Extracting knowledge patterns with a social network analysis approach: an alternative methodology for assessing the impact of power inventors

Massimiliano Ferrara[1,2,4] · Roberto Mavilia[2,3] (iD) ·
Bruno Antonio Pansera[1,4]

Abstract This paper proposes a new, alternative analysis of patent data in order to extract knowledge patterns from inventors' collaboration networks. Indeed, moving from a basic network analysis, we provide new developments to map and study co-inventorship. The goal of this research is to provide an overall understanding of the dynamics concerning knowledge flows in inventive activities. We show how the network of inventors is, on average, increasing in size: more and more inventors are contributing to technology innovations and they are more connected to each other. We also show to what extent inventors from different countries tend to cooperate with their local peers or internationally. Furthermore, an analysis of the clustering of inventors is carried out to show differences across countries in the structure of inventors' communities, with a particular focus on the dynamics of collaboration for power inventors (i.e. star inventors).

✉ Roberto Mavilia
  roberto.mavilia@unibocconi.it

1   DIGIEC, Mediterranean University of Reggio Calabria, Reggio Calabria, Italy

2   ICRIOS, Bocconi University, Milan, Italy

3   MEDAlics, University for Foreigners "Dante Alighieri", Reggio Calabria, Italy

4   DECISIONS_Lab, Mediterranean University of Reggio Calabria, Reggio Calabria, Italy

Introduction

This paper proposes a new, alternative approach for the analysis of patent data based on social network analysis (SNA). Our quantitative assessment allows us to fully exploit the richness of patent data as well as to extract knowledge patterns from patent inventors' network of collaborations. Furthermore, our approach describes a methodology in order to address several issues related to inventive activities and it ultimately provides a more complete understanding of the patterns that characterize the geography of innovation.

In particular, the study proposes a new SNA-based approach (with simple and informative empirical applications) to extract knowledge patterns about patent inventors and their structure of relationships. Indeed, in our network, nodes represent inventors while edges denote co-inventorship relations. Moving from a basic network analysis, we provide new developments to map and study co-inventorship.

The advent of SNA has provided an important tool that helps researchers to better understand the impact of scientific contributions to the existing body of literature. Specifically, the analyses developed using citations data provide meaningful metrics for gauging the impact of scholars' contributions in a certain field (Griffin et al. 2016). Wasserman and Faust (1994) emphasized that the usefulness of SNA lies in the successful recognition of the existing relationships between social entities, models, and the associations between the two. From its introductory developments in sociology in the early 1970s (Wellman and Whitaker 1974) the use of SNA has increased significantly in popularity throughout the social sciences (Otte and Rousseau 2002). Since then, SNA has been applied in many disciplines, ranging from anthropology (Boissevain 1979) to zoology and animal biology (e.g., dolphin social networks; Lusseau 2003).

Recent advances in citation indexing and bibliometric studies provide new tools for network analysis methods and their applications across social sciences as well. For instance, communication researchers have widely relied on SNA to understand publications' patterns and trends in specific content areas such as health communication (Kim et al. 2010). Scientific publications have also been used to predict job placements in the communication sector (Feeley et al. 2011). As the current study illustrates, SNA is a tool that can be meaningfully incorporated into tailored methodology in order to rank authors. As an example, Griffin et al. (2016) chose to use eigenvector centrality rather than simply tally up the number of publications to create a ranking of authors. Kim et al. identify the structural patterns of theoretical relationships in communication sciences by using both webometric and SNA methods. In particular, they analyzed the networked structure of theories in social sciences represented by co-occurrences on the World Wide Web. To do so, co-occurrences in communication science theories were retrieved from the Web and analyzed using SNA tools. Although relational bibliometric usually examines relationships within science mainly through the Institute for Scientific Information (ISI), it has undergone a revolutionary change through the wide diffusion of new important sources of scientific communication such as scholars' webpages and online article databases (aka Google Scholar) (Delgado and Repiso 2013; Lee et al. 2009; Park 2010; Thelwall 2009). On the one hand, the use of webometrics as a new type of bibliometric source has flatly emerged. Webometrics is widely understood as the study of web-based contents and web phenomena through quantitative methods (Thelwall 2009). SNA methods, on the other hand, constitute a set of analytical techniques for understanding the structure of scholarly communication across many academic disciplines (Chung et al. 2013). More precisely, SNA is used to identify structural properties based on relationships or patterns of relationships among interacting components (Wasserman and Faust 1994). The intuition

behind such methodology is that the social ties among which actors are embedded in has important consequences for such actors indeed (Freeman 2004). Several scholars have argued that the combination of the SNA methods and bibliometric tools can provide a richer picture of the status of a given field (Leydesdor 2007). On top, webometrics offers interesting analytical developments to provide a more comprehensive understanding of the status of communication sciences. (Anderson 1996: 8).

Before going back to specific content of the current paper it is appropriate to very quickly review the notion of gatekeepers. Following Le Gallo and Plunket, we can adopt two different definitions: Giuliani and Bell (2005) and Morrison (2008) consider gate-keepers as local actors with strong connections outside their cluster and with a central position within their cluster, enabling them to exploit externally produced knowledge within their local context; instead Gould and Fernandez (1989) consider gatekeepers as individuals with unique and non-redundant ties to external actors (Breschi and Lenzi 2015; Graf and Kruger 2011).

Le Gallo and Plunket have further explored whether and how gatekeepers contribute to the performance of innovations within clusters. Using patent data and inventor networks in genomic they analyzed whether gatekeepers affect the quality of patents produced both by the teams which they belong to as well as by the teams which they are socially connected to, within regional networks. Hence they were able to separate for each individual its intra and inter-regional connections. This micro-level approach can help us to understand how the embeddedness within regional and global networks might influence inventive performance.

As a matter of fact, the analysis of patent data through SNA methods can represent a building block in investigating scientific developments and understanding the nature of innovative processes of a country. In addition, such analysis can lead policymakers to make better and more efficient investment choices. Finally, investigating inventions via patent data allows decision makers to monitor the evolution of R&D activities over time among different organizations, institutions and countries. By looking at patent networks, indeed, we may also provide a precise and detailed picture of cooperation occurring among different entities and/or countries in the R&D sector, an indicator of geopolitical devel-opments and possibly a forecast of future patterns of innovative activities.

In several fields the amount of data available to researchers and scholars has rapidly increased over the last few years. Patents data have been growing as well. To deal with this huge quantity of information, scientists have been relying more and more on a wide range of statistical methods that belong to the domain of data science and big-data analytics. Despite new statistical approaches are starting to spread, most of the analyses in the patent literature still rely on classical methodologies.

In fact, approaches based on classical statistical methods present several difficulties for the investigation of phenomena that involve large amount of data and where variables are strictly related to each other (Tsvetovat and Kouznetsov 2011); whereas SNA appears particularly adequate to it.[1]

Bearing in mind such methodological issues, differently from most other studies, we try to exploit methods drawn from data science to provide a meaningful and comprehensive elaboration of such large quantity of data. In particular, we propose a new, general approach to the extraction of knowledge patterns by looking at patent inventors and their relationships.

---

[1] See, for example, (Wasserman and Faust 1994; Barabási et al. 1999, 2002; Barabasi and Albert 1999; Albert and Barabási 2002; Clauset et al. 2004, 2009; Leicht et al. 2006; Newman and Leicht 2007).

Thanks to this approach we can describe many features of scientific and R&D collaboration among different inventors and across countries, characterize new measures and concepts, such as the notions of "power inventor" and "clique of inventors", as well as adapting to our analysis some SNA metrics already employed in different domains.

The paper is structured as follows. After reviewing the relevant literature in the following section, "Data, research questions and methodology" Section illustrates our data and methodology showing how we address several questions about inventors' networks and the geography of innovation. Furthermore, we present the results of our measurements and computations for a set of countries that in our opinion yield the most interesting and representative outcomes. Then, we briefly discuss our contribution to the literature. Finally, "Concluding remarks" Section concludes.[2]

Literature review

Since the seventies many researchers have expressed some considerations about the influence that many companies can have on research and on the fields of expertise. Just think of what was published by Crawford and Biderman (1970), Blumer (1967a, b), Green (1971), Horowitz (1967), McCartney (1971), Orlans (1967), Platt (1971), Rist (1973, 1975), Galliher and McCartney (1973), Furstenberg (1971) where the authors' relationship between sponsorship by various governmental agencies and commissions, as well as private foundations, has influenced the "objectivity" of scientific investigations.

Broadhead and Rist (1976) speculated and synthesized the effect of the sponsorship of the research activity, summarizing this action in three fundamental steps:

1. Through detailed specification of the research issue so that the eventual problem is cast within a framework congruent with the sponsor's perspective.
2. Through emphasis upon a positivistic style of research thought more susceptible to manipulation for the purposes of controlling the results.
3. Through the threat of withdrawing present funding and denying future support should the researcher move into areas "not in the best interest" of the sponsor.

They point out that: "The actual manifestations of social control over social research can be shown as they appear within organizational frameworks. A key component of that control organization is the small group of managers and administrators within a formal organization who screen prospective researchers seeking funding, entering into the organization itself, or accessing to data already collected. This small group of "gatekeepers" has a central role in deciding the fate of those who desire to conduct social research with someone else's money, data, or organization." (Broadhead and Rist 1976)

Recent studies have shown how the situation after forty years has not changed much. In Zsindely et al. (1982a, b) it was showed that, at least in chemistry, the editorial bodies of international journals are tree gatekeepers and their professional status is positively correlated with the scientific quality of the international journals in this discipline. The person of the editor-in-chief alone seems to have a much lesser in influence. They lead back this gap by the lower participation of scientists of certain countries in the editorial boards of

---

[2] There are also two appendixes in the additional material: Appendix A includes formal definitions of several concepts introduced in the main text, whereas Appendix B presents statistical facts about patents and inventors' collaboration in relation to technological sectors (which in our dataset are captured by IPC classes), not discussed in the main text due to space limitations.

international chemistry journals is frequently compensated by the high citation frequency of the editors coming from these countries. The higher than average citation frequency of editors from countries with scientific life of a more closed character shows that in order to receive an invitation to the editorial boards of international chemistry journals scientists from such countries must acquire a higher than average international prestige.

In Zsindely et al. authors point out the significant correlation between the number of science journal editors from different countries and the number of scientists. They argue using "the extent of participation in the editorial hoard of international science journals as a new science indicator. The deviations from the regression lines between the new publication indicator and other indicators allow one to assess the "open" or "closed" character of the scientific life of a given country". They also considered that the critical mentality and decisions of the editors have so far protected and will also warrant in the future the social and intellectual integrity of science. Moreover Crane (1965) rightfully considers the members of such boards the publication gatekeepers of a given field.

In the field of economics and management of innovation (EMIT) there is a strong evidence that research collaborations across firms and regions are a key factor for the acquisition of external knowledge (Singh 2005), as well as for promoting innovation (Miguélez and Moreno 2013).

Several papers analyze various aspects connected to researchers' professional relationships and their impact on innovative activities. With this regard our work aims at providing an original approach for the analysis of these relationships and for the extraction of knowledge patterns exploiting patent data. As already mentioned, the base of our analysis is on inventors and not applicants, as in Lissoni (2012). Differently though from Hingley and Bas (2009) for instance who investigate the evolution of both the number of applicants and the distribution of applicant size.

Patent data and innovation

Economic studies of innovation have for long made use of patent data (Griliches 1990; Nagaoka et al. 2010). Assisted by digitalization of records and increasing computational power economists and other social scientists have extracted increasing quantities of information from patent documents: the applicants' identity and location, the technological contents of the invention, or the latter's impact, as measured by citations. More recently information on inventors has attracted a good deal of attention. Identifying inventors allows studying their mobility patterns, both in space and across companies (Agrawal et al. 2006; Marx et al. 2009) as well as their social capital, as measured by their position in co-inventor networks (Fleming et al. 2007; Breschi and Lissoni 2009; Lissoni et al. 2010). Inventor data can also be matched to additional information at the individual level, ranging from professional identities (does the inventor appear also on a list of R&D employees or a list of academic scientists?) to other type of archival data on knowledge related activities (such as scientific publications; see Azoulay et al. 2009; Breschi et al. 2008; Lissoni et al. 2008). Identifying inventors within any given set of patent data, as well as matching them to any other list of individuals, requires the elaboration of complex disambiguation algorithms. They are necessary to analyze in a non-trivial way the text strings containing the inventor's names, surnames, and addresses. Yet, it is only of late that users of inventor data have started discussing openly about the disambiguation techniques they employ and examine their implications in terms of data quality and reliability of the evidence produced (Raffo and Lhuillery 2009; Li et al. 2014).

