International Journal of u- and e- Service, Science and Technology
Vol.6, No.6 (2013), pp.31-40
http://dx.doi.org/10.14257/ijunesst.2013.6.6.04

Scientific Collaboration Networks in China's System Engineering Subject

Sen Wu[1], Jiaye Wang[1,*], Xiaodong Feng[1] and Dan Lu[1]

[1] *Dongling School of Economics and Management, University of Science and Technology Beijing, Beijing, China*
wusen@manage.ustb.edu.cn, leaf_01@126.com, fengxd1988@gmail.com, ludan_email@163.com

Abstract

The collaboration of scientific research is becoming more intensive, especially in interdisciplinary subjects. Complex network theory can be explored to demonstrate the characteristics of the scientific collaboration. This paper analyzes the co-author networks from 3 important journals in Chinese system engineering field in recent years using complex network theory. Scientific papers published on selected journals are used to construct the scientific collaboration networks. Statistical properties measuring the clusters of the constructed networks and influence of authors are studied. Empirical results show that the networks have two important features: small world effect and scale-free property, derived from the small average shortest path length, large average clustering coefficients and the power law of degree distributions of these networks. Besides, the structure of graphs in scientific collaboration networks is revealed and analyzed. Based on this analysis, the research groups that are active in the field of system engineering in China can be discovered in a qualitative way.

Keywords: *Social networks, scientific collaboration, small world, scale-free*

1. Introduction

One of the main characteristics of current scientific research is the high penetration of cooperation among researchers[1]. The cooperation can be found in almost all the academic fields due to the development of communication tools and easiness of information interaction. Some research has also showed that coauthored articles have better quality than the articles finished alone in some subjects [2, 3]. In modern social and scientific research, massive information is presented in various ways. With the fast-speed spread and update, the information's great value can always spur the spark of scientific research.

One aspect of collaboration among researchers manifests as co-authored publications. Specifically, it has been realized that the co-authorship of publications in academic journals provides a visualized way to analyze the collaboration patterns within the academic community. Co-authorship of a paper can be viewed as the collaboration record of the two or more authors, and these collaborations form a co-authorship network, in which the network nodes represent authors, and two authors are connected by an edge if they have coauthored one or more papers.

The analysis of such scientific collaboration network (SCN) is useful to reveal many interesting features of academic communities. According to this analysis, authors that seem separated may be clustered into groups through the relationships of coauthors. And their relationships can be visualized by graphs or other ways. We can find the author communities in some research areas judging by some presented indicators.

ISSN: 2005-4246 IJUNESST

This paper is focused on the characteristic analysis of scientific collaboration network in China's systems engineering subject. Firstly, scientific collaboration networks are constructed according to scientific papers published in recent years from three selected typical journals .Then clustering features of the constructed networks and influence of the authors are demonstrated based on statistical properties.

The rest of the paper is organized as follows. Section 2 reviews the related work of the scientific collaboration network. Section 3 constructs the scientific collaboration networks in China's systems engineering subject according to the papers published in recent years from three selected typical journals and the statistical properties and structure of the scientific collaboration network are explored. Finally, section 4 concludes the research findings.

2. Related Work about Scientific Collaboration Network

Complex network is the subject that studies the networks' common properties and is devoted to discovering and processing their universality method [5].

Since the 1990s, complex network research has gradually been developed into a subject that has its own discipline and rules. And on some academic occasions it is even called "the new science of networks" by some scientists [6, 7]. Research in complex network subject has developed more vigorously, and lots of monographs [6, 8] and article comments [9, 10] have been published. The studies of the complex network have revealed the internal essence of many phenomena in real world and attracted a great deal of interest from researchers in different scientific fields. Normally, a complex network can be visualized in the form of graph. In the graph, there are two basic elements, namely node and edge. A node can represent a person, a cell or any entity in the real world, while the edge connecting two nodes represents a certain relationship between these two specific nodes, such as a conversation, a contract or a cooperation relationship of two researchers.

