TECHNOLOGICAL AND ECONOMIC DEVELOPMENT OF ECONOMY
ISSN 2029-4913 print/ISSN 2029-4921 online

2014 Volume 20(1): 45–64
doi:10.3846/20294913.2014.825460

FEATURES OF THE DISCIPLINE KNOWLEDGE NETWORK: EVIDENCE FROM CHINA

Wei SHAN[a], Chen LIU[b], Jing YU[c]

[a]*School of Economics and Management, Beijing University of Aeronautics and Astronautics, Xueyuan Road 37, 100191 Beijing, China*
[b]*Business School, University of Shanghai for Science and Technology, Jungong Road 516, 200093 Shanghai, China*
[c]*Department of Political Science, East China Normal University, Dongchuan Road 500, 200241 Shanghai, China*

Received 04 February 2012; accepted 17 June 2012

Abstract. Interdisciplinary knowledge exchange constitutes a network with discipline nodes and knowledge flow edges. Using data on Chinese academic literature, the current paper establishes a discipline knowledge network and analyses its structural features. Citation analysis is first used to measure the flow of knowledge between disciplines to build a discipline knowledge network. Subsequently, the features of the network, such as degree distribution, degree correlation, knowledge flow mode and other structure properties, are then analysed based on complex networks and social network theory. The tail of the degree distribution of this discipline knowledge network is in concordance with exponential distribution. The network has also a distinct hierarchical structure. Moreover, the knowledge flow between disciplines is directional. It flows from certain basic and academic disciplines to the applied disciplines.

Keywords: knowledge network, discipline knowledge network, complex network, citation analysis.

Reference to this paper should be made as follows: Shan, W.; Liu, C.; Yu, J. 2014. Features of the discipline knowledge network: evidence from China, *Technological and Economic Development of Economy* 20(1): 45–64.

JEL Classification: A12, C15, C80, I21.

Introduction

Modern science has been divided into different categories of tiny disciplines, and scientists have always been limited in certain areas. But nowadays, they have to handle knowledge

Corresponding author Wei Shan
E-mail: *shanwei2000@gmail.com*

form multiple disciplines to solve complex problems. These cause a reverse process, and interdisciplinary cooperation is becoming more and more common in the areas of science and technology (Klein 2008). Therefore, theory and practice of interdisciplinary collaboration has been frequently studied (Klein 2006, 2008; Yang *et al.* 2010). Research becomes interdisciplinary when it involves several fields (Huutoniemi *et al.* 2010). Furthermore, interdisciplinary collaboration inevitably results in knowledge flows between researchers or between research fields, which compose the knowledge network. Although there are many scholarly works on interdisciplinary collaboration, attention has been drawn on the structure and dynamics of the discipline knowledge network, especially in China.

With the rise of the Chinese economy, research papers published by Chinese researchers ranked second only to the US in 2006 (Zhou, Leydesdorff 2008). This achievement is inseparably connected with China's reform and opening-up policy. However, compared with its economy, the pace of reforms in China's educational system lags far behind. China's educational system is significantly influenced by the former Soviet Union, wherein the division of disciplines and professions is subject to strict supervision. This division makes Chinese researchers more likely to be limited to a fixed field compared with their western counterparts. Nevertheless, the flow of knowledge between disciplines is inevitable.

The interdisciplinary flow of knowledge forms a unique network system that takes subjects as nodes and knowledge flow between disciplines as connections. Citation analysis theory and social and complex network analysis provide the possibility and the specific methods to analyse the network.

Citation analysis originates from the landmark study of Dr Garfield (1955) and the establishment of the Science Citation Index (SCI). The SCI is often used to evaluate researchers, research institutions, academic papers, and journals according to a variety of indicators, or to follow the developments in a research field. Price, who was honoured as the "father of scientometrics", creatively made a diagram of a network of scientific papers based on the cite-and-been-cited relations of scientific papers, and studied in-degree and out-degree distribution (Price 1965). Preferential attachment in scientific co-authorship networks is different for authors with different forms of centrality (Abbasi *et al.* 2012). Large scale databases, such as SCI, enable citation networks to be used in research in different fields, research statuses and trends in different countries and regions (Uzun 1996; Kim 2001; Leydesdorff, Zhou 2005). Structural indices in an ego citation network are introduced to describe ego article citation networks in a graph-theoretic setting (Hu *et al.* 2012). However, these studies on knowledge interaction are based on citation analysis either emphasis on certain research field (Bassecoulard *et al.* 2007; Yu *et al.* 2010; Ortega, Aguillo 2010), a specific journal (Ronda-Pupo, Guerras-Martín 2010), or a certain research organization (Tomassini, Luthi 2007). Besides, most of these studies take researchers as nodes of the network (Haythornthwaite 2005; Sorenson *et al.* 2006; Fiala 2012). Discipline or research area is seldom considered as the study element.

As a special network structure, citation networks have an inseparable connection with the social network and complex network theories. Social network uses graph theory to study the complex social structure formed by the social interaction between members. Its representative theories include the strength of weak ties (Granovetter 1973) and the structural hole theory (Burt 1995). These two well-known network theories were used to identify characteristic elements of network theorizing (Borgatti, Halgin 2011). Another method in network research is based on the random graph theory (Erdős, Rényi 1960). With the rapid increase in efficiency in computer data processing, large-scale networks can now be handled. Special features of complex network, such as small world (Watts, Strogatz 1998) and scale-free (Barabási, Albert 1999), are being studied intensively. The effect of three topological characteristics, clustering, modularity and degree correlations, have been studied (Pósfai *et al.* 2013). Citation networks have also been found to have the characteristics of complex networks (Newman 2001a, b) and that they have a power law distribution with an index of about 3 (Redner 1998). At present, information propagation of online social networks comes into the notice of network researchers (Campbell, Kwak 2010; Kumar *et al.* 2010; Bakshy *et al.* 2012).

Science citations and cited documents tend to have links on the subject matter, which represents journals of different disciplines cited interdisciplinarily (Leydesdorff 2004; Narin *et al.* 1972). That is to say, citation networks include information related to cross and pervasion between disciplines. It can be used to analyse the development profile, ground-breaking achievements, mutual penetration, and direction of future development of various disciplines to reveal the overall structure of disciplinary development. Therefore, the present paper establishes the discipline knowledge network in China and studies that show how disciplines connect to each other. Then, this paper examines the role of each subject and its status in the network. Moreover, the characteristics and relationship of knowledge flow between disciplines in the discipline knowledge network are analysed. To be more precise about the network of subject knowledge in China, the present paper divides the disciplines following the Chinese education sector and the data from the databases of Chinese scientific papers. The methods of analyses used are social network analysis and complex networks analysis.

