A New Scale Free Evolving Network Model with Community Structure

CAI Jun
School of Electronic and Information
Guang Dong Polytechnic Normal University
Guangzhou, China
e-mail: gzhcaijun@gmail.com

LIU Wai Xi
Department of Electronic and Information Engineering
Guang Zhou University, Guangzhou, P. R. China
Guangzhou, China
e-mail: liuwaixi@gmail.com

***Abstract*—Understanding and modeling the structure of a complex network can lead to a better knowledge of its evolutionary mechanisms, and to a better cottoning on its dynamic and functional behavior. The nodes within a network not only tend to connect the nodes with high degree (scale-free property), and tend to connect with their relatively close distance nodes (community structure property), and the high-degree nodes are easier to connect with their relatively far nodes comparing with the low-degree nodes in the process of network evolution. This phenomenon has been commonly observed from many real-world networks, ranging from social to biological even to technical networks. To model this kind of networks, the present letter proposes a scale free network model with community structure (SFC) to capture and describe their essential topological properties. Numerical simulations indicate that the generated network based on SFC model has scale-free and community structure property. Under the control of the parameters of the model, the community structure of network can be adjustable.**

Keywords- Scale-free networks; community structure; network model; complex network.

I. INTRODUCTION

In the past 10 years, complex network research has made great progress and has become one of the most important areas of interdisciplinary research[1,2]. Many real networks can be described as complex network, for example: biological networks (such as metabolic networks, protein - protein networks, genetic networks, etc), transport network (such as airport networks, urban transport networks, railway networks, etc.) and information networks (such as WWW networks, Internet networks, telephone call networks, etc.). These research results show that modern complex network research has deepened our understanding of these complex systems.

To model these systems and capture the structure properties observed in real networks, a lot of work has been done in this field. Small world and scale-free are two common properties of complex network, in order to represent this two properties, Watts Strogatz and Barabási respectively proposed small-world network model (denoted as W-S model) [3] and B-A scale-free Network model (denoted as B-A model)[4]. Recently, further studies showed that the community structure[5] is one of the most important topology properties following the small world and scale-free properties of complex network. Communities are defined as collections of nodes within which connections are denser, but among which connections are sparser. There are many real-world networks which exhibit community structure, and community structure is supposed to play an import role in many real networks, for example: communities in social network represent real social groups based on interest or background; communities in a citation network might represent related papers on a single topic; communities on the web might mean pages on related topics; communities in a biochemical network or neuronal system might correspond to functional units of some types; communities in electronic circuit can be a functional unit; communities in information network also has an important role[6]. However, the current research focusing on the community structure is how to design algorithms to quickly and accurately detect community structure in static or dynamic complex networks [7].

To date, many network models can generate scale-free features based on different ideas and mechanisms[7,8,9]. However, the community structure of most existing evolving network models is unobvious. The community feature about the W-S model and B-A model is discussed, and the result showed W-S model has good community structure, while the community structure of B-A model is unobvious18. How to construct an evolving network model with community structure is an attractive question. Several evolving models with community structure have been proposed in social networks, biological networks and polymer melts networks. Watts et al.[10] and Motter et al.[11] proposed some good network models, but with a fixed number of nodes therefore not evolving. M.Kimura et al.[12] proposed a growing network model with community structure, however, it turned out that there is a possibility that a node belonging to a community may have no connections with other nodes in the same community but only has connections with nodes in other communities, which is unacceptable. More recently, Ch. Li et al.[13] proposed a community-based evolution model considering that the connection probability among communities is less than one within communities. Following this theory, J. Zhang et al.[14] proposed a polymer melts model with community structure. Riitta.T et al.[15] introduced and analyzed a community-based evolving network model in order to demonstrate this phenomenon of “richer get richer” in communities scale, as well as the mechanism of community size preferential attachment which means when establishing new links between communities or adding a new node to an existing community, communities with larger sizes are selected with higher probabilities. Fan et al[16]

© the authors

developed a multi-community weight-driven bipartite network model. A common feature of these four models is that the network evolves with time but the number of communities in the network is always constant. However, recent research[17] has shown that the number of communities may also change as the network evolves, furthermore, almost all of these models with community grow from a small network with obvious community structure, but the community structure in real networks is formed in the process of network evolution. From the above analysis, one clearly needs a better network model that can precisely describe the topology of an evolving network with a community structure and scale free property.

In this paper, a new evolving networks model (denoted as SFC model) is proposed with three steps: the addition of new nodes, new links, and the rewiring of links. The preferential connection with high degree nodes is adopted in the process of the new node addition. The preferential connection with close nodes is applied in the process of the new link addition and the rewiring of link, which is closer to the evolving of real network, and the high-degree nodes are easier to connect with their relatively far nodes comparing with the low-degree nodes. The community structure in SFC model can be automatically formed in process of the evolution of the model without setting the number of communities in advance. Simulation results show that SFC model not only has obvious community structure and scale-free properties, and its community structure can be adjustable by changing the model generating parameters.

