

# The role of external linkages and gatekeepers for the renewal and expansion of U.S. cities' knowledge base, 1990-2004\*

Stefano Breschi<sup>§</sup> and Camilla Lenzi<sup>#</sup>

<sup>§</sup> Bocconi University,  
Department of Management and Technology and CRIOS  
&

<sup>#</sup> Politecnico di Milano  
Department of Architecture, Built Environment and Construction Engineering

## Abstract

This paper examines the role of external linkages and gatekeepers for the renewal and expansion of cities' knowledge base, by presenting new evidence about co-invention networks in U.S. metropolitan areas based on European Patent Office (EPO) data for the period 1990-2004. We argue that the relative importance of direct external linkages and external relations mediated by gatekeepers varies according to specific local conditions. In particular, our findings suggest that external relations are on average a chief conduit to inject non-redundant knowledge at the local level and contribute to broadening and rejuvenating the local knowledge base. However, cities are quite heterogeneous in how they benefit from external relations. Whereas *direct* external connections outperform, on average, external links *mediated* by gatekeepers, the latter are especially important in cities with a localized and specialized knowledge base, as they enable the trans-coding and absorption at the local level of externally sourced knowledge.

JEL codes: O31, R11

Keywords: networks, gatekeepers, knowledge base

## 1. INTRODUCTION

The role of social networks for the creation and the spatial diffusion of scientific and technological knowledge has received a great deal of attention in the last decade. A rather wide consensus has emerged in the literature about the importance of social proximity within well-defined knowledge communities *vis à vis* spatial proximity to explain knowledge flows and their spatial reach (BOSCHMA, 2005; BRESCHI and LISSONI, 2001; CAPELLO, 2009). In addition, the debate on the most suitable network architecture to stimulate regional innovation has revealed the paramount importance not only of internal structural properties, but also of the embeddedness of local actors in broader networks (FLEMING et al., 2007a; LOBO and STRUMSKY, 2008). Dense interactions at the local level (i.e. local buzz) combined with embeddedness in global knowledge networks (i.e. global pipelines) allow to exploit the advantages deriving from trusted and repeated ties at the local level with non-redundant information deriving from external sources (BATHELT et al., 2004; MORRISON et al., 2013; STORPER and VENABLES, 2004), thereby leading to a superior innovative performance.

Building upon these insights, the most recent literature has thus focused on the roles played and the positions held by specific actors in local networks, by investigating their characteristics, attributes and performances. In this respect, *gatekeepers* have attracted much attention, especially in studies on industrial clusters, as they perform a crucial interfacing function between the local and the external knowledge systems, such as screening external sources, accessing them, and conveying new knowledge to local actors (GIULIANI and BELL, 2005; GRAF, 2011; MUNARI et al., 2012). However, most of the existing studies are small-scale investigations on single clusters, providing thus only limited evidence on the actual contribution of these actors to the overall innovative performance of the local contexts in which they are embedded.

This paper proposes an assessment of the importance of external relations and gatekeepers for inventive performance at the local level, by introducing two important novelties with respect to the extant literature. At the conceptual level, the paper argues that external linkages, albeit important, are not all alike; rather, their value and

contribution to broadening and rejuvenating the local knowledge base can vary considerably according to specific local conditions. In particular, the paper distinguishes between direct external ties and external linkages mediated by gatekeepers and argues that gatekeepers play a more fundamental role in contexts characterized by a specific and localized knowledge base because of their ability to access, convey, trans-code external knowledge and to make it accessible and meaningful to local actors. At the methodological level, the paper examines the impact of external linkages and gatekeepers by departing from the ego-network perspective, prevailing in the existing literature, which focuses on the attributes of single actors in the network and their performance. Rather, the paper aims at assessing the importance of connections to external sources of knowledge and the specific role of gatekeepers in a large-scale framework, by examining their impact on the inventiveness of the cities in which they are located. In other words, the paper analyzes the meso-level outcomes, in terms of innovativeness, of micro-level knowledge exchanges among individuals. In so doing, the paper also provides a methodological contribution by explicitly modeling the structure of external relations and, accordingly, the channels through which external knowledge flows into a city.

We test our expectations on a large dataset of patents and their inventors in 196 U.S. Metropolitan Statistical Areas (MSAs)<sup>2</sup> in the period 1990-2004. The focus on urban settings particularly suits the study of the link between knowledge network structure and innovation, as invention in the U.S. has ever been and still is a predominantly metropolitan phenomenon (CARLINO et al., 2007; FELLER, 1971; LAMOREAUX and SOKOLOFF, 2000; PRED, 1973).

The next section reviews the literature and elaborates the hypotheses to be tested. Then, the paper discusses the methodological issues related to the construction of appropriate indicators to capture the intensity of a city's external linkages and the identification of gatekeepers, and it describes the data sources and the econometric framework. The results of the empirical analysis are then presented and, lastly, the final section offers some concluding remarks and outlines future research directions.

<sup>2</sup> Metropolitan Statistical Areas (MSA) are defined by the U.S. 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.

## **2. THE IMPORTANCE OF EXTERNAL LINKAGES AND GATEKEEPERS FOR THE RENEWAL AND EXPANSION OF CITIES' KNOWLEDGE BASE**

The importance of external linkages as sources of innovation is a fertile research topic in management, innovation and regional studies. Broadly speaking, the recombination of diverse pieces of knowledge is at the heart of any knowledge creation and dynamics process (ANTONELLI, 2008; FLEMING, 2001; FLEMING and SORENSON, 2001; SAVIOTTI, 2007). Given the increasing complexity of technologies and knowledge, the creation of novel solutions relies more and more upon the integration and recombination of different knowledge sources (BRESNAHAN et al., 2001; GITTELMAN, 2007; OWEN-SMITH and POWELL, 2004). Openness to external repositories of knowledge and embeddedness in knowledge networks reaching actors beyond local boundaries are therefore crucial and can provide some shelter to the risk of over-embeddedness, lock-in to obsolete sets of technologies and decrease in the variety of technological approaches and solutions, by favoring a continuous expansion and rejuvenation of the existing knowledge base (UZZI, 1996, 1997).

At the regional level, the importance of connecting to spatially dispersed sources of knowledge has been emphasized in several studies. In particular, it has been pointed out that, given the highly localized nature of knowledge, knowledge exchanges among the same set of (co-located) actors may lose value over time as information becomes redundant and opportunities for re-combination may fade out (BOSCHMA and FRENKEN, 2010). A disproportionate inward orientation in knowledge exchanges may crystallize the existing knowledge base and reduce technological heterogeneity and the potential for knowledge exploration and re-combination, thus leading to decreases in creativity and possible losses of positions in the spatial ranking (BATHELT et al., 2004; GRAF, 2011; STORPER and VENABLES, 2004).

In this respect, the recent literature suggests that *gatekeepers* play a fundamental role as mediators of the knowledge exchanges between external and local actors (GIULIANI and BELL, 2005; MORRISON, 2008; MORRISON et al., 2013; MUNARI et al., 2012). Originally proposed by ALLEN and COHEN (1969) and ALLEN (1977), gatekeepers (also, and frequently interchangeably, termed boundary spanners) are individuals, which guarantee access to external sources of knowledge by establishing unique linkages with

outside actors and, more importantly, ensure knowledge absorption and diffusion within their proximate working and social environments. In the definition put forward by GOULD and FERNANDEZ (1989, p. 92), gatekeeping is a specific form of brokerage that corresponds to structural position in transaction networks in which “an actor can selectively grants outsiders access to members of his or her own group”. Likewise, in regional studies, gatekeepers are generally intended as individuals (and sometimes organizations) that enable knowledge transfer among different spatial units, i.e. industrial clusters, cities or regions (GIULIANI and BELL, 2005; GRAF, 2011; MORRISON et al., 2013).

Because of their key intermediation role, gatekeepers may be crucial to overcome any communication impedance that may occur in cross-boundary knowledge exchanges, regardless of how boundaries are defined (namely, in geographical, organizational, or institutional terms). Knowledge and technologies, in fact, especially those related to specific artifacts, products and manufacturing processes tend to evolve in a cumulative, path-dependent and highly localized way, within defined technological, organizational and geographical spaces (ANTONELLI, 1995; BOSCHMA and FRENKEN, 2006; DOSI, 1982; FRENKEN and BOSCHMA, 2007). Specific coding schemes, standard concepts and technical jargons develop within and across firms located in a regional cluster (LISSONI, 2001). These *codebooks*, which are highly specific to the epistemic communities of engineers and technicians working in a given cluster, are instrumental to facilitate internal communication and information processing, but may also impose barriers to the exchange of knowledge across regional boundaries (COWAN et al., 2000). In other words, a highly localized and specialized knowledge base provides efficiency gains in internal knowledge exchanges within firms as well as within regions and clusters, whilst it may hinder cross-boundary exchanges due to mismatches in interpretation, cognitive mapping or, more simply, misunderstanding of contents (TUSHMAN and KATZ, 1980; TUSHMAN and SCANLAN, 1981). In this regard, gatekeepers may not only search for and collect relevant information outside their proximate professional and social environments, but they are also able to trans-code this information and to diffuse it within their organizations and geographical areas. In other terms, integrating the external knowledge into the local knowledge base requires the absorptive capacity to read the diverse external knowledge and to create mental maps

and codes to translate it at the local level. Gatekeepers are those best positioned to dispatch the externally sourced knowledge at the local level by framing, coding and separating it into pieces that can be easily understood, processed and used by others thanks to shared language and symbols. Differently, absent the intermediary role played by gatekeepers, individuals with ‘the same specialized language and symbolic representations [...] will not be able to tap into diverse external knowledge sources even if the relevant pipelines are in place’ (BATHELT et al., 2004, p.45).