The first part is introduction. The second part summarizes the important literature on the emergence and the development of citation network, social network, and complex network. The second part also states the purpose of this study, the research methods, and the data sources. The methodology introduces the division of disciplines in China, data collection and processing, and principal methods used in this study. Subsequently, the results and discussion on the features and characteristics of knowledge flow in discipline knowledge network in China are presented. The last part is conclusion.

1. Methodology

The current study establishes the discipline knowledge network in China based on China's discipline division and the relationship of literature citation between different disciplines. To accomplish this, network analysis is used to study the structural characteristics of the Chinese scientific research system. The network nodes of subject knowledge are the disciplines. The relationship between different nodes is established through interdisciplinary citation. We use alternative methods, given that gathering all the citation relationship in the vast academic literature is unnecessary and impossible, and that accurately determining the membership of each subject literature is a contentious issue. Each discipline has representative authoritative journals; hence, by using the citation relationship among these journals, we establish an alternative network of discipline knowledge. The citation between these authoritative journals can sufficiently reflect the citation relationship between their respective disciplines.

1.1. Disciplines in China

The division of disciplines in the educational and research system in China is significantly influenced by the former Soviet Union. Compared with Europe and the US, China has a centralized administrative directive nature and emphasizes disciplines rather than professions. The disciplinary system in China is composed of higher education sector and basic research sector, where higher education includes two division systems: undergraduate education system and postgraduate education system. The former is marked by the "College Undergraduate Course Catalog", the goal of which is to cultivate personnel with basic theoretical knowledge. The latter is marked by the "Course Catalog of Awarding Doctor's Degree and Master's Degree and Educating Graduate Students", the goal of which is to train high-level personnel to conduct basic disciplinary research. Basic research is governed by the National Natural Science Foundation of China (NSFC) as regards the division of disciplines. Among the divisions, the college undergraduate course catalogue is mainly for university undergraduate programs. The division of NSFC is related to the application of a national natural science foundation. The most influential and most closely related to the scientific research division is the "specialty catalog of degree conferment and educating graduate students" issued by the Academic Degree Committee of China's State Council in 1997. The present study intends to establish interdisciplinary knowledge network based on that catalogue.

Although this method has many drawbacks and is subject to much criticism from those in the education and research sectors, this somewhat rigid division method and system make the boundary between disciplines more clear cut. Moreover, they provide a more reliable classification of subject for this research.

This catalogue includes 12 branches of subjects, 88 first-level disciplines, and 382 second-level disciplines. The present study focuses on the first-level disciplines, which is similar to the Classification of Instructional Programs in US.

1.2. Data collection

The discipline knowledge network of this paper refers to first-level disciplines as nodes. Military science is a special field of study; hence, the important results are not published in academic journals. Moreover, for the sake of confidentiality, this field is closed to some degree; thus, its citation relationship cannot reflect the flow of knowledge in this field. For this reason, the category of military science is taken as a single node. There are 81 nodes in the discipline knowledge network. We select two or three authoritative academic journals for each subject to gather data on the citation relationship between different disciplines. The choice of authoritative journals mainly refers to the national first-level journals category identified by the Office of the State Council Academic Degree Committee and "A Guide to the Core Journals of China (Zhu *et al.* 2008)". The entire discipline knowledge network is based on 198 magazines belonging to 81 subjects. Some important comprehensive Chinese journals, such as Chinese Science Bulletin and

Progress in Natural Science and Social Science in China, are not included. The reason is that each network node is a discipline, but these journals cannot be classified into a specific discipline. Thus, they cannot accurately reflect the knowledge flow relationship between different disciplines.

Literature reference data come from China National Knowledge Infrastructure (CNKI) from 1999 to 2008. CNKI is a full-text database of Chinese literature from which we can refer to the citation relationship between literature and journals. The result of the data statistics is an 81×81 matrix:

$$G=\begin{bmatrix} g_{1,1} & g_{1,2} & \cdots & g_{1,81} \\ g_{2,1} & g_{2,2} & \cdots & g_{2,81} \\ \cdots & \cdots & \cdots & \cdots \\ g_{81,1} & g_{81,2} & \cdots & g_{81,81} \end{bmatrix}, \quad (1)$$

where $g_{i,j}(i,j=1,2,\cdots 81)$ is the citation quantity of the i^{th} discipline cited from the j^{th} discipline.

Given that the matrix and its adjacent network have a one to one relationship, we do not distinguish them. For example, in proper circumstances, matrix *G* can be referred to as network *G*.

Network *G* consists of *N* and *E*, that is, $G=(N,\Phi)$. $N=\{n_1,n_2,n_3,\cdots,n_N\}$ is the collection of nodes in the network. $E=\{e_{i,j}\,|\,i,j=1,2,\cdots,N\}$, where $e_{i,j}$ is an orderly relationship formed by n_i and n_j (i.e. the direct edge between n_i and n_j), and the weight is $g_{i,j}$. The degree of a node $n_i(i=1,2,\cdots,N)$ is k_i, which is the number of edges connected to the node. In a direct network, the degree of a node can be divided into in-degree and out-degree. In-degree k_i^{in} is the quantity of edge $e_{j,i}$ that points to the node, whereas out-degree k_i^{out} is the quantity of edge $e_{i,j}$ that starts from the node. In discipline knowledge network, the in-degree k_i^{in} of node *i* means that the number of disciplines citing discipline *i* is k_i^{in} and that it is related to knowledge outflow. Conversely, out-degree k_i^{out} means that discipline *i* cites another k_i^{out} discipline and it has a knowledge inflow relationship with k_i^{out} disciplines.

1.3. Data processing

Matrix *G* is the adjacency matrix of the discipline knowledge network. However, it cannot be used directly in the analysis of the features of discipline knowledge network in China due to the following problems:

(a) The number of selected journals for each discipline is different. Moreover, each journal contains different number of academic papers. This difference in the number of journals and academic papers makes the citation relationship between disciplines incomparable;

(b) Some occasional citations exist. These citations do not indicate the exchange of knowledge between the two disciplines. These relationships may also interfere with the real structure of subject knowledge, especially in analysing the structure without considering network weight.

We can solve problem (a) by standardizing the number of citations. The main diagonal elements of matrix *G* are the self-citations of academic papers within the discipline. Usually, it is the maximum element of each row or column in the matrix. Thus, the largest exchange and flow of knowledge occurs inside the discipline, which is logical. This occurrence proves that certain structural features do exist between disciplines. The elements of each row of matrix *G* are divided by the diagonal elements of the line, i.e.:

$$W = \left[w_{ij} \right] = \left[{g_{i,j}}/{g_{i,i}} \right]. \tag{2}$$

In this way, the elements in *G* are standardized. The elements in matrix *W* indicate the strength of citation of one discipline from other disciplines. This eliminates the influence of the number of academic journals and documentations, making the citation relationship between different disciplines comparable.

