Received: 19 April 2021 | Revised: 30 July 2021 | Accepted: 15 September 2021

DOI: 10.1111/pirs.12640

FULL ARTICLE

The role of inventors' networks and variety for breakthrough inventions

Niccolò Innocenti[1] | Francesco Capone[2] | Luciana Lazzeretti[2] | Sergio Petralia[3]

[1]Robert Shuman Centre for Advanced Studies, Florence School of Regulation, European University Institute, Florence, Italy

[2]Department of Economics and Management, University of Florence, Florence, Italy

[3]Department of Economic Geography, Utrecht University, Utrecht, CB, The Netherlands

Correspondence
European University Institute, Robert Shuman Centre for Advanced Studies, Florence School of Regulation, Via Boccaccio, 121–50133, Florence, Italy.
Email: niccolo.innocenti@eui.eu.

Abstract

This paper studies the role of the inventors' networks and the diversity of the local industrial structure for regional innovation, differentiating between regions' production of normal and breakthrough innovations. Our results suggest that, on the one hand, local related variety enhances the overall innovation rate, while unrelated variety supports the rate of breakthrough innovations. On the other hand, we find evidence that inventors' networks are determinant for regions' patenting of normal and breakthrough innovations. To do so we constructed a unique database of Italian patenting activity at the United States Patent and Trademark Office (USPTO) dating back to 1972.

KEYWORDS

breakthrough invention, inventors' networks, Italy, USPTO, variety

JEL CLASSIFICATION

033, R12, D85

1 | INTRODUCTION

In recent innovation literature, several studies analyse the role of inventors' networks on regional inventive activity. This line of research has developed rapidly in recent years thanks to newly available data sources, the use of algorithms that allow the disambiguation of inventors (Fleming, Mingo, et al., 2007; Li et al., 2014; Ventura et al., 2015), and the digitization of patents since the 1800s (Petralia et al., 2017).

In particular, the role of inventors' networks for regional innovation has gained considerable interest with recent works in journals of Regional Sciences, Economic Geography and Innovation studies (Capone et al., 2021; Innocenti et al., 2020; van der Wouden & Rigby, 2019). A considerable strand of this literature has focused on understanding how inventors' networks promote the exchange of knowledge at the local level and how they affect places' innovative productivity (Acs et al., 2002; Singh, 2005; Fleming, Mingo, et al., 2007b), shifting from a firm-centred paradigm to one where areas and regions were the primary unit of analysis (Breschi & Lenzi, 2016; Lobo & Strumsky, 2008; Strumsky & Thill, 2013).

Furthermore, the hype developed from the technological "relatedness" concept has allowed broadening this field of study by analysing how (diversified) places with a higher technological relatedness could excel in the production of inventions (Ebersberger et al., 2014; Van der Wouden, 2020; van der Wouden & Rigby, 2019; etc.).

The present research aims to investigate whether the structure of knowledge networks and the variety of the local industrial structure are important for regional innovation and if there are different effects between normal and breakthrough inventions. The idea is that local related variety might enhance the overall innovation rate due to knowledge spillovers that occur when there is a fair degree of variety (Boschma et al., 2012), while local unrelated variety is expected to support the surge of breakthrough innovations. This is because a wider range of diverse activities may favour unusual connections and spillover among distant sectors, producing more radical inventions (Castaldi et al., 2015). However, what seems particularly interesting is how the inventor network and industrial structure interact and affect overall and breakthrough patenting production (Boschma & Frenken, 2010; Broekel & Boschma, 2012).

The focus of our analysis is at the smallest geographic level of investigation offered by the EU classification system for units with comparable population size, the NUTS 3 level (i.e., Italian provinces). In this work, USPTO co-patenting activities are analysed in the 103 Italian provinces covering a long period, 1972–2011.[1] We built inventors' networks using data on patent collaborations and used it to measure different characteristics these networks. In addition, the industrial structure of the provinces is measured following the methodology first developed by Frenken et al. (2007) building the variables of related and unrelated variety based on employment data. We estimated a negative binomial regression to investigate whether the inventors' network and industrial structure of a region favour production of normal or breakthrough patents.

The paper contributes to several debates. First of all, it contributes to the general debate on the role of inventors network structure for regional inventions production (Breschi & Lenzi, 2016; Lobo & Strumsky, 2008; Strumsky & Thill, 2013). Second, it improves the previous contributions considering also the role of variety/diversity (Capone et al., 2021; Giannini et al., 2019; Innocenti et al., 2020; Van der Wouden, 2020; van der Wouden & Rigby, 2019). Finally, it enhances the previous streams of research analysing the role of inventors' networks *and* variety differentiating for overall and breakthrough inventions. While there are some works on the role of variety/diversity for breakthrough inventions (Castaldi & Los, 2012; Castaldi et al., 2015; Silverberg & Verspagen, 2007), to the best of our knowledge, this is the very first paper analysing the role of inventors' networks *and* variety for overall and breakthrough inventions.

2 | INVENTORS' NETWORKS, DIVERSITY AND REGIONAL INVENTION

The interest around inventors' networks actually dates back in time to Granovetter (1973), Burt (2000), Powell (1990) and Powell et al. (1996). The first contributions are the work of Singh (2005) and Fleming, Mingo, et al. (2007b), who studied the invention productivity of firms according to the structure of knowledge networks of inventors.

However, the recent revitalization of this line of research increased thanks to the development of algorithms that allowed disambiguating inventors. Thanks to the work of Fleming, King, et al. (2007), Li et al. (2014) and Ventura et al., (2015), it is possible to uniquely identify individual inventors in US patents. In Europe the disambiguation efforts where mostly linked to the Massacrator tool (Pezzoni et al., 2014) and the CRIOS Lab activities on EPO patents (Coffano & Tarasconi, 2014; Morrison et al., 2017).[2]

A recent line of research, connected to this field, renewed this interest in inventors' networks and urban context to investigate the innovative capacity of cities and regions. This stream of research has mainly investigated how inventors' networks support the inventive performances of cities and it has been thriving in journals such as *Journal of Urban Economics*, *Journal of Regional Science* and recently in *Papers in Regional Sciences* (Breschi & Lenzi, 2016; Hermans, 2021; Innocenti et al., 2020; Lobo & Strumsky, 2008; Strumsky & Thill, 2013; van der Wouden & Rigby 2019).

Due to the growing availability of data and this renewed interest, an increasing number of contributions started to focus on the structure of inventors' networks in US cities in particular considering a long period. For instance, Petralia et al. (2017) thanks to the use of a machine learning algorithm developed a long-range database (HistPat) based on USPTO data containing information on all patens from 1836 until 1975. More recently, Van der Wouden (2020) merged the Histpat database from Petralia et al. (2017) and USPTO to investigate complexity and collaborations patterns in the US. In this resurge of this stream of research, the discussion mainly focuses on urban areas and regions, involving then regional and urban sciences. Some initial contributions focused on the debate on the primacy of agglomeration economies of inventors on inventor networks (Lobo & Strumsky, 2008; Strumsky & Thill, 2013). The most recent works, on the other hand, focused on the role of social proximity in the inventors' networks and on their contribution to supports knowledge flows and the renewal of the city knowledge base (Breschi & Lenzi, 2016).

The present work contributes to this debate, but it also aims to consider the local variety/diversity for regional inventions production. This subsequent line of research developed from these works started to focus on the theme of diversity and variety of cities and regions and in particular investigating how the role of inventor networks is greater in places with high variety or complexity. A first work for example is that of van der Wouden and Rigby (2019) that connects the debate on knowledge networks with the diversified structure of cities. They investigate the role of the structure of networks of inventors in diversified and specialized cities in the US. They find that the structure of knowledge network is relevant for innovation, in particular in diversified cities with high relatedness. In addition, Rigby (2018) underlines that the variety of territories is an important element that influences inventors' networks in fostering innovation. They both emphasize the potential of this stream of research for the future. Two other contributions (Capone et al., 2021; Innocenti et al., 2020) investigate the role of inventors' networks for innovation moderated by the related/unrelated variety of the local industrial structure. Giannini et al. (2019) in the same line presents an investigation on textile and clothing industry, pointing out that unrelated variety is more important than related variety in explaining the innovation performance in the industry.[3]

Finally, the main novelty of the paper is to investigate the role of the structure of inventors' networks *and* its variety, not (only) for inventions in general, but also the so-called breakthrough inventions. In this respect, the literature suggests that these forces may have a different impact on the production of inventions and the "quality" of these inventions. Several contributions focus on the connection between variety and breakthrough innovations (Castaldi & Los, 2012; Silverberg & Verspagen, 2007). The hypothesis is that local related variety enhances the overall innovation rate and therefore recombination or incremental innovation, while local unrelated variety supports the rate of breakthrough innovations. Frenken et al. (2007) underline that regions with higher related variety, benefiting from Jacob externalities, can more easily produce recombination and therefore incremental innovation. On the other side, regions with high unrelated variety are expected to outperform regions with a low level of unrelated variety in producing breakthrough inventions (Castaldi et al., 2015). In this context, Castaldi and Los (2012) analyse the geography of patents in the US and found that regional clustering of breakthrough patents is higher than general patents, while Ejermo (2009) underlines that regional innovative performance in breakthrough patents is explained by different hypotheses than explaining regional innovative performance in general. Silverberg and Verspagen (2007) analyse the distribution of breakthrough and general patents, underlining that the distribution of highly cited patents implies a Pareto distribution of citations and highly cited patents can be considered as technological developments in new technological paradigms.

