Physica A 388 (2009) 4267–4272

Contents lists available at ScienceDirect

Physica A

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

Characteristics of the network of scientific journals pertaining to Chinese patents

Xia Gao [a,b,*], Jiancheng Guan [c]

[a] School of Economics and Management, Inner Mongolia University, 010021 Hohhot, PR China
[b] School of Management, Beijing University of Aeronautics and Astronautics, 100083 Beijing, PR China
[c] School of Management, Fudan University, 200433 Shanghai, PR China

ARTICLE INFO

Article history:
Received 20 July 2008
Received in revised form 25 November 2008
Available online 24 June 2009

Keywords:
Network
Scientific journals
Chinese patents

ABSTRACT

We present the network of scientific journals pertaining to Chinese patent data in the period of 1995–2002, in which two journals are considered linked if they have been cited by a common patent. We study a variety of statistical properties of the network of scientific journals of China (NSJC), including degree distribution, local clustering and average degree of the nearest neighbors in the one-mode projected network. Besides this, we generate a weighted network for the NSJC in which the number of common patents citing two particular journals is mapped to line weights. For such a weighted network, distributions of weight, vertex strength, weight per degree, and the relationship between the vertex strength and degree have been analyzed. The above findings show that for the NSJC, small-world behavior is not distinct, while properties of random networks are observed.

1. Introduction

In recent years, the complex network theory has been extensively used to describe a vast number of natural, physical and social systems, such as metabolic networks [1], the World Wide Web [2], scientific collaboration networks [3,4], bus-transport networks [5] and so on. However, there are few studies on the relationship between science and technology in social science based on networks. In fact, the relationship between science and technology as a complex interplay between them is now more acceptable. A substantial body of empirical studies relevant to the linkage between science and technology have been done mainly in a quantitative and especially bibliometric manner. In these studies, patents and publications are often used as proxy indicators of technological and scientific activity. The majority of the quantitative contributions focus on analyzing the scientific literature cited in patents, so-called non-patent references (NPRs) [6–8]. As pointed out by Ref. [9], citation links between patents and papers signify, if not explicitly, at least implicitly the contribution of science to technology.

In this paper, we propose a network of scientific journals, resulting from the Chinese patent data of United States Patent and Trademark Office (USPTO) between 1995 and 2002. However, we are not interested in exploring the interaction between science and technology in China, but rather in obtaining some general statistical properties of the structure of the scientific journal network of China. We have discussed the former topic in another paper.

This paper is organized as follows. In Section 2, we give an introduction to the collection of the data and the construction of the network of scientific journals of China (NSJC). In Section 3, we present the one-mode projection of the NSJC. Network properties, including the degree distribution, clustering coefficient and average degree of the nearest neighbors, will be

* Corresponding author at: School of Economics and Management, Inner Mongolia University, 010021 Hohhot, PR China.
E-mail address: gaoxia1976@sina.com (X. Gao).

0378-4371/$ – see front matter
doi:10.1016/j.physa.2009.06.037

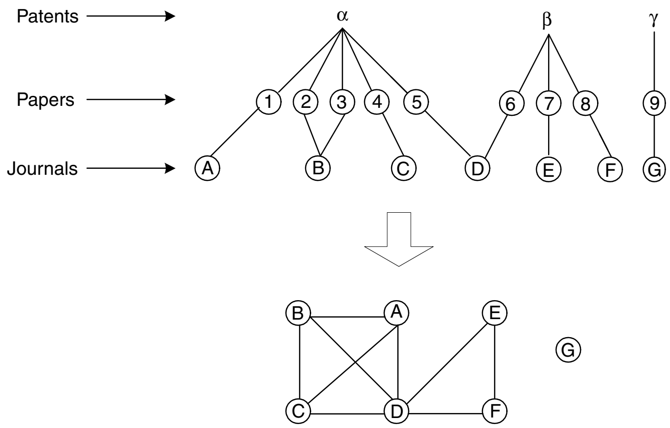

Fig. 1. Bipartite graph representation of the network of scientific journals. *Top:* Bipartite graph of patents (α, β, γ), papers (1 to 9) and journals (A to G), with lines linking each paper to the corresponding journals and patents, respectively. *Bottom*: The one-mode projection of the same network onto just journals.

discussed. In Section 4, we construct a weighted network representation for the NSJC, and some statistical quantities will be studied. Some concluding remarks are given in the last part, Section 5.

