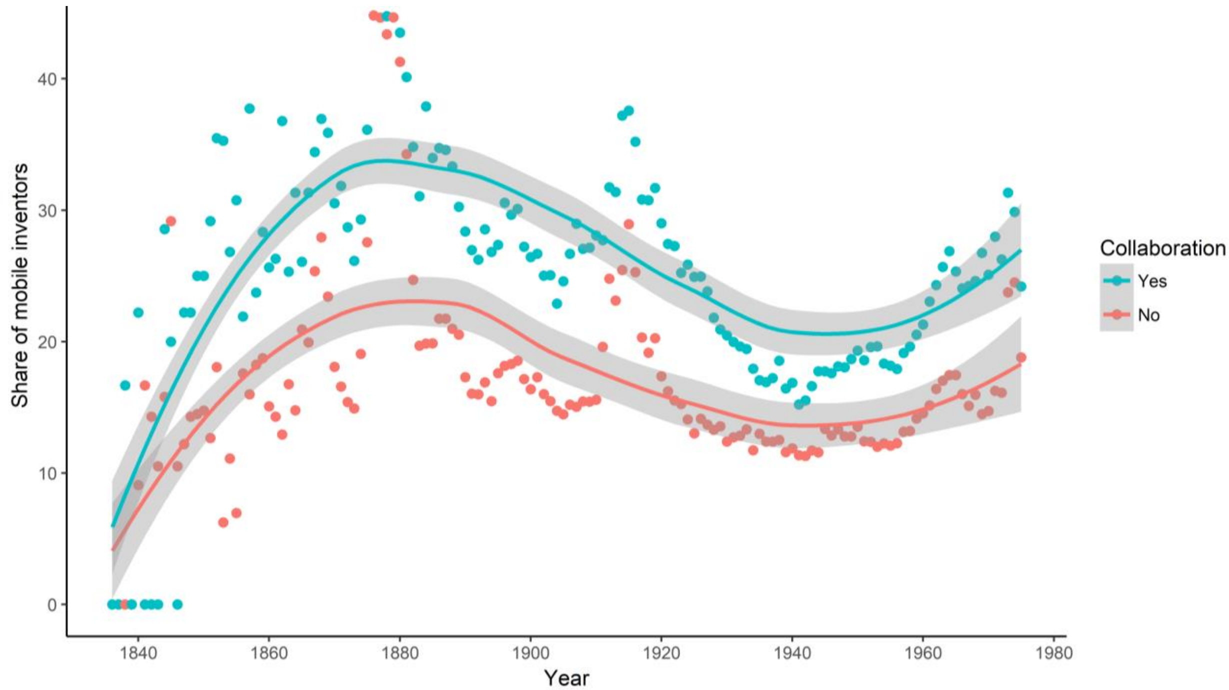
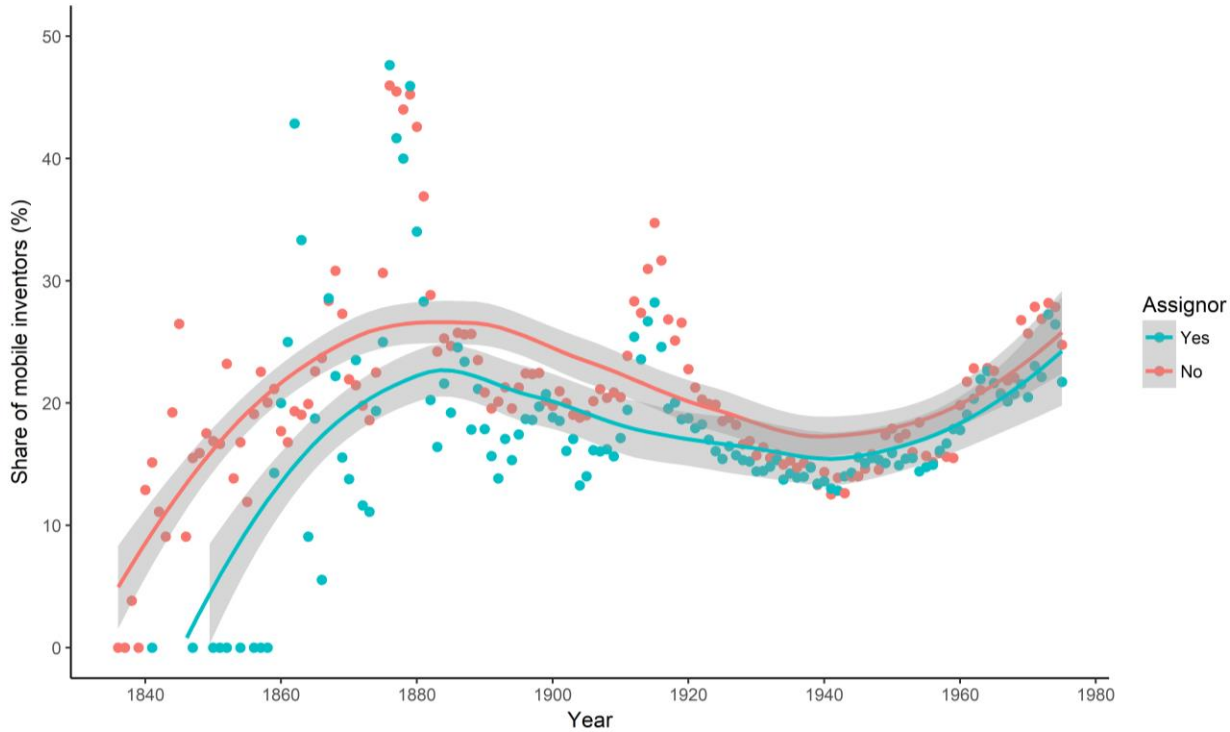


Figure 4.3 Share of mobile inventors by collaboration



Inventors with patents assigned to corporations tend to be less mobile than inventors with unassigned patents. Figure 4.4 demonstrates that only between 1930 and 1950, the mobility of both groups of inventors was relatively similar. After the 1950s, inventors with unassigned patents tend to be more mobile. This likely indicates the greater independence of inventors who are not linked to firms or other organizations. However, the number of independent inventors was dropping dramatically at this time. It is important to note that only individuals are allowed to apply for patents, but the rights to the patent can be assigned to others, including corporations. However, in early years, inventors frequently licensed the use of their patents to others without legally assigning the patent rights. As large corporations began to develop in-house R&D labs questions regarding the ownership of patents produced by inventors on corporate pay-rolls started to emerge. In the 1930s the US Supreme Court ruled that patents produced by inventors on corporate pay-rolls belonged to the firm. These patents were applied for and granted to individual inventors, but assigned to the corporation, on issue. Figure 4.5 shows that the share of patents with a corporate assignee increased significantly during the 1930s. By 1975 about 80 percent of the patents by repeating US inventors were assigned to a corporation, most on issue.

Figure 4.4 Share of inventor mobility by corporate assignee



The patterns of geographical movement of mobile inventors have shifted over time. The most prominent flows of inventors between US cities in consecutive periods of roughly 40 years are plotted in Figures 4.6 to 4.8. In the earliest period (Figure 4.6) the movement occurred mostly within cities in the North-Eastern part of the United States. Of course, this was the most populated area of the US and the region with the highest densities of firms and patents. New York, Boston and Philadelphia are the cities with the largest flows of inventors. Figure 4.7 shows that Chicago had become the second largest source of and destination for inventors, pushing Philadelphia out the top three, in the second period. However, in the third period Philadelphia is back in second place, resulting from the rapid growth of the pharmaceutical industry, followed by Chicago. Figure 4.8 demonstrates that New York is still the location with the largest inventor flows, followed by Philadelphia and Chicago, between 1930 and 1975. Note that Los Angeles became more important than Detroit and Poughkeepsie in terms of the size of inventor flows over this period, indicating the rise of California as a font of knowledge creation after the Second World War.

