Papers in Evolutionary Economic Geography

13.11

rKnowledge: The Spatial Diffusion of rDNA Methods

Maryann P. Feldman, Dieter F. Kogler, David L. Rigby

Utrecht University
Urban & Regional research centre Utrecht

rKnowledge: The Spatial Diffusion of rDNA Methods

Maryann P. Feldman
Department of Public Policy, University of North Carolina, Chapel Hill, United States

Dieter F. Kogler
School of Geography, Planning & Environmental Policy, University College Dublin, Ireland

David L. Rigby
Department of Geography, University of California, Los Angeles, United States

Abstract:

The 1980 patent granted to Stanley Cohen and Herbert Boyer for their development of rDNA technology played a critical role in the establishment of the modern biotechnology industry. From the birth of this general purpose technology in the San Francisco Bay area, rDNA-related knowledge diffused across sectors and regions of the U.S. economy. The local absorption and application of rDNA technology is tracked across metropolitan areas with USPTO patent data. The influence of cognitive, geographical and social proximity on the spatial diffusion of rDNA knowledge is explored using event history and panel models.

JEL Codes: M13, O31, O32, and O34

Keywords: Evolutionary Economic Geography, Technology Evolution, Knowledge Recombination and Diffusion, Patent Analysis, General Purpose Technology, rDNA Method.

Correspondence Address: Maryann P. Feldman, Department of Public Policy, 209 Abernethy Hall, University of North Carolina, Chapel Hill, United States. Email: maryann.feldman@unc.edu. Authors are listed alphabetically.

Introduction

In December 1980, the United States Patent and Trademark Office (USPTO) issued a patent entitled *Process for Producing Biologically Functional Chimeras* (#4237224). The patent covered the recombinant DNA (rDNA) technique developed by Dr. Stanley Cohen of Stanford University, California and Dr. Herbert Boyer of the University of California, San Francisco. In the evaluation of the patent the USPTO introduced a new technology sub-class to the patent classification system – a relatively rare occurrence signaling the birth of a new technology. While most new technological innovation is incremental, certain discoveries provide fundamental breakthroughs that transform industrial activity and provide a platform for increased productivity throughout the economy. rDNA has characteristics of a general-purpose technology in its ability to be incorporated with a variety of other technologies and sectors in subsequent discoveries (Feldman and Yoon, 2012). Theory argues that innovative activity tends to cluster in regions where resources relevant to the performance and survival of firms are most abundant, including the presence of skilled labor and access to transportation (Krugman, 1991), proximity to markets and input suppliers (Baum and Haveman, 1997; Storper and Christopherson, 1987), the presence of universities and research organizations (Zucker et al., 1998), and cultural and institutional supports for entrepreneurial activity (Saxenian, 1994; Sorenson and Audia, 2000). Yet the influence of prior technological expertise and the ways in which information is incorporated into existing expertise has not been considered in a systematic way.

This paper begins to fill an empirical gap in the understanding of both innovative clusters and technological change by examining one technology and its diffusion over time and geographic space. Working from an evolutionary economic geography framework, we explore the influence of geographical proximity, social proximity and cognitive proximity on the diffusion of new knowledge across U.S. cities. Our focus on Cohen-Boyer's rDNA technology has the advantage that the boundaries of the technology are well defined. The analysis is based on approximately 9,000 patents in the USPTO class 435/69.1, which was created in 1989 as a new technology class. Since the patent for rDNA was granted to Cohen and Boyer in 1980, rDNA technology diffused across the United States, spawning new biotechnology firms, new specializations within emerging firms, and new regional clusters of biotechnology research and economic development. Our results suggest that geographical proximity had little impact on the flow of rDNA knowledge. Diffusion of the Cohen-Boyer technology was directed mainly by social contacts developed among co-inventors and by the absorptive capacity of host cities, captured in the form of their knowledge cores and the cognitive proximity of those cores to the knowledge in class 435/69.1.

Technology Evolution and Diffusion in an Evolutionary Framework

It has become convention to characterize the history of technological change as comprising long periods of more or less constant incremental improvement punctuated by aperiodic bursts of basic discovery and innovation that usher in new knowledge systems and that shift parts of the economy to new planes of development (Nelson and Winter, 1982; Schumpeter, 1911). The temporal lumpiness of basic innovation is mirrored by its uneven geography, with islands of innovation emerging from the economic landscape, sometimes remote and sometimes connected via heterogeneous social and economic networks, that bloom and wither as economic agents compete within, and simultaneously shape the evolution of, the capitalist space economy. Schumpeter (1942) recognized that basic innovation created surges of investment activity that generated long cycles of economic growth and resulted in the simultaneous creative and destructive effects associated with innovation. These creative gales change the technological and economic advantage both within, and between, places.

Over the course of history there have been many attempts to identify and define technologies that are radical and to separate them from innovations that are more incremental (Sahal, 1981; Dosi, 1982; Nelson and Winter, 1982; Abernathy and Clark, 1985; Clark, 1985). Interest in breakthrough inventions focuses on their role in creating private wealth (Harhoff et al., 1999) while at the same time generating social benefits (Trajtenberg, 1990), but more fundamentally on the way in which they hold the potential to transform the economic landscape (Christensen 1997). Helpman (1998) and Lipsey et al. (2005) argue that these transformative powers reside in the broad applicability of many breakthrough innovations that they characterize as General Purpose Technologies (GPT). Although the history of development of some GPT is well-known (see for example, Fogel, 1964; Fishlow 1965), isolating the introduction of a GPT to the economy and studying its subsequent diffusion across the economy has proven difficult (Phene et al., 2006; Kerr, 2010). Feldman and Yoon (2012) argue that the Cohen-Boyer class of patents provides an example of a GPT in routine science. To date, the factors that influenced the diffusion of this breakthrough technology remain largely unexplored.

How do we explain the inconstant geography and history of technological advance? A starting point is acknowledging the difficulty of knowledge creation. Ideas and knowledge are complex goods and Edison's aphorism aside, a precise recipe for their production is unknown. However, with the advent of intellectual property rights protection, knowledge production has become increasingly commodified (Lamoreaux and Sokoloff 1996), and a critical dimension of competition (Lichtenberg and Philipson, 2002). Nonetheless, the risk and the attendant high cost of knowledge creation cannot be borne by all firms (Audretsch et al., 2002). The search for new technology is highly specialized reflecting the resources and knowledge capabilities of individual economic agents and their partners (Wenerfelt 1984; Barney 1991; Kogut and Zander 1992), the maturity of the industries within which they compete (Abernathy and Utterback 1978; Klepper 1997), and the broader ecology of the places where they are located (Cooke et al. 1997; Morgan 1997; Storper 1997; Gertler 2003; Asheim and Gertler, 2005).

Spatial variations in the creation of knowledge and competitive advantage are well-known (Feldman 1994; Maskell and Malmberg, 1999, Feldman and Kogler, 2010). This heterogeneity reflects the pool of private assets and capabilities created by distinct assemblages of firms, workers and institutions in different locations, and by the capacity of these assemblages to develop localized forms of social capital (Saxenian 1994; Storper 1997; Feldman and Zoller 2012). In relatively thin geographical extensions of these claims the region is viewed as little more than the spatial analog of the strategic firm partnership. More robust geographical models examine the ways in which spatial proximity increases the flow of tacit knowledge directly through face-to-face contact (Malmberg and Maskell 2002; Asheim and Gertler, 2005), and indirectly through enhancing other forms of proximity within localized clusters of economic actors (Gertler, 2003).

Arguments about spatial proximity have long played a role in the diffusion of knowledge within geography (Hägerstrand 1953; Brown 1981) and beyond (Griliches 1957). Recent empirical evidence of the localization of knowledge flows by Jaffe et al. (1993), Maruseth and Verspagen (2002) and Sonn and Storper (2008) reinforce those earlier claims. At the same time, growing recognition of different forms of proximity and relatedness (Noteboom 2000; Boschma 2005; Boschma and Frenken 2010) has raised questions about the role of distance in regulating both the creation and flow of knowledge. Attention is increasingly directed at the role of social proximity and cognitive proximity in the diffusion of knowledge (Huber 2012).

Social proximity refers to the strength of inter-personal relationships that exist between individual actors (Boschma, 2005). These relationships may take a variety of forms, though they tend to cohere around the concept of trust borne by repeated interaction in common work-places, industrial organizations or related institutions. Autant-Bernard et al. (2007) also note that social proximity can be developed among actors well beyond the local scale often through work-related collaboration, regular meetings, through conferences and trade fairs. Once trust-based social relationships are in place, it is much more likely that actors will engage in interactive learning processes and knowledge sharing, guided by an open attitude towards communicative rationality rather than purely market-driven considerations (Lundvall, 1992). Social proximity is much more likely to develop when actors are connected through short social chains. Formal collaboration among individuals, as in the case of co-inventorship, or common employment with the same company, adds to the development of such short chains, that in turn enhances the strength of social proximity (Breschi and Lissoni, 2009).