Social network as one kind of complex network is a collection of people, each of whom is related with some subset of the others through a particular relation. Such a network can be represented as a set of nodes denoting people, joined in pairs by edges denoting this relationship [4]. Recently, numerous initiatives have been launched with the aim of promoting collaboration among individual researchers and bringing them together. For instance, in new or large centers of excellence, or alternatively interdisciplinary research groups [11], scientific collaboration network is with a relatively stable structure. This form of collaboration should be encouraged. For scientific collaboration aiming to produce new scientific knowledge, scientific collaboration group is a good community for its members to share the same goal. The spread of knowledge in a scientific collaboration network can make full use of the strength of researchers and enhance efficiency of collaboration. In return, the collaboration can stabilize and strengthen cooperation relationships in the scientific collaboration network [12].

The main characteristics that a scientific collaboration network emphasizes include degree, degree distribution, clustering coefficient, distance, average shortest path-length and betweenness [11]. Generally, the authors concerned can be viewed as nodes of the network and their collaborative relations can be viewed as edges of the network [2]. That is, two nodes are connected if the two authors have published a paper together and explicit networks of such connections can be constructed using data from some academic databases. And the weight of the edge between two nodes represents the collaboration frequency between two authors. The author's influence can be measured by the node degree in the scientific collaboration network, and the collaboration closeness between authors is measured by edges' weights. If both the value of the node degree and the weights of edges are large enough, the current node tends to be important [13]. In other words, the author represented by this node is very important in terms of his status and collaborative relations in the scientific collaboration network. In this way, the academic

influence evaluation problem can be transformed into node influence evaluation problem in the scientific collaboration network.

In recent years, scientific collaboration has been widely accepted and there is lots of initiatives studying the cooperation of authors in different research fields. Some research results have also proved that the rate of coauthor is improving gradually [14, 15]. The study of collaboration network in quantitative and empirical aspects attracts much attention among scholars in different academic areas, including literature metrology, scientific metrology and science and technology management [2, 8].

3. Empirical Analysis of Scientific Collaboration Networks

Our work focuses on scientific collaboration network in systems engineering subject in China. This section first presents how the network is built by extracting information from several typical Chinese journals in system engineering subject, then analyzes the global property of the whole scientific collaboration network and the local characteristics of the sub-networks. average

3.1. Scientific Collaboration Network Extracted from CNKI

We first define the scope of this scientific collaboration network as system engineering subject in China and get data from CNKI (China National Knowledge Infrastructure), which is one of the most widely-used and influential Chinese electric library. The journals in CNKI are selected for the following reasons.

We choose the phrase 'systems engineering' as the keyword when we select journals from section N (list of natural science topic journals) in the GCJC (the Core Journals of China) (6th edition, 2011). We also take a list of 30 authoritative journals recognized by NFSC (National Natural Science Foundation of China) into consideration.

Moreover, we take the impact factor (IF) into consideration as well. The concept of IF was first introduced by E. Garfield in 1972. The IF of a journal refers to the frequency of citation by journal articles in a period [16, 17]. To some extent, the impact factor can be seen as a valid reflection of the influence of a journal. That is to say, the higher a journal's IF is, the more influential the journal is. In modern scientific research, an impact factor is not only an index measuring the academic level of a journal, but also an index which is of vital importance to measure the validity of the papers published in journals.

All these factors above can ensure the reliability of collected data. Considering the above two main reasons, we chose three journals, which are Systems Engineering-theory & Practice (*SE&P*), Journal of Systems Engineering (*JSE*), and Systems Engineering (*SE*). They are all in section N in GCJC, in the list recognized by NFSC. Their IFs are 0.854, 0.588 and 0.537, all higher than average value. To balance the number of papers of the 3 journals we selected, we decided to capture datasets from different years as shown in the following Table 1.

Table 1. The Datasets we Captured from CNKI

Journal	SE&P	JSE	SE
Time periods	2010-2012	2007-2012	2009-2012
Number of papers	938	581	936

Now we can model the scientific collaboration network as a graph G(V, E), where V denotes all the authors in the scientific collaboration network and E represent the relations between them. If any two authors v_i, v_j have ever coauthored at least one paper, an edge e_{ij} = (v_i, v_j) is added to E. For simplicity of description, we use the following notations in subsequent sections. Neg(v): the neighbor set of v; Inc(v): the set of edges that are attached to v, the number of incident edges of v.

So, we will construct 3 scientific collaboration networks, each of which is based on one journal.

3.2 Analysis on the Global Characteristic

In this section, we will analyze the properties and characteristics of the whole 3 scientific collaboration networks in terms of degree distribution, clustering coefficient and author influence. Firstly, the related concepts are introduced as follows.