Figure 4.5 Share of US patents with a corporate assignee

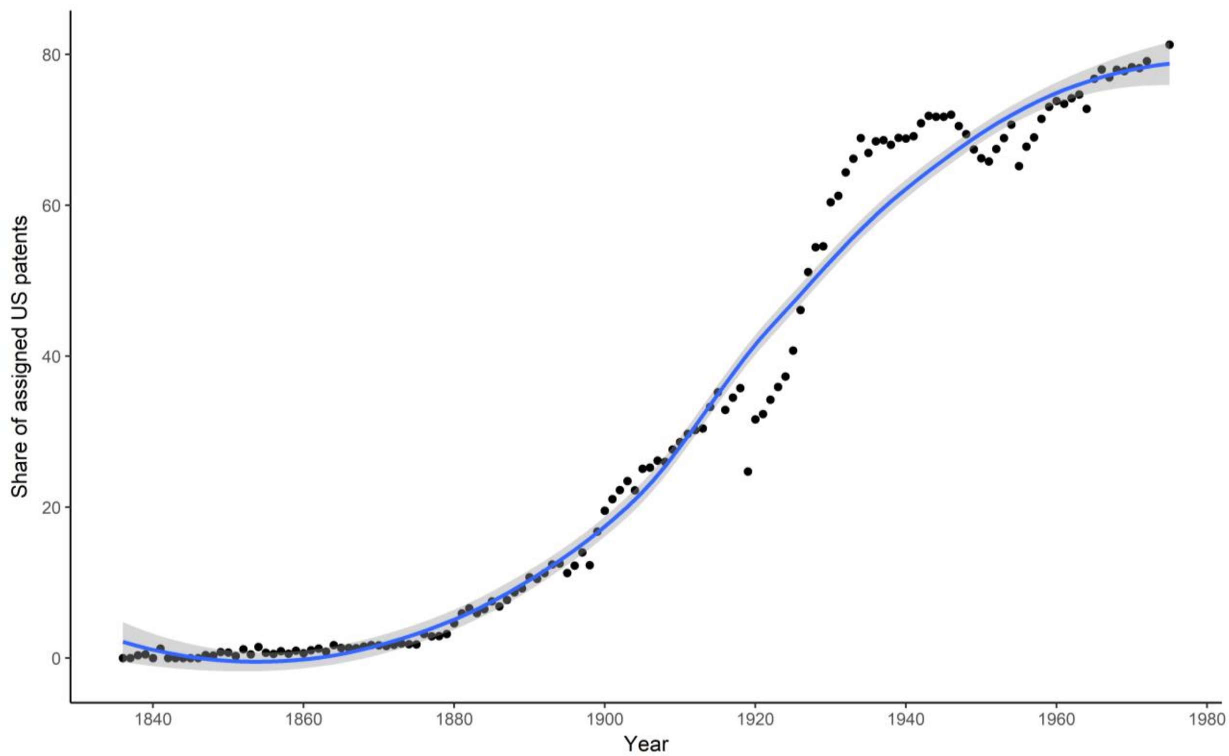
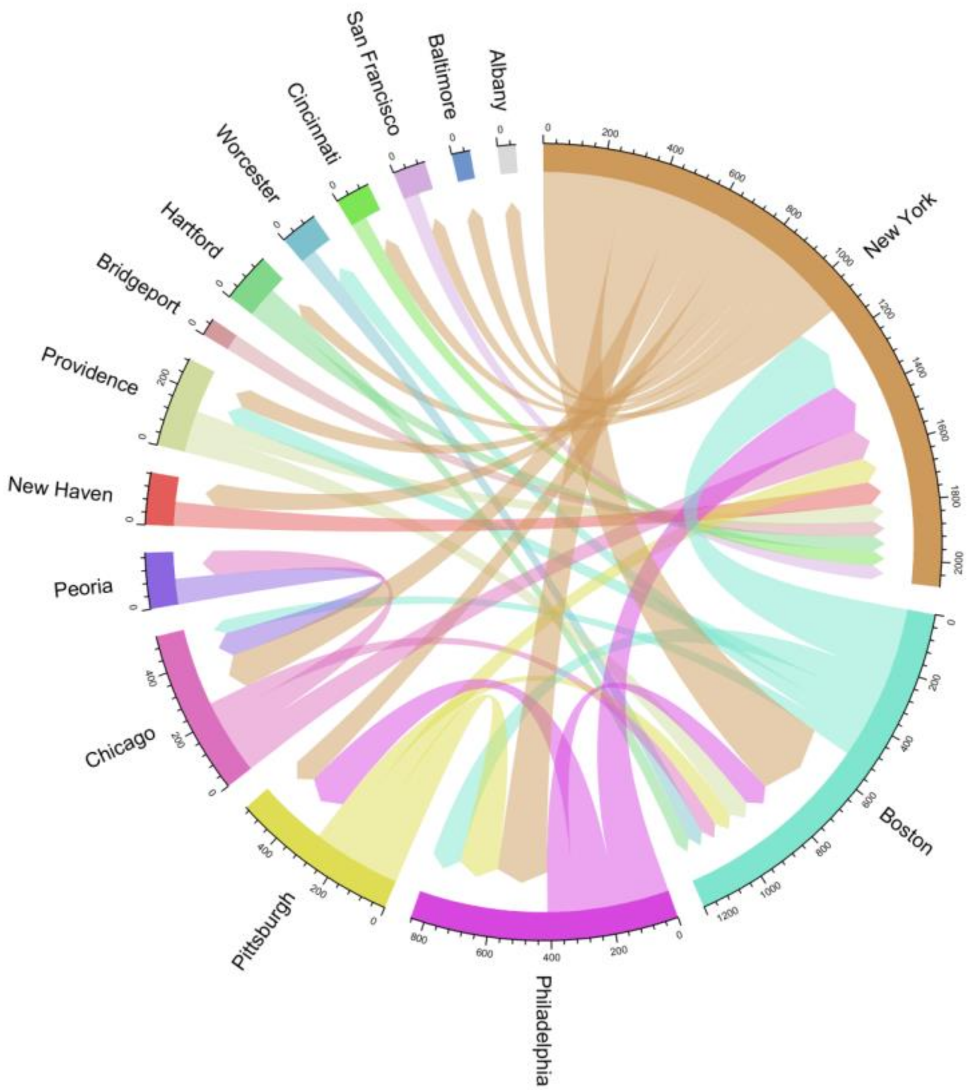
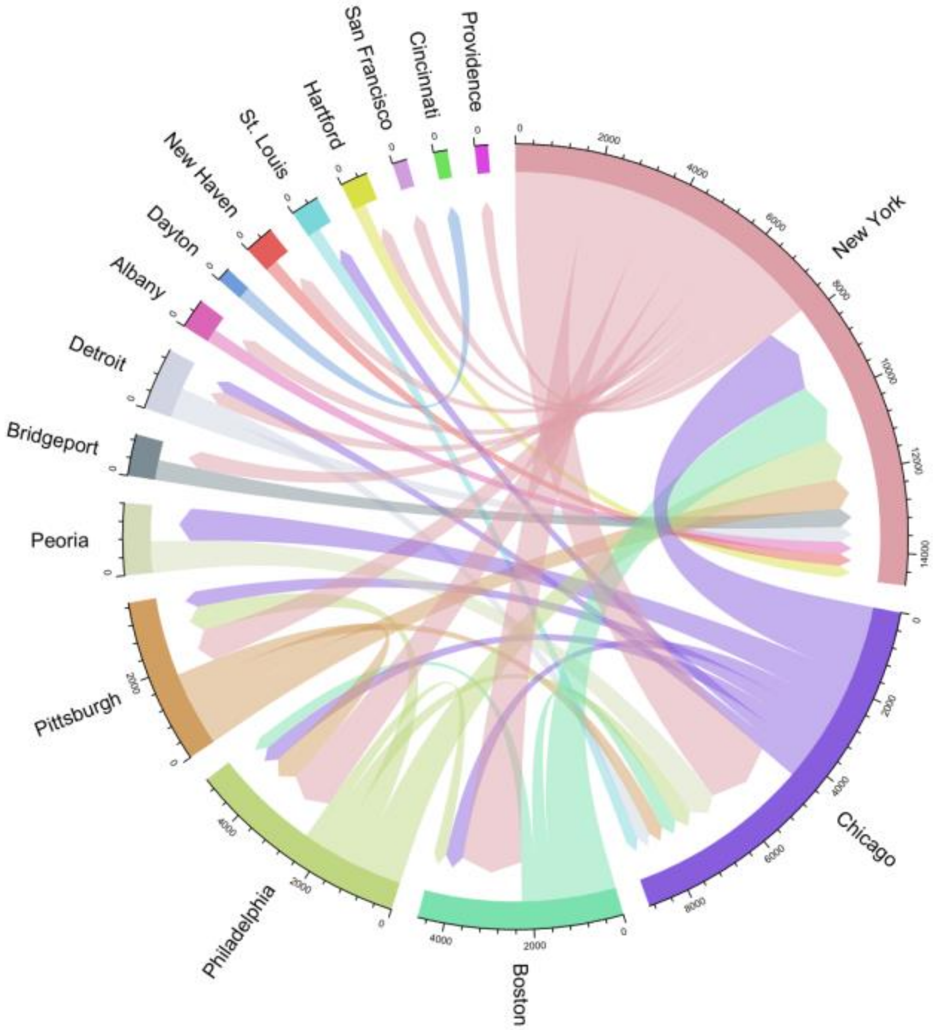


