Self-organized Collaboration Network Model Based on Module Emerging

Hongyong Yang, Lan Lu, and Qiming Liu

School of Computer Science and Technology, Ludong University, Yantai 264025, China
hyyang_ld@yahoo.com.cn

Abstract. Recently, the studies of the complex network have gone deep into many scientific fields, such as computer science, physics, mathematics, sociology, etc. These researches enrich the realization for complex network, and increase understands for the new characteristic of complex network. Based on the evolvement characteristic of the author collaboration in the scientific thesis, a self-organized network model of the scientific cooperation network is presented by module emerging. By applying the theoretical analysis, it is shown that this network model is a scale-free network, and the strength degree distribution and the module degree distribution of the network nodes have the same power law. In order to make sure the validity of the theoretical analysis for the network model, we create the computer simulation and demonstration collaboration network. By analyzing the data of the network, the results of the demonstration network and the computer simulation are consistent with that of the theoretical analysis of the model.

Keywords: Module emerging, Scientific Collaboration, Self-organized, Network model, Complexity.

1 Introduction

The studies of the complex network have opened out the internal essence of many phenomena in real world. Erdos and Renyi presented the random network model (called ER model) [1] since 1960, ER random graph has been used as a basic model to study complex network. In 1998, a paper, written by Watts and Strogatz on small-world network, was appeared in Nature [2] to depict the feature that there exist the shorter paths among nodes in a large network. Another seminal paper was written by Barabasi and Albert on scale-free networks in Science in 1999[3]. The scale-free network (also called BA model) is a network model inspired to the formation of the World Wide Web and is based on two basic ingredients: growth and preferential attachment. The degree distribution of BA model produces the form $p(k) \propto k^{-\gamma}$. Because the power distribution absents distinct characteristic length, this class network is called scale-free network. It has been shown that many real complex networks are small-world and scale-free network, these include transportation networks, phone call networks, Internet, WWW, the actors collaboration networks, scientific coauthorship and citation networks so on [4,5,6].

J. Zhou (Ed.): Complex 2009, Part I, LNICST 4, pp. 798 – 805, 2009.

The presentation of these two original results affords new theory to realize the complex network, and the upsurge of reaching complex network is raised in many science fields. The statistic of the static state parameter is analyzed for Internet, WWW, cinema and television actors' collaboration network, scientists' collaboration network, people relation network, and linguistics network so on. Based on the network model, the dynamics features are studied to infectious disease, percolation model, network searching, and network navigation. Especially, for the research of the coauthorship network in scientific articles, many new results have been obtained recently [7-9]. These researches enrich the realization for complex network, and increase understands for the new characteristic of complex network.

The collaboration network of scientific researcher is a network describing the collaboration relation of scientific researcher. In this network, each node is denoted as a scientific researcher of the network, and each link between two nodes is created if two scientific researchers have published an article together. In this paper, based on the complex network, we establish an evolution network model to study the collaboration relation of scientific researchers.

The remainder of this paper is organized as follows. In section 2, the definition of the module (also called motif) is presented, and a scientific collaboration evolving model is taken on based on the motif emerging. The complex characteristics of the network model are calculated in the section 3, and a computer simulation is created to validate the theoretic results. In section 4, a demonstration network is constructed to show the efficiency of the network model. Finally, some conclusions are made in section 5.

2 Self-organized Network Evolution Model

The scientific cooperation network is used to describe the collaboration relation among the scientific thesis authors. Generally, every scientific researcher is depicted as a node of the network, and the collaboration relation of the people (publishing a paper together) is pictured as the link between the nodes. When analyzing the collaboration network, we found there are a lot of authors in a paper and these authors compose of a full-connected sub-network. Here we call this sub-network as a module (also called motif).

Since there are a lot of the same authors in the different papers, a module is embedding into the scientific collaboration network by the same author. Then a self-organized collaboration network model based on motif emerging is created. Following the analysis, we present a new scientific collaboration network evolution model:

Initially, suppose m_0 articles and k_0 authors in the network. The authors of a paper will compose of a motif; let the sum of the initialized motif be n_0. With the time t changing, we increase an article every time, and author collaboration network evolves as following (Here, we suppose K is the most author number in a paper):

(1) There are i authors in a new paper with a probability q_i, and these authors will group a full-connected sub-network as a motif, where $\sum_{i=1}^{K} q_i = 1$. When $i=1$, there is only one node in the motif, i.e., this paper is written by one author.

