Journal of Informetrics 16 (2022) 101238

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Regular article

Patent collaborations: From segregation to globalization

Maria Tsouchnika [a,*], Alex Smolyak [b], Panos Argyrakis [a], Shlomo Havlin [b]

[a] Department of Physics, University of Thessaloniki, and Center of Complex Systems, Thessaloniki 54124, Greece
[b] Department of Physics, Bar-Ilan University, Ramat-Gan 52900, Israel

ARTICLE INFO

Keywords:
EPO Patents network
Giant component
Percolation
Network evolution
Globalization

ABSTRACT

We studied the evolution of the European Patent Office (EPO) patents applicants' collaborations network, within a 35 years span of data (1978–2013). Focusing on the Giant Component (GC) formation process over many time-windows, distributed throughout the data timeline, we found that the features governing this phenomenon are indicative of emerging globalization in the applicants' collaborations. The timeline appears to be divided into three regimes, corresponding to three states of the network's evolution. In the early years state, the GC takes long to form and the instant of its creation is easily pinpointed, while it features geographically segregated groups of applicants with technologically similar activities. In contrast, in the late years state, the GC forms quickly, the exact point of its creation is harder to spot, the applicants' activities are more disparate technologically, while their inter-regional collaborations are significantly increased. The middle years are an intermediate state between the two extreme of early and late years. Moreover, we concluded that the critical patents, which induce the GC's formation, are typically introduced by large-sized applicants and also that top patent-producing applicants are likely to submit critical patents, albeit at a lower rate than their overall patent submission. Lastly, we uncovered the crucial role that Japan plays in the network's coherence, through its prominent participation in the GC and the critical patents.

1. Introduction

Innovation is widely regarded as a major contributing factor to economy growth and job creation. A plethora of theoretical and empirical studies support the notion that innovation, in many ways, has a positive effect on productivity and growth (Caballero and Jaffe, 1993; Cainelli et al., 2004; Cassiman et al., 2010; Crépon et al., 1998; Geroski et al., 1993; Geroski, 1989; Griliches, 1958; 1964; 1980; 1986; Griliches and Mairesse, 1983; Hall, 2011; Hall et al., 2009; Harhoff, 1998; Klette, 1996; Klette and Johansen, 2000; Klomp and Van Leeuwen, 1999; Landau and Rosenberg, 1986; Lööf and Heshmati, 2002; Mansfield, 1961; 1962; 1965; Medda and Piga, 2014; Mohnen and Hall, 2013; Parisi et al., 2006; Raymond et al., 2015; Romer, 1986; Solow, 1957). Consequently, fostering innovation is important in improving the living standards and welfare of people and nations

However, innovation alone does not suffice for promoting well-being; adequate diffusion and adoption of the corresponding knowledge is also required (Damijan and Kostevc, 2015; Hall and Khan, 2003; Mansfield, 1961; Vasilyeva et al., 2021). Diffusion of the innovation's outcome is a prerequisite for boosting productivity and, consequently, growth (Gurbiel, 2002; Moreno and Suriñach, 2014; Suriñach et al., 2011). An effective way to achieve diffusion of knowledge is through the collaboration between all potential carriers of innovation, i.e. inventors, firms, research institutes, etc. Becheikh et al. (2006), performed a systematic review of the relevant literature, from 1993 to 2003, and concluded that networking is a remarkably good determinant of innovation. This conclusion was

* Corresponding author.
E-mail address: mtsou@auth.gr (M. Tsouchnika).

https://doi.org/10.1016/j.joi.2021.101238
Received 23 May 2021; Received in revised form 25 November 2021; Accepted 1 December 2021
Available online 25 December 2021

based on the fact that the majority of the studies examined, found networking to have a beneficial effect on innovation. The rest of the studies reviewed by Becheikh et al. found the effect to be insignificant, while none revealed a notable negative effect. Furthermore, knowledge flow is often trapped within regional and firm boundaries. Promoting collaborative ties between regions and firms has been shown to significantly increase the chance of knowledge escaping these boundaries (Singh, 2005). Moreover, there is a growing need of forming collaborative ties between the carriers of innovation, which have become more collective and distributed over the past decades (Kerl and Moehrle, 2015; Khan et al., 2013; Teece, 1992; Tether, 2002). Plausible motives driving this behaviour include common purpose, pooling resources, tackling difficulties and reducing the risks related to the innovation process, as well as keeping up with the fast pace of technological change and increasing competitiveness (Brown and Eisenhardt, 1995; Kerl and Moehrle, 2015; Stock et al., 2002; Tether, 2002).

Apart from networking and forming collaborative ties, another feature which effect on innovation output has been widely investigated is firm size. While in 1934, Schumpeter (1934) suggested that small size favours innovation output, he contradicted it eight years later (Schumpeter, 1942). This triggered a major debate that led to a series of theoretical and empirical studies, which reach far into the recent years. Indicatively, Tether in Tether (1998) questions the validity of a belief that small-sized firms are more efficient innovators than large ones (Acs and Audretsch, 1990), which emerged in the early90ıs. This belief was prompt by a group of empirical studies (Acs and Audretsch, 1990; Cogan, 1993; Kleinknecht et al., 1993; Pavitt et al., 1987; Santarelli and Piergiovanni, 1996), which all find that small-sized firms produce a higher number of innovations per thousand employees, than large ones do. Moreover, in Becheikh et al. (2006), while the majority of the relevant studies reviewed are suggestive of a positive correlation between size and innovation output, there is also a number of studies indicative of a negative, negligent, or even a complex relationship.

Overall, the subject of size vs. innovation output is regarded rather inconclusive, and the relationship between these two features is not considered straightforward (Revilla and Fernández, 2012; Rogers, 2004), but complex, in spite of the fact that most studies reveal a positive correlation (Becheikh et al., 2006). Therefore, many studies tend to examine a more specific aspect of the subject, as the effect of the size on innovation, with respect to networking (Rogers, 2004; Tether, 2002), or with respect to the technological regimes (Revilla and Fernández, 2012). Another aspect that is particularly interesting is the effect of size on how radical the produced innovations are. In Minguela-Rata et al. (2014), it is found that small-sized firms commonly engage in more radical innovations, while large-sized in more incremental ones. This result is attributed to the fact that small-sized firms are considered more flexible and adaptable to technological changes than the more rigid large-sized ones. Similarly, Stock et al. (2002) used data from the computer modem industry and inferred that small-sized, adjustable firms exhibit higher rate of technological change in the performance of their products.

All of the above motivated us to analyze the patents data from a network perspective, highlighting aspects that promote - or inhibit - collaboration. Patents data, although not short of weaknesses, are recognized as a useful means of determining technological advance and innovation (Griliches, 1990; Lööf and Heshmati, 2002; Pavitt, 1985) and have thus been used in many pertinent studies (Caballero and Jaffe, 1993; Crépon et al., 1998; Fleming et al., 2007; Hall, 2011; Ma and Lee, 2008; Pilkington, 2004; Schilling and Phelps, 2007; Zhang et al., 2017). A patent serves as a reflection of the underlying innovative procedures that led to the corresponding invention. Furthermore, patent data are stored in readily available and up to date databases, comprising information about the inventions, such as their date and type, and about the people, institutes and firms involved, as well as their collaborations and interactions. Therefore, by studying the network of collaborations formed by all those involved in the production of a patent, we implicitly investigate aspects of the relationship between these collaborative activities and innovation. The nodes of the network are the patent applicants, representing the firms/inventors working on a project that led to a patent application, filed at the EPO. A link between a pair of nodes (applicants) is drawn, when the applicants have at least one joint patent application. The network of patent collaborations studied is a social network and, more specifically, an affiliation network (Albert and Barabási, 2002; Barabâsi et al., 2002; Newman, 2001b; Newman et al., 2002).

Clearly, having a network with many small isolated components does not promote collaboration and the diffusion of knowledge as much as a network having a giant component (GC) would[1] Forging co-operative arrangements for innovation is found to correlate positively with higher levels of innovation (Kerl and Moehrle, 2015; Tether, 2002). There is also evidence implying that a network with short path lengths and increased aggregation of isolated components into bigger ones, has a positive effect on future patent production and therefore on innovation (Fleming et al., 2007; Schilling and Phelps, 2007). Furthermore, it has been found that the most productive European researchers increase their research output, by forming a dense core of strong ties between them, coupled with a large network of collaborations with geographically distant researchers (Hâncean et al., 2021). Moreover, Bettencourt et al. (2009) introduce and explore the very interesting idea that critical (red-bond) patents - the addition of which results in the formation of the network's GC, at the percolation threshold - represent a highly innovative moment in the course of scientific events.

In view of the above, the GC of the patents collaboration network is an excellent standpoint from which to study the interplay between collaboration and innovation. The availability of many years of patent data enables us to perform a dynamical analysis, and study the GC's formation in many different snapshots of the network, over time. Thus, we can examine questions pertaining to the GC's existence, as well as to the conditions that could possibly influence - promote or hinder - its formation over time. On that basis, the basic questions explored in this study are:

- Does the GC exist in every snapshot and if yes, how long does it take to form?
- Do the conditions the GC's formation change over time?

[1] The GC of a network is a connected component which size is proportional to the number of the network's nodes (Albert and Barabási, 2002; Barabâsi et al., 2002; Newman, 2001b; Newman et al., 2002).

  - Do the technological areas of the patents influence the GC's formation?
  - Does the geographical origin of the applicants influence the GC's formation?
- Are the major patent contributors (large-sized applicants) more likely than the ones with medium to small contributions (small/medium-sized applicants) to produce a critical patent?

Probing into the latter question would be an indirect contribution to the size vs. innovation output problem, assuming that the larger the size of an applicant the more patents is capable of producing (Tether, 2002). Overall, investigating the above questions would be one more step towards unraveling the complex role that social collaboration networks play in promoting economic growth and innovation.

