An evolving network model with community structure

Chunguang Li[1,2] **and Philip K Maini**[2]

[1] Centre for Nonlinear and Complex Systems, School of Electronic Engineering,
University of Electronic Science and Technology of China, Chengdu 610054,
People's Republic of China
[2] Centre for Mathematical Biology, Mathematical Institute, University of Oxford,
Oxford OX1 3LB, UK

E-mail: cgli@uestc.edu.cn

Abstract
Many social and biological networks consist of communities—groups of nodes
within which connections are dense, but between which connections are sparser.
Recently, there has been considerable interest in designing algorithms for
detecting community structures in real-world complex networks. In this
paper, we propose an evolving network model which exhibits community
structure. The network model is based on the inner-community preferential
attachment and inter-community preferential attachment mechanisms. The
degree distributions of this network model are analysed based on a mean-
field method. Theoretical results and numerical simulations indicate that this
network model has community structure and scale-free properties.

PACS numbers: 89.75.Hc, 02.50.−r

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Complex networks are currently being studied across many fields of science and engineering
[1], stimulated by the fact that many systems in nature can be described by models of complex
networks. A complex network is a large set of interconnected nodes, in which a node is a
fundamental unit usually with specific dynamical or information content. Examples include the
Internet, which is a complex network of routers and computers connected by various physical
or wireless links; the World Wide Web, which is an enormous virtual network of web sites
connected by hyperlinks; and various communication networks, food webs, biological neural
networks, electrical power grids, social and economic relations, coauthorship and citation
networks of scientists, cellular and metabolic networks, etc. The ubiquity of various real and

artificial networks naturally motivates the current intensive study of complex networks, on both theoretical and application levels.

Many properties of complex networks have currently been reported in the literature. Notably, it is found that many complex networks show the small-world property [2], which implies that a network has a high degree of clustering as in some regular network and a small average distance between nodes as in a random network. Another significant recent discovery is the observation that many large-scale complex networks are scale free. This means that the degree distributions of these complex networks follow a power-law form $P(k) \sim k^{-\gamma}$ for large network size, where $P(k)$ is the probability that a node in the network is connected to k other nodes and γ is a positive real number determined by the given network. Since power laws are free of characteristic scale, such networks are called 'scale-free networks' [3]. The scale-free nature of many real-world networks can be generated by a mechanism of growing with preferential attachment [3].

Communities are defined as collections of nodes within which connections are dense, but between which connections are sparser. There are many real-world networks which exhibit community structure, and community structures are supposed to play an important role in many real networks. For example, communities in a citation network might represent related papers on a single topic [4]; communities on the web might represent pages on related topics [5]; communities in a biochemical network or neuronal system might correspond to functional units of some kinds [6, 7]; and communities also play an important role in information networks [8]. There have been several investigations into designing algorithms for detecting community structure in large-scale complex networks [9, 10]. To study the effects of community structure on network properties and dynamics, the modelling of real networks with community structure is very important. However, most of the existing evolving network models do not take the community structure into account. In [11], a networked seceder model was proposed to illustrate group formation in social networks. In [12], a bipartite growing network model for social community with group structures was proposed. In [13], a social network model based on social distance attachment was proposed, which can exhibit community structure. In [14], the authors proposed a growing network model with community structure. However, there is the possibility within the network model that a node belongs to a community but has no connections with nodes in this community but has connections with nodes in other communities, which is unacceptable.

In this paper, we propose an evolving network model with community structure based on the inner-community preferential attachment and inter-community preferential attachment mechanics. We use a mean-field method to analyse the degree distributions of this network model. Numerical simulations are also performed to investigate the properties of this network model. Some more realistic generalizations and extensions are also discussed in section 4.

2. Network model

In this section, we describe the growing mechanics of the proposed network model. For simplicity, we consider only undirected network models in this paper. We assume that there is a total of M ($M \geqslant 2$) communities in the network. The proposed model is defined by the following scheme.