Nevertheless, standardizing the number of citations is not simple. For instance, the citation in *Applied Economics* from *Theoretical Economics* exceeds its self-citation (the element in *W* is greater than 1). This is also logical, given that *Theoretical Economics* and *Applied Economics* are inseparable and that the literature in *Applied Economics* is often cited from *Theoretical Economics*. This is related to the division of economic disciplines by the education and scientific research departments in China. Some scholars questioned this division of economic disciplines in China (Fu 2008). There is only one particular element in the whole matrix. Thus, we adopt a method that is somewhat arbitrary but does not affect the following analysis, i.e. by making it equal to 1. Hence, in the matrix *W*, elements $w_{i,j} \le 1\,(i, j = 1,2,\cdots 81)$ are the intensity of flow of knowledge from discipline *j* to discipline *i*. The main diagonal element is 1, indicating that the intensity of flow of knowledge within the discipline is 1. Matrix *W* is an adjacency matrix that reflects the network of knowledge flow between disciplines. The weight of the network is the flow intensity of knowledge.

Some smaller elements in matrix *G* exist. These elements can be neglected unlike the citation quantity within the discipline. Compared with most other elements, the differences are relatively large. These smaller elements imply that the citations of relative disciplines have been few in 10 years. Thus, we can consider these citations as incidental citations. Incidental citation is simply the citing of literature of one discipline from another literature of another discipline. However, this form of citation does not mean that there is knowledge exchange between the two disciplines. Moreover, these incidental citations are the only few non-zero elements in *G*.

To eliminate incidental citation, a critical value γ is set in matrix *W*; all elements less than γ are classified as incidental citations. When testing the numeral value in $0.01 \le \gamma \le 0.05$, we find that $\gamma = 0.02$ is a proper critical value, which can effectively eliminate incidental citation.

In this way, problems (a) and (b) are solved. The adjacency network of the new matrix *W* that removes incidental citation is the discipline knowledge network, as shown in Fig. 1. The discipline knowledge network is a connected network that includes 81 nodes and 1744 edges.

W is a direct network, but some network analysis methods require it to be indirect. Thus, the symmetrical treatment of the network is required. There are many ways to apply the symmetrical treatment to network analysis, including *Maximum*, *Minimum*, *Average*, etc. We use averaging in the present study, making the new symmetric network adjacent to the matrix:

$$S = \left[s_{ij} \right] = \left[{w_{ij} + w_{ji}}/{2} \right]. \tag{3}$$

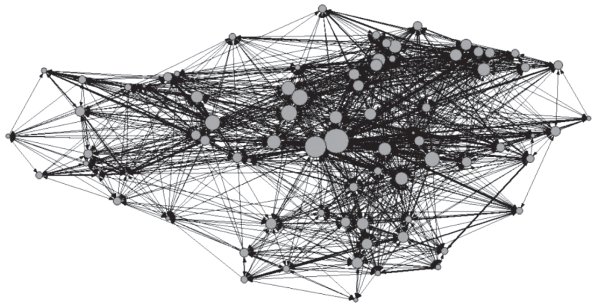

Fig. 1. Discipline knowledge network in China

Although more mature network analysis methods can be used to analyse the network after symmetrical treatment, the symmetrical treatment is an irreversible process. Thus, some information in the network may be lost. This paper uses multi-methods to analyse the direct network W and the indirect network S after symmetrisation.

1.4. Methods

The network analysis method is used to analyse discipline knowledge network in China. It includes three parts: descriptive characteristics analysis, assortative analysis, and structural analysis.

Descriptive characteristics analysis describes the basic features of the discipline knowledge network in China, including its density, average degree, average shortest path, diameter, degree distribution of network nodes, and the betweenness of network nodes.

Assortative analysis examines the degree correlation of network nodes. Based on the direction of knowledge flow in the network, this analysis divides the disciplines represented by nodes in the network into three types: upstream disciplines, downstream disciplines, and intermediate disciplines.

Structural analysis, beginning from the clustering coefficient of the network, investigates the structural features of the network, such as its hierarchy and cyclic topology.

2. Results and discussion

2.1. Descriptive characteristics

(a) Density and average degree

Network density and average degree are the indicators used to measure the number of connections between nodes in the network. Network density m is the ratio of the number of edges in the network and the number of possible edges. The density of direct network is:

$$m_{direct} = \frac{|E|}{N(N-1)}. \tag{4}$$

The average degree of network $<k>$ is the mean value of the degree of all nodes in the network:

$$<k> = \frac{1}{N}\sum_{i=1}^{N} k_i \ . \tag{5}$$

Direct network has the same average in-degree and out-degree. Hence, this value is indiscriminately called average degree of direct network.

The density of network W is $m_W = 0.269$. Network S is obtained by the symmetrical treatment of W. We adopt the average method; thus, the one-way connection and two-way connection between nodes are all considered unidirectional edges, enlarging the density of network S ($m_S = 0.376$). The average degree of W is $< k_W > = 21.531$. For the same reason, the average degree of network S becomes larger ($< k_S > = 30.074$). The larger nodes in W and S are shown in Table 1. Compare to most researched networks in Table 2 (Albert, Barabási 2002), W has a small size and great density.

Table 1. Some of the largest nodes in networks W and S

W	In Degree	Out Degree	S	Degree
Environmental Science and Engineering	28	62	Physics	71
Physics	14	71	Environmental Science and Engineering	63
Agricultural Engineering	53	32	Agricultural Engineering	57
Management Science and Engineering	34	45	Management Science and Engineering	52
System Science	32	38	Forestry Engineering	48

Table 2. Features of the network that have been studied

Networks	Size	Average Degree	Average Shortest Path Length	Clustering Coefficients
WWW (site level)	153,127	35.21	3.10	0.18
Internet (domain)	3,015–6,029	3.52–4.11	3.70–3.76	0.18–0.30
Movie actors	225,226	61	3.65	0.79
Words, synonyms	22,311	13.48	4.50	0.70
Power grid	4,941	2.67	18.70	0.08

(b) Average shortest path length and diameter

In unweight networks, the distance between node i and node j is the number of edges of the shortest paths between them, which is denoted as t_{ij}. The weight of the weighted network is divided into dissimilarity weight and similarity weight. Assume that node i is connected to node j through node k (in a dissimilarity weight network). The distance between i and j is $t_{ij}^{s} = w_{ik} + w_{kj}$. Similarity weight network uses harmonic mean $t_{ij}^{d} = w_{ik}w_{kj} / (w_{ik} + w_{kj})$. In the discipline knowledge network, the greater the quantity of citation, the more likely that knowledge flows between them. Thus, the similarity weight network is adopted:

$$t_{ij}^{d} = 1 \Big/ \sum_{w_p \in T_{ij}} \frac{1}{w_p}, \tag{6}$$

where T_{ij} is the collection of edges of the shortest paths between node i and node j.