(1) Degree. The degree of author v is defined as the cardinality of the neighbor set Neg(v), that is, Deg(v)=|Neg(v)|. Degree accounts for the number of connections of an author with other authors, which is a straight forward measurement of vertex importance [18].

(2) Degree Distribution. The degree distribution P(k) means the proportion of the number of authors which satisfy Deg(v)= k, and in many real-world complex networks, the degree distribution follows the power law:

$$P(k)=c*k^{-a} \quad (1)$$

Where, c is a constant, a denotes power law index. In 1999, Barabasi and Albert defined this power law as the network's scale-free property. They tracked the evaluation process of World Wide Web with the viewpoint and background of statistical physics, and then discovered that there are many complex networks whose distributions of degree comply with the power law [19,20]. This kind of property is the so-called scale-free, which means many authors have few neighbors while few authors have many neighbors.

(3) Clustering Coefficient. Clustering coefficient C is to describe the gathering property of all authors in a network. The clustering coefficient of author v is the fraction of possible triangles that exists. Therefore C(v) is defined as:

$$C(v) = \frac{2T(v)}{Deg(v) * (Deg(v) - 1)} \quad (2)$$

Where, T(v) is the number of triangles contain author v and Deg(v) is the above defined degree of v. In scientific collaboration networks, clustering coefficient means the probability of the collaboration of two authors both having collaboration with the same author. The average clustering coefficient of a network is the average clustering coefficient of all nodes in the network, evaluating how well connected the neighbors of a vertex in the whole graph are [11].

(4) PageRank. PageRank is widely used for ranking web pages that are returned as the answers to a query on search engine [21]. The principle behind Page Rank is that every link is a vote and votes from important webs or authors have significant contributions to the importance of this web or author. Hence, the more links from important authors to an author, the higher Page Rank is assigned to the author. An author's importance ranked by Page Rank, PR(v), is given by the following equation:

$$PR(v) = (1 - \alpha) + \alpha \sum_{u \in Neg(v)} \frac{PR(u)}{Deg(u)} \quad (3)$$

Where, PR(u) is the PageRank of author u and α is a damping parameter, which can be set between 0 and 1, denoting the probability at each page(author) the "random surfer" will get bored and request another random page(author)[21]. In the Google's research of PageRank[21], it was usually set to 0.85, so in our following experiments using networkx tool, α is also set the default value.

Table 2 gives a summary of the above characteristics of the 3 scientific collaboration networks described in the previous section.

Table 2. Summary of the Analysis Results of Three Scientific Collaboration Networks

	SE&P	*JSE*	*SE*
number of papers	938	581	936
number of authors	2118	1162	1714
maximum degree	57	13	24
average degree	2.71	2.31	2.32
average clustering coefficient	0.75	0.66	0.63
maximum PageRank	11.80	4.25	6.71
number of connected subnetworks	518	307	454

Table 2 demonstrates the differences between the 3 scientific collaboration networks. The average degree of *SE&P* is much larger than other two, which suggests that the level of collaboration in *SE&P* scientific collaboration network is stronger than those in *JSE* and *SE*. Moreover, the difference of collaboration rate among the 3 scientific collaboration networks can also be demonstrated from the difference among their average clustering coefficients, that are 0.75, 0.66 and 0.63 respectively. The average clustering coefficients of the three scientific collaboration networks are relatively large, meaning all the 3 scientific collaboration networks have strong collaboration and therefore have the characteristics of the small world network.

To illustrate the details of the degree distribution, we make the following three tables (Table 3 to Table 5).

Table 3. Degree Distribution in the Network of Systems Engineering-theory & Practice

degree	1	2	3	4	5	6	7	8	9
the number of nodes	387	794	542	205	89	45	18	3	16
degree	10	11	12	13	15	17	20	24	57
the number of nodes	3	5	2	1	2	3	1	1	1

Table 4. Degree Distribution in the Network of Journal of Systems Engineering

degree	1	2	3	4	5	6	7	8	9	10	12	13
the number of nodes	304	456	281	67	21	13	8	4	1	2	4	1

Table 5. Degree Distribution in the Network of Systems Engineering

degree	1	2	3	4	5	6	7	8
the number of nodes	510	697	332	94	54	15	6	12
degree	9	10	12	13	15	17	19	24
the number of nodes	8	4	2	1	1	3	1	1

Looking into degree distribution of the 3 journals, the phenomenon that nodes' degree are mostly 1, 2 or 3 can be clearly shown. Comparing scientific collaboration network with *SE* and *JSE*, we can see that there are more nodes with degree of 2 or 3 in the scientific collaboration network of *SE&P*. That is to say, authors of papers published in SE&P have a closer relationship in scientific collaboration.

