Co-invention networks and inventive productivity in US cities

Stefano Breschi [a,1], Camilla Lenzi [b,*]

[a] Department of Management and Technology and CRIOS, Bocconi University, Via Sarfatti 25, I-20136 Milan, Italy
[b] Department of Architecture, Built Environment and Construction Engineering, Politecnico di Milano, Piazza Leonardo da Vinci 32, I-20133 Milan, Italy

ABSTRACT

The role of collaboration networks within and across cities as drivers of urban creativity and new knowledge creation is increasingly acknowledged in the literature. We propose that the combination of (1) high internal social proximity between co-inventors within a city and (2) local cliques of inventors in which interaction is dense allows a city to achieve greater inventive creativity. Internal social proximity allows knowledge to circulate quickly across a larger pool of sources; dense cliques promote trust, cooperation, and a more effective use of the acquired knowledge. Moreover, social proximity between a city's inventors and inventors outside the city contributes to enriching and renewing a city's knowledge base by facilitating faster access to fresh external knowledge. We find evidence to support these propositions in a study of the inventive productivity of 331 US cities.

1. Introduction

The role of collaboration networks as drivers of regional and urban innovation has gained paramount attention during the last decade. The idea that social ties and inter-personal contacts mediate the transmission of knowledge and are key explanatory factors of the urban concentration of innovative activities is not a new one and has been widely debated in the literature on agglomeration economies (Beaudry and Schiffauerova, 2009; Glaeser and Gottlieb, 2009). However, until recently, a lack of micro-level data and for-mal network models has prevented a rigorous empirical evaluation of social networks' effects on innovation. According to Duranton and Puga (2004) and Puga (2010), in fact, the micro-foundations of the learning mechanisms upon which knowledge spillovers are based remain relatively less developed with respect to the devel-opment of the theoretical and empirical micro-foundations of other agglomeration economies. The use of patent data as relational data can offer an empirical contribution in this direction, as patent data can be employed to map the socio-professional networks in which inventors are embedded (Ter Wal and Boschma, 2009).

The extant literature does not offer conclusive evidence on the importance of social networks for inventive performance. Social proximity has been found to explain a great deal of the tendency for knowledge to diffuse locally (Agrawal et al., 2006 and

2008; Breschi and Lissoni, 2004; Singh, 2005). However, several studies have consistently shown that knowledge creation and inventive performance in a metropolitan area depend more on the agglomeration of inventors and creative individuals than on any structural property of the co-invention network (Bettencourt et al., 2007; Fleming et al., 2007; Lobo and Strumsky, 2008).

This paper provides an empirical contribution to the litera-ture on knowledge-related agglomeration economies (Rosenthal and Strange, 2001) and the micro-foundations of learning mechanisms and knowledge spillovers (Duranton and Puga, 2004) by proposing a more careful examination of the relative importance of agglomeration forces versus social networks (and their struc-tural properties) on a city's inventive performance. Importantly, this work accounts for not only the network structure within a metropolitan area but also for the ties and the related knowledge flows linking inventors located in different cities. Non-local sources of knowledge have been found to provide a significant contribution to the diffusion of ideas and to patenting growth at the local level (Agrawal et al., 2008, 2010; Kerr, 2010).

In particular, we argue that a city will achieve higher inventive productivity when its co-invention network presents a combination of two key properties: high social proximity between network members in the city, i.e., *internal social proximity*, and local cliques of co-inventors in which interaction is dense, i.e., *clique density*. In this context and throughout the paper, social proximity is high when interactions quickly link back to individuals participating in the network. For example, suppose that A interacts with B, B inter-acts with C, and C interacts with D. In this instance, two members of the network, B and C, separate A from D. Therefore, the social

* Corresponding author. fax: +39 02 2399 9477.
  E-mail addresses: stefano.breschi@unibocconi.it (S. Breschi), camilla.lenzi@polimi.it (C. Lenzi).
[1] fax: +39 02 58363399.

proximity between A and D is much greater than if, for example, ten other individuals were to separate A and D. Using this same example, additional cross-interactions between network members increase the sense in which the network is more prone to the formation of cliques and in that respect is denser. This would be the case if, for instance, A and C also interact directly. In this case, A and C would both have a direct tie to an additional network member, more than in the original example. Moreover, A, B, and C would represent a clique, namely a cohesive group of individuals who are directly connected to one another. In addition, we propose that the impact of internal social proximity on a city's inventive performance also critically interacts with the social proximity of metropolitan inventors to inventors in other cities, i.e., *external social proximity*.[1]

We test our hypotheses on a database covering 331 US Metropolitan Statistical Areas (MSAs)[2], their inventors, and the respective patents applied for at the European Patent Office (EPO). A major concern in our empirical exercise regards the possible endogeneity between the patenting rate and network variables, which may arise if inventors are attracted to and develop net-work ties in cities that are already highly innovative. To address this issue, we adopt the following strategy. First, we attempt to mitigate the problem by taking a significant time lag between our dependent and independent variables. Specifically, the depen-dent variable measures patent productivity in 2009, whereas net-work variables are computed in the time window 1995–1999. Nev-ertheless, this approach cannot completely rule out the possibil-ity of endogeneity and the resulting bias in the estimated coef-ficients. For this reason, after providing OLS estimates, we check and control for possible endogeneity using the instrumental vari-able technique proposed by Lewbel (2012). This approach identifies structural parameters in regression models affected by endogeneity by supplementing available external instruments with generated ones that are uncorrelated with the product of heteroskedastic errors.

The remainder of the paper proceeds as follows. In the next section, we conceptually derive from the literature the research hy-potheses to be verified. Section 3 describes the construction of our key network variables. Section 4 presents the empirical models to be tested and the data. Section 5 discusses the empirical results and comments on the robustness checks to detect and control for possible endogeneity. Finally, Section 6 concludes.

2. Co-invention networks and inventive productivity

Inventive activity in the US is a predominantly urban phe-nomenon (Carlino et al., 2007; Feller, 1971). Approximately 94% of all patent applications made by US organizations in the period 1990–2009 were generated within MSAs, with the ten most prolific cities accounting for approximately 48% of all patenting activity in the period. The tendency for innovative activities to cluster in se-lected cities has been attributed to the importance of agglomera-tion economies. Agglomeration economies, especially knowledge-related ones, are the basis of enhanced economic performance and

creativity in cities (Glaeser, 1999; Glaeser et al., 1992; Henderson, 2003; Rosenthal and Strange, 2001; Rosenthal and Strange, 2008; Glaeser and Gottlieb, 2009).

Metropolitan settings are key engines and incubators of new knowledge creation processes because they facilitate intellectual linkages among individuals through social proximity and face-to-face contacts. As discussed by Glaeser and Gottlieb (2009), the thick web of social interactions in cities creates agglomeration economies, which can lead to considerable variation over time and space in innovative episodes. The co-location of creative individu-als within the same region or urban environment is credited with facilitating both formal interactions and informal or serendipitous encounters in which the tacit knowledge relevant for inventive cre-ativity is transmitted and exchanged. This network of relationships generates pervasive localized knowledge flows among individuals and firms and guarantees the rapid diffusion of ideas at the local level, which in turn boosts the inventive productivity of all local actors (Jaffe et al., 1993).

Recent literature suggests that two specific network structural properties (and their combination) are particularly desirable for knowledge diffusion and creation (Schilling and Phelps, 2007; Uzzi and Spiro, 2005). First, the actors in the network are able to reach other actors in the network through a relatively low number of intermediaries; i.e., on average, they are socially proximate. Sec-ond, the actors in the network are locally clustered in the sense that they tend to create tightly knit groups (i.e., cliques) charac-terized by a relatively high density of ties (Watts and Strogatz, 1998).

