

# Measuring knowledge persistence: a genetic approach to patent citation networks

Arianna Martinelli · Önder Nomaler

Published online: 9 April 2014  
© Springer-Verlag Berlin Heidelberg 2014

**Abstract** The aim of this paper is to propose a new empirical method for identifying technologically important patents within a patent citation network and to apply it to the telecommunication switching industry. The method proposed is labelled the genetic approach, as it is inspired by population genetics: as geneticists are interested in studying patterns of migration and therefore the common origins of people, in innovation studies we are interested in tracing the origin and the evolution of today knowledge. In the context of patent and citation networks, this is done by calculating the patent's persistence index, i.e., decomposing patent's knowledge applying the Mendelian law of gene inheritance. This draws on the idea that the more a patent is related (through citations) to “descendent” patents, the more it affects future technological development and therefore its contribution persists in the technology. Results show that the method proposed is successful in reducing the number of both nodes and links considered. Furthermore, our method is indeed successful in identifying technological discontinuities where previous knowledge is not relevant for current technological development.

**Keywords** Patent data · Patent citation network · Technology dynamics · Telecommunication manufacturing industry

**JEL Classifications** 030 031

---

A. Martinelli (✉)  
LEM (Laboratory of Economics and Management), Scuola Superiore Sant'Anna Piazza Martiri della Libertá 33, 56127 Pisa, Italy  
e-mail: a.martinelli@sssup.it

Ö. Nomaler  
School of Innovation Sciences, Eindhoven University of Technology, 5600 MB Eindhoven,  
The Netherlands

## 1 Introduction

The use of patents as an indicator of innovative activities is rather established in the innovation literature (Yoon and Kim 2011; Griliches 1991). However, one of the usual caveats pointed out by scholars is that patents are not informative about the importance (and the economic value) of the invention they disclose. The usual way to overcome this issue is to look at forward citations.<sup>1</sup> In fact, as the number of citations signals the importance of scientific articles, we have evidence that important patents receive as well more citations.

Recently, exploiting the similarity between scientific and patent citations a new stream of literature has adopted a network approach to the study of patent and citation data, and analytical tools previously used for publication networks have been used in identifying important patents (von Wartburg et al. 2005; Fontana et al. 2009). The aim of this paper is to contribute to this stream of research and to propose a new method for identifying technologically important patents in a patent citation network.

In order to make sense of the new indicator and method, we need to define our notion of importance. First of all, we need to stress that this work focuses on the technological importance of patents, as a patent citation network pinpoints only the technological relation between inventive steps (i.e. patents). In the next section, we will discuss at length the nature and the interpretation of such links. For the moment, it is enough to consider that being cited as *prior art*<sup>2</sup> implies a technological relation between two inventions.

In this paper, the importance of a patent relates to the persistence of its technological contribution. This draws on the idea that the more a patent is related (through citations) to “descendent” patents, the more it affects future technological development and therefore its contribution persists in the technology. The choice of a genetic jargon is not accidental. In fact, from a purely conceptual perspective, the exercise carried out here is similar to the one made by population geneticists. As they trace our (geographical) origins by looking at genetic mutation in today’s population (Cavalli-Sforza et al. 1994), in this paper we propose a method to “decompose” the technological content of today’s patents into the technological contribution of the previous ones. We can therefore investigate the origin and development of today’s knowledge, identifying technological lineages representing cumulative chains of technological advances leading to today’s knowledge. For this reason, this new method is labelled the “genetic approach” (GA) to patent citation networks.

As we will see, differently from common approaches to patent and citation data, the GA (and, in particular, the knowledge persistence indicator) assesses patents’ citation structure from a broader perspective, considering not only direct citations but also all the direct and indirect “descendants”. Therefore, importance is not a “local” characteristic of the node (i.e. the number of citations a patent has) but a global characteristic.

---

<sup>1</sup> Scholars have addressed the problem of evaluating patent values using several indicators such as forward citations and renewal fees (Bessen 2009; Harhoff et al. 2005; Gambardella et al. 2008).

<sup>2</sup> In most patent systems the *prior art* constitutes all information that has been made available to the public that might be relevant to a patent’s claims of originality.

In this article, we propose an application of the methodology in the context of the telecommunication switching industry. This represents an interesting case study because, in the last decades, this industry has gone through numerous incremental and radical technological changes. Results show that the method proposed is successful in reducing the number of both nodes and links considered. Furthermore, our method is indeed successful in identifying technological discontinuities where previous technological knowledge is not relevant for current technological development.

As regards the broader impact of this new methodology, we can expect two lines of generalization. The first (obvious) one is the possibility to apply the persistence index to other technological fields in order to characterize their cumulative and disruptive dynamics. The scientific importance of such an exercise is to foster quantitative research and appreciative theorizing defined as rigorous storytelling (Nelson 1989) in studies related to technology dynamics and its effect on industry dynamics.<sup>3</sup> The second one regards the possibility of applying the persistence outside the context of patent citation networks. We postpone some speculations on this aspect until the last section.

The paper is structured as follows: Section 2 introduces patent and citation networks and provides the theoretical background for our notion of importance; Section 3 presents the details of the new method; Section 4 provides the comparison to existing alternative method; Section 5 introduces the industry here examined and the dataset; and Section 6 presents the empirical results. Conclusions follow.

## 2 Theoretical background

### 2.1 A network approach to patent and citation data

Patents and citations have been already used for building several types of knowledge networks the nodes of which are firms, inventors (Balconi et al. 2004), or technological classes (Bottazzi and Pirino 2010). The network analyzed in this work is composed of patents (the vertexes) and citations (the arcs). In particular, the forward citation between two patents establishes a directed link going from the cited to the citing patents. If we adopt a stylized (but realistic) view of patents, considering them as a collection of “technical problems and newly proposed solutions”, the advantage of the network approach is the possibility to appreciate the interrelated nature of technological developments.

As regards the nature of such relations, this depends on the interpretations of citations (e.g. knowledge spillovers, knowledge flows, etc.) (Marco 2007). If, on the one hand, we prefer to remain rather agnostic about the exact nature of this relation, on the other hand, such a technical relation is codified in the definition of *prior art*. The citation to an older patent means that the invention disclosed in the old patent can be relevant to the new patent’s claims of originality. This means that the definition of the

---

<sup>3</sup>As regards this point, a MATLAB code for the calculation of the persistence index is available on request.

*prior art* allows us to limit and to circumscribe the novelty of an invention. Following (Strandburg et al. 2009) "... A citation from one patent to another may indicate either that the later patent builds upon the technology of the earlier patent or simply that the earlier technology was closely enough related to be material to determining whether the later patent should be issued... (page 108)". In this work, we adopt a "minimal" connotation of such links: if patent A cites patent B, we can think of a technological relation between the inventions disclosed by the two patents. In a conservative way we can claim that a patent citation network built as just described allows us to map the technological relations between all the inventions carried out within a specify technology.<sup>4</sup> The analysis of its structure and topology (i.e. the layout of interconnections of the various elements) allows unfolding patterns of technical changes.

The challenge posed by patent citation networks is not only their size,<sup>5</sup> but also other characteristics such as directionality and acyclicity. The former refers to the fact that ties are not symmetric. The latter means that no cycles are present in the network, as patents can only cite previous patents and therefore the time dimension is embedded into the ties' direction. These characteristics make difficult to apply standard network analysis techniques, as these mainly deal with undirected networks and most of the usual indicators cannot be applied in a straightforward way to a directed network.<sup>6</sup>

For these reasons, the interest on patent citation networks is not confined to innovation studies. Researchers in complex systems began to be interested in the patent citation network as a whole, as it constitutes an example of a complex system emerging from human activities. They apply mechanical statistics in order to study network dynamics (i.e. the study of the organizing principle) and to model a system so as to replicate specific properties such as the power law distribution of the number of forward citations (Valverde et al. 2007).<sup>7</sup> They explain the emergence of such a property by the combination of two (necessary) forces: network expansion and "preferential attachment" (Barabási and Albert 1999; Albert and Barabási 2002).<sup>8</sup> In particular, they adapt the latter principle for the context of patent and citation network, making the probability of a patent to be cited dependent on the number of current citations and the age of the patent. These studies are rather recent and if, on the one hand, they shed some light on the structure and dynamics of patent networks, on the other hand, their approach is poorly grounded in innovation studies. For instance, they are not interested in estimating (and explaining) sectoral differences in the parameters fitting their models. For these scholars, the patent citation network is just an example of a

<sup>4</sup>Note that this highlights the fact that the method here proposed is meaningful only for technology appropriated through patents.

<sup>5</sup>Notice that the whole NBER patent dataset has 3 774 768 nodes and 16 522 438 links.

<sup>6</sup>For instance, centrality betweenness in a directed network is based on the assumption that directional ties can be transformed into unidirectional (Wasserman and Faust 1994).

<sup>7</sup>The number of forward citations follows a scale-free distribution, meaning that a few patents have most of the links, whereas the large majority has just a few.

<sup>8</sup>According to this organizational principle, each new node probability of having  $k$  links decays following a power law distribution, namely  $P(k) \sim k^{-\gamma}$ . Therefore, new nodes have a "preference" in connecting to already well-connected nodes.

complex system emerging from individual (innovative) behaviors and their aim is to find potential commonalities with natural complex systems.

In the innovation studies domain, there has been a recent<sup>9</sup> interest in patent citation networks, and in a short time several empirical articles have been published. All of them use the bibliometric indicators developed by Hummon and Doreian (1989) for identifying the main flow of knowledge within a patent citation network (Mina et al. 2007; Verspagen 2007; Fontana et al. 2009; Barberá et al. 2010; Martinelli 2012).

