Computer Physics Communications 182 (2011) 1875–1878

Contents lists available at ScienceDirect

Computer Physics Communications

www.elsevier.com/locate/cpc

The diffusion of technological knowledge through interlaced networks

Leonardo Costa Ribeiro [a,1], Ricardo Machado Ruiz [a], Eduardo da Motta e Albuquerque [a], Américo T. Bernardes [b,*]

[a] Departamento de Economia, FACE, Universidade Federal de Minas Gerais, 30161-970, Belo Horizonte, MG, Brazil
[b] Departamento de Física, ICEB, Universidade Federal de Ouro Preto, 35400-000, Ouro Preto, MG, Brazil

ARTICLE INFO

Article history:
Received 16 June 2010
Received in revised form 29 November 2010
Accepted 2 December 2010
Available online 4 December 2010

Keywords:
Scale-free networks
Patents
Technological diffusion

ABSTRACT

In the last decade, many papers discussed the basic features of the networks formed by knowledge diffusion. In this paper, we show that a network formed by the spreading of technological information, represented by the patents citation network, does not obey the expected power law behavior and, therefore, is not a scale-free network. We mapped a network formed by almost 600 000 patents, covering a 40 years period. Although the complete network is not scale-free, small portions within the whole network can be described by power laws. The network combines several fields of knowledge, but those power law portions belong to specific fields of technological knowledge. A mathematical model is introduced, which can explain the basic dynamics of the formation of this network.

1. Introduction

The study of complex networks and the mechanisms underlying their formation gave an important direction to the research on the diffusion of information [1,2]. It has been suggested that the network created by scientific citations follow a well-established picture of a scale-free network [3]. Thus, citations should obey the basic mechanisms of preferential attachment, as in other examples of social networks [4–6]. Technological development may be represented by the deposit of patents [7], and the citation of patents by patents also creates a network.

The patent concept was created to protect the intellectual production, ensuring, for a limited period of time, the exclusive rights of an author's discovery. Such protection was of great value to various sectors, stimulating the creation of new products or new uses of old discoveries. A patent must be registered at the United States Patent and Trademark Office (USPTO) in order to be valid within the North American territory, which may generate economical benefits to its author. Once registered, the patent becomes public and its content is made available at the USPTO website.

Jaffe and Trajtenberg [8] presented a broad review of the literature regarding patents as indicators of technological innovation and patent citations as a research tool. This literature is the starting point for our effort to investigate networks since patent citations have the property of network formation (or are expressed in the form of networks). Valverde et al. [9] studied some proprieties of the Trajtenberg's network, confirming the basic assumptions on the structure of this kind of network presented in other papers.

In this paper, instead of working on the whole network, we selected a specific one, which starts with a patented revolutionary technological innovation: the first application of NMR for image diagnosis of cancer in humans, USPTO patent number 3 789 832, authored by Damadian [10]. This is the first registered patent in USPTO regarding the application of nuclear magnetic resonance (NMR) for image diagnosis of cancerous tissues in humans. Once this information was recorded, we obtained the list of all patents that cited Damadian's patent and recorded the data of all those patents, constituting the first network layer. Next, we scanned the list of all patents which cited at least one of those in the first layer and recorded their information, thus constituting the second network layer. The subsequent layers were built in the same fashion until we reached the border of the network, the layer in which the patents have no citations, thus finishing the construction of this citation network. This is an unidirectional network given that this network was created following a time direction. Since USPTO made available the information on the year of patent register, it is possible to obtain the dynamics of the network's formation.

2. Analysis of the patent citation network

The "NMR network" consists of 32 layers of citations, 598 360 patents (nodes) and 2 720 450 citations (links) between patents. To

* Corresponding author.
E-mail address: atb@iceb.ufop.br (A.T. Bernardes).

[1] Permanently at Instituto Nacional de Metrologia, Normalização e Qualidade Industrial – Inmetro, 25250-020, Duque de Caxias, RJ, Brazil.

0010-4655/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.cpc.2010.12.005

investigate the structure of this network we calculated its connectivity distribution, shown in Fig. 1. Contrary to usual assumptions, it is not a power-law shaped tail, which usually allows one to classify the network as scale-free. The points on this logarithmic plot can be fitted by a second degree polynomial (continuous line in the major plot) with a correlation coefficient of 0.98.

To understand the whole network's structure, we have identified networks interlaced within it. We analyzed the connectivity distribution of these small networks, which we call subnetworks. The first analyzed was the one formed from the most connected site, the network's hub.

