Scientometrics (2010) 83:115–131
DOI 10.1007/s11192-009-0027-9

Mapping library and information science in China: a coauthorship network analysis

Erjia Yan · Ying Ding · Qinghua Zhu

Received: 7 November 2008 / Published online: 19 June 2009

Abstract This paper aims to identify the collaboration pattern and network structure of the coauthorship network of library and information science (LIS) in China. Using data from 18 core source LIS journals in China covering 6 years, we construct the LIS coauthorship network. We analyze the network from both macro and micro perspectives and identify some key features of this network: this network is a small-world network, and follows the scale-free character. In the micro-level, we calculate each author's centrality values and compare them with citation counts. We find that centrality rankings are highly correlated with citation rankings. We also discuss the limitation of current centrality measures for coauthorship network analysis.

Keywords Coauthorship network · Complex network · Scientific collaboration · Library and information science · China

Introduction

Starting from late 1990s, we have witnessed a new movement in network analysis with the focus shifting from the analysis of small graphs to large-scale statistical properties of graphs (Newman 2003). This new approach has been driven by the availability of computers and networks which allow us to gather and analyze large scale data and also by the breakdown of boundaries between disciplines, allowing us to uncover the generic properties of complex networks (Albert and Barabási 2002).

Coauthorship network, an important form of social network, has been intensively studied in this movement. Currently, most articles studied "hard science" coauthorship networks; less focus has been paid to "soft science" coauthorship networks, such as the

E. Yan (✉) · Y. Ding
School of Library and Information Science, Indiana University, Bloomington, USA
e-mail: eyan@indiana.edu

Q. Zhu
Department of Information Management, Nanjing University, Nanjing, China

coauthorship networks of library and information science (LIS). Meanwhile, few researches place their focus on the coauthorship network of LIS in China. China is currently one of the fast developing countries in the world. Recently China has issued several research policies to boom up its national and international research power. Sufficient funding has been provided nationally to encourage large-scale collaborative projects (Jin 1999; Wang et al. 2005). Library and Information Science in China is also experiencing dramatic changes during recent years along with the "open" policy of China and the exponential growth of Chinese economy (Wu and Yuan 1994). The modern research equipments and new Chinese research policies enable more and more Chinese researchers to collaborate with a broader scope. It becomes an interesting research topic to identify the key features of the current coauthorship network of LIS in China.

This paper aims to deploy the network analysis methods to identify the key features of the coauthorship network of LIS in China. We not only analyze this network with macro-level methods which capture the global features of the networks, but also obtain three sets of centrality values from micro-level analysis which illustrates the local features of the individual elements in the networks, and apply these values to impact analysis and discuss the correlation between centrality measures and citation counts.

In the first part of this article, we conduct a literature review on coauthorship network analysis, tracing its origin and current development. We then introduce the data set and method of this research. In the next part, we construct the coauthorship network of LIS in China, and apply both macro-level and micro-level metrics to analyze this network. We find out some structural properties of this network and also discuss the relationship between citation counts and centrality measures. We conclude this article by discussing the limitation of this research and proposing some future researches.

Related work

In 1998, Watts and Strogatz (1998) showed that many real-world networks have a small average shortest path length, but also a clustering coefficient significantly higher than expected by random chance. A year later, Barabási and Albert (1999) found that some nodes had many more connections than others and that the network as a whole had a power-law distribution. In 2002 and 2003, two review articles were published in top physics (Albert and Barabási 2002) and mathematics (Newman 2003) journals, symbolizing the modeling of social network has gained its first-phase fruition.

As for the coauthorship networks that we are investigating in this paper, an early example of this kind is the Erdös Number Project, in which the smallest number of coauthorship links between any individual mathematician and the Hungarian mathematician Erdös are calculated (Castro and Grossman 1999). Newman studied and compared the coauthorship graph of arXiv, Medline, SPIRES, and NCSTRL (Newman 2001a, b) and found a number of network differences between experimental and theoretical disciplines. By mapping the graph containing all relevant publications of members in an international collaboration network COLLNET, Yin et al. (2006) found that this scientific community displays many aspects of a small-world network and is vulnerable to disruption. Using the Science Citation Index (SCI) data for 1990 and 2000, Wagner and Leydesdorff (2003) found that in the 10 years between 1990 and 2000, the global network has expanded to include more nations and it has become more interconnected. Cronin and Shaw (2007) found physical location played an important role in collaboration in their study of Rob

Kling's intellectual impact and influences. A recent paper by Hou et al. (2008) has revealed the coauthorship pattern of *Scientometrics* using the data from SCI. According to Liu et al. (2005), coauthorship analysis has also been applied to various ACM conferences: Information Retrieval (SIGIR), Management of Data (SIGMOD) and Hypertext, as well as mathematics and neuroscience, and information systems.

Methodology

Data collection

We select 18 LIS journals from Chinese Social Science Citation Index (CSSCI: http://cssci.nju.edu.cn/CSSCIlyqk2008.htm) with a time span of 6 year (2002–2007). Parallel to Social Science Citation Index of Institute for Scientific Information, CSSCI's source journals contain the high quality journals published in China. Thus, these 18 journals are capable of revealing the collaboration patterns of China's LIS research.

These source journals are (translated into English based on CSSCI translations): (1) Archives Science Bulletin, (2) Archives Science Study, (3) Document, Information & Knowledge, (4) Information and Documentation Services, (5) Information Science, (6) Information Studies: Theory & Application, (7) Journal of Academic Libraries, (8) Journal of Information, (9) Journal of Library Science in China, (10) Journal of the China Society for Scientific and Technical Information, (11) Library, (12) Library and Information, (13) Library and Information Service, (14) Library Journal, (15) Library Theory and Practice, (16) Library Tribune, (17) Library Work and Study, (18) New Technology of Library and Information Service.