Our results are indicative of a three-part division of the available timeline regarding specific characteristics of the GC formation and of the two major groups of applicants (largest and second largest connected components of the growing network) immediately before the percolation threshold. The system appears to undergo a shift in its state as it advances through time. It appears to move from a state of slow, clear-cut percolation transitions, marked by technological similarity and geographical confinement of the two groups, to a state featuring faster, although more incremental, transitions that hallmark technological complementarity and geographical interplay, having passed through a middle, transitive state. Furthermore, our analysis suggests that top patent-producing applicants are likely to introduce critical patents, although not as likely as would be expected according to their overall patent output. Also the vast majority of red-bond applicants are found to be large-sized firms. Lastly, our findings highlight the key-role played by Japan to the EPO patent network and predominantly to its GC.

The remainder of this paper is structured as follows. Section 2 offers basic information on percolation theory, which lies at the heart of this analysis. Section 3 outlines our analysis and the methodology we followed, and presents all our results. Section 4 explores the interpretations and certain implications of our results. Finally, Section 5 summarizes our key findings.

2. Percolation theory fundamentals

In physics, percolation theory models the behaviour of systems with metric structure, such as lattices embedded in space, or non embedded networks, in which their sites/bonds, or nodes/links, are occupied with a probability p (Ben-Avraham and Havlin, 2000; Bunde and Havlin, 2012; Cohen and Havlin, 2010; Dorogovtsev and Mendes, 2002; Grimmett, 1999; Newman, 2018; Stauffer and Aharony, 2018). Lattices are mainly used to model phase transitions, e.g. the conversion of water into ice and vice versa. Networks model many infrastructures like the Internet, power grids and transportation systems, as well as social relations. In such systems, as the probability p increases the occupied sites start to form clusters. At low p values, there are many small-sized, isolated clusters; as p increases, the clusters become fewer and greater in size, as many of them have merged together. When many clusters have merged into one, at a certain $p = p_c$ called the percolation threshold, the largest cluster's size becomes proportional to the size of the system and the cluster which becomes a "spanning cluster" is called a giant component. The emergence of the spanning cluster, signifies that the system has underwent a phase transition, analogous to the transition from water to ice. Percolation theory studies the emergence of the spanning cluster with respect to the probability p. In percolation phase transition, the spanning cluster emerges at the percolation threshold of a critical probability p_c; below that threshold it does not exist, whereas above the threshold it does.

A similar concept is also applicable in non embedded networks (Barabâsi et al., 2002; Cohen and Havlin, 2010; Dorogovtsev and Mendes, 2002; Newman, 2018; 2001b). As in the case of lattices, in static networks the analogous of the spanning cluster would be a connected component that contains a finite constant fraction of the network's nodes. This connected component is called the Giant (connected) Component (GC). Similarly, site percolation in networks can be described as the random removal of a network's nodes and corresponding edges, with probability p. As more nodes are being removed, the network disintegrates into a sea of finite clusters. The phase transition (also known as percolation transition) between these two phases, i.e., the one in which a giant component exists, and the one in which the network is entirely fragmented and non-functional, is defined by the percolation threshold p_c.

3. Method of calculation - Results

In this section we outline our investigation and the methodology we followed in order to explore all the questions posed in the Introduction, and present all our results.

3.1. Basic analysis of the static network

One of the hallmarks of the patents data[2] is the fact that the vast majority of the patents are filed by individual applicants. Out of the 2,502,311 patents included in the 35 years of data, just 6.2% (154,474 patents) result from collaborative activities. A basic analysis was performed on the extracted aggregated, static network. It was found to have 429,359 nodes (applicants), consisting of 306,238 isolated nodes (components of size 1) and 123,121 nodes that are connected by 151,474 links (patents). The degree distribution, with a slope of -2.33 and the component size distribution, with a slope of -3.89 are depicted in Fig. 1. The maximum and average degree are 852 and 4.05, respectively, whereas the average path length and the network diameter are 5.53 and 27. The Largest Connected Component (LCC) is made up of 34,214 (28%) nodes and 69,246 links. The Second Largest Connected Component (SLCC) consists of a mere 93 applicants. The isolated applicants of the patents network were omitted in all parts of the subsequent analysis for which they were impertinent.

[2] For details about the data used, see Appendix B.

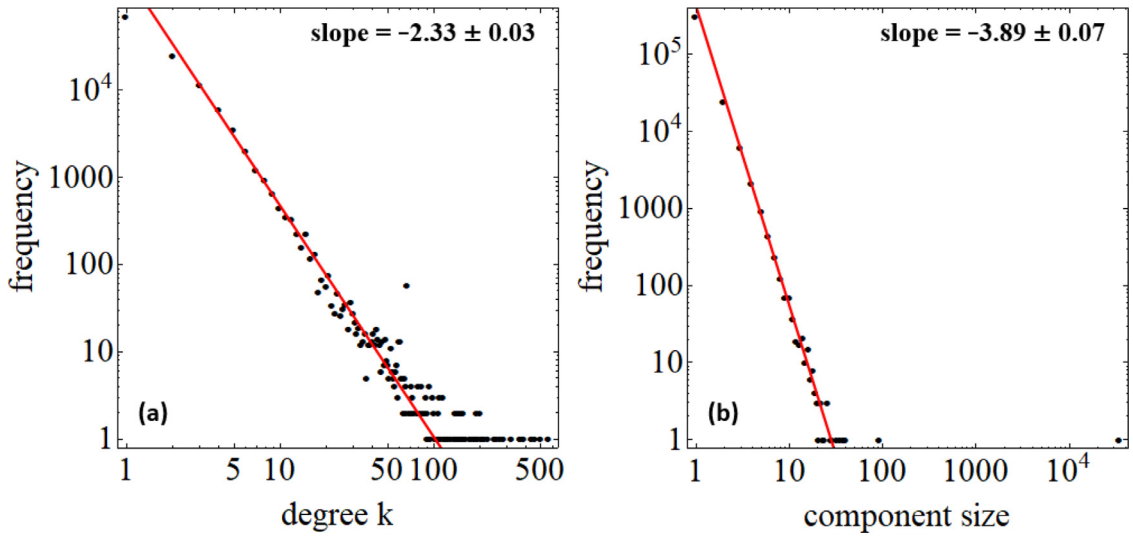

Fig. 1. (a) Degree and (b) component size distributions of the aggregated, static network.

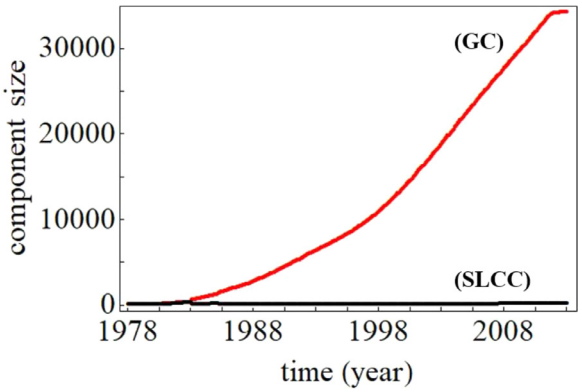

Fig. 2. Size of a) the largest (red) and b) the second largest (black) connected components of the growing network, throughout the available timeline (1978 - 2013).

3.2. Basic analysis of the growing network

In this subsection we introduce the "time-window of sliding origin" concept that enables us to perform a preliminary dynamic analysis on the network. Specifically, we study the conditions of the GC formation in multiple points in the 35 years of the data timeline and find the first evidence of changes in said conditions over time.

3.2.1. The emergence of the giant component (GC)

The massive difference between the LCC and the SLCC hints that the LCC is also the Giant Component (GC) of the network. The GC of a network is a connected component which size is proportional to the number of the network's nodes, in other words the GC grows with the network (Albert and Barabási, 2002; Barabâsi et al., 2002; Newman, 2001b; Newman et al., 2002). Fig. 2 depicts the sizes of the largest and second largest connected components of the growing network, over the 35 years of the data. To obtain this result, we replicated the network's growth, i.e. we progressively added the patents to the system, in a non-descending chronological order. Once a new patent, with at least two applicants[3] is submitted, it contributes to the network's growth by either adding new nodes (applicants) to existing components, forming new components, or merging existing components into bigger ones. Fig. 2 confirms that the LCC is in fact the GC of the network, as it clearly grows with the network.

The growth process was found to be consistent with that reported in similar previous studies (Bettencourt et al., 2009; Liu et al., 2015; Liu and Xia, 2015; Newman, 2001b; Perc, 2010). Early on, a collection of small-sized components are formed. During this phase

[3] All one-applicant patents were previously extracted from the data set, as these represent island nodes or self-loops that do not add to any component of the growing network.

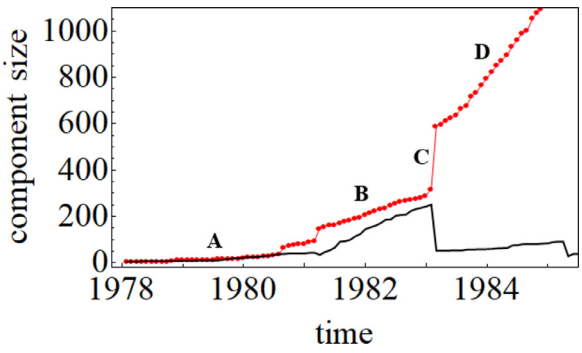

Fig. 3. The emergence of the giant component. Starting date of the system's temporal reconstruction: 1978, LCC: red circles, SLCC: black line. Phase A: Small-sized, indistinguishable components, phase B: the two major components become discernible, phase C: the percolation threshold, the GC forms, phase D: the GC outgrows all other components.

all components are indistinguishable with respect to their size. A second phase follows that leads up to the percolation threshold, during which one can discern two major components of comparable size, growing at approximately the same rate.

