Papers in Evolutionary Economic Geography

17.15

Co-inventor Networks and Knowledge Production in Specialized and Diversified Cities

Frank van der Wouden, David L. Rigby

Utrecht University
Urban & Regional research centre Utrecht

Co-inventor Networks and Knowledge Production in Specialized and Diversified Cities

Frank van der Wouden* **and David L. Rigby**
Department of Geography
UCLA

Department of Geography, University of California – Los Angeles, 1255 Bunche Hall, Los Angeles, Box 951524, California 90095, The United States of America
Email: fvanderwouden@ucla.edu

Abstract

Why do some cities produce more knowledge than others? The standard explanation rests upon the social networks that connect economic actors, within and between cities, and that structure the quantity and the quality of interactions from which new ideas are generated. These interactions are increasingly understood as shaped by different forms of proximity that congeal, at different times in different places, in complex assemblies that give rise to different forms of competitive advantage. Recent research focusing on the U.S. urban system has shown that metropolitan regions characterized by more extensive local and non-local network ties outperform cities where economic agents are isolated. However, across most of this work, little attention is given to the character of the local knowledge base and whether that is related to the structure of co-inventor networks. In this paper, we show that the social networks linking co-inventors differ between cities that produce specialized knowledge and those that produce diversified knowledge. These ideas are extended in models of tie-formation that show inventors in specialized cities value spatial proximity less and cognitive proximity more than inventors in diversified cities as they partner with collaborators from other urban areas.

Keywords: knowledge production, patent, co-inventor, network, specialization, collaboration

JEL: O33, R11, R12, R15, D83, D85

1. Introduction

Knowledge production is increasingly imagined as an interactive task through which economic agents recombine existing ideas in novel ways (Arthur, 1999; Kauffman, 1993; Singh & Fleming, 2010; Wuchty, Jones, & Uzzi, 2007). Thus, the pace of invention relies upon access to multiple subsets of knowledge along with the capacity to translate those knowledge stocks into new technologies (Cohen & Levinthal, 1990). For economic geographers, these constraints on invention have historically focused attention on industrial districts or clusters within which actors are assumed to generate economies from the reduced cost of interaction and from spillovers that are bounded by co-location (Jacobs, 1969; Marshall, 1920). Jaffe et al. (1993) and Audretsch and Feldman (1996) provide supporting empirical evidence. For economic sociologists, competitive advantage is "located" in the structure of social connections that economic agents cultivate (Powell, 1990; Uzzi, 1996). In this sense, social proximity is seen as independent of spatial proximity and perhaps more important in regulating the fortunes of firms and the flows of knowledge between them (see also Agrawal et al., 2006; Breschi & Lissoni, 2009).

Investigation of the geography of knowledge production problematizes the relationship between spatial, social and other forms of proximity, illustrating the conditions under which proximity is advantageous, but also when it becomes a liability (Boschma, 2005; Grabher, 1993). A good deal of this work contests the separation of spatial and social relations, seeking to understand how co-location affects the structure of social ties (Broekel & Boschma, 2012; Gertler, 2003; Healy & Morgan, 2012). At the same time, the primacy of the local in the formation of social networks is questioned by Bathelt et al. (2004) and Amin and Cohendet (2005) who suggest that spatial embeddedness is less and less important to the inter-organizational linkages that enhance firm and regional performance.

Apart from a few notable exceptions, there is relatively little research that examines how the characteristics of the social networks and the spatial clusters within which economic agents are embedded both influence behavior and economic outcomes. In recent work on geographical variations in knowledge production, Fleming et al. (2007) and Lobo and Strumsky (2008) explore how the social networks that link inventors influence the pace of invention independent of place-specific characteristics including agglomeration. Breschi and Lenzi (2016) update these papers. Whittington et al. (2009) push this analysis a little further and begin to examine the interaction between social and spatial relationships that influence innovation in knowledge-intensive industries.

In this paper we extend these ideas, exploring how co-inventor networks influence knowledge production in U.S. cities after controlling for a number of location-based covariates. Building up from the case-studies of Whittington et al. (2009) we provide broad evidence that social networks and localized processes of agglomeration are positively related to knowledge production within cities. At the scale of individual metropolitan areas we find that the advantages of spatial proximity and social proximity are substitutes for one another. In addition, we add value to existing research by showing that characteristics of the networks linking co-inventors within and between U.S. metropolitan areas themselves depend on the nature of knowledge produced in different places (see also Cantner et al., 2010). In particular, we show that specialized cities, characterized by high levels of cognitive proximity across local knowledge stocks, develop co-inventor networks with significantly greater indices of centrality than those found in diversified cities and that such centrality enhances knowledge production. Finally, we show that the pipelines connecting inventors between cities also vary with the level of metropolitan technological specialization: on average inventors in specialized cities are less impacted by geographical distance and more impacted by cognitive distance in their search for knowledge production partners in other urban areas.

The paper is organized in four following sections. Section 2 provides a brief review of the literature that motivates our research. In Section 3 we explore the operationalization of the core theoretical concepts and we discuss the sources of the data employed in our empirical analysis. The results from that analysis occupy Section 4 of the paper and we offer a number of concluding remarks in Section 5.

2. Literature Review

Across the market economy, the heterogeneity of firm characteristics suggests a multiplicity of competitive strategies. Since the pioneering work of Cyert and March (1963) this heterogeneity is thought to express the firm-specific assets that undergird resource-based visions of firm performance developed by Wernerfelt (1984) and Barney (1991). Kogut and Zander (1992) were among the first to emphasize the critical role of knowledge within this framework. What is clear from related empirical work is that firms search for efficiency and for knowledge in many different ways (Baily et al., 1992; Baldwin & Gorecki, 1998; Saxenian, 1994). Within economic geography, a standard distinction is made between those firms that seek competitive advantage internally and those that search for efficiency through co-location with others.

For the firms that agglomerate in space, the collective resources that sustain the industrial district have long been envisioned, after Marshall (1920), as lower-cost access to specialized suppliers and buyers, to the associated pools of labor that clusters exploit and nurture, through to spillovers of knowledge. A somewhat different vision is offered by Jacobs (1969) who does not contest the efficiency of Marshall's (1920) districts, but who rather imagines the long-run prospects of firms to rest more squarely on the diversity that cities provide. A modern update is advanced by Duranton & Puga (2001). Glaeser et al. (1992) present empirical evidence of more rapid industrial growth within diversified local economies, while Henderson (2003) and Baldwin et al. (2010) provide firm-level evidence of higher levels of productivity in specialized urban economies. More recent work still suggests that even within industrial districts the characteristics and behaviors of firms remain highly variable and that not all firms generate efficiencies in the same way (Neffke et al., 2011; Potter & Watts, 2011; Rigby & Brown, 2015).

For economic sociologists, these differences might be explained by the structure of the social networks that link firms and other economic agents (Ronald S Burt, 2000; W.W. Powell, 1990). Social networks are broadly seen as an organizational form that enhance the sharing of knowledge and other resources in technologically complex industries where novel ideas are widely distributed and the rapidity of innovation generates considerable uncertainty (Hagedoorn, 1993; Powell et al., 1996). Though networks are broadly seen as raising efficiency, precisely how firms are embedded within networks is critical to their performance (Granovetter, 1973; Uzzi, 1997). There is increasing evidence that networks with weak ties promote exploration and technological discovery, while networks with strong ties promote exploitation (Burt, 1992; Rowley et al., 2000; Walker et al., 1997).