Method

Given that we have established a social network, we can describe its properties on two levels, by macro-level metrics (global graph metrics) and micro-level metrics (individual actor properties). Macro-level metrics seek to describe the global characteristic of a social network as a whole (Liu et al. 2005) with the aim to capture the overall structural features of a network. Commonly used measures are diameter, mean distance, components, clusters, etc. Micro-level metrics relate to the analysis of the individual properties of network actors, for example, actor position, actor status, and distance to others, which informs us about "the differential constraints and opportunities facing individual actors which shape their social behavior" (Yin et al. 2006, p. 1600). It zooms in to capture the features of the individual nodes/actors in a network with the consideration of the topology of the network. Micro-level metric usually refers to centrality, which indicates how central the actor is to the network. Central actors are well connected to other actors (Liu et al. 2005), and metrics of centrality will measure an actor's degree (degree centrality), average distance (closeness centrality), or the degree to which geodesic paths between any pair of actors passes through the actor (betweenness centrality).

Macro-level metrics

Macro-level metrics are useful to identify the global structural features of the network. There are many ways of characterizing the structure of a network. In this study, we will focus five key elements of networks: component, distance, cluster, degree distribution, and error and attack tolerance of the network.

Component In social network analysis, connected graphs are called components. A component of a graph is a subset with the characteristic that there is a path between any node and any other one of this subset (Nooy et al. 2005). A coauthorship network usually consists of many disconnected components, and usually what we focus on is the largest component. Nascimento (Nascimento et al. 2003) reported that the largest component in SIGMOD's coauthorship graph has about 60% of all authors. In the four coauthorship networks studied by Newman (2001b), Medline has the largest component, with 92.6% of all the authors, while NCSTRL has the smallest largest component, containing 57.2% of all authors. After some comparison studies on coauthorship networks, Kretschmer (2004) suggests that the largest components usually have a ratio of more than 40% of all the authors. A same finding has also been presented by Grossman (2002), who finds that this ratio ranges from 40% to more than 80%. The ratio of the largest components is useful in identifying the collective collaboration pattern of a field, and making comparisons across different disciplines.

Mean distance A geodesic is the shortest path between two vertices. The distance from vertex u to vertex v is the length of the geodesic from u to v. As defined formally by Watts and Strogatz (1998), and informally by Milgram (1967), many social networks display structures where most individuals are at very few degrees of distance from one another. According to Yin et al. (2006), short mean distance is a sign allowing authors to share information more rapidly, and thus is a desirable property of coauthorship networks. In this study, we will calculate the mean distance of LIS in China, and compare this figure with those of other disciplines (see Table 1).

Cluster Real networks are also clustered, which indicates that nodes in these networks have a higher chance to be linked than those in random networks. Coauthorship networks are likely to be characterized by local clusters of individuals who are tied to most of the others (Newman 2001a). Clustering coefficient is a standard way to identify how clustered these networks are:

Table 1 Typical macro-level metrics results

Subjects	No. of authors	No. of papers	Largest component (%)	Mean distance	Clustering coefficient	References
SPIRES	56,627	66,652	89	4.0	0.73	Newman (2001a, b)
NCSTRL	11,994	13,169	57	9.7	0.50	Newman (2001a, b)
SIGMOD	2,394	–	60	5.7	0.69	Nascimento et al. (2003)
Biology	1,520,251	2,163,923	92	4.6	0.07	Newman (2004)
Physics	52,909	98,502	85	5.9	0.43	Newman (2004)
Mathematics	253,339	–	82	7.6	0.15	Newman (2004)
Quantum-Hall	381	385	40	3.4	–	Kretschmer (2004)
Digital library	1,567	759	38	6.6	0.89	Liu et al. (2005)
COLLNET	64	223	–	3.0	0.64	Yin et al. (2006)
Info. system	639	1,597	37	–	–	Vidgen et al. (2007)
Sensor network	291	560	96	–	0.33	Rodriguez and Pepe (2008)

$$C = \frac{3 \times \text{number of triangles on the graph}}{\text{number of connected triples of vertices}}$$

Here a "triangle" is a trio of authors, each of whom is connected to both of the others, and a "connected triple" is a single author connected to two others. It is the probability of collaborating if both have collaborated with a third author. If a network has a clustering coefficient of 0.5, it indicated that there is 50% of a chance that two authors both collaborating with a third author would also collaborate with each other. In scientific collaboration networks, it is the tendency for authors to form local clusters with people with whom they share common interests. To the extent that the relations among authors display a tendency towards transitivity, collaboration networks are likely to be characterized by local clusters of individuals who are tied to most of the others (Yin et al. 2006). This definition corresponds to the concept of the "fraction of transitive triples" used in sociology (Wasserman and Faust 1994).

Degree distribution The degree of a vertex is the number of lines incident with it. Vertices with higher degrees play as the hub in the network. These hubs have many more lines than other vertices, and function to connect among nodes that would have otherwise been unable to reach one another. They are crucial to the robustness of the network as well as the transmission of information. p(k) is defined as the fraction of nodes in the network that have degree k. Equivalently, p(k) is the probability that a random chosen node in a network has degree k (Yin et al. 2006). Albert and Barabási (2002) have discovered that power-law distribution is related to their degree distribution in many real-world networks: $p(k) \sim k^{-\gamma}$. Networks with power-law distributions are often referred to as "scale-free" networks. In this study, we will calculate the exponent $-\gamma$ to detect the degree distribution of this network (see Table 3).

