

# **The Dynamics of the Inventor Network**

## **in German Biotechnology:**

### **Geographical Proximity versus Triadic Closure**

**Anne L.J. Ter Wal**

Innovation Studies Centre  
Imperial College Business School  
South Kensington Campus, London SW7 2AZ  
United Kingdom  
[a.terwal@imperial.ac.uk](mailto:a.terwal@imperial.ac.uk)  
[www.anneterwal.com](http://www.anneterwal.com)

**Abstract:**

Economic geography has developed a stronghold analyzing how geography impacts innovation. Yet, despite increased interest in networks, a critical assessment of the role of geography in the evolution of networks is still lacking. This paper attempts to explore the interplay between geographic distance and triadic closure as two main forces that drive the evolution of collaboration networks. Analyzing the evolution of inventor networks in German biotechnology, the paper theoretically argues and empirically demonstrates that – as the technological regime of an industry changes over time – inventors increasingly rely on network resources by forming links to partners of partners, whilst the direct impact of geographic distance on tie formation decreases. Whereas initially triadic closure reinforces the geographic distance effect by closing triads among proximate inventors, over time triadic closure becomes an increasingly powerful vehicle to generate longer-distance collaboration ties as the effect of geographic proximity decreases.

**Keywords:** network evolution, geographic proximity, triadic closure, inventor networks

**JEL codes:** D85, L14, L65, R11

**Acknowledgements:**

An earlier version of this paper was published as a chapter in my PhD thesis at the Department of Economic Geography at Utrecht University. I would like to thank Ron Boschma, Koen Frenken, Ammon Salter, Paola Criscuolo, Bill McEvily, and Simone Ferriani for their comments on earlier versions of this paper. I am also grateful to the members of my PhD committee – and in particular to Robin Cowan, Stefano Breschi, and Bart Nooteboom – for their feedback and extensive discussion on this work. Further, I would like to acknowledge Utrecht University and the UK Engineering and Physical Sciences Research Council's Innovative Manufacturing Programme for its sponsorship of the research that underpins this article.

## **1. Introduction**

From the literature on the geography of innovation it is known that flows of information decay with increasing distance (Audretsch and Feldman, 1996b; Breschi and Lissoni, 2001; Boschma, 2005). Information about the availability, suitability and reliability of potential partners decreases in quality and quantity with geographic distance. As a result, this tradition of research predicts that patterns of collaboration are strongly driven by the geographic distribution of individuals and organizations and the ease at which they can exchange knowledge at distance. This causes networks to evolve largely along geographic lines. Alternatively, the literature on collaboration networks argues that high-quality information about potential partners resides in the network of prior collaboration ties (Gulati, 1995; Sorenson et al., 2006). Consequently, inventors are strongly inclined to connect to partners of partners. As a result, networks exhibit a tendency towards ‘triadic closure’, i.e. that partners of partners become directly connected, closing a triad (a set of three nodes) in the network. This theoretical perspective predicts that the spatial evolution of networks is primarily path-dependent, as the prior structure of the network largely determines how the network will evolve.

This paper attempts to unite both perspectives, exploring the interplay between geographic distance and triadic closure as two main forces that drive the evolution of collaboration networks between individual inventors. The paper argues that there is a shifting balance from geographic distance to triadic closure as the dominant force that drives network dynamics. In the early stages of industry evolution, it is expected that a relatively weak triadic closure effect reinforces the localized nature of collaboration produced by a relatively strong geographic distance effect. In the later stages of industry evolution, however, a stronger triadic closure effect may stimulate the formation of longer-distance collaboration ties by closing triads with inventors from locations further apart, which could emerge due to a relatively weaker geographic distance effect. This relative shift is attributed to changes in the technological regime of an industry (Malerba and Orsenigo, 1997).

More specifically, taking biotechnology in Germany between 1970 and 2002 as an example of an emerging, spatially agglomerated industry, it is proposed that geographic proximity is most important for

engaging in collaboration when knowledge is predominantly ‘basic’ and ‘tacit’ in the earliest stages of the industry, and gradually loses relevance in later stages where knowledge is increasingly targeted at the development of specific commercial applications (Gilsing and Nooteboom, 2006; Nesta and Saviotti, 2006). Conversely, it is argued that triadic closure gains importance as a mechanism of network evolution as an industry becomes more established. Closed triads act as vehicles of trust (Coleman, 1988; Kilduff and Tsai, 2003), which become increasingly relevant in later stages of industrial evolution where the growing focus on the development of commercial applications heightens concerns about involuntary knowledge spillovers (Liebeskind et al., 1996; Gilsing and Nooteboom, 2006). Logistic regression models of tie formation and stochastic estimation models of network evolution applied to USPTO and EPO co-inventor networks largely confirm these predictions.

Unveiling some of the basic principles that underpin the spatial evolution of collaborative innovation networks, this paper connects to two streams of research. First, it adds to an emerging body of research on network dynamics that exists outside economic geography (e.g. Gulati, 1995; Ahuja, 2000; Powell et al., 2005; Orsenigo et al., 2001). Although these studies have certainly increased our understanding of the evolution of networks, they tend to disregard the role of geography in the evolution of networks. Second, economic geography has developed a stronghold analysing how geography impacts innovation (Audretsch and Feldman, 1996b; Boschma, 2005; Breschi and Lissoni, 2009). Yet, despite increased interest in networks, a critical assessment of geography’s role in the evolution of innovation networks is still lacking (Ter Wal and Boschma, 2011). Only recently have dynamic studies of networks emerged in economic geography (Balland, 2012; Giuliani, 2010; Li et al., 2012; Ter Wal, 2011). Taking a dynamic view on the conditions under which either geographic or non-geographic principles dominate tie formation in innovation networks, the paper contributes to strengthening this body of work on the collaborative nature of innovation and the geography of innovation.

This paper is structured as follows. First, the next section will introduce the specific context of the biotechnology industry, describing how its technological regime has changed over time. The third section, then, reviews the existing literature and formulates hypotheses regarding the changing role of geographic

distance and triadic closure in network evolution. Section 4 explains the data and method. Section 5 performs the analyses in three parts: descriptive analyses plotting trends in the role of geographic distance and triadic closure over time; logistic regression models of tie formation that explore the interaction between those factors; and a stochastic estimation model that simulates how geographic distance and triadic closure jointly drive the evolution of the network. Section 6 concludes.

### INSERT FIGURE 1 ABOUT HERE ###

## **2. The changing technological regime in biotechnology**

Biotechnology can be considered an archetypical science-based industry (Pavitt, 1984; Tamada et al., 2006). Today's commercial applications in the field – ranging from medical drugs and food-processing to chemical substances – rely heavily on relatively recent scientific advancements in molecular and cellular biology (Powell et al., 1996). The origins of modern biotechnology date from the discovery of the double helix structure of DNA in the 1950s, and the subsequent discoveries of recombinant DNA and monoclonal antibody technology in the 1970s. In the 1980s scientists made considerable progress in the development of genetic engineering (Liebeskind et al., 1996). These new discoveries had enormous technological potential across industries, though particularly in the pharmaceutical industry.

Until the 1960s the knowledge base of the pharmaceutical industry was dominated by organic chemistry (Gilsing and Nooteboom, 2006), and drug development and food processing were largely based on random screening and trial and error practices (Gambardella, 1995). In Germany large pharmaceutical companies like BASF, Bayer, and Hoechst prospered in this period (Lehrer, 2005). The revolutionary discoveries in biotechnology enabled a more rational approach to the development of new chemical substances and drug design (Powell et al., 1996). This development was initially driven by small biotech firms, generally referred to as Dedicated Biotech Firms (DBFs) (Audretsch, 2001; Powell et al., 2005). These small firms were mostly university spin-offs and were closely connected to academic research laboratories (Zucker et al., 1998; Lehrer, 2005), specialized in biotechnology research and the development of products and techniques with potential commercial value. However, they lacked the resources for extensive clinical tests and complex regulatory approval procedures (Gilsing and

Nooteboom, 2006). From the mid-1980s onwards large established pharmaceutical firms started to enhance their role by giving financial support to DBFs, developing new technologies into safe and effective products, and bringing them to the market (Audretsch, 2001). Although in Germany large companies like BASF and Boehringer Mannheim entered the field of biotechnology rather late, large investments in research and development ensured that they could catch up and maintain leading positions alongside newly emerging DBFs (Krauss and Stahlecker, 2001). The increased role of large pharmaceutical companies has spurred substantial changes in the technological regime of the industry (Malerba and Orsenigo, 1997).

First, there has been a shift from a predominantly generic to a more specialized knowledge base. This development is illustrated by what is often referred to as the second biotechnology revolution (Gambardella, 1995)<sup>1</sup>. Initially, the biotechnology industry was characterized by a high level of technological uncertainty, typical for the exploration stage of an emerging technology (March, 1991). In the emergence of the biotechnology industry, progress was made in the development of basic knowledge without clear direction of where and how the new set of technologies would be applied (Liebeskind et al., 1996). This changed from the 1980s, when the combination of new genetic engineering techniques and existing insights from molecular biology were increasingly used “*as a research tool to enhance the speed and efficiency of the discovery process of new drugs*” (Gilsing and Nooteboom, 2006: 8). To this end, the knowledge involved in these new methods – such as automatic gene sequencing – became increasingly codified in commercially available documents and instrumentation (Rothaermel and Thursby, 2007) and the knowledge embodied in biotech inventions was of a more specialized character. Along these lines, Nesta and Saviotti (2005) find that in the 1980s it was knowledge diversity rather than knowledge integration that drove innovation in biotechnology, whereas knowledge integration was a more important determinant of innovative activity in the 1990s. Similarly, Audretsch (2001) and Hopkins et al. (2007) note that from the late 1980s onwards large experienced pharmaceutical firms replaced their broad

---

<sup>1</sup> It is acknowledged that the ‘biotech revolution’ might in fact not be as revolutionary as is often thought. Like most other emerging technology fields biotechnology follows a pattern of continuous incremental technological change (Hopkins et al., 2007).

learning strategies during the exploration phase, with a focused approach, targeting specific technologies and applications<sup>2</sup>. In addition, the experimentation process itself became increasingly industrialized over the 1980s and 1990s (Hopkins et al., 2007: 569). The trend from basic, generic knowledge towards increasingly specialized knowledge is illustrated by trends in USPTO biotechnology patents since the 1970s. The generality of an invention is expressed as the extent to which the follow-up technological advances (as captured by the patents that cite the focal invention) are spread across different technological fields, rather than concentrated in just a few of them (Trajtenberg et al., 1997: 27). As Figure 1 (left) shows, the generality of biotech inventions decreased substantially from the 1980s onwards.

A second and related development of the knowledge base in the biotechnology industry is a shift towards increasing appropriability, defined as the ability to reap financial benefits from an invention (Teece, 1986). In the early stage of the biotech industry, high levels of technological uncertainty made it difficult to judge the commercial value of new scientific developments and to develop industrial applications from which firms could derive value (Liebeskind et al., 1996). The ability to capture value from their inventions was substantially augmented with the increasingly applied and specialized character of patented biotech inventions, stimulating DBFs to focus more explicitly on certain technologies and spurring large pharmaceutical companies to target the development of industrial applications associated with these technologies (Audretsch, 2001; Nesta and Saviotti, 2006). At the same time, the increased codification of the technology's knowledge base accompanying this trend (see also Saviotti, 1998) implied a decreased ability to control knowledge flows and, hence, a greater risk of unintended knowledge spillovers, and imitation by competing firms (Gilsing and Nooteboom, 2006; García-Muiña et al., 2009). This increased risk is reflected in intense competition for patentable know-how. As Liebeskind et al. (1996: 429) note, strict property right regimes make that "*only firms that are the first to discover a process or product can reap any financial rewards from it*". The increasingly competitive environment for new knowledge creation heightens the need for high-trust collaboration (Gargiulo and Benassi, 2000).