Cognitive proximity focuses upon the extent to which different actors, or in aggregate industries and regions, share common knowledge structures. High cognitive proximity implies greater correspondence between knowledge sets, skills, routines and institutions of knowledge creation and sharing and, thus, a higher potential for absorptive capacity (Cohen and Levinthal 1990; Noteboom 2000). Higher levels of cognitive proximity also lead to enhanced collaboration as well as knowledge sharing. In similar fashion, recombinant models of technological progress rest on the cognitive proximity of technological subsets and of the economic agents that shape their integration (Weitzman, 1998; Fleming and Sorenson, 2001). Kogler et al. (2013) and Rigby (2013) extend these arguments in an explicitly spatial framework.

These different forms of proximity are finding purchase in a variety of empirical applications. Thus, Breschi and Lissoni (2001 and 2004) express a good deal of skepticism regarding the measurement of localized knowledge spillovers, suggesting that empirical estimates are unreliable, at least in part, because they do not separate social from spatial proximity. Maggioni et al. (2007) develop a variety of econometric models exploring knowledge production and co-patenting within and across European regions. They show that geographical proximity is always more important than social networks measured by participation within the EU Fifth Framework Programme and EPO co-patent applications. Using similar data Autant-Bernard et al. (2007) find strong evidence of spatial and social proximity in R&D cooperation across Europe. Fischer et al. (2006) examine patent citations across European regions in an extended gravity model, revealing that spatial and cognitive proximity regulate knowledge flows. In the United States, Agrawal et al. (2008) use the knowledge production function to explore how spatial proximity and social proximity influence access to knowledge. Using patent citations structured by MSA and the co-ethnicity of inventors, they show that the two forms of proximity are statistically significant and that they act as substitutes. Strumsky and Lobo (2008) report that the agglomeration of inventors is more important than inventor networks in regulating the pace of invention in metropolitan areas. Rigby and van der Wouden (2013) find that cognitive proximity trumps both spatial and sectoral proximity in this regard. Huber (2012) provides an excellent summary of much of this work and reports a more nuanced set of results regarding the importance of the different measures of proximity operating within the Cambridge technology cluster.

Specific work that relates to knowledge diffusion in the biotechnology industry outlines that the "growth and diffusion of intellectual human capital was the main determinant of where and when the American biotechnology industry developed" (Zucker et al., 1998:302). Johnson and Lybecker (2012) examine knowledge flows within the biotechnology sector by means of patent citation analysis. The results confirm prior findings elsewhere, i.e. that inter-firm knowledge transfers decrease with distance (Jaffe et al., 1993), but also provide evidence that the impact of physical distance has been diminishing in this sector over time. Spatial proximity is especially important in the early stages of an industry life cycle (Klepper, 1996), and for knowledge and R&D active industries in general (Audretsch and Feldman, 1996; Breschi, 1999). Mariani (2004) confirms that regional competencies along with localized spillovers play a more dominant role in the development of significant innovations in the early stages of the biotechnology industry life cycle compared to the more established research-intensive industry such as traditional chemicals.

The biotechnology industry "is characterized by a rapid knowledge diffusion and intense technological competition" (Gittelman and Kogut, 2003: 369). The link between scientific knowledge and innovation outputs are especially strong considering that this industry engages in the commercialization of scientific discoveries in the realm of basic science. Thus, there are particular strong linkages between technological innovation and local scientific knowledge in the biotechnology industry (Cohen et al., 2002). Studies that review the emergence of the US biotechnology industry emphasize the important role of knowledge spillovers from universities as drivers of firm start-ups (Zucker and Darby, 1996; Zucker et al., 1998; Feldman, 2001; Prevezer, 2001). Zucker and Darby (1996) found that the agglomeration of star scientists (defined as highly productive individuals who have discovered a major scientific breakthrough) in the biotechnology field directly results in a high concentration of new biotech ventures at the same location. Both, Almeida and Kogut (1999) and Zucker et al. (1998) provide arguments that stress the importance of the labor market (mobility of scientists and engineers), and there especially the mobility of those individuals between organizations, as a driver of knowledge diffusion. However, both studies also point to the localization of such labor mobility, i.e. people move from one organization to another, but not necessarily between places.

The aim of this paper is to explore the roles of spatial, social and cognitive proximity in guiding the spatial diffusion of rDNA technology across metropolitan areas of the United States.

The Cohen-Boyer rDNA Patents and the Creation of a New Sub-Class

The Cohen-Boyer discovery builds upon prior advances in biochemistry and genetics, but it is notable in that it was patented because it had immediately apparent commercial applications (Feldman et al., 2008). The patent was controversial when it was filed in 1974 and then was subject to three continuations and a six-year delay. Two factors delayed the granting of a patent (Feldman and Yoon, 2012). First, academic patents were rare at the time and ownership for discoveries under federally funded research was not automatically assigned to universities. Second, rDNA was highly controversial (Smith Hughes 2001). The scientific community agreed to a voluntary moratorium on rDNA research until its safety could be investigated. The original patent application claimed both the process of making recombinant DNA and any products that resulted from using that method. When the USPTO initially denied the product claims Stanford divided the claims into two divisional product applications, one that claimed recombinant DNA products produced in prokaryotic cells, and the other, which claimed recombinant DNA products produced in eukaryotic cells.[1] The rDNA patents referred to process and product patents: the product claims cover compositions of matter (recombinant DNA plasmids) that were then used to make proteins and are a basic component of the production method.

[1] A prokaryotic cell is one without a contained nucleus. The Prokaryotic patent is US 4468464, issued on August 28, 1984. A eukaryotic cell has a contained nucleus. The Eukaryotic patent is US 4740470, issued on April 26, 1988.

Every patent is placed into one or more distinct technology classes that are designed to reflect the technological characteristics of the underlying knowledge base that they embody. On Dec. 5, 1989, the USPTO issued Classification Order Number 1316, which created a new patent class 435/69.1 - Chemistry: molecular biology and microbiology (Recombinant DNA technique included the method of making a protein or polypeptide). When the set of technology codes is revised, as in this example, the USPTO reviews all granted patents and reclassifies those meeting the criteria of the new codes. This provides the researcher with a consistent set of all of the patents that use a specific technology. Strumsky et al. (2012) provide a review regarding the use of patent technology codes to study technological change, and point to their usefulness in tasks that relate to the identification of technological capabilities, the definition of technology spaces, or as an indicator of the arrival of technological novelty.

The data used in our analysis are patent records made available through the United States Patent and Trademark Office (USPTO). Patents have become an analytic staple for scholars interested in the geography and history of knowledge production (Lamoreaux and Sokoloff, 1996; O'hUallachain, 1999; Jaffe and Trajtenberg, 2002; O'hUallachain and Lee, 2011), on the various types of technical knowledge produced as indicated by patent classes (Hall *et al*., 2001; Strumsky *et al*., 2012) and on the factors that regulate knowledge flow (Jaffe *et al*., 1993; Breschi and Lissoni 2001; Sonn and Storper 2008). The popularity of patent data is related to their ready availability and to the wealth of information that they provide. At the same time, the disadvantages of patent as overall measures of economic and inventive activity are well known (Pavitt, 1985; Griliches, 1990). It is clear that patents do not represent all forms of knowledge production within the economy and that patents do not capture all produced knowledge. Patents, however, do provide insights into the organizations actively engaged in inventive activity in technologies, like rDNA, where the protection of intellectual property is important.

We focus on patents that make knowledge claims in USPTO class 435/69.1, regardless of whether 435/69.1 is the primary class or not[2] In total, there are 8,947 patents used in our analysis. All patents in our sample contain at least one inventor residing in a U.S. Core-Based Statistical Area (CBSA) that is classified as a Metropolitan Statistical Area (MSA)[3]. Patents are allocated to the metropolitan area that corresponds to the address of the first U.S. inventor listed.

[2] When we refer to rDNA or Cohen-Boyer patents, we are explicitly referring to patents that make claims to producing knowledge in class 435/69.1.
[3] Refer to OMB (2009) for a detailed list and definitions of CBSAs and MSAs.

It is possible to locate patents across multiple MSAs on the basis of the locations of co-inventors or to allocate fractions of a patent across MSAs when multiple co-inventors exist. In our experience, this has little impact on the subsequent results. The USPTO patent data does not identify inventors that can be linked across patents. Fortunately, Ronald Lai and colleagues at Harvard University have produced a list of individual inventors and their co-inventors that can be linked to individual patent records in the USPTO (Lai et al. 2011). Any inventors located outside the United States, or not located within one of the 366 U.S. core metropolitan statistical areas (CMSAs) are dropped from our data.

The start of our study is 1976, the year of the first USPTO patent application in USPTO class 435/69.1. Three patents pre-date the application of the rDNA patent (#4237224) in 1978 because their knowledge claims were adjudicated to belong to class 435/69.1 in a process of reclassification. We focus on the year of application rather than patent grant year to capture the time of invention. Because many patents are not granted for several years after application, we end the analysis in 2005 to dampen the impact of right censoring in the data.[4]

[Insert Figure 1 Here]

The Spatial Diffusion of rDNA

Diffusion of rDNA technology over both time and space can be traced by expansion in the number of patents placed in this technology class. Figure 1 shows the growth of r-DNA knowledge claims over time, recording the annual count of rDNA patents and the number of CMSAs , or cities, where inventors using rDNA resided.

The number of patents associated with class 435/69.1 increased rapidly through the late 1980s and early 1990s, following the classic S-shaped diffusion curve. The counts of rDNA patents remain level throughout the late 1990s at around 800 applications per year, although some significant fluctuations are visible. The number of patent applications that utilize class 435/69.1 has subsequently leveled off, and fell below 300 in 2005, the final year in our timeframe.

