

# Emergence of High clustering in Random Evolving Networks\*

Shengli Zhou<sup>1,2</sup>, Xianmin Geng<sup>1</sup>, and Zhiyuan Yao<sup>2</sup>

<sup>1</sup> College of Science Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu Province, China

<sup>2</sup> State Key Laboratory of Mechanics and Control of Mechanical Structures Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu Province, China  
zslmath@sohu.com, zslmnn@163.com

**Abstract** – To find the formation mechanism of the complex network possess high clustering property, the paper proposes a new network model based on the neighboring nodes conglomerative connection mechanism. By analytical analysis, the paper gets a lower bound of the clustering coefficient and proves that the model satisfies the scale-free behaviour as well as the high clustering property. Furthermore, when the model loses the neighboring nodes conglomerative connection mechanism, it also loses the high clustering property.

**Index Terms** - Complex systems, statistical physics, Probability theory.

## 1. Introduction

Many real complex systems can be described by the networks with a large number of nodes and edges among the nodes, whose nodes represent the elements of the system and whose edges represent the interactions between them. For example, in the World Wide Web (WWW), whose nodes are HTML documents and the edges are the links pointing from one page to another; in the movie actor collaboration network, each actor is represented by a node, two actors being connected if they were cast together in the same movie. In the study of the complex network, scientists have revealed some common features of the real networks, such as small-world property, high clustering property, scale-free behaviour and so on. However, because of their large size and the complexity of their interactions, the formation mechanisms of these features are mostly unknown, it has become a topic of interest in recent years.

Scale-free behavior means that the probability of a node's degree equal to  $k$  is  $\alpha k^{-\gamma}$ , where  $\alpha$  is a constant. In the formation mechanism of this behaviour, Barabasi and Albert have done much highly effective work, they show that growing together with preferential attachment is a sufficient condition. However, in the formation mechanism of high clustering property, there remains some difficulties.

Traditionally, the clustering coefficient is used to measure the connectivity in the random networks. If node  $A$  is connected to node  $B$  and node  $B$  to node  $D$ , then the clustering coefficient of node  $B$  is the connectivity of node  $A$  and node  $D$ . The clustering coefficient of network is defined as [5]: Let us focus first on a selected node  $l$  in the

network, having  $k_l$  edges which connect it to  $k_l$  other nodes. The ratio among the number  $E_l$  of edges that actually exist between these  $k_l$  nodes and the total number  $k_l(k_l - 1)/2$  gives the value of the clustering coefficient of node  $l$ ,

$C_l = \frac{E_l}{k_l(k_l - 1)/2}$ . The clustering coefficient of the whole

network is the average of all individual  $C_l$ 's.

Scientists have got many real complex network's clustering coefficients, when contrast the values to the clustering coefficients of the ER random model which with the same number of nodes and edges, they found that the former is typically larger than latter. For example, in a movie actor collaboration network with 225226 nodes,  $C = 0.79$  vs  $C_{rand} = 0.00027$ ; in a WWW with 153127 nodes,  $C = 0.79$  vs  $C_{rand} = 0.00023$ ; in a electrical power grid of the western United States with 4941 nodes,  $C = 0.08$  vs  $C_{rand} = 5.5 \times 10^{-5}$  and so on. Scientists name this property of real network after the high clustering property.

In the formation mechanism of high clustering property, though Watts and Strogatz have proposed some models which possess high clustering property, they have not proposed an explicit formation mechanism of this property. Besides, the WS model does not satisfy the scale-free behavior, which contrast to the BA model does not satisfy the high clustering property. In 2002, Holme and Kim provided a complex network model with the two properties, fly in the ointment, they could not propose an explicit formation mechanism to this property too. Besides, the result is just numerical, it lacks of an analysis result.