According to these tables above, we can draw Log_Log figure of degree distribution and PageRank distribution. Since the PageRank values are continuous, to draw

PagerRank(PR) distribution figures, we firstly group the PageRank values with 0.1 as an interval from the minimum to the maximum.

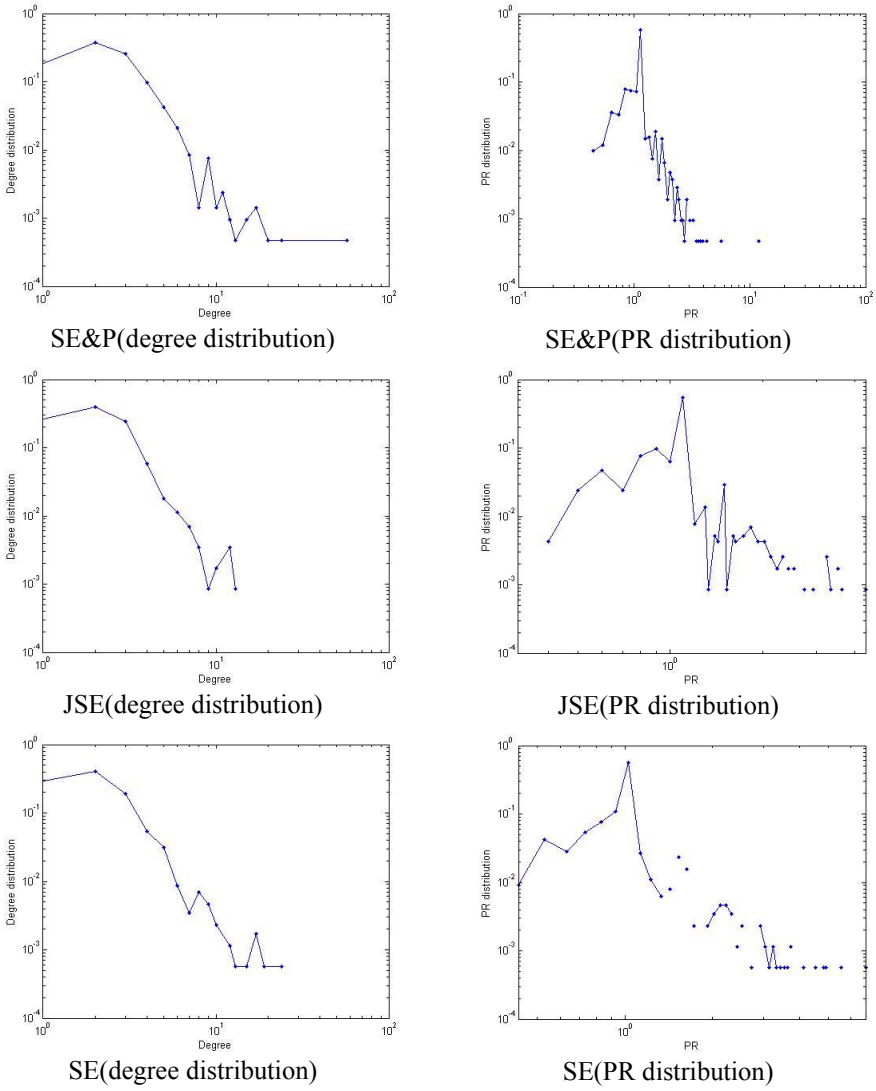


Figure 1. Degree Distribution and PR Distribution of the Three SCNs

Figure 1 above shows that the degree distribution drawn by Log_Log figure decreases with the increase of degree, demonstrating that these networks generally follow the power law, and their power law indexes are similar (respectively are -2.25, -2.657, -2.534). Obviously, these 3 networks can be seen as examples of scale-free networks. However the PR distribution has different trend that the PR distribution firstly increase slowly and then decrease sharply with a long tail when PR is large enough.

3.3. Structure of Connected Sub-networks

When constructing networks of the co-author relationships and conducting further study, each scientific collaboration network is unconnected as a whole. Therefore, the sub-networks are obtained through dividing the whole unconnected network into some connected ones.