Step 1: initialization

Start from a small number m_0 ($m_0 > 1$) of fully connected nodes in each community. Use $\frac{M(M-1)}{2}$ inter-community links to connect each community to the other $M - 1$ communities,
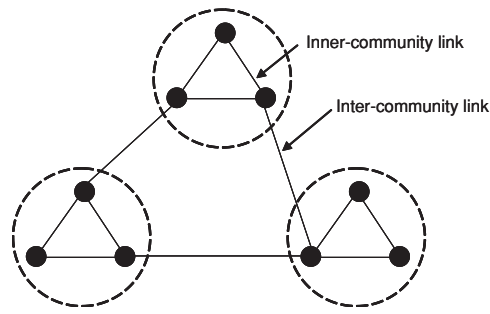
Figure 1. An initial network example with $M = 3$ and $m_0 = 3$.

so that there is a link between each community. The nodes to which the inter-community links connect are selected randomly in each community. For example, figure 1 shows an initial network with $M = 3$ and $m_0 = 3$.

Step 2: growth

At each time step, a new node is added to a randomly selected community. The new node will be connected to m $(1 \leqslant m \leqslant m_0)$ nodes in the same community through *m inner-community links* (defined as the links that connect nodes in the same community), and with probability α connected to n $(1 \leqslant n \leqslant m)$ nodes (none with probability $1 - \alpha$) in the other $M - 1$ communities through *inter-community links* (defined as the links that connect nodes among different communities).

Step 3: preferential attachments

(a) *Inner-community preferential attachment.* When choosing nodes in the same community (we denote it as the jth community) to which the new node connects through inner-community links, we assume that the probability Π that a new node will be connected to node i in community j depends on the inner-degree s_{ij} (defined as the number of inner-links connected to node i) of that node, such that

$$\Pi(s_{ij}) = \frac{s_{ij}}{\sum_k s_{kj}}. \tag{1}$$

(b) *Inter-community preferential attachment.* When choosing the nodes in other communities to which the new node connects through inter-community links, we assume that the probability Π that a new node will be connected to node i in community k $(k \neq j)$ depends on the inter-degree l_{ik} (defined as the number of inter-links connected to the node), such that

$$\Pi(l_{ik}) = \frac{l_{ik}}{\sum_{m,n,n \neq j} l_{m,n}}. \tag{2}$$

The motivation for using the inner-community preferential attachment mechanics is the same as that in many other growing network models. There also exist inter-community preferential attachment phenomena in some real networks with community structure. For example, in scientific collaboration networks, a multidisciplinary researcher is more likely to be willing to study other unknown fields to him/her, and has stronger desire to collaborate
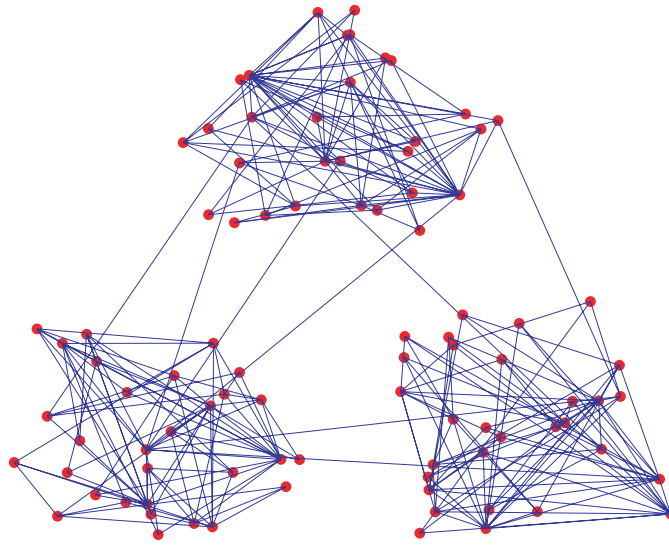
Figure 2. A network generated by the proposed scheme: $N = 90, M = 3, m_0 = 3, m = 3,$ $\alpha = 0.1$ and $n = 1$ (the nodes in each community are randomly placed).

with researchers in other fields than other single-topic researchers. In friendship networks, an individual with many inter-community links is more likely to make friends with different kinds of people, and has higher probability of making new friends with people in other communities than other people with less inter-community links. These are all inter-community preferential attachment phenomena.