[4] The average time-lag between the application and actual grant of patents that contain USPC 435/69.1 is about 2.5 years at the beginning of the investigated time period, i.e. in the late 1970s and early 1980s, and then increases to an average of just over 3.5 year towards the end of the analyzed timeframe. The database utilized in this study provides data for USPTO patents granted up to the end of 2010 (Lai et al., 2012), and therefore right censoring the data in 2005 is considered a conservative approach.

Few cities were engaged in the production of rDNA inventions before the mid-1980s. By 1987, 11 years after the initial rDNA patent application, only 20 MSAs were producing patents in this technology field. Over the following 10 years geographic diffusion accelerated with approximately 90 MSAs developing rDNA technologies in the early 1990s. After stabilizing at this number for about five years, the number of MSAs participating in rDNA invention activities started to decline in 2001. In the final two observed years, 2004 and 2005, the number of MSAs where inventors reside remained around 65. Note, however, that the right censoring in the data series likely means this number is actually somewhat higher.

[Insert Table 1 Here]

The geographical spread of rDNA technology is further detailed in Table 1, along with the total number of rDNA patent applications for the study period 1976 – 2005; the year of the first rDNA patent in each place and the year when inventors in the city applied for a total of 10 patents. The cities listed are well-known centers of invention associated with academic research and subsequently the biotechnology industry. The first ranked city on this list, based on the total number of rDNA related patent applications from 1976-2005, is the metropolitan area around San Francisco. Home to Stan Cohen and Herbert Boyer, the two initial inventors of the rDNA technology, this is certainly no surprise. Within San Francisco there is evidence of rapid diffusion as a number of different inventors there produced ten patents within three years of the development of rDNA applications. From this initial lead, the city developed a well-known center of biotechnology research and commercialization activities.

The Boston-Cambridge-Quincy CBSA, which again is considered one of the key biotechnology centers in the nation, is a close second, with the same first rDNA patent application but a longer lag of six years to achieve ten patent applications. Total application counts start slightly later in the second tier metropolitan areas of Philadelphia, Washington, New York and San Diego.

All of the cities on this list record their first patent application before 1989. From this descriptive data there are no clear trends between the year of the first application and the total number of applications. The correlation between total patent applications and the year of first rDNA application is -0.53 while the correlation between the total number of patents and year when 10 patents were applied for is -0.76. There is clearly not a deterministic relationship in the pattern of diffusion. The correlation between the year of the first rDNA patent application and the year when 10 patent applications were filed is 0.75, and thus a plausible association.

[Insert Figure 2 Here]

As knowledge of rDNA technology has expanded both over space and time, this technology has found broader application as an input to invention across related patent classes. Figure 2 illustrates the patent classes that have been most frequently combined with the rDNA technology over three year periods running from 1976 to 2005. Accordingly, technology in class 435/69.1 is most closely associated with its parent technology class 435: Chemistry: Molecular Biology and Microbiology. While USPC 930 (Peptide or protein Sequence) is frequently combined with the rDNA technology in the initial time periods, its significance rapidly declines over time. Contrary to this, USPC 424 (Drug, Bio-Affecting and Body Treating Compositions) appears to become an increasingly more important ingredient, as measured by its co-classification share, in the development of rDNA technology related inventions over time. Overall, over time the combinations of other technologies used with USPC 435/69.1 expand. In addition to USPCs 536 (Organic Compounds), 530 (Chemistry: Natural Resins or Derivatives; Peptides or Proteins), 424 (Drug, Bio-Affecting and Body Treating Compositions), 800 (Multicellular Living Organisms and Unmodified Parts Thereof and Related Processes) and 514 (Drug, Bio-Affecting and Body Treating Compositions), which represent the largest shares of combined patent classes, more recently rDNA technology is also combined with such diverse technology fields as cleaning compositions (USPC 510), nanotechnology (USPC 977), or data processing (USPC 702). The remainder of this section describes the construction of measures of spatial, social and cognitive proximity.

Spatial Proximity of rDNA

Invention incorporating rDNA technology depends upon access to knowledge of rDNA. Codified rDNA information may be broadly available, but tacit knowledge of rDNA technology depends upon the ability of a set of potential adopters who have the relevant absorptive capacity. We operationalize the role of geographical proximity in constraining the flow of rDNA knowledge between U.S. metropolitan areas in two ways. First, data on the latitude and longitude of each MSA determine the Euclidean distance between each pair of metropolitan areas. For each city, the average distance to all other 365 metro areas is calculated. A simple hypothesis is that, ceteris paribus, metropolitan areas on average closer to all other MSAs are more likely to develop rDNA related capacity in the form of inventions in class 435/69.1. This physical measure of network geography is fixed over time. A second measure of geographical proximity combines distances between metropolitan areas with a binary (0/1) indicator of whether each city has generated an rDNA patent in class 435/69.1. For each city in a given year, the resulting products indicate distances only to other cities that have generated an rDNA patent. The minimum of such distances is recorded by MSA. If a given metro area has already patented in class 435/69.1, the minimum distance to a city with knowledge of rDNA is set at zero. This measure of geographical proximity may change value for an individual city over time. We hypothesize that the smaller this minimum distance, the greater the likelihood of a city generating an rDNA patent in a subsequent year. For empirical analysis this independent variable is lagged one year.

Social Proximity of rDNA

To measure social proximity for each MSA and year we used a database of all patents generated by inventors of rDNA class patents and all the co-inventors on those patents. This measure is operationalized in the following way. First, we construct a city social proximity matrix with dimension 366x366. All cells in this matrix are coded zero. Second, we identify all inventors of rDNA patents with an application year t. Each metro area with an inventor of a patent in class 435/69.1 has the value one added to the cell on the principal diagonal of the city-social proximity matrix corresponding to that city. If a city has two inventors of an rDNA patent in year t it would be given the value two on its principal diagonal. Third, we identify all non-rDNA patents generated by Cohen-Boyer inventors over the prior 5-year period (t-5 to t-1). Fourth, we list all co-inventors on those non-rDNA patents and we note the cities where those co-inventors are located. If the inventor of a Cohen-Boyer patent is located in city 1, and that inventor develops a non-Cohen-Boyer patent with a co-inventor located in city 50, then cell (1,50) of the city social proximity matrix is given the value 0.5. We thereby discount the social proximity of non-rDNA co-inventor relationships across cities. These social proximity values for all cities are generated across all years from 1980-2005. Finally, a summary measure of social proximity for each MSA is generated from the city social proximity matrix for each year. Values are lagged one year in the model.

[Insert Table 2 Here]

Over the first few years after the introduction of the patent for rDNA, the city with the highest social proximity to this technology was Bridgeport, Connecticut, home of Yale University (1980), then Chicago, Illinois (1981) and then San Francisco (1982). Not surprisingly perhaps, San Francisco, the location of the original Cohen-Boyer patent, has the highest measure of social proximity by 1985 and maintains the top-ranked position over the next 20 years. After 1985, larger MSAs, most known for their biotechnology industry clusters, fill out the remaining top ranks of the social proximity measure.

Descriptive statistics on the rDNA metropolitan social proximity measure are reported in Table 2 for three time periods ten years apart. The relatively high maximum value of social proximity in 1985 indicates the concentration of this technology in patent class 435/69.1, soon after its development. Thereafter, the increase over time in the mean social proximity measure for U.S. metropolitan areas, together with a decline in the maximum value, is evidence of the spatial diffusion of knowledge regarding the Cohen-Boyer technology. The data in Table 2 suggest that between 1995 and 2005 there is little change in the geographical spread of rDNA knowledge.

Cognitive Proximity of rDNA

By cognitive or technological knowledge we refer to subsets of knowledge that are associated with particular classes of inventions, technologies, or even industries. The proximity of a region to such knowledge subsets refers to local facility or expertise with specific technologies or to how close, in a technological sense, the economic agents of a region are to having such expertise. The knowledge subset with which we are most interested is that circumscribed by patent class 435/69.1. While rDNA patents are a subset of the broader class 435, we separate the sub-class in what follows and map it in technology space as a distinct set of knowledge along with 438 other unique primary patent classes.

In order to construct a U.S. knowledge space we need information on the number of patents in each technology class along with measures of proximity, the technological distance, between each pair of classes. Co-class information on individual patents is employed to measure the technological proximity of technology classes, following the earlier work of Jaffe (1986), Engelsman and van Raan (1994), Verspagen (1997), Breschi et al. (2003) and Nesta and Saviotti (2005). The number of primary patent classes on which we focus is considerably larger than that employed in most prior studies and thus the technology space outlined below is of higher resolution than those reported to date.