There is growing interest in the relationships between social networks and spatial clusters of economic agents Whittington et al. (2009). In part, this is motivated by the interaction of different forms of proximity (Boschma, 2005). It is also driven by empirical work that illustrates how the structure of social networks vary over space (Cantner & Graf, 2006; Sorenson, 2005) and how this structure is impacted by distance (Bathelt et al., 2004; Knoben & Oerlemans, 2012; Malmberg & Maskell, 2006). Linking the literature on social networks with studies of agglomeration, Fleming et al. (2007) explore how co-inventor networks impact invention across U.S. cities. They show that shorter path lengths and larger connected components are positively correlated with patent production, but that small-world networks fail to generate the expected productivity gains in local knowledge production. Lobo and Strumsky (2008) and Breschi and Lenzi (2016), in closely related work, report the positive influence of network size and connectedness on urban invention.

While this research illustrates how social networks influence the pace of knowledge production within cities, it does not consider whether the architecture of regional knowledge stocks might shape the structure of social networks. The stocks of knowledge that accumulate in particular places may be characterized by their age and size, by their diversity across scientific, technological or industrial fields (Kogler et al., 2016), by complexity (Balland & Rigby, 2017) and by the ease of their recombination (Fleming & Sorenson, 2001). It seems reasonable to assume that the structure of social networks might vary between cities and regions that have knowledge stocks with different characteristics. In particular, when economic agents in regions share relatively high levels of cognitive proximity, the potential for interaction is higher than where a region's knowledge base is fragmented across different scientific or industrial fields. Thus, we might expect, following Cantner et al. (2010), that more specialized regional knowledge bases may be associated with larger and denser co-inventor networks that, in turn, may shape the character of local knowledge production. The empirical work that follows takes up this issue in more detail.

3. Data & Methods

The aim of this paper is to explain variations in knowledge production across U.S. metropolitan areas between 1975 and 2005, and to explore the roles of the structure of knowledge and social networks in such explanation. We measure knowledge production using patent data derived from the United States Patent and Trademark Office (USPTO). Thus, our dependent variable is the annual number of patents produced within each U.S. metropolitan area. Many patents are generated by more than one inventor. When these teams of inventors are located in the same metropolitan area, the individual patent is fully assigned to that same location. In the case of patents produced by multiple inventors located across different metropolitan areas, individual patents are fractionally split across those areas with shares determined by the geographical distribution of co-inventors. Patents developed only by foreign inventors are excluded from our data. Fractional counts of patents imply that the dependent variable is not a "count variable". Our fractional counts focus on the application year of patents as is customary in the literature.

Two independent variables play a central role in our analysis of metropolitan knowledge production. The first of these is a measure of urban co-inventor networks and the second is a measure of the specialization of a city's knowledge core. In fact, we measure two social networks of co-inventors for each metropolitan area in each year - an intra-city network comprising the links between inventors located in the same city and an inter-city network comprising the links between inventors located in different cities. The specialization of the knowledge base of each U.S. metropolitan area is built up annually through analysis of the stocks of patents in each technology class and by analysis of the relatedness distance between those classes. We also gather data for a series of additional covariates that are defined below.

The USPTO lists the names of all inventors on patents. These inventors form the nodes of potential collaboration networks that vary year-by-year according to whether inventors have applied for a patent in a given time-period. When two or more inventors are listed on the same patent then a link is established between the inventor-nodes. The addresses (city and county) are listed for all inventors on patents. We use the inventor county to assign individual patents (either fully or fractionally) to the corresponding CBSA[1]. In turn, the largest 366 CBSAs form the metropolitan statistical areas (MSAs) upon which our analysis is focused. When co-inventors on

[1] We use the December 2009 classification of CBSAs by the U.S. Bureau of the Census.

a patent are located in the same metropolitan area then we have an intra-city network link. When inventors on a patent are located in different metropolitan areas then we have an inter-city network link. Patents with more than three co-inventors can simultaneously represent intra- and inter-city network linkages. Intra-city and inter-city networks are examined below.

The number of nodes in our networks is given by the number of distinct inventors. Unfortunately, the USPTO does not uniquely identify individual inventors. Thus, it is impossible to tell from USPTO records whether an inventor on patent *i*, in application year *t*, named John Smith is the same inventor as John Smith listed on patent *j* from the same application year *t*. To resolve such ambiguities, we utilize disambiguated inventor IDs made available by Li et al. (2014) and link these to the inventors on all patents. With the disambiguated inventor data we can define the size of active inventor networks for all cities in all years. Here we follow Fleming et al. (2007) and Lobo and Strumsky (2008).

Co-invention networks play a central role in the diffusion of ideas and knowledge amongst inventors (Singh, 2005). Several descriptive measures have been developed to characterize the structural aspects of regional co-inventor networks and used to explain the observed variance in regional knowledge production. Fleming et al. (2007) find that the number of local and non-local inventors has a positive effect on subsequent patenting activity. Examining the largest connected component (LC) in the network, measured as the share of inventors associated with the LC, they find that the size of the LC and the inverse path length between inventors is positively related to subsequent patenting. Lobo & Strumsky (2008) report that inventor density (inventors per square mile), network aggregation and the ratio of non-local inventors in metropolitan co-inventor networks have a positive and significant relationship with the rate of metropolitan patenting. Unlike Fleming et al. (2007), they find a significant negative relationship between the size of the LC and patenting. Strumsky and Thill (2013) examine the relationship between a series of metropolitan co-inventor network statistics and four metropolitan economic performance indicators (wage, income, jobs and GDP). Their results show that the relationships between these network statistics and metropolitan performance indicators are inconsistent, indicating the delicate nature of the relationship between network connectivity, knowledge production and regional economic performance. Breschi & Lenzi

(2016) explicitly attempt to measure the structure of internal and external co-inventor networks using the average inverse geodesic distance between any pair of inventors linked to an urban area. They find no significant relationship between greater internal or external social proximity and the rate of patent production. However, they find a positive and significant effect of the interaction between internal social proximity and clique density on the rate of patenting. Moreover, they report a positive and significant relationship between the interaction of internal and external social proximity, and patent production. These findings suggest that the effect of external social proximity on the rate of metropolitan patenting is complementary to the effect of internal social proximity.

Comparing different metropolitan co-inventor networks is difficult for at least two reasons. First, these networks tend to be disconnected. This means that within each urban area not every inventor is connected (directly or indirectly) to other inventors through a co-inventor patent linkage. A number of network-level statistics, especially centrality measures, behave poorly for disconnected networks rendering them of questionable utility (De Nooy et al., 2011). Second, a number of measures of network characteristics do not scale well. Thus, it is difficult to determine whether the observed value of the network-level measure is a direct result of the structural network characteristics the researcher is trying to capture or whether it is an indirect effect of network size and density (Anderson et al., 1999). As a consequence, scholars interested in regional co-inventor networks have often limited their analyses to the largest component of networks, or used elementary descriptive statistics to characterize those networks. In both cases, the effects of co-inventor network structures on regional knowledge production may be biased.