In the social networks, two nodes can get in touch with each other by one edge or a path through other nodes and edges, between which there exists a shortest path to connect each two nodes in the connected networks. Therefore, for each connected network we can

get the average shortest path length for the whole network and betweenness centrality for each node.

(1) Average shortest path length is denoted as

$$ASPL = \frac{\sum_{u \in V, v \in V} d(u, v)}{n(n-1)} \tag{5}$$

Where, V is the set of nodes in a network, d(u, v) is the length of the shortest path from u to v, and n is the number of nodes in the network . Average shortest path length is to describe how compact the network is. The shorter the average shortest path length of a network is, the more compact the network is.

(2) Betweenness centrality of a node v is the sum of the ratio of the number of all-pairs shortest paths that pass through v to the total number of all-pairs shortest paths [22]:

$$BC(v) = \sum_{s,t \in V} \frac{\sigma(s, t \mid v)}{\sigma(s, t)} \tag{6}$$

Where V is the set of nodes, $\sigma(s, t)$ is the number of shortest paths between s and t, and $\sigma(s, t|v)$ is the number of those shortest paths passing through node v other than s, t. If $s=t$, $\sigma(s, t)=1$, and if $v \in \{s,t\}$, $\sigma(s, t|v)=0$. Like PageRank and degree, betweenness centrality can also be used to decide the authors' influence in the scientific collaboration networks.

Here we choose top 5 largest connected sub-networks of *SE&P*, *JSE* and *SE*, and present part of their characteristics as follows in Table 6 to Table 8.

Table 6. Index of Largest 5 Sub-networks of Systems Engineering-theory & Practice

Sub-network	the number of nodes in the sub-network	max degree	maximum betweenness centrality	average clustering coefficient	diameter	average shortest path-length
1	153	57	0.6965	0.7820	9	3.7967
2	37	24	0.6995	0.8292	4	2.2598
3	21	11	0.6895	0.6565	5	2.5714
4	20	12	0.5429	0.7907	4	2.0579
5	17	15	0.7104	0.7224	3	1.8162

Table 7. Index of Largest 5 sub-networks of Journal of Systems Engineering

Sub-network	the number of nodes in the sub-network	max degree	maximum betweenness centrality	average clustering coefficient	diameter	average shortest path-length
1	16	13	0.7698	0.8208	3	1.9417
2	15	7	0.6923	0.8425	4	2.3143
3	15	10	0.7802	0.7089	4	2.3905
4	14	5	0.6667	0.5571	5	2.6923
5	14	10	0.8590	0.4484	4	2.1209

Table 8. Index of Largest 5 sub-networks of Systems Engineering

Sub-network	the number of nodes in the sub-network	max degree	maximum betweenness centrality	average clustering coefficient	diameter	average shortest path-length
1	33	19	0.6035	0.6836	5	2.6742
2	28	24	0.8362	0.8225	4	2.0053
3	23	17	0.9221	0.3650	4	2.2372
4	22	17	0.8429	0.6559	5	2.1991
5	21	17	0.8737	0.8328	3	2.0333

Table 6-8 show that the average shortest path-length of sub-networks is quite small. Combining it with high clustering coefficient of the whole network, which is also mentioned above, we can find that the whole network is under the small world effect.

To illustrate the structure of these sub-networks visually, Figure.2 shows the graphs of the largest connected sub-networks in each of the 3 scientific collaboration networks.