2. Chinese patent data and the scientific journal network

We use data on Chinese patents that were applied for at USPTO and disclosed between 1995 and 2002. Here we focus only on utility patents because they represent the progress of technology and more closely connect with scientific research. From those utility patents we firstly extract citations to scientific references, and further collect journal publications covered by the Science Citation Index (SCI). Altogether we could identify 3560 papers from the citations of 2546 utility patents, which are distributed over 724 SCI journals. Thus, we set up a patent–journal database, which is composed of patent numbers and the names of the corresponding SCI journals.

The above patent–journal database permits us to construct a network of SCI journals. The following hypothetical example illustrates the main idea (see Fig. 1). A reasonable assumption to make at this point is that non-patent citations within a patent can be taken as "representative for the scientific base of the citing patents" [10]. Due to the providers of the scientific source for a common citing patent, the relevant journals are 'linked' to each other by some kind of knowledge relation. As shown in the bottom part of Fig. 1, we end up with a map representing the network of linkage among all journals. In other words, we obtain the undirected network of scientific journals pertaining to Chinese patent data in the period of 1995–2002.

3. Unipartite projection of the network of scientific journals

From the original bipartite network, we can get the one-mode projected network where only the nodes representing SCI journals remain and are tied to each other whenever they are cited by the same patent. The one-mode projected network of scientific journals does not have weight or directions. In the following, we present some general measurements for the NSJC, including degree distribution, clustering coefficient, average degree of the nearest neighbors etc.

The cumulative degree distribution $P(k)$ of the one-mode projected network of scientific journals is plotted in Fig. 2(a). It can be presented as an exponential function:

$$P(k) = Ae^{-\alpha k}$$

where the exponent $\alpha = 0.0249 \pm 0.0012$. Note that the power-law forms of the statistical distributions are followed by most social networks [11,12], while only a few networks have the exponential distribution [13,14]. Growth and preferential attachment play an important role in the development of the power-law scaling. Here, the cumulative degree distribution $P(k)$ of our network is characterized by an exponential decay, indicating that the absence of preferential attachment eliminates the scale-free feature of the distribution [15]. Furthermore, as pointed out by Ref. [16], the exponential distribution occurs for evolving networks when nodes are attached completely randomly. This implies that the evolution of the NSJC possesses an accidental property. Besides, the average degree of the NSJC is $\langle k \rangle = 21.09$, which indicates that each journal has 'collaborated' with 21.09 other journals on average. This is quite close to the MEDLINE network case in Ref. [3]. The journal *Proceedings of the National Academy of Sciences of the United States of America (PNAS)* is obtained with the maximal degree $k_{\max} = 306$. In Fig. 2(b), we plot the degree for the NSJC as function of the number of times that each journal is cited by patents. It can be described by a linear function, which implies that journals receiving more citations from patents tend to have larger degree. The journal most cited by patents is also the *PNAS*.