Over the past decade universities and public research organizations have devoted increasing attention to the issue of Intellectual Property (IP), as part of a general trend towards a greater commitment to work with industry, market their research results and promote entrepreneurship based on science. The Academic Patents in Europe (APE-INV) coordinated the collection of data on micro- inventors, scientists and patents. Moreover, it promoted the production of studies on a national basis across Europe. APE-INV has also promoted an operational notion of "Patent academia", centered on the identity of the inventor (the academic researcher) rather than that of the assignee (the Researcher University). Most academic patents are assigned to be commercial companies, both as a result of the search sponsorship or market transactions. APE-INV has also actively promoted a debate on the economic value of academic patents, the role of IP law and university autonomy in determining the propensity of universities to invest in patents, or relationship between the productivity of scientists and their activities related to patents.

Miguélez and Moreno (2013) document the existence of a positive correlation between cross-regional networking and innovation in Europe. Similarly, Breschi and Lissoni (2009) find that network activity across firms and locations is in large part responsible for the localization of knowledge flows. Nicholas (2009) analyses the relationship between inventors and the locations of R&D laboratories of firms, identifying a strong positive effect of distance from a laboratory where the average quality of externally available inventions is high.

Thelwal (2014) explores the relationship between geographic distance and triadic closure, which represent two important forces driving the evolution of collaboration networks. By analyzing the dynamics of inventor networks in German biotechnology sector she is able to show—both theoretically and empirically—that inventors increasingly rely on network resources, forming links with partners of partners. On the other hand, the direct impact of geographic distance on the formation of ties is decreasing: initially the triadic closure reinforces the effect of geographic distance by closing triads among proximate inventors, then it becomes an increasingly powerful vehicle to generate longer-distance collaboration ties, and so the effect of geographic proximity decreases.

Guellec and Pottelsberghe de la Potterie (2001) present three new patent-based indicators of technology internationalization that reflect international cooperation in research and the location of research facilities of multinational firms. Exploiting these indicators, they find that there is an increasing trend towards technological globalization in the OECD area and that Nordic countries have a particularly high propensity to collaborate together.

Ponds et al. (2010) investigate the effect of knowledge spillovers from academic research on regional innovation. They found that this effect is mediated not only by geographic proximity, but also by network stemming from university-industry collaboration. They also show that this type of collaboration is not limited to a regional scale.

Patent data and international collaboration

The importance of knowledge spillovers is well known in the literature (see, for instance, Griliches (1992). Especially at the international level they can play an important role in reducing the growing technological gap among the most and the least sophisticated economies. To investigate the impact of international knowledge flows and their effects many studies focused on international R&D collaborations, cross-border inventions (i.e., inventions involving people coming from different countries) and on their impact on innovation quality.

At the theoretical level, there are some authors who argued that these collaborations could induce higher-quality innovations thanks to the combinations of different skills and knowledge (Levinthal and March 1993; March 1991). Other authors suggest instead that international collaborations may be not efficient due to high coordination costs and difficulties in integrating knowledge coming from different research teams, Furman et al. (2006); Grant (1996); Singh (2008).

At the empirical level, results are mixed. Branstetter et al. (2014) investigate Indian and Chinese inventors and finds that cross-border inventions receive more citations - and, hence, are more valuable—than patents filed by inventors of only one country. Montobbio and Sterzi (2011) show that innovative activities of Latin American countries have largely benefited from R&D performed on some OECD countries. Recently, Giuliani et al. (2016) find that cross-border inventions between BRICS firms and EU actors are growing and are more valuable than domestic ones, representing an opportunity to accumulate technological capabilities. Alnuaimi et al. (2012) also finds that patent quality is positively influenced by international collaborations. However, it also evidences the difficulties of research teams in the absorption of external knowledge. Hence it is not clear whether such collaborations may enhance the accumulation of technological competencies in developing countries.

As for developed countries Kim et al. (2009) investigates the international transmission of knowledge in USA. It finds that the number of US firms that collaborate with researchers having foreign experience has increased. It also shows that these firms have an easier access to non-US technological know-how than firms not collaborating with inventors characterized by foreign research experience.

Menon (2014) also analyses the US scenario, and in particular the effect of top inventing companies on local inventors. His aim is to check whether the patenting activity of the most inventive companies have any causal effect on other local inventors: his results show that this effect is not bounded within narrow technological categories. Developing an original approach based on SNA, our work goes in this direction.

Patent data, international collaboration and SNA

The literature that studies patent data and dynamics of innovative activities using SNA is still small yet rapidly growing, indicating that it may be a promising and insightful approach. In an early study, Balconi et al. (2002) show that both classical and advanced contributions to the economics of knowledge—as well as most of the empirical research on contributions of universities to commercial innovation—can be investigated using SNA.

Lissoni and Miguelez (2014) study the geography of innovation analyzing patents. They show how localized knowledge flows are largely mediated by labor and technology markets and how social distance among inventors, along with inter- and intra-national borders, affects the diffusion of knowledge. Lubango (2015) finds that both the links and the H-indexes of co-inventors and co-authors highly enhanced the flows of academic knowledge into industrial patents in South African firms, as well as the diffusion of knowledge in large R&D and innovation clusters. Hsueh and Wang (2009) also propose a network-based analysis of patent data, applying it to the field of Liquid Crystal Display. They aim at understanding how knowledge flows spread among institutions and countries. In particular, their approach identifies the key influential players, knowledge spillovers patterns among them and the overall knowledge spillover efficiency. Wagner and Leydesdorff (2005) hypothesize that international collaboration is a self-organizing network, exploiting SNA to show that growth of international co-authorship can be explained on the

basis of the organizing principle of preferential attachment, even if the attachment mechanism deviates from an ideal power-law.

Furthermore, Chen and Guan (2016) use patent citations data and SNA to show that linkages among core and peripheral nodes are sparse, more so for linkages between peripheral nodes only. Linkages between core and peripheral nodes though are reported to be important to building innovation capabilities and technological catch-up.

Forti et al. (2013) analyze academic inventors' neighbors to investigate the contribution of social networks to the generation of inventive ideas within university. They show that the ego-networks of inventors are more cohesive than the ones of non-inventors. This is probably due to higher climate of trust. They also find that both inventors and non-inventors extend their network and become more central over time, as well as no evidence that after patenting inventors close their networks. Ejermo and Karlsson (2006) study the structure of interregional inventor networks in Sweden, showing that several factors influence the spatial affinity of regions. In particular, they show that spatial affinity extends beyond a region if it has less own R&D-related resources, it is close to the other region and it is relatively small.

Similar to our work, Sternitzke et al. (2008) analyze cooperation networks between inventors and applicants—in the field of optoelectronics—emphasizing bibliometric measures and network theoretical ones. Furthermore, they investigate "boundary spanner" inventors and, in a general way, the role played by applicant position within citation networks.

Another notable work using SNA was carried out by Cammarano et al. (2017). They focused on bio-pharmaceutical companies and point at a relationship between network structure and innovation strategies[3] In particular, their research finds a rationale behind R&D international collaborations with SNA validating the technique for innovation studies.

Wang et al. (2014) use patents data to show how inventor's network structure affects researchers' exploratory innovation. In particular, their study reveals a *U* shaped relationship between the average degree centrality of inventors and their exploratory attitude.

Building upon Wang et al. (2014), Tseng et al. (2016) focus on semiconductor patents data to investigate the relationship between inventors' network properties and their innovation capabilities. They show that firms that are more central in the network and have more linkages within the network are more likely to generate new knowledge.

Data, research questions and methodology

In this section, we present our approach, along with its support data structures and metrics, showing how it can address several questions related to patents and inventors and how it can ultimately provide an understanding of patterns about the geography of innovation.

An output of a EU-funded project (Id: 506022 under FP6-2002-CITIZENS-2 call) called Knowledge-Based Entrepreneurship: Innovation, Networks and Systems (KEINS) leaded by the Invernizzi Center for Research on Innovation, Organization, Strategy and Entrepreneurship at Bocconi University (ICRIOS) was the EP-INV database produced by ICRIOS which implemented the so called Massacrator algorithm (Pezzoni et al. 2014), containing all EPO applications, reclassified by applicant and inventor; and from three lists of university professors of all ranks (from assistant to full professors), one for each

---

[3] We thank an anonymous reviewer for pointing at this important reference.

countries (PROFLISTs). Academic inventors have been identified by matching names+-surnames of inventors in the EP-INV database with those in the PROFLISTs and by checking through the e-mail and phone the identity of the matches in order to exclude homonyms. This data-base was largely used world-wide, and in particular the methodology used to build a database on academic inventors from France, Italy, and Sweden (1978–2004), which was delivered to the European Commission as part of the KEINS project (Knowledge-Based Entrepreneurship: Innovation, Networks and Systems) and which will provide the basis for future publications was described in Lissoni et al. (2006a, b). It provides an overview of the database contents, as well as information on access rules and on related datasets by CESPRI-Università Bocconi. The database is the result of joint efforts by CESPRI—Bocconi University (IT) and other European universities; in Pezzoni et al. following Raffo and Lhuillery (2009) the authors described disambiguation as a three step process: cleaning & parsing, matching, and filtering. Based on Monte-Carlo simulations, by means of sensitivity analysis, we show how various filtering criteria can be manipulated in order to obtain optimal combinations of precision. They also show how these different combinations generate different results for applications to studies on inventors' productivity, mobility, and networking and discuss quality issues related to linguistics. The criteria are sensitive to data quality, while those based upon co-inventorship networks are always effective. Details on data access and data quality improvement via feedback collection are also discussed. In Lissoni et al. (2008) the authors described networks of inventors and the position of academic inventors in France; in particular, they built upon Balconi's methodology in order to map the networks of inventors in France. They also exploited the French section of the KEINS database complemented with the novel dataset on CNRS inventors collected following the KEINS methodology.

Patents have been one of the main topics investigated in several fields,[4] as they provide a wealth of useful information on the dynamics of innovative processes, on the state of art of the Research & Development (R&D) in different sectors and its main characters, i.e. researchers and inventors (Balconi et al. 2002; Breschi and Lissoni 2009; Guellec and Pottelsberghe de la Potterie 2001; Hingley and Bas 2009; Hsueh and Wang 2009; Singh 2008).[5] As a matter of fact, patent submission is usually the first public claim of a new invention that could become an innovation (Garcia and Calantone 2002).

The data used in this research comes from the PATSTAT-ICRIOS database. Hence, we have access to a large amount of information: the priority, the number of applications and publications and the related dates, the title and abstract, the designated states for protection, the status of application, the main and secondary International Patent Classification (IPC) codes, the applicants' and inventors' names and addresses, and the references (citations) to prior art patents and to non-patent literature.[6] As Lissoni (2012), we use data regarding inventors rather than applicants.

Thanks to the wealth of details about patents and inventors present in our dataset we can provide a first, intuitive answer to several questions:

---

[4] See, for example, Abbas et al. (2014), Lissoni and Miguelez (2014), Lubango (2015), Wagner and Leydesdorff (2005), Forti et al. (2013), Ejermo and Karlsson (2006), Sternitzke et al. (2008).

[5] See also the more recent contributions of Lissoni (2012), Landini et al. (2015), Miguélez and Moreno (2013) and Lubango (2015).

[6] We also have access to File Index (FI) concordance tables to convert IPC codes into more aggregated and manageable technological classes or Nomenclature of Units for Territorial Statistics (NUTS3).

(a)   What is the degree of cooperation among inventors coming from different countries?
(b)   Does it evolve over time?
(c)   Are there groups of inventors that cooperate intensively with each other?
(d)   Do countries maintain research collaborations with specific other countries?
(e)   Are some inventors more "important" than others?
(f)   Which is the distribution of patents against inventors and IPC classes?