Error and attack tolerance Error tolerance indicates that local failures rarely lead to the loss of the global information-carrying ability of the network. It's the ability to transmit information when a fraction of random vertices are deleted. Attack tolerance, on the other hand, means the ability of retain network's transmission when removing of a few nodes that play a vital role in maintaining the network's connectivity. Albert et al. (2000) finds that scale-free networks display an unexpected degree of robustness, the ability of their nodes to communicate being unaffected even by unrealistically high failure rates. This is attributed to the fact that there are some redundancy of ties between local clusters and central hubs. However, error tolerance comes at a high price in that these networks are extremely vulnerable to attacks. Such error tolerance and attack vulnerability are generic properties of social networks (Albert and Barabási 2002). The ability to sustain network ties in collaboration network is important to the cumulative development of deep specialist knowledge (Yin et al. 2006). Two experiments will be conducted to identify the robustness of the LIS coauthorship network in China.

Table 1 shows some typical macro-level metrics results in coauthorship network analysis. The ratio of the largest component of each coauthorship network ranges from 40% to above 90%, which is consistent with Grossman's (2002) findings. The mean distances of these coauthorship range from 3 to 10. Most of the networks' clustering coefficients are at the 0.1 level, whereas Biology coauthorship network only has a value of 0.07. Newman (2001c) concluded that this is the result of top-down organization of biology laboratories, which tends to produce tree-like collaboration networks. Such tree-like networks possess low clustering coefficients.

Micro-level metrics

Macro-level metrics are useful for identifying the overall collaboration pattern of scientific collaboration network. However, for individual actors, they may not be helpful. Micro-level metrics are designed for solving such problems. We apply three centrality measures (degree, closeness, betweenness centrality) to investigate different contribution of actors in the coauthorship network, which are important to understanding power, stratification, ranking, and inequality in social structures (Wasserman and Faust 1994).

Degree centrality Degree centrality is equal to the number of ties (connections) that a vertex has with other vertices. The equation of it is as following where $d(n_i)$ is the degree of n_i:

$$C_D(n_i) = d(n_i)$$

Generally, authors with higher degree or more connections are more central to the structure and tend to have a greater capacity to influence others. Yet, for some authors with high degree, it is because they co-authored with many authors in a single paper, rather than co-authored in many papers.

Closeness centrality A more sophisticated centrality measure is closeness (Freeman 1979) which emphasizes the distance of a vertex to all others in the network by focusing on the geodesic distance from each vertex to all others.

$$\mathrm{C_c}(n_i) = \sum_{i=1}^{N} \frac{1}{d(n_i, n_j)}$$

In above equation, $\mathrm{C_c}(n_i)$ is the closeness centrality, and $d(n_i, n_j)$ is the distance between two vertices in the network. Authors with high closeness centrality possess either of the following characters: first, authors collaborate with authors from diverse subgroups, and thus shortened their distance to each author from each of the subgroups; meanwhile, authors collaborate with authors in the first case would also result in a high closeness centrality. Authors in the former case are the central authors we intend to identify, whereas those in the latter case are the noises we try to avoid.

Betweenness centrality Betweenness centrality is based on the number of shortest paths passing through a vertex. Vertices with a high betweenness play the role of connecting different groups. In the following formula, g_{jik} are all geodesics linking node j and node k which pass through node i; g_{jk} is the geodesic distance between the vertices of j and k.

$$C_B(n_i) = \sum_{j,k \neq i} \frac{g_{jik}}{g_{jk}}$$

In social networks, vertices with high betweenness are "pivot points of knowledge flow in the network" (Yin et al. 2006, p. 1603). For coauthorship networks, vertices with high betweenness connect authors who share similar research interest. Therefore, authors with high betweenness usually engage in research of different fields and thus show interdisciplinarity.

Table 2 summarizes the macro and micro level metrics for coauthorship network analysis discussed in above paragraphs.

Results and analysis

An overview

A simple and widely used coauthorship network model is based on an undirected, binary graph in which each edge represents a coauthorship relationship. If any two authors coauthored an article, an edge without weight is created. In this study, we take such an approach in constructing the coauthorship network of LIS studies in China. The whole coauthorship network contains 11,067 vertices and 14,471 edges (authors without collaboration are excluded). What needs to be pointed out is that there is a chance that a vertex stands for multiple authors due to the fact that different authors may have the same name. Although the probability of it is not high (around 0.01%), yet it has been magnified in degree centrality as we will discuss later.

Macro-level structure analysis

In this study, we will focus four key elements of networks: component, distance, degree distribution, and error and attack tolerance of the network. Similar to observations from previous research in coauthorship networks, the LIS coauthorship network is not a single connected graph. The largest component of the network has 5,408 authors, and the second largest component has 42 authors. The dot graph of the relationship between Lg (Author per Component) and Lg (No. of Component) is shown as Fig. 1.

In this study, the largest component has a ratio of 48.87% of all the authors, which is in consistent with the previous findings. Two factors affect this ratio: (1) the number of the papers under study, the inadequacy of time span or source journals would result in a lower ratio. Our primary research only included the data from 2002 to 2006, and as a result the largest component only has a ratio of 31.01%; (2) the nature the discipline under study, more authors would be involved if it is an experimental research, thus disciplines like biology science and chemistry science would have a bigger size of largest component.

We calculate the distance of the largest component which is 8.8414. The distribution of the mean distance of each author is shown in Fig. 2.