The present work contributes to the above debates investigating whether the structure of knowledge networks *and* the variety of the local industrial structure are important for regional innovation and in particular on the capacity to produce *breakthrough inventions*.

2.1 | Hypothesis of the research

As the above literature review shows, there is an increasing number of contributions showing that the structure of local knowledge networks and local variety is important for facilitating invention in cities and regions (Capone et al., 2021; Innocenti et al., 2020; Van der Wouden, 2020; van der Wouden & Rigby 2019).

We adopt this approach in our paper for investigating the structure of local inventors' networks and variety for overall innovation and breakthrough innovation. As said, we expect differentiated results for overall and breakthrough innovations. Regarding the structure of the local network, we mainly focus on three aspects related to the structure of inventors' networks: *density*, *cohesion* and *social proximity* for overall innovation and breakthrough inventions.

2.1.1 | Density

In innovation studies, there is an open debate on the role of networks' density for innovation. From a perspective, a more connected network (dense) should promote the knowledge flows within its structure. In fact, according to Coleman (1990) a dense and connected network would facilitate knowledge exchange and diffusion.

However, first Granovetter (1973) and later Burt (2000) asserted that (over) connected networks would become stale and less creative because they would circulate only redundant knowledge (the well-known "strength of weak ties") (Granovetter, 1973). Creating links with cognitively distant actors is crucial for (radical) innovations and the acquisition of new knowledge. Some previous works have found that density is negatively correlated with overall innovation (Galaso & Kovářík, 2021; Innocenti et al., 2020). This could be also expected for breakthrough innovations where too much density of the network could impede knowledge transfer and exchange also for radical inventions.

Hypothesis 1. A more dense regional inventors' network will be negatively correlated with overall invention and breakthrough innovations.

2.1.2 | Cohesion

According to the previous discussion, van der Wounden and Rigby (2019) assert in particular that co-patents networks are disconnected networks and network measures like (density) centrality behave poorly for them. Moreover, they underline that size and density are found to interact strongly with network measures. Consequently, they advise using some measure of network internal cohesion to investigate networks of different sizes and density, and surpassing the problem of disconnected networks.

One of the major cohesion measurement of Social Network Analysis (SNA) is the identification of cohesive subgroups of actors within a network (Wasserman & Faust, 1994). Cohesive subgroups are subsets of actors among whom there are relatively strong, direct, intense, frequent or positive ties. One expects greater homogeneity among organizations who have relatively frequent context and direct contact and less homogeneity among persons who have less contact. Several authors underline the importance to investigate the cohesion of sub-groups of inventors' networks focusing on actor's micro-behaviours and the sub-structure of networks (Lobo & Strumsky, 2008; Strumsky & Thill, 2013). These studies have all confirmed that a more cohesive network will be positively correlated with overall inventions, this can also be expected for breakthrough innovation as strong, direct, intense, frequent or positive ties can also support the production of radical innovations. A higher impact of cohesive inventors' networks could be expected for breakthrough innovation production than for overall innovation.

Hypothesis 2. A more cohesive inventors network will be positive correlated with local invention and breakthrough innovation.

2.1.3 | Social proximity

Finally, social proximity is another process (at micro level) that can help to promote knowledge exchange and innovation. Several authors underline that social ties and trust can help to collaborate and to be more innovative (Boschma, 2005; Broekel & Boschma, 2012; etc.). In particular, on the knowledge network of inventors, Breschi and Lenzi (2016) confirm that internal social proximity supports knowledge flows and the renewal of the local knowledge base.

Social proximity, trust and consequently mutual relationships, would promote local innovation. This means that collaborations are promoted among partners of partners and follow trust and social proximity processes (Breschi & Lenzi, 2016). Several authors have investigated social proximity in inventors' networks, not always confirming a positive relation between social proximity and innovation. It is then expected that higher social proximity will be positively correlated with overall invention production. Regarding breakthrough innovation, radical inventions can be considered as developed outside from the usual connection of social proximity, favouring more weak ties and structural holes and unusual relations (Burt, 2000; Granovetter, 1973). We then expect that social proximity could be negatively correlated with breakthrough innovations.

Hypothesis 3. An inventor networks with higher social proximity will be positively correlated with overall inventions, while it will be negatively correlated with breakthrough inventions.

2.1.4 | Related and unrelated variety

As already discussed in the Introduction, since the 1990s, many studies investigated the role of knowledge networks on the innovativeness of cities or regions. And even if more recently, one other relevant stream of research gathered growing attention looking at the effect of diversity and variety for innovativeness (Aarstad et al., 2016; Castaldi et al., 2015; Miguelez & Moreno, 2017) and particularly the role of industrial variety (Tavassoli & Carbonara, 2014; Leppälä, 2020). However, what seems to be under-investigated is the interaction between the two forces and the overlaps among these two streams of research. While from a theoretical perspective the connection between the inventors network and the industrial structure of the area for the innovative capacity is quite straightforward (Boschma & Frenken, 2010), the empirical applications are lacking and this stream started to develop only in the last very few years (Capone et al., 2021; van der Vouden & Rigby, 2019).

The literature regarding variety and particularly industrial variety as a relevant force able to favour the growth of territories in terms of employment, GDP, innovativeness, etc., dates back in time (Glaeser et al., 1992). However, it experienced extraordinary growth since the seminal work of Frenken et al. (2007), where the authors, starting from the idea of diversity (Jacobs, 1969) suggested that this diversity of spatially close firms able to promote transfers of knowledge and growth of productivity may be additionally divided in two different measures, namely related variety and unrelated variety. Regarding the first part, related variety, they suggest that the most interesting outcomes in terms of growth, innovation and learning are achieved when firms and plants in an area have a degree of variety allowing exchange and cross-connections about their technological knowledge bases. Related variety, should capture a definite degree of cognitive proximity, and thus measures the presence of sectors that are not too close nor too distant from each other (Boschma et al. 2012).

While many works already analysed the effect of related variety for the growth of territories in terms of employment, productivity and GDP (Bishop & Gripaios, 2010; Frenken et al., 2007), few previous works investigated the effect of related variety on the production of patents (Aarstad et al., 2016; Castaldi et al., 2015). However, different perspectives are leading to point out the attention to the diversification into related industries as an element capable to favour innovation. Several scholars suggest that to produce innovative ideas a fair degree of cognitive proximity is needed, this will allow the knowledge to spill-over between related sectors, favouring the recombination of knowledge into new processes, products and favouring innovations (Boschma et al. 2012; Content & Frenken, 2016). However, what is expected when the sectors that are present in an area are related, is that the innovation produced tends to be incremental rather than radical (Solheim et al., 2018). Nevertheless, given the fact that the interaction among related sectors is expected to favour the insurgence of incremental innovations and these tend to be more frequent than the radical, we may expect related variety to favour the general innovativeness of the area. Given the aforementioned reasoning, we put forward the following hypothesis:

Hypothesis 4. Regions characterized by a higher level of related variety are associated with higher overall patens production.

As discussed, related variety is considered to favour the growth and innovativeness of the area, even if not in term of breakthrough innovations. The unrelated variety defines the degree to which cognitively distant industries do not share the same competencies and belong to different technological fields that are present in the region. During the first period of the development of these measures and concepts, this force was considered mainly as an indicator of portfolio effect, thus able to defend the area from sector-specific external shocks. For this reason, the main applications of this measure were in studies on the resistance and resilience of territories to external shocks (Nyström, 2018; Sedita et al., 2017). However, more recently following the idea that it is in terms of knowledge spill-overs between different sectors that growth and radical innovations are encouraged. While knowledge spill-overs within the same sector only foster incremental innovations (Broekel & Boschma, 2012), started to rise the idea that unrelated variety may act not only as a force of portfolio effect but also as a determinant of growth in the long run (Bishop & Gripaios, 2010; Grillitsch et al., 2018; Pinheiro et al., 2018; Saviotti & Frenken, 2008).

The few empirical works aimed at disentangling the effect of unrelated variety for the innovativeness of the area leads to different results, Tavassoli and Carbonara (2014) find positive effects of unrelated variety in terms of absolute production of patents. While other authors (Castaldi et al., 2015; Miguelez & Moreno, 2017; Solheim et al., 2018) find no evidence of this association, but strong evidence of the positive impact of diversification into unrelated industries for the production of breakthrough innovations. The studies on the effect of variety also highlight how, more technologically diversified regions will yield better performances, as a result of the transmission of innovations and knowledge between firms belonging to different sectors. What is of import then, are cross-fertilization and cross-cutting processes, resulting from the interplay of ideas that follow different technological trajectories. The combinations of technologies coming from unrelated domains tend to fail more easily, however when they are successful tend to have a more radical dimension concerning innovations coming by the interaction of related fields (Castaldi et al., 2015; Saviotti & Frenken, 2008). In a given territory, the capacity to generate radical innovations can be established when there are sufficiently different enterprises since variety allows the exchange of knowledge coming from different sectors and stimulates the generation of new ideas leading to new applications new paths of development and new functions (Grillitsch et al., 2018; Saviotti & Frenken, 2008).