In the context of a co-invention network, knowledge and in-formation tend to diffuse more rapidly, and with less noise, when relatively few intermediaries separate inventors (i.e., when internal social proximity is high), than when members are connected by longer chains of ties (i.e., when internal social proximity is low). As a consequence, new information or ideas generated within the network may rapidly reach (or flow to) all other members of the network and be recombined with their own knowledge, thereby improving inventive productivity.

Furthermore, when inventors are embedded in cohesive cliques, in which an actor's partners also collaborate with one another, in-formation spreads quickly, and more important, its usefulness and reliability is verified along multiple pathways (Schilling and Phelps, 2007). Moreover, the high density of linkages within a clique cre-ates conformity (Patacchini and Venanzoni, 2014) and a common code of communication, which stimulates collective learning, an argument also suggested in the debate on knowledge-related ag-glomeration economies (Glaeser and Gottlieb, 2009). Finally, dense cliques may allow network members to monitor opportunistic behavior, which promotes trust and reciprocity among partners, thereby encouraging higher levels of collaboration (Schilling and Phelps, 2007; Uzzi, 1996). Therefore, denser cliques allow knowl-edge to be shared and used rapidly, spurring greater knowledge creation.

Thus, we would expect metropolitan inventive performance to improve where there is a combination of (1) high internal social proximity between network members and (2) local cliques of ac-tors in which interaction is dense. When a network's internal so-cial proximity is low, dense collaborative cliques may find it dif-ficult to maintain high levels of invention because there are few between-clique or between-team links that promote the transfer of knowledge and ideas generated elsewhere in the larger network. Similarly, when internal social proximity is high, the lack of dense cliques and of the redundancy of ties may be equally detrimental to invention. Although information may circulate rapidly, it is not being spread through known and trusted sources, which may lead to a less effective exploitation of new ideas. Thus, we posit the fol-lowing:

---

[1] Following Fleming et al. (2007) and Lobo and Strumsky (2008), we will use the term metropolitan network or internal co-invention network to denote the sub-set of inventors, and the ties among them, in a given city; we will refer to the metropolitan network's structural properties as its internal structure. Accordingly, social proximity among inventors in a city is labeled *internal social proximity*. We will use the term external ties to denote links connecting metropolitan inventors with inventors located in different cities. Accordingly, social proximity among in-ventors across cities is labeled *external social proximity*.

[2] MSAs are defined by the US Office of Management and Budget (OMB) as urban core areas of at least 50,000 people, plus adjacent counties that have a high degree of social and economic integration with the core, as measured by commuting ties. The choice of using EPO rather than USPTO data is fully discussed in Appendix A.

Hypothesis 1: *Within a city's co-invention network, a combination of high internal social proximity between inventors and high clique density will be positively associated with higher rates of invention in the city.*

Local networking, however, may also present "a dark side" (Burt, 2004). Regions and cities that are too inwardly oriented may actually reduce their inventive potential and lose position in the urban ranking (Neal, 2010). Actors in the local economy tend to converge toward a common and homogenous pool of knowledge, such that the variety of technological approaches and solutions de-creases and the risk of lock-in increases. Moreover, external ties can provide considerable resources and information advantages that would not otherwise be available locally: new information concerning market opportunities and unmet demand (Bresnahan et al., 2001); the awareness of specialized skills and human capital (Gittelman 2007, Kerr 2010); and access to a larger repertoire of technological and organizational solutions (Owen-Smith and Powell, 2004). More generally, external ties forged between individuals, such as between co-inventors, have been identified as an effective vehicle of information transmission across "geographical holes" in a network (Bell and Zaheer, 2007).

Therefore, the degree to which a city's co-invention network is embedded into the broader (national or global) knowledge net-work may be as important as the structure of local networking ties. The easier and faster the access to external knowledge is be-cause of closer social ties among one city's actors with other cities' actors (i.e., higher external social proximity), the more knowledge creation will be enhanced. However, excessive reliance upon ex-ternal sources of knowledge can pose the risk of technological dependence and may undermine endogenous innovation capabil-ities. Moreover, external knowledge will have a limited impact on a city's inventive productivity unless the metropolitan network has a pool of actors able to absorb and diffuse that knowledge (Cohen and Levinthal, 1990). When a network's internal social proximity is low, knowledge received through external ties is more likely to be lost through transmission leakages and gaps. Conversely, if internal social proximity is high and the circulation of ideas among actors in the city is rapid, knowledge received from external sources is likely to benefit a larger number of local individuals, with limited noise, thus allowing for more effective knowledge absorption and recombination at the local level. Hence, we posit the following:

Hypothesis 2: *The combination of (1) high internal social proximity among inventors within a focal city and (2) high external social proximity among inventors in the focal city and inventors in other cities will be positively associated with higher rates of invention in the focal city.*

3. Measuring co-invention networks

3.1. Patents as relational data

The hypothesis that social networks and personal interactions affect innovative productivity has only recently become amenable to direct empirical testing. The adoption of patent data as a source of relational data now makes it possible to examine networks us-ing the tools of social network analysis and graph theory (Breschi and Lissoni, 2004; Singh, 2005). In such a framework, the nodes of the network are inventors and the edges of the network link co-inventors listed on the same patent document. In other words, we assume that a pair of inventors is connected if they are jointly designated as inventors in one or more patent documents.

To test our hypotheses and construct the co-invention network, we extracted data on all patent applications made by US organiza-tions at the EPO, using the CRIOS-PATSTAT database. Next, we ex-tracted the names and addresses of inventors as they were listed

on each patent. Because the validity of results derived from social network analysis techniques depends on the correct identification of individual inventors (i.e., nodes), we thoroughly cleaned and standardized their names and addresses.[3] We used the reported addresses to assign each inventor to one of the 370 US MSAs, us-ing the definition files available on the US Census Bureau website.[4]

We excluded from the analysis the MSAs of Hawaii, Puerto Rico, and Alaska, along with 30 other MSAs, either because there were no patents filed in the MSA in our sample period or because the data used to control for other economic characteristics of the city were missing or incomplete. The final data set includes informa-tion on 331 MSAs. The selected MSAs account for approximately 94% of all EPO patent applications made by US organizations in the period 1990–2009. Additional details on our methodology for network construction are provided in Appendix B.

3.2. Measuring social proximity

Internal social proximity among inventors in a network is com-puted as the average across all nodes (i.e., inventors) in the net-work of the sum of the reciprocal social distances to all other nodes k she can reach in the co-invention network within the city:

$$\text{Internal social proximity}_{\text{MSA}_i} = \frac{\sum_{j=1}^{n} \sum_{\substack{k=1 \\ j \neq k}}^{n} \frac{1}{d_{jk}}}{n} \qquad (1)$$

where d_{jk} is the geodesic social distance (i.e., the smallest number of intermediaries) that separates inventor j from inventor k in the co-invention network internal to a city[5], and n is the number of inventors located in city i. The index ranges between 0 (every in-ventor in a city is an isolate, i.e., when she does not collaborate with any other inventor in the city) and n (every inventor directly collaborates with every other inventor in the city).[6]

The existence of local cliques of inventors in which interaction is dense is captured by the extent to which the collaborators of an inventor are also partners with one another. More formally, this is measured as follows:

$$\text{Clique density}_{\text{MSA}_i} = \frac{(\text{number of closed paths of length two})_i}{(\text{number of paths of length two})_i} \qquad (2)$$

In particular, a path of length two is any set of three nodes (e.g., u, v, and w) that are connected by two edges (e.g., u collaborates with v, and v collaborates with w). A path of length two is consid-ered closed if the first and third nodes are also directly connected (i.e., if u also collaborates with w). Clique density is measured as the fraction of paths of length two in the network that are closed. This index varies between 0 (there are no closed paths of length two) and 1 (all paths of length two are closed).[7]

---

[3] Lissoni et al. (2006) provide fuller details on the routine implemented to uniquely identify inventors.

