Intermetropolitan Networks of Co-invention in American Biotechnology

by

Der-Shiuan Lee

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved August 2011 by the
Graduate Supervisory Committee:

Breandán Ó hUallacháin, Chair
Luc Anselin
Michael Kuby
José Lobo

ARIZONA STATE UNIVERSITY

December 2011

ABSTRACT

Regional differences of inventive activity and economic growth are important in economic geography. These differences are generally explained by the theory of localized knowledge spillovers, which argues that geographical proximity among economic actors fosters invention and innovation. However, knowledge production involves an increasing number of actors connecting to non-local partners. The space of knowledge flows is not tightly bounded in a given territory, but functions as a network-based system where knowledge flows circulate around alignments of actors in different and distant places. The purpose of this dissertation is to understand the dynamics of network aspects of knowledge flows in American biotechnology. The first research task assesses both spatial and network-based dependencies of biotechnology co-invention across 150 large U.S. metropolitan areas over four decades (1979, 1989, 1999, and 2009). An integrated methodology including both spatial and social network analyses are explicitly applied and compared. Results show that the network-based proximity better defines the U.S. biotechnology co-invention urban system in recent years. Co-patenting relationships of major biotechnology centers has demonstrated national and regional association since the 1990s. Associations retain features of spatial proximity especially in some Midwestern and Northeastern cities, but these are no longer the strongest features affecting co-inventive links. The second research task examines how biotechnology knowledge flows circulate over space by focusing on the structural properties of intermetropolitan co-invention networks. All analyses in this task are conducted using social network analysis.

Evidence shows that the architecture of the U.S. co-invention networks reveals a trend toward more organized structures and less fragmentation over the four years of analysis. Metropolitan areas are increasingly interconnected into a large web of networked environment. Knowledge flows are less likely to be controlled by a small number of intermediaries. San Francisco, New York, Boston, and San Diego monopolize the central positions of the intermetropolitan co-invention network as major American biotechnology concentrations. The overall network-based system comes close to a relational core/periphery structure where core metropolitan areas are strongly connected to one another and to some peripheral areas. Peripheral metropolitan areas are loosely connected or even disconnected with each other. This dissertation provides empirical evidence to support the argument that technological collaboration reveals a network-based system associated with different or even distant geographical places, which is somewhat different from the conventional theory of localized knowledge spillovers that once dominated understanding of the role of geography in technological advance.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure | Page

Figure Page

Chapter 1

INTRODUCTION

1.1 Research Problem Statement

In the last few decades, regional differences of inventive activity and economic growth have become important issues in economic geography. These differences are generally explained by the theory of localized knowledge spillovers (herafter LKSs). Geography influences the nature and strength of information flows, as the diffusion of knowledge is often more local than global. A central claim for knowledge exchange and collaboration tending to be localized is that face-to-face contacts are essential for effective knowledge transfer (Howells 2002). Geographical proximity among economic actors (e.g., inventors, firms, or research institutions) fosters invention and innovation. Empirical evidence for the presence of LKSs is widespread in cities and regions across the U.S. and Europe (Jaffe 1989; Acs et al. 1992; 1994; Audretsch and Feldman 1996; Anselin et al. 1997; 2000; Almeida and Kogut 1999; Varga 2000). Geographical concentration of inventive firms in clusters enhances the possibility of interaction and lowers costs through trade in goods and services, labor mobility, research collaboration, and interpersonal communication (Bathelt 2005; Ponds et al. 2007).

While studies of LKSs and geographical proximity are prominent in the geography of invention and innovation, other scholars stress the role of collaborative networks in which individuals and groups are embedded in webs of social relationships through direct connections and indirect linkages (Rondé and

Hussler 2005; Maggioni et al. 2007; Knoben 2009; Wilhelmsson 2009). In collaborative networks, people exchange information, develop vicariously perceptions and opinions, and reduce uncertainty about events, ideas, or phenomena in pursuit of particular goals (Rice and Aydin 1991; Amin and Cohendet 2005). Breschi and Lissoni (2003) argued that collaborative networks are channels for knowledge flows that are not limited to local boundaries but can span long distances (also Maggioni et al. 2007; Ponds et al. 2007). A high degree of network-based proximity that links scientists and engineers from different cities and regions extends the scope and geography of cooperation. This is especially apparent in high-technology industries such as biotechnology where research collaboration through global networks has become crucial for inventive performance (Coe and Bunnell 2003; McKelvey et al. 2003; Gertler and Levitte 2005; Birch 2007; Cooke 2007; Ponds et al. 2007). Combining both local and complementary non-local skills and competencies are considered a major strategy for "firms evolving in a dynamic environment that requires rapid adaptation" (Rondé and Hussler 2005, p. 1151) and can reduce the possibility of negative technological lock-in (Asheim and Isaksen 2002; Boschma 2005; Gertler and Levitte 2005).

Collaborative networks enhance the geographical complementarities of complex relationships (Storper 1997; Amin 2002; Boggs and Rantisi 2003; Bathelt et al. 2004; Yeung 2005; Sunley 2008). Studies of the geography of invention and innovation are shifting from focusing on closed territorial relationships, towards an emphasis on markets and technological collaboration that increasingly occur between distant clusters (Coe and Bunnell 2003; Bathelt 2005, 2007; Yeung 2005; Vallance 2007; Kroll and Mallig 2009). Phrases such as “local sticky and global ubiquitous” (Asheim and Isaksen 2002), “local buzz and global pipelines” (Bathelt et al. 2004; Gertler and Levitte 2005), “local circuit and global circuit” (Malmberg 2003), and “local nodes in global networks” (Coenen et al. 2004; Gertler and Levitte 2005) stress knowledge exchange within and between economic actors at varying geographical scales. The space of knowledge flows is not tightly bounded in a territory, but regarded as a network-based system where knowledge flows circulate around alignments of economic actors in different places (Amin 2002; Amin and Cohendet 2005). Network-based systems of knowledge flows provide an alternative way to conceptualize the geography of cooperation (McCann and Simonen 2005). Cities in such a spatially stretched economic sphere are immersed in global networks where knowledge collaboration and exchange are decisive forces for technological advance (Maskell et al. 2006; Sunley 2008; Autant-Bernard et al. 2010).

Knowledge flows occur in various forms and they have diverse geographical characteristics. The existence of collaborative networks raises two critical challenges for the investigation and understanding of the geography of information exchange. Is the increasingly network nature of technological collaboration likely to modify geographical structures of knowledge flows? Are knowledge flows via networks becoming less constrained by geography? The purpose of this dissertation is to understand the dynamics of network aspects of knowledge flows in American biotechnology. I investigate knowledge exchange by concentrating on inter-territorial (e.g., intermetropolitan, intercity, or interregional) and not intra-territorial knowledge collaboration and exchange. This particularly occurs in biotechnology co-invention because it involves "over the distance" interactions between inventors from different locations (Breschi and Lissoni 2004; Maggioni et al. 2007), and has a high dependence on global networking relationships (Feldman 2001; Cortright and Mayer 2002; Owen-Smith and Powell 2004; Coenen et al. 2004; Gertler and Levitte 2005; Fontes 2005; Coenen et al. 2006; Cooke 2006). In short, biotechnology co-invention is an ideal case for investigating collaborative networks, the longitudinal dynamics of network-based systems, and associated local and global interactions.

1.2 Research Topics

The main argument of the dissertation is that the space of knowledge flows in biotechnology co-invention is not tightly bounded within territories and neighboring areas, but circulates around alignments of economic actors in different or even distant locations.

Two pivotal research topics are investigated in the dissertation:

1. Does biotechnology co-invention reveal significant differences in its spatial compared with its network-based dependency across the U.S. urban system?
2. How and to what extent do biotechnology flows circulate in network-based systems?

Spatial and network-based dependencies

While spatial proximity has a significant influence on the effects of localized knowledge flows, network-based proximity facilitates distinctive patterns of knowledge circulation between distant actors. The first research task is to assess the relative importance of spatial versus network-based proximity on the biotechnology co-invention urban system. An important related goal is to identify certain longitudinal dynamics in the spatial and network structures of intermetropolitan knowledge flows. I use an integrated methodology in which both exploratory spatial data analysis (ESDA) and social network analysis techniques are explicitly applied and compared. Detailed discussion of the methods is provided in the methodology chapter (see Chapter 4.1).

Properties of intermetropolitan co-invention networks

The second research task is to understand how and to what extent biotechnology flows circulate in network-based systems by focusing on the structural properties of intermetropolitan co-invention networks from three distinct perspectives: components and cohesive subgroups of metropolitan areas, intermetropolitan network centralization and centrality, and positions established within the co-invention network-based system. Intermetropolitan co-invention networks are constructed by tracking inventors who participate in biotechnology co-patenting and attributing each co-patent to metropolitan areas where the inventors reside.

1. Components and cohesive subgroups of metropolitan areas

The characteristics of network components and their cohesive subgroups reveal patterns of connection among metropolitan areas. Specific research questions tackled in this perspective include: Is the network-based space structured into groups of metropolitan areas centered on several U.S.

biotechnology centers? Do the member metropolitan areas within each group demonstrate certain types of spatial associations? Are these metropolitan areas intensely connected? I answer these questions by identifying components and cohesive subgroups of metropolitan areas within the co-invention network. A *component* is defined as a maximal connected group of nodes in a network. Metropolitan areas within the same component are assumed to underpin and facilitate knowledge flows through direct connections or indirect linkages. A *cohesive subgroup* is defined as a set of nodes with relatively strong or frequent ties in a network (Wasserman and Faust 1994).

By applying the concept of nested components, two procedures are conducted to identify cohesive subgroups of metropolitan areas underlying each component. One focuses on the number of neighbors of each metropolitan area titled the *k*-cores procedure. The other focuses on the frequency of interaction between each pair of metropolitan areas and is titled the *m*-slices procedure. The analysis of nested components in an intermetropolitan co-invention network, by either the number of neighbors or the frequency of interaction, reveals whether some metropolitan areas collaborate with one another intensely. Detailed discussion of the methods is provided in the methodology chapter (see Chapter 4.2).

2. Intermetropolitan network centralization and centrality

The center of a network identifies metropolitan areas with the best access to knowledge flows. Specific research questions addressed include: How tightly organized is the network around its most central metropolitan area(s)? How important is a metropolitan area in transferring knowledge to other areas? To what extent does a metropolitan area control or mediate knowledge flows in the network? Two levels of network center measure – centralization and centrality – are calculated to answer these questions. *Centralization* is a global-level measure used to assess the extent to which a whole network has a centralized structure (Scott 2000). *Centrality* is a local-level measure that reveals the visibility of an individual node to other nodes.

Both centralization and centrality are calculated and interpreted by three different perspectives: degree, closeness, and betweenness. *Degree*-based measures describe the extent to which a metropolitan area directly connects to other areas. *Closeness*-based measures assess how a metropolitan area accesses knowledge flows not only by directly connecting to its neighbors but also through chains of intermediaries to the entire network. *Betweenness*-based measures explore how a metropolitan area controls or mediates interaction between nonadjacent areas. Detailed discussion of the methods is provided in the methodology chapter (see Chapter 4.2).

3. Positions of metropolitan areas within the co-invention network-based system

Network positions show the co-invention network-based system these metropolitan areas form and the roles played by different types of areas within the system of knowledge exchange. Ponds et al. (2007) argued that a city's network position has a significant effect on regional inventive activity. A *position* refers to a set of nodes having a similar pattern of relations to the rest of the network (Wasserman and Faust 1994). Specific questions addressed in this part include:

Do some U.S. metropolitan areas have similar network positions? Do the varying positions of metropolitan areas reveal a hierarchical cluster structure? What are the relationships among these positions established within the network-based system?

This dissertation uses regular equivalence as the criterion for partitioning individual areas into network positions. *Regular equivalence* identifies nodes that have similar patterns of ties to equivalent (rather than identical) others (Wasserman and Faust 1994). Metropolitan areas are regularly equivalent if they have the same pattern of ties with members of other positions that are also regularly equivalent. The method for identifying regularly equivalent positions uses two key procedures: (1) the UCINET social network analysis package to estimate degrees of regular equivalence for pairs of areas (Borgatti and Everett 1993; Wasserman and Faust 1994; Borgatti et al. 2002), and (2) hierarchical clustering to identify patterns of similarity and simplification in the system. Detailed discussion of the methods is provided in the methodology chapter (see Chapter 4.2).

Analysis of temporal stability and instability

Central questions include: Has knowledge transmission via network-based proximity become more or less influential compared with spatial proximity over time? Will network components and cohesive subgroups become more or less dominated by certain metropolitan areas over time? What are the temporal changes in network relationships between major biotechnology centers (e.g., New York, Boston, and San Francisco)? What are the changes in network positions of minor biotechnology centers? These research questions are investigated in1979, 1989, 1999, and 2009 to roughly coincide with important advances in information and communication technologies that most likely influence the co-invention urban system. Initiating the analysis in 1979 captures the start of the personal computer (PC) age, the rise of e-mail and PC networking occurred around 1989, and 1999 ushered in the use of search engines (e.g., Google, Yahoo) to obtain information on the Internet.

1.3 Research Purpose

The theory of localized knowledge spillovers (LKSs) emphasizes that geographical proximity among economic actors fosters invention and innovation. Cities are seen as a space of territorial embeddedness and local networking. However, the argument of network-based proximity recognizes that interactive learning and sharing are not simply locally bounded. Through collaborative activities between non-local partners, cities act as functional nodes immersed in wider networks where knowledge exchange are decisive forces for technological advance (Amin 2002; Cowan and Jonard 2004; Amin and Cohendet 2005; Maskell et al. 2006; Sunley 2008; Autant-Bernard et al. 2010). The geography of cooperation is not constrained by aspects of co-location but has a broad range of collaborations and interactions over space. This dissertation examines the role of network-based proximity in the biotechnology co-invention urban system.

The purpose of this dissertation is to understand the dynamics of network aspects of knowledge flows in American biotechnology. The first research task is to compare both spatial and network-based dependencies of biotechnology co-invention across U.S. metropolitan areas. Results of this task provide insights into the relative importance of spatial and network-based proximities in the space of knowledge flows. The second research task is to investigate the structural properties of intermetropolitan co-invention networks by focusing on the roles of metropolitan areas from the following three perspectives. The characteristics of network components reveal patterns of connection among metropolitan areas. The center of a network identifies metropolitan areas with the best access to knowledge flows. Network positions of metropolitan areas show the kind of co-invention network-based system these member areas form and the roles played by different types of areas within the system of knowledge exchange. The aims of conducting these empirical analyses in the second research task are to understand network properties of co-invention, the diverse positions of metropolitan areas in systems of knowledge exchange, and how these properties and positions change over time.

1.4 Organization of the Dissertation

This dissertation is divided into eight chapters. Following this introduction chapter, Chapter 2 reviews the literature on the role of proximity in knowledge exchange in spatial and network-based systems. Chapter 3 reviews the literature on social network analysis used to describe intermetropolitan co-invention networks. Chapter 4 outlines the methodology used in this investigation. Methods of spatial and network-based dependencies are used to identify intermetropolitan relationships. Social network analysis techniques detect structural properties of intermetropolitan co-invention networks. Chapter 5 describes the data and presents preliminary descriptive tabulations. Chapter 6 presents results showing differences in spatial and network-based dependencies of biotechnology co-invention across U.S. metropolitan areas. Chapter 7 discusses results showing properties of intermetropolitan co-invention networks. Chapter 8 concludes the dissertation by summarizing the findings, linking the results to the literature, and pointing to future research directions.

Chapter 2

KNOWLEDGE, PROXIMITY, AND NETWORK-BASED SPACE

This chapter reviews the literature on the role of proximity in knowledge exchange in both spatial and network-based systems in order to propose a conceptual framework for examining the space of knowledge flows. Scholars have long recognized that knowledge is a key to regional economic development. Nelson and Winter's (1982) book *An Evolutionary Theory of Economic Change* stimulated interest in how the tacit nature of knowledge shapes technological change. Maskell and Malmberg (1999) suggested that tacit knowledge such as intuition, know-how, and personal skills is a prime determinant of the geography of invention and innovation. Knowledge flows tend to be restricted in space when the process of interactive learning reinforces local inventive activity (Gertler 2003; Gertler and Levitte 2005). However, the implications of knowledge exchange on spatial proximity have been criticized by scholars from relational and global perspectives (Breschi and Lissoni 2001a; 2001b; Gallaud and Torre 2004; Autant-Bernard et al. 2007b). They recognized that interactive learning and sharing are able to spread globally via relational networks (e.g., Breschi and Lissoni 2003; Maggioni et al. 2007; Ponds et al. 2007). Knowledge is not simply locally bounded and substantial knowledge production and exchange occurs between geographically distant partners.

This chapter is organized as follows. Section 2.1 introduces the theory of localized knowledge spillovers (LKSs) with an emphasis on assumptions, empirical studies, and recent criticism. Section 2.2 discusses the importance of external knowledge resources for regional invention and innovation. Section 2.3 concludes the chapter by proposing a conceptual framework for summarizing biotechnology network-based systems of knowledge exchange.

2.1 Geographical Knowledge Flows

Gertler (2007) outlined three related elements to explain why inventive activity tends to occur more efficiently among nearby economic actors. First, knowledge transfer requires extensive communication and trust, which makes it spatially sticky. Second, more knowledge exchange occurs when economic actors share common social contexts, which are mostly locally defined. Third, the dynamic nature of invention process requires intense learning by doing and collective understanding. Geographical proximity makes this type of interaction easier and speeds up the flows of ideas. These three elements lead to the development of LKSs, which constitute knowledge flows bounded in space (Breschi and Lissoni 2001a).

This section is organized as follows. The first part outlines the theoretical framework of LKSs. The second part reviews empirical studies of LKSs and practical strategies for regional development. Keeping knowledge flows in limited space has been questioned by some scholars who argue that knowledge can also be transferred to long-distance partners via relational networks. The last part discusses recent criticism of LKSs to develop an alternative way of conceptualizing geographical structures of knowledge flows.

Theoretical framework of LKSs

The theory of LKSs is based upon two fundamental arguments. First, because of the tacit aspects of knowledge, spillovers occur more easily over short compared with long distances (McCann and Simonen 2005). Geographical proximity is the best way to benefit from knowledge externalities. Second, because of geographical proximity, firms are better able to identify and interact with potential partners (Rallet and Torre 1999), which enable "rich" local interactive learning and sharing. Breschi and Lissoni (2001a; 2001b) proposed a mechanism with a three-step logical chain to depict the LKSs concept (2001a, p. 980; 2001b, p. 258):

1. Knowledge generated within innovative firms and/or universities is somehow transmitted to other firms.
2. Knowledge that spills over is a (pure) public good, i.e. it is freely available to those wishing to invest to search it out (non-excludability), and may be exploited by more than a few users at the same time (non-rivalry).
3. Despite this, knowledge that spills over is mainly "tacit," i.e. highly contextual and difficult to codify, and is therefore more easily transmitted through face-to-face contacts and personal relationships, which require spatial proximity; in other words, it is a public good, but a local one.

Within this logical chain, knowledge is considered a "local" public good spreading pervasively within a spatially bounded area (Callon and Bowker 1994; Markusen 1996; Breschi and Lissoni 2001a; 2001b; 2003). The role of geographical proximity facilitates knowledge production and exchange by providing opportunities for interacting and sharing of experience between firms (Audretsch 1998; Howells 2002). Co-located individuals, inventive firms, and other institutions receive more positive benefits from nearby knowledge resources to practice invention and innovation compared with those firms located in distant places (Jaffe 1989; Acs et al. 1992; Audretsch and Feldman 1996).

Empirical studies of LKSs

Considerable research effort has been made to identify the nature and strength of LKSs. Jaffe (1989), Acs et al. (1992; 1994), and Audretsch and Feldman (1996) found indirect evidence of LKSs in their investigations of patent and innovation counts.[1] Approaches that are more direct in finding local knowledge spillovers have tracked the geography of patent citations (Jaffe et al 1993; Almeida and Kogut 1999; Thompson and Fox-Kean 2005). Estimating regional knowledge production functions permits incorporation of spatial dependence and/or spatial heterogeneity to understand the distinctive roles of industrial and university research on the geography of invention and innovation (Anselin et al. 1997; 2000; Autant-Bernard 2001; Varga 2000). Overall, these empirical studies have found that knowledge flows are geographically localized by showing that higher rates of research and development (R&D), invention and innovation, entrepreneurial activity, and high-technology production are bounded in space (Feldman 1999).

The theory of LKSs has been further applied within the broad realm of the new industrial geography (NIG) (Martin and Sunley 1996). The term NIG stresses the development of localized networks as a decisive force in the generation of dynamic regional growth processes (Amin and Thrift 1992). Silicon

[1] Jaffe's (1989) empirical work found that industrial patenting (an indicator of innovative output) responds positively to knowledge spillovers from university research (an indicator of innovative input) conducted in the same U.S. state. Building upon Jaffe's work, Acs et al. (1992) also found significant evidence in favor of the agglomeration advantages by using innovation counts instead of patent data.

Valley, Boston's Route 128, and Italian industrial districts are the most debated cases of NIG, where interactive learning and invention are particularly strong (Saxenian 1994; Asheim 1996). NIG-related districts are also titled innovative milieux (Camagni 1991), high-technology clusters (Porter 1990), learning regions (Florida 1995; Asheim 1996), technopoles or science parks (Luger and Goldstein 1991; Massey et al. 1992), and regional innovation systems (Cooke 2001; Asheim and Isaksen 2002). In these territorial production complexes, localized supply chain networks including both backward and forward links are dense, and related firms benefit from agglomeration economies of nearby suppliers and business services (Ettlinger 1990).

Recent criticism on LKSs

Many scholars acknowledge the importance of geographical proximity to knowledge flows and spillovers, but others are unconvinced and call for clarification. First, the evidence of knowledge flows is largely circumstantial and seldom investigated explicitly with regard to the nature of boundaries (Breschi and Lissoni 2001a; 2001b; Gallaud and Torre 2004; Ponds et al. 2007). Predefined geographical units of analysis limit a thorough understanding of how and to what extent knowledge flows circulate in space (Autant-Bernard et al. 2007b; Maggioni et al. 2007). For example, Breschi and Lissoni criticized Jaffe's (1989) work on the choice of the U.S. states as the geographical units of analysis as (2001b, p. 260):

> [S]tate boundaries are a very poor proxy for the geographical units within which knowledge ought to circulate. U.S. states simply are too large geographical units to allow us to assume that inventors, entrepreneurs and managers living in one

state will have more chances to have face-to-face contacts between each other than with people living elsewhere. Similarly, there is no reason to presume the existence of a common cultural background, nor a close set of parental or friendship ties, which ought to make mutual understanding and trust easier, and reduce transaction costs.

Autant-Bernard et al. (2007b), Boschma and Ter Wal (2007), and Ponds et al. (2007) also argued that some studies of LKSs only investigate the effects of geographical proximity on collaboration choices and the reality of spatial externalities has rarely been convincingly demonstrated.

Second, studies of LKSs claim that the tacit aspects of knowledge are mostly a matter of face-to-face contacts, which requires geographical proximity (Amin and Cohendet 2005; Breschi and Lissoni 2001a; 2001b; Coe and Bunnell 2003). Increasing emphasis is placed on tacit knowledge as opposed to codified knowledge, in that the former is seen as more valuable to local inventive activity, while the latter is associated more with standardized mass production, which can be communicated across distance (Martin and Sunley 2003). However, the dualisms of tacit versus codified knowledge and local versus global geographies are criticized by some scholars as an over-simplistic view. Howells (2002) argued that people could not perfectly separate knowledge and its spatial features into "tacit-local" and "codified-global" binary relations. Torre and Rallet (2005) also argued that this bipolar distinction is too simple, which the theory of LKSs is consequently assumed as *tacit knowledge = face-to-face transmission = need for geographical proximity = co-location of economic actors*. In fact, although the nature of tacitness differs from that of codification, both knowledge domains are interdependent and complementary (Polanyi 1966; Bathelt et al. 2004). Breschi and Lissoni (2001a) suggested that "tacitness is a key exclusionary mean, which can be willfully manipulated to prevent a number of actors (even local ones) from understanding the content of scientific and technical messages" (p. 980). Knowledge is considered "tacit" not because it cannot be fully articulated in abstract contexts, but because it is highly specific. Although some information can be codified by developing appropriate vocabulary with supporting formats such as academic articles, codebooks, or manuals, messages that transmit the "information" are still often tacit and dynamic (Breschi and Lissoni 2001a; 2001b). Additionally, since tacitness and codification are mutually compatible, tacit messages can be transmitted along with codified documents across longer distances through a broad range of advanced information and communication technologies.

In short, a rethinking of tacit and codified knowledge realizes that it is difficult and not necessary to organize both types of knowledge along neat geographical scales and domains. Recent studies of knowledge flows are centered on issues of how easily knowledge can be shared and transferred across distance. In the dissertation, I develop an alternative framework for describing knowledge production and exchange over space.

2.2 External Relations and Proximity

The notion of LKSs emphasizes that geographical proximity between economic actors fosters local knowledge exchange and diffusion because human skills and know-how are bounded in space. However, some scholars argue that studies of LKSs overly stress local effects. They highlight external knowledge resources via collaborative networks. This section discusses the important role of external knowledge resources for regional invention and innovation. The first part outlines the importance of external knowledge resources in technological advance. The second part discusses external links with non-local partners and how these links are critical in forming a network-based system. This is followed by a discussion of knowledge exchange over space, particularly on how long-distance collaborations can be built and maintained by communities of practice. The last part concentrates on the role of collaboration by proposing a conceptual framework for understanding inter-territorial networks of biotechnology co-invention.

External knowledge resources

Increasingly, studies stress that access to external knowledge resources is critical to triggering successful invention and innovation and regional development (Powell 1996; Oinas 1999; Audretsch 2001; Bathelt 2002; McKelvey et al. 2003; Coenen et al. 2004; McKelvey 2004; Gertler and Levitte 2005). Breschi and Malerba (2001) argued that strong external links are vitally important to regional competitiveness. Bathelt et al. (2004) showed that dynamic firms in successful clusters build and maintain a variety of internal and external knowledge resources. In addition, advanced information and communication technologies reduce costs of moving knowledge and increase access and availability of universal resources (Torre and Rallet 2005). Understanding the role of knowledge in driving regional economic growth is shifting from a focus on closed territorial relationships, towards an emphasis on extra-local links with distant markets and technological clusters (Coe and Bunnell 2003; Bathelt 2005; 2007; Yeung 2005; Vallance 2007; Kroll and Mallig 2009). The concepts of extra-local links and external knowledge resources provide new ways for explaining the geography of invention and innovation. Technological and commercial successes of many firms in Silicon Valley, for example, are closely tied to partners located in other regions and countries (Saxenian and Hsu 2001).

External links and network-based system

Concern for extra-local relationships is influenced by the "relational turn" in contemporary economic geography (Storper 1997; Boggs and Rantisi 2003; Bathelt et al. 2004; Yeung 2005; Sunley 2008). The relational turn refers to a greater awareness of social, cultural, and ethnic dimensions of economic systems in different places and how these attributes shape complex relationships among diverse economic actors (Amin 2002; Yeung 2005). It concerns the ways in which social and cultural forces influence levels and growth of urban and regional economies (Boggs and Rantisi 2003). Sunley (2008) argued that "the origins of relational thinking lie partly in economic sociology and its view of the network embeddedness of economic life, and partly in the learning processes and untraded assets that are typical of institutionalist approaches" (p. 2). Innovation and invention not only require local interactions between firms within a cluster, but they also need ties among distant actors that provide access to complementary information, skills, and technologies (Asheim and Isaksen 2002; Boschma 2005; Gertler and Levitte 2005; Maggioni et al. 2007). Knowledge flows are not tightly bounded within a given territory. Relational patterns function as network-based systems associated with different geographical sites. Knowledge flows in these systems are dependent upon shifting alignments of economic actors in different locations in pursuit of particular corporate goals (Amin 2002; Amin and Cohendet 2005). Many ties between firms in Silicon Valley and Taiwan, for example, are shaped by interpersonal connections between Taiwanese nationals with educational and working experience in both places (Saxenian and Hsu 2001).

A network-based system, as Amin and Cohendet (2005) argued, is "a disconnected spatial ecology of knowledge that can be held in place as relational knowledge" (p. 472). It is composed of relational ties between economic actors from different places allowing for the possibility of seeing a broad range of knowledge flows that have two crucial dimensions. First, it holds knowledge to be place-specific for different markets where inventive activities are specifically aligned with the distinctive needs of local clients and customers (Vallance 2007). Second, it translates varied ideas and practices into a corporate template where interactive learning takes place among distant actors through collaborative projects or other joint activities (Allen 2000). Amin and Cohendent (2005) noted that "without doubt, one of the achievements of corporate form…is to hold varied knowledge architectures in place and establish knowledge coherence across different spatial sites" (p. 471).

Building extra-local links is not a simple task, especially when connecting with geographically distant partners. Differences of cultural and socio-institutional contexts limit mutual understanding of outsiders. Knowledge transfer problems may be aggravated with distance, especially when linking to remote areas, which results in considerable uncertainties on the level of investment in research and development (Gertler 2001).

Knowledge exchange through communities of practice

Knowledge flows are not neatly organized into separate bundles of local and global geographies. People with appropriate levels of expertise and experience are able to share and understand technological know-how, even if they are geographically dispersed. Without personal relationships and skills for exchanging information, neighboring people may learn nothing from each other. People engaged in a common language, basic understanding, and mutual interaction are referred to as a community where knowledge can be shared, conveyed, and utilized effectively based on their relationships. This is particularly a case in a *community of* practice – a group whose members regularly engage in sharing and learning by their common interests (Gertler 2001; Coe and Bunnell 2003; Amin and Cohendet 2004; Conene et al. 2004; Bathelt 2007; Vallance 2007). Nooteboom et al. (2007) inferred the concept of a community of practice to the cognitive distance between firms that:

> [T]hey (people in different firms) need to share certain basic perceptions and values to sufficiently align their competencies and motives…, established by means of shared fundamental categories of perception, interpretation and evaluation inculcated by organizational culture. Differences in such organizational focus yield cognitive distance between firms (p. 1017).

Wenger et al. (2002) illustrated the features of communities of practice as:

> [M]any communities start among people who work at the same place or live nearby. But co-location is not a necessity. Many communities of practices are distributed over wide areas. Some communities meet regularly…Other are connected primarily by e-mail and phone and may meet only once or twice a year. What allows members to share knowledge is not the choice of a specific form of

communication (face-to-face as opposed to Web-based, for instance) but the existence of shared practice – a common set of situations, problems and perspectives (p. 25).

In short, members of a community of practice develop their experiences, perceptions and opinions, and reduce uncertainty about each other in pursuit of corporate goals (Rice and Aydin 1991; Amin and Cohendet 2005). It envisions the possibility of scientific collaboration resulting in various forms and spatial levels so long as the members are mediated within a community. This dissertation concentrates on co-inventive activity, which involves repeated exchanges of both tacit and codified knowledge through a series of “face-to-face” and “over the distance” interactions (Breschi and Lissoni 2004; Maggioni et al. 2007). Since co-inventors are linked by a common set of meanings, understandings, and learning processes, geographical proximity is neither a necessary nor a sufficient condition for interactive learning (Boschma 2005). Amin and Cohendet (2004; 2005) argued that through regular and frequent contacts – including video- or teleconferences, telephone conversations, or e-mail exchanges, as well as occasional on-site meetings (also Sapsed et al. 2005) – dispersed inventors within a community of practice can find ways to collaborate.

Inter-territorial networks of biotechnology co-invention

Network-based approaches provide useful analytical tools to assess knowledge flows (Oinas 1999; Gertler 2003). Cowan (2005) argued that current interests in collaborative networks has been fueled by three different developments: (1) the growth of network technologies in measuring social networks, knowledge exchange, and the inventing performance of firms, (2) the increasing multiple relationships of firms in webs of global alliances, and (3) the expanding knowledge and technological progress underlying corporate inventive strategies. The rise of inter-territorial collaborative networks, as Amin and Cohendet (2005) explained, "is precisely what has made trust, intimacy, and familiarity possible at a distance and thereby allows learning to take place" (p. 470). Interactive learning and deliberate collaboration with non-local partners are crucial for invention and innovation success (Cooke 2001; McKelvey et al. 2003; Gertler and Levitte 2005; Birch 2007; Trippl et al. 2009; Balland et al. 2011).

Inter-territorial networks explored here with an emphasis on biotechnology co-invention might reveal the possibility of a core/periphery structure that Borgatti and Everett (1999) originally characterized. This is a relational system where core areas play an active role in the network as they are strongly connected with each other and to some outsiders. Conversely, areas in the periphery play a passive role in the network as they are loosely connected or even disconnected from one another (Alderson and Beckfield 2004; Cattani and Ferriani 2008; Alderson et al. 2010; Rubí-Barceló 2010). Cooke (2006) argued that a large portion of biotechnology value-chains (e.g., venture capital, R&D, human resources) occurs in several global cities such as Boston and San Francisco in North America, and Cambridge, Munich and Stockholm in Europe. Knowledge sharing and collaboration in these places originates in centers of excellence – leading academic research and large pharmaceutical companies (Coenen et al. 2004). These global cities generally composed of the core region serve as "megacenters" by operating with relatively open science conventions and integrating other cities into a system of “open innovation” that stretches biotechnology knowledge domains over space (Coenen et al. 2004; Cooke 2006; Moodysson et al. 2008).

On the other hand, peripheral areas have distinctive knowledge architectures that support local clients and customers. Global circulation of knowledge through non-local links provides exchange opportunities for local invention and invention. This may lead to certain “club” characteristics in a network where member cities enjoy strong knowledge production and exchange ties on a global scale, which can reduce the possibility of negative technological lock-in (Asheim and Isaksen 2002; Boschma 2005; Gertler and Levitte 2005; Cooke 2007). Access to inter-territorial networks is of particular importance for inventive biotechnology firms located in peripheral areas that are remote from main research and market centers. Successful networking strategies assist in accessing external knowledge. Most remote inventive firms crucially rely on non-local knowledge partners and global networking relationships (Simmie 2003; Fontes 2005; Trippl et al. 2009; Balland et al. 2011).

In short, the space of knowledge flows in biotechnology co-invention shows a strong concentration of inventive activity in several megacenters. A high degree of network-based proximity links cities and regions and extends the scope and geography of cooperation over space. Is the U.S. intermetropolitan network of biotechnology co-invention consistent with a relational core/periphery structure or is it more complex? This dissertation investigates individual metropolitan areas' network positions to provide useful insights into the structural properties of intermetropolitan co-invention networks.

2.3 Intermetropolitan Co-Invention Network-Based Systems

A broad perspective on knowledge circulation in a spatially stretched economic sphere offers an alternative way to conceptualize the space of knowledge flows. This dissertation investigates inter-territorial knowledge flows by concentrating on intermetropolitan networks of biotechnology co-invention. These networks are constructed by tracking inventors who participate in biotechnology co-patenting and attributing each co-patent to metropolitan areas where the inventors reside. A network-based system consists of co-patenting ties between inventors from different metropolitan areas with three key features: (1) areas are mostly tied by network-based proximity (Coenen et al. 2004; Vallance 2007), (2) both local and global knowledge flows underpin collaboration and exchange (Breschi and Lissoni 2003; Bathelt 2007; Kroll and Mallig 2009), and (3) a collaborative ecology are aligned as knowledge coherence across different geographical sites (Amin and Cohendent 2004; Sunley 2008; Autant-Bernard et al. 2010).

Social network analysis has become an important tool to analyze the ways that individuals, firms, cities, regions, and countries are interconnected (e.g., Cattani and Ferriani 2008; Boschma and Frenken 2009; Alderson et al. 2010; Rubí-Barceló 2010). A network and its structure is essentially a sociogram, in which entities are nodes, and the relationships among pairs of entities are connecting lines. In the intermetropolitan co-invention networks of this dissertation, the nodes show the observed metropolitan areas with varying intensity of biotechnology capacity, while the links reveal the existence of knowledge flows between connecting areas.

Chapter 3

APPLYING SOCIAL NETWORK ANALYSIS TO CO-INVENTION NETWORKS

Social network concepts and analytical techniques are popular in studies of invention networks (e.g. Breschi and Lissoni 2003; Singh 2005; Ejermo and Karlsson 2006; Cantner and Graf 2006; Fleming and Frenken 2007, Ponds et al. 2007; Boschma and Frenken 2009). Empirical evidence shows that networks facilitate knowledge exchange and influence inventive performance (Cooke 2001; Coe and Bunnell 2003; McKelvey et al. 2003; Gertler and Levitte 2005; Birch 2007). This chapter reviews the literature on social network analysis used to describe intermetropolitan networks of biotechnology co-invention. Section 3.1 presents channels of knowledge exchange through networks. Section 3.2 links interfaces between intermetropolitan and social networks. Section 3.3 interprets terminology and concepts of social network analysis used in the dissertation.

3.1 Channels of Knowledge Exchange through Networks

Knowledge flows circulate in space in various forms and through different channels. Cassi and Morrison (2007) presented a simple four-way classification of channels of knowledge exchange, which is modified and shown in Figure 3.1. These channels are classified into two dimensions: directions of knowledge flows, and relationships between inventors. The former refers to the flows of knowledge, either in one-way or two-way directions, while the latter refers to the relationships between inventors, either through formal or informal agreements. The channels in the quadrants differ according to the influence of social networks on knowledge flows.[2] The upper left quadrant shows unilateral flows that are transmitted through formal relationships. These interactions typically occur in inter-organizational communities such as licensing and consulting agreements whose one-way ties occur between firms. The lower left quadrant is characterized by formal relationships and bilateral flows. Examples include co-patenting, research and development alliances, and cross-firm task forces. These corporate relationships do not differentiate between directions of knowledge flow. The upper right quadrant occurs at the individual level. Job mobility is the most common example of this type of knowledge exchange. Since human skills and know-how are tacit, the processes of job mobility forge knowledge flows between firms (Agrawal et al. 2006). The lower right quadrant captures informal social networks of individuals, but with an emphasis on interpersonal communication. For example, people from different communities interact by sharing experience and expertise. Common understanding and mutual trust are essential in these exchanges. This dissertation focuses on co-inventive activity over time and space with an emphasis on bilateral knowledge flows in formal co-patenting relationships of biotechnology inventors, as shown in the lower left quadrant of Figure 3.1.