Hypothesis 5. Regions characterized by a higher level of unrelated variety are associated with higher production of breakthrough patents.

Finally, an additional reflection can be developed on how network properties and variety may interact influencing overall and breakthrough innovations. The intention here is to understand if the effect of inventors' networks' measures varies on the basis of the industrial variety.

Regarding density, it generally negatively influences innovation, as density is associated with homogeneous and recurrent knowledge flows. In general, we expect this homogeneity to be particularly negative when the territory has low industrial variety (less negative with high variety). Moreover, for breakthrough innovations, then a combination of high density and low variety could be particularly detrimental for innovation.

Regarding cohesion, we expect that the positive effect of cohesion decrease with a higher level of (unrelated) variety. For overall innovation, the cohesion of inventors' networks would be more important in territory with low industrial variety, while it would be less relevant with a higher level of variety. For radical innovations, casual and new relationships are more important than usual and habitual ties. Therefore, the cohesion of inventors' networks decreases in supporting the exchange of knowledge as the (unrelated) variety in the area increases. We expect that the positive effect of cohesion decreases with a higher level of unrelated variety.

3 | DATA SOURCE AND EXPLANATION OF THE DATABASE

The main sources of data to build the networks of inventors are the HistPat database (Petralia et al., 2017), which contains all the historical patent data granted by the United States Patent and Trademark Office (USPTO) for the period 1836–1975 and the database of Li et al., (2014) which contains disambiguated information on inventors for the period 1975–2010.

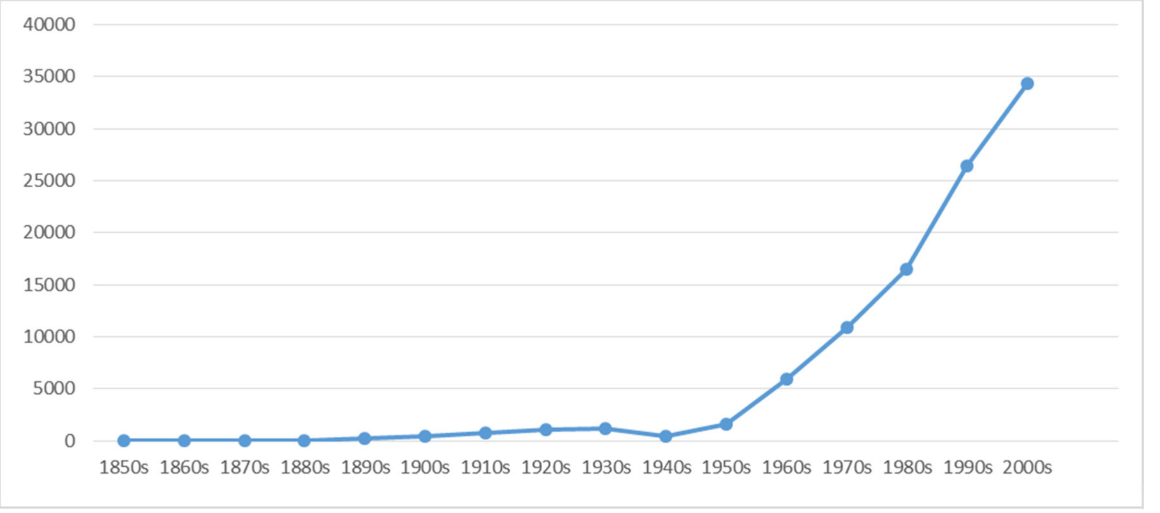
First, we included detailed geographical information of patents granted to Italian inventors residing in Italy during the period. Patents realized by multiple inventors located across different provinces have been fractionally split across areas.[4]

Figure 1, represents the evolution of the patents granted by USPTO where at least one inventor is Italian. The trend is similar to the general trend of USPTO (Van der Wouden, 2020) with an exponential increase in the last 50 years. It is interesting also to note that Italian inventors are deeply involved in applying for patents in the US as the last numbers show. The database registers more than 5,000 patents in the 1960s reaching almost 35,000 in the 2000s.

Based on this increase starting from the 1970s, and since employment data before 1981 are not homogenous and available, we decided to focus the empirical analysis on the period 1972–2011. Then for the following step, we manually disambiguated[5] the data regarding the inventors for the period 1972–1974 coming from the HistPat database to have four periods of ten years of disambiguated data that permitted us to investigate collaboration among inventors and inventors' networks.

It is then possible to investigate the geography of inventions realized by Italian inventors in USPTO (Figure 2). A clear pattern of regional differentiation is evident in terms of the three major areas of the country (North, Centre, and South). Besides the cities of the industrial triangle, the other provinces characterized by high densities of patenting activities are urbanized provinces with large populations (Rome, Palermo, Naples, etc.). Initially, the distribution of patents is strongly concentrated in the North and the northern provinces of the Centre (especially in Tuscany). The industrial triangle (Turin, Genoa, and Milan) is already clearly delineated, although in a somewhat embryonic shape.

In the beginning, the distribution of patents is rather skewed with few provinces Turin, Genoa, Milan, Florence, etc., holding the major bulk of patents. Then an increasing spread of patenting activities is happening with a process of spatial diffusion and growing concentration of the patenting activities in a few selected areas of the country. The main concentration is the industrial triangle, which becomes clearly visible.



Source: Author's elaboration on USPTO data.

FIGURE 1 Patent granted by USPTO per decade where at least one inventor is Italian *Source*: Author's elaboration on USPTO data

4 | THE ROLE OF THE STRUCTURE OF INVENTORS' NETWORKS AND VARIETY FOR OVERALL INNOVATION AND BREAKTHROUGH INVENTIONS

This section analyses more than 83,000 patents developed by Italian inventors in the USPTO in the period 1972–2011.

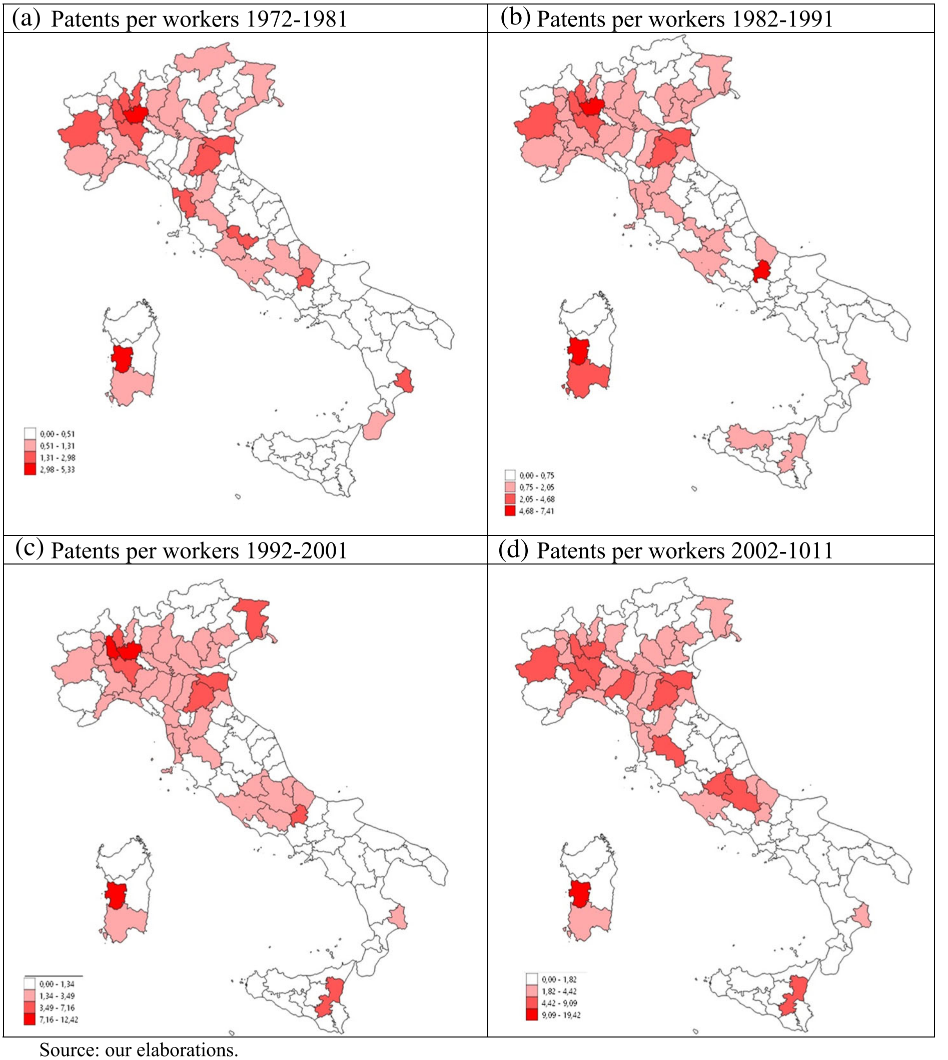
As discussed above, we divided the considered period into four sub-periods to analyse knowledge networks and diversity using a dynamic perspective. We chose four periods of the same length: 1972–1981, 1982–1991, 1992–2001 and 2002–2011. The choice of these four periods each of them closing in the year of Italian Census (1981, 1991, 2001 and 2011) is connected to the larger availability of data at the year of the Census. Data that are used to build the other independent and control variables.