To measure the cognitive proximity, or knowledge relatedness, between patent technology classes in a single year we employ the following method. Let *P* indicate the total number of patent applications in the chosen year. Then, let $F_{ip} = 1$ if patent record p lists the classification code *i*, otherwise $F_{ip} = 0$. Note that *i* represents one of the 438 primary technology classes into which the new knowledge contained in patents is classified. In a given year, the total number of patents that list technology class *i* is given by $N_i = \sum_p F_{ip}$. In similar fashion, the number of individual patents that list the pair of co-classes *i* and *j* is identified by the count $N_{ij} = \sum_p F_{ip}F_{jp}$. Repeating this co-class count for all pairs of 438 patent classes yields the (438x438) symmetric technology class co-occurrence matrix ***C*** the elements of which are the co-class counts N_{ij}. The co-class counts measure the technological proximity of all patent class pairs, but is also influenced by the number of patents found within each individual patent class N_i. Thus, we standardize the elements of the co-occurrence matrix by the square root of the product of the number of patents in the row and column classes of each element, or

$$S_{ij} = {}^{N_{ij}}\Big/\sqrt{N_i^2 * N_j^2}$$

where S_{ij} is an element of the standardized co-occurrence matrix (***S***) that indicates the technological proximity, or knowledge relatedness, between all pairs of patent classes in a given year. The elements on the principal diagonal of ***S*** are set to 1. We prefer this simple form of standardization to calculation of the cosine index between all pairs of classes for the reasons outlined by Joo and Kim (2010).

With the aid of UCINET (Borgatti *et al*. 2002), the network of technological relatedness across the 438 primary patent classes and class 435/69.1 is mapped. The technological relatedness network is generated with the Gower-scaling metric, itself derived to examine patterns of similarity across network nodes (Gower 1971). The nodes in the network correspond to each of the 438 distinct technological classes within the USPTO, and class 435/69.1. The relative positions of the nodes are fixed by the standardized co-occurrence class counts (S_{ij}). Note that the standardized co-occurrence matrix ($\boldsymbol{S}$) is symmetric. The principal diagonal plays no role in the relative locations of the nodes.

[Insert Figure 3 Here]

The knowledge relatedness networks for 1975-2005 are shown in Figure 3. The node colors in Figure 3 represent the aggregate technology (6 class) grouping of Hall *et al*. (2001): Black = Chemicals (1), Green = Computers & Communications (2), Yellow = Drugs & Medical (3), Red = Electronics (4), Blue = Mechanical (5), Grey = Miscellaneous (6). There is clear evidence of the clustering of individual patent categories within most of these classes, indicating that our technological proximity or relatedness measure is capturing what may be considered as a common knowledge base within these more aggregate technology groupings. The size of each node illustrates the number of patents in that technology class in the given year. Node sizes have been scaled to allow comparison over time. All network links are not included in Figure 3 for their density would render the network largely unreadable. The network links shown are illustrative of the total, representing the strongest links in the network at each time period.

rDNA patents in class 435/69.1 are illustrated with the small, yellow triangle in each of the slides of Figure 3. In early years, rDNA patents are closely linked, in technology space, with chemicals and with drugs and medical patent classes. By 1995 the close association between class 435/69.1 and its parent class 435, illustrated in Figure 3, is already apparent.

In order to measure the cognitive proximity of the knowledge base of a metropolitan area to the rDNA patent class we find the average relatedness of a city's patents to class 435/69.1. In technology space, nodes that are close together have a high relatedness score. These are the technology classes that tend to co-occur with relatively high frequency on individual patents. In terms of rDNA, the average relatedness value for metropolitan area *m* in year *t* is calculated as:

$$AR^{mt} = \frac{\sum_j S_{CBj}^t * D_j^{mt}}{N^{mt}}$$

where S_{CBj}^t represents the technological relatedness between rDNA (class 435/69.1) patents and patents in all 439 technology classes *j*, where the vector *j* includes class 435/69.1. This term is the (row or column) vector of the standardized co-occurrence matrix noted above for the rDNA technology class. D_j^{mt} is the count of the number of patents in technology class *j* in metro area *m* in year *t*, and N^{mt} is a count of the total number of patents in city *m* in year *t*.

[Insert Table 3 Here]

Table 3 presents descriptive statistics for metropolitan area cognitive proximity to the rDNA patent class. The mean average relatedness value of patents in general to the rDNA class of patents within U.S. metropolitan areas was approximately three times higher in 2005 than in 1985, indicating the diffusion of rDNA-related technology. The metro areas with the highest cognitive proximity values to patent class 435/69.1 are, perhaps, not those we might have expected. Most metro areas listed in Table 3 have relatively small numbers of patents, but those patents are in patent class 435/69.1 or close to it in the technology space of Figure 3. Indeed, Madison, WI, Kennewick, WA, Durham-Chapel Hill, NC, Blacksburg VA, Flagler, FL, Athens, GA and Iowa City, IA are all university towns and sites of rDNA inventions over the period investigated.

Furthermore, all the metro area listed in Table 1 as key centers of rDNA invention have average cognitive proximity that are greater than average for U.S. cities. We hypothesize that metropolitan areas with higher levels of cognitive proximity are more likely to patent in rDNA. In the statistical analysis reported below we control for a number of other covariates that likely influence the spatial diffusion of rDNA technology, in addition to the influence of geographical, social and cognitive proximity. The number of patents generated in each metropolitan area provides a proxy for city-size/inventiveness. Insofar as patenting in a specialized field of biotechnology is likely associated with basic research in universities and hospitals, typically though not always found in larger urban areas, we hypothesize that patent counts in general will be positively related to the probability of a city patenting in class 435/69.1. Note that the city-size variable is positively correlated with our social proximity variable, as might be expected. However, that correlation is not cause for undue concern as co linearity renders estimators inefficient rather than biased. We return to this issue later. Levels of bio-medical research funding in universities, in industry and in total were also constructed from NIH records for each city across the time period under study. Higher levels of biomedical research are expected to increase the probability of patenting in rDNA. All independent variables are lagged one year.

Model and Estimation Results

Our primary research question focuses on the probability of a metropolitan area generating an rDNA patent in class 435/69.1. We have time series panel data for 366 MSAs over 30 years. The limited (binary) nature of the dependent variable suggests use of a logit or probit regression model. There is a right-censoring issue in our data that may generate significant bias in estimated coefficients (Allison 1984). Armed with repeated observations on the same set of metropolitan areas over time enables exploration of a fixed effects panel model to deal with potential problems of unobserved heterogeneity. Another possibility that does not control for unobserved heterogeneity, but that more explicitly handles censored data, is the event history model. We use the Cox non-proportional (extended) hazard model, incorporating time-varying covariates, to examine the date of a first rDNA invention within a metropolitan area, while we turn to the panel form of the logit model to examine the probability of repeated invention in patent class 435/69.1 across all years in the study period.

In theory, endogeneity should not present significant challenges to estimation. Nonetheless, in order to dampen such concerns, all time-varying independent variables are lagged by one year. We do not have a clear theoretical rationale for employing a lag of only one period, we seek only to ensure that the characteristics of patenting in year t do not influence the value of independent variables employed to explain the probability of a city developing a rDNA patent in that same year. We turn attention first to the event history model and attempts to identify the date at which a metropolitan area first develops a patent in class 435/69.1. Our patent data by city are not left-censored for our data series start with the introduction of the first rDNA patent in 1980. However, there are right-censoring issues with our data, as a number of metropolitan areas, 165 out of 366, do not develop a Cohen-Boyer invention by 2005 when our study-period ends. The Cox semi-parametric survival model is the most widely used of the family of hazard models, largely because it does not assume a particular form of probability distribution for survival times. The cost of this flexibility is the assumption of the proportionality of hazards, an assumption that we violate because of the time-varying covariates that enter our model. Thus, we make use of the extended Cox model (Blossfeld et al. 2007).

[Insert Table 4 Here]

Table 4 presents the estimation results of the extended Cox hazard model for our patent data against our time-varying covariates. Model (1) presents results for all metropolitan areas. The geographical proximity of a metropolitan area to cities that have developed rDNA patents has no significant effect on the hazard ratio. Social proximity exhibits a significant and positive influence on the hazard ratio: a one-unit increase in social proximity raises the probability that a metropolitan area will generate a first rDNA patent over the baseline hazard by 3.3%. Cognitive proximity also has a positive and significant effect, raising the hazard ratio by a little more than 1% for every one-unit increase in this variable.

The patent count variable is employed as a proxy for city-size and also has a significant, positive influence on the hazard ratio: large cities are more likely to adopt. Note that metropolitan patent counts are highly correlated with social proximity (Pearson coefficient = 0.6) as might be expected. As the number of patents increase within a metropolitan area, the social proximity of the city also increases. Removing the patent count variable from the Cox model doubles the size of the hazard ratio for social proximity, while leaving all other covariates essentially unchanged.

The amount of university and industrial research and development conducted within a city also exert significant influence on the hazard ratio, though in different directions. University R&D acts to lower the hazard ratio, while industrial R&D increases the hazard ratio. In both cases, the influence of R&D on the hazard ratio is relatively small in size. Turning to the time-fixed variable, average distance to other cities, it is surprising that this variable has a significant, positive influence on the hazard ratio. The coefficient suggests that increasingly remote cities are more likely to develop rDNA patents. We have more to say about this result below.