---

<sup>2</sup> After extensive clinical testing and long approval procedures the first biotechnology products reached the market in the late 1980s and early 1990s (Audretsch, 2001).

Trends signaling increased efforts to derive financial benefit from their patented biotech inventions are also apparent in patent-based measures of appropriability at the aggregate industry level (Trajtenberg et al., 1997). As Figure 1 (right) shows, appropriability<sup>3</sup> – measured as the extent to which citing patents are held by the same organization as the focal patent – increases dramatically after 1985, providing further support for the conjecture that appropriability increased over the development of the biotechnology industry.

### **3. Hypotheses**

The previous section has demonstrated that the technological regime in biotechnology changed in the course of time, shifting towards a specialized knowledge base, more strongly focused on value appropriation. The central argument in the paper is that the changing technological regime has implications for the spatial dynamics of the inventor network. Two alternative mechanisms through which information about potential partners is available – and that hence underlie the formation of networks – are juxtaposed: geographic distance and triadic closure.

#### ***3.1 Geographic distance***

Economic geography and regional science have an established tradition of studying the importance of geographic proximity – or, inversely, distance – for innovation and the formation of networks. In this stream of research the concepts of geographic clusters and knowledge spillovers are strongly intertwined. When Alfred Marshall laid the foundations of the cluster concept at the end of the 19<sup>th</sup> century, he identified access to local knowledge externalities as one of the main benefits firms derive from their location in a cluster. Consequently, various studies point towards the positive effect of a firm's location in a cluster on its innovative performance (e.g. Baptista and Swann, 1998). High-tech industries, in particular, show a strong tendency to cluster (Audretsch and Feldman, 1996a). For those industries,

---

<sup>3</sup> The logic behind taking self-citations as an approximation for appropriability is that self-citations to a focal patent signal an organization's subsequent developments of the focal patent and hence, the importance the organization holding the focal patent attaches to it in terms of (potential) financial value (Trajtenberg et al., 1997).

knowledge externalities are of particular relevance, as timely access to knowledge on recent technological developments and scientific progress is a key competitive advantage.

Through analysis of knowledge flows as evidenced by patent citations, the literature on knowledge spillovers (Jaffe et al., 1993; Breschi and Lissoni, 2001) has shown that the quality and quantity of information flows are subject to decay with distance. Whilst initial research on localized knowledge spillovers leaves the mechanisms through which knowledge is transmitted unexplored, later research has explicitly searched for explanations for the localized character of knowledge transfer between individuals or organizations. Under the surface of direct observation various channels, other than direct collaboration, serve as a conduit for knowledge flows in a local system (Owen-Smith and Powell, 2004; Ibrahim et al., 2009). In particular, Singh (2005) and Breschi and Lissoni (2009) show empirically that the localized nature of knowledge spillovers is largely due to knowledge transmission in strongly localized social networks and localized flows of mobile labor. Often, by sharing a common educational background and work experience, entrepreneurs and technicians in a local area form communities exhibiting strong personal relationships across organizational boundaries (Grabher and Ibert, 2006). Particularly when communities are located in specialized geographic clusters, these social networks among inventors are strongly localized, providing at least a partial explanation for localized knowledge spillovers (Dahl and Pedersen, 2004; Suire and Vicente, 2009). Mobility of labor constitutes another important means of knowledge transfer across firms (Song et al., 2003). Since mobile labor is strongly inclined to stay in its home region, this mechanism of knowledge transfer also contributes to the localization of knowledge flows (Almeida and Kogut, 1999).

As mechanisms of knowledge transfer between agents are strongly localized, agents are more likely to know about each other in close geographic proximity than if they were located further apart. Therefore, information about the availability, suitability and reliability of potential partners is subject to decay with distance. In consequence, the geographic distance between two inventors negatively affects the probability that they will engage in collaboration. This intuitive relationship between geographic distance and network formation has been proved to exist in a variety of research contexts (e.g. Bell and Zaheer,



2007; Maggioni et al., 2007; Abramovsky and Simpson, 2011), but cannot be assumed to be generally applicable (Gordon and McCann, 2005; Huber, 2012). Particularly, the extent to which geographic distance matters for network formation will depend on the nature of the knowledge.

The importance of geographic distance in the formation of collaborative relations among inventors depends on the level of ‘basicness’ of the knowledge involved. Zucker et al. (1998) demonstrate that the initial development of technology by both DBFs and large pharmaceutical companies is strongly supported by the co-presence of academic scientists who are actively contributing to the basic science underlying the technology. Such company-scientist links are particularly prone to be localized when they are formed around ‘star scientists’ who are most likely associated to important scientific discoveries (Audretsch and Stephan, 1996). This suggest that the local component of collaborative innovation is particularly important when basic, early-stage tacit knowledge of scientists has not yet diffused to the broader community and is not yet available in readily accessible codified form. Along the same lines, Gittelman (2007) observed in a study of research collaboration in biotechnology that local partnerships are more likely to lead to patented outcomes, whereas distant collaborations instead result in scientific impact. Strongly codified knowledge for the production of scientific knowledge is more easily communicated at distance than knowledge geared towards the development of innovations. In fact, scientific knowledge is typically purposefully disseminated, whereas knowledge for commercial innovation is deliberately protected. Yet, although a patent itself is a codified piece of knowledge, the tacit knowledge that was needed to develop it is more easily transferred and co-produced at geographic proximity (Fleming et al., 2007).

Tacit knowledge, strongly embedded in human capital and a crucial building block of generating basic knowledge, is most easily exchanged through repeated face-to-face interaction or the mobility of people, which are easier and more frequent at short geographic distances (Zander and Kogut, 1995; Gertler, 2003; Faulconbridge, 2006; Torre, 2008). The knowledge base of an industry is predominantly basic and tacit in young emerging industries. Along those lines, Audretsch and Feldman (1996a) argue that in the early stages of a new technology, when knowledge tends to be highly tacit, firms and

individuals benefit most from geographic proximity. When an industry grows and matures, knowledge gets more codified and is more easily transferable over larger distances (Cowan et al., 2004). As described in the previous section, the development from tacit to increasingly codified knowledge – that accompanied the trend from a generic to more specialized knowledge base – could also be observed in biotechnology. Accordingly, it is expected that the importance of geographic distance in the formation of networks to decline over time, and the distance over which inventors collaborate in German biotechnology to increase over time. Therefore Hypothesis 1 is formulated as follows:

*Hypothesis 1: The role of geographic distance in tie formation declines over time, as the technological regime in biotechnology experiences a shift from tacit to increasingly codified knowledge.*

### **3.2 Triadic closure**

Triadic closure is the tendency for new links to be formed between the direct network neighbors of a node, resulting in closed triads in the network (Davis, 1970). That is, two direct partners  $j$  and  $k$  of a node  $i$  get directly connected, producing a completely connected triad among nodes  $i$ ,  $j$ , and  $k$  in the network. At the dyad level this implies that a prior indirect tie between  $j$  and  $k$  via  $i$  – at network distance two – turns into a direct tie. A tendency towards closure produces dense cliques of strongly interconnected actors in the network (Skvoretz, 1991).

Triadic closure is mechanism of tie formation that relates to the broader force to form connections to those with whom one is indirectly connected. Networks of prior ties act as a “*repository of information on the availability, competencies, and reliability of prospective partners*” (Gulati and Gargiulo, 1999: 1440). As this type of information is of the utmost relevance for those seeking new collaboration partners, it can be argued that the shorter the network distance between any two inventors in the network of prior ties, the higher the likelihood that they become directly connected, i.e. starting to collaborate themselves. As such, inverse network distance has been operationalized as a form of social proximity (Sorenson et al., 2006), which itself is defined as the extent to which individuals or organizations are socially embedded in

a network of friendship relations and repeated interactions (Boschma, 2005) that carry information about potential partners (Boschma and Frenken, 2010).

Triadic closure represents the specific case where there is only a single intermediary between two nodes in a network, i.e. when two nodes are separated by a path of length two in the network of prior ties and, hence, have a level of social proximity of 0.5. The fundamental mechanism here is the *tertius iungens* or the ‘third who joins’, “*connecting people in one’s social network by either introducing disconnected individuals or facilitating new coordination between connected individuals*” (Obstfeld, 2005: 102). Having a shared intermediary provides both parties with detailed information about the availability, suitability and reliability of that common partner. Given the direct experience the intermediary has had collaborating with both parties, it acts as an alternative, and more precise channel of such information on geographic proximity. As such, inventors who have a shared intermediary are more likely to form a collaboration tie than those who do not, as a result of which networks exhibit a tendency towards triadic closure.

This effect is further strengthened because the closed structures that triadic closure produces foster the development of trust in the relation (Uzzi, 1997). In sociological research the presence of closed triads is generally interpreted as a sign of social capital (Coleman, 1988; Kilduff and Tsai, 2003). Typical properties of closed and cohesive network structures are reciprocal, repeated and frequent interactions between the actors, with the possibility to cross-check information obtained from direct ties through indirect paths in the network. Each of these properties stimulates the creation of trust in the collaboration (McEvily et al., 2003; Walker et al., 1997; Buskens, 2002). In this regard, Reagans and McEvily (2003) demonstrate that strong social cohesion around a relationship reinforces the willingness and motivation to invest time, energy and effort in sharing knowledge with others. Hence, trust in dense parts of the network facilitates intensive exchange of complex or sensitive knowledge (Zaheer and Bell, 2005). As a result, inventors will be even more inclined to connect to partners of partners when sensitive knowledge and risk of unintended spillovers are at stake. Thus, in situations of high risk and high cost to opportunistic behavior, actors have a clear preference to form ties embedded in dense structures, which may result in

network closure (Gargiulo and Benassi, 2000). Several studies empirically demonstrate the preference of organizations and individuals to form ties embedded in cohesive network structures. In the context of US venture capital networks Sorenson and Stuart (2008) found that, at higher risks, actors will be inclined to form network relationships with socially proximate individuals. Beckman et al. (2004) argued that in a situation of strong market uncertainty at the (early) exploitation stage the need for trust is high.

In German biotechnology a heightened need for trust through triadic closure arises as the technological regime in biotechnology evolves. As the focus on the development of commercial applications becomes more salient and, at the same time, knowledge codification increases the risk of involuntary knowledge spillovers (Saviotti, 1998), it is argued that the tendency to form closed triads in the network increases. Unintended knowledge spillovers are costly in high-tech fields like biotechnology, where it matters to be the first to bring new industrial applications to the market and strict patent regimes ensure the first-mover to reap the benefit associated with them (Liebeskind et al., 1996)<sup>4</sup>. To be clear, the argument is not that trust is not important in early stages of the industry where knowledge was predominantly tacit. Yet, it is proposed that the necessity of trust for collaboration is heightened when the technological regime in biotechnology is characterized by an increased desire to appropriate returns from the R&D efforts. Therefore, triadic closure is expected to become increasingly important in the German biotechnology inventor network over time.

*Hypothesis 2: The role of triadic closure in tie formation increases over time, as the technological regime in biotechnology experiences a shift from tacit to increasingly codified knowledge.*

### **3.3 Interplay between geographic distance and triadic closure**

Triadic closure in network evolution introduces an element of path dependence, where the “current state of affairs cannot be derived from current conditions only, since “the current state of affairs has emerged

---

<sup>4</sup> Notwithstanding the fact that patents legally protect the innovation, they by no means overcome the risk of involuntary knowledge spillovers. In actual fact, the codification of the new knowledge embodied in the innovation by means of a detailed description on the patent, facilitates the use of that knowledge by others, albeit in a slightly modified form.

from and has been constrained by previous states of affairs” (Boschma and Frenken, 2006, p. 280). The prior structure of the network largely drives how the network will further evolve. Given the expected shift with the decreasing influence of geographic distance on tie formation and the increasing influence of triadic closure, this gives way to an interesting interplay between both forces.