To address these issues we consider a social network whose nodes and edges represent respectively inventors and co-inventorship relations. We use the universe of data, including all inventors filing for patent applications between 1980 and 2010 in all technology fields represented by the 2-digits technology classification according to WIPO (2011).[7] On top of that we define other support data structures and accompanying metrics—both establishing new ones and adapting those already existing to make them suitable for this type of analysis.

For instance, among those structures we introduce and characterize the concept of "power inventor", which aims at capturing those inventors that display a high impact in terms of both number of patents and number of collaborations—including at the international level—and hence have the potential to favor further developments and improvements of R&D activities within their countries. As another example we define a new metric to quantitatively understand how much inventors tend to cooperate—intensively—within compact groups.

In the rest of the paper we introduce several metrics and additional data structures in the additional material (more formally characterized in Appendix A), addressing different questions about inventors and patents and coming up with a series of statistical facts about knowledge patterns among the countries in our dataset.

Distribution of patents and inventors' network

Here we provide a characterization of patents and inventors present in our dataset, showing how patents are distributed among inventors for a large number of countries and how the inventors' network has evolved over time.

Distribution of patents among inventors

Given a node n_i in a social network G where the nodes correspond to inventors,[8] we define the metric M_1 such that M_{1i} denotes how many patents were filed by inventor n_i; this metric coincides with the classical weighted degree centrality (Hanneman and Riddle 2005). We measure M_1 for all countries in the European Union, all Mediterranean and North African countries, BRICS (Brazil, Russia, India, China, South Africa), South Korea, Japan, Vietnam and Taiwan.

For most of these countries M_1 follows a power law distribution, implying that there is a low number of inventors filing a high number of patents and large number of inventors filing a very low number of patents. Some other countries—i.e., Greece, Croatia, Principate of Monaco, Slovenia, Turkey and the BRICS—show a slightly disturbed power law distribution, as the highest number of inventors and filed patents is quite low. Other

---

countries—i.e. Egypt, Lebanon, Malta, Morocco and Tunisia—present a disturbed power law distribution. Again, this is due to the low number of inventors and filed patents in the corresponding countries. Finally, for some countries—i.e. Albania, Algeria, Libya and Montenegro—the distribution of M_1 is totally different from a power law. In some cases, even a linear distribution can be observed, as in the case of Algeria.[9]

Figure 1 displays the distribution of M_1 for France, Greece, Egypt and Algeria, each one representing the distribution of the different aforementioned categories of countries.
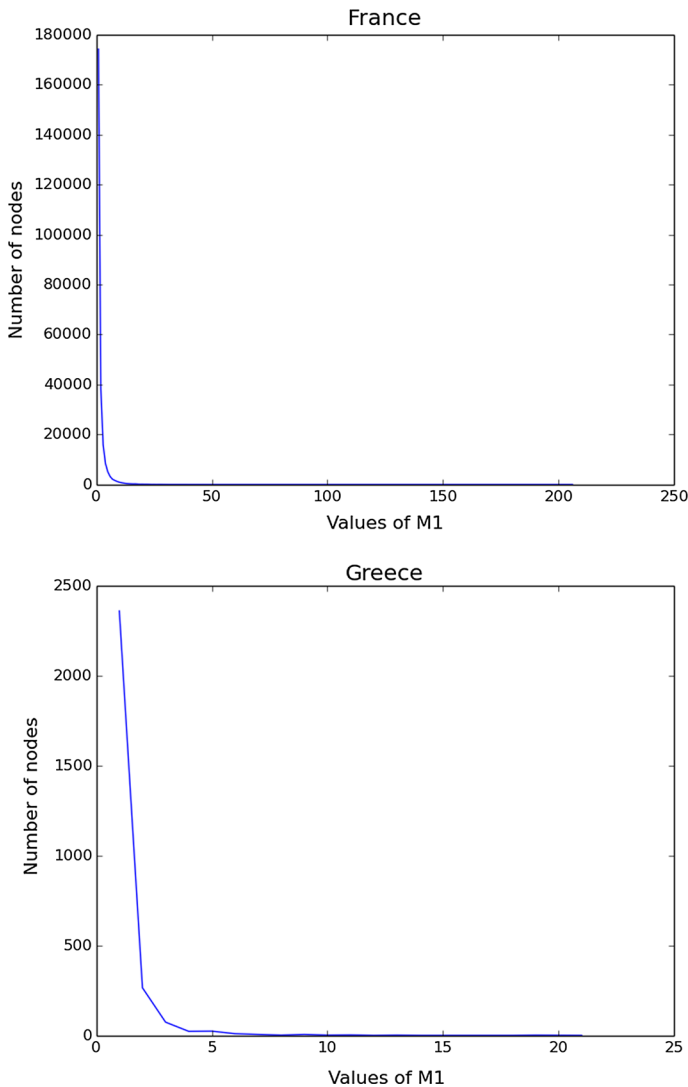

Fig. 1 Distribution of M_1 for selected countries

---

[9] For all those countries whose distribution of M_1 is totally different from a more or less disturbed power law the number of inventors and filed patents is so scarce to make any investigation about them unreliable and not really meaningful.
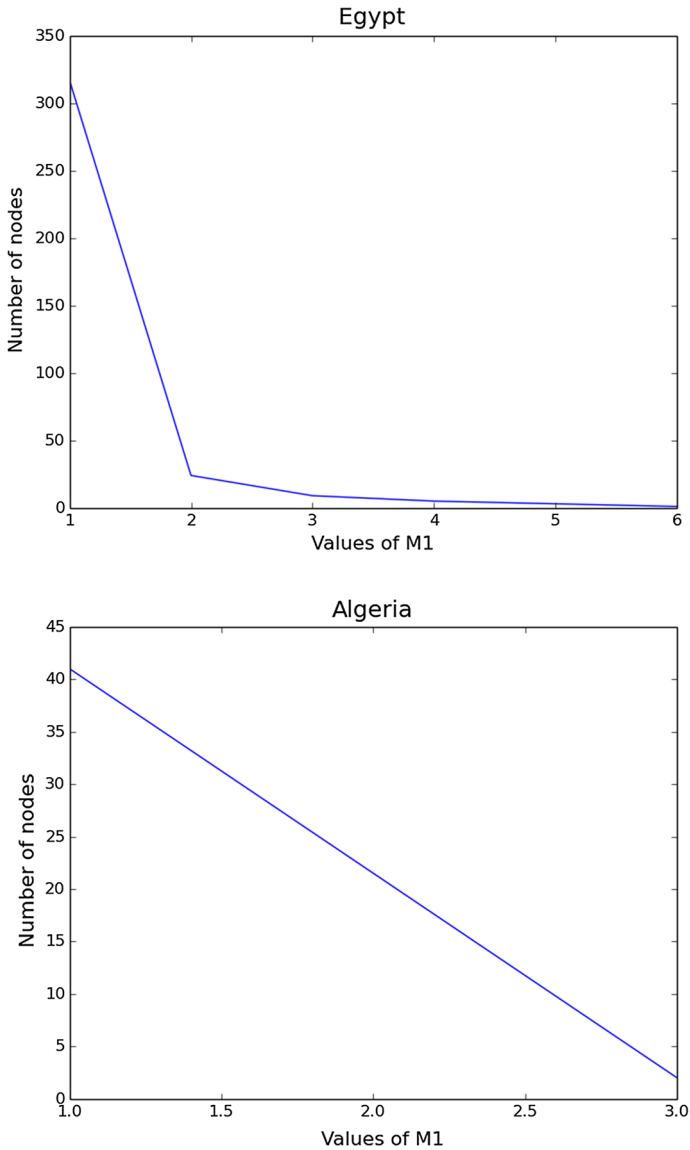
Fig. 1 continued

Evolution of inventors' networks over time

With the supports of social network G_k,[10] we now present the temporal evolution of the number of nodes $|N_k|$, which is the set of inventors of a given country k, the number of

---

[10] See Appendix A, Section A.4. in the additional material.
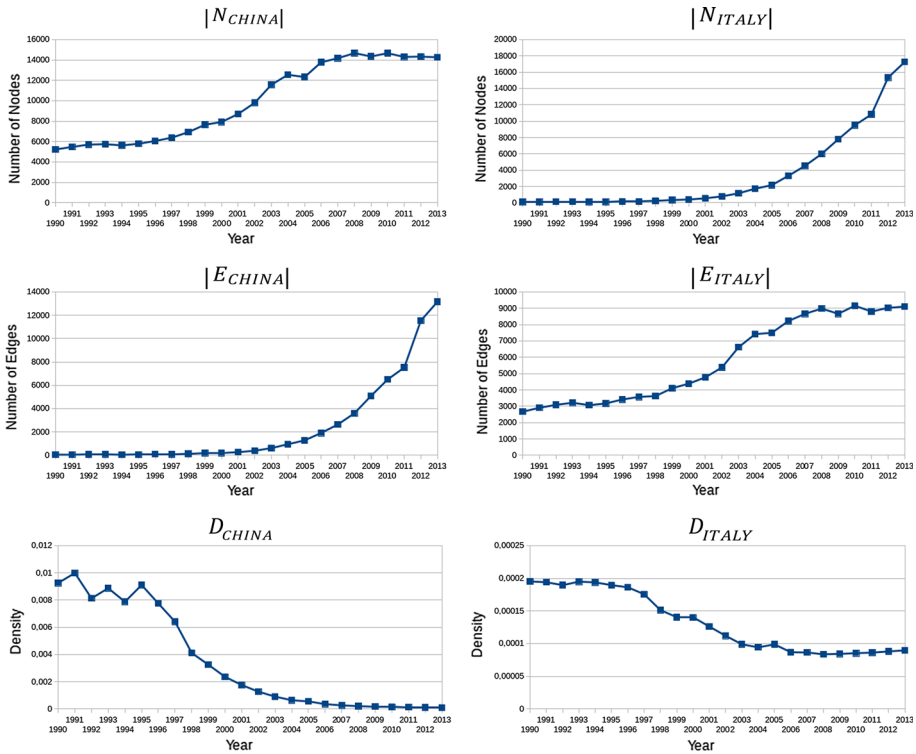
Fig. 2 Trends of $|N_k|$, $|E_k|$ and D_k for selected countries (China and Italy)

edges $|E_k|$, which corresponds to collaborations established by inventors in country k, and the density[11] D_k.

Generally, all countries into consideration present a growing number of nodes and edges over the years. Some countries—e.g. those of BRICS, South Korea, Taiwan, Turkey and Israel—present an exponential growth. The majority of EU countries instead show an increasing linear trend, although there are countries that with the North African ones show a growing but irregular trend over time.

As an example, Fig. 2 displays the temporal evolution of these variables for China and for Italy. In case of China notice that for both nodes and edges the rate of growth starts growing exponentially in 2002, whereas in case of Italy we can see that both nodes and edges show an increasing trend until 2008 and that thereafter both numbers have substantially remained constant.

As for density D_k, it generally decreases, as the density of a network is inversely proportional to the square of the number of nodes. To obtain a constant trend against time it would be necessary that the number of edges grows proportionally to the square of the number of nodes, which is unthinkable in real scenarios.

In Fig. 2 we also show the trend of D_k for China and Italy. Observe that the decrease is more pronounced in China than in Italy: this is due to the fact that the increase of the number of nodes is exponential for China and linear for Italy.

---

[11] Recall that in SNA the density of a network is equal to the ratio of actual connections and all potential connections in the network.

Cooperation patterns among inventors

We now address the issue of cooperation among inventors. As a first step into our investigation, we define a new metric M_2, such that M_{2i} indicates the dimension of the neighborhood of n_i.

This is useful to understand how much the inventors of a given country k tend to cooperate for filing patents. We measure M_2 for all Mediterranean and North African countries, the BRICS and some EU member states. For most of them it follows a quite disturbed power law distribution.