Then we constructed the inventors' networks for each Italian province (103) in the four periods using the disambiguated ID of the inventor.[6] In these networks, "nodes" are inventors and the linkages are collaborations on patents. Our final database was composed of 412 (103 in each period) networks of more than 310,000 nodes as inventors located in the Italian provinces.

4.1 | Dependent variable

As already discussed, the aim of the paper is twofold. First, to investigate the effect of inventors network and industrial structure on the overall innovation production and second, on the capacity to produce those that are defined as breakthrough inventions. For this reason, we will test two different dependent variables. The first one is the count of all patents and is defined as the simple number of the patents produced during the following four three-years intervals (1982–1984, 1992–1994, 2002–2004, 2012–2014) for every Italian province. While the second dependent variable, represent a subset of the first one, is limited to the breakthrough patents produced by the Italian provinces during the same four three-year intervals.

To define breakthrough patents we followed the literature that suggests identifying these patents using the number of citations received and considering the top cited (Ahuja & Morris Lampert, 2001; Castaldi et al., 2015;



Source: our elaborations.

FIGURE 2 Patents productivity per workers along time in Italy 1972–2011 *Source*: Author's elaboration on USPTO data

Singh & Fleming, 2010). In this context, citations received by a patent can be considered as an indicator of the patent value (Trajtenberg, 1990). Of course, this is a tricky issue, and to facilitate a fair distribution of patents falling in different technological categories it is important to identify these top-cited patents separately for each technological category and divided by years intervals. The average number of citations largely differ among technological categories and also among periods.[7]

We then measured the top cited patents (5%) separately for every 2 digit class in each five-year interval following the approach of several authors (Ahuja & Morris Lampert, 2001; Singh & Fleming, 2010).

4.2 | Operationalization of the independent and control variables

The hypotheses of this study suggest that the local invention capacity is driven by how knowledge flows within the local inventors' networks. The structure of inventors' networks, then, can promote or impede the production of inventions. This will be then fostered, or hindered, by the variety and diversity of the local industrial structure concerning overall patents and breakthrough patents production.

The structure of the inventors network is analysed from different measures of social network analysis (Wasserman & Faust, 1994). First, the connectedness of the network is measured by its density. As Granovetter (1973) asserts, (over) connected networks would become stale and less creative because they would circulate only redundant knowledge. As said, according to this hypothesis, a higher level of network density will be associated with a negative effect on invention production (Table 1).

Second, we investigate the cohesion of the subgroups structure of the inventors network. As mentioned, cohesive subgroups may have a better exchange of knowledge and a higher level of innovation. A method in social network analysis to measure cohesive sub-groups is the use of *k*-cores (Wasserman & Faust, 1994). A *k*-core is a simple tool for identifying well-connected structures within large graphs. It is a way of grouping nodes, all of which are connected to some number (*k*) of other nodes; in other words, they share the same degree of *k* (Seidman, 1983).

We calculate the *k* value of the k-core with the highest value (max *k*-core). We expect that the max *k*-core values will be associated with higher invention production in a province. Max *k*-core would express the maximum number of repeated collaborations within the province of *k*-cores, which would indicate that there is a more cohesive structure to the inventors network. This would support a higher local innovation capacity. Also, other works apply *k*-cores to measure network cohesion (Innocenti et al., 2020; van der Wouden & Rigby, 2019).

TABLE 1 Hypotheses of the research

Variables	Proxy (hypothesis)	Overall innovations	Breakthrough innovations
Hypothesis 1: Density of the network	Density (−)	A more dense regional inventors network will be negatively correlated with overall invention and also breakthrough innovation	
Hypothesis 2: Cohesion	K-core index (+)	A more cohesive inventors network will be positive correlated with local invention and breakthrough innovation	
			A higher cohesion will be expected for breakthrough innovations than overall patents
Hypothesis 3: Social proximity	Triadic closure (+/−)	An inventor networks with higher social proximity will be positively correlated with overall inventions	An inventor networks with higher social proximity will be negatively correlated with breakthrough inventions
Hypothesis 4: Related variety	Related variety (+)	Positive effect on the overall patents	
Hypothesis 5: Unrelated variety	Unrelated variety (+)		Positive effect on the production of breakthrough patents

Source: our elaboration.

Finally, social proximity can also be relevant for knowledge exchange and innovation. Generally speaking, social proximity could be measured based on family, kinship of friendship ties (Boschma, 2005). As we do not have a direct measure of social proximity between inventors, we measure social proximity with triadic closure. In general, transitivity refers to the extent to which the tie which relates two nodes in a network that are connected by an edge is transitive. In an undirected network (as co-patenting), the transitivity coefficient corresponds to the fraction of triads that are closed, which is the fraction of pairs of inventors with common ties to another inventor. This means that collaborations are promoted among partners of partners and follow trust and social proximity processes. Besides, other works adopt triadic closure for social proximity (Ter Wal, 2014). A higher level of triadic closure will be expected to promote the local inventive capacity.

Regarding external linkages, we calculate external co-patenting collaborations between local inventors and inventors located in other provinces. We elaborate the share of patenting collaborations with inventors located outside the province. To capture the effect of agglomeration economies of inventors and the clustering of inventors, we calculate the share of inventors on population. We expect that a higher concentration of inventors will promote local patenting activity (Lobo & Strumsky, 2008).

Then, to study the effect of the industrial structure, we calculate the related and unrelated variety indexes for each province (Frenken et al., 2007). For this purpose, we use the entropy measures following the rules adopted in Frenken et al. (2007) and Hartog et al. (2012). The related and unrelated elements are the decompositions of the variety, that is, the sum of the entropy at the four-digit level (following the NACE industrial classification). The unrelated variety is then measured as the total amount of entropy at the two-digit level, following the idea that sectors which do not share the same two digits are unrelated to each other. A higher value of this index indicates that the province is largely diversified in dissimilar industrial sectors. The other element, called related variety, is the weighted sum of the entropy within each two-digit sector. In this case, the idea is that sectors belonging to the same two-digit class are technologically related to each other, so the interactions among them are easier and the sharing of knowledge is favoured. In this vein, a higher level of this index favours knowledge spillovers between industrial sectors and thus innovation (Hartog et al. 2012).

Several control variables are included in the empirical models. The first one is gross domestic product (GDP) *per capita*, a measure available in Eurostat, used to capture the economic development of each province. GDP can be considered as a proxy for the general economic prosperity of a province (Lacalle-Calderon et al., 2017).

The following variables are drawn from the Italian National Institute of Statistics (ISTAT) Census. The first is a variable controlling for the concentration of medium-large firms in the province, which is added to control for the stronger capacity to produce patents recognized in larger firms (Gambardella, 2021; Mazouz & Zhao, 2019). It is measured here as the proportion of firms with more than 50 employees present in a province divided by the same measure at the country level. We also control our estimates for the accumulation of human capital per province, proxied by the percentage of residents with at least a tertiary education qualification.

The next variable included in the models is population density, used to control for urbanization levels and measured as the population and area ratio of provinces.

Finally, a variable controlling for R&D expenditure is also included in the model. Since this variable is not available at the provincial level (NUTS 3) for the entire period, we used regional data (NUTS 2) available in EUROSTAT Regional database and rescaled the values for the working forces of the provinces.[8] For this reason, the outcomes need to be evaluated carefully. Table 2 presents the variables used in the study.

4.3 | Panel analysis

In this section, the estimation results are presented utilizing panel regression techniques with year and province fixed effects. This method is used to identify the relationship between patent production and different measures of inventor networks structure as well as the effect of industrial variety of the province.

TABLE 2 The variables in the study

Dependent variables	Explanations
Patents$_{(1982–1984), (1992–1994), (2002–2004), (2012–2014)}$	Patents in the period
Breakthroughs$_{(1982–1984), (1992–1994), (2002–2004), (2012–2014)}$	Breakthroughs patents of the period.
Independent and control variables	**Explanations**
Density$_{(1972–1981), (1982–1991), (1992–2001), (2002–2011)}$	Density of the network
Transitivity $_{(1972–1981), (1982–1991), (1992–2001), (2002–2011)}$	Share of triads that are closed
Max k-core$_{(1972–1981), (1982–1991), (1992–2001), (2002–2011)}$	*k* value of the *k-core* with the highest value
Share external ties$_{(1972–1981), (1982–1991), (1992–2001), (2002–2011)}$	Share of collaboration with firms located outside the province
Inventors per population$_{(1981, 1991, 2001, 2011)}$	Share of inventors on population
Related variety$_{(1991, 2001, 2011)}$	$\sum_{g=1}^{G} Pg\, Hg$
Unrelated variety$_{(1991, 2001, 2011)}$	$\sum_{g=1}^{G} Pg \log_2 \left(\frac{1}{Pg}\right)$
GDP *per capita*$_{(1991, 2001, 2011)}$	GDP *per capita* at constant prices in EUR
Human Capital$_{(1981, 1991, 2001, 2011)}$	Share of residents with tertiary education or higher
Medium-large firms concentration$_{(1991, 2001, 2011)}$	Proportion of firms with more than 50 workers in a province divided by the same measure at the country level
Population density$_{(1981, 1991, 2001, 2011)}$	Population on province's area (squared km)
Expenditure in R&D$_{(1985, 1991, 2001, 2011)}$	Expenditure in R&D of the province

Source: our elaboration.