In the early stages of the industry evolution, the effect of geographic distance on tie formation is strong, whereas the effect of triadic closure is relatively weak. As a result, the majority of open triads in the network – which embed for potential for triadic closure – are likely to be among proximate inventors. Triadic closure then reinforces the localized patterns that are produced by the strong effect of geographic distance. In a similar fashion that initial location decisions of firms may lead to path-dependent creation of geographic clusters when these firms generate spin-offs (Arthur, 1994), the initial collaboration decisions among proximate inventors may lead to a path-dependent process where dense clusters of collaboration mostly emerge among locally concentrated groups of inventors. In effect, the predominance of localized collaboration lingers on, although the direct effect of geographic distance on tie formation starts to fade.

In later stages of industry evolution, however, the effect of geographic distance on tie formation is dampened, whereas the effect of triadic closure is amplified. This changes the interplay between the geographic distance and triadic closure effects and as such moderates the path dependency of triadic closure on the geographic distance. That is, the weakened effect of geographic proximity increases the generation of ties among more distant inventors. As a result, the power of triadic closure to close triads among distant inventors also increases, functioning as an endogenous driver of change (Martin, 2010) that breaks the stable localized pattern of collaboration produced by the earlier phase in which triadic closure amplified the effect of geographic distance. Gradually, triadic closure then transitions into a mechanism that facilitates the formation of longer-distance collaborations.

*Hypothesis 3: The geographic distance of the ties are formed through triadic closure increase, as the technological regime in biotechnology experiences a shift from tacit to increasingly codified knowledge.*

## 4. Data and Method

### 4.1 Patent data

The empirical analysis is based on patent data. Patent data is increasingly used in scientific research as relational data (Breschi and Lissoni, 2004). This study uses patent data to reconstruct inventor networks in retrospect. Biotech firms have always exhibited a strong tendency to protect their innovations through patents (Blind et al., 2006). This makes patent data a reliable source of longitudinal data on innovation for this sector. The study relies on two sources of patent data: USPTO patents from the publicly available NBER Patent Citations Data File (Hall et al., 2001) and EPO patents from the OECD REGPAT database (Maraut et al., 2008). Relying on two datasets not only provides ways to validate findings across datasets, it also extends the time frame of the analysis. The USPTO dataset covers patents from 1963 to 1999<sup>5</sup> and the EPO dataset from 1978 to 2002. All USPTO patents with application year 1996 or later have been excluded from the dataset; the dataset is incomplete as not all patents applied for in these years were granted before 1999. The application date is used for dating patents, since this is closest to the time of invention. Data from 1970 onwards was analyzed. Although it is acknowledged that 1970 is an early start for analyzing biotechnology, particularly outside the US, it is fair to say that the early patents cover discoveries lying at the roots of biotechnology, as the inventors involved in those early patents recur on later patents that can be more convincingly labeled as biotechnology.

From the USPTO dataset, all patents in subcategory 33 (biotechnology, as defined by Hall et al. 2001) were selected, encompassing the USPTO-defined patent classes 435 (molecular biology and microbiology) and 800 (multicellular living organisms and parts thereof). For the EPO dataset, patents were selected on the basis of the IPC-code of the patents, relying on Schmoch's (2003) IPC-to-industry concordance table to determine which patents can be classified as biotechnology. Then, from both datasets all patents with at least one inventor resident in Germany were retrieved. Foreign inventors that co-occurred with German inventors on a patent were excluded from the database. Since the information on

---

<sup>5</sup> Since the inventor-level USPTO database starts from 1974, the information on inventors for the years 1963-1974 has been added manually from the USPTO website's Patent Full-Text Database.

their co-invention links to other foreign inventors is lacking, the study disregards foreign inventors and limits the spatial scale of analysis to German-based inventors. All patent data were checked thoroughly for obvious typing errors in inventors' names (USPTO) and coding errors in inventor IDs (EPO). This is crucial for reconstructing the networks in the software package UCINET (Borgatti et al., 2002), in which the linking algorithm is based on unique inventor names or IDs.

Figure 2 compares the EPO and USPTO databases, making use of priority dates to identify which USPTO patents have equivalent patents at EPO and vice versa. For the overlapping years between the datasets (1978-1995), 77% of the USPTO patents have an EPO equivalent and 55% of the EPO patents have a USPTO equivalent. The actual number of patents with an equivalent patent in the other dataset, however, is considerably lower than that. Due to the difference in which patents are classified as biotechnology in both patent systems some USPTO patents which have an equivalent at EPO are not included in the EPO dataset and the other way around. The number of patents at EPO is higher than at USPTO, resulting in a higher share of EPO patents without a USPTO equivalent than vice versa. The upper two graphs in Figure 3 compare the aggregate number of patents and inventors over time for both datasets, confirming that – apart for the first overlapping years – the number of patents and inventors is consistently higher for EPO than for USPTO. Further, the graphs show a dramatic increase in EPO patents and inventors after 1995, the final year for which USPTO data was available. With the number of inventors per patent and the number of patents per inventor being fairly constant (see the lower two graphs of Figure 3), the growth of the network is mainly the result of the increasing number of patents over time. A final observation regarding the set of overlapping patents across both datasets is that EPO patents tend to be applied for earlier than the equivalent USPTO patents. The application year differs for 41.6% of overlapping patents, with an average delay of USPTO patents vis-à-vis EPO patents of 2.6 years.

### INSERT FIGURES 2 AND 3 ABOUT HERE ###

#### ***4.2 Generating networks from patent data***

The USPTO and EPO datasets were used to construct collaboration networks at the level of the individual inventor. In such co-inventorship networks, two individual inventors are linked if they have worked on the

same patent. In other words, I used a one-mode representation at the inventor level of a two-mode network with patents and inventors. A five-year moving window procedure was applied to reconstruct the inventor networks: each yearly network observation contains all co-invention links for that year and the preceding four years. In line with other studies on inventor networks (Fleming et al., 2007), it is assumed that co-invention links exist during five years. The underlying assumption is that the collaboration process which ultimately led to a patent application took five years on average, which may be justified by anecdotal evidence that collaborative research projects in biotechnology are typically long-term and are underpinned by open-ended contractual arrangements (Shan et al., 1994; Powell et al., 1996).

In the specific context of the study of the geographic and social determinants of tie formation, analyzing networks at the level of the individual inventor is preferable over analyzing them at the level of the patent applicant (i.e. most typically the organization). The inventor level is the most detailed and pure level of collaborative innovation available through patent data<sup>6</sup>. For example, considering the fact that patents developed by subsidiaries are often assigned to the company's headquarters, applicant-level network analysis is often problematic for studying the spatial structure of the network. The analysis is limited to the networks among incumbent inventors, who form the "core" inventors in German biotechnology. Incumbent inventors are defined as those occurring at least twice in the data, and in different years. The selection on incumbent inventors is made in order to decrease the volatility of inventors entering and exiting within a short timeframe, which may conflate effects that happen among the core of inventors in the network.

---

<sup>6</sup> Certainly, in a co-invention network at the individual level, it is likely that inventors that co-occur on a patent work for the same company. Yet, for various reasons it cannot be automatically assumed that all inventors mentioned on a patent work for the patent's applicant. First, Giuri et al. (2007) demonstrated on the basis of a large-scale survey of European inventors that on average more than 20 percent of all patents involved some form of collaboration with an external organization, usually not mentioned on the patent; about 15 percent of the surveyed patents included external co-inventors. Second, quite often inventors appear on patents of more than one applicant. In a survey among European biotechnology firms Laforgia and Lissoni (2006) found out that about 20 percent of these cases of 'multiple-applicant-inventorship' are due to labor mobility. The remaining 80 percent are largely the result of mergers and acquisitions, or inventors that also occur on the patents of universities and public research institutes. In addition, many patents are sold on the market for technology. Particularly small firms, including DBFs, often decide not to make the substantial investment to commercially exploit the patent but to sell the patent to larger firms (Giuri et al., 2007).



### ***4.3 Measures and estimation***

The analysis will be carried out in three steps. First, descriptive statistics will shed light on how the average geographic distance over which inventors collaborate and the tendency towards triadic closure have changed over time. Second, logistic regression techniques will estimate how the probability that a collaborative tie is formed between two inventors depends on the geographic distance, social proximity and the interaction between them. Third, as a final step to understand how geographic distance and triadic closure jointly drive the dynamics of the network, the study applies a stochastic network simulation procedure, using the program SIENA (Simulation Investigation for Empirical Network Analysis), developed by Snijders et al. (2001; 2007). Recent examples in economic geography where SIENA models have been applied to the evolution of networks are Balland (2012) and Giuliani (2010). Choice of methods will be justified and explained in the Results section.

**Tie formation** is the dependent variable in the analysis, albeit defined slightly differently in the regression models and the SIENA models. Conceptually, the dependent variable is defined as whether or not a tie exists between any pair of these inventors at time  $t$ , given the set of inventors that is part of the co-invention network in German biotechnology at time  $t-5$ . Correspondingly, the logistic regression models aim to explain why certain connections in the network exist as a function of prior ties. The stochastic estimation model, by contrast, aims to explain why certain connections appear, remain or disappear in between two time periods. It aims to reveal the forces that drive the evolution of a network in between two observations and as such makes the “evolution of the network” the dependent variable in the network.

**Geographic distance** between inventors is the first independent variable. The place of residence of the patent’s inventors is used to determine the location of innovation in biotechnology, deliberately disregarding the location of the patent applicant (see also Nicholas, 2009). Large companies tend to assign the patent to the headquarters, even when the patent might have been developed in one of the company’s subsidiaries outside the headquarters’ region. Notwithstanding the possibility that some inventors might live in another region than where they work, inventor location is generally agreed to be a more reliable

approximation of where the innovation was developed (Acs et al., 2002; Ejeremo and Karlsson, 2006). The distance between two inventors is expressed in distance “as the crow flies” between their places of residence, calculated on the basis of city geographic coordinates<sup>7</sup>.

**Triadic closure** and social proximity are related constructs that form the second set of independent variables. For each pair of inventors at time  $t$ , social proximity is defined as the inverse of the path length between them in the co-invention network at time  $t-5$ . For those inventors between whom no prior path exists (i.e. who belong to separate components of network), social proximity is set to 0. In essence, triadic closure is a specific case of the social proximity effect on tie formation. Triadic closure is defined as the tendency that a tie forms between two inventors at time  $t$  who were connected with a path of length two at time  $t-5$  (i.e. at social proximity = 0.5). In other words, triadic closure is conceptualized as the tendency that a tie forms at time  $t$  between partners of partners at time  $t-5$ .<sup>8</sup>

### INSERT FIGURES 4 AND 5 ABOUT HERE ###

## 5. Analysis

### 5.1 Descriptive analysis

Figure 4 displays the pattern of spatial concentration in German biotechnology, as expressed in the number of inventors per spatial unit per year (439 German NUTS3-districts), based on USPTO data from 1970 to 1995. Through the course of time five main clusters of biotechnology inventors have emerged: the Rhineland area with Wuppertal, Cologne, and Düsseldorf as its main centres; the Rhine-Neckar triangle around Heidelberg, Darmstadt, and Mannheim; Munich and the area around the Starnberger See; the capital city of Berlin; and the small university city of Marburg. This pattern of concentration is consistent with earlier studies on spatial concentration of German biotechnology (Zeller, 2001; Krauss and

---

<sup>7</sup> For EPO data from the REGPAT database inventor location is coded with NUTS3 codes. Hence, I took the distance between capitals of those NUTS3 regions to calculate the (approximate) distance between inventors.