3.2 Interfaces between Intermetropolitan and Social Networks

This section links interfaces between intermetropolitan and social networks. As shown in Figure 3.2, the upper part of the figure illustrates a simple geographical space with cities A to F. The lower part of the figure shows its

[2] Note that these channels are ideally separated for the purpose of illustration, and they may tend to affect and overlap with one another in reality.

		Relationships between inventors	
		Formal	Informal
Directions of knowledge flows	Unilateral	Licensing/ Consulting	Job mobility
	Bilateral	**Co-patenting/** R&D alliances	Interpersonal know-how exchange

Figure 3.1 Cross-classification of channels of knowledge exchange via networks (adapted from Cassi and Morrison 2007)

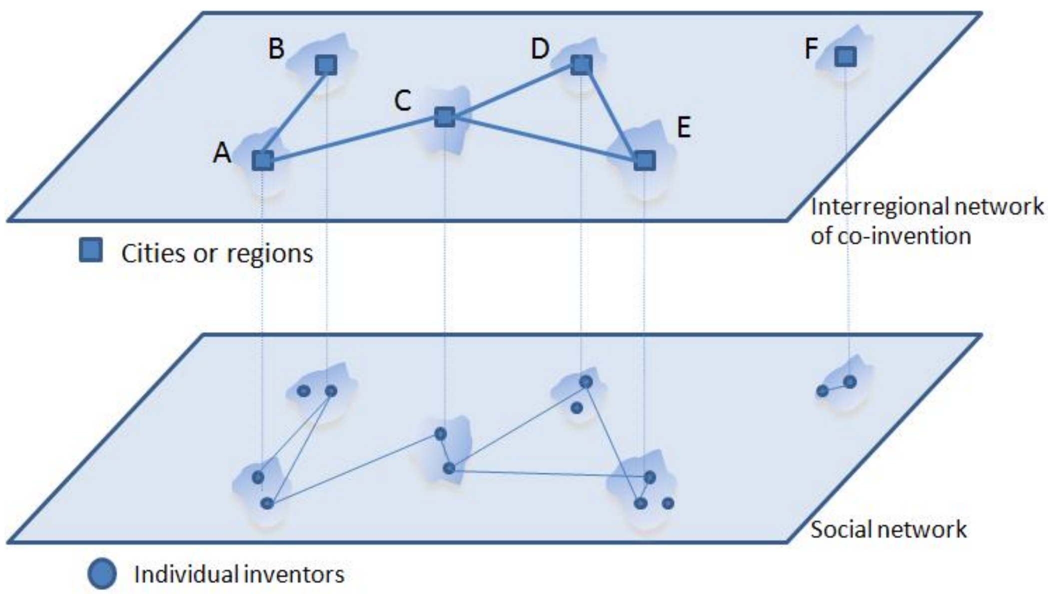


Figure 3.2 Intermetropolitan network of co-invention, social network, and knowledge circulation in space

social network counterpart, as individual inventors are located in different cities. The links in the lower part refer to co-inventive activities between individual inventors. Some inventors have extra-local relationships with inventors located in other cities, while other inventors only collaborate with local partners. Social relations are assumed to underpin and facilitate interactions and communication. Co-patenting is viewed as evidence of groups of inventors with social relations participating in cooperative invention activities. Knowledge is embedded in individuals who reside in areas (Polanyi 1966). Inventor ties shape the intermetropolitan network shown in the upper part of the figure. City C is directly connected to A, D, and E, indirectly connected to B, and it has no connection with F. Inventors in city F only co-invent with local partners. These inventors are isolated from direct involvement with non-local inventors in the network and it is assumed that they have no knowledge exchange with inventors in other cities. In this network-based system, city C occupies a favorable position since it has the most connections to other areas. Cities A and B are the most intensely tied. Two of A's inventors co-patent with an inventor in B. In this simple system, no other pair of cities has more than one link.

3.3 Terminology and Concepts of Social Network Analysis

This section discusses terminology and concepts of social network analysis used in this study. Focus is placed on non-directional relationships between nodes.

Nodes, lines, and graphs

A network is simply defined as a set of entities linked by relational ties (Cassi and Morrison 2007). Structural features of a network are commonly depicted as a sociogram, in which entities are nodes (or points) in a two-dimensional space, while the relationships between pairs of entities are connecting lines (or edges, arcs). In a simple sociogram, as shown in Figure 3.3, the nodes indicate the observed metropolitan areas, while the lines signify co-inventive ties between these areas.[3] In this figure, nodes A, B, C, D, and E are connected by a set of lines. Nodes B and A are the most connected pair. Node F has no connection with other nodes.

Valued graphs

Network data may consist of valued relations, in which the frequency of interaction between each pair of nodes is recorded, as compared with a line in a binary graph that represents only the presence of a tie. Valued graphs are the appropriate representation of the intensity of relations in a network. The frequency of interaction can be visualized either by labeling its magnitude along a line, or by depicting a relative thick or thin line. In Figure 3.3, for example, the value attached to the line connecting nodes A and B coded 2 is referred to as the frequency of interaction between them. The higher valued line can also be depicted as a thicker line compared with thinner lines that have lower tie-strength.

[3] Note that a sociogram is a hypothetical graph and there is no single "correct" way to depict a co-invention network and its features.

Geodesic distance

Geodesic distance is the smallest number of lines (or edges, arcs) connecting any two distinct nodes. For example, in Figure 3.3, there are three different paths connecting nodes A and D: A-B-D, A-B-E-D, and A-B-C-E-D with the lengths of these three paths 2, 3, and 4, respectively. The shortest path between A and D is A-B-D, which yields a geodesic distance of 2. A geodesic path can be regarded as the optimal or most efficient way for connecting two nonadjacent economic actors (Henneman and Riddle 2005). This dissertation uses a geodesic distance-based approach to measure the closeness of an intermetropolitan network. Metropolitan areas are able to exchange new ideas or pass information if and only if they are either directly or indirectly connected. Geodesic distance-based approaches provide ways of tracing connections between areas even if they are not geographically proximate. It is not meaningful to measure two areas' geodesic distance if there is no path connecting these two areas. When considering knowledge exchange, all areas need to be embedded in the same network. For example, in Figure 3.3, there are no lines connecting with node F, so the geodesic distances from other nodes to F are infinite. This indicates that no knowledge flows circulate between this isolated node and others.

Using measures of geodesic distance to investigate intermetropolitan networks has several purposes. First, if the average geodesic distance among areas is small, it suggests that knowledge flows occur directly and quickly. Second, the longest geodesic distance among all pairs of areas is defined as the *diameter* of the network, which could refer to the domain of knowledge flows

(Wasserman and Faust 1994; Balconi et al. 2004). Third, a clustering analysis of geodesic distances among areas produces a dendogram that is helpful in visualizing the hierarchical structure of similarity (or dissimilarity) relationships among metropolitan areas.

In short, these concepts of social network analysis are used to construct intermetropolitan co-invention networks. The nodes show the U.S. metropolitan areas and the links reveal the existence of biotechnology co-patenting among intermetropolitan pairs. Geodesic distance-based approaches are applied to measure the degrees of closeness and regular equivalence between member areas of the co-invention network.

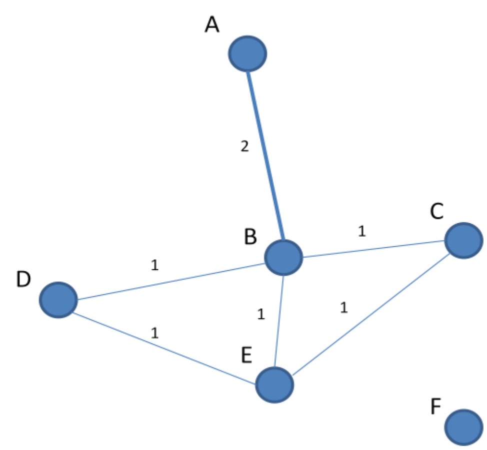


Figure 3.3 Simple graph of social network

Chapter 4

RESEARCH METHODS

This chapter outlines the methodology used in these two research topics of the dissertation. Section 4.1 introduces methods of spatial and network-based dependencies used to identify intermetropolitan relationships in biotechnology co-invention. Section 4.2 outlines methods of social network analysis used to detect structural properties of intermetropolitan co-invention networks. These research methods are summarized in Section 4.3.

4.1 Measures of Dependence

The first research task asks whether the U.S. biotechnology co-invention urban system reveals significant differences between spatial and network-based intermetropolitan dependencies. The longitudinal changes in these dependencies are also explored. Maggioni et al. (2007) tested whether non-spatial networks between geographically distant clusters prevail over patterns based on spatial contiguity. By comparing spatial versus relational dependence in European patenting activity, they identified several strong relational clusters among geographically distant centers. These relational associations had stronger ties compared to those extracted from analysis of spatial dependencies. This dissertation focuses on co-inventive activity within and across American biotechnology communities. The focus here is to identify differences in spatial and network-based dependencies across the U.S. urban system by comparing patterns revealed in global- and local-level measures of association. Each metropolitan area's co-invention rate (or co-patenting rate, both terms are used interchangeably) is estimated by dividing its annual biotechnology co-patent counts by the number of wage and salary jobs. This ratio is multiplied by a scaling factor of 1,000. Labor force and not total population is used to standardize co-patent counts because not all people generate inventions. Moreover, as independent inventors account for a relatively small portion of total patenting activity especially in biotechnology (Adelman and DeAngelis 2007), the number of wage and salary workers better relates co-inventive activity to potential inventors.[4] The data for wage and salary jobs were retrieved from the Bureau of Economic Analysis' Regional Economic Accounts (Bureau of Economic Analysis 2010). The estimated metropolitan co-invention rates of some areas with few skilled workers and rare co-patenting events may be spuriously identified as "outliers" (Messner and Anselin 2004). To compensate for co-patenting rate instability in these metropolitan areas, original rates are smoothed using an Empirical Bayes Smoother (see Anselin et al. 2006a for more details). Anselin et al. (2006a) argued that the Empirical Bayes Smoother is referred to as "shrinkage in the sense that the crude rate is moved (shrunk) towards an overall mean, as an inverse function of the inherent variance" (p. 39). The Geoda software package generates the empirical estimates (Anselin 2004).

This section consists of two parts. The first part presents a global-level measure of dependence. The second part discusses the procedure of modeling a local-level measure of dependence. The aim of estimating both measures is to

[4] U.S patent data show that corporations account for an average of 80% of the assignees (or owners) of biotech patents awarded from 1990 to 2004, following by universities, non-profit organizations, and U.S. government (Adelman and DeAngelis 2007).

provide a complete analysis of the relative influence of spatial and network-based proximities in American biotechnology co-invention.

Global-level measure of dependence

Moran's *I* is used to detect global-level spatial and network-based dependencies in the U.S. biotechnology co-invention urban system. This statistic provides an overall measure of the strength of cross-sectional autocorrelation in a data distribution (Moran 1950). It is calculated by comparing the co-patenting rate of each metropolitan area and co-patenting rates of its "spatial" or "network-based" neighbors. The global measure of Moran's *I* is defined as:

$$I = \frac{m}{\sum_{i=1}^{m}\sum_{j=1}^{m} w_{ij}} \frac{\sum_{i=1}^{m}\sum_{j=1}^{m} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{m}(x_i - \bar{x})^2}$$

where m is the number of metropolitan areas; w_{ij} is an element of an $m \times m$ weights matrix W; x_i and x_j are the biotechnology co-patenting rates in areas i and j, respectively; and $\bar{x}$ is the average of all x values. The interpretation of Moran' *I* statistic is similar to the Pearson's correlation coefficient in that both values range between +1 and -1. When *I* > 0, the overall pattern indicates positive autocorrelation, meaning that areas with similar co-patenting rates, either high or low, are spatially (or network-based) located "near" each other. When I < 0, on the other hand, it shows negative autocorrelation, meaning that areas with dissimilar co-patenting rates are located "near" each other. When *I* = 0, the overall pattern is random, indicating that metropolitan biotechnology co-invention is independent of either spatial or network-based proximity.

Local-level measure of dependence

The local-level measure of dependence is based on the *local indicators of spatial association* (also called LISA), which allows for the decomposition of Moran' *I* into the contribution of each individual area. The Local Moran statistic is defined as:

$$I_i = \frac{(x_i - \bar{x})}{\frac{1}{m}\sum_{k=1}^{m}(x_k - \bar{x})^2} \sum_{j=1}^{m} w_{ij}(x_j - \bar{x})$$

It provides a means to assess significance of local spatial patterns (Anselin 1995). A map combining the information on the locations and the significance of Local Moran statistics is referred to as a *LISA cluster map*. Ó hUallacháin and Lee's (2011) approach is used to distinguish between the following possible local association patterns in both the spatial and network-based systems.

1. *Co-invention Cores* (high-high): These are metropolitan areas with high co-patenting rates and are significantly similar to their neighbors.
2. *Co-invention Peripheries* (low-low): These are metropolitan areas with low co-patenting rates and are significantly similar to their neighbors.
3. *High Co-invention Islands* (high-low): These are metropolitan areas with high co-patenting rates but are significantly different from their neighbors.
4. *Low Co-invention Islands* (low-high): These are metropolitan areas with low co-patenting rates but are significantly different from the co-patenting rates of their neighbors.

5. *Non-significant Areas*: Based on a conditional permutation approach, these are metropolitan areas with non-significant Local Moran statistics ($p > 0.05$), indicating a failure to reject the null hypothesis of spatial randomness. Co-patenting rates in these metropolitan areas are not significantly similar to or different from co-patenting rates of their neighbors.

Two types of LISA cluster maps – spatial and network-based – are compared to identify significant cores, peripheries, and islands across the U.S. biotechnology co-invention urban system. If a co-inventive core (high-high) appears in the spatial LISA cluster map, it indicates that intermetropolitan spatial dependence is important in biotechnology co-patenting. If the same neighboring areas constitute a co-inventive core in the network-based LISA cluster map, collaborative relationships are defined by both spatial and network-based associations. More than likely differences in spatial and network-based dependencies occur. In particular, network-based LISA cluster maps with non-significant spatial dependence should show co-inventive cores that are geographically scattered. Intermetropolitan collaborative networks often favor co-patenting by inventors living in geographically dispersed locations. Low co-inventive islands (low-high) in the network-based system are probably common as several major biotechnology centers exist and most metropolitan areas' ties favor these centers.

To capture the neighboring structure of each observation, whether from the aspect of spatial or network-based proximity, I establish a weights matrix W specifying the interaction strength between each pair of metropolitan areas. Each row i of matrix W has elements w_{ij} corresponding to the columns j. Three principal spatial weight choices exist: (1) contiguity (w_{ij} =1 for i and j sharing a common boundary), (2) distance (w_{ij} =1 for $d_{ij} < \delta$) where d_{ij} is the distance between areas i and j and δ is the threshold, and (3) the number of nearest neighbors. Owing to the "island" nature of U.S. metropolitan areas, contiguity-based weights are inappropriate. The wide variation in metropolitan spacing also renders the distance between areas problematic (Ó hUallacháin and Lee 2011). In this dissertation, the number of nearest neighbors – an area's k values – is perhaps the best choice in identifying neighboring metropolitan areas in the continental U.S. Since there are no objective rules to determine the appropriate number of nearest neighbors, several alternatives are considered ranging from five to 17. Concern for the stability of the LISA cluster maps in the Monte Carlo simulations led to the selection of ten nearest neighbors or 7 percent of all possible 149 metropolitan neighbors. This number of nearest neighbors defines discernible regional groupings using the smallest k value.

The network-based weights matrix W_n is based on the number of times each pair of metropolitan areas jointly involves in biotechnology co-patenting. For any two metropolitan areas, an intermetropolitan tie is established if inventors from both areas co-invent the same patent. The more often inventors from two areas co-invent, the stronger are intermetropolitan relational ties. This network-based weights matrix W_n is obtained by converting the intermetropolitan valued matrix into a set of binary relations (see Chapter 5 for details). The original valued matrix is dichotomized using the average number of times that metropolitan areas are tied in biotechnology co-patenting (e.g., Maggioni et al. 2007). When the annual frequency of co-invention between two metropolitan areas i and j is greater than or equal to an average-based cut-off point then w_{ij} =1, indicating that both areas are relationally connected as neighbors; otherwise w_{ij} =0. In summary, an intermetropolitan network is constructed to obtain a network-based weights matrix. The simplest form of an intermetropolitan network consists of a square actor-by-actor matrix, where the rows and the columns represent the same set of metropolitan areas. The relationships between every possible pair of metropolitan areas provide a way to assess the structure of connections within which these metropolitan areas are embedded. In this analysis of 150 metropolitan areas, a 150×150 symmetrical matrix is generated for each chosen year of analysis. Elements on the off-diagonal are the annual number of times each pair of metropolitan areas join in biotechnology co-patenting. These indicate the nature and strength of intermetropolitan ties that facilitate knowledge exchange across metropolitan boundaries. Elements on the main-diagonal are excluded from the analysis since metropolitan self co-patenting is not considered in this study.

4.2 Methods of Social Network Analysis

The second research task concerns how and to what extent knowledge flows of biotechnology circulate through co-invention networks over space. Autant-Bernard et al. (2007) argued that there is a limited understanding of how knowledge flows circulate among related firms through collaborative networks. Coenen et al. (2004) argued that the advantage of network-based flows among a group of cities is to stretch knowledge domains over space. Ponds et al. (2007) emphasized a city's network position as an indicator of inventive capacity. Cooke (2007) argued that the biotechnology industry leads to a global hierarchical structure of knowledge flows. By combining these arguments above, this dissertation explores the structural properties of intermetropolitan co-invention networks by focusing on the roles of metropolitan areas from three different perspectives: components and cohesive subgroups of metropolitan areas, intermetropolitan network centralization and centrality, and positions established within the co-invention network-based system. The characteristics of network components reveal patterns of connection among metropolitan areas. The centers of a network identify metropolitan areas with the best access to knowledge flows. Network positions of metropolitan areas show the co-invention network-based system these member areas form and the roles played by different types of areas within the system. The purposes of conducting these empirical analyses are to understand network properties of co-invention and the diverse positions of metropolitan areas in systems of knowledge exchange. All analyses are performed by using the UCINET program for social network analysis (Borgatti et al. 2002).

This section is organized as follows. The first part introduces methods used to identify network components and cohesive subgroups of metropolitan areas. The second part outlines methods used to investigate the importance of metropolitan areas in exchanging knowledge. The third part discusses procedures used to partition individual metropolitan areas into network positions.

Identify components and cohesive subgroups

This part introduces the theoretical background for investigating network components. It also defines and illustrates the structural properties of a network as an approach for identifying cohesive subgroups where metropolitan areas are intensely connected to each other.

1. Components

The concept of a *component* is defined as a maximal connected set of nodes in a network that can trace a direct or indirect linkage to one another. Components are isolated from each other, so there are no paths between member nodes of different components. The size of a component is measured by the number of nodes that are linked to one another. In Figure 3.3, for example, the size of the major component is five, which consists of five nodes including A, B, C, D and E. Metropolitan areas linked in a component are able to transfer knowledge through co-patenting ties and each has a competitive niche that is integral to the network-based system (Coenen et al. 2004; Cooke 2006). Isolated metropolitan areas do not have information exchange opportunities because they lie outside the major component and only have intra-metropolitan knowledge flows between local inventors. The characteristics of components identified in an intermetropolitan co-invention network are taken as an initial step in describing knowledge exchange in networks of American biotechnology.

2. Cohesive subgroups

A cohesive subgroup is defined as a set of nodes with relatively strong or frequent ties in a network (Wasserman and Faust 1994). Traits of nested components are used to identify cohesive subgroups of nodes underlying each component. This approach applies progressively stronger cut-off criteria of connectedness to draw a component boundary into a series of concentric circles. A nested component is depicted as having a core, which consists of the most connected nodes, along with outer circles of the core, being gradually extended to include more and more nodes with weaker levels of connectedness. Figure 4.1 shows a simple case of nested components where nodes located in circle A are the most closely connected to one another. This is the core of a component. Nodes in circle B are extracted with a weaker criterion of connectedness. They include all nodes of circle A together with additional nodes in circle B, which are connected at a weaker level of connectedness (Scott 2000). The boundary of circle C includes nodes at the weakest level of connectedness. This boundary contains all nodes of a component.

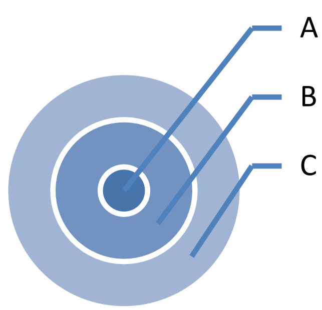


Figure 4.1 Conceptual graph of nested components

Two procedures identify cohesive subgroups of nodes in a network. One emphasizes the number of neighbors of each node titled the *k*-cores procedure. The other focuses on the frequency of connection between each pair of nodes and is titled the *m*-slices procedure.[5] In graph theory, the number of neighbors directly tied to a node refers to the node's degree. Analysis of different levels of degrees among nodes illustrates certain members with relatively high cohesion. These nodes are called *k*-cores subgroups and *k* indicates the minimum node's degree in a subgroup. For example, a *1*-core subgroup contains nodes directly tied to at least one other node (indicating a degree of at least one). A *2*-core subgroup contains nodes connected to at least two other nodes (indicating a degree of at least two), while the nodes with degree one are ignored. A *3*-core subgroup contains even fewer nodes because only nodes with degree three or more are included. Since nodes in a *3*-core are also part of a *2*-core, but not all member nodes of a *2*-core belong to a *3*-core, *k*-cores subgroups are nested meaning that higher *k*-cores are always contained in lower *k*-cores. Iterating the *k*-cores procedure is a way to detect the denser substructures of a network. In short, a *k*-core subgroup must have at least *k*+1 nodes and all nodes in that subgroup have a degree greater than or equal to *k*.

In contrast to the *k*-cores procedure, the *m*-slices procedure focuses on valued relations. Multiple or repeated ties between nodes are sometimes considered more important than the number of neighbors that directly connect to a node. The more often two nodes connect with one another, the stronger and closer are their interactions. It perhaps occurs in an intermetropolitan co-invention network where several major cities have repeated ties to each other. In short, an *m*-slices procedure is regarded as a chain of nodes connected by lines of

[5] This concept was introduced by John Scott (2000), who called it an *m-core*, but recent scholars renamed it the *m-slice* (Hanneman and Riddle 2005; Nooey et al. 2005; Knoke and Yang 2008).

a specified multiplicity (Scott 2000). For example, a *1*-slice subgroup contains all nodes that are connected with at least one tie. A *2*-slice subgroup contains nodes that are connected by at least two ties, and so forth. In an *m*-slice subgroup, all nodes tied with each other by ties greater than or equal to *m* are retained. Connections of nodes that are lower than *m* ties are disregarded. An *m*-slice subgroup must have at least *m*+1 nodes and all ties in that subgroup have a frequency of connection greater than or equal to *m*.

In summary, the analysis of nested components provides a powerful set of analytical tools for examining cohesive subgroups of metropolitan areas. Either by using the *k*-cores or *m*-slices procedure, the patterns of cohesive groupings underlying U.S. intermetropolitan networks of biotechnology co-invention reveal spaces with dense knowledge flows.

Investigate intermetropolitan network centrality and centralization

Two levels of network center measure – centrality and centralization – are calculated to identify the importance of metropolitan areas in transferring knowledge. Centrality is a local-level measure used to reveal the visibility of an individual node to other network nodes. Centralization is a global-level measure that assesses the extent to which a whole network has a centralized structure (Scott 2000). Both centrality and centralization are calculated and interpreted by three different perspectives: degree, closeness, and betweenness.

1. Degree-based centrality and centralization

Degree-based centrality describes the extent to which a node directly connects to other nodes. It is the simplest and most straightforward way to identify a network center (Wasserman and Faust 1994). The degree centrality for node *i*, $C_D(n_i)$, is the number of neighbors that directly connected to node *i*. It is defined as:

$$C_D(n_i) = \sum_{j=1}^{g} x_{ij}$$

where the subscript *D* refers to "degree-based," and $\sum_{j=1}^{g} x_{ij}$ counts the total number of direct ties that node *i* links to the remaining *g*-1 nodes ($j \neq i$). This measure is sensitive to the number of network nodes (i.e., the number of *g*), which makes cross-comparison difficult. Wasserman and Faust (1994) proposed a standardization modification by dividing $C_D(n_i)$ by the maximum number of possible connections to all *g*-1 other nodes. This standardized measure of degree centrality for node *i* is defined as:

$$C_D^{'}(n_i) = \frac{C_D(n_i)}{g-1}$$

where $C_D(n_i)$ is standardized by the maximum number of possible connections (*g*-1) to calculate the proportion of network neighbors that are adjacent to node *i*. As a proportion, the standardized degree centrality score ranges between zero and one. The minimum value of zero indicates that node *i* has no connection with other nodes. The maximum value of one occurs when node *i* directly ties to all *g*-1 other nodes. Wasserman and Faust (1994) argued that:

> [A]n actor with a large degree…should be recognized by others as a major channel of relational information…, occupying a central location. In contrast,...actors with low degrees are clearly peripheral in the network. Such

actors are not active in the relational process. In fact, if the actor is completely isolated (so that $d\ (n_i) = 0$), then removing this actor from the network has no effect on the ties that are present (p. 179-80).

In analyzing an intermetropolitan co-invention network, a metropolitan area with the highest degree centrality score is regarded as the most active place in transferring knowledge to other areas. The present analysis focuses on bilateral knowledge flows and does not distinguish between incoming flows (referred to as in-degree) and outgoing flows (referred to as out-degree) held by an individual metropolitan area.

Degree-based centralization assesses the extent to which an entire network has a centralized structure (Scott 2000). Many scholars adopt Freeman's (1979) network degree centralization, C_D, which reveals the variability of all nodes' degree centralities around the largest degree centrality. It is defined as:

$$C_D = \frac{\sum_{i=1}^{g}\left[C_D(n^*) - C_D(n_i)\right]}{\max \sum_{i=1}^{g}\left[C_D(n^*) - C_D(n_i)\right]}$$

where $C_D\,(n^*)$ denotes the largest degree centrality score observed in a network, and $C_D\ (n_i)$ refers to various degree centralities of all g-1 other nodes. The numerator sums the observed differences in degree centralities for a node with the largest value and every other node, while the denominator measures the theoretical maximum possible sum of these differences. Freeman (1979) proposed a simplified form as:

$$C_D = \frac{\sum_{i=1}^{g}\left[C_D(n^*) - C_D(n_i)\right]}{(g-1)(g-2)}$$

The theoretical maximum possible sum of the differences in the denominator of this equation occurs in a star graph where a central node connects to all other nodes and these nodes only link with the central node. The central node has the highest degree centrality of $C_D(n^*) = (g\text{-}1)$, while each of other nodes has the same degree centrality of $C_D(n_i) = 1$. The difference in degree centralities between the most central $C_D(n^*)$ and any other $C_D(n_i)$ is g-2. Because this difference occurs g-1 times (i.e., g-1 other nodes), the sum of these differences is (g-1) (g-2). The range is from zero and one. The minimum value of zero occurs in a regular graph where each node has the same degree. At the other extreme, the maximum value of one occurs in a star graph. In comparing intermetropolitan co-invention networks over time, a wide range of network degree centralization most likely occurs.

2. Closeness centrality and centralization

Closeness-based centrality assesses how a metropolitan area accesses knowledge flows through direct and indirectly connections. It is an inverse function of geodesic distances from a given node to all others (Freeman 1979; Wasserman and Faust 1994; Scott 2000; Knoke and Yang 2008). Geodesic distance is the smallest number of lines connecting any two distinct nodes. A metropolitan area is viewed as a closeness-based network center if it has the shortest geodesic distances to all other areas, meaning that this area is the most critical to facilitating knowledge flows that access to the entire network. The closeness centrality for node i, $C_C(n_i)$, is the inverse of the total geodesic distances from node i to g-1 other nodes. It is defined as:

$$C_C(n_i) = \left[\sum_{j=1}^{g} d(n_i, n_j) \right]^{-1}$$

where the subscript C refers to "closeness-based," and $\sum_{j=1}^{g} d(n_i, n_j)$ denotes the geodesic distance from node i to g-1 other nodes ($j \neq i$). Closeness centrality also depends on the number of network nodes. Wasserman and Faust (1994) proposed a standardized closeness centrality measures for node i as:

$$C_C^{'}(n_i) = \left[\sum_{j=1}^{g} d(n_i, n_j) \right]^{-1} (g-1) = (g-1) C_C(n_i)$$

where $C_C(n_i)$ is simply multiplied by the number of other g-1 nodes except node i itself. This standardized measure assesses the inverse of average geodesic distance between node i and its fellow nodes to determine the node's integration within the network. The range is from zero to one. A city with high closeness centrality score means that it can influence all other cities easily via shorter geodesic distances compared with cities that have lower scores.

Freeman's (1979) *network closeness centralization*, C_C, assesses the variability of all nodes' closeness centralities around the largest closeness centrality. It is defined as:

$$C_C = \frac{\sum_{i=1}^{g} \left[C_C^{'}(n^*) - C_C^{'}(n_i) \right]}{\max \sum_{i=1}^{g} \left[C_C^{'}(n^*) - C_C^{'}(n_i) \right]}$$

where $C'_C(n^*)$ denotes the largest closeness centrality after standardized, and $C'_C(n_i)$ refers to the closeness centralities of all g-1 other nodes. The numerator sums the observed differences in closeness centrality for a node with the largest value and every other node in the network, while the denominator measures the theoretical maximum possible sum of these differences. Freeman (1979) further modified the denominator of the equation above to a simplified form defined as:

$$C_C = \frac{\sum_{i=1}^{g}\left[C_C^{'}(n^*) - C_C^{'}(n_i)\right]}{[(g-2)(g-1)]/(2g-3)}$$

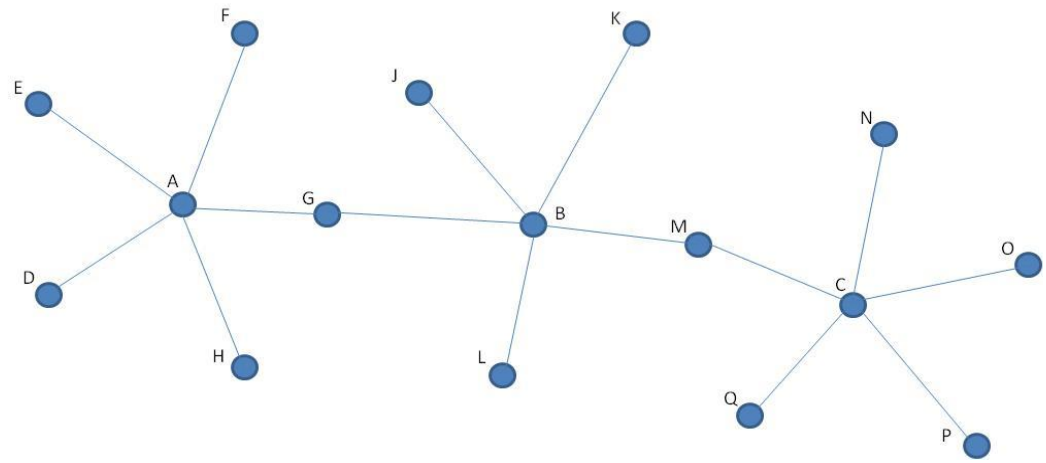
The denominator of the above equation reaches its maximum possible value in a star graph. The star node links to all g-1 other nodes with the same geodesic distance of one, while the g-1 nodes each have geodesic distance of two to the remaining g-2 nodes. The range is from zero to one. The minimum value of zero occurs in a regular graph, while the maximum value of one occurs in a star graph.

3. Betweenness centrality and centralization

Scott (2000) presented a conceptual betweenness-based network shown in Figure 4.2. Nodes G and M lie between three different star graphs viewed as two betweenness-based network centers in controlling knowledge flows.[6] Betweenness-based centrality assesses the extent to which a node's position falls on the geodesic paths as a go-between among other network nodes. It determines whether a node plays a relatively important role as a "broker" or "gatekeeper" of knowledge flows with a high potential of control on the indirect relations of other nodes. A geodesic path between two nodes is the shortest and most efficient

[6] The three star graphs shown in Figure 4.2 are {A, D, E, F, G, H}, {B, G, J, K, L, M}, and {C, M, N, O, P, Q}.

channel for transferring knowledge. When there is more than one geodesic path linking two nodes, all paths with the same number of lines are equally selected.



Note. From Social Network Analysis: A Handbook, by J. Scott, 2000, p. 84.

Figure 4.2 Conceptual betweenness-based network

The betweenness centrality for node *i*, $C_B(n_i)$, is the sum of the proportions of node *i* located on geodesic paths between any other pair of nodes. It is defined as:

$$C_B(n_i) = \sum_{j<k} [g_{jk}(n_i) / g_{jk}]$$

where the subscript *B* refers to "betweenness-based," g_{jk} represents the number of geodesic paths between nodes *j* and *k* , and $g_{jk}(n_i)$ denotes the number of geodesic paths between nodes *j* and *k* that contain node *i* ($i \neq j \neq k$). This index has the minimum value of zero when node *i* falls on no geodesic path for all pairs of *g*-1 nodes. It reaches a maximum possible value of (*g*-1) (*g*-2)/2 if node *i* appears on every geodesic path for all pairs of g-1 nodes.[7] Wasserman and Faust (1994) suggested a standardized betweenness centrality to remove the size effect, which is defined as:

$$C_B^{'}(n_i) = C_B(n_i)/[(g-1)(g-2)/2]$$

As a proportion, the range is also from zero to one. The closer the standardized betweenness centrality is one, the more likely a given metropolitan area controls or mediates knowledge flows in the network.

Wasserman and Faust's (1994) *network betweenness centralization*, C_B, is the variability of all nodes' betweenness centralities around the largest betweenness centrality, which is defined as:

$$C_B = \frac{\sum_{i=1}^{g}\left[C_B(n^*) - C_B(n_i)\right]}{\max \sum_{l=1}^{g}\left[C_B(n^*) - C_B(n_i)\right]}$$

where $C_B(n^*)$ denotes the largest betweenness centrality observed in a network. The numerator sums the differences in betweenness centrality for a node with the largest value and every other node, while the denominator measures the theoretical maximum possible variation of these differences. Actor betweenness centrality attains its theoretical maximum at (g-1) (g-2)/2 in a star graph. Because this maximum value occurs at least g-1 times, the denominator can be simplified as $(g\text{-}1)^2(g\text{-}2)/2$. The measure of simplified network betweenness centralization is defined as:

[7] Knoke and Yang (2008) explain that the total number of geodesic paths among g-1 nodes (excluding node *i*) is $C_{g-1}^{2} = \frac{(g-1)!}{2!(g-1-2)!} = \frac{(g-1)!}{2!(g-3)!} = \frac{(g-1)(g-2)}{2}$, assuming that each pair has only one geodesic path.

$$C_B = \frac{\sum_{i=1}^{g}\left[C_B(n^*) - C_B(n_i)\right]}{(g-1)^2(g-2)/2}$$

Network betweenness centralization reaches the maximum value of one when a single dominant node sits on all geodesic paths in a star graph. Its minimum value of zero occurs when every node has the same betweenness centrality score in a regular graph. The closer a network betweenness centralization approaches one, the more unequally distributed is betweenness-based centralized in a network.

Procedures of finding network positions in the co-invention network-based system

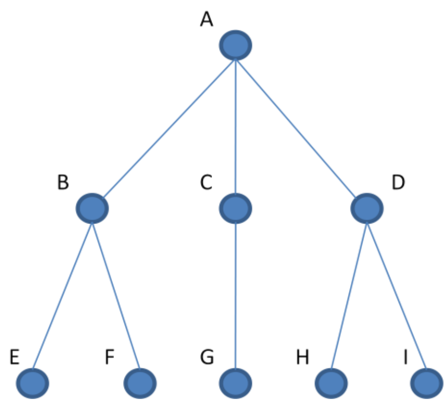
This part focuses on similarities between metropolitan areas based on their network positions. A network position refers to a set of nodes that have a similar pattern of relations to the rest of the network (Wasserman and Faust 1994). Scott (2000) argued that the underlying structure of a network is more apparent in the relations of positions than among individual nodes themselves. Understanding how network positions of metropolitan areas form provides an empirical glimpse into the U.S. network-based system of knowledge flows and uncovers whether the system is consistent with a relational core/periphery structure that Borgatti and Everett (1999) originally characterized. The research questions addressed include: Do some U.S. metropolitan areas have similar network positions? Do the varying network positions of metropolitan areas reveal a hierarchical cluster structure? What are the relationships among these positions?

This part first defines and illustrates the concept of equivalence. It is followed by an introduction of methods used to measure degrees of similarity of metropolitan areas. Finally, this part outlines approaches used to identify patterns of similarity and simplification.

1. Define and illustrate the concept of equivalence

In social network analysis, nodes that have similar patterns of ties constitute an equivalent position in a network (Nooy et al. 2005). Three types of equivalence (structural, automorphic, and regular) define nodes as "equivalent" in terms of their relations with others. Figure 4.3 illustrates the basic ideas of these three types of equivalence by using a non-directional, binary example.