The insert in Fig. 1 shows the distribution of connectivity for this subnetwork. In contrast to what was observed for the network as a whole, in this case the connectivity distribution follows a power law. Following this procedure, and in order to obtain a general description of the network, we analyzed several subnetworks, starting from randomly chosen nodes belonging to the whole citation's network. We observed that these subnetworks can be divided into two different groups:

(i) in the first group are those subnetworks which present a connectivity distribution which is well represented by a power law; an example of this subnetworks distribution is that shown in the insert in Fig. 1;

(ii) in the second group are the subnetworks that do not present connectivity distributions that might be represented by a power law. Similar to the whole network, the points of their connectivity distribution on a logarithmic plot may be represented by a second degree polynomial. However, it is important to note that in this second group small subsubnetworks can be identified within a subnetwork, and some of them are described by power laws.

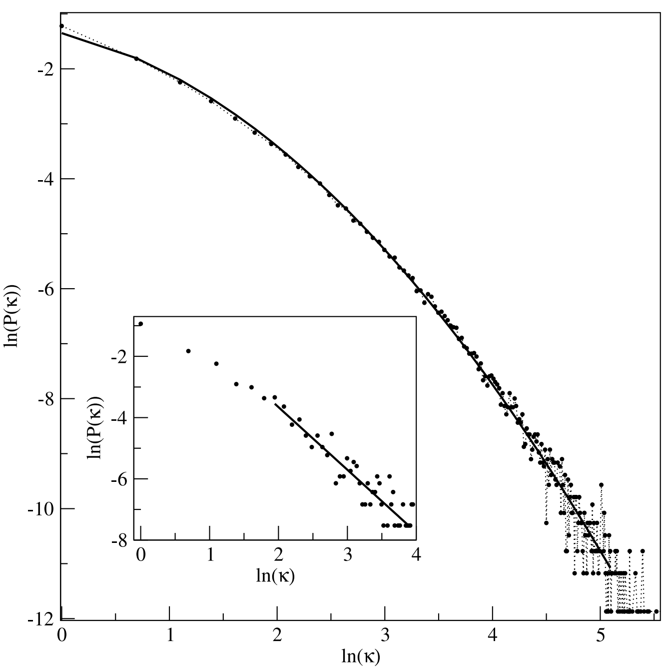

Fig. 1. Logarithm of the connectivity distribution $P(\kappa)$ as a function of the logarithm of the connectivity κ for NMR patent citation's network (filled circles). A second degree polynomial has been used (continuous line) to fit the data points. The insert shows the logarithm of the connectivity distribution $P(\kappa)$ of the subnetwork built from the network's hub (filled circles). A power law function has been used to fit the points (continuous line).

This study allows us to create a mental model for the network as being formed by many scale-free networks connected to each other. In this case, when the chosen site is such that the subnetwork formed starting from it is simple scale-free, we would have the distribution in the form of a power law for the subnetwork. Otherwise, when the chosen site is such that the subnetwork formed starting from it contains many scale-free subnetworks, we would have the distribution in the form of a second degree polynomial (like the whole network, which is supposedly also formed by many scale-free subnetworks). Fig. 2 shows schematically this idea.

In Fig. 3 we show the mean connectivity per site $\langle M \rangle$ versus the number of sites for several subnetworks. One can observe that the larger the subnetwork (higher the value of N), the smaller its mean connectivity per site is. Or, the higher the connectivity, the smaller the subnetwork is.

Since a patent represents an innovation and since its content is public, diffusion of knowledge may occur. A patent can lead to the creation of another patent. Thus, the citation of a patent by another one means the spreading of the original knowledge: knowledge diffusion. In order to understand the dynamics of this diffusion, let us return to the process which formed the network. Initially, Damadian's patent represented a technological innovation of the greatest relevance within a certain field of knowledge. Since it was a very important innovation, several patents cited it. The formation of a network starts from this point. Many of them may represent improvements of a product or a process. But at a given point of this dynamic process, another patent appeared as a relevant innovation. This innovation might be either in a technological field, as is the case of Damadian's patent, or in a different area.

The concept of technological sub-domains was proposed by the *Observatoire des Sciences et des Techniques* (OST) [11]. It allows us to classify patents in 30 areas of technological knowledge. The complete classification of these sub-domains can be found at [11]. Since a patent in the USPTO database is classified according to a US scheme, we thus converted the US classes and subclasses of each patent to the technological sub-domains defined by OST.