<sup>8</sup> This conceptualization of triadic closure does not suffer from the two-mode nature of the data that typically inflates the number of closed triads in a one-mode representation of a network as any group of three inventors on a single patent forms a closed triad. In the stochastic estimation models this may, however, be a concern. Yet, as the number of patents per inventor and the number of patents per inventor are fairly stable over time (see Figure 2) the bias towards detecting significant triadic closure effects this may create can be assumed to be constant over time.

Stahlecker, 2001). The pattern of spatial concentration in these five clusters has been remarkably stable over time. All five clusters were already present in the 1970s, when biotechnology was not even known as such.

Although the pattern of spatial concentration has been fairly stable over time, the distance over which inventors collaborate has changed. Figure 5 (left) indicates the changing geographic distance of collaboration over time, expressed in kilometers, for USPTO and EPO data. Both lines show a significant increase over time in the average geographic distance between collaborating inventors, from under 20km in the late 1970s to 70-100 km from the 1990s onwards. Only in the late 1980s (EPO) or early 1990s (USPTO) do we observe a break in the upward trend. The fact that this spike is observed earlier for EPO than for USPTO patents may be partly due to the fact that EPO patents are applied for earlier than their equivalent USPTO patents.

Figure 5 (right) demonstrates the role of triadic closure in the dynamics of the network. If triadic closure plays a role in network evolution, it is expected that a high number of triads with two connections at  $t-5$  will be closed at time  $t$ . Every pair of nodes that are connected by a path of length 2 (through one shared intermediary) has potential for triadic closure. The tendency for triadic closure is expressed as the ratio of the observed number of closed triads over the number of random expected closed triads<sup>9</sup>. As may be expected, at any point in time the number of observed closed triads among incumbent inventors in the network is higher than the number of random expected ones. However, the extent to which this is the case – i.e. the tendency for triadic closure – differs over time. The ratio of triadic closure fluctuates between 5 and 10 from 1975 to 1990, after which an upward trend sets in towards a ratio around the year 2000 where observed closed triads are 20 times more likely than could be expected at random.

### TABLE 1 AND FIGURE 6 ABOUT HERE ###

---

<sup>9</sup> The latter is obtained by calculating the share of new possible links that close a triad among all possible new links in the network. Then, if new links are formed randomly, this share would be equal for actual new ties, and hence, the random expected number of closed triads is expressed as the product of the share of potential new ties that close a triad and the actual number of new ties among incumbent inventors that were formed between  $t$  and  $t-5$ .

## 5.2 Regression analysis

The descriptive statistics have suggested that the evolution of network dynamics is characterized by a decreasing role of geographic distance and an increased role of triadic closure over time. This section applies logistic regression and graphical techniques to explore the interplay between geographic distance and triadic closure in the network's dynamics. As dyads (inventor pairs) are the unit of analysis, each inventor occurs in multiple observations leading to non-independence of observations. To counter this problem, I use conditional logit models with random intercepts for each dyad (see also Reagans, 2011). Dyads, which co-occur over different years, thus form the higher-order groups for which standard errors are adjusted. An additional advantage of this method is that it accounts for unobserved heterogeneity. As the regression models lack control variables these may suffer from omitted variable bias. Assuming that the relationship between geographic distance, social proximity and tie formation is the same across dyads, the variable intercept captures the heterogeneity in dyad-specific aptitude to engage in collaboration with one another caused by unobserved factors.

Table 1 shows the separate models for USPTO and EPO data, for the entire period for which data are available and for separate time periods, and with sample splits by period<sup>10</sup>. Observations were grouped for five-year periods for the USPTO data. For EPO data observations were grouped for three- to five-year periods to ensure the grouping of years corresponded with the EPO periods to allow for benchmarking of the results. The results with three- or four-year periods were robust to the ones presented.

The USPTO-models show that the coefficient of geographic distance changes from -0.063 in 1976-80 to -0.010 in 1986-1990 after which the value stabilizes around that value. The EPO models show very similar coefficients for the overlapping years (1987-1990 and 1991-1995). The last two EPO models also demonstrate that the effect of geographic distance on tie formation is relatively stable from the late 1980s onwards. Taken together, the regression analyses lend support to Hypothesis 1 which predicted that

---

<sup>10</sup> In logit models comparison of coefficients of separate equations for different groups (in this case time periods) is preferable over models with time-interactions (Hoetker, 2007), although in this case models where all variables were interacted with time periods yielded results with very similar interpretation to the ones presented in the paper.

the role of geographic distance would decrease over time as the technological regime of the biotechnology industry experiences a shift from tacit to increasingly codified knowledge.

In terms of social proximity, the USPTO models show a gradual increase in coefficients from 1976-1980 to 1991-1995. The EPO models show slightly different coefficients for the overlapping years (lower than for USPTO for 1987-1990 and higher than for USPTO for 1991-1995), yet comparable in order of magnitude. Only for 2000-2002 can we observe a substantial increase in coefficient size. Overall, the logistic regressions provide support for Hypothesis 2.

The interaction between geographic distance and triadic closure is only consistently significant for models until 1990. For more recent years, the interaction is either statistically insignificant or significant with small coefficient size. This suggests that – as predicted by Hypothesis 3 – triadic closure initially reinforces the effect of geographic distance, but that this effect fades out over the course of time. However, as sign and significance of interaction coefficients have limited interpretability for logistic models (Hoetker, 2007), graphical techniques will be used to further interpret the interaction between geographic proximity and triadic closure. Given the coefficient estimates of the full regression model<sup>11</sup>, the technique predicts the likelihood that a tie forms between inventors at varying levels of the variables of interest. The procedure (see Zelner, 2009) assumes that the regression coefficients are drawn from a multivariate normal distribution, predicting the probability of tie formation over 1000 iterations for each combination of values of the variables involved. In this case, I estimate and compare the probability of tie formation at increasing levels of geographic distance at multiple periods in time, for inventors involved in a potentially closed triad (i.e. at social proximity = 0.5) versus inventors who do not have a prior network path between them (i.e. at social proximity = 0).

Figure 6 visualizes how the probability that two inventors collaborate depends on geographic distance and triadic closure and how these effects change over time. The upper two graphs plot the

---

<sup>11</sup> The graphing of interaction using this method does not support conditional logit models. Therefore, the graphs are based on standard logistic models with robust standard errors, separately for each time period. These regression models are largely consistent with the higher-order conditional logit estimates in coefficient sign, size and significance, and are available in Appendix A.

probability of tie formation at time  $t$  for inventor pairs between which no prior path exists at time  $t-5$ . The upper left graph shows for USPTO data that in the first time period (1976-1980) the probability of tie formation drops sharply with increasing distances. Only ties at less than 50 km distance are as likely to occur as a random tie in the network. Over 1981-1985 and 1996-1990 this effect of geographic distance is dampened, after which we observe a modest increase for 1991-1995. For overlapping years, the EPO graph (upper right) shows a similar pattern, although overall probabilities relative to random are somewhat higher. It also shows that the relative reappraisal of the role of geography – for those who were prior unconnected through indirect paths – observed with USPTO data for 1991-1995 sets through for the final two EPO observation periods (1996-1999 and 2000-2002).

The middle two surface graphs plot the probabilities that two inventors connect through triadic closure, i.e. that those who are connected through one intermediary at time  $t-5$  connect directly at time  $t$ . At first sight, the way in which the surface depends on geographic distance seems strongly similar to the situation where social proximity is set to 0, yet the pattern is more pronounced. In early periods the probability of tie formation quickly drops with increasing distance. This effect is dampened over 1981-85 and 1986-90, after which the trend reverses<sup>12</sup>. On closer inspection, however, we observe that social proximity indeed moderates the effect of geographic distance on tie formation. Despite the fact that the likelihood of ties at close geographic distance increases from 1986 onwards (i.e. the slope of the depth axis at distance is 0), the slope of the distance axis of the surface flattens over time. In fact, the distance at which a tie at social proximity is 0.5 is as likely as a random tie increases over time: from around 40 km in 1976-1980 to 200 km in 1991-1995 for USPTO data, and from 200 km in 1991-1995 to 300 km in 1996-1999 for EPO data, lending support to Hypothesis 3. This trend is not observed at low levels of social proximity. These differences show more evidently in the lowest two graphs, which plot the change in probability associated with an increase of social proximity from 0 (no prior path) to 0.5 (triadic closure). In the first observation period (1976-1980), triadic closure fails to foster the creation of ties at greater

---

<sup>12</sup> For overlapping periods of USPTO and EPO data, the probabilities relative to random ties are again higher for EPO data than for USPTO data.

geographic distance. Already in 1981-1985, however, we observe that in particular for inventors who are distant from each other having an indirect tie through a shared intermediary dramatically increases the odds of directly connecting. This effect is slightly weaker for subsequent periods, yet remains stable for both USPTO and EPO data across the 1990s and early 2000s. In particular inventors at greater distance from each other are much more likely to start collaborating, if they are connected through an indirect path of prior collaboration.

### INSERT TABLE 2 AND FIGURE 7 ABOUT HERE ###

### ***5.3 A stochastic model of network evolution***

The explanatory power of standard regression techniques as models of tie formation is limited. In particular, regression models of tie formation fail to account for the prior structure of the network. For example, the effect of triadic closure is dependent on the number of indirect paths (i.e. partners of partners) a node has available. As a result, some networks offer more potential for triadic closure than others. The probabilities are also conflated due to the fact that networks grow bigger over time and that the general tendency to form ties in a network may fluctuate over time. For example, the effect of network size may explain the observation that the probabilities of tie formation relative to a random tie were consistently higher for the larger EPO networks than for the smaller USPTO networks. It may also explain the finding that ties at very small distance were becoming more likely over time, as the proportion of these close-distance ties over all possible ties decreases when networks increase in size. Stochastic estimation models of network evolution alleviate some of those concerns (although they inevitably also raise others).

SIENA has been specifically designed for the statistical analysis of dynamic networks<sup>13</sup> and the first applications now have found their way to economic geography (Balland, 2012; Giuliani, 2010). It simulates how a network has evolved between subsequent network observations, in this case with five-year intervals. Through repeated simulations of how a network may have evolved between  $t$  and  $t+5$ , the

---

<sup>13</sup> It is acknowledged that the main power of stochastic estimation models in SIENA lies in the modelling of the co-evolution of network structure and actor attributes, in which the effects of social selection (attributes influence network) and social influence (network influences attributes) can be disentangled. However, the technique is equally valuable and appropriate for gaining insight into the principles of network dynamics without co-evolution.

model iteratively varies the strength of factors that may drive the evolution of the network, searching for the parameter values that best predict the actual evolution. In this case, it varies the parameters “geographic distance” and “triadic closure” (alongside with two control variables) in order to estimate those parameter values for each of these forces that can best predict how the network has evolved from its state at time  $t$  to its state at time  $t+5$ . Since the sequence of events that have made the network evolve between two observations is unknown, the SIENA program simulates how the network has evolved from one state into the next. This simulation process takes place on the basis of Monte Carlo repetition, the default number of repeated simulations of the network evolution process being 1000. In the simulation of the network evolution process, ties can be created or dissolved and node/dyadic characteristics (e.g. distance between inventors) can change. Each of these changes can take place multiple times. Hence, the possibility that links are created and again dissolved between two observation points is left open<sup>14</sup>.

A Methods of Moments estimation procedure is used to estimate parameter values for the selected network evolution mechanisms. Each parameter is associated to a target statistic, which describes the visible outcome of the effect. The target statistic of the triadic closure effect, for instance, is the observed number of closed triads. SIENA iteratively searches the parameter values that lead to a minimal deviation between the generated and observed values for these target statistics. This estimation is a stochastic process, since the repeated network simulation runs are not fully identical. Repeated estimations might lead to slightly different outcomes. Therefore the simulation process has to be rerun at least twice in order to check whether stable outcomes have been obtained. The extent to which model estimation converges to stable outcomes is specified for each parameter by a convergence t-statistic. Values under 0.100 generally indicate good convergence (Snijders et al., 2007). Only estimation models for which this condition has been met for all parameters are reported.