In order to perform this fundamental function, gatekeepers have to possess a number of key attributes, which distinguish them from other individuals. In general, they are more productive (BURT, 1992), more creative and capable to develop alternative solutions and original visions by framing problems from a different and fresh perspective (BURT, 2004; FLEMING et al., 2007b; HARGADON and SUTTON, 1997; OBSTFELD, 2005). Moreover, they also tend to occupy a highly central and influential position in the social structures in which they are embedded (FERNANDEZ and GOULD, 1994; PADGETT and ANSELL, 1993).

From the individual perspective, the uniqueness and importance of such gatekeeping positions do provide gatekeepers with competitive advantages with respect to other network members (BURT, 2008), which in turn can lead to higher (private) returns in terms of economic and innovative outcomes. However, the competitive advantages and superior performance of gatekeepers parallel their control power (and the relative rents they may accrue from) on the bridging ties and knowledge exchanges that they enable between internal and external actors (BURT, 2008; GOULD and FERNANDEZ, 1989). If social ties are channels for the exchange of knowledge or other resources, gatekeepers are therefore in the position to choose whether to grant access or not to information flows from outside. Whereas this power to restrict knowledge exchanges may have negligible effects on a single gatekeeper’s performance (and possibly it may provide a sufficient basis for the generation of rents), it can have substantial impact when moving from the individual to the aggregate level of analysis, such as the firm or the regional one.

In fact, gatekeepers may not always be willing to share the valuable knowledge they source from outside with other actors nearby and frequently prefer to act as external

stars rather than as true gatekeepers. Given their control power on external knowledge sources, gatekeepers can restrict strategically the diffusion and circulation of valuable knowledge. This is, for example, documented in the case of leading firms in industrial clusters that frequently play a significant role in facilitating the flow of outside information into the cluster (GIULIANI and BELL, 2005; MORRISON, 2008; MORRISON et al., 2013).

Moreover, gatekeepers may have trouble in managing and matching multiple external and internal connections due to increasing coordination costs, thus reducing the amount and efficiency of information flows (WHITTINGTON et al., 2009). Similarly, the larger the number of external sources managed by each gatekeeper, the more complex becomes to process, code, interpret and absorb information; too much information can be harmful by decreasing the efficiency of knowledge exchanges (DAHLANDER and FREDERIKSEN, 2012). More generally, social structures where external linkages (and the associated knowledge flows) are predominantly mediated by gatekeepers, are more exposed to the disruption of such links than social structures in which external ties are mostly direct, and possibly redundant. In addition to this, direct links could be generally preferable than indirect links as they allow faster knowledge diffusion (i.e. in a shorter number of steps), possibly with less knowledge leakages and risks of distortion of the message contents because of a lower number of intermediaries (TUSHMAN and SCANLAN, 1981).

In brief, though external ties look crucial for the renewal and expansion of the local knowledge base and, more generally, for innovation and creativity, the importance of gatekeepers for mediating the access to external sources of knowledge cannot be considered uncontroversial at an aggregate level of analysis such as the regional or urban one. On the one hand, direct external links may outperform gatekeepers-mediated links as they allow a faster and less noisy access to external knowledge. On the other hand, gatekeepers are more open-minded, creative thinker, innovative and influential than other actors and, even more importantly, they perform the translation and trans-coding of external information, which may be necessary for successfully transfer and apply the externally sourced knowledge at the local level. Thus, under what conditions, are gatekeepers a more effective way to link local actors to external sources of knowledge?

As argued by TUSHMAN and KATZ (1980), the answer to this question depends on the nature of the tasks and activities performed and, accordingly, the characteristics of the knowledge base. In fact, similar problems may be approached with very different solutions, also by firms operating in the same industry, as a result of the evolution of the different knowledge bases and coding schemes (TUSHMAN and KATZ, 1980; TUSHMAN and SCANLAN, 1981). Therefore, communication impedance and barriers in knowledge exchanges and learning are likely to be more frequent the greater the insularity, localism and dissimilarity between the knowledge base of internal firms and external counterparts (LISSONI, 2001; COWAN et al., 2000). Especially in these contexts, gatekeepers can improve knowledge transfer by translating the external knowledge into meaningful and useful knowledge for local actors (GRAF, 2011). On the other hand, knowledge flows are less likely to find major impedance when trading partners share similar cognitive maps, notions and languages; in these case, the trans-coding role of gatekeepers is less valuable and single actors can rely on their own direct contacts to access external knowledge.

Therefore, although external ties are, on average, important to sustain the expansion and renewal of the existing knowledge base, they are not all alike but rather they may have different impact according to specific context conditions, namely the nature of the local knowledge base. By extending the argument by TUSHMAN and KATZ (1980) into a spatial perspective, the role of gatekeepers is likely to be more important in regions and cities with a more *localized* and, thus, specialized knowledge base as gatekeepers can ensure the translation and circulation of external knowledge at the local level. Differently, regions and cities with a more *general* and diversified knowledge base are more likely to possess the coding schemes and technical languages needed to interpret external knowledge, having developed approaches and solutions more similar to externally located actors. In these cases, the trans-coding function of gatekeepers becomes less valuable and direct connections can be more efficient in accessing external knowledge in faster and a less noisy fashion.

Accordingly, we posit that:

*The intensity of external linkages is positively associated to the renewal and expansion of the local knowledge base in a city (Hypothesis 1)*



*External linkages mediated by gatekeepers are relatively more important in cities with a localized and specialized knowledge base, whereas external direct connections play a more significant role in cities with a general and diversified knowledge base. (Hypothesis 2).*

### **3. MEASURING EXTERNAL RELATIONS AND IDENTIFYING GATEKEEPERS**

The use of patents as relational data has made amenable to study the impact of social networks and personal interactions on innovation and creative processes through the tools of social network analysis and graph theory (BRESCHI and LISSONI, 2004, 2009; SINGH, 2005; TER WAL and BOSCHMA, 2009). 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, i.e. any two inventors are connected if they are designated together as inventors in one or more patent documents.

For this study, we extracted data on all patent applications made by U.S. organizations at the European Patent Office (EPO)<sup>3</sup> in 1990-2004, using the CRIOS-PATSTAT database. As the proper application of social network analysis techniques depends on the correct identification of individual inventors (i.e., nodes), we thoroughly cleaned and standardized their names and addresses.<sup>4</sup> We used the reported addresses to assign each inventor to one of the 370 U.S. Metropolitan Statistical Areas (MSAs), using the definition files available on the U.S. Census Bureau website.<sup>5</sup>

The basic assumption behind the use of patents as relational data is that co-invention ties work as pipes through which knowledge is transmitted. As the effectiveness with which a pipe performs this function is likely to decay with its age, we followed the current practice and adopted a five-year moving window to construct the co-invention

<sup>3</sup> The USPTO patent data might appear a more natural choice for a study on invention in U.S. cities. However, the average quality of USPTO patents has declined considerably in the period under consideration due to a series of concomitant factors. Because of these trends, patents of insufficient quality or with inadequate search of prior-art were issued more often (Hall et al. 2004, Jaffe and Lerner 2004). The use of EPO data reduces this noise as only more valuable patents are extended to Europe. Patents are dated by using their priority year, i.e., the first date at which the patent was applied for anywhere in the world as this date is closest to the actual time of the invention.

<sup>4</sup> Fuller details on the routine implemented for data cleaning and standardization are provided in Lissoni et al. (2006).

<sup>5</sup> We used the June 2003 definition of MSAs issued by OMB. Details and maps are available at: <http://www.census.gov/population/www/metroareas/metrodef.html>.

network (FLEMING et al., 2007a; LOBO and STRUMSKY, 2008; SCHILLING and PHELPS, 2007). In other words, the network observed at any time  $t$  is built using the co-invention ties formed during the time period  $[t-1, t-5]$  and excluding older ties. Adopting different time windows did not substantively change the results.