The clustering coefficient for LIS coauthorship network is 0.425, which means that two authors are much more likely to have collaborated if they have a third common collaborator than are two authors chosen at random from the community. We also construct a random network, sharing same number of vertices and edges with LIS coauthorship network. The mean distance is 13.9, which is 36.69% longer than LIS coauthorship network; and cluster coefficient is 0.012, which is only a thirtieth of LIS coauthorship. This finding is consistent with former research done by Albert and Barabási (2002) who find that the cluster coefficient of real-world network is higher than that of random networks. Based on this, we can conclude that the LIS coauthorship in China is a small-world network.

Table 3 shows the degree distribution of all components in the LIS coauthorship network.

The table shows that about half (43.04%) of the authors only collaborate with another author. 80% authors in the network only collaborate with less than three authors. Eight authors collaborate with more than 30 authors. They are hubs of the network. The

Table 2 A summary of macro and micro level metrics

	Measure	Function	Characteristics	Coauthorship network	Limitation
Macro-level	Component	Detecting to what degree a network scatters	Many measures can only be applied to a single component	Useful for comparison across disciplines	Usually only focusing on the largest component
	Distance and cluster	Measuring the density and organization of a network	Estimating small-world effect	Detecting collaboration pattern of fields	Only valid for connected networks
	Degree distribution	Detecting the structure of a network	Estimating power-law distribution and scale-free effect	Stratify authors according to degree	Papers co-authored by too many authors would add noise to it
	Error and attack tolerance	Detecting the robustness of a network	Useful for studying evolving networks	Detecting the robustness of coauthorship networks	May not be comparable with other coauthorship networks
Micro-level	Degree centrality	The most active and most visible actor in the network	Each author's degree centrality constitutes degree distribution	Detecting the most collaborative authors	Disproportionate to papers co-authored by too many authors
	Closeness centrality	Detecting actors that are closest to all others in the network	The reciprocal of each author's mean distance	Detecting authors with extensive collaborating scope	Indirect to academic performance
	Betweenness centrality	Detecting actors who lies on a great number of shortest paths in the network	Resembles to weighted degree centrality, which takes distance into account	Detecting the brokers and connectors	Authors involved in interdisciplinary research would result in an immoderate value

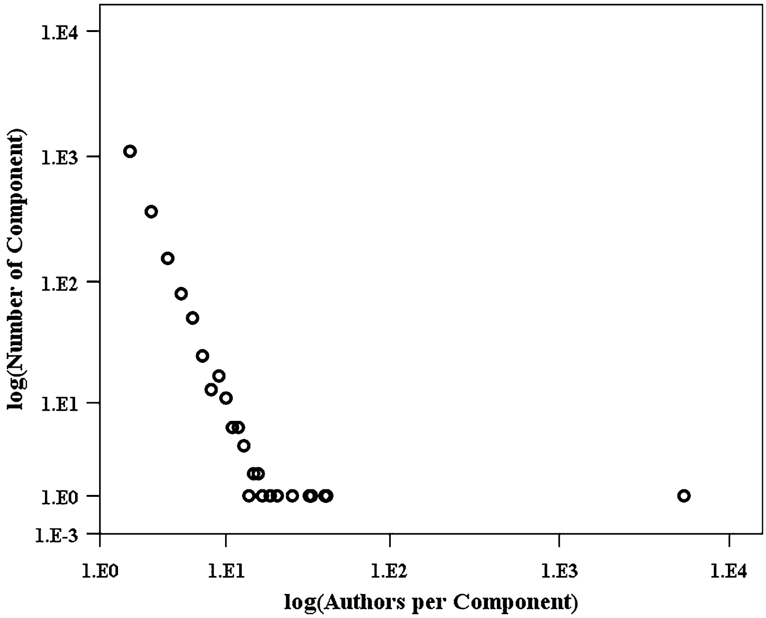

Fig. 1 Distribution of components

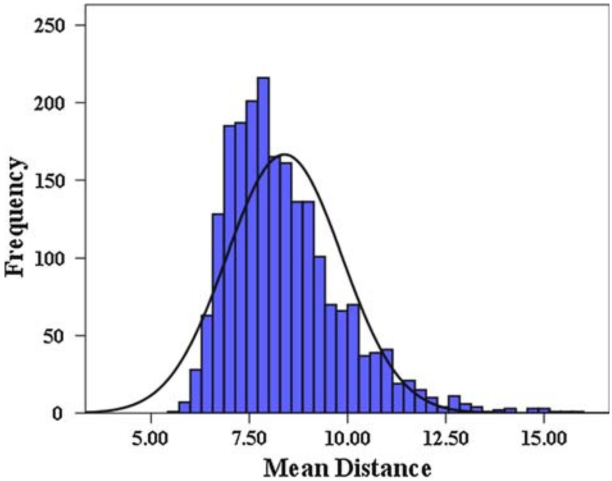

Fig. 2 Distance distribution of largest component

relationship between the two variables fit the curve: $p(k) = 2.0020\ k^{-2.6992}$, with $R^2 = 0.9611$. This shows that the LIS coauthorship network possess certain character of scale-free network. Furthermore, this result is also consistent with Price's network of citations, which is probably the earliest published example of a scale-free network (Price 1965). He quoted a value of $\alpha = 2.5$–3 for the exponent of his network. Other relevant research on scale-free network also confirmed his assumption (Newman 2003).

In order to test the error and attack tolerance of the LIS coauthorship network, we conduct two sets of experiment as follows. Firstly, we delete a certain number of vertices from the largest component of LIS coauthorship network randomly, with the decrement shown on the right column of Table 4. The decrement is based on the last largest component remained. In each of the steps, we record the size of largest component remained, the number of component, mean distance of the largest component remained, and its diameter. Then we reuse the largest component, and delete the hubs of the network. The decrements shown in Table 5 represent the deletion of vertices with more than 30 ties, more than 20 ties, more than 15 ties, more than 10 ties, and more than 6 ties respectively, and record the same set of data as in the random deletion.