---

<sup>14</sup> For undirected networks there are various algorithms that define the decision rules of a single simulation run. The model is based on the ‘unilateral initiative and reciprocal confirmation’ algorithm, in which a new link is created or dissolved when one actor takes the initiative and the other actor confirms. This algorithm is closest to reality, more so than the ‘forcing model’, for instance, in which a link change that is proposed by one actor is automatically accepted by the other.



Table 2 shows the outcomes of the stochastic estimation model in SIENA for the non-overlapping network observations, starting from 1975, separately for USPTO and EPO data. Figure 7 graphically depicts the parameter estimates for triadic closure and geographic distance also for the intermediate, overlapping network observations. The dotted lines indicate the lower and upper bound of the 95 percent confidence interval. For the observation 1970-1975 the stability in the network – expressed by the number of links retained – was too low for the estimation to converge. The “Network change” sections of Table 1 shows that for the remaining network observations the number of links retained account for approximately half of all links at time  $t$ . All reported models are based on repeated estimations and convergence is good ( $t < |0.1|$ ) for all models. Robustness of the results has been tested; estimation models on the basis of the “Forcing model” algorithm have yielded very similar results to those reported.

Four parameters were estimated. The first two parameters are generally included in any estimation model (Snijders et al., 2007). The rate of change parameter accounts for the number of links that are created or dissolved. This parameter is consistently higher for the EPO analyses than for the USPTO analysis, which corresponds with the fact that the number of ties retained is a smaller proportion of all new ties for EPO than for USPTO. The second parameter is degree. This parameter accounts for the observed density in the network and is generally considered to be the ‘baseline’ parameter that indicates the general tendency of nodes in the network to increase or decrease the number of direct links they have (Snijders et al., 2007). It can be interpreted as the ‘cost’ or ‘benefit’ of having additional links, irrespective of other mechanisms, that make nodes decide to create or dissolve links. For the German biotechnology inventor network the parameter is consistently negative and significant, implying that inventors find it ‘costly’ to increase the number of collaboration partners.

The third parameter concerns the importance of geographic distance as a force driving network dynamics. Figure 7 shows that – in line with the descriptive and regression analysis and in support with Hypothesis 1 – that the effect of geographic distance diminishes from the 1970s to the early 1990s (i.e. the parameter values move closer to zero), after which it stabilizes from the early 1990s onwards. For overlapping years, the EPO and USPTO analyses give very consistent estimates of the geographic distance

effect. This parameter is negative and significant for nearly all network simulations, apart from the observations from 1985-1990 until 1988-1993 of the USPTO analysis. This temporary spike in the trend is observed earlier for EPO than for USPTO data, which may be driven by the fact that EPO patents may have been applied for earlier than the (equivalent) USPTO patents.

The fourth parameter is triadic closure. The parameter is positive and significant for all observations in both the EPO and USPTO analyses. As Figure 7 shows, the effect increases over time. Only in the late 1980s (for EPO data) or early 1990s (for USPTO) data, do we observe a trend break, where the effect of triadic closure on tie formation is stronger for a few years and then drops back to previous levels. A similar spike was observed in the descriptive analysis. Yet, overall we discern an increasing trend for the role of triadic closure in network evolution. This is in line with the expectations formulated by Hypothesis 2<sup>15</sup>.

## **6. Discussion**

### ***6.1 Synthesis of results and theoretical contribution***

The paper has examined the interplay between geographic distance and triadic closure as two forces that underpin the evolution of collaborative inventor networks. It has been argued that information about the suitability and reliability of potential collaboration partners typically decays with increasing geographic distance, but can alternatively be available through the network of prior collaboration ties resulting in the tendency of partners of partners to become connected, closing a triad in the network. Taking collaboration networks of inventors in German biotechnology as an example, it has demonstrated that the role of geographic distance and triadic closure as mechanisms of tie formation and network evolution shifts over time as the technological regime of the industry changes (March, 1991; Malerba and Orsenigo, 1997; Audretsch, 2001; Nesta and Saviotti, 2006).

---

<sup>15</sup> Please note that I do not include an interaction effect between geographic distance and triadic closure in these models. Unlike the regressions models where such an interaction was used to detect how triadic closure may reinforce or weaken geographic distance effects, the two forces are ‘independent’ drivers of change in the stochastic model. As a logical extension of the observation that their strength shifts over time, it can reasonably be assumed that triadic closure indeed fosters the creation of longer-distance collaborations over time, as predicted in Hypothesis 3 and supported by the logistic models.

More precisely, through a combination of conventional regression analyses and stochastic estimation models of network evolution, this paper shows that geographic proximity between inventors is mostly relevant for tie formation in the early stage of the industry, when its knowledge base is largely ‘basic’ and ‘generic’ as opposed to specialized and applied. By contrast, triadic closure gains relevance once the industry becomes more established, with a growing focus on the development of specific applications, higher levels of knowledge specialization and codification and the associated risk of unintended and costly knowledge leakages. Whilst inventors initially tend to collaborate with geographically proximate partners, they increasingly direct their partner selection towards the principle of the *tertius iungens* (Obstfeld, 2005). That is to say, inventors increasingly utilized the network’s resources by forming collaborative links with the partners of their partners. The high-quality information two partners can gather about each other through their shared intermediary as well as the closed network structure that comes about once they become directly connected foster the development of a high-trust environment. Ties that close a triad thus become increasingly attractive in an industry where appropriation of value from research is a primary concern (Uzzi, 1997; Buskens, 2002; Zaheer and Bell, 2005). As the industry grows, inventors decide to collaborate not necessarily with local partners. Instead, they increasingly select new partners they come to know through their current partners, paying less attention to whether they are geographic proximate or distant.

The shifting emphasis produces path-dependency for how the spatial configuration of collaboration among inventors evolves. Triadic closure promotes the formation of local ties only to the extent that prior ties are localized, but equally spurs the formation of non-local ties when inventors involved in a triad are geographically dispersed. Initially, triadic closure tends to reinforce the geographic distance effect by closing triads among proximate inventors. However, as the effect of geographic proximity decreases and the effect of triadic closure increases, the latter becomes an increasingly powerful vehicle to generate longer-distance collaboration ties. That is, initially triadic closure reinforces early collaboration decisions driven by a strong geographic distance effect, consistent with the central notion of path-dependence thinking that initial conditions often have long lasting effects (Boschma and Frenken,

2006). Yet, over time triadic closure transforms into a mechanism for endogenous change that at least to a certain extent breaks the path dependency in the spatial dynamics of collaboration patterns, consistent with the more recent stance taken by Martin (2010) that truly evolutionary models of path dependence should include endogenous drivers of change that run alongside typical path dependence processes.

Taken together, by unveiling basic principles that underpin the evolution of collaborative innovation networks, this study provides support for existing theoretical principles related to the collaborative nature of innovation, as described in the literature on the geography of innovation (Audretsch and Feldman, 1996b) and collaboration networks (Powell et al., 1996). More specifically, these results bring two broad contributions. First, this paper adds to the growing literature on network dynamics through its distinctive geographic angle. The way geography impacts on the evolution of networks is shown to be inconstant over time. Interestingly, even when the direct effect of geographic proximity starts fading, geography leaves a lasting imprint on spatial patterns of collaboration since triadic closure – at least initially – may reinforce existing localized collaboration patterns. Second, the study extends proximity theories that lay at the foundation of our current understanding of spatial patterns of collaboration. It shows that triadic closure – as a form of social proximity – is an important principle guiding tie formation. The effect of other forms of proximity – in this case geographic proximity – is largely contingent on such effects, as the legacy of the prior collaboration structure may alter how other proximity effects may become apparent. Spatial patterns of collaboration cannot be exclusively explained from pure tendencies to collaborate with proximate or distant partners, but are the outcome of path-dependent processes in which geographic factors and past structures in the network interplay.

## ***6.2 Limitations and future research***

The study is subject to various limitations, which also open perspectives for future research. First, the shifts observed in the forces that drive network evolution have been attributed to changes in the technology regime of the biotechnology industry. Yet, it is hard to make any causal claims; the fact that changes in the technology regime and in the network dynamics of an industry run in parallel does not

necessarily mean that one is causing the other. Future research could extend existing work about the relation between knowledge characteristics and geographic proximity from a static perspective (Gittelman, 2007; Zander and Kogut, 1995) to dynamics studies that further scrutinize the relationship between knowledge and networks from a dynamic perspective. Content-based network analysis techniques could be an interesting way forward. Such network analysis does not only analyze the pattern of interaction among agents, but also treats the knowledge they have in common as a network of ideas or a network of knowledge itself (Criscuolo et al., 2007).

Second, the findings of the study are limited to the idiosyncratic nature of the biotechnology sector. This raises the question to what extent the proposed model of network dynamics can be extended to other emerging sectors. The way in which the interplay between geographic distance and triadic closure unfolds over time may well apply to other sectors, as the nature of knowledge changes during these sectors' lifecycles. Currently, it is still largely unknown what drives the dynamics of knowledge networks and how network dynamics differ across industries.

Third, this paper has only provided a 'stylized model' contrasting two main mechanisms of network dynamics, an important direction for future work would be to broaden the range of endogenous and attribute-related drivers of network dynamics. Sociology offers a much wider array of endogenous network effects than triadic closure alone that have potential theoretical relevance for the dynamics of inventor networks. In that regard, one could think of betweenness effects that express actors' preference to position themselves between unconnected others (Burt, 2004; Snijders et al., 2007). In terms of attribute-related effects, one could think of other forms of proximity (Boschma, 2005). For instance, the inclination of inventors to collaborate with cognitively similar or dissimilar peers could be captured as an attribute-related parameter, provided good data on individuals' competences and knowledge bases is available (see Balland, 2012 for a recent example).

Finally, despite the merits of stochastic estimation models of network evolution, such models have clear limitations. For instance, the size of the networks that can currently be analyzed in the program is limited, particularly when repeated network observations are analyzed at once. More importantly, a

considerable level of stability among subsequent network observations is required for the model to converge towards stable outcomes. Notwithstanding these limitations, this study has shown the value of stochastic modeling techniques in getting to grips with the forces and mechanisms that underlie the evolution of networks. Descriptive methods as well as more conventional regression methods have produced largely consistent findings, suggesting that stochastic estimation models of network evolution are well-suited to detect forces of network evolution empirically, creating vast potential for future research. An interesting direction for future research in this regard could be to investigate the mutual dependence of networks and proximities that cannot be modeled with more conventional statistical approaches. Not only does proximity positively affect tie formation, individuals may also become more proximate once they are connected (see also Baum et al., 2010; Cowan et al., 2006), and once they have become exceedingly proximate the need to further collaborate may disappear (Broekel and Boschma, 2012). Such mutual dependence can be disentangled by stochastic estimation models of network evolution that can separate forces of social selection (proximity affecting tie formation) from social influence (network ties affecting proximity), further exploiting the opportunities that such recent advances in dynamic network analysis have brought.

## References

- Abramovsky, L., and Simpson, H. (2011) Geographic proximity and firm–university innovation linkages: Evidence from great Britain. *Journal of Economic Geography*, 11:949-77.
- Acs, Z. J., Anselin, L., and Varga, A. (2002) Patents and innovation counts as measures of regional production of new knowledge. *Research Policy*, 31:1069-85.
- Ahuja, G. (2000) Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative Science Quarterly*, 45:425-55.
- Almeida, P., and Kogut, B. (1999) Localization of knowledge and the mobility of engineers in regional networks. *Management Science*, 45:905-17.
- Arthur, W. B. (1994) *Increasing returns and path dependence in the economy*. Ann Arbor: University of Michigan Press.
- Audretsch, D. B. (2001) The role of small firms in us biotechnology clusters. *Small Business Economics*, 17:3-15.
- Audretsch, D. B., and Feldman, M. P. (1996a) Innovative clusters and the industry life cycle. *Review of Industrial Organization*, 11:253-73.