[4] In this paper, we have adopted the June 2003 definition of MSAs. For fuller details and maps, please see http://www.census.gov/population/www/metroareas/metrodef.html. When information on an inventor's MSA was unavailable, we used the state and the zip code to assign the inventor to the corresponding MSA us-ing ZIPList5, a commercial database listing every active ZIP code currently defined by the US Postal Service (http://www.zipinfo.com/products/z5cbsa/z5cbsa.htm). For each ZIP code that is part of a metro area, the database identifies the MSA in which the ZIP code lies. This database uses the MSA definitions issued by the OMB in June 2003.

[5] The reciprocal of an infinite social distance, i.e., when two inventors in the net-work are disconnected, is set at 0.

[6] In the network literature, this index is commonly called the average social dis-tance weighted reach. A detailed discussion of the key properties and advantages of this index is available in Appendix C.

[7] In the network literature, this index is commonly referred to as the clustering coefficient. This index is problematic in the case of affiliation networks, such as

According to Hypothesis 1, we expect that the interaction between internal social proximity and clique density will be positively associated with higher levels of inventive productivity.

Finally, we measured the extent to which metropolitan inventors have external ties by computing the external social proximity between inventors located in a given city and all other inventors located in all other cities. Formally, this variable is computed as follows:

$$\text{External social proximity}_{MSA_i} = \frac{\sum_{j=1}^{n_i} \sum_{h=1}^{n_h} \frac{1}{d_{jh}}}{n_i} \qquad (3)$$

where n_i denotes the number of inventors located in city i, n_h denotes the number of inventors located in other cities (i.e., not located in city i), and d_{jh} denotes, as above, the geodesic social distance in the US global co-invention network between inventor j and inventor h. The index ranges from zero (i.e., all inventors in city i are not connected to any external inventor) to n_h in the (theoretical) case in which every inventor in city i directly collaborates with every other inventor in every other city.[8]

According to Hypothesis 2, we expect the effect of external social proximity on metropolitan inventive productivity to be complementary to the effect of internal social proximity.

4. Data and estimation strategy

4.1. Model specification

Our basic regression equations are the following:

$$\log(\text{patents/workers})_i = \alpha + \beta_1 \ln(\text{ISP}_i) + \beta_2 \text{CD}_i$$
$$+ \beta_3 \ln(\text{ISP}_i)\text{CD}_i + \sum_h \delta_h Z_{ih} + \varepsilon_i \qquad (4a)$$

$$\log(\text{patents/workers})_i = \alpha + \gamma_1 \ln(\text{ISP}_i) + \gamma_2 \ln(\text{ESP}_i)$$
$$+ \gamma_3 \ln(\text{ISP}_i)\ln(\text{ESP}_i) + \sum_h \delta_h Z_{ih} + \varepsilon_i \qquad (4b)$$

where, for each city i, the dependent variable is patent productivity measured as the log of the number of patents per worker in 2009, ISP is *internal social proximity*, CD is *clique density*, and ESP is *external social proximity*. Moreover, the equations include a series of other control variables (described in Section 4.4) (Z_{ih}), which are intended to capture the impact of agglomeration economies on inventive productivity, plus an idiosyncratic error term (ε_i).

4.2. Dependent variable

Our dependent variable is the innovative productivity of a city, and it is measured as the log of the total number of patents per worker employed in the private sector in a given city in 2009. Patents are dated according to their priority year, i.e., the first date at which the patent was applied for anywhere in the world, which is closest to the actual time of the invention. We took all EPO patents with priority year 2009 and assigned them to MSAs using the address of inventors as reported in the patent document. The number of patents has been normalized to account for a city's size by employment rather than by population, as the former most

likely represents the segment of the population engaged in inventive activities. We used the data on the average annual employment in the private sector released by the Bureau of Labor Statistics (BLS), Quarterly Census of Employment and Wages (QCEW)[9], available at the county level and aggregated to the MSA level.[10]

4.3. Network variables

We constructed our network variables by considering the co-invention network in the period 1995–1999. By taking a significant time lag between our dependent variable and measures of network structure, we are attempting to minimize concerns of reverse causation and endogeneity in our estimated coefficients. Section 3 has already outlined our key network variables: *internal social proximity, clique density*, and *external social proximity*.

4.4. Agglomeration variables

Our regressions include several variables intended to capture the importance of agglomeration economies for inventive activities (see Carlino et al., 2007 for a similar approach).

Inventor density in each MSA accounts for the city scale effect associated with the agglomeration of inventive individuals, which is well documented in the literature (Bettencourt et al., 2007; Lobo and Strumsky, 2008). This variable is also lagged and has been computed as the ratio between the total number of inventors active in a given MSA in the period 1995–1999 and the MSA land area measured in square miles. It has been inserted in logs in the regressions.

To control further for the scale of local inputs available for the inventive process, our regressions also include a measure of R&D intensity and of human capital at the MSA level. The measure of R&D intensity used is the amount of R&D expenditures in science and engineering at local colleges and universities divided by the number of full-time students enrolled at colleges and universities in the MSA in 1999 (i.e., *University R&D per student* as in Carlino et al. (2007)). Data on university R&D were derived from the NSF WebCaspar search engine.[11]

College educated per worker, instead, captures the local availability of human capital; it is computed as the ratio between the number of individuals (over 25 years of age) living in MSA i with a college degree or more education   in 2000 and the total number of   employees in the same year. Data on education were derived from the 2000 Census at the county level and aggregated to the MSA level.[12]

Income per capita controls for a city's economic wealth.[13] Moreover, *population density*, measured as the metropolitan population

---

co-invention networks, and tends to become inflated and biased as the number of inventors per patent increases (Newman, 2001, 2010). If patents contain more than two inventors, closed paths of length two will arise automatically, inflating its value. Therefore, we excluded from its computation those paths formed by inventors who are connected because of a joint patent and only counted those paths that are the outcome of independent interactions between pairs of inventors, as recommended by Opshal (2010).

[8] A detailed discussion of the key properties and advantages of this index is available in Appendix C.

[9] http://www.bls.gov/cew.

[10] For all data used in the empirical analysis, which are available at the county level, we used the June 2003 delineation file of MSAs to aggregate them at the MSA level. The delineation file is available at http://www.census.gov/population/metro/files/lists/2003/03cbsa.txt.

[11] The ideal R&D measure would be private R&D expenditures. Unfortunately, these data are not available at the MSA level. However, the simple correlation co-efficient between the number of inventors active in each of the 50 US States and business funds for industrial R&D performance in 1999 (see Table A-32, Research and Development in Industry: 1999, National Science Foundation, available at http://www.nsf.gov/statistics/nsf02312/pdf/a32.pdf) is approximately 0.96. For these rea-sons, we are quite confident that our variables represent an acceptable proxy for the scale of inventive inputs at the MSA level. Data on R&D were retrieved from the *NSF Survey of Research and Development Expenditures at Universities and Colleges* and data on students from the *Integrated Postsecondary Education Data System (IPEDS) Enrollment Survey* as archived at https://ncsesdata.nsf.gov/webcaspar.

[12] For details, see http://www.census.gov/support/USACdata.html.

[13] Data on income per capita was taken from BEA Regional Accounts (http://www.bea.gov/regional/). Income was deflated by using the Consumer Price Index (CPI) as reported by the US BLS (http://www.bls.gov/cpi/), where the reference year is 1982–84.