Models (2) and (3) of Table 4 provide results from the extended Cox model for small cities with relatively few patents and for large cities with relatively high patent counts, respectively. These two groups fall just inside the quartiles of the distribution of metropolitan areas by patent count, corresponding to the 30th and 70th percentiles. (Trying to estimate the model for the lower quartile generated a very small number of failures (patents in class 435/69.1) and no model convergence). Cities in the bottom quartile of metro areas by patent count generated only 17 of the 201 total first-time patents examined in the Cox model. Cities in the top quartile were responsible for 87 Cohen-Boyer patents. The key differences between these two sub-samples are found in the values of the hazard ratios for social and cognitive proximity. On the one hand, for small cities, the hazard ratio for social proximity is very large, indicating that a one-unit increase in the connectedness of the city's inventors to inventors of rDNA patents raises the probability of patenting in class 435/69.1 by about 30% over the baseline rate. Cognitive proximity has no significant influence on the hazard ratio in small cities. On the other hand, in large cities, social proximity has no significant influence on the hazard ratio, while cognitive proximity exerts a significant, positive effect. We suspect that cities over a certain size threshold have a sufficient level of social proximity to generate rDNA technologies and that further increases in social proximity make little difference to the probability of such events.

The hazard ratios for social proximity and cognitive proximity are significantly different between metropolitan areas in the lower and upper quartiles by patent count. No other covariates are significantly different between these two groups. Also note that the time-fixed measure of distance to other cities is insignificant in both small and large city sub-samples. This might help explain the unexpected sign and significance of this variable in the full sample. Smaller cities are on average closer together than large cities and with most rDNA patents generated in larger cities, the association of size and distance gives the unexpected sign on the simple distance variable.

[Insert Table 5 Here]

Table 5 presents results from examining the independent variables in a longitudinal panel framework using a logit model incorporating fixed effects to treat unobserved heterogeneity that is constant over time. Our key finding is that the results are broadly consistent with those already reported for the event history model. One marked difference between the Cox hazard model and the logistic model is that we examine the probability of repeated patents over time in the latter, while we focused only on time to first rDNA patent in the former. The fixed-time measure of average distance between cities also drops out of the fixed effects logit model. In the logit model of Table 5 we include time fixed effects though we do not report them. To reduce incidental parameters issues, this conditional form of the logit model eliminates 167 cities from analysis because the value of the dependent variable in these cities is unchanged. In this instance, these cities never develop an rDNA patent.

The partial logistic regression coefficients reported in Table 5 are log odds ratios, reporting how a one-unit increase in the independent variable influences a change in the log odds of the dependent variable. The lagged value of geographic proximity, distance to the nearest city that has generated a Cohen-Boyer patent has no significant influence on the log odds of a patent in class 435/69.1. Social proximity and cognitive proximity have a significant effect on the log odds ratio and both exhibit the anticipated positive sign. For example, a one-unit increase in social proximity raises the log odds of an rDNA patent being invented in a metropolitan area by 0.0457. This is an increase in the odds ratio of a Cohen-Boyer patent of 1.046, after transforming the coefficient. City-size, as proxied by the sum of patent counts, has no significant influence on the log odds ratio. Removing the patent count variable yields no change on the social proximity measure in this model. Research and development in the university and in industry significantly influence the log odds ratio, though again in different directions. Industry R&D increases those odds. University R&D reduces the log odds of an rDNA patent, suggesting that further work might consider the technology transfer orientation and operations at different institutions. While the Bayh-Dole passed in 1980 it was not until the later 1990s that the majority of research universities had established tech-licensing offices. Attitudes towards technology transfer were even slower to change to encourage active patenting. Note that marginal effects are not reliably produced for the panel form of the fixed effects logit model.

Reflective Conclusions

In this paper we trace the spatial diffusion of a significant new technology, the knowledge base of rDNA, represented by the creation of USPTO patent class 435/69.1. rDNA was developed by Cohen and Boyer in the San Francisco Bay area in the mid-1970s and the patent for this technology was granted in 1980. Between 1980 and 2005 multiple patents in class 435/69.1 were developed in a relatively small number of metropolitan areas across the United States. The pace of rDNA diffusion followed the standard logistic form. Our primary interest was in the factors that regulated the spatial spread of this new knowledge class and, in particular, the relative roles of geographical proximity, social proximity and cognitive proximity. Understanding the relative importance of these proximities illustrates the mechanisms by which knowledge is transmitted and the use of new technology diffuses.

rDNA techniques spread rapidly from San Francisco to a number of relatively large metropolitan areas and to a few small cities around the U.S. These areas were relatively far from one another and thus little evidence was found to support the role of spatial proximity in facilitating the flow of knowledge regarding rDNA. Even after controlling for the location of academic and industrial research and development, geographic proximity, as measured by distance, was not statistically significant.

Social proximity, measured by the network of rDNA co-inventors within the US, played a positive and significant role in the spread of rDNA technology. Inventors associated with patents in the technology class 435/69.1 passed information on this new knowledge set to their co-inventors located in the same city or in different cities across the country. This suggests that co-inventing relationships provided a mechanism for the diffusion of the technology.

Absorption and application of this new technological information was not automatic, however, and also dependent on cognitive proximity -- the technological profile of knowledge in different cities and on the closeness of that profile to the knowledge base of rDNA. The non-local nature of rDNA knowledge flow indicates that social networks of co-inventors associated with this technology class do not have a strong local component and/or that few cities within close proximity of one another have the absorptive capacity to develop this knowledge subset. The specialization of rDNA technology appears to limit diffusion to those few areas with strong concentrations of biotechnology related activity.

Results from our event history model suggest that in smaller, less inventive U.S. cities, even where cognitive proximity to rDNA technology was not strong, social proximity played the critical role in the diffusion of knowledge in patent class 435/69.1. This indicates that attracting a few individuals who have strong social ties outside the city allows for a greater likelihood of being able to participate in the new technology. When the technology under consideration is a significant breakthrough or a General Purpose Technology the ability to engage with the technology is critical. Conversely, in larger, more inventive cities, where we might assume that social proximity is always relatively high, absorptive capacity played the lead role in diffusion.

References

Abernathy W. J. and Clark K. B. (1985) Innovation: mapping the winds of creative destruction, *Research Policy* **14**(1), 3–22.

Abernathy W. J. and Utterback J. M. (1978) Patterns of Industrial Innovation, *Technology Review* **80**(7), 40-47.

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

Allison P. (1984) *Event History Analysis: Regression for Longitudinal Event Data*. Sage, Newbury Park.

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

Asheim B. and Gertler M. S. (2005) The geography of innovation: regional innovation systems, in Fagerberg J. Mowery D. and Nelson R. (Eds) *The Oxford Handbook of Innovation*, pp. 291–317. Oxford University Press, Oxford.

Audretsch D. B., Bozeman B., Combs K. L., Feldman M. P., Link A. N., Siegel D. S., Stephan P., Tassey G. and Wessner C. (2002) The economics of science and technology, *Journal of Technology Transfer* **27**(2), 155-203.

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

Autant-Bernard C., Mairesse J. and Massard N. (2007) Spatial knowledge diffusion through collaborative networks, *Papers in Regional Science* **86**(3), 341-350.

Barney J. B. (1991) Firm resources and sustained competitive advantage, *Journal of Management* **17**(1), 99-120.

Baum J. and Haveman H. (1997) Love they neighbor: differentiation and agglomeration in the Manhattan hotel industry, 1898-1990, *Administrative Science Quarterly* **42**, 304-338.

Blossfeld H.-P., Golsch K. and Rohwer G. (2007) *Event History Analysis with STATA*. Lawrence Erlbaum Associates Publishers, Mahwah, NJ.

Borgatti S., Everett M. and Freeman L. (2002) *UCINET for Windows: Software for Social Network Analysis*. Analytic Technologies, Harvard, MA.

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

Boschma R. A. and Frenken K. (2010) The spatial evolution of innovation networks: a proximity perspective, in R. Boschma and R. Martin (Eds) *The Handbook of Evolutionary Economic Geography*, pp. 120-135. Edward Elgar, Cheltenham.

Breschi S. (1999) Spatial patterns of innovation: evidence from patent data, in A. Gambardella and F. Malerba (Eds) *The Organization of Innovative Activity in Europe*, pp. 71-102. Cambridge University Press, Cambridge, MA.

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

Breschi S. and Lissoni F. (2004) Knowledge networks from patent data: methodological issues and research targets. KITeS Working Paper #150, KITeS, Centre for Knowledge, Internationalization and Technology Studies, Universita' Bocconi, Milano, Italy.

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

Breschi S., Lissoni F. and Malerba F. (2003) Knowledge-relatedness in firm technological diversification, *Research Policy* **32**(1), 69–87.

Brown L. A. (1981) *Innovation Diffusion*. Methuen, London.

Christensen C. M. (1997) *The Innovator's Dilemma*. Harvard Business School Press, Boston, MA.

Clark K. B. (1985) The interaction of design hierarchies and market concepts in technological evolution, *Research Policy* **14**(5), 235–251.

Cohen W. M. and Levinthal D. A. (1990) Absorptive capacity: a new perspective on learning and innovation, *Administrative Science Quarterly* **35**, 128-152.

Cohen W., Nelson R. and Walsh J. (2002) Links and impacts: The influence of public research on industrial R&D, *Management Science* **28**(1), 1–23.

Cooke P., Uranga M. and Etxebarria G. (1997) Regional innovation systems: institutional and organizational dimensions, *Research Policy* **26**(4–5), 475–491.

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

Engelsman E. C. and van Raan A. F. J. (1994) A patent-based cartography of technology, *Research Policy* **23**(1), 1–26.