(2) There are *j* authors in the motif who are identical as that in the primary network with a probability p_j, where $\sum_{j=0}^{K} p_j = 1$. The new motif will be inset the primary network with the same authors. When *j=0*, the nodes in the motif will have none of the identical nodes, and the motif will set into the network without connected. When *j=K*, all node in the motif have been included in the network, and the nodes will be invariable after the motif inset.

Rule 1: The motif degree of the node is defined as 1 if an author publishes a paper; and the motif degree will add 1 if this author publishes a new paper again, and so on.

Rule 2: The link strength between the nodes is defined as the strength is 1 if two nodes are linked firstly; and the strength adds 1 if these two nodes are linked again, and so on.

Rule 3: The node is selected to inset the network, following the method of motif degree preference, the selecting probability

$$\Pi(i) = n_i / \sum_j n_j \tag{1}$$

where the numerator n_i as the motif degree of node *i*, which is the summation of the paper of node *i*, the denominator is the summation of the motif degree of all nodes in the network.

3 Degree Distribution of Network Model

3.1 Analysis on the Characteristics of the Network Model

In this section, we will research the complex feature of the evolution model. In the scientific collaboration, one author can publish several papers, i.e., the module degree of one node can be very large. One author can cooperate with many authors, i.e., the link strength of the node can be very large. In the description of the network, the degree distribution *p(k)* is used to denote the probability of a node with degree *k*. Next, we calculate the module degree distribution and the strength distribution of the node in the new evolution model.

The module degree of a node is the number of the paper published by the author. Following the algorithm description, the new module insets the network connected with *j* nodes that are decided by the probability p_j. The diversification of the module degree of the node *i*

$$\frac{\partial n_i}{\partial t} = \sum_{l=1}^{K} q_l \sum_{j=0}^{l} p_j j \Pi(i) . \tag{2}$$

Suppose *l*=1, ……,K, and let $\sum_{j=0}^{l} p_j j = J_l$, we have

$$\sum_{l=1}^{K} q_l J_l = \bar{J} .$$

Since the sum of the module degree satisfying

$$\sum_j n_j = n_0 + t(\sum_{l=1}^{K} q_l l), \tag{3}$$

Let $\sum_{l=1}^{K} q_l l = L$, when t is large enough, we get

$$\frac{\partial n_i}{\partial t} = \bar{J}\Pi(i) = \frac{\bar{J}}{L}\frac{n_i}{t}. \tag{4}$$

From $n_i(t_i) = 1$ at t=t, we obtain

$$n_i(t) = (\frac{t}{t_i})^{\beta}, \tag{5}$$

where $\beta = \bar{J}/L$. We can calculate the module degree distribution using the method in [4,5]

$$p(n) = \frac{t}{(n_0 + t)\beta} n^{-(\frac{1}{\beta}+1)} \propto n^{-(\frac{1}{\beta}+1)}. \tag{6}$$

Now we will calculate the strength distribution of the node. In the scientific collaboration network, if two scientific researchers publish one paper together, the link strength of the nodes is 1 with a line between two authors. If these two researchers collaborate once again, the link strength will increase 1. The strength of a node is equal to the sum of the link strengths of its neighbors. Following the evolving process of the model, the diversification of the strength of the node will satisfy

$$\begin{aligned}\frac{\partial s_i}{\partial t} &= \sum_{l=1}^{K} q_l \sum_{j=0}^{l} p_j j(l-1)\Pi(i) \\ &= \sum_{l=1}^{K} q_l J_l (l-1)\Pi(i) \\ &= (M - \bar{J})\Pi(i),\end{aligned}$$

where $M = \sum_{l=1}^{K} q_l J_l l$, since

$$\frac{\partial n_i}{\partial t} = \bar{J}\Pi(i),$$

we have

$$\frac{\partial s_i}{\partial t} = \frac{M - \bar{J}}{\bar{J}} \frac{\partial n_i}{\partial t} .$$

From Eq.(5), we obtain

$$s_i = \frac{M - \bar{J}}{\bar{J}} n_i = \frac{M - \bar{J}}{\bar{J}} (\frac{t}{t_i})^{\beta} , \tag{7}$$

where $\beta = \bar{J} / L$. Let $A = \frac{M - \bar{J}}{\bar{J}}$, We can calculate the strength distribution

$$p(s) = \frac{tA^{\frac{1}{\beta}}}{(n_0 + t)\beta} s^{-(\frac{1}{\beta}+1)} \propto s^{-(\frac{1}{\beta}+1)} . \tag{8}$$

From the theories analysis, it can be obtained that the module degree distribution of the node has the same power law as the strength degree distribution, and the scale-free feature is put up by the module degree and strength degree.

