Title	Understanding a Large Scientific Project from Network Perspective
Author(s)	Kun, NIE; Xijin, TANG; Jifa, Gu; NAKAMORI, Yoshiteru
Citation	
Issue Date	2005-11
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/3853
Rights	©2005 JAIST Press
Description	The original publication is available at JAIST Press http://www.jaist.ac.jp/library/jaist-press/index.html, IFSR 2005 : Proceedings of the First World Congress of the International Federation for Systems Research : The New Roles of Systems Sciences For a Knowledge-based Society : Nov. 14-17, 2063, Kobe, Japan, Symposium 1, Session 4 : Technology Creation Based on Knowledge Science Knowledge/Technology Management(1)

Understanding a Large Scientific Project from Network Perspective

Kun NIE[1], Xijin TANG[2], Jifa Gu[2] and Yoshiteru NAKAMORI[1]
[1]School of Knowledge Science, Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan
[2]Academy of Mathematics and Systems Science
Chinese Academy of Sciences, Beijing, 100080, P.R. CHINA

ABSTRACT

Complex network and social network analysis (SNA) are two representative branches in the field of network research. In this paper, four kinds of network are built on the basis of a practical large scientific project: namely coauthorship network, activity network, citation network and keyword network. After analyzing these networks, some useful information is acquired, for example, who share information with whom, how the project is organized and undertaken, ideas growing process, etc., which give us some ideas on how to organize and manage a large scientific project, how to promote knowledge sharing and knowledge creation during the process of scientific research.

Keywords: scientific collaboration, coauthorship network, activity network, citation network, keyword network

1. INTRODUCTION

Nowadays, many scientists who have different academic bacgrounds: physicists, mathematicians, biologists, computer scientists, sociologists etc., do research on various types of networks from different perspectives.

Barabasi et al. (2002) have studied the coauthorship network based on the electronic database containing all relevant journals in mathematics and neuro-science for an 8-year period[1,2], Newman (2001) have studied the coauthorship networks in the fields of physics, biomedical and computer science using data from computer databases of scientific papers over a 5-year period[3,4,5], and researches on other types of network, such as WWW, Internet, metabolic network, film actors network, and so on. The empirical studies above show that most of these networks are not consistent with ER model[6] proposed by Erdos and Renyi in 1959[2,5,7,8,9]. These networks are found to be highly clustered, like regular lattices, yet have small characteristic path lengths, like random graphs, so called 'small-world' network[10], moreover, the scientists found that a common property of many large networks is that the vertex connectivities follow a scale-free power-law distribution, Growth and preferential attachment (BA model) is used to explain why scale-free phenomena happens[11].

Just as what we have explained above, complex network studies the structure of network, while SNA focused on the relations within a network[12,13,14,15]. SNA has been sucessfully helped to tackle social problems and promote management innovation[16,17,18,19]. Some basic concepts related to SNA: degree, closeness, betweeness, component, subgroup, cutpoint, etc., are used in this paper.

Inspired by the researches above, we find network providing an efficient way to describe, understand and solve some difficult complex problems. Fortunately, we have undertaken a major project of NSFC (National Natural Science Foundation of China) from 1996-2004 whose purpose is to solve macroeconomic decision problem. we wonder whether network theory is useful in analyzing the large project, but our emphasis is not to study the topological structure of network, but to know how the members collaborate with each other, how the large project is organized and undertaken, how the new scientific ideas are incubated and developed into validated new theory, etc. So four kinds of network are constructed, which are introduced as follows:

2. COAUTHORSHIP NETWORK

Coauthorship network is a set of authors and their relations when writing one or more papers together. For a paper with several authors, there are relations between any two authors of them. Obviously, coauthoring is an important way of scientific collaboration, by which the authors can share information and knowledge with each other, especially when the authors are from different backgrounds, they can benefit greatly from each other.