Figure 4.6 Geographical mobility of repeating US inventors between 1836-1889



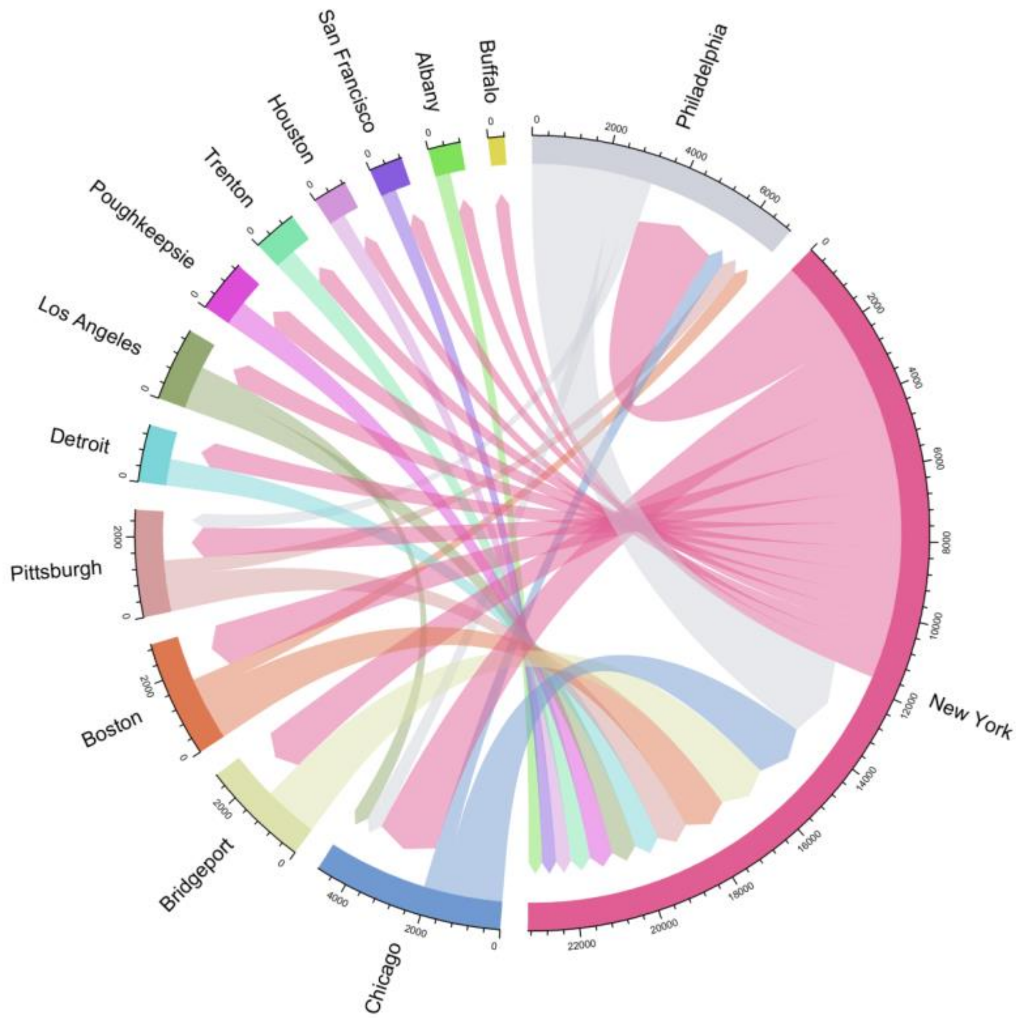
Mobility within previous 5 years & between MSA's
Only flows greater than 50 are plotted

Figure 4.7 Geographical mobility of repeating US inventors between 1890-1929



Mobility within previous 5 years & between MSA's
Only flows greater than 300 are plotted

Figure 4.8 Geographical mobility of repeating US inventors between 1930-1975



Mobility within previous 5 years & between MSA's
Only flows greater than 400 are plotted

The inter-firm mobility of US repeating inventors is dominated by a relatively small number of firms. Figure 4.9 plots the inter-firm mobility flows after 1930 of repeating US inventors who patented at least twice in a time period of five years. DuPont and EW Bliss have exchanged the most inventors. Other companies with relatively large exchange flows are Bell Telephone Laboratories and Western Electric Company. Western Electric was one of the major suppliers of Bell Telephone Laboratories. Components of both companies later became part of AT&T. The flow between Esso Research and Engineering Company and Standard-Oil Development Company is related to the break-up of Standard-Oil.

Figure 4.9 Inter-firm mobility of inventors between 1930-1975

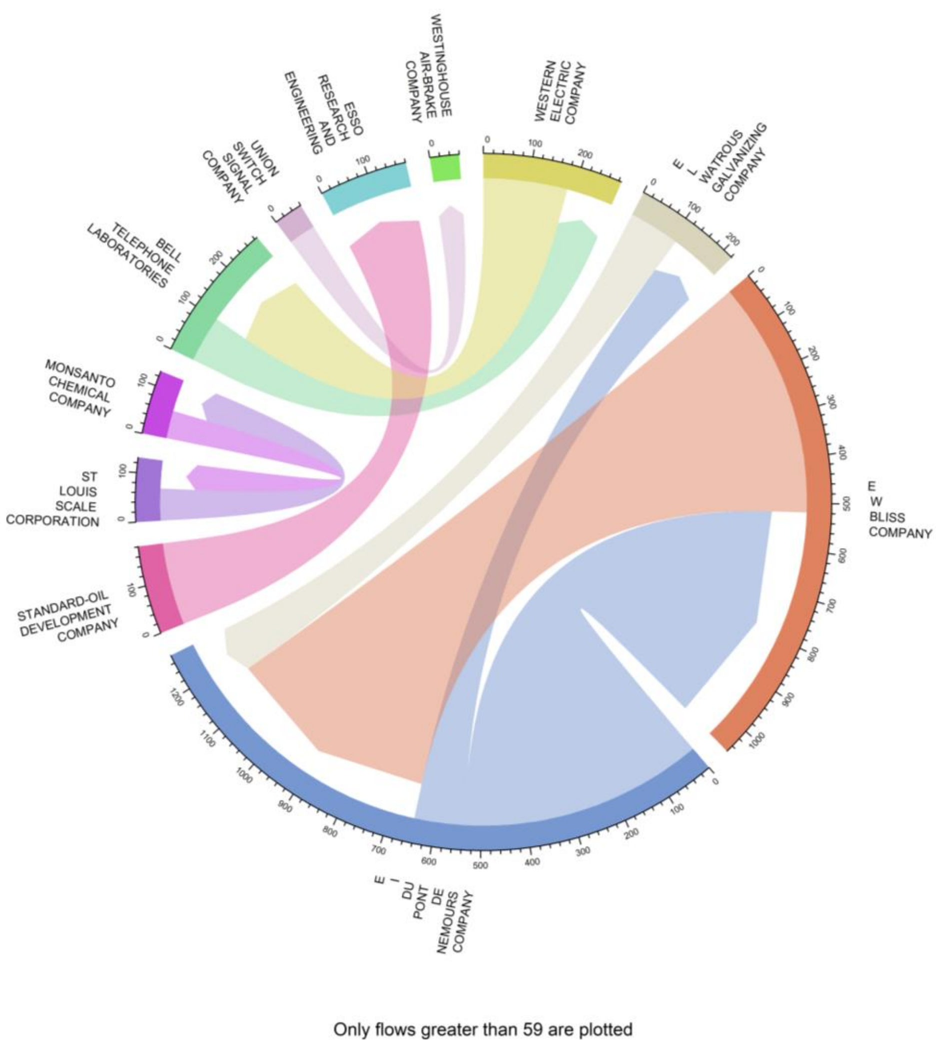
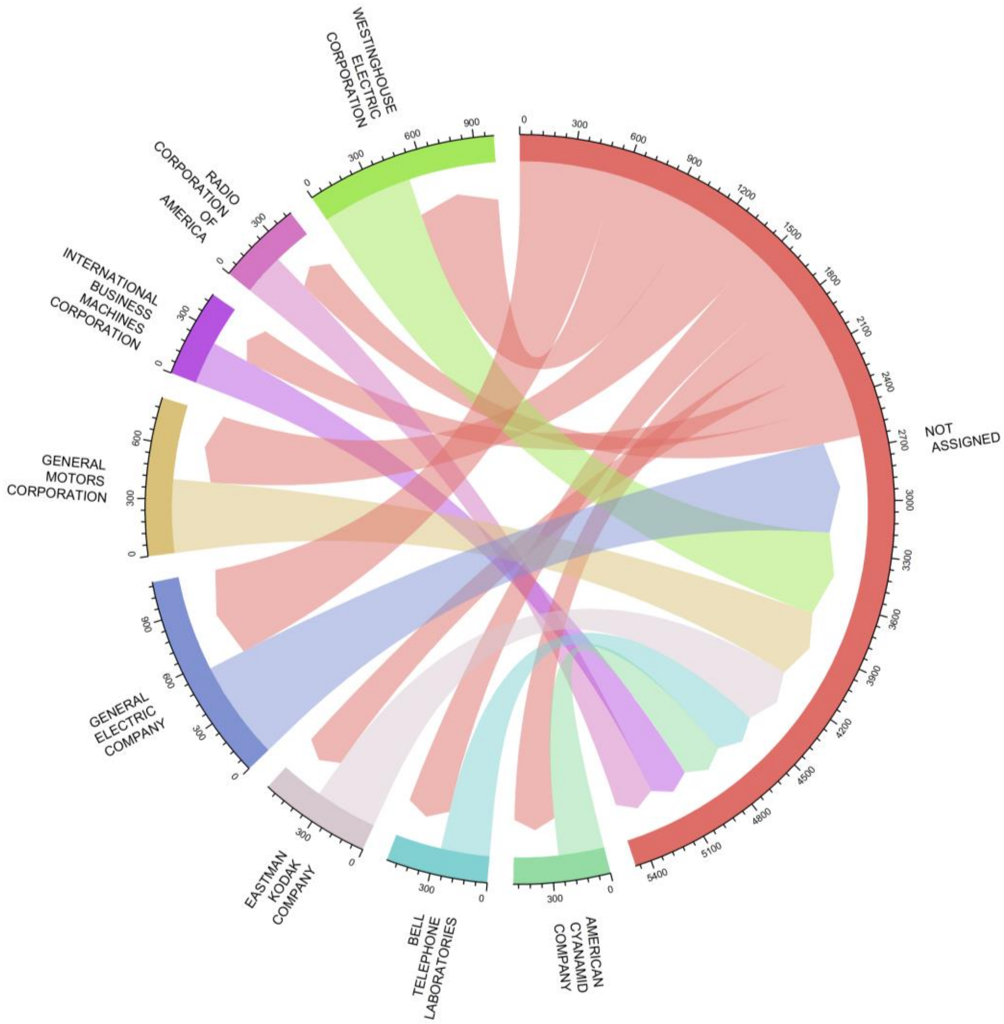