3.2 Simulation

Next, we apply the computer simulation to study the complex characteristic of the network model. Suppose *K=4*, $q_1 = 0.2$, $q_2 = 0.3$, $q_3 = 0.3$, $q_4 = 0.2$;

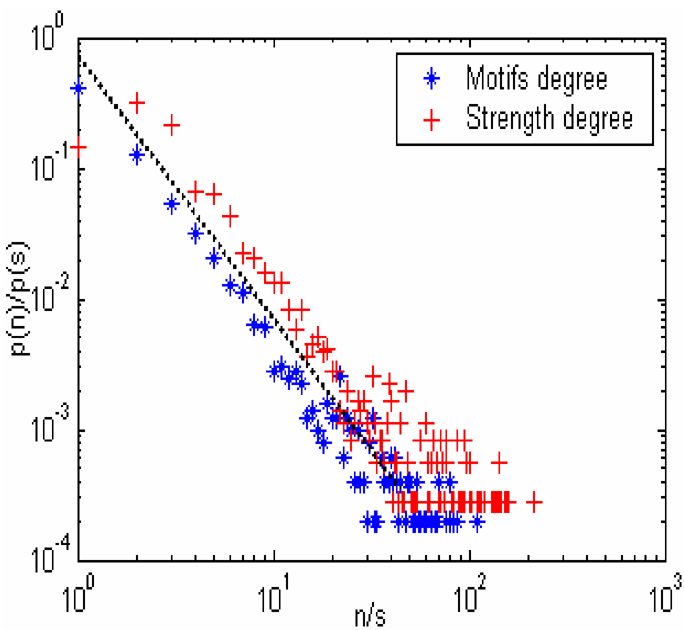

Fig. 1. The distribution plots (log-log plots) of the model

and $p_0 = 0.2$, $p_1 = 0.15$, $p_2 = 0.25$, $p_3 = 0.25$, $p_4 = 0.15$. In the initialized network, there are 10 authors with 3 paper published. Based on the network evolving model, a scientific collaboration network with 500 papers is built by the computer simulation, and the motif degree distribution and the strength degree distribution are calculated. The distribution plots (Log-Log) are shown Fig. 1 whose data are obtained from the average of 20 times simulations. These two distributions have the same power law and take on the feature of the scale-free network, which is accord with the theoretical results.

4 Analysis on Demonstration Network

We have been working a statistic of the papers in Journal of Information from January 2001 to December 2006 in database of China National Knowledge Infrastructure (CNKI). A coauthorship database is created on the articles and their authors; there are 801 articles and 1078 authors in the database. A scientific collaboration network is made from the coauthorship database, where a node denotes au author, link of two nodes denotes the cooperation between two authors (they have vended an article together). The module is defined as the collaboration authors in an article, which is full connected sub-network. It is easy to know that the cooperation network is a sort of self-organized network. The information of the database is shown in Table 1.

Table 1. The information of the database of the Journal of Information

Author Number	1	2	3	≥4
Paper number	223	325	212	41
Percent	27.8%	40.6%	26.5%	5.1%

Then, we analyze the characteristic of this network by calculating the module degree of the node in the coauthorship network, which is the paper number of every author. The statistic result is shown in Table 2, where a lots of nodes have small module degree (there are 1018 authors with module degree less 3) and few nodes with large module degree (only 6 authors with module degree greater than 10).

Table 2. The module degree of the authors

Module degree	1	2	3	4	5	6	7	8	9	10	11	12	15	22
Number of the authors	805	157	56	23	9	8	4	5	5	1	1	2	1	1

In the demonstration network, we calculate the strength degree of the node, i.e., the cooperation times with other authors (Table 3). There are many nodes with little strength degree (792 authors with module degree less than 3), however, there exists a larger node with strength degree 37 who has a very large workgroup and has more cooperation researchers.