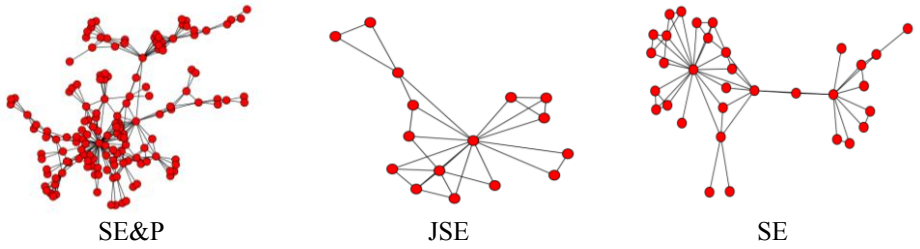
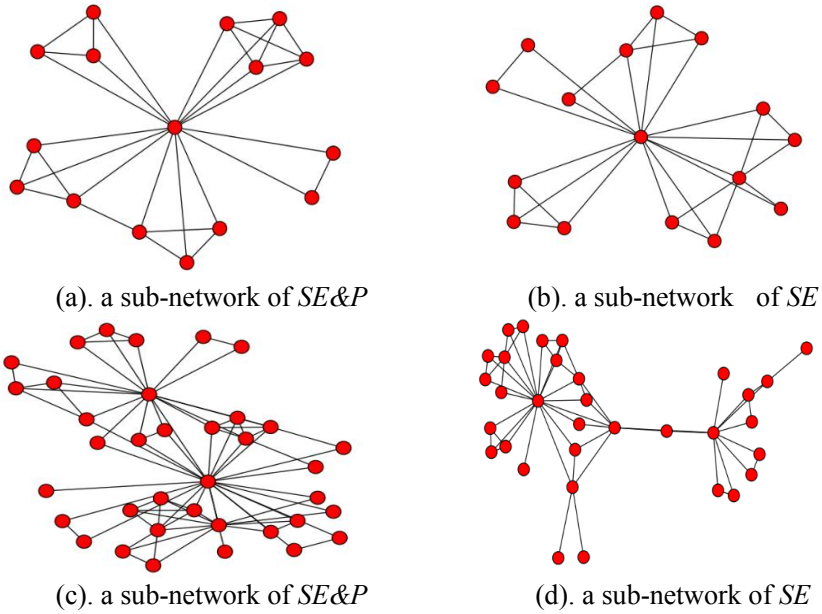


Figure 2. The Largest Connected sub-networks in each SCN

From Figure 2, it can be seen that the structure of the graphs in SCNs seems more like stars, meaning that little number of authors are surrounded by the large number of authors. Moreover, the farther the authors are from the central authors, the larger the number of authors is. This is an example of the power law that less nodes has more connected nodes (defined as degree in above).

In complex networks, the betweenness of a node is usually used to measure the node's influence and importance [22]. The structure of nodes' betweenness in smaller sub-networks is simpler. Here we mainly analyze several representative sub-networks.

In Figure. 3, subfigure (a) and (b) are both graphs of 16-node-network. As can be seen from the graphs, they are also typical star networks with a core author, namely the leader of the group, surrounded by all other authors. The average clustering coefficient of them are 0.906 and 0.892. The average shortest path-lengths of them are 1.73 and 1.75. Their diameters are both 2. With shorter average shortest path-length, smaller diameter and higher average clustering coefficient, both of them are typical small world networks. In subfigure (c), except for two apparent scattered groups, others are interconnected and the core nodes are connected directly too. So researchers in this network can exert bigger influence on each other with the spirit of teamwork, which is good for spreading and controlling new ideas. Subfigure (d) is the largest sub-network in network of Systems Engineering and obviously there are two groups in it. One group has 10 nodes, which is connected to the other 22-node group by a key scientific researcher. The maximum degree of this sub-network is 19. The betweenness centrality of the core authors in the four subfigures are 0.82, 0.81, 0.70 and 0.60 respectively, which are big and mean that most authors are connected by each core author in each scientific collaboration network.



(a). a sub-network of *SE&P* (b). a sub-network of *SE*

(c). a sub-network of *SE&P* (d). a sub-network of *SE*

Figure 3. Some Graphs of Representative sub-networks

4. Conclusion

In this paper, we construct scientific collaboration networks using academic papers published in recent years from 3 selected typical Chinese journals of system engineering subject. The networks of the selected journals are compared through several statistical properties of the constructed networks. The empirical results show that each network has small average shortest path length and large average clustering coefficients, the degree distributions of all the networks generally comply with the power law and their power law indexes are similar, which means the networks present two important characteristics: small world effect and scale-free property. This means that the scientific collaboration network has characteristics of a complex network. We also demonstrate structure of the largest connected sub-networks of the 3 scientific collaboration networks and other representative sub-networks, finding that graphs in SCNs seems more like stars. The core nodes, or the leader researchers in the networks can exert bigger influence on others. Based on the analysis of several representative sub-networks, the specific characteristics, such as how the influence disseminates and how close the collaboration is of some research groups in system engineering subject in China can also be discovered.