By analyzing coauthorship network using network methods and theory, we can dig deep into the large project, while it is impossible if we use traditional data analysis. We can know:

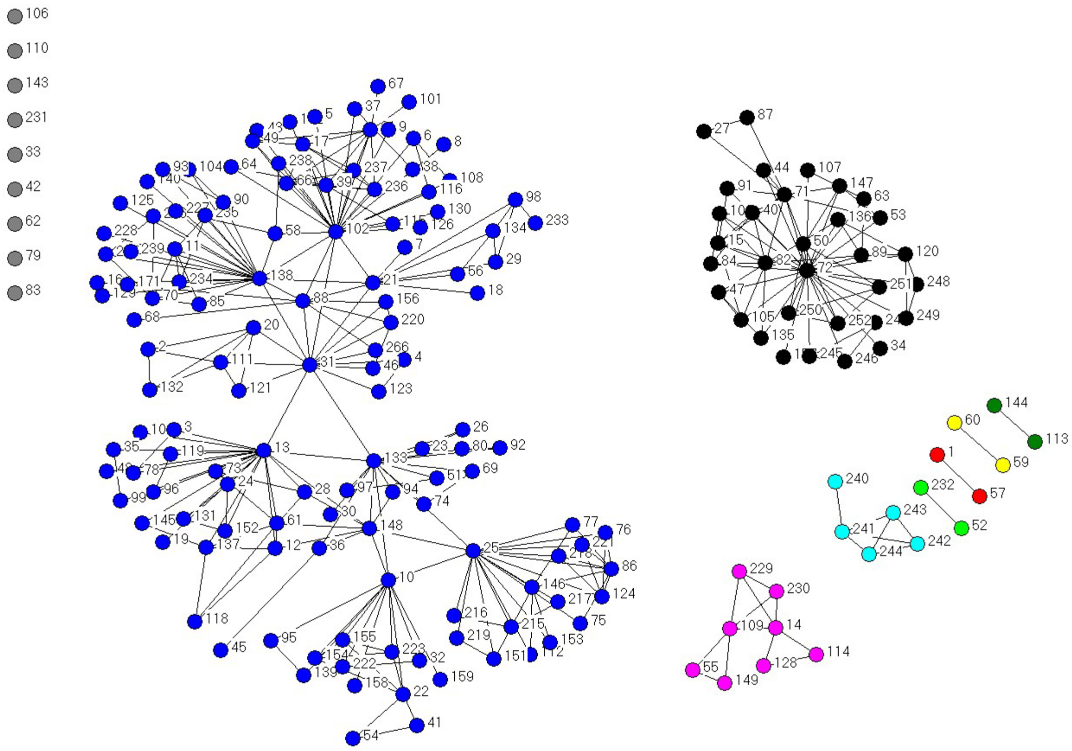

Fig.1: Coauthorship network by 2004

Note: In this graph, the codes stand for the authors' name. All together there are 192 authors, and 17 components which are separated from each other. The largest component has 130 authors, the second largest component has 32 authors, and the 9 isolated authors each is a component respectively. Within a component there are strong relations and weak relations between these authors which are distinguished by the number of collaboration.

1) The degree of connection. For the largest components with 130 authors, the average shortest path is 3.55 which mean one author should take 3.55 steps to any other author on average. The smaller the value is, the more quickly the knowledge is shared by the members of the whole network.
2) The most powerful authors in the network. Using different concepts or methods: degree, closeness, betweeness and cutpoint, we can find who are key scientists in the network. Node 31, 133, 13, 102, 138, 21, 88, etc., are the most important authors in our coauthorship network, but different methods lead to different order of importance, for example, node 31 has a low value of degree but the highest value of closeness and betweeness, it is because the authors connected to author 31 are also very important, therefore, author 31 lies in the center of the whole network even though he doesn't have many collaborators. Actually, the authors 31, 133, 13 are principal investigators in our large project.
3) The evolution of the network. Fig. 1 shows all the authors by 2004, it is also necessary to map the authors and their relations year by year, and find the most important authors and most connected relations in a certain year. As a matter of fact, the results of our analysis are different from year to year. For instance, the strongest relations are between author 13, 31 and 133 from 1997 to 1999, but it change to between 31, 88 and 117 in 2000, and change to between 31,88 and 68 from 2001 to 2003 again. Our purpose is not to get the results, but to know what has caused this to happen, and whether it is good or bad for knowledge sharing and knowledge creation in the project.
4) Research teams. We can divide a large component into many small groups where there are dense connections within the groups and sparse connections between the groups using various kinds of algorithms. We conclude that the subgroups can be used to identify the many research teams in our large project.