In this paper, we suggest that the neighboring nodes conglomerative connection which means when a new node with  $m$  edges is added to the network, the edges are linked to node  $i$  and its neighbors clustering is a coefficient condition propose the high clustering property, and propose a new network model. By analytical analysis, we give a lower bound of the clustering coefficient, prove that it satisfies the scale-free behaviour and the high clustering property. Furthermore, when the model loses the neighboring nodes conglomerative

\* This work is supported by NNSF (grant No.11171215), NNSF (grant No.51275229) in China, and the National Basic Research Program (973 Program) (2011CB707602).

connection mechanism, it also loses the high clustering property.

## 2. The mode with high clustering property

The neighboring nodes conglomerative connection is very common in the real systems. For example, in the friendship network of social networks each node represents a person, and two persons are connected if they are friends. A new number  $V$  adds into the group, he becomes a friend of  $W$  at first. Then, since  $W$  may introduce his friend to  $V$ , so contrast with other members in the group,  $V$  becomes the friend of  $W$ 's friend easily. In the citation network, the nodes represent published articles and an edge represents a reference to a previously published article. When a new article  $A$  cites article  $B$ , for the relevance of the study, the article  $A$  cites the articles which is cited by  $B$  easily. We speculate the neighboring nodes conglomerative connection is the key to the high clustering property of network. A model based on the neighboring nodes conglomerative connection will be proposed, and our speculation is verified to be rational by calculating the clustering coefficient.

To incorporate the conglomerative connection character of the network, the model starts with a complete graph with  $m_0$  ( $m_0 \geq 2$ ) nodes, at every time step, we add a new node with  $m$  ( $m \leq m_0$ ) edges. The  $m$  edges link to the existing nodes by the following ways: we choose an edge from the edges of the new node randomly, assume the edge links to an existing node  $i$  with preferential probability  $\frac{k_i(t)}{\sum_j k_j(t)}$  (In order to describe

simply, we name this operation after the first operation); each of the rest  $m-1$  edges randomly links to a neighbour node of  $i$  with probability  $p$ , and with probability  $1-p$  links to an existing node besides node  $i$  preferentially (we name this operation after the second operation).

## 3. Clustering coefficient in the model

Intuitively, the clustering coefficient  $C$  is proportional to  $p$  and  $m$ . But simulations of the model show that  $C$  increases as  $p$  increases (see figure 1), decreases as  $m$  increases (see figure 2).

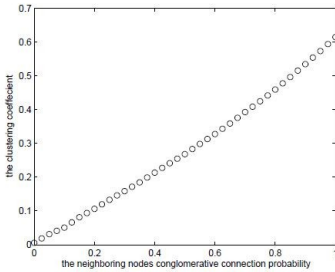


Figure 1:  $C$  increases as  $p$  increases. Clustering coefficient  $C$  for networks of 10000 nodes and the number of the initial nodes is  $m_0 = 5$ ; the edge number of new adding node is  $m = 3$ , for  $\circ$ .

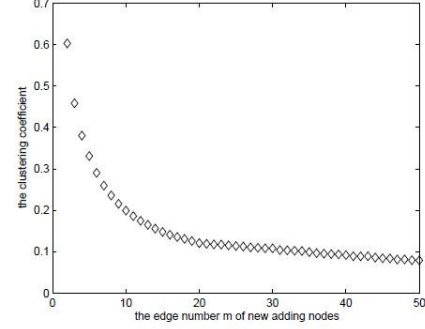


Figure 2:  $C$  increases as  $m$  increases. Clustering coefficient  $C$  for networks of 10000 nodes and the number of the initial nodes is  $m_0 = 50$ ; the edge number of new adding node is  $p = 0.8$ , for  $\diamond$ .

This is a mix model which possess the neighboring nodes conglomerative connection and all nodes preferential attachment. When  $p = 0$ , the model is the BA model. We consider the case  $p \neq 0$ .

Firstly, we calculate the degree of node  $i$  at time  $t$  and the degree distribution of the network.