Table 1
Summary statistics[a].

Variable	Mean	Std. dev.	Min	Max
Patents per worker (2009)	0.0004	0.0005	0.00	0.0035
University R&D per student (1999)	1.17	2.08	0.00	13.73
College educated per worker (2000)	0.25	0.07	0.08	0.56
Income per capita (1999)	15,262.3	2577	8545.4	32,144.8
Population density (1999)	2628.62	692.42	865.28	6606.58
Herfindahl index on 35 IPC classes (1995–1999)	0.10	0.09	0.01	1.00
RTA[b] Electronics (1995–1999)	−0.22	0.34	−0.92	0.46
RTA Scientific Instruments (1995–1999)	−0.09	0.31	−0.89	0.68
RTA Chemicals (1995-1999)	−0.06	0.33	−0.94	0.68
RTA Pharma-Biotech (1995–1999)	−0.18	0.38	−0.94	0.63
RTA Industrial Processes (1995–1999)	0.10	0.34	−0.84	0.81
RTA Mechanical Engineering (1995-1999)	0.07	0.39	−0.90	0.85
RTA Consumer Goods (1995–1999)	0.09	0.37	−0.82	0.88
Average age of prior art (1995–1999)	6.99	2.06	0	15.33
Share of employment in manufacturing (1999)	0.18	0.09	0.03	0.55
Share of employment in professional and business services (1999)	0.12	0.04	0.04	0.27
Herfindahl index at the firm level (1995–1999)	0.15	0.16	0.00	1.00
Establishments per worker (1999)	0.07	0.01	0.04	0.14
Index of enforcement of non-compete laws (2004)	4.49	2.03	0	9
Inventor density (1995–1999)[c]	0.22	0.51	0.001	5.11
Internal social proximity (1995–1999)[d]	7.22	19.56	0.00	158.6
Clique density (1995–1999)[e]	0.18	0.22	0.00	1.00
External social proximity (1995–1999)[f]	348.87	391.20	0.17	2274.5

[a] Number of MSA observations = 331.
[b] RTA = Revealed technological advantage index (standardized between −1 and +1).
[c] Inventor density is the ratio between the total number of inventors active in a given MSA and the MSA land area measured in square miles.
[d] Equation [1] provides the formal definition of internal social proximity.
[e] Equation [2] provides the formal definition of clique density.
[f] Equation [3] provides the formal definition of external social proximity.

divided by the land area (in square miles) of the MSA, roughly accounts for the presence of urbanization (Jacobs) externalities. To the extent that agglomeration economies outweigh congestion effects, we would expect this variable to positively affect metropolitan patenting productivity (Carlino et al., 2007). Both variables have been measured in 1999 and have been inserted in logs in the regressions.

In addition, the Herfindahl index at the technology level, computed using the share of patents made in each of 35 technolog-ical fields (i.e., *Herfindahl index on 35 IPC classes*) in the period 1995–1999[14], captures to what extent a city is specialized in a narrow set of technological fields. This variable controls for the presence of externalities arising from technological specialization that several empirical studies suggest prevail over urbanization effects (Henderson, 2003; Rosenthal and Strange, 2003, 2004). The revealed technological advantage (RTA) index in each MSA in the period 1995–1999 measures specialization in seven broad techno-logical areas, namely electronics, scientific instruments, chemicals, pharmaceuticals and biotechnology, industrial processes, mechanical engineering, and consumer goods, and accounts for variations in patent propensity across technologies. Moreover, the *average age of prior art* patents cited by patents in each MSA captures the extent to which metropolitan inventors are engaged in newer or more fertile technological areas; it is measured in years, by considering patents with priority date in the period 1995–1999, and it is introduced in logs in the regressions. Additionally, our regressions include the *share of employment in manufacturing* and *in profes-sional and business services* in 1999, computed on data taken from the BLS - QCEW (codes 1013 and 1024, respectively).

The Herfindahl index at the patent-assignee level (i.e., *Herfind-ahl index at the firm level*) in the period 1995–1999 controls for the extent to which inventive activities within a city are concen-

trated in the hands of few firms. Similarly, the number of *establish-ments per worker* in each MSA in 1999 accounts for the local mar-ket structure and captures whether more competitive MSAs enable greater knowledge creation (Glaeser et al., 1992). Data on the number of establishments were taken from BLS – QCEW and aggre-gated from the county to the MSA level; it has been introduced in logs in the regressions.

We also added a variable accounting for the degree of enforcement of non-compete laws in the state where each city is located (i.e., *index of enforcement of non-compete laws*), as the enforcement of such laws may prevent knowledge circulation and inhibit creative efforts (Gilson, 1999). The index considers 12 questions for each jurisdiction and assigns one point to each jurisdiction for each question if the jurisdiction's enforcement of that dimension of noncompetition law exceeds a given threshold. Thus, the index varies from 0 to 12, with higher values signaling a stronger enforcement of non-compete agreements (Garmaise, 2011; Malsberg, 2004).

Finally, all regressions include BEA region dummies.
Table 1 provides descriptive statistics of the variables used in our study, while Table 2 illustrates the values of the dependent and of the network variables for six selected MSAs.

On average, the number of patents per worker is equal to 0.0004. Moreover, Table 2 reveals the existence of sizeable spatial heterogeneity across cities with respect to inventive productivity. Regarding the key network variables are concerned, each inventor can reach on average approximately 7 other inventors within the city and nearly 350 outside the city.

5. Empirical results

5.1. The impact of co-invention network structure on inventive productivity

This section comments on the results of the estimation of equations [4a] and [4b]. For all models, coefficients on agglomera-tion variables (see Section 4.4) were consistent with estimates in

[14] To this purpose, we exploited the reclassification table of International Patent Classification (IPC) codes into 35 technology fields compiled by the World Intellec-tual Property Organization (WIPO) and available at http://www.wipo.int/ipstats/en/statistics/technology_concordance.html.

Table 2
Values of network variables for selected MSAs[a].

Patents per worker	Number of inventors	Inventor density[b]	Internal social proximity[c]	Clique density[d]	External social proximity[e]
Boston-Cambridge-Quincy, MA-NH					
0.00145	9633	2.75	36.74	0.84	516.12
San Francisco-Oakland-Fremont, CA					
0.00196	8660	3.50	90.71	0.73	880.24
Chicago-Naperville-Joliet, IL-IN-WI					
0.00052	7033	0.98	40.75	0.73	588.46
Cleveland-Elyria-Mentor, OH					
0.00060	1771	0.88	7.75	0.24	241.34
Austin-Round Rock, TX					
0.00063	1628	0.39	4.63	0.52	237.77
Orlando, FL					
0.00017	443	0.13	3.34	0.88	219.02

[a] Patents per worker is measured in 2009. The other variables are measured over the period 1995–1999.
[b] Inventor density is the ratio between the total number of inventors active in a given MSA and the MSA land area measured in square miles.
[c] Equation [1] provides the formal definition of internal social proximity.
[d] Equation [2] provides the formal definition of clique density.
[e] Equation [3] provides the formal definition of external social proximity.

the literature and are not reported with the exception of inven-tor density[15]. The full set of estimates is available in Table D1 in Appendix D. Our primary results are reported in Table 3. Models 1 and 2 report ordinary least-squares (OLS) estimates with standard errors robust to arbitrary heteroscedasticity.

Inventor density shows, as expected, a consistently positive and significant effect across all models. It captures the positive effect of the inventive potential available in a city, which is primarily related to the concentration of skilled and creative individuals and to the positive scaling effect between patenting and population detected by Bettencourt et al. (2007).