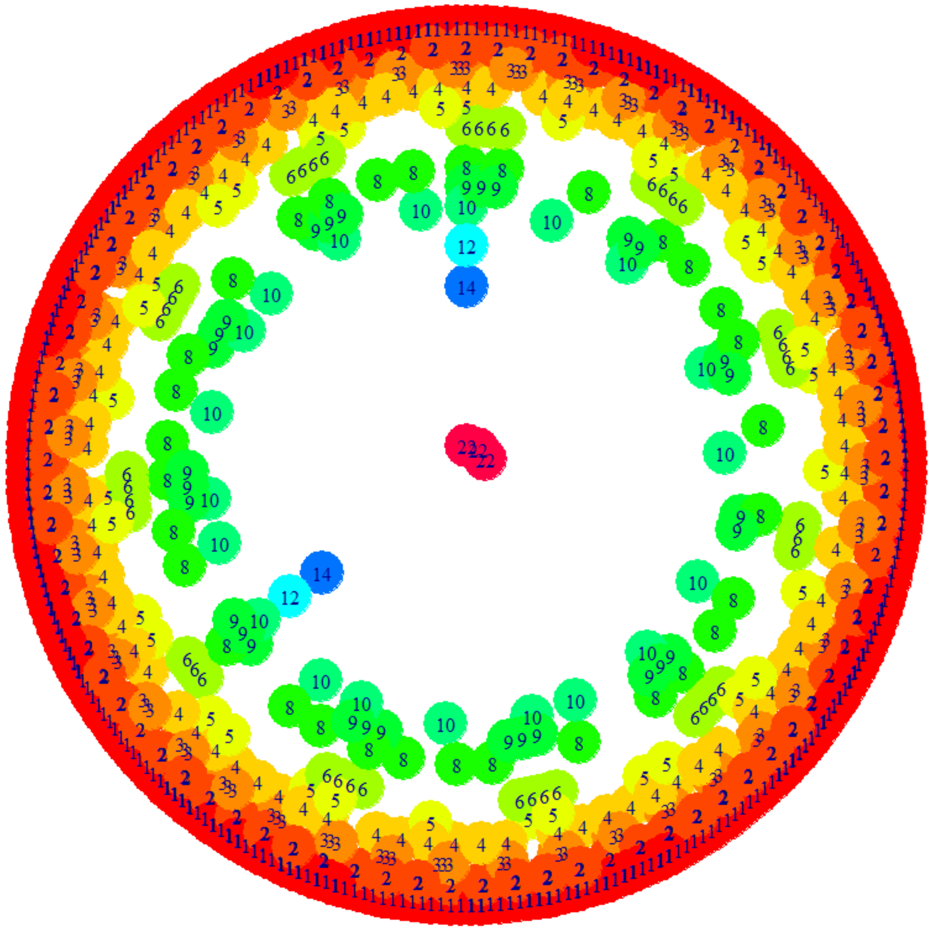
Fortunately, the k-core network measure developed by Seidman (1983) allows comparison of networks of different size and density and it is also applicable to disconnected networks (Butts et al., 2012). The k-core measure is a nodal degree based approach to identify cohesive (linked) subgroups across a network. A k-core is a subgraph in which each node is connected to a minimum *k* other nodes in the subgraph (Seidman, 1983). Thus, k-core subgraphs contain nodes that have a specified number of ties to other nodes in the subgraph. Formally, a subgraph is a k-core when $d_s(i) \geq k$ for all $n_i \in N_s$, where $d_s(i)$ denotes the number of connections (*degree*) of every node n_i in the subset of vertices N_s, and k represents the order of the core.

Matula and Beck (1983) offer an algorithm to degenerate a full network into different k-cores. We use the number of k-cores to characterize the structure of inter-city and intra-city co-inventor networks. In general, networks with a larger number of k-cores have greater variability amongst the number of connections of nodes than in networks with a smaller number of k-cores. And, of course, networks with larger k-cores tend to have a higher density of linkages between individual nodes. Thus, we hypothesize that metropolitan areas with larger numbers of k-cores across inter-city and intra-city co-inventor networks will generate more patents.

Note that we follow a different approach than Strumsky and Thill (2013), who also employ k-core measures to characterize co-inventor network structure. They focus on the percentage of inventors within a metropolitan area that are part of the largest k-core. We argue that this is not so much a measure of the structure of the co-inventor network, as it is a measure of repeated collaboration amongst a (small) subset of co-inventors. For instance, Figure 1 shows the 2005 network of co-inventors that are internal to Chicago. The highest k-core number (22) corresponds to three co-inventors collaborating together on 22 patents. We measure the structure of the co-inventor network in Chicago by the number of different k-cores found in the city. In Chicago, this number is 12.

The architecture of the knowledge base of cities might influence regional knowledge production. In particular, we are interested in whether metropolitan areas with more specialized or more diverse knowledge stocks generate more patents. This question has a long history within economic geography, stimulated by the arguments of Marshall (1920) and Jacobs (1969), as outlined above. Across much of the literature, the standard measure of specialization (or diversity) is the Herfindahl index. While this index is widely used, it has one major failing, namely its inability to control for varying "distances" between the economic categories across which specialization is measured. Here, we calculate the specialization (or diversity) of the knowledge base of cities by examining the distribution of patents across the 438 primary technological classes of the USPTO. For each pair of these classes we measure the technological distance or the cognitive proximity between them using patent co-classification data. We then compute the average relatedness or the average cognitive proximity between all pairs of patents that are generated within a city. This measure of average relatedness is bounded by the interval

0 – 1. Higher values of average relatedness indicate greater specialization.



[COLOR] Figure 1: The Number of k-cores in the Internal Co-Inventor Network of Chicago, 2005.

The details of these calculations are outlined below. Co-class information on individual patents is employed to measure the technological proximity of technology classes, following the earlier work of Jaffe (1986), Engelsman and van Raan (1994), Nesta and Saviotti (2005) and Kogler et al. (2013). To measure the proximity, or knowledge relatedness, between patent technology classes in a single year we employ the following method. Let P indicate the total number of patent applications in the chosen year. Then, let $F_{ip} = 1$ if patent record p lists the classification code i, otherwise $F_{ip} = 0$. Note that i represents one of the 438 primary technology classes into which the new knowledge contained in patents is classified. In a given year, the total number of patents that list technology class i is given by $N_i = \sum_p F_{ip}$. In similar fashion, the number of individual patents that list the pair of co-classes i and j is identified by the count $N_{ij} = \sum_p F_{ip} F_{jp}$. Repeating this co-class count for all pairs of 438 patent classes yields the (438x438) symmetric technology class co-occurrence matrix ***C*** the elements of which are the co-class counts N_{ij}. The co-class counts measure the technological proximity of all patent class pairs, but they are also influenced by the number of patents found within each individual patent class N_i. Thus, we standardize the elements of the co-occurrence matrix by the square root of the product of the number of patents in the row and column classes of each element, or

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

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

The average relatedness value for a metropolitan area *m* in year *t* is calculated as:

$$AR^{m,t} = \frac{\sum_i \sum_j S_{ij}^t * D_{ij}^{m,t} + \sum_i S_{ii}^t * 2D_{ii}^{m,t}}{N^{m,t} * (N^{m,t} - 1)} \qquad for\ i \neq j$$

where S_{ij}^t represents the technological relatedness between patents in technology classes *i* and *j*, $N^{m,t}$ is a count of the total number of patents in region *m* in year *t*, and where $D_{ij}^{m,t}$ counts the number of pairs of patents that can be located in technology classes *i* and *j* in region *m* in year *t*. To clarify the meaning of $D_{ij}^{m,t}$, imagine a region with three patents, one in technology class 1 and two in technology class 2. Then, the pair counts $D_{ij}^{m,t}$ represent elements in the (438x438) symmetric matrix

$$\boldsymbol{D}^{t,r} = \begin{bmatrix} 0 & 2 & \dots & 0 \\ 2 & 1 & & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}$$

with three patents, there are 3x2 = 6 unique distance measures to calculate, the distance between the patent in class 1 and each of the patents in class 2, the distances from both patents in class 2 to the patent in class 1 and the distance between the two patents in class 2. Note that the latter distance is counted twice. These routines are repeated for each of the 31 years in our analysis across all 366 metropolitan areas.