3. ACTIVITY NETWORK

Activity network is a set of people and their relations when attending one or more activities together. There are a large number of people from different research institutes or universities in a large project, so various kinds of activities are absolutely beneficial to unite all the members, to encourage them to exchange good ideas, and to help to achieve "what you know is what I know and what I know is what you know".

In activity network, two kinds of scientific collaboration should be discerned. One is the activities within the project, that is, only the members in the project are involved; the other is the activities which are open to all participants, such as domestic or international meetings. The first kind of activities is to improve academic exchanges among the members from different sub-projects within a large project; the second kind of activities is to absorb good or advanced ideas from outside of the project.

In short, activity network exhibits a different function from that of coauthorship network. It spreads more knowledge to much more people. We had 296 activities and 116 participants all together by 2004, after collecting such two kinds of activities and all the participants, similar analysis to coauthorship network can be done, so no more about activity network will be said here.

4. CITATION NETWORK

"If I have seen further, it is by standing on the shoulders of giants." What Newton has said tells us human knowledge and understanding is a very long cumulative affair. Therefore, some good ideas of former scientists may be the basis of scientific innovation of our project.

In order to find out the original ideas in our project, we construct citation network that is set of papers and their relations when one paper refers to many other papers. In our project, the most cited papers is "A new discipline of science: the study of open complex giant system and its methodology", then comes "Open complex giant system". More details are listed in Table 1 below. As to the authors of these papers, "X.S. Qian" is ancestor of systems science in China; he is not a member of our project. "J.Y. Yu", "R.W. Dai" and "J.F. Gu" are principal investigator in charge of our project, and "I. Nonaka", "P.B. Checkland", "Y. Nakamori", "J. Holland", etc. are very famous scientists in the fields of knowledge science, system science, management science and intelligent system. In this sense, we can find which papers have greater influence on our project as original ideas, and who the authors of the papers are, no matter they are members of our project or not, no matter they are domestic or international experts.

Table 1: the papers with higher value of in-degree

No.	The title of papers	The authors' name	in-degree
1	A new discipline of science: the study of open complex giant system and its methodology	X.S. Qian, J.Y. Yu, R.W. Dai	70
2	Open complex giant system	S.Y. Wang, J.Y. Yu, R.W. Dai	25
3	Building System methodology	X.S. Qian	11
4	A technology for organizing management-system engineering	X.S. Qian, G.Z. Xu, S.Y. Wang	9
5	The Wu-li Shi-li Ren-li approach: an oriental systems methodology	J.F. Gu, Z.C. Zhu	9
6	Systems science	G.Z. Xu, J.F. Gu, H.A. Che	8
7	The knowledge-creating Company	I. Nonaka, H. Takeuchi	7
8	Systems thinking, systems practice	P.B. Checkland	6
9	Knowledge management system toward sustainable society	Y. Nakamori	6
10	Hidden order	J. Holland	6
11	Emergence	J. Holland	6
…			

5. KEYWORD NETWORK

Keyword network is a set of keywords and their relations when they appear in the same papers. As we know, the keywords in a paper contain very important information like problem being discussed, methods and methodologies to solve problem, etc., and from the combination of the keywords it is possible to make a guess of what the paper is discussing.

Table 2 shows that frequency of the keywords which tells us how many times a keyword is discussed repeatedly in all the papers. The keyword “Metasynthesis” appears 59 times which is the highest value of frequency, then comes “Complex system” which appears 26 times, etc.

Table 2: Frequency of the keywords

Keyword	Freq.	Keyword	Freq.
Metasynthesis	59	Venture enterprise	14
Complex system	26	CAS	11
HWMSE	25	System science	11
DSS	21	System modeling	9
Complexity	20	Model integration	9
MAS	18	Text mining	8
WSR	16	Electronic common brain	8
Consensus	16	Systems engineering	8
Macro economy	15	Systems methodology	8
Venture capital	14	Web mining	8