After t time steps, this scheme generates a network with $Mm_0 + t$ nodes, and $[Mm_0(m_0 - 1) + M(M - 1)]/2 + mt + \text{integer}(\alpha nt)$ links in the sense of mathematical expectation. The parameters α and n control the ratio between inter- and inner-community links.

We performed a numerical simulation with a total of $N = 90$ nodes, $m_0 = m = 3, n = 1$ and the probability $\alpha = 0.1$, that is, at each time step, we connect a new node with three nodes in a selected community and with probability $\alpha = 0.1$ connect it to a node in another community. For the purposes of clarity, the generated network is shown in figure 2, in which the nodes in each community are randomly placed. As we can see, the network generated by the proposed scheme exhibits community structure. A detailed study of the properties of this network model will be presented in the next section.

3. Network properties

In this section, firstly we analytically calculate the degree distribution $P(k)$, which is defined as the probability that a randomly selected node has degree k (a node has k connections), and then we numerically study some properties of this network model.

We use a mean-field method [15] to analyse the scaling property of the network model. Firstly we analyse the inner-degree distribution, and then extend the results to that of the inter-degree and the total degree (defined as the sum of the inner-degree and inter-degree). Similar to [15], we assume that s_{ij} is continuous, and thus the probability $\Pi(s_{ij}) = s_{ij} / \sum_k s_{kj}$ can be

interpreted as a continuous rate of change of s_{ij}. Consequently, for a node i in community j, we have

$$\frac{\partial s_{ij}}{\partial t} = \frac{1}{M} m \frac{s_{ij}}{\sum_k s_{kj}} \tag{3}$$

and, noting that $\sum_k s_{kj} = 2mt \frac{1}{M} + m_0(m_0 - 1) \approx 2mt \frac{1}{M}$ for large t in the sense of mathematical expectation, we have

$$\frac{\partial s_{ij}}{\partial t} \approx \frac{s_{ij}}{2t}. \tag{4}$$

The solution of this equation, with the initial condition that node i in community j was added to the system at time t_i with connectivity $s_{ij}(t_i) = m$, is

$$s_{ij}(t) \approx m \left(\frac{t}{t_i} \right)^{0.5}. \tag{5}$$

Using (5), the probability that a vertex has a connectivity $s_{ij}(t)$ smaller than k can be written as

$$P(s_{ij}(t) < k) = P \left(t_i > \frac{m^2 t}{k^2} \right). \tag{6}$$

Assuming that we add the nodes (including the initial nodes) at equal time intervals to the network, the probability density of t_i is

$$P_i(t_i) = \frac{1}{Mm_0 + t}. \tag{7}$$

By substituting (7) into (6), we have

$$P \left(t_i > \frac{m^2 t}{k^2} \right) = 1 - P \left(t_i \leqslant \frac{m^2 t}{k^2} \right) = 1 - \frac{m^2 t}{k^2(Mm_0 + t)}. \tag{8}$$

The probability density for $P(k)$ can be obtained using

$$P(k) = \frac{\partial P(s_{ij}(t) < k)}{\partial k} = \frac{2m^2 t}{Mm_0 + t} k^{-3}, \tag{9}$$

predicting that the inner-degree distribution obeys a power-law distribution $P(k) \sim k^{-\gamma}$ with $\gamma = 3$ independent of M and m.

Similarly,

$$\frac{\partial l_{ik}}{\partial t} = \frac{M-1}{M} \alpha n \frac{l_{ik}}{\sum_{m,n,n \neq j} l_{m,n}} \tag{10}$$

in which

$$\sum_{m,n,n \neq j} l_{m,n} = 2 \frac{M-1}{M} \alpha n t + [M(M-1) - (M-1)] \tag{11}$$

in the sense of mathematical expectation. If $\alpha = 0$, then no inter-community links are added in the evolution of the network, and the total number of inter-community links is always $\frac{M(M-1)}{2}$. If $\alpha \neq 0$, then the solution of this equation, with the initial condition that node i in community k was added to the network at time t_j with inter-community connectivity $l_{ik}(t_j) = \alpha n$ in the sense of mathematical expectation, is

$$l_{ik}(t) = \alpha n \left(\frac{t + \beta}{t_j + \beta} \right)^{0.5} \tag{12}$$