Cities and regions that build knowledge stocks around particular industries and technologies will likely record different numbers of patents over time as some sectors of the economy heat up and others cool down. Patents generated in very dynamic technology classes likely build incrementally on recent patents in the same sector. One way of controlling for the distribution of urban knowledge stocks across more or less dynamic classes is to capture the average age of citations on the patents generated each year. Cities active in newer technologies will likely have citations that are more recent than cities where invention is in older technologies. As patents are indexed by USPTO numbers that track the timing of their introduction to the economy, we calculate the mean age of citations on patents by averaging the USPTO numbers of the patents that they cite. When this average number is higher it references recent patents or newer technologies. We anticipate that metropolitan areas that are over represented in newer technologies will thus cite patents that have higher USPTO numbers on average. Including this mean age of citations should control for the degree to which urban areas are active in more dynamic technological sectors. Other authors in this field have used similar approaches (Fleming et al. 2007; Strumsky and Thill 2013; Breschi and Lenzi 2016).

Cities that devote a lot of effort in producing inventions are more likely to produce more patents than cities that don't make such investments. Typically R&D spending and venture capital funding are obvious indicators of such efforts. Unfortunately, there are no R&D or venture capital data available at the city level for our time frame (see Sorenson & Stuart, 2001). Instead, we construct a proxy based on the metropolitan distribution of grants allocated by the

National Science Foundation. These data are available for individual years[2]. We calculate the ratio of NSF funding per worker for each city, and hypothesize that higher levels of R&D spending, as captured by NSF grants, should be associated with higher levels of patent production. We focus on NSF spending per worker to try and capture an R&D effect that is independent of the size of cities that soaks up a good deal of the variance in our dependent variable. Earlier work on regional knowledge production across U.S. metropolitan areas has not controlled for R&D spending.

The level of inter-firm competition within a metropolitan area might affect inventive activity. There is significant disagreement as to whether larger firms with more monopolistic control over markets generate more or less new knowledge than would be found in more competitive markets comprising larger numbers of smaller firms. The differences between an early and late Schumpeter are well-known (Nooteboom, 1994). On the one hand, the monopoly argument holds that larger firms with greater market control are more likely to invent because they can more fully appropriate the economic benefits from their efforts. On the other hand, the competition argument suggests that firms' inventive activity benefits from knowledge externalities that rise with the number of firms (Rogers, 2004). We control for the level of economic competition within a metropolitan area by calculating the ratio of the number of firms to employment. Higher levels of this ratio signify greater competition. Counts of the number of firms and employment at the county level may be found in the County Business Patterns data generated by the U.S. Bureau of the Census. County figures are summed across the regional units that comprise each MSA. We have no explicit hypothesis on how competition impacts knowledge production, reflecting ambiguity in the existing literature (Acs & Audretsch, 1988).

Clearly larger MSAs are expected to generate more patents than smaller MSAs. We use employment within urban areas, obtained from the County Business Patterns (U.S. Census Bureau), to control for urban scale or size effects. We also use the density of inventors (inventors/land area) as a proxy to control for the level of agglomeration at the MSA level. We hypothesize that larger cities and cities with higher levels of inventor density will generate larger numbers of patents. In some of the regression models presented in Section 4 we make use of a

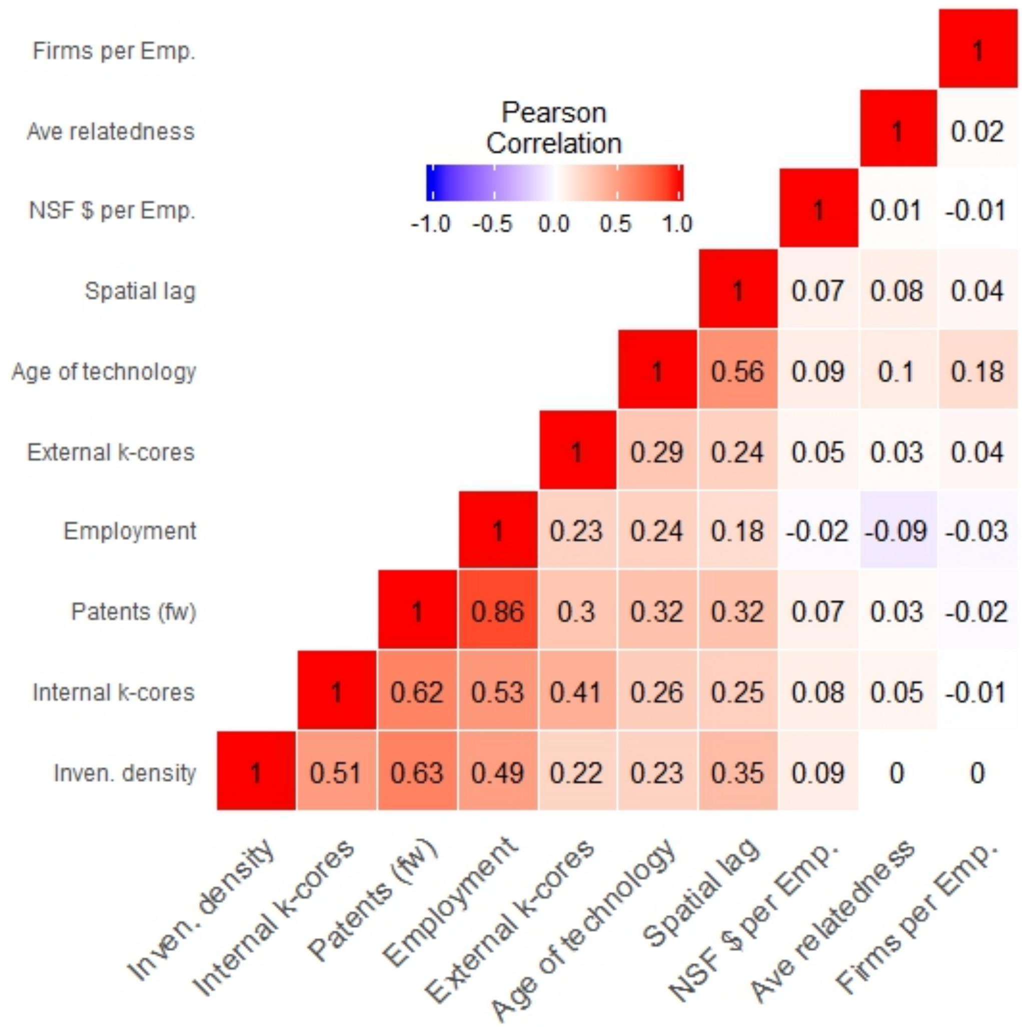
[2] Data available at https://www.nsf.gov/awardsearch/download.jsp

“spatial lag” variable that captures for every MSA the average number of patents generated by all other cities and where that average is weighted by the inverse distance to the focal city.

Descriptive statistics for all variables are shown in Table 1 for three time periods spanning most of the period under investigation. Variables that exhibit significant skew are augmented by the value 1 and then logged. Approximately 1.8 million patents were generated in the 366 U.S. metro areas over the period 1975 to 2005. The New York MSA produced most patents since 1975 accounting for 141,000 of the total. In second place, San Jose inventors produced approximately 107,000 patents over the study period. In third place, Los Angeles inventors generated approximately 96,000 patents. Chicago, San Francisco and Boston occupy the next ranks in terms of urban knowledge production since 1975. Hinesville-Fort Stewart, GA produced the fewest patents of any metropolitan area since 1975, registering a little over 9. Laredo, Texas and Jacksonville, NC occupy ranks 365 and 364 in the urban knowledge production hierarchy generating 37 and 50 patents respectively over the 31 years examined.