To measure the extent to which metropolitan inventors are connected with inventors external to the city, we computed the *average distance-weighted external reach* between inventors located in a given city and all other inventors located in all other cities. This index captures the proximity between inventors in the global network. In particular, for an individual inventor  $i$ , it is defined as the sum of the reciprocal distances to all other inventors she can reach in the global U.S. co-invention network.<sup>6</sup> Accordingly, the average distance-weighted external reach of city  $c$  is the distance-weighted external reach averaged across all inventors located in the city.

Formally, this index is defined as follows:

$$(A) \quad h) \quad \frac{\sum_{i=1}^{n_c} \sum_{j \in h} \frac{1}{d_{ij}}}{\sum_{i=1}^{n_c} \sum_{j \in h} 1}$$

where  $n_c$  denotes the number of inventors located in city  $c$  and  $n_h$  denotes the number of inventors located in other cities (i.e., not located in city  $c$ ), and  $d_{ij}$  denotes the geodesic distance (i.e., shortest path) in the global co-invention network between inventor  $i$  (located in city  $c$ ) and inventor  $j$  (not located in city  $c$ )<sup>7</sup>.

The index takes a minimum value of zero (i.e., all inventors in city  $c$  are not connected to any external inventor). In the (theoretical) case in which every inventor in city  $c$  directly collaborates with every other inventor in every other city (i.e. when every node

<sup>6</sup> It is worth remarking that the adjectives “internal” and “external” used to characterize the network ties only refer to the spatial boundaries delimiting a city. Thus, a co-invention tie between two inventors located within the same city is *internal* to that city, whereas a co-invention tie between two inventors located in different cities is *external* to them, *regardless of the organizational affiliation of the two inventors*. In particular, in the case of firms with plants located in different cities, the external reach indicator considers also those links that may arise between inventor  $i$  working in city  $c$  for applicant  $a$  and all inventors working for applicant  $a$  in every other U.S. city. Although one may question that such types of links are more likely to be established (and possibly mediated by firm-level practices and strategies), from a spatial perspective, they do not differ from the links established by inventor  $i$  working in city  $c$  for applicant  $a$  with all the other inventors in any other U.S. city and working for other firms than  $a$ .

<sup>7</sup> For disconnected (i.e. not reachable) pairs of inventors  $d_{ij} = \infty$  and therefore  $1/d_{ij}$  is equal to  $1/\infty$  which reduces to zero.

is connected by a path length one to all other nodes in the network), the numerator of the index takes value  $(n_c n_h)$  and thus the theoretical maximum value of the index is  $n_h$ . Higher values of the external reach index imply that a city has faster access to a larger pool of external knowledge and resources.

This index shows several desirable features for modeling a city's external ties with respect to the measures used in previous studies, such as the number of co-inventors external to the city (FLEMING et al., 2007a; LOBO and STRUMSKY, 2008) or centrality indexes (GIULIANI and BELL, 2005). First, it captures both a network's size and connectivity simultaneously, as both aspects matter in explaining knowledge flows across cities and their impact on a city's creative potential. Higher values of the external reach index imply that a city's inventors are socially close to many inventors located in other cities (i.e., they are separated by a low number of intermediaries) and thus have *faster* access to a *larger* pool of external knowledge and resources, with less noise. Second, it enables us to consider not only first-order ties (i.e., external inventors directly connected to a city's inventors because of joint patents), but also second-order and higher-order co-invention links, and thus the overall scope of a city's external ties. The amount of external knowledge that a city can reach is likely to depend not only on the number of external direct ties but also on the number of external indirect ties, though the value and impact of the knowledge received is also likely to decay with the social distance between sender and receiver (AHUJA, 2000). Third, contrary to other studies that adopt an ego-network approach and focus on the performance of individual actors, we adopt a meso-level perspective focusing on the overall network structural properties and their impact upon regional performance.

Following the discussion above, we expect that external reach has a positive impact on the expansion and renewal of a city's knowledge base. Yet, given that the importance of external ties is likely to decay quite rapidly with the *social* distance among nodes, the external reach index has been computed by considering only inventor pairs for which

4 (BRESCHI and LISSONI, 2004; SINGH, 2005)<sup>8</sup>. Fuller details on the computation of this index are available in the Appendix to the paper.

<sup>8</sup> Precisely, for pairs of inventors  $(ij)$  where  $d_{ij} > 4$ , has been set equal to zero. Namely, we assumed that no knowledge flow is taking place for distance above that threshold.

It is important to point out that external reach comprises different types of relationships between local and external inventors. In particular, external linkages can be either *direct*, whenever a local inventor is directly connected to an external inventor, or *indirect*, whenever a local inventor needs the intermediation of another co-localized inventor to reach external inventors.

In this second case, external reach is mediated by specific actors, i.e. gatekeepers, that occupy distinctive positions in the U.S. global co-invention network, by establishing bridges across cities. In this paper, the identification of gatekeepers is based on GOULD and FERNANDEZ (1989) that elaborate on the original definition proposed by ALLEN (1977) and TUSHMAN and KATZ (1980). Accordingly, an inventor  $i$  located in city  $c$  is defined as a gatekeeper when the shortest path leading from any other inventor  $j$  in city  $c$  to any other inventor  $h$  in city  $d$  (different from  $c$ ) passes through  $i$ .

In order to account for the importance of gatekeepers in mediating external linkages, we computed the share of the overall distance weighted external reach, which is mediated by gatekeepers. This corresponds to how much the external reachability would decrease for local inventors, should one remove all gatekeepers from the city (BORGATTI, 2006; VALENTE and FUJIMOTO, 2010). Higher values of the index corresponds to cities in which external linkages mostly rely upon gatekeepers. On the other hand, lower values of the index imply that most of the linkages with externally located inventors are direct and do not need any intermediation (i.e. removing the gatekeepers would not diminish substantially the external reachability).

The index shows some attractive features for capturing the importance of gatekeepers in mediating knowledge flows across cities. First, it refines and improves on centrality-based indexes or ratios between internal and external linkages used in previous spatial analyses (GIULIANI and BELL, 2005; GRAF, 2011; MORRISON, 2008). In fact, these measures either do not define gatekeepers as the necessary and single bridge able to link actors in two different spatial units or neglect the inflation bias that may arise in co-invention networks due to inventors listed in the same patent document or both. More interestingly, it provides a measure of the relative contribution of direct vs. indirect (i.e. gatekeepers-mediated) relations to the total external reach of a city and how much its

external relations are robust to the removal of the links established by gatekeepers (i.e. how much gatekeepers control the flows of knowledge across cities).

Following the discussion above, we expect that external reach mediated by gatekeepers has a positive effect on the expansion and renewal of the knowledge base in cities with a localized and specialized knowledge base.

#### **4. DATA AND ESTIMATION FRAMEWORK**

##### *4.1. The dependent variable: definition and measurement*

To account for the renewal and expansion of a city's knowledge base, we exploit information on the technological classification of patent documents. In fact, patents can be described as a bundle of different technologies, defined by the technological codes in which each patent is classified; accordingly a city's knowledge base can be decomposed and portrayed through the technology fields, and their possible combinations, associated to the locally produced patents (STRUMSKY et al., 2012).

Following the conceptual discussion, we define our dependent variable as the number of *new pairs of technological codes* introduced in a city at time  $t$ , in which either one or both technological codes are new to the city, i.e. no local patent before time  $t$  had been classified in one or both of those fields. It is worth stressing that, differently from other studies (e.g. FLEMING et al., 2007b), we exclude the new pairs of technological classes that are purely re-combination of existing knowledge (i.e. technological classes) and do not expand a city's knowledge base<sup>9</sup>. As shown in the next section, however, our results are robust to alternative measurements of the dependent variable. Importantly, this variable does not only improve on measures such as the count of new combinations in a city; it is also preferable to other (simpler) indicators such as the number of patents introducing a new technical field or the simple count of the newly introduced codes, as it emphasizes the intrinsic re-combinatorial nature of technical change and inventive processes (FLEMING, 2001; FLEMING et al., 2007b; KATILA and AHUJA, 2002).

<sup>9</sup> To be more specific, consider two technological classes  $i$  and  $j$ , in which a city's patents before time  $t$  have been classified, but that were not combined together in any patent. Suppose that a patent at time  $t$  is co-classified in  $i$  and  $j$ . This patent introduces a new combination of technologies. Yet, this re-combination does not expand or renew the existing knowledge base of a city. For this reason, it is not included in the computation of new pairs of technological codes.

As patenting activities may show erratic patterns on a short time basis and the dependent variable can take on a positive value only in cities with a number of patents greater than zero, the analysis has been carried out on a sub-sample of U.S. cities (196 out of 370) showing persistent inventive activity (i.e. with a positive number of patents for each year in the period 1990-2004).