- Audretsch, D. B., and Feldman, M. P. (1996b) R&d spillovers and the geography of innovation and production. *American Economic Review*, 86:630-40.
- Audretsch, D. B., and Stephan, P. E. (1996) Company-scientist locational links: The case of biotechnology. *The American Economic Review*, 86:641-52.
- Balland, P.-A. (2012) Proximity and the evolution of collaboration networks: Evidence from research and development projects within the global navigation satellite system (gnss) industry. *Regional Studies*, 46:741-56.
- Baptista, R., and Swann, P. (1998) Do firms in clusters innovate more? *Research Policy*, 27:525-40.
- Baum, J. A. C., Cowan, R., and Jonard, N. (2010) Network-independent partner selection and the evolution of innovation networks. *Management Science*, 56:2094-110.
- Beckman, C. M., Haunschild, P. R., and Phillips, D. J. (2004) Friends or strangers? Firm-specific uncertainty, market uncertainty and network partner selection. *Organization Science*, 15:259-75.
- Bell, G. G., and Zaheer, A. (2007) Geography, networks, and knowledge flow. *Organization Science*, 18:955-72.
- Blind, K., Edler, J., Frietsch, R., and Schmoch, U. (2006) Motives to patent: Empirical evidence from germany. *Research Policy*, 35:655-72.
- Borgatti, S. P., Everett, M. G., and Freeman, L. C. (2002) Ucinet for windows: Software for social network analysis: Harvard, MA: Analytic Technologies.
- Boschma, R., and Frenken, K. (2010) The spatial evolution of innovation networks. A proximity perspective. In *Handbook of evolutionary economic geography*, edited by R. Boschma and R. Martin. Cheltenham: Edward Elgar.
- Boschma, R. A. (2005) Proximity and innovation: A critical assessment. *Regional Studies*, 39:61-74.
- Boschma, R. A., and Frenken, K. (2006) Why is economic geography not an evolutionary science? Towards an evolutionary economic geography. *Journal of Economic Geography*, 6:273-302.
- 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. In *Handbook of quantitative science and technology research: The use of publication and patent statistics in studies on s&t systems*, edited by H. Moed, W. Glänzel and U. Schmoch. Berlin/Heidelberg/New York: Springer.
- Breschi, S., and Lissoni, F. (2009) Mobility of skilled workers and co-invention networks: An anatomy of localized knowledge flows. *Journal of Economic Geography*, 9:439-68.
- Broekel, T., and Boschma, R. (2012) Knowledge networks in the dutch aviation industry: The proximity paradox. *Journal of Economic Geography*, 12:409-33.
- Burt, R. S. (2004) Structural holes and good ideas. *American Journal of Sociology*, 110:349-99.
- Buskens, V. (2002) *Social networks and trust*. Boston/Dordrecht/London: Kluwer Academic Publishers.
- Coleman, J. S. (1988) Social capital in the creation of human capital. *American Journal of Sociology*, 94:95-120.
- Cowan, R., Jonard, N., and Özman, M. (2004) Knowledge dynamics in a network industry. *Technological Forecasting and Social Change*, 71:469-84.

- Cowan, R., Jonard, N., and Zimmermann, J. B. (2006) Evolving networks of inventors. *Journal of Evolutionary Economics*, 16:155-74.
- Criscuolo, P., Salter, A., and Sheehan, T. (2007) Making knowledge visible: Using expert yellow pages to map capabilities in professional services firms. *Research Policy*, 36:1603-19.
- Dahl, M. S., and Pedersen, C. Ø. R. (2004) Knowledge flows through informal contacts in industrial clusters: Myth or reality? *Research Policy*, 33:1673-86.
- Davis, J. A. (1970) Clustering and hierarchy in interpersonal relations: Testing two theoretical models on 742 sociograms. *American Sociological Review*, 35:843-52.
- Ejermo, O., and Karlsson, C. (2006) Interregional inventor networks as studied by patent coinventorships. *Research Policy*, 35:412-30.
- Faulconbridge, J. R. (2006) Stretching tacit knowledge beyond a local fix? Global spaces of learning in advertising professional service firms. *Journal of Economic Geography*, 6:517-40.
- Fleming, L., King III, C., and Juda, A. I. (2007) Small worlds and regional innovation. *Organization Science*, 18:938-54.
- Gambardella, A. (1995) *Science and innovation in the us pharmaceutical industry*. Cambridge: Cambridge University Press.
- García-Muiña, F. E., Pelechano-Barahona, E., and Navas-López, J. E. (2009) Knowledge codification and technological innovation success: Empirical evidence from spanish biotech companies. *Technological Forecasting and Social Change*, 76:141-53.
- Gargiulo, M., and Benassi, M. (2000) Trapped in your own net? Network cohesion structural holes, and the adaptation of social capital. *Organization Science*, 11:183-96.
- Gertler, M. S. (2003) Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there). *Journal of Economic Geography*, 3:75-99.
- Gilsing, V., and Nooteboom, B. (2006) Exploration and exploitation in innovation systems: The case of pharmaceutical biotechnology. *Research Policy*, 35:1-23.
- Gittelman, M. (2007) Does geography matter for science-based firms? Epistemic communities and the geography of research and patenting in biotechnology. *Organization Science*, 18:724-41.
- Giuliani, E. (2010) Network dynamics in regional clusters: The perspective of an emerging economy. *Papers in Evolutionary Economic Geography (PEEG)*.
- Giuri, P., Mariani, M., Brusoni, S., Crespi, G., Francoz, D., Gambardella, A., Garcia-Fontes, W., Geuna, A., Gonzales, R., Harhoff, D., Hoisl, K., Le Bas, C., Luzzi, A., Magazzini, L., Nesta, L., Nomaler, Ö., Palomeras, N., Patel, P., Romanelli, M., and Verspagen, B. (2007) Inventors and invention processes in europe: Results from the patval-eu survey. *Research Policy*, 36:1107-27.
- Gordon, I. R., and McCann, P. (2005) Innovation, agglomeration, and regional development. *Journal of Economic Geography*, 5:523-43.
- Grabher, G., and Ibert, O. (2006) Bad company? The ambiguity of personal knowledge networks. *Journal of Economic Geography*, 6:251-71.

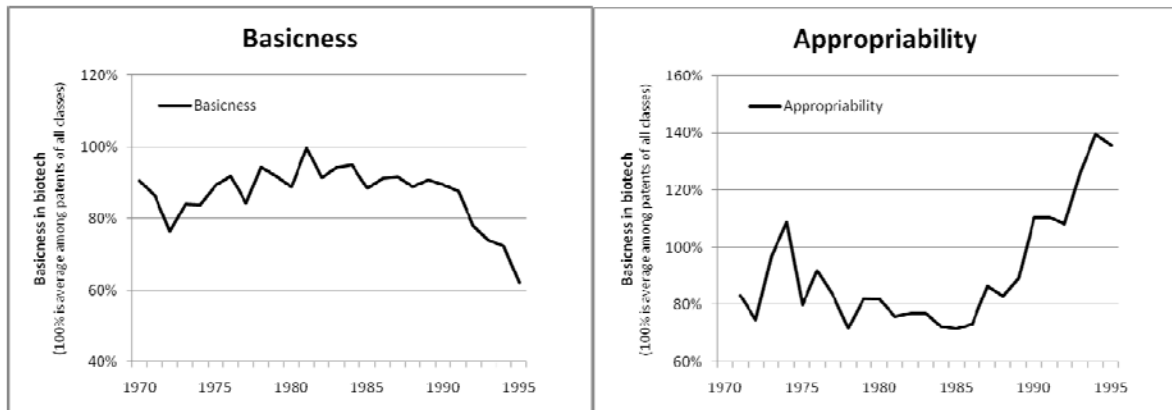


- Gulati, R. (1995) Social structure and alliance formation patterns: A longitudinal analysis. *Administrative Science Quarterly*, 40:619-52.
- Gulati, R., and Gargiulo, M. (1999) Where do interorganizational networks come from? *American Journal of Sociology*, 104:1439-93.
- Hall, B. H., Jaffe, A. B., and Trajtenberg, M. (2001) The nber patent citations data file: Lessons, insights and methodological tools: NBER Working Paper 8498.
- Hoetker, G. (2007) The use of logit and probit models in strategic management research: Critical issues. *Strategic Management Journal*, 28:331-43.
- Hopkins, M., Martin, P., Nightingale, P., Kraft, A., and Mahdi, S. (2007) The myth of the biotech revolution: An assessment of technological, clinical and organisational change. *Research Policy*, 36:566-89.
- Huber, F. (2012) Do clusters really matter for innovation practices in information technology? Questioning the significance of technological knowledge spillovers. *Journal of Economic Geography*, 12:107-26.
- Ibrahim, S. E., Fallah, M. H., and Reilly, R. R. (2009) Localized sources of knowledge and the effect of knowledge spillovers: An empirical study of inventors in the telecommunications industry. *Journal of Economic Geography*, 9:405-31.
- Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993) Geographic localisation of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*, 108:577-98.
- Kilduff, M., and Tsai, W. (2003) *Social networks and organizations*. London: Sage Publications.
- Krauss, G., and Stahlecker, T. (2001) New biotechnology firms in germany: Heidelberg and the bioregion rhine-neckar triangle. *Small Business Economics*, 17:143-53.
- Lehrer, M. (2005) Science-driven vs. Market-pioneering high tech: Comparative german technology sectors in the late nineteenth and late twentieth centuries. *Industrial and Corporate Change*, 14:251-78.
- Li, P. F., Bathelt, H., and Wang, J. (2012) Network dynamics and cluster evolution: Changing trajectories of the aluminium extrusion industry in dali, china. *Journal of Economic Geography*, 12:127-55.
- Liebeskind, J. P., Oliver, A. L., Zucker, L., and Brewer, M. (1996) Social networks, learning, and flexibility: Sourcing scientific knowledge in new biotechnology firms. *Organization Science*, 7:428-43.
- Maggioni, M. A., Nosvelli, M., and Uberti, T. E. (2007) Space versus networks in the geography of innovation: A european analysis. *Papers in Regional Science*, 86:471-93.
- Malerba, F., and Orsenigo, L. (1997) Technological regimes and sectoral patterns of innovative activities. *Industrial and Corporate Change*, 6:83-116.
- Maraut, S., Dernis, H., Webb, C., Spiezia, V., and Guellec, D. (2008) The oecd regpat database: A presentation. *OECD Science, Technology and Industry Working Papers*.
- March, J. G. (1991) Exploration and exploitation in organizational learning. *Organization Science*, 2:71-87.
- Martin, R. (2010) Roepke lecture in economic geography—rethinking regional path dependence: Beyond lock-in to evolution. *Economic Geography*, 86:1-27.
- McEvily, B., Perrone, V., and Zaheer, A. (2003) Trust as an organizing principle. *Organization Science*, 14:91-103.
- Nesta, L., and Saviotti, P. P. (2005) Coherence of the knowledge base and the firm's innovative performance: Evidence from the us pharmaceutical industry. *Journal of Industrial Economics*, 53:123-42.