Table 4 shows that the size of largest component remaining decreases gradually, and the number of component, mean distance of the largest component remained, and its diameter increases gradually or fluctuantly as more vertices are deleted. The network breaks down when about 78% of its vertices are deleted. The same results can be applied to the deletion of hubs in the network shown in Table 5. However, comparing to the small amount of vertices deleted in the second time, the decrement of the size of largest component remained and the increment of the number of components, mean distance of the largest

Table 3 Degree distribution of LIS coauthorship

Degree	Freq.	Freq.%	Degree	Freq.	Freq.%	Degree	Freq.	Freq.%
1	4,763	43.0379	13	22	0.1988	25	4	0.0361
2	2,814	25.4269	14	26	0.2349	26	2	0.0181
3	1,334	12.0539	15	16	0.1446	27	6	0.0542
4	737	6.6594	16	14	0.1265	29	2	0.0181
5	423	3.8222	17	9	0.0813	30	1	0.0090
6	257	2.3222	18	9	0.0813	33	1	0.0090
7	165	1.4909	19	6	0.0542	34	1	0.0090
8	136	1.2289	20	5	0.0452	35	1	0.0090
9	98	0.8855	21	3	0.0271	36	1	0.0090
10	87	0.7861	22	3	0.0271	38	1	0.0090
11	70	0.6325	23	8	0.0723	42	1	0.0090
12	34	0.3072	24	5	0.0452	49	1	0.0090

Table 4 Results of random deletion

Size of largest component	Ration to 5408	No. of component	Mean distance	Diameter	No. of vertices	Fraction deleted	Decrement
5,257	0.9721	22	9.0875	25	5,358	0.0092	50
5,066	0.9368	23	9.1695	25	5,207	0.0372	50
4,175	0.7720	98	9.5236	24	4,866	0.1002	200
3,351	0.6196	82	11.5659	35	3,975	0.2650	200
2,374	0.4390	103	13.4171	33	3,151	0.4173	200
1,369	0.2531	81	16.7354	41	2,174	0.5980	200
372	0.0688	82	11.3480	29	1,169	0.7838	200

Table 5 Results of the deletion of hubs

Size of largest component	Ration to 5408	No. of component	Mean distance	Diameter	No. of vertices	Fraction deleted	Decrement
5,306	0.9811	29	9.5040	26	5,400	0.0015	8
4,845	0.8959	57	11.1686	31	5,273	0.0250	35
4,397	0.8131	51	13.389	40	4,807	0.1111	38
1,825	0.3375	280	20.5037	53	4,236	0.2167	161
39	0.0072	280	5.1331	13	1,628	0.6990	197

component remained, and its diameter are sharper in the deletion of hubs. The figures indicate that the loss of hubs in the LIS coauthorship network results in considerable fragmentation; this suggests that the community is fragile, and dependent on central figures to maintain connection across the entire network. Removal of additional, less central vertices, however, the speed through fragmentation is significantly slow.

Micro-level structure analysis

We apply three centrality measures (degree, closeness, betweenness centrality) to investigate different contribution of vertices in the coauthorship network. We also retrieve accumulative citation counts of each author's publications in the database of CSSCI from 2002 to 2006 (Due to some problem inherent to CSSCI, citations to a paper can only be counted for the first author).

The effect of identical names has been magnified in degree centrality analysis. Such as Li Yong and Zhou Ningli in Table 6, they are normal names in China, which suggests that they may stands for more than one author in the network. Meanwhile, some abnormal data occurs, the citations of Hu Tiejun, Zhao Pengmo, Hu Dehua, and Xie Yangqun are too small comparing to their degree centrality, this may due to the fact that they are not first authors for most papers they wrote, accordingly, CSSCI cannot calculate the citations to these papers for them.

In Table 7, we display the top 20 authors with the highest closeness centrality.

Some abnormal data also show up in this table: there are discrepancies between closeness centrality and citation counts for Jiang Enbo and Feng Ying. This is the result of the algorithm of closeness centrality. Closeness centrality measures the distance of an author to the rest authors in the network. Thus, if an author coauthored with an author with high closeness centrality, this author would also have a high closeness centrality. For this analysis, Feng Ying only coauthored once with Chen Ling who has a high closeness centrality; Jiang Enbo coauthored once with Zhang Xiaolin who also has a high closeness centrality. Thus, they have high closeness centrality but low citation counts. Based on this, we can conclude that central authors usually have high closeness centrality (Yin et al. 2006), but high closeness centrality does not necessarily indicate that an author is central to the network.

In Table 8, we display the top 20 authors with the highest betweenness centrality in the largest component.

The largest component contains 5,408 vertices, while 2,134 vertices have betweenness centrality other than zero. The result indicates that the removal of these vertices will increase the distance of the network, while the removal of the other 3,274 will not.