Structural equivalence is the most rigorous form of equivalence, where two nodes are perfectly equivalent if they have identical ties with the same other nodes (Knoke and Yang 2008). Figure 4.3 shows nine nodes that are divided into seven structural equivalent positions: {A}, {B}, {C}, {D}, {E, F}, {H, I}, and {G}. Nodes A, B, C, D, and G form distinctive equivalent positions because each of them has a unique tie to one of the other nodes. Nodes E and F fall in the same structural equivalent position because both nodes have identical ties to node B. Structural equivalence is the most widely used criterion of equivalence for the analysis of network position particularly in studying competitive relationships among nodes (Hanneman and Riddle 2005). However, it is very unusual to find nodes that are perfectly structurally equivalent in a real network system (Scott 2000). Some authors argue that less-restrictive equivalence criteria might be more appropriate in the study of large and complicated social networks (Wasserman and Faust 1994; Knoke and Yang 2008).



Note. Modified from Social Network Analysis, by S. Wasserman and K. Faust, 1994, p. 468.

Figure 4.3 Conceptual graph of equivalence

Automorphic equivalence loosens the requirement of structural equivalence by considering that nodes are located in the same position if they have identical patterns and numbers of ties with others, but are not necessarily exactly tied with the same other nodes (Knoke and Yang 2008). Figure 4.3 reveals five different automorphic equivalent positions: {A}, {B, D}, {C}, {E, F, H, I}, and {G}. Nodes B and D occupy the same equivalent position because they have the same patterns of ties to both the position {A} and the position {E, F, H, I}. In general, the criterion of automorphic equivalence argues that nodes within the same position can be replaced with each other without modification of the overall relational structure. This type of equivalence focuses on sets of nodes that are substitutable having similar relations with other sets of nodes (Hanneman and Riddle 2005).

Regular equivalence is the least restrictive criterion for partitioning individual nodes into positions. It does not require nodes to have identical ties to the same other nodes based on the structural equivalence criterion, or to be substitutable for each other based on the automorphic equivalence criterion. Nodes are regularly equivalent if they have ties to other nodes that are also regularly equivalent (Wasserman and Faust 1994; Knoke and Yang 2008). If a node occupying the first position has a relation with a node in the second position, then the other regularly equivalent nodes in the first position must also have relations with other nodes in the second position (White and Reitz 1983). The conceptual graph in Figure 4.3 reveals three different regular equivalent positions: {A}, {B, C, D}, and {E, F, G, H, I}. Nodes E, F, G, H, and I are regularly equivalent because they have no tie with any node in the position {A}, but all have a tie to the position {B, C, D}. In short, regular equivalent nodes need to have the same types of relationships with nodes in other regular equivalent positions, but are not necessarily tied to the same others (Scott 2000; Knoke and Yang 2008). This dissertation uses the regular equivalence criterion to partition individual metropolitan areas into network positions.

2. Measure degrees of regular equivalence for pairs of metropolitan areas

The method for identifying regular equivalent positions consists of two key procedures: (1) measuring degrees of regular equivalence for pairs of metropolitan areas (as discussed below), and (2) identifying patterns of similarity and simplification established within a system (as discussed in the next subsection). White and Reitz's (1985) regular graph equivalence (REGE) algorithm is used to estimate degrees of regular equivalence for pairs of metropolitan areas (also Borgatti and Everett 1993; Wasserman and Faust 1994). Given a focal pair nodes *i* and *j*, the algorithm is defined as:

$$M_{ij}^{t+1} = \frac{\sum_{k=1}^{g} \max_{m=1}^{g} \sum_{r=1}^{R} M_{km}^{t} ({}_{ijr}M_{kmr}^{t} + {}_{jir}M_{kmr}^{t})}{\sum_{k=1}^{g} \max_{m=1}^{*} \sum_{r=1}^{R} ({}_{ijr}Max_{kmr} + {}_{jir}Max_{kmr})}$$

where M_{ij}^{t+1} is an estimate of the degree of regular equivalence for nodes *i* and *j* at iteration *t*+1. This measure is a function of how well node *I*'s ties to all other nodes are matched by node *j*'s ties to all other nodes and *vice versa*. The term ${}_{ijr}M_{kmr}$ estimates how well node *i*'s ties with node *k* match the profile of node *j*'s ties to node *m*. Since nodes *k* and *m* might not be perfectly regular equivalent, M_{km}^{t} is the estimated regular equivalence of *k* and *m* from the previous iteration. The numerator calculates the *best* matching set of ties between node *i*'s ties with its neighborhood and node *j*'s ties with its neighborhood and weighted by M_{km}^{t}. The denominator is "the maximum possible equivalence, which would occur if every tie from node *i* to its neighborhood could be perfectly matched by a tie from node *j* to its neighborhood, and the two neighborhoods were perfectly equivalent" (Mahutga and Smith 2011, p. 260).

The REGE algorithm is an iterative process by initially setting all estimates of pair-wise regular equivalence at the highest value of one. All of node *i*'s ties to its neighborhood and all of node *j*'s ties to its neighborhood are perfectly “matched” and all of neighboring nodes of this pair are regularly equivalent. This scenario leads to the similarity score for this pair of one, which occurs when the numerator is equal to the maximum possible score of the denominator. For each iteration round, the algorithm re-calculates the estimated degrees of regular equivalence that are weighted by the previous iteration’s equivalence between the matched neighbors. Since the estimated values get smaller in each iteration round, the similarity scores decrease in successive iterations (Borgatti and Everett 1993). The iterative procedure is continued until the revised estimate of regular equivalence makes little difference to previous estimates (Scott 2000). A three-time iteration REGE algorithm is suggested in the literature and is set as the default in the UCINET software package (Faust 1988; Borgatti and Everett 1993; Wasserman and Faust 1994; Scott 2000; Hanneman and Riddle 2005).

3. Identify patterns of similarity and simplification in the system

Having computed a regularly equivalent matrix containing similarity scores for all pairs of metropolitan areas, the focal point here is to identify patterns of similarity and simplification in a network-based system. Performing a hierarchical clustering is a common way to partition metropolitan areas into equivalent positions. Each metropolitan area is treated initially as a singleton position. The most similar areas with the highest degree of regular equivalence are successively joined until all areas are merged into a single all-inclusive position. During the joining process, a threshold value of α is selected as a ceiling to determine which areas should be joined at a particular position (Knoke and

Yang 2008). In the case that d_{ij} is the similarity score of regular equivalence between nodes *i* and *j*, both nodes occupy the same position only if $d_{ij} \geq \alpha$. Nodes within the same position have higher similarity scores compared with nodes located in different positions. A tree diagram (also called dendrogram) depicts the hierarchical results (Alderson and Beckfield 2004).

Positions formed in a dendrogram depend upon the choice of clustering methods. Generally, there are three methods for merging positions. The *single-link method* (also titled the nearest neighbor method) uses the minimum similarity between nodes in different positions. The *complete-link method* (also titled the farthest neighbor method) uses the maximum similarity. Another variation that uses the average similarity between nodes in different positions is titled the *average-link method*. Scott (2000) argued that the single-link method tends to "chain" new nodes into existing positions. Knoke and Yang (2008) pointed out that the complete-link method is more likely to produce large numbers of homogeneous and tightly bounded positions but the probability of chaining is low. Computer programs for hierarchical clustering are widely available in both the standard statistical analysis (e.g. SPSS and SAS) and the UCINET social network analysis packages.

4.3 Summary

This chapter outlines the methodology used in these two pivotal research topics of the dissertation. The first topic asks whether the U.S. biotechnology co-invention urban system reveals significant differences between spatial and network-based dependencies. The longitudinal changes in these dependencies are also explored. I identify differences in spatial and network-based dependencies across the U.S. urban system by comparing patterns revealed in global- and local-level measures of association. Moran's *I* is used to detect global-level spatial and network-based dependencies, while the local-level measure of dependence is based on the *local indicators of spatial association* (also called LISA).

The second topic is to understand how and to what extent biotechnology flows circulate in network-based systems by focusing on the structural properties of intermetropolitan co-invention networks from three distinct perspectives: components and cohesive subgroups of metropolitan areas, intermetropolitan network centralization and centrality, and positions established within the co-invention network-based system. First, the characteristics of network components reveal patterns of connection among metropolitan areas. Cohesive groupings underlying U.S. intermetropolitan networks of biotechnology co-invention reveal spaces with dense knowledge flows. Second, the center of a network identifies metropolitan areas with the best access to knowledge flows. Two levels of network center measure – centrality and centralization – are calculated to identify the importance of metropolitan areas in transferring knowledge. Both centralization and centrality are calculated and interpreted by three different perspectives: degree, closeness, and betweenness. Third, network positions of metropolitan areas show the kind of co-invention network-based system these member areas form and the roles played by different types of areas within the system of knowledge exchange. The regular equivalence criterion is used to partition individual metropolitan areas into network positions, which enables to determine whether the U.S. co-invention network-based system is consistent with a relational core/periphery structure. The method for identifying regularly equivalent positions uses two key procedures: (1) the UCINET social network analysis package to estimate degrees of regular equivalence for pairs of areas, and (2) hierarchical clustering to identify patterns of similarity and simplification in the system.

Chapter 5

DATA

This chapter describes the data and presents preliminary descriptive tabulations. Section 5.1 briefly overviews the development of biotechnology, and outlines biotechnology co-patent data used as a proxy of co-invention for measuring intermetropolitan knowledge flows. Section 5.2 defines the choice of geographical units of analysis. Section 5.3 discusses the process of allocating co-patent data to geography. Section 5.4 presents several geographical structures of U.S. biotechnology co-patenting.

5.1 Biotechnology and Co-Patent Data in Biotechnology

The United Nations Convention on Biological Diversity defines biotechnology as "any technological application that uses biological systems, living organisms, or derivatives thereof, to make or modify products or processes for specific use" (United Nations, 1993). Under this definition, biotechnology includes a diverse collection of technologies that manipulate cellular, or molecular components in living things to make products, discover new knowledge, or modify plants, animals, and microorganisms to carry desired trails (U.S. Department of Commerce, 2003). Firms involved in biotechnology are not separately classified as a single industry within the North American Industry Classification System (NAICS) or its predecessor, the Standard Industrial Classification (SIC). Instead, more than 60 four-digit NAICS categories are engaged in biotechnology-related activities (U.S. Department of Commerce, 2003). Despite this breadth of industries involved, most biotechnology-related firms are assigned in either NAICS 54171 (research and development in the physical, engineering, and life sciences) or NAICS 32541 (pharmaceutical and medicine manufacturing) (Cortright and Mayer 2002).

Advances in the U.S. biotechnology were marked by the establishment of the first biotechnology company, Genentech, in 1977 in San Francisco in developing commercially useful products such as drugs. In 1980, the U.S. Supreme Court provided an important incentive for the development of biotechnology companies by ruling that biological materials could be patented. Thus, "private companies could look forward to making substantial profits from therapies and products that they developed through genetic engineering" (Wasserman, 2009). The biotechnology industry has registered a fast growth since the 1990s. According to a study by the Biotechnology Industry Organization (BIO) released in 2010, total employment in the U.S. biotechnology-related sectors added more than 193,000 jobs or 15.8 percent from 2001 to 2008. This rapid rate of job growth was about 4.5 times as much as the overall growth rate for the national private sector (3.5 percent). Average wage growth in biotechnology has increased by 10.1 percent since 2001, compared with 3.2 percent for the overall private sector.

This study explores network structures of biotechnology knowledge flows by examining patent data on co-inventorships (e.g., Cantner and Graf 2006; Ejermo and Karlsson 2006; Fleming et al. 2007). Patent co-inventorship refers to the situation where a patent either is invented by more than one individual, or lists more than one individual as a designed inventor (Breschi and Lissoni 2004).

According to the U.S. Patent Act, a co-patent requires some level of "joint manner" between co-inventors, but is not necessary for the co-inventors to physically work together at the time of invention (Title 35 of the United States Code). Co-patent data are useful in assessing knowledge flows for two reasons: (1) it tabulates knowledge production and exchange occurring in geography, and (2) it provides a way to assess inter-territorial (e.g., intermetropolitan, interregional, or international) knowledge flows. In most of the U.S. Patent and Trademark Office's (USPTO) compilation of areal patent data (counties, metropolitan areas, states), the geographical origin of a patent is indicated by the place of residence of the first-named inventor at the time of application. This approach potentially underestimates (or overestimates) biotechnology invention in some areas because non-local co-patenting is increasingly common (see Section 5.4 for details). This study allocates co-patent data to geography by explicitly attributing co-inventors' contributions to their residential locations (see Section 5.3 for details). Generally, non-local co-patenting occurs in two forms. First, co-inventors may affiliate with the same corporation but are located in different areas. In other cases, co-inventors affiliated with different firms or research institutions from different locations collaborate and share knowledge with each other (Frietsch and Jung 2009). Both forms of co-patenting represent intentional knowledge flows circulating across space. Information on the quality of patents is not tackled here because the focus is on the role of network-based proximity between inventors (Cantner and Graf 2006). Frivolous patents are less likely in biotechnology and there is no evidence that a meaningful geography of patent frivolity or superiority exists across metropolitan areas.

The co-patent data were obtained from two data sources: (1) the U.S. Patent and Trademark Office (USPTO) database, and (2) the patent citations data package in the National Bureau of Economic Research (NBER) (U.S. Patent and Trademark Office 2010; Hall et al. 2001). The USPTO issues patents by technological categories, but there are no patent classes or subclasses for biotechnology *per se*. The realm of biotechnology-related patents extends to several classes of the U.S. patent classification (USPC) system. A broad interpretation of biotechnology invention was made by including patents awarded in the USPC classes 424, 435, 514, and 800 (Hall et al. 2001; Cortright and Mayer 2002; Hevesi and Bleiwas 2005).[8] Classes 424 and 514 are drugs, particularly bio-affecting and body treating compositions. Class 435 is a chemical class and includes molecular biology and microbiology. Class 800 encompasses multicellular living organisms, unmodified parts thereof, and related processes (U.S. Patent and Trademark Office 2011).

The following procedure was used to assembly the biotechnology co-patent data investigated. Biotechnology patents in the four USPC classes were first extracted from the USPTO and NBER databases. Patents awarded in the year 1979, 1989, 1999, and 2009 were subsequently identified. Each patent must

[8] Hall et al. (2001) provide an aggregated classification scheme for main technological fields identifying the USPC classes 435 and 800 as "biotechnology." Cortright and Mayer (2002), and Hevesi and Bleiwas (2005) argue that the USPC classes 424, 435, 514, and 800 are likely to encompass most of the patented biotechnology inventions. Although these classes are not a complete list of biotechnology-related patents, this study takes a broad collection of data by including biotechnology patents in the USPC classes 424, 435, 514, and 800.

have at least one inventor geographically located in one U.S. metropolitan area. The data were split into two groups. Patents invented by *multiple* inventors were distinguished from *solo* inventors. As shown in Table 5.1, co-patents were a large and growing proportion of all American biotechnology patents. The four selected USPC classes accounted for more than 5,000 biotechnology co-patents awarded in 1999 and in 2009. Only 484 biotechnology co-patents were awarded in 1979. The percentage of co-patenting increased from 55 to 81 percent of all biotechnology patents in the period 1979-2009. Figure 5.1 shows that the average team size of co-patenting in American biotechnology has steadily increased – from 2.54 in 1979 to 4.07 in 2009. Knowledge production and exchange among U.S. biotechnology inventors have clearly increased since 1979.

5.2 Geographical Units of Analysis

The choice of geographical units of analysis affects research results and implications. Scholars in U.S.-related studies often rely on jurisdictional units (e.g., states, counties, cities, and towns), as data are often tabulated by government agencies using geographical boundaries (Ratanawaraha and Polenske 2007). Most of these boundaries are arbitrary and rarely represent a well-integrated local economy. The geographical units of analysis chosen for this study are U.S. census-defined Metropolitan Statistical Areas (MSAs). Metropolitan statistical areas are collections of counties that constitute integrated labor markets. Acs and Armington (2006) argued that "these geographical units do a better job of ensuring that people both live and work within their boundaries" (p. 10). Using U.S. metropolitan statistical areas as units in locational analysis

Table 5.1 Number of U.S. biotechnology patents

	1979	1989	1999	2009
Co-patents	484	1596	5870	5125
Solo patents	399	786	1763	1231
Total patents	883	2382	7633	6356
Percentage of co-patenting *	54.81%	67.00%	76.90%	80.63%

* Percentage of co-patenting = (co-patents / total patents) $\times$ 100%

Source: The NBER patent database (Hall et al. 2001) and The USPTO patent database

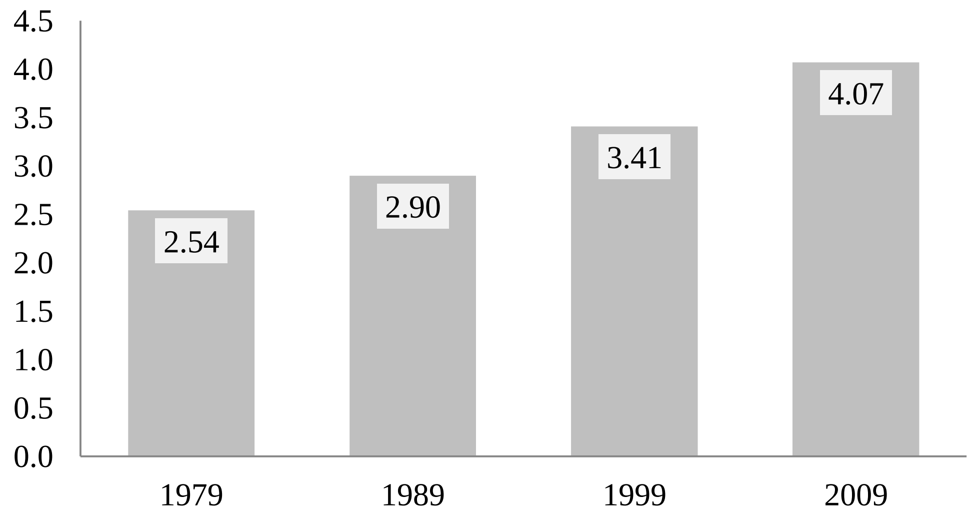


Figure 5.1 Average team size of biotechnology co-patenting

of patents is widely accepted (e.g., Jaffe at al. 1993; Ó hUallacháin 1999; Acs et al. 2002; Cortright and Mayer 2002; Bettencourt et al. 2007; Lobo and Strumsky 2008). The 1999 definition of metropolitan statistical areas was selected for all years of analysis. Using the same metropolitan definition – assignment of counties to areas – is essential in longitudinal analysis and 1999 sets a common standard for the period 1979 to 2009. Based on the 1999 definition, 275 MSAs (including 17 consolidated metropolitan statistical areas) were identified in the continental U.S. The analysis requires a stable set of areas across the decades and 1999 is a useful intermediate year. However, co-inventive activities are unevenly distributed across space with the result that some small metropolitan areas without any co-patented awards may be identified as outliers in spatial and network properties. The number of metropolitan observations was reduced to 150 by focusing on those areas having at least one biotechnology co-patent awarded in 1999. This avoids swamping the analysis with cases that have no co-patenting.

5.3 Process of Allocating Co-Patent Data to Geography

The co-patent data focus on the share of co-invention contributed by inventors from different locations. Evidence suggests that the volume of non-local co-patenting has steadily increased. Multinational corporations are driving this process as they draw on collaboration of inventors from various locations. Improvements in information and communication technologies also facilitate knowledge flows over space. Co-patents that include inventors that live in more than one areas are allocated fractionally to each location (e.g., Felix 2006; Ejermo and Karlsson 2006; Maggioni et al. 2007; Maraut et al. 2008). This is achieved in a two-step process. First, an inventor's residential location is matched to his or her hometown MSA.[9] Second, the *fractional counting method* used recognizes the respective contribution of each metropolitan area where the co-inventors of a patent live and avoids multiple counting of any co-invented patent (Felix 2006). If a co-patent has four inventors located in four different metropolitan areas, for example, one-quarter of the patent is allotted to each of the four metropolitan areas. If the four inventors are evenly located in two different metropolitan areas, half of the patent is allocated to each area.

An example of allocating co-patent data to geography is shown in Table 5.2. The first patent (#6009450) is co-invented by six inventors, while the second patent (#6009451) by four inventors. In the first step (as shown in the upper part of the table), each inventor's hometown is matched to a corresponding metropolitan area. Six inventors in the first patent are located in three metropolitan areas, and four inventors in the second patent are evenly located in four different metropolitan areas. In the second step (as shown in the lower part of the table), the fractional counting method is used to allocate the co-patent to each of the inventors' corresponding metropolitan areas. The first patent is proportionally allocated one-sixth of the patent to Boston, one-half to Phoenix, and one-third to San Francisco because there are one, three, and two out of six inventors located in these areas, respectively. The second patent is allocated one-

[9] In the original NBER patent database, an inventor's residential information is shown on the patent application including nationality, state, municipality, and zip code. Generally, it is much easier to identify a MSA on a basis of a zip code rather than of a name of municipality. However, the Zip codes of inventors are often missing in the NBER database, so I develop an approach to identify MSAs based on the names of municipalities.

Table 5.2 Example of allocating co-patent data to geography

Step 1 – match a MSA to each inventor's hometown

Patent ID	Last name	First name	Middle name	Municipality	State	Inventorship order	MSA
Patent 1							
600945	Dworkin	James	Douglas	Chandler	AZ	1	Phoenix
600945	Glaser	P.	Michael	Tempe	AZ	2	Phoenix
600945	Torla	Michael	John	Chandler	AZ	3	Phoenix
600945	Vadekar	Ashok		Sunnyvale	CA	4	S. F.
600945	Lambert	Robert	John	Mountain	CA	5	S. F.
600945	Vanston	Scott	Alexande	Marlboro	MA	6	Boston
Patent 2							
600945	Haq	Samuel	F.	Chandler	AZ	1	Phoenix
600945	Malone	Patrick	F.	Sunnyvale	CA	2	S.F.
600945	Fortney	John	H.	Portland	OR	3	Portland
600945	Varner	Donald	P.	Scotch Plains	NY	4	N.Y.

Step 2 – proportionally allocate each co-patent to the inventors' corresponding MSAs

Patent	Boston	N.Y.	Philadelphia	Phoenix	Portlan	Sacramento	S.F.	Seattle
600945	1/6	0	0	3/6	0	0	2/6	0
600945	0	1/4	0	1/4	1/4	0	1/4	0
…	…	…	…	…	…	…	…	…
Total	1/6	1/4	0	3/4	1/4	0	7/12	0

Table 5.3 Geographical structures of biotechnology co-patenting (at the MSA scale)

	1979	1989	1999	2009
Intra-metropolitan flows Co-patenting in the same MSA	83.2%	79.0%	73.7%	62.3%
Intermetropolitan flows	16.8%	21.0%	26.3%	37.7%
Co-patenting across 2 MSAs	16.3%	18.6%	22.4%	30.0%
Co-patenting across 3 MSAs	0.5%	1.8%	3.2%	6.7%
Co-patenting across 4 MSAs	0	0.6%	0.6%	0.8%
Co-patenting across 5 MSAs (and above)	0	0	0.1%	0.2%

Source: The NBER patent database (Hall et al. 2001) and The USPTO patent database quarter of the patent to each of the four metropolitan areas. After all of the patents have been proportionally and geographically allocated, the fractional contributions of each column are summed to obtain the co-patent counts of each metropolitan area.

5.4 Geographical Structures of U.S. Biotechnology Co-Patenting

Table 5.3 shows the geographical structures of U.S. biotechnology co-patenting at the MSA scale. While co-patenting was mainly localized within the same metropolitan area (referred to as *intra-metropolitan flows*), the overall pattern clearly shows that the share of non-local co-patenting (referred to as *intermetropolitan flows* – links between inventors who co-invent a patent across metropolitan boundaries) increased over the decades, from 17 percent in 1979 to 38 percent in 2009. Moreover, a growing number of metropolitan areas jointly participated in co-patenting activities, leading to a wider geography of cooperation. In 1979, co-patenting activity never exceeded three metropolitan areas. In 1989, the maximum co-patenting breadth extended to four metropolitan areas. In 1999 and 2009, more and more co-patents tied inventors of five metropolitan areas and above. This widening geographical span of co-invention confirms the broadening of geographical co-operative knowledge flows.

Co-patenting varies across metropolitan areas.

Table 5.4 shows biotechnology co-patent counts in the top 30 MSAs in the four years of analysis. Biotechnology co-patenting rates among U.S. metropolitan areas are discussed in the next chapter. With nearly 700 biotechnology co-patents in 2009, San Francisco stood out as the most prolific co-invention center in the

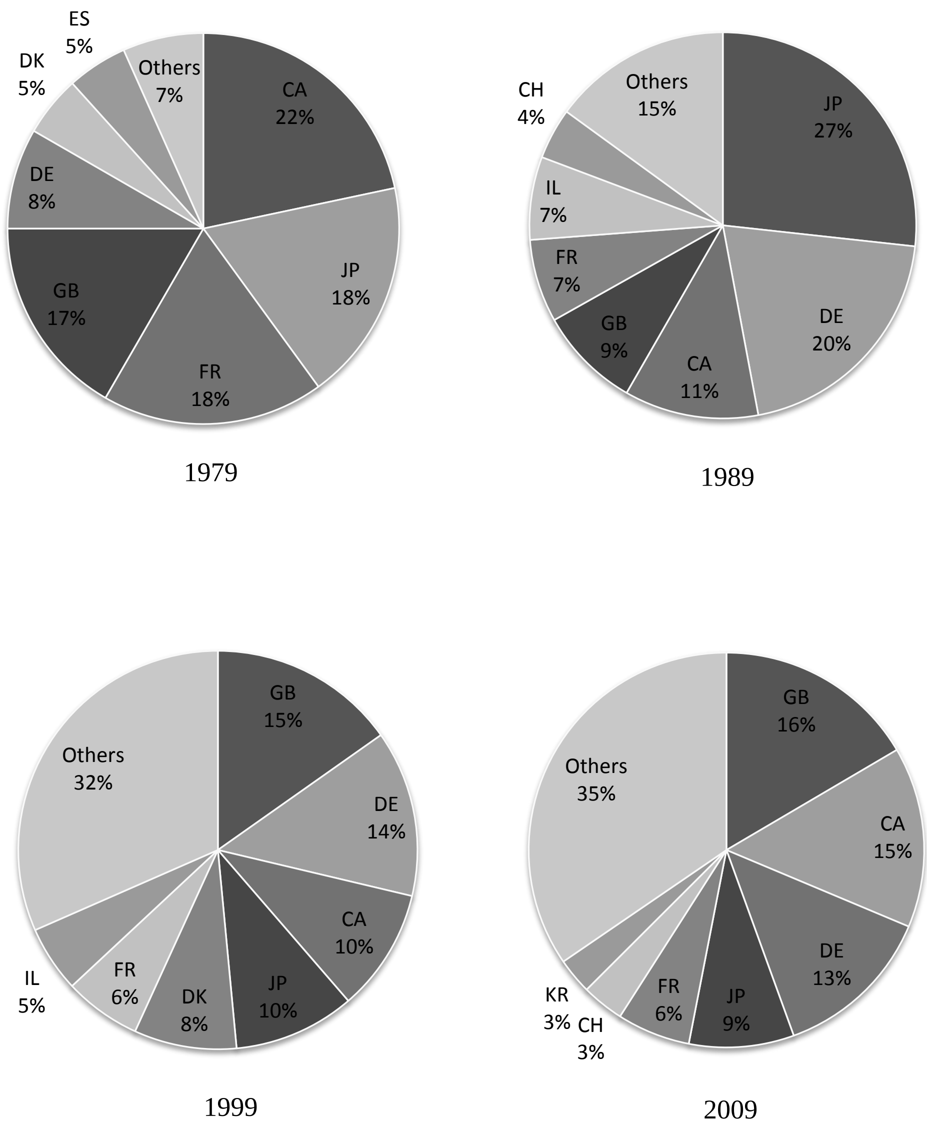
Table 5.4 Biotechnology co-patenting in the top 30 U.S. MSAs

	Metropolitan Area (ranked by 2009 data)	1979	1989	1999	2009
1	San Francisco—Oakland—San Jose, CA CMSA	23.2	146.5	753.8	692.0
2	New York—Northern New Jersey—Long Island, NY-NJ-CT-PA CMSA	142.3	372.1	696.2	542.6
3	Boston—Worcester—Lawrence, MA-NH-ME-CT CMSA	8.8	68.3	486.2	445.2
4	Philadelphia—Wilmington—Atlantic City, PA-NJ-DE-MD CMSA	65.8	146.4	446.9	366.4
5	San Diego, CA MSA	4.0	29.7	276.6	273.2
6	Washington D.C.—Baltimore, DC-MD-VA-WV CMSA	11.5	50.5	359.8	250.8
7	Los Angeles—Riverside—Orange County, CA CMSA	3.3	29.1	146.6	203.1
8	Seattle—Tacoma—Bremerton, WA CMSA	0.0	25.2	110.9	144.9
9	Raleigh—Durham—Chapel Hill, NC MSA	3.8	26.2	107.2	128.0
10	Chicago—Gary—Kenosha, IL-IN-WI CMSA	17.1	59.6	150.8	105.2
11	Madison, WI MSA	1.0	11.1	63.8	85.7
12	Denver—Boulder—Greeley, CO CMSA	0.5	2.8	69.7	85.4
13	Des Moines, IA MSA	1.3	1.2	56.7	76.3
14	St. Louis, MO-IL MSA	8.8	13.9	77.5	73.3
15	New Haven—Bridgeport—Stamford CT, NECMA	3.3	19.2	74.8	60.0
16	Minneapolis—St. Paul, MN-WI MSA	21.1	17.0	46.1	53.6
17	Indianapolis, IN MSA	20.5	49.8	112.5	50.4
18	Atlanta, GA MSA	1.0	6.4	29.1	47.2
19	Houston--Galveston--Brazoria, TX CMSA	3.0	28.3	82.0	45.0
20	Dallas—Fort Worth, TX CMSA	1.8	9.1	43.4	41.1
21	Detroit—Ann Arbor—Flint, MI CMSA	5.6	36.4	76.4	39.1
22	Rochester, NY MSA	7.7	26.5	26.3	23.3
23	Sacramento—Yolo, CA MSA	0.0	10.9	57.9	23.1
24	Hartford, CT MSA	0.3	6.3	23.3	22.6
25	Salt Lake City—Ogden, UT MSA	0.0	6.0	33.3	20.6
26	Portland—Salem, OR-WA CMSA	1.5	2.2	16.0	20.5
27	Cincinnati—Hamilton, OH-KY-IN CMSA	9.8	38.2	134.0	19.4
28	Pittsburgh, PA MSA	0.7	2.9	26.2	19.2
29	New London—Norwich, CT-RI MSA	6.5	18.8	37.3	19.0
30	Rochester, MN MSA	0.0	0.6	11.4	48.2
	Average of the top 30 MSAs	12.5	42.0	154.4	133.1

Source: The NBER patent database (Hall et al. 2001) and The USPTO patent database

U.S., followed by New York with 542.6 co-patents, Boston with 445.2 co-patents, and Philadelphia with 366.4 co-patents. These four major centers accounted for more than 44 percent of the American biotechnology co-patenting. San Francisco and Boston established themselves as the leaders in biotechnology invention by the 1970s and continued to sustain their first-mover advantages (Cortright and Mayer 2002). New York and Philadelphia are the traditional centers of the pharmaceutical industry, particularly along the Northeast Corridor where many leading companies are headquartered. Inventors in San Diego (273.2 co-patents) and Washington D.C.-Baltimore (250.8 co-patents) have significantly increased their biotechnology co-patenting. San Diego is well known for its rapid growth of biological research over the past several years and particularly successful in "securing venture capital and research contracts with pharmaceutical firms" (Cortright and Mayer 2002, p. 14). The Washington D.C.-Baltimore metro area is home to several important research centers that facilitate biotechnology invention, including the National Institutes of Health (NIH) and the Biotechnology Industry Organization (BIO). Los Angeles (203.1 co-patents), Seattle (144.9 co-patents), Raleigh-Durham (128 co-patents), and Chicago (105.2 co-patents) rounded out the top ten list of leading biotechnology co-patenting centers. A few small metropolitan areas were also ranked in the top 30 for biotechnology co-patenting including New Haven (Connecticut), Rochester (New York), Salt Lake City (Utah), and New London (Connecticut). The remaining metropolitan areas had some co-patenting activity but their levels were below the average (133.1 co-patents) of the top 30 metropolitan areas in the sample.

The disaggregated co-patent data also show that American biotechnology heavily relies on global sources of knowledge. Co-invention spans national and international territories. Figure 5.2 shows foreign-based inventors share of U.S. biotechnology co-patents. International co-patenting mostly occurs with inventors located in large developed countries, particularly Canada, Japan, United Kingdom, Germany, and France. However, this emphasis is being gradually attenuated by inventors from Switzerland and South Korea. The present analysis does not investigate international collaboration.



CA: Canada, CH: Switzerland, DE: Germany, DK: Denmark, ES: Spain,
FR: France, GB: United Kingdom, IL: Israel, JP: Japan, KR: South Korea

Figure 5.2 Share of geographical origin of foreign-based inventors participating in U.S. biotechnology co-patents

Chapter 6

RESULTS OF SPATIAL AND NETWORK-BASED DEPENDENCIES

This chapter presents results showing differences in spatial and network-based dependencies of biotechnology co-invention across U.S. metropolitan areas. The focus here is to compare patterns of metropolitan co-patenting rates revealed in global- and local-level measures of association. Each metropolitan area's co-patenting rate is calculated by dividing its annual biotechnology co-patent counts by the number of wage and salary jobs. This ratio is multiplied by a scaling factor of 1,000.

This chapter is organized as follows. Section 6.1 describes the spatial distribution of biotechnology co-patenting rates. Section 6.2 discusses the results of global association using Moran's *I* statistics. Section 6.3 interprets local associations generated by LISA cluster maps. Section 6.4 provides a summary of these results.

6.1 Spatial Distribution of U.S. Biotechnology Co-Patenting Rates

To compensate for small MSAs' biotechnology co-patenting rates with few inventors and rare co-patenting events that may be spuriously identified as "outliers," the original co-patenting rates are smoothed using an Empirical Bayes Smoother (Anselin et al. 2006a). Anselin et al. (2006a) argued that the Empirical Bayes Smoother is referred to as "shrinkage in the sense that the crude rate is moved (shrunk) towards an overall mean, as an inverse function of the inherent variance" (p. 39). Note that smoothed data could be potentially misinterpreted especially for those areas with small or zero co-patenting rates. I only focus on

150 large U.S. metropolitan areas and compare both the original and the smoothed co-patenting rates in global and local levels to identify whether any problematic areas or errors occur. As expected, the relationships between metropolitan co-inventive activities using the original and the smoothed co-patenting rates were statistically significant and strongly correlated in 2009 ($r = 0.998$, $p < 0.001$) and in 1999 ($r = 0.997$, $p < 0.001$). In 1979 and in 1989, the correlations between these two types of rates were relatively lower, but still significant ($r = 0.897$ in 1979, $p < 0.001$; $r = 0.945$ in 1989, $p < 0.001$).

The spatial distribution of biotechnology co-patenting rates is illustrated by the "box map" and the linked "cartogram," as shown in Figures 6.1-6.4. In the Geoda software package, the box map is a mapping function that transforms information from a box plot into a choropleth map domain, allowing for easy identification of extreme observations or outliers within the overall distribution. The circular cartogram is also a mapping function of Geoda where the size of each geographical area is made proportional to the value of a given variable (Anselin 2004; Anselin et al. 2006b). Figures 6.1-6.4 show a series of box maps with the linked cartograms for U.S. metropolitan co-invention by both the original and the smoothed co-patenting rates. The map for the original co-patenting rates in 1979 is shown in the upper part of Figure 6.1. Twenty metropolitan areas with extremely high co-patenting rates were categorized as High Outliers (red cycles). One cluster of these High Outliers emerged in the Northeast centered on New York and nearby areas in southern New England. A few High Outliers were scattered throughout the Midwest. Other High Outliers were San Francisco, Raleigh-Durham (North Carolina), Auburn-Opelika (Alabama), and McAllen (Texas). The lower part of Figure 6.1 shows a smoothed rate map where St. Louis was newly added to the list of areas with high co-patenting rates. Note that some of these outliers are difficult to identify in the box map owing to their small physical size. In order to make these small areas recognizable and fit them together in a layout, two circular cartograms with different rates are shown in the right part of Figure 6.1, where the size of each circle is made proportional to the value of its co-patenting rate. Lawrence (Kansas), Kalamazoo (Michigan), and New London (Connecticut) were regarded as the most densely co-inventive cities in the early development of U.S. biotechnology.