- Nesta, L., and Saviotti, P. P. (2006) Firm knowledge and market value in biotechnology. *Industrial and Corporate Change*, 15:625-52.
- Nicholas, T. (2009) Spatial diversity in invention: Evidence from the early r&d labs. *Journal of Economic Geography*, 9:1-31.
- Obstfeld, D. (2005) Social networks, the tertius iungens and orientation involvement in innovation. *Administrative Science Quarterly*, 50:100-30.
- Orsenigo, L., Pammolli, F., and Riccaboni, M. (2001) Technological change and network dynamics - lessons from the pharmaceutical industry. *Research Policy*, 30:485-508.
- Owen-Smith, J., and Powell, W. W. (2004) Knowledge networks as channels and conduits: The effects of spillovers in the boston biotechnology community. *Organization Science*, 15:5-21.
- Pavitt, K. (1984) Sectoral patterns of technological change: Towards a taxonomy and a theory. *Research Policy*, 13:343-73.
- Powell, W. W., Koput, K. W., and Smith-Doerr, L. (1996) Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative Science Quarterly*, 41:116-45.
- Powell, W. W., White, D. R., Koput, K. W., and Owen-Smith, J. (2005) Network dynamics and field evolution: The growth of inter-organizational collaboration in the life sciences. *American Journal of Sociology*, 110:1132-205.
- Reagans, R. (2011) Close encounters: Analyzing how social similarity and propinquity contribute to strong network connections. *Organization Science*, 22:835-49.
- Reagans, R., and McEvily, B. (2003) Network structure and knowledge transfer: The effects of cohesion and range. *Administrative Science Quarterly*, 48:240-67.
- Rothaermel, F. T., and Thursby, M. (2007) The nanotech versus the biotech revolution: Sources of productivity in incumbent firm research. *Research Policy*, 36:832-49.
- Saviotti, P. P. (1998) On the dynamics of appropriability, of tacit and of codified knowledge. *Research Policy*, 26:843-56.
- Schmoch, U., Laville, F., Patel, P., and Frietsch, R. (2003) Linking technology areas to industrial sectors: Final report to the ec commission. Karlsruhe/Paris/Brighton: DG Research.
- Shan, W., Walker, G., and Kogut, B. (1994) Interfirm cooperation and startup innovation in the biotechnology industry. *Strategic Management Journal*, 15:387-94.
- Singh, J. (2005) Collaborative networks as determinants of knowledge diffusion patterns. *Management Science*, 51:756-70.
- Skvoretz, J. (1991) Theoretical and methodological models of networks and relations. *Social Networks*, 13:275-300.
- Snijders, T. A. B. (2001) The statistical evaluation of social network dynamics. In *Sociological methodology*, edited by M. E. Sobel and M. P. Becker. Boston and London: Basil Blackwell.
- Snijders, T. A. B., Steglich, C. E. G., Schweinberger, M., and Huisman, M. (2007) Manual for siena version 3: University of Groningen / University of Oxford.
- Song, J., Almeida, P., and Wu, G. (2003) Learning-by-hiring: When is mobility more likely to facilitate interfirm knowledge transfer? *Management Science*, 49:351-65.

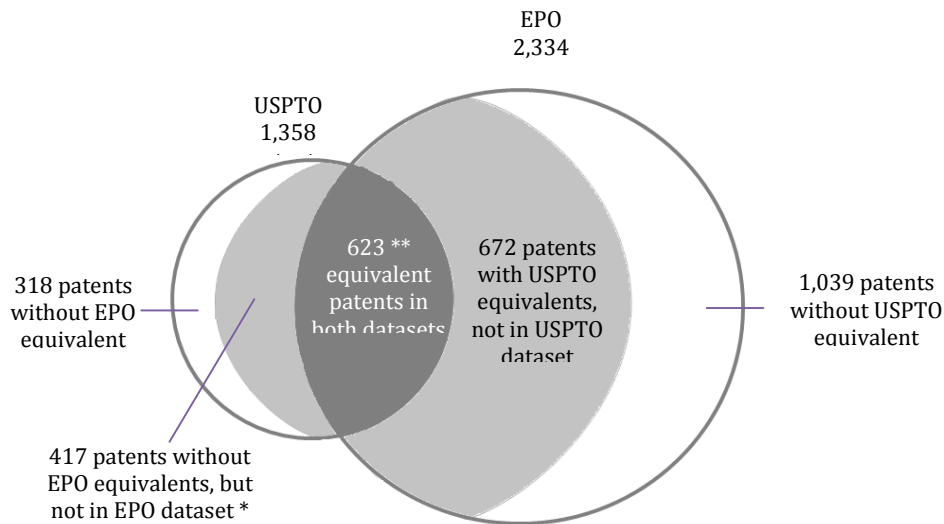
- Sorenson, O., Rivkin, J. W., and Fleming, L. (2006) Complexity, networks and knowledge flow. *Research Policy*, 35:994-1017.
- Sorenson, O., and Stuart, T. E. (2008) Bringing the context back in: Settings and the search for syndicate partners in venture capital investment networks. *Administrative Science Quarterly*, 53:266-94.
- Suire, R., and Vicente, J. (2009) Why do some places succeed when others decline? A social interaction model of cluster viability. *Journal of Economic Geography*, 9:381-404.
- Tamada, S., Naito, Y., Kodama, F., Gemba, K., and Suzuki, J. (2006) Significant difference of dependence upon scientific knowledge among different technologies. *Scientometrics*, 68:289-302.
- Teece, D. J. (1986) Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy*, 15:285-305.
- Ter Wal, A. L. J. (2011) Cluster emergence and network evolution: A longitudinal analysis of the inventor network in sophia-antipolis. *Regional Studies*, in press.
- Ter Wal, A. L. J., and Boschma, R. A. (2011) Co-evolution of firms, industries and networks in space. *Regional Studies*, 45.
- Torre, A. (2008) On the role played by temporary geographical proximity in knowledge transmission. *Regional Studies*, 42:869-89.
- Trajtenberg, M., Henderson, R., and Jaffe, A. (1997) University versus corporate patents: A window on the basicness of invention. *Economics of Innovation and New Technology*, 5:19 - 50.
- Uzzi, B. (1997) Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly*, 42:35-67.
- Walker, G., Kogut, B., and Shan, W. J. (1997) Social capital, structural holes and the formation of an industry network. *Organization Science*, 8:109-25.
- Zaheer, A., and Bell, G. G. (2005) Benefiting from network position: Firm capabilities, structural holes, and performance. *Strategic Management Journal*, 26:809-25.
- Zander, U., and Kogut, B. (1995) Knowledge and the speed of the transfer and imitation of organizational capabilities: An empirical test. *Organization Science*, 6:76-92.
- Zeller, C. (2001) Clustering biotech: A recipe for success? Spatial patterns of growth of biotechnology in munich, rhineland and hamburg. *Small Business Economics*, 17:123-41.
- Zelner, B. A. (2009) Using simulation to interpret results from logit, probit, and other nonlinear models. *Strategic Management Journal*, 30:1335-48.
- Zucker, L. G., Darby, M. R., and Brewer, M. B. (1998) Intellectual human capital and the birth of us biotechnology enterprises. *American Economic Review*, 88:290-306.

**Figure 1: Evolution of level of ‘basicness’ and appropriability in biotechnology (USPTO data)**



These figures depict the average measures of ‘generality’ and ‘appropriability’ as proposed by Trajtenberg et al. (1997) for worldwide USPTO patents in biotechnology. The measure is indexed (where 100% is set to the average among all patents in all classes in a certain year) to offset non-sector-specific trends in patenting and to circumvent problems related to yearly variations in number of citations and the time lag between a patent and its citations.

**Figure 2: Overlapping patents in USPTO and EPO datasets**



Patent priority numbers allow identifying the USPTO patents that have an equivalent patent at EPO and vice versa. These figures refer to the period that both datasets overlap (i.e. 1978-1995).

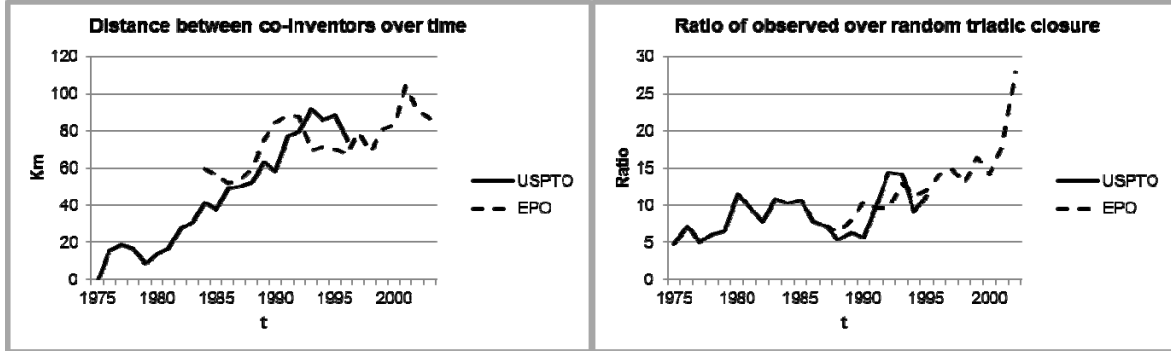
The classification system in which patents are indicated as biotechnology is different for EPO and USPTO patents. As a result, there are USPTO patents which have an equivalent at EPO but do not occur in the EPO dataset and vice versa.

\* 219 out of 417 USPTO patents with an EPO-equivalent that is not in the EPO dataset have IPC-classes G01N and C07K. However, including these in the EPO dataset nearly doubles the total set of EPO patents to 4,583, the majority of which do not correspond with patents in the USPTO dataset.

\*\* Important: 41.6% of equivalent patents are applied for later at USPTO than at EPO, whereas virtually no patents were applied for earlier at EPO than at USPTO. For those equivalent patents for which the application years was different, USPTO patents were on average applied for 2.6 years after the EPO patent.



Figure 5: Descriptive statistics of realized co-invention ties over time (USPTO and EPO data)



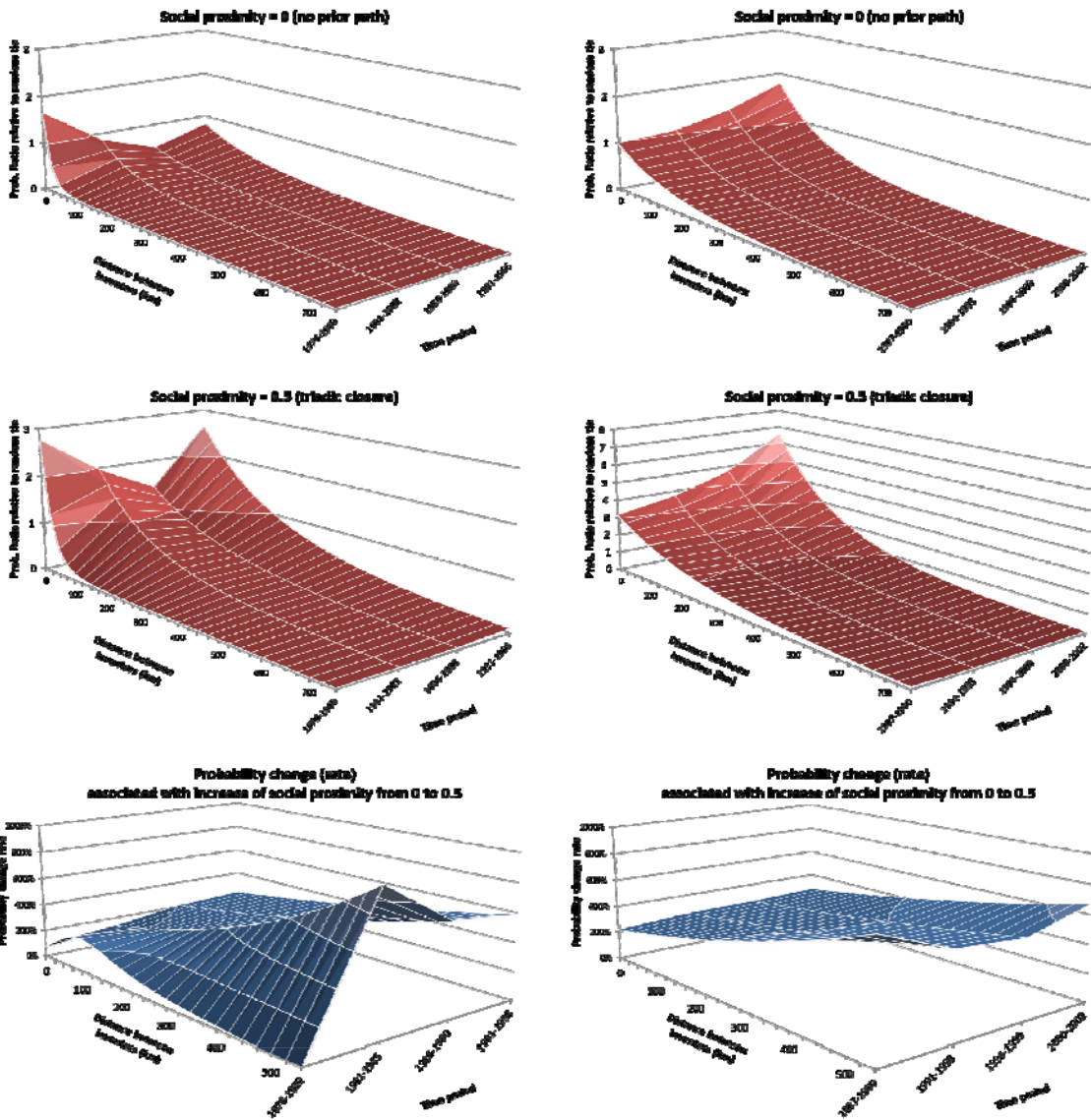
Figures are based on co-invention ties in networks with a five-year moving window. For example, on the horizontal axis “1975” incorporates ties from 1971 up to and including 1975. Random expected triadic closure is calculated on the number of two-paths (potential closed triads) at time  $t$  that would be closed at time  $t+5$  if the new ties formed in between these periods were random. Observed triadic closure is the number of these closing triads that were indeed observed.