Table 6 Degree centrality (top 41 authors)

Author	Degree	Citation	Author	Degree	Citation	Author	Degree	Citation
Qiu Junping	49	388	Wang Zizhou	27	210	Hu Dehua	23	8
Jing Jipeng	42	75	Bi Qiang	27	49	Li Yong	23	132
Hou Hanqing	38	31	Zhou Ning	26	128	Wang Yan	23	68
Zhang Xiaolin	36	403	Qin Tiehui	26	54	Zhao Pengmo	23	1
Su Xinning	35	144	Dong Li	25	19	Zang Guoquan	23	75
Mao Jun	34	58	Jiang Airong	25	30	Zhu Zhongming	23	27
Zhu Qinghua	33	73	Jiao Yuying	25	81	Li Guangjian	23	50
Wang Zhijin	30	194	Zhang Yufeng	25	26	Zhou Ningli	22	18
Sun Chengquan	29	14	Sheng Xiaoping	24	169	Liu Lei	22	75
Ma Haiqun	29	130	Zhang Zhixiong	24	49	Chen Chuanfu	22	160
Hu Changping	27	189	Dong Hui	24	52	Liang Zhanping	21	89
Lai Maosheng	27	64	Gan Liren	24	56	Ma Feicheng	21	347
Liu Wei	27	76	Yang Yi	24	58	Xie Yangqun	21	10
Wang Yuefen	27	35	Hu Tiejun	23	0			

Table 7 Closeness centrality (top 20 authors)

Author	Closeness centrality	Mean distance	Citation	Author	Closeness centrality	Mean distance	Citation
Mao Jun	0.1782	5.6113	58	Ma Haiqun	0.1653	6.0494	130
Zhang Xiaolin	0.1746	5.7289	403	Chen Chuanfu	0.1645	6.0804	160
Qiu Junping	0.1727	5.7901	388	Li Dan	0.1644	6.0826	85
Wang Xin	0.1713	5.8376	66	Gao Fan	0.1643	6.0855	31
Liu Wei	0.1681	5.9501	76	Li Guangjian	0.1643	6.0872	50
Sheng Xiaoping	0.1672	5.9796	169	Jiang Enbo	0.1642	6.0903	9
Chun Jingli	0.1669	5.9901	210	Zhou Ningli	0.1640	6.0976	18
Yu Yuan	0.1669	5.9932	19	Jing Jipeng	0.1639	6.1004	75
Zhang Xuefu	0.1657	6.0332	20	Feng Ying	0.1638	6.1061	4
Chen Ling	0.1657	6.0351	42	Bi Qiang	0.1629	6.1402	49

Applying centrality to impact analysis

Before the advent of mature network models and theories, scientific evaluation is based on comparatively simple and static data, such as citation counts, publication counts, impact factor, and so on. Now, with the help of high performance computer as well as advanced network theories, we are able to construct networks with thousands even millions of vertices. Accordingly, many characteristics and properties of networks can be applied to scientific evaluation. Specifically, centrality analysis displayed appropriate characteristics as indicators for scientific evaluation.

In the interest of verifying the correlation of centrality with citation counts, we conduct correlation analysis for the two values of authors in Tables 6, 7, 8. The results are shown in Table 9.

Table 8 Betweenness centrality (top 20 authors)

Author	Betweenness centrality	Citation	Author	Betweenness centrality	Citation
Qiu Junping	0.1378	388	Wang Xin	0.0472	66
Zhang Xiaolin	0.1058	403	Bi Qiang	0.0463	49
Mao Jun	0.0925	58	Liang Zhanping	0.0454	89
Jing Jipeng	0.0720	75	Li Guangde	0.0430	3
Ma Haiqun	0.0697	130	Li Wei	0.0422	162
Li Dan	0.0631	85	Sheng Xiaoping	0.0409	169
Hu Changping	0.0540	189	Wang Lin	0.0385	63
Chen Ling	0.0527	42	Zhang Min	0.0376	95
Chen Chuanfu	0.0493	160	Liu Wei	0.0376	76
Qin Tiehui	0.0489	54	Zhang Xufeng	0.0359	26

Table 9 shows that the three centrality measures correlate with citation counts, with betweenness centrality has the most significant correlation. The high correlation of citation counts with centrality suggests that centrality measures in certain degree also assess author's scientific productivity and quality. They can be indicators, or at least supplementary indicators for impact evaluation, providing alternative perspectives for current methods.

Table 10 shows the rankings of citations and three centrality measures (R_D stands for the rank results of degree centrality, R_C for closeness centrality, R_B for betweenness centrality, and R_{CC} for citation counts within top 20 authors).

Generally, citation counts and citation counts ranking match the ranking acquired by centrality measures. Yet, there are some discrepancies, especially for each individual author. Although the motivation for citation varies, citation counts measure the quality and impact of articles (Garfield and Sher 1963; Frost 1979; Lawani and Bayer 1983; Baird and Oppenheim 1994). While centrality measures both article impact and author's field impact: degree centrality measures author's collaboration scope, closeness centrality measures author's position and virtual distance with others in the field, and betweenness centrality measures author's importance to other authors' communication. Hence, centrality has its value in impact evaluation, since it integrates both article impact and author's field which is usually difficult to measure. Their relationship can be illustrated in Fig. 3.

Table 9 Partial correlation between each pair of centrality measures

	Degree	Closeness	Betweenness	Citation
Degree	1	−0.2129 $p < 0.0000$	−0.4687 $p < 0.0000$	0.4223 $p = 0.0060$
Closeness	−0.2129 $p < 0.0000$	1	0.2049 $p < 0.0000$	0.5614 $p = 0.0100$
Betweenness	−0.4687 $p < 0.0000$	0.2049 $p < 0.0000$	1	0.7282 $p = 0.0003$
Citation	0.4223 $p = 0.0060$	0.5614 $p = 0.0100$	0.7282 $p = 0.0003$	1