We observe that the scale-free portions of the "NMR network" belong to specific technological sub-domains. More clearly, we selected two subnetworks: N_{sf} presents the connectivity distribution in the form of a power law; and N_{p2} with the points in the loga-

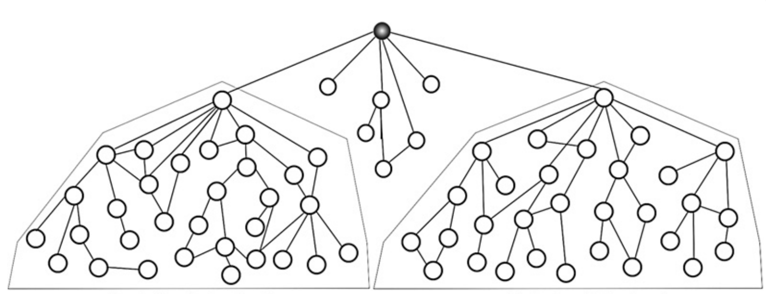

Fig. 2. Representation of a patent citation network's portion, in which the circles represent the patents and the lines the citations between them. This picture illustrates the case in which the subnetwork generated from the initial site (filled circle) consists of many scale-free networks.

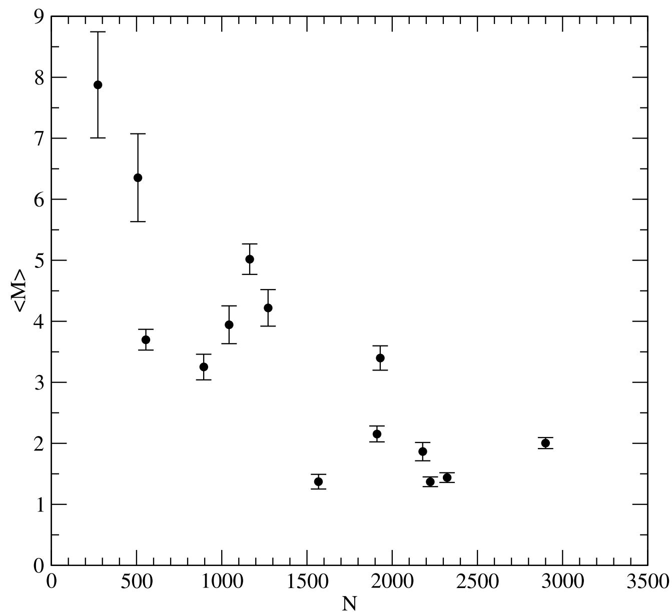

Fig. 3. Average connectivity per node $\langle M \rangle$ as a function of the number of nodes N for several subnetworks generated from randomly chosen nodes in the whole patents citation's network.

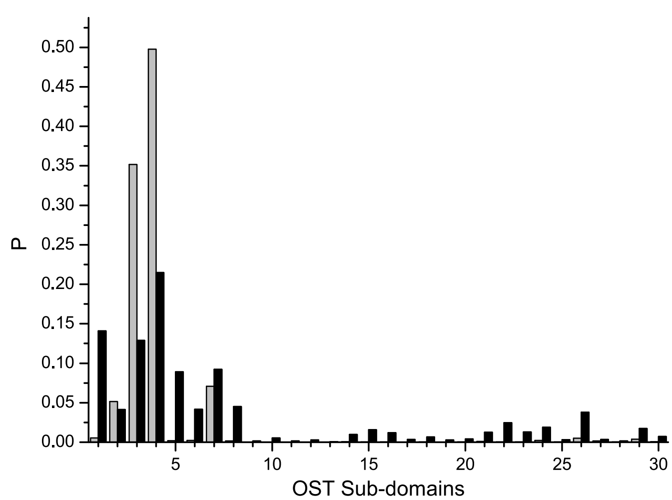

Fig. 4. Relative frequency P of patent's technological sub-domains obtained for: (a) a scale-free subnetwork (gray bars); and (b) a subnetwork formed by several interlaced scale-free $\{\text{sub}\}^n$networks (filled bars). See text for discussion.

rithm plot represented by a second degree polynomial. The latter, as discussed above, is formed by several interlaced subnetworks, some of which are scale-free. In Fig. 4 the gray bars represent the relative frequency of patents by technological sub-domains for the scale-free subnetwork N_{sf}. A small dispersion is noted in the distribution: 90% of the patents are classified into sub-domains 3 and 4. Filled bars in Fig. 4 represent the distribution of patents by sub-domains for the N_{p2} subnetwork. A higher diversity of technological sub-domains is observed.