At the percolation threshold, a single patent that enters the system introduces a critical link that joins these two components into one; the growing GC (Albert and Barabási, 2002; Barabâsi et al., 2002; Newman, 2001b; Newman et al., 2002). From this point onward, the vast majority of new patents join new or existing nodes/components to the GC. Consequently, the evolving GC rapidly outgrows any other component in the system and a vast gap between its size and that of the Second Largest Connected Component (SLCC) is quickly forged, as seen in Fig. 2. Fig. 3 highlights the growth process outlined above, depicting the emergence of the GC, up until approximately two years after the percolation threshold. It appears that it takes approximately five years for the two major independent groups of applicants to connect and form the GC.

What are the implications of the network's percolation during the growth process? In a static network, each component represents an independent group of applicants. The links in a component represent the collaborations between its applicants. Through these links, diffusion and flow of knowledge can be realized and, consequently, applicants can influence each other, either directly or indirectly. Nonetheless, knowledge is trapped within the bounds of a component, therefore, the more fragmented a network is, the less it can facilitate collaboration and knowledge spreading. In a growing network, the addition of the critical patent at the percolation threshold invokes the birth of the Giant Component, which marks the beginning of the transformation of the system from a sea of small-sized, independent components - islands of "localized" patent-induced collaborations - to a tangible network that forms the grounds for more "globalized" collaborations. Thus, it is of great interest to study the circumstances that pertain to the GC formation and how these change in time.

Performing a temporal analysis on the GC-formation specifics would enable statistical inference and the detection of patterns or trends likely to reveal useful information about its history and evolution. To this end, we employed the "time-window of sliding origin" concept, which allows us to exploit all of the available 35 years of patents data. This concept enables us to study aspects of the GC formation like the variation of time and number of patents required for reaching the percolation threshold, as well as certain characteristics of the two major groups of applicants, immediately before their union, i.e. immediately before the birth of the GC.

3.2.2. The "time-window of sliding origin" concept and the GC formation throughout the timeline

By starting the temporal reconstruction of the system from June 1978[4] we make the tacit - and erroneous - assumption that no patents were filed before this date. Nonetheless, this false assumption allows us to open up a "window" on the temporal evolution of the system and examine the circumstances pertaining to the emergence of the GC at this particular point in the data timeline.

Clearly, one could start the reconstruction from any arbitrarily chosen date and open up a new window from that date to study the GC formation. Thus, by progressively moving forward (sliding) the starting date (origin) of the time-window, we end up with a set of time-windows that form a chain of snapshots of the evolving system. This notion embodies the "time-window of sliding origin" concept that permits us to exploit all of the available 35 years of data and perform a temporal analysis on the circumstances under which the GC forms, throughout the available time span.

Fig. 4 depicts the emergence of the GC captured on a sample group of time-windows[5] It is evident that the time required for the GC formation varies with the starting date of the window. Specifically, it appears that the GC takes significantly longer to form in the early windows than in the windows that start later in the timeline.

3.2.3. Variation of time and patents required for the GC formation through time

The findings of Fig. 4 clearly warrants a more detailed analysis. Therefore, we examined the time required to reach the percolation threshold on a larger set of time-windows densely distributed in the timeline of the data. The results are shown in Fig. 5(a), which

[4] The earliest date in the data.

[5] All percolation thresholds were confirmed by the method described in Appendix C.

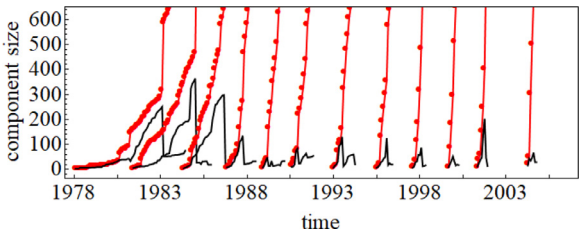

Fig. 4. The emergence of the GC, in different points in time. Each GC corresponds to a different "time-window", opened up on the system's timeline. Red circles: LCC, black lines: SLCC. The circumstances of the GC emergence appear to be changing over time.

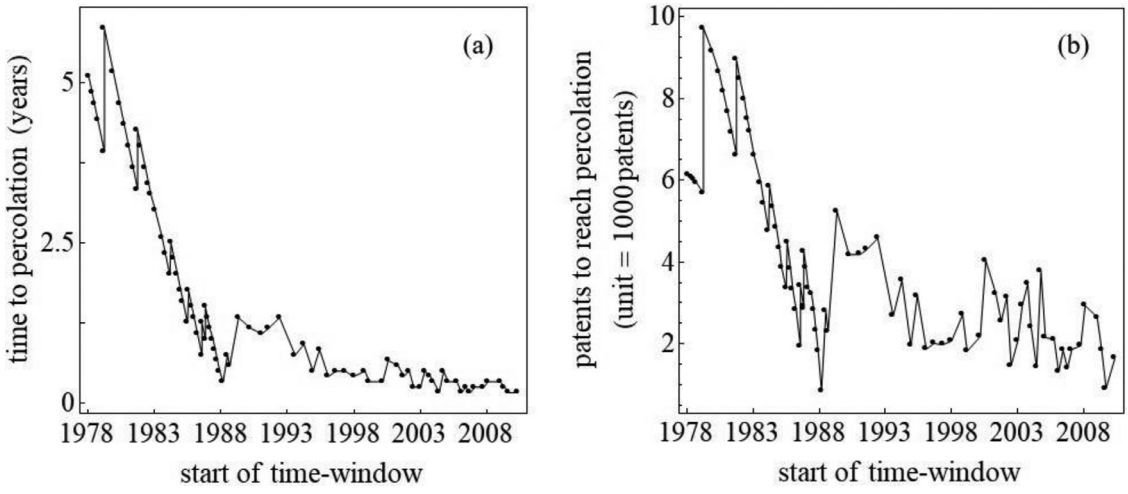

Fig. 5. (a) Elapsed time to percolation and (b) patents filed until percolation vs. starting date of sliding time-window. Both figures hint on a division of the timeline into at least two - if not three - regimes.

depicts the time elapsed between the starting date of each time-window and the date on which the GC forms vs. the starting date of the time-window.

It is evident that it takes a considerably larger amount of time for the GC to form when the origin of the time-window falls into the early years of the timeline, i.e. roughly the first decade (late 70s to late 80s). Regarding the remaining years (late 80s to early 10s), there is an indication of further division into two more regimes (late 80s to middle 90s and middle 90s to early 10s), albeit a weaker one. The results for the number of patents required to reach the percolation threshold are very similar, as shown in Fig. 5(b).

3.2.4. Qualitative differences during the growth process

The system's behaviour, during the growth phase leading up to percolation, also reveals qualitative variations that separate the timeline into at least two - possibly three - regimes. This phase was portrayed in detail in Fig. 3.2.1, for the time-window whose origin is the first date of the data. In that case, the percolation transition requires a considerable amount of time (circa 5 years), during which the two largest evolving components grow independently of each other, until the critical patent enters the system and they merge abruptly, Fig. 3. The resulting component is the network's GC, which continues to grow and quickly becomes vastly larger than any other component. As mentioned before, similar descriptions of this growth process have been previously recorded (Bettencourt et al., 2009; Liu et al., 2015; Liu and Xia, 2015; Newman, 2001b). The exact same qualitative behaviour is observed in all time-windows with starting dates within the early period (late-70s to late-80s, Fig. 4).

However, this regularity breaks down during the second period (late-80s to mid-90s), as time-windows that reveal qualitatively different behaviour are noticed sporadically for the first time. In those windows, it appears that the LCC merges with the SLCC[6] on more than one occasion, long before the merge that brings on the actual percolation, when both components are still small-sized. As a consequence, when the percolation threshold is reached, the LCC is already noticeably larger than any other component and therefore the percolation transition is rather unremarkable.

Overall, the evolving system's behaviour regarding the percolation transitions, in the three time-periods, can be described as follows: During the first period all percolation transitions are "clean" and unambiguous and require considerable amounts of time. In the second and third period, the percolation transitions are typically shorter than those of the first period. Additionally, the second

[6] Each time the LCC merges with the SLCC, the component which at that time ranks third (with respect to its size) becomes the new SLCC of the network.

period introduces time-windows, which feature step-wise and therefore less dramatic percolation transitions. These gradual transitions become even more frequent during the third period. Given that we were mainly interested in comparing the characteristics of the two groups of applicants represented by the two largest components at the percolation threshold, we only considered time-windows with marked percolation transitions at which the two largest components are of similar size, in all subsequent calculations. At any rate, these windows are densely distributed throughout the timeline and therefore there is no loss of generality.

3.3. Advanced analysis of the growing network: Investigating the role of the technological areas of the patents and the geographical origin of the applicants

Three features of the GC formation process have already been examined and were found to change with time. These are: the amount of time and patents required for the GC to emerge (Fig. 5) and - from a qualitative standpoint - the percolation transitions and the growth process leading up to it.

To further examine the conditions of the GC's formation, we investigate the possible involvement of two other features, i.e., the technological areas of the submitted patents and the geographical origin of the applicants. We define the "adjacent-pre-percolation" state as the state of the LCC and the SLCC, immediately before the percolation threshold, i.e. exactly before the critical (red-bond) patent - the patent which joins them into the GC - is added. Then, we compare the two groups of applicants that correspond to the LCC and the SLCC in their adjacent-pre-percolation state, in terms of a. the technological areas of the patents and b. the geographical origin of the applicants, in multiple time-windows.