Note that the relatively large values for the average age of citations in Table 1 reflects the fact that we estimate the mean citation age of patents within an urban area by examining the USPTO numbers on all patents that are cited by inventors in a particular city and year. Utility patent issue numbers start at 3858241 in 1975. Thus, for 1980, the average age of citations (3732129) corresponds to an average date of issue of 1973 (an average age of 7 years). The average relatedness value (index of knowledge specialization) across U.S. metropolitan areas was 0.032 in 1980. This value increased to 0.036 in 1990 and 0.043 in 2000. Knowledge production is becoming more specialized at the urban level across the United States. This means that the average “technological distance” between all pairs of patents generated within a metropolitan area is declining over time.

Figure 2 illustrates the correlation coefficients between our variables. While the Pearson correlation coefficients are reasonably large in a few cases, the coefficients in our regression models with/without core variables are relatively stable. The reader is reminded that multicollinearity does not bias estimators it merely makes then inefficient. Inefficiency does not appear to be a problem in the results presented.



[COLOR] Figure 2: Correlation Coefficients Between Variables (all years)

Variables	1980				1990				2000			
	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
Patents (fw) ‡	3.17	1.53	0	8.24	3.53	1.53	0	8.3	3.98	1.66	0.85	9.04
Employment ‡	11.63	1.09	8.35	15.93	11.83	1.09	9.08	16.05	12.03	1.09	9.57	16.13
Inven. density ‡	0.02	0.04	0	0.45	0.03	0.05	0	0.49	0.05	0.1	0	0.89
Ave relatedness ‡	0.03	0.03	0	0.21	0.03	0.04	0	0.6	0.04	0.04	0	0.23
Age of technology	3732129	449067.1	0	4215443	4253089	340550.4	0	4685821	5257239	168177	4720271	5720329
NSF $ per Emp.	16.82	94.61	0	1342.35	24.83	116.87	0	1666.88	40.95	165.35	0	2393.8
Firms per Emp. ‡	0.04	0.01	0	0.06	0.04	0.01	0	0.08	0.04	0.04	0	0.75
Internal k-cores	0.85	1.34	0	9	1.27	1.71	0	9	2.34	3.46	0	22
External k-cores	1.34	1.71	0	14	2.02	2.12	0	11	4.17	5.19	0	33
Spatial lag ‡	8.19	0.49	6.29	9.78	8.5	0.44	6.64	10.02	9.14	0.37	7.41	10.42

‡ **Natural log of variable**

Table 1: Descriptive Statistics

4. Results

We anticipate that knowledge production in U.S. urban areas might be influenced by the inventive activity of neighboring cities. Indeed, statistical tests reveal that there is significant positive spatial autocorrelation in MSA patent output. It is important to control for this spatial autocorrelation in order to generate unbiased estimates of the influence of the independent variables on knowledge production in U.S. metropolitan areas. We introduce spatial autocorrelation into our models using the *spdep* and *splm* packages in R. Estimation makes use of fixed effects panel models covering 31 years and 366 metro regions. These models control for unobserved variables that are fixed at the MSA level. We control for time-specific shocks by adding time fixed-effects and as a crude "control" for concerns with endogeneity we lag all independent variables by one-period. We employ White's robust standard errors in case of heteroscedasticity.

Table 2 presents our first results, exploring whether cities that are specialized or diversified in terms of knowledge production produce more patents. All RHS variables in these models are lagged one period, save for the spatial lag term in the autocorrelation models. Model 1 in Table 2 is offered as a baseline, ignoring concerns with spatial autocorrelation and not including co-inventor networks. The independent variables included in Model 1 function largely as hypothesized. We control for the influence of MSA size with the employment variable. Not surprisingly, larger urban areas with higher levels of employment on average generate significantly more patents than smaller urban centers. Our simple measure of the strength of agglomeration within urban areas is inventor density. Increases in density raise the number of patents produced, as hypothesized. The age of technology is also significant and positively related to patent output. Thus, cities producing newer forms of knowledge, captured through the date of their citations generate more patents. In line with most models of knowledge production, as R&D spending per worker increases inventive output also increases. Most importantly, perhaps, the average relatedness variable is significant and has a positive sign suggesting that on average more specialized cities produce more patents than more diversified cities. Our measure of competition, the number of firms in an MSA per worker, is insignificant in model 1[3].

[3] Model 1 has slightly fewer observations than our other models because it is fitted using OLS and incomplete records are handled differently than in the other models that are fit with maximum likelihood techniques.

Adding spatial autocorrelation in model 2 revealed that both spatial lag and error terms in the autocorrelation model were significant. Lagrangian multiplier tests suggested the lag form of autocorrelation was stronger and so spatial lags were added to all models. A comparison of models 1 and 2 indicates that most independent variables have similar coefficients after the introduction of the spatial lag term. The only real exception is the measure of competition which remains negative but becomes significant at the 0.1 level after controlling for spatial autocorrelation. This result suggests that urban areas with large firms tend to generate more patents than cities with fewer large firms. Indeed, Klepper (1996) and Acs & Audretsch (1988) suggest that more competitive regional economies make it difficult for firms to appropriate the returns from patenting. Note that the pseudo R-squared term is much larger in models with the spatial lag form of autocorrelation added, as is often the case.

Models 3-5 introduce network measures to our analysis of urban knowledge production. Like Strumsky and Thill (2013), we capture the structure of internal and external city networks using k-core degeneracy, but use a different measure for the reasons indicated above. In line with existing studies (Fleming et al. (2007), Lobo and Strumsky (2008) and Breschi and Lenzi (2016)), Model 3 shows that the structure of co-inventor networks, those that are internal to the city and those that link collaborators within a city to inventors elsewhere ("external networks") have a positive and significant influence on patent production. Indeed, denser webs of collaboration amongst inventors (either internal or external) foster the production of patents. These network effects are independent of our measure of urban agglomeration that is captured by inventor density. Note that the internal network measure has a stronger influence on knowledge production than the external network measure. It seems reasonable to anticipate some interaction between the measures of agglomeration and co-inventor networks (see Whittington et al. 2009). This concern is the focus of models 4-6. Thus in model 4, we interact inventor density (our measure of urban agglomeration) with the number of internal co-inventor k-cores in the city to examine whether or not internal collaboration networks are a complement or a substitute for agglomeration. The negative coefficient on the internal interaction variable in Model 4 indicates a substitution effect and suggests that cities with large urban agglomerations gain less from local networks than cities where such agglomeration is rather poorly developed. Model 5 supports a similar story of substitution between the forces of agglomeration within cities and external collaboration networks. These results are somewhat surprising. We had suspected that agglomeration and networks would act as complements, combining to raise the overall volume of urban knowledge production, especially in the case of external knowledge networks.

Dependent variable: No. of patents (fw)	(1)	(2)	(3)	(4)	(5)	(6)
Spatial Autocorr.		0.406^{***}	0.395^{***}	0.478^{***}	0.395^{***}	0.391^{***}
		(0.041)	(0.040)	(0.039)	(0.041)	(0.040)
Employment ‡	1.390^{***}	1.232^{***}	1.209^{***}	1.157^{***}	1.200^{***}	1.206^{***}
	(0.032)	(0.030)	(0.030)	(0.029)	(0.030)	(0.030)
Inventor density ‡	2.921^{***}	3.117^{***}	2.243^{***}	4.313^{***}	3.815^{***}	2.242^{***}
	(0.116)	(0.104)	(0.116)	(0.152)	(0.138)	(0.116)
Ave. relatedness ‡	1.368^{***}	2.111^{***}	1.993^{***}	1.870^{***}	2.064^{***}	1.994^{***}
	(0.099)	(0.089)	(0.088)	(0.087)	(0.088)	(0.088)
Age of technology	0.000^{***}	0.000^{***}	0.000^{***}	0.000^{***}	0.000^{***}	0.00000^{***}
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
NSF $ per Emp.	0.0001^{***}	0.0001^{***}	0.0001^{***}	0.0001^{***}	0.0001^{***}	0.0001^{***}
	(0.00004)	(0.00004)	(0.00004)	(0.00003)	(0.00004)	(0.00004)
Firms per Emp. ‡	-0.284	-0.375^{*}	-0.329^{*}	-0.396^{**}	-0.358^{*}	-0.326
	(0.219)	(0.202)	(0.199)	(0.196)	(0.200)	(0.199)
Internal k-cores			0.044^{***}	0.078^{***}		0.051^{***}
			(0.003)	(0.003)		(0.004)
External k-cores			0.007^{***}		0.025^{***}	0.011^{***}
			(0.002)		(0.002)	(0.002)
Interaction internal				-0.191^{***}		
				(0.009)		
Interaction external					-0.121^{***}	
					(0.014)	
Interaction int. * ext.						-0.001^{***}
						(0.0003)
N	10.980	11.346	11.346	11.346	11.346	11.346
CBSA	366	366	366	366	366	336
R-Squared	0.22	0.96	0.96	0.96	0.96	0.96