The GA here proposed heavily differentiates from the studies just mentioned. Differently from the complex system approach, we are interested in grounding the indicators in the innovation literature and the persistence indicator is going to shed some light about technology evolution. Differently from the Hummon and Doreian approach (HDA), the GA is more flexible in its construction as it does not stress the importance of direct links. As the HDA is now increasingly used in the literature, we are going to discuss in detail their differences in Section 4.

## 2.2 Importance as knowledge persistence

The aim of this section is to provide a theoretical background for the notion of importance proposed in this paper. As already anticipated in the introduction, important patents are the ones the technological contribution of which is found in today's patents. How is knowledge transmitted through the network and then found in recent patents? Endorsing a neutral outlook on the meaning of the links, we can think of citations as pipes through which pieces of knowledge are *inherited* from cited to citing patents. Therefore, as a population geneticist compares the genotype of contemporary populations, we could study the "genetic" structure of today's knowledge<sup>10</sup> and rigorously trace their origin.

The genetic decomposition is explained in the next section. However, we can anticipate that it is operationalized using the Mendelian notion of genetic inheritance. We are, therefore, able to identify "knowledge lineages" depending on the citation structure between a patent and all its direct and indirect "descendants". In fact, the contribution of each patent depends on the topological structure of the network composed of all the paths generated in it.<sup>11</sup>

In this way, a patent citation network represents a system of knowledge generation where the inventive step is provided by the recombination of existing (as inherited)

---

<sup>9</sup>This field of research is relatively new, excluding some explorative papers such as Ellis et al. (1978). In this study, the analysis of patent citation networks is presented as a study about technological relatedness, which can be interpreted as an application to technology of the approach pioneered by Garfield (1979) for the study of links among scientific disciplines. However, those scholars already admit the difficulty of an extensive use of this method because of the lack of a systematic database of patent citations (unlike the possibility of using the ISI database for publications). Probably it is for this reason that, until recently, there has been limited scientific work on patent citation networks.

<sup>10</sup>Section 3 clarifies how "today's knowledge" is defined.

<sup>11</sup>It is important to stress that, if on the one hand, the conceptual similarity with genetics is evident, on the other, the employment of the same methodologies and statistical tools for the analysis is problematic. In fact, geneticists often use frequencies of the alleles for the clustering exercise, which cannot straightforwardly be translated in the context of patents.

pieces of knowledge. Following the Mendelian decomposition mechanism, it appears that a building block of the evolutionary process, which is random mutation, is missing. In this respect, the genetic patent decomposition might look like a deterministic representation, where new inventions are simply a sum of proportion of previous knowledge and where nothing genuinely new is created. However, knowledge is recombined and therefore, by definition, is transformed into something different and therefore new. Undoubtedly, this concept of innovation used in this work is rather similar to the concept of “recombining knowledge” put forward by Weitzman (1998, 1996), according to which . . . *new ideas arise out of existing ideas in some kind of cumulative interactive process* . . . (Weitzman 1996, page 209).

We expect that, in a dense network, few patents are successful in spreading their knowledge, displaying a high level of persistence and generating a large lineage of descendent patents. These lineages represent chains of technical change with a certain extent of cumulativeness. Therefore, the use of both the genetic decomposition and the persistence index allows us to answer questions related to patents technological importance and technology dynamics.

### 3 A genetic approach to patent citation networks: genetic decomposition and knowledge persistence<sup>12</sup>

Figure 1 represents a very simple patent citation network structure with five *startpoints*, two *intermediates* and two *endpoints*.<sup>13</sup> *Endpoints* are generally recent patents and they represent the set of what in the previous section was labeled “today’s knowledge”. The genetic decomposition corresponds to decompose the knowledge content of the *endpoints* in function of the *startpoints*. After the decomposition is performed, it is possible to quantify the degree in which the *startpoints*’ prospective knowledge is retained in the *endpoints* and to look at how much knowledge the two *endpoints* actually share.

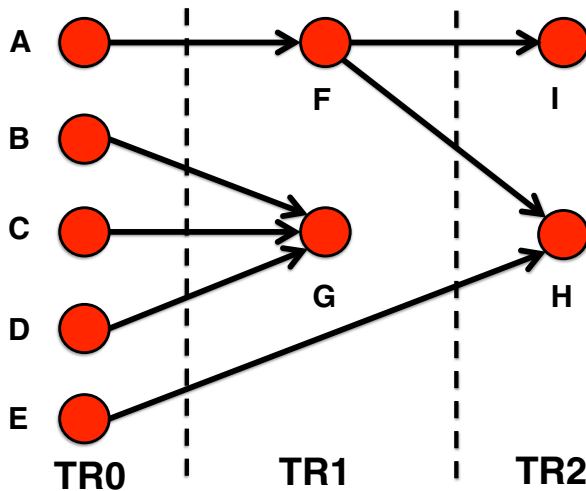
Clearly, both these aspects depend on the structure of the forward citations and on the number of *intermediates* and their citations. This means that, for calculating the persistence index, we should not focus only on *startpoints* but we also need to account for the *intermediates*’ contribution that represents the “new” knowledge injected in the system. In fact, in the framework of the genetic decomposition in a patent citation network, both knowledge persistence and knowledge creation coexist. Summarizing, Fig. 1 represents a process of knowledge creation, transmission, and transformation.

The simple network displayed in Fig. 1 is composed of three layers of patents, indicated by TR0, TR1, and TR2. The genetic decomposition of the network in Fig. 1 is performed using the following heuristics:

1. *Endpoints* are identified, and working backwards, each patent is assigned to a layer;

<sup>12</sup>As already mentioned in the introduction, a MATLAB code for performing the calculations of this section is available on request.

<sup>13</sup>Note that directionality follows the direction of the knowledge flow between cited and citing patents.



**Fig. 1** Simple patent citation network structure

2. For each *startpoint* belonging to the first layer (TR0 in Fig. 1), the persistence index is calculated. This is going to quantify how much of their knowledge is retained in the *endpoints* (TR2 in Fig. 1);
3. The *startpoints* are deleted (the network is truncated) and therefore a new layer of *startpoints* is created (TR1 in Fig. 1);
4. Calculation of the persistence index for the new group of *startpoints*;
5. Deletion of the layer and repetition of step 2 and 3 up to the last layer.

This procedure is repeated for each layer and the number of layers depends on the length of the largest geodesic distance in the network.<sup>14</sup>

Steps 2 and (recursively) 4 represent the core of the new method, which corresponds to the application of the Mendelian law of genetic inheritance to citations.

Looking at the first layer (TR0) in Fig. 1, we can see that the only patent cited by patent **F** is patent **A**; thus 100 % of the inherited knowledge embodied in patent **F** is the knowledge of patent **A**. Instead, patent **G** makes three citations to patents **B**, **C**, and **D**; thus the inherited knowledge embodied in patent **G** consists 33.3 % of patent **B**, 33.3 % of patent **C** and 33.3 % of patent **D**.

In the second layer (TR1), the endpoint **I** makes only one citation and that is directed to patent **F**. Since the embodied knowledge in patent **F** is 100 % that of patent **A**, the inherited knowledge embodied in patent **I** is again 100 % that of patent **A**. The endpoint **H** makes three citations. The first is to patent **F** that has 100 % patent **A** knowledge; thus  $\frac{1}{3}100\% = 33.3\%$  of the inherited knowledge embodied in patent **H** is the knowledge of patent **A**. The second citation of patent **H** is to patent **G** that embodies 33.3 % of each of the respective knowledge of patents **B**, **C**, and **D**.

<sup>14</sup>The geodesic is the longest shortest distance between two nodes.

**Table 1** Genetic decomposition and persistence of knowledge for Truncation 0

		Startpoints				
		A	B	C	D	E
Endpoints	I	1	0	0	0	0
	H	0.33	0.11	0.11	0.11	0.33
Sum		1.33	0.11	0.11	0.11	0.33

Since patent **H** inherits only  $\frac{1}{3}$  of its knowledge from patent **G**, it inherits indirectly  $\frac{1}{3}33.3\% = 11.1\%$  of each of the knowledge of the *startpoints* **B**, **C**, and **D**. Finally 33.3% of the inherited knowledge in patent **H** comes directly from *startpoint* **E**. These results are given in Table 1 in matrix form.

Focusing on the TR0 level, the genetic decomposition answers the question: *How much knowledge of (startpoints) A, D, C, B, and E is retained in (endpoints) H and I?* The answer is displayed in Table 1, where each other row decomposes the inherited knowledge embodied in another endpoint down into the shares of the *startpoints* which have supplied the knowledge. For this reason, each row adds up to 1.0.

The persistence index can be found in the last row, the column sum. It supplies a fractional count that is the basis of the persistence index: effectively, 1.33 out of the 2 *endpoints* are the pure descendants of the *startpoint* **A**, 0.33 out of 2 are the pure descendants of the *startpoint* **E**, and each of the *startpoints* **B**, **C**, and **D** has 0.11 pure descendants. Clearly, *startpoint* **A** is the most important *startpoint* since it is  $\frac{1.33}{0.33} = 4$  times as important as patent **E** and  $\frac{1.33}{0.11} = 12$  times as important as patents **B**, **C**, and **D**. Patent **E** is the second most important *startpoint* and *startpoints* **B**, **C**, and **D** look not so important on their own.