The rest of this paper is organized as follows: In section Ⅱ, we propose a new scale free evolving network model with community structure. In section Ⅲ, the numerical results about node degree distribution and community structure are analyzed. The whole paper is finally concluded in section Ⅳ.

II. MODEL DESCRIPTION

The generating algorithm for the proposed SFC model is initialized with m_0 isolated nodes, and at each time step we perform one of the following three operations.

(ⅰ)With probability p_1 we add a new node: The new node has m new links that with probability $\prod(k_i)$ are connected to node i already present in the network, which is the same with B-A model .

$$\prod(k_i) = \frac{k_i + 1}{\sum_j (k_j + 1)} \tag{1}$$

Where k_i is the degree of node i .

(ⅱ)With probability p_2 we add m ($m \le m_0$) new links: For this a node i is randomly selected as the starting point of the new link. If i has no neighbor, nothing will be done. Otherwise we do the following: The other end j' of the link is selected with probability $\prod(d_{j'})$ given by Eq. (2)

$$\prod(d_{j'}) = \frac{d_{ij'}^{-(1+\delta_i \lg d_{ij'})}}{\sum_{j'} d_{ij'}^{-(1+\delta_i \lg d_{ij'})}} \, ,$$

$$(d_{ij'} \neq 1 , \quad \delta_i = 1 - \frac{k_i}{k_{\max}} , \ \delta_i \in [0,1]) \tag{2}$$

Where d_{ij} is the shortest distance between node i and the node j . $k_{\max}$ is the maximum degree in network. Eq. (2) incorporates the fact that new links are more likely to be generated among close nodes, and the high-degree nodes (popular nodes) are easier to connect with their relatively far nodes comparing with the low-degree nodes in the process of network evolution. This process is repeated m times.

(ⅲ)With probability $1 - p_1 - p_2$ we rewire m links: For this we randomly select a node i and cut off randomly one link l_{ij} connected to i and then rewire i to j' . j' is a node selected with probability $\prod(d_{j'})$ given by Eq. (2). This process is repeated m times.

In the SFC model, the parameters should satisfy $0 \le p_1 \le 1$ and $0 \le p_2 \le 1 - p_1$. Repeat step (ⅰ)-(ⅲ) until the network has grown to desired size. The network evolution is illustrated through a simple example in Fig.1. The new vertex V possibly links to popular nodes (here M , N), and the vertex W possibly add new links to its neighbors (here adjacent nodes are h , i , j , k) and the vertex V possibly rewire links to its neighbors (here adjacent nodes are e , f , g) if the link l_{VM} is removed. Roughly speaking, the mechanism of adding new links and rewiring links contribute to the formation of communities in network.

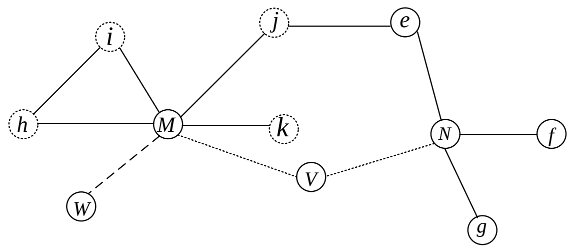


Figure 1. Growth process of the network

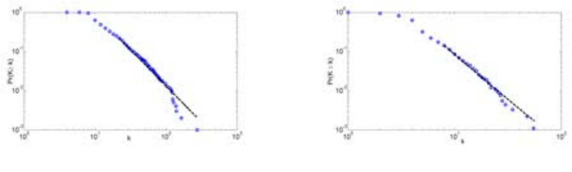
III. NUMERICAL RESULTS AND ANALYSIS

A. Node degree distribution

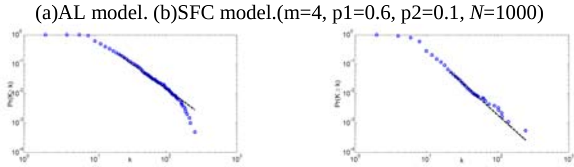
Owing to utilizing node degree to choose the model parameter δ , it is difficult to theoretically directly analyze the node degree distribution of SFC model. In this section, we numerically study the node distribution properties of SFC model through comparing it with AL model[18] that considers global degree preferential attachment mechanism in addition of new links and rewiring of links. First, we perform a numerical simulation to separately generate many different networks based on SFC model and AL model under different model parameter (m , p_1) and network size (N). Effect of network size is studied by varying N , when model parameter