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

‡ Natural log of variable

All independent variables (except the spatial autocorrelation term) are lagged

Table 2: Determinants of the Pace of Patenting in U.S. Metropolitan Areas

Model 6 explores the interaction between internal and external knowledge networks at the city-level. The negative coefficient for the interaction variable indicates that the number of internal and external k-cores act as substitutes. Again, we expected that external collaborations (pipelines) should feed the internal inventor pool (local buzz) with non-local knowledge thus boosting overall knowledge output. Perhaps it is the case that inventors can only collaborate with a finite number of partners. Hence, collaborating with co-inventors located in other metropolitan areas limits the opportunities to collaborate within the city and vice versa. Note that these findings run counter to the results of Breschi & Lenzi (2016) who report a positive interaction between internal and external network effects on urban invention. We note here that limiting our analysis to census years, for which we have educational data, and including the share of the MSA population with a bachelor's degree as a measure of human capital, produces broadly similar results to those reported in Table 1 and yields a positive coefficient on the human capital variable.

We now shift toward examination of the influence of networks on knowledge production in metropolitan areas that are characterized as either relatively specialized or relatively diverse in terms of the range of technologies generated. Our analysis of this question was prompted by exploration of the data and a suspicion that knowledge networks exhibit significant differences within specialized and diversified urban areas. We illustrate these differences in Figures 3 and 4. Overall, specialized cities tend to have much more well-developed internal and external co-inventor networks than diversified metropolitan areas and this finding holds for cities of different size. For example, Figure 3 clearly shows the differences in the internal collaboration network structure of the medium sized cities Boise and Pittsburgh in 2005. On average, inventors in Boise, a specialized city, are much more connected to other local inventors than inventors in a more diversified city such as Pittsburgh. In much larger cities we see the same pattern, with a much more well-developed internal network in San Jose, a specialized metropolitan area, than in Chicago which is technologically more diversified. Figure 4 illustrates these same differences in the structure of external co-invention networks in the smaller MSAs of Poughkeepsie and Cleveland and again in the larger MSAs of San Jose and Chicago.

Though these figures suggest differences in network structure between specialized and diversified urban areas, more careful examination is required to substantiate this claim. To engage this issue, we separated MSAs into two groups around the median value of average relatedness or technological specialization. This grouping was performed year-by-year and yielded a set of cities more technologically specialized than the median and a set of cities more technologically diversified than the median. We then rerun our models of urban knowledge production across the two sets of cities. The results are presented in Table 4. Note that we also explored sub-setting cities into the upper and lower quartiles of the distribution of average relatedness and found similar results to those we report below. We focus on the results either side of the median for that increases the number of observations in our two datasets.

First note that we are unable to control explicitly for spatial autocorrelation in the models of Table 4 because of the unbalanced nature of our panel data following its separation into specialized and diversified city-time components. As a crude proxy for spatial autocorrelation we add another variable to the models that represents the spatial lag term in the autocorrelation model. This variable measures the inverse distance weighted value of patents generated in all cities save for the focal MSA. Excluding this variable has no significant difference on our results.

Specialized Metropolitan Areas	Diversified Metropolitan Areas

[COLOR] Figure 3: Internal (within-city) Co-Inventor Networks in Technologically Specialized and Technologically Diversified Urban Areas

Specialized Metropolitan Areas	Diversified Metropolitan Areas

[COLOR] Figure 4: External (between-city) Co-Inventor Networks in Technologically Specialized and Technologically Diversified Urban Areas

This spatial lag term is positive and significant and appears to operate much like the lag term in the models with spatial autocorrelation. Model 7 in Table 4 reports the coefficients for our standard model of knowledge production for the set of cities that are less specialized or more diversified than the median city. Employment size, inventor density, average relatedness and the number of firms per worker are all statistically significant and exert a positive influence on the volume of patents generated within technologically diversified urban areas. The age of technology and R&D spending have no significant influence on patent production. Most important, perhaps, the size of internal and external networks in these diversified cities have no bearing on knowledge production. As we switch to technologically specialized cities in Model 8, scale, inventor density, average relatedness and the age of technology all exert the anticipated positive influence on patenting. R&D spending has no significant effect and the number of firms per worker is negative in line with the findings reported earlier. Most importantly, the size of internal and external co-inventor networks exert a significant positive influence on knowledge production for specialized cities in contrast to the results for diversified cities.

The results in Table 4 provide confirmation that the nature and importance of co-inventor collaboration networks vary with the technological profiles of urban areas. We suspect that in diversified knowledge cities the breadth of the cognitive overlap between groups of inventors is not sufficiently high for dense networks of collaborating agents to form. In contrast, specialized cities channel knowledge development along relatively narrow trajectories that engender greater cognitive overlap and more readily hasten a shared division of labor in the knowledge production process. In turn, the efficiency of greater specialization and interaction sustain higher levels of knowledge output in cities with higher levels of cognitive proximity among inventors. Though our data support this notion, clearly more work is required to bolster this claim.

These issues are explored a little further by examining the factors that shape tie-formation among the pool of inventors distributed across U.S. metropolitan areas. In particular, we seek to analyze whether the factors that influence external collaboration, or collaboration between inventors located in different metropolitan areas, vary between specialized and diversified cities. This analysis begins with a simple gravity model framework where we anticipate that the number of external collaborations recorded for a specific pair of cities is a positive function of the size of those cities, the number of inventors in each city, and a negative function of the distance between them. We add to this simple specification a measure of the cognitive proximity between all pairs of cities, measured as the average relatedness between all patents generated across each city pair in a given year. We hypothesize that as the cognitive proximity between cities increases, so inventors in those cities should be more likely to collaborate. Finally, we classify cities, again using the average relatedness of the patents that they produce, into two subsets – specialized cities and diversified cities, as in the analysis for Table 4. We cut the data set in half so that we are not estimating a model using the paired city-city collaboration observations twice. We are left with a little over 1 million observations. We run one model with the null category of a dummy variable representing diversified cities and we interact all RHS variables with that same dummy variable, the non-zero observations representing specialized cities. This specification allows us to test whether tie formation between specialized cities is significantly different from that in diversified cities across all the independent variables in the model.