As explained in the step list before, the decomposition algorithm puts only the *startpoints* into competition. In other words, the intermediate patents **F** and **G** do not show up in Table 1 as knowledge suppliers as they just transmit the knowledge from the *startpoints* to the *endpoints*. Therefore, after step 2, *startpoints* in layer TR0 are removed from the network (i.e., truncate the network from the left), and a new set of *startpoints* is created. In the example of Fig. 1, one left truncation (i.e., removal of patents **A**, **B**, **C**, **D**, and **E**) brings patents **F** and **G** forward as *startpoints*. Table 2

**Table 2** Genetic decomposition and persistence of knowledge for Truncation 1

		Startpoints	
		F	G
Endpoints	I	1	0
	H	0.5	0.5
Sum		1.5	0.5



shows the results for the genetic decomposition and the persistence index for the TR1 level.

*How much knowledge of **F** and **G** is retained in **H** and **I**?* Patent **F** appears to have 1.5 pure descendants and patent **G** has 0.5 pure descendants; thus patent **F** is  $\frac{1.5}{0.5} = 3$  times as important as patent **G**.

In the simple network displayed in Fig. 1, only three layers are present since another step of truncation leaves us only with the *endpoints*. In real samples, the network is left truncated and analyzed as long as it is possible to truncate further. It follows that, for each layer, a matrix such as Table 1 and 2 is calculated and the persistence index is calculated as the sum of each contribution (the last row in Tables 1 and 2). Furthermore, the persistence index is then normalized using the maximum, meaning that for each truncation the persistence index takes a value between 0 and 1.<sup>15</sup>

#### 4 The genetic approach vs. other approaches

Before moving to the empirical details, it is worth to spend few words to compare this new method to existing comparable methods. In particular, we are going to discuss differences with citation count, the originality indicator proposed by Trajtenberg et al. (1997), and the connectivity approach (Hummon and Doreian 1989). Furthermore, in the empirical section (Section 6.1), we will compare the persistence index to all these indicators in order to show their relations.

The network approach to patent and citation data represents a shift of perspective respect to citation counts. The two have already shown some complementarities: having a large number of citations is not a sufficient condition for becoming an important connection in the main flow of knowledge within the network (Fontana et al. 2009). Citation counts can be easily performed from a network perspective. In fact, it is always possible to count the direct number of ties a patent has. However, any network approach allows us to enlarge this local perspective and to evaluate the whole citation structure.

Trajtenberg et al. (1997) use a similar jargon for introducing some patent indicators of basicness and generality of invention. In particular, their measurement of patent importance does not consider only the forward citations but also the importance (i.e. the forward citation) of the citing patents. They compute patent importance as a sum of forward citation and a fraction of the forward citation of the citing patents. As they evaluate importance looking at two rounds of forward citations, they broaden the local perspective of citation counts. However, the choice of the weight is rather arbitrary (as well as the choice to stop at the second round) and it does not account for possible cross citations between all these subsequent patents. In this respect, the network approach here implemented allows us to account for all the patent citation structures.

---

<sup>15</sup>It is worth noting that by construction, the persistence index can be calculated only for *startpoints* and *intermediates*.

Finally, as already anticipated in Section 2.1, recently scholars (among other see: Mina et al. (2007); Verspagen (2007); Fontana et al. (2009); and Barberá et al. (2010)) have used the bibliometric method proposed by Hummon and Doreian for identifying the main flow of knowledge in a patent citation network.

Ultimately, both the genetic approach (GA) and the Hummon and Doreian approach (HDA) have the same aim, which is to trace “important” technological advances. However, they deeply differ in the underlying rationale and definitions. An example from population genetics can be helpful. Summarizing, we can say that the work of geneticists is: given the observed population, genetic differences with ancestors<sup>16</sup> highlight streams of migration. The application of the GA to *endpoints* follows exactly the same rationale and therefore it can be considered a backward mapping of successful (persistent) technologies. The application of the HDA to populations would work in a completely different way, that is: starting from the ancestors at each (population) bifurcation, follow the future development of the largest stream. This would correspond to tracing just the largest population and ignoring the remainders and their future development. Of course, the application of the HDA by geneticists would be nonsense. However, this example clarifies the basic differences between the two approaches that is, the direction of mapping: backwards for the GA and forward for the HDA. As a consequence, the HDA might be very sensible to “local” peaks and therefore discarding chains of innovations which from an ex-post perspective (the importance of the *endpoints*) is relevant. Furthermore, the HDA uses a search algorithm that selects highly valuable subsequent links, which introduces a certain bias towards a particular definition of cumulateness and over-emphasizes the notion of incremental progress.<sup>17</sup> By contrast, the implementation of the GA presented in this work is rather flexible, as the analysis of the *persistence weighted network* does not impose any specific structure.

## 5 Case study and dataset

### 5.1 The telecommunication switching industry

Before moving to the empirical section, we are going to discuss briefly the industry under examination. We do not want to summarize the milestones of technical change in the telecommunication switching industry<sup>18</sup> but rather desire to convince the reader on the validity of this industry as a testbed. This consideration is based on two facts: first, the industry is an innovative one, and second, patents are representative of such advances. The study of technical change in the industry clearly show that the period between the 1975 and 2001 is characterized by both “normal” and “disruptive” technical change. The former refers to the development and the

---

<sup>16</sup>Please note that this is a rough simplification of the principles and statistical methods used by geneticists. However, it would not be of any help to discuss these in further detail, given the impossibility of using them in the context of patent genetic decomposition.

<sup>17</sup>For further details, see Martinelli (2012).

<sup>18</sup>For a detailed account see Martinelli (2010) and Martinelli (2012).

consolidation of digital switches, whereas the latter refers to the transition from circuit to packet switching.<sup>19</sup> The presence of such waves of technical change makes this industry a suitable testbed for a new method for studying technology dynamics. Moving to the second point, interviews with engineers active in the period under examination support the idea that patents were used for protecting inventions and enhancing intense cross-licensing (Chapuis and AEJ 1990; Martinelli 2010). Therefore, patents provide complete and comprehensive data about technological change in the telecommunication switching industry.

## 5.2 Data

The patent sample was retrieved from the USPTO website using technological subclasses that all belong to technological class 370 (“Multiplex Communication”).<sup>20</sup>

The selection of the technological class and technological subclasses was driven by the reading of their descriptions and by the analysis of firms’ patent portfolio for companies highly specialized in switch production. In order to account for the complexity of a switch and to consider important technologies that might not have been captured by the first search, the first round of cited patents were added to the original set. These cited patents were retrieved from the NBER patent database (Hall et al. 2001); for patents granted before 1975, the citations were taken from the patent documents<sup>21</sup> and manually added. The final sample includes 6214 patents covering the period 1924–2003.

Citations for patents issued before 1975 were manually collected. It is important to note that the citations considered are only the “internal” ones, meaning that once the patent sample is selected, only citations to patents included in the starting sample are considered.

---

<sup>19</sup>These two switching methods differ in the allocation of the bandwidth, which brings about inefficiencies in the use of the infrastructure depending on the type of information switched. A circuit switching network establishes a circuit (or channel) between senders and receivers (which might be telephone, computers, etc.) before the users may communicate, as if the nodes were physically connected with an electrical circuit. In this case, the bandwidth is statically allocated also when no information is transmitted (for instance, when there is a silence during a phone call). By contrast, a characteristic feature of packet switching is that no connection is established and information is chopped into packets that are individually routed in the network; this implies that each packet can take a different route between the same sender and receiver. These two switching modes are differently efficient depending on the type of information transmitted. The former allows us to guarantee a minimum level of quality of service for voice communication (often required by control agency on the telecommunication operator). However, it is extremely inefficient in dealing with bursts typical of data communication.

<sup>20</sup>These subclasses are:

1. Having space switch as intermediate stage (e.g., T-S-T, T-S-S, or S-S-T) (370/370);
2. Having details of control storage arrangement (370/371 and 370/378);
3. Using time slots (370/458);
4. Synchronization information is distributed over multiple frames (510/370).

<sup>21</sup>The pdf version of original patent documents are available at: <http://patft.uspto.gov/netahtml/PTO/search-adv.htm>.

## 6 Empirical analysis<sup>22</sup>

This section is going to present the results obtained using the GA introduced in the previous sections. In particular, two types of analyses are performed: (i) the persistence index is used to identify important patents and to assess firm's patent portfolios and (ii) the persistence index is used to weight the citation network in order to shed some light about technology dynamics in the telecommunication switching industry.

### 6.1 Persistent patents

As mentioned in Section 3, the persistence index is calculated only for the *startpoints* generated after each truncation; therefore the persistence index can be calculated only for a subsample of the patents. Table 3 shows the number of *startpoints* evaluated in each of the 25 truncation levels present in our network.<sup>23</sup>

The persistence index calculated at each truncation presents a very skewed distribution displayed in Appendix A. Even if graphs are small, they clearly show the high left skewness, meaning that only a handful of patents are successful in spreading their knowledge, whereas the contribution of the high majority is diluted over time. This is consistent with the evidence on strongly skewed distribution of other indexes for patent importance and value (e.g. forward citations, license fees, etc. etc.) indicating that, generally, only few patents are important and valuable (Marsili and Salter 2005; Silverberg and Verspagen 2007).

As high persistence indicates high technological importance, we can rank patents and identify "important" ones. Given a rank, we can expect a certain extent of arbitrariness in deciding which patents are important, those corresponds to set a minimum threshold. However, in the case of persistence index, arbitrariness is very low because of its extremely skewed distribution. Using a rather conservative cut off point, which is 0.5, we extract 79 "important patents".<sup>24</sup>

The rationale about using a method for reducing network complexity is the possibility to infer some properties on the whole network just using that subsample. In order to test whether our subsample (representing 1.27 % of the full sample) is a representative one, we compare some patent indicators and test the persistent subsample against the full population. Table 4 reports some summary statistics for patents characteristics and citations indicators for the set of persistent patents and the remaining sample.<sup>25</sup> The table includes:

1. the patent citation count proposed by Jaffe and Trajtenberg (2005), which consists in the number of forward citations plus 1;
2. the average issue year of the patent;

<sup>22</sup>For the results obtained using the HDA on the same data, see Martinelli (2010).