Regarding the network variables, the coefficient on the inter-action term between internal social proximity and clique density (i.e., β_3 in [4a]) shows a positive and highly statistically signif-icant effect (Model 1, Table 3) on inventive productivity, thereby supporting Hypothesis 1. It is important to note that the negative coefficient on internal social proximity (i.e., β_1 in [4a]) has to be interpreted as the effect of internal social proximity on the depen-dent variable when the value of clique density is zero (Jaccard and Turrisi, 2003). In the absence of dense cliques of interaction (i.e., a high level of clique density), cities may lose the ability to en-sure the rapid diffusion and recombination of ideas and knowledge and, consequently, inventive performance may decline and ultimately fail.[16] Similarly, the effect of clique density (i.e., β_2 in [4a]) is indeed negative when the (log of) internal social proxim-ity is equal to zero, i.e., when internal social proximity is equal to one, but its effect becomes positive when internal social proximity is equal to 4.5 (= $\exp(0.286/0.190)$) (i.e., for relatively small values). When internal social proximity is below a minimum thresh-old, the network is characterized by dense cliques of collaborators without any direct or indirect links to other teams of inventors. In those cases, information spreads rapidly within cliques, but highly redundant knowledge is likely to circulate, which is detrimental to inventive creativity.[17]

Turning to our Hypothesis 2, the coefficient on the interac-tion term between internal and external social proximity (i.e., γ_3 in [4b]) shows a positive and highly statistically significant effect (Model 2, Table 3), suggesting that the impact of internal social proximity on inventive productivity increases with the amount of external knowledge available to local inventors. The negative and significant effect of internal social proximity (i.e., γ_1 in [4b]) in-dicates that in the absence of non-redundant and fresh external knowledge, inventive productivity may be hindered, as the local knowledge base may crystallize around a common and homoge-nous pool of knowledge, thus inhibiting the variety of technologi-cal approaches and solutions and increasing the risk of technologi-cal lock-in. Similarly, the negative sign of external social proximity (i.e., γ_2 in [4b]) suggests that external knowledge will have a lim-ited impact on a city's inventive productivity unless the metropoli-tan network has a pool of actors able to trans-code, absorb, and diffuse that knowledge. Disproportionate reliance upon external sources of knowledge may not have effects on productivity if it is not associated with an internal network structure facilitating knowledge circulation.

5.2. Detecting and controlling for endogeneity

Despite having substantially lagged the various variables (by up to 15 years in the case of the network- and patent-based vari-ables), we nevertheless cannot rule out the possibility that se-rial correlation in MSA-level activity could cause our network and other control measures (chiefly, inventor density) to be correlated with the model error term. Indeed, highly innovative cities in 2009 were likely to also be highly innovative in 1999; furthermore, cities may exhibit persistent differences and heterogeneity in terms of inventors' productivity and their ability to forge new ties (Lee, 2010).[18] In an attempt to check the robustness of our OLS es-timates to this issue, we adopted an instrumental variable ap-proach using amenities as instruments for the suspect endogenous

---

[15] The unreported variables include: university R&D per student, college educated per worker, income per capita, population density, the Herfindahl index on 35 IPC classes, the seven RTA indexes, the average age of prior art, the share of employ-ment in manufacturing, the share of employment in professional and business ser-vices, the Herfindahl index at the firm level, the number of establishments per worker, the index of enforcement of non-compete laws, and dummy variables for BEA regions.

[16] Although internal social proximity presents a negative (not significant) sign, its marginal effect becomes positive when clique density is equal to 0.07 (= 0.013/0.190), well below the sample average value.

[17] Several authors have highlighted the negative effects of cliquish network struc-tures (Uzzi, 1996). Cliquishness can cause isolation and reduce exposure to alterna-

tive ideas, limiting the absorption and recombination of knowledge sourced from external links and thus reducing creative drive.

[18] Because of their superior capabilities, more productive inventors are expected to form more, new, non-redundant ties. More prolific inventors are also more vis-ible to their peers and are generally considered better partners, thereby becoming more attractive for new collaborations. In short, inventors who are more productive are more likely to develop new and more ties both within and across cities, which affect both the structural properties of the co-invention network and increase the innovative performance of the city in which these individuals are located. At the city level of analysis, this process may eventually lead inventors to develop net-works in cities that are already highly productive (i.e., host more prolific inventors).

Table 3
The impact of co-invention network structure on metropolitan inventive productivity[a].

Dependent variable: Log of patents per worker (2009)	Model 1 OLS	Model 2 OLS	Model 3 IV-GMM[f]	Model 4 IV-GMM[f]
Inventor density (1995–1999)[b] (log)	0.151*** (0.036)	0.148*** (0.033)	0.074*** (0.015)	0.077*** (0.016)
Internal social proximity (1995–1999)[c] (log)	−0.013 (0.017)	−0.060** (0.026)	0.003 (0.009)	−0.013 (0.012)
Clique density (1995–1999)[d]	−0.286** (0.135)		−0.063 (0.079)	
Internal social proximity × Clique density (1995–1999)	0.190* (0.105)		0.217*** (0.037)	
External social proximity (1995–1999)[e] (log)		−0.006 (0.008)		−0.001 (0.005)
Internal social proximity × External social proximity (1995–1999)[f]		0.013** (0.005)		0.008*** (0.002)
Hansen J overidentification test statistic[g]			$\chi^2 = 96.477$ $p = 0.712$	$\chi^2 = 108.022$ $p = 0.400$
C statistic (orthogonality of external excluded instruments)[g]			$\chi^2 = 7.063$ $p = 0.631$	$\chi^2 = 3.808$ $p = 0.924$
Kleibergen–Paap rk under identification test LM statistic[g]			$\chi^2 = 146.165$*** $p = 0.006$	$\chi^2 = 145.301$*** $p = 0.007$
Kleibergen–Paap rk weak instruments F statistic[g]			4.791	6.560
Endogeneity test[g]			$\chi^2 = 5.796$ $p = 0.215$	$\chi^2 = 9.898$ $p = 0.042$
Number of MSA observations	331	331	319[h]	319[h]
R-square	0.637	0.634	0.602	0.598
F[g]	12.847***	12.578***	39.704***	24.819***

Standard errors are in parentheses and robust to arbitrary heteroscedasticity.

*, **, *** Denote statistical significance at the 10, 5, and 1 % level. Each model includes additional control for agglomeration variables, namely: university R&D per student; college educated per worker; income per capita (log); population density (log); Herfindahl index on 35 IPC classes; revealed technological advantage index in seven broad technological areas (i.e., electronics, scientific instruments, chemicals, pharmaceuticals and biotechnology, industrial processes, mechanical engineering, and consumer goods); the average age of prior art (log); the share of employment in manufacturing; the share of employment in professional and business services; the Herfindahl index at the firm level; the number of establishments per worker (log); the index of enforcement of non-compete laws; dummy variables for BEA regions.