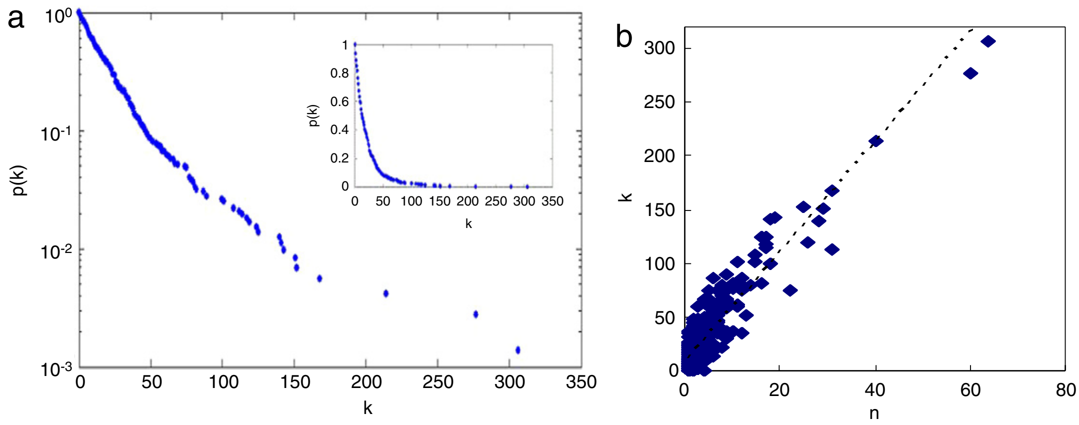

Fig. 2. (a) A log-linear plot of the cumulative degree distribution for the NSJC; a linear plot of data fitting points appears in the inset. (b) Degree versus the number of times that each journal is cited by patents.

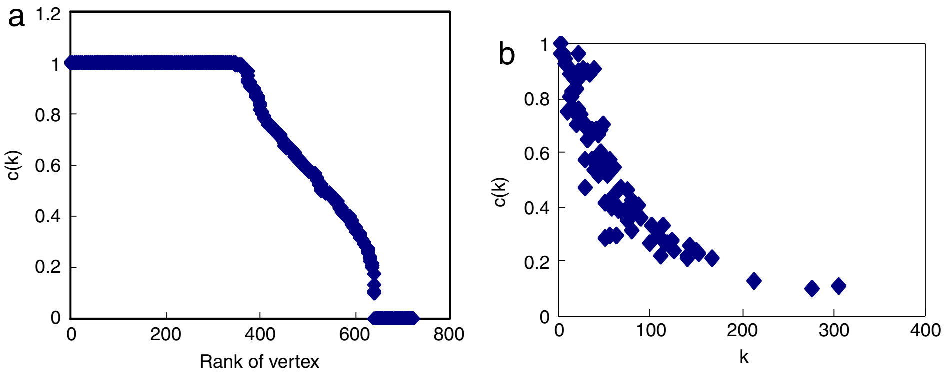

Fig. 3. (a) Zipf plot of the clustering coefficient for the NSJC. (b) Average local clustering $c(k)$ as a function of the degree.

The clustering coefficient is the measurement of the triangles a node has. Node i's clustering coefficient is defined as [13] $c_i = 2e_i/k(k-1)$, where k is the degree and e_i is the number of links connecting the k neighbors to each other. Fig. 3(a) gives the distribution of the local clustering coefficient for the NSJC in Zipf plots. We can observe a flat region, followed by a rapid decrease to zero. The average local clustering $\langle c(k)\rangle$ as a function of degree is depicted in Fig. 3(b). For small k the average local clustering remains close to unity. As k grows, $\langle c(k)\rangle$ decays as an exponential function $\langle c(k)\rangle \sim \mathrm{e}^{-\alpha k}$with $\alpha = 0.0089$. In many real networks, the clustering exhibits a scaling-law behavior, $\langle c(k)\rangle \sim k^{-\beta}$. A hierarchical and modular structure would be observed if the value of the exponent β was closer to 1 [17]. In our case, there is clear evidence that $\langle c(k)\rangle$ does not follow the scaling law. This means that the NSJC lacks a hierarchical and modular structure. The clustering coefficient is $c = 0.0425$. Besides, the average length is equal to 2.67. The above findings show that small-world behavior is not observed in the NSJC. The NSJC seems to have properties of random networks.