<sup>23</sup>The longest geodesic in the network is 25; therefore, the network here analyzed has 25 levels of truncation.

<sup>24</sup>See Appendix B for the complete list and some relevant information.

<sup>25</sup>Note that, because of data availability, some observations are missing for the number of claims and generality measure.

**Table 3** Number of *startpoints* per truncation

Truncation	Number of <i>startpoints</i>
0	1787
1	485
2	239
3	160
4	131
5	127
6	120
7	139
8	148
9	135
10	144
11	178
12	219
13	219
14	225
15	244
16	231
17	242
18	217
19	157
20	102
21	52
22	33
23	14
24	6
TOTAL	5754

3. the SPLC (Search Path Link Count) indicator introduced by Hummon and Doreian (1989) in their main path analysis. Without entering into the details, these indicators evaluate the connectivity of a citation by measuring how many downstream and upstream patents are connected throughout such citation;<sup>26</sup>
4. the number of claims reported in the patent;
5. the generality index proposed by Trajtenberg et al. (1997) calculated on the IPC patent classes.

<sup>26</sup>For details see Hummon and Doreian (1989). It is worth noting that the SPLC indicator evaluates citations and not patents, and therefore they can be assigned to either the cited or the citing patent of each citation. The interpretation does not vary a lot and in this case the citation weight is assigned to the citing patent.

**Table 4** Summary statistics for citation indicators

Variable	N	Mean	Median	SD	Min	Max
Citation	5675	4.502	3	3.623	2	39
Count	79	13.303	11	8.386	2	58
Issue year	5675	1987.293	1989	10.121	1946	2003
	79	1984.57	1987	10.972	1956	2001
SPLC	5674	0.01	0.001	0.046	0	.955
	79	0.288	0.098	0.42	0	2
Claims	4219	15.489	12	13.518	1	236
	54	20.185	18.5	15.102	1	86
Generality	5458	0.579	0.682	0.312	0	1
	75	0.548	0.677	0.301	0	1

Table 4 shows that the patents extracted using genetic decomposition are older, receive more citations, have higher connectivity, and display more claims. Interestingly, figures related to the generality indicator are rather similar, suggesting some potential similarity between the two indicators (however see Fig. 2). These results hold both for the mean and the median.

Table 5 shows the results of the comparison between the two samples. The non parametric Wilcoxon-Mann-Whitney used for accounting for the high skewness of the variables rejects the hypothesis of samples extracted from the same distribution for the citation count and the SPLC.

As anticipated in Section 4, we now focus on the comparison between the persistence index and other patent indicators there exposed.

Figure 2 shows the scatterplots of the persistence index against the number of forward citations, the generality index (calculated on the IPC classes of backward citations), and the SPLC indicator. What clearly appears from the graphs is the loose relations between the new proposed indicator and the already existing ones. Indeed, this points to the fact that the former unfolds a different aspect of individual inventions covered by patents. In particular, we can see that high persistence is not systematically correlated to high frequency of forward citations, high level of generality, and high level of cumulateness. In fact, going back to the discussion of the previous sections, persistence as explained and operationalized relates to both “long term” and widespread influence of a patent on subsequent innovations.

As inventions covered by patents are developed within a firm, persistence can be used to evaluate firm’s patent portfolio. In this respect, we use the persistence index for dividing the patent set into four equal groups and we look at the 10 top assignees for each quartile. Table 6 reports (for each quartile) the name of the assignee, the number and the percentage of its patents in that quartile, and the average issue year of these patents. The latter is displayed in order to give an idea of the “vintage” of the portion of the patent portfolio in each quartile.

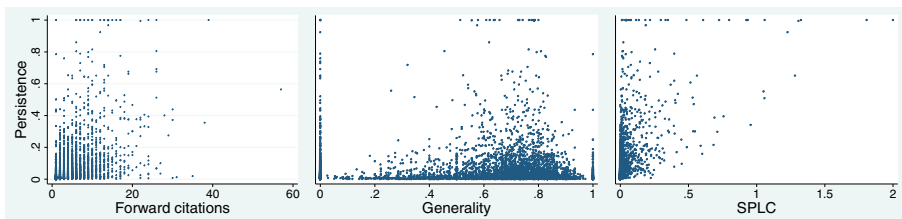
**Table 5** Results of the Wilcoxon-Mann-Whitney test on the median

Variables	z	P-value
Weighted patent count	-9.015	0.000
SPLC	-8.845	0.000
Claims	-0.965	0.3346
Generality	1.949	0.0513

Table 6 shows not great differences in the assignees' names over quartiles. However, when we look at the concentration and the dispersions of the shares of patents, we can see that the upper and the top quartiles are more concentrated as the top four companies account for a larger share of (more persistent) patents. Furthermore, the increase in the HHI index over quartiles support the idea that persistent patents are developed by a lower number of companies. In a nutshell, companies developed both persistent and not-so persistent inventions, However, only a few companies are able to produce inventions the impact of which are found in subsequent inventions. In the final part of next session we will elaborate more on the relation between these companies and technological evolution.

## 6.2 The *persistence weighted network*

Following the previous section, we can see that the persistence index is associated with patents and, therefore, from a network perspective, it refers to a characteristic of the node. However, this indicator can be also used to weight the links of the patent citation network in order to build a *persistence weighted network*. In such a network, each link (i.e. the citations) is weighted using the product of the normalized (by the maximum) persistence index of the citing and cited patents to which the link connects. The persistence index is therefore used for transforming the binary patent citation network into a weighted one, where the values of the links inform the persistence level of the knowledge transferred by that citation. It is worth to point out

**Fig. 2** Persistence index and patent indicators

**Table 6** Assignees over patent persistence quartiles

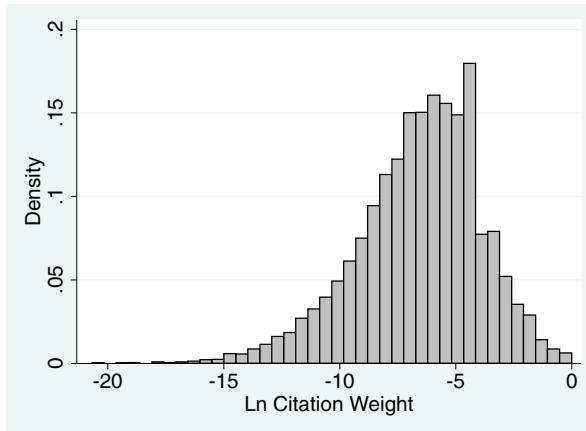
Assignee	First Quartile			Assignee	Second Quartile		
	Num. patents	%	Av. Issue year		Num. patents	%	Av. Issue year
Bell Labs	93.0	6.5%	1977.83	Bell Labs	86.5	6.1%	1978.15
IBM	72.0	5%	1985.18	IBM	70.0	4.9%	1985.10
Motorola	61.0	4.2%	1993.93	Motorola	58.5	4.1%	1993.12
NEC	43.7	3.1%	1989.91	NEC	42.5	3%	1989.28
Scientific Atlanta	34.0	2.4%	1995.09	Ericsson	39.0	2.7%	1992.31
Siemens	32.0	2.2%	1982.22	Siemens	39.0	2.7%	1983.67
AT&T	26.0	1.8%	1992.81	Alcatel	32.0	2.2%	1991.30
Hitachi	25.5	1.8%	1988.81	GTE	31.0	2.2%	1982.87
Alcatel	24.0	1.7%	1990.13	Fujitsu	28.0	1.9%	1993.00
TOTAL	1439				1438		
C4		0.1874				0.1791	
HHI		0.0149				0.0150	
Assignee	Third Quartile			Assignee	Fourth Quartile		
	Num. patents	%	Av. Issue year		Num. patents	%	Av. Issue year
Bell Labs	127.0	6.5%	1977.88	Bell Labs	167.8	11.7%	1978.63
IBM	69.0	6.5%	1983.33	IBM	75.0	5.2%	1984.73
NEC	56.2	6.5%	1990.30	NEC	58.5	4.1%	1989.55
Ericsson	52.0	6.5%	1990.33	Motorola	57.0	4%	1993.04
Alcatel	48.0	6.5%	1993.69	Ericsson	49.0	3.4%	1991.20
Siemens	45.5	6.5%	1983.02	AT&T	46.8	3.2%	1992.10
Fujitsu	39.7	6.5%	1992.81	Alcatel	45.0	3.1%	1991.66
AT&T	38.0	6.5%	1991.73	Nortel	43.0	3%	1989.66
Motorola	37.0	6.5%	1992.78	Fujitsu	43.0	2.9%	1992.15
TOTAL	1439				1438		
C4		0.2114				0.2490	
HHI		0.0205				0.0286	

The number of patents has decimal because it is a fractional count; for instance, if a patent has three assignees they all account for 0.33 each

that, if on the one hand, this use of the persistence index is rather straightforward, on the other hand, it is just one possibility, leaving many opportunities for alternative applications. In fact, we think the persistence index is a rather flexible indicator and it can be used in several ways both for network analysis and visualization. We leave for future research further explorations and methodological extensions.

Given the fact that the persistence index is available for a subsample of patents, it is not possible to weight all the citations. Therefore, the weighted network includes 16,747 citations, corresponding to 80.3 % of the full sample.