© the authors

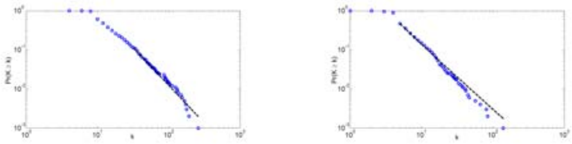
is unchanged. Effect of model parameter is studied by varying m and p_1, when network size is unchanged. Our simulation work shows that the node degree distributions are approximate power-law, which exponential is only affected by model parameter p_1. The degree distribution of three group networks selected from SFC model and AL model under different network size and model parameter is shown in fig. 2..The model parameter in one group is set as $m_0 = 4$, $m = 4$, $p_1 = 0.6$, $p_2 = 0.1$, $N = 1000$ and $m_0 = 4$, $m = 4$, $p_1 = 0.6$, $p_2 = 0.1$, $N = 2000$. Another is set as $m_0 = 4$, $m = 4$, $p_1 = 0.8$, $p_2 = 0.1$ and $N = 1000$. The third is set as $m_0 = 4$, $m = 2$, $p_1 = 0.8$, $p_2 = 0.1$ and $N = 1000$. The networks based on AL model are shown in fig.2 (a), (c), (e) and (j), whose cumulative degree distributions obey power-law distribution $P(k) \propto k^{-\gamma}$ with $\gamma \approx 2.82$, $\gamma \approx 2.75$, $\gamma \approx 2.92$ and $\gamma \approx 2.77$ respectively. The networks based on SFC model are shown in fig.2 (b), (d), (f) and (h), whose cumulative degree distributions also obey power-law distribution $P(k) \propto k^{-\gamma}$ with $\gamma \approx 3.22$, $\gamma \approx 3.17$, $\gamma \approx 3.02$ and $\gamma \approx 2.96$ respectively. The degree distribution shape generated from SFC model and AL model is similar, which is approximate power-law distribution. The difference is that the slope of node distribution in SFC model is slightly larger than one in AL model under the same model parameter. Similar results are obtained from other networks generated from SFC model and AL model under different model parameters and network size. The main reason for this phenomenon is that the rewiring and addition link among "old" nodes in SFC model is based on the distance among nodes, not on degree preferential attachment mechanism, which lead to the percentage of relatively large degree in network generated form SFC model is less than that from AL model.



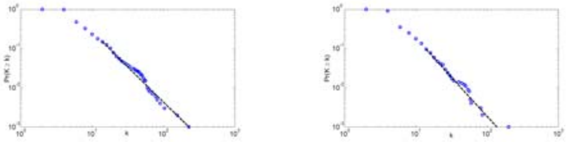
(a)AL model. (b)SFC model.(m=4, p1=0.6, p2=0.1, *N*=1000)



(c)AL model. (d)SFC model.(m=4, p1=0.6, p2=0.1, *N*=2000)



(a)AL model. (b)SFC model.(m=4, p1=0.8, p2=0.1, *N*=1000)



(j)AL model. (h)SFC model.(m=2, p1=0.8, p2=0.1, *N*=1000)

Figure 2. Cumulative degree distributions of SFC model and AL model.

B. Modularity of network model

In order to quantify the community structure, we utilize the modularity method of Newman[19]. Let $A = (a_{ij})$ be a $k \times k$ symmetric matrix that denotes network with n nodes and k communities. Thus

$$A_{uv} = \begin{cases} 1 & \textit{if community u and v are connected,} \\ 0 & \textit{otherwiese,} \end{cases}$$

V_r and V_s respectively represent the number of nodes in community r and s. We define $\|A\| = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}$, $\|A_{sr}\| = \sum_{i\in V_s}\sum_{j\in V_r} e_{ij}\ (i \neq j)$, $\|A_{rr}\| = \sum_{i\in V_r}\sum_{j\in V_r} e_{ij}$ and $e_{rs} = \|A_{rs}\| / \|A\|$, then the modularity Q is to be

$$Q = \sum_{r=1}^{k}[(e_{rr} - (\sum_{s=1}^{k} e_{rs})^2] \qquad (3)$$

Basically, Q is the fraction of all links within communities subtracts the expected value of the same quantity in a graph whose nodes have the same degrees but links are distributed randomly, and the higher modularity *Q*, the better network community structure. The maximum of *Q* is 1. It is also found that *Q* value above 0.3 is an indicator of good community structure in a network.

The modularity of SFC model changes with the model parameters *m* and the network size *N*, as is shown in fig.3. As *m* value increases, the network density increases, *Q* value in network gradually become small and the community structure become unclear. It indicates that the community structure characteristic of SFC model is affected by network density. For the same *m* value and different network size, the same law can also be found, but the impact of network sizeN is not so apparent. Thus, *m* is one of the important parameters that affect the community structure of SFC model.