The identification of the most appropriate level of technological aggregation is quite an important issue in such a definition of the dependent variable. Following STRUMSKY et al. (2012) and FLEMING et al. (2007b), we classified patents at the *group* level, as it corresponds to the lowest hierarchical level in the International Patent Classification (IPC) adopted at the EPO. As we will show below, however, our results are robust to the use of a more aggregated technological classification.<sup>10</sup> Looking at the distribution of the 504,400 patents in our sample, we observe that around 20% of them are classified in only one IPC technological group. By definition, those patents do not enter into the calculation of our dependent variable. Moreover, 95% of all patents in our sample are classified in less than or equal to eight technological groups, and 99% in less than or equal to fifteen groups. The remaining 1% of all patents (4702 observations) represent outliers classified in a number of technological groups going from fifteen to sixty three. As the presence of such extreme observations may bias the count of new *combinations* observed in a city, in the construction of our dependent variable, we chose to limit our attention to patents classified in less than or equal to eight groups<sup>11</sup>. Yet, robustness checks carried out by including in the computation also patents classified in a greater number of technological groups show that our basic findings hold and are qualitatively similar.

Finally, for the construction of the dependent variable, we considered patents in which *only* inventors located in the focal MSA were reported in the document. By excluding

<sup>10</sup> Groups are next divided in subgroups; however, subgroups are nested into groups (i.e. their hierarchical level varies across groups) and therefore cannot be exploited in a study such as ours. Further details are available at: [http://www.wipo.int/export/sites/www/classifications/ipc/en/guide/guide\\_ipc.pdf](http://www.wipo.int/export/sites/www/classifications/ipc/en/guide/guide_ipc.pdf)

<sup>11</sup> An illustrative case of the problems that may arise from the inclusion of these outliers is that of patent EP1632470, which has been applied for by NPS Pharmaceuticals, located in Salt Lake City, UT. This patent, with priority year 1991, is classified in 40 different technological groups. Of the 780 combinations of technological groups associated to this patent, 702 were new to Salt Lake City in 1991. Given that in the same year, this company applied for other three patents classified in a similarly large number of different technological groups, the total number of new combinations introduced in Salt Lake City in 1991 was equal to 2978, which is an order of magnitude higher than the number of new combinations in 1990 (210) and 1992 (212).

new combinations that are the outcome of cross-city collaborative patents, this variable not only mitigates possible endogeneity concerns with reference to the external reach and gatekeeping indicators, but it can also be considered as a measure of a city's autonomous re-combinatorial and inventive capabilities.

#### 4.2. *Explanatory variables*

The empirical model was designed to assess the importance of external linkages and gatekeepers, while controlling for other factors affecting the renewal and expansion of a city's knowledge base, namely:

1. the importance of agglomeration economies (DURANTON and PUGA, 2004; ELLISON et al., 2010; GLAESER, 1999);
2. the nature of the local knowledge base (FRENKEN and BOSCHMA, 2007; FRENKEN et al., 2007; NEFFKE et al., 2011; Ó HUALLACHÁIN and LEE, 2011);
3. the structure of the co-invention network within the city (BETTENCOURT et al., 2007; FLEMING et al., 2007a; LOBO and STRUMSKY, 2008).

In terms of agglomeration, the empirical model includes two variables to measure the scale of inventive inputs and the relevance of agglomeration economies. First, the number of *internal patents* in the city at time  $t$  (i.e. by excluding patents with inventors external to the city) captures both the scale effect associated with the agglomeration of inventive activities at the city level, as suggested by BETTENCOURT et al. (2007) and LOBO and STRUMSKY (2008), and the potential for technological re-combination<sup>12</sup>. Second, we also controlled for the degree of *concentration* of inventive activities among firms by computing the Herfindahl index at the level of patent assignees. This accounts for the local market structure and captures whether more competitive cities enable greater knowledge creation and re-combination (BEAUDRY and SCHIFFAUEROVA, 2009).

<sup>12</sup> We tried to include in the analysis also the total number of technological groups in which the internal patents made in a city in the period  $(t-1, t-5)$  are classified. Yet, we were forced to exclude it because of problems of multicollinearity with the other independent variables. We thank two anonymous referee for raising this issue.

We included several variables to capture the nature of the knowledge base in the city. First, we computed for each city and for each year the Herfindahl index using the share of patents made in IPC four-digit (i.e. subclass) technological fields. This is an index of *absolute* specialization. It captures to what extent a city is specialized into a narrow set of fields, thereby controlling for the presence of externalities arising from technological specialization (FELDMAN and AUDRETSCH, 1999).<sup>13</sup> Second, we also computed the average number of *citations to the non-patent literature* made by a city's internal patents. This variable captures the scientific orientation and generality of the local knowledge base. On the one hand, one may argue that a more generic and science-based knowledge base could sustain higher opportunities of technological re-combination; more universal notions are more likely to find multiple applications at the junction of different technological domains (FLEMING and SORENSON, 2001, 2004). On the other hand, one may also argue that science-based knowledge is more distant from technological applications, with a negative effect on the re-combinatorial capabilities of a city.

Whereas the Herfindahl index captures the *absolute* technological specialization of a city, our main hypothesis on the role of the local knowledge base refers to the similarity across cities in the profile of technological specialization. To this purpose, we included in the model an index of *relative* technological specialization between the technological profile of city  $c$  and all the other cities with which inventors of city  $c$  have collaborative linkages. In particular, we computed the so-called Krugman index (KI) at the level of technological *groups* in which a city's patents are classified. For each city  $c$  a time  $t$ , we computed the following index:

$$\sum_i \left| \frac{P_{ci}}{P_c} - \frac{P_i}{P} \right|$$

where  $P_{ci}$  is the number of patents of city  $c$  in technological field  $i$ ,  $P_c$  is the total number of patents of city  $c$ ,  $P_i$  is the total number of patents made by all US cities

<sup>13</sup> Although the adoption of a relatively aggregated technological classification level may not allow to fully exploit the details about the technological content of patent data, the measurement of this variable is based on technological *subclasses* (i.e. 4-digit IPC) and not technological *groups* (i.e. the lower technological aggregation level), as the use of groups would disproportionately and artificially inflate its value.



(excluding city  $c$ ) in technological field  $i$ , weighted by the frequency of co-inventing links between inventors of city  $c$  and inventors of city  $d$ , and  $P$  is the total number of patents made by all US cities (except city  $c$ ), weighted by the frequency of co-inventing links between inventors of city  $c$  and inventors of city  $d$ .<sup>14</sup> The index ranges from 0 to 2, taking value 0 for cities whose technological profile is perfectly identical to the average technological profile of the other cities with which it has external linkages, and taking value 2 in cities which are specialized in completely different fields. We expect that the more specialized and dissimilar the knowledge base in a city (i.e. its technological specialization profile) from the average knowledge base of all other cities with which it has linkages, the more the re-combination of internal and external knowledge may be hindered by communication and learning impedances in knowledge exchanges. In this case, the mediating role of gatekeepers can be especially important to access, code, and transfer the external knowledge into the city.

As far as the network variables are concerned, in addition to the variables capturing the importance of external linkages and gatekeepers, we included two further controls for other structural properties of the co-invention network *within* a city (by considering only ties among inventors located in the same city).

First, the fraction of a city's inventors in the *largest connected component* is the ratio between the number of inventors that are in the largest component of the network and the total number of metropolitan inventors. It ranges from zero (all inventors are isolates) to one (all inventors are directly or indirectly connected). Previous studies used this indicator to capture the size and degree of internal connectivity in the co-invention network (e.g. LOBO and STRUMSKY, 2008). Second, the *clustering coefficient* captures the extent to which the partners of an inventor, within the city, are also partners with each other. This index ranges from zero to one, with higher values indicating that

<sup>14</sup> More formally,  $P_i$  is defined as:

$$\sum$$

where  $P_{di}$  is the number of patents that city  $d$  has made in technological field  $i$ , and  $w_{dc}$  is the weight of city  $d$  on all external collaborative links between inventors of city  $c$  and inventors in all other cities. This implies that cities with which inventors of city  $c$  have more collaborative links (and thus exchange knowledge) weight more in the computation of the Krugman index of city  $c$ . Similarly, only cities with which inventors of city  $c$  have external linkages enter into the computation of the index.

the internal city network is composed of dense cliques of collaboration<sup>15</sup>. As argued by some authors, cliquishness can cause isolation and localism, reduce exposure to alternative ideas, limit the access, absorption and re-combination of externally sourced and ultimately re-combination opportunities (UZZI and SPIRO, 2007).