**Fig. 3** Distribution of citation weights

Figure 3 represents the distribution of the logarithm of the citations weight. It highlights the fact that very few links transmit persistently knowledge.<sup>27</sup> In fact, in this graph, the last column represents the number of citations with the highest weight, directly connecting the patents with the highest persistence index.<sup>28</sup>

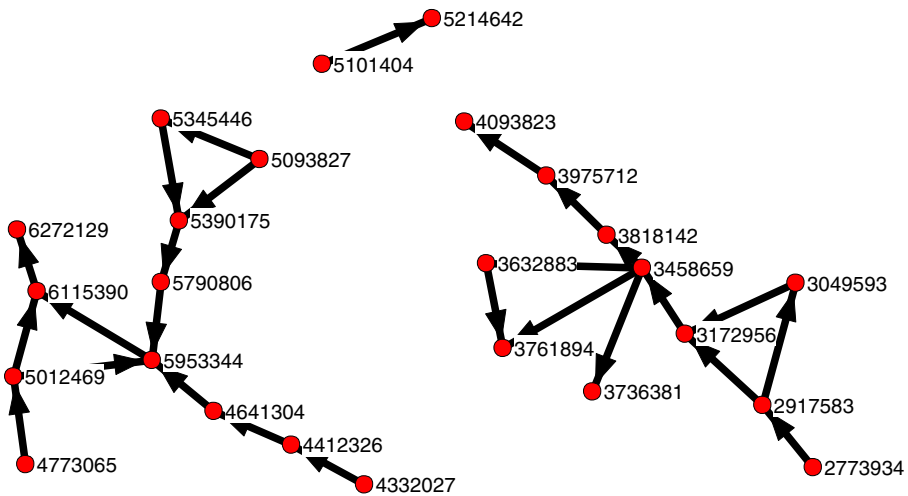
In the following pages, the *persistence weighted network* is analyzed looking at how its structure evolves considering different cutoff points. Using a visual metaphor, we can conceive the *persistent weighted network* as the technological landscape, where patent height depends on the persistence index. Using a very high cut off point corresponds to deleting unimportant links and consider only citations with a (relative) high persistence measure. In this logic, the first step is to set a very high threshold such as 0.9 and to look at citations transmitting the most persistent knowledge within the network.

The resulting network structure (Fig. 4) presents two separated components indicating two disconnected areas of highly persistence knowledge unfolding from a network of more diluted knowledge. Undoubtedly, this fragmentation is dependent on the chosen threshold. However, this two component structure is stable down to 0.75 threshold.

The existence of two separate components suggests the presence to two separate technologies that did not intensively interbreed. In fact, the lack of a bridge between the components indicates that the two components do not share any persistence knowledge. A way to validate this finding is by looking at some characteristics of the two components such as their vintage, their technological contents, and their assignees.

<sup>27</sup>The summary statistics for the weights are: mean 0.0169, standard deviation 0.0655, minimum 1.05e-09, and maximum 1.

<sup>28</sup>Because of the normalization by the maximum the maximum weight is 1, which becomes 0 using the logarithmic transformation.



**Fig. 4** Network with cut off point 0.90

At first, components 1 and 2 differ in their vintage: component 1 comprises patents granted between 1949 and 1977, whereas component 2 are patents granted between 1980 and 1999. Beyond the years of grant, substantial differences emerge. Looking at their technical contents, component 1 deals with the development of a reliable digital telecommunication circuit switching, while component 2 is composed of three branches converging to patent 5953344 and finally to the endpoint 6272129. These patents address technical solutions responding to increasing demand for data communication and the development of packet switching and the use of Internet protocol.

According to the literature, the industry underwent a wave of disruptive technological change from circuit switching to packet switching, occurring in the period under examination. Following the innovation literature, this discontinuity represents a paradigmatic change (Dosi 1982) as it affects not only the design of telecommunication switches but also technological competences (both at firm and inventor level) needed for their development (Martinelli 2012). Therefore, patents in the two separated components in Fig. 4 disclose radically different inventions. The lack of highly persistent connections suggest that the later technology (i.e. the packet switching indicated with 2) is loosely built on previous technological development.

Given these two peaks represent such different technologies it is interesting to look at how they are connected and therefore to what extent knowledge from an early paradigm is retained in a subsequent one. This means to look at how the network structure changes, lowering the cutoff points. For instance, we can imagine two opposite scenarios: on the one hand, these two peaks may be connected by several links or they may be connected by unique link. Both examples would correspond to two different ways through which the technology evolves and a technological discontinuity emerges.

Figure 5 shows citations with a weight larger than 0.75, which allows us to include in the structure more patents and citations. The additionally included patents are indicated in white. Looking at the structure, we can observe that most of the citations newly included are in the area of the two separate components (where some triads are closed). This means that, in this case, persistent (and therefore important) knowledge links tend to cluster around the peaks rather than connecting them. Second, the citation network is by no means broken anywhere between the earlier and the latest patents and a single semipath connect the two components. More on this point: it is interesting to notice that the connection between the two components takes place through one single patent (patent 4245341).

As the structure is dependent on the choice of cut off points, Fig. 6 shows the resulting network when the threshold is lowered to 0.5.

The two structures are different but still comparable, despite the increase in the number of nodes and edges.

Looking at the whole structure, we can notice: (i) the emergence of a few short paths (indicated with A in Fig. 6), and (ii) the emergence of a shortest path connecting the two isolated components of Fig. 4 (indicated with B in Fig. 6). Following the analysis of the previous figures, it is interesting to look at the technical contents of these patents.

Despite the different vintage of the newly added patents indicated in A and B, they all disclose similar technological developments. They relate to the design of an hybrid switch (the ATM), including features of both circuit and packet switching. The early patents in A put forward the idea of packets, but still in a “connection-oriented” framework, whereas the patent

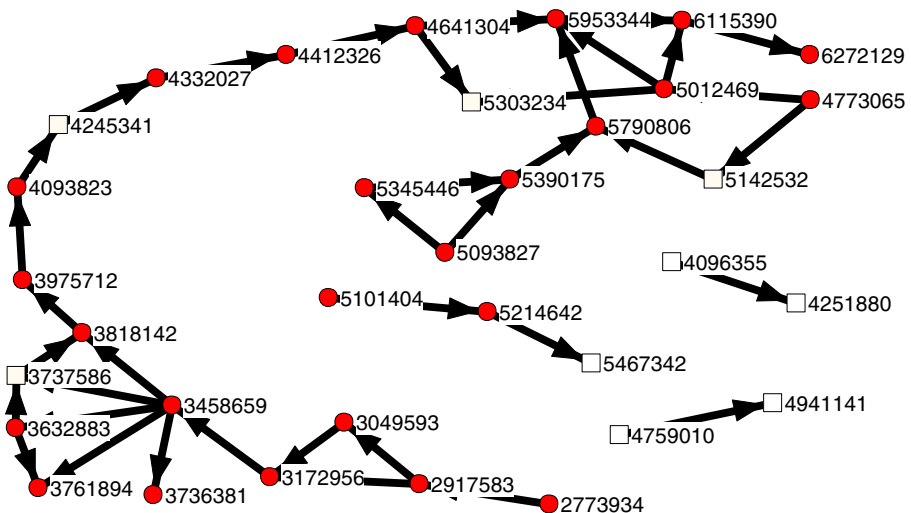
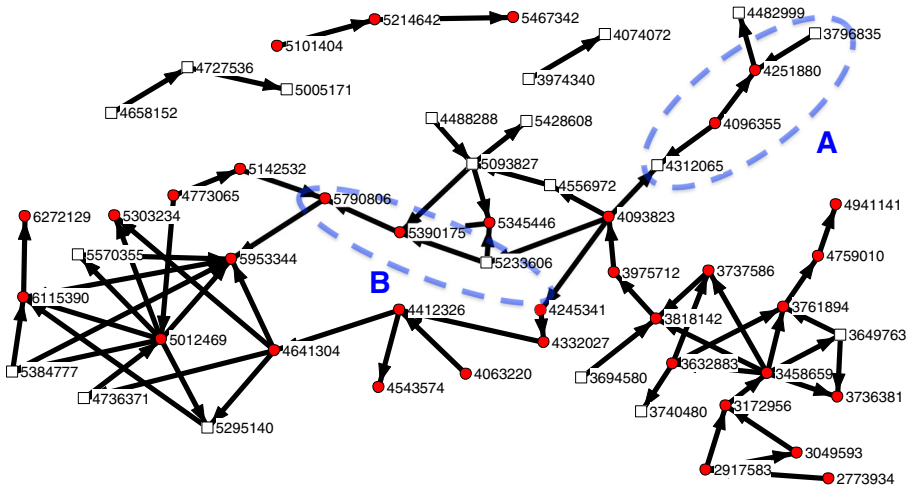


Fig. 5 Network with cut off point 0.75



**Fig. 6** Network with cut off point 0.5

in group B includes later patents related to the early development of ATM. ATM switches data are still divided into packets but the path is conceived at the outset, and all the packets are then sent through the same circuit. Despite its potential in mitigating some problems of packet switching (such as the decrease in the Quality of Service), it was still not optimal for quick demand increase for data communication, and therefore quickly dismissed by manufacturers.

The technical content of the patents included by lowering the threshold sheds some light on chains of less-and-less persistent innovation, detecting ways through which the technological space was explored. In the case of the telecommunication switching industry, this corresponds to abandoned technologies. In a dynamic perspective, this method highlights patents that contained relatively persistent knowledge but that are made obsolete with the emergence of new knowledge.