3.3.1. Technological proximity of the two major GC groups of applicants (LCC and SLCC in the adjacent-pre-percolation state) over time

To assess the technological proximity of the two groups, we utilized the International Patent Classification (IPC) codes International patent classification (IPC) of the patents. The IPC codes are used by the European Patent Office (EPO) as a means for classifying the patents with respect to their relative technological areas and conversely, have been used as a tool for determining the technological areas of a patent (Choi et al., 2015; Jun and Lee, 2014; Park et al., 2015). There are four types of IPC categorization (from the most coarse-grained to the most refined): Section, Class, Subclass and Group (Main and Subgroup). Each of these categorizations features a set of IPC codes. The Section IPC categorization was used in all calculations, as it is the most coarse-grained of all, with many patents falling into each of the eight sections and therefore yields the most (statistically) reliable results. The databases used in this study provide the 8th edition IPC codes, for each patent.

We determined the frequency distribution of the eight IPC sections for the two groups of applicants (LCC and SLCC in the adjacent-pre-percolation state) that join to form the GC, for multiple time-windows. The degree of similarity/dissimilarity of these distributions serves as a measure of the technological proximity between these two groups. It was found that the technological proximity varies over time (Fig. 12 in D). Specifically, in most of the early period time-windows the distributions overlap significantly, in the middle period they tend to be less comparable and finally in the majority of the late period windows they tend to differ considerably.

In order to quantify the technological proximity of the two groups of applicants, we employed the normalized Euclidean distance, d, between the two 8-dimensional vectors u and v corresponding to the IPC sections frequencies of the two groups:

$$d = \frac{1}{2}\frac{|u' - v'|^2}{|u'|^2 + |v'|^2} \tag{1}$$

where $u' = u - \bar{u}$, $v' = v - \bar{v}$, $\bar{u} = \frac{1}{N}\sum_{i=1}^{N} u_i$, $\bar{v} = \frac{1}{N}\sum_{i=1}^{N} v_i$ and $N = 8$, the eight IPC sections.

The calculated Euclidean distance, depicted in Fig. 6, varies over time in a way that is once more suggestive of a three-period division of the timeline. This division is by no means clear-cut, nonetheless, it is evident that high technological distance between the two components became progressively more probable through the three periods. A plausible approximate partition would be 1978–1986, 1986–1995 and 1995–2013, shown in Fig. 6 in red, blue and green, respectively, for visualization purposes. This provides us with a rough estimation of the Euclidean distance in the three periods: from an average of 0.086 with a very low variance (0.001) in the early windows, to 0.196 with a higher than tenfold leap in its variance (0.013) during the middle ones, to 0.253, which is almost three times higher than that of the early period and with a higher still variance (0.022), in the late windows.

Both the number of patents required to reach percolation (Fig. 5b) and the technological distance of the two major components in the adjacent-pre-percolation state (Fig. 6) have been shown to change over time in a manner that separates the timeline into three regimes. To probe into the relation between these two quantities, the two figures were combined into one, Fig. 7, by eliminating the common variable, i.e. the time. We used the same approximate partition of the timeline as in Fig. 6 to denote the period of each time-window (early, middle and late).

In Fig. 7, the points representing time-windows of the same region tend to cluster together, which further supports the findings of Fig. 5 and 6 regarding the division of the timeline into at least two - if not three - regimes. Moreover, Fig. 7 hints at an association between the two factors (the technological proximity of the two groups of applicants and the amount of patents required for the GC formation), which is strong for the time-windows of the early period and quite weaker for the rest of the timeline. In the early period time-windows, requiring a high amount of patents to reach percolation most likely co-exists with low Euclidean distance (high technological proximity) between the two major groups of applicants that constitute the GC. For the rest of the timeline, the lower the amount of patents the more likely high technological distance becomes, however there is still a good chance for low technological proximity in many time-windows of low amount of patents. Overall, as the system evolves, it appears that it undergoes notable changes that are reflected in both these factors pertaining to the GC formation. Namely, the average value of both quantities shifts, as the system moves through the three regimes, according to the following pattern: from a period characteristic of low Euclidean

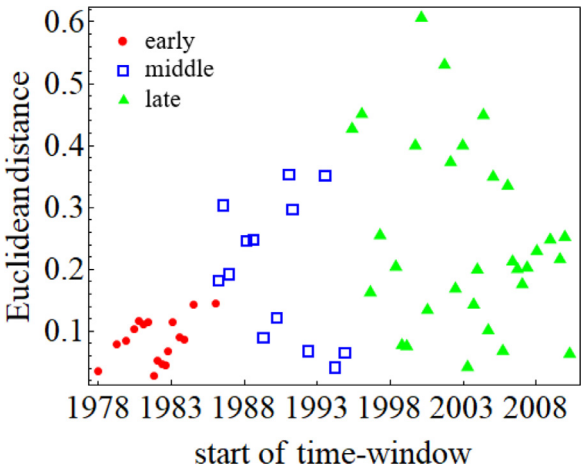

Fig. 6. Normalized Euclidean distance of the IPC codes distribution, of the two major components that make up the GC, in their adjacent-pre-percolation state for multiple time-windows in the available timeline. Red disks, blue squares and green triangles are used to visualize a rough division of the timeline into the early (1978–1986), middle (1986–1995) and late (1995–2013) periods of the timeline. Low technological distance is highly likely in the early years. High technological distance appears to be increasingly more likely in the middle and late years.

distance and high number of patents, to a period of medium-high Euclidean distance and number of patents, and, finally, to a period of higher Euclidean distance and lower number of patents.

Similar results were obtained with the use of Pearson correlation coefficient, r, as a measure of the technological proximity between the two components that form the GC, when plotted against the number of patents required to reach percolation, Fig. 13 in Supplementary Figures. The null hypothesis is that the two sets (of IPC codes distribution) are independent, with the significance level set to 0.05.

3.3.2. Geographical interplay between the two major groups of applicants (LCC and SLCC in the adjacent-pre-percolation state) over time

The geographical breakdown of the applicants in both groups in the adjacent-pre-percolation state, for multiple time-windows, revealed a series of interesting facts. First, the bulk of the applicants in both groups were found to be operating in just three countries, namely, France (FR), Germany (DE) and Japan (JP), at all times. Applicants from the United States (US) are also present, with a much lower yet non-negligible contribution. Second, for all time-windows investigated, one of the two groups mostly comprises FR-DE applicants, while the other of JP ones. Third, the level of geographical confinement of the two groups appears to change over time, in a way that is once again implying a three-region division of the timeline (Fig. 14 in Supplementary Figures).

In light of these findings, we classified the country occurrences into four categories: all residing in Europe (EU), Japan (JP), United States (US) and all the rest countries in the world (REST). The percentages of these four categories, calculated for both groups, illustrate the geographical interplay between the two applicants' groups. The results, shown in Fig. 8, corroborate the three-part aforementioned division of the 35-year timeline.

During the first regime (ca. 1978–1986), the applicants in the two groups are clearly segregated, confined either in EU or JP, while US and rest of the world applicants are practically absent. Unlike the first regime that is effortlessly distinguishable from the rest, the distinction between the second (ca. 1986–1995) and the third (ca. 1995–2013) regimes is significantly less prominent. Both regimes exhibit geographical interactions between the two groups, in stark contrast to the complete segregation witnessed during the first regime. The intensity of these interactions however is much lower in the second regime. In particular, while there is a notable rise in both EU and US percentages in the "mostly JP" group, only the US percentage rises in the mostly "EU" group; Japanese applicants are yet absent from this group during the second period. Lastly, applicants from the rest of the world join in the network for the first time in the last two regimes. The contribution of these applicants is larger in the third regime in both groups, nonetheless it remains rather small at all times.

These results prompted a supplementary statistical analysis on the geography of the raw patent data from which the network examined so far was derived. As mentioned before, this is the data set that comprises all patents, excluding the ones with just one applicant[7] First, we determined the monthly average of the number of applicants[8] per patent (Fig. 16 in Supplementary Figures), which remains remarkably constant, averaging ~2.25 throughout the timeline. This quantity was subsequently broken down into the four aforementioned geographical categories, EU, JP, US and the rest of the world (REST), in Fig. 9.

During the first ~12-14 years of the timeline, there is a sharp decline in the EU applicants, which coincides with a corresponding rise in the JP applicants. During the same period, the number of US applicants is pivoting around a low but non trivial value, while

[7] These patents result in isolated (island) nodes or self-loops. For results for the network with one-applicant patents included see Fig. 15 in Supplementary Figures.

[8] Given that these are the raw patent data, the applicants in this analysis are not unique, since an applicant can participate in multiple patents, throughout the timeline.

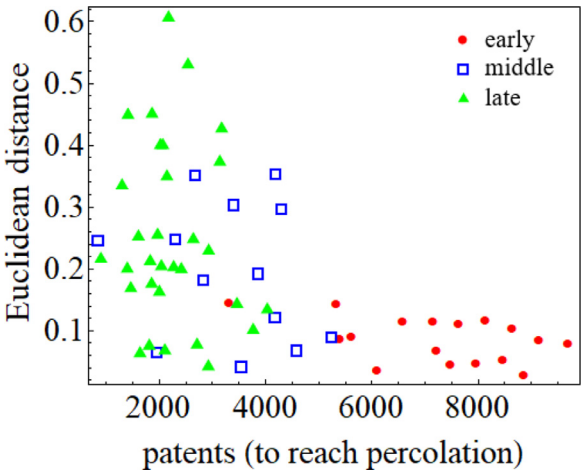

Fig. 7. Normalized Euclidean distance of the IPC codes distribution, of the two major clusters that make up the GC, in their adjacent-pre-percolation state, vs. the required number of patents to reach the percolation state, for multiple time-windows in the available timeline. Red, blue and green colour is used to visualize a rough division of the timeline into the early (1978–1986), middle (1986–1995) and late (1995–2013) periods. A conjunction of low technological distance and high number of patents is highly likely in the early years. The opposite conditions appear to be increasingly more likely in the middle and late years.