with $\beta = \frac{[M(M-1)-(M-1)]M}{2\alpha n(M-1)}$. If α is not so small, such that $2\frac{M-1}{M}\alpha nt \gg [M(M-1)-(M-1)]$ for large t (usually, this is the case), then

$$\frac{\partial l_{ik}}{\partial t} \approx \frac{l_{ik}}{2t}. \tag{13}$$

The solution of this equation, with the initial condition that node i in community k was added to the network at time t_j with inter-community connectivity $l_{ik}(t_j) = \alpha n$ in the sense of mathematical expectation, is

$$l_{ik}(t) \approx \alpha n \left(\frac{t}{t_j}\right)^{0.5}. \tag{14}$$

Similar to the above analysis of the inner-degree distribution, the inter-degree distribution $P(k)$ can be written as

$$P(k) = \frac{2(\alpha n)^2 t}{Mm_0 + t} k^{-3}, \tag{15}$$

predicting that the inter-degree distribution also obeys a power-law distribution $P(k) \sim k^{-\gamma}$ with $\gamma = 3$ independent of M, α and n.

In this case, the total degree of node i in community j is

$$k_{ij}(t) = s_{ij}(t) + l_{ij}(t) \approx (m + \alpha n)\left(\frac{t}{t_i}\right)^{0.5}. \tag{16}$$

It is easy to show that the total degree distribution is

$$P(k) = \frac{2(m + \alpha n)^2 t}{Mm_0 + t} k^{-3} \tag{17}$$

which indicates that the total degree also obeys a power-law distribution $P(k) \sim k^{-\gamma}$ with $\gamma = 3$.

Because $\alpha n < m$ (usually $\alpha n \ll m$), from (5) and (14) (or (12)) we know that the inter-degrees of network nodes are smaller than the inner-degrees of network nodes in the sense of statistics. So, networks generated by this model will have community structure.

Next, we numerically study the properties of the network model. We consider a network generated by the proposed scheme with $M = 3, m_0 = 3, m = 3, \alpha = 0.3, n = 1$ and $N = 3000$. In figures 3–5, the slopes of the lines were obtained by cumulative distributions of 10 runs. The inner-degree distribution of this network is shown in figure 3. As we can see, the inner-degree obeys a power-law distribution $P(k) \sim k^{-\gamma}$ with $\gamma \approx 3$ (the slope of the line is -3). The inter-degree distribution is shown in figure 4, and the inner-degree obeys a power-law distribution $P(k) \sim k^{-\gamma}$ with $\gamma \approx 2.9$ (the slope of the line is -2.9). The total degree distribution is shown in figure 5, and we observe that the total degree also obeys a power-law distribution $P(k) \sim k^{-\gamma}$ with $\gamma \approx 3$. So, all these three distributions are of power-law type, indicating that the generated network is a scale-free network, and these distributions verify the above analytical results. In figure 6 we plot the time evolution of the total degree of two nodes, in which one is an initial node and the other one is added to the network at $t = 60$. This is in good agreement with (16).

We also calculated the diameter D, the average path length l and the clustering coefficient C for this network. The detailed definitions of these parameters can be found in [1]. The values of these parameters are $D = 7, l = 4.0274$ and $C = 0.0270$, respectively. So the network has small average path length and relatively large clustering coefficient (compared to a random graph).
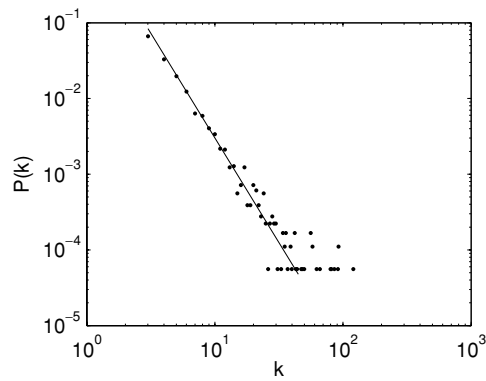
Figure 3. The inner-degree distribution of a network with $N = 3000$, $M = 3$, $m_0 = 3$, $m = 3$, $\alpha = 0.3$ and $n = 1$. As can be seen, the distribution is well approximated by a straight line with gradient -3.
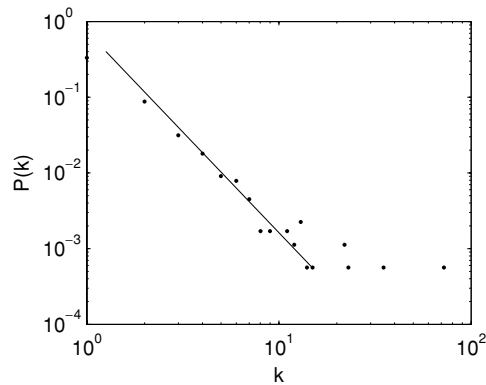