Figure 6.2 shows the box maps and the link cartograms for metropolitan co-invention in 1989. High Outliers were mostly located in the Northeast and the Great Lakes. San Francisco and Raleigh-Durham were two new appearances in this list. Visual inspection of the linked cartograms clearly reveals that Lawrence and New London remained the top two co-inventive cities. Figure 6.3 shows the same sets of maps for 1999. While the overall pattern was similar, there were some changes in these High Outliers compared with those in a decade earlier. First, Boston and San Diego had emerged as two of the national top biotechnology co-patenting centers. Second, the Midwestern co-inventive cities had broadened to include Madison (Wisconsin), Iowa City (Iowa), and Des Moines (Iowa). Third, there was less variation in metropolitan co-patenting rates as the sizes of the High Outliers in the cartogram became more evenly distributed across space. Figure 6.4 shows the 2009 spatial distribution of metropolitan co-

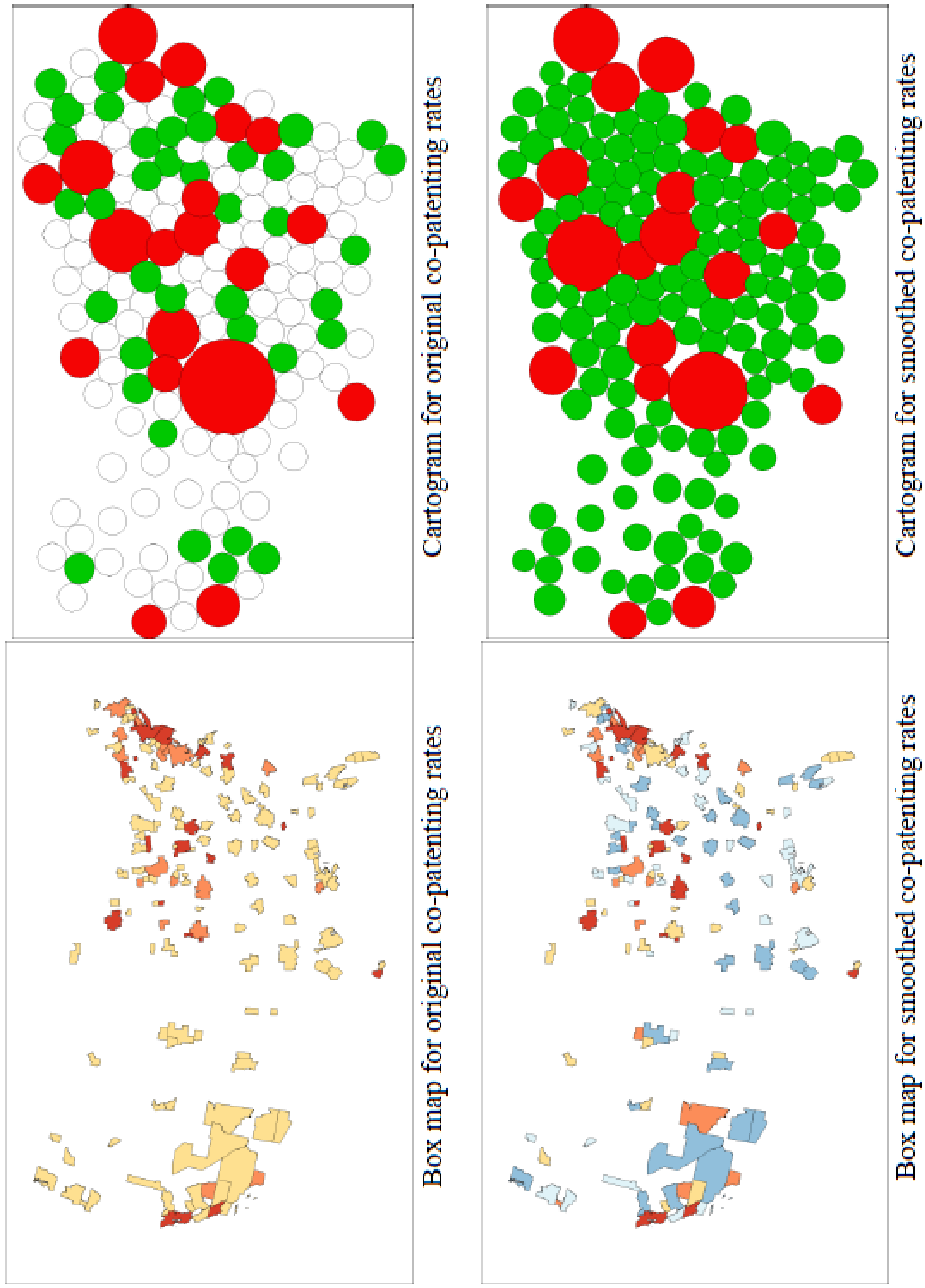


Figure 6.1 Box maps with linked cartograms for metropolitan co-patenting rates, 1979

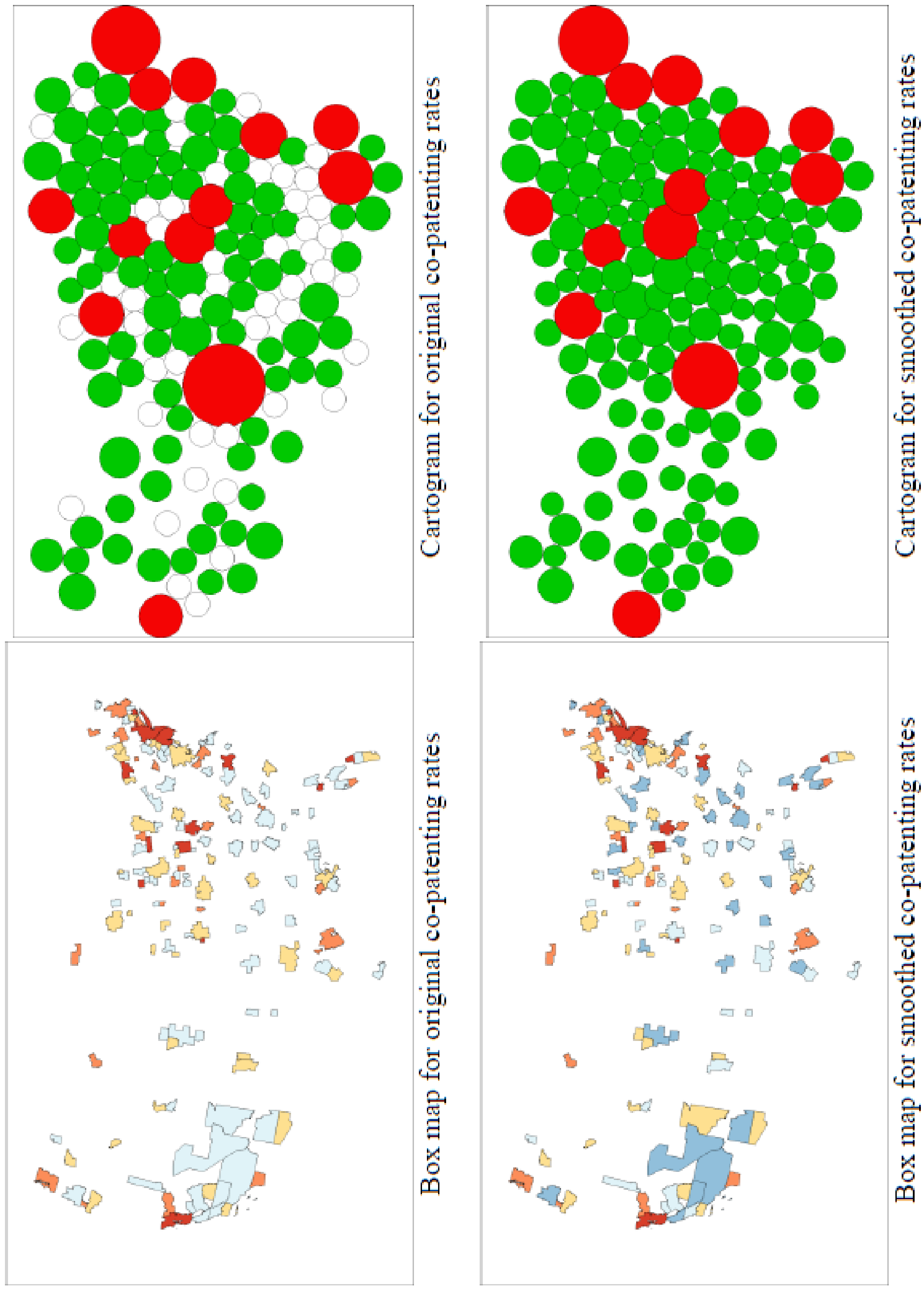


Figure 6.2 Box maps with linked cartograms for metropolitan co-patenting rates, 1989

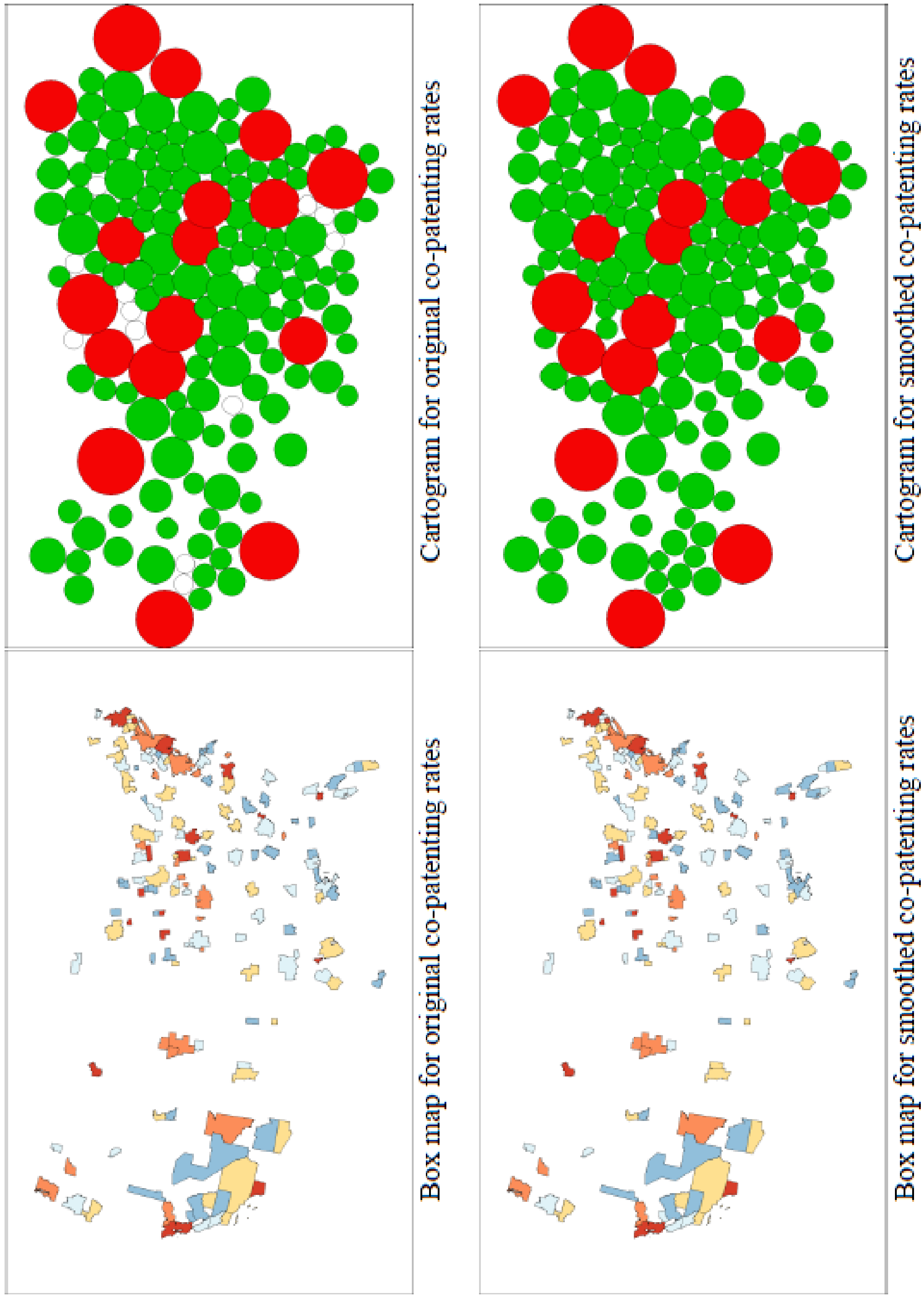


Figure 6.3 Box maps with linked cartograms for metropolitan co-patenting rates, 1999

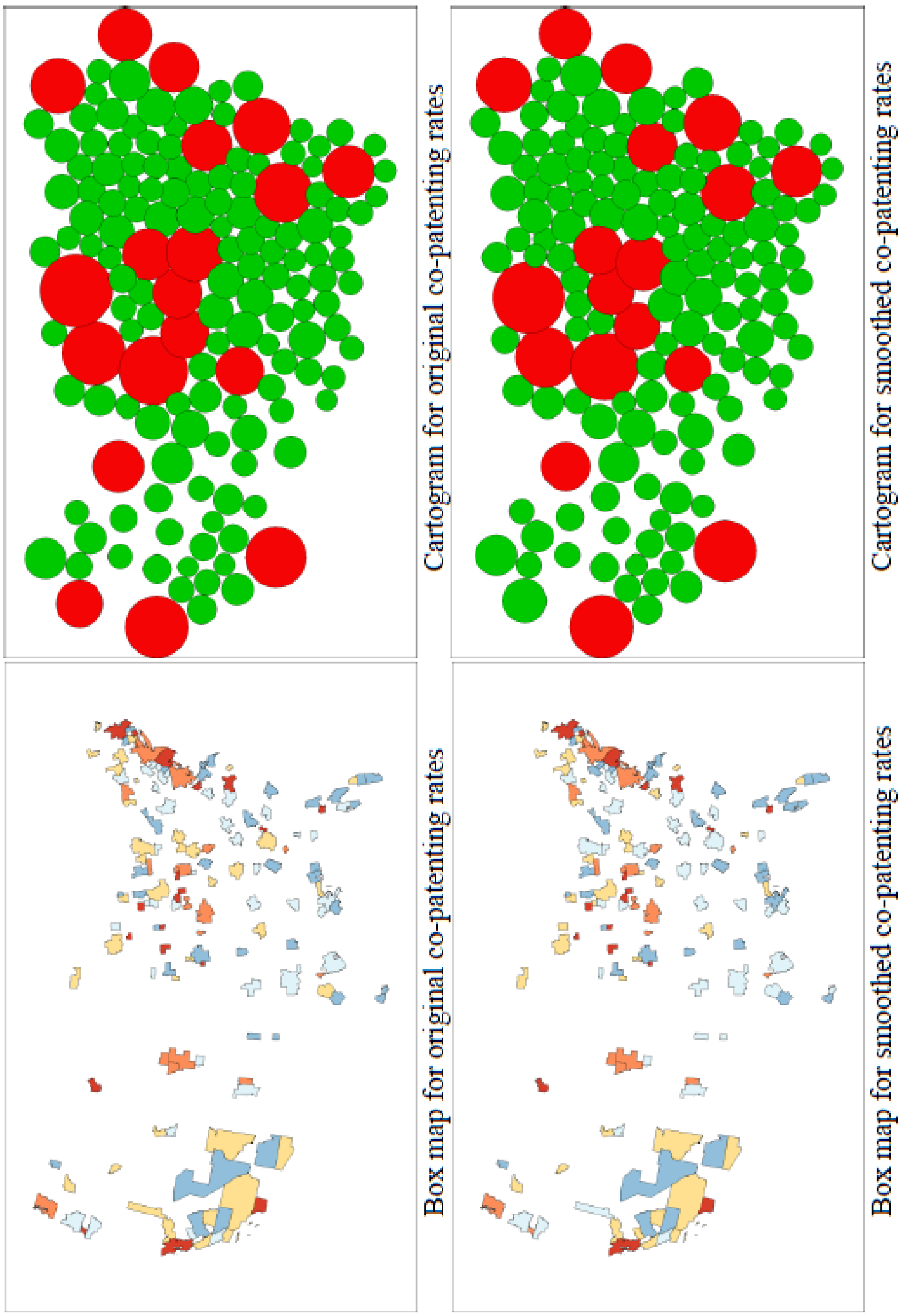


Figure 6.4 Box maps with linked cartograms for metropolitan co-patenting rates, 2009

invention. High co-inventive metropolitan areas were most evident in the Northeast Corridor, from Boston to Washington D.C.-Baltimore. A few high co-patenting Midwestern areas included Madison, Rochester (Minnesota), Champaign-Urbana (Illinois), Lafayette (Indiana), and Bloomington (Indiana). In addition, San Francisco, San Diego, Seattle, and Corvallis (Oregon) also had high biotechnology co-patenting rates.

6.2 Global Association

Moran's *I* is a global-level measure of dependence to detect the presence of spatial and network-based dependencies in metropolitan co-invention. Metropolitan co-invention is computed using the smoothed co-patenting rates. Two types of weights matrices (spatial and network-based) are used to assess the extent to which the overall structure is significant clustering or mostly random. The spatial weights matrix is selected on the basis of 7 percent of all possible 150 metropolitan areas, which yields ten nearest neighbors. The network-based weights matrix is constructed by using the average intermetropolitan co-patenting ties in each chosen year as a cut-off point to convert the observed co-invention network into a set of binary relations. The weights are adjusted to deal with the increase in co-patenting in the period 1979-2009. The average intermetropolitan co-patenting ties in 1979 was 1.6, which was rounded down to 1.0 as the cut-off point of the network-based binary matrix. Intermetropolitan pairs with one or more co-patent ties are assigned a unity weight and all other pairs are assigned a zero weight. The same rounding down rule is applied to the 1989 data and the cut-off point remains at one. The co-patenting average in 1999 rose to 2.8, which was rounded down to 2.0 as the cut-off point. Intermetropolitan pairs with two or more co-patenting ties are assigned a unity weight. Otherwise, their weights are set at zero. In 2009, the average co-patenting ties increased to 3.2. Intermetropolitan pairs with three or more co-patenting ties are assigned a unity value and all other weights are set to zero.[10]

The Moran's *I* results for metropolitan co-patenting rates using ten nearest neighbors as spatial weights and the co-patenting frequencies as network-based weights are shown in Table 6.1. The pseudo significance values are based on a permutation approach. None of the spatial results is significant. Only the 2009 network-based system shows significant global dependence. The latter coefficient is negative and small indicating that U.S. metropolitan areas with dissimilar co-

Table 6.1 Moran's *I* statistics for metropolitan co-patenting rates

(type of weights matrix)	1979	1989	1999	2009
10 nearest neighbors as spatial weights	0.0104 (0.224)	-0.011 (0.465)	-0.0137 (0.418)	0.0247 (0.141)
Dichotomized network with average co-patenting tie ≥ 1	0.0262 (0.131)	0.0049 (0.394)		
Dichotomized network with average co-patenting tie ≥ 2			0.0256 (0.211)	
Dichotomized network with average co-patenting tie ≥ 3				-0.0848 (0.015)

Note: Numbers in parentheses are two-tailed significance levels.

[10] The average intermetropolitan co-patenting ties were 1.6 in 1979, 1.9 in 1989, 2.8 in 1999, and 3.2 in 2009.

patenting rates are significantly network-based associated. This finding for the 2009 network-based system alone might be interpreted as evidence that areas with low co-patenting rates are significantly dependent on ties to a major biotechnology center. Inventors in minor co-inventive cities perhaps actively seek to establish relations with partners in high co-inventive cities. The absence of significant positive global network-based dependence implies that most minor co-inventive cities have few links with other low co-inventive cities. Moreover, most major co-inventive cities are not significantly linked to the others.

6.3 Local Association

While the Moran's *I* results mostly suggest randomness of metropolitan co-invention in both the spatial and network-based systems, a focus on local associations reveals useful detail. LISA cluster maps identify local groupings and classify them into five patterns: Co-invention Cores (high-high), Co-invention Peripheries (low-low), High Co-invention Islands (high-low), Low Co-invention Islands (low-high), and Non-significant Areas ($p > 0.05$). Figures 6.5-6.6 show spatial LISA cluster maps of metropolitan co-invention using ten nearest neighbors as spatial weights. Significant spatial LISA clusters were largely absent in 1979, as shown in the upper part of Figure 6.5. However, a minor Co-invention Core (high-high) of metropolitan areas with high co-patenting rates occurred in the Midwest focused on Lafayette (Indiana), St. Louis (Missouri), St. Joseph (Missouri), and Des Moines (Iowa). This core had several neighbors categorized as Low Co-invention Islands (low-high) including Sioux City (Iowa), Omaha (Nebraska), Lincoln (Nebraska), Columbia (Missouri), Toledo (Ohio), and

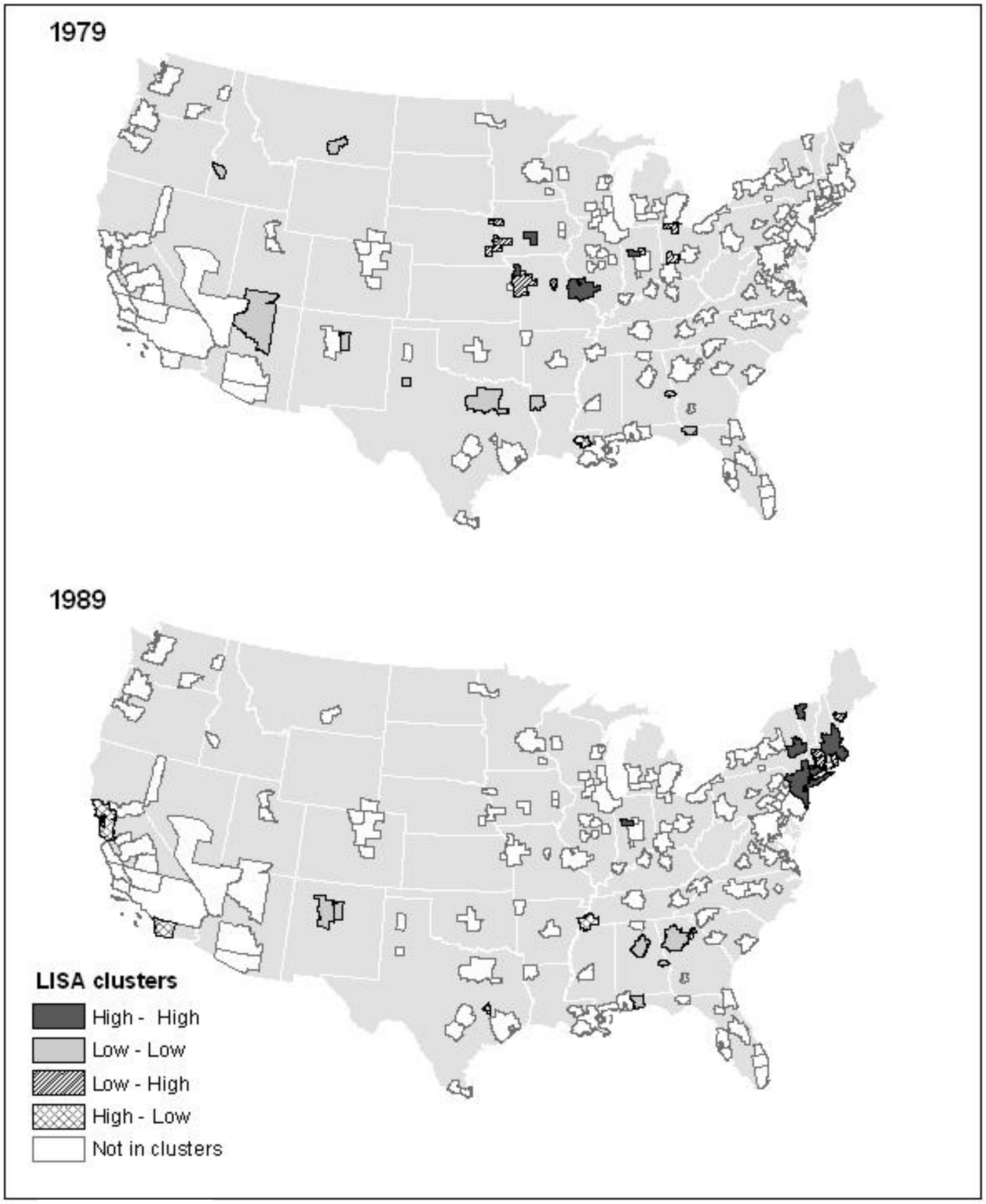


Figure 6.5 Spatial LISA cluster maps of metropolitan co-invention
Note. Neareast neighbor k = 10

Dayton-Springfield (Ohio). An indeterminate Co-invention Periphery (low-low) emerged throughout the Intermountain West and the Southeast with focal centers in Billings (Montana), Boise City (Idaho), Flagstaff (Arizona), Santa Fe (New Mexico), Lubbock (New Mexico), Dallas (Texas), Shreveport-Bossier (Louisiana), and Tallahassee (Florida). This region was inconsequential in American biotechnology co-patenting.

In 1989, as shown in the lower part of Figure 6.5, a distinct Co-invention Core occurred in the Northeast centered on New York, Boston, Albany-Troy (New York), New Haven (Connecticut), and Burlington (Vermont) where co-invention was largely tied to pharmaceutical technologies. Several Low Co-invention Islands also occurred in this region including Portland (Maine), Springfield (Massachusetts), Hartford (Connecticut), and Providence (Rhode Island). Two noticeable new appearances in the West – San Francisco and San Diego – were identified as High Co-invention Islands (high-low). These areas had high co-patenting rates, but their neighbors were significantly far less engaged in biotechnology co-invention. Memphis (Tennessee) and State College (Pennsylvania) also joined the list of High Co-invention Islands. A sizeable medical center in the former and a large public university (Pennsylvania State University) in the latter led to unusually high co-patenting rates compared with their nearest neighbors.

The 1999 spatial LISA cluster map is shown in the upper part of Figure 6.6. San Francisco and San Diego were High Co-invention Islands and a noticeable Co-invention Periphery was evident in the Southeast. The latter was

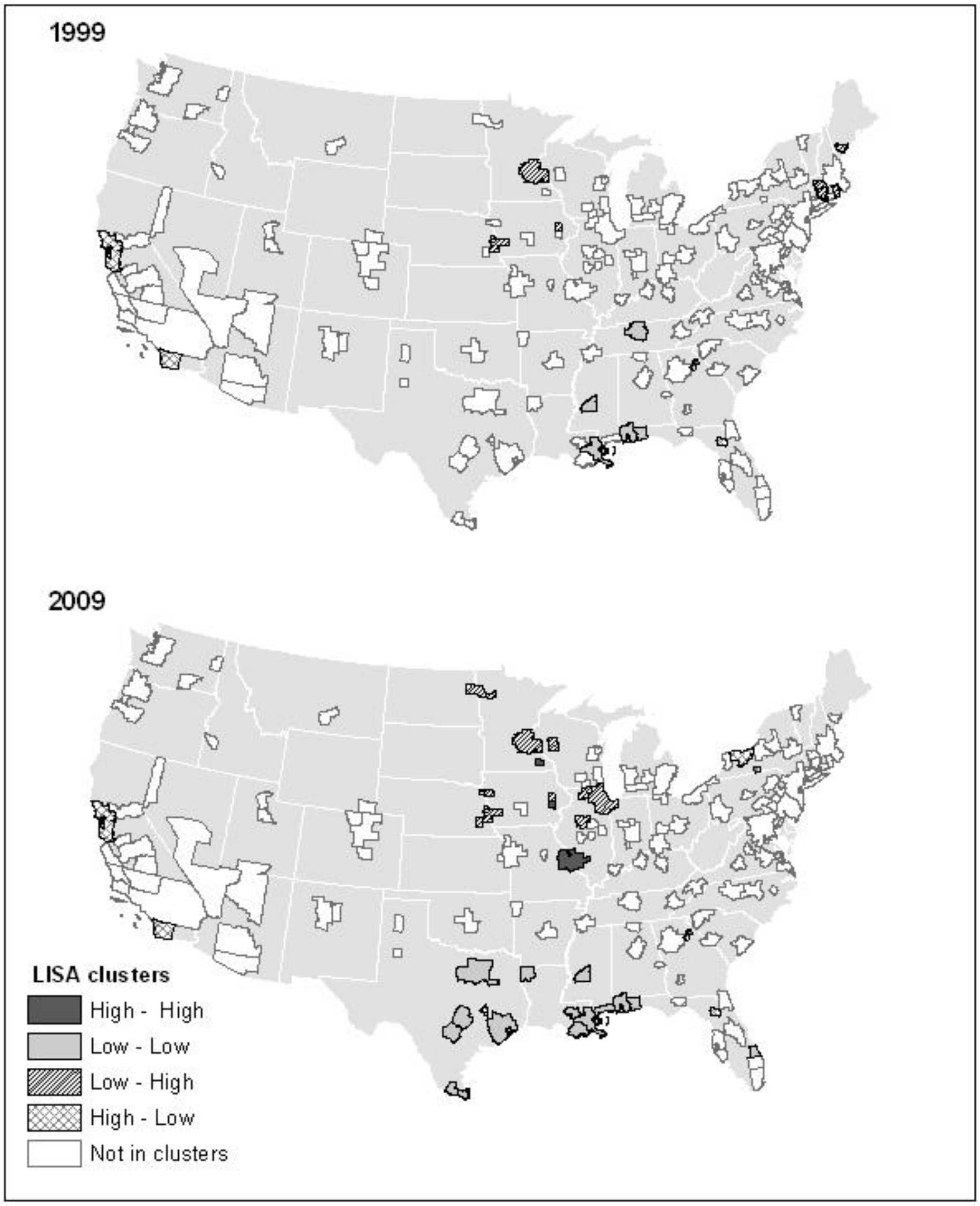


Figure 6.6 Spatial LISA cluster maps of metropolitan co-invention
Note. Neareast neighbor k = 10

focused on Jackson (Mississippi), Mobile (Alabama), New Orleans (Louisiana), and Pensacola (Florida). In 2009, this Co-invention Periphery noticeably expanded, stretching from east Texas to Alabama, as shown in the lower part of Figure 6.6. A small Co-invention Core was evident in the Midwest around St Louis (Missouri), Iowa City (Iowa), and Rochester (Minnesota).

Network-based LISA cluster maps provide an alternative perspective on co-patenting ties. It is important to stress that these maps shown in Figures 6.7 and 6.9 depict network and not spatial associations of metropolitan areas. Note that when interpreting a network-based local association, one should look at the neighboring structure from its network-based weights matrix to determine the dependence of each observation with others. Geographical distant areas can be closely tied in a co-invention network. Moreover, nearby areas may have little association. Over the course of the period 1979-2009, the network-based collaborative patterns of biotechnology based on U.S. intermetropolitan co-patenting activities are mostly composed of Low Co-invention Islands (low-high), and a small number of prominent Co-invention Cores (high-high). In the early years of the period, most co-patenting activities were conducted by local inventors as 83.2 percent of co-patenting activities were mainly localized within the same metropolitan area (see Table 5.3 for details). As shown in the upper part of Figure 6.7, Raleigh-Durham (North Carolina) in 1979 was the only metropolitan area categorized as a Co-invention Core. In 1989, as a growing number of inventors joined non-local co-patenting ties, more discernible Co-invention Cores emerged. New York, collaborating with 37 metropolitan areas, was the national

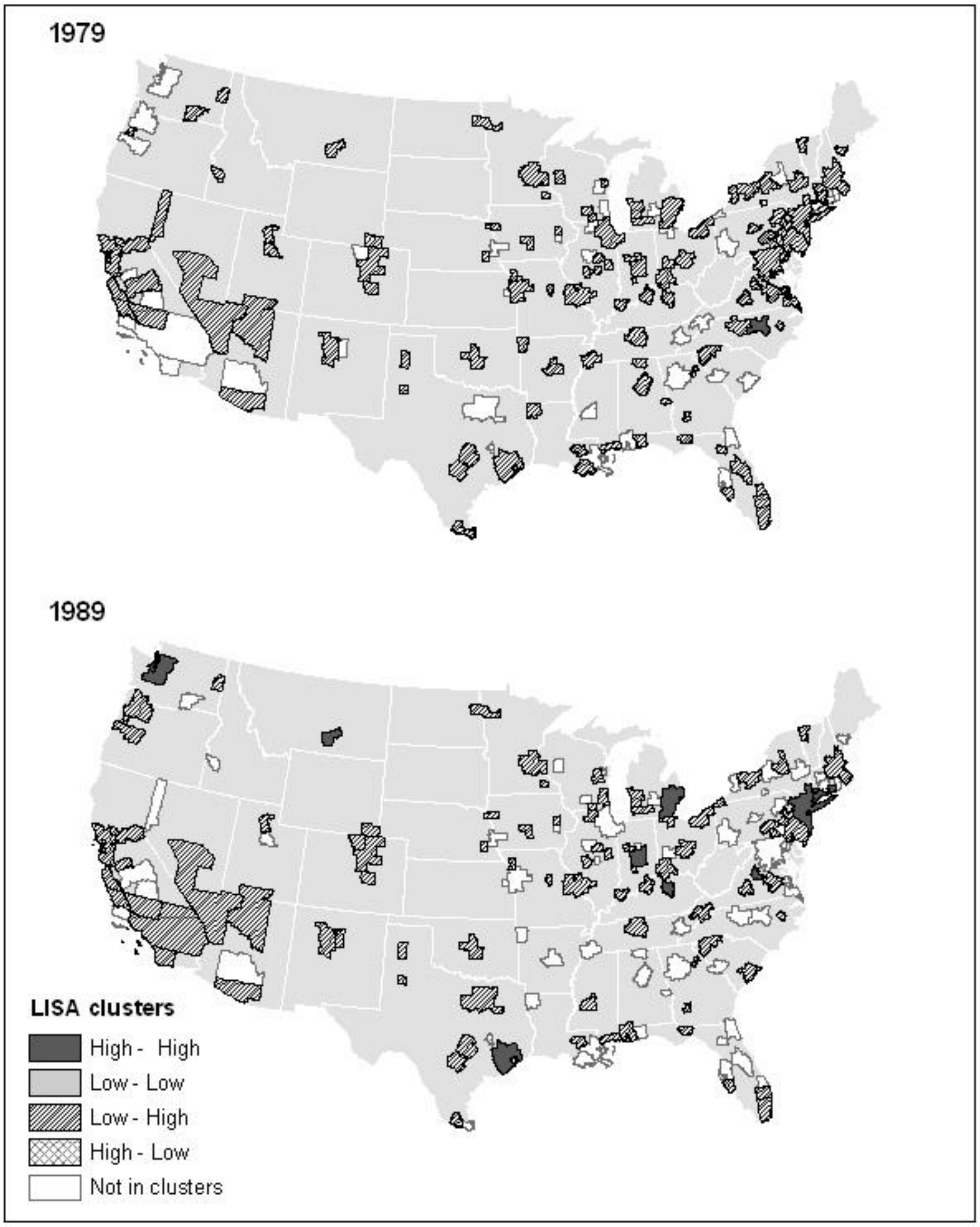


Figure 6.7 Network-based LISA cluster maps of metropolitan co-invention
Note. Average-based cut-off point in 1979 and 1989 = 1

core for biotechnology co-invention. Detroit-Ann Arbor (Michigan) and Indianapolis (Indiana) were two distinct Co-invention Cores with each having nine metropolitan partners. Figure 6.8 shows that both cities have different types of neighboring structures. Detroit-Ann Arbor was strongly related to Boston, New York, and Philadelphia, as well as neighboring Lansing-East Lansing (Michigan) and Cincinnati (Ohio), while Indianapolis was primarily tied to New York, Washington D.C.-Baltimore, Cincinnati, as well as Austin (Texas) and New Orleans (Louisiana) in the South. Seattle (Washington) and Houston (Texas) were also categorized as Co-invention Cores. Seattle was mainly associated with San Francisco, and Houston was relationally close to Philadelphia. These results indicate that their network-based associations have regional biases.