[a] Coefficients have been multiplied by 1000 to ease readability.
[b] Inventor density is the ratio between the total number of inventors active in a given MSA and the MSA land area measured in square miles.
[c] Equation [1] provides the formal definition of internal social proximity.
[d] Equation [2] provides the formal definition of clique density.
[e] Equation [3] provides the formal definition of external social proximity.
[f] Estimates are obtained by applying the two step IV GMM estimator and the ivreg2h routine available in STATA12 (Lewbel, 2012). IV GMM instruments in models 3 and 4 include three dummy variables for topography to account for the slope of land (plains with hills or mountains, open hills and mountains, hills and mountains) (1999); a dummy variable taking value 1 if the MSA is on the ocean (1999); surface water as share of total MSA's land (1999); inverse of minimum annual temperature (1999); the number of museums per inhabitant (1999), the number of property crimes per 10000 inhabitant (1999) and the share of urbanized land on total MSA's land (1999).
[g] Test statistic is robust to arbitrary heteroscedasticity.
[h] Due to some missing in the instruments, the number of MSA observations included in the analysis is 319.

regressors (i.e., inventor density and network structural character-istics). The importance of amenities for individual location choices has been richly described and documented in the literature since the seminal paper by Roback (1982), and differences in (consump-tion) amenities can cause differences in population density across space (Rappaport, 2008). Hence, if amenities positively affect inventor density (as argued by Carlino et al., 2007), by extension, amenities can also impact network structural characteristics.[19]

Following Albouy (2008, 2009) and Albouy and Stuart (2014), we accounted for both natural and artificial amenities as instruments for inventor density and network variables. As natural amenities, we considered the following variables: three dummy variables for topography to account for the slope of land (plains

with hills or mountains, open hills and mountains, hills and moun-tains); a dummy variable taking value 1 if the MSA is on the ocean; surface water as share of total area; and the inverse of min-imum annual temperature.[20] Several studies emphasize the impor-tance of geographic conditions as predictors of population den-sity (Burchfield et al., 2006; Rosenthal and Strange, 2008; Combes et al., 2009). Mountainous and coastal areas have been found to experience higher income and greater leisure amenities (Saiz, 2010) and higher quality of life (Albouy, 2008). Moreover, Partridge (2010) documents sizeable migrations toward more mountainous and coastal areas in the US, possibly because they offer more pleasant places to stay. Similarly, Rappaport (2007) illustrates the relevance of clement weather in directing population migration flows.

[19] Ultimately, an increase in the density of the population of inventors due to their attraction to specific locations could plausibly spur interaction among them (Brueckner and Largey, 2008). As noted by Glaeser et al. (2001), crowding makes encounters easier, meaning that density increases socialization (Liben-Nowell et al., 2005; Pan et al., 2013).

[20] Data on average temperatures were derived from http://wonder.cdc.gov/NASA-NLDAS.html and data on topographic characteristics from http://ers.usda.gov/data-products/natural-amenities-scale.aspx.

As artificial amenities, we included the following variables: the number of museums per inhabitant, the number of property crimes per 10,000 inhabitants, and the share of urbanized land in an MSA's total land (Carlino et al., 2007; Albouy 2008, 2009; Glaeser et al., 2001; Wu, 2010; Wang and Wu, 2011).[21] Artificial amenities such as leisure activities, safety, and the architectural beauty of local buildings, increase attractiveness, density, and in turn, the possibility to get in contact with an attractive mix of so-cial partners (Albouy, 2008; Glaeser et al., 2001).

In particular, we applied the estimation method proposed by Lewbel (2012), which serves to identify structural parameters in regression models affected by endogeneity by supplementing available external instruments through constructed ones obtained as simple functions of the model's data. Identification is achieved by having regressors that are uncorrelated with the product of heteroskedastic errors. The greater the degree of heteroskedasticity in the error process, the higher will be the correlation of the gener-ated instruments with the included endogenous variables, which are the regressands in the auxiliary ('first-stage') regressions. Es-timations have been obtained by applying the robust two-step GMM estimator as recommended by Baum et al. (2007) and Baum (2006), as for an overidentified equation, IV-GMM robust estimates will be more efficient than 2SLS estimates. First-stage results are available in Table D2 and D3 in Appendix D.

Regarding Models 3 and 4 in Table 3, which reports the main results of the second-stage regressions, our key findings on the interaction variables, i.e., internal social proximity × clique density and internal social proximity × external social proximity, seem to be supported, albeit with reduced magnitude in the case of the interaction between internal and external proximity. This conclu-sion is however subject to demonstrating the instruments' valid-ity. However, performing such a diagnostic is more complex with multiple endogenous variables, as each instrument plays a role in each first-stage regression, making identification more problematic (Baum et al., 2007).

Table 3 reports the usual battery of diagnostic tests for instru-ment validity (i.e., orthogonality and strength) at the bottom. Tests of overidentifying restrictions (the Hansen tests and the related J statistics), in both cases, do not allow us to reject the null that the excluded instruments are independent of the error process, al-though these tests are known to suffer from weak power (Baum et al., 2003). Moreover, the C statistic on the external excluded instru-ments confirms that the specified orthogonality conditions are sat-isfied. The underidentification tests strongly reject the null hypoth-esis, suggesting that the Lewbel's approach is satisfactory in this case. The assessment of instrument relevance is generally based on the comparison of the Kleinberg–Paap rk Wald statistic with the tabulated Stock and Yogo (2005) critical values. Unfortunately, this comparison cannot be implemented in the present case because the Stock and Yogo (2005) critical values have been tabulated for up to 100 excluded instruments and three endogenous regressors. In terms of the F statistic[22], the results in Tables D2 and D3 show F statistics greater than 10 with the exception of the inventor density variable. Finally, the endogeneity test on all suspect endogenous variables (i.e., inventor density and the network variables) does not reject the null of exogeneity in the case of equation [4a] in Model

3, whereas exogeneity is rejected in the case of equation [4b] in Model 4.[23]

Despite the warnings highlighted above and the possible esti-mation bias, the main results are nevertheless consistent with the OLS estimates. On balance, therefore, we can reasonably conclude that the results concerning our key hypotheses are confirmed and sufficiently robust, suggesting that any remaining endogeneity in our regressions is unlikely to explain our central findings.

6. Conclusions

Our results provide a contribution to the debate on agglomer-ation, networks, and inventive performance. First, our findings in-dicate that urban settings are key engines and incubators of new knowledge creation processes not only because they provide in-divisibility (i.e., agglomeration) benefits but also because they fa-cilitate synergies between individuals (i.e., social networks). While our results do not contradict prior findings that agglomeration and scale effects are important, we show that network structure is im-portant for inventive activities in cities. Second, we explicitly mod-eled the interaction between the internal structure of a metropoli-tan co-invention network and the embeddedness of a city's inven-tors in the global collaboration network. Whereas previous stud-ies have indicated the importance of external linkages, these have been either nearly neglected or poorly modeled in empirical anal-yses. Interestingly, our results show that external linkages are im-portant for sustaining higher rates of inventive productivity, es-pecially when combined with an internal network structure that facilitates faster knowledge diffusion and recombination. Third, in terms of methodology, we have also shown that social network analysis techniques are useful for grasping the micro-level founda-tion (i.e., the social interaction processes aimed at new knowledge creation) of meso-level outcomes (i.e., inventive productivity at the city level).

Finally, our results should be interpreted with some caution. Patent data capture only a subset of links relevant to knowledge exchange, and we observe only those links stemming from the sub-set of US patents applied for through the EPO, which nonetheless are likely to channel more valuable knowledge (see the discussion in Appendix A). In this respect, our results supply conservative es-timates of the effect of a city's internal network structure and its interplay with external connections.

A few final comments on the robustness of our estimates are also in order. We acknowledge that identification in a sample as small as ours is a critical issue and that the choice of valid instru-ments, especially of artificial amenities, can be subject to skepti-cism. Notwithstanding, diagnostic tests tend to support the IV es-timates obtained and the attempt to verify the robustness of our OLS estimates. On balance, therefore, we can conclude that the key findings of this paper are sufficiently sound and based on robust evidence.

---

[21] As discussed by Albouy (2008), even if artificial amenity variables tend to present a high degree of collinearity and could be questioned on the grounds of endogeneity, as recently documented by Rosenthal and Ross (2010) in the case of crime, oftentimes there is no other recourse, due to the unavailability of natural experiments and confounding factors in urban dynamics.