The shortest path of the network plays an important role in the dissemination of internal material and information as well as provides the highest efficiency and lowest cost. The average shortest path of the network is the average value of the nearest distance of all nodes pair, which is denoted as *l*.

The diameter of the network d is the longest length of all the shortest paths, i.e. $d = \max l_{ij}$. In unweight networks, $d = \max l_{ij}$ means starting from a node to reach any node through most d steps. In weighted networks, it means starting from a node to reach any node in that network through the farthest d. Hence, the number of nodes a weighted network goes through may not be the least, but the cost is minimal.

Without considering the weights of the edges of the network, the average shortest path of network *W* is 1.872, with a diameter of 4. This means that in the discipline knowledge network, nodes go 1.872 steps on average; only then can the two nodes meet. Starting from a node, nodes go 4 steps at most to reach another node. Considering the weights of the edges of the network, by using a similarity weight calculation, the average shortest path of network *W* is 0.029, with a diameter of 1.000. This average shortest path can be regarded as the average similarity degree between disciplines or the intensity of knowledge dissemination. Diameter is the proximity of two least close disciplines. The average shortest path of network *S* is 1.63, with a diameter of 3.

(c) Degree distribution

The degree can measure the importance of a node to a certain extent. As more nodes are connected to it, the greater is its effect on the network. The degree distribution of network $P(k)$ means randomly selecting a node in the network, with its degree being the probability of k. For the direct network, $P(k^{in})$ and $P(k^{in})$ (i.e. two kinds of distribution) are considered. Degree distribution can also be represented by the function of cumulative degree distribution (Newman 2003):

$$P_k = \sum_{k'=k}^{\infty} P(k'). \tag{7}$$

The equation implies that the probability distribution of a degree is no less than k. If the degree distribution is a power law distribution, i.e. $P(k) \sim k^{-\gamma}$, the cumulative degree distribution, therefore, is in accordance with the power law distribution with an exponent $\gamma - 1$. If $P(k)$ is an exponential distribution, P_k thus have an exponential distribution with same exponent. Power law distribution is a line in the double logarithmic coordinates, whereas exponential distribution is a line in the semi-logarithmic coordinates.

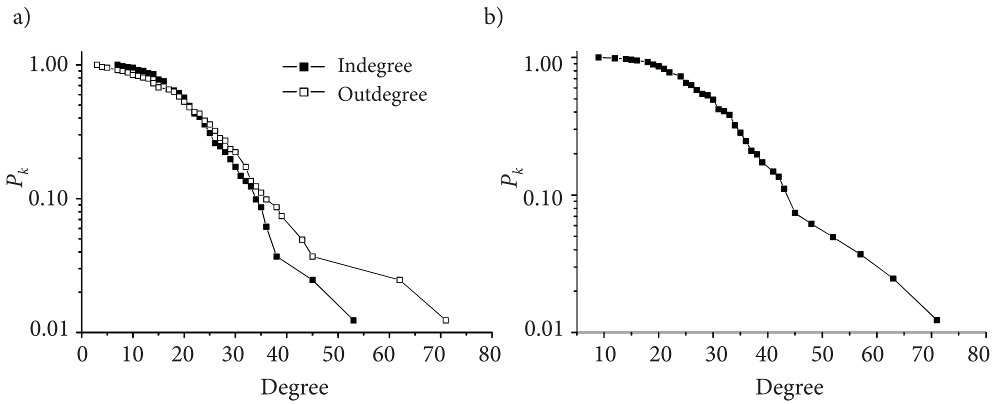

Fig. 2. Degree distribution (a) and direct network; (b) of a symmetrical network in the discipline knowledge network in China

Fig. 2 shows that the in-degree and out-degree of network W and the tail of the cumulative degree distribution of S in the semi-logarithmic coordinates have a nearly straight line. Thus, they are exponentially distributed. Regression results show that the in-degree distribution of network W is $P_k^{in} \propto e^{-\frac{k}{8.319}}$ ($R^2 = 0.996$), the out-degree distribution is $P_k^{in} \propto e^{-\frac{k}{10.432}}$ ($R^2 = 0.992$), and the degree distribution of network S is $P_k^{in} \propto e^{-\frac{k}{13.920}}$ ($R^2 = 0.973$). Compared with the other networks, discipline knowledge network does not have the characteristics of power law distribution caused by its formation mechanism. Barabási and Albert (1999) observed that the power law degree distribution network is built on the basis of two mechanisms: growth and priority connection. The formation of discipline knowledge network does not have these features. Although there is also a large number of nodes with a small degree and a small number of nodes with a large degree in the exponential degree distribution, the distribution is relatively homogeneous compared with the power law degree distribution.

(d) Betweenness centrality

Disciplines also assume the function of the flow of knowledge intermediaries. This function can be measured by the betweenness of network nodes. In a network, the shortest path has a special significance to the dissemination of information and materials in networks. The transformation of a node in the shortest path between node i and j may lengthen the distance between two nodes. The number of shortest paths that go through the nodes determines the ability of the node to act as an intermediary. The betweenness of node i is the number of shortest paths that go through the node. Given that there are multiple shortest paths between some nodes, only a part of the paths goes through i; hence, the betweenness of that node is defined as:

$$b_i = \sum_{j,k=1, j \neq k}^{N} \frac{n_{jk}(i)}{n_{jk}}, \tag{8}$$

where n_{jk} is the number of shortest paths linking *j* and *k*, and $n_{jk}(i)$ is the number of shortest paths linking *j* and *k* through node *i*.

The node with relatively large betweenness plays an important role in the spread of knowledge in networks. If that node is lost, all the shortest paths that go through that node may change. For the nodes with multiple paths, losing that node means losing a shortcut to transfer knowledge. However, for a node that has only one path going through it, the transfer of knowledge needs to go through more steps. The average betweenness of nodes in network *W* is 69.765, whereas the average betweenness of nodes in network *S* is 25.235. The nodes with larger betweenness are shown in Table 3.