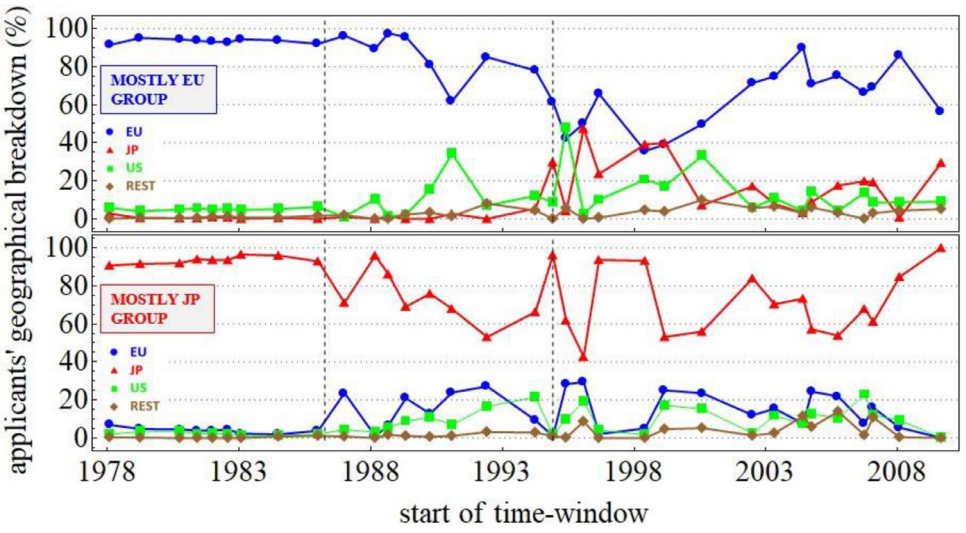

Fig. 8. Geographical breakdown of the groups of applicants corresponding to the LCC and SLCC, as captured exactly before the addition of the critical bond which joins them into the GC (adjacent-pre-percolation state). At (almost) all times, one group consists of mostly EU applicants, while the other of mostly JP. The depicted rough division of the timeline (two vertical lines) is the same as in Fig. 6 and 7 (EU: blue disks, JP: red triangle, US: green squares, rest of the world: brown rhombi) The geographical overlap between the two groups of applicants appears to increase over time.

the REST applicants barely exist. For the next 15 years, US and JP applicants participate at approximately the same - quite stable - rate, while the EU applicants continue to decline, albeit at a much slower rate, which now coincides with a slow but steady rise in the REST applicants. Finally, for the last 5-6 years the US and JP applicants count fall just enough for the rest of the world applicants to reach them, whose participation continues its steady rise. Overall, it appears that non-European applicants' participation in the EPO patents increases over time, which leads to more opportunities for the formation of inter-continental collaborations.

Finally, to complete this geographical analysis, we look into the static network, i.e. the network that results from all the patents aggregated over the whole 35 years of data, excluding only the patents with just one applicant[9] The geographical breakdown of the applicants in that network and its GC is shown in Fig. 10.

[9] As mentioned before, these patents result in isolated (island) nodes or self-loops. For results for the network with these patents included see Fig 17 in Supplementary Figures.

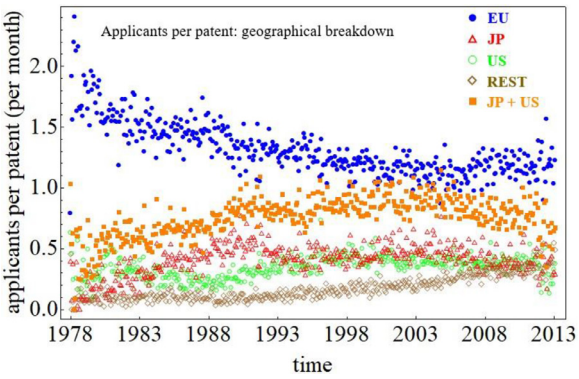

Fig. 9. Geographical breakdown of the number of the applicants over submitted patents, per month, throughout the timeline. (EU: blue disks, JP: red triangles, US: green circles, rest of the world: brown rhombi, JP + US: orange squares) Non-European applicants' participation in applications submitted to the EPO appears to increase over time.

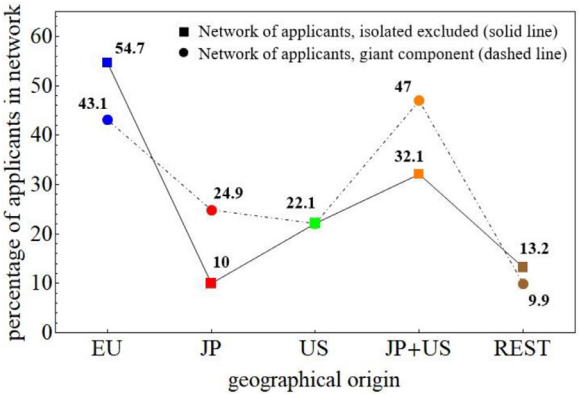

Fig. 10. Geographical breakdown of the applicants in the network resulting from the time-aggregated data, spanning all available timeline (squares: the network of applicants, isolated nodes excluded, disks: the GC of the network). US and JP play a significant role in the EPO applicants network. JP persistently contributes to the network's coherence.

Fig. 10 complements the findings of Fig. 8 and 9 regarding JP as it corroborates the crucial part that it plays in the core of the EPO patents network. Unlike the US applicants that submit patent applications indiscriminately, JP applicants principally participate in patent collaborations that contribute to the GC of the network.

3.4. Are the major patent-contributors more likely to participate in a critical patent?

Last but not least, we explored the question of whether the likelihood of participating in a critical patent increases with the total amount of patents that an applicant contributes to the network or with its size. The latter is also related to the firm size vs. innovation output problem (Schumpeter, 1934; 1942) (see Introduction).

To this end, we ranked the applicants with respect to their overall patents production, over several time-windows distributed in the 35-year timeline and assessed the relative position of applicants participating in the critical-patents. Thus, for each time-window, we brought the network into the percolation threshold state and determined the percentage of the patents for all applicants having filed at least one patent[10] from the starting date of the time-window to the percolation threshold. For each time-window we calculated the percentage p_a for applicant a according to the formula:

$$p_a = \frac{patents\ filed\ by\ applicant\ a\ in\ t_p}{total\ of\ patents\ filed\ in\ t_p}\ x\ 100 \tag{2}$$

where t_p is the time elapsed from the starting date of the time-window to the date of the critical patent (percolation threshold). We subsequently evaluated the applicants' rank and percentile rank, according to their patent percentage, which are shown in Table 1 for the red-bond firms of nine representative time-windows in the data timeline.

[10] All patents are included in this analysis, even those with only one applicant.

Table 1
Rank and percentile rank according to patent percentage, for the red-bond firms of nine representative time-windows in the data timeline.

year of time-window	critical patent applicants	rank	percentile (%)
1978	*applicant 1*	17	99.75
	applicant 2	115	98.30
1983	*applicant 1*	77	98.91
	applicant 2	87	98.77
	applicant 3	91	98.71
	applicant 4	374	94.70
	applicant 5	2044	71.02
1989	*applicant 1*	2	99.93
	applicant 2	10	99.64
1993	*applicant 1*	38	90.28
	applicant 2	521	99.64
1997	*applicant 1*	29	98.99
	applicant 2	586	79.50
2000	*applicant 1*	189	96.60
	applicant 2	584	89.49
2004	*applicant 1*	58	97.09
	applicant 2	109	94.54
2007	*applicant 1*	5	99.80
	applicant 2	249	90.01
2010	*applicant 1*	71	94.95
	applicant 2	71	94.95

The firms in the top 1% were found to submit nearly half the patents ($\sim$ 48%) on average, yet, they participate in the red-bond patents on an average of only $\sim$ 24%. Therefore, we conclude that the top patent-producing applicants do not dominate the production of red-bond links that bring the GC into existence and induce the network's coherence, as one would expect, with respect to their overall patent output. Our results suggest that the reason behind this discrepancy is that many of the firms in this category (the top 1%), have a low percentage of collaborative patents. Thus, the likelihood of them yielding a critical patent is reduced in spite of them being highly productive on the whole.

Finally, the plausible assumption that the larger a firm's size the more patents it is capable of producing (Tether, 2002) is overall corroborated by our results, as the bulk of the top patent-producing applicants are large-sized ones. Additionally, the majority of the red-bond applicants (Bosch, Toyota, Hitachi, Honda, Nissan, Phillips, to name a few) in this network are found to be large-sized firms, regardless of their total patent output in each time-window.

These findings have further implications re the firm size vs. innovation-output debate (see Discussion).

4. Discussion

Overall, our results reveal that as the patent network evolves through time, the percolation threshold (on average) comes at shorter times, requires fewer patents, features increasing inter-regional collaborations and increasing technological distance between the two major groups of applicants involved. How do all these fit together?

It is plausible to presume that the basic underlying factor in this change is the rise in the inter-continental collaborations and the resulting globalization, through the decades. Thus, in the early years - when advanced communication technologies were not yet widespread - geographical distance was most likely hindering inter-continental collaborations and therefore the two groups were segregated for many years until their eventual union. It is likely that this is the reason behind the high technological similarity between the two groups as they were forced to function independently for long periods of time. During the transitional phase of the middle years, a change develops that appears to mitigate the negative effect of distance on inter-continental collaborations. Therefore, windows of shorter time and higher technological distance between the two groups start to emerge. Lastly, the shortest intervals to percolation and higher technological distance are seen recurrently in late-years windows, when geographical interplay is at its highest.