We will next discuss the average degree for neighbors. The mean degree of the nearest neighbors has a very informative magnitude in aspects of the network topology [18,19]. For a vertex i, it is defined as

$$k_{nn,i} = \frac{1}{k_i}\sum_{j\in\nu_i} k_j.$$

This quality may also be expressed as function of the degree, $k_{nn}(k)$, by averaging over nodes of the network with degree k [18]. If there are positive degree correlations between neighboring nodes (assortative mixing), $k_{nn}(k)$ should grow with increasing k. The contrary trend should be observed if the network shows anticorrelation between the degree of the neighbors (disassortative mixing). Fig. 4 shows the average degree of the network of all nodes. The horizontal axis shows the degree k, and the vertical axis shows the degree correlation $k_{nn}(k)$. Fig. 4 has a null correlation. So, there is no assortative mixing, and a hierarchical structure for the degree is not observed in the network.

4. Weighted representation for the network of scientific journals

A one-mode projection of the original bipartite network neglects multiple common collaborations, which loses some valuable information on the original network. A way to partially avoid such a loss of information is by the use of a weighted

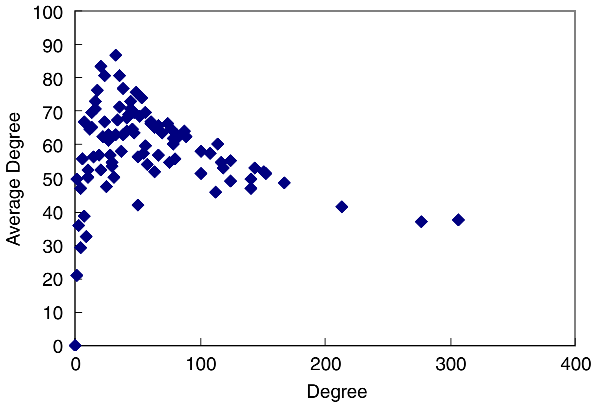

Fig. 4. Average degree of the nearest neighbors as a function of the degree k, $\overline{k_{nn}}(k)$, for the NSJC.

network for the projection [18]. In this work, the link weight w_{ij} is the number of patents that cite both journal i and journal j, simultaneously. The cumulative weight distribution, $P(w)$, is plotted in Fig. 5(a). This distribution follows a power-law behavior $P(w) \sim w^{-\delta}$ with $\delta \approx 2.7$. This finding confirms the predications made from the model for the growth of weighted networks in Ref. [20].

Next, we investigate the vertex strength, which measures a combination of weight and number of edges. For a given vertex i, the strength of i is defined as follows [18]:

$$s_i = \sum_{j \in \nu(i)} w_{ij}$$

where the sum runs over the set of all neighbors of i, $\nu(i)$. The strength of a vertex denotes the total number of partnerships (patents) in which a particular actor (journal) has been involved. The relationship between the vertex strength $s(k)$ and degree k can often be expressed as follows:

$$s(k) \sim k^{\beta}.$$

Fig. 5(b) is a plot of $s(k)$ versus k for the NSJC. In our case, $\beta \approx 1.0$, which means that the strength is proportional to its degree. This reconfirms that the weight-driven dynamics in Ref. [20] generates an effective degree preferential attachment that is parameter independent. A similar result is also observed in Refs. [5,21]. Fig. 5(c) presents the cumulative distribution of the strength $P(s)$. It shows that $P(s)$ decays slowly for small s while the decay is dramatic for large s. $P(s)$ can be described by an exponential function $P(s) \sim e^{-\gamma s}$, where $\gamma \approx 0.0016$. In fact, this expression is straightforward to obtain since $P(k) \sim e^{-\alpha k}$ and $s \sim k$.

In Ref. [22], the concept of weight per degree is introduced. WPD is defined as a ratio between the strength and the degree for each vertex. It measures the tendency of actors to keep on "collaborating" with previous partners. Fig. 5(d) gives the distribution of the WPD in a Zipf plot for the NSJC. We can observe a rapid decrease to WPD values close to 4, extending up to a flat region with WPD values of 2, followed by a sudden drop to zero. The average weight per degree for the NSJC is equal to 2.39, which is larger than that of some collaboration networks [12,18].