Scale-free networks have the small world feature, i.e., starting from a node, we can reach another node with few steps (that famous "six degrees of separation"). So, different from other types of networks, diffusion occurs rapidly in scale-free networks. From this point of view we can conclude that the knowledge of an innovation belonging to a specific technological sub-domain diffuses rapidly by means of a network in which all the nodes belong nearly to the same specific sub-domain. Hence, it requires the creation of a scale-free network of citations. But it hardly permeates into other technological sub-domains. A certain relevant patent generates a scale-free citation network within the same sub-domain, in which the efficiency of the diffusion process is very high. The dynamics of creation of the whole network is thus dominated by the formation of several scale-free subnetworks which separately diffuse the knowledge in an efficient way.

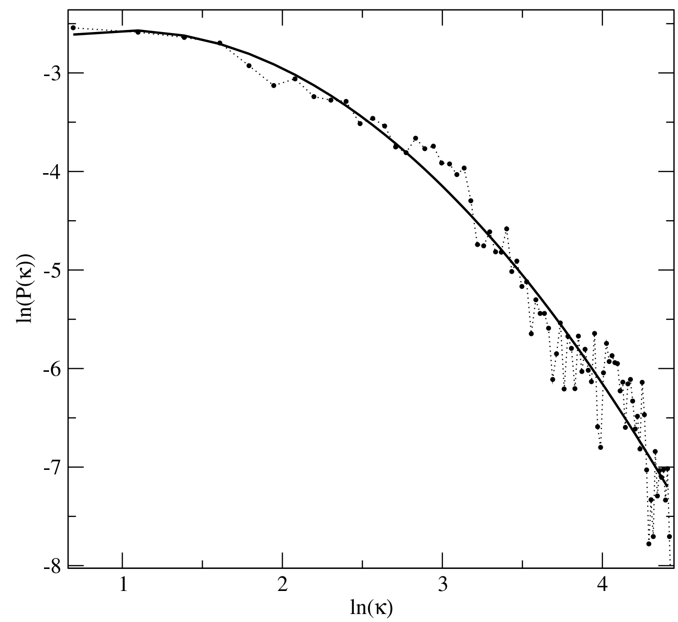

Fig. 5. Logarithm of the connectivity distribution $P(\kappa)$ as a function of the logarithm of the connectivity κ of the proposed model for the patent citation network (filled circles). Regression, in $\ln \times \ln$ scale, of a second degree polynomial for the connectivity distribution (continuous line).

3. Model for the patent network

We now introduce a model for the patent citation's network. Our model creates a network formed by many scale-free networks of different sizes and connectivities, independently built and then connected to each other. All of them have been built following Barabasi's prescription [12,13]. This well known procedure starts from a nucleus of m_0 connected nodes and, at each step of the construction, a new node i is added and connected to m nodes of the network. The probability of choosing a site j is proportional to the connectivity K_j of the node j. The link between two networks N_α and N_β is made by creating a link between two randomly chosen sites i_α and j_β. Once again, the probabilities of choosing those sites are proportional to their connectivities $K_{i,\alpha}$ and $K_{j,\beta}$. This process creates a network consisting of coupled subnetworks.

The values for m_ι, $\iota = \alpha, \beta, \ldots$, were randomly chosen from a uniform probability distribution, defined between a given interval of values. Given m_ν, we define the number of sites to be added, following the observation described above: the greater the connectivity per site m_ν, the smaller the subnetwork. Thus, the size of the network was chosen to be inversely proportional to m_ν. The value of m_0 is the same for all subnetworks, and its choice does not interfere in the properties of the network.

Fig. 5 shows the connectivity distribution of a network obtained by our model. In this case, a network was built with 5000 subnetworks. m_ι were chosen within the interval $[2, 22]$ and the size N_ι of each subnetwork is given by $N_\iota = 2000/m_\iota$. As observed in the patent citation's network, the connectivity distribution of this model in a logarithmic plot is represented by a second degree polynomial. The correlation coefficient is $\sim$0.95.

The diffusion process has also been simulated in our model. In order to achieve a better understanding of the dynamics of diffu-