Figure 4.10 plots the major firm mobility flows of inventors with one unassigned and one assigned patent in a five year time-window. General Electric has been granted 540 patents granted to inventors whose previous patent was unassigned. Roughly the same number of inventors (561) produced an unassigned patent after having been granted a patent assigned to General Electric. Although these data don't tell us whether the incoming inventors are the same as the exiting inventors, they do suggest that these inventors are not recruited to patent for General Electric. In the case these inventors are recruited, a similar sized group of inventors is leaved the company and produced their next patent unassigned. This pattern is not unique to General Electric but is observed by all firms in Figure 4.10 and throughout most large firms in the data.

Figure 4.10 Mobility of inventors with unassigned patents between 1930-1975



Only flows greater than 210 are plotted

5. Empirical results

The key result of this paper is that mobility positively affects US inventor productivity over the period 1836 to 1975. Inventors who move between firms, who move between metropolitan areas, or who do both, produce significantly more patents than non-movers in the five years after they are tracked. However, this treatment effect varies with the control variables in the models. In Table 5.1 the main results from the four analyses with matched samples listed in Table 3.2 are presented. In Tables 5.2 through Table 5.5 the results are presented for each of the four analyses and separated for different time periods.

The reader is reminded that the dependent variable in the analysis is the productivity of inventors. Because productivity is measured as a count of the patents generated by repeat inventors, after a focal year when they are identified, the analysis makes use of a Poisson regression model estimated with maximum likelihood techniques. The difference between the mean and variance of productivity values was not sufficiently large to call for use of the negative binomial model. The coefficients in the Poisson model are estimated on the log scale and so need to be exponentiated to be read as those from a standard regression model.

In all four models presented in Table 5.1 mobility has a significant and positive impact on productivity. In model 1 the coefficient for Geographical Move indicates that repeat inventors who change their location, between the time of the patent identified in a focal year and the next earlier patent, generate on average 1.27 more patents in the five years after the focal year compared to inventors who do not change their location. In model 2, productivity for inventors who move between firms is 1.47 patents greater than inventors who do not move between firms across the five years after which the focal patents are identified. These results suggest that movement between firms has a bigger impact on future patent productivity than geographical mobility.

However, it is unlikely that the distribution of covariates in the treatment and control samples for model 1 and 2 are comparable. Therefore, it is difficult to establish precisely whether geographical or firm mobility has a stronger positive impact on productivity. To answer this question, model 3 and model 4 are estimated using samples in which only the control groups differ – the treatment groups are the same. When Double Mobility (firm and geographic move) takes the value of one, representing the treatment, the coefficient indicates the expected change in future patent production compared to the control group. In this case, the control group observations in model 3 are inventors who only experienced geographical mobility. In model 4 the control group consists of inventors who only moved between firms. The larger coefficient in model 3 suggests that firm mobility boosts future patent production more than geographical mobility, because inventors experiencing both have a greater increase in productivity compared to geographical movers (model 3) than to firm movers (model 4).

The effect of the control variables varies along the treatment and control groups in the different models. The coefficients of the three control variables can be interpreted as the log change in the productivity of the control group inventors associated with a one unit increase in the variables. In turn, the coefficients of the interacted control variables correspond to the log change in the productivity of the treatment group inventors.

Table 5.1 Results from statistical models examining the effect of mobility on productivity

Dependent Variable: Productivity	1	2	3	4
Treatment	Geographical Movement	Firm Movement	G & F Movement	G & F Movement
Control	No Geographical Movement	No Firm Movement	Only G Movement No F Movement	Only F Movement No G Movement
Geographical Move (0/1)	0.24***			
	(0.02)			
Firm Move (0/1)		0.39***		
		(0.02)		
Double Move (0/1)			0.36***	0.29***
			(0.10)	(0.08)
Sum of Collaborators	-0.02***	0.02***	0.02***	-0.04***
	(0.001)	(0.001)	(0.01)	(0.002)
Assignor (0/1)	0.50***	0.04**	-0.28***	0.46***
	(0.02)	(0.02)	(0.10)	(0.06)
Previous Patents	0.05***	0.01***	0.04***	0.06***
	(0.001)	(0.0003)	(0.004)	(0.001)
Movement * Sum of Collaborators	0.01***	-0.001	-0.02***	0.02***
	(0.001)	(0.001)	(0.01)	(0.002)
Movement * Assignor	-0.15***	0.21***	0.005	-0.28***
	(0.02)	(0.02)	(0.10)	(0.08)
Movement * Previous Patents	-0.01***	-0.0000	0.01***	-0.01***
	(0.001)	(0.0004)	(0.004)	(0.002)
Fixed effects:				
- MSA	✓	✓	✓	✓
- Year	✓	✓	✓	✓
- Primary Tech. Class	✓	✓	✓	✓
N	29,264	32,026	2,117	7,417

*p < .1; **p < .05; ***p < .01

Across all four models in Table 5.1, inventors in the control groups are more productive if they produced more patents in the past. For these inventors, an increase of one *Previous Patent* is associated with an increase in productivity after the focal year of between 1% and 6%. The influence of higher productivity in the past on the treatment groups in models 1-4 is quite different. Only in model 3 for double-movers does higher past patent productivity lead to greater productivity gains for the treatment group relative to the control. In all other models, higher past productivity does not raise the future productivity of the treatment group over the control group. Looked at differently, these results suggest that more experienced inventors do not benefit as much from mobility as less experienced inventors.

The *Sum of Collaborators* has a negative impact on productivity for control group inventors in models 1 and 4, but a positive effect for the treatment group in these models. These control groups are characterized by inventors without geographical mobility. This indicates that the productivity of inventors without geographical mobility does not increase as a result of having a large number of collaborators, while it does for geographical mobile inventors. Interestingly, this effect is the opposite for firm mobility. Models 2 and 3 report that inventors without firm mobility benefit from a large number of collaborators, while inventors with firm mobility do not. These findings suggest that the benefits of a social network are resilient to geographical mobility but not to firm mobility.

Control group inventors working on patents with an *Assignor* in Model 1 and 4 are more productive than their counter-parts in the treatment group. The inventors in these control groups are characterized by no geographical mobility, suggesting that inventors with no geographical mobility benefit (in terms of increased productivity) from patenting for corporations. Corporate inventors with geographical mobility face decreased productivity. This negative effect is strongest for geographically mobile corporate inventors who remained within the same firm.

The effects of firm and spatial mobility on productivity have changed over time. In Tables 5.2 through Table 5.5 the results are presented for each of the four analyses listed in Table 3.2 and separated for different time periods. Note that in Table 5.2 the analysis of 1910 is missing. In Table 5.3 the year 1910 and 1970 are missing. These years are missing because the Poisson models could not be fit for these samples – there is not enough variation among the covariates in these years.

Table 5.2 shows that the positive effect of geographical mobility on productivity is strongest in the post-World War 2 era. In earlier years there are no significant differences in productivity between inventors who changed location and those who did not. On the other hand, Table 5.3 indicates that firm mobility has had a significant positive effect on inventor productivity throughout the whole time-period, except for the two focal years following the Second World War. The benefits of double mobility compared to only geographical mobility are reported in Table 5.4. The positive effect of additional firm mobility has been the strongest during the early 20th century, save for the Great Depression. Outside this time-window the benefits from additional firm mobility are insignificant. Table 5.5 illustrates that the gains in productivity experienced by inventors with double mobility compared to inventors with only firm mobility is limited to the 1970 sample. This is the only period in which double mobility inventors have significantly greater productivity than those inventors who switched between firms but not geographical locations. This finding, combined with the evidence reported in Table 5.2, suggests that the positive effect of geographical mobility on productivity is strongest in later years of the analysis.

The number of *Previous Patents* has a constant and positive effect on patent productivity for the inventors in the control groups throughout the entire time-period of this study. Table 5.2 through Table 5.5 report that the treatment group produced fewer patents when the number of Previous Patents increased. This finding suggests, again, that experienced inventors benefit less from firm or geographical mobility than inexperienced inventors do.

The effect of *Sum of Collaborators* on the productivity has changed over time. The Sum of Collaborators has a mostly negative impact on productivity for control group inventors in Table 5.2. The treatment group (geographically mobile) inventors benefit from greater numbers of previous collaborators from 1940 onwards. An additional previous collaborator is associated with an expected increase in the patent production of mobile inventors of 5 percent in 1970. In Table 5.3 through 5.5 the effect of the Sum of Collaborators on productivity tends to be negative for the treatment group, but no clear trends can be observed.