The 1999 network-based LISA cluster map is shown in the upper part of Figure 6.9. San Francisco and Boston categorized as Co-invention Cores were two leading co-patenting centers where many metropolitan areas aligned their resources with these two centers in biotechnology co-invention. Both co-inventive cores had the highest degree of co-patenting activity with each other, indicating that inventors located in both centers were network-based dependent. As shown in Figure 6.10, San Francisco was extensively engaged in co-patenting with major biotechnology centers across the U.S. including New York, San Diego, Los Angeles, and Washington D.C.-Baltimore. In contrast, Boston's major partners were largely concentrated in the Northeast including New York, Philadelphia, Providence (Rhode Island), and Washington D.C.-Baltimore. Seattle, Denver, Indianapolis, and New London were medium-size Co-invention

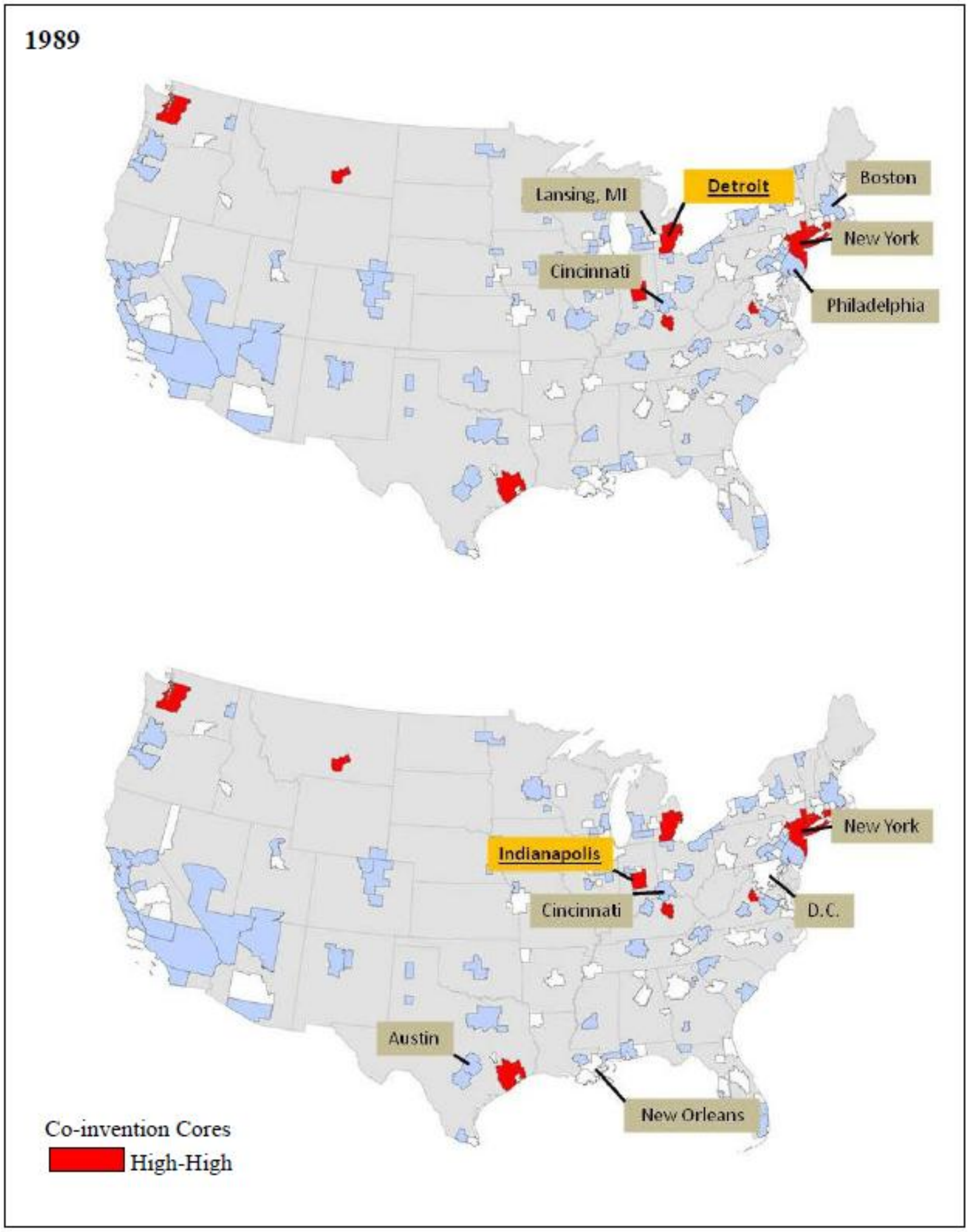


Figure 6.8 Comparison of Detroit and Indianapolis's closest network-based metropolitan partners in 1989

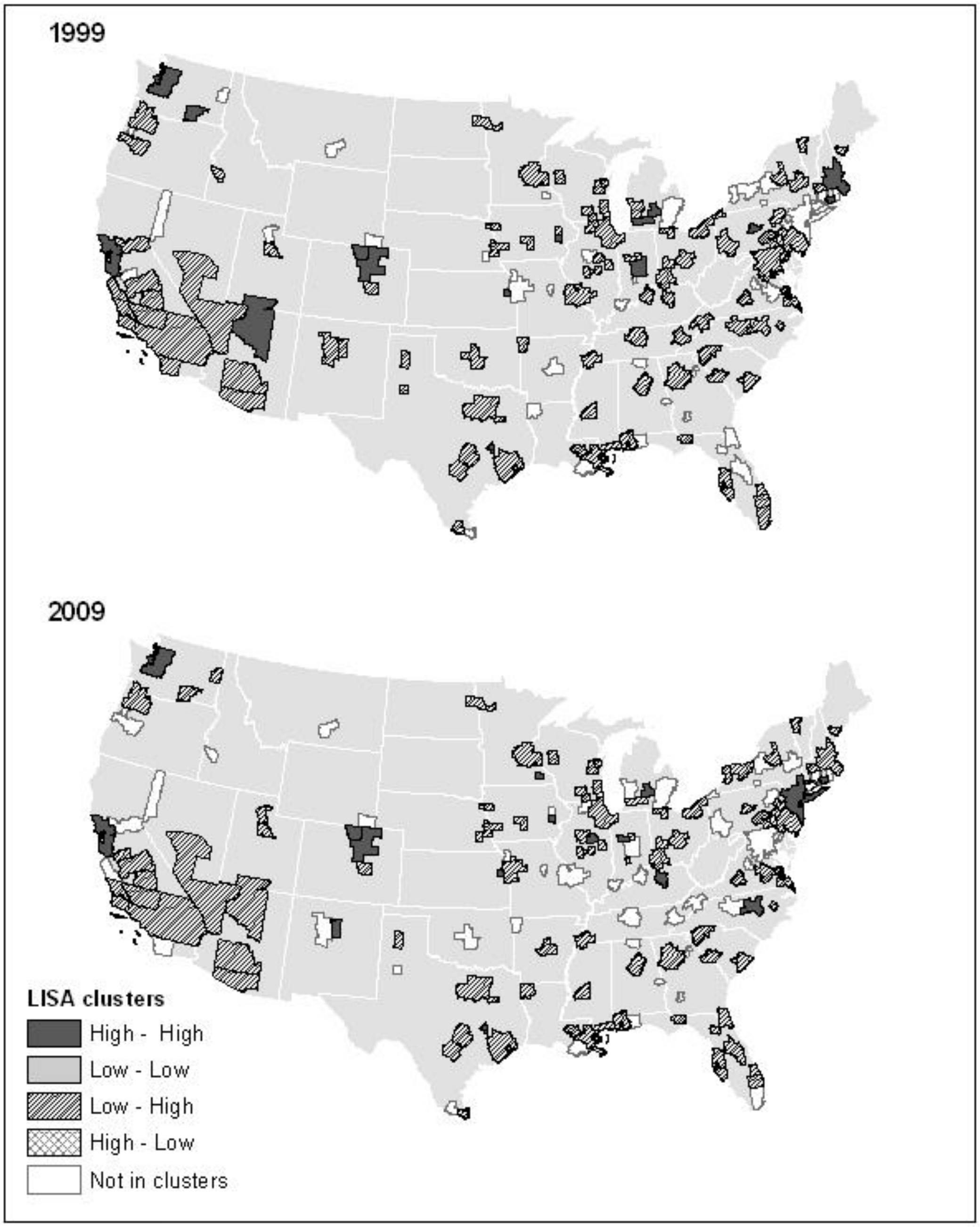


Figure 6.9 Network-based LISA cluster maps of metropolitan co-invention
Note. Average-based cut-off point in 1999 = 2, in 2009 = 3

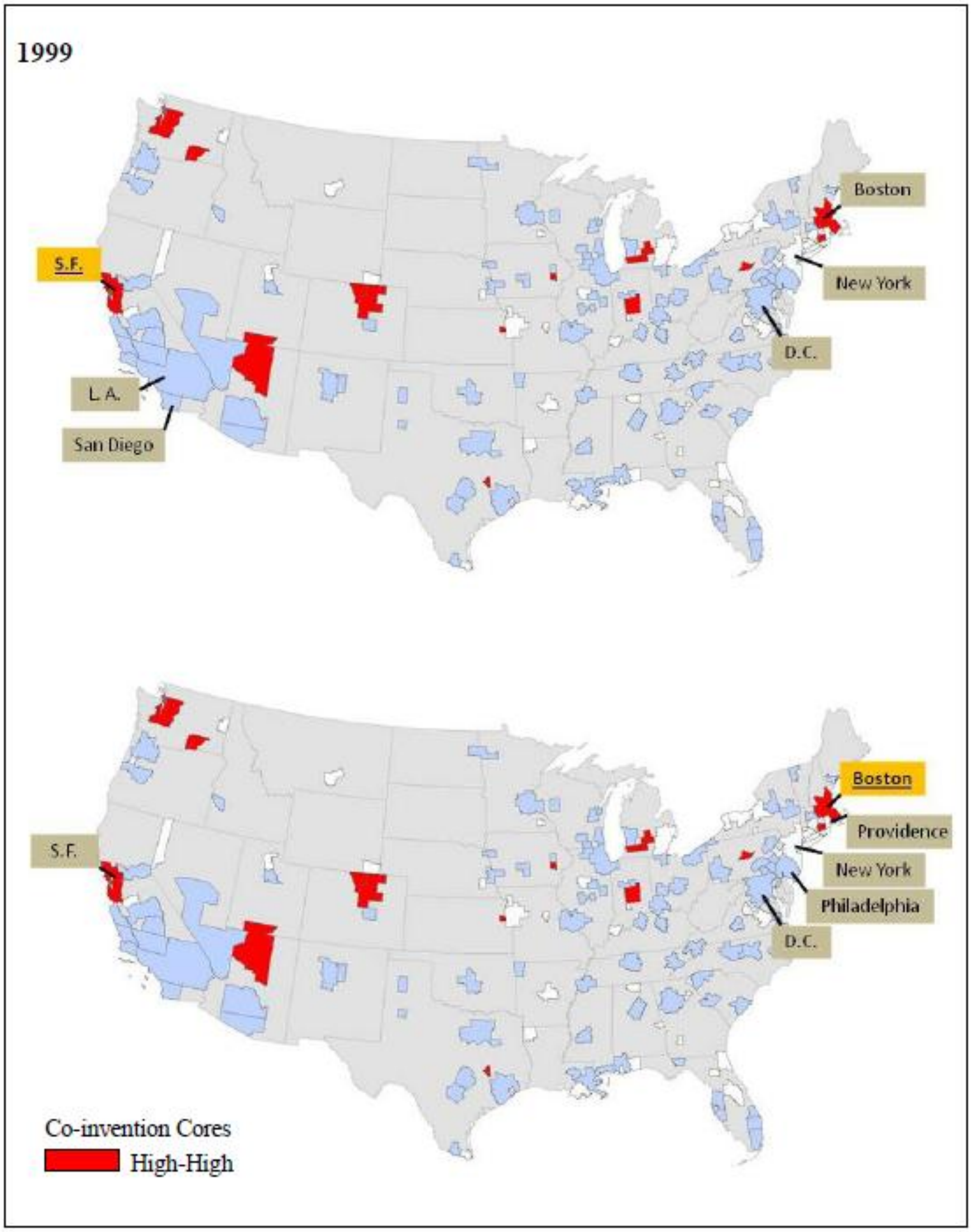


Figure 6.10 Comparison of San Francisco and Boston's closest network-based metropolitan partners in 1999

Cores linked to major biotechnology centers. In comparing their neighboring structures, one noticeable geographical-related feature among these metropolitan areas is shown in Figure 6.11. Inventors located in Indianapolis and New London tended to primarily team up with inventors in Midwestern cities (i.e., Chicago, Lafayette, Bloomington) and Northeastern cities (i.e., Boston, New Haven, Hartford, Providence), respectively. Inventors located in Seattle and Denver had mostly national ties. Regional biases also occurred in some small co-inventive core areas. For example, Figure 6.12 shows that State College (Pennsylvania) had a high degree of co-invention with several Northeastern cities (e.g., Boston, Philadelphia, Washington D.C.-Baltimore), and Kalamazoo-Battle Creek (Michigan) had its strongest links with several Midwestern cities (e.g., Chicago and Grand Rapids). On the other hand, Lawrence (Kansas) widely cooperated with New York and San Francisco; Bryan-College Station (Texas) had partners from Gainesville (Florida), Knoxville (Tennessee), to Washington D.C.-Baltimore.

As shown in the lower part of Figure 6.9, major biotechnology concentrations that form the 2009 network-based Co-invention Cores included New York (New York), San Francisco (California), Washington D.C.-Baltimore (District of Columbia), Boston (Massachusetts), Denver (Colorado), Seattle (Washington), and Raleigh-Durham (North Carolina). Smaller metropolitan areas were also focuses of these Co-invention Cores including the university towns of Fort Collins (Colorado State University), Iowa City (University of Iowa),

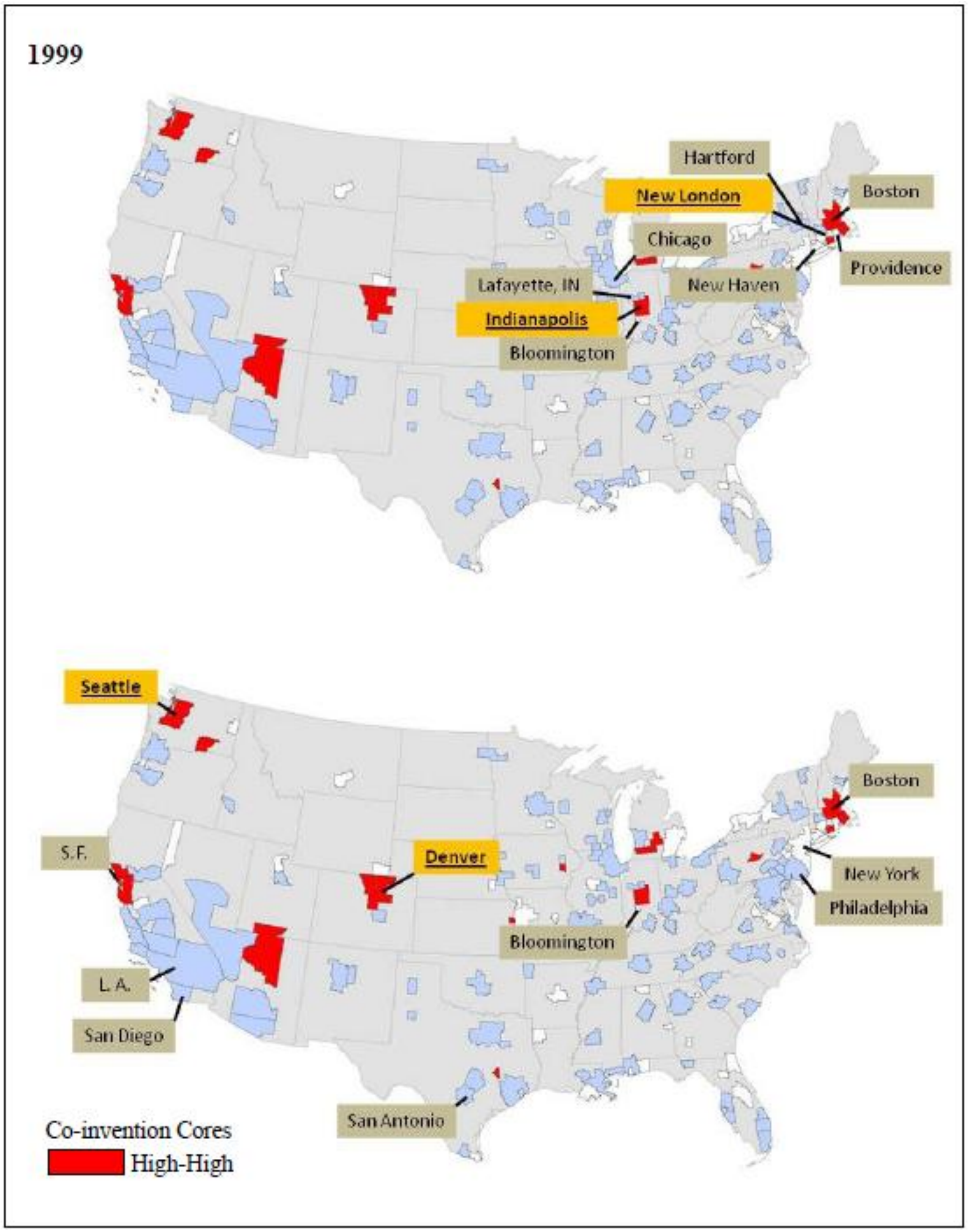


Figure 6.11 Comparison of New London and Indianapolis with Denver and Seattle's closest network-based metropolitan partners in 1999

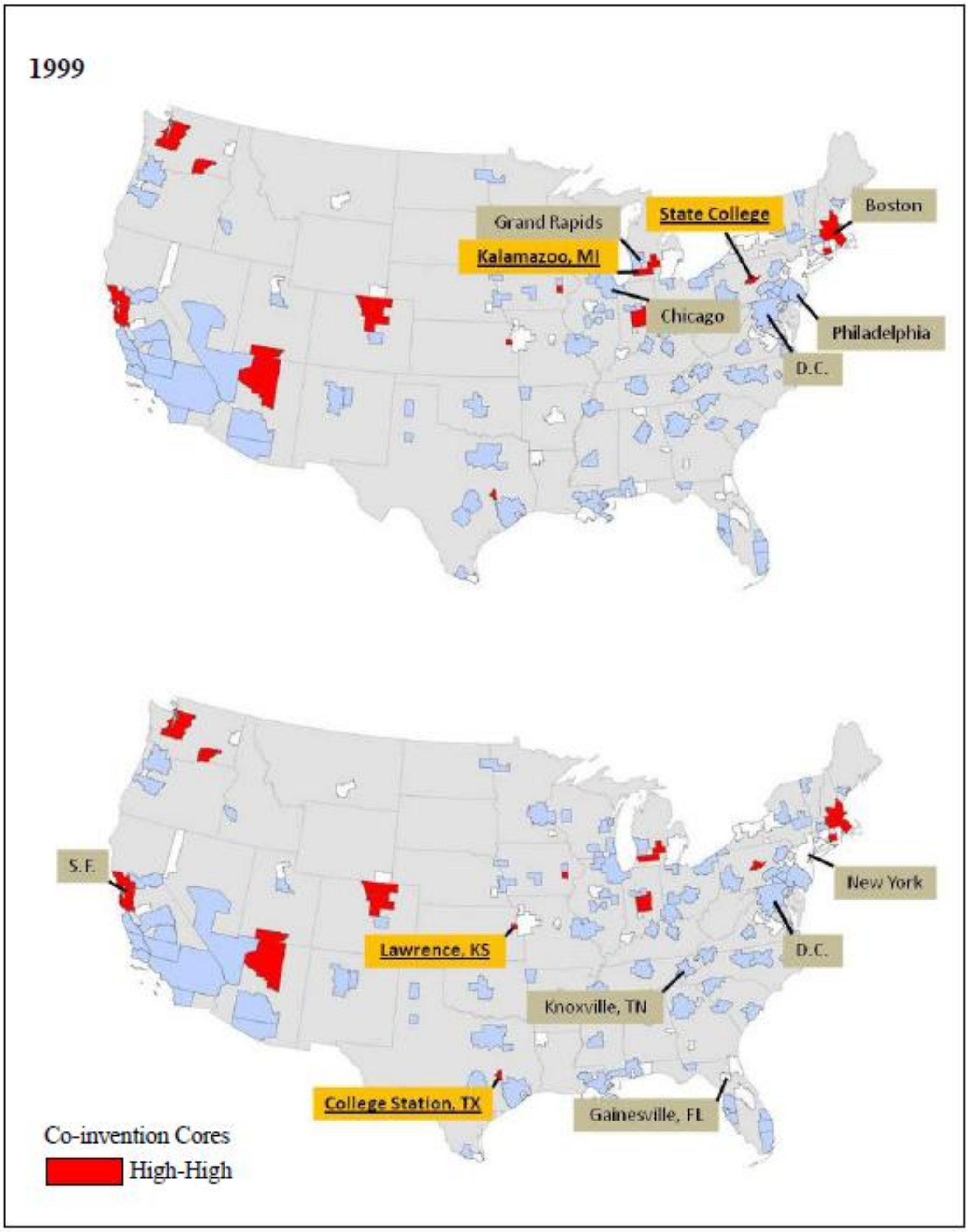


Figure 6.12 Comparison of State College and Kalamazoo with Lawrence and College Station's network-based closest metropolitan partners in 1999

Lafayette (Purdue University), Lansing-East Lansing (Michigan State University), Lexington (University of Kentucky), Bryan-College Station (Texas A&M University), and Bloomington-Normal (Illinois State University). In addition, Santa Fe (New Mexico) with its large national laboratory (i.e., Los Alamos National Laboratory) and research center (i.e., Santa Fe Institute), Rochester (Minnesota) known as a "Med City" with the headquarters of Mayo Clinic, and New London that has a cluster of pharmaceutical companies all belonged to the 2009 Co-invention Core. These areas had high co-patenting rates and their closely network-based associations also had high co-patenting rates. Some studies argue that biotechnology inventive firms have largely concentrated around major universities and research centers (e.g., Audretsch 2001; Niosi and Banik 2005; Cooke 2007). Small and medium Co-invention Cores identified here occurred in areas having these two essential requirements for biotechnology invention.

Comparison of the spatial and network-based LISA cluster maps, especially in 2009, suggests that the latter better define co-patenting relationships. The 2009 spatial LISA cluster map did not identify any striking spatial associations. A small Co-invention Core occurred in the Midwest, a periphery was evident from east Texas to Alabama, and San Francisco and San Diego were categorized as High Co-invention Islands. The 2009 network-based LISA cluster map identified the biotechnology Co-invention Cores and remaining areas that were less engaged in co-patenting. Inspection of co-patenting ties of some of the major biotechnology centers shows that their network-based associations had regional biases. San Francisco and New York dominated the network-based Co-invention Core. Their collaborators spanned broadly across more than 70 major metropolitan areas nationwide. San Francisco's strongest partners were national including San Diego, Los Angeles, New York, Boston, Philadelphia, and Washington D.C.-Baltimore. As shown in Figure 6.13, its ties were truly national. In contrast, New York's strongest partners were mainly eastern including Boston, New Haven, Hartford, Philadelphia, and Washington D.C.-Baltimore. New York was also strongly tied to San Francisco but its links with Los Angeles and San Diego were weaker. A third example shown in Figure 6.14 illustrates the evidence of regional effects. Seattle's collaborations spanned 40 metropolitan areas but its strongest links were with San Francisco, Los Angeles, and San Diego. Raleigh's closest partners, however, were relatively dispersed. These results suggest that the co-patenting relationships of major biotechnology centers are national and regional but not spatial. Spatial proximity is less important for intermetropolitan collaboration compared with both network and regional relationships.

6.4 Summary

This chapter presents results showing differences in spatial and network-based dependencies of biotechnology co-invention across U.S. metropolitan areas. The spatial distribution of biotechnology co-patenting rates in 1979 shows that 20 metropolitan areas with extremely high co-patenting rates were identified as High Outliers including New York and nearby areas in southern New England, some Midwestern cities, San Francisco, Raleigh-Durham, Aubum-Opelika (Alabama),

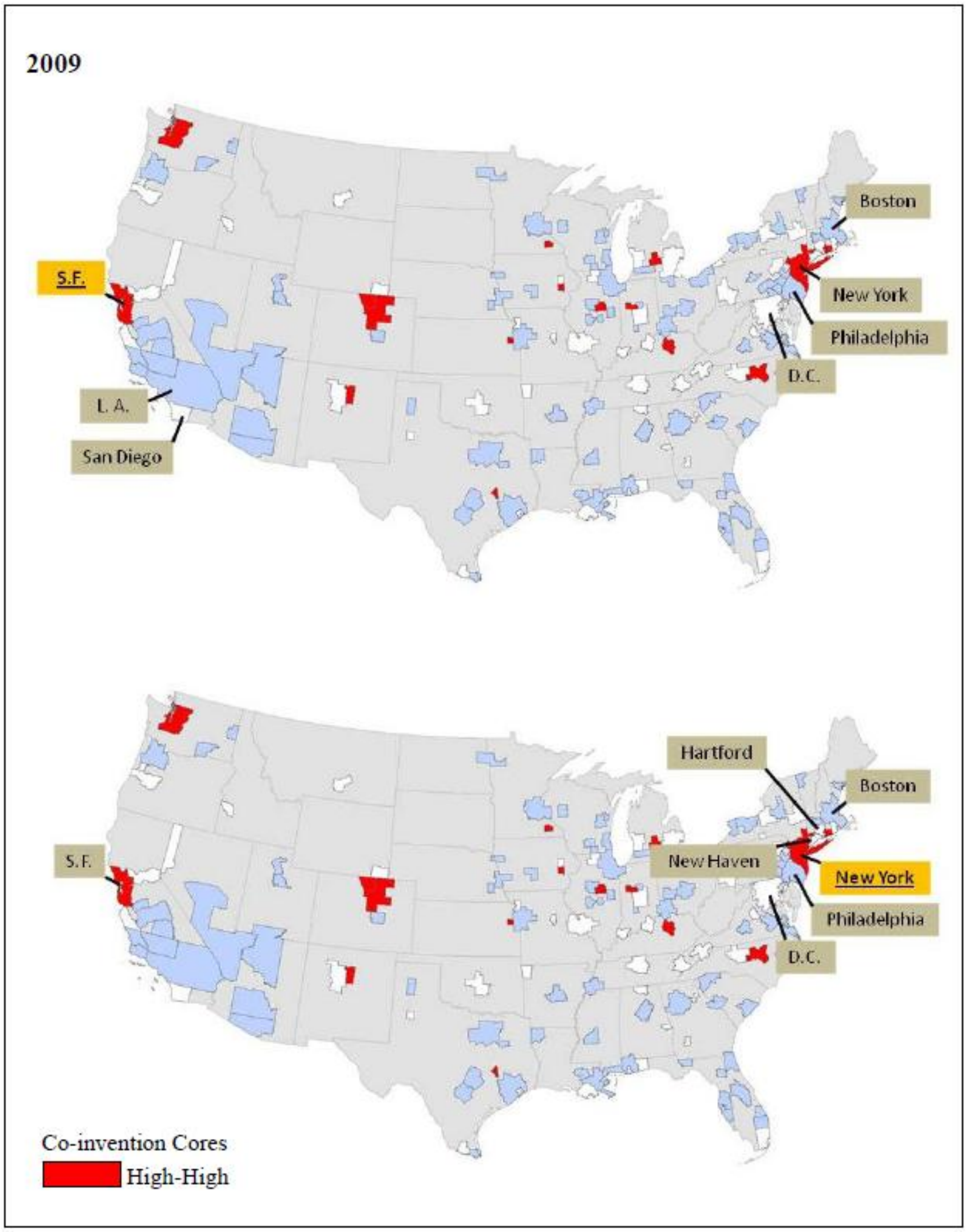


Figure 6.13 Comparison of San Francisco and New York's network-based closest metropolitan partners in 2009

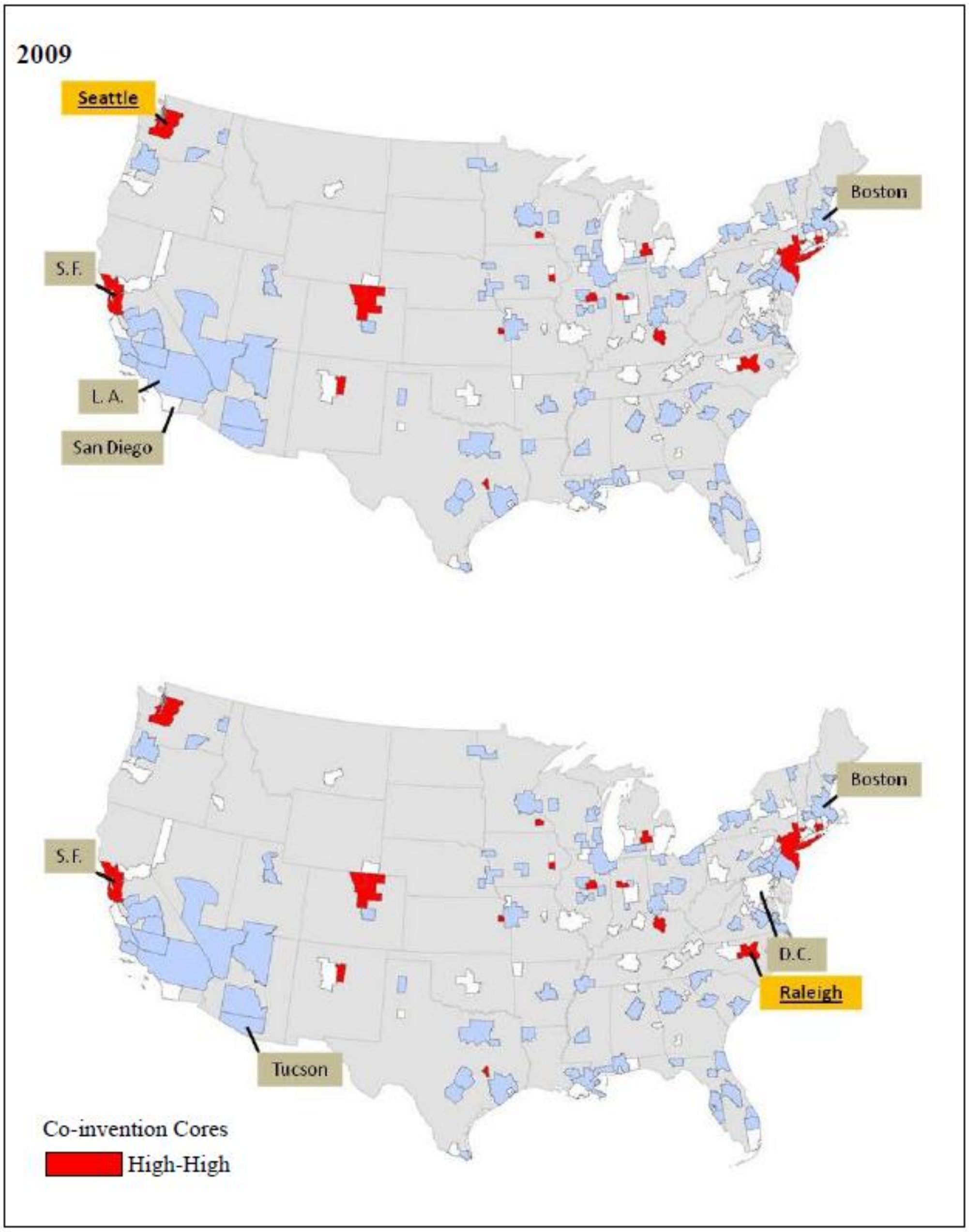


Figure 6.14 Comparison of Seattle and Raleigh's network-based closest metropolitan partners in 2009

and McAllen (Texas). In 1989, High Outliers were mostly located in the Northeast and the Great Lakes. San Francisco and Raleigh-Durham were two new appearances in this list. In 1999, Boston and San Diego had emerged as two of the national top biotechnology co-patenting centers. High Outliers in the Midwest had broadened to include Madison (Wisconsin), Iowa City (Iowa), and Des Moines (Iowa). In 2009, High Outliers were most evident in the Northeast Corridor, the Midwest, and the West including San Francisco, San Diego, and Corvallis (Oregon).

Results of the Moran'*I* statistics in the period 1979-2009 show that only the 2009 network-based system has a statistically significant negative coefficient. This is interpreted as evidence that areas with low co-patenting rates are significantly dependent on ties to areas with high co-patenting rates. Analysis of local dependence using the spatial LISA cluster maps shows few discernable spatial associations in any of the years of analysis. In 1979, a few Co-invention Cores and several Low Co-invention Islands occurred in the Midwest. In 1989, Co-invention Cores emerged in the Northeast. San Francisco and San Diego were two noticeable High Co-invention Islands in the West. In 1999, a distinct Co-invention Periphery was evident in the Southeast. In 2009, this Co-invention Periphery noticeably expanded, stretching from east Texas to Alabama. A small Co-invention Core was evident in the Midwest. San Francisco and San Diego remained as High Co-invention Islands.

Analysis of local dependence using the network-based LISA cluster maps over the period 1979-2009 reveals that a few metropolitan areas have emerged as

Co-invention Cores with significant ties to distant partners, but evidence of some regional biases is also noted. Raleigh-Durham (North Carolina) was the only significant network-based Co-invention Core in 1979. More discernible network-based Co-invention Cores including New York, Detroit-Ann Arbor, and Indianapolis emerged in 1989. In 1999, San Francisco and Boston were two noticeable Co-invention Cores with different types of neighboring structures. The former was extensively engaged in co-patenting with major cities across the U.S., while the latter's closest network-based partners were largely concentrated in the Northeast. In 2009, network-based Co-invention Cores expanded including New York, San Francisco, Washington D.C.-Baltimore, Boston, Denver, Raleigh-Durham, and some college towns. In investigating their neighboring structures, these network-based Co-invention Cores showed national and regional but not spatial associations. Spatial proximity is less important for intermetropolitan collaboration compared with both network and regional relationships.

Chapter 7

RESULTS OF THE PROPERTIES OF INTERMETROPOLITAN NETWORKS OF BIOTECHNOLOGY CO-INVENTION

This chapter provides results showing the structural properties of intermetropolitan networks of biotechnology co-invention from three perspectives: components and cohesive subgroups of metropolitan areas, intermetropolitan network centralization and centrality, and positions of metropolitan areas established within the co-invention network-based system. Scott (2000) argued that the underlying structure of a network is more apparent in the relations of positions than among individual nodes themselves. Understanding how network positions of metropolitan areas form provides an empirical glimpse into the U.S. network-based system of knowledge flows and uncovers whether the system is consistent with a relational core/periphery structure.

This chapter is organized as follows. Section 7.1 presents patterns of connections among U.S. metropolitan areas using the concept of nested components and the *k*-cores and *m*-slices procedures to identify cohesive subgroups. Section 7.2 interprets global-level centralization and local-level centrality using three types of measures – degree, closeness, and betweenness – to determine the importance of metropolitan areas in transferring knowledge. Section 7.3 discusses positions of metropolitan areas using the regular equivalence criterion to reveal the kind of co-invention network-based system these areas form and the roles played by different types of areas within the system. Section 7.4 provides a summary of these results.

7.1 Patterns of Connection among U.S. Metropolitan Areas

A QAP correlation procedure (Quadratic Assignment Procedure) is used to compare the co-invention networks for the four chosen years (1979, 1989, 1999, and 2009). Based on a permutation approach, QAP computes a correlation coefficient between two networks by comparing the similarity of their square actor-by-actor matrices (Borgatti et al. 2002; Hanneman and Riddle 2005). Table 7.1 shows that the QAP correlation coefficients over the four years are all statistically significant and positively correlated. Before 1999, few co-patenting activities crossed metropolitan boundaries, resulting in relatively simple, non-interconnected intermetropolitan networks (see Figures 7.1 and 7.3 for details). The intermetropolitan co-invention networks in 1979 and in 1989 have a moderate correlation (r = 0.777). As a growing number of inventors participated in non-local co-patenting between 1989 and 1999, a more organized and less fragmented co-invention network structure emerged (see Figures 7.5 and 7.7 for details). The overall correlation coefficient between the two networks increased to 0.781. The strongest relationship between the 1999 and 2009 networks (r = 0.903) indicates the evidence of the increasingly network aspects of biotechnology co-invention by connecting and collaborating with large numbers of inventors across metropolitan boundaries.

Figure 7.1 visualizes the 1979 intermetropolitan co-invention network where the U.S. metropolitan areas are nodes and the co-invention ties among pairs

Table 7.1 QAP correlation coefficient matrix between the four co-invention networks

	The 1979 network	The 1989 network	The 1999 network
The 1989 network	.777		
The 1999 network	.741	.781	
The 2009 network	.665	.734	.903

Note: all correlation coefficients are statistically significant at the 0.01 level (two-tailed)

of areas are connecting lines.[11] The width of a line segment is set by frequency of biotechnology co-patenting, indicating the strength of co-invention between each intermetropolitan pair. Isolates that do not co-invent with other metropolitan areas are placed in the upper left corner of the figure. Seven components were evident in the 1979 co-invention network where the main component consisted of 29 metropolitan areas, along with 107 isolates. With an emphasis on the cohesive subgroup, Figure 7.2 shows the dense part of the network consisting of metropolitan areas with more frequent and direct connections. A total of 15 areas were derived by a *2*-core procedure (i.e., each selected MSA is directly tied to at least two other MSAs). The annual frequency of co-patenting between New York and Philadelphia was 17, which was the strongest connection among all intermetropolitan pairs. The next strongest pairs with frequencies of two were New York and Detroit, New York and Miami, and Chicago and Minneapolis. Co-patenting relationships in some minor components were geographically defined. Spatial proximity occurred in three strong local associations: New London (Connecticut) and Hartford (Connecticut); Des Moines (Iowa) and Omaha

[11] The network visualization for this and the following figures was performed using NetDraw as implemented in the UCINET 6 software package (Borgatti et al. 2002).