All explanatory variables (with the exclusion of internal patents) are computed on five-year moving window and are one-year lagged (i.e. computed over the period  $t-1, t-5$ ) with respect to the dependent variable to mitigate endogeneity concerns. Summary statistics for all variables are available in Table 1 and the correlation matrix is reported in the Annex. In order to reduce concerns about collinearity among variables and to facilitate interpretation of results, we mean-centered all dependent variables.<sup>16</sup>

[Table 1 about here]

#### 4.3. *The estimation framework*

To account for the integer and over-dispersed nature of the dependent variable, we used a conditional negative binomial framework with fixed-effects by controlling for time effects, as done in similar studies (FLEMING et al., 2007a; FLEMING et al., 2007b; SCHILLING and PHELPS, 2007). Since the conditional fixed-effects negative binomial model has been recently criticized for not being a ‘true fixed-effect’ model (ALLISON and WATERMAN, 2002), we tested the robustness of our results by using alternative estimation models. Moreover, we checked for the possibility of spatial dependence in our dependent variable, i.e., correlation in the dependent variable among neighboring cities. To this purpose, we computed for each year the Moran  $I$  index by using a continuous row-standardized distance matrix. Moran  $I$  is generally very small and never

<sup>15</sup> In affiliation networks as co-invention networks, this coefficient tends to increase with the number of inventors per patent (NEWMAN, 2010). In fact, any three inventors listed in the same patent document by construction form a triangle, i.e. all paths of length two are automatically closed. By extension, the greater the number of patents with more than two inventors, the greater the coefficient. To correct for this possible bias and inflationary mechanism, we computed the generalized version of the clustering coefficient proposed by OPSAHL (2010) that excludes those triples of inventors connected because of a joint patent and only counts the number of triples (i.e., closed paths of length two) that are the outcome of independent interactions between pairs of inventors.

<sup>16</sup> Diagnostic tests tend to exclude serious risks of collinearity. The variance inflation ratio (VIF) is below 2.5 for all variables with the exception of the (log) of internal patents (3.52), the share of external reach mediated by the gatekeepers (3.12) and the Krugman index (2.56). However, even for these variables, the VIF is well below 10, which is the rule of thumb value usually considered in order to detect serious problems of multicollinearity.

statistically significant, supporting the null hypothesis of zero spatial correlation in the dependent variable. As an additional control, we also computed for each year and for each model, the Moran  $I$  index on the regression residuals. Also in this case, Moran  $I$  is generally very small and largely not statistically significant.

## 5. RESULTS

Results of regression estimates are reported in Table 2. The first two columns report estimates of models, which include only our variables of interest, plus the total number of internal patents. The third column reports instead the full model, which includes also the control variables. All models report estimated coefficients transformed to incidence rate ratios (IRR), defined as  $\exp(\hat{\beta})$ .<sup>17</sup> Starting with the control variables, the coefficient on the number of internal patents shows, as expected, that the scale of experimentation matters. More specifically, estimates indicate that, all else equal, a unit increase in the (log) number of internal patents is associated to a doubling in the number of new combinations of technological groups introduced in a city. Regarding the other control variables, the IRR lower than one on the Herfindahl index at the level of IPC 4-digits suggests that *absolute* specialization on a narrow set of technological fields is associated to a decreased ability to re-combine existing technologies with new ones and to expand the knowledge base. Similarly, the coefficients on the fraction of inventors in the largest component and on the clustering coefficient seem to suggest that in cities characterized by dense cliques of collaborators, knowledge may circulate rapidly within cliques, but it may also be highly redundant, which is detrimental for re-combination opportunities. Furthermore, the value of the coefficient on the Krugman index signals that a city with a technological profile dissimilar from that of the other cities, with which it exchanges knowledge, is more likely to expand and renew their knowledge base, possibly because the lower cognitive overlapping enhances the opportunities for learning.

<sup>17</sup> For a unit change in  $\beta_k$ , the expected count in the dependent variable changes by a factor of  $\exp(\hat{\beta}_k)$ , holding all other variables constant. For a standard deviation change in  $\beta_k$ , the expected count changes by a factor of  $\exp(\hat{\beta}_k)$ . Alternatively, IRR can be interpreted in terms of percentage change in the expected count. Thus, for a unit change in  $\beta_k$ , the percentage change in the expected count is given by  $[\exp(\hat{\beta}_k) - 1] \times 100$ .

Turning the attention to our main hypotheses, we first observe that the average external reach presents a larger than one and statistically significant IRR, thereby confirming our hypothesis 1, even though the magnitude of the effect is not extremely large: keeping all other variables constant, a standard deviation increase in the average external reach (for  $\leq 4$ , see above) brings around 6.4% ( $=[\exp(0.0061317 \times 10.105)-1] \times 100$ ) more new combinations in a city<sup>18</sup>. In addition to this, the coefficient of the interaction between the Krugman index and the share of external reachability mediated by gatekeepers provides support to hypothesis 2: the more similar (dissimilar) the technological profile of a city to the other cities with which it exchanges knowledge, the smaller (greater) the importance of gatekeepers to access external information and, on parallel, the greater (smaller) the importance of direct linkages. This result is broadly consistent with the view that gatekeepers not only perform a bridging function, but also trans-code and diffuse the external knowledge at the local level. In order to appreciate the magnitude of this effect, it may be useful to consider two polar cases: on the one hand, a city with a technological profile relatively similar to the other cities (i.e. with a value of the Krugman index, KI, equal to 25<sup>th</sup> percentile) and, on the other hand, a city with a relatively dissimilar technological profile (i.e. with a value of the Krugman index, KI, equal to 75<sup>th</sup> percentile). For each city, one can compute the effect of a similar increase in the share of external reach mediated by gatekeepers on the expected number of new technological combinations. In particular, our estimates show that when the share of external reachability mediated by gatekeepers goes from 6% to 36%<sup>19</sup> the expected number of new combinations *decreases* by 11% for a city with a relatively similar profile, whereas the same variation *increases* the expected number of new combinations by 5.8% for a city with a relatively dissimilar profile.

[Table 2 about here]

<sup>18</sup> Please note that  $\exp(0.0061317) = 1.0061505$  is the IRR associated to the average external reach.

<sup>19</sup> The two values correspond, respectively, to the 25<sup>th</sup> and 75<sup>th</sup> percentile of the empirical distribution of the variable. In this way, one can simulate what happens to the expected number of new combinations when the share of external reach mediated by the gatekeepers increase from small to large values.

As these results might depend on the assumptions made to construct the dependent and independent variables, we carried out several robustness checks. In the first place, we tested the robustness of results with respect to alternative estimation methods. The model of Table 2 was estimated using both Poisson fixed effects and unconditional negative binomial regression (GUIMARÃES, 2008). Results (not reported for brevity) were remarkably stable. Secondly, we tested the extent to which the results might depend on the assumptions made to build the measures of gatekeeping (see above fn. 8). In particular, Model 1 in Table 3 considers only pairs of inventors at  $d_{ij} \leq 2$  in order to construct the external reach indicator and the share of it mediated by gatekeepers, while Model 2 in Table 3 considers only pairs of inventors at  $d_{ij} \leq 3$ . All major results are again confirmed. Interestingly, the estimates indicate that the impact of external reach decreases as the social distance between internal and external inventors increases. If we compare the coefficient of the external reach variable in Table 2 and Table 3 (Models 1 and 2), we observe that its value decreases as the threshold value for  $d_{ij}$  increases, thus suggesting that socially closer interactions are more valuable for broadening the existing knowledge base, as knowledge passes through a lower number of intermediaries and it is thus more trusted and less distorted. The third set of robustness checks concerned the dependent variable. To this purpose, we tested alternative and more restrictive measures of our dependent variable. Specifically, Model 3 in Table 3 estimates a model in which the dependent variable is defined as the number of new pairs of technological *groups*, in which *both* groups in the pair were never used previously in the city (i.e. we count the number of totally new technological combinations). In addition to that, as our results might be affected by the highly disaggregated technological level used to build our dependent variable and to measure the dissimilarity in the technological specialization profile, we re-computed these two variables (i.e. the number of new combinations introduced in a city and the Krugman index) by using a higher level of aggregation (i.e. so-called IPC 4-digit technological *subclasses*) (Model 4 in Table 3). In both cases, our key findings are largely confirmed.