However, the characteristics of the dependent variable (discrete and non-negative) require the use of estimation methods appropriate for count data (Hausman et al., 1984) and, thus, falling in the area of the Poisson family.

Utilizing the Poisson regression model requires that the mean of the data is restrained to be equal to the variance (Demidenko, 2013; Hilbe, 2011), which means $\mathrm{Var}(Y_i) = \mathrm{E}(Y_i) = \mu i$. While in this case, the strong variability of the number of patents between provinces called for additional tests on the over-dispersion of the dependent variable. Hence, we conducted a likelihood ratio test, as reported at the bottom of each model; these tests clearly indicated the NBR might be expected to perform better than the Poisson estimator (Greene, 2003).

Mathematically, the estimated model takes the following form:

$$Y_{i,t+3} = \alpha_i + \lambda_i + \beta_1 Density_{i,t} + \beta_2 Transitivity_{i,t} + \beta_3 MaxKcores_{i,t} + \beta_4 Depth_{i,t} + \beta_5 ShareInventors_{i,t} + \beta_6 RelVar_{i,t} + \beta_7 UnrelVar_{i,t} + \beta_8 Controls_{i,t} + \epsilon_{i,t}$$

where $y_{i,t+3}$ represents the dependent variable, which is the number of patents (breakthrough patents) produced during the four three-years intervals (1982–1984, 1992–1994, 2002–2004, 2012–2014) for every Italian province *i*.

The independent variables are computed at the beginning of each period (1981, 1991, 2001 and 2011). Finally, α_i represents the province dummies, λ_i represents the time dummies and $\epsilon_{i,t}$ the error term. To enable comparisons between the effects of the variables, these are standardized at zero mean and unit standard deviation.

The use of negative binomial panel regression is also helpful to avoid endogeneity problems and reverse causality between the dependent and independent variables. Finally, most of the acknowledged variables which may influence the patent production are also included to avoid misspecification of the model due to missing variable bias. In addition as a robustness check, we perform the full model using a scaled dependent variable number of patents/ breakthrough patents divided by the population of the province and in this case, we utilized panel regression techniques with year and province as fixed effects which showed similar results.

We also controlled for the collinearity of the models, and each one report the mean of the variance inflation factor (VIF) measure, showing that collinearity is not an issue for our results. All the values are lower than 5 and the means are all lower than 3, largely below the suggested threshold of 10 (Neter et al., 1989).

The estimation of the models followed a stepwise approach. The models report standardized coefficient to facilitate comparison, only independent variables are standardized since the dependent variable is a count variable. Table 3 reports a selection of the most representative models using as dependent variable the count of all the patents granted, while Table 4 show the results of the same models using as dependent variable, the count of the breakthrough patents.

The first model (model (1) in Table 3) is the simplest and aims to show the relationship between the knowledge networks variables and the overall production of patents. The density of the network is significant and negative, meaning that a denser and more connected network reduces the capacity to produce patents in a province. Similar results were found by Lobo and Strumsky (2008) and Innocenti et al. (2020) pointing out that in high density networks, re-circulation of existing, and even redundant, knowledge may occur and therefore hindering innovation.

The highest positive coefficient among the networks variable is *Max K-cores*, while transitivity is not significant as well as *External ties* (co-patent activities with other provinces). *Max K-cores* indicates that the cohesion of subgroups plays an important role in favouring the overall production of patents.

Among the control variables, it is worth mentioning the positive role played by the share of inventors, meaning that the agglomeration economies effect of inventors favours the growth of patent production showing similar results to those found by Strumsky and Thill (2013). While *population density*, *human capital* and *R&D* are not significant. Model (2) reflects the results of the previous model except for the population density variable that turns out to be significant and positive. The newly added control variable related to the concentration of medium-large firms, is significant and positive meaning that the provinces showing a higher concentration of medium large firms tend to produce more patents, while the *GDP per capita* is not significant.

Model (3) introduces related and unrelated variety variables. Focusing on the two newly included variables, it is interesting to note that only related variety is significant and positively associated with the overall production of patents, while unrelated variety is significant in only two models. With the addition of these industrial variables, also GDP and R&D show positive and significant results, meaning that regions with higher values are capable of producing more patents.

The following four models (models (4)–(7)) are aimed at the addition of interaction terms. The intention here is to understand if the effect of inventors' networks measures depends on the level of the industrial variety. Model (5) is the only one with a significant (and negative) interaction term that is *Max K-cores* and *Unrelated variety*. As expected, the positive effect of *Max-K-cores* decreases with a higher level of unrelated variety. This is in line with the fact that for overall innovation, related variety is more relevant and also the role of the cohesion of inventors' networks reduces its importance with a higher level of unrelated variety.

In conclusion Hypothesis 4 is then confirmed, a higher level of related variety is positively correlated with the overall patents production, with a positive role of the structure of inventors' networks.

Table 4 presents the estimations on the breakthrough patents. These models are organized as in Table 3. However, there are some relevant and interesting differences, density is significant and negative, but only in the first two models in Table 4 ((1)–(2)). So Hypothesis 1 is partly confirmed as in the last models on breakthrough patents density is not significant.

Transitivity is significant and negative in all the models ((1)–(7)). This confirms Hypothesis 3 only for breakthrough patents, while transitivity was not significant in overall patents production (Table 3). The negative sign of transitivity and social proximity for breakthrough patents, points out that for the production of radical innovations, casual and new relationships are more important than the usual connections of the traditional inventors network, favouring more *weak ties* and unusual relations (Granovetter, 1973).

Max k-cores shows to be significant and positive in all the models, confirming the role of cohesion of inventors' networks also for breakthrough patents production. The coefficients are in general higher in the overall patents

TABLE 3 Estimations on overall patens production (all patents)

Variables	Model (1) NBR	Model (2) NBR	Model (3) NBR	Model (4) NBR	Model (5) NBR	Model (6) NBR	Model (7) NBR
Density	−0.0806° (0.0463)	−0.117° (0.0627)	−0.152* (0.0638)	−0.264* (0.106)	−0.133* (0.0652)	−0.142* (0.0615)	−0.142* (0.0646)
Transitivity	0.00283 (0.0430)	−0.0315 (0.0517)	−0.0674 (0.0459)	−0.0696 (0.0450)	−0.0283 (0.0496)	−0.0714 (0.0459)	−0.0682 (0.0459)
Max K-cores	0.201*** (0.0504)	0.237*** (0.0615)	0.265*** (0.0517)	0.270*** (0.0508)	0.328*** (0.0565)	0.260*** (0.0516)	0.255*** (0.0535)
Share of external ties (*depth*)	−0.0336 (0.0326)	−0.00852 (0.0400)	0.0188 (0.0409)	0.0183 (0.0409)	0.0249 (0.0396)	0.0254 (0.0411)	0.0177 (0.0410)
Share of Inventors	0.309*** (0.0482)	0.412*** (0.0520)	0.381*** (0.0431)	0.374*** (0.0421)	0.343*** (0.0497)	0.375*** (0.0432)	0.376*** (0.0436)
Pop density	0.0699 (0.0442)	0.116° (0.0644)	0.0593 (0.0509)	0.0659 (0.0496)	0.0452 (0.0570)	0.0587 (0.0502)	0.0606 (0.0514)
Human capital	0.0401 (0.0551)	0.0923 (0.0589)	0.0212 (0.0533)	0.0146 (0.0500)	−0.0272 (0.0577)	0.0156 (0.0527)	0.0179 (0.0539)
R&D	0.0113 (0.0629)	−0.0417 (0.0789)	0.166° (0.0892)	0.185* (0.0877)	−0.00796 (0.0779)	0.173* (0.0768)	0.168* (0.0784)
Medium-large firms		0.432* (0.232)	0.455* (0.257)	0.421° (0.254)	0.0958 (0.225)	0.383 (0.256)	0.367 (0.256)
GDP *per capita*		0.167 (0.482)	0.123* (0.0501)	0.0995* (0.0506)	0.325*** (0.0581)	0.133** (0.0500)	0.144* (0.0589)
Rel. Var.			0.134** (0.0417)	0.132** (0.0404)	0.0607° (0.0368)	0.131** (0.0415)	0.121** (0.0459)
Unrel. Var.			0.127 (0.0772)	0.139* (0.0628)	0.111* (0.0485)	0.127 (0.0863)	0.124 (0.0971)
Density*Unrel. Var.				−0.0760 (0.0521)			
Max K-cores*Unrel. Var.					−0.216*** (0.0381)		
Density*Rel. Var.						−0.0702 (0.0614)	
Max K-cores*Rel. Var.							0.0233 (0.0349)
Constant	0.556*** (0.130)	1.259*** (0.295)	1.066*** (0.270)	1.086*** (0.268)	1.191*** (0.245)	1.095*** (0.269)	1.086** (0.272)
Obs.	412	309	309	309	309	309	309
Log Likelihood	1747.79	−1420.46	−1416.25	−1415.267	−1399.76	−1415.53	−1416.03
LR test on overdispersion	158.83***	70.21***	35.50***	29.96***	48.13***	36.73***	28.24***
VIF (mean)	1.49	1.71	1.72	1.94	1.78	1.77	1.77
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors are in parenthesis.
***p < 0.01. **p < 0.05. *p < 0.1.