The rate at which node  $i$  acquires edges is

$$\begin{aligned} \frac{\partial k_i(t)}{\partial t} &= C_m^1 \frac{1}{m} \frac{k_i(t)}{\sum_j k_j(t)} + \sum_{h \in \Omega_i(t)} C_m^1 \frac{1}{m} \cdot \frac{k_h(t)}{\sum_j k_j(t)} \\ &\cdot (p^{m-1} C_{m-1}^1 \frac{1}{k_h(t)} + (1-p)^{m-1} \cdot C_{m-1}^1 \frac{k_h(t)}{\sum_j k_j(t)}) \\ &+ \sum_{l=1}^{m-2} C_{m-1}^l p^l (1-p)^{m-1-l} (C_l^1 \frac{1}{k_h(t)} + C_{m-1-l}^1 \frac{k_i(t)}{\sum_j k_j(t)}) \\ &+ \sum_{g \in \Gamma_i(t)} C_m^1 \frac{1}{m} \frac{k_g(t)}{\sum_j k_j(t)} \cdot \sum_{l=0}^{m-2} C_{m-1}^l \cdot p^l (1-p)^{m-1-l} C_{m-1-l}^1 \frac{k_i(t)}{\sum_j k_j(t)} \\ &= \frac{k_i(t)}{2t} + o\left(\frac{1}{t}\right), \end{aligned}$$

which gives:  $k_i(t) = m \cdot \left(\frac{t}{t_i}\right)^{\frac{1}{2}}$ . Where  $k_i(t)$  is the degree of node  $i$  at time  $t$ ,  $\Omega_i(t)$  is the set of the neighbors of node  $i$  at time  $t$ ,  $\Gamma_i(t)$  is the set of the node which is neither node  $i$  nor the neighbors of  $i$  at time  $t$ ,  $t_i$  is the time at which node was added to the network. So

$$P(k_i(t) < k) = P\left(m \cdot \left(\frac{t}{t_i}\right)^{\frac{1}{2}} < k\right) = P\left(t_i > \frac{m^2 t}{k^2}\right) = 1 - \frac{m^2 t}{k^2 (t + m_0)}.$$

When  $k > m$ , the probability density  $P(k)$  can be obtained from  $P(k) = \frac{\partial P(k_i(t) < k)}{\partial k}$ , and as  $t \rightarrow \infty$ , we have

$$P(k) = 2m^2 k^{-3}. \text{ When } k = m, P(k) = 1 - \sum_{k=m+1}^{\infty} 2m^2 k^{-3}.$$

Let  $e_i(t)$  be the number of edges among the neighbors of node  $i$  at time  $t$ . Now, we calculate  $e_i(t)$  and the clustering

coefficient of the network. The rate equation of  $e_i(t)$  can be obtained from the following 4 cases:

The new node does not link to node  $i$ . In this case,  $e_i(t)$  can not increase. Thus

$$\left(\frac{\partial e_i(t)}{\partial t}\right)_1 = 0.$$

The new node preferentially links to node  $i$  at the first operation. Let  $P_2(l)$  be the probability of  $e_i(t)$  increasing  $l$  under this case:

$$P_2(l) = C_m^1 \frac{1}{m} \frac{k_i(t)}{\sum_j k_j(t)} \sum_{a=0}^l C_{m-1}^a p^a \cdot (1-p)^{m-1-a} \cdot \frac{\sum_{i_1 \dots i_{l-a}} k_{i_1}(t) \dots k_{i_{l-a}}(t)}{(\sum_j k_j(t))^{l-a}} \cdot \left(1 - \frac{k_i(t) + \sum_{h \in \Omega_i(t)} k_h(t)}{\sum_j k_j(t)}\right)^{m-1-l},$$

where  $a$  is the number of the randomly linked edges,  $\Omega_i(t)$  is the set of the neighbour nodes of  $i$ . If we want  $e_i(t)$  to increase  $l$ ,  $a$  must satisfy  $0 \leq a \leq l$ . Summing all the  $l$ , we get the following equation:

$$\left(\frac{\partial e_i(t)}{\partial t}\right)_2 = \sum_{l=1}^{m-1} l \cdot P_2(l)$$

The new node preferentially links to one node  $i$  at the first operation, and links to  $i$  at the second operation.