This conclusion is in agreement with similar results found by studying patent collaborations with statistical methods (Guan and Chen, 2012; Ma and Lee, 2008). A number of factors could be identified as plausible contributors to this globalization phenomenon, such as the vastly available air-travel, the remarkable upsurge in the use of internet and various other socio-economic factors that forced the applicants to reach out further to build complementing collaborative ties and meet the demands of an increasingly competitive and fast-paced world.

Moreover, regarding the firm size vs. innovation-output debate, our results imply that large firm size favours innovation activities. First, the assumption that the top patent-producing applicants are also likely to be large-sized ones (Tether, 2002) was corroborated by our analysis. Second, the majority of the red-bond applicants (Bosch, Toyota, Hitachi, Honda, Nissan, Phillips, to name a few) in this network were found to be large-sized firms regardless of the total amount of patents they contribute in each time-window. Since a patent is in itself an innovation output indicator and furthermore, the critical bonds of a network are believed to correspond to a highly innovative events (Bettencourt et al., 2009), we infer that in the EPO patents network, it is the large-sized applicants (irrespective of the magnitude of their total patent contribution) that are most likely to introduce radical innovation. Therefore

our findings are suggestive of a positive correlation between size and innovation output, as in the majority of the studies reviewed in Becheikh et al. (2006), as well as in Schumpeter (1942); Tether (1998).

Lastly, the striking presence of Japan in the EPO patent network should not go unnoticed. In all time-windows examined, the Japanese applicants make up a significant portion - nearly half - of the GC, at the moment of its emergence, i.e. at the percolation threshold (Fig. 8). Furthermore, the fact that Japan seems to forge strategic alliances that are key to the network's coherence also surfaced in two more of our findings. Firstly, Japan's participation in the static, time-aggregated network never exceeds 10% (see Fig. 10 and Fig. 17), while it reaches 25% in its GC. Secondly, an abundance of Japanese applicants (Toyota, Hitachi and subsidiaries, Honda, Nissan, Ube industries, JSR corporation, Riken, the National Institute of Advanced Industrial Science And Technology etc.) consistently appear in red-bond patents, namely the patents that induce the GC emergence. Interestingly enough, judging from the type of activities of these companies and from the red-bond patents themselves, it appears that they are all related, in one way or another, to the automobile industry. Thus, we believe that the automobile industry has played a crucial part in the EPO patent network and consequently to the introduction and - even more so - to the diffusion of innovation.

5. Conclusions

After identifying the existence Giant Component (GC) as key to the diffusion and promotion of innovation procedures, we posed the following questions:

- Does the GC exist in every snapshot and if yes, how long does it take to form?
- Do the conditions the GC's formation change over time?
  - Do the technological areas of the patents influence the GC's formation?
  - Does the geographical origin of the applicants influence the GC's formation?
- Are the major patent contributors (large-sized applicants) more likely than the ones with medium to small contributions (small/medium-sized applicants) to produce a critical patent?

To explore these questions, we performed a temporal analysis, on the patent applicants' collaboration network derived from the REGPAT patents data, which span 35 years, from year 1978 (June) to 2013 (July). Specifically, we studied the network over a collection of time-windows, which allowed us to exploit the whole timeline of 35 years and study the evolution of the applicant's collaboration network dynamically, instead of limiting the investigation to the aggregated, static network. This approach enabled us to open up multiple windows of observation onto the system's evolution, with starting dates distributed throughout the timeline. We focused our analysis on the GC, and specifically on certain characteristics of its formation, i.e. the network's percolation, in multiple time-windows. This analysis, uncovered evidence of qualitative and quantitative differentiations of characteristics such as, the amount of time/patents required for percolation, the technological similarity and the geographical overlap of the major groups of applicants which make up the GC and the abruptness of the percolation transition. All these aspects of the GC formation are found to change in time in a way that is suggestive of a three-regime division of the timeline.

Specifically, during the first period (ca. 1978–1987, early-years), we observe clear-cut percolation transitions, in which the two largest, separately growing components are merged abruptly into the network's GC. The groups of applicants represented by these two components are found to exhibit high technological resemblance, based on the IPC codes of their corresponding patents. Furthermore, in all early-years windows, the two groups are almost completely geographically segregated into two regions, EU (principally FR and DE) and JP, with a very small, but nonetheless measurable US fraction of applicants being present in both communities. Moreover, this period is marked by the highest amounts of time/patents required for the GC formation.

During the second period (ca. 1987–1995, middle years), there is a shift in the percolation transition process, in many windows. In such windows, the transition appears to be shorter in time, more gradual and less striking. Overall, during this period, the two groups of applicants that make up the GC are more complementary and less similar technologically, and begin to exhibit some geographical overlap. Specifically, the community comprising mostly JP applicants, starts to welcome collaboration with both the EU and the US, however, the respective EU community still remains mostly detached. Additionally, the amounts of both time and patents required for the GC formation lessens, on average.

In the third period (ca. 1995–2013, late years), even more windows display the less-abrupt, less outstanding percolation transition behaviour. Overall, it is the period in which the two groups of applicants demonstrate the greatest technological distance, as well as the greatest geographical overlap. There is a notable rise in both frequency and intensity of collaborations between Europe, Japan and the US, in both groups. Furthermore, in this period the windows displaying the least amount of time/patents required for percolation have become recurrent.

Moreover, our results indicate that top patent-producing applicants are likely to yield a critical patent, however at a rate significantly lower than their overall patent production. Additionally, the top patent-producing applicants are predominantly large-sized firms, supporting the assumption that the largest the firm size the more patents it is capable of producing (Tether, 2002). Finally, the red-bond applicants (Bosch, Toyota, Hitachi, Honda, Nissan, Phillips amongst others) are also large-sized firms, notwithstanding the amount of patents they contribute in each time-window. Therefore, regarding the GC formation and the network's coherence, the significant applicants are typically large-sized ones, but not necessarily amongst the top patent-producing.

Last but certainly not least, essentially all of our findings stress the vital importance of Japan to the GC of the EPO patent network. Firstly, in all time-windows, one of the two major groups that join into the GC consists mainly of Japanese applicants. Secondly, Japanese applicants make up one quarter of the static, aggregated network's GC (while they constitute just 10% of the whole network). Thirdly, Japanese firms and research institutes are found to systematically participate in red-bond patents. Toyota,

Hitachi and subsidiaries, Honda, Nissan, Ube industries, JSR corporation, Riken, the National Institute of Advanced Industrial Science And Technology among others are consistently featured in the critical patents.

CRediT authorship contribution statement

Maria Tsouchnika: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Alex Smolyak:** Methodology, Investigation, Writing – review & editing. **Panos Argyrakis:** Conceptualization, Writing – review & editing, Supervision. **Shlomo Havlin:** Conceptualization, Writing – review & editing, Supervision.

Acknowledgment

This work was supported by the European Commission FET Project MULTIPLEX No. 317532. We thank Michael Kanetidis for his valuable comments.

Appendix A. List of acronyms and abbreviations

DE	Germany
EPO	European Patent Office
EU	Europe
FR	France
GC	Giant Component
IPC	International Patent Classification
JP	Japan
JPO	Japan Patent Office
LCC	Largest Connected Component
NUTS	Nomenclature of Territorial Units for Statistics
OECD	Organization for Economic Co-operation and Development
PCT	Patent Cooperation Treaty
REST	All countries except Japan, US and all European ones
SLCC	Second Largest Connected Component
US	United States
USPTO	United States Patent and Trademark Office

Appendix B. The European Patent Office (EPO) patent application data

The theme of this study is to investigate the evolution of the collaborations resulting from - and in - any kind of innovation focusing mainly on the European region. To this end, we employed the data set of patent applications (which may be granted, rejected or withdrawn) to the European Patent Office (EPO) in order to build and study the characteristics of the EPO patent applicants' network through time. Patent applications data are more suited than granted patents data for this analysis.

The specific goals of the project involved gathering and processing certain pieces of information, such as: which applicants contributed to which patent, the geographical origin of the applicants, the patent's filing date and the corresponding technological areas of the patent. This data was derived from a combination of the EPO-subsets of two distinct databases: a) the "OECD, REGPAT database, July2014" and b) the "OECD, Triadic Patent Families database, July2014".

The OECD REGPAT is a rich database which comprises two data sets: the set of patent applications to the EPO and those filed under the Patent Cooperation Treaty (PCT). The Triadic database consists of patents filed at all three of the following patent-granting organizations: the EPO, the United States Patent and Trademark Office (USPTO) and the Japan Patent Office (JPO), by the same applicant, while referring to the same invention.

At first glance, the REGPAT database fits perfectly to the needs of this study, providing all the required information: a list of all the applicants of each patent, regional data for each applicant, the patent filing and priority dates and information about the technological areas of a patent, in the form of International Patent Classification (IPC) codes. However, the OECD REGPAT database had two major limitations that we needed to address.

The first one is that the applicants are not uniquely identified. Each new entry is assigned a surrogate key derived by the combination of three fields: name, address and country code, which are entered by the applicant. Two entries, corresponding to the same applicant, with the slightest difference in either one of these three fields - e.g. an extra comma - appear as two distinct applicants in the database (Lissoni et al., 2010). Therefore, as multiple keys are very often assigned to the same applicant, a network derived directly from this database would be very different than the real one. To find the unique applicants, we chose three fields of each entry (name, address and NUTS code) as similarity indicators. Thankfully, the address and NUTS code of an applicant of the REGPAT database are very reliable entries (Maraut et al., 2008).

We first prepared the database by applying preliminary cleaning procedures, such as removing corrupted or unwanted characters and normalizing the text by replacing umlauts, accents, etc. Then, we separated the applicants into groups according to their country code. For each applicant entry in a group, we calculated the Levenshtein distance of the three chosen fields with every other applicant

entry. Whenever a pair of applicants scored lower than a certain threshold in all three distances, the applicants were considered to be identical. By running the algorithm on a large number of random samples and carefully inspecting the results, we managed to determine a threshold value that ensures avoidance of false positives. This procedure resulted in a remarkable reduction (∼26%) in the number of unique applicants.