In order further to validate our finding, we can look at the assignees for these highly persistent patents in the network. The lower-right pane of Table 6 displays the top 10 assignees of the patents in the top quartile of the persistence index. All these companies have been top-players in telecommunication switches, however, specialized in different market segments and therefore on slightly different technologies. The average issue year of their persistent patents help to place such patents (and therefore firm's technological choices) in the technological evolution timeline. Bell Laboratories was the most advanced institution as regarding telecommunication switching and in particular in the use of circuit switching. Not surprisingly, its share of persistent patents are rather old and reflect the commercial introduction of digital switches based on circuit switching. As regarding

digital switches, Nortel also represents a case of success being able to internationalize rapidly in the early 1990s (Sutton 1998). By contrast, IBM has never active in the telephone switching market, as its main interest was in computer networking and data transmission. In this respect, its patent portfolio is slightly younger than that of Bell Laboratories. Consistently with what emerges from the history of telecommunication switching industry (Fransman 1995; Martinelli 2010), only a limited number of companies (among which the Japanese NEC and Fujitsu, and Nortel) were involved in developing of ATM switches (the branches A and B in Fig. 6) in the late 1980s. Finally, Motorola has been active in a specific and later-developed submarket designing and selling wireless network infrastructure equipment (for instance: cellular transmission base stations and signal amplifiers).

Going back to Section 4, we can now return on one of the difference between the GA and the HDA. In the empirical exercise just carried out, there is no reason to assume that technology evolution developed only along one path, especially in case of a technological discontinuity where lot of “search around” is performed. In the HDA, the search algorithm sequentially searches for a path between each *start-point* edges with the highest connectivity measure are sequentially selected up to an *endpoints*. It follows that, in the case of the HDA approach, the single ridge connecting the two peaks is endogenous in the search algorithm that does not allow for the emergence of multiple paths. On the contrary, in the analysis so far performed using the GA approach, no “greedy” algorithm is used and multiple connecting semi-paths might emerge. From these different algorithms, different considerations about the time structure emerge. If, on the one hand, in both cases we have the representation of technology evolution, on the other hand, in the GA, the time dimension is less constraining, but is still present in the direction of the arcs and in the numbering of patents.

## 7 Conclusion

The availability of patents and citations data has increased their use as innovation indicators. Recently, we observe the emergence of a complementary approach to patent and citation counts, which is to consider them from a network perspective. In fact, it is possible to exploit citations to map the technological relation between inventions (i.e. patents). In this setting, a patent citation network represents the space of the “technologically possible solutions”, the structure and dynamics of which characterize technical change in a specific technology.

The aim of this paper is to propose a new empirical method for identifying technologically important patents within a patent citation network and to apply it to the telecommunication switching industry.

The method proposed is inspired by population genetics: as geneticists are interested in studying patterns of migration and therefore the common origins of people, in innovation studies we are interested in tracing the origin and the evolution of today knowledge. In this respect, the genetic parallel is rather clear: as patent A

cites patent B, we can say that patent A inherits some knowledge from patent B. At the heart of this genetic method for analyzing patent and citation data is the persistence index that measures how much knowledge from older patents is retained (and therefore persists) into the recent patents. Accordingly, to the genetic parallel, the persistence index is calculated decomposing patent's knowledge applying the Mendelian law of genes inheritance. In this framework, the novelty of a patent (i.e. its inventive step) derives from the recombination of its "inherited" knowledge.

The empirical exercise of this paper consist in the use of this persistence index for identifying important patents and citations within the network. In particular, the analysis of the structure and evolution of the *persistence weighted network* allows us to unfold specific patterns of technological change in the case of the telecommunication switching industry.

If a network approach on patent data is not new, this paper suggests an alternative method overcoming some limitations in the existing approaches. In particular, the genetic approach is a flexible method that places very little assumptions on the possible outcomes of the emerging patterns of technical change. First of all, this approach is designed to account for differences in technology evolution, meaning that new technologies characterized by different levels of cumulateness or radicalness can display different network structures. In this respect, this method allows for testing hypothesis that are not confined to structural network properties but that are rooted in innovation studies. Second, the genetic approach does not make any assumptions about optimality or efficiency structure in the pattern of network evolution. Again, this means that it accommodates for differences in patterns of technical change. Finally, as regards the Hummond and Doreian approach, the method here presented has a different rationale with less emphasis on direct links and uniqueness of the emerging pattern of technical change.

Summarizing, we can conclude that the method proposed is successful in reducing the number of both nodes and links considered. This reduction might look arbitrary as it is based on cut-off points. However, the persistence index displays a highly left-skewed distributions that mimics a scale-free distribution. This makes easier to justify cut off choices and confirm the idea that few patents are "important" both in terms of economic value and in terms of knowledge contribution.

Furthermore, our method is indeed successful in identifying technological discontinuities where previous knowledge is not relevant for current technological development. Indeed, we can show that our methodology can be successfully used for identifying different technological paradigms are defined by Dosi (1982).

As regarding the broader impact of this new methodology, we can expect two lines of generalization. The first obvious one is the application of the persistence index to other technological fields in order to characterize their cumulative and disruptive dynamics. The scientific importance of such exercise is to foster

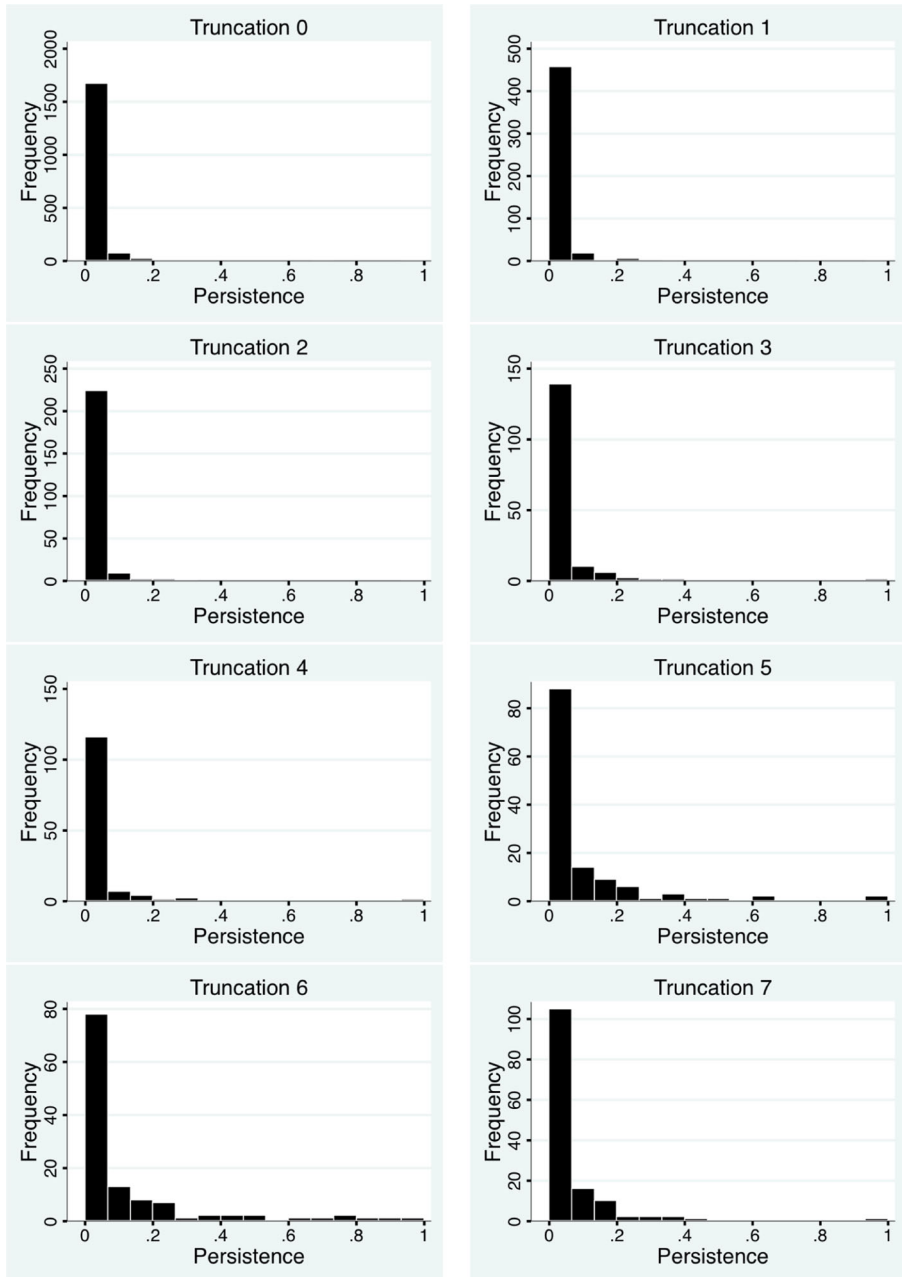
quantitative research and appreciative theorizing defined as a rigorous storytelling (Nelson 1989) in studies related to technology dynamics and its effect on industry dynamics.