Assume that there are  $a$  randomly linked edges when the new node links to  $h$  ( $h \in \Omega_i(t)$ ) at the first operation. Let  $\Omega_{ih}(t)$  be the set of the common neighbours of node  $h$  and  $i$ ,  $p_{ih}$  be the probability of the edge which links to  $i$  is randomly linked, thus

$$p_{ih} = \frac{\frac{a}{k_h(t)}}{\frac{(m-1-a)k_i(t)}{\sum_j k_j(t)} + \frac{a}{k_h(t)}}.$$

Let  $(P_3(l))_h$  be the probability that  $e_i(t)$  increasing  $l$  under this case. Now we get the equation of  $(P_3(l))_h$  from the following 4 cases.

1) When  $|\Omega_{ih}(t)| \geq l-1$  and the edge which links to  $i$  is randomly linked.

$$(P_3(l))_{h_1} = C_m^1 \frac{1}{m} \frac{k_h(t)}{\sum_j k_j(t)} \sum_{a=1}^{m-1} C_{m-1}^a p^a (1-p)^{m-1-a} p_{ih} \cdot C_a^1 \frac{1}{k_h(t)} \sum_{b=0}^{\min(a, l-1)} C_{a-1}^b \sum_{u_1 \dots u_b \in \Omega_{ih}(t)} \left(\frac{1}{k_h(t)}\right)^b \left(1 - \frac{1 + |\Omega_{ih}(t)|}{k_h(t)}\right)^{a-1-b} \cdot C_{m-1-a}^{l-1-b} \sum_{i_1 \dots i_{l-1-b} \in \Omega_{ih}(t) \setminus \{u_1 \dots u_b\}} \frac{k_{i_1}(t) \dots k_{i_{l-1-b}}(t)}{(\sum_j k_j(t))^{l-1-b}} \left(1 - \frac{k_i(t) + \sum_{s \in \Omega_i(t)} k_s(t)}{\sum_j k_j(t)}\right)^{m-1-a-l+c},$$

where  $b$  is the number of nodes in the set  $\Omega_{ih}(t)$  which randomly linked edges link to under this case.

2) When  $|\Omega_{ih}(t)| \geq l-1$  and the edge which links to  $i$  is preferentially linked.

$$(P_3(l))_{h_2} = C_m^1 \frac{1}{m} \frac{k_h(t)}{\sum_j k_j(t)} \sum_{a=0}^{m-2} C_{m-1}^a p^a (1-p)^{m-1-a} (1-p_{ih}) \cdot C_{m-1-a}^1 \frac{k_i(t)}{\sum_j k_j(t)} \sum_{c=0}^{\min(a, l-1)} C_a^c \sum_{u_1 \dots u_c \in \Omega_{ih}(t)} \left(\frac{1}{k_h(t)}\right)^c \left(1 - \frac{1 + |\Omega_{ih}(t)|}{k_h(t)}\right)^{a-1-c} \cdot C_{m-2-a}^{l-1-c} \sum_{i_1 \dots i_{l-1-c} \in \Omega_{ih}(t) \setminus \{u_1 \dots u_c\}} \frac{k_{i_1}(t) \dots k_{i_{l-1-c}}(t)}{(\sum_j k_j(t))^{l-1-c}} \left(1 - \frac{k_i(t) + \sum_{s \in \Omega_i(t)} k_s(t)}{\sum_j k_j(t)}\right)^{m-1-a-l+c},$$

similarly, where  $c$  is the number of nodes in the set  $\Omega_{ih}(t)$  which randomly linked edges link to under this case.