For instance, in Fig. 3 we show the distribution of M_2 for Brazil that presents a peak in the value of M_2 between 0 and 10. The only country presenting a perfect power law distribution is Austria, as shown in the same Figure.
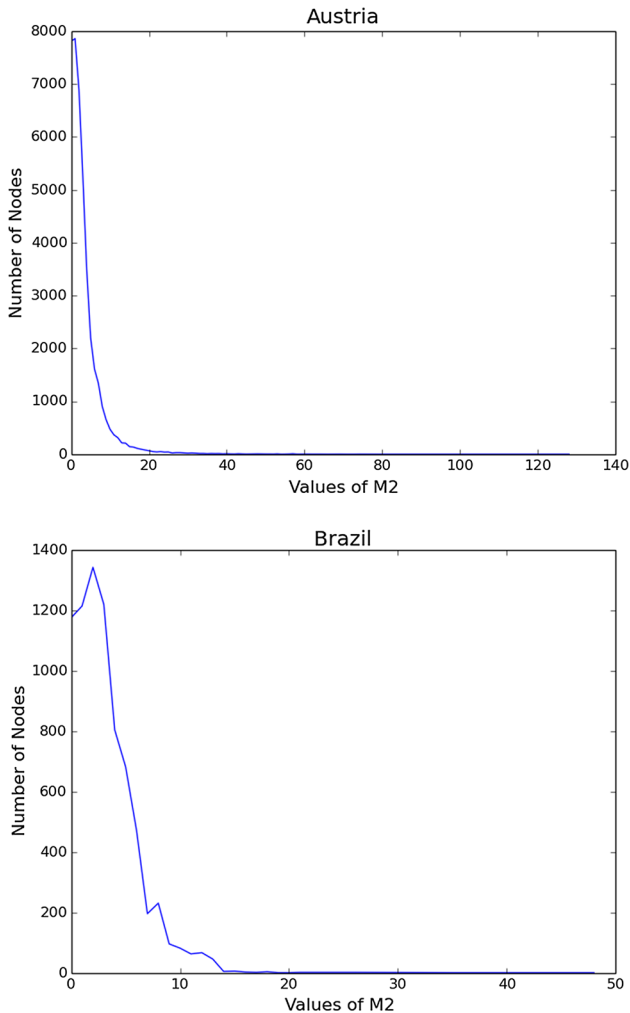

Fig. 3 Distribution of M_2 for Brazil and Austria

Cliques of inventors

To understand the dynamics of inventorship relations, and hence how cooperation occurs among inventors, we resort to the concept of clique. In SNA clique corresponds to a subgraph where every node is adjacent to all the others. In our scenario it indicates the presence of a compact group of inventors who cooperate intensively with each other. We compute the distribution of the cliques' dimension for the social network G_k.

Since in a clique all nodes are totally connected to each other its dimension can be considered a valid metric to quantify inventors' tendency, in a given country, to form more or less large working groups. The general trend we find for this phenomenon is the one of a power law distribution, where the dimension of the largest clique being different from country to country.

For instance, in Fig. 4 we report the distribution of cliques for Japan and United Kingdom. In both cases we can observe a power law distribution, with a different maximum number of cliques: although Japan has far more nodes than the UK the dimension of its largest clique is lower than the one of UK. A case of particular interest is represented by Israel (Fig. 4), which presents a maximum dimension of cliques equal to 7. This is a very
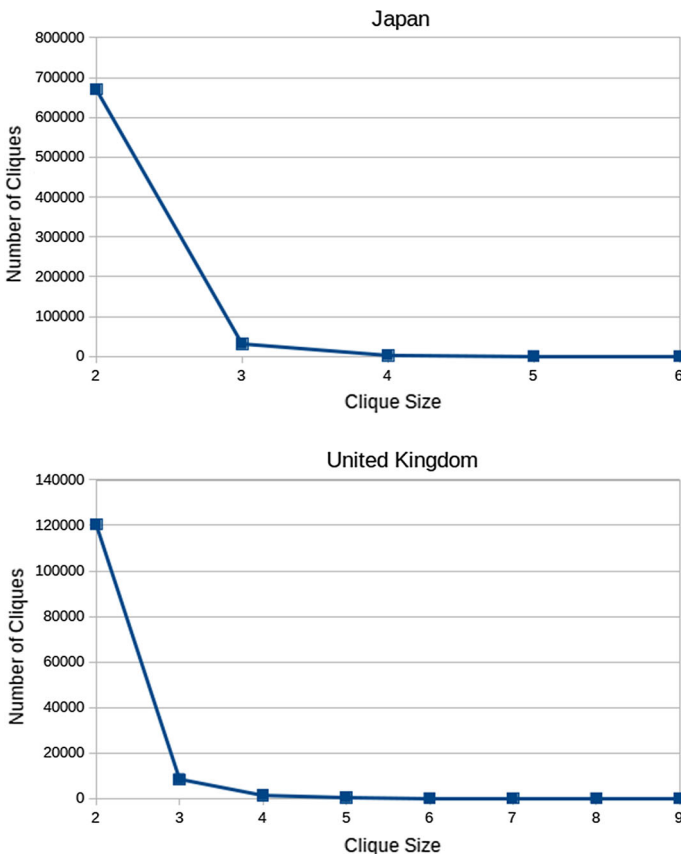

Fig. 4 Distribution of clique size for selected countries

high value considering that the number of Israelis inventors (30,358) is much lower than the one of Japan (924,554) and the UK (231,128).

To capture and quantify this last observation we define a parameter that indicates how much the inventors of a country k are aggregated in cliques. The previous result suggests that the difficulty to have larger and larger cliques grows exponentially. Hence, to define a corresponding index, we deem appropriate to consider only the cliques of maximum, sub-maximum and sub–sub-maximum dimension, as well as assigning an exponentially decreasing weight to these cliques.

Specifically, the aggregation index Agg_k of inventors on cliques for country k is defined as:

$$Agg_k = \frac{1}{|N_k|} \cdot \sum_{l=0}^{2} 2^{(|C_k|-l)} \cdot v_{(|C_k|-l)}$$

where C_k indicates the largest clique of country k, $|C_k|$ denotes its dimension and v_x is the number of cliques having dimension x.

In Table 1, we report the dimensions of the maximum, sub-maximum and sub–sub-maximum cliques for several (selected) countries, along with the corresponding aggregation index. We have chosen to split the table in three parts, grouping countries that share a similar number of inventors: the first includes countries with 35,000 inventors, the second those having a number of inventors between 40,000 and 65,000, and the third countries having more than 110,000 inventors.

In the first group of countries, Taiwan has the highest Aggregation Index, since, despite displaying a number of nodes quite low compared to the other countries in the same group, it has a high number of cliques of maximum, sub-maximum and sub–sub-maximum dimension. In the same group, India has the lowest value of Agg_k: notwithstanding the high

Table 1 Values of Agg_k for several countries

| Country	$|N_k|$	$|C_k|v_{(|C_k|)}$	$|C_k|-1v_{(|C_k|-1)}$	$|C_k|-2v_{(|C_k|-2)}$	Agg_k
Brazil	7721	5 (1)	4 (8)	3 (118)	0.143
Russia	12,813	7 (2)	6 (5)	5 (16)	0.085
Taiwan	18,729	7 (7)	6 (20)	5 (118)	0.317
India	26,516	6 (1)	5 (0)	4 (101)	0.063
Israel	30,358	7 (8)	6 (12)	5 (42)	0.103
Denmark	30,762	9 (1)	8 (0)	7 (22)	0.108
Finland	31,903	7 (1)	6 (17)	5 (72)	0.110
Austria	40,734	7 (7)	6 (20)	5 (118)	0.146
Spain	43,131	9 (6)	8 (13)	7 (41)	0.270
Belgium	48,073	8 (1)	7 (4)	6 (33)	0.059
China	54,419	8 (5)	7 (17)	6 (37)	0.107
Sweden	63,593	9 (10)	8 (5)	7 (6)	0.113
South Korea	115,272	9 (2)	8 (16)	7 (58)	0.109
United Kingdom	231,128	9 (4)	8 (11)	7 (7)	0.025
Japan	924,554	6 (26)	5 (244)	4 (2294)	0.049
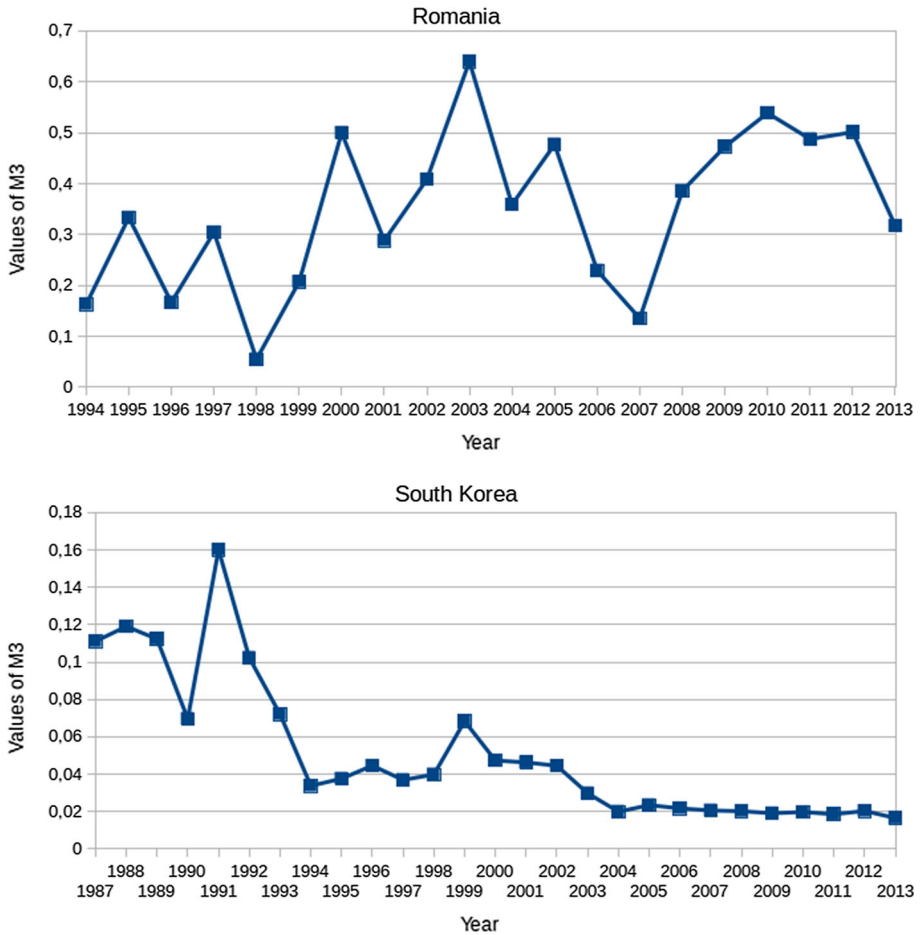
Fig. 5 Temporal evolution of M3 for selected countries

number of cliques of sub-maximum dimension, there is only one clique of maximum dimension and no clique of sub–sub-maximum dimension.

Analogous reasoning can be made for the second group of countries, where the highest lowest value of Agg_k belongs to Spain and the lowest to Belgium. As for the third group reported in Table 1, the values of Agg_k are generally small. The highest value is observed for South Korea that, in spite of having less inventors than the UK and Japan, display a dimension of maximum clique equal to the one of UK and even higher than the one of Japan.