Figure 4. The inter-degree distribution of a network with $N = 3000$, $M = 3$, $m_0 = 3$, $m = 3$, $\alpha = 0.3$ and $n = 1$. As can be seen, the distribution is well approximated by a straight line with gradient -2.9.
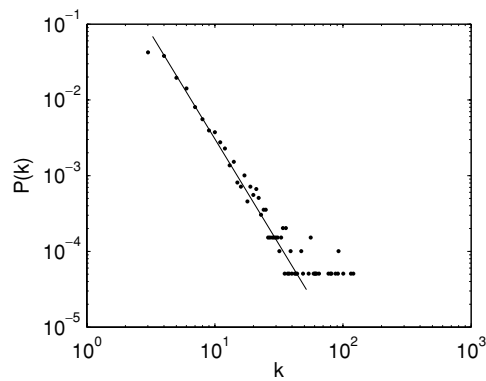

Figure 5. The total degree distribution of a network with $N = 3000$, $M = 3$, $m_0 = 3$, $m = 3$, $\alpha = 0.3$ and $n = 1$. As can be seen, the distribution is well approximated by a straight line with gradient -3.
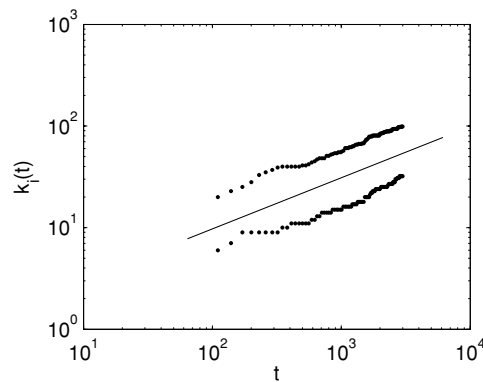
Figure 6. The time evolution of the degree of two nodes, one is an initial node (the upper one) and the other one was added to the network at $t = 60$ (lower one). As can be seen, the time evolution of the degrees is well approximated by a straight line with gradient 0.5.

4. Conclusions and remarks

In this paper we proposed an evolving network model with community structure. We have theoretically analysed the scaling properties of the network by using a mean-field approach. The analytical and numerical results indicate that the network can produce community structure, and the inner-degree, inter-degree and total degree all obey power-law distributions, so the network has scale-free properties. The network also has small average path length and relatively large clustering coefficient (compared to a random graph). Note that in [16] the authors proposed a growing network model of two coupled networks—this idea is a little similar to that in the present paper, but the authors did not consider community structure.

To make our model more understandable and to avoid unnecessarily complicated notations, we made some simplifications in our model. Some more realistic cases and extensions can be considered with minor modification of our model, for example:

(1) We did not consider the formation of new communities during the evolution of the network model. In some real-world networks, some new communities may appear during the evolution. With minor modification, our model could allow from time to time the introduction of new communities: if a node i is the first of a new community, one could only set links between i and the old communities at the time step of its introduction, and follow the same evolving rules in the proceeding time steps as defined in section 2. It is easy to show that, with this modification, the main results will not change.

(2) In the preferential attachment mechanism, if a node i is not connected to a node j (in the same community or different communities) at the beginning of its introduction, it will never be. In considering the resulting networks, this is of no problem, but in considering the evolution process, it is somewhat unrealistic. We can overcome this problem as follows: in each time step, besides the steps mentioned in section 2, with probability β, an existing node is selected, and we perform the same inner- and inter-community attachments of this node to all the other existing nodes that have no links with it. This will surely not change the community structure of the network model.