3) When  $|\Omega_{ih}(t)| < l-1$  and the edge which links to  $i$  is randomly linked.

$$(P_3(l))_{h_3} = C_m^1 \frac{1}{m} \frac{k_h(t)}{\sum_j k_j(t)} \sum_{a=1}^{a_h^1} C_{m-1}^a p^a (1-p)^{m-1-a} p_{ih} \cdot C_a^1 \frac{1}{k_h(t)} \sum_{c=0}^{\min(a, |\Omega_{ih}(t)|)} C_{a-1}^b \sum_{u_1 \dots u_b \in \Omega_{ih}(t)} \left(\frac{1}{k_h(t)}\right)^b \left(1 - \frac{1 + |\Omega_{ih}(t)|}{k_h(t)}\right)^{a-1-b} \cdot C_{m-1-a}^{l-1-b} \sum_{i_1 \dots i_{l-1-b} \in \Omega_{ih}(t) \setminus \{u_1 \dots u_b\}} \frac{k_{i_1}(t) \dots k_{i_{l-1-b}}(t)}{(\sum_j k_j(t))^{l-1-b}} \left(1 - \frac{k_i(t) + \sum_{s \in \Omega_i(t)} k_s(t)}{\sum_j k_j(t)}\right)^{m-1-a-l+b},$$

Where  $a_h^1$  is the maximum of the randomly linked edges' number. For  $|\Omega_{ih}(t)| < l-1$  and the edge which links to  $i$  is randomly linked, if we want  $e_i(t)$  to increase  $l$ , there must be at least  $l-1-|\Omega_{ih}(t)|$  preferentially linked edges, thus there are at most

$$a_h^1 = m-1-(l-1-|\Omega_{ih}(t)|) = m-l+|\Omega_{ih}(t)|$$

randomly linked edges. Is the number of nodes in the set  $\Omega_{ih}(t)$  which randomly linked edges link to under this case, similarly.

When  $|\Omega_{ih}(t)| < l-1$  and the edge which links to  $i$  is preferentially linked.

$$(P_3(l))_{h_4} = C_m^1 \frac{1}{m} \frac{k_h(t)}{\sum_j k_j(t)} \sum_{a=0}^{a_h^2} C_{m-1}^a p^a (1-p)^{m-1-a} (1-p_{ih}) \cdot C_{m-1-a}^1 \frac{k_i(t)}{\sum_j k_j(t)} \sum_{c=0}^{\min(a, |\Omega_{ih}(t)|)} C_a^c \sum_{u_1 \dots u_c \in \Omega_{ih}(t)} \left(\frac{1}{k_h(t)}\right)^c \left(1 - \frac{1 + |\Omega_{ih}(t)|}{k_h(t)}\right)^{a-c} \cdot C_{m-2-a}^{l-1-c} \sum_{i_1 \dots i_{l-1-c} \in \Omega_{ih}(t) \setminus \{u_1 \dots u_c\}} \frac{k_{i_1}(t) \dots k_{i_{l-1-c}}(t)}{(\sum_j k_j(t))^{l-1-c}} \left(1 - \frac{k_i(t) + \sum_{s \in \Omega_i(t)} k_s(t)}{\sum_j k_j(t)}\right)^{m-1-a-l+c},$$

where  $a_h^2 = m+|\Omega_{ih}(t)|-1-l$  is the maximum of the randomly linked edges' number and  $c$  is the number of nodes in the set  $\Omega_{ih}(t)$  which randomly linked edges link to under this case, so

$$(P_3(l))_h = ((P_3(l))_{h_1} + (P_3(l))_{h_2}) \Big|_{|\Omega_{ih}(t)| \geq l-1} \\ + ((P_3(l))_{h_3} + (P_3(l))_{h_4}) \Big|_{|\Omega_{ih}(t)| < l-1} \\ \left(\frac{\partial e_i(t)}{\partial t}\right)_3 = \sum_{h \in \Omega_{ig}(t)} \sum_{l=1}^{m-1} l \cdot (P_3(l))_h$$

The new node preferentially links to a node which is neither  $i$  nor the neighbour of  $i$  at the first operation, and links to  $i$  at the second operation.