Acknowledgements

This work is supported by the National Natural Science Foundation of China under Grant No.71271027, by the Fundamental Research Funds for the Central Universities of China under Grant No. FRF-TP-10-006B and by the Research Fund for the Doctoral Program of Higher Education under Grant No. 20120006110037.

References

[1] R. H. Lin and J. H. Fan, "A Study of the Characteristics of Co-authorship Network of Chinese Scholars in Management and Their Impacts on Cooperative Performance", J. Research and Develop Management, vol. 04, **(2012)**, pp. 81-92.
[2] J. S. Katz and D. Hicks, "How much is a collaboration worth? A calibrated bibliometric model", Scientometrics, vol. 40, no. 3, **(1997)**, pp. 541-554.
[3] W. Glanzel and A. Schubert, "Double Effort = Double Impact? A Critical View at International Co-authorship in Chemistry", Scientometrics, vol. 50, no. 2, **(2001)**, pp. 199-214.

[4] N. Eagle, M. Macy and R. Claxton, "Network Diversity and Economic Development.Science", vol. 328, no. 1029, **(2010)**.
[5] R. A. Olsen, "Book Reviews: A Review of Linked: The New Science of Networks and Nexus: Small Worlds and the Groundbreaking Science of Networks", J. Journal of Behavioral Finance, vol. 41, **(2003)**.
[6] R. P. Kumar and S. Saramaki, "The strength of strong ties in scientific collaboration networks", Europhysics Letters, vol. 97, no. 1, **(2012)**, pp. 18007.
[7] D. J. Watts, "The 'New' Science of Networks. Annual Review of Sociology", vol. 30, **(2004)**, pp. 243-270.
[8] G. R. Chen, "Introduction to Complex Networks and Their Recent Advances", Advances in Mechanics, vol. 06, **(2008)**, pp. 653-662.
[9] S. H. Strogatz, "Exploring Complex Network", Nature, vol. 410, **(2001)**, pp. 268-176.
[10] R. Albert and A. L. Barabasi, "Statistical Mechanics of Complex Networks. Reviews of Modern Physics", vol. 71, **(2002)**, pp. 27-97.
[11] N. S. Sara and V. Alexei, "Network Clustering Coefficient without Degree-correlation Biases", Phys. Rev., vol. 71, no. 057101, **(2005)**.
[12] B. He, Y. Ding and J. Tang, "Mining diversity subgraph in multidisciplinary scientific collaboration networks: A meso perspective", Journal of Informetrics, vol. 7, no. 1, **(2013)**, pp. 117-128.
[13] P. C. Zhang and H. Peng, "On the Relationship between Scientific Cooperative Networks' Characteristics and Teams' Knowledge Creation", Science Research Management, vol. 07, **(2011)**, pp. 104-112.
[14] B. F. Jones, S. Wuchty and B. Uzzi, "Multi-university research teams: Shifting Impact, Geography, and Stratification in Science", J. Science, vol. 322, no. 5909, **(2008)**, pp. 1259-1262.
[15] A. Gazni, R. Cassidy and F. Didegah, "Mapping world scientific collaboration: Authors, institutions, and countries", Journal of the American Society for Information Science and Technology, vol. 63, no. 2, **(2012)**, pp. 323-335.
[16] E. Garfield, "The history and meaning of the journal impact factor", JAMA: the journal of the American Medical Association, vol. 295, no. 1, **(2006)**, pp. 90-93.
[17] P. Huang, Y. Q. Luo, Y. Wang, H. M. Tan and S. Ma, "Analysis of Influencing of Impact Factor of Scientific Periodical and its Improving Measures", Acta Editologica. S1, **(2006)**, pp. 180-181.
[18] S. Wasserman and K. Faust, "Social Network Analysis: Methods and Applications", Cambridge: Cambridge University Press, **(1994)**.
[19] D. Watts and S. Strogatz, "Collective Dynamics of "Small World" Networks", Nature, vol. 393, no. 4, **(1998)**, pp. 440-442.
[20] D. Watts, "Networks, Dynamics, and the Small World Phenomenon", J. American Journal of Sociology, vol. 105, **(1999)**, pp. 493- 527.
[21] S. Brin and L. Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine", Computer Networks and ISDN Systems, vol. 30, **(1998)**, pp. 107-117.
[22] B. Ulrik, "On Variants of Shortest-Path Betweenness Centrality and their Generic Computation", Social Networks, vol. 30, no. 2, **(2008)**, pp. 136-145.