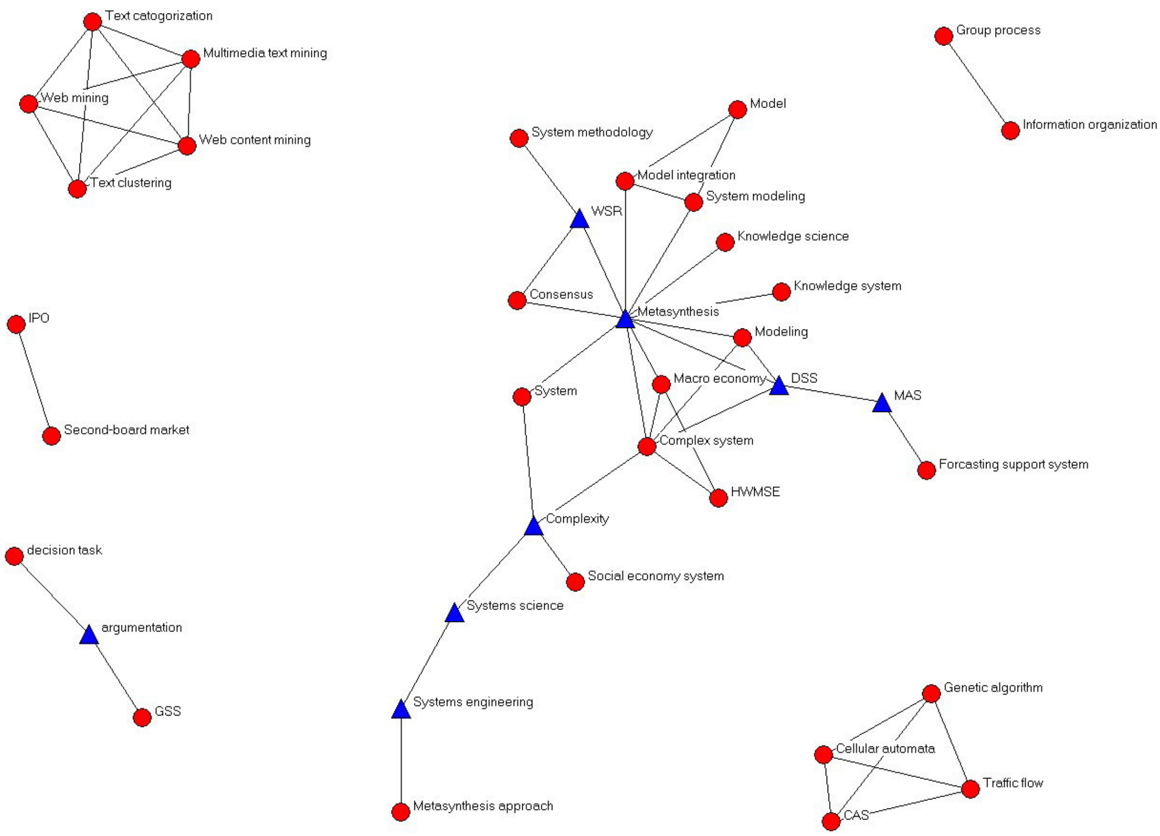

Fig.2: keyword network by 2004

But Table 2 ignores the relations between the keywords which are shown in Fig.2. In the center of the graph is the keyword “metasynthesis”, and “macro economy”, “complex system”, “knowledge science”, “knowledge system”, “consensus”, “WSR”, etc., are connected to it. Here maybe we can get a better understanding that our project focuses on solving macro economy complex system problem using metasynthesis approach on the whole, while a large number of methods, methodologies and theories make a great support to metasynthesis approach.

Another aspect we can know is the ideas growing process from evolution of keyword network with time. That is, each keyword stands for an idea in our project; then new keywords represent new emerging ideas in this year. From the year 1999 to 2004, the number of keyword increase from 21 to 298. The change of keywords with time tells us how such a project incubates new scientific ideas and enables those ideas developed into validated theory.

6. CONCLUDING REMARKS

Differently from traditional data analysis, network analysis pays attention to the relations among the objects. The application of network analysis in this paper allows us to have an insight into a large scientific project after combining quantitatively analysis with computerized visualization: who works with whom, who shares information with whom, then quantitatively identify who are principal contributors in different times, where the original ideas come from, and ideas growing process, etc.

All of our endeavors are supposed to give some support to project manager and project sponsor, further to improve knowledge sharing and knowledge creation in a large project, to manage and organize a large project in a better way. But our efforts is not enough, SECI model [20], *i*-system [21, 22], WSR [23], etc., are playing a great role in knowledge creation, knowing these different models and methodologies is helpful to enhance our research.