Let  $\Gamma_i(t)$  be the set of the node which is neither  $i$  nor the neighbor of  $i$ . Similarly, assume that there are  $a$  randomly linked edges at which the new node links to  $g$  ( $g \in \Gamma_i(t)$ ) at the first operation. For the edge links to  $i$  must be a preferentially linked edge, satisfies  $0 \leq a \leq m-1$ . Let  $\Omega_{ig}(t)$  be the set of common neighbors of node  $g$  and  $i$ ,  $(P_4(l))_g$  be the probability of  $e_i(t)$  increasing under this case, we get the equation of  $(P_4(l))_g$  from the following 2 cases.

When  $|\Omega_{ih}(t)| < l$ ,

$$(P_4(l))_{g_1} = C_m^1 \frac{1}{m} \frac{k_g(t)}{\sum_j k_j(t)} \sum_{a=0}^{b_g} C_{m-1}^a p^a (1-p)^{m-1-a} \\ \cdot C_{m-1-a}^1 \frac{k_i(t)}{\sum_j k_j(t)} \sum_{d=0}^{\min(a, |\Omega_{ih}(t)|)} C_a^d \sum_{v_1 \dots v_d \in \Omega_{ig}(t)} \left(\frac{1}{k_g(t)}\right)^d \left(1 - \frac{|\Omega_{ig}(t)|}{k_g(t)}\right)^{a-d} \\ \cdot C_{m-2-a}^{l-d} \sum_{i_1 \dots i_{l-d} \in \Omega_i(t) \setminus \{v_1 \dots v_d\}} \frac{k_{i_1}(t) \dots k_{i_{l-d}}(t)}{(\sum_j k_j(t))^{l-d}} \left(1 - \frac{\sum_{h \in \Omega_{ig}(t)} k_h(t)}{\sum_j k_j(t)}\right)^{m-2-a-l+d},$$

where

$$b_g = m-1-(l-|\Omega_{ig}(t)|+1) = m-2-l+|\Omega_{ig}(t)|$$

is the maximum of the randomly linked edges' number and  $d$  is the number of nodes in the set  $\Omega_{ig}(t)$  which randomly linked edges link to under this case.

When  $|\Omega_{ih}(t)| \geq l$

$$(P_4(l))_{g_2} = C_m^1 \frac{1}{m} \frac{k_g(t)}{\sum_j k_j(t)} \sum_{a=0}^{m-2} C_{m-1}^a p^a (1-p)^{m-1-a} \\ \cdot C_{m-1-a}^1 \frac{k_i(t)}{\sum_j k_j(t)} \sum_{d=0}^{\min(a, l)} C_a^e \sum_{v_1 \dots v_d \in \Omega_{ig}(t)} \left(\frac{1}{k_g(t)}\right)^e \left(1 - \frac{|\Omega_{ig}(t)|}{k_g(t)}\right)^{a-e} \\ \cdot C_{m-2-a}^{l-e} \sum_{i_1 \dots i_{l-e} \in \Omega_i(t) \setminus \{v_1 \dots v_d\}} \frac{k_{i_1}(t) \dots k_{i_{l-e}}(t)}{(\sum_j k_j(t))^{l-e}} \left(1 - \frac{\sum_{h \in \Omega_{ig}(t)} k_h(t)}{\sum_j k_j(t)}\right)^{m-2-a-l+e},$$

where  $e$  is the number of nodes in the set  $\Omega_{ig}(t)$  which randomly linked edges link to under this case, similarly.