Working on an assigned patent has a constant positive effect for inventors without geographical mobility throughout the entire time-window. Table 5.2 and Table 5.5 show that inventors in the control group significantly benefit from an *Assignor* on their patents (except for year 1900 and 1960 in Table 5.5). In the Tables 5.3 and 5.4 no clear trends over time or across groups are discovered.

The findings of Table 5.1 broadly sustained two robustness checks. In the first robustness check the time-window for mobility to occur was changed from 5-years to 3-years and then to 10-years. The results were qualitatively the same as those reported above. In the second robustness check the matching algorithm is changed from coarsened exact matching to an exact matching algorithm. This last robustness check resulted in smaller sample sizes yet the overall results were also consistent with those reported.

Table 5.2 Results from statistical models examining the effect of geographical mobility on productivity

Dependent variable: Productivity

Treatment: Geographical Mobility

Control: No Geographical Mobility

	Total	1900	1920	1930	1940	1950	1960	1970
Geographical Move (0/1)	0.24***	0.07	-0.07	-0.09	-0.01	1.03***	0.22***	1.00***
	(0.02)	(0.06)	(0.05)	(0.07)	(0.09)	(0.07)	(0.07)	(0.11)
Sum of Collaborators	-0.02***	-0.02**	-0.05***	0.0004	-0.02***	-0.01***	-0.01**	-0.04***
	(0.001)	(0.01)	(0.01)	(0.004)	(0.003)	(0.002)	(0.003)	(0.002)
Assignor (0/1)	0.50***	0.24***	0.42***	0.40***	0.74***	0.49***	0.23***	0.57***
	(0.02)	(0.06)	(0.04)	(0.05)	(0.06)	(0.06)	(0.06)	(0.09)
Previous Patents	0.05***	0.07***	0.07***	0.04***	0.06***	0.04***	0.05***	0.08***
	(0.001)	(0.004)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Move * Sum of Collaborators	0.01***	-0.01	0.01	-0.01*	0.02***	0.03***	0.03***	0.05***
	(0.001)	(0.01)	(0.01)	(0.01)	(0.005)	(0.004)	(0.004)	(0.003)
Move * Assignor	-0.15***	-0.06	0.23***	0.04	0.01	-0.81***	-0.07	-0.89***
	(0.02)	(0.08)	(0.06)	(0.07)	(0.09)	(0.07)	(0.07)	(0.11)
Move * Previous Patents	-0.01***	0.004	-0.01***	0.005**	-0.02***	-0.02***	-0.02***	-0.05***
	(0.001)	(0.01)	(0.004)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Fixed effects:								
- MSA	✓	✓	✓	✓	✓	✓	✓	✓
- Year	✓	×	×	×	×	×	×	×
- Primary Tech. Class	✓	✓	✓	✓	✓	✓	✓	✓
N	29,264	1,388	2,562	2,848	3,201	4,158	5,745	6,447

*p < .1; **p < .05; ***p < .01

Table 5.3 Results from statistical models examining the effect of firm mobility on productivity

Dependent variable:	Productivity								
Treatment:	Firm Mobility								
Control:	No Firm Mobility								
	Total	1900	1910	1920	1930	1940	1950	1960	1970
Firm Move (0/1)	0.39^{***}	0.41^{***}	0.34^{***}	0.41^{***}	0.96^{***}	0.72^{***}	0.01	-0.002	0.24^{**}
	(0.02)	(0.07)	(0.10)	(0.04)	(0.06)	(0.06)	(0.07)	(0.06)	(0.10)
Sum of Collaborators	0.02^{***}	-0.05^{***}	-0.05^{***}	-0.11^{***}	-0.01	0.03^{***}	-0.003	-0.0003	0.01^{***}
	(0.001)	(0.01)	(0.01)	(0.01)	(0.01)	(0.003)	(0.003)	(0.003)	(0.004)
Assignor (0/1)	0.04^{**}	0.13^{*}	0.52^{***}	-0.09^{**}	0.34^{***}	0.39^{***}	0.03	-0.26^{***}	-0.88^{***}
	(0.02)	(0.07)	(0.08)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.08)
Previous Patents	0.01^{***}	0.12^{***}	0.05^{***}	0.07^{***}	0.07^{***}	0.01^{***}	0.03^{***}	0.05^{***}	0.02^{***}
	(0.0003)	(0.01)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.005)
Move * Sum of Collaborators	-0.001	-0.003	-0.004	-0.002	-0.02^{***}	-0.02^{***}	-0.01^{*}	0.02^{***}	-0.06^{***}
	(0.001)	(0.01)	(0.01)	(0.01)	(0.01)	(0.004)	(0.004)	(0.003)	(0.004)
Move * Assignor	0.21^{***}	0.07	-0.21^{*}	0.18^{***}	-0.24^{***}	-0.30^{***}	0.38^{***}	0.43^{***}	1.13^{***}
	(0.02)	(0.09)	(0.11)	(0.05)	(0.06)	(0.07)	(0.07)	(0.06)	(0.10)
Move * Previous Patents	-0.0000	-0.03^{***}	0.0001	-0.001	-0.004	0.004^{***}	-0.001	-0.002	0.06^{***}
	(0.0004)	(0.01)	(0.003)	(0.001)	(0.003)	(0.001)	(0.003)	(0.003)	(0.01)
Fixed effects:									
- MSA	✓	✓	✓	✓	✓	✓	✓	✓	✓
- Year	✓	×	×	×	×	×	×	×	×
- Primary Tech. Class	✓	✓	✓	✓	✓	✓	✓	✓	✓
N	32,026	879	3,158	2,448	3,307	4,037	4,901	6,483	6,813

$^{*}p < .1$; $^{**}p < .05$; $^{***}p < .01$

Table 5.4 Results from statistical models examining the effect of double mobility on productivity

Dependent variable:	Productivity						
Treatment:	Firm and Geographical Mobility						
Control:	Only Geographical – No Firm Mobility						
	Total	1900	1920	1930	1940	1950	1960
Double Move (0/1)	0.36***	-0.26	0.93***	0.48	2.37***	1.07**	-0.41
	(0.10)	(0.56)	(0.25)	(0.39)	(0.54)	(0.51)	(0.26)
Sum of Collaborators	0.02***	-0.16	0.18***	-0.13***	0.05***	-0.01	-0.03*
	(0.01)	(0.17)	(0.05)	(0.05)	(0.02)	(0.02)	(0.02)
Assignor (0/1)	-0.28***	-15.83	-35.28	17.24	2.47***	-0.67	-0.30
	(0.10)	(1,343)	(3,398)	(1,957)	(0.71)	(0.69)	(0.74)
Previous Patents	0.04***	0.08	0.03	0.14***	-0.02*	0.09***	0.11***
	(0.004)	(0.09)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)
Move * Sum of Collaborators	-0.02***	0.05	-0.11	0.11**	-0.06***	-0.04	0.02
	(0.01)	(0.27)	(0.08)	(0.05)	(0.02)	(0.02)	(0.02)
Move * Assignor	0.005	-1.85***	-0.83***	0.17	-1.81***	-0.89*	0.81***
	(0.10)	(0.69)	(0.25)	(0.41)	(0.55)	(0.50)	(0.26)
Move * Previous Patents	0.01***	0.02	0.06**	-0.03*	0.01	0.02	-0.02
	(0.004)	(0.10)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
Fixed effects:							
- MSA	✓	✓	✓	✓	✓	✓	✓
- Year	✓	×	×	×	×	×	×
- Primary Tech. Class	✓	✓	✓	✓	✓	✓	✓
N	2,117	43	151	144	196	345	481

$^{*}p < .1$; $^{**}p < .05$; $^{***}p < .01$

Table 5.5 Results from statistical models examining the effect of double mobility on productivity

Dependent variable:	Productivity								
Treatment:	Firm and Geographical Mobility								
Control:	Only Firm – No Geographical Mobility								
	Total	1900	1910	1920	1930	1940	1950	1960	1970
Double Move (0/1)	0.29***	-0.73	0.21	0.32	-0.16	-0.04	0.08	0.06	0.62**
	(0.08)	(1.00)	(0.48)	(0.20)	(0.21)	(0.20)	(0.27)	(0.21)	(0.28)
Sum of Collaborators	-0.04***	-0.02	-0.03	-0.06***	-0.01	0.003	0.0003	-0.02***	-0.08***
	(0.002)	(0.08)	(0.03)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.003)
Assignor (0/1)	0.46***	-3.54***	1.26**	0.58***	0.44**	0.54***	0.58**	0.23	0.73***
	(0.06)	(1.17)	(0.53)	(0.21)	(0.18)	(0.18)	(0.24)	(0.20)	(0.25)
Previous Patents	0.06***	0.04	0.05***	0.09***	0.05***	0.06***	0.03***	0.08***	0.12***
	(0.001)	(0.03)	(0.01)	(0.01)	(0.004)	(0.004)	(0.01)	(0.01)	(0.003)
Move * Sum of Collaborators	0.02***	-0.29***	0.16***	-0.01	0.03***	0.01	-0.01	0.04***	0.09***
	(0.002)	(0.11)	(0.04)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Move * Assignor	-0.28***	0.84	-0.57	-0.24	-0.10	0.06	-0.07	-0.10	-0.66**
	(0.08)	(0.92)	(0.48)	(0.20)	(0.21)	(0.20)	(0.27)	(0.21)	(0.28)
Move * Previous Patents	-0.01***	0.08**	0.03**	0.02*	0.01**	-0.04***	0.01	-0.01	-0.10***
	(0.002)	(0.04)	(0.01)	(0.01)	(0.005)	(0.01)	(0.01)	(0.01)	(0.01)
Fixed effects:									
- MSA	✓	✓	✓	✓	✓	✓	✓	✓	✓
- Year	✓	×	×	×	×	×	×	×	×
- Primary Tech. Class	✓	✓	✓	✓	✓	✓	✓	✓	✓
N	7,417	100	460	480	952	855	895	1,425	2,250