Feldman M. P. (1994) *The Geography of Innovation*. Kluwer Academic, Dordrecht.

Feldman M. P. (2001) Where science comes to life: university bioscience, commercial spin-offs, and regional economic development, *Journal of Comparative Policy Analysis: Research and Practice* **2**, 345–361.

Feldman M. P., Colaianni A. and Liu C. K. (2008) Chapter D, 17.22: lessons from the commercialization of the Cohen-Boyer patents: the Stanford University licensing program, in A. Krattiger, R. T. Mahoney, U. Nelsen, J. A. Thompson, A. B. Bennett, K. Satyanarayana, G. D. Graff, C. Fernandez and S. P. Kowalski (Eds) *Intellectual Property Management in Health and Agricultural Innovation: A Handbook of Best Practices*. Vol. 1, pp. 1797–1808. MIHR, Oxford; PIPRA, Davis, CA; Oswaldo Cruz Foundation, Brazil; and bioDevelopments-Internation Institute, Ithaca, NY.

Feldman M. P. and Kogler D. F. (2010) Stylized facts in the geography of innovation, in B. Hall and N. Rosenberg (Eds) *Handbook of the Economics of Innovation*, pp. 381–410. Elsevier, Oxford.

Feldman M. P. and Yoon J. W. (2012) An empirical test for general purpose technology: an examination of the Cohen-Boyer rDNA technology, *Industrial and Corporate Change* **21**(2), 249-275.

Feldman M.P. and Zoller T. D. (2012) Dealmakers in place: social capital connections in regional entrepreneurial economies, *Regional Studies* **46**(1), 23-37.

Fischer M., Scherngell T. and Jansenberger E. (2006) The geography of knowledge spillovers between high-technology firms in Europe: evidence from a spatial interaction modeling perspective, *Geographical Analysis* **38**, 288-309.

Fishlow A. (1965) *American Railroads and the transformation of the Ante-Bellum Economy.* Harvard University Press, Cambridge, MA.

Fleming, L. & Sorenson, O. (2001) Technology as a complex adaptive system: evidence from patent data, *Research Policy* **30**(7), 1019–1039.

Fogel R. (1964) *Railroads and American Economic Growth: Essays in Econometric History*. Johns Hopkins University Press, Baltimore.

Gertler M. S. (2003) Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there), *Journal of Economic Geography* **3**(1), pp. 75–99.

Gittelman M. and Kogut, B. (2003) Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citation patterns, *Management Science* **49**(4): 366-382.

Gower J. C. (1971) A general coefficient of similarity and some of its properties, *Biometrics* **27**(4), 857–871.

Griliches Z. (1957) Hybrid corn: an exploration in the economics of technological change, *Econometrica* **25**(4), 501-522.

Griliches Z. (1990) Patent statistics as economic indicators: A survey, *Journal of Economic Literature* **28**(4), 1661–1707.

Hägerstrand T. (1953) *Innovation Diffusion as a Spatial Process*. University of Lund, Lund.

Hall B. H., Jaffe A. B. and Trajtenberg M. (2001) The NBER Patent Citations Data File: Lessons, Insights and Methodology Tools, *National Bureau of Economic Research* (Cambridge, MA) Working Paper No. 8498.

Harhoff D., Narin F., Scherer M. and Vopel K. (1999) Citation frequency and the value of patented inventions, *Review of Economics and Statistics* **81**(3), 511-515.

Helpman E. (1998) Introduction, in Helpman E. (Ed.) *General Purpose Technologies and Economic Growth*, pp. 1-14. MIT Press, Cambridge, MA.

Huber F. (2012) On the role and iterrelationship of spatial, social and cognitive proximity: personal knowledge relationships of R&D workers in the Cambridge information technology cluster, *Regional Studies* **46**(9), 1169-1182.

Jaffe A. (1986) Technological opportunity and spillovers of R&D, *American Economic Review* **76**(5), 984–1001.

Jaffe A., Trajtenberg M. and Henderson R. (1993) Geographic localization of knowledge spillovers as evidenced by patent citations, *Quarterly Journal of Economics* **108**, 577-98.

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

Johnson D. K. N. and Lybecker K. M. (2012) Does distance matter less now? The changing role of geography in biotechnology innovation, *Review of Industrial Organization* **40**(1), 21-35.

Joo S. H. and Kim Y. (2010) Measuring relatedness between technological fields, *Scientometrics* **83**(2), 435–454.

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

Klepper S. (1996) Entry, exit, growth, and innovation over the product life cycle, *American Economic Review* **86**, 562–583.

Klepper S. (1997) Industry life cycles, *Industrial and Corporate Change* **6**(1), 145-181.

Kogler D. F., Rigby D. L. and Tucker I. (2013) Mapping knowledge space and technological relatedness in U.S. cities, *European Planning Studies,* forthcoming, pp. 19

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

Krugman P. (1991) Increasing returns and economic geography, *Journal of Political Economy* **99**, 483-499.

Lai R. D, Amour A., Yu A., Sun Y. and Fleming L. (2011) Disambiguation and co-authorship networks of the U.S. Patent Inventor Database (1975–2010), Working Paper, Harvard University.

Lamoreaux N. and Sokoloff K. (1996) Long-term change in the organization of inventive activity, *Proceedings of the National Academy of Sciences* **93**(23), 12686–12692.

Lichtenberg F. R. and Philipson T. J. (2002) The Dual Effects of Intellectual Property Regulations: Within- and Between-Patent Competition in the US Pharmaceuticals Industry, *National Bureau of Economic Research* (Cambridge, MA) Working Paper No. 9303.

Lipsey R. G., Carlaw K. and Bekar C. T. (2005) *Economic Transformations: General Purpose Technologies and Long Term Economic Growth*. Oxford University Press, Oxford.

Lundvall B.Å. (1992) *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*. Pinter, London.

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

Malmberg A. and 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.

Mariani M., 2004. What determines technological hits? Geography versus firm competencies. *Research Policy* **33**, 1565–1582.

Maskell P. and Malmberg A. (1999) The competitiveness of firms and regions: “ubiquitification” and the importance of localized learning, *European Urban and Regional Planning Studies* **6**(1), pp. 9–25.

Maurseth P.-B. and Verspagen B. (2002) Knowledge spillovers in Europe: A patent citations analysis, *Scandinavian Journal of Economics* **104**, 531-545.

Morgan K. (1997) The learning region: Institutions, innovation and regional renewal, *Regional Studies* **31**(5), 491–503.

Nelson R. R. and Winter S. G. (1982) *An Evolutionary Theory of Economic Change*. Belknap Press, Cambridge.

Nesta L. and Saviotti P. P. (2005) Coherence of the knowledge base and the firm's innovative performance: Evidence from the U.S. pharmaceutical industry, *Journal of Industrial Economics* **53**(1), 123–142.

Nooteboom B. (2000) *Learning and Innovation in Organizations and Economies*. Oxford University Press, Oxford.

O'hUallachain B. (1999) Patent places: Size matters, Journal of Regional Science 39(4), 613–636.

O'hUallachain B. and Lee D. (2011) Technological variety and specialization in urban invention, *Regional Studies* **45**(1), 67–88.

OMB (2009) *OMB Bulletin No. 10-02: Update of Statistical Area Definitions and Guidance on Their Uses*. United States Office of Management and Budget, Washington, D.C.

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

Phene A., Fladmoe-Lindquist K. and Marsh L. (2006) Breakthrough innovations in the U.S. biotechnology industry: the effects of technological space and geographic origin, *Strategic Management Journal* **27**, 369-388.

Prevezer, M. (2001) Ingredients in the early development of the U.S. biotechnology industry, *Small Business Economics* **17**(1), 17–29.

Rigby D. L. (2012) The geography of knowledge relatedness and technological diversification in U.S. cities. Forthcoming, *Regional Studies*.

Rigby D. L. and van der Wouden F. (2013) Abandoning the Herfindahl: Knowledge relatedness and the rate of invention in U.S. cities. Department of Geography, UCLA. Paper available on request.

Sahal D. (1981) *Patterns of Technological Innovation*. Addison Wesley, Reading, MA.

Saxenian A. (1994) *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Harvard University Press, Cambridge, MA.

Schumpeter J. A. (1911) *A Theory of Economic Development*. Harvard University Press, Cambridge, MA.

Schumpeter J.A. (1942) *Capitalism, Socialism and Democracy*. Harper, New York.

Smith Hughes S. (2001) Making dollars out of DNA: the first major patent in biotechnology and the commercialization of molecular biology, 1974–1980, *Isis* **92**(3), 541–575.

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

Sorenson O. and Audia P. G. (2000) The social structure of entrepreneurial activity: footwear production in the United States, 1940–1989, *American Journal of Sociology* **106**, 424–462.

Storper M. and Christopherson S. (1987) Flexible specialization and regional agglomerations: the case of the U.S. motion picture industry, *Annals of the Association of American Geographers* **77**, 104–117.

Storper M. (1997) *The Regional World: Territorial Development in a Global Economy*. Guilford, New York.

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

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

Trajtenberg M. (1990) *Economic Analysis of Product Innovation: The Case of CT Scanners*. Harvard University Press, Cambridge, MA.

Verspagen B. (1997) Measuring intersectoral technology spillovers: Estimates from the European and US Patent Office Databases, *Economic Systems Research* **9**(1), 47–65.