## 6. CONCLUSIONS

This paper aimed at assessing the importance of external sources of knowledge and gatekeepers for the expansion and renewal of a city's knowledge base. Differently from most of the literature on the subject, the paper has adopted a meso-level perspective and discussed under what local conditions gatekeepers are more likely to generate positive returns on the innovative outcomes (measured through the introduction of new and previously uncombined technology pairs) of the geographical contexts in which they are embedded. Interestingly, the results indicate that although external relations and their structure play a pivotal role to renew, expand and regenerate a city's knowledge base, not all types of relations are alike. Importantly, albeit considered as more imaginative, inspired and open to novelties and radical innovation, gatekeepers *per se* do not necessarily contribute to enrich the knowledge base of the cities in which they are located. We argue that, at an aggregate level of analysis such as the urban level, the control power that gatekeepers can exert on the knowledge flows they govern can more than offset the benefits accruing from their superior inventive performance. Still, given their ability to interpret and trans-code externally sourced knowledge into locally meaningful information, gatekeepers can have quite a substantial role in spurring the expansion and renewal of the knowledge base in cities characterized by a highly localized and specialized knowledge base, dissimilar from the average technological specialization profile of the other cities, with which it exchanges knowledge. These cities, in fact, require an interface to access, absorb and use external knowledge. On the other hand, those cities that have a more general and universal knowledge base prefer to rely on direct connections (and thus less noisy information exchanges) to source knowledge from outside the city boundaries.

The paper adds two main contributions to the extant literature. On conceptual grounds, the paper helps qualifying the role and function of gatekeepers. Gatekeepers have been often invoked as the most important means to access and exploit external knowledge. In this perspective, strategy and policy recommendations have been drawn that aim at increasing their number and importance in mediating knowledge flows both across organizational and geographical boundaries. Not only our paper questions this view, but it also shows that on average direct linkages outperform connections mediated by

gatekeepers by allowing faster, more trusted and less noisy knowledge exchanges. On methodological grounds, the paper proposes an operational method to quantify the importance of gatekeepers in brokering knowledge flows across cities that emphasizes the meso-level effects (i.e. on the metropolitan knowledge base) of individual behavior and interactions (i.e. knowledge exchanges and gatekeeping roles) and that, hopefully, will be useful and deployed in future research.

Some cautionary words should be finally mentioned. First, patent data capture only a subset of the links relevant to knowledge exchanges within and across cities, albeit one should also note that the network of collaborators is the most immediate and influential environment from which inventors draw ideas and information. In this respect, we believe that our work is supplying rather conservative estimates on the impact of external linkages on a city's knowledge base. Second, in order to corroborate and boost confidence about our findings on the role of external linkages and gatekeepers for the rejuvenation and expansion of the knowledge base, the meso-level approach adopted in this paper should be complemented by empirical work at a different level of analysis, such as the firm or the inventive team level. In particular, the analysis at the team level would enable to directly link the technological profile of potential gatekeepers with the inventive output and recombinatorial capabilities of the team she is collaborating with, whereas the analysis at the metropolitan level captures not only this direct effect but also the indirect effects that gatekeepers may exert on the recombinatorial capabilities of the other inventors in the city although they might not be direct team-workers. Finally, the dependent variable used in this paper does not take into account the degree of relatedness between different technologies and thus the ease of recombination between them, nor it distinguishes the newly introduced combinations according to their radical *vs.* incremental nature. In this sense, we believe that the contribution of this paper is complementary to the most recent studies exploring the role of knowledge relatedness for the dynamics of technological specialization in cities. Overall, the three limitations above represent equally promising avenues for further research.

Table 1  
Summary statistics

Variable	Mean	Std. Dev.	Min	Max
Number of new combinations	98.454	119.686	0	727
Number of internal patents	136.245	309.367	1	2916
Herfindahl (firms)	0.129	0.147	0.004	0.914
Herfindahl (technologies; IPC 4-digit)	0.068	0.057	0.012	0.488
Average number of citations to NPL	1.278	1.070	0.000	6.263
Krugman index (KI)	1.381	0.344	0.370	1.971
Fraction of inventors in the largest component	12.223	10.351	1.005	67.262
Clustering coefficient	0.245	0.227	0.000	1.000
Average (distance-weighted) external reach ( $d = 4$ )	7.302	10.105	0.071	162.995
Share of external reach mediated by gatekeepers ( $d = 4$ )	0.235	0.201	0.000	0.862

Note: 2940 observations (196 MSA  $\times$  15 years).



Table 2  
Determinants of knowledge base renewal and expansion in U.S. cities, 1990-2004  
Dependent variable: Number of new combinations of IPC technological *groups*

Variables	(1)	(2)	(3)
Number of internal patents (log)	2.152** (0.037)	2.176** (0.037)	2.158** (0.037)
Herfindahl index (firm level)			0.853 (0.127)
Herfindahl index (technology level)			0.199** (0.081)
Average number of citations to NPL			0.984 (0.020)
Fraction of inventors in largest component			0.994** (0.002)
Clustering coefficient			0.979 (0.041)
Average external reach	1.002 <sup>†</sup> (0.001)	1.004** (0.001)	1.006** (0.001)
External reachability mediated by gatekeepers	0.573** (0.041)	0.655** (0.048)	0.878 (0.074)
Krugman index (KI)	1.271** (0.091)	1.192* (0.085)	1.307** (0.097)
External reachability mediated by gatekeepers × KI		2.706** (0.375)	3.153** (0.445)
Constant	5.852** (0.229)	6.142** (0.240)	6.207** (0.244)
Observations (196 cities × 15 years)	2940	2940	2940
Log-likelihood	-12136.4	-12110.7	-12085.6
$\chi^2$	3150.8	3147.9	3204.7

Notes: <sup>†</sup>  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ . Year dummies included. The table reports the estimated coefficients transformed to incidence rate ratios (IRR) defined as  $\exp(\hat{\beta})$ . Standard errors in parentheses.

Table 3  
Estimates for different thresholds of distance (Models 1 and 2) and for different definitions of the dependent variable (Models 3 and 4)

	(1) Threshold distance	(2) Threshold distance	(3) Completely new combinations of IPC <i>groups</i>	(4) New combinations of IPC <i>subclasses</i> (4-digit)
Number of internal patents (log)	2.172** (0.037)	2.161** (0.037)	1.931** (0.042)	1.884** (0.042)
Herfindahl index (firm level)	0.888 (0.132)	0.869 (0.130)	0.669* (0.127)	0.709† (0.139)
Herfindahl index (technology level)	0.188** (0.077)	0.195** (0.080)	0.157** (0.078)	0.608 (0.321)
Average number of citations to NPL	0.986 (0.020)	0.985 (0.020)	0.991 (0.024)	1.016 (0.026)
Fraction of inventors in largest component	0.995** (0.002)	0.995** (0.002)	0.996† (0.002)	0.991** (0.002)
Clustering coefficient	0.975 (0.041)	0.978 (0.041)	0.944 (0.057)	0.886† (0.055)
Average external reach	1.034** (0.007)	1.011** (0.002)	1.009** (0.002)	1.007** (0.002)
External reachability mediated by gatekeepers	0.632** (0.093)	0.814* (0.083)	0.777* (0.091)	0.808† (0.096)
Krugman index (KI)	1.362** (0.101)	1.328** (0.098)	1.567** (0.147)	1.441** (0.146)
External reachability mediated by gatekeepers × KI	5.858** (1.429)	3.640** (0.609)	3.428** (0.704)	4.338** (1.091)
Constant	6.226** (0.247)	6.178** (0.244)	3.649** (0.174)	4.027** (0.214)
N	2940	2940	2940	2940
Log-likelihood	-12086.0	-12087.3	-10397.5	-8709.3
$\chi^2$	3205.6	3202.4	1470.7	1082.1

Notes: †  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ . The table reports the estimated coefficients transformed to incidence rate ratios (IRR) defined as  $\exp(\beta)$ . Standard errors in parentheses. Year dummies included. Model 1 includes only pairs of inventors at distance  $\leq 1$  in order to calculate the external reach variable and the share of it mediated by gatekeepers. Model 2 includes only pairs of inventors at distance  $\leq 2$ . Model 3 defines the dependent variable as the count of new combinations of IPC technological *groups*, in which *both* groups are new to the city (i.e. they represent completely new technological combinations). Model 4 defines the dependent variable as the count of new combinations of IPC technological subclasses (IPC 4-digit).