Table 3. Nodes with the largest betweenness in discipline knowledge network

W	**Betweenness**	*S*	**Betweenness**
Environmental Science and Engineering	322.384	Physics	243.074
Management Science and Engineering	294.916	Environmental Science and Engineering	171.973
Agricultural Engineering	286.510	Management Science and Engineering	98.689
Biomedical Engineering	237.389	Agricultural Engineering	93.095
Geography	197.211	Biomedical Engineering	74.735

2.2. Assortative characteristics

(a) Degree correlation

The degree distribution of a network completely determines the statistical properties of non-correlated networks (Boccaletti *et al.* 2006). Most networks are correlated. That is, nodes with large degree tend to link to other nodes with large degree (called assortative), or nodes with large degree tend to link to nodes with small degree (called disassortative). According to Newman, social networks are often assortative, whereas technical networks and biological networks are disassortative (Newman 2002). The quantitative indicators used to judge network correlation were proposed by Newman, who defined a Pearson correlation coefficient (Newman 2002) to judge network correlation.

$$r = \frac{M^{-1}\sum_i j_i k_i - \left[M^{-1}\sum_i \frac{1}{2}(j_i + k_i)\right]^2}{M^{-1}\sum_i \frac{1}{2}(j_i^2 + k_i^2) - \left[M^{-1}\sum_i \frac{1}{2}(j_i + k_i)\right]^2}, \tag{9}$$

where *M* is the number of network edges, and j_i and k_i are degree of the nodes that link to the *i*th edge ($-1 \le r \le 1$). When $r > 0$, the network is assortative. This means that the nodes tend to link to other nodes with similar degree. When $r < 0$, the network is disassortative. This means that the nodes with large degree tend to link to nodes with small degree. The Pearson correlation coefficient of network *S* is –0.036, which indicates that it has a non-significant degree correlation. Another intuitive approach to measure the degree correlation of a network is to use the correlation figure (Pastor-Satorras *et al.* 2001) of a node degree and its neighbour's average degree (Fig. 3). Fig. 3(a) also shows that the degree of nodes in the network does not have a non-significant correlation.

In direct networks, the correlation between nodes is far more complex. Some factors that must be considered include whether there is correlation between the in-degree and out-degree, and whether there is a correlation between the in-degree/out-degree and the in-degree/out-degree of their neighbours. As is shown in Fig. 3(b), there is no significant correlation between the in-degree and out-degree in the discipline knowledge network.

A node in direct network has two kinds of neighbours: out-neighbour and in-neighbour. For the node n_i, if there is a direct edge e_{ij} pointing to node n_j, then n_j is the out-neighbour of n_i. In the discipline knowledge network, it means the literature of discipline n_i cited the literature of discipline n_j. Conversely, if there is a node n_j pointing to n_i through edge e_{ji}, then n_j is the in-neighbour of n_i. The correlation between in-degree and average in-degree of out-neighbour and average out-degree of in-neighbour are shown in Fig. 3(c). Their correlations with out-degree are shown in Fig. 3(d). These figures show that the average out-degree

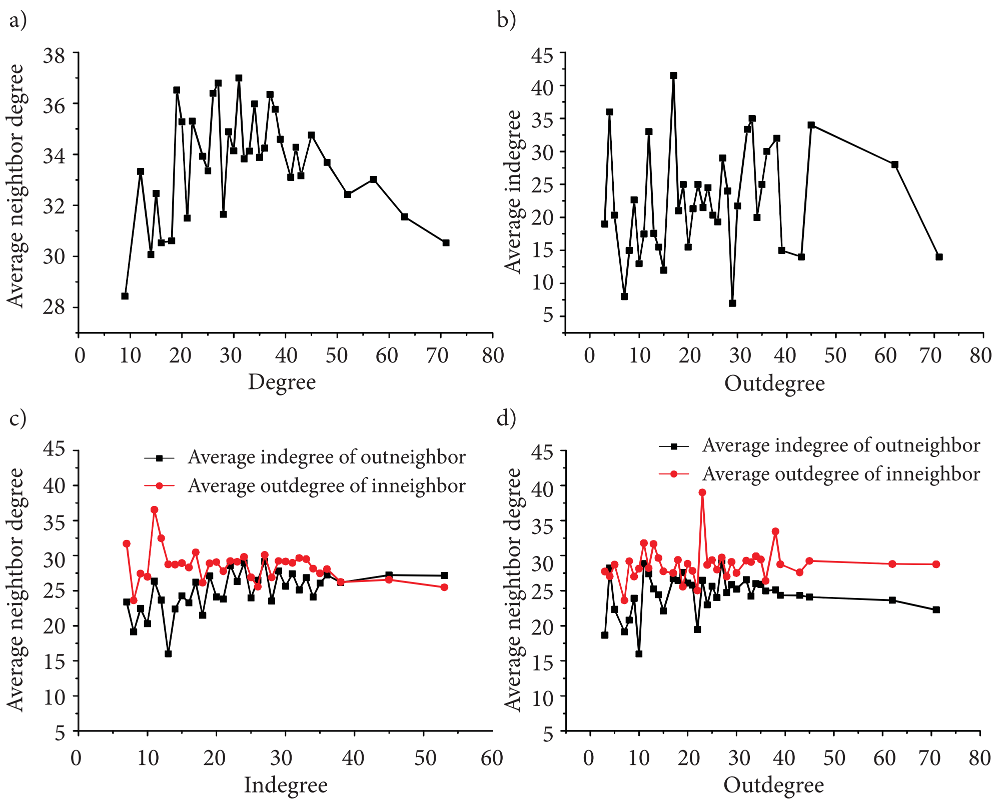

Fig. 3. Degree of correlation of the discipline knowledge network. (a) degree correlation of network *S*; (b) correlation of in-degree and out-degree in network *W*; (c) correlation of the in-degree of a node and its in-neighbour's average out-degree in network *W*; (d) and correlation between the out-degree of a node and its out-neighbour's average in-degree in network *W*

and average in-degree of neighbours have an average trend, which hardly changes with the in-degree or out-degree of nodes.

(b) Role of disciplines in knowledge flow

In the discipline knowledge network, if one discipline cites literature from another discipline, there is an inflow of knowledge in that discipline. Otherwise, there is an outflow of knowledge. Although all disciplines in the discipline knowledge network have both inflow and outflow of knowledge, they do not have same roles in the process of knowledge flow. In some disciplines, the outflow of knowledge accounts for a major position, and in other disciplines, the inflow of knowledge presents important status, whereas some disciplines have roughly the same amount of inflow and outflow, which means they assume the role of knowledge transfer. In discipline knowledge network, some disciplines influence others through the dissemination of knowledge. The disciplines that tend to outflow knowledge are situated in the "upstream" of the network. These disciplines are influential and are usually cited by a number of other disciplines. Moreover, these disciplines are less affected by others, including some basic disciplines such as *mathematics* and *physics*. The disciplines that tend to inflow knowledge are situated in the "downstream" of the network, which have little influence. They cite large amounts of knowledge from other disciplines, whereas the amount of information cited from them is small. Discipline knowledge network is a weighted direct network. Hence, the position of nodes in the knowledge flow network can be measured by the ratio g_i of the in-degree and out-degree of node *i*, and the ratio g'_i of the in-weight and out-weight. (Note that the discipline knowledge flow have opposite direction compare to the pointing of edges.)

$$g_i = \frac{k_i^{in}}{k_i^{out}}; \tag{10}$$

$$g'_i = \frac{l_i^{in}}{l_i^{out}}, \tag{11}$$

where: k_i^{out} is the out-degree of node *i*, k_i^{in} is the in-degree of node *i*; l_i^{out} is the out-weight of node *i*; and l_i^{in} is the in-weight of node *i*. Some disciplines' g_i and g'_i are obtained based on these two formula (shown in Table 4 and Table 5).