*p < .1; **p < .05; ***p < .01

Conclusion

The main objective of this paper was to examine whether inventor mobility influences the productivity of inventors. Historical patent data from 1836 to 1975 were used to identify all U.S. inventors on patents, the locations of those inventors and whether the patents were assigned on issue. Mobile inventors were defined as those who changed assignee or location over time. Matching procedures were used to link mobile inventors to non-mobile inventors with similar covariates. Comparisons of the patent productivity among these matched samples of mobile and immobile knowledge producers reveals that, in general, firm mobility and geographical mobility raise the future number of patents produced by inventors. Firm mobility has a greater impact than geographical mobility on inventor productivity. However, the impact of both forms of mobility shifts over time. The evidence presented in the paper suggests that geographical mobility raises inventor productivity in the later years of the analysis, while firm mobility played a more active role in raising inventor productivity in the first half of the 20th century.

The main results of this historical analysis of inventor mobility and productivity confirms the findings of those like Hoisl (2007) who report a positive association between mobility and productivity. This paper contributes to the broader literature on inventors and mobility in two ways. First, the analysis describes and examines inventor mobility over a period that has previously been undocumented. Analysis of mobility over the long-run yielded new insights. For instance, the trends in inventor spatial mobility indicate that during the Great Depression and Second World War inventors tended to be less mobile, suggesting that the uncertainties surrounding crises influence the behavior of inventors. Although it is well known that uncertainty affects human behavior (i.e. Becker, 2013; Deci & Ryan, 1985), it has not been documented how uncertainty affects patterns of mobility in U.S. inventors.

Second, using a matched case-control longitudinal sample design, this research presents the first long-run systematic evidence on the positive and significant causal relationship between U.S. inventor mobility and future patent productivity. The design of the analyses facilitates causal claims, because the mobile and non-mobile groups are matched such that there is no relationship or bias between the treatment variable (*Mobility*) and the control variables (Ho et al., 2007; Iacus et al., 2011, 2012).

The presented findings are of relevance to knowledge producers, managers and policy-makers. Individual knowledge producers might be interested to learn that their productivity may rise with mobility. Switching jobs and/or locations has generally boosted the future outputs of U.S. inventors between 1900 and 1970. Managers and other corporate executives might use the evidence of this research to motivate decisions or explicitly design recruitment processes to hire knowledge workers from outside firm or regional boundaries. For policy-makers, the significant and positive causal link between inventor mobility and productivity, reported in this paper, can be used to argue and justify institutional frameworks with liberal migration policies or to question the use of non-compete legislation. More research is required to examine how the productivity gains associated with mobility are distributed across firms.

It is important to note that the findings reported above are subject to a number of limitations. The focus of this research has been limited to U.S. urban inventors. Inventors located outside current U.S. MSAs are excluded from the analysis. Although current U.S. patenting occurs primarily in metropolitan areas, this might not have been the case throughout the eighteenth and nineteenth centuries. Lamoreaux & Sokoloff (2000) report that patenting in the early U.S. glass industry primarily followed the urbanized locations of production, but there was significant variation.

Another limitation originates with the quality of the raw historical patent text files. As these files form the input for the searching, matching and machine learning algorithms, unreadable texts produce errors in the identification of inventors, geography and firms on patents. These errors, in terms of geography or assignee, have immediate consequences for designation of the mobility status of inventors. Although the algorithms tasked with identifying the geography and assignor on a patent are carefully designed and evaluated, mistakes do occur and are likely to inflate the mobility rate. While the mobility rates reported are consistent with the findings of Trajtenberg (2005, 2006) in the U.S. NBER patent data files, they are considerably higher than those generated for the last thirty years or so of U.S. patenting. It is unclear if these differences are the result of recent shifts in inventor mobility or whether they reflect errors in the use of different inventor and assignee disambiguation algorithms.

The evidence presented in this paper indicates there is a causal relationship between inventor mobility and productivity, but does not explain why mobility occurs, which inventors are more likely to be mobile and what the pull and push factors of inventor mobility are. These issues are obvious directions for future research. The current research could be extended by linking the current database to the individual level Census data that cover the first few decades of the twentieth century. Matching inventors to the individual records from the U.S. Decennial Census would also allow additional controls for inventor age, and to proxy measures of their education.

Literature

Abowd, J., Kramarz, F. and D. Margolis (1999). High wage workers and high wage firms. *Econometrica* 67: 251-333.

Acs, Z. and A. Varga (2005). Entrepreneurship, agglomeration and technological change. *Small Business Economics* 24: 323-334.

Agarwal, R. and A. Ohyama (2013). Industry or academia, basic or applied? Career choices and earnings trajectories of scientists. *Management Science* 59: 950-970.

Almeida, P. and B. Kogut (1999). Localization of knowledge and the mobility of engineers in regional n etworks. *Management Science* 45: 905-917.

Arrow, K. (1962). Economic welfare and the allocation of resources for invention. In R.Nelson (ed.), *The Rate and Direction of Inventive Activity: Economic and Social Factors.* Princeton, NJ: Princeton University Press: pp.609-626.

Asheim, B. and M. Gertler (2005). The geography of innovation: regional innovation systems. In Fagerberg, J., Mowery, D. and R. Nelson (eds.) *The Oxford handbook of Innovation*. Oxford: Oxford University Press: pp291-317.

Audretsch, D. and M. Feldman (1996). R&D spillovers and the geography of innovation and production. *American Economic Review* 86: 630-640.

Balsvik, R. (2011). Is labor mobility a channel for spillovers from multinationals? Evidence from Norwegian manufacturing. *Review of Economics and Statistics* 93: 285-297.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management* 17: 99-120.

Becker, G. S. (2013). *The economic approach to human behavior*. University of Chicago press.

Behrens, C., Duranton, G. and F. Robert-Nicoud (2014). Productive cities: sorting, selection, and agglomeration. *Journal of Political Economy* 122: 507-553.

Boschma, R., Eriksson, R. and U. Lindgren (2009). How does labor mobility affect the performance of plants? The importance of relatedness and geographical proximity. *Journal of Economic Geography* 9: 169-190.

Borjas, G. and K. Doran (2012). The collapse of the Soviet Union and the productivity of American mathematicians. *Quarterly Journal of Economics* 127: 1143-1203.

Bosetti, V., Cattaneo, C. and E. Verdolini (2015) Migration of skilled workers and innovation: A European perspective. *Journal of international Economics* 96: 311-322.

Breschi, S. and F. Lissoni (2009). Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *Journal of Economic Geography* 9: 439-468.

Cappelli, R., Czarnitzki, Doherr, T. and F. Montobbio (2018). Inventor mobility and productivity in Italian regions. Regional Studies doi=10.1080/00343404.2018.1445847.

Clemens, M. (2010). A labor mobility agenda for development. Center for Global Development, Working Paper 201. www.cgdev.org.

Dauth, W., Findeisen, S., Moretti, E. and J. Suedekum (2016). Spatial wage disparities – workers, firms, and assortative matching.

Deci, E., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Springer Science & Business Media.

Edler, J., Fier, H. and C. Grimpe (2011). International scientist mobility and the locus of knowledge and technology transfer. *Research Policy* 40: 791-805.

Faggian, A. and P. McCann (2008). Human capital, graduate migration and innovation in British regions. *Cambridge Journal of Economics* 33: 317-333.

Fallah, M., Choudhury, P. and T. Daim (2012). Does movement of inventors between companies affect their productivity? *Technology in Society* 34: 196-206.

Franzoni, C., Scellato, G. and P. Stephan (2014). The mover's advantage: The superior performance of migrant scientists. *Economics Letters* 122: 89-93.

Gorg, H. and E. Strobl (2005). Spillovers from foreign firms through worker mobility: an empirical investigation. *Scandinavian Journal of Economics* 107: 693-709.

Grant, R. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal* 17: 109-122.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis*, *15*(3), 199–236.

Hoekman, J., K. Frenken and R. Tijssen (2010). Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Research Policy* 39: 662-673.

Hoisl, K. (2017). Tracing mobile inventors. The causality between inventor mobility and inventor productivity. *Research Policy* 36: 619-636.