The second one concerns the relevance of the methodology to other types of networks. The GA method could be used to analyze any graph representation of dynamic evolutionary processes (that follows the arrow of calendar time) where entities evolve into different/other entities by mutation and/or recombination while retaining various properties of their predecessors. In fact, all these processes can, in principle, be represented as a directed and acyclical graph on which it is possible to apply our method. Nature provides other examples of networks with these characteristics, such as: family trees, phylogenetic networks, food webs, feed-forward neural networks, and software call (Karrer and Newman 2009). Any such graph is a system that is potentially analyzable by our GA, which is essentially a fractional system of compound inheritance accounting. Related to the field of innovation studies, publication citation networks are another acyclical network “commonly” studied; the application and the interpretation of the persistence index in such context would be rather straightforward. In particular, our indicator could be used as a scientometrics tool to evaluate the long-run impact of individual papers and indirectly their authors/affiliated organizations. In historical perspective, the persistence approach differentiates between fashion of the day and long lasting influence, also refereed in network analysis as the difference between popularity and prestige. Finally, an other interesting application related to innovation could be for studying hierarchical organizations of learning (i.e., successions of master(s) and apprentices or supervisors and PhD. students) to find out the most (i.e., persistently) influential/prolific individual or different lineages of styles/paradigm.

Looking at the limitations of this method, we can identify two main caveats. If, on the one hand, this method can potentially be applied for comparative studies about technology dynamics in different technology, on the other hand, this is only possible for sectors in which patents are used for innovation appropriability. Limitations on the methodology itself relate to the (necessary) assumptions of the use of the Mendelian laws for the operationalization of the persistence indexes. However, in our connotation of the links between patents, the network under examination represents a system of both knowledge creation and retention. Therefore, it sounds plausible to assume that the innovative step in each patent is in the recombination of existing knowledge (i.e. inherited by the previous patents).

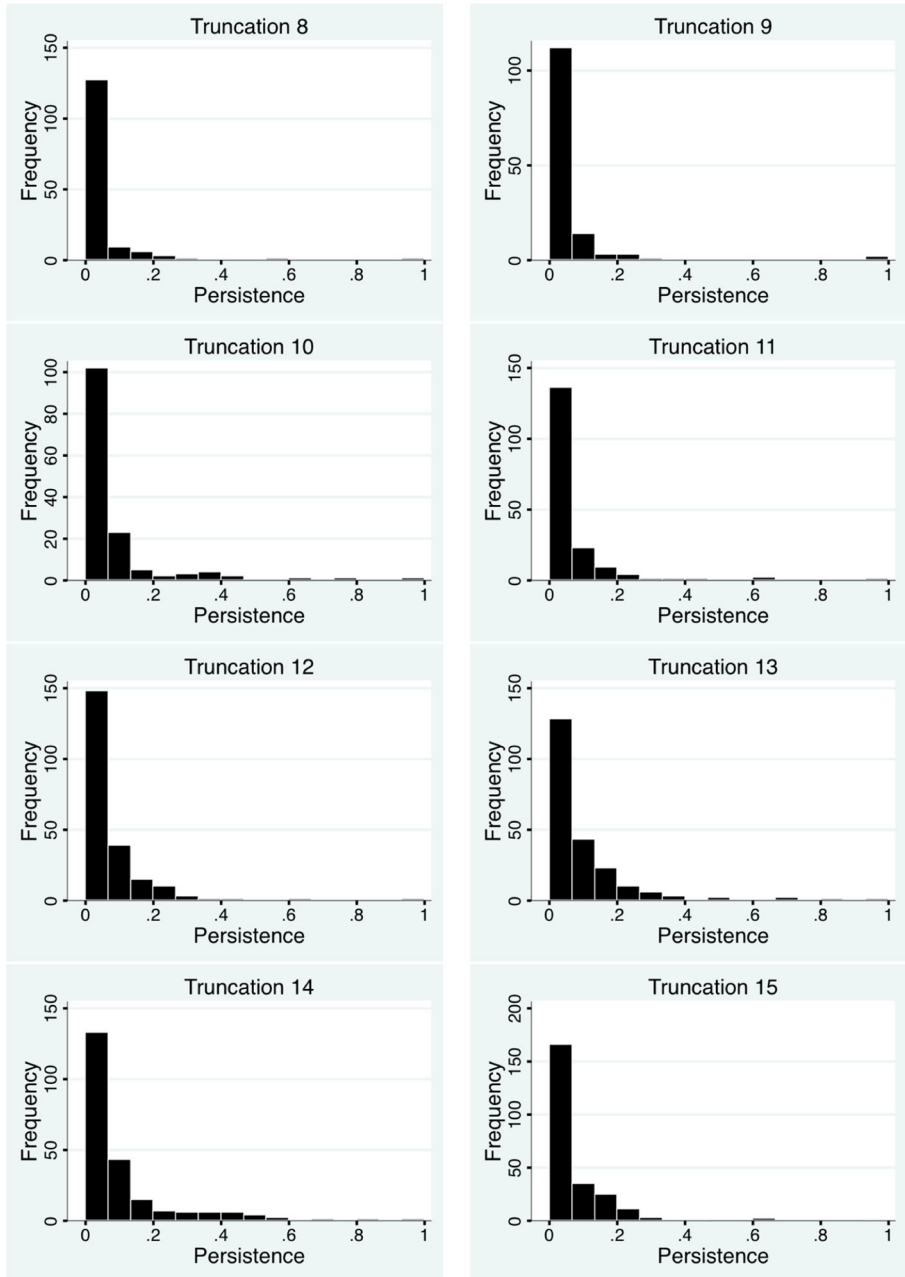
**Acknowledgements** This paper benefited from comments received at the DIME Workshop “Mapping Technological Change” held in Pisa in May 2010 and at the “Ratio Colloquium for Young Social Scientists 2013” held in Stockholm in August 2013. The authors would like to thank Bart Verspagen for helpful discussions and comments on previous drafts of the paper, as well as two anonymous referees for giving further suggestions. Arianna Martinelli would like to express endless gratitude to her colleague Alessandro Nuvolari for his support and help not to give up. Usual disclaims apply.

### Appendix A: Distribution of persistence index

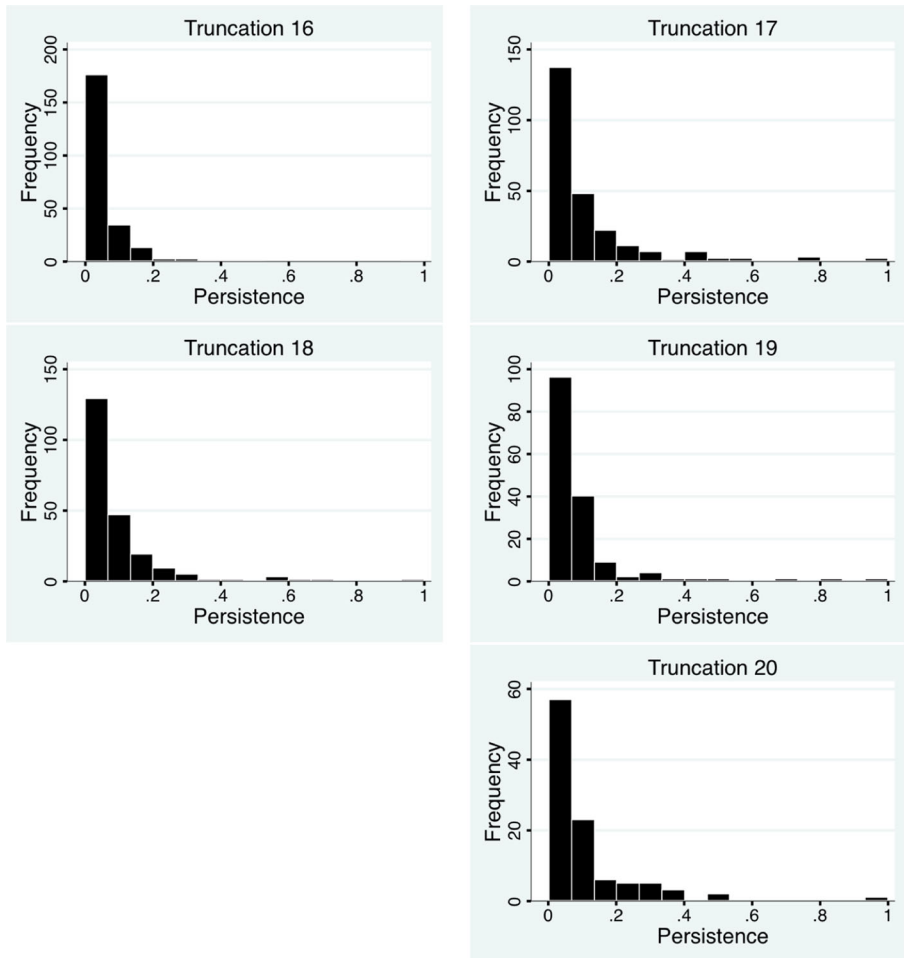


Histogram of the frequency distribution of the persistence index for the first 6 truncations





Histogram of the frequency distribution of the persistence index for the second 6 truncations



Histogram of the frequency distribution of the persistence index for the last 3 truncations