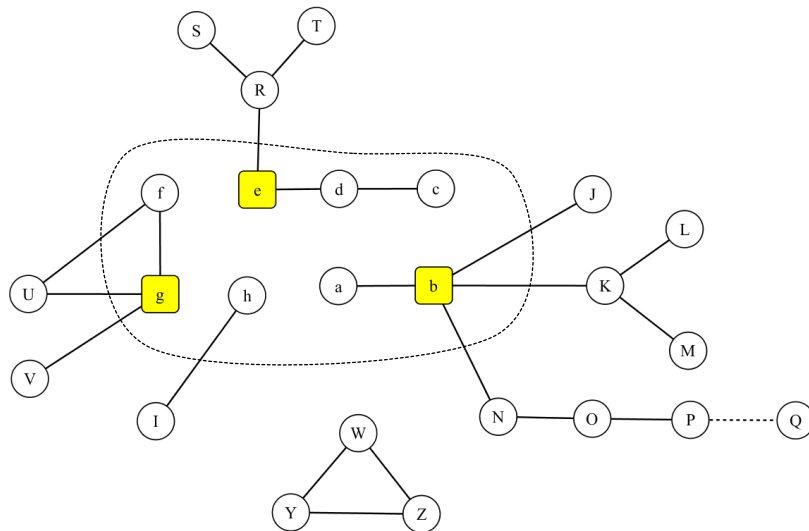
# ANNEX

Table A1. Correlation matrix

		1	2	3	4	5	6	7	8	9
Number of new combinations	1									
Internal patents	2	0.855								
Herfindahl (firms)	3	-0.194	-0.184							
Herfindahl (technologies; IPC 4-digit)	4	-0.319	-0.313	0.523						
Average number of citations to NPL	5	0.107	0.119	-0.249	-0.039					
Krugman index	6	-0.652	-0.738	0.248	0.313	-0.242				
Largest component	7	-0.112	-0.116	0.540	0.367	-0.150	0.178			
Clustering coefficient	8	0.369	0.424	-0.125	-0.167	0.098	-0.341	-0.024		
Average (distance-weighted) external reach ( $d = 4$ )	9	0.043	0.136	0.145	0.084	0.065	-0.225	0.306	0.108	
Share of external reach mediated by gatekeepers ( $d = 4$ )	10	0.496	0.582	0.263	0.110	0.045	-0.443	0.462	0.317	0.401

This Appendix illustrates in detail the construction of the measures used in the paper to capture the intensity of external linkages and the role gatekeepers play in mediating the access to the external sources of knowledge. To this end, we use the example reported in Figure 1. The example illustrates the hypothetical case of a city, in which eight inventors are located (denoted in the figure by the small letters from *a* to *h*). Outside the boundaries of the hypothetical city there are eighteen inventors (denoted in the figure by the capital letters from *I* to *Z*).

Figure 1 – External linkages and gatekeepers



Notes: small letters from a to h denote inventors located in a city, whose boundaries are defined by the dashed line. Capital letters from I to Z denote inventors located outside the boundaries of the city. Rounded square, yellow nodes denote gatekeepers, namely inventors located in the city which mediate the access to externally located inventors for other co-located inventors.

If we take inventor  $a$  located in the hypothetical city  $c$ , it takes two steps to reach inventor  $J$  outside her city, three steps to reach inventor  $L$ , four steps to reach inventor  $P$ , five steps to reach inventor  $Q$ , and so on. Her distance-weighted external reach equals the sum of the reciprocal distances to all other external inventors she can reach, as follows:

$$\sum - \quad - \quad - \quad - \quad - \quad - \quad -$$

Note that since  $d_{aQ} = 5 > 4$ , we set  $1/d_{aQ} = 0$  (see footnote 8 above). The average distance-weighted external reach (ADWR) for the hypothetical city is easily obtained by computing the DWR for each inventor located in that city and then taking the average across all inventors. In the example, this becomes:

$$- \left( \frac{DWR}{2} + \frac{DWR}{1} \right)$$

The above calculation assumes that indirect linkages whose distance is greater than 4 do not convey any knowledge flow. This threshold is somewhat arbitrary, though it has been shown that it corresponds to the distance at which knowledge flows start decaying very rapidly (BRESCHI and LISSONI, 2004; SINGH, 2005). In order to test the robustness of results, the ADWR index can be recalculated by assuming a faster rate of depreciation of knowledge with respect to social distance. For example, one can easily check that, assuming a threshold at distance 3, the ADWR index in the example takes value 1.98.

As far as gatekeepers are concerned, we adapted the definition given by (GOULD and FERNANDEZ, 1989). Informally, they are actors that mediate the access to externally located inventors for local inventors. Formally, we define an inventor  $i$  located in city  $c$  as a gatekeeper when the shortest path leading from any other inventor  $j$  in city  $c$  to any other inventor  $h$  in city  $d$  (different from  $c$ ) passes through  $i$ . Following this definition, there are three gatekeepers in the hypothetical city illustrated in Figure 1. They are identified by the rounded square, yellow nodes. For example, the shortest path from node  $f$  to node  $V$  is through node  $g$ , even though  $f$  can directly reach node  $U$ . Similarly, the shortest paths from node  $a$  to nodes  $N$ ,  $O$  and  $P$  pass through node  $b$ . On the other hand, inventor  $h$  does not need any intermediation to reach the external inventor  $I$ .

Given the above definition of gatekeepers, one way to measure the relative importance of gatekeepers vs. direct external linkages in mediating the access to external sources of knowledge is by computing to what extent the average distance-weighted external reach would decrease, should one remove them from the city (BORGATTI, 2006; VALENTE

and FUJIMOTO, 2010). To this purpose, we simply computed for each city the share of total distance-weighted external reachability, which is mediated by gatekeepers.

With reference to the example above, the (distance-weighted) external reachability of inventor  $a$  is entirely mediated by node  $b$ . Similarly, inventors  $c$  and  $d$  can reach other externally located inventors only through node  $e$ . As far as inventor  $f$  is concerned, she can reach inventor  $U$  in a direct way, so that  $g$  does not perform any gatekeeping role.<sup>20</sup> Yet, inventor  $f$  needs to pass through inventor  $g$  in order to reach inventor  $V$ . Given that the total external reachability of inventor  $f$  is equal to  $1/2 = 1.5$ , the share of it that is mediated by gatekeepers is  $5/1.5 = 1/3$ . In other words, the external reachability of node  $f$  would decrease by 0. (or 1/3 of the total ) should one remove all gatekeepers from the city. If one applies the same logic to all inventors in our hypothetical city, it can be easily checked that the share of total external reachability that is mediated by gatekeepers is equal to  $5.75/16.83 = 0.34$ . In other words, 34% of the total external reachability in the city is accounted for by indirect linkages, which are mediated by gatekeepers.

<sup>20</sup> To put it in a different way, the shortest path from  $f$  to  $U$  does not pass through  $g$ . More generally, there are cases in which an inventor reaches an external inventor both through direct connections and through indirect links, mediated by other inventors. As indirect links imply, by definition, a greater distance than direct links, these triples are obviously excluded from the computation of the external reach mediated by gatekeepers. This happens, particularly, in the case of patents with more than two inventors, one of which located outside the boundaries of the city.

## References

- AHUJA G. (2000) Collaboration Networks, Structural Holes, and Innovation: A Longitudinal Study, *Administrative Science Quarterly* **45**, 425-55.
- ALLEN T. J. (1977) *Managing the Flow of Technology*. MIT Press, Cambridge, Mass.
- ALLEN T. J. and COHEN S. I. (1969) Information flow in research and development laboratories, *Administrative Science Quarterly* **14**, 12-9.
- ALLISON P. D. and WATERMAN R. P. (2002) Fixed-effects negative binomial regression models, *Sociological methodology* **32**, 247-65.
- ANTONELLI C. (1995) *The economics of localized technological change and industrial dynamics*. Kluwer Academic Pub, Dordrecht.
- BATHELT H., MALMBERG A. and MASKELL P. (2004) Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation, *Progress in Human Geography* **28**, 31-56.
- BEAUDRY C. and SCHIFFAUEROVA A. (2009) Who's right, Marshall or Jacobs? The localization versus urbanization debate, *Research Policy* **38**, 318-37.
- BETTENCOURT L. M. A., LOBO J. and STRUMSKY D. (2007) Invention in the city: Increasing returns to patenting as a scaling function of metropolitan size, *Research Policy* **36**, 107-20.
- BORGATTI S. P. (2006) Identifying sets of key players in a social network, *Computational and Mathematical Organization Theory* **12**, 21-34.
- BOSCHMA R. (2005) Proximity and Innovation: A Critical Assessment, *Regional Studies* **39**, 61-74.
- BOSCHMA R. and FRENKEN K. (2010) The spatial evolution of innovation networks. A proximity perspective, in BOSCHMA R. and MARTIN R. (Eds) *Handbook of evolutionary economic geography*, pp. 120-35. Edward Elgar Cheltenham.
- BOSCHMA R. A. and FRENKEN K. (2006) Why is economic geography not an evolutionary science? Towards an evolutionary economic geography, *Journal of Economic Geography* **6**, 273-302.
- BRESCHI S. and LISSONI F. (2001) Knowledge Spillovers and Local Innovation Systems: A Critical Survey, *Industrial and Corporate Change* **10**, 975-1005.
- BRESCHI S. and LISSONI F. (2004) Knowledge networks from patent data: Methodological issues and research targets, in GLANZEL W., MOED H. and SCHMOCH U. (Eds) *Handbook of Quantitative S&T Research: The Use of Publication and Patent Statistics in Studies of S&T Systems*, pp. 613-43. Springer Verlag, Berlin.
- BRESCHI S. and LISSONI F. (2009) Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows, *Journal of Economic Geography* **9**, 439-68.
- BRESNAHAN T., GAMBARDILLA A. and SAXENIAN A. (2001) 'Old Economy' Inputs for 'New Economy' Outcomes: Cluster Formation in the New Silicon Valleys, *Industrial and Corporate Change* **10**, 835-60.
- BURT R. (2004) Structural Holes and Good Ideas, *American Journal of Sociology* **110**, 349-99.
- BURT R. S. (1992) *Structural holes: the social structure of competition*. Cambridge Mass: Harvard University Press.
- BURT R. S. (2008) Information and structural holes: comment on Reagans and Zuckerman, *Industrial and Corporate Change* **17**, 953-69.