Hoisl, K. and G. de Rassenfosse (2014). Knowledge fit and productivity gains from mobility. Paper presented at the Druid Society Conference, CBS, Copenhagen, 2016.

Hunter, R., Oswald, A., and B. Charlton (2009). The elite brain drain. *Economics Journal* 119: F231-251.

Iacus, S. M., King, G., & Porro, G. (2011). Multivariate Matching Methods That Are Monotonic Imbalance Bounding. *Journal of the American Statistical Association*, *106*(493), 345–361.

Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, *20*(1), 1–24.

Jackson, C. (2013). Match quality, worker productivity, and worker mobility: Direct evidence from teachers. *Review of Economics and Statistics* 95: 1096-1116.

Jaffe, A., Trajtenberg, M. and R. Henderson. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics* 108: 577-598.

Kaiser, U., Kongsted, H. and T. Ronde (2015). Does the mobility of R&D labor increase innovation? *Journal of Economic Behavior and Organization* 110: 91-105.

Kerr, W. (2008). Ethnic scientific communities and international technology diffusion. *Review of Economics and Statistics* 90: 518-537.

Kogut, B. and U. Zander (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science* 3: 383-397.

Lamoreaux, N. R., & Sokoloff, K. L. (2000). The Geography of Invention in the American Glass Industry, 1870–1925. *The Journal of Economic History*, *60*(03), 700–729.

Lawson, C. and E. Lorenz (1999). Collective learning, tacit knowledge and regional innovative capacity. *Regional Studies* 33: 305-317.

Maliranta, M., Mohnen, P. and P. Rouvinen (2009). Is inter-firm mobility a channel of knowledge spillovers? Evidence from a linked employer-employee panel. *Industrial and Corporate Change* 18: 1161-1191.

Marx, M., Strumsky, D. and L. Fleming (2009). Mobility, skills, and the Michigan non-compete experiment. *Management Science* 55: 875-889.

Maskell, P. and A. Malmberg (1999). The competitiveness of firms and regions: 'ubiquitification' and the importance of learning. *European Urban and Regional Studies* 6: 9-25.

Mawdsley, J. and D. Somaya (2016). Employee mobility and organizational outcomes: An integrative conceptual framework and research agenda. *Journal of Management* 42: 85-113.

Miguelez, E. and R. Moreno (2013). Research networks and inventors' mobility as drivers of innovation: Evidence from Europe. *Regional Studies* 47: 1668-1685.

Nathan, M. (2014). The wider economic impacts of high-skilled migrants: a survey of the literature for receiving countries. IZA Journal of Development and Migration. https://doi.org/10.1186/2193-9039-3-4.

Niebuhr, A. (2010) Migration and innovation: does cultural diversity matter for regional R&D activity? *Papers in Regional Science* 89: 563-585.

Owen-Smith, J. and W. Powell (2004). Knowledge networks as channels and conduits: The effects of spillovers in the Boston biotechnology community. *Organization Science* 15: 5-21.

Palomeras and Melero (2010). Markets for inventors: Learning-by-hiring as a driver of mobility. *Management Science* 56: 881-895.

Peri, G. (2012). The effect of immigration on productivity: Evidence from U.S. states. *Review of Economics and Statistics* 94: 348-358.

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*.

Poole, J. (2009). Knowledge transfers from multinational to domestic firms: evidence from worker mobility. *Review of Economics and Statistics* 95: 393-406.

Rahko, J. (2017). Knowledge spillovers through inventor mobility: the effect on firm-level patenting. *Journal of Technology Transfer* 42: 585-614.

Romer, P. (1986). Increasing returns and long-run growth. *Journal of Political Economy* 94: 1002-1037.

Romer, P. (1990). Endogenous technological change. *Journal of Political Economy* 98: S71-S102.

Rosenkopf, L. and P. Almeida (2003). Overcoming local search through alliances and mobility. *Management Science* 549: 751-766.

Samila, S. and O. Sorenson (2011). Noncompete covenants: Incentives to innovate or impediments to growth. *Management Science* 57: 425-438.

Saxenian, A. (1994). *Regional Advantage*. Cambridge, MA: Harvard University Press.

Saxenian, A. (2005). From brain drain to brain circulation: Transnational communities and regional upgrading in India and China. *Studies in Comparative International Development* 40: 35-61.

Saxenian, A. (2007). *The New Argonauts. Regional Advantage in a Global Economy*. Cambridge, MA: Harvard University Press.

Schilling, M. and C. Phelps (2007) Interfirm collaboration networks: The impact of large-scale network structure on innovation. *Management Science* 53: 1113-1126.

Song, J., Almeida, P. and G. Wu (2003) Learning-by-hiring: When is mobility more likely to facilitate interfirm knowledge transfer? *Management Science* 49: 351-365.

Stephan, P. and S. Levin (2001). Exceptional contributions to US science by the foreign-born and foreign-educated. *Population Research and Policy Review* 20: 59-79.

Stephan, P., Gurmu, S., Black, G. and A. Sumell (2007). Who's patenting in the university? *Economics of Innovation and New Technology* 16: 71-99.

Stuen, E., Mobarak, A. and K. Maskus (2012). Skilled immigration and innovation: evidence from enrollment fluctuations in U.S. doctoral programs. *Economic Journal* 122: 1143-1176.

Topel, R. and P. Ward (1992). Job mobility and the careers of young men. *Quarterly Journal of Economics* 107: 439-479.

Trajtenberg, M (2005). Recombinant ideas: the mobility of inventors and the productivity of research. *Proceedings of the CEPR Conference - Munich.*

Trajtenberg, M., Shiff, G. And R. Melamed (2006). The "names game": harnessing inventors' patent data for economic research. *NBER Working Paper No. 12479, Cambridge, MA.*

Van der Wouden, F. (2018). Co-inventors on US patents: Changing patterns in Collaboration, Complexity and Geography. *Working Paper – Under Review*

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal* 5: 171-180.

Zucker, L., Darby, M. and J. Armstrong (1998).Geographically localized knowledge: spillovers or markets? *Economic Inquiry* 36: 65-86.

Zucker, L. and M. Darby (2006). Movement of star scientists and engineers and high-tech firm entry. National Bureau of Economic Research Working Paper No. 12172.

Zwick, T., Frosch, K., Hosil, K. and D. Harhoff (2017). The power of individual-level drivers of inventive performance. *Research Policy* 46: 121-137.

Conclusion

This dissertation aimed to shed light on the structures of inventor collaboration in U.S. cities between 1836 and 1975. It has examined these structures from different perspectives. The first analytical chapter examined the relationships between complexity, collaboration and geography. The second of the analytical chapters investigated the mechanisms structuring collaboration between US inventors. In the third main chapter of the dissertation, the links between inventor mobility were explored. The results from this research have generated a number of novel insights into the characteristics of collaboration among metropolitan inventors in the U.S. These findings are discussed below before a brief report on the shortcomings of the analysis and some directions for future research.

The research presented in the first core chapter of the dissertation shows increasing rates of collaboration on US patents between 1930 and 1975. Before the 1930s, rates of collaboration were relatively stable over time. The increasing rates between 1930 and 1975 correspond to the rates of collaboration reported for post-1975 US patents by Wuchty et al. (2007). Although this dissertation did not investigate why collaboration increased after the 1930s, developments in the US economy and legal institutions might explain this trend. In 1933 the US Supreme Court ruled that the rights of technologies developed by inventors hired-to-invent should be directly assigned to the firm. This provided incentives for firms to recruit inventors and assign them to technological issues, instead of purchasing licenses of patented technologies by privately operating inventors. As a result, the firm becomes the organizational unit in which most inventions are developed and many individual inventors become team scientists collaborating within the firm rather than competing in the private market selling their technologies. In addition to these institutional changes, infrastructure and transportation developments throughout the 19th and 20th century eased the friction of distance. By the 1920s almost all major US cities had access to the railroads. In 1930 about 40% of US homes had a telephone connection and 20% of Americans had access to an automobile. All these developments fostered collaboration by lowering the costs of transport for inventors and for their ideas.

The observed positive relationship between knowledge complexity and collaboration on US patents provides empirical evidence for the theoretical claim that complexity and collaboration are significantly correlated. The evidence suggests that the impact of complexity on the odds of a patent being generated through collaboration increases markedly after the 1940s. This could be interpreted as the shift to 'big science' after World War II, in which the complexity of knowledge production accelerated. The growing complexity of new knowledge demanded rising levels of specialization and a deeper division of labor. Individual inventors did not embody all the knowledge required to develop these complex new ideas and inventor collaboration increased rapidly.

Another result of the first analytical chapter concerns the geography of collaboration. Although metropolitan collaboration occurs mostly between inventors from the same city, increasing complexity decreases the odds of between-city collaboration. This suggests that the production of complex knowledge relies more strongly on the geographical co-location of inventors than does less complex knowledge. Local collaboration allows spontaneous encounters, repeated face-to-face meetings, and other interactions facilitated by the *local buzz*, more easily than non-local collaboration. This local buzz might be particularly important for the production of complex knowledge because it relies on difficult to diffuse tacit knowledge: geography matters.