[22] According to Staiger and Stock (1997)'s 'rule of thumb', an F statistic greater than 10 excludes serious weak identification problems (meaning that weak instru-ment bias is small); yet, this criterion is safely applicable only to the cases of a single endogenous variable.

[23] To shed further light on the instrument relevance and exogeneity tests, we per-formed an additional robustness check (results not shown) and re-ran all the es-timates by excluding the inventor density variable, which is the one that is likely most affected by a possible weak instrument problem on the basis of the F statistic of the excluded instruments in the first stage. Diagnostic tests and estimates are substantially unchanged. Interestingly, having three endogenous regressors enables the comparison with the Stock and Yogo (2005) critical values. In both cases, the Kleinberg–Paap rk Wald statistic allows us to reject the null of weak instruments with a bias of 20% with respect to the OLS. Moreover, the joint test for endogeneity on all suspect endogenous variables does not allow us to reject the null of exogene-ity for both models. Even if we cannot definitively exclude the possibility that these results can be partly related to the weak power of endogeneity tests, and their sen-sitiveness to the treatment of errors and details of model specification, and even if we cannot claim with absolute certainty that our instruments are fully exogenous and strong, these results are quite encouraging in supporting our main findings.

Acknowledgements

Appendix A. Patent data

By providing the name of each inventor involved with an invention and each inventor's address, patent data enable us to assemble and analyze the full set of co-invention ties linking individuals and to distinguish (co-invention) ties linking inventors located in the same city from co-invention ties linking individuals located in different cities. In this paper, we restrict our attention to the ties formed between US inventors.

Although co-invention data may capture a subset of all relevant connections linking individuals within and across cities, these are neither accidental nor unchecked. The network of inventors is the most immediate and influential environment from which ideas and information can be drawn, at least for the technical contents of their patents (Fleming et al., 2007; Breschi and Lissoni, 2004).

USPTO patent data might appear a more natural choice for a study on the inventive productivity of US cities. However, the use of EPO data can be defended on methodological grounds. First, by using patent applications of US organizations at the EPO, we are able to drop from the analysis patents of low quality or with low commercial value, which would not be cost-effective to extend to Europe. Several scholars have argued that, precisely in the period of our investigation, the average quality of patents issued by the USPTO had declined due to a series of concomitant factors, lead-ing to a dramatic surge in the number of patents. As a conse-quence of these trends, patents of insufficient quality or with in-adequate search of prior art were issued more often (Hall et al., 2004; Jaffe and Lerner, 2004). Second, EPO patent records provide more information than USPTO records, such as the inventors' ad-dresses (street and zip code). This is crucial information for the correct identification of individual inventors and to avoid problems of homonymy. For example, the NBER database on USPTO patents provides information on the street address for only 11.5% of the ob-servations in the 'Inventors' file. Conversely, nearly all EPO patents report the inventors' street address, with a few exceptions due to some companies' policy to provide the address of their headquar-ters regardless of where an invention was produced. Third, tests comparing the distribution of patents per MSA per year applied for at the EPO with the distribution of patents per MSA per year ap-plied for at the USPTO reject that the two differ at any statistically significant level.

Appendix B. Methodological notes on the network construction

From a technical perspective, the co-invention network is the one-mode projection of an affiliation (or two-mode) network. As such, it is a valued network. In this paper, we binarize the adja-cency matrices, an approach that is consistent with most of the existing studies in the management and economic geography lit-erature and graph models. The co-invention network consists of all co-invention ties among inventors who received patents in the period 1995–1999. Specifically, we constructed a binary adjacency matrix, where the generic cell (i,j) is equal to one if inventor i cre-ated one or more patents with inventor j and 0 otherwise. The choice of a five-year period to construct the co-invention network is consistent with previous studies (Fleming et al., 2007; Lobo and Strumsky, 2008; Schilling and Phelps, 2007). Adopting different time windows did not substantively change the results.

Appendix C. Network variables

The index used to measure social proximity/distance is the so-called average social distance weighted reach. This index particu-larly suits our empirical setting because it can be computed for sparse and fragmented networks such as co-invention networks and captures both the size and connectivity of the network. Alternative indicators of the topological structure of the network—such as the average social distance or the density—suffer from im-portant shortcomings. First, the computation of these indexes re-quires inventors to be connected to obtain finite social proximity values between any two inventors. This is problematic in the case of co-invention networks, which are generally composed of several components and tend to be highly fragmented (i.e., co-invention networks are generally composed of relatively small sub-groups of nodes called network components disconnected from one another). Moreover, they provide useful insights on how well connected actors within a city are and therefore on how smoothly knowledge flows but do not tell much about the scale of such effects. Finally, the so-called average of reciprocal path length discussed in Jackson (2008) could be used as an alternative measure. This indicator can also be computed for networks that are not fully connected, but it captures only a network's connectivity and not its size. In our view, both dimensions are important in explaining the inventive produc-tivity of a city. Given two networks with the same average social proximity among nodes, the amount of knowledge flowing within the city will be higher the larger the number of nodes in the net-work. We therefore preferred the so-called average social distance weighted reach to the average of reciprocal path length index.

Regarding external social proximity, the chosen indicator is an improvement relative to the measures used in previous studies, such as the mere number of external co-inventors to the city (Fleming et al., 2007; Lobo and Strumsky, 2008) or a centrality index (Giuliani and Bell, 2005), which can be computed only for connected networks (see the discussion above).

Appendix D. Additional empirical results

See Tables D1, D2 and D3.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jue.2015.12.003.

References

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

Agrawal, A., Cockburn, I., Rosell, C., 2010. Not Invented Here? Innovation in company towns. Journal of Urban Economics 67, 78–89.

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

Albouy, D., 2008. Are big cities bad places to live? Estimating quality of life across metropolitan areas NBER Working Paper No. 14472.

Albouy, D., 2009. What are cities worth? Land rents, local productivity, and the cap-italization of amenity values NBER Working Paper No. 14981.

Albouy, D., Stuart, B., 2014. Urban population and amenities NBER Working Paper No. 19919.

Baum, C., 2006. An Introduction to Modern Econometrics using Stata. StataCorp LP, College Station, Texas.

Baum, C., Schaffer, M., Stillman, S., 2003. Instrumental variables and GMM: Estima-tion and testing. Stata Journal 3, 1–31.

Baum, C., Schaffer, M., Stillman, S., 2007. Enhanced routines for instrumental vari-ables/generalized method of moments estimation and testing. Stata Journal 7, 465–506.

Beaudry, C., Schiffauerova, A., 2009. Who's right, Marshall or Jacobs? The localiza-tion versus urbanization debate. Research Policy 38, 318–337.

Bell, G.G., Zaheer, A., 2007. Geography, networks, and knowledge flow. Organic Sci-ence 18, 955–972.

Bettencourt, L.M.A.A., Lobo, J.J.J., Strumsky, D., 2007. Invention in the city: Increasing returns to patenting as a scaling function of metropolitan size. Research Policy 36, 107–120.

Breschi, S., Lissoni, F., 2004. Knowledge networks from patent data: Methodological issues and research targets. In: Glanzel, W., Moed, H., Schmoch, U. (Eds.), Handbook of Quantitative S&T Research: The Use of Publication and Patent Statistics in Studies of S&T Systems. Springer Verlag, Berlin, pp. 613–643.

Bresnahan, T., Gambardella, A., Saxenian, A., 2001. Old economy" inputs for "new economy" outcomes: Cluster formation in the New Silicon Valleys. Industrial Corporation Chang. 10, 835–860.