## Appendix B: List of persistent patents

**Table 7**

Number	Issue year	Technological class	Persistence index	Assignee
2773934	1956	179/15	1	GENERAL DYNAMICS
2917583	1959	370/362	1	BELL LABORATORY
3049593	1962	370/360	1	INT STANDARD ELECTRIC US
3172956	1965	370/376	1	BELL LABORATORY
3458659	1969	370/370	1	NEW NORTH ELECTRIC
3632883	1972	370/368	1	PHILIPS
3736381	1973	370/370	1	BELL LABORATORY
3818142	1974	370/370	1	ERICSON
3975712	1976	714/800	1	MOTOROLA
4093823	1978	370/535	1	CHUWESLEYW
4229792	1980	370/447	1	HONEYWEL
4332027	1982	370/448	1	BUROUGHS
4412326	1983	370/448	1	BELL LABORATORY
4641304	1987	370/447	1	RCA
4773065	1988	370/362	1	TOSHIBA
4941141	1990	370/376	1	NEC
5012469	1991	370/322	1	SARDANA KARAMVIR
5093827	1992	370/354	1	BELL LABORATORY
5214642	1993	370/471	1	HITACHI
5345446	1994	370/358	1	BELL LABORATORY
5390175	1995	370/398	1	ATT
5790806	1998	709/252	1	SCIENTIFIC ATLANTA
5953344	1999	370/443	1	LUCENT TECHNOLOGY
6115390	2000	370/443	1	LUCENT TECHNOLOGY
6272129	2001	370/356	1	3COM
4096355	1978	370/458	0.998	IBM
4074072	1978	370/388	0.994	BELL LABORATORY
5101404	1992	370/398	0.966	HITACHI
3761894	1973	710/53	0.924	BELL LABORATORY
4727536	1988	370/468	0.859	GENERAL DATACOM
5467342	1995	370/253	0.815	SCIENTIFIC ATLANTA
4251880	1981	370/468	0.804	BELL LABORATORY
4759010	1988	370/379	0.804	NEC
5161152	1992	370/463	0.797	ALCATEL
3737586	1973	370/370	0.788	BELL LABORATORY
4245341	1981	370/535	0.785	NATIONALE INDUSTRIEL AROSPATIALE
5303234	1994	370/442	0.774	NEC

**Table 7** (continued)

Number	Issue year	Technological class	Persistence index	Assignee
5142532	1992	370/432	0.768	BRITISH TELECOMUNICATION
3770897	1973	370/510	0.758	ITT
3770897	1973	370/510	0.758	LINK LABORATORY
3456242	1969	709/251	0.739	DIGIAC
3694580	1972	370/371	0.728	BELL LABORATORY
4658152	1987	370/535	0.717	BELL COMUNICATION RESEARCH
4556972	1985	370/354	0.705	BELL LABORATORY
3796835	1974	370/355	0.691	IBM
5327428	1994	370/353	0.682	IBM
5570355	1996	370/352	0.676	LUCENT TECHNOLOGY
4736371	1988	370/236	0.675	NEC
4482999	1984	370/452	0.661	IBM
5894477	1999	370/353	0.652	NORTEL
3649763	1972	370/372	0.65	BELL LABORATORY
4312065	1982	370/230	0.65	TEXAS INSTRUMENT
4408323	1983	370/389	0.65	BELL LABORATORY
3854011	1974	370/510	0.645	GENERAL DYNAMICS
5953330	1999	370/352	0.643	LUCENT TECHNOLOGY
5005171	1991	370/522	0.64	BELL LABORATORY
3740480	1973	370/370	0.633	BELL LABORATORY
4763319	1988	370/397	0.623	BELL COMUNICATION RESEARCH
5355362	1994	370/222	0.618	NEC
4488288	1984	370/393	0.607	BELL LABORATORY
5428608	1995	370/261	0.594	ATT
5623491	1997	370/397	0.593	DSC COMUNICATION
4852089	1989	370/468	0.571	DATA GENERAL
5384777	1995	370/337	0.568	IBM
4063220	1977	340/825	0.563	XEROX
5351236	1994	370/358	0.563	BELL LABORATORY
5982767	1999	370/352	0.555	MITEL
5233606	1993	370/418	0.55	BELL LABORATORY
4621357	1986	370/370	0.549	BELL LABORATORY
5276678	1994	370/267	0.539	INTELECT
3974340	1976	370/228	0.536	ERICSON
5144619	1992	370/353	0.526	NORTEL
5295140	1994	370/443	0.525	MOTOROLA
4679190	1987	370/355	0.523	IBM
4543574	1985	340/825	0.522	NTT

**Table 7** (continued)

Number	Issue year	Technological class	<i>Persistence index</i>	Assignee
3766322	1973	370/422	0.519	PLESSY
4764921	1988	370/510	0.515	NORTEL
4764921	1988	370/510	0.515	NTT
4947388	1990	370/411	0.512	HITACHI
2754367	1956	179/15	0.509	GENERAL ELECTRIC
5784369	1998	370/358	0.5	ERICSON

## References

- Albert R, Barabási AL (2002) Statistical mechanics of complex networks. *Rev Mod Phys* 74(1):47–98
- Balconi M, Breschi S, Lissoni F (2004) Networks of inventors and the role of academia: an exploration of italian patent data. *Res Policy* 33(1):127–145
- Barabási AL, Albert R (1999) Emergence of scaling in random networks. *Sci* 286(5439):509–512
- Barberá D, Jiménez F, Castelló I (2010) The strategic importance of the real world: connectivity analysis of the evolution of the artificial disc. *Res Policy* (Forthcoming)
- Bessen J (2009) Estimates of patent rents from firm market value. *Res Policy* 38(10):1604–1616. doi:10.1016/j.respol.2009.09.014. <http://www.sciencedirect.com/science/article/pii/S0048733309001905>
- Bottazzi G, Pirino D (2010) Measuring industry relatedness and corporate coherence, IEM Working paper 2010/10
- Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton
- Chapuis RJ, AEJ J (1990) Electronics, computers and telephone switching. North-Holland Publishing Company, Amsterdam
- Dosi G (1982) Technological paradigms and technological trajectories. *Res Policy* 11(3):147–162
- Ellis P, Hepburn G, Oppenheim C (1978) Studies on patent citations networks. *J Doc* 34(1):12–20
- Fontana R, Nuvolari A, Verspagen B (2009) Mapping technological trajectories as patent citation networks. An application to data communication standards. *Econ Innov New Tech* 18(4):311–336
- Fransman M (1995) Japan's computer and communications industry. Oxford University Press, Oxford
- Gambardella A, Dietmar H, Verspagen B (2008) The value of European patents. *Eur Manag Rev* 5:69–84
- Garfield E (1979) Citation indexing: its theory and application in science, technology and humanities. Wiley Interscience, New York
- Griliches Z (1991) Patent statistics as economic indicators: a survey. *J Econ Lit* 28(4):1661–1707
- Hall B, Jaffe AB, Trajtenberg M (2001) The NBER patent citation data file: lessons, insights and methodological tools. NBER Working Paper (8498)
- Harhoff D, Scherer FM, Vopel K (2005) Patents: economics, policy and measurement, Cheltenham, U.K. and Northampton, Mass.: Elgar, cig working papers Exploring the Tail of Patented Invention Value Distributions. FS IV 97–27. <http://ideas.repec.org/p/wzb/wzebiv/fsiv97-27.html>
- Hummon NP, Doreian P (1989) Connectivity in a citation network: the development of dna theory. *Soc Network* 11:39–63
- Jaffe AB, Trajtenberg M (2005) Patents, citations, and innovations: a window on the knowledge economy. The MIT Press, Cambridge
- Karrer B, Newman MEJ (2009) Random graph models for directed acyclic networks. *Phys Rev E* 80(046110)
- Marco AC (2007) The dynamics of patent citations. *Econ Lett* 94(2)

- Marsili O, Salter A (2005) Inequality of innovation: skewed distributions and the returns to innovation in dutch manufacturing. *Econ Innov New Tech* 14(1–2):83–102
- Martinelli A (2010) The dynamics of technological discontinuities: a patent citation network analysis of telecommunication switches. PhD Thesis, Eindhoven University of Technology, Eindhoven, The Netherlands
- Martinelli A (2012) An emerging paradigm or just another trajectory? understanding the nature of technological changes using engineering heuristics in the telecommunications switching industry. *Res Policy* 41(2):414–429
- Mina A, Ramlogan R, Tampubolon G, Metcalfe JS (2007) Mapping evolutionary trajectories: applications to the growth and transformation of medical knowledge. *Res Policy* 36(5):789–806
- Nelson RR (1989) The tension between process stories and equilibrium models. In: Langlois, R (ed) *Economics as a process: essays in the new institutional economics*, Cambridge University Press, chap economics as a process: essays in the new institutional economics
- Silverberg G, Verspagen B (2007) The size distribution of innovations revisited: An application of extreme value statistics to citation and value measures of patent significance. *J Econom* 139(2):318–339
- Strandburg KJ, Csárdi G, Tobochnik J, Érdi P, Zálányi L (2009) Patent citation network revisited: signs of a twenty-first century change? *North Carolina Law Rev*:87
- Sutton J (1998) *Technology and market structure*. The MIT Press
- Trajtenberg M, Jaffe AB, Henderson R (1997) University versus corporate patents: a window on the basicness of invention. *Econ Innov New Tech* 5(1):19–50
- Valverde S, Solé RV, Bedau MA, Packard N (2007) Topology and evolution of technology innovation networks. *Phys Rev E (Stat Nonlinear Soft Matter Phys)* 76(5):056, 118
- Verspagen B (2007) Mapping technological trajectories as patent citation networks: a study on the history of fuel cell research. *Adv Complex Syst* 10(1):93–115
- von Wartburg I, Teichert T, Rost K (2005) Inventive progress measured by multi-stage patent citation analysis. *Res Policy* 34(10):1591–1607. doi:10.1016/j.respol.2005.08.001. <http://www.sciencedirect.com/science/article/B6V77-4H57JH3-2/2/d9672939678a99e9591964265e0f2425>
- Wasserman S, Faust K (1994) *Social network analysis*. Cambridge University Press
- Weitzman ML (1996) Hybridizing growth theory. *Am Econ Rev* 86(2):207–212
- Weitzman ML (1998) Recombinant growth. *Q J Econ* 113(2):331–360
- Yoon J, Kim K (2011) Identifying rapidly evolving technological trends for r&d planning using sao-based semantic patent networks. *Scientometrics* 88(1):213–228