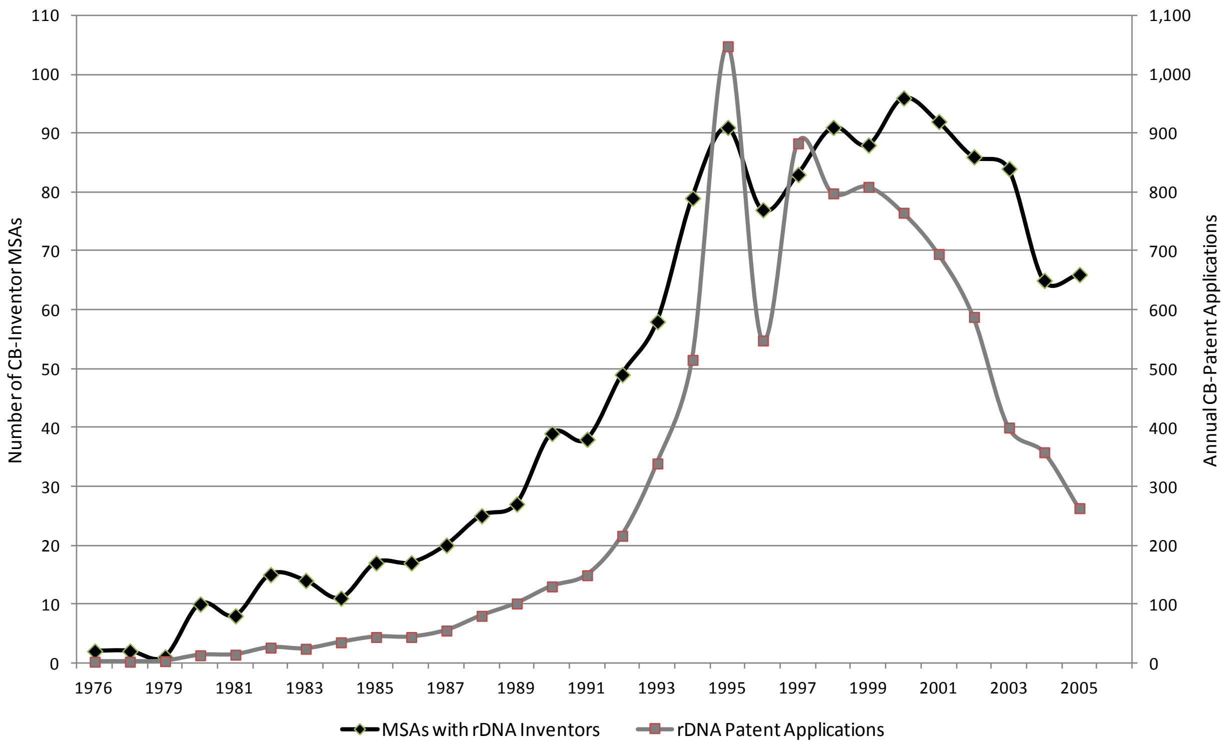
Weitzman M. (1998) Recombinant growth, *Quarterly Journal of Economics* **113**(2), 331-360.

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

Zucker L. G. and Darby M. R. (1996) Star scientists and institutional transformation: patterns of invention and innovation in the formation of the biotechnology industry, *Proceedings of the National Academy of Sciences of the United States of America* **93**(23), 12709-12716.

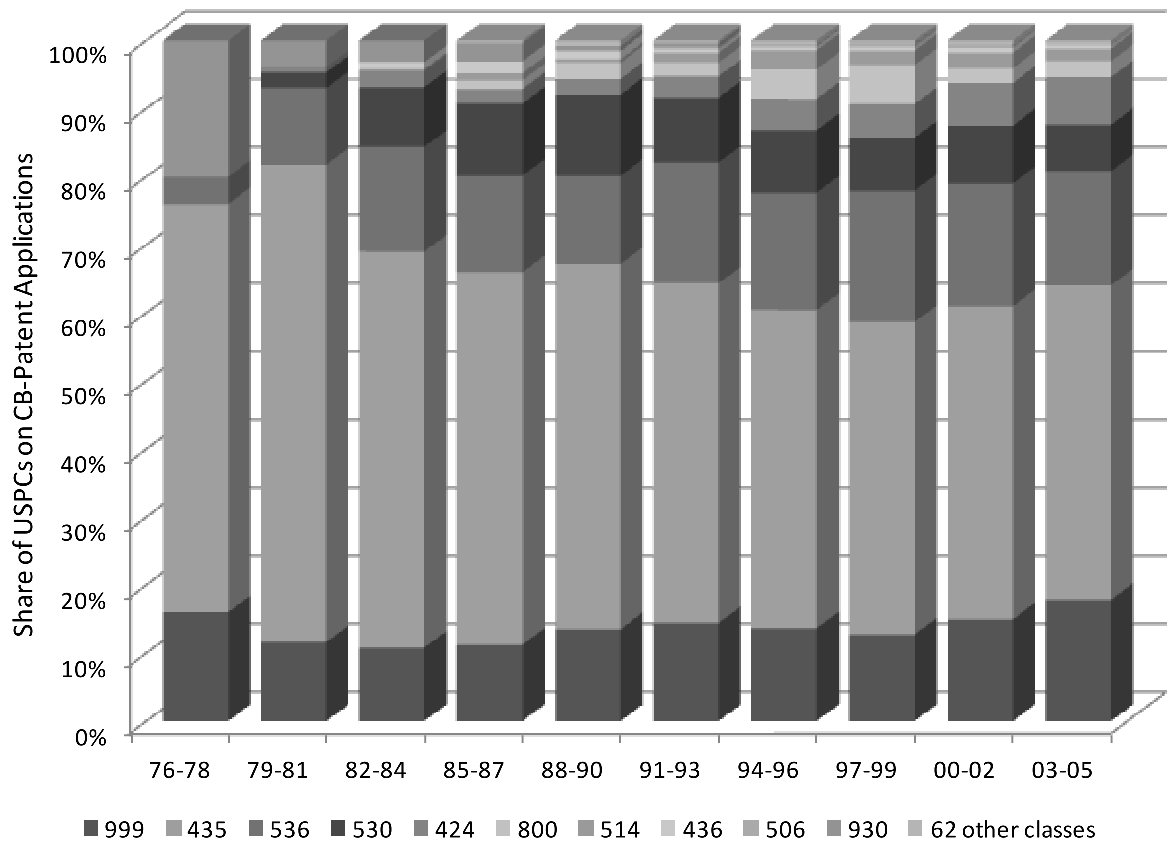
Zucker L. G., Darby M. R. and Brewer M. B. (1998) Intellectual human capital and the birth of U. S. biotechnology enterprises, *American Economic Review* **88**(1), 290-306.

Figure 1: Annual number of rDNA patent applications and corresponding count of MSAs where their respective inventors reside, 1976-2005.



Notes: The analysis is based on rDNA patents developed by inventors residing in one of the 366 Metropolitan Statistical Areas (MSAs) of the U.S. The 576 Micropolitan Statistical Areas (μSAs) that make up the remainder of the 942 Core Based Statistical Areas (CBSAs) as definite by the U.S. Office of Management and Budget (OMB, 2009) are ignored due to their marginal contributions to rDNA patenting.