Foreign collaborations

We are now interested in the international dimension of inventorship relationships. We start looking at the level of cooperation with foreign colleagues for inventors of a given country k.
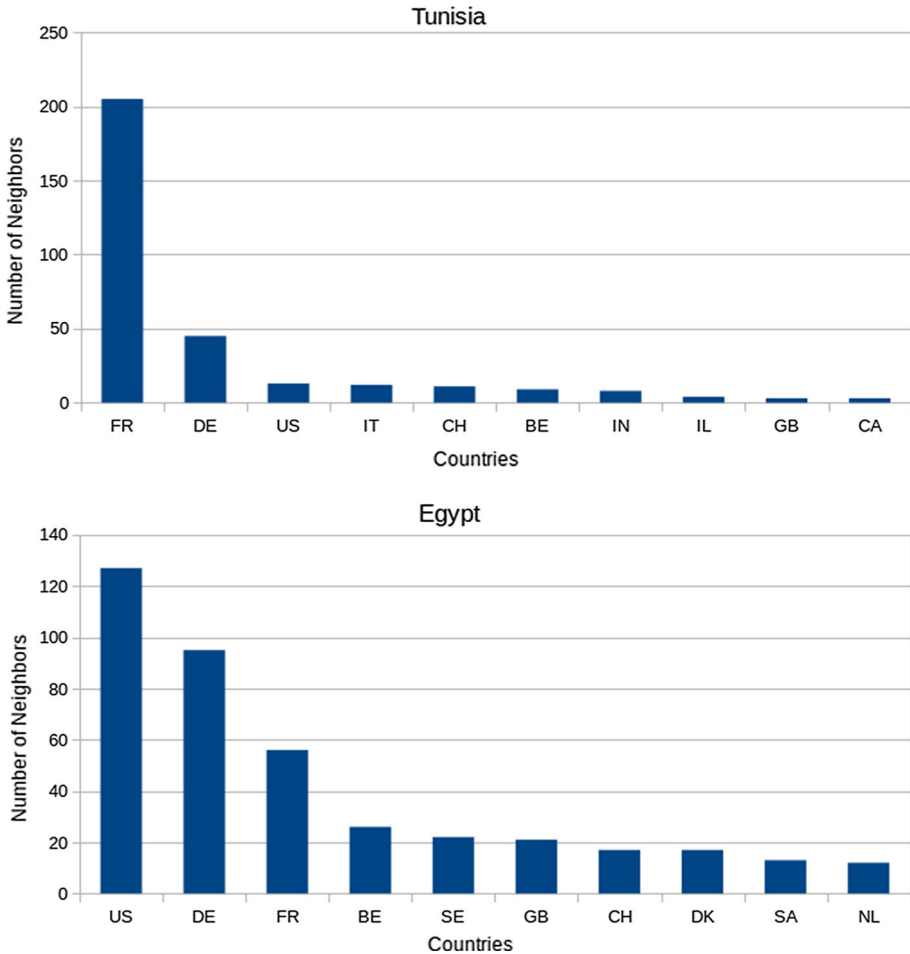
Fig. 6 Distribution of foreign collaborations for Algerian, Moroccan, Tunisia and Egyptian inventors

For this purpose, we define a metric M_3 such that M_{3k} represents the average fraction of foreign collaborations established by investors of k. M_3 ranges between 0 and 1 and allows us to understand how much the inventors of k tend to work with foreign colleagues: the higher M_3 the higher the tendency to establish scientific collaborations abroad.

We compute M_3 over a large period of time for several countries to understand their degree of internationalization and how it has evolved over time. We consider some EU member states, some countries of BRICS, some North African countries, as well as South Korea and Taiwan.

Findings are heterogeneous. Some countries—such as Spain, Taiwan, South Korea, China and Brazil—present a generally decreasing trend. Others—such as Austria, Italy and South Africa—display an increasing trend for this measure. There are also countries that show quite an irregular trend for M_3, characterized by the presence of peaks and decays over time.

In Fig. 5 we report the trend of M_3 for South Korea, Austria and Romania. We can observe that, while South Korean inventors have cooperated more and more with each
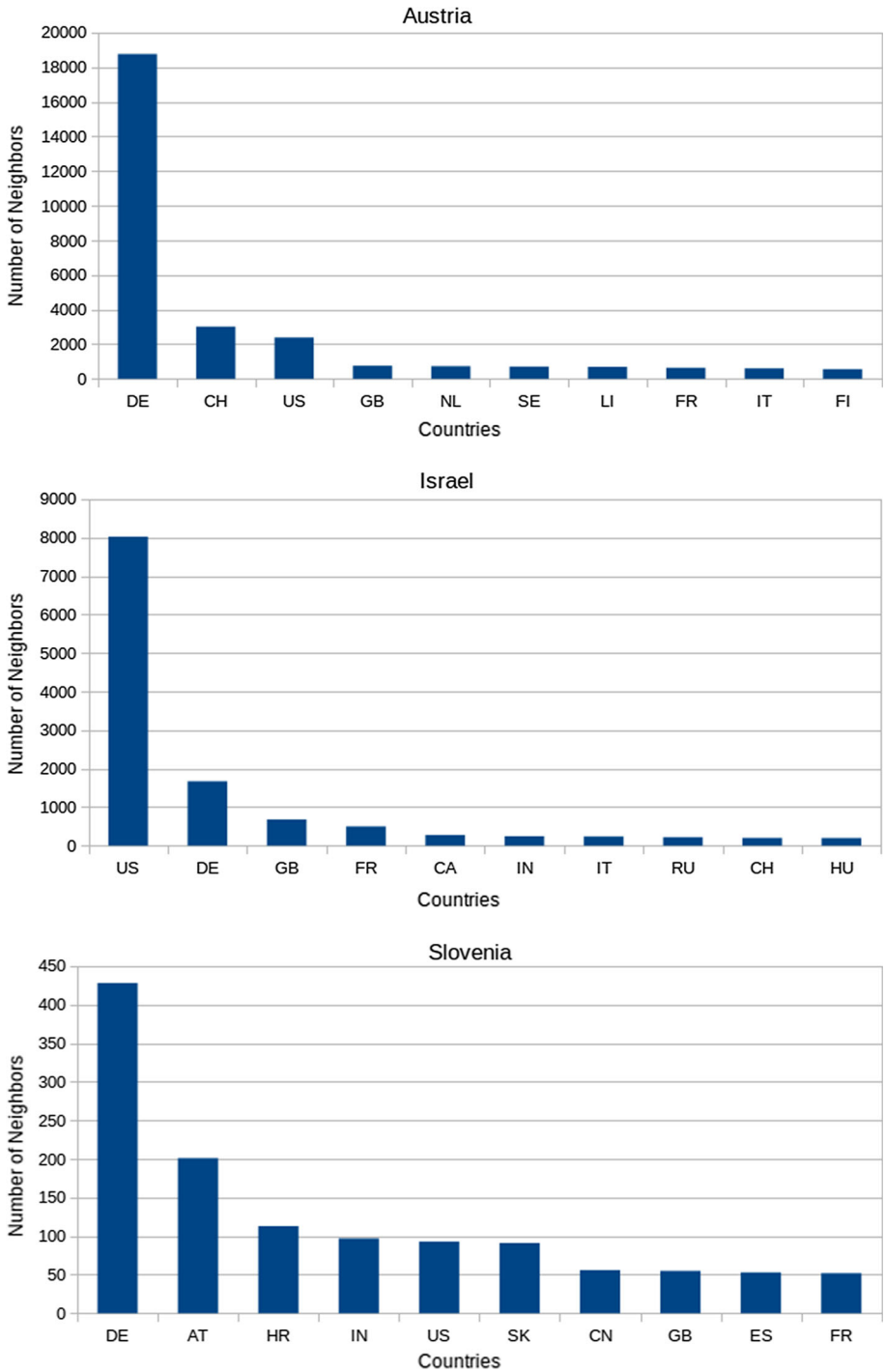
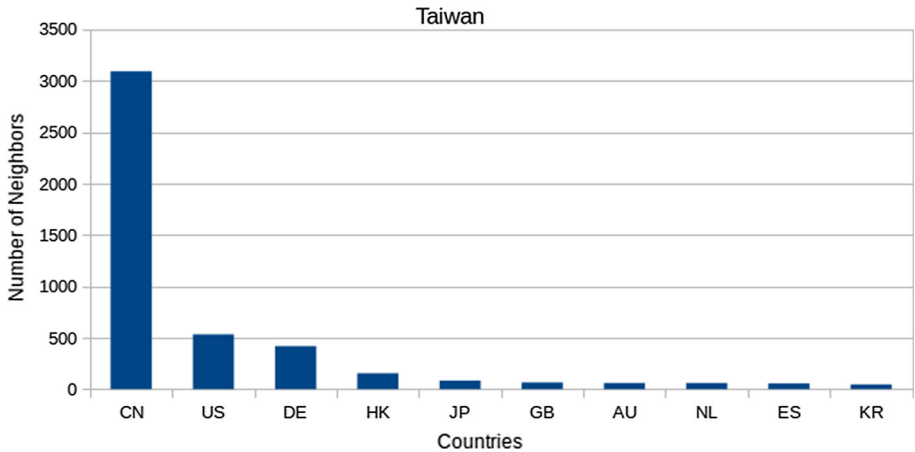Fig. 7 Distribution of foreign collaborations for selected countries

Fig. 7 continued

other within South Korea over the last years, Austrian inventors have grown their number of foreign collaborations, becoming more internationalized. In case of Romania, we can see that no clear trend emerges.

After the characterization of inventors' tendency to cooperate with foreign colleagues, we now aim at understanding the states a given country mostly cooperates with—as far as patents are concerned.

Given a country k, for each node $n_i \in N_k$, we consider the set nbh_i^F, which is the set of direct neighbours of n_i that belong to a country other than the one of n_i, and then we compute—for all Mediterranean and North African countries, BRICS and for some EU member states—the distribution of countries to which the nodes associated with these neighborhoods belong to. In particular, we focus our attention on some past colonies. In Fig. 6, we show the results obtained for some former French colonies—e.g. Algeria, Morocco and Tunisia—and for a former British colony—e.g., Egypt.

As one would expect, a large fraction of inventors in Algeria, Morocco and Tunisia mainly cooperate with French inventors. However, for both Morocco and Tunisia, there is a notable degree of cooperation with Germany as well; on the other hand, Algeria does not show significant inventorship relations with any other country. Looking at Egypt, which is a British past colony, we do not observe such high degree of cooperation with its former motherland, the United Kingdom. In fact, the highest number of collaborations can be found with inventors of United States, Germany and France. Interestingly, there is a good contribution with inventors coming from Saudi Arabia.[12]

In Fig. 7, we show the results obtained for Israel, Austria, Slovenia and Taiwan. In case of Israel, we can observe a strong cooperation with US inventors and with German ones as well. Germany is also an important source of scientific collaborations also for Austria and Slovenia.

---

[12] These results are in line with Landini et al. (2015), who observe how Egypt is one of the most active Northern African countries in terms of amount and variety of international collaborations and research output.

In addition, for Slovenian inventors we can observe a high concentration of links with Austrian ones—as evidence of the very strong links between Slovenia and German-speaking countries dating back to the Austro-Hungarian Empire period—and, to a lesser extent, with Croatian inventors. The same figure displays also the foreign collaborations entertained by Taiwanese inventors. Somehow unexpectedly, given the controversial political relations between the two countries, Taiwan has a high degree of scientific cooperation with China.

To gain additional insights on the international collaborations existing among countries, we compute the variety level of the countries inventors of a country k cooperating with. Drawing on the measure of biodiversity introduced by Simpson (1949), we build an indicator of the internationalization level of inventor teams, which we exploit to conduct an explorative empirical analysis of the trends and features in research groups' internationalization level by using cross-patent data from the Worldwide Patent Statistical Database.

It also allows us to measure the size of firms in relation to the industry and the degree of competition among them. This is known as the HI, and in this context tells us if inventors of a given country privilege collaborations with inventors of one or more foreign countries. The higher the HI_k of a country k, the more concentrated the external collaborations of k.

Table 2 reports the value of the Herfindahl Index (HI) for several countries. Among the North African countries, the highest value of HI is obtained by Algeria. This result is due to the fact that Algerian inventors cooperate mainly with inventors of one country, France. An analogous observation can be drawn for Tunisia and Morocco.