The second of the core research chapters presents evidence that geographical distance has a negative effect on tie-formation between inventors. This finding has already been well documented in the literature (Cassi & Plunket, 2015; Crescenzi, Nathan, & Rodríguez-Pose, 2016; Ter Wal, 2013). However, what is novel is that the evidence presented here shows a large decrease in the negative impact of the spatial separation between inventors throughout the 19th century, but little change in the 20th century. This is surprising because the rapid developments in telecommunication and transport technologies during this century should have reduced the friction of geographical distance. Throughout the entire time-window of this study, geographical distance negatively affects tie-formation. It is important to note that these transport developments were taking place at the same time that new ideas were becoming increasingly complex. Thus, it is possible to conclude that advances in communication and transportation developments seem not to have been able to match the benefits of geographical proximity at this time of rising knowledge complexity (see Audretsch & Feldman, 1996; Feldman & Kogler, 2010; Jaffe et al. 1993; Storper & Venables, 2004).

Not all types of social connections between inventors have a positive effect on collaboration. Interestingly, only *after* the 1930s does having second-order connections consistently affect collaboration positively. The formation of direct ties between second-order connections is referred to as the process of triadic closure and leads to clustering in the inventor network. One possible explanation can be found in the rapid increase in US inventor collaboration after the 1930s, reported in the first chapter. This trend generated a large number of connections throughout an expanding pool of inventors that slowly becomes more experienced with collaboration. This increases the overall likelihood of cooperation across the pool of potential collaborators.

Another finding reported in chapter three is related to technological or cognitive proximity. In times of uncertainty, U.S. inventive collaborative behavior is shown to be different from other years. Inventors are more likely to collaborate with inventors with similar knowledge stocks than in times with less uncertainty. During crises, individuals might be less likely to experiment because it is unclear if society will appreciate their experiments. Instead, inventors (and their companies) are more likely to make traditional, less risky decisions and collaborate with inventors with similar technological skills. Indeed, inventors might choose to exploit familiar technologies rather than to experiment with new technologies (Granovetter, 2005; March, 1991; Schumpeter, 1934).

These results are relevant for policymakers and corporate executives concerned with innovation policies. Geographical proximity is still an important driver of collaboration. This suggests that the significance of agglomerating economic activities in space, even after controlling for telecommunication and transportation technologies, social connections and technological proximity. However, simply clustering economic activities might not be enough. Promoting interactions between inventors, either first- or second-order, significantly increases the probability of collaboration. Thus, policy-makers and executives should focus their efforts to facilitate stimulating environments that promote dense social networks. Promoting individuals to interact and engage with previous collaborators and their close connections will most likely lead to successful future collaborative output, ceteris paribus. Policy-makers can boost experimentation and possibly influence the direction of knowledge production by providing environments with low degrees of uncertainty. The evidence presented here suggests that in situations of low uncertainty, inventors are more risk-taking in terms of experimental technological collaborations. Thus, policies reducing the uncertainties of knowledge production could result in less path dependent and more diversified outcomes.

The main result of the fourth chapter of the dissertation, the third of the analytical chapters, is that there is a significant positive relationship between inventor mobility and productivity. This evidence confirms the findings of those like Hoisl (2007) who report similar associations. Analysis of mobility over the long-run showed trends in inventor spatial mobility indicating that during the Great Depression and Second World War inventors tended to be less mobile. This suggests, again, that the uncertainties surrounding crises influence the behavior of inventors. Although it is well known that uncertainty affects human behavior (i.e. Becker, 2013; Deci & Ryan, 1985), the evidence presented here is the first showing how uncertainty affects patterns of mobility among U.S. inventors.

Moreover, chapter four presented evidence on a causal relationship between mobility and productivity. Using a matched case-control longitudinal sample design, the evidence presented reports a positive and significant causal relationship between U.S. inventor mobility and future patent productivity. The future knowledge production of individual knowledge producers (i.e. inventors) benefits from mobility. Switching jobs and/or locations has generally boosted the future outputs of U.S. inventors between 1900 and 1970. Managers and other corporate executives can use this evidence to construct policies motivating decisions or explicitly design recruitment processes to hire knowledge workers from outside firm or regional boundaries. The significant and positive causal link between inventor mobility and productivity can be used by policy-makers to argue for institutional frameworks with liberal migration policies or to question the use of non-compete legislation.

This dissertation has a number of limitations. Perhaps the most important limitation is that not *all* patents granted by the USPTO are included in the new inventor-patent database. In some years only 80 percent of the patents have at least one identified inventor. Although this likely has no direct effects on the results presented here, for it does not look like the omitted patents are significantly different from those examined, the data are still incomplete. There are numerous reasons why some patents have missing data, but most common is the quality of the raw patent text files. Some of the original patent files are handwritten and difficult to decipher, even by the advanced optical character recognition software of Google. As a result, the search and (fuzzy) match algorithms are not able to detect names or locations in (parts) of the text files. Advances in both OCR and text mining software will most likely overcome these issues in the future.

The focus of this dissertation has been on U.S. metropolitan collaboration. This limitation is also related to the poor quality of the input data. It has proven to be more complicated to get non-urban geographies correct. The main reason is that are a lot of duplicate geographical locations with the same name and sometimes in the same state. To disambiguate address data, information on the county is required, but not always present or readable. Urban geographies tend to be more uniquely named and easier to correctly identify. The urban focus introduces at least two biases. The first is straightforward – the results can only be generalized to urban innovative processes. The second bias is that due to this urban focus, probably more early-year than later-year patents are ignored. While current patent production is predominately an urban phenomenon, this is not necessarily the case in the 19^{th} century or early 20^{th} century. Work by Lamoreaux & Sokoloff (2000) and others suggest that non-urban patenting has been more common in early years than it is towards the post-Second World War.

This dissertation has excluded patents from where at least one inventor resides outside the United States. As such, this dissertation focuses solely on patenting and collaboration taking place within the United States. In early years, this focus has little impact because international collaboration is not likely to occur due to the limited transportation and communication technologies. However, these technologies improve considerable over the time-window of this study, resulting in an increasing bias over time. At this point it is unclear how the bias against patents with some foreign collaboration affects the results of this study. Future research could extend this work by including rural and foreign patenting and collaboration to the analysis.

The disambiguation of inventors has been one of the most challenging tasks of this dissertation. The availability of individual level characteristics that can be used to disambiguate inventors has been severely limited. The disambiguation could have been done faster and more accurately if more individual level data could have been mined from the patent records. Unfortunately, at the time of the completion of this work, no additional inventor data were available. Newer work on patent assignment might help improve future disambiguation efforts.

The complex interplay between individual, firm and regional level processes provide interesting future research opportunities. While this dissertation has shown that there is a distinct geography to urban patenting and collaboration, the role of the firm in urban patenting and collaboration has received minimum attention and could be explored further. It is unclear how corporations mitigate the mobility of inventors, influence the (geographical) patterns of collaboration and steer directions of knowledge production. Although firms are associated with roughly 80 percent of the patents in the 1970s their motives, behavior and influence in the knowledge production processes remain largely a black-box. Other interesting paths for future research could examine the long-run structures and patterns of collaboration in other fields of knowledge production. Finally, there is much scope in the data already assembled to explore sectoral, geographical and historical changes in the nature of inventor collaboration.

Literature

Audretsch, D. B., & Feldman, M. P. (1996). R & D Spillovers and the Geography of Innovation and Production. *The American Economic Review*, *86*(3), 630–640.

Becker, G. S. (2013). *The economic approach to human behavior*. University of Chicago press.

Cassi, L., & Plunket, A. (2015). Research Collaboration in Co-inventor Networks: Combining Closure, Bridging and Proximities. *Regional Studies*, *49*(6), 936–954.

Crescenzi, R., Nathan, M., & Rodríguez-Pose, A. (2016). Do inventors talk to strangers? on proximity and collaborative knowledge creation. *Research Policy*, *45*(1), 177–194.

Deci, E., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Springer Science & Business Media.

Feldman, M. P., & Kogler, D. F. (2010). *Stylized facts in the geography of innovation*. *Handbook of the Economics of Innovation* (1st ed., Vol. 1). Elsevier BV.

Granovetter, M. (2005). The impact of social structure on economic outcomes. *Journal of Economic Perspectives*, 33–50.

Hoisl, K. (2007). Tracing mobile inventors—The causality between inventor mobility and inventor productivity. *Research Policy*, *36*(5), 619–636.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics*, *108*(3), 577–598.

Lamoreaux, N. R., & Sokoloff, K. L. (2000). The Geography of Invention in the American Glass Industry, 1870–1925. *The Journal of Economic History*, *60*(03), 700–729.

March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, *2*(1), 71–87.

Schumpeter, J. A. (1934). *The Theory of Economic Development: An Inquiry Into Profits, Capital, Credit, Interest, and the Business Cycle*. Harvard University Press.

Storper, M., & Venables, A. J. (2004). Buzz: Face-to-face contact and the urban economy. *Journal of Economic Geography*, *4*(4), 351–370.

Ter Wal, A. L. J. (2013). The dynamics of the inventor network in german biotechnology: Geographic proximity versus triadic closure. *Journal of Economic Geography*, *14*(3), 589–620.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036–9.