$$(P_4(l))_g = ((P_4(l))_{g_1}) \Big|_{|\Omega_{ig}(t)| < l} + ((P_4(l))_{g_2}) \Big|_{|\Omega_{ig}(t)| \geq l},$$

$$\left(\frac{\partial e_i(t)}{\partial t}\right)_4 = \sum_{g \in \Gamma_i} \sum_{l=1}^{m-1} l \cdot (P_4(l))_g.$$

When only taking  $l = a = m-1$ , we get

$$\left(\frac{\partial e_i(t)}{\partial t}\right)_2 \geq p^{m-1}(m-1) \frac{k_i(t)}{\sum_j k_j(t)}.$$

When only taking  $l = a = 1$ , we get

$$\left(\frac{\partial e_i(t)}{\partial t}\right)_3 \geq p^{m-1}(m-1) \frac{k_i(t)}{\sum_j k_j(t)}.$$

So

$$\frac{\partial e_i(t)}{\partial t} = \left(\frac{\partial e_i(t)}{\partial t}\right)_1 + \left(\frac{\partial e_i(t)}{\partial t}\right)_2 + \left(\frac{\partial e_i(t)}{\partial t}\right)_3 + \left(\frac{\partial e_i(t)}{\partial t}\right)_4 \\ \geq \left(\frac{\partial e_i(t)}{\partial t}\right)_2 + \left(\frac{\partial e_i(t)}{\partial t}\right)_3 \\ = 2p^{m-1}(m-1) \frac{k_i(t)}{\sum_j k_j(t)}.$$

Let  $x_l$  be the number of edges among the neighbours of node  $i$  when  $i$  was added into the network and there are  $l$  edges which link randomly, thus  $x_l \geq l$ . So

$$e_i(t_i) = \sum_{l=0}^{m-1} x_l C_{m-1}^l p^l (1-p)^{m-1-l} \geq (m-1)p.$$

From  $k_i(t) = m \cdot \left(\frac{t}{t_i}\right)^{\frac{1}{2}}$  we obtain

$$e_i(t) = e_i(t_i) + \int_{t_i}^t \frac{\partial e_i(x)}{\partial x} dx \\ \geq (m-1)p + 2p^{m-1}(m-1) \left(\frac{k_i(t)}{m} - 1\right).$$

Let  $e_{i,k}(t)$  be the number of edges among the neighbors of  $i$  at time  $t$  under the case  $\lim_{t \rightarrow \infty} k_i(t) = k$ ,

$e_{i,k}(t) \in \{0, 1, 2, \dots, \frac{k(k-1)}{2}\}$  is a monotone increasing sequence, thus the limit  $e_i(t) = \lim_{t \rightarrow \infty} e_{i,k}(t)$  exists, and we have

$$e_i(k) \geq (m-1)p + 2p^{m-1}(m-1) \left(\frac{k}{m} - 1\right).$$

We get

$$C_i(k) = \frac{2e_i(k)}{k(k-1)} \geq \frac{2(m-1)p + 4p^{m-1}(m-1) \left(\frac{k}{m} - 1\right)}{k(k-1)}.$$

And

$$\begin{aligned}
C &= \sum_{k=m} C_i(k) p(k) \\
&\geq \frac{2p}{m} \left(1 - \sum_{k=m+1} 2m^2 k^{-3}\right) \\
&\quad + 4m^2 \sum_{k=m+1} \frac{(m-1)p + 2p^{m-1}(m-1)\left(\frac{k}{m} - 1\right)}{k^4(k-1)} \\
&\sim O(1).
\end{aligned}$$