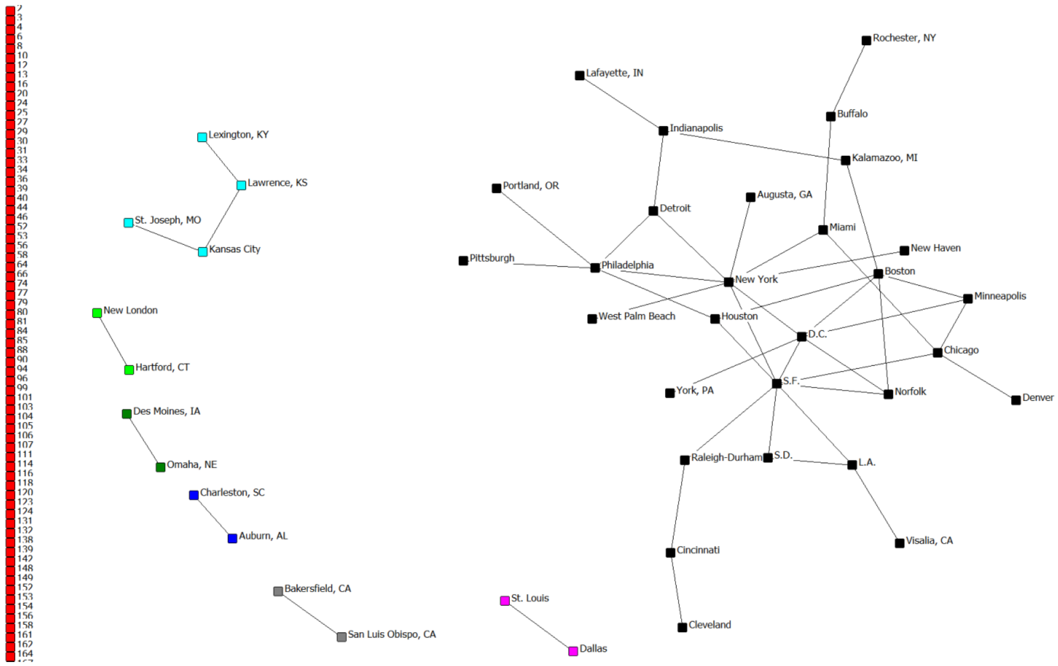


Figure 7.1 Intermetropolitan co-invention network in 1979

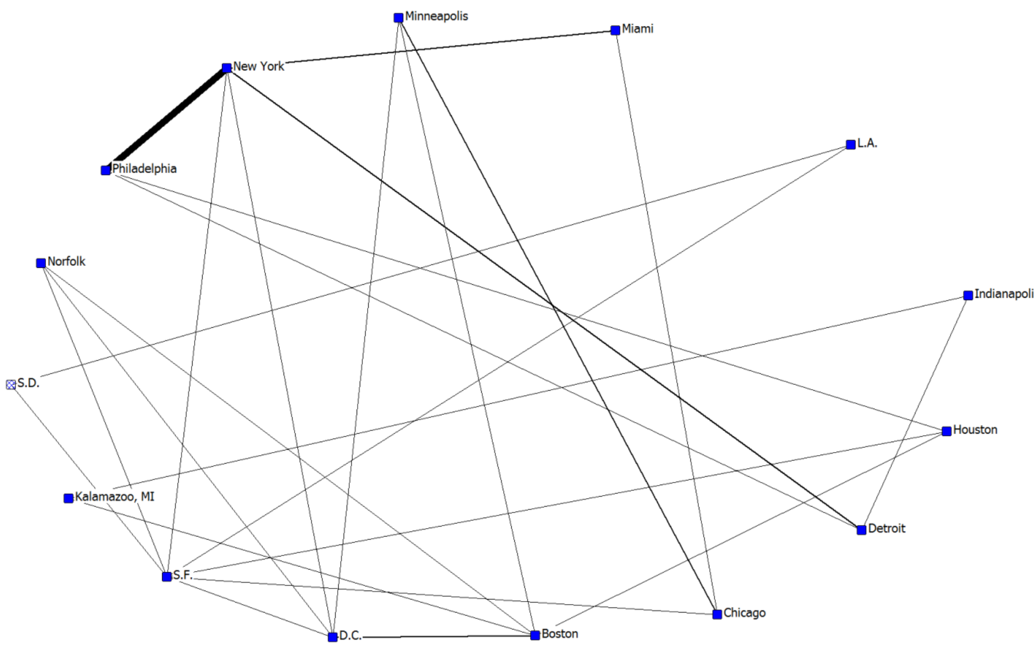


Figure 7.2 Core of the 1979 co-invention network (tied to two and more others)

(Nebraska); and St. Joseph (Missouri), Kansas City (Missouri-Kansas), and Lawrence (Kansas).

In the period 1979-1989, the total number of intermetropolitan co-patenting ties increased from 47 to 187, which was approximately a fourfold expansion. Figure 7.3 shows the 1989 intermetropolitan co-invention network, which had a more complicated structure and less fragmentation compared with a decade earlier. Beyond having two minor components and 60 isolates, the main network component consisted of 86 metropolitan areas. New York, co-patenting with 37 metropolitan partners, remained the national hub for biotechnology co-invention, followed by San Francisco with 24 partners, and Washington D.C.-Baltimore and Philadelphia with each having 17 partners. Figure 7.4 presents the dense part of the network with 13 metropolitan areas that were adjacent to at least five other members of the cohesive group. By conducting an *m*-slices procedure, connections between these 13 metropolitan areas are progressively removed as the frequency of co-patenting is increased. When the multiplicity of the lines (i.e., the frequency of co-invention) is increased from six to eight, for example, only three intermetropolitan pairs remain. New York and Philadelphia (39 ties), New York and Boston (8 ties), and New York and Indianapolis (8 ties) had the strongest co-patenting relationships. While Champaign-Urbana (Illinois) and Port St. Lucie (Florida) were excluded from the main component, both metropolitan areas collaborated with one another by means of biotechnology co-patenting. A large public university (University of Illinois at Urbana-Champaign) in the former

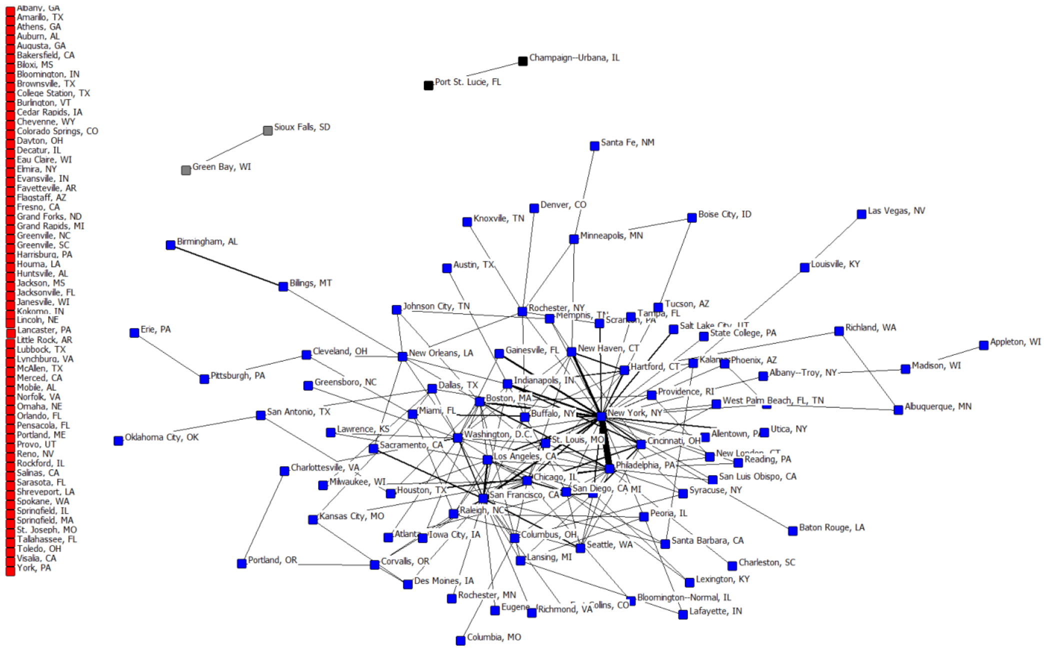


Figure 7.3 Intermetropolitan co-invention network in 1989

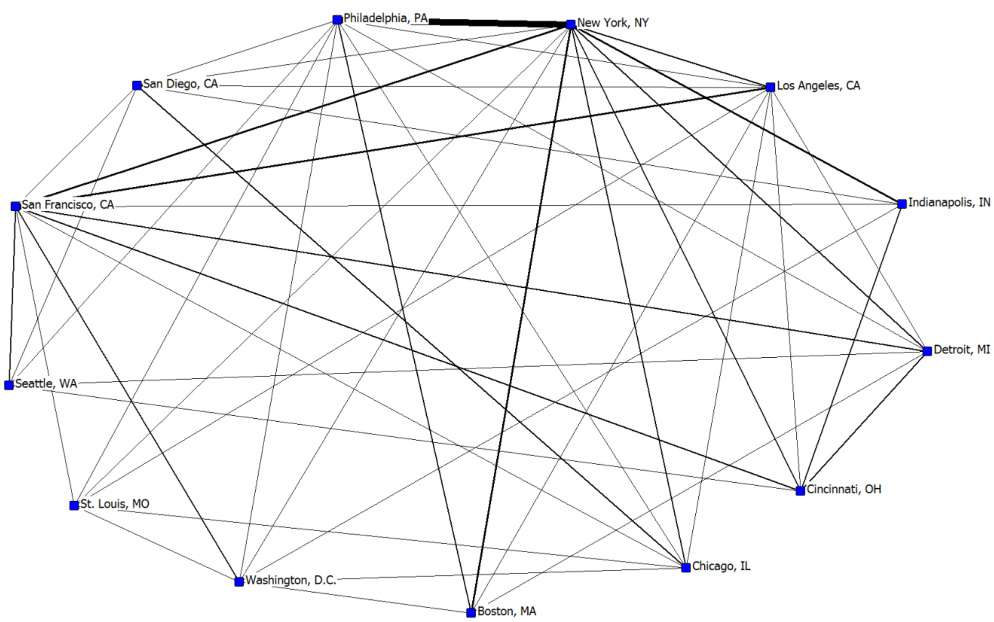


Figure 7.4 Core of the 1989 co-invention network (tied to five and more others)

and a cluster of medical and clinic laboratories in the latter has sustained this long distance collaborative relationship.

In 1999, a growing number of inventors participated in non-local co-patenting, causing intermetropolitan ties to increase to 690 from 187 ten year earlier. The co-invention network, as shown in Figure 7.5, had a dense main component, along with one minor component and 19 isolates. With a total of 129 metropolitan areas engaged in the main network-based system, New York and San Francisco were the two leading centers for biotechnology co-invention with 63 and 59 metropolitan partners, respectively. Philadelphia and Boston had 52 partners. Figure 7.6 shows a cohesive group of 22 metropolitan areas whose minimum network degrees were 13. When the multiplicity of the lines is increased to 23, eight metropolitan areas remain. The closest relationship occurred between New York and Philadelphia with 132 co-patenting ties. The connections between New York and New Haven (39 ties), San Francisco and Boston (39 ties), and Boston and Washington D.C.-Baltimore (37 ties) occupied a less connected second tier. Only one co-patenting pair separated from the main component. These two metropolitan areas – Brownsville-San Benito and McAllen-Mission – were geographically contiguous at the southern tip of Texas.

Turning to 2009, the co-invention network included a main component containing 138 metropolitan areas and 12 isolates. Figure 7.7 shows a dense web of intermetropolitan relationships. Co-patenting ties increased nearly 30 percent over the previous decade, from 690 in 1999 to 887 in 2009. San Francisco, collaborating with 75 metropolitan areas, replaced New York (69 neighbors) as

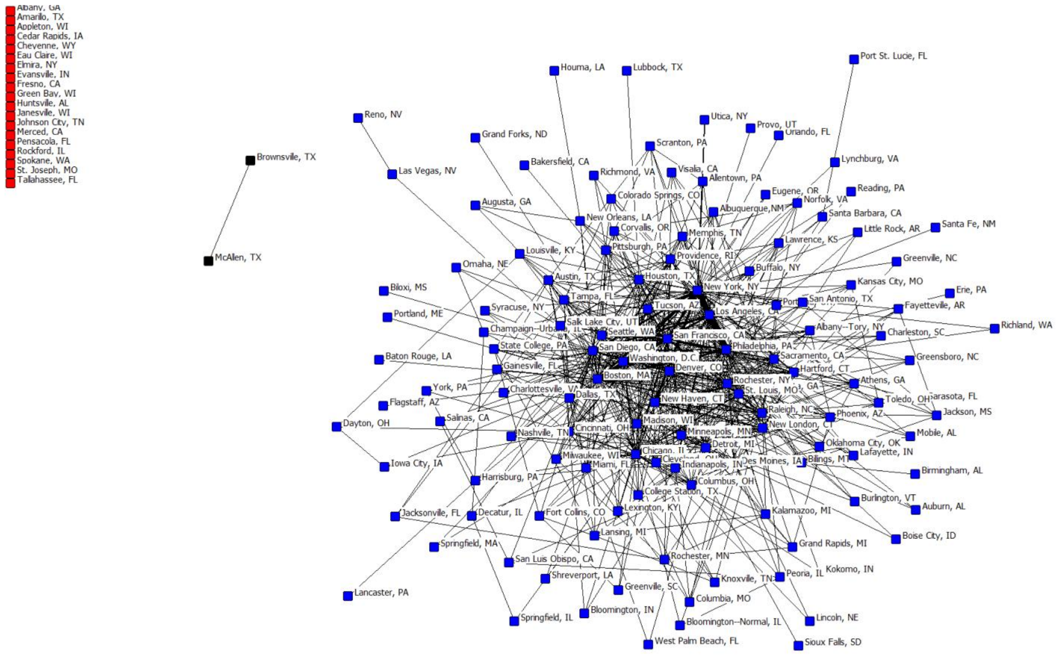


Figure 7.5 Intermetropolitan co-invention network in 1999

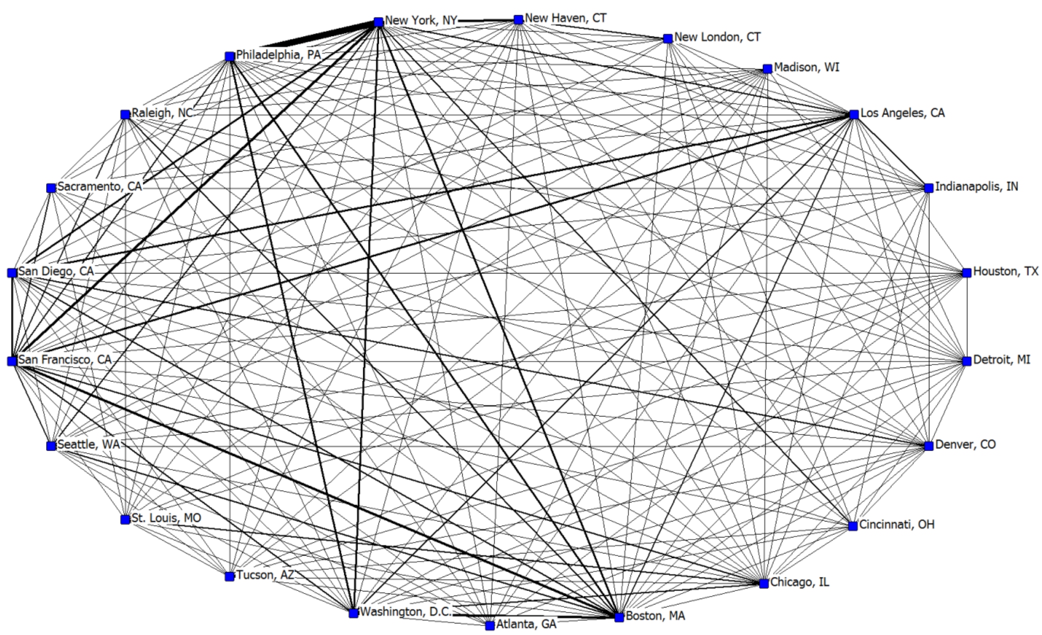


Figure 7.6 Core of the 1999 co-invention network (tied to 13 and more others)

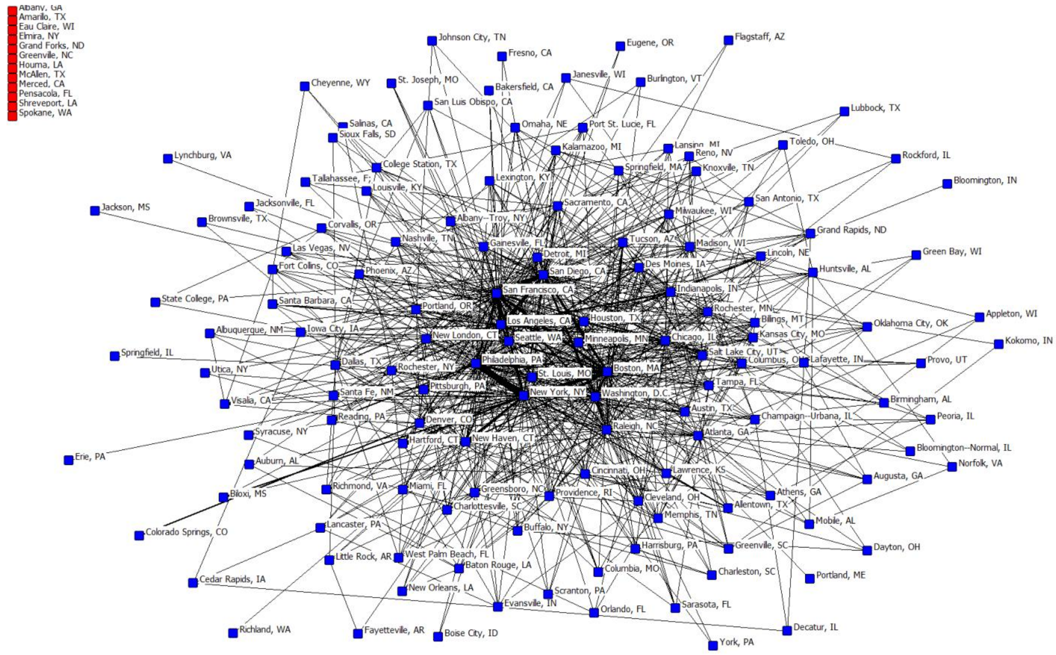


Figure 7.7 Intermetropolitan co-invention network in 2009

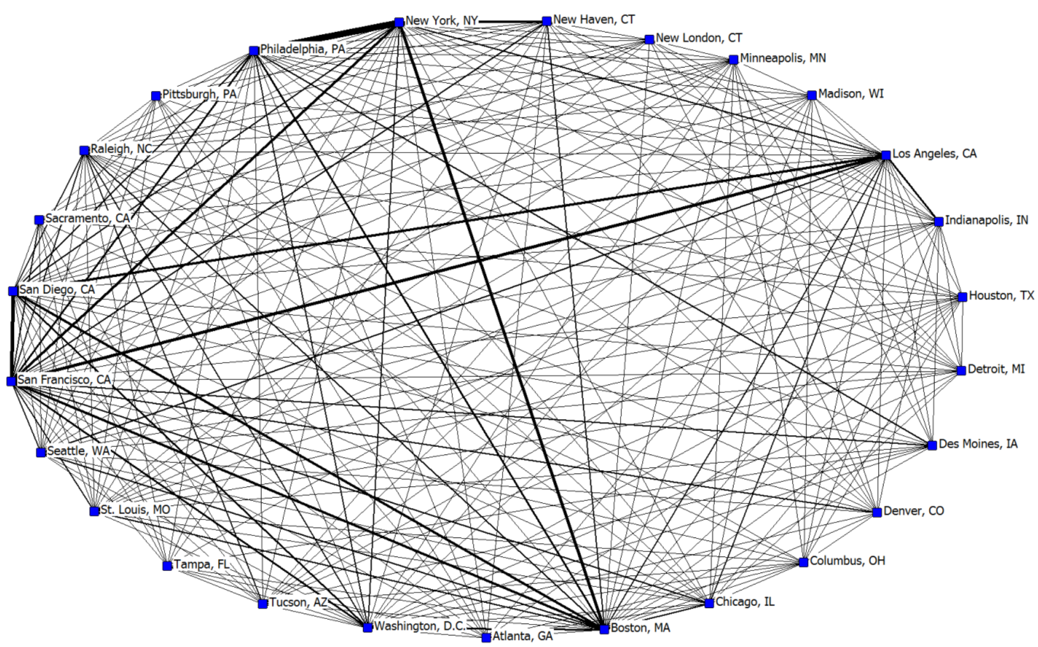


Figure 7.8 Core of the 2009 co-invention network (tied to 13 and more others)

the national leading co-invention center. Boston (66 neighbors), San Diego (64 neighbors), and Los Angeles (60 neighbors) were also national leaders in biotechnology co-invention. Figure 7.8 shows that 26 metropolitan areas were interconnected within the cohesive part of the network where each area was tied to at least 13 other members. More intense co-patenting ties between metropolitan areas emerged along the Northeast Corridor including the pairs of New York and Philadelphia (162 ties), and New York and Boston (65 ties). In the West, intense co-patenting ties were evident between San Francisco and San Diego (78 ties), following by San Francisco and Los Angeles (64 ties). In addition, San Francisco was strongly tied to New York.

7.2 Intermetropolitan Network Centralization and Centrality

The purpose of this section is to determine which metropolitan areas are central to the co-invention network in transferring knowledge. Three types of network centrality measures – degree, closeness, and betweenness – are calculated and interpreted. These measures describe the locations of individual metropolitan areas in terms of how close they are to the "center" of the action in a network. Degree centrality assesses the number of direct ties that an area has to other areas. It is the most important and straightforward way to identify a network center. A metropolitan area is viewed as the center of a co-invention network if it has the most directly connected neighbors. Closeness and betweenness centralities measure individual areas' reachabilities across the entire network by both direct and indirect ties. Closeness centrality implies that a metropolitan area with shorter geodesic distances to all other areas is more engaged in intermetropolitan knowledge exchange. This type of centrality is a distance measure that cannot be computed if the network is not fully connected. This requires excluding isolates and minor components from the analysis and focusing on the main component alone. Betweenness centrality reflects the extent to which a metropolitan area located on geodesic paths connecting other areas plays a relatively important role as an intermediary or gatekeeper for information exchange.

Each type of centrality measure is also applied to global-level centralization to examine the extent to which an entire network has a centralized structure. Network centralization is calculated by assessing variability of individual nodes' centralities. When the measure is large, it indicates that few metropolitan areas are highly central and the remaining areas occupy much less central positions in the network. In contrast, if the measure is low, it means that metropolitan areas are connected with others that have similar central positions in the network (Kang 2007).

Table 7.2 summarizes the descriptive statistics of network centralization for the intermetropolitan co-invention networks using degree, closeness, and betweenness measures. The co-invention network in 2009 had the highest degree centralization score indicating that few metropolitan areas with a greater proportion of direct ties were highly central to the overall network structure. Knowledge flows were increasingly concentrated in a small number of major biotechnology centers. The closeness measure shows a slight variation across the four years of analysis. The betweenness trend has a profile that differs from the degree and closeness measures. The co-invention network in 1989 had the highest betweenness centralization score indicating that central metropolitan areas served as critical gatekeepers for controlling knowledge flows to other areas. In contrast, the co-invention network in 2009 had the lowest betweenness centralization score indicating that relatively more metropolitan areas were interconnected. Knowledge flows in the 2009 co-invention network were less likely to be controlled by a small number of intermediaries.

Previous research shows that a city's inventive performance is closely tied to its position in networks of intercity exchange (e.g., Alderson and Beckfield 2004; Ponds et al. 2007; Varga and Parag 2009; Alderson et al. 2010; Neal 2011). This dissertation explores this relationship further by comparing three types of network centrality measures with co-patenting performance in biotechnology across U.S. metropolitan areas. Table 7.3 shows the correlations among these three centrality measures (degree, closeness, and betweenness), the natural logarithms of biotechnology co-patents (noted *ln_co-patents*), the natural logarithms of the number of wage and salary jobs (noted *ln_labor force*), and the biotechnology co-patenting rate[12]. Degree, closeness, and betweenness centralities were strongly correlated in 1999 and in 2009 ($r = 0.922\sim0.720$; $p < 0.01$). Metropolitan areas with high degree centrality in the network are also likely to have high closeness and betweenness centralities. Major co-inventive cities occupy central network positions allowing them to spread or receive

[12] Each observed MSA's biotechnology co-patenting rate is calculated by dividing its annual biotechnology co-patents by the total number of wage and salary jobs. In order to compensate for co-patenting instability in small MSAs, original rates are smoothed using an Empirical Bayes Smmother (see Section 4.1 for detail).

influence from the entire network, and control knowledge flows between other cities. An area's biotechnology co-patents and the number of its wage and salary

Table 7.2 Descriptive statistics of network centralization

	1979	1989	1999	2009
Number of MSAs	29	86	129	138
Number of ties	39	185	689	887
Degree centralization	.204	.394	.415	.460
Closeness centralization	.348	.468	.421	.449
Betweenness centralization	.400	.492	.163	.129

Note: * all measures are calculated based on each year's main component.
* Networks have been dichotomized.

Table 7.3 Pearson correlation coefficient matrix between three indicators of network centrality, ln_co-patents[1], ln_labor force[2], and co-patenting rate in biotechnology[3]

	Degree	Closeness	Betweenness	ln(co-patents)	ln(labor force)
1999 (n=129)					
Closeness	.900*				
Betweenness	.900*	.720*			
ln_co-patents	.853*	.872*	.675*		
ln_labor force	.706*	.679*	.601*	.680*	
Co-patenting rate	.508*	.495*	.407*	.581*	-.040
2009 (n=138)					
Closeness	.900*				
Betweenness	.922*	.749*			
ln_co-patents	.901*	.867*	.766*		
ln_labor force	.695*	.692*	.600*	.710*	
Co-patenting rate	.508*	.462*	.457*	.568*	-.065

Note: * correlation is significant at the .01 level (two-tailed).
[1] ln_co-patents is the natural logarithm of biotechnology co-patents.
[2] In_labor force is the natural logarithm of the number of wage and salary jobs.
[3] co-patenting rate is calculated by dividing its annual biotechnology co-patents by the number of wage and salary jobs.

jobs are positively correlated ($r = 0.680$ in 1999, $r = 0.710$ in 2009; $p < 0.01$), as are the co-patenting rate and the number of biotechnology co-patents ($r = 0.581$ in 1999, $r = 0.568$ in 2009; $p < 0.01$). However, there is no significant relationship between an area's co-patenting rate and the size of its employment. While several large cities such as San Francisco, Boston, San Diego, and Philadelphia have high co-patenting rates, some small areas with fewer jobs are heavily engaged in co-patenting. These smaller co-inventive cities include Des Moines (Iowa), Rochester (Minnesota), Athens (Georgia), Lawrence (Kansas), and Corvallis (Oregon). Degree centrality has a strong correlation with biotechnology co-patent counts ($r = 0.853$ in 1999, $r = 0.901$ in 2009; $p < 0.01$) and the number of wage and salary jobs ($r = 0.706$ in 1999, $r = 0.695$ in 2009; $p < 0.01$), but has a smaller significant relationship with the co-patenting rate ($r = 0.508$ in 1999 and in 2009; $p < 0.01$). Major co-inventive cities with high degree centralities have high co-patenting rates, while some small areas with less degree centrality scores are also engaged in co-patenting activity with others. Closeness and betweenness centralities show positive and significant relationships with co-patents, the size of employment, and the co-patenting rate, but the correlation coefficients are relatively low. A city's degree centrality appears to have stronger influence on its co-inventive performance compared with closeness and betweenness centralities.

Table 7.4 lists the top 15 metropolitan areas ranked by degree, closeness, and betweenness centralities in the period 1979-2009. The areas listed in the degree and closeness rankings mostly overlap. Since 1999, San Diego, Los Angeles, and Seattle in the West and Raleigh-Durham in the East had replaced

Table 7.4 Ranking of MSAs on measures of network centrality

1979	Degree	Share%	1979	Closeness	Share%	1979	Betweenness	Share%
San Francisco	0.286	10.3	New York	0.500	5.1	New York	0.464	20.4
New York	0.286	10.3	San Francisco	0.483	5.0	San Francisco	0.447	19.6
D.C.	0.214	7.7	D.C.	0.444	4.6	Miami	0.160	7.0
Boston	0.179	6.4	Houston	0.394	4.1	Philadelphia	0.154	6.8
Philadelphia	0.179	6.4	Philadelphia	0.389	4.0	D.C.	0.146	6.4
Chicago	0.143	5.1	Detroit	0.384	3.9	Raleigh	0.138	6.0
Detroit	0.107	3.8	Chicago	0.384	3.9	Chicago	0.128	5.6
Miami	0.107	3.8	Miami	0.378	3.9	Detroit	0.117	5.1
Houston	0.107	3.8	Boston	0.373	3.8	Boston	0.088	3.8
Norfolk	0.107	3.8	Norfolk	0.368	3.8	Indianapolis	0.087	3.8
Minneapolis	0.107	3.8	Minneapolis	0.359	3.7	Cincinnati	0.071	3.1
Indianapolis	0.107	3.8	Raleigh	0.346	3.6	Los Angeles	0.071	3.1
Los Angeles	0.107	3.8	Los Angeles	0.341	3.5	College Station	0.071	3.1
Buffalo	0.071	2.6	San Diego	0.337	3.5	Houston	0.066	2.9
Cincinnati	0.071	2.6	W. Palm Beach	0.337	3.5	Kalamazoo	0.028	1.2

1989	Degree	Share%	1989	Closeness	Share%	1989	Betweenness	Share%
New York	0.435	10.0	New York	0.586	1.9	New York	0.509	25.9
San Francisco	0.282	6.5	San Francisco	0.497	1.6	San Francisco	0.209	10.6
D.C.	0.200	4.6	D.C.	0.491	1.6	D.C.	0.131	6.7
Philadelphia	0.200	4.6	Chicago	0.459	1.5	New Orleans	0.082	4.2
Chicago	0.153	3.5	Indianapolis	0.455	1.5	Rochester, NY	0.071	3.6
Los Angeles	0.141	3.2	Philadelphia	0.455	1.5	Chicago	0.065	3.3
Boston	0.129	3.0	Los Angeles	0.452	1.5	Philadelphia	0.062	3.1
San Diego	0.118	2.7	Boston	0.443	1.4	Miami	0.054	2.7
Indianapolis	0.106	2.4	San Diego	0.440	1.4	Indianapolis	0.053	2.7
Cincinnati	0.106	2.4	Cincinnati	0.434	1.4	Kalamazoo	0.050	2.5
Detroit	0.106	2.4	Detroit	0.434	1.4	Boston	0.048	2.5
New Haven	0.094	2.2	Miami	0.431	1.4	Cleveland	0.048	2.4
New Orleans	0.094	2.2	St. Louis	0.431	1.4	Albany-Troy	0.047	2.4
Raleigh	0.082	1.9	New Haven	0.427	1.4	Dallas	0.041	2.1
Rochester, NY	0.082	1.9	Buffalo	0.423	1.4	New Haven	0.036	1.8

1999	Degree	Share%	1999	Closeness	Share%	1999	Betweenness	Share%
New York	0.492	4.6	New York	0.650	1.1	New York	0.172	12.8
San Francisco	0.461	4.3	San Francisco	0.640	1.1	San Francisco	0.128	9.5
Philadelphia	0.414	3.8	Philadelphia	0.618	1.1	Boston	0.107	8.0
Boston	0.406	3.8	Boston	0.615	1.1	Philadelphia	0.094	7.0
D.C.	0.375	3.5	D.C.	0.601	1.1	Chicago	0.093	6.9
San Diego	0.352	3.3	San Diego	0.595	1.0	San Diego	0.077	5.7
Chicago	0.344	3.2	Chicago	0.590	1.0	D.C.	0.074	5.5
Raleigh	0.313	2.9	Los Angeles	0.579	1.0	Raleigh	0.061	4.6
Los Angeles	0.313	2.9	Raleigh	0.569	1.0	Los Angeles	0.053	3.9
Denver	0.227	2.1	Denver	0.542	1.0	Minneapolis	0.039	2.9
Detroit	0.203	1.9	Detroit	0.531	0.9	Salk Lake City	0.034	2.5
New Haven	0.203	1.9	New Haven	0.531	0.9	Des Moines	0.030	2.2
St. Louis	0.195	1.8	Tucson	0.529	0.9	Houston	0.027	2.0
Des Moines	0.195	1.8	Sacramento	0.525	0.9	Cincinnati	0.025	1.9
Seattle	0.188	1.7	Seattle	0.525	0.9	Detroit	0.023	1.7

2009	Degree	Share%	2009	Closeness	Share%	2009	Betweenness	Share%
San Francisco	0.547	4.2	San Francisco	0.688	1.1	San Francisco	0.137	11.3
New York	0.518	4.0	New York	0.675	1.0	New York	0.130	10.8
Boston	0.482	3.7	Boston	0.656	1.0	Boston	0.099	8.2
San Diego	0.467	3.6	San Diego	0.643	1.0	San Diego	0.099	8.2
Los Angeles	0.438	3.4	Los Angeles	0.637	1.0	Los Angeles	0.079	6.6
D.C.	0.431	3.3	Philadelphia	0.628	1.0	D.C.	0.076	6.3
Philadelphia	0.423	3.3	D.C.	0.623	1.0	Philadelphia	0.066	5.5
Raleigh	0.328	2.5	Raleigh	0.585	0.9	Chicago	0.042	3.4
Chicago	0.314	2.4	Chicago	0.578	0.9	Indianapolis	0.039	3.2
Seattle	0.292	2.3	Seattle	0.573	0.9	Raleigh	0.036	3.0
St. Louis	0.270	2.1	St. Louis	0.564	0.9	Seattle	0.035	2.9
Indianapolis	0.226	1.7	New Haven	0.555	0.9	Madison	0.030	2.5
Madison	0.226	1.7	Madison	0.548	0.9	Atlanta	0.025	2.1
New Haven	0.219	1.7	Indianapolis	0.548	0.9	Dallas	0.022	1.8
Atlanta	0.212	1.6	Houston	0.539	0.8	Denver	0.019	1.6

Detroit, Miami, Houston, and Cincinnati as the most centralized areas in the network. Biotechnology is a new industry that is knowledge-based and is predominantly produced by new start-ups and small firms. Some studies argue that new biotechnology firms have largely concentrated around major universities and research centers from which company founders spun off their commercial organizations (e.g., Audretsch 2001; Niosi and Banik 2005; Cooke 2007). These emerging biotechnology centers all occur in areas containing major research universities, existing technological infrastructure, and entrepreneurial culture. For example, the existence of the University of Washington and research institutions in the South Lake Union neighborhood has been a major impetus to foster Seattle's leadership in biotechnology invention (The Bioscience Brief 2007).

Below the top ten of the 2009 list shown in Table 7.4, there were minor discrepancies between the degree and closeness rankings. For example, Indianapolis ranked 12th highest in the degree and 14th highest in the closeness. Atlanta only appeared in the degree, while Houston was solely evident in the closeness. The betweenness measure was more skewed compared with the two other measures. Forty percent of betweenness centrality was contributed by New York and San Francisco in 1979 and 37 percent in 1989. Although the share of betweenness centrality dropped to 22 percent in 2009, both major biotechnology centers are still situated on the geodesic paths between most pairs of metropolitan areas and crucial to transmit information throughout the entire network. Several differences in the betweenness compared with the degree and closeness rankings are noteworthy. Raleigh-Durham, for example, ranked sixth highest in the 1979 betweenness, and was not included in the top ten areas in the two other measures in that year. More cases were evident in 1989 where New Orleans (Louisiana), Rochester (New York), and Kalamazoo (Michigan) were only listed in the betweenness ranking. Indianapolis's ranking position in the 2009 betweenness was higher than that in the degree and closeness measures. These metropolitan areas with high betweenness centrality scores play an intermediary role in knowledge exchange as they most likely fall on the geodesic paths between other members of the network. Compared with some major biotechnology centers, these metropolitan gatekeepers have a high potential of control on the indirect relations of the other members. For example, Raleigh-Durham (in the 1979 network); New Orleans, Rochester (New York), Kalamazoo (in the 1989 network); and Indianapolis (in the 2009 network) occupy a favorable position of go-between in the network by wielding power over interactions between non-adjacent cities. It is important to identify metropolitan gatekeepers when studying intermetropolitan co-invention networks to increase knowledge flow opportunities for other areas. In short, no matter which of network centralities is concerned, San Francisco, New York, Boston, San Diego ranking at the top of the lists are most central to the contemporary biotechnology co-invention network across American cities.

7.3 Structures of Biotechnology Network-Based Systems

The previous section established a ranking of U.S. metropolitan areas in terms of network centrality. This section describes the co-invention network-based system these areas form and the positions played by different types of areas within the system of knowledge exchange. The focus is to trace the structural properties of the co-invention network-based system based on the positions played by metropolitan areas in the network. In the previous section, the 2009 co-invention network reveals that a small number of biotechnology centers dominate the network-based system, and that minor areas with little co-patenting tend to collaborate with these major areas but not to co-invent with other minor areas. A common image in social network analysis and particularly in collaboration networks that illustrates these findings is that of the core/periphery structure (Alderson and Beckfield 2004; Alderson et al. 2010; Rubí-Barceló 2010). In a simple core/periphery structure, a group of core actors is densely connected and the complementary set of peripheral actors is simply connected to some members of the core (Borgatti and Everett 1999; Cross et al. 2001; Rubí-Barceló 2010). If the U.S. co-invention network-based system is consistent with a core/periphery structure, the system should be able to be characterized by: (1) core metropolitan areas that play an active role in the system as they are internally connected with each other and to some outsiders, and (2) peripheral metropolitan areas that play a passive role in the system as they are loosely connected or even disconnected from one another.

To answer this question, I focus on the 2009 network data and use regular equivalence as the criterion to partition individual metropolitan areas into relational sets (or positions) composed of areas that have the same relations to members of other equivalent sets (Alderson and Beckfield 2004; Cattani and Ferriani 2008; Alderson et al. 2010; Balland et al. 2011). The REGE algorithm is calculated to obtain a symmetric similarity matrix for degrees of regular equivalence between pairs of metropolitan areas by their geodesic distances (Borgatti and Everett 1993; Wasserman and Faust 1994; Borgatti et al. 2002).[13] I apply a single-link hierarchical clustering to partition metropolitan areas into positions and employ a blockmodeling approach to generalize about the nature of relations between positions in the 2009 co-invention network-based system.[14]

Figure 7.9 shows the dendrogram for hierarchical clustering of the similarity matrix using the REGE algorithm. The similarity level at which any pair of metropolitan areas is aggregated is the point at which both can be reached by tracing from left to right. The figure identifies various levels of regular equivalent positions composed of metropolitan areas that have the same relation to members of other equivalent positions. The similarity level of 99.998 percent is the most inclusive level at which members of a position have a high degree of similarity. Table 7.5 lists five regular equivalent sets identified via hierarchical clustering with an explanation of the roles they perform in the network. San Francisco and New York stand out as National Co-invention Centers forming the *primary* position (titled Position 1). These two powerful centers connect with over 70 metropolitan areas nationwide. The next equivalent set contains 38 metropolitan areas categorized as Regional Co-invention Centers including Boston, Philadelphia, Indianapolis, San Diego, Chicago, Seattle, and Washington

[13] For the sake of consistency with the previous section, I only focus on the 138 metropolitan areas located in the main component of the 2009 co-invention network.
[14] In essence, blockmodeling is a data reduction technique that systematically searches for relational pattern in a network by regrouping actors and presenting condensed aggregate-level information (Knoke and Yang 2008).