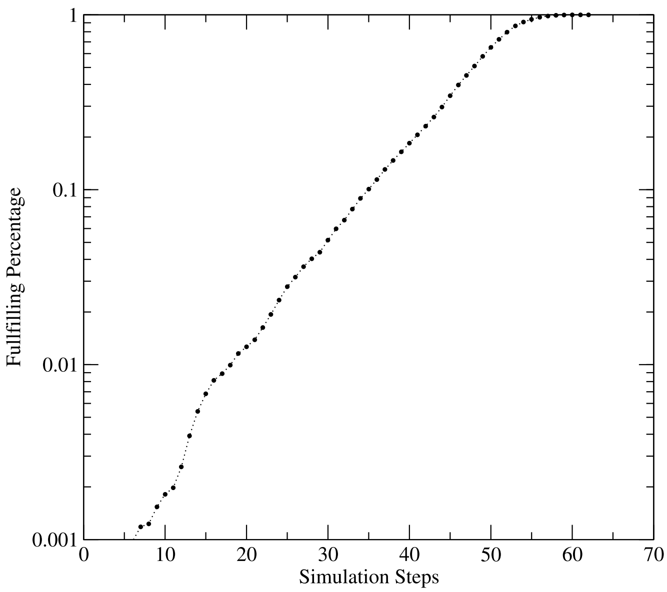

Fig. 6. Evolution of the percentile of filling of the network by sites with label 1 in function of the number of steps of the simulation (circles). For intermediary steps we have the regression of an exponential function (continuous line).

sion, we compare two networks: the one described above $N_{p2,sim}$ and a scale-free $N_{sf,sim}$ with the same number of nodes. Initially, all sites of both networks were labelled as 0. In the first step, we set the initial node to 1. Next, at each new step, we set to 1 the label of all nodes which are connected to a site of label 1, and so on.

Fig. 6 shows the dynamics of labelling. In the present case, 62 steps were necessary to cover the whole network. We performed simulations starting from different nodes and the results are similar.

For a scale-free network $N_{sf,sim}$, starting from a hub, 7 steps are required to cover the whole network. Similar numbers are obtained if one starts at different nodes. On the other hand, almost 60 steps were required to completely cover $N_{p2,sim}$. The network labelling follows an exponential law for times smaller than a certain value, above which the process moves slowly to saturation, meaning that certain points are difficult to reach. Thus, as $N_{p2,sim}$ does not present the small world feature, it has a low diffusion efficiency at a global scale.

The concept we propose to explain the behavior of the diffusion is that these technologies are very sectoral, that is, any given innovation related to an area X is intensely used and diffused only within its specific area and will hardly permeate into other areas. Thus, a certain relevant patent generates a citation network within its area in the form of a scale-free network, in which the efficiency of the diffusion process is very high. We have then several scale-free subnetworks which separately diffuse the knowledge in an efficient way. However, when we analyze the network as a whole, since the connectivity between these subnetworks is low, we would have little diffusion.

4. Conclusion

In conclusion, we have shown evidence that the process of technological diffusion (through a patent citation's network) is efficient at the level of specific technological sub-domains, by means of creating a scale-free network of patent citations. We have also shown that diffusion through the whole network does not present the same efficiency. These conclusions are supported by the fact that we have described the topology of the patent citations network, and observed that, in contrast to other previous works, its connectivity distribution cannot be represented by a power law. In addition, from the fact that our observations show that the efficient spread of knowledge takes place at the level of specific sub-domains, we argue that other networks found in literature and that represent processes of knowledge diffusion are in fact scale-free.

Acknowledgements

This work was supported in part by the Brazilian agencies CAPES, CNPq, and FAPEMIG.

References

[1] R. Albert, A.L. Barabasi, Rev. Mod. Phys. 74 (2002) 47–97.
[2] S.N. Dorogovtsev, A.V. Goltsev, J.F.F. Mendes, Rev. Mod. Phys. 80 (2008) 1275–1335.
[3] A.L. Barabasi, et al., Physica A 311 (2002) 590–614.
[4] G. Palla, I. Derenyi, I. Farkas, T. Vicsek, Nature 435 (2005) 814–818.
[5] G. Kossinets, D.J. Watts, Science 311 (2006) 88.
[6] J.M. Kertesz, et al., Comput. Phys. Commun. 180 (2009) 517–522.
[7] United States Patent and Trademark Office (USPTO), USPTO is the world's largest deposit of patents. Once registered at USPTO, a patent becomes public and its content is made available at the USPTO website: Some information are available: date of registration, name and country of the authors, assignee, US class, patent and non-patent references, http://www.uspto.gov.
[8] A. Jaffe, M. Trajtenberg, Patents, Citations & Innovation: A Window on the Knowledge Economy, The MIT Press, Cambridge, 2002.
[9] S. Valverde, et al., Physical Review E 76 (2007) 056118.
[10] USPTO Patent Number 3 789 832, page www.uspto.org.
[11] Observatoire des Sciences et des Techniques, Science & Technologie: Indicateurs 2006, Economica, Paris, 2006.
[12] A.L. Barabasi, R. Albert, Science 286 (1999) 509–512.
[13] R. Albert, H. Jeong, A.L. Barabasi, Nature 406 (2000) 378–382.