5. Conclusions

In this paper, we have analyzed the statistical properties of the network of scientific journals of China (NSJC). In the one-mode projected network, we find that the cumulative degree distribution follows an exponential distribution, which implies that the evolution of the NSJC possesses an accidental property. Also, the distribution of local clustering obeys an exponential distribution, which means that the NSJC lacks a hierarchical and modular structure. Besides, the average length for the NSJC is relatively low, but the clustering coefficient is not high. The above findings show that small-world behavior is not observed distinctly in NSJC. The NSJC seems to have properties of random networks. Furthermore, we regard the NSJC as a weighted network with link weight of the number of patents that cites two particular journals simultaneously. In such a network representation, the cumulative weight distribution follows a power law while the distribution of strength obeys an exponential distribution. The weight per degree and the relationship between the vertex strength and degree are also analyzed. A linear dependence between strength and degree is observed. We also find that the average weight per degree for NSJC is larger than that of some collaboration networks. In summary, analogously to collaboration networks (authors, actors in movies, corporate board members), the NSJC can be viewed as a special sort of collaboration network. The above collaboration networks generally show apparent small-world behavior while the NSJC seems to have more properties of random networks.

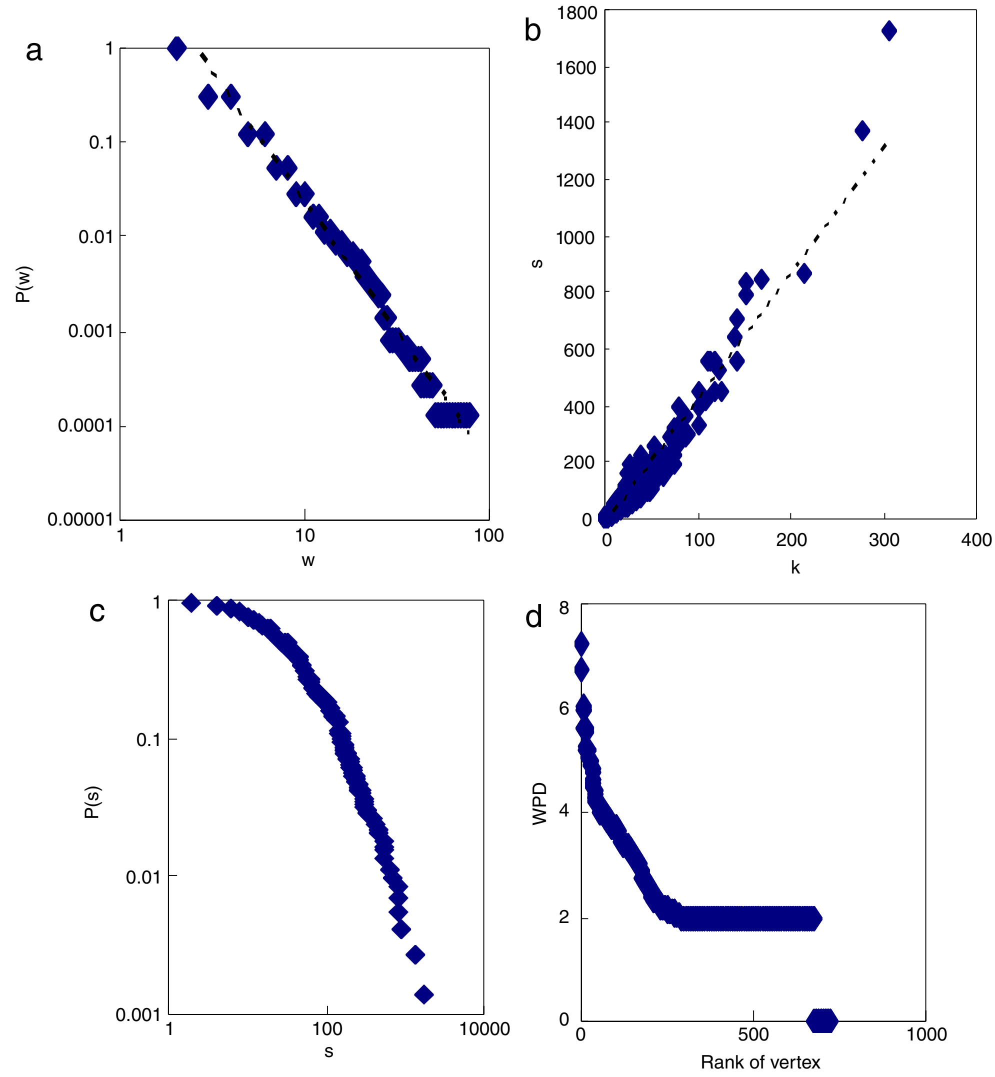

Fig. 5. (a) Log–log plot of the cumulative weight distribution. (b) Log–log plot of the cumulative strength distribution. (c) Strength s as a function of the degree k for the NSJC. The dashed line indicates that s versus k fits to the form $s(k) \sim k$. (d) Linear-scale Zipf plot of weight per degree for the NSJC.