When processing our research, we followed the rule of “Confident hypothesis, rigorous validation” proposed by Shi HU, which is quoted as the essential idea of metasynthesis approach by X.S. QIAN[24], that is, we kept in touch with the members and principal investigator of the project constantly. They put forward some good suggestions and requirements to progress our research when we submitted them our reports. At last, a long way is waiting for us, because our research still is at the beginning.

REFERENCES

[1] Barabasi A.L, et al, Evolution of the Social Network of Scientific Collaborations [J], *Physica A 311*, 590-614, 2002.

[2] Barabasi A.L. and Albert R., Statistical mechanics of complex networks[J], *Reviews of Modern Physics 74*, 47-97, 2002.

[3] Newman M E J. Scientific Collaboration Networks: I. Network Construction and Fundamental Results [J], *Phys. Rev. E 64*, 016131, 2001.

[4] Newman M.E.J., Scientific Collaboration Networks: II. Shortest Paths, Weighted Networks, and Centrality [J], *Phys. Rev. E 64*, 016132, 2001.

[5] Newman M.E.J., The structure and function of complex networks [J], *SIAM Review 45*, 167-256, 2003.

[6] Erdos P. and Renyi A., On the evolution of random graphs [J], *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 5, 17-61, 1960.

[7] Hoffman P., The Man Who Loved Only Numbers: The Story of Paul Erdos and the Search for Mathematical Truth [M]. New York: Hyperion, 1998.

[8] Liljeros, F., Edling, C.R., et al., The web of human sexual contacts [J], *Nature 411*, 907-908, 2001.

[9] Milgram S., The small world problem [J]. *Psychology Today*, (2):60--67, 1967.

[10] Watts D.J and Strogatz S.H., Collective dynamics of ‘small-world’ networks [J], *Nature, 393*: 440-442, 1998.

[11] Barabasi A.L. and Albert R., Emergence of scaling in random networks[J], *Science 286*, 509-512, 1999.

[12] Borgatti, S.P., What is social network analysis? 1998. Http://www.analytictech.com/networks/whatis.htm.

[13] Borgatti, S.P., Everett M G, and Freeman L C. UCINET 6.0 Version 1.00. Natick: Analytic Technologies, 1999.

[14] Hanneman R.A., Introduction to Social Network Methods, On-line Textbook[M], 2001, http://faculty.ucr.edu/~hanneman/SOC157/NETTEXT.PDF

[15] Wasserman S, and Faust.K. Social Network Analysis: Methods and Applications [M]. *Cambridge University Press*, New York, 1994.

[16] Rauch J.E., Business and Social Networks in International Trade [J], *Journal of Economic Literature*, Vol. XXXIX, 1177-1203, 2001.

[17] Cross R, Parker A, and Borgatti S.P., Making Invisible Work Visible: Using Social Network Analysis to Support Strategic Collaboration [J]. *California Management Review*. 44(2): 25-46, 2002.

[18] Cross R, Parker A, Prusak L. and Borgatti S.P., Knowing What We Know: Supporting Knowledge Creation and Sharing in Social Networks [J]. *Organizational Dynamics* 30(2): 100-120, 2001.
[19] Cross R.L, Parker A, and Borgatti S.P., A Birds-eye View: Using Social Network Analysis to Improve Knowledge Creation and Sharing [J]. Knowledge Directions, 2(1): 48-61, 2000.
[20] Nonaka, I and Toyama R., The knowledge-creating theory revisited: knowledge creation as a synthesizing process, *Knowledge Management Research & Practice*, 1(1): 2-10, 2003.
[21] Nakamori. Y and Sawaragi Y, Complex Systems Analysis and Environmental Modeling, *European Journal of Operational Research*, 122(2): 178-189, 2000.
[22] Nakamori. Y, Systems Methodology and Mathematical Models for Knowledge Management, *Journal of Systems Science and Systems Engineering*, 12(1): 49-72, 2003.
[23] Gu J.F and Tang X.J., Wu-li Shi-li Ren-li System Approach to a Major Project on the Research of Meta-synthesis System Approach [J]. *International Journal of Knowledge and Systems Sciences*, 1(1):70-77, 2004.
[24] Qian, X.S., Establishing Systematology. Shanxi Science and Technology Press, Taiyuan (in Chinese). 2001.