The second limitation of the REGPAT database is that only the year of the patent filing and priority dates are explicitly provided. This was a major drawback, as we wanted to perform a temporal analysis, which requires a fine date granularity. We managed to successfully address this limitation by matching the EPO REGPAT patent set with the EPO TRIADIC patent set, which provides the full filing and priority dates, and by taking advantage of the implicit time-stamp of the EPO application ID.

In particular, we implemented the following procedure: First, we matched the TRIADIC entries to the REGPAT entries, by using the EPO application ID and appended the first priority date and the first EPO filing date of the TRIADIC entries to the REGPAT ones. We also adequately cleaned the database of the patent families, by removing the non-prior members. Clearly, the TRIADIC EPO patent set is a subset of the REGPAT one, therefore after this step there were REGPAT entries still left unmatched. We determined the date of the unmatched entries by utilizing the fact that the EPO application ID is a seven digit serial number that holds implicit information about the chronological order of the patents and that the first EPO filing date is more consistent with the date mentioned in the EPO application ID than the first priority date.

Under these assumptions, we sorted all the entries - matched and unmatched - by two fields: the 11,546 matched first-EPO-filing dates and their EPO application ID. The result was a database of chronologically sorted patents, with a mix of matched entries and some sporadic blocks of a few unmatched entries. We inspected the distance between the consecutive sorted matched dates and noticed that the vast majority of them were only one day apart. This was clearly a very satisfactory level of granularity, therefore it was reasonable to consider the blocks of unmatched entries as having being filed on the same day of the last preceding matched entry. The network was built using this order.

Appendix C. Confirmation of the percolation threshold

For random, static networks the percolation threshold can be determined by examining the size of the second largest cluster of the network (Bunde and Havlin, 2012). Specifically, as nodes are continuously removed from the network, the size of the second largest cluster varies, and it reaches a maximum at the percolation threshold, coinciding with the decomposition of the giant component and the complete fragmentation of the network into a sea of small-sized clusters. This criterion has been repeatedly used as a means to pinpoint the percolation threshold of real networks, including scale-free ones, such as the Internet (Karrer et al. 2014; Kawamoto et al. 2015; Li et al. 2015b).

In growing (evolving) networks, percolation is regarded as the emergence of the giant component from the coalescence of dominant clusters that grew out of an initial sea of small island clusters. In the evolving network of patents that we studied, there are always two dominant clusters, the union of which marks the birth of the giant component. We regard the patent that joins these two major clusters as the “critical” or “red-bond” patent. The giant component, once formed, quickly outgrows any other cluster in the system (Fig. 2 and Fig. 3). In order to pinpoint the red-bond patent, we follow the evolution of two largest clusters at any time-step, throughout the timeline, as in Bettencourt et al. (2009); Liu et al. (2015); Liu and Xia (2015); Newman (2001a); Tomassini and Luthi (2007).

To confirm the validity of the red-bonds thus pinpointed, we employed the following criterion. The timeline is divided into two regimes: one starting from the beginning of the timeline up until right before the critical point and the other from the critical point up to the end of the timeline. We take snapshots of the evolving network, distributed in both areas, including one on and one exactly before the criticality. These snapshots are static (equilibrium) networks, on which we can apply the method of the second largest cluster size maximum, to determine the percolation threshold. Clearly, there should be a percolation threshold, p_c, denoted by the

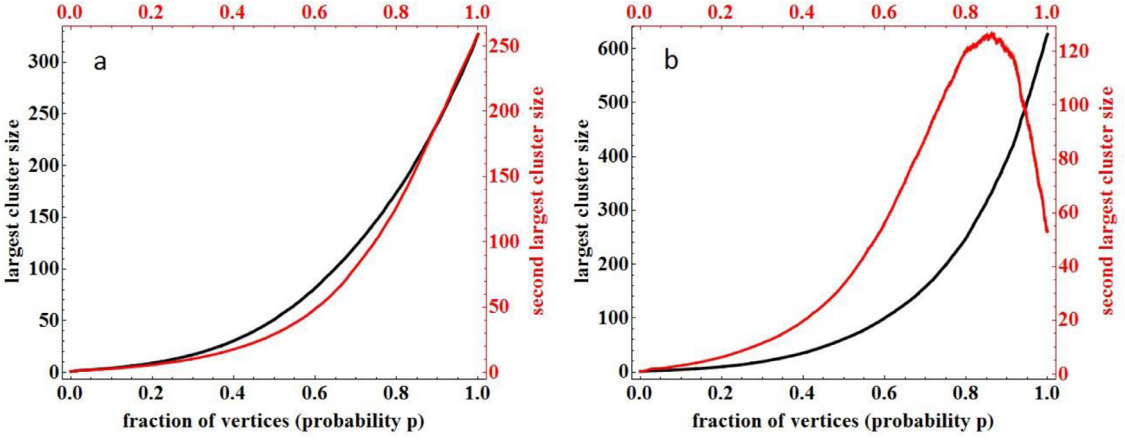

Fig. 11. Size of largest and second largest components, corresponding to snapshots of the growing network at points (a) below and (b) at and above the addition of the red-bond bond patent.

second largest cluster size maximum, for all snapshots in the second period. However, the second largest cluster should not have any maximum in any snapshot of the first period. All previously pinpointed red-bonds in this study were validated using this method. In all snapshots below the critical point, i.e. before the addition of the red-bond patent, both clusters grow continuously with p, Fig. 11a. In contrast, the size of the second largest cluster increases until it reaches a maximum and then falls, in all snapshots at and above the red-bond patent Fig. 11b. This method was also used in other studies involving growing networks (Do Yi et al., 2013; Li et al., 2015a).

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.joi.2021.101238.

References

Acs, Z. J., & Audretsch, D. B. (1990). *Innovation and small firms*. Mit Press.

Albert, R., & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of modern physics, 74*(1), 47.

Barabâsi, 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*(3–4), 590–614.

Becheikh, N., Landry, R., & Amara, N. (2006). Lessons from innovation empirical studies in the manufacturing sector: A systematic review of the literature from 1993–2003. *Technovation, 26*(5–6), 644–664.

Ben-Avraham, D., & Havlin, S. (2000). *Diffusion and reactions in fractals and disordered systems*. Cambridge university press.

Bettencourt, L. M. A., Kaiser, D. I., & Kaur, J. (2009). Scientific discovery and topological transitions in collaboration networks. *Journal of informetrics, 3*(3), 210–221.

Brown, S. L., & Eisenhardt, K. M. (1995). Product development: Past research, present findings, and future directions. *Academy of management review, 20*(2), 343–378.

Bunde, A., & Havlin, S. (2012). *Fractals and disordered systems*. Springer Science & Business Media.

Caballero, R. J., & Jaffe, A. B. (1993). How high are the giants' shoulders: An empirical assessment of knowledge spillovers and creative destruction in a model of economic growth. *NBER macroeconomics annual, 8*, 15–74.

Cainelli, G., Evangelista, R., & Savona, M. (2004). The impact of innovation on economic performance in services. *The Service Industries Journal, 24*(1), 116–130.

Cassiman, B., Golovko, E., & Martínez-Ros, E. (2010). Innovation, exports and productivity. *International Journal of Industrial Organization, 28*(4), 372–376.

Choi, J., Jang, D., Jun, S., & Park, S. (2015). A predictive model of technology transfer using patent analysis. *Sustainability, 7*(12), 16175–16195.

Cogan, D. J. (1993). The irish experience with literature-based innovation output indicators. In *New concepts in innovation output measurement* (pp. 113–137). Springer.

Cohen, R., & Havlin, S. (2010). *Complex networks: Structure, robustness and function*. Cambridge university press.

Crépon, B., Duguet, E., & Mairessec, J. (1998). Research, innovation and productivi [ty: An econometric analysis at the firm level. *Economics of Innovation and new Technology, 7*(2), 115–158.

Damijan, J. P., & Kostevc, v. (2015). Learning from trade through innovation. *Oxford bulletin of economics and statistics, 77*(3), 408–436.

Do Yi, S., Jo, W. S., Kim, B. J., & Son, S.-W. (2013). Percolation properties of growing networks under an achlioptas process. *EPL (Europhysics Letters), 103*(2), 26004.

Dorogovtsev, S. N., & Mendes, J. F. F. (2002). Evolution of networks. *Advances in physics, 51*(4), 1079–1187.

Fleming, L., King, C., & Juda, A. I. (2007). Small worlds and regional innovation. *Organization Science, 18*(6), 938–954.

Geroski, P., Machin, S., & Van Reenen, J. (1993). The profitability of innovating firms. *The Rand journal of economics*, 198–211.

Geroski, P. A. (1989). Entry, innovation and productivity growth. *The review of economics and statistics*, 572–578.

Griliches, Z. (1958). Research costs and social returns: Hybrid corn and related innovations. *Journal of political economy, 66*(5), 419–431.

Griliches, Z. (1964). Research expenditures, education, and the aggregate agricultural production function. *The American economic review, 54*(6), 961–974.

Griliches, Z. (1980). Returns to research and development expenditures in the private sector. In *New developments in productivity measurement and analysis NBER Chapters* (pp. 419–462). National Bureau of Economic Research, Inc.

Griliches, Z. (1986). Productivity, r&d, and basic research at the firm level in the 1970s. *The American economic review, 76*(1), 141–154.

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

Griliches, Z., & Mairesse, J. (1983). Comparing productivity growth: An exploration of french and US industrial and firm data. *European economic review, 21*(1–2), 89–119.

Grimmett, G. (1999). What is percolation? In *Percolation* (pp. 1–31). Springer.

Guan, J., & Chen, Z. (2012). Patent collaboration and international knowledge flow. *Information Processing & Management, 48*(1), 170–181.