Table 1: Regression model explaining tie formation between inventors  
Higher-order conditional logistic regression with random intercepts at dyad-level

	Entire period	1976-1980	1981-1985	1986-1990	1991-1995		
<b>USPTO</b>							
Geographic distance	-0.013 * (0.001)	-0.063 * (0.010)	-0.026 * (0.003)	-0.010 * (0.001)	-0.010 * (0.001)		
Social proximity	4.961 * (0.124)	3.765 * (0.494)	6.518 * (0.412)	7.477 * (0.320)	9.491 * (0.307)		
Geographic distance * Social proximity	0.002 † (0.001)	0.047 * (0.011)	0.040 * (0.005)	0.010 * (0.002)	-0.001 (0.001)		
Constant	-4.223 * (90.98)	-3.858 * (0.408)	-8.297 * (0.352)	-9.114 * (0.308)	-10.345 * (0.290)		
Observations (dyad-level)	238092	7808	31126	57266	141340		
Groups (unique dyads)	101928	3886	14627	29544	74640		
Log-likelihood	-9692.83	-689.61	-1731.20	-2422.25	-4017.88		
Wald $\chi^2$	4135.04	183.42	547.16	937.49	1448.57		
<b>EPO</b>							
	Entire period			1987-1990	1991-1995	1996-1999	2000-2002
Geographic distance	-0.011 * (0.000)			-0.012 * (0.001)	-0.009 * (0.000)	-0.012 * (0.001)	-0.011 * (0.001)
Social proximity	5.243 * (0.094)			9.125 * (0.361)	7.952 * (0.173)	9.995 * (0.241)	12.505 * (0.298)
Geographic distance * Social proximity	-0.000 (0.001)			0.014 * (0.002)	-0.003 * (0.001)	0.002 (0.001)	0.006 * (0.001)
Constant	-9.655 * (0.101)			-10.865 * (0.386)	-9.790 * (0.163)	-12.643 * (0.232)	-14.37 * (0.290)
Observations (dyad-level)	2044711			113835	458949	615347	856580
Groups (unique dyads)	874529			63570	230930	315533	526290
Log-likelihood	-32595.37			-2877.08	-7758.82	-6560.94	-8840.62
Wald $\chi^2$	8681.72			889.52	3644.20	2588.04	2187.96

\* Parameter is significant at 0.01 level - † Parameter significant at 0.05 level

Figure 6: Graphing the interaction between geographic distance and social proximity  
 USPTO 1976-1995 EPO 1987-2002



Estimations are based on standard logistic regression models (Appendix A), as the simulation procedure through which the probabilities of tie formation are estimated does currently not support higher-order conditional logit models (Table 1).

The lower two graphs show the change in probability, expressed as a percentage:

$$\frac{p(\text{tie formation})_{\text{socialprox}=0.5} - p(\text{tie formation})_{\text{socialprox}=0}}{p(\text{tie formation})_{\text{socialprox}=0}}$$

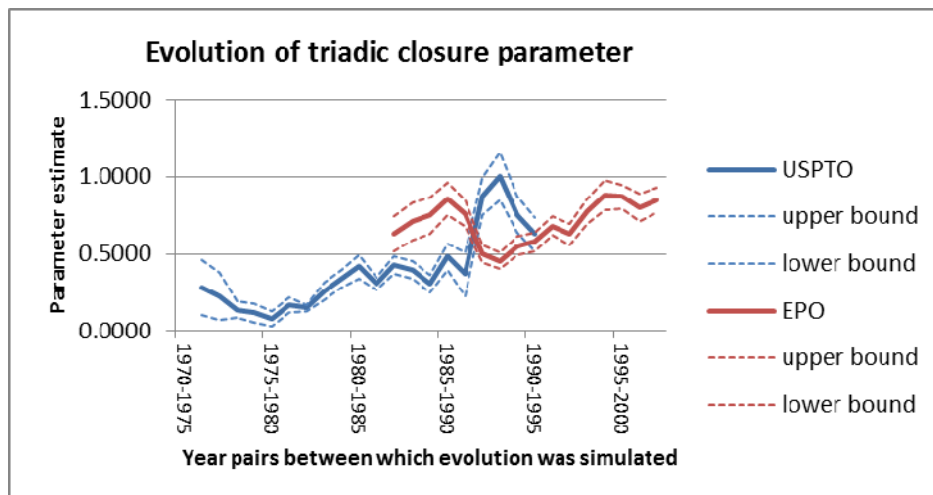
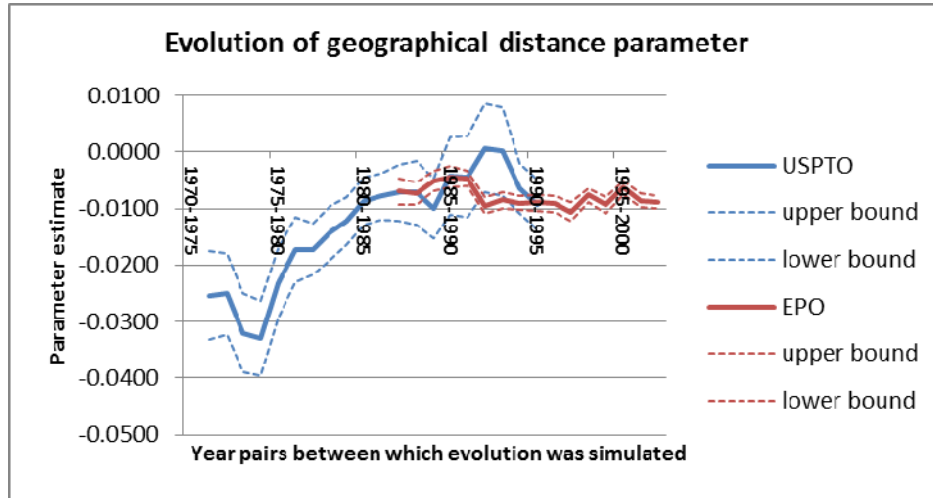
**Table 2: Determinants of network evolution: a stochastic estimation model in SIENA**

<b>USPTO</b>	<b>1975 →1980</b>	<b>1980 →1985</b>	<b>1985 →1990</b>	<b>1990 →1995</b>	<b>1995 → 2000</b>
<b>Network change</b>					
Number of nodes	49	96	122	214	
Links created	17	64	45	87	
Links dissolved	47	76	77	109	
Links retained	50	78	111	178	
Links $t - t+1$	97 → 67	154 → 142	188 → 156	287 → 265	
<b>Parameter estimates</b>					
Rate of change	1.4676 * (0.1866)	1.5069 * (0.1174)	1.0932 * (0.0937)	0.9571 * (0.0629)	
Degree	-2.2926 * (0.0992)	-1.8339 * (0.0638)	-1.7881 * (0.0718)	-2.0348 * (0.0512)	
Geographic distance	-0.0233 * (0.0031)	-0.0087 * (0.0021)	-0.0042 (0.0034)	-0.0094 * (0.0022)	
Triadic closure	0.0785 * (0.0236)	0.4152 * (0.0376)	0.4812 * (0.0423)	0.6259 * (0.0546)	
<b>Model</b>					
Number of iterations	1651	1647	1823	1660	
Convergence t	0.042	0.037	0.060	0.021	
<b>EPO</b>	<b>1975 →1980</b>	<b>1980 →1985</b>	<b>1985 →1990</b>	<b>1990 →1995</b>	<b>1995 → 2000</b>
<b>Network change</b>					
Number of nodes			294	500	670
Links created			194	317	305
Links dissolved			169	438	488
Links retained			218	345	351
Links $t - t+1$			387 → 412	783 → 662	839 → 656
<b>Parameter estimates</b>					
Rate of change			11.4775 * (1.2160)	16.5200 * (1.2226)	16.2754 * (1.2401)
Degree			-1.9275 * (0.0418)	-2.2403 * (0.0392)	-2.3896 * (0.0317)
Geographic distance			-0.0044 * (0.0009)	-0.0090 * (0.0007)	-0.0063 * (0.0006)
Triadic closure			0.8573 * (0.0529)	0.5788 * (0.0297)	0.8687 * (0.0372)
<b>Model</b>					
Number of iterations			1665	2079	1767
Convergence t			0.084	0.068	0.044

\*\*\* Parameter is significant at 0.01 level



**Figure 7: Parameters for geographic distance and triadic closure over time in stochastic estimation model in SIENA**



**Appendix A: Logistic regression explaining tie formation between inventors**

<b>USPTO</b>	1976-1980	1981-1985	1986-1990	1991-1995
Geographic distance	-0.049 (2.95)*	-0.009 (12.62)*	-0.007 (11.88)*	-0.006 (13.90)*
Social proximity	2.543 (8.47)*	3.545 (30.39)*	4.528 (35.12)*	4.854 (51.29)*
Geographic distance * Social proximity	0.043 (2.58)*	0.011 (8.94)*	0.006 (7.84)*	0.005 (8.19)*
Constant	-1.982 (7.42)**	-3.382 (37.78)*	-4.640 (45.92)*	-4.515 (63.94)*
Observations (dyads)	7808	31126	57266	141340
Log-likelihood	-872.37	-2656.93	-3187.74	-5512.48
McFadden's R <sup>2</sup>	0.586	0.484	0.460	0.462

<b>EPO</b>	1987-1990	1991-1995	1996-1999	2000-2002
Geographic distance	-0.006 (12.02)*	-0.007 (18.09)*	-0.007 (20.98)*	-0.007 (20.02)*
Social proximity	4.701 (43.12)*	5.080 (69.37)*	5.123 (70.40)*	5.297 (69.53)*
Geographic distance * Social proximity	0.007 (9.60)*	0.005 (8.40)*	0.004 (7.63)*	0.005 (10.10)*
Constant	-4.500 (55.52)**	-4.977 (88.77)*	-5.261 (98.45)*	-5.441 (94.92)*
Observations (dyads)	113835	458949	615347	856580
Log-likelihood	-3935.31	-10153.24	-9423.78	-8840.62
McFadden's R <sup>2</sup>	0.437	0.434	0.403	0.406

\* Parameter is significant at 0.01 level