Acknowledgements

This research is funded by the National Natural Science Foundation of China (Project No. 70773006). The authors are grateful for the valuable comments and suggestions of anonymous reviewers and the editors, which significantly improved the article. The authors, naturally, bear responsibility for any remaining shortcomings.

References

[1] J. Stelling, S. Klamt, K. Bettenbrock, S. Schuster, E.D. Gilles, Nature 420 (2002) 190.
[2] R. Albert, H. Jeong, A.-L. Barabási, Nature 401 (1999) 130.
[3] M.E.J. Newman, Proc. Natl. Acad. Sci. 98 (2001) 404.
[4] M.H. Li, Y. Fan, J.W. Chen, L. Gao, Z.R. Di, J.S. Wu, Physica A 350 (2005) 643.
[5] X.P. Xu, J.H. Hu, F. Liu, L.S. Liu, Physica A 374 (2007) 441.
[6] A. Verbeek, K. Debackere, M. Luwel, P. Andries, E. Zimmermann, F. Deleus, Scientometrics 54 (2002) 399.
[7] M. Meyer, Scientometrics 66 (2006) 425.
[8] J.C. Guan, Y. He, Scientometrics 72 (2007) 403.
[9] M. Meyer, Research Policy 29 (2000) 409.
[10] K. Pavitt, L. Soete, in: K. Pavitt (Ed.), Technical Innovation and British Economic Performance, Macmillan, London, 1980.
[11] M.E.J. Newman, Phys. Rev. E 64 (2001) 016131.
[12] X.P. Xu, Y. Li, F. Liu, Physica A 376 (2007) 738.
[13] D.J. Watts, S.H. Strogatz, Nature 393 (1998) 440.
[14] LAN Amaral, A. Scala, M. Barthélémy, et al., Proc. Natl. Acad. Sci. USA 97 (2000) 11149.

[15] A.-L. Barabási, R. Albert, Science 509 (1999) 286.
[16] A.L. Barabási, R. Albert, H. Joeng, Physica A 272 (1999) 173.
[17] A.L. Barabási, Z.N. Oltvai, Nature Reviews Genetics 5 (2004) 101.
[18] J. Ramasco, S.A. Morris, Physical Review E 73 (2006) 016122.
[19] M.E.J. Newman, Phys. Rev. E 89 (2002) 208701.
[20] A. Barrat, M. Barthélémy, A. Vespignani, Phys. Rev. Lett. 92 (2004) 228701.
[21] A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, Proc. Natl. Acad. Sci. USA 101 (2004) 3747.
[22] Y. Fan, M. Li, J. Chen, L. Gao, Z. Di, J. Wu, International J. Mod. Phys. B 18 (2004) 2505.