- CAPELLO R. (2009) Indivisibilities, synergy and proximity: the need for an integrated approach to agglomeration economies, *Tijdschrift voor economische en sociale geografie* **100**, 145-59.
- CARLINO G. A., CHATTERJEE S. and HUNT R. M. (2007) Urban density and the rate of invention, *Journal of Urban Economics* **61**, 389-419.
- COWAN R., DAVID P. A. and FORAY D. (2000) The explicit economics of knowledge codification and tacitness, *Industrial and Corporate Change* **9**, 211-53.
- DAHLANDER L. and FREDERIKSEN L. (2012) The core and cosmopolitans: A relational view of innovation in user communities, *Organization Science* **23**, 988-1007.
- DOSI G. (1982) Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change, *Research Policy* **11**, 147-62.
- DURANTON G. and PUGA D. (2004) Micro-foundations of urban agglomeration economies, in HENDERSON J. V. and THISSE J. F. (Eds) *Handbook of regional and urban economics*, pp. 2063-117. Elsevier, Amsterdam.
- ELLISON G., GLAESER E. and KERR W. (2010) What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns, *American Economic Review* **100**, 1195-213.
- FELDMAN M. P. and AUDRETSCH D. B. (1999) Innovation in cities: Science-based diversity, specialization and localized competition, *European Economic Review*, pp. 409-29.
- FELLER I. (1971) The urban location of United States invention, 1860-1910, *Explorations in Economic History* **8**, 285-303.
- FERNANDEZ R. M. and GOULD R. V. (1994) A Dilemma of State Power: Brokerage and Influence in the National Health Policy Domain, *The American Journal of Sociology* **99**, 1455-91.
- FLEMING L. (2001) Recombinant Uncertainty in Technological Search, *Management Science* **47**, 117-32.
- FLEMING L., KING C. and JUDA A. I. (2007a) Small Worlds and Regional Innovation, *Organization Science* **18**, 938-54.
- FLEMING L., MINGO S. and CHEN D. (2007b) Collaborative Brokerage, Generative Creativity, and Creative Success, *Administrative Science Quarterly* **52**, 443-75.
- FLEMING L. and SORENSON O. (2001) Technology as a complex adaptive system: evidence from patent data, *Research Policy* **30**, 1019-39.
- FLEMING L. and SORENSON O. (2004) Science as a map in technological search, *Strategic Management Journal* **25**, 909-28.
- FRENKEN K. and BOSCHMA R. A. (2007) A theoretical framework for evolutionary economic geography: industrial dynamics and urban growth as a branching process, *Journal of Economic Geography* **7**, 635-49.
- FRENKEN K., VAN OORT F. and VERBURG T. (2007) Related Variety, Unrelated Variety and Regional Economic Growth, *Regional Studies* **41**, 685-97.
- GITTELMAN M. (2007) Does Geography Matter for Science-Based Firms? Epistemic Communities and the Geography of Research and Patenting in Biotechnology, *Organization Science* **18**, 724-41.
- GIULIANI E. and 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**, 254-77.



- GOULD R. V. and FERNANDEZ R. M. (1989) Structures of mediation: A formal approach to brokerage in transaction networks, *Sociological methodology* **19**, 89-126.
- GRAF H. (2011) Gatekeepers in regional networks of innovators, *Cambridge Journal of Economics* **35**, 173-98.
- GUIMARÃES P. (2008) The fixed effects negative binomial model revisited, *Economics Letters* **99**: 63-66.
- HARGADON A. and SUTTON R. I. (1997) Technology brokering and innovation in a product development firm, *Administrative Science Quarterly* **42**, 716-49.
- KATILA R. and AHUJA G. (2002) Something old, something new: A longitudinal study of search behavior and new product introduction, *Academy of Management Journal* **45**, 1183-94.
- LAMOREAUX N. R. and SOKOLOFF K. L. (2000) The Geography of Invention in the American Glass Industry, 1870-1925, *Journal of Economic History* **60**, 700-29.
- LISSONI F. (2001) Knowledge codification and the geography of innovation: the case of Brescia mechanical cluster, *Research Policy* **30**, 1479-500.
- LOBO and STRUMSKY D. (2008) Metropolitan patenting, inventor agglomeration and social networks: A tale of two effects, *Journal of Urban Economics* **63**, 871-84.
- MORRISON A. (2008) Gatekeepers of Knowledge within Industrial Districts: Who They Are, How They Interact, *Regional Studies* **42**, 817-35.
- MORRISON A., RABELLOTTI R. and ZIRULIA L. (2013) When Do Global Pipelines Enhance the Diffusion of Knowledge in Clusters?, *Economic Geography* **89**, 77-96.
- MUNARI F., SOBRERO M. and MALIPIERO A. (2012) Absorptive capacity and localized spillovers: focal firms as technological gatekeepers in industrial districts, *Industrial and Corporate Change* **21**, 429-62.
- NEFFKE F., HENNING M. and BOSCHMA R. (2011) How do regions diversify over time? Industry relatedness and the development of new growth paths in regions, *Economic Geography* **87**, 237-65.
- NEWMAN M. (2010) *Networks: An Introduction*. Oxford University Press, USA.
- Ó HUALLACHÁIN B. and LEE D.-S. (2011) Technological Specialization and Variety in Urban Invention, *Regional Studies* **45**, 67-88.
- OBSTFELD D. (2005) Social Networks, the Tertius Iungens Orientation, and Involvement in Innovation, *Administrative Science Quarterly* **50**, 100-30.
- OPSAHL T. (2010) Triadic closure in two-mode networks: Redefining the global and local clustering coefficients, *arXiv:1006.0887v2*.
- OWEN-SMITH J. and POWELL W. W. (2004) Knowledge Networks as Channels and Conduits: The Effects of Spillovers in the Boston Biotechnology Community, *Organization Science* **15**, 5-21.
- PADGETT J. F. and ANSELL C. K. (1993) Robust Action and the Rise of the Medici, 1400-1434, *American Journal of Sociology* **98**, 1259-319.
- PRED A. R. (1973) *Urban Growth and the Circulation of Information: The United States System of Cities, 1790-1840*. Harvard University Press, Cambridge, Mass.
- SCHILLING M. A. and PHELPS C. C. (2007) Interfirm Collaboration Networks: The Impact of Large-Scale Network Structure on Firm Innovation, *Management Science* **53**, 1113-26.
- SINGH J. (2005) Collaborative Networks as Determinants of Knowledge Diffusion Patterns, *Management Science* **51**, 756-70.
- STORPER M. and VENABLES A. J. (2004) Buzz: face-to-face contact and the urban economy, *Journal of Economic Geography* **4**, 351-70.

- STRUMSKY D., LOBO J. and VAN DER LEEUW S. (2012) Using patent technology codes to study technological change, *Economics of Innovation and New Technology* **21**, 267-86.
- TER WAL A. and BOSCHMA R. (2009) Applying social network analysis in economic geography: framing some key analytic issues, *Annals of Regional Science* **43**, 739-56.
- TUSHMAN M. L. and KATZ R. (1980) External communication and project performance: An investigation into the role of gatekeepers, *Management Science* **26**, 1071-85.
- TUSHMAN M. L. and SCANLAN T. J. (1981) Boundary spanning individuals: Their role in information transfer and their antecedents, *Academy of Management Journal* **24**, 289-305.
- UZZI B. (1996) The Sources and Consequences of Embeddedness for the Economic Performance of Organizations: The Network Effect, *American Sociological Review* **61**, 674-98.
- UZZI B. (1997) Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness, *Administrative Science Quarterly* **42**, 35-67.
- UZZI B. and SPIRO J. (2007) Collaboration and Creativity: The Small World Problem. *American Journal of Sociology* **111**, 447-504.
- VALENTE T. W. and FUJIMOTO K. (2010) Bridging: locating critical connectors in a network, *Social Networks* **32**, 212-20.
- WHITTINGTON K., OWEN-SMITH J. and POWELL W. (2009) Networks, Propinquity, and Innovation in Knowledge-intensive Industries, *Administrative Science Quarterly* **54**, 90-122.