TABLE 4 Estimations on breakthrough patents

	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)	Model (7)
Variables	NBR	NBR	NBR	NBR	NBR	NBR	NBR
Density	−0.444* (0.178)	−0.313° (0.160)	−0.207 (0.144)	−0.274 (0.205)	−0.202 (0.158)	−0.201 (0.143)	−0.198 (0.145)
Transitivity	−0.141* (0.0690)	−0.191**(0.0731)	−0.166* (0.0709)	−0.165*(0.0706)	−0.161*(0.0701)	−0.170* (0.0714)	−0.168*(0.0715)
Max K-cores	0.299*** (0.0757)	0.238** (0.0808)	0.172* (0.0764)	0.174* (0.0760)	0.285*** (0.0799)	0.167* (0.0765)	0.164* (0.0812)
Share of external ties (*depth*)	−0.0460 (0.0762)	0.0188 (0.0816)	0.00600 (0.0827)	0.00456 (0.0826)	0.0215 (0.0852)	0.0122 (0.0831)	0.00856 (0.0829)
Share of Inventors	0.401*** (0.0639)	0.446*** (0.0710)	0.357*** (0.0660)	0.354*** (0.0660)	0.326*** (0.0620)	0.350*** (0.0668)	0.355*** (0.0668)
Pop density	0.0734 (0.0649)	0.140° (0.0740)	0.0859 (0.0734)	0.0877 (0.0732)	0.0855 (0.0693)	0.0842 (0.0732)	0.0847 (0.0741)
Human capital	−0.0122 (0.0686)	−0.0680 (0.0721)	−0.0403 (0.0668)	−0.0400 (0.0665)	−0.106 (0.0653)	−0.0408 (0.0668)	−0.0403 (0.0674)
R&D	0.189** (0.0703)	0.206** (0.0704)	0.262*** (0.0569)	0.260*** (0.0561)	0.266*** (0.0523)	0.263*** (0.0570)	0.262*** (0.0587)
Medium-large firms		0.357 (0.391)	0.125 (0.394)	0.119 (0.394)	0.00498 (0.385)	0.0913 (0.396)	0.124 (0.393)
GDP *per capita*		0.134° (0.0686)	0.0847 (0.0792)	0.0760 (0.0811)	0.197* (0.0864)	0.0954 (0.0806)	0.0997 (0.0922)
Rel. Var.			0.147 (0.0957)	0.142 (0.0946)	0.105° (0.0623)	0.130 (0.0893)	0.137 (0.0878)
Unrel. Var.			0.272** (0.0951)	0.264** (0.0961)	0.369*** (0.0967)	0.273**(0.0954)	0.270** (0.0957)
Density*Unel. Var.				−0.0584 (0.120)			
Max K-cores*Unrel. Var.					−0.182*** (0.0522)		
Density*Rel. Var.						−0.108 (0.156)	
Max K-cores*Rel. Var.							0.0180 (0.0562)
Constant	−0.739* (0.294)	−0.240 (0.470)	−0.0398 (0.451)	−0.0438 (0.450)	0.0452 (0.446)	−0.00118 (0.452)	−0.0108 (0.460)
Obs.	412	309	309	309	309	309	309
Log Likelihood	−627.35	−531.32	−522.92	−522.81	−516.73	−522.66	−522.87
LR test on overdispersion	43.45***	24.43***	18.22***	17.27***	13.93**	18.61***	15.28***
VIF (mean)	1.49	1.71	1.72	1.94	1.78	1.77	1.77
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors are in parenthesis.
***p < 0.01. **p < 0.05. *p < 0.1.

productions models (Table 3), but unfortunately, it is not possible to directly compare the magnitude of the parameters in the two models. However, following the results, also Hypothesis 2 can be confirmed.

Regarding the control variables, share of inventors is still significant and positive showing the highest coefficient, meaning that the agglomeration effect of inventors also favours the growth of breakthrough patents production. Also in the same direction, R&D expenditure shows to be significant and positive in all models, underlining that R&D is more important for radical innovation than for combinatorial or incremental innovation as shown by the mixed results of this variable in Table 3.

Conversely, in this case, no significance is detected in any model related to the concentration of medium-large firms, meaning that concerning the overall production of patents, for breakthrough the size of firms in the province is less relevant.

Starting form model (3) in Table 4 the related and unrelated variety variables are introduced. The main difference with the previous set of models is that unrelated variety is now significant and positive, while related variety is significant only in one model (5). Hypothesis 5 is then confirmed underlining that a higher level with unrelated variety is positively correlated with the breakthrough patents.

The following four models (models (4)–(7)) are aimed at the addition of interaction terms of knowledge networks measures and related/unrelated variety. Model (5) is still the only one with a significant (and negative) interaction term, that is, *Max K-cores* and *Unrelated variety*. As expected, the positive effect of *Max-K-cores* decreases with a higher level of unrelated variety. The cohesion of inventors network decreases in supporting the exchange of knowledge as the unrelated variety in the area increases. This confirms the fact that for radical innovations casual and new relationships are more important than usual and habitual ties.

5 | CONCLUSIONS

In this paper, we investigated how the structure of the local inventors' network and industrial diversity affect the regular and breakthrough patenting activity. We expected that the role of the structure of inventors' networks and diversity influence the overall patents and breakthrough patents production, but in a differentiated way following the articulation of the proposed hypotheses.

Results show a clear association between both, the characteristics of the inventor network and industrial diversity of Italian provinces with respect to the production of regular and breakthrough patents. More precisely, our results suggest that the structure of knowledge networks are positively correlated with the innovative capacity, with particular reference to the cohesion of inventor networks, while social proximity offers a different, and precisely negative, contribution for breakthrough innovation.

However, the most interesting result concerns the analysis of the related and unrelated variety. A high degree of related variety mainly affects the overall patents production, permitting recombination of local existing (related) knowledge. While high values of unrelated variety favour the production of breakthrough innovation allowing the connection of unrelated local knowledge. Similar results are also found by previous works on breakthrough inventions such as Castaldi et al. (2015) or Castaldi and Los (2012). However, to the best of our knowledge, this is the first work that links this phenomenon with inventors' networks.

In this context, social proximity (measured as triadic closure) shows a different result and appears to have a (negative) impact mainly in breakthrough innovation and in regions with a high degree of unrelated variety. In this situation, it is not the recurring relationships that promote the creation of radical innovations, but rather the unusual, distant ties, in other words, "the strength of weak ties" seems to prevail here. Another important variable of the network structure is related to the cohesion of the inventor networks, as confirmed also by other works on the topic (Capone et al., 2021; Innocenti et al., 2020; van der Wouden & Rigby, 2019), increasing the possibility of developing innovations in general and also breakthrough inventions. At a preliminary analysis, it seems that cohesion and the hierarchical structure of local inventors' networks are more relevant for the overall production of innovations, while for breakthrough innovations seems that there are also other mechanisms in place that potentially cannot be grasped with the analysis carried out. Other authors have also pointed out that explaining regional performance in terms of breakthrough innovation requires a different hypothesis than explaining regional innovative performance in general terms (Ejermo, 2009).

In conclusion, therefore, the general hypotheses of a positive contribution of network structure and diversity to the overall production of innovations and breakthrough innovations are confirmed and diversified. This certainly opens up new interesting questions which will require further efforts in the future.

It is then possible to underline also some limits of our work. First of all, the use of the Italian case and the patents included in the USPTO limit the study. An extension of the study to other European countries would allow further generalizability of the results even if here it would be necessary to make an advance on the EPO database through disambiguation algorithms of the inventors. Furthermore, even from the point of view of the operationalization of the variables, it is possible to develop some further advances. Perhaps one of the most important, concerns the result of social proximity (triadic closure), which should be confirmed with more precise measurements of social phenomena, for example through kinship, parental bonds, etc. Unfortunately, this is not possible for analysis on patents however, a positive result of social proximity could be hypothesized at least for the overall patents' production as other authors reach through other different types of measurement (Breschi & Lenzi, 2016).

Finally, most literature on these themes prefers to measure diversity according to technological knowledge bases (based for instance on patents IPC) and through the use of more complex methods (i.e., relatedness). Our choice was instead to focus on the diversity of the local structure of the regions and for our point of view, it is a strength of the paper, however, further development of the work could be realized in this direction.

In conclusion, this study represents an interesting advancement in trying to differentiate the role of knowledge networks and diversity for overall and breakthrough innovations. Further studies in the future are certainly necessary, but this line of research seems to be interesting and promising since this first work.

ORCID

Niccolò Innocenti https://orcid.org/0000-0001-8421-5479
Francesco Capone https://orcid.org/0000-0003-2000-3033
Luciana Lazzeretti https://orcid.org/0000-0002-9759-2289
Sergio Petralia https://orcid.org/0000-0001-5449-2758

ENDNOTES

[1] The choice of using the USPTO database is motivated by two main reasons. First, the patenting in the US being more complex and expensive can allow considering only the best innovations of the Italian inventors. Secondly, it is the only database that presents homogeneous data, partly geo-referenced and disambiguated for such a long period (Petralia et al., 2017).

[2] However, all these works do not have geographically referenced patents, while in this work we are able to geo-localise the Italian inventors (in USPTO) place of residents in Italy and realise an analysis at the provincial level in Italy.