Future extensions of this work include considering other kinds of inner- and inter-linking mechanisms other than preferential attachment, such as preferential linking [17] and distance preference [18]. Future extensions also include the modelling of directed and weighted network

models with community structure [19] because many real-world networks with community structures are directed and/or weighted [20]. Based on this network model, we can study the effects of community structure on network dynamics, such as stability, synchronization, disease and rumour spreading and robustness. We can also introduce different types of dynamical node models in different communities to study the dynamical behaviour of some real networks, for example, biochemical networks and brain networks.

Acknowledgments

References

[1] Strogatz S H 2001 *Nature* **410** 268
   Albert R and Barabási A L 2002 *Rev. Mod. Phys.* **74** 47
   Dorogovtsev S N and Mendes J F F 2002 *Adv. Phys.* **51** 1079
   Newman M E J 2003 *SIAM Rev.* **45** 167
[2] Watts D J and Strogatz S H 1998 *Nature* **393** 440
   Watts D J 1999 *Small-Worlds: The Dynamics of Networks Between Order and Randomness* (Princeton, NJ: Princeton University Press)
[3] Barabási A L and Albert R 1999 *Science* **286** 509
[4] Redner S 1998 *Eur. Phys. J.* B **4** 131
[5] Flake G W, Lawrence S R, Giles C L and Coetzee F M 2002 *IEEE Comput.* **35** 66
[6] Holme P, Huss M and Jeong H 2003 *Bioinformatics* **19** 532
[7] Sporns O, Chialvo D R, Kaiser M and Hilgetag C C 2004 *Trends Cogn. Sci.* **8** 418
[8] Xie H, Yan K-K and Maslov S 2004 *Preprint* cond-mat/0409087
[9] For a recent review, see Newman M E J 2004 *Eur. Phys. J.* B **38** 321
[10] For some more recent papers see, for example, Massen C P and Doye J P K 2005 *Phys. Rev. E* **71** 046101
   Young M, Sager J, Csárdi G and Hága P 2004 *Preprint* cond-mat/0408263
   Bagrow J P and Bollt E M 2004 *Preprint* cond-mat/0412482
   Latapy M and Pons P 2004 *Preprint* cond-mat/0412368
   Clauset A, Newman M E J and Moore C 2004 *Phys. Rev. E* **70** 066111 (*Preprint* cond-mat/0408187)
   Duch J and Arenas A 2005 *Preprint* cond-mat/0501368
   Clauset A 2005 *Preprint* physics/0503036
   Costa L da F 2004 *Phys. Rev. E* **70** 056106 (*Preprint* cond-mat/0312712)
[11] Grönlund A and Holme P 2004 *Phys. Rev. E* **70** 036108
[12] Noh J D, Jeong H-C, Ahn Y-Y and Jeong H 2004 *Preprint* cond-mat/0412149
[13] Boguna M, Pastor-Satorras R, Diaz-Guilera A and Arenas A 2004 *Phys. Rev. E* **70** 056122
[14] Kimura M, Saito K and Ueda N 2004 *Neural Netw.* **17** 975
[15] Barabási A-L, Albert R and Jeong H 1999 *Physica* A **272** 173
[16] Zheng D and Ergün E 2001 *Preprint* cond-mat/0112052
[17] Dorogovtsev S N, Mendes J F F and Samukhin A N 2000 *Phys. Rev. Lett.* **85** 4633
[18] Jost J and Joy M P 2002 *Phys. Rev. E* **66** 036126
[19] Li C and Chen G Modelling of weighted evolving networks with community structures (submitted)
[20] Li C and Chen G 2003 *Preprint* cond-mat/0311333
   Barrat A, Barthélémy M, Pastor-Satorras R and Vespignai A 2004 *Proc. Natl Acad. Sci. USA* **101** 3747
   Zheng D, Trimper S, Zheng B and Hui P M 2003 *Phys. Rev. E* **67** 040102
   Barrat A, Barthélémy M and Vespignai A 2004 *Phys. Rev. Lett.* **92** 228701
   Li C and Chen G 2004 *Physica* A **343** 288
   Morelli L G 2003 *Phys. Rev. E* **67** 066107