Table 4. Ratio of knowledge inflow and outflow of some nodes (1)

Node	g_i	Node	g_i
Physics	5.071	Textile Science and Engineering	0.111
Mathematics	4.143	Ethnology	0.143
Metallurgical Engineering	3.071	Military Science	0.150
Chemistry	2.786	Art Theory	0.188
Computer Science and Technology	2.500	Surveying and Mapping	0.208

Table 4 is the ratio of the in-degree and out-degree of nodes. The five largest nodes are on the left column, whereas the five smallest nodes are on the right column.

Table 5. Ratio of knowledge inflow and outflow of some nodes (2)

Node	Out-S/In-s	Node	Out-S/In-S
Physics	5.071	Military Science	0.011
Chemistry	4.143	Textile Science and Engineering	0.016
History	3.071	Ethnology	0.033
Theoretical Economics	2.786	Surveying and Mapping	0.039
Clinical Medicine	2.500	Agricultural Resources	0.085

Table 5 is the ratio of the in-weight and out-weight of nodes. The five largest nodes are on the left column, whereas the five smallest nodes are on the right column.

The results show that basic disciplines are in the upstream of the network knowledge flow. The number of citations from other disciplines is very small. Some applied sciences are situated in the downstream of the knowledge flow. The nature of the discipline determines its position in the process of knowledge flow. Thus, different investment policies should be adopted based on different types of disciplines. Basic research on the disciplines in the upstream should be increased, whereas the knowledge absorption and application capacity of the disciplines in the downstream should be enhanced.

2.3. Structural characteristics

(a) Hierarchical structure

Networks in the real world consist of a large number of modular called subgroups. Inside these subgroups, the nodes (or members) of this network are closely linked to each other, with only a few links connected outside the network. This constitutes a network hierarchy, which can be measured by the relationship between node clustering coefficients and degree (Ravasz, Barabási 2003).

The clustering coefficient C_i of node i has multiple definitions. The most intuitive definition is the ratio of all edges of neighbouring nodes and the number of edges that may exist (Albert, Barabási 2002).

$$C_i = \frac{2L_i}{k_i(k_i - 1)}, \tag{12}$$

where L_i is the number of edges between neighbours of node i, and k_i is the number of neighbours of node i. The clustering coefficient of the entire network is the average value of the clustering coefficient of each node, i.e. $C = \sum C_i / N$, where N is the total number of nodes in the network. The clustering coefficient of network *S* is 0.631, which indicates a large aggregation of the network. Considering that network *S* has a smaller average shortest path, the discipline knowledge network has the features of small world network. Table 6 shows the nodes with the largest clustering coefficients in network *S*.

In discipline knowledge network, nodes with large clustering coefficients have small degree (from 9–15) and close connections with neighbouring nodes. However, the nodes with small clustering coefficients have relatively large degree. The work of Ravasz and Barabási (2003) shows that nodes with greater degree always results in smaller clustering coefficient. Possibly,

Table 6. Clustering coefficients of some nodes in network *S*

Node	C	Node	C
Stomatology	0.972	Physics	0.391
Political Science	0.857	Environmental Science and Engineering	0.417
Veterinary Medicine	0.848	Forestry Engineering	0.461
Law	0.810	Management Science and Engineering	0.467
Electrical Engineering	0.800	Agricultural Engineering	0.477

more adjacent nodes have less likelihood of connecting in-between, but the number of existing edges between neighbouring nodes increases sharply. They indicate that in a hierarchical network, the clustering coefficient of nodes is inversely proportional to the degree of nodes, i.e. $C(k) \sim k^{-1}$. Based on this property, actor networks and the Web were studied and found that these networks have obvious hierarchy characteristic. The relationship between clustering coefficient and the degree in the discipline knowledge network is presented in Fig. 4.

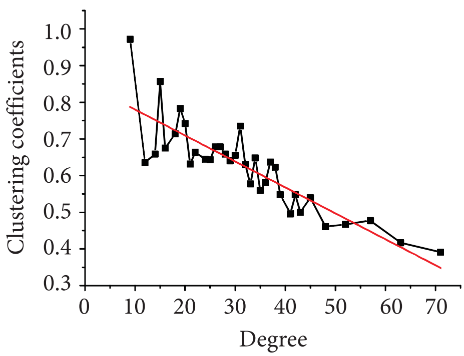

Fig. 4. Relationship between clustering coefficient and degree in network *S*

Fig. 4 shows an obvious linear relationship between degree and clustering coefficient in network *S*. Hence, *S* is a network with obvious hierarchy.

(b) Cyclic structure

Clustering coefficient only considers the circle with three edges, and ignores the influence from nodes that are quite remote. The nodes with the same degree may have significant different clustering coefficients. To measure the relationship between network nodes better, H.-J. Kim and J. M. Kim (2005) provide an indicator to calculate the local cyclic coefficient of network nodes:

$$r_i = \frac{2}{k_i(k_i-1)} \sum_{<lm>} \frac{1}{S^i_{lm}}, \tag{13}$$

where: k_i is the degree of node *i*; $<lm>$ is all the neighbour pairs of node *i*; and S^i_{lm} is the length of the smallest circle that goes through node i and neighbour l and m. The cyclic coefficient of network is $R = \langle r_i \rangle$ (the average value of local cyclic coefficient of all nodes). r_i reaches the maximum (1/3) when node i, l, and m form a triangle. In this case, the network is a complete network, and all pairs of nodes have direct connections. When $R = 0$, there is no loop in the network. In this case, the network is a tree. Therefore, we can get $0 \le R \le 1/3$. The distribution of nodes' local cyclic coefficient in discipline knowledge network S is presented in Fig. 5.

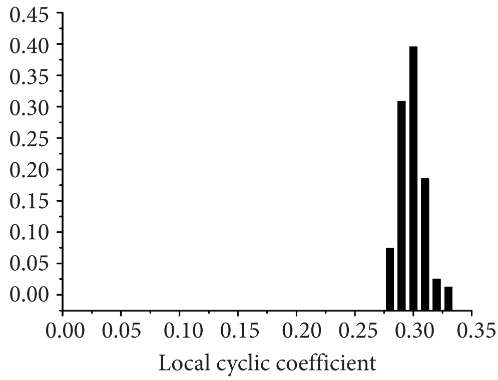

Fig. 5. Local cyclic coefficient distribution of network *S*

In the discipline knowledge network, the local cyclic coefficients of nodes are concentrated in the narrow range of 0.27–0.33. Nodes with local cyclic coefficients are greater than 0.3 account for 60% of all the nodes. The cyclic coefficient of the entire network is 0.306, which is close to 1/3. The cyclic coefficient shows that network *G* is a network with a large number of circles.