Brueckner, J.K., Largey, A.G., 2008. Social interaction and urban sprawl. Journal of Urban Economics 64, 18–34.

Burchfield, M., Overman, H.G., Puga, D., Turner, M.A., 2006. Causes of sprawl: A portrait from space, Q. Journal of Economics 121 (2), 587–633.

Burt, R.S., 2004. Structural holes and good ideas. American Journal of Sociology 110, 349–399.

Carlino, G.A., Chatterjee, S., Hunt, R.M., 2007. Urban density and the rate of invention. Journal of Urban Economics 61, 389–419.

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

Combes, P.P., Duranton, G., Gobillon, L., Roux, S., 2009. Estimating agglomeration economies with history, geology, and worker Effects. In: Glaeser, E.L. (Ed.), Agglomeration Economies. The University of Chicago Press, Chicago, pp. 15–66.

Duranton, G., Puga, D., 2004. Micro-foundations of urban agglomeration economies. In: Henderson, V., Thisse, J.F. (Eds.), Handbook of Regional and Urban Eco-nomics. Elsevier, Amsterdam, pp. 2063–2117.

Feller, I., 1971. The urban location of United States invention, 1860-1910. Explorations in Economic History 8, 285–303.

Fleming, L., King, C., Juda, A.I., 2007. Small worlds and regional innovation. Organic Science 18, 938–954.

Garmaise, M.J., 2011. Ties that truly bind: noncompetition agreements, executive compensation, and firm investment. Journal of Law, Economics and Organiza-tion 27, 376–425.

Gilson, R.J., 1999. The legal infrastructure of high technology industrial districts: Sil-icon Valley, Route 128, and covenants not to compete. New York University Law Review 74, 575–629.

Gittelman, M., 2007. Does geography matter for science-based firms? Epistemic communities and the geography of research and patenting in biotechnology. Organic Science 18, 724–741.

Giuliani, E., Bell, M., 2005. The micro-determinants of meso-level learning and innovation: evidence from a Chilean wine cluster. Research Policy 34, 47–68.

Glaeser, E.L., 1999. Learning in cities. Journal of Urban Economics 46, 254–277.

Glaeser, E.L., Gottlieb, J.D., 2009. The wealth of cities: Agglomeration economies and spatial equilibrium in the United States. Journal of Economic Literature 47, 983–1028.

Glaeser, E.L., Kallal, H.D., Scheinkman, J.A., Shleifer, A., 1992. Growth in cities. Journal of Political Economics 100, 1126–1152.

Glaeser, E.L., Kolko, J., Saiz, A., 2001. Consumer city. Journal of Economic Geography 1, 27–50.

Hall, B.H., Graham, S., Harhoff, D., Mowery, D.C., 2004. Prospects for improving US patent quality via postgrant opposition. Innovation Policy and the Economy 4, 115–143.

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

Jaccard, P.J.J., Turrisi, R., 2003. Interaction Effects in Multiple Regression, 2nd ed. Sage Publications, Inc, Thousand Oaks, CA.

Jackson, M.O., 2008. Social and Economic Networks. Princeton University Press.

Jaffe, A.B., Lerner, J., 2004. Innovation and Its Discontents: How Our Broken Patent System Is Endangering Innovation And Progress, And What To Do About It. Princeton University Press, Princeton, NJ.

Jaffe, A.B., Trajtenberg, M., Henderson, R., 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. Quarterly Journal of Economics 108, 577–598.

Kerr, W.R., 2010. Breakthrough inventions and migrating clusters of innovation. Journal of Urban Economics. 67, 46–60.

Lee, J., 2010. Heterogeneity, brokerage, and innovative performance: Endogenous formation of collaborative inventor networks. Organic Science 21, 804–822.

Lewbel, A., 2012. Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. Journal of Business Economic Statistics. 30, 67–80.

Liben-Nowell, D., Novak, J., Kumar, R., Raghavan, P., Tomkins, A., 2005. Geographic routing in social networks. Proceedings of National Academy of Sciences U. S. A. 102, 11623–11628.

Lissoni, F., Sanditov, B., Tarasconi, G., 2006. The Keins database on academic inventors: Methodology and contents. Cespri Work Paper.

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

Malsberger, B.M., 2004. Covenants Not To Compete: A State-by-state Survey, Bna Books, Washington, DC.

Neal, Z.P., 2010. From central places to network bases: A transition in the U.S. urban hierarchy, 1900-2000. City Community 10, 49–74.

Newman, M., 2010. Networks: An Introduction. Oxford University Press, USA.

Newman, M.E.J., 2001. The structure of scientific collaboration networks. Proceedings of National Academy of Science USA. 98, 404–409.

Opsahl, T., 2010. Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. arXiv:1006.0887v2.

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

Pan, W., Ghoshal, G., Krumme, C., Cebrian, M., Pentland, A., 2013. Urban characteristics attributable to density-driven tie formation. National Community 4. Partridge, M.D., 2010. The duelling models: NEG vs amenity migration in explaining US engines of growth. Pap. Reg. Sci. 89, 513–536.

Patacchini, E., Venanzoni, G., 2014. Peer effects in the demandfor housing quality. Journal of Urban Economy 83, 6–17.

Puga, D., 2010. The magnitude and causes of agglomeration economies. Journal of Regional Science 50, 203–219.

Rappaport, J., 2007. Moving to nice weather. Regional Science of Urban Economy 37, 375–398.

Rappaport, J., 2008. Consumption amenities and city population density. Regional Science of Urban Economy 38, 533–552.

Roback, J., 1982. Wages, rents, and the quality of life. Journal of Political Economy 90, 1257–1278.

Rosenthal, S.S., Ross, A., 2010. Violent crime, entrepreneurship, and cities. Journal of Urban Economy 67, 135–149.

Rosenthal, S.S., Strange, W.C., 2001. The determinants of agglomeration. Journal of Urban Economy 50, 191–229.

Rosenthal, S.S., Strange, W.C., 2003. Geography, industrial organization, and agglomeration. Review of Economic Statistics 85, 377–393.

Rosenthal, S.S., Strange, W.C., 2004. Evidence on the nature and sources of agglomeration economies. In: Henderson, V.J., Thisse, J.F. (Eds.), Handbook of Regional and Urban Economics. Elsevier, Amsterdam, pp. 2119–2171.

Rosenthal, S.S., Strange, W.C., 2008. The attenuation of human capital spillovers. Journal of Urban Economy 64, 373–389.

Saiz, A., 2010. The geographic determinants of housing supply. Quarterly Journal Economics 125, 1253–1296.

Schilling, M.A., Phelps, C.C., 2007. Interfirm collaboration networks: The impact of large-scale network structure on firm innovation. Management Science 53, 1113–1126.

Singh, J., 2005. Collaborative networks as determinants of knowledge diffusion patterns. Management Science 51, 756–770.

Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. Econometrica 65, 557–586.

Stock, J., Yogo, M., 2005. Testing for weak instruments in linear IV regression. In: Andrews, D.W.K. (Ed.), Identification and Inference for Econometric Models. Cambridge University Press, New York, pp. 80–108.

Ter Wal, A., Boschma, R., 2009. Applying social network analysis in economic geography: framing some key analytic issues. Annual Regional Science 43, 739–756.

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

Uzzi, B., Spiro, J., 2005. Collaboration and creativity: The small world problem. American Journal of Sociology 111, 447–504.

Wang, C., Wu, J., 2011. Natural amenities, increasing returns, and urban development. Journal of Economic Geography. 11 (4), 687–707.

Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of small-world networks. Nature 393, 440–442.

Wu, J., 2010. Economic fundamentals and urban-suburban disparities. Journal of Regional Science 50, 570–591.