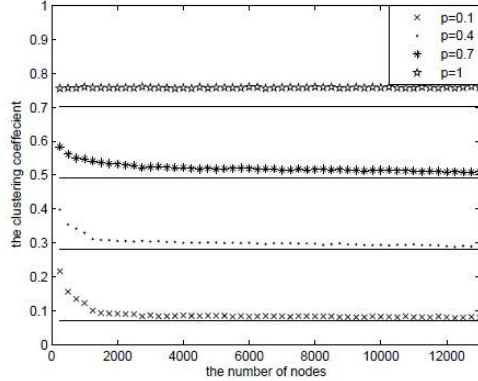


Figure 3: Clustering coefficient for networks of continuously variable nodes. Where the abscissa axis represents the size of the network and the ordinate axis represents the clustering coefficient of the whole network. The solid lines represent the analytical result. The number of the initial nodes is  $m_0 = 5$  and the edge number of the new adding node is  $m = 2$ . The neighboring nodes conglomerative connection probability is  $p = 0, 0.4, 0.7, 1$  for  $\times$ ,  $\cdot$ ,  $*$  and  $\star$ , respectively.

Figure 3 shows the comparison between the simulation results and the analytical results. From the figure we can see that the simulation result is consistent with the analytical result, and our analytical result is a lower bound of the network's clustering coefficient.

From above we have showed that this model has high clustering property. If  $p = 0$ , namely the neighboring nodes conglomerative connection disappear, from [10] we know

$$C = \frac{m}{8} \cdot \frac{\ln N(t)^2}{N(t)}, \text{ and } C \rightarrow 0 \text{ as } N(t) \rightarrow \infty, \text{ the model does}$$

not have high clustering property. Thus we have proved that the neighboring nodes conglomerative connection is the key of the network with high clustering property.

#### 4. Conclusion

This paper speculates the reasons that the real networks produce the high clustering property, and proposed a new evolving mechanism: neighbouring nodes conglomerative connection, which is prevalent in many real networks. Furthermore, we verified that the neighbouring nodes conglomerative connection is the key of the network producing high clustering property indeed. The model in this paper was relatively simple, to be described convenience, it only considered the operation of adding nodes with neighbouring nodes conglomerative connection leading to the high clustering in the network. Similarly, we can also prove that the operations of adding edges with neighbouring edges conglomerative connection leading to the high clustering in the network; when the model loses neighbouring nodes or edges, conglomerative connection, even it has the mechanism of randomly add node or edge, preferentially add node or edge, the mode still loses the property of high clustering coefficient. We have already done some research about these aspects; these results prove our speculations are right.

#### 5. References

- [1] M. Durzinsky, A. Wagler, R. Weismantel, W. Marwan, "Automatic reconstruction of molecular and genetic networks from discrete time series data," *Biosystems*, vol. 93, no. 3, pp. 181-190, 2008.
- [2] J. Shao, S. Buldyrev, L. Braunstein, S. Havlin, H. Stanley, "Structure of shells in complex networks," *Phys. Rev. E*, vol. 80, pp. 036105 1-13, 2009.
- [3] J. Ramasco, S. Dorogovtsev, R. Satorras, "Self-organization of collaboration networks," *Phys. Rev. E*, vol. 70, pp. 036106 1-10, 2004.
- [4] A. Barabási, R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5839, pp. 509-512, 1999.
- [5] D. Watts, S. Strogatz, "Collective dynamics of small-world networks," *Nature*, vol. 393, pp. 509-512, 1998.
- [6] P. Erdős, A. Rényi, "On singular radii of power series," *Publ. Math.* vol. 10, pp. 290-302, 1959.
- [7] P. Erdős, A. Rényi, "On the evolution of random graphs" *Publ. Math. Inst. Hung. Acad. Sci.*, pp. 17-61, 1960.
- [8] R. Albert, A. Barabási, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 74, pp. 47-92, 2002.
- [9] L. Adamic, B. Huberman, "Growth dynamics of the world wide web," *Nature*, vol. 401, 1999.
- [10] V. Eguiluz, K. Klemm, "Epidemic Threshold in Structured Scale-Free Networks," *Phys. Rev. Lett.*, vol. 89, pp. 108701 1-4, 2002.
- [11] P. Holme, B. Kim, "Growing scale-free networks with tunable clustering," *Phys. Rev. E*, vol. 65, pp. 026107 1-4, 2002.