Table 10 Overall ranking of top 20 authors in the network

Author	Citations	R_{CC}	R_D	R_C	R_B
Zhang Xiaolin	403	1	4	2	2
Qiu Junping	388	2	1	3	1
Wang Zhijin	194	3	8	297	30
Hu Changping	189	4	14	36	7
Sheng Xiaoping	169	5	26	6	16
Chen Chuanfu	160	6	37	12	9
Su Xinning	144	7	5	479	46
Ma Haiqun	130	8	10	11	5
Liang Zhanping	89	9	39	148	13
Li Dan	85	10	56	13	6
Liu Wei	76	11	16	5	19
Jing Jipeng	75	12	2	18	4
Zhu Qinghua	73	13	7	387	34
Wang Xin	66	14	49	4	11
Lai Maosheng	64	15	15	92	21
Mao Jun	58	16	6	1	3
Qin Tiehui	54	17	17	23	10
Bi Qiang	49	18	11	20	12
Chen Ling	42	19	78	10	8
Hou Hanqing	31	20	3	135	23

The quality of an article is subjective, yet we can measure it indirectly through article impact which can be quantified by citation counts. Similar to article quality, author's field reputation is also difficult to assess, but we can assess it through social capital (Burt 1980; Burt 2002; Cronin and Shaw 2002). Accordingly, centrality measures integrate both article impact—citation counts and social capital—author's field impact, as displayed in Fig. 3.

Another factor contributed to the discrepancy lies on the algorithm of centrality measures. The current algorithm has some inherent drawbacks. Authors from papers coauthored by multiple authors have high degree centrality. This may be magnified when coauthored by many authors. For instance, if a paper is coauthored by 10 authors, each of these authors would have a degree centrality of 9. This is equivalent to 45 papers if they were coauthored by just two authors. It is obvious that they have quite different academic impact. Closeness centrality is a measure of network property rather than a direct measure of academic impact. Authors involved in interdisciplinary research would have a high betweenness centrality

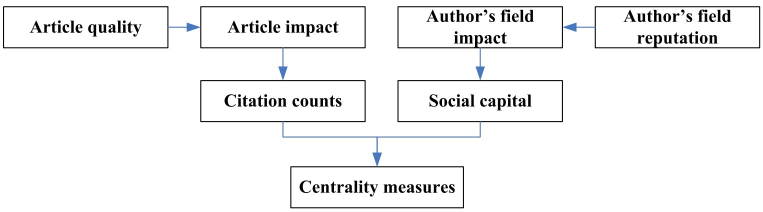

Fig. 3 Relation between citation and centrality

even through their role in this specific discipline may not be that significant. In future studies, it will be necessary to improve the algorithm of centrality measure to improve the reliability of centrality measures. Some scholars have already embarked on this. Newman (2005) proposed a new betweenness measure that includes contributions from essentially all paths between nodes, not just the shortest, and meanwhile giving more weight to short paths. Brandes (2008) introduced variants of betweenness measures, as endpoint betweenness, proxies betweenness, and bounded distance betweenness.

Conclusion

Using the data from 18 core source journals in the field of LIS covering 6 years, we construct the coauthorship network for LIS in China. This network provides us with rich information about the collaboration patterns of LIS studies in China. We find that this network is a small-world network, which indicates that there are many shortcuts in the network, where authors can reach others or be reached with much less ties. It also indicates that information flows more quickly and more directly in this network, inspiring more collaboration in the foreseeable future. Through degree distribution analysis, we find that this network also possesses the characteristics of scale-free networks in which quite a few authors collaborate widely while most authors collaborate with limited number of authors.

We also analyze the network at the micro-level through degree, closeness and betweenness centrality. We conduct correlation analysis between citation counts and centrality values, and discover that they are highly correlated. The centrality is originated from sociology (Freeman 1977, 1979). Yet centrality analysis is relatively new to coauthorship network analysis. Currently several papers have applied centrality measures to coauthorship network analysis (Mutschke 2003; Liu et al. 2005; Yin et al. 2006; Liu et al. 2007), they all found that centrality measures are useful for impact evaluation. From our study, we find that citation counts are correlated with centrality measures, yet some discrepancies occur for individual author. On one hand, this is due to the fact that citation and centrality evaluate different content; on the other hand, some drawbacks of centrality algorithms also add noises to the results.

The data and method used in this study only reflect the collaboration patterns of LIS studies in China for a given time period. It is a static rather than a dynamic one. In future studies, we intended to apply PageRank to impact analysis. PageRank creates a new synergy to information retrieval for a better ranking of Web pages. It is query independent and also content free. PageRank has been deployed in bibliometrics to evaluate research impact (Liu et al. 2005; Yin et al. 2006; Chen et al. 2007). The PageRank algorithm can be viewed as a state space system of the form:

$$x(n+1) = dTx(n) + b$$

The state $x(n)$ is a Markov vector of length v, and T is a Markov matrix on $\mathbb{R}^v$, while d the so called *damping* is a scalar in [0, 1]. Furthermore, b is a vector of length v consisting of all positive numbers which sum to $1 - d$. PageRank provides an integrated algorithm to combine simple counting (the b part) and the topology of the network (the $dTx(n)$ part) in a simple and efficient way. The damping factor allows us to tune the algorithm based to the specific needs of the applications and whether the topology of the network should be considered or not and to which level. In the future, we are interested to use PageRank and different weighted PageRank algorithms to analyze co-authorship networks.

Acknowledgements The authors would like to thank Blaise Cronin, Lokman Meho, Elin Jacob and Alice Robbin for their review of this article. The authors would also like to thank Lijiang Guo and Hui Fang for their assistance.

References

Albert, R., & Barabási, A. (2002). Statistical mechanics of complex networks. *Review of Modern Physics, 74*(1), 47–97.

Albert, R., Jeong, H., & Barabási, A. (2000). Error and attack tolerance of complex networks. *Nature, 406*, 378–382.