Conclusions

This paper considers disciplines and the relationship between them as a network and studies connective characteristics. In this network, disciplines are taken as nodes and the citation relationship between disciplines as edges. Size of this network is small compare to other social networks or complex networks (Albert, Barabási 2002), but it is highly connected. This means that interactive, which is knowledge exchange, between disciplines is more frequently than other networks. Even so, the discipline knowledge network has the ubiquitous network features of small world and heterogeneity. The small average shortest paths and large clustering coefficients imply that it is a small world network. Different form most heterogeneity networks, which have power-law degree distribution, the degree distribution of discipline knowledge network have an exponential distribution tail. This means that although some of the disciplines have a higher connection, there are no super connected nodes like power-law distribution networks. Moreover, the discipline knowledge network has an obvious hierarchy. The large number of loops in the network indicates that the knowledge flows between disciplines are highly cyclical. Another special feature of discipline knowledge network is that the flows on it are directive. It can be measured by comparison of in-degree and out-degree or comparison of in-weight and out-weight. Results indicate that knowledge tends to flow from certain basic subjects or academic disciplines to non-basic applied science.

Discipline knowledge network results in knowledge propagation, and it is a kind of information transmission network. In information transmission networks, information exchange between network nodes is impacted by complex factors like influence, homophily and social contagion (Anagnostopoulos *et al.* 2008; Aral *et al.* 2009; Shalizi, Thomas 2011). This is the basic problem of information transmission networks (Bakshy *et al.* 2012), and discipline knowledge network also has to be studied from this point of view. Moreover, measuring knowledge and flow of knowledge is not an easy task. This makes the establishment and quantitative analysis of knowledge networks relatively difficult. Citation analysis provides a convenient way to establish knowledge network. However, the determination of network weight is still subject to in-depth studies. Discipline knowledge network is evolving. The connection of nodes and the evolution of edge weights need further research. Finally, this paper is based on Chinese literature. Hence, the establishment of a more general subject network still needs further research.

Acknowledgment

The authors are very grateful for the insightful comments and suggestions of the anonymous reviewers and Associate Editor Jonas Šaparauskas, which have helped to significantly improve this article. Furthermore, this research was supported by the National Natural Science Foundation of China (No. 70901023, No. 71371025), the Research Fund for the Doctoral Program of Higher Education of China (No. 20101102120024), the Humanity and Social Science Foundation of Ministry of Education of China (No. 12YJCZH126), and the Beijing Municipal Science and Technology Foundation (Z131100004613018).

References

Abbasi, A.; Hossain, L.; Leydesdorff, L. 2012. Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks, *Journal of Informetrics* 6(3): 403–412. http://dx.doi.org/10.1016/j.joi.2012.01.002

Albert, R.; Barabási, A.-L. 2002. Statistical mechanics of complex networks, *Reviews of Modern Physics* 74: 47–97. http://dx.doi.org/10.1103/RevModPhys.74.47

Anagnostopoulos, A.; Kumar, R.; Mahdian, M. 2008. Influence and correlation in social networks, in *Proceedings of the 14th ACM SIGKDD Internal Conference on Knowledge Discover* & *Data Mining*, New York, USA, 7–15. http://dx.doi.org/10.1145/1401890.1401897

Aral, S.; Muchnik, L.; Sundararajan, A. 2009. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks, *PNAS* 106(51): 21544–21549. http://dx.doi.org/10.1073/pnas.0908800106

Bakshy, E.; Rosenn, I.; Marlow, C.; Adamic, L. 2012. The role of social networks in information diffusion, in the *Proceedings of ACM WWW 2012*, April 16–20, 2012, Lyon, France. 10 p.

Barabási, A.-L.; Albert, R. 1999. Emergence of scaling in random networks, *Science* 286(5439): 509–512. http://dx.doi.org/10.1126/science.286.5439.509

Bassecoulard, E.; Lelu, A.; Zitt, M. 2007. Mapping nanosciences by citation flows: a preliminary analysis, *Scientometrics* 70(3): 859–880. http://dx.doi.org/10.1007/s11192-007-0315-1

Boccaletti, S.; Latora, V.; Moreno, Y.; Chavez, M.; Hwang, D.-U. 2006. Complex networks: structure and dynamics, *Physical Reports* 424(4/5): 175–308. http://dx.doi.org/10.1016/j.physrep.2005.10.009

Borgatti, S. P.; Halgin, D. S. 2011. On network theory, *Organization Science* 22(5): 1168–1181. http://dx.doi.org/10.1287/orsc.1100.0641

Burt, R. S. 1995. *Structural holes: the social structure of competition*. Cambridge: Harvard University Press.

Campbell, S.W.; Kwak, N. 2010. Mobile communication and civic life: linking patterns of use to civic and political engagement, *Journal of Communication* 60(3): 536–555. http://dx.doi.org/10.1111/j.1460-2466.2010.01496.x

Erdős, P.; Rényi, A. 1960. On the evolution of random graphs, *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 5: 17–61.

Fiala, D. 2012. Bibliometric analysis of CiteSeer data for countries, *Information Processing* & *Management* 48(2): 242–253. http://dx.doi.org/10.1016/j.ipm.2011.10.001

Fu, R. M. 2008. Comparative study of economics discipline in Chinese and foreign universities, *China University Teaching* (1): 88–91 (in Chinese).

Garfield, E. 1955. Citation indexes for science: a new dimension in documentation through association of ideas, *Science* 122(3159): 108–111. http://dx.doi.org/10.1126/science.122.3159.108

Granovetter, M. S. 1973. The strength of weak ties, *American Journal of Sociology* 78(6): 1360–1380. http://dx.doi.org/10.1086/225469

Haythornthwaite, C. 2005. Knowledge flow in interdisciplinary teams, in *Proceedings of the 38th Hawaii International Conference on System Sciences,* 3–6 January, 2005, Hawaii, USA. http://dx.doi.org/10.1109/HICSS.2005.372

Hu, X.; Rousseau, R.; Chen, J. 2012. Structural indicators in citation networks, *Scientometrics* 91(2): 451–460. http://dx.doi.org/10.1007/s11192-011-0587-3

Huutoniemi, K.; Klein, J. T.; Bruun, H.; Hukkinen, J. 2010. Analyzing interdisciplinarity: typology and indicators, *Research Policy* 39(1): 79–88. http://dx.doi.org/10.1016/j.respol.2009.09.011

Kim, H.-J.; Kim, J. M. 2005. Cyclic topology in complex networks, *Physical Review E* 72: 036109. http://dx.doi.org/10.1103/PhysRevE.72.036109

Kim, M.-J. 2001. A bibliometric analysis of physics publications in Korea: 1994–1998, *Scientometrics* 50(3): 503–521. http://dx.doi.org/10.1023/A:1010514932626

Klein, J. T. 2006. Afterword: the emergent literature on interdisciplinary and transdisciplinary research evaluation, *Research Evaluation* 15(1): 75–80. http://dx.doi.org/10.3152/147154406781776011

Klein, J. T. 2008. Evaluation of interdisciplinary and transdisciplinary research: a literature review, *American Journal of Preventive Medicine* 35(2s): 116–123. http://dx.doi.org/10.1016/j.amepre.2008.05.010

Kumar, R.; Novak, J.; Tomkins, A. 2010. Structure and evolution of online social networks, in *Book Chapter of Link Mining: Models, Algorithms, and Applications*, Part 4, 337–357.