Dependent variable: No. of patents (fw)	(7) *Diversified Cities*	(8) *Specialized Cities*
Spatial Lag	0.301***	0.299***
	(0.056)	(0.078)
Employment ‡	1.023***	1.251***
	(0.051)	(0.068)
Inventor density ‡	1.740***	2.309***
	(0.267)	(0.202)
Ave. relatedness ‡	3.226***	0.607***
	(1.193)	(0.138)
Age of technology	0.000	0.000***
	(0.000)	(0.00000)
NSF $ per Emp.	0.0001	0.0001
	(0.0001)	(0.0001)
Firms per Emp. ‡	6.090***	-0.413*
	(1.456)	(0.236)
Internal k-cores	0.009	0.017***
	(0.006)	(0.005)
External k-cores	0.001	0.011***
	(0.004)	(0.003)
N	3890	3890
CBSA	334	335
R-Squared	0.15	0.27

F-Statistic	67.5*** (df=9; 3518)	141.0*** (df=9; 3517)

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

‡ Natural log of variable

All independent variables (except the spatial lag term) are lagged

Table 4: Knowledge Production in Specialized and Diversified Cities

Table 5 reports the results. The dependent variable reports whether inventors in a pair of cities collaborate with one another or not. With this variable taking a categorical form our base model is fit in logit form using maximum likelihood techniques. When the dummy variable, specialization, takes the value 0, the model generates the coefficients for external tie formation for diversified cities. The coefficients in the logit model are to be read as the log odds of the probability of collaboration between a pair of cities. The results show that inventors in diversified cities collaborate more when the number of inventors in the pair of cities under consideration increases, and they collaborate less as the geographical distance between the cities increases. These results are just as we might expect. In addition, as the technological profiles of the pair of cities becomes more similar, as their cognitive proximity increases, then collaboration between inventors in the two cities is more likely. As the index of city specialization (the dummy variable) turns to 1, we see that specialized cities in general engage in significantly less collaboration than their diversified partners (β_1 = -0.153 in table 5). The interactions in the model now reveal how the independent variables influence tie formation for specialized cities relative to diversified cities. These results show that as the size of potential partner cities increases, the effects on the probability of external inventor collaboration is significantly lower in specialized cities than in diversified cities. This might be read as suggesting that size alone is a less important factor for collaboration in technologically specialized urban areas as compared to diversified cities. The positive coefficient on the interaction of geographic distance and specialized cities indicates that inventors in specialized cities are less impacted than inventors in diversified cities by increases in the distance separating them from potential collaborators. Finally, the positive coefficient on the interaction between cognitive proximity and specialized cities shows that technological relatedness is more important to inventors in specialized cities when forming their collaborations than it is for inventors in diversified cities. These results are robust when running a linear probability model and when explicitly estimating the number of between city collaborations in a negative binomial specification.

Overall, these results establish that the forces influencing between-city tie-formation for inventors in specialized urban areas and those in diversified urban areas are significantly different. Tie formation across all cities is a positive function of the size of potential interacting partner cities, a positive function of the similarity of the knowledge base across cities and a negative function of the geographical distance between them. However, inventors in specialized cities are significantly more selective than inventors in diversified cities when it comes to choosing their collaborative partners. They are more likely to engage with co-inventors in other cities when those partners exhibit greater technological similarity and they are less dissuaded by the friction of distance when doing so. The size of interacting partner cities is significantly less important for inventors in specialized cities than inventors in diversified cities. This result adds to the pipelines literature.

Dependent variable: Collaboration (0/1)	
Dummy: Specialization	- .1530685***
	(.0824852)
Inventor city *i* ‡	.9672761***
	(.0060137)
Inventor city *j* ‡	.9824665***
	(.0070511)
Geographical distance ‡	- 1.028428***
	(.0109944)
Cognitive distance ‡	46.52678***
	(2.075717)
Interact. dummy * inventors city *i* ‡	- .0801405***
	(.0093588)
Interact. dummy * inventors city *j* ‡	- .0620131***
	(.0092803)
Interact. dummy * geographical distance ‡	.1836709***
	(.0176951)
Interact. dummy * cognitive distance ‡	9.342122***

	(2.357164)
N	1.034.161
Prob. > Chi^2	0.0000
Pseudo R-Squared	0.4584

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

‡ Natural log of variable

Year fixed effects included, but not shown

Table 5: Collaborative Tie Formation for Inventors Located in Diversified and Specialized Cities

Conclusion

Knowledge production is concentrated in cities where the density of economic agents is relatively high. That density encourages interaction and fuels processes of agglomeration that reinforce urban advantage at least for some economic agents. Where clusters of firms and other economic actors combine to form social networks so the economic advantages of cities are multiplied. We show that urban networking speeds invention within U.S. metropolitan areas after controlling for the influence of agglomeration. Social networks built from alliances of co-inventors within cities and social networks emerging from inventor collaborations between cities accelerate urban invention. In general, internal city networks exert a stronger influence on the pace of urban invention than external networks that link co-inventors across cities. Both internal and external co-inventor networks act as substitutes for agglomeration or the positive influence of inventor density on the pace of knowledge production. Internal and external networks also substitute for one another.

Perhaps most important, the influence of social networks on urban invention is strongly conditioned by the architecture of knowledge found within cities. Metropolitan areas with specialized knowledge cores tend to be associated with more robust or denser social networks of co-inventors that are significantly and positively related to the pace of invention. This is true for both internal social networks and external social networks. Metropolitan areas with diversified knowledge cores have social networks that are much less well-developed than specialized cities and which are not significantly related to urban patenting.

Finally, we report that the social ties linking co-inventors found in different cities are also shaped by the technological characteristics of the knowledge cores in which they reside. Inventors in metropolitan areas that have specialized knowledge cores are significantly less constrained by geographical proximity and significantly more tightly constrained by cognitive proximity in their search for collaborators than are inventors located in urban areas with diversified knowledge cores. This makes sense as specialized places seek to partner with other similarly specialized locations irrespective of distance. The pipelines that connect diversified cities are shorter and less focused in terms of technology.

References

Acs, Z. J., & Audretsch, D. B. (1988). Innovation in large and small firms: an empirical analysis. *The American Economic Review*, *78*(4), 678–690.

Agrawal, A., Cockburn, I., & McHale, J. (2006). Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships. *Journal of Economic Geography*, *6*(5), 571–591.

Amin, A., & Cohendet, P. (2005). Geographies of knowledge formation in firms. *Industry and Innovation*, *12*(4), 465–486.

Anderson, B. S., Butts, C., & Carley, K. (1999). The interaction of size and density with graph-level indices. *Social Networks*, *21*(3), 239–267. http://doi.org/10.1016/S0378-8733(99)00011-8

Arthur, W. B. (1999). Complexity and the economy. *Science*, *284*(5411), 107–109.

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

Baily, M. N., Hulten, C., Campbell, D., Bresnahan, T., & Caves, R. E. (1992). Productivity Dynamics in Manufacturing Plants. *Brookings Papers on Economic Activity. Microeconomics*, *1992*, 187–267. http://doi.org/10.2307/2534764

Baldwin, J. R., Brown, W. M., & Rigby, D. L. (2010). Agglomeration economies: Microdata panel estimates from Canadian manufacturing. *Journal of Regional Science*, *50*(5), 915–934.

Baldwin, J. R., & Gorecki, P. (1998). The dynamics of industrial competition. *Cambridge Books*.

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

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

Bathelt, H., Malmberg, A., & Maskell, P. (2004). Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography*, *28*(1), 31–56. http://doi.org/10.1191/0309132504ph469oa

Boschma, R. (2005). Proximity and Innovation: A Critical Assessment. *Regional Studies*, *39*(1), 61–74. http://doi.org/10.1080/0034340052000320887

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

Breschi, S., & Lissoni, F. (2009). Mobility of skilled workers and co-invention networks: An anatomy of localized knowledge flows. *Journal of Economic Geography*, *9*(4), 439–468. http://doi.org/10.1093/jeg/lbp008

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

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

Burt, R. S. (2000). The network structure of social capital. *Research in Organizational Behavior*, *22*, 345–423.