Baird, L. M., & Oppenheim, C. (1994). Do citations matter. *Journal of Information Science, 20*(1), 2–15.

Barabási, A., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*, 509–512.

Barabási, A. L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A—Statistical Mechanics and Its Applications, 311*(3–4), 590–614.

Brandes, U. (2008). On variants of shortest-path betweenness centrality and their generic computation. *Social Networks, 30*, 136–145.

Burt, R. S. (1980). Autonomy in a social topology. *American Journal of Sociology, 85*, 892–925.

Burt, R. S. (2002). The social capital of structural holes. In M. F. Guillén, R. Collins, P. England, & M. Russell (Eds.), *New directions in economic sociology* (pp. 203–247). Thousand Oaks, CA: Sage Foundation.

Castro, R. D., & Grossman, J. (1999). Famous trails to Paul Erdös. *MATHINT: The Mathematical Intelligencer, 21*, 51–63.

Chen, P., Xie, H., Maslov, S., & Redner, S. (2007). Finding scientific gems with Google. *Journal of Informetrics, 1*, 8–15.

Cronin, B., & Shaw, D. (2002). Banking (on) different forms of symbolic capital. *Journal of the American Society for Information Science and Technology, 53*(14), 1267–1270.

Cronin, B., & Shaw, D. (2007). Peers and spheres of influence: Situating Rob Kling. *Information Society, 23*(4), 221–233.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry, 40*(1), 35–41.

Freeman, L. C. (1979). Centrality in social networks. Conceptual clarification. *Social Networks, 1*, 215–239.

Frost, C. O. (1979). The use of citations in literary research: A preliminary classification of citation functions. *Library Quarterly, 49*(4), 399–414.

Garfield, E., & Sher, I. H. (1963). New factors in evaluation of scientific literature through citation indexing. *American Documentation, 14*(3), 195–202.

Grossman, W. J. (2002). The evolution of the mathematical research collaboration graph. *Congressus Numerantium, 158*, 201–212.

Hou, H., Kretschmer, H., & Liu, Z. (2008). The structure of scientific collaboration networks in Scientometrics. *Scientometrics, 75*(2), 189–202.

Jin, X. (1999). A comparison study on the confusion and threat of the United States and Chinese library education. *International Information & Library Review, 31*, 1–18.

Kretschmer, H. (2004). Author productivity and geodesic distance in bibliographic co-authorship networks and visibility on the Web. *Scientometrics, 60*(3), 409–420.

Lawani, S. M., & Bayer, A. E. (1983). Validity of citation criteria for assessing the influence of scientific publications—new evidence with peer assessment. *Journal of the American Society for Information Science, 34*(1), 59–66.

Liu, X., Bollen, J., Nelson, M. L., & Van de Sompel, H. (2005). Co-authorship networks in the digital library research community. *Information Processing and Management, 41*, 1462–1480.

Liu, L. G., Xuan, Z. G., Dang, Z. Y., Guo, Q., & Wang, Z. T. (2007). Weighted network properties of Chinese nature science basic research. *Physica A—Statistical Mechanics and Its Applications, 377*(1), 302–314.

Milgram, S. (1967). The small world problem. *Psychology Today, 2*, 60–67.

Mutschke, P. (2003). Mining networks and central entities in digital libraries. A graph theoretic approach applied to co-author networks. *Advances in Intelligent Data Analysis V, 2810*, 155–166.

Nascimento, M. A., Sander, J., & Pound, J. (2003). Analysis of SIGMODs coauthorship graph. *SIGMOD Record, 32*(3), 8–10.

Newman, M. E. J. (2001a). Scientific collaboration networks: I. Network construction and fundamental results. *Physical Review E, 64*, 016131.

Newman, M. E. J. (2001b). Scientific collaboration networks: II. Shortest paths, weighted networks, and centrality. *Physical Review E, 64*, 016132.

Newman, M. E. J. (2001c). The structure of scientific collaboration networks. *Proceedings of the National Academy of Science of the United States of America, 98*(2), 404–409.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review, 45*(2), 167–256.

Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America, 101*, 5200–5205.

Newman, M. E. J. (2005). A measure of betweenness centrality based on random walks. *Social Networks, 27*, 39–54.

Nooy, W., Mrvar, A., & Batagelj, V. (2005). *Exploratory social network analysis with pajek*. Cambridge, UK: Cambridge University Press.

Price, J. D. S. (1965). Networks of scientific papers. *Science, 149*, 510–515.

Rodriguez, M. A., & Pepe, A. (2008). On the relationship between the structural and socioacademic communities of a coauthorship network. *Journal of Informetrics, 2*(3), 195–201.

Vidgen, R., Henneberg, S., & Naude, P. (2007). What sort of community is the European conference on information systems? A social network analysis 1993–2005. *European Journal of Information Systems, 16*(1), 5–19.

Wagner, C. S., & Leydesdorff, L. (2003). Mapping global science using international co-authorships: A comparison of 1990 and 2000. In *Proceedings of ninth international conference on scientometrics and informetrics*, Beijing, China.

Wang, Y., Wu, Y., Pan, Y., Ma, Z., & Rousseau, R. (2005). Scientific collaboration in China as reflected in coauthorship. *Scientometrics, 62*(2), 183–198.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge, UK: Cambridge University Press.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*, 440.

Wu, X., & Yuan, F. (1994). Library and information education in the People's Republic of China: The impact of reform and "opening-up". *Education for Information, 12*, 247–257.

Yin, L., Kretschmer, H., Hanneman, R. A., & Liu, Z. (2006). Connection and stratification in research collaboration: An analysis of the COLLNET network. *Information Processing and Management, 42*, 1599–1613.