Differently from these three countries, Egypt has a much lower value of HI, because the collaborations of Egyptian inventors are more distributed among several countries. An

	Country	HI_k	$HI_k Top80\%$	$HI_k^* Top80\%$
Table 2 Values of HI, HI Top 80% and HI* Top 80% for North African countries	Egypt	0.144	0.148	0.147
	Morocco	0.358	0.367	0.365
	Tunisia	0.434	0.442	0.441
	Algeria	0.502	0.527	0.521
	Libya	0.180	0.250	0.143
	Spain	0.138	0.138	0.137
	Estonia	0.128	0.130	0.129
	Poland	0.118	0.119	0.119
	Austria	0.343	0.343	0.343
	Bulgaria	0.176	0.179	0.179
	Romania	0.173	0.175	0.175
	Slovenia	0.095	0.096	0.095
	Croatia	0.132	0.135	0.134
	Brazil	0.138	0.138	0.137
	India	0.128	0.130	0.129
	China	0.118	0.119	0.119
	South Africa	0.159	0.160	0.160
	Israel	0.369	0.370	0.370
	Taiwan	0.415	0.442	0.441
	Turkey	0.253	0.255	0.254
	Bosnia-Herzegovina	0.176	0.187	0.156

analogous reasoning can be drawn for EU countries—where the highest value of HI is obtained for Austria.

As for the BRICS, the value of HI is generally small, since Brazil, South Africa, China and India tend to cooperate with several countries. High values of HI can be observed also for Taiwan, Israel and Turkey.

However, it should be noted that there is one caveat that applies to the HI as we defined it: its value could be strongly distorted by the presence of a large number of extemporaneous collaborations between an inventor of country k and a foreign one who cooperated for one or two patents only.
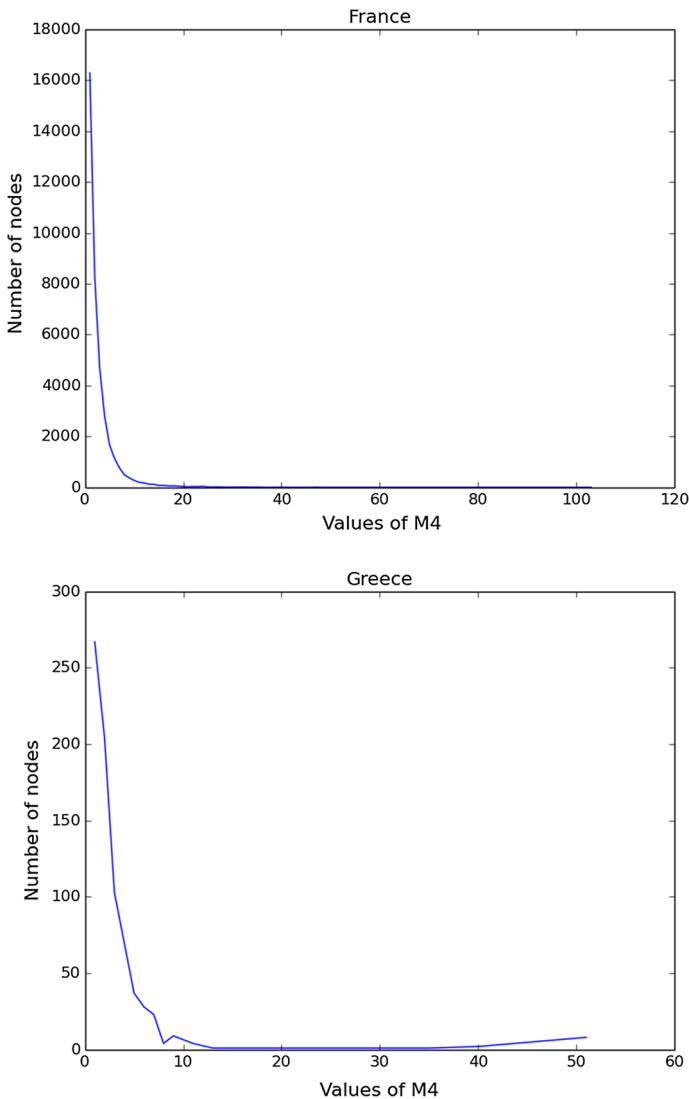

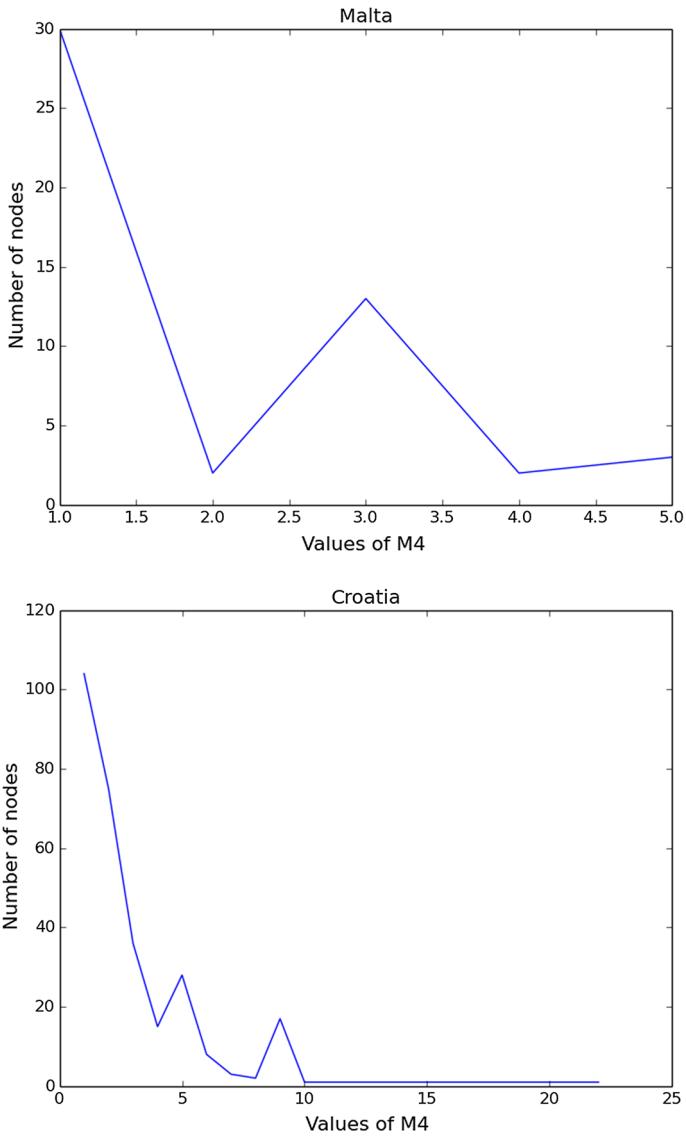**Fig. 8** Distribution of M4 for selected countries

Fig. 8 continued

Considering the power law trend, usual for the measures in our reference scenario, we use a modified version of the HI, limiting the countries into consideration to the top 80%. In this way the main cause of the distortion—the tail of the power law distribution—is removed. The results obtained with this new measure of the HI are reported in the third column of Table 2.

Power inventors

Before characterizing the notion of "power inventor", we define another metric, namely M_4, which is analogous to M_2 with the difference that it considers only the neighborhood of foreign inventors rather than the whole neighborhood,[13] since we are now interested in collaborations with foreign countries.

We measure M_4 for Euro-Mediterranean and North African countries, and the BRICS. For many of them M_4 follows a power law distribution even if some, such as Greece, Turkey and Principate of Monaco, present a slightly disturbed power law distribution.

For other countries, such as Morocco, Slovenia, Tunisia, Croatia, Cyprus and Egypt, M_4 follows a more disturbed power law distribution.

Finally, for other countries such as Syria, Albania, Algeria, Bosnia-Herzegovina, Lebanon, Libya and Malta, M_4 does not follow a power law distribution.[14] Figure 8 displays all the different cases. Hence, M_1, M_2 and M_4 generally follow a (potentially disturbed) power law distribution.

We are now ready to introduce the concept of "power inventor", which indicates an inventor with the following characteristics:

- $C1$: filing of many patents;
- $C2$: many collaborations;
- $C3$: having an international stature, which implies that she cooperates a lot with inventors of foreign countries.

To evaluate these conditions, we can use, respectively, M_1, M_2 and M_4.[15]

Then, the set P_k^X of power inventors of a country k is the set of those inventors simultaneously belonging to the top $X\%$ percentile of the distribution of M_1, M_2 and M_4 in each country k.

Since available data are huge, and since the power law distributions characterizing M_1, M_2 and M_4 are generally steep, we choose a low value for X and we set $X = 5$.[16]

With regard to this definition of power inventor we would like to point out that our aim is not proposing a new concept characterized by a precise supporting mathematical foundation.

Instead, we would like to introduce an informal and empirical, yet reasonable, concept that can capture the presence of inventors more important than others along several dimensions and that can support the extraction of knowledge patterns about this phenomenon.

The notion is similar to the one of "star scientist" employed in several studies that analyze the dynamics underlying the generation of new ideas and the value that some elite scientists and researchers bring to these processes, as well as their impact on other researchers.[17]

The idea behind the concept of power inventor is that of an inventor that has a notable scientific productivity (demonstrated through the filing of many patents) that can

---

[13] Essentially, nbh_i is substituted by nbh_i^F.

[14] This last case refers to countries that have a very low number of collaborations with foreign countries, which makes any analysis for them unreliable.

[15] Interestingly, M_2 and M_3 are equivalent to E-I index Hanneman and Riddle (2005).

[16] In the following, when X is not specified, we intend it to be equal to 5.

[17] Appendix A, Section A.7, in the additional material, includes a discussion on our notion of power inventor and the concept of star scientist.

catalyze the growth of innovative activities and help the internationalization process of his own country.

Hence, all the aforementioned criteria are necessary in order to be considered a power inventor. On the one hand, if an inventor filed very few patents, even if all of them were in cooperation with foreign inventors, he could not have enough weight to influence the R&D scenario of his country.

If an inventor filed many patents, but all of them were in cooperation with few inventors, she would not manage to stimulate through collaborations other inventors to file patents.

Finally, if an inventor filed many patents, but all of them were in cooperation with inventors of his country, he would certainly play a prominent role in his country but he would not have the capability to stimulate contacts with foreign countries.

Characteristics of power inventors' networks

After having established what power inventors are, we are now interested in understanding the features of the networks they operate within.

We construct another support data structure, the power inventor's social network, and two related metrics $rAgg_k$ and rf_k, which are useful in that they allow us to understand whether power inventors tend to cooperate more or less intensively among each other compared to other inventors in the same country, i.e. whether there is a backbone of power inventors or not in a given country.[18] Values for the two metrics are reported in Table 3.

Looking at Table 3 we can see that, for countries like Austria and Italy, power inventors are much more aggregated in cliques than all other inventors—they display the highest value of $rAgg_k$. As for rf_k, the values are quite homogeneous across countries and range between 3 and 7.

Since both $rAgg_k$ and rf_k present values higher than 1, we can conclude that there exists a backbone of power inventors in each of these countries.

To expand our inquiry about this issue, we consider a new data structure, the *clique social network*.[19] In Fig. 9, we plot the clique social network of Spain and Israel. We can see in case of Spain the network is characterized by a particular dense core implying that in this country there is a group of particularly active inventors who often cooperate with each other. In case of Israel there are instead several nodes whose dimension is generally smaller than those belonging to the core of Spain.

In Table 4 we report some characteristics of clique social networks, namely the number of nodes and edges, and density.

Looking at it, we can see that, for instance, France and Italy have the lowest values for density. The reason is that both these countries have many nodes that are only partially connected to each other. By contrast, Brazil has a very low number of nodes and edges but, at the same time, the highest density among the countries under consideration, implying that Brazilian power inventors are strongly connected to each other.

The impact of power inventors

Underlying the notion of power inventor there is the idea that these inventors have the ability of stimulate further inventions and improve also the scientific production of their

---

[18] See Appendix A, Section A.4, in the additional material.

[19] See Appendix A, Section A.6, in the additional material.