Figure 7.9 Hierarchical clustering of similarity matrix from the REGE algorithm

Table 7.5 Characteristics of regularly equivalent sets and positions

Set name	Ave. degree centrality	Ave. closeness centrality	Ave. Betweenness centrality	Self-ties[1]/All ties[2]	Position
National co-invention center ($n = 2$)	0.533	0.682	0.134	1/73	P1
Regional co-invention center ($n = 38$)	0.194	0.535	0.022	277/504	P2
Median co-invention MSA ($n = 9$)	0.062	0.470	0.001	5/38	P3
Minor co-invention MSA ($n = 58$)	0.057	0.457	0.002	41/226	P4
Non-significant co-invention MSA ($n = 31$)	0.022	0.384	0	0/46	P5

Note: three indicators of network centrality are calculated based on the 2009 main component.
[1] Self-ties are the number of ties that members of the set are self-connected.
[2] All ties are the number of ties involved by members of the set.

D.C.-Baltimore. These metropolitan areas occupy the *major* position (titled Position 2) in the co-invention network because they highly collaborate with each other. Nearly 55 percent of the co-patenting ties (277 out of 504 ties) in this position are contributed by the members themselves. Nine metropolitan areas from the third equivalent set are categorized as Median Co-invention MSAs occupying the *median* position in the network (titled Position 3). With limited co-inventive activities, these nine metropolitan areas such as Buffalo, Santa Barbara, and Albuquerque rank just above the average of each centrality measure and only 13 percent of the ties (5 out of 38 ties) in this position are internally connected. Fifty-eight metropolitan areas from the fourth equivalent set are categorized as Minor Co-invention MSAs occupying the *minor* position in the network (titled Position 4) as their centrality-status within the network is well below the national average. These areas are less connected with each other and the proportion of internal ties is less than 18 percent. The last equivalent set consists of 31 metropolitan areas with lowest degree, closeness, and betweenness centralities. They are categorized as Non-significant Co-invention Areas and occupy the outskirts of the network-based system (titled Position 5). Table 7.5 shows that metropolitan areas in this position are isolated from one another (self-ties = 0). In a core/periphery structure, these areas are arrayed on the periphery in the sense that all their relations are with areas located in more central positions, but not with each other.

What kind of structural feature would be most appropriate to describe the U.S. co-invention network-based system? I employ a blockmodeling approach to detect and characterize the nature of relations between these defined positions. The term *block* refers to a square submatrix of regularly equivalent metropolitan areas that have similar relations to the areas occupying the other blocks (Wasserman and Faust 1994; Knoke and Yang 2008). A block filled completely with "1s" is titled *1-block*, while a block filled completely with "0s" is titled *0-block*. In a classic core/periphery interaction model, the core/core submatrix is a 1-block indicating the presence of ties between core areas. The periphery/periphery submatrix is a 0-block indicating the absence of ties between peripheral areas. The core/periphery submatrix representing ties between the core and the peripheral areas can be either 1-block or 0-block. I assume that metropolitan areas in both the primary and major positions (Positions 1 and 2) comprise the core of the network because they are highly connected with each other and to some other areas on the outskirts. Metropolitan areas in the median and the minor positions, as well as non-significant co-inventive MSAs (Positions 3, 4, and 5) comprise the periphery because they have substantial relationships with core metropolitan areas, but not with each other.

Given a partition of metropolitan areas into core and peripheral blocks a priori, I use a QAP correlation procedure as a basis for testing the presence of a core/periphery structure in the co-invention network-based system (Borgatti and Everett 1999). It is conducted by checking the independence of two proximity matrices – the ideal core/periphery pattern matrix and the 2009 real binary adjacent matrix. Two idealized versions of pattern matrices that correspond to a core/periphery interaction model are created (Borgatti and Everett 1999). The first is that the ties exist in both core/core and core/periphery blocks (seen as 1-blocks) and no ties are found in the periphery/periphery block (seen as a 0-block), as shown in Figure 7.10. The alternative version is that the ties are only found in the core/core block (seen as a 1-block) and no ties are found in all other blocks (seen as 0-blocks), as shown in Figure 7.11. In the 2009 binary adjacent matrix, rows and columns are permuted according to the arrays of core and peripheral metropolitan areas. The relationships between the adjacent matrix and both the ideal core/periphery pattern matrices are statistically significant, but the correlation coefficient for the alternative version of pattern matrix ($r = 0.387$, $p < 0.001$) is larger than that for the first version ($r = 0.264$; $p < 0.001$). This indicates that the U.S. co-invention network-based system is much closer to the alternative core/periphery structure than the first structure. It suggests that most major

		Core				Periphery					
		1	2	3	4	5	6	7	8	9	10
Core	1		1	1	1	1	1	1	1	1	1
	2	1		1	1	1	1	1	1	1	1
	3	1	1		1	1	1	1	1	1	1
	4	1	1	1		1	1	1	1	1	1
Periphery	5	1	1	1	1		0	0	0	0	0
	6	1	1	1	1	0		0	0	0	0
	7	1	1	1	1	0	0		0	0	0
	8	1	1	1	1	0	0	0		0	0
	9	1	1	1	1	0	0	0	0		0
	10	1	1	1	1	0	0	0	0	0	

Figure 7.10 Ideal core/peripheral pattern matrix (the first version)

		Core				Periphery					
		1	2	3	4	5	6	7	8	9	10
Core	1		1	1	1	0	0	0	0	0	0
	2	1		1	1	0	0	0	0	0	0
	3	1	1		1	0	0	0	0	0	0
	4	1	1	1		0	0	0	0	0	0
Periphery	5	0	0	0	0		0	0	0	0	0
	6	0	0	0	0	0		0	0	0	0
	7	0	0	0	0	0	0		0	0	0
	8	0	0	0	0	0	0	0		0	0
	9	0	0	0	0	0	0	0	0		0
	10	0	0	0	0	0	0	0	0	0	

Figure 7.11 Ideal core/peripheral pattern matrix (the alternative version)

biotechnology centers are densely connected in co-patenting; some minor metropolitan areas are significantly dependent on major centers for invention collaboration, but most of them are loosely connected.

Beyond testing for a core/periphery structure based on a hypothesis testing of a priori partitions, I examine the structure underlying the co-invention network-based system from the relational data. Blockmodeling is used to detect the relations between and within positions and generates three forms of output: density table, image matrix, and reduced graph. A *density table* is a square matrix that has positions (rather than individual actors) as its rows and columns. The value of each cell titled *cell density* is the proportion of ties that are present from metropolitan areas in each pair of positions. For the density within a position, main diagonal elements in that submatrix are excluded from calculations since self-ties are not considered in the present analysis. An *image matrix* is also a square matrix, obtained from the density table by recoding each cell density to either one or zero, representing the presence or absence of relation between each pair of positions. The density of the full matrix is used as an average-based cut-off point to dichotomize the image values. When a cell density is greater than or equal to the cut-off point, a unity value is assigned to the matrix cell; otherwise, the cell is set to zero. A *reduced graph* is a graphical representation of the image matrix where positions are nodes and the relationships among pairs of positions are connecting lines. A “1” in a cell of the image matrix indicates that the two corresponding positions are adjacent to each other; otherwise, they are nonadjacent.

Figures 7.12-7.13 show the density table and the image matrix of the five positions. For determining the 1-blocks, a density of the entire matrix of 0.094 or higher is used. It is apparent that both the primary and major positions (Positions 1 and 2) occupying the core of the system are mutually dependent (cell density = 1.0) and are also tied to the median and minor positions (Positions 3 and 4). In contrast, Positions 3 and 4 clearly belong to the periphery of the network as they are only tied to the core but not to each other. Their respective cell densities with Positions 1 and 2 are all greater than the cell density of themselves (density between Positions 3 and 4 = 0.015). Position 5 consisting of 31 non-significant co-inventive MSAs is isolated from the other positions in the system.

Figure 7.14 is the reduced graph corresponding to the image matrix shown in Figure 7.13. This result is similar to that obtained from the a priori partition. The only difference identified between the two resulting structures is that Position 5 arrayed on the outskirts of the network is not detected as a part of the core/periphery structure that is included in the peripheral region a priori. The disconnect between Position 5 and the core region (including Positions 1 and 2) is because their respective densities (0.00 and 0.05) are all lower than the cut-off density value of 0.094, causing a “no-tie” condition in the structure. In fact, examination of the density table in Figure 7.12 indicates that these non-significant co-inventive MSAs in Partition 5 are mostly tied to the core metropolitan areas, particularly the members of Position 2 (cell density 0.050). They have few connections with each other or other peripheral areas (cell density 0.017~0).

	P1	P2	P3	P4	P5
P1	1.000	1.000	1.000	0.431	0.000
P2	1.000	0.394	0.111	0.127	0.050
P3	1.000	0.111	0.139	0.015	0.007
P4	0.431	0.127	0.015	0.025	0.017
P5	0.000	0.050	0.007	0.017	0.000

Note: The entire network density (cut-off point) is 0.094.

Figure 7.12 Density table of co-invention network-based system (at the position level)

	P1	P2	P3	P4	P5
P1	1	1	1	1	0
P2	1	1	1	1	0
P3	1	1	1	0	0
P4	1	1	0	0	0
P5	0	0	0	0	0

Figure 7.13 Image matrix of co-invention network-based system (at the position level)

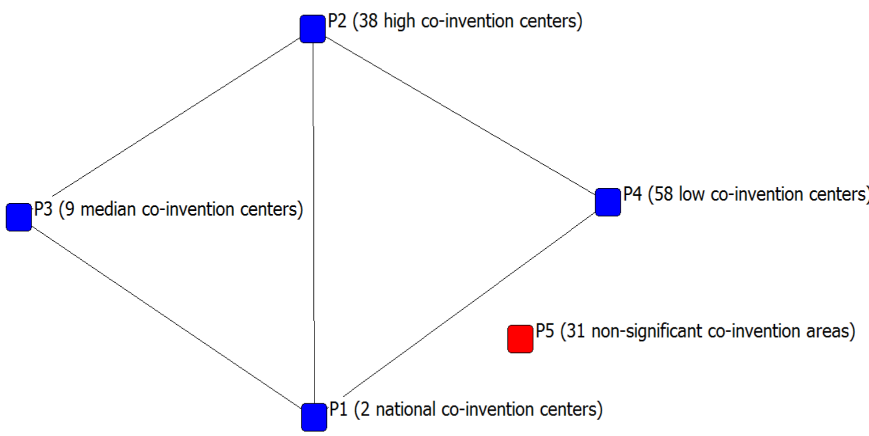


Figure 7.14 Reduced graph of co-invention network-based system

Figure 7.15 shows a 2×2 density table in the core/periphery partition basis with Position 5 added to the periphery, along with Positions 3 and 4. The density of the core/core block (0.433) is much higher compared with the core/periphery block (0.114), while the periphery/periphery block is the lowest (0.018). By modifying the previous image matrix of Figure 7.13 to account for the revised core/periphery structure, Figure 7.16 shows the image matrix of the co-invention network-based system, in which the core/core block is a 1-block, the core/periphery block is a partial 1-block, and the periphery/periphery block is nearly a 0-block.

In short, given this core/periphery partition layout, the results suggest that the U.S. co-invention network-based system is consistent with a relational core/periphery structure that Borgatti and Everett (1999) originally characterized. Figure 7.17 shows the characteristics of the core/periphery network-based system. The core of the system is the locus of a high level of knowledge sharing and collaboration. It comprises 40 metropolitan areas that occupy the primary and major positions in the co-invention network. These core metropolitan areas are the most active in the system as they account for 40 percent of the total co-patenting ties that are internally connected and 50 percent of the total ties that are involved in some levels of collaborations with outsiders. On the other hand, the periphery of the network-based system is composed largely of 98 metropolitan areas that occupy the median and minor positions in the network, as well as non-significant co-inventive MSAs. These metropolitan areas play a passive role in the system, even though 50 percent of their total co-patenting ties are connected to

	Core	Periphery
Core	0.433	0.114
Periphery	0.114	0.018

Note: The core includes P1 and P2 and the periphery includes P3, P4, and P5.

Figure 7.15 Density table of co-invention network-based system
(at the core/periphery partition basis)

		Core		Periphery		
		P1	P2	P3	P4	P5
Core	P1	1	1	1	1	0
	P2	1	1	1	1	0
Periphery	P3	1	1	1	0	0
	P4	1	1	0	0	0
	P5	0	0	0	0	0

Note : The core includes P1 and P2 and the periphery includes P3, P4, and P5.

Figure 7.16 Image matrix of co-invention network-based system
with the core/periphery partition basis (at the position level)

	Core	Periphery
Core	354 ties (40%)	446 ties (50%)
Periphery		87 ties (10%)

Note: there were 887 intermetropolitan co-patenting ties in 2009

Figure 7.17 Characteristics of the core/periphery network-based system the core. They are more or less located at the periphery with weaker levels of connectivity (10 percent of the total co-patenting ties are internally connected).

7.4 Summary

This chapter provides results showing the structural properties of intermetropolitan networks of biotechnology co-invention from three perspectives. The first perspective is to identify components and cohesive subgroups of metropolitan areas within each year's co-invention network. The major component of the 1979 co-invention network consisted of 29 metropolitan areas, along with six minor components and 107 isolates. In 1989, the largest connected set of the network included 86 metropolitan areas, along with 2 minor components and 60 isolates. More inventors in a few metropolitan areas participated in non-local co-patenting in 1999, causing that the largest connected set of metropolitan areas had expanded to 129 MSAs, along with one minor component and 19 isolates. A more organized web of intermetropolitan relationships was evident in the 2009 co-invention network where 138 MSAs were engaged in the main network-based system, along with 12 isolates.

The second perspective is to identify the importance of metropolitan areas in transforming knowledge using centralization and centrality measures. The 1989 network had the most betweenness-based centralized feature with selected metropolitan areas serving as critical gatekeepers for controlling knowledge flows to other areas. The co-invention network in 2009 had the lowest betweenness centralization score. Knowledge flows in that year were less likely to be controlled and mediated by a small number of intermediaries. Results of the analysis of the local network centrality measures suggest that metropolitan areas with high degree centrality in the network are also likely to have high closeness and betweenness centralities. These three centralities strongly correlate with a metropolitan area's biotechnology co-inventive performance. San Francisco, New York, Boston, and San Diego are identified as four major American biotechnology concentrations. Some small and medium metropolitan areas with high betweenness centrality scores such as Raleigh-Durham, New Orleans, Rochester, Kalamazoo, Des Moines, and Indianapolis were in a favorable position to mediate knowledge flows compared with other major biotechnology centers.

The third perspective is to trace the structural properties of the co-invention network-based system using the regular equivalence criterion and the blockmodeling approach. The result suggests that the 2009 network-based system is consistent with a relational core/periphery structure. Core metropolitan areas play an active role in the system as they are strongly connected to one another and to some outsiders. Peripheral metropolitan areas play a passive role as they are loosely connected or even disconnected with each other.

Chapter 8

SUMMARY AND CONCLUSIONS

This chapter provides a general summary of the dissertation, its findings, and conclusions. Section 8.1 gives a brief overview of the research topics, conceptual framework, and methodologies. Section 8.2 summarizes the main research findings. Section 8.3 presents the contributions of the dissertation. Section 8.4 points out the limitations and future research directions.

8.1 Overview

This dissertation provides insights into the network structures of intermetropolitan knowledge flows in co-invention networks of American biotechnology. Conventionally, the theory of localized knowledge spillovers (LKSs) emphasizes that geographical concentration of inventive firms in clusters enhances knowledge exchange and diffusion because human skills and know-how are bounded in space. However, other scholars stress the role of collaborative networks in which individuals and groups are embedded in webs of social relationships through direct connections and indirect linkages (Rondé and Hussler 2005; Maggioni et al. 2007; Knoben 2009; Wilhelmsson 2009). This is especially apparent in high-technology industries such as biotechnology where scientists and engineers from different locations collaborate to advance inventive performance (Coe and Bunnell 2003; McKelvey et al. 2003; Gertler and Levitte 2005; Birch 2007; Cooke 2007; Ponds et al. 2007). The existence of collaborative networks raises two critical challenges for the investigation and understanding of the geography of information exchange. Is the increasingly network nature of technological collaboration likely to modify geographical structures of knowledge flows? Are knowledge flows via networks becoming less constrained by geography? The purpose of this dissertation is to understand the dynamics of network aspects of knowledge flows in American biotechnology. I investigate knowledge flows by concentrating on intermetropolitan and not intra-metropolitan knowledge collaboration and exchange. This particularly occurs in biotechnology co-invention because it involves "over the distance" interactions between inventors from different locations (Breschi and Lissoni 2004; Maggioni et al. 2007), and has a high dependence on global networking relationships (Owen-Smith and Powell 2004; Coenen et al. 2004; Gertler and Levitte 2005; Fontes 2005; Coenen et al. 2006; Cooke 2006). Intermetropolitan co-invention networks are constructed by tracking inventors who participate in American biotechnology co-patenting and attributing each co-patent to metropolitan areas where the inventors reside (e.g., Felix 2006; Ejermo and Karlsson 2006; Maggioni et al. 2007; Maraut et al. 2008).

The main argument of the dissertation is that the space of knowledge flows in biotechnology co-invention is not tightly bounded within territories and neighboring areas, but circulates around alignments of economic actors in different or even distant locations. The first research task asks whether the biotechnology co-invention urban system reveals significant differences between its spatial and network-based dependencies. The longitudinal changes in these dependencies are also explored. I compare spatial and network-based dependencies revealed in global- and local-level measures of association across

150 large U.S. metropolitan areas over four decades (1979, 1989, 1999, and 2009) to roughly coincide with important advances in information and communication technologies that most likely influence the co-invention urban system. Initiating the analysis in 1979 captures the start of the personal computer (PC) age, the rise of e-mail and PC networking occurred around 1989, and 1999 ushered in the use of search engines (e.g., Google, Yahoo) to obtain information on the Internet. Techniques in exploratory spatial data analysis (ESDA) are used to assess the relative importance of spatial versus network-based proximity on the biotechnology co-invention urban system. Particularly, Moran's *I* is used to detect global-level spatial and network-based dependencies, while the local-level measure of dependence is based on the *local indicators of spatial association* (also called LISA). I use LISA cluster maps to identify local groupings of biotechnology co-invention into five distinct patterns: Co-invention Cores (high-high), Co-invention Peripheries (low-low), High Co-invention Islands (high-low), Low Co-invention Islands (low-high), and Non-significant Areas. Two types of LISA cluster maps – spatial and network-based – are compared to identify significant cores, peripheries, and islands across U.S. metropolitan areas. Results of this task provide insights into the relative importance of spatial and network-based proximities in the space of knowledge flows. The second research task is to understand how and to what extent biotechnology flows circulate in network-based systems by focusing on the structural properties of intermetropolitan co-invention networks from three distinct perspectives: components and cohesive subgroups of metropolitan areas, intermetropolitan network centralization and

centrality, and positions established within the co-invention network-based system.

The characteristics of network components reveal patterns of connection among metropolitan areas. Specific questions addressed include: Is the network-based space structured into groups of metropolitan areas centered on several U.S. biotechnology centers? Do the member metropolitan areas within each group demonstrate certain types of spatial associations? Are these metropolitan areas intensely connected? I answer these questions by identifying components and cohesive subgroups of metropolitan areas within the co-invention network. Both the *k*-cores and the *m*-slices procedures are conducted to determine cohesive subgroups of metropolitan areas with dense knowledge flows underlying each period's co-invention network.

The center of a network identifies metropolitan areas with the best access to knowledge flows. Specific questions tackled in this part include: How tightly organized is the network around its most central metropolitan area(s)? How important is a metropolitan area in transferring knowledge to other areas? To what extent does a metropolitan area control or mediate knowledge flows in the network? I answer these questions by identifying the importance of metropolitan areas in transferring knowledge via two levels of network center measure – centralization and centrality. The former examines the extent to which an entire network has a centralized structure, while the latter describes the locations of individual metropolitan areas in terms of how close they are to the "center" of the action in a network. Both centralization and centrality are calculated and interpreted by three different perspectives: degree, closeness, and betweenness. Degree-based measures describe the extent to which a metropolitan area directly connects to other areas. Closeness-based measures assess how a metropolitan area accesses knowledge flows not only by directly connecting to its neighbors but also through chains of intermediaries. Betweenness-based measures explore how a metropolitan area controls or mediates interaction between nonadjacent areas.

Network positions show the co-invention network-based system these metropolitan areas form and the roles played by different types of areas within the system of knowledge exchange. Scott (2000) argued that the underlying structure of a network is more apparent in the relations of positions than among individual nodes themselves. Understanding how network positions of metropolitan areas form provides an empirical glimpse into the U.S. network-based system of knowledge flows and uncovers whether the system is consistent with a relational core/periphery structure. Specific questions addressed in this part include: Do some U.S. metropolitan areas have similar network positions? Do the varying positions of metropolitan areas reveal a hierarchical cluster structure? What are the relationships among these positions established within the network-based system? The regular equivalence criterion that identifies metropolitan areas having similar patterns of co-patenting ties is used to partition individual areas into network positions. A blockmodeling approach is further employed to detect and characterize the nature of relations between these defined positions. The aims of conducting these empirical analyses in the second research task are to understand network properties of co-invention, the diverse positions of metropolitan areas in systems of knowledge exchange, and how these properties and positions change over time.

8.2 Research Findings

Results of the spatial and network-based proximities in biotechnology co-patenting rates over the four years of analysis suggest that the latter better define American co-invention relationships. The negative global Moran's *I* coefficient found in 2009 particularly indicates that U.S. metropolitan areas with dissimilar co-patenting rates are significantly network-based associated. This global finding might be interpreted as evidence that areas with low co-patenting rates are significantly dependent on ties to areas with high co-patenting rates. Inventors in minor co-inventive cities perhaps actively seek to establish relations with partners in major co-inventive cities for biotechnology collaboration. The absence of significant positive global network-based dependence implies that most minor co-inventive cities have few links with other minor co-inventive cities. Most major co-inventive cities are not significantly linked to the others.

Analysis of local dependence using the spatial LISA cluster maps shows few discernable spatial associations in any of the years of analysis. While significant spatial LISA clusters were largely absent in 1979, a minor Co-invention Core (high-high) occurred in the Midwest focused on Lafayette (Indiana), St. Louis (Missouri), St. Joseph (Missouri), and Des Moines (Iowa). Several Low Co-invention Islands (low-high) were also identified in this region. In 1989, a distinct Co-invention Core emerged in the Northeast centered on New

York, Boston, Albany-Troy (New York), New Haven (Connecticut), and Burlington (Vermont) where co-invention was largely tied to pharmaceutical technologies. San Francisco and San Diego categorized as High Co-invention Islands (high-low) were two noticeable new appearances in the West, along with Memphis (Tennessee) and State College (Pennsylvania). A sizeable medical center in the Memphis metropolitan area and a large public university (Pennsylvania State University) in State College led to unusually high co-patenting rates compared with their nearest neighbors. The 1999 spatial LISA cluster map shows that San Francisco and San Diego were High Co-invention Islands. A distinct Co-invention Periphery (low-low) was evident in the Southeast centered on Jackson (Mississippi), Mobile (Alabama), New Orleans (Louisiana), and Pensacola (Florida). In 2009, this Co-invention Periphery noticeably expanded, stretching from east Texas to Alabama. A small Co-invention Core was evident in the Midwest around St Louis (Missouri), Iowa City (Iowa), and Rochester (Minnesota). San Francisco and San Diego remained as High Co-invention Islands.

Analysis of local dependence using the network-based LISA cluster maps over the four years reveals that a few metropolitan areas have emerged as Co-invention Cores with significant ties to distant partners, but evidence of some regional biases is also noted. In the early years of the period, most co-patenting activities were conducted by local inventors as 83.2 percent of co-patenting activities were mainly localized within the same metropolitan area (see Table 5.3 for details). Raleigh-Durham (North Carolina) was the only significant network-based Co-invention Core in 1979. As a growing number of inventors joined non-local co-patenting ties in 1989, more discernible network-based Co-invention Cores emerged. New York, collaborating with 37 nationwide metropolitan areas, was the primary Co-invention Core for biotechnology co-invention. Detroit-Ann Arbor (Michigan) and Indianapolis (Indiana) were also identified as Co-invention Cores but their neighboring structures showed some regional differences. Detroit-Ann Arbor was strongly related to Boston, New York, and Philadelphia, as well as neighboring Lansing-East Lansing (Michigan) and Cincinnati (Ohio), while Indianapolis was primarily tied to New York, Washington D.C.-Baltimore, Cincinnati, as well as Austin (Texas) and New Orleans (Louisiana) in the South.

In 1999, San Francisco and Boston were categorized as Co-invention Cores where many metropolitan areas aligned their resources with these two centers for biotechnology collaboration. Although both centers were highly dependent with each other, their neighboring structures showed different types of associations. San Francisco was extensively engaged in co-patenting with major cities across the U.S. including New York, San Diego, Los Angeles, and Washington D.C.-Baltimore. In contrast, Boston’s closest network-based partners were largely concentrated in the Northeast including New York, Philadelphia, Providence (Rhode Island), and Washington D.C.-Baltimore. The evidence of regional effects was also evident in some small and medium co-inventive core areas. For example, inventors located in New London and Indianapolis tended to primarily team up with inventors in Northeastern cities (i.e., Boston, New Haven, Hartford, Providence) and Midwestern cities (i.e., Chicago, Lafayette, Bloomington), respectively.

In 2009, network-based Co-invention Cores expanded including New York, San Francisco, Washington D.C.-Baltimore, Boston, Denver, Seattle, and Raleigh-Durham. Several college towns were also focuses of the defined core areas including Fort Collins (Colorado State University), Iowa City (University of Iowa), Lafayette (Purdue University), Lansing-East Lansing (Michigan State University), Lexington (University of Kentucky), Bryan-College Station (Texas A&M University), and Bloomington-Normal (Illinois State University). In addition, metropolitan areas such as Santa Fe (New Mexico), Rochester (Minnesota), and New London (Connecticut) that host large medical centers or pharmaceutical companies all belonged to the 2009 network-based Co-invention Cores. Some studies argue that biotechnology inventive firms have largely concentrated around major universities and research centers (e.g., Audretsch 2001; Niosi and Banik 2005; Cooke 2007). Small and medium Co-invention Cores identified here occurred in areas having these two essential requirements for biotechnology invention.

In comparing the neighboring structures of intermetropolitan collaboration, several noticeable regional features among Co-invention Cores were found. First, San Francisco's strongest partners were national including San Diego, Los Angeles, New York, Boston, Philadelphia, and Washington D.C.-Baltimore. In contrast, New York's strongest partners mostly resided in the East including Boston, New Haven, Hartford, Philadelphia, and Washington D.C.-

Baltimore. Second, inventors in Seattle had strong ties with partners in San Francisco, Los Angeles, and San Diego, while inventors located in Raleigh-Durham had mostly national ties. In short, these recent network-based neighboring structures suggest that the co-patenting relationships of major biotechnology centers are national and regional but not spatial. Spatial proximity is less important for intermetropolitan collaboration compared with both network and regional relationships.

The architecture of the U.S. intermetropolitan co-invention networks reveals a trend toward more organized structures and less fragmentation over the four years of analysis. The major component of the 1979 co-invention network consisted of 29 metropolitan areas, along with six minor components and 107 isolates. The strongest partnership occurred between New York and Philadelphia with an annual frequency of co-patenting of 17. The next closest pairs were New York and Detroit, New York and Miami, and Chicago and Minneapolis. In 1989, intermetropolitan ties increased from 47 to 187. The largest connected set of the network included 86 metropolitan areas, along with 2 minor components and 60 isolates. New York, containing 37 metropolitan partners, remained the national co-patenting hub, followed by San Francisco with 24 partners, and Washington D.C.-Baltimore and Philadelphia with each having 17 partners. The strongest co-patenting relationships remained centered on New York, extending to Philadelphia, Boston, and Indianapolis. In 1999, more inventors in a few metropolitan areas participated in non-local co-patenting, causing annual intermetropolitan ties to increase to 690. The largest connected set of metropolitan areas had expanded to 129 MSAs, along with one minor component and 19 isolates. The closest link occurred between New York and Philadelphia. New York and New Haven, San Francisco and Boston, and Boston and Washington D.C.-Baltimore occupied a less connected second tier.

A more organized web of intermetropolitan relationships was evident in the 2009 co-invention network where 138 MSAs were engaged in the main network-based system, along with 12 isolates. San Francisco, collaborating with 75 metropolitan areas, replaced New York (69 neighbors) as the national leading co-invention center. Boston (66 neighbors), San Diego (64 neighbors), and Los Angeles (60 neighbors) were also leaders in biotechnology co-invention. By connecting and collaborating with more and more inventors across metropolitan boundaries, the network of biotechnology co-invention became more complex. The most intense intermetropolitan co-patenting ties consistently existed along the Northeast Corridor including the pairs of New York and Philadelphia, New York and Boston. A considerable proportion of co-patenting ties were also accounted for by San Francisco and San Diego, and San Francisco and Los Angeles. In addition, San Francisco was strongly tied to New York.

Using the global centralization measures to investigate the tightness of the intermetropolitan co-invention networks, I found that degree centralization scores increased constantly, indicating that knowledge flows were gradually concentrated in a small number of major biotechnology centers. The closeness centralization shows a slight variation along the four years' co-invention networks. As for the betweenness trend, the 1989 network had the most betweenness-based centralized feature with selected metropolitan areas serving as critical gatekeepers for controlling knowledge flows to other areas. These metropolitan areas included New York, San Francisco, Washington D.C.-Baltimore, and New Orleans had a high potential of control on the indirect relations of the other member areas. In contrast, the co-invention network in 2009 had the lowest betweenness centralization score indicating that relatively more metropolitan areas were interconnected. Knowledge flows in the 2009 co-invention network were less likely to be controlled and mediated by a small number of intermediaries.

Results of the analysis of the local network centrality measures lead to several conclusions. First, metropolitan areas with high degree centrality in the network are also likely to have high closeness and betweenness centralities. Most major co-inventive centers directly connecting many metropolitan partners occupy central positions, which allow them to reach influence throughout the entire network, and to control knowledge flows between other cities. These three (degree, closeness, and betweenness) centralities strongly correlate with a metropolitan area's biotechnology co-inventive performance (i.e., co-patenting counts, the number of wage and salary jobs, and the co-patenting rate). This result corroborates previous findings that a city's inventive performance is closely tied to its central position in networks of knowledge exchange (e.g., Alderson and Beckfield 2004; Ponds et al. 2007; Varga and Parag 2009; Alderson et al. 2010; Neal 2011). Second, a small number of metropolitan areas occupy the central positions in the co-invention network. Particularly, San Francisco, New York, Boston, and San Diego ranking at the top of the three network centralities are most central to the contemporary biotechnology co-invention network across American cities. These four areas are often identified as major American biotechnology concentrations, which corroborates Cooke's (2006; 2007) findings that these four are megacenters in the global network of biotechnology collaboration. Cooke (2006) elaborated that global megacenters integrate individual cities into a system of "open innovation" that stretches biotechnology knowledge domains of academic research and large pharmaceutical companies over space (also Coenen et al. 2004; Moodysson et al. 2008). Third, some small and medium metropolitan areas with high betweenness centrality scores play an intermediary role in knowledge exchange as they most likely fall on the geodesic paths between other members of the network. For example, Raleigh-Durham (ranked sixth in the 1979 network); New Orleans, Rochester (New York), Kalamazoo (ranked fourth, fifth, and tenth in the 1989 network, respectively); Des Moines (ranked 12th in the 1999 network); and Indianapolis (ranked ninth in the 2009 network) are in a favorable position to mediate knowledge flows compared with some major biotechnology centers. They generally occupy the position of go-between in the network by wielding power over interactions between non-adjacent cities (Müller-Prothmann 2007). It is important to identify metropolitan gatekeepers when studying intermetropolitan co-invention networks to increase knowledge flow opportunities for other areas.

The regular equivalence criterion and the blockmodeling approach to describing network structure suggest that the 2009 U.S. co-invention network-based system is consistent with a relational core/periphery structure that Borgatti and Everett (1999) originally characterized. This is a system in which core metropolitan areas are strongly connected to one another and to some peripheral areas. Conversely, metropolitan areas in the periphery are loosely connected with each other (Alderson and Beckfield 2004; Cattani and Ferriani 2008; Alderson et al. 2010; Rubí-Barceló 2010). The core of the system is the locus of a high level of knowledge sharing and collaboration. It consists of 40 metropolitan areas with two types of regular equivalent positions in the network. One is composed of two national co-invention centers – San Francisco and New York – occupying the primary position in the network-based system. The other contains 38 regional co-invention centers occupying the major position in the network. These areas include Boston, Philadelphia, Indianapolis, San Diego, Chicago, Seattle, and Washington D.C.-Baltimore. Members of these two positions in the core region are the most active in the system as they account for 40 percent of the total co-patenting ties that are internally connected and 25 percent of the total ties that are involved in some levels of collaborations with outsiders. On the other hand, the periphery is composed of 98 metropolitan areas with three types of regular equivalent positions in the network. These include nine median co-invention MSAs occupying the median position in the network, 58 minor co-invention MSAs occupying the minor position in the network, and 31 metropolitan areas having non-significant co-inventive links with others. These peripheral metropolitan areas play a passive role in the system, even though 25 percent of their total co-patenting ties are connected to the core. They are more or less located at the periphery with weaker levels of connectivity (10 percent of the total co-patenting ties are internally connected).

This core/periphery property is typical of many network structures and does not have to be interpreted as a disadvantage or limitation of biotechnology development. Barabási (2005) argued that most scientific networks follow a trend of core reinforcement and growth through the periphery. Members of the core establish conventions and norms that favor knowledge exchange and circulation, while those in the periphery constitute a pool of potential recruits who can bring fresh and new ideas into the system (Cattani and Ferriani 2008; Balland et al. 2011). Conventionally, geographical clustering of economic actors has been seen as an efficient structure that favors the geography of invention and innovation. The data show that biotechnology co-patenting in 2009 accounted for 62.3% intra-metropolitan knowledge flows (see Table 5.3 for details). However, as Balland et al. (2011) argued that "focusing only on geographical clusters is a narrow view of innovations occurring in most technological fields" (p. 3). Technological and commercial successes of many firms in clusters are generally embedded in larger network structures where close ties to non-local knowledge partners are also critical (Trippl et al. 2009; Balland et al. 2011).

8.3 Contributions of the Dissertation

Differences of inventive activity in space has been evidenced and largely explained by the tacit nature of knowledge and the theory of localized knowledge spillovers. An important issue is to go beyond distance-based concepts of relationships to understanding how the geography of invention and innovation is shaped by structural properties of intermetropolitan networks. This dissertation contributes to this challenge, focusing on the network structures of knowledge flows in American biotechnology at the metropolitan scale. I exploit a dataset that details key relationships linking U.S. metropolitan areas in co-invention networks. This dissertation focuses on three central issues. First, exploratory spatial data analysis is explicitly applied to assess the relative importance of spatial versus network-based proximity on the biotechnology co-invention urban system. Second, network analysis techniques are applied to examine the geographical patterns of collaboration, especially in relation to the structural properties of co-invention networks, and to describe its spatial heterogeneity. Third, an empirical assessment of the network-based system is conducted to describe precisely what kind of co-invention system metropolitan areas form, and the positions played by different types of areas within the system of knowledge exchange.

This dissertation extends understanding of the urban geography of knowledge flows in two main directions. First, it works with useful new insights into the geography of invention and innovation by emphasizing the role of intermetropolitan networks in knowledge collaboration and exchange. While spatial clustering enhances knowledge production and diffusion, biotechnology invention and innovation increasingly connect non-local partners (Owen-Smith and Powell 2004; Coenen et al. 2004; Coke 2006). Interactive learning and collaboration are not tightly bounded within given territories. Cities act as functional nodes immersed in wider networks where knowledge exchange are decisive forces for technological advance (Amin 2002; Cowan and Jonard 2004; Amin and Cohendet 2005; Maskell et al. 2006; Sunley 2008; Autant-Bernard et al. 2010). This dissertation provides empirical evidence to support the argument that technological collaboration reveals a network-based system associated with different or even distant geographical places (Coenen et al. 2004; Cooke 2006; Bathelt 2007; Kroll and Mallig 2009). The U.S. biotechnology communities have demonstrated national and regional associations of invention since the 1990s. Associations retain features of spatial proximity especially in some Midwestern and Northeastern cities, but these are no longer the strongest features affecting co-inventive links. This evidence of the increasingly network aspects of knowledge exchange provides a new conception of the geography of cooperation, which is somewhat different from the conventional theory of localized knowledge spillovers that once dominated understanding of the role of geography in technological advance (Jaffe 1989; Acs et al. 1992; Audretsch and Feldman 1996).

Second, this dissertation investigates knowledge flows using a core/periphery model of intermetropolitan co-patenting, which reflects the presence of regional disparities within the U.S. biotechnology communities. The spatial distribution of co-inventive activities is highly concentrated in a small number of leading cities. Inventors cooperate in both sets of leading cities and certain lagging cities. This trend has implications for public policies, particularly in the peripheral areas. The ability of economic actors (e.g., inventors, firms, and research institutions) to participate in collaborative work depends on their attractiveness as a potential partner. Policy makers should identify the characteristics of collaborative networks and enhance local research capability, with the aim of entering into wider networks. Economic actors in the peripheral areas need to develop sufficient internal skills and competencies (e.g. internal research and development, diversity of available competencies) in order to benefit from external knowledge flows for local development (Cohen and Levinhal 1990; Simmie 2003; Fontes 2005; Autant-Bernard et al. 2010).