[3] Giannini et al. (2019) point out that these results differ from most of the empirical literature that stresses the importance of related variety for the innovative performance of firms. They believe that one of the main problems is how related and unrelated varieties are defined and measured.

[4] This was done according to the inventors' residence and deliberately disregarding the location of the patent applicant (as in Ter Wal, 2014). Large companies usually assign patents to their headquarters, even when the patent is developed in one of the company's subsidiaries. Notwithstanding the possibility that some inventors might live in another region than where they work, the inventor location is generally agreed to be a more reliable approximation of where the invention was developed (Acs et al., 2002; Ejermo & Karlsson, 2006).

[5] We have followed the most used approach to disambiguate the inventors, which usually apply a disambiguation algorithm (Ventura et al., 2015; Van der Wouden, 2020). Investigating firstly if the inventor in the period 1972–1974 were also present in the disambiguated database of 1975–2010 and then verified each inventor patent production in the decade, analysing if two inventors with the same name work in the same address and within the same subclass of IPC. If two

inventors with the same name, same location and same subclass of IPC worked in two different patents were identified as the same person. This was possible as patents developed by Italian inventors in the 3 years were only 2,800 patents with 4,500 inventors.

[6] As the database by Li et al. (2014) regards the period 1975–2011, we manually disambiguated the inventors in the period 1972–1974.

[7] These sub-populations of patents are usually selected as cohorts of patents in a technological field to provide a fair comparison between patents of the same age.

[8] The data of R&D in 1981 was not available and we used the data in 1985.

REFERENCES

Aarstad, J., Kvitastein, O. A., & Jakobsen, S.-E. (2016). Related and unrelated variety as regional drivers of enterprise productivity and innovation: A multilevel study. *Research Policy*, *45*(4), 844–856. https://doi.org/10.1016/j.respol.2016.01.013

Acs, Z. J, Anselin, L., & Varga, A. (2002). Patents and innovation counts as measures of regional production of new knowledge. *Research Policy*, *31*(7), 1069–1085. https://doi.org/10.1016/s0048-7333(01)00184-6

Ahuja, G., & Morris Lampert, C. (2001). Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions. *Strategic Management Journal*, *22*(6–7), 521–543. https://doi.org/10.1002/smj.176

Audretsch, D. B., & Feldman, M. P. (1996). R&D spillovers and the geography of innovation and production. *The American Economic Review*, *86*(3), 630–640.

Balland, P. A., & Rigby, D. (2017). The geography of complex knowledge. *Economic Geography*, *93*(1), 1–23. https://doi.org/10.1080/00130095.2016.1205947

Bishop, P., & Gripaios, P. (2010). Spatial Externalities, Relatedness and Sector Employment Growth in Great Britain. *Regional Studies*, *44*(4), 443–454. https://doi.org/10.1080/00343400802508810

Boix, R., & Trullén, J. (2007). Knowledge, networks of cities and growth in regional urban systems. *Papers in Regional Science*, *86*(4), 551–574. https://doi.org/10.1111/j.1435-5957.2007.00139.x

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

Boschma, R., & Frenken, K. (2009). Technological relatedness and regional branching. In H. Bathelt, M. Feldman, & D. Kogler (Eds.), *Beyond territory: Dynamic Geographies of Knowledge Creation and Innovation*. (64–68). Routledge.

Boschma, R., & Frenken, K. (2010). The spatial evolution of innovation networks. A proximity perspective. In R. Boschma & R. Martin (Eds.), *The handbook of evolutionary economic geography* (pp. 120–135). Edward Elgar Publishing.

Boschma, R., & Martin, R. (2010). The aims and scope of evolutionary economic geography. In R. Boschma & R. Martin (Eds.), *The handbook of evolutionary economic geography* (pp. 3–40). Edward Elgar Publishing.

Boschma, R., & Ter Wal, A. (2007). Knowledge networks and innovative performance in an industrial district: The case of a footwear district in the South of Italy. *Industry and Innovation*, *14*(2), 177–199. https://doi.org/10.1080/13662710701253441

Boschma, R., Minondo, A., & Navarro, M. (2012). Related variety and regional growth in Spain. *Papers in Regional Science*, *91*(2), 241–256.

Breschi, S., & Lenzi, C. (2014). The role of external linkages and gatekeepers for the renewal and expansion of US cities' knowledge base, 1990–2004. *Regional Studies*, *49*(5), 782–797. https://doi.org/10.1080/00343404.2014.954534

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

Broekel, T., & Boschma, R. (2012). Knowledge networks in the Dutch aviation industry: the proximity paradox. *Journal of Economic Geography*, *12*(2), 409–433. https://doi.org/10.1093/jeg/lbr010

Burt, R. S. (2000). The network structure of social capital. *Research in Organizational Behavior*, *22*, 345–423. https://doi.org/10.1016/S0191-3085(00)22009-1

Capello, R. (2016). What makes Southern Italy still lagging behind? A diachronic perspective of theories and approaches. *European Planning Studies*, *24*(4), 668–686. https://doi.org/10.1080/09654313.2015.1128402

Capone, F., Lazzeretti, L., & Innocenti, N. (2021). Innovation and diversity: The role of knowledge networks in the inventive capacity of cities. *Small Business Economics*, *56*(2), 773–788. https://doi.org/10.1007/s11187-019-00268-0

Castaldi, C., Frenken, K., & Los, B. (2015). Related variety, unrelated variety and technological breakthroughs: An analysis of US state-level patenting. *Regional Studies*, *49*(5), 767–781. https://doi.org/10.1080/00343404.2014.940305

Castaldi, C., & Los, B. (2012). Are new 'Silicon Valleys' emerging? The changing distribution of superstar patents across US states. In *Danish Research Unit for Industrial Dynamics (DRUID) Summer Conference*.

Coffano, M., & Tarasconi, G. (2014). CRIOS-Patstat database: sources, contents and access rules. Center for Research on Innovation, Organization and Strategy, CRIOS working paper, (*1*), DOI: https://doi.org/10.2139/ssrn.2404344

Coleman, J. S. (1990). Foundations of social theory. Harvard University Press: Cambridge, MA.

Content, J., & Frenken, K. (2016). Related variety and economic development: a literature review. *European Planning Studies*, *24*(12), 2097–2112. https://doi.org/10.1080/09654313.2016.1246517

Crescenzi, R., Nathan, M., & Rodríguez-Pose, A. (2016). Do inventors talk to strangers? On proximity and collaborative knowledge creation. *Research Policy*, *45*(1), 177–194. https://doi.org/10.1016/j.respol.2015.07.003

Demidenko, E. (2013). *Mixed models: Theory and applications with R*. John Wiley & Sons.

Ebersberger, B., Herstad, S. J., & Koller, C. (2014). Does the composition of regional knowledge bases influence extra-regional collaboration for innovation? *Applied Economics Letters*, *21*(3), 201–204. https://doi.org/10.1080/13504851.2013.848019

Ejermo, O. (2009). Regional Innovation Measured by Patent Data—Does Quality Matter? *Research Paper. Industry and Innovation*, *16*(2), 141–165.

Ejermo, O., & Karlsson, C. (2006). Interregional inventor networks as studied by patent coinventorships. *Research Policy*, *35*(3), 412–430. https://doi.org/10.1016/j.respol.2006.01.001

Fagerberg, J. (1988). International competitiveness. *Economic Journal*, *98*, 355–374. https://doi.org/10.2307/2233372

Fagerberg, J., & Srholec, M. (2008). National innovation systems, capabilities and economic development. *Research Policy*, *37*(9), 1417–1435. https://doi.org/10.1016/j.respol.2008.06.003

Feldman, M. P., & Kogler, D. (2010). Stylized facts in the geography of innovation. In B. H. Hall & N. Rosenberg (Eds.), *Handbook of the economics of innovation* (pp. 381–410). Elsevier. https://doi.org/10.1016/S0169-7218(10)01008-7

Fleming, L., King, C., & Juda, A. (2007). Small worlds and regional innovation. *Organization Science*, *18*(6), 938–954. https://doi.org/10.1287/orsc.1070.0289

Fleming, L., Mingo, S., & Chen, D. (2007). Collaborative brokerage, generative creativity, and creative success. *Administrative Science Quarterly*, *52*(3), 443–475. https://doi.org/10.2189/asqu.52.3.443

Frenken, K., Van Oort, F., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies*, *41*(5), 685–697. https://doi.org/10.1080/00343400601120296

Galaso, P., & Kovářík, J. (2021). Collaboration networks, geography and innovation: Local and national embeddedness. *Papers in Regional Science*, *100*(2), 349–377. https://doi.org/10.1111/pirs.12578

Gambardella, A. (2021). *The functions of patents in our societies: innovation, markets, and new firms*. Available at SSRN 3789554.

Giannini, V., Iacobucci, D., & Perugini, F. (2019). Local variety and innovation performance in the EU textile and clothing industry. *Economics of Innovation and New Technology*, *28*(8), 841–857. https://doi.org/10.1080/10438599.2019.1571668

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

Granovetter, M. S. (1973). The Strength of Weak Ties. *American Journal of Sociology*, *78*(6), 1360–1380. https://doi.org/10.1086/225469

Greene, M. (2003). *Econometric Analysis*. New York, NY: Prentice Hall.