Table 3 Values of *rAggk* and *rfk* for selected countries

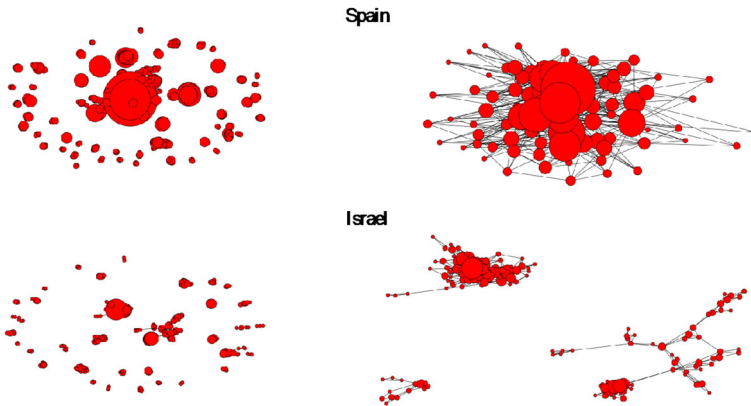| Country	rAggk	rfk
Brazil	5.595	6.763
Taiwan	29.012	6.541
China	23.483	5.846
Austria	44.084	4.193
Italy	47.472	5.100
Israel	22.583	3.674
Russia	10.034	5.149
Spain	25.498	6.624

Fig. 9 Visualization of the clique social network (left) and a zoomed portion of it (right) for two countries

Table 4 Number of nodes, number of edges and density of the clique social networks of some countries

Country	Number of nodes	Number of edges	Density
Brazil	19	23	0.135
India	209	417	0.019
China	577	1299	0.008
Israel	336	744	0.013
France	3452	8615	0.001
Italy	2125	5763	0.003
Spain	618	1672	0.009

colleagues. To quantitatively assess this intuition, we compute two different measures of scientific activity, namely the average number of patents filed by the neighbors of power inventors of a country k, denoted with $AvgPatNumNbh_k^P$, and the average number of patents filed by neighbors of generic inventors in the same country, denoted with $AvgPatNumNbh_k$. The values of these metrics are reported in Table 5.

A comparison between the two sets of values can support us in quantifying the benefit induced by cooperating with a power inventor, since high values of $AvgPatNumNbh_k^P$

Table 5 Average number of patents of the neighbours of a power inventor, of a generic inventor and values of the parameter *rPatNumNbh*

Country	$AvgPatNumNbh_k^P$	$AvgPatNumNbh_k$	$rPatNumNbh_k$
Brazil	10.979	3.329	3.299
Taiwan	3.437	1.344	2.557
China	6.554	2.680	2.445
India	3.876	2.264	1.712
Italy	7.210	3.397	2.122
Spain	7.668	2.639	2.906
South Africa	2.986	1.466	2.037
Israel	4.675	3.153	1.482

relative to $AvgPatNumNbh_k$ indicate that being in the neighbourhood of a power inventor stimulates patent filings.

From the Table it is evident, as we would expect, that being in the neighborhood of a power inventor is associated with an increase of filed patents. To sharpen our understanding we also define the following ratio:

$$rPatNumNbh_k = \frac{AvgPatNumNbh_k^P}{AvgPatNumNbh_k}$$

A value of $rPatNumNbh_k$ higher than 1 indicates that belonging to the neighbourhood of a power inventor is beneficial for filing patents. These values are reported in Table 5, and we can observe that values for this ratio are greater than 1 for all countries, in some cases even greater than 3—Brazil and Spain. Thus, we can conclude that cooperating with power inventors is beneficial for other inventors.

We now focus on the dimension of the neighborhood of power inventors. We already know, by definition, that a power inventor has a lot of scientific collaborations, but we would like to quantify its magnitude. We exploit the metrics $AvgDimNbh_k^P$, which is the average dimension of the neighbourhood of a power inventor, and $AvgDimNbh_k$, which is the average dimension of the neighbourhood of a generic inventor. A comparison between these two measures can tell us about the importance and centrality of a power inventor. We compute their variation over time for the same countries as in Table 3. Figure 10 displays, as an example, the time series of $AvgDimNbh_k^P$ and $AvgDimNbh_k$ for Spain. Note that the former measure remained almost constant from 1990 to 2003 and, then, had an increase, whereas the latter show an increasing trend.

Finally, we also report the values of the following ratio:

$$rDimNbh_k = \frac{AvgDimNbh_k^P}{AvgDimNbh_k}$$

When $rDimNbh_k$ is higher than 1, the average dimension of neighbourhoods of power inventors is higher than the corresponding one of generic nodes. In Table 6 we report the values of $rDimNbh_k$ for several countries in the year 2013, and we can observe that the values obtained for all the countries into consideration are always higher than 1 and range from 1.597 (for Brazil) to 2.870 (for Taiwan).
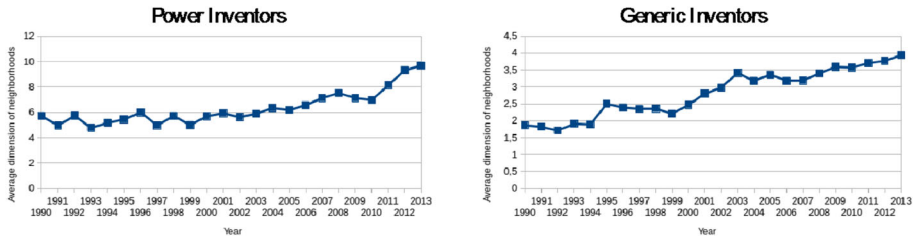
Fig. 10 Temporal evolution of the average neighbourhood dimension for Spain

Table 6 Values of *rDimNbh* for several countries in the year 2013

Country	rDimNbh
Austria	2.214
Italy	2.249
Spain	2.454
South Korea	2.360
Russia	2.561
Brazil	1.597
China	2.579
India	2.135
Poland	2.235
South Africa	2.311
Taiwan	2.870
Israel	1.974

Discussion

With respect to the existing literature, our study allows to characterize inventors' relationships from a quantitative standpoint and understand the cooperation patterns within inventors' networks in an innovative way. To this end, we have developed several new metrics and introduced new concepts, such as the notion of "power inventor", the aggregation coefficient as well as the parameters based on the modified HI for the computation of both the variety of external collaborations entertained by a given country and of the variability of IPC classes.

We have also redefined concepts already existing in SNA, such as the "internal" and "external neighborhood" of a node or the "clique" of nodes, adapting them so that they can be useful to gain insights on research collaborations and their dynamics. As in other studies, we exploit SNA to analyze relationships among inventors within and across countries. While we are not the first to employ an approach based on SNA, the originality of our contribution consists in presenting a general methodology for the extraction of several knowledge patterns with a reference to the geography of innovation. The suggested approach can then be applied to any country or to any sector of interest.

Concluding remarks

The goal of this research was to provide an overall understanding of the dynamics concerning knowledge flows in inventive activities. We show how the network of inventors is on average increasing in size, more and more inventors are in fact contributing to technology innovations and they are more connected to each other. We also show to what extent inventors from different countries tend to cooperate with their local peers or internationally. Furthermore, an analysis of the clustering of inventors is carried out to show differences across countries in the structure of inventors' communities, with a particular focus on the dynamics of collaboration for power inventors (i.e. star inventors).

From a methodological standpoint, new and alternative approaches to the extraction of knowledge patterns about the geography of innovation based on a SNA of patent data are proposed. These approaches are extremely general and can be exploited to analyze patents, inventors and their relationships within and among countries. First, an indicator of the internationalization proposed by Simpson (1949) to measure biodiversity is applied here to investigate trends and features in inventive teams' internationalization. Second, an index that measures the dimension of the neighborhood of power inventors is proposed to see how power inventors differentiate themselves from a generic inventor. Third, the benefits of cooperating with a power inventor are quantified with an appropriate index that highlights the expected increase in productivity that such a collaboration would mean for inventors in different countries. Finally, a parameter that indicates how much the inventors of a country are aggregated in cliques is proposed.

This research has implications for policy makers and R&D managers. Firstly, it provides an understanding of how inventive activities are carried out in different countries in terms of openness to cooperation. Secondly, the dynamics of knowledge flows for power inventors are distinguished from the dynamics of knowledge flows when a power inventor is not involved. This provides insights on how to nurture national innovation systems that can help generate more knowledge. Finally, insights on innovation cliques in different countries have implications for new market entry strategies definition and R&D internationalization endeavors organization.

It is important to stress that our work is aimed at providing a first, intuitive characterization of some patterns related to innovations and inventors' collaborations dynamics, since no causality statements can be made on the basis of our analysis; yet, many facts can be deduced from the data we have. Furthermore, even if the quality of the dataset used reduces the concerns about disambiguation of inventors, still different optimization algorithms may produce different results, a further effort carried out proved those results not to be significantly different from the ones presented here. More rigorous analyses and more precise answers will be object of future research, as well as providing a taxonomy of all the countries in our database depending on nodes and cliques, but this requires more time due to the vast amount of data to be analyzed. However, this work represents a promising starting point.

Deepening our understanding of power inventors could be a promising line of research, given their importance within the innovation system and the inventors' network of a given country. For instance, it would be possible to construct the profiles of power inventors of a given country using some classification methodology, as well as applying techniques of information diffusion to understand how much the mobility from one country to another of power inventors can impact on the filing of new patents in these two countries. Moreover, it would be interesting to inquire on the main causes that lead power inventors to leave

their homeland, or the reasons that induce power inventors to file patents in another country.

Answering these questions could give important strategic suggestions to policymakers for the definition of strategic actions aimed at retaining inventors in their country and to favor the arrival of new ones from abroad.

In conclusion, forthcoming research will focus on how the structure of innovation networks is related to innovation performance and internationalization strategies. This would help to set guidelines for policymakers that want to better understand how to nurture national and regional systems that take advantage of global value chains.

References

Abbas, A., Zhang, L., & Khan, S. U. (2014). A literature review on the state-of-the-art in patent analysis. *World Patent Information, 37,* 3–13.

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

Albert, R., & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics, 74,* 47–97.

Alnuaimi, T., Singh, J., & George, G. (2012). Not with my own: Long-term effects of cross-country collaboration on subsidiary innovation in emerging economies versus advanced economies. *Journal of Economic Geography, 12*(5), 943–968.

Anderson, J. A. (1996). *Communication theory: Epistemological foundations.* New York: The Guilford Press.

Azoulay, P., Ding, W., & Stuart, T. (2009). The impact of academic patenting on the rate, quality and direction of (public) research output. *The Journal of Industrial Economics, 57,* 637–676.

Balconi, M., Breschi, S., & Lissoni, F. (2002). Networks of inventors and the location of university research: An exploration of italian data. In *Proceedings of the International Conference "Rethinking Science Policy"*, Brighton, UK, SPRU.

Barabasi, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*(5439), 509–512.

Barabási, A.-L., Albert, R., & Jeong, H. (1999). Mean-field theory for scale-free random networks. *Physica A: Statistical Mechanics and its Applications, 272*(1), 173–187.

Barabási, A.-L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications, 311*(3), 590–614.

Blumer, H. (1967a). Sociological analysis and the variable. In J. G. Manis & B. N. Meltzer (Eds.), *Symbolic interaction.* Boston: Allyn and Bacon.

Blumer, H. (1967b). Threats from agency-determined research: The case of Camelot. In I. L. Horowitz (Ed.), *The rise and fall of project Camelot.* Cambridge: M.I.T. Press.

Boissevain, J. (1979). Network analysis: A reappraisal. *Current Anthropology, 20,* 392–394.

Branstetter, L., Li, G., & Veloso, F. (2014). The rise of international coinvention. In A. B. Jaffe & B. F. Jones (Eds.), *The changing frontier: Rethinking science and innovation policy* (pp. 135–168). Chicago: University of Chicago Press.

Breschi, S., & Lenzi, C. (2015). 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.

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

Breschi, S., Lissoni, F., & Montobbio, F. (2008). University patenting and scienti c productivity: A quantitative study of Italian academic inventors. *European Management Review, 5*(2), 91–109.

Broadhead, R. S., & Rist, R. C. (1976). Gatekeepers and the social control of social research. *Social Problems, 23*(3), 325–336.

Cammarano, A., Caputo, M., Lamberti, E., & Michelino, F. (2017). R&D collaboration strategies for innovation: An empirical study through social network analysis. *International Journal of Innovation and Technology Management, 14*(01), 1740001.

Chen, Z., & Guan, J. (2016). The core-peripheral structure of international knowledge flows: Evidence from patent citation data. *R&D Management, 46*(1), 62–79.

Chung, C. J., Barnett, G. A., Kim, K., & Lacka, D. (2013). An analysis on communication theory and discipline. *Scientometrics, 95*(3), 985–1005.

Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E, 70,* 066111.

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review, 51*(4), 661–703.

Crane, D. (1965). Scientists at major and minor universities: A study of productivity and recognition. *American Sociological Review, 30,* 699–714.

Crawford, E. T., & Biderman, A. D. (1970). Paper money: Trends of research sponsorship in American sociology journals. *Social Science Information, 9,* 51–77.

Delgado, E., & Repiso, R. (2013). The impact of scientific journals of communication: Comparing google scholar metrics, web of science and scopus. *Comunicar, 41,* 45–52.

Ejermo, O., & Karlsson, C. (2006). Interregional inventor networks as studied by patent coinventorships. *Research Policy, 35*(3), 412–430.

Feeley, T. H., LaVail, K. H., & Barnett, G. A. (2011). Predicting faculty job centrality in communication. *Scientometrics, 87,* 303–314.

Fleming, L., King, C., & Juda, A. I. (2007). Small worlds and regional innovation. *Organization Science, 18,* 938–954.

Forti, E., Franzoni, C., & Sobrero, M. (2013). Bridges or isolates? investigating the social networks of academic inventors. *Research Policy, 42*(8), 1378–1388.

Freeman, L. C. (2004). *The development of social network analysis: A study in the sociology of science.* Vancouver: Empirical Press.

Furman, J. L., Kyle, M. K., Cockburn, I. M., & Henderson, R. (2006). Public and private spillovers, location and the productivity of pharmaceutical research. Working paper 12509, National Bureau of Economic Research.

Furstenberg, F. F. (1971). Political intrusions and governmental confusion: The case of the National Institute of Law Enforcement and Criminal Justice. *American Sociologist, 6,* 59–62.

Galliher, J. F., & McCartney, J. C. (1973). The influence of funding agencies on juvenile delinquency research. *Social Problems, 21,* 77–90.

Garcia, R., & Calantone, R. (2002). A critical look at technological innovation typology and innovativeness terminology: A literature review. *Journal of Product Innovation Management, 19*(2), 110–132.

Giuliani, E., & Bell, M. (2005). The micro-determinants of meso-level learning and innovation: Evidence from a Chilean wine cluster. *Research Policy, 34*(1), 47–68.

Giuliani, E., Martinelli, A., & Rabellotti, R. (2016). Is co-invention expediting technological catch up? a study of collaboration between emerging country firms and eu inventors. *World Development, 77,* 192–205.

Gould, R. V., & Fernandez, R. M. (1989). Structures of mediation: A formal approach to brokerage in transaction networks. *Sociological Methodology, 19,* 89–126.

Graf, H., & Kruger, J. J. (2011). The performance of gatekeepers in innovator networks. *Industry and Innovation, 18*(1), 69–88.

Grant, R. M. (1996). Prospering in dynamically-competitive environments: Organizational capability as knowledge integration. *Organization Science, 7*(4), 375–387.

Green, P. (1971). The obligations of American social scientists. *The Annals, 394,* 13–27.

Griffin, D. J., Bolkan, S., Holmgren, J. L., & Tutzauer, F. (2016). Central journals and authors in communication using a publication network. *Scientometrics, 106*(1), 91–104.

Griliches, Z. (1990). Patent statistics as economic indicators: A survey. *Journal of Economic Literature, 28*(4), 1661–1707.

Griliches, Z. (1992). Introduction to "output measurement in the service sectors". In Z. Griliches (Ed.), *Output measurement in the service sectors* (pp. 1–22). Chicago: University of Chicago Press.

Guellec, D., & van Pottelsberghe de la Potterie, B. (2001). The internationalization of technology analyzed with patent data. *Research Policy, 30*(8), 1253–1266.

Hanneman, R. A. & Riddle, M. (2005). Introduction to social network methods. Published in digital form at http://faculty.ucr.edu/hanneman/, University of California, Riverside, CA.

Hingley, P., & Bas, S. (2009). Numbers and sizes of applicants at the European patent office. *World Patent Information, 31*(4), 285–298.

Horowitz, I. L. (1967). The rise and fall of project Camelot. In I. L. Horowitz (Ed.), *The rise and fall of project Camelot*. Cambridge: M.I.T. Press.

Hsueh, C. C. & Wang, C. C. (2009). The use of SNA in knowledge diffusion research from patent data. In *Proceedings of the International Conference on Advances in SNA and Mining (ASONAM)*, (pp. 393–398).

Kim, J., Lee, S. J., & Marschke, G. (2009). International knowledge flows: Evidence from an inventor-firm matched data set. In *Science and engineering careers in the United States: An analysis of markets and employment*, (pp. 321–348). National Bureau of Economic Research, Inc.

Kim, J. N., Park, S. C., Yoo, S. W., & Shen, H. (2010). Mapping health communication scholarship: Breadth, depth, and agenda of published research in health communication. *Health Communication, 25,* 487–503.

Landini, F., Malerba, F., & Mavilia, R. (2015). The structure and dynamics of networks of scientific collaborations in Northern Africa. *Scientometrics, 105*(3), 1787–1807.

Lee, S., Kim, J. H., & Rosen, D. (2009). A semantic network and categorical content analysis of Internet and online media. *The Open Communication Journal, 3,* 15–28.

Leicht, E. A., Holme, P., & Newman, M. E. J. (2006). Vertex similarity in networks. *Physical Review E, 73,* 026120.

Levinthal, D. A., & March, J. G. (1993). The myopia of learning. *Strategic Management Journal, 14*(S2), 95–112.

Leydesdor, L. (2007). Betweenness centrality as an indicator of the interdisciplinary of scientific journals. *Journal of the American Society for Information Science, 58*(9), 1303–1319.

Li, G. C., Lai, R., Damour, A., Doolin, D. M., Sun, Y., Torvik, V. I., et al. (2014). Disambiguation and co-authorship networks of the US patent inventor database. *Research Policy, 43*(6), 941–955.

Lissoni, F. (2012). Academic patenting in Europe: An overview of recent research and new perspectives. *World Patent Information, 34*(3), 197–205.

Lissoni, F. & Miguelez, E. (2014). Patents, innovation and economic geography. Technical Report 2014–16, Groupe de Recherche en Economie Théorique et Appliquée.

Lissoni, F., Coffano, M., Maurino, A., Pezzoni, M., & Tarasconi, G. (2010). APE-INV's name game algorithm challenge: A guideline for benchmark data analysis and reporting. Mimeo.

Lissoni, F., Llerena, P., McKelvey, M., & Sanditov, B. (2008). Academic patenting in Europe: new evidence from the KEINS database. *Research Evaluation, 17*(2), 87–102.

Lissoni, F., Llerena, P., & Sanditov, B. (2006a). Networks of inventors and academics in France. Italy and Sweden: Evidence from the Keins Database.

Lissoni, F., Sanditov, B., & Tarasconi, G. (2006b). The Keins database on academic inventors: Methodology and contents. CESPRI working paper 181.

Lubango, L. M. (2015). The effect of co-inventors' reputation and network ties on the diffusion of scientific and technical knowledge from academia to industry in South Africa. *World Patent Information, 43,* 5–11.

Lusseau, D. (2003). The emergent properties of a dolphin social network. *Proceedings of the Royal Society of London, Series B: Biological Sciences, 270,* 186–188.

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science, 2*(1), 71–87.

Marx, M., Strumsky, D., & Fleming, L. (2009). Mobility, skills, and the Michigan non-compete experiment. *Management Science, 55*(6), 875–889.

McCartney, J. L. (1971). The financing of sociological research: trends and consequences. In Edward A. Tiryakian (Ed.), *The phenomenon of sociology*. New York: Appleton-Century-Crofts.

Menon, C. (2014). Spreading big ideas? The effect of top inventing companies on local inventors. *Journal of Economic Geography, 15,* 743–768.

Miguélez, E., & Moreno, R. (2013). Research networks and inventors' mobility as drivers of innovation: Evidence from Europe. *Regional Studies, 47*(10), 1668–1685.

Montobbio, F., & Sterzi, V. (2011). Inventing together: Exploring the nature of international knowledge spillovers in Latin America. *Journal of Evolutionary Economics, 21*(1), 53–89.

Morrison, A. (2008). Gatekeepers of knowledge within industrial districts: Who they are how they interact. *Regional Studies, 42*(6), 817–835.

Nagaoka, S., Motohashi, K., & Goto, A. (2010). Patent statistics as an innovation indicator. *Handbook of the Economics of Innovation, 2,* 1083–1127.

Newman, M. E. J., & Leicht, E. A. (2007). Mixture models and exploratory analysis in networks. *Proceedings of the National Academy of Sciences, 104,* 9564–9569.

Nicholas, T. (2009). Spatial diversity in invention: Evidence from the early R&D labs. *Journal of Economic Geography, 9*(1), 1–31.

Orlans, H. (1967). Ethical problems in the relations of research sponsors and investigators. In Gideon Sjoberg (Ed.), *Ethics, politics, and social research*. Cambridge: Schenkman.

Otte, E., & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. *Journal of Information Science, 28,* 441–453.

Park, H. W. (2010). Mapping the e-science landscape in South Korea using the webometrics method. *Journal of Computer-Mediated Communication, 15,* 211–229.

Pezzoni, M., Lissoni, F., & Tarasconi, G. (2014). How to kill inventors: Testing the Massacrator algorithm for inventor disambiguation. *Scientometrics, 101,* 477–504.

Platt, A. M. (1971). *The politics of riot commissions*. New York: Collier.

Ponds, R., van Oort, F., & Frenken, K. (2010). Innovation, spillovers and university–industry collaboration: An extended knowledge production function approach. *Journal of Economic Geography, 10*(2), 231–255.

Raffo, J., & Lhuillery, S. (2009). How to play the name game: Patent retrieval comparing different heuristics. *Research Policy, 38*(10), 1617–1627.

Rist, R. C. (1973). Polity, politics, and social research: A study in the relationship of federal commissions and social science. *Social Problems, 21,* 113–128.

Rist, R. C. (1975). Ethnographic techniques and the study of an urban school. *Urban Education, 10,* 86–108.

Simpson, E. (1949). Measurement of diversity. *Nature, 163,* 688.

Singh, J. (2005). Collaborative networks as determinants of knowledge diffusion patterns. *Management Science, 51*(5), 756–770.

Singh, J. (2008). Distributed R&D, cross-regional knowledge integration and quality of innovative output. *Research Policy, 37*(1), 77–96.

Sternitzke, C., Bartkowski, A., & Schramm, R. (2008). Visualizing patent statistics by means of SNA tools. *World Patent Information, 30*(2), 115–131.

Thelwall, M. (2009). *Introduction to webometrics*. New York: Morgan and Claypool.

Tseng, C. Y., Lin, S. C., Pai, D. C., & Tung, C. W. (2016). The relationship between innovation network and innovation capability: A social network perspective. *Technology Analysis & Strategic Management, 28*(9), 1029–1040.

Tsvetovat, M., & Kouznetsov, A. (2011). *SNA for startups: Finding connections on the social web*. Sebastopol: O'Reilly Media Inc.

Wagner, C. S., & Leydesdorff, L. (2005). Network structure, self-organization, and the growth of international collaboration in science. *Research Policy, 34*(10), 1608–1618.

Wang, C., Rodan, S., Fruin, M., & Xu, X. (2014). Knowledge networks, collaboration networks, and exploratory innovation. *Academy of Management Journal, 57*(2), 484–514.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge: Cambridge University Press.

Wellman, B., & Whitaker, M. (1974). Community, network, communication: An annotated bibliography. Toronto: Centre for urban and community studies, University of Toronto World Intellectual Property Organization (2009). *Patent-Based Technology Analysis Report—Alternative Energy Technology*.

WIPO. (2011). World intellectual property report: The changing face of Innovation. WIPO Press.

Zsindely, S., Schubert, A., & Braun, T. (1982a). Citation patterns of editorial gatekeepers in international chemistry journals. *Scientometrics, 4*(1), 69–76.

Zsindely, S., Schubert, A., & Braun, T. (1982b). Editorial gatekeeping patterns in international science journals. A new science indicator. *Scientometrics, 4*(1), 57–68.