Gurbiel, R. (2002). Impact of innovation and technology transfer on economic growth: The central and eastern europe experience. *Warshaw School of Economics, 162*, 1–18.

Hall, B. H. (2011). Innovation and productivity. *Technical Report*. National bureau of economic research.

Hall, B. H., & Khan, B. (2003). Adoption of new technology. *Working Paper 9730*. National bureau of economic research Cambridge, Mass., USA.

Hall, B. H., Lotti, F., & Mairesse, J. (2009). Innovation and productivity in SMEs: Empirical evidence for italy. *Small business economics, 33*(1), 13–33.

Hâncean, M.-G., Perc, M., & Lerner, J. (2021). The coauthorship networks of the most productive european researchers. *Scientometrics, 126*(1), 201–224.

Harhoff, D. (1998). R&d and productivity in german manufacturing firms. *Economics of Innovation and New Technology, 6*(1), 29–50.

International patent classification (IPC), https://www.wipo.int/classifications/ipc/en.

Jun, S., & Lee, S.-J. (2014). A small world network for technological relationship in patent analysis. In *Soft computing in big data processing* (pp. 91–99). Springer.

Karrer, B., Newman, M. E. J., & Zdeborová, L. (2014). Percolation on sparse networks. *Physical review letters, 113*(20), 208702.

Kawamoto, H., Takayasu, H., Jensen, H. J., & Takayasu, M. (2015). Precise calculation of a bond percolation transition and survival rates of nodes in a complex network. *PloS one, 10*(4), e0119979.

Kerl, A., & Moehrle, M. G. (2015). Initiatives for multi cross industry innovation: The case of universal home. In *2015 portland international conference on management of engineering and technology (PICMET)* (pp. 2223–2229). IEEE.

Khan, A., Moehrle, M. G., & Böttcher, F. (2013). Initiatives for multi cross industry innovation: The case of future_bizz. In *2013 proceedings of PICMET'13: Technology management in the IT-driven services (PICMET)* (pp. 616–622). IEEE.

Kleinknecht, A., Reijnen, J. O. N., & Smits, W. (1993). Collecting literature-based innovation output indicators. the experience in the netherlands. In *New concepts in innovation output measurement* (pp. 42–84). Springer.

Klette, T. J. (1996). R&d, scope economies, and plant performance. *The Rand journal of economics*, 502–522.

Klette, T. J., & Johansen, F. (2000). Accumulation of r&d capital and dynamic firm performance: A not-so-fixed effect model. In *The economics and econometrics of innovation* (pp. 367–397). Springer.

Klomp, L., & Van Leeuwen, G. (1999). The importance of innovation for company performance. *Netherlands Official Statistics, 14*(2), 26–35.

Landau, R., & Rosenberg, N. (1986). *The positive sum strategy: Harnessing technology for economic growth*. Washington, DC: The National Academies Press.

Li, D., Fu, B., Wang, Y., Lu, G., Berezin, Y., Stanley, H. E., & Havlin, S. (2015a). Percolation transition in dynamical traffic network with evolving critical bottlenecks. *Proceedings of the National Academy of Sciences, 112*(3), 669–672.

Li, D., Zhang, Q., Zio, E., Havlin, S., & Kang, R. (2015b). Network reliability analysis based on percolation theory. *Reliability Engineering & System Safety, 142*, 556–562.

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.". *Technical Report*. Technical Report, Academic Patenting in Europe-APE-INV.

Liu, L., Han, C., & Xu, W. (2015). Evolutionary analysis of the collaboration networks within national quality award projects of china. *International Journal of Project Management, 33*(3), 599–609.
Liu, P., & Xia, H. (2015). Structure and evolution of co-authorship network in an interdisciplinary research field. *Scientometrics, 103*(1), 101–134.
Lööf, H., & Heshmati, A. (2002). Knowledge capital and performance heterogeneity:: A firm-level innovation study. *International Journal of Production Economics, 76*(1), 61–85.
Ma, Z., & Lee, Y. (2008). Patent application and technological collaboration in inventive activities: 1980–2005. *Technovation, 28*(6), 379–390.
Mansfield, E. (1961). Technical change and the rate of imitation. *Econometrica: Journal of the Econometric Society*, 741–766.
Mansfield, E. (1962). Entry, gibrat's law, innovation, and the growth of firms. *The American economic review, 52*(5), 1023–1051.
Mansfield, E. (1965). Rates of return from industrial research and development. *The American economic review, 55*(1/2), 310–322.
Maraut, S., Dernis, H., Webb, C., Spiezia, V., & Guellec, D. (2008). The OECD REGPAT database: A presentation. *OECD Science, Technology and Industry Working Papers*.
Medda, G., & Piga, C. A. (2014). Technological spillovers and productivity in italian manufacturing firms. *Journal of productivity analysis, 41*(3), 419–434.
Minguela-Rata, B., Fernández-Menéndez, J., & Fossas-Olalla, M. (2014). Cooperation with suppliers, firm size and product innovation. *Industrial Management & Data Systems*.
Mohnen, P., & Hall, B. H. (2013). Innovation and productivity: An update. *Eurasian Business Review, 3*(1), 47–65.
Moreno, R., & Suriñach, J. (2014). Innovation Adoption and Productivity Growth: Evidence for Europe. *IREA Working Papers 201413*. University of Barcelona, Research Institute of Applied Economics.
Newman, M. (2018). *Networks: An introduction*. Oxford university press.
Newman, M. E. J. (2001a). Scientific collaboration networks. i. network construction and fundamental results. *Physical review E, 64*(1), 016131.
Newman, M. E. J. (2001b). The structure of scientific collaboration networks. *Proceedings of the national academy of sciences, 98*(2), 404–409.
Newman, M. E. J., Watts, D. J., & Strogatz, S. H. (2002). Random graph models of social networks. *Proceedings of the National Academy of Sciences, 99*(suppl 1), 2566–2572.
Parisi, M. L., Schiantarelli, F., & Sembenelli, A. (2006). Productivity, innovation and r&d: Micro evidence for italy. *European economic review, 50*(8), 2037–2061.
Park, S., Lee, S.-J., & Jun, S. (2015). A network analysis model for selecting sustainable technology. *Sustainability, 7*(10), 13126–13141.
Pavitt, K. (1985). Patent statistics as indicators of innovative activities: Possibilities and problems. *Scientometrics, 7*(1–2), 77–99.
Pavitt, K., Robson, M., & Townsend, J. (1987). The size distribution of innovating firms in the UK: 1945–1983. *The Journal of industrial economics*, 297–316.
Perc, M. (2010). Growth and structure of slovenia's scientific collaboration network. *Journal of informetrics, 4*(4), 475–482.
Pilkington, A. (2004). Technology portfolio alignment as an indicator of commercialisation: An investigation of fuel cell patenting. *Technovation, 24*(10), 761–771.
Raymond, W., Mairesse, J., Mohnen, P., & Palm, F. (2015). Dynamic models of r & d, innovation and productivity: Panel data evidence for dutch and french manufacturing. *European economic review, 78*, 285–306.
Revilla, A. J., & Fernández, Z. (2012). The relation between firm size and r&dproductivity in different technological regimes. *Technovation, 32*(11), 609–623.
Rogers, M. (2004). Networks, firm size and innovation. *Small business economics, 22*(2), 141–153.
Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of political economy, 94*(5), 1002–1037.
Santarelli, E., & Piergiovanni, R. (1996). Analyzing literature-based innovation output indicators: The italian experience. *Research policy, 25*(5), 689–711.
Schilling, M. A., & Phelps, C. C. (2007). Interfirm collaboration networks: The impact of large-scale network structure on firm innovation. *Management science, 53*(7), 1113–1126.
Schumpeter, J. A. (1934). *The theory of economic development; an inquiry into profits, capital, credit, interest, and the business cycle*. Cambridge, Mass: Harvard University Press.
Schumpeter, J. A. (1942). *Socialism, capitalism and democracy*. Harper and Brothers.
Singh, J. (2005). Collaborative networks as determinants of knowledge diffusion patterns. *Management science, 51*(5), 756–770.
Solow, R. M. (1957). Technical change and the aggregate production function. *The review of economics and statistics*, 312–320.
Stauffer, D., & Aharony, A. (2018). *Introduction to percolation theory*. CRC press.
Stock, G. N., Greis, N. P., & Fischer, W. A. (2002). Firm size and dynamic technological innovation. *Technovation, 22*(9), 537–549.
Suriñach, J., Manca, F., Moreno, R., et al. (2011). Extension of the Study on the Diffusion of Innovation in the Internal Market. *European Economy - Economic Papers 2008 - 2015 438*. Directorate General Economic and Financial Affairs (DG ECFIN), European Commission.
Teece, D. J. (1992). Competition, cooperation, and innovation: organizational arrangements for regimes of rapid technological progress. *Journal of economic behavior & organization, 18*(1), 1–25.
Tether, B. S. (1998). Small and large firms: Sources of unequal innovations? *Research policy, 27*(7), 725–745.
Tether, B. S. (2002). Who co-operates for innovation, and why: An empirical analysis. *Research policy, 31*(6), 947–967.
Tomassini, M., & Luthi, L. (2007). Empirical analysis of the evolution of a scientific collaboration network. *Physica A: Statistical Mechanics and its Applications, 385*(2), 750–764.
Vasilyeva, E., Kozlov, A., Alfaro-Bittner, K., Musatov, D., Raigorodskii, A. M., Perc, M., & Boccaletti, S. (2021). Multilayer representation of collaboration networks with higher-order interactions. *Scientific reports, 11*(1), 1–11.
Zhang, G., Duan, H., & Zhou, J. (2017). Network stability, connectivity and innovation output. *Technological forecasting and social change, 114*, 339–349.