Grillitsch, M., Asheim, B., & Trippl, M. (2018). Unrelated knowledge combinations: The unexplored potential for regional industrial path development. *Cambridge Journal of Regions, Economy and Society*, *11*(2), 257–274. https://doi.org/10.1093/cjres/rsy012

Hartog, M., Boschma, R., & Sotarauta, M. (2012). The Impact of Related Variety on Regional Employment Growth in Finland 1993–2006: High-Tech versus Medium/Low-Tech. *Industry & Innovation*, *19*(6), 459–476. https://doi.org/10.1080/13662716.2012.718874

Hausman, J., Hall, B., & Griliches, Z. (1984). Econometric models for count data with an application to the patents-R&D relationship. *Econometrica*, *52*(4), 909–938. https://doi.org/10.2307/1911191

Hermans, F. (2021). The contribution of statistical network models to the study of clusters and their evolution. *Papers in Regional Science*, *100*(2), 379–403. https://doi.org/10.1111/pirs.12579

Hidalgo, C., & Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, *106*, 10570–10575. https://doi.org/10.1073/pnas.0900943106

Hilbe, J. (2011). *Negative Binomial Regression*. Cambridge: Cambridge University Press.

Innocenti, I., Capone, F., & Lazzeretti, L. (2020). Knowledge networks and industrial structure for regional innovation. An analysis of patents collaborations in Italy. *Papers in Regional Science*, *9*(1), 55–72. https://doi.org/10.1111/pirs.12478

Jacobs, J. (1969). The economy of cities. New York: Vintage.

Lacalle-Calderon, M., Perez-Trujillo, M., & Neira, I. (2017). Fertility and economic development: Quantile regression evidence on the inverse J-shaped pattern. *European Journal of Population*, *33*(1), 1–31. https://doi.org/10.1007/s10680-016-9382-4

Leppälä, S. (2020). Innovation, R&D spillovers, and the variety and concentration of the local industry structure*. *The Scandinavian Journal of Economics*, *122*(3), 1231–1255. https://doi.org/10.1111/sjoe.12368

Li, G.-C., Lai, R., D'Amour, A., Doolin, D. M., Sun, Y., Torvik, V. I., Yu, A. Z., & Fleming, L. (2014). Disambiguation and co-authorship networks of the US patent inventor database (1975– 2010). *Research Policy*, *43*(6), 941–955. https://doi.org/10.1016/j.respol.2014.01.012

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

Malmberg, A., & Maskell, P. (2006). Localized learning revisited. *Growth and Change*, *37*(1), 1–18. https://doi.org/10.1111/j.1468-2257.2006.00302.x

Mazouz, K., & Zhao, Y. (2019). CEO incentives, takeover protection and corporate innovation. *British Journal of Management*, *30*(2), 494–515. https://doi.org/10.1111/1467-8551.12330

Miguelez, E., & Moreno, R. (2017). Networks, diffusion of knowledge, and regional innovation performance. *International Regional Science Review*, *40*, 331–336.

Morescalchi, A., Pammolli, F., Penner, O., Petersen Alexander, M., & Riccaboni, M. (2013). *Networks of innovators within and across borders. Evidence from patent data. IMT Lucca EIC Working Paper 04/2013*.

Morrison, G., Riccaboni, M., & Pammolli, F. (2017). Disambiguation of patent inventors and assignees using high-resolution geolocation data. *Scientific Data*, *4*, 170064. https://doi.org/10.1038/sdata.2017.64

Neter, J., Wasserman, W., & Kutner, M. (1989). *Applied linear regression models*. Homewood, IL: Irwin.

Nuvolari, A., & Vasta, M. (2015). The ghost in the attic? The Italian national innovation system in historical perspective, 1861-2011. *Enterprise and Society*, *16*, 290–270. https://doi.org/10.1353/ens.2015.0042

Nuvolari, A., & Vasta, M. (2017). The geography of innovation in Italy 1861-1913 evidence from patent data. *European Review of Economic History*, *21*, 326–356. https://doi.org/10.1093/ereh/hex006

Nyström, K. (2018). Regional resilience to displacements. *Regional Studies*, *52*(1), 4–22. https://doi.org/10.1080/00343404.2016.1262944

Petralia, S., Balland, P., & Morrison, A. (2017). Climbing the ladder of technological development. *Research Policy*, *46*(5), 956–969. https://doi.org/10.1016/j.respol.2017.03.012

Pezzoni, M., Lissoni, F., & Tarasconi, G. (2014). How to kill inventors: Testing the Massacrator© algorithm for inventor disambiguation. *Scientometrics*, *101*(1), 477–504. https://doi.org/10.1007/s11192-014-1375-7

Pinheiro, F. L., Alshamsi, A., Hartmann, D., Boschma, R., & Hidalgo, C. (2018). *Shooting high or low: Do countries benefit from entering unrelated activities?* Working Paper. Cambridge, MA: MIT Media Lab.

Powell, W. (1990). Neither market nor hierarchy. *The Sociology of Organizations: Classic, Contemporary, and Critical Readings*, *315*, 104–117.

Powell, W. W., Koput, K. W., & Smith-Doerr, L. (1996). Inter-organizational collaboration and the locus of innovation: networks of learning in biotechnology. *Administrative Science Quarterly*, *41*, 116–145. https://doi.org/10.2307/2393988

Rigby, D. (2018). You have to (Br)enter to (Br)exit: The EU collaboration space. *Environment and Planning a*, *50*, 493–496. https://doi.org/10.1177/0308518X17750878

Saviotti, P. P., & Frenken, K. (2008). Export variety and the economic performance of countries. *Journal of Evolutionary Economics*, *18*, 201–218. https://doi.org/10.1007/s00191-007-0081-5

Sedita, S. R., De Noni, I., & Pilotti, L. (2017). Out of the crisis: An empirical investigation of place-specific determinants of economic resilience. *European Planning Studies*, *25*(2), 155–180. https://doi.org/10.1080/09654313.2016.1261804

Seidman, S. B. (1983). Network structure and minimum degree. *Social networks*, *5*(3), 269–287.

Silverberg, G., & Verspagen, B. (2007). The size distribution of innovations revisited: An application of extreme value statistics to citation and value measures of patent significance. *Journal of Econometrics*, *139*(2), 318–339. https://doi.org/10.1016/j.jeconom.2006.10.017

Singh, J. (2005). Collaborative networks as determinants of knowledge diffusion patterns. *Management Science*, *51*(5), 756–770. https://doi.org/10.1287/mnsc.1040.0349

Singh, J., & Fleming, L. (2010). Lone inventors as sources of breakthroughs: Myth or reality? *Management Science*, *56*(1), 41–56. https://doi.org/10.1287/mnsc.1090.1072

Solheim, M. C., Boschma, R., & Herstad, S. (2018). *Related variety, unrelated variety and the novelty content of firm innovation in urban and non-urban locations (No. 1836)*. Utrecht University, Department of Human Geography and Spatial Planning, Group Economic Geography.

Strumsky, D., & Thill, J. C. (2013). Profiling US metropolitan regions by their social research networks and regional economic performance. *Journal of Regional Science*, *53*(5), 813–833. https://doi.org/10.1111/jors.12048

Tavassoli, M., & Carbonara, N. (2014). The role of knowledge variety and intensity for regional innovation. *Small Business Economics*, *43*, 1–17. https://doi.org/10.1007/s11187-014-9547-7

Ter Wal, A. L. (2014). The dynamics of the inventor network in German biotechnology: Geographic proximity versus triadic closure. *Journal of Economic Geography*, *14*(3), 589–620. https://doi.org/10.1093/jeg/lbs063

Tiffin, M. A. (2014) *European productivity, innovation and competitiveness: The case of Italy*, International Monetary Fund Working Paper 14-79.
Trajtenberg, M. (1990). A penny for your quotes: patent citations and the value of innovations. *The Rand journal of economics*, 172–187.
van der Wouden, F., & Rigby, D. L. (2019). Co-inventor networks and knowledge production in specialized and diversified cities. *Papers in Regional Science*, *98*(4), 1833–1853. https://doi.org/10.1111/pirs.12432
Van der Wouden, F. (2020). A history of collaboration in US invention: Changing patterns of co-invention, complexity and geography. *Industrial and Corporate Change*, *29*, 599–619. https://doi.org/10.1093/icc/dtz058
Ventura, S. L., Nugent, R., & Fuchs, E. R. H. (2015). Seeing the non-stars: (Some) sources of bias in past disambiguation approaches and a new public tool leveraging labeled records. *Research Policy*, *44*(9), 1672–1701. https://doi.org/10.1016/j.respol.2014.12.010
Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications. Cambridge: Cambridge University Press.
Whittington, K. B. (2018). A tie is a tie? Gender and network positioning in life science inventor collaboration. *Research Policy*, *47*(2), 511–526. https://doi.org/10.1016/j.respol.2017.12.006
Whittington, K. B., Owen-Smith, J., & Powell, W. W. (2009). Networks, propinquity, and innovation in knowledge-intensive industries. *Administrative Science Quarterly*, *54*(1), 90–122. https://doi.org/10.2189/asqu.2009.54.1.90

How to cite this article: Innocenti, N., Capone, F., Lazzeretti, L., & Petralia, S. (2021). The role of inventors' networks and variety for breakthrough inventions. *Papers in Regional Science*, 1–21. https://doi.org/10.1111/pirs.12640