Table 3. The strength degree of the authors

Strength Degree	0	1	2	3	4	5	6	7	8	9	10	11
Authors Number	135	340	317	107	79	31	16	11	10	4	4	4

Strength Degree	12	13	15	16	17	20	30	37
Authors Number	5	3	4	2	3	1	1	1

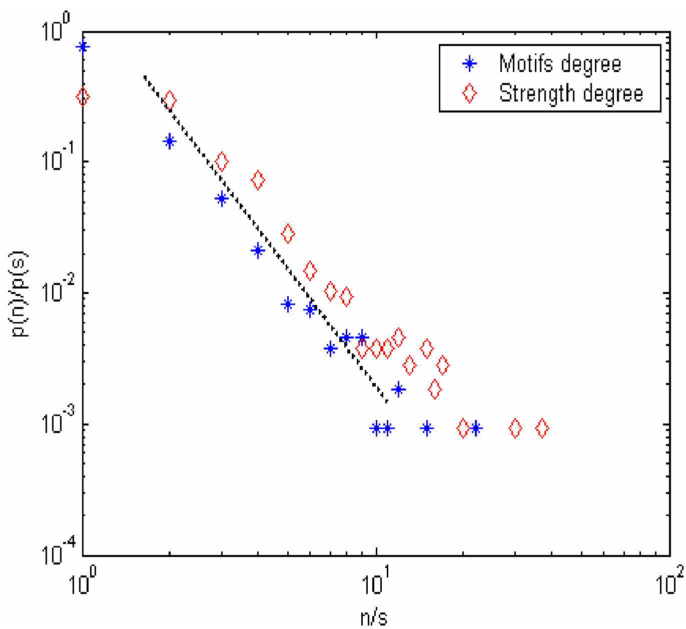

Fig. 2. The plot of the distributions in the demonstration network (Log-Log)

Based on the results of statistic, a plot of the distributions is drawn by using the data of Table 2 and Table 3 with log-log figure, where "*" defines the motif degree distribution of node and "◇" defines the strength degree of the node in the demonstration network. In Fig. 2, the power law of the motifs degree is the same as that of the strength degree, which is consensus with the results in section 3.

5 Conclusions

In this paper, based on the collaboration ways of the authors, a self-organized collaboration network model is presented. By defining the module with the authors sub-network full-connected in a paper, the network is evolved by inset the module. It is validated by the theoretical analysis and computer simulation that the distributions of the module degree and strength degree satisfy the same power law and this network is a scale-free network. By analyzing the data of the demonstration network in CNKI, the demonstration results are consistent with that of the theoretical analysis. Therefore, this model can be applied to study the evolvement of the author collaboration network.

Acknowledgements

This work is supported by Chinese National Natural Science Foundation (under the grant 60774016, 60875039), the Science Foundation of Office of Education of Shandong Province (under the grant J08LJ01) and the Natural Science Foundation of Ludong University (under the grant L20074102, Z0704) of China.

References

1. Erdos, P., Renyi, A.: On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci. 5, 17–60 (1960)
2. Watts, D., Strogatz, S.: Collective dynamics of "small world" networks. Nature 393(4), 440–442 (1998)
3. Barabási, A., Albert, R.: Emergence of scaling in random networks. Science 286(5439), 509–512 (1999)
4. Albert, R., Barabási, A.: Statistical mechanics of complex networks. Rev. Modern Phys. 74(1), 47–97 (2002)
5. Newman, M.: The structure and function of complex network. SIAM Review 45(2), 167–256 (2003)
6. Boccaletti, S., Latora, V., Moreno, Y., et al.: Complex Networks: Structure and dynamics. Physics Reports 424(4-5), 175–308 (2006)
7. Newman, M.: Scientific collaboration. I. Network construction and fundamental results. Phys. Rev. E 64(1), 016131 (2001)
8. Newman, M.: Scientific collaboration. II. Shortest paths, weighted networks, and centrality. Phys. Rev. E 64(1), 016132 (2001)
9. Wang, F.-S., Yang, H.-Y.: Analysis on Scientific Collaboration Network of Author. Journal of the China Society for Scientific and Technical Information 26(5), 659–663 (2007)