Butts, C. T., Acton, R. M., Hipp, J. R., & Nagle, N. N. (2012). Geographical variability and network structure. *Social Networks*, *34*(1), 82–100. http://doi.org/10.1016/j.socnet.2011.08.003

Cantner, U., & Graf, H. (2006). The network of innovators in Jena: An application of social network analysis. *Research Policy*, *35*(4), 463–480.

Cantner, U., Meder, A., & Ter Wal, A. L. J. (2010). Innovator networks and regional knowledge base. *Technovation*, *30*(9), 496–507.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly. Mar1990*, *35*(1), 128–152.

Cyert, R. M., & March, J. G. (1963). *A behavioral theory of the firm*. NJ: Englewood Cliffs.

De Nooy, W., Mrvar, A., & Batagelj, V. (2011). *Exploratory social network analysis with Pajek* (Vol. 27). Cambridge University Press.

Duranton, G., & Puga, D. (2001). Nursery cities: Urban diversity, process innovation, and the life cycle of products. *American Economic Review*, *91*(5), 1454–1477. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-0001149837&partnerID=tZOtx3y1

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

Fleming, L., King, C., & Juda, A. (2007). Small Worlds and Regional Innovation. *Organization Science*, *18*(6), 938–954. http://doi.org/10.1287/orsc.1070.0289

Fleming, L., Mingo, S., & Chen, D. (2007). Collaborative Brokerage, Generative Creativity, and

Creative Success. *Administrative Science Quarterly*, *52*(3), 443–475.

Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: Evidence from patent data. *Research Policy*, *30*(7), 1019–1039. http://doi.org/10.1016/S0048-7333(00)00135-9

Gertler, M. S. (2003). Tacit knowledge and the economic geography of context, or The undefinable tacitness of being (there). *Journal of Economic Geography*, *3*(1), 75–99. http://doi.org/10.1093/jeg/3.1.75

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., & Shleifer, A. (1992). Growth in cities. *Journal of Political Economy*, *100*(6), 1126–1152.

Grabher, G. (1993). The weakness of strong ties: the lock-in of regional development in the Ruhr area. *The Embedded Firm: On the Socioeconomics of Industrial Networks*, 255–277.

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

Hagedoorn, J. (1993). Understanding the rationale of strategic technology partnering: Interorganizational modes of cooperation and sectoral differences. *Strategic Management Journal*, *14*(5), 371–385.

Healy, A., & Morgan, K. (2012). Spaces of innovation: Learning, proximity and the ecological turn. *Regional Studies*, *46*(8), 1041–1053.

Henderson, J. V. (2003). Marshall's scale economies. *Journal of Urban Economics*, *53*(1), 1–28.

Jacobs, J. (1969). *The economy of cities*. Random House. Retrieved from https://books.google.nl/books/about/The_economy_of_cities.html?id=GgyxAAAAIAAJ&pgis=1

Jaffe, A. B. (1986). Technological opportunity and spillovers of R&D: evidence from firms' patents, profits and market value. *National Bureau of Economic Research - Working Paper*.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics*, *108*(3), 577–598. Retrieved from http://www.nber.org/papers/w3993

Kauffman, S. A. (1993). *The origins of order: Self organization and selection in evolution*. Oxford university press.

Klepper, S. (1996). Entry , Exit , Growth , and Innovation over the Product Life Cycle. *The American Economic Review*, *86*(3), 562–583.

Knoben, J., & Oerlemans, L. A. G. (2012). Configurations of inter-organizational knowledge links: Does spatial embeddedness still matter? *Regional Studies*, *46*(8), 1005–1021.

Kogler, D. F., Essletzbichler, J., & Rigby, D. (2016). The evolution of specialization in the EU15 knowledge space. *Journal of Economic Geography*.

Kogler, D. F., Rigby, D. L., & Tucker, I. (2013). Mapping Knowledge Space and Technological Relatedness in US Cities. *European Planning Studies*, *21*(9), 1374–1391. http://doi.org/10.1080/09654313.2012.755832

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

Li, G.-C., Lai, R., D'Amour, A., Doolin, D. M., Sun, Y., Torvik, V. I., … Fleming, L. (2014). Disambiguation and co-authorship networks of the US patent inventor database (1975–2010). *Research Policy*, *43*(6), 941–955.

Lobo, J., & Strumsky, D. (2008). Metropolitan patenting, inventor agglomeration and social networks: A tale of two effects. *Journal of Urban Economics*, *63*(3), 871–884. http://doi.org/10.1016/j.jue.2007.07.005

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

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

Matula, D. W., & Beck, L. L. (1983). Smallest-last ordering and clustering and graph coloring algorithms. *Journal of the ACM*, *30*(3), 417–427. http://doi.org/10.1145/2402.322385

Neffke, F., Henning, M., & Boschma, R. (2011). How Do Regions Diversify over Time? Industry Relatedness and the Development of New Growth Paths in Regions. *Economic Geography*, *87*(3), 237–265. http://doi.org/10.1111/j.1944-8287.2011.01121.x

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

Nooteboom, B. (1994). Innovation and diffusion in small firms: theory and evidence. *Small Business Economics*, *6*(5), 327–347.

Potter, A., & Watts, H. D. (2011). Evolutionary agglomeration theory: increasing returns, diminishing returns, and the industry life cycle. *Journal of Economic Geography*, *11*(3), 417–455.

Powell, W. W. (1990). Neither market nor hierarchy. *Research in Organizational Behavior*, *12*, 295–336.

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

Rigby, D. L., & Brown, W. M. (2015). Who benefits from agglomeration? *Regional Studies*, *49*(1), 28–43.

Rogers, M. (2004). Networks, firm size and innovation. *Small Business Economics*, *22*(2), 141–153.

Rowley, T., Behrens, D., & Krackhardt, D. (2000). Redundant governance structures: An analysis of structural and relational embeddedness in the steel and semiconductor industries. *Strategic Management Journal*, 369–386.

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

Seidman, S. B. (1983). Network Structure and Minimum Degree. *Social Networks*, *5*, 269–287.

Singh, J. (2005). Collaborative Networks as Determinants of Knowledge Diffusion Patterns. *Management Science*, *51*(5), 756–770. http://doi.org/10.1287/mnsc.1040.0349

Singh, J., & Fleming, L. (2010). Lone Inventors as Sources of Breakthroughs: Myth or Reality? *Management Science*, *56*(1), 41–56. http://doi.org/10.1287/mnsc.1090.1072

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

Sorenson, O., & Stuart, T. E. (2001). Syndication networks and the spatial distribution of venture capital investments. *American Journal of Sociology*, *106*(6), 1546–1588. http://doi.org/10.1086/321301

Strumsky, D., & Thill, J.-C. (2013). Profiling Us Metropolitan Regions By Their Social Research Networks and Regional Economic Performance. *Journal of Regional Science*, *53*(5), 813–833. http://doi.org/10.1111/jors.12048

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

Uzzi, B. (1997). Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness. *Administrative Science Quarterly*, *42*(1), 35–67. http://doi.org/10.2307/2393808

Walker, G., Kogut, B., & Shan, W. (1997). Social capital, structural holes and the formation of an industry network. *Organization Science*, *8*(2), 109–125.

Wernerfelt, B. (1984). A Resource based view of the firm. *Strategic Management Journal*, *5*(2), 171–180. http://doi.org/10.1002/smj.4250050207

Whittington, K. B., Owen-Smith, J., & Powell, W. W. (2009). Networks, Propinquity, and Innovation in Knowledge-intensive Industries. *Administrative Science Quarterly*, *54*(1), 90–122. http://doi.org/10.2189/asqu.2009.54.1.90

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036–9. http://doi.org/10.1126/science.1136099