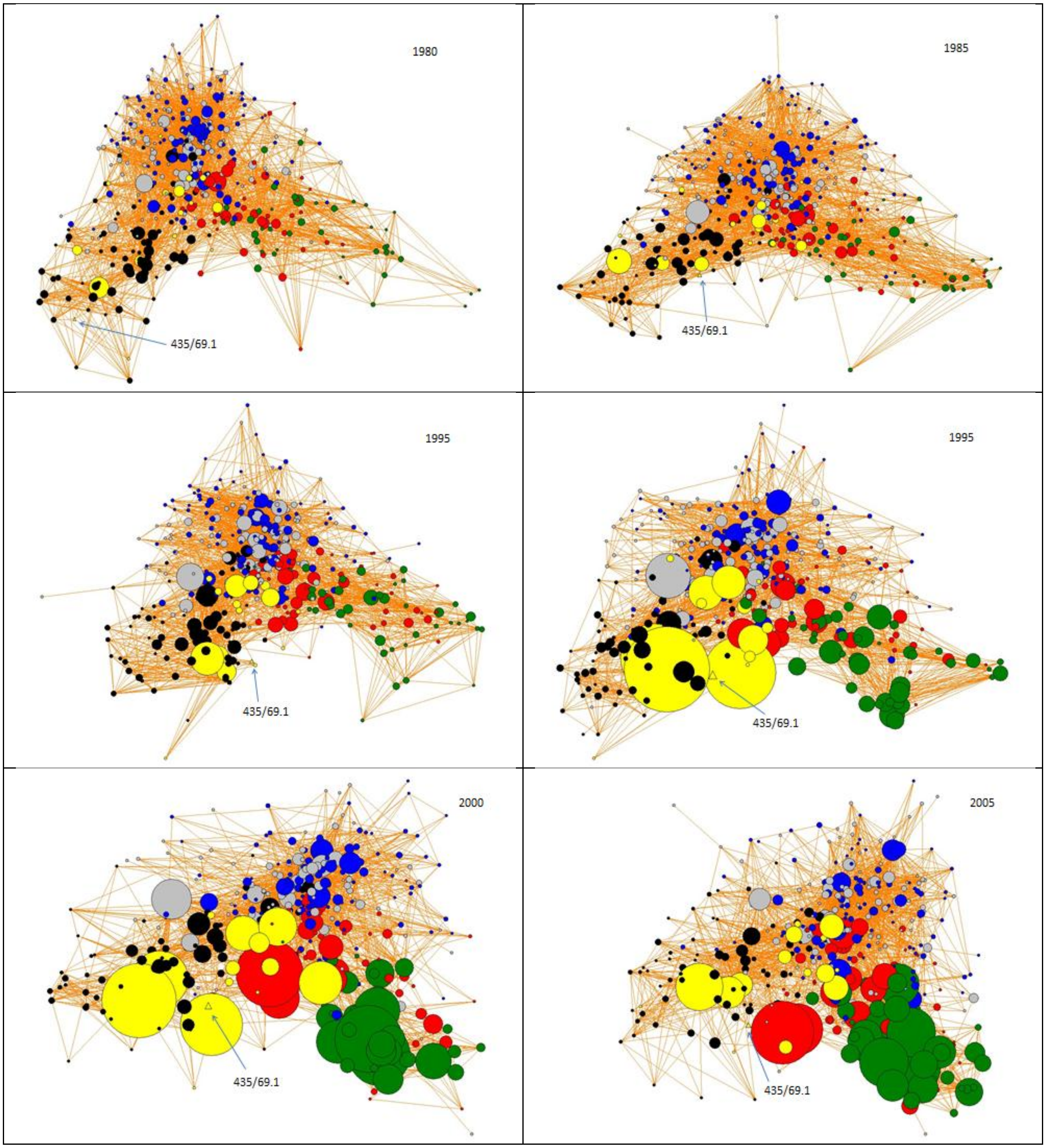
Figure 2: Distribution of US patent classes (USPC) listed on rDNA related patent documents; three-year shares based on application year, 1976-2005.



Notes: 999 = rDNA USPC 435/69.1, 435 = Chemistry: Molecular Biology and Microbiology, 536 = Organic Compounds, 530 = Chemistry: Natural Resins or Derivatives; Peptides or Proteins, 424 = Drug, Bio-Affecting and Body Treating Compositions, 800 = Multicellular Living Organisms and Unmodified Parts Thereof and Related Processes, 514 = Drug, Bio-Affecting and Body Treating Compositions, 436 = Chemistry: Analytical and Immunological Testing, 506 = Combinatorial Chemistry Technology: Method, Library, Apparatus, 930 = Peptide or Protein Sequence;[5]

[5] The "62 other classes" refers to USPCs that are either rarely combined with class 435/69.1, or to classes that only have been combined with the rDNA technology in more recent time periods, incl. USPCs 510 (Cleaning Compositions), 977 (Nanotechnology), 426 (Food and Edible Material), and 702 (Data Processing); USPC 514 is an integral part of class 424.

Figure 3: The U.S. technology space incorporating rDNA (USPC 435/69.1)



Notes: Patent class 435/69.1 is the yellow triangle in the lower left of the technology space. The nodes represent all 438 primary classes of utility patents and node sizes reflect the number of patents in each class, scaled for comparability over the years 1980, 1985, 1990, 1995, 2000 and 2005. The colors of the nodes represent the six aggregate technology classes of Hall *et al*. (2001): Black = Chemicals (1), Green = Computers & Communications (2), Yellow = Drugs & Medical (3), Red = Electronics (4), Blue = Mechanical (5), Grey = Miscellaneous (6).

Table 1: Key Places (MSAs) of rDNA invention

	Metropolitan Statistical Area (MSA)	rDNA Patent Applications 1976-2005	Year of First rDNA Patent Application	Year When MSA Reached 10 Applications
1	San Francisco-Oakland-Fremont, CA	1,133	1978	1981
2	Boston-Cambridge-Quincy, MA-NH	990	1978	1984
3	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	691	1981	1988
4	Washington-Arlington-Alexandria, DC-VA-MD-WV	639	1980	1986
5	New York-Northern New Jersey-Long Island, NY-NJ-PA	617	1980	1985
6	San Diego-Carlsbad-San Marcos, CA	585	1982	1985
7	San Jose-Sunnyvale-Santa Clara, CA	483	1985	1990
8	Seattle-Tacoma-Bellevue, WA	400	1981	1988
9	Los Angeles-Long Beach-Santa Ana, CA	260	1982	1989
10	St. Louis, MO-IL	150	1976	1989
11	Chicago-Joliet-Naperville, IL-IN-WI	147	1980	1990
12	Sacramento--Arden-Arcade--Roseville, CA	127	1987	1992
13	Baltimore-Towson, MD	126	1988	1993
14	Houston-Sugar Land-Baytown, TX	123	1983	1992
15	Madison, WI	122	1982	1987
16	Indianapolis-Carmel, IN	116	1981	1984
17	Durham-Chapel Hill, NC	113	1984	1992
18	Des Moines-West Des Moines, IA	97	1989	1995
19	Oxnard-Thousand Oaks-Ventura, CA	90	1985	1994
20	Dallas-Fort Worth-Arlington, TX	79	1983	1992

Table 2: Descriptive statistics for city social proximity to rDNA inventors

Year	1985	1995	2005
Minimum	0	0	0
Maximum	50.787	34.786	36.163
Mean	2.214	3.485	3.460
Std Dev	7.063	6.528	6.541
Top-Ranked Cities	San Francisco New York Chicago Cleveland Boston	San Francisco San Diego New York Boston San Jose	San Francsico San Jose San Diego Boston Philadelphia

Notes: Values are centrality measures from UCINET (Borgatti *et al*. 2002)

Table 3: Descriptive statistics for cognitive proximity of metropolitan areas to rDNA

Year	1985	1995	2005
Minimum	0	0	0
Maximum	0.0241	0.0449	0.1160
Mean	0.0016	0.0055	0.0046
Std Dev	0.0029	0.0083	0.0094
Top-Ranked Cities	Madison Kennewick Elkhart College Stn. Charleston	Honolulu Shreveport Durham-Chapel Hill Madison Blacksburgh	Flagler Athens Auburn Iowa City Decatur

Table 4: Estimating the influence of different forms of proximity on the likelihood of a city inventing a first Cohen-Boyer patent in relation to the baseline hazard (single failure estimated with the extended Cox Semi-Parametric Hazard Model with time-varying covariates)

	Hazard Ratios		
Time-Fixed Covariates	Model (1) Full Sample	Model (2) Low Patent Cities (<10 patents)	Model (3) High Patent Cities (>90 patents)
Average Distance to Other Cities	1.03451*** (0.0107)	1.01377 (0.0484)	1.02711 (0.0248)
Time-Varying Covariates			
Lag Geographic Proximity	0.99059 (0.0146)	1.05948 (0.0456)	0.99897 (0.0243)
Lag Social Proximity	1.03280*** (0.0095)	1.29740*** (0.0916)	1.00837 (0.0099)
Lag Cognitive Proximity	1.01147*** (0.00243)	1.00282 (0.0048)	1.04668* (0.0269)
Lag Patent Count	1.00157*** (0.0002)	1.02773 (0.0186)	1.00096*** (0.0002)
Lag University R&D	0.99999** (2.47E=07)	0.99999 (8.18E-07)	0.99999* (4.03E-07)
Lag Industry R&D	1.00000*** (2.31E-06)	1.00000 (9.38E-06)	1.00000 (5.93E-06)
	n = 6573 Failures = 201 LL = -1048.071 LR Chi^2 = Prob > Chi^2 = 0.000	n = 2397 Failures = 17 LL = -65.944 LR Chi^2 = Prob > Chi^2 = 0.000	n = 740 Failures = 87 LL = -292.007 LR Chi^2 = Prob > Chi^2 = 0.000

Notes: All time-varying covariates are lagged one period and are interacted with log(time). Breslow method is used for ties. Robust standard errors reported in parentheses.
*** significant at the 0.01 level, ** significant at the 0.05 level, * significant at the 0.1 level.
LL = log pseudo-likelihood. Note that converting the hazard ratios to a regression coefficient by logging and then dividing by the standard error yields the usual *p*-scores.

Table 5: Estimating the influence of different forms of proximity on the probability of a city inventing a Cohen-Boyer patent (repeated events estimated with a Conditional Fixed Effects Panel Logit Model)

Independent variables	Partial Regression Coefficient (Log Odds)
Lag Geog Proximity	0.00974 (0.0169)
Lag Social Proximity	0.04570*** (0.0122)
Lag Cognitive Proximity	0.00767** (3.1092)
Lag Patent Count	0.00032 (0.0004)
Lag University R&D	-1.17E-06** (4.93E-07)
Lag Industry R&D	0.00001** (6.47E-06)
	n = 4975 Log Likelihood = -1207.355 LR Chi^2 = Prob > Chi^2 = 0.000

Notes: All independent variables are lagged one period. *** significant at the 0.01 level, ** significant at the 0.05 level, * significant at the 0.1 level. Year fixed effects included but not reported. 167 cities (4175 observations) dropped by the conditional logit because of no change in the dependent variable.