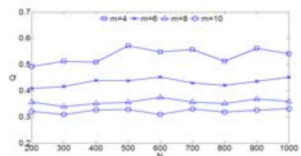


Figure 3. The modularity of SFC model with different m and N.

In this experiment, we study the impact of the process choose parameter p_1 and p_2 on community structure. We perform a numerical simulation with $m_0 = 4$, $m = 4$ and the network size $N = 1000$ under different p_1 and p_2. The results show that the modularity of SFC model is also rather sensitive to the model generation parameter p_1 and p_2, as is shown in fig.4. If one selects $p_1 = 1$, $p_2 = 0$, then the model is just the B-A model. For the B-A model, with the increase of network size, the mechanism that links preferentially point to popular nodes leads to the nature of the node connection with random choice. So the community structure becomes unobvious, the Q value is below 0.3 in fig.4. However, the emergence of community structure in the networks generated by SFC model can be contributed to the mechanism of rewiring and addition link among "old" nodes. With the decrease of p_1, Q in network increases gradually, which is shown in fig.4. These results make sense because p_1 and p_2 are crucial for the SFC model to change the nature of the node connection with random choice. As a result, we can generate different community structure networks applying to different real world network circumstance by changing p_1 and p_2 values.

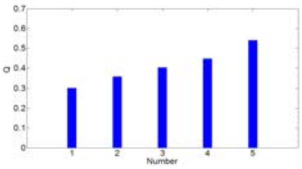


Figure 4. The Modularity under the cases with p_1 and p_2 value at $p_1$and p_2 are (p_1=1, p_2 =0) , (p_1=0.8, p_2 =0.1),(p_1=0.7, p_2 =0.2),(p_1=0.6, p_2 =0.1) and (p_1=0.5, p_2 =0.1).

IV. CONCLUSIONS

In this paper, we have proposed a model which combines degree preferential attachment mechanism and the rewiring and addition link mechanism based on the distance among nodes. It can produce very efficiently networks resembling real-world networks in that they not only have power-law distribution, but good community structure, which is capable of generating different networks from social to technological fields through selecting different model parameters. Based on this network model, we can study the effects of scale free and community structure on network dynamics, such as the stability, synchronization, disease and rumor spreading, and information communication or transfer.

However, in this paper, we just analyze the degree distribution and the community structure of the SFC model. Future extensions of this work should include studying other parameters of SFC model, such as: degree-degree correlations, clustering coefficient and average path length. Future extensions work also include the modeling of directed and weighted network models with scale-free and community structure, because many real-world networks are directed and /or weighted.

ACKNOWLEDGEMENTS

We would like to thank all the members in our research group for the valuable discussions about the ideas presented in this paper. This work is funded by the National High Technology Research and Development Program of China (Grant No.2007AA01Z449), the Key Program of NSFC-Guangdong Joint Funds under Grant No.U0735002 and the National Natural Science Foundation of China under Grant No. 61202271.

REFERENCES

[1] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez et al., Phys. Rep. 424, 4 (2006).
[2] Andrew J. Evans, Complexity16, 2(2010).
[3] D.J. Watts, S.H. Strogatz, Nature 393, 440(1998).
[4] A.L. Barabási, R. Albert, Science 286, 5439(1999).
[5] M.E.J. Newman, SIAM Rev. 45, 2(2003).
[6] M.Gustafsson, M.Hörnquist, A. Lombardi, Physica A 367, 15(2006).
[7] S.Fortunato, Phys. Rep. 486, 3(2010).
[8] Y.Y Gu, J.T. Sun, Phys. Lett. A 372, 4564(2008).
[9] D. Hai-Feng, L.Shu-Zhuo, W.F.Marcus, Chin.Phys.Soc. 56, 12(2007).
[10] D.J.Watts, P.S. Dodds, M.E.J. Newman, Science 296, 1302(2002).
[11] A.E. Motter, T.Nishikawa, Y.C.Lai, Phys. Rev.E 68, 036105(2003).
[12] M.Kimura,K.Saito,N.Ueda, Neural Networks 17, 7(2004).
[13] Li, Ch., Maini, Ph.K, J. Phys. A: Math. Gen. 38,975 (2005).
[14] J.Zhang, H.Huang, Eur.Phys.J.B. 68, 247(2009).
[15] Riitta.T, Jukka-Pekka.O, Jari.S et al., Physica A: Statistical and Theoretical Physics 371, 2(2006).
[16] H.Fan, et al., Phys.Rev.E 78, 026103(2008).
[17] G. Palla, A.L. Barabási, T.Vicsek, Nature 446, 664(2007).
[18] R.Albert, A.L. Barabási, Rev.Mod.Phys. 74, 1(2002).
[19] Newman M E J and Grivan M, Phys. Rev. E 69, 026113(2004).

© the authors