The methodological contribution of the dissertation is the application of social network concepts and techniques to trace out exactly how the structure of intermetropolitan co-invention network has changed over time and space. Social network analysis techniques are a powerful method for investigating the interplay between geographical and relational structures in studies of invention and innovation (Boschma and Frenken 2009). Most previous studies have examined the structure of inter-organizational knowledge communities (e.g. Owen-Smith and Powell 2004) or networks of local interpersonal relations (e.g. Fleming and Frenken 2006), but lack a broad geographical context (Glückler 2007). This study explores network structures of knowledge flows by examining co-patenting links between inventors who reside in different locations of the U.S. urban system. It improves on earlier analysis by incorporating understanding of the structure of collaborative networks at the metropolitan scale, and by assessing the roles of metropolitan areas in network-based systems. Results of these structural properties of intermetropolitan networks are suited to go beyond the conventional theory of localized knowledge spillovers as a dominated source of agglomeration of inventive activity in space.

8.4 Limitations and Future Research Directions

In this dissertation, I have mapped and described important intermetropolitan co-invention networks, but there are at least three limitations and future research directions. First, the co-invention networks are constructed by attributing each co-patent to metropolitan areas where the inventors reside. Co-patenting links in the networks are used as a way to account for bilateral knowledge flows between metropolitan areas, but the directionality of connections is ignored. However, non-directional data is not available to distinguish between incoming flows (in-degree) and outgoing flows (out-degree) held by an individual network member. The former can define an area's popularity or prestige, and the latter an area's power or influence. Areas with many incoming ties are considered particularly prominent or have high levels of expertise, while areas with many outgoing ties are also influential. To further understand the characteristics of individual metropolitan areas in a collaborative network, future research should consider directional data analysis.

Second, this dissertation has investigated the network aspects of knowledge exchange by connecting collaborative inventors from different metropolitan areas. However, it did not take into account the majority of co-patenting within the same metropolitan areas (accounted for around 80% in 1979 and 60% in 2009) that may be likely to influence inventors' collaboration choices and the structure of intermetropolitan co-invention networks. Bala and Goyal

(2004) argued that the decision to collaborate is made by weighing the costs and benefits of interacting. Recently Autant-Bernard et al. (2007) used a model of cooperation choice including three sets of determinants (firm's collaborative capacity, network positions, and geographical locations) to test the presence of spatial effects relative to network effects in European collaboration in nanotechnologies. Their results confirm that geographical proximity is no longer regarded as the main determinant of agglomeration. Local network effects can contribute to reinforcing the geographical concentration at the European and intra-national levels. Future research on this subject should attempt to identify the relevant factors driving distant collaboration, which will bring forth new insights in the geography of invention and innovation.

Finally, American biotechnology heavily relies on global sources of knowledge. Co-patenting activities span not only national but also international territories, which corroborates Wagner and Leydesdorff's (2005) and Roijakkers and Hagedoorn's (2006) findings that networks of collaboration are becoming more international. While this dissertation primary focuses on the domestic U.S. context and does not investigate international collaboration, future research should assess the extent to which U.S. inventors establish distant relationships with foreign partners. Other future research directions also include: (1) cross-sectoral comparison of co-invention networks between biotechnology and other high-technology industries such as semiconductors, and (2) comparison of collaborative networks between co-inventors in patent applications and co-authors in academic publications. The first direction concerns that the conditions for generating inventions may vary between different technologies. Some technologies mostly rely on narrow science specialties, while others are more interdisciplinary. The second direction distinguishes between scientific communities (e.g., scholars) and engineering communities (e.g., inventors) by arguing that the former have wide latitude in choosing with whom to collaborate while the latter are more locally tied with one another (Gittelman 2007). The argument of *invisible colleges* emphasizes that academic networks between individual scholars are not tightly bounded by their institutions, but are instead driven by a wide variety of others from a broad range of research institutions and locations (Price and Beaver 1966; Crane 1969). This future research will provide more insights into different types of network-based spaces.

REFERENCES

Acs, Z. J. and C. Armington. 2006. *Entrepreneurship, Geography, and American Economic Growth*. Cambridge, New York: Cambridge University Press.

Acs, Z. J., D. B. Audretsch, and M. P. Feldman. 1992. Real effects of academic research: comment. *American Economic Review 82* (1): 363-67.

Acs, Z. J., D. B. Audretsch, and M. P. Feldman. 1994. R&D spillovers and recipient firm size. *Review of Economics and Statistics 76* (2): 336-40.

Acs. Z. J., L. Anselin, and A. Varga. 2002. Patents and innovation counts as measures of regional production of new knowledge. *Research Policy 31*: 1069-85.

Adelman, D. E. and K. L. DeAngelis. 2007. *Patent metrics: The mismeasure of innovation in the biotech patent debate*. Arizona Legal Studies. Discussion Paper No. 06-10. The University of Arizona. Retrieved from http://ssrn.com/abstract=881842

Agrawal, A., I. Cockburn, and J. McHale. 2006. Gone but not forgotten: Knowledge flows, labor mobility, and enduring social relationships. *Journal of Economic Geography 6*: 571-91.

Alderson, A. S. and J. Backfield. 2004. Power and position in the world city system. *American Journal of Sociology 109* (4): 811-51.

Alderson, A. S., J. Backfield, and J. Sprague-Jones. 2010. Intercity relations and globalisation: the evolution of the global urban hierarchy, 1981-2007. *Urban Studies 47*: 1861-1877.

Allen, J. 2000. Power/economic knowledge: symbolic and spatial formations. In *Knowledge, Space, Economy*, eds. J. Bryson, P. W. Daniels, N. Henry, and J. Pollard, 15-33. London: Routledge.

Almeida, P. and B. Kogut 1999. Localization of knowledge and the mobility of engineers in regional networks. *Management Science 45* (7): 905-17.

Amin, A. 2002. Spatialities of globalization. *Environment and Planning A 34*: 385-99.

Amin, A. and P. Cohendet. 2004. *Architectures of Knowledge: Firms, Capabilities and Communities*. Oxford, U.K.: Oxford University Press.

Amin, A. and P. Cohendet. 2005. Geographies of knowledge formation in firms. *Industry and Innovation 12* (4): 465-86.

Amin, A. and N. Thrift. 1992. Neo-Marshallian nodes in global networks. *International Journal of Urban and Regional Research 16* (4): 571-87.

Anselin, L. 1995. Local indicators of spatial association—LISA, *Geographical Analysis 27*, 93-115.

Anselin, L. 2004. *GeoDa 0.9.5-i Release Notes*. Spatial Analysis Laboratory, Department of Agricultural and Consumer Economics. University of Illinois, Urbana-Champaign, IL. Retrieved from http://geodacenter.org/downloads/pdfs/geoda095i.pdf

Anselin, L., N. Lozano, and J. Koschinsky. 2006a. *Rate transformations and smoothing*. In Spatial Analysis Laboratory Department of Geography. University of Illinois: Urbana—Champaign. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.87.1996&rep=rep1&type=pdf

Anselin, L., I. Syabri, and Y. Kho. 2006b. GeoDa: an introduction to spatial data analysis. *Geographical Analysis 38* (1): 5-22.

Anselin, L., S. Sridharan, and S. Gholston. 2007. Using exploratory spatial data analysis to leverage social indicator databases: the discovery of interesting patterns. *Social Indicators Research 82*: 287-309.

Anselin, L., A. Varga, and Z. Acs. 1997. Local geographic spillovers between university research and high technology innovations. *Journal of Urban Economics 42*: 422-48.

Anselin, L., A. Varga, and Z. Acs. 2000. Geographical spillovers and university research: a spatial econometric perspective. *Growth and Change 31*: 501-15.

Asheim, B. T. 1996. Industrial districts as 'learning regions': a condition for prosperity? *European Planning Studies 4*: 379-400.

Asheim, B. T. and A. Isaksen 2002. Regional innovation system: the integration of local 'sticky' and global 'ubiquitous' knowledge. *Journal of Technology Transfer 27* (1): 77-86.

Audretsch, D. B. 1998. Agglomeration and the location of innovative activity. *Oxford Review of Economic Policy 14* (2): 18-29.

Audretsch, D.B. 2001. The role of small firms in U.S. biotechnology clusters. *Small Business Economics 17*: 3-15.

Audretsch, D. B. and M. P. Feldman. 1996. R&D spillovers and the geography of innovation and production. *The America Economic Review 86* (3): 630-52.

Autant-Bernard, C. 2001. Science and knowledge flows: evidence from the French case. *Research Policy 30*: 1069-78.

Autant-Bernard, C., P. Billand, and D. Frachisse. 2007a. Social distance versus spatial distance in R&D cooperation: empirical evidence from European collaboration choices in micro and nanotechnologies. *Papers in Regional Science 86* (3): 495-519.

Autant-Bernard, C., J. Mairesse, and N. Massard. 2007b. Spatial knowledge diffusion through collaborative networks. *Papers in Regional Science 86* (3): 341-50.

Autant-Bernard, C., M. Fadairo, and N. Massard. 2010. Knowledge diffusion and innovation polices within the European regions: challenges based on recent empirical evidence. *GATE Working Paper No. 1010*. Retrieved from http://ssrn.com/abstract=1625246

Balconi, M., S. Breschi, and F. Lissoni. 2004. Networks of inventors and the role of academia: an exploration of Italian patent data. *Research Policy 33*: 127-45.

Balland, P.-A., R. Suire, and J. Vicente. 2011. The Economic Geography of Technological Standards: Evidence on the European GNSS Industry. Retrieved from http://ssrn.com/abstract=1619855

Bathelt, H. 2002. The re-emergence of a media industry cluster in Leipzig. *European Planning Studies 10* (5): 583-611.

Bala, V. and S. Goyal. 2004. A non-cooperative model of network formation. *Econometrica 68*: 1181-229.

Barabási, A. I. 2005. Network theory – the emergence of the creative enterprise. *Science 5722*: 639-41.

Bathelt, H. 2005. Geographies of production: growth regimes in spatial perspective (II)—knowledge creation and growth in cluster. *Progress in Human Geography 29* (2): 204-16.

Bathelt, H. 2007. Buzz-and-pipeline dynamics: towards a knowledge-based multiplier model of clusters. *Geography Compass 1* (6): 1282-98.

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

Bettencourt, L. M. A., J. Lobo, and D. Strumsky. 2007. Invention in the city: increasing returns to patenting as a scaling function of metropolitan size. *Research Policy 36*: 107-20.

Birch, K. 2007. Knowledge space and biotechnology. *Geography Compass 1* (5): 1097-117.

Boggs, J. S. and N. M. Rantisi. 2003. The 'relational turn' in economic geography. *Journal of Economic Geography 3*: 109-16.

Borgatti, S. P. and M. G. Everett. 1993. Two algorithms for computing regular equivalence. *Social Networks 15*: 361-76.

Borgatti, S. P. and M. G. Everett. 1999. Models of core/periphery structures. *Social Networks 21*: 375-95.

Borgatti. S. P., M. G. Everett, and L. C. Freeman. 2002. *UCINET 6.0 for Windows: Software for Social Network Analysis*. Harvard, MA: Analytic Technologies.

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

Boschma, R. A. and K. Frenken. 2009. The spatial evolution of innovation networks: a proximity perspective. *Papers in Evolutionary Economic Geography* 09/05, Utrecht University. Retrieved from http://econ.geo.uu.nl/peeg/peeg0905.pdf

Boschma, R. A. and A. L. J. Ter Wal. 2007. Knowledge networks and innovative performance in an industrial district: the case of a footwear district in the south of Italy. *Industrial and Innovation 14* (2): 177-99.

Breschi, S. and F. Lissoni. 2001a. Knowledge spillovers and local innovation systems: a critical survey. *Industrial and Corporate Change 10* (4): 975-1005.

Breschi, S. and F. Lissoni. 2001b. Localized knowledge spillovers vs. innovative milieux: knowledge 'tacitness' reconsidered. *Papers in Regional Science 80* (3): 255-73.

Breschi, S. and F. Lissoni. 2003. Mobility and social networks: localized knowledge spillovers revisited. *CESPRI Working Paper, No. 142*. Universitá Bocconi, Milano.

Breschi, S. and F. Lissoni. 2004. Knowledge network from patent data: methodological issues and research targets. *CESPRI Working Paper 150*. Universitá Bocconi, Milano.

Breschi, S. and F. Malerba. 2001. The geography of innovation and economic clustering: some introductory notes. *Industrial and Corporate Change 10* (4): 817-33.

Bureau of Economic Analysis. 2010. Regional economic accounts local area personal income. Retrieved from http://www.bea.gov/regional/reis/

Callon, M. and G. Bowker. 1994. Is science a public good? Fifth Mullins Lecture, Virginia Polytechnic Institute, 23 March 1993. *Science, Technology, & Human Values 19* (4): 395-424.

Camagni, R. 1991. Local milieu, uncertainty and innovation networks: towards a new dynamic theory of economic space. In *Innovation Networks: Spatial Perspectives*, ed. R. Camagni. London: Belhaven Press.

Cantner, U. and H. Graf. 2006. The network of innovators in Jena: an application of social network analysis. *Research Policy 35*: 463-80.

Cassi, L. and A. Morrison. 2007. Social networks and innovation: concepts, tools, and recent empirical contribution. ISCTE. No. 2007/59. Lisbon, Portugal.

Cattani, G. and S. Ferriani. 2008. A core/periphery perspective on individual creative performance: social networks and cinematic achievements in the Hollywood film industry. *Organization Science 6*: 824-44.

Coe, N. M. and T. G. Bunnell. 'Spatializing' knowledge communities: towards a conceptualization of transnational innovation networks. *Global Networks 3* (4): 437-56.

Coenen, L., J. Moodysson, and B. T. Asheim. 2004. Nodes, networks and proximities: on the knowledge dynamics of the Medicon Valley biotech cluster. *European Planning Studies 12* (7): 1003-16.

Coenen, L., J. Moodysson, C. D. Ryan, B. Asheim, and P. Phillips. 2006. Comparing a pharmaceutical and agro-food bioregion: on the importance of knowledge bases for socio-spatial patterns of innovation. *Industry and Innovation 13* (4): 393-414.

Cohen, W. M. and D. A. Levinthal. 1990. Absorptive capacity: a new perspective on learning and innovation. *Administrative Science Quarterly 35*: 128-52.

Cooke, P. 2001. New economy innovation system: biotechnology in Europe and the U.S.A. *Industry and Innovation 8*: 267-89.

Cooke, P. 2003. Networks and hierarchies in bioscientific knowledge management. Presented at DRUID Summer Conference, June 12-14, 2003, Aalborg, Denmark.

Cooke, P. 2006. Global bioregional networks: a new economic geography of bioscientific knowledge. *European Planning Studies 14* (9): 1265-85.

Cooke, P. 2007. *Growth Cultures: The Global Bioeconomy and its Bioregions*. New York, NY: Routledge.

Cortright, J. and H. Mayer. 2002. *Signs of life: the growth of biotechnology centers in the U.S.* The Brookings Institution, Washington D.C. Retrieved from www.brookings.edu/~/media/Files/rc/reports/2002/06labormarkets_joseph%20cortright%20and%20heike%20mayer/biotech.pdf

Cowan, R. 2005. Networks models of innovation and knowledge diffusion. In *Clusters, Networks and Innovation*, eds. S. Breschi and F. Malerba, 29-53. Oxford, U.K.: Oxford University Press.

Cowan, R. and N. Jonard. 2004. Networks structure and the diffusion of knowledge? *Journal of Economic Dynamics and Control 28*: 1557-575.

Crane, D. 1969. Social structure in a group of scientists: a test of the "invisible college" hypothesis. *American Sociological Review 34*: 335-52.

Cross, R., A. Parker, and L. Prusak. 2001. Knowing what we know: supporting knowledge creation and sharing in social networks. *Organizational Dynamics 30* (2): 100-120.

Ehrlich, K and I. Carboni. 2005. *Inside Social Network Analysis*. IBM Technical Report 05-10.

Ejermo, O. and C. Karlsson. 2006. Interregional inventor networks as studies by patent coinventorships. *Research Policy 35*: 412-30.

Ettlinger, N. 1990. Worker displacement and corporate restructuring: a policy-conscious appraisal. *Economic Geography 66* (1): 67-82.

Faust, K. 1988. Comparison of methods for positional analysis: structural and general equivalences. *Social Networks 10*: 313-41.

Feldman, M. P. 1999. The new economics of innovation, spillovers and agglomeration: a review of empirical studies. *Economics of Innovation and New Technology 8* (1):5-25.

Feldman, M. P. 2001. Where science comes to life: university bioscience, commercial spin-offs, and regional economic development. *Journal of Comparative Policy Analysis: Research and Practice 2*: 345-61.

Felix, B. 2006. Patent procedures and statistics: an overview. Statistics in Focus, European Commission, Eurostat, Brussels. Retrieved from http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-NS-06-019/EN/KS-NS-06-019-EN.PDF

Fleming, L. and K. Frenken. 2007. The evolution of inventor networks in the Silicon Valley and Boston regions. *Advances in Complex System 10* (1): 53-71.

Fleming, L., C. King, and A. Juda. 2007. Small worlds and regional innovation. *Organization Science 18* (6): 938-54.

Florida, R. 1995. Calibrating the region. In *Local and Regional Systems of Innovation*, eds. J. de la Mothe and G. Paquet. New York: Kluwer Academic.

Fontes, M. 2005. Distant networking: the knowledge acquisition strategies of 'out-cluster' biotechnology firms. *European Planning Studies 13* (6): 899-920.

Freeman, L. C. 1979. Centrality in social networks: I. conceptual clarification. *Social Networks 1*: 215-39.

Frietsch, R. and T. Jung. 2009. Transnational patents—Structures, trends and recent development. Studien zum deutschen Innovationssystem Nr. 7-2009. Berlin: EFI.

Gallaud, D. and A. Torre. 2004. Geographical proximity and the diffusion of knowledge. In *Rethinking Regional Innovation*, eds. by G. Fuchs, P. Shapira, and A. Koch. New York: Springer.

Gertler, M. S. 2001. Tacit knowledge and the economic geography of context, or the indefinable tacitness of being (there). DRUID Summer Conference, Aalborg, Denmark, June.

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.

Gertler, M. S. 2007. Tacit knowledge in production system: how important is geography? In *The Economic Geography of Innovation*, ed. K. R. Polenske, 87-111. New York: Cambridge University Press.

Gertler, M. S. and Y. M. Levitte. 2005. Local nodes in global networks: the geography of knowledge flows in biotechnology innovation. *Industry and Innovation 12* (4): 487-507.

Gittelman, M. 2007. Does geography matter for science-based firms? Epistemic communities and the geography of research and patenting in biotechnology. *Organization Science 18* (4): 724-41.

Glaeser, E. L., H. D. Hallal, J. S. Scheinkman, and A. Shleifer. 1992. Growth in cities. *The Journal of Political Economy 100*: 1126-52.

Glaeser, E. L., J. A. Scheinkman, and A. Shleifer. 1995. Economic growth in a cross-section of cities. *Journal of Monetary Economics 36*: 117-43.

Glückler, J. 2007. Economic geography and the evolution of networks. *Journal of Economic Geography 7* (5): 619-34.

Granovetter, M. 1974. *Getting a Job. A Study of Contacts and Careers*. Chicago: The University of Chicago Press (1995 second edition).

Griliches, Z. 1990. Patent statistics as economic indicators: a survey. *Journal of Economic Literature 28*: 1661-707.

Grossman, G. and E. Helpman. 1991. *Innovation and Growth in a Global Economy*. Cambridge, MA: MIT Press.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg. 2001. The NBER patent citation data file: Lessons insights and methodological tools. National Bureau Economic Research. Working Paper 8498. Retrieved from http://www.nber.org/papers/w8498

Hanneman R. and M. Riddle. 2005. *Introduction to Social Network Methods*. University of California, Riverside, CA.

Haythornthwaite, C. 2005. Social network methods and measures for examining E-learning. http://www.wun.ac.uk/elearning/seminars/seminars/seminar_two/papers/haythornthwaite.pdf

Henderson, V., A. Kuncoro, and M. Turner. 1995. Industrial development in cities. *The Journal of Political Science 103*: 1067-90.

Henderson, R. M. and I. M. Cockburn. 1996. Scale, scope and spillovers: the determinants of research productivity in drug discovery. *RAND Journal of Economics 27* (1): 32-59.

Hevesi, A. G. and K. B. Bleiwas. 2005. The economic impact of the biotechnology and pharmaceutical industrial in New York. Office of the State Comptroller, report 11-2005. http://www.nyba.org/documents/biotechreport.pdf

Howells, J. R. L. 2002. Tacit knowledge, innovation and economic geography. *Urban Studies 39* (5-6): 871-84.

Isard, W. 1956. *Location and Space-Economy*. Cambridge, MA: Technology Press of MIT and London, UK: Chapman & Hall.

Jacobs, J. 1969. *The Economy of Cities*. New York: Vintage Books.

Jaffe, A. B. 1989. Real effects of academic research. *American Economic Review 79* (5): 957-79.

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

Kang, S. M. 2007. Equicentrality and network centralization: a micro-macro linkage. *Social Network 29*: 585-601.

Knoben, J. 2009. Localized inter-organizational linkages, agglomeration effects, and the innovative performance of firms. *The Annals of Regional Science 43*: 757-8-79.

Knoke, D. and S. Yang. 2008. *Social Network Analysis*. London, UK: Sage Publications.

Kroll, H. and N. Mallig. 2009. Regional patterns of co-patenting by technological fields, a Europe-US comparison. Presented at Science and Innovation Policy Atlanta Conference, October 2-3, 2009, Atlanta, GA.

Lissoni, F. and F. Montobbio. 2006. Co-inventorship and co-authorship in academic quantitative analysis of patent-publication pairs. Presented at 2nd Annual Workshop on the Law and Economics of Intellectual Property and Information Technology, July 27-28, 2006, Manchester, UK.

Lobo, J. and D. Strumsky. 2008. Metropolitan patenting, inventor agglomeration and social networks: a tale of two effects. *Journal of Urban Economics 63* (3): 871-84.

Lucas, R. E. 1988. On the mechanics of economic development. *Journal Monetary Economics 22*: 3-42.

Luger, M. I. and H. A. Goldstein. 1991. *Technology in the Garden: Research: Research Parks and Regional Economic Development*. Chapel Hill, NC: University of North Carolina Press.

Maggioni, M. A., M. Nosvelli, and T. E. Uberti. 2007. Space versus networks in the geography of innovation: a European analysis. *Papers in Regional Science 86* (3): 471-93.

Mahutga, M. C. and D. A. Smith. 2011. Globalization, the structure of the world economy and economic development. *Social Science Research 40*: 257-72.

Malmberg, A. 2003. Beyond the cluster: local milieus and global connections. In *Remaking the Global Economy*, eds. J. Peck and H. W. –C. Yeung. London: Sage.

Malmberg, A. and P. Maskell. 2002. The elusive concept of localization economies: towards a knowledge-based theory of spatial clustering. *Environment and Planning A 34*: 429-49.

Maraut, S., H. Dernis, C. Webb, V. Spiezia, and D. Guellec. 2008. The OECD Regpat Database: A Presentation. STI Working Paper. Retrieved from http://www.oecd.org/dataoecd/22/19/40794372.pdf

Markusen, A. 1996. Sticky places in slippery space: a typology of industrial districts. *Economic Geography 72*: 293-313.

Marshall, A. 1920. *Principles of Economics*. London: Macmillan.

Martin, R. and P. Sunley. 1996. The new ‘geography turn’ in economics: some critical reflections. *Cambridge Journal of Economics 23*: 65-91.

Maskell, P., H. Bathelt, and A. Malmberg. 2006. Building global knowledge pipelines: the role of temporary clusters. *European Planning Studies 14* (8): 997-1017.

Maskell, P. and A. Malmberg. 1999. Localised learning and industrial competitiveness. *Cambridge Journal of Economics 23* (2): 167-85.

Massey, D., P. Quintas, and D. Wield. 1992. *High-Tech Fantasies: Science Parks in Society, Science and Space*. New York, NY: Routledge.

McCann, P. and J. Simonen. 2005. Innovation, knowledge spillovers and local labor markets. *Papers in Regional Science 84* (3): 377-405.

McKelvey, M. 2004. Evolutionary economics perspective on the regional-national-international dimensions of biotechnology innovations. *Environmental and Planning C: Government and Policy 22*: 179-97.

McKelvey, M., H. Alm, and M. Riccaboni. 2003. Does co-location matter for formal knowledge collaboration in the Swedish biotechnology-pharmaceutical sector? *Research Policy 32*: 483-501.

Messner, S. F. and L. Anselin. 2004. Spatial analyses of homicide with areal data. In *Spatial Integrated Social Science*, eds. M. F. Goodchid and D. Janelle, 127-44. New York: Oxford University Press.

Moodysson, J., L. Coenen, and B. Asheim. 2008. Two sides of the same coin? Local and global knowledge flows in medicon valley. CIRCLE Paper no. 2008/15, Lund University, Sweden. Retrieved from www.lu.se/upload/CIRCLE/workingpapers/200815_Moodysson_et_al.pdf

Moran, P.A.P. 1950. Notes on continuous stochastic phenomena. *Biometrika* 37: 17-33.

Müller-Prothmann, T. 2007. Social network analysis: a practical method to improve knowledge sharing. In *Hands-on Knowledge Co-creation and Sharing: Practical methods and Techniques*, eds. A. S. Kazi and L. Wohlfart, and P. Wolf. 219-33. Knowledge Board, Stuttgart.

Neal, Z. P. 2011. The causal relationship between employment and business networks in U.S. cities. *Journal of Urban Affairs 33* (2): 167-84.

Nelson, R. R. and S. G. Winter. 1982. *An Evolutionary Theory of Economic Change*. Cambridge, MA: Harvard University Press.

Niosi, J. and M. Banik. 2005. The evolution and performance of biotechnology regional systems in innovation. *Cambridge Journal of Economics 29*: 343-57.

Nooteboom, B., W. V. Haverbeke, G. Duysters, V. Gilsing, and Ad van den Oord. 2007. Optimal cognitive distance and absorptive capacity. *Research Policy 36*: 1016-34.

Nooy, W. d., A. Mrvar, and V. Batagelj. 2005. Exploratory Social Network Analysis with Pajek. New York, NY: Cambridge University Press.

Oinas, P. 1999. Activity-specificity in organizational learning: implication for analyzing the role of proximate. *GeoJournal 49*: 363-72.

Ó hUallacháin, B. 1999. Patent places: Size matters. *Journal of Regional Science 39*: 613-36.

Ó hUallacháin, B and D.-S. Lee. 2011. *Technological specialization and variety in urban invention*. *Regional Studies 45* (1): 67-88.

Owen-Smith, J. and W. W. Powell. 2004. Knowledge networks as channels and conduits: The effects of spillovers in the Boston biotechnology community. *Organization Science 15* (1): 5-21.

Polanyi, M. 1966. The logic of tacit inference. *Philosophy 41* (155): 1-18.

Ponds, R. 2007. Interregional knowledge networks and spatial differences in innovation. Presented at DRUID-DIME Academy Winter Conference, 25-27 January 2007, Aalborg, Denmark.

Ponds, R., F. von Oort, and K. Frenken. 2007. The geographical and institutional proximity of research collaboration. *Papers in Regional Science 86* (3): 423-43.

Porter, M. 1990. *The Comparative Advantage of Nations*. New York: Free Press.

Powell, W. 1996. Inter-organizational collaboration in the biotechnology industry. *Journal of Institutional and Theoretical Economics 152* (1): 197-215.

Price, D. J. D. S. and D. D. B. Beaver. 1966. Collaboration in an invisible college. *American Psychologist 21*: 1011-18.

Rallet, A. and A. Torre. 1999. Is geographical proximity necessary in the innovation networks in the era of global economy? *GeoJurnal 49*: 373-80.

Ratanawaraha, A. and K. R. Polenske. 2007. Measuring the geography of innovation: A literature review. In *The Economic Geography of Innovation*, ed. K. R. Polenske, 30-59. New York: Cambridge University Press.

Rees, J., G. J. D. Hewings, and H. A. Stafford. 1981. *Industrial Location and Regional Systems: Spatial Organization in the Economic Sector*. New York, NY: Bergin Publishers.

Rice, R. E. and C. Aydin. 1991. Attitudes toward new organizational technology: network proximity as a mechanism for social information processing. *Administrative Science Quarterly 36* (2): 219-44.

Roijakkers, N. and J. Hagedoorn. 2006. Inter-firm R&D partnering in pharmaceutical biotechnology since 1975: Trends, patterns, and networks. *Research Policy 33*: 1153-165.

Romer, P. 1986. Increasing returns and economic growth. *American Economic Review 94*: 1002-37.

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

Rondé, P. and C. Hussler. 2005. Innovation in regions: what does really matter? *Research Policy 34*: 1150-72.

Rubí-Barceló, A. 2010. Core/periphery scientific collaboration networks among very similar researchers. Documentos de Trabajo FUNCAS No. 547. Retrieved from http: //dialnet.unirioja.es/servlet/articulo?codigo=3250499

Sapsed, J., D. Gann, N. Marshall, and A. Salter. 2005. From here to eternity?: The practice of knowledge transfer in dispersed and co-located project organizations. *European Planning Studies 13* (6): 831-51.

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

Saxenian, A. and J.-Y. Hsu. 2001. The Silicon Valley—Hsinchu connection: technical communities and industrial upgrading. *Industrial and Corporate Change 10* (4): 893-920.

Scott, J. 2000. *Social Network Analysis: A Handbook*. London: Sage Publications.

Simmie, J. 2003. Innovation and urban regions as national and international nodes for the transfer and sharing of knowledge. *Regional Studies 37* (6-7): 607-20.

Singh, J. 2005. Collaborative networks as determinants of knowledge diffusion patterns. *Management Science 51* (5): 756-70.

Smith, D. M. 1981. Industrial location: An economic Geographical Analysis. New York, NY: Wiley.

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

Sunley, P. 2008. Relational economic geography: a partial understanding or a new paradigm? *Economic Geography 84* (1): 1-26.

The Bioscience Brief. 2007. June 6, 2011. Retrieved from http://nationalbbr.org/brief/bioscience_brief_vol3_no1.pdf

Thompson, P. and M. Fox-Kean. 2005. Patent citations and the geography of knowledge spillovers: A reassessment. *American Economic Review 95*(1): 450-60.

Torre, A. and A. Rallet. 2005. Proximity and localization. *Regional Studies 39* (1): 47-59.

Tripple, M, F. Tödtling, and L. Lengauer. 2009. Knowledge sourcing beyond buzz and pipelines: evidence from the Vienna software cluster. *Economic Geography 85* (4): 443-62.

U.S. Patent and Trademark Office. 2010. USPTO patent full-text and image database. Retrieved from http://patft.uspto.gov/netahtml/PTO/search-adv.htm

United Nations. 1993. *The Convention on Biotechnology Diversity*. September 12, 2011. Retrieved from http://www.cbd.int/convention/articles/?a=cbd-02

Vallance, P. 2007. Rethinking economic geographies of knowledge. *Geography Compass 1* (4): 797-813.

Varga, A. 2000. Local academic knowledge spillovers and the concentration of economic activity. *Journal of Regional Science 40*: 289-309.

Varga, A. and A. Parag. 2009. Academic knowledge transfers and the structure of international research networks. *IAREG Working Paper 1.3d*. Retrieved from http://www.iareg.org/fileadmin/iareg/media/papers/WP1_03d.pdf

Wagner, C. S. and L. Leydesdorff. 2005. Mapping the network of global science: comparing international co-authorships from 1990 to 2000. *International Journal of Technology and Globalisation 1*(2): 185-208.

Wasserman, A. G. 2009. *GW Medicine Notes 13* (12).The George Washington University Medical Faculty Associates. December 2009. Retrieved from http://gwmed.com/docs/File/Newsletter/December%202009%20Newsletter.pdf

Wasserman, S. and K. Faust. 1994. *Social Network Analysis: Methods and Application*. Cambridge, New York: Cambridge University Press.

Wenger, E., R. McDermott, and W. M. Snyder. 2002. Cultivating communities of practices. Working paper. Boston, MA: Harvard Business School Press.

White, D. R. and K. P. Reitz. 1983. Graph and semigroup homomorphisms on networks of relations. *Social Networks 5*: 193-234.

Wilhelmsson, M. 2009. The spatial distribution of inventor networks. *The Annals of Regional Sciences 43*: 645-68.

Yeung, H. W.-C. 2005. Rethinking relational economic geography. *Transactions of the Institute of British Geographers NS 30*: 37-51.

APPENDIX A

THE LIST OF 150 LARGE U.S. METROPOLITAN STATISTICAL AREAS

2 Albany, GA
3 Albany--Schenectady--Troy, NY
4 Albuquerque, NM
6 Allentown--Bethlehem--Easton, PA
8 Amarillo, TX
10 Appleton--Oshkosh--Neenah, WI
12 Athens, GA
13 Atlanta, GA
14 Auburn--Opelika, AL
15 Augusta--Aiken, GA--SC
16 Austin--San Marcos, TX
17 Bakersfield, CA
20 Baton Rouge, LA
24 Billings, MT
25 Biloxi--Gulfport--Pascagoula, MS
27 Birmingham, AL
29 Bloomington, IN
30 Bloomington--Normal, IL
31 Boise City, ID
32 Boston, MA
33 Brownsville-San Benito, TX
34 Bryan--College Station, TX
35 Buffalo--Niagara Falls, NY
36 Burlington, VT
39 Cedar Rapids, IA
40 Champaign--Urbana, IL
42 Charleston--North Charleston, SC
44 Charlottesville, VA
46 Cheyenne, WY
47 Chicago--Gary--Kenosha, IL--IN--WI
49 Cincinnati--Hamilton, OH--KY--IN
51 Cleveland--Akron, OH
52 Colorado Springs, CO
53 Columbia, MO
56 Columbus, OH
58 Corvallis, OR
60 Dallas--Fort Worth, TX
64 Dayton--Springfield, OH
66 Decatur, IL
67 Denver--Boulder--Greeley, CO
68 Des Moines, IA
69 Detroit--Ann Arbor--Flint, MI
74 Eau Claire, WI
77 Elmira, NY
79 Erie, PA
80 Eugene--Springfield, OR
81 Evansville--Henderson, IN--KY
84 Fayetteville--Springdale, AR
85 Flagstaff, AZ--UT
88 Fort Collins--Loveland, CO
90 Fort Pierce--Port St. Lucie, FL
94 Fresno, CA
96 Gainesville, FL
99 Grand Forks, ND--MN
101 Grand Rapids-Holland, MI
103 Green Bay, WI
104 Greensboro-Winston, NC
105 Greenville, NC
106 Greenville--Anderson, SC
107 Harrisburg--Lebanon, PA
108 Hartford, CT
111 Houma, LA
112 Houston--Galveston, TX
114 Huntsville, AL
115 Indianapolis, IN
116 Iowa City, IA
118 Jackson, MS
120 Jacksonville, FL
123 Janesville--Beloit, WI
124 Johnson City-Bristol, TN-VA
128 Kalamazoo--Battle Creek, MI
129 Kansas City, MO--KS
131 Knoxville, TN
132 Kokomo, IN
134 Lafayette, IN
138 Lancaster, PA
139 Lansing--East Lansing, MI
142 Las Vegas, NV--AZ
143 Lawrence, KS
146 Lexington, KY
148 Lincoln, NE
149 Little Rock--North Little Rock, AR
151 Los Angeles, CA
152 Louisville, KY--IN
153 Lubbock, TX
154 Lynchburg, VA
156 Madison, WI
158 McAllen--Edinburg--Mission, TX
161 Memphis, TN--AR--MS
162 Merced, CA
163 Miami--Fort Lauderdale, FL
164 Milwaukee--Racine, WI

165 Minneapolis--St. Paul, MN--WI
167 Mobile, AL
174 Nashville, TN
175 New Haven---Bridgeport--Stamford, CT
176 New London--Norwich, CT-RI
177 New Orleans, LA
178 New York--Northern New Jersey, NY-NJ-CT
179 Norfolk--Virginia Beach--Newport, VA-NC
182 Oklahoma City, OK
183 Omaha, NE--IA
184 Orlando, FL
188 Pensacola, FL
189 Peoria--Pekin, IL
190 Philadelphia--Wilmington—Atlanta City, PA
191 Phoenix--Mesa, AZ
193 Pittsburgh, PA
196 Portland, ME
197 Portland--Salem, OR--WA
198 Providence-Warwick, RI-MA
199 Provo--Orem, UT
202 Raleigh--Durham--Chapel Hill, NC
204 Reading, PA
206 Reno, NV
207 Richland--Kennewick--Pasco, WA
208 Richmond--Petersburg, VA
210 Rochester, MN
211 Rochester, NY
212 Rockford, IL
214 Sacramento--Yolo, CA
216 Salinas, CA
217 Salt Lake City--Ogden, UT
219 San Antonio, TX
220 San Diego, CA
221 San Francisco--Oakland--San Jose, CA
222 San Luis Obispo-Paso Robles, CA
223 Santa Barbara, CA
224 Santa Fe, NM
225 Sarasota--Bradenton, FL
227 Scranton--Wilkes-Barre--Hazleton, PA
228 Seattle--Tacoma--Bremerton, WA
232 Shreveport--Bossier City, LA
234 Sioux Falls, SD
236 Spokane, WA
237 Springfield, IL
238 Springfield, MA
241 St. Joseph, MO
242 St. Louis, MO--IL
243 State College, PA
247 Syracuse, NY
248 Tallahassee, FL
249 Tampa--St. Petersburg, FL
252 Toledo, OH
254 Tucson, AZ
258 Utica--Rome, NY
260 Visalia--Tulare--Porterville, CA
262 Washington D.C.--Baltimore, DC-MD-VA-WV
265 West Palm Beach--Boca Raton, FL
272 York, PA