Leydesdorff, L. 2004. Clusters and maps of science journals based on bi-connected graphs in the Journal Citation Reports, *Journal of Documentation* 60(4): 371–427. http://dx.doi.org/10.1108/00220410410548144

Leydesdorff, L.; Zhou, P. 2005. Are the contributions of China and Korea upsetting the world system of science?, *Scientometrics* 63(3): 617–630. http://dx.doi.org/10.1007/s11192-005-0231-1

Narin, F.; Carpenter, M.; Berlt, N. C. 1972. Interrelationships of scientific journals, *Journal of the American Society for Information Science* 23(5): 323–331. http://dx.doi.org/10.1002/asi.4630230508

Newman, M. E. J. 2001a. Scientific collaboration networks. I. Network construction and fundamental results, *Physical Review E* 64: 016131. http://dx.doi.org/10.1103/PhysRevE.64.016132

Newman, M. E. J. 2001b. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality, *Physical Review E* 64: 016132. http://dx.doi.org/10.1103/PhysRevE.64.016132

Newman, M. E. J. 2002. Assortative mixing in networks, *Phsical Review Letters* 89(20): 208701. http://dx.doi.org/10.1103/PhysRevLett.89.208701

Newman, M. E. J. 2003. The structure and function of complex networks, *SIMA Review* 45(2): 167–256. http://dx.doi.org/10.1137/S003614450342480

Ortega, J. L.; Aguillo, I. F. 2010. Shaping the European research collaboration in the 6th Framework Programme health thematic area through network analysis, *Scientometrics* 85(1): 377–386. http://dx.doi.org/10.1007/s11192-010-0218-4

Pastor-Satorras, R.; Vázquez, A.; Vespignani, A. 2001. Dynamical and correlation properties of the internet, *Physical Review Letters* 87(25): 258701. http://dx.doi.org/10.1103/PhysRevLett.87.258701

Pósfai, M.; Liu, Y.-Y.; Slotine, J.-J.; Barabási, A.-L. 2013. Effect of correlations on network controllability, *Scientific Reports* 3: 1067. http://dx.doi.org/10.1038/srep01067

Price, D. J. S. 1965. Networks of scientific papers, *Science* 149(3683): 510–515. http://dx.doi.org/10.1126/science.149.3683.510

Ravasz, E.; Barabási, A.-L. 2003. Hierarchical organization in complex networks, *Physical Review E* 67: 026112. http://dx.doi.org/10.1103/PhysRevE.67.026112

Redner, S. 1998. How popular is your paper? An empirical study of the citation distribution, *The European Physical Journal B - Condensed Matter and Complex Systems* 4(2): 131–134. http://dx.doi.org/10.1007/s100510050359

Ronda-Pupo, G. A.; Guerras-Martín, L. Á. 2010. Dynamics of the scientific community network within the strategic management field through the Strategic Management Journal 1980–2009: the role of cooperation, *Scientometrics* 85(3): 821–848. http://dx.doi.org/10.1007/s11192-010-0287-4

Sorenson, O.; Rivkin, J. W.; Fleming, L. 2006. Complexity, networks and knowledge flow, *Research Policy* 35(7): 994–1017. http://dx.doi.org/10.1016/j.respol.2006.05.002

Shalizi, C. R.; Thomas, A. C. 2011. Homophily and contagion are generically confounded in observational social network studies, *Sociological Methods and Research* 40(2): 211–239. http://dx.doi.org/10.1177/0049124111404820

Tomassini, M.; Luthi, L. 2007. Empirical analysis of the evolution of a scientific collaboration network, *Physica A - Statistical Mechanics and its Applications* 385(2): 750–764. http://dx.doi.org/10.1016/j.physa.2007.07.028

Uzun, A. 1996. A bibliometric analysis of physics publications from Middle Eastern countries, *Scientometrics* 36(2): 259–269. http://dx.doi.org/10.1007/BF02017319

Watts, D. J.; Strogatz, S. H. 1998. Collective dynamics of 'small-world' networks, *Nature* 393: 440–442. http://dx.doi.org/10.1038/30918

Yang, C. H.; Park, H. W.; Heo, J. 2010. A network analysis of interdisciplinary research relationships: the Korean government's R&D grant program, *Scientometrics* 83(1): 77–92. http://dx.doi.org/10.1007/s11192-010-0157-0

Yu, G.; Wang, M.-Y.; Yu, D.-R. 2010. Characterizing knowledge diffusion of Nanoscience & Nanotechnology by citation analysis, *Scientometrics* 84(1): 81–97. http://dx.doi.org/10.1007/s11192-009-0090-2

Zhou, P.; Leydesdorff, L. 2008. China ranks second in scientific publications since 2006, *ISSI Newsletter* 4(1): 7–9.

Zhu, Q.; Dai, L. J.; Cai, R. H. 2008. *A guide to the core journals of China*. Beijing: Peking University Press (in Chinese).

Wei SHAN. He is an Associate Professor of Management at School of Economics and Management, Beijing University of Aeronautics and Astronautics, Beijing, China. He received his PhD in Technological Economics and Management from Harbin Institute of Technology. He has published more than 30 papers in journals and conferences. His researches have been sponsored by the National Natural Science Foundation of China and the Research Fund for the Doctoral Program of Higher Education of China. The research results have received several academic honors. His current research interests focus on Knowledge Management and Technological Innovation.

Chen LIU. He is an Assistant Professor at Business School, University of Shanghai for Science and Technology, Shanghai, China. He received his PhD in Management Science and Engineering from Harbin Institute of Technology. He has published more than 10 papers in journals and conferences. He has been sponsored by the National Natural Science Foundation of China and the Humanity and Social Science Foundation of Ministry of Education of China. His current research interests focus on Information Management and Online Social Network.

Jing YU. She is an Assistant Professor at Department of Political Science, East China Normal University, Shanghai, China. She received her PhD in Communication Studies from Fudan University. She has published more than 10 papers in journals and conferences. She has been sponsored by the Humanity and Social Science Foundation of Ministry of Education of China. Her current research interests focus on Information Propagation.