# High clustering coefficient of networks based on mean-field method

Caifeng Du [a], Pengzhen Ren [b] and Kai Gao [c]

College of Science, China University of Petroleum, Qingdao 266580, China

[a]ducaif@amss.ac.cn, [b]renpengzhen@foxmail.com, [c]1498061604@qq.com

## Abstract

**Due Many real networks in nature and society are characterized by a high degree of clustering,in addition to the small-world and the scale-free property.We propose a model of a growing network in which preferential linking is combined with local clustering organization.The model is solved exactly in the limit of a large network size.We demonstrate that both the degree distribution and the triangle distribution depend on the network size.When the size of the network tends to infinity,there is a stationary power law degree distribution and a high clustering coefficient.**

## Keywords

**random network, mean-field method, high clustering coefficient.**

## 1. Introduction

With the increasing attention of network representing many real networks, new statistical measures have been motivated to describe random networks [1, 2, 3]. These measures has revealed that, as a difference with the classical Erdos Renyi [4] model, real networks are characterized by three property :a power law distribution of vertex degrees [1, 5, 6], a high clustering coeffcient [1,7, 8] and a short average path length[1,9,10,11].

Consider undirected simple graphs on $i=1,2,\cdots,N$ nodes.The clustering coeffcient for node *i* with $k_i$ is defined as $C_i=\frac{T_i}{k_i(k_i-1)}$ , where $T_i$ is the number of links between the $k_i$ neighbors of $i$ ,in a simple word, we call it the triangle number of node *i*. And the average clustering coefficient defined as $C=\frac{\sum_i C_i}{N}$. Empirical results indicate that *C* is significantly higher for most real networks than for a random one of similar size[1,2,9]. Furthermore,the clustering coefficient of real networks is to a high degree independent of the number of the nodes in the network[1].

Yet, most models of complex networks have difficulty capturing this feature.For example,the random graph model[4] predicts the average clustering coefficient $C(N)$ decreases as $N^{-1}$ .The scale-free model[12] predicts a larger clustering coefficient than a random one of the same size.But numerical simulations of it show that the average clustering coefficient depends on the system size as $C(N)\propto N^{-0.75}$ ,and this is still not consistent with that many real networks have the $C(N)$ independent of the network size. Recently many works have been done to predict the case of the real network by simulations and approximation.

In the present paper we propose a networks model with local clustering organization which means that whenever a new node point to an old one it also prefer to pointing to the neighbors of the old node. The model is solved exactly in the limit of a large network size. We demonstrate that both the degree distribution and the triangle distribution depend on the network size. When the size of the network tends to infinity,the degree distribution behaves as $P(k)\propto k^{-3}$ and the average clustering coefficient $C(N)$ independent of the network size *N*.

## 2. The Scale-free model

Let us introduce the model of scale free growing networks with undirected links.

Initially($t = m-1$),there is a completed graph with $m$ nodes, $s = 0,1,\cdots,m-1$. At each increment of time , a new node with $m$ edges is added. The first edge is linked to the existed nodes of the network with preferential attachment: the probability that a new node will be linked to node $i$ depend on the degree $k_i$ of that node, such that $\frac{k_i}{\sum_j k_j}$.Then, the rest $m-1$ edges are uniformly linked to the neighbors of node i that has been linked to by the first edge.

We solve the model by mean-field method and rate equation.

From the model, a node $s$ receives links from a new node by two case:

(a)it can be linked by the first link with preferential probability $\frac{k_i}{\sum_j k_j}$;

(b)if one of it'neighbors $j$ get the first link, it also can be linked with uniformly probability $\frac{m-1}{k_j}$.

We can easily obtain the differential equation about degree $k_s(t)$:

$$\frac{dk_s(t)}{dt} = \frac{k_s}{\sum_j k_j} + \sum_{j\in(s,j)} \frac{k_j}{\sum_j k_j}\frac{m-1}{k_j} \tag{1}$$

where $j \in (s, j)$ denote all nodes $j$ linked to node $s$. From the initial condition of the equation(1), $k_s(s) = m$ ,and we know $\sum_j k_j = 2mt$ ,we get the function

$$k_s(t) = m(t/s)^{1/2} \tag{2}$$

We can see that the above function is as same as that of the scale-free model[12].And we can get similarly the power law degree distribution $P(k) \approx 2m^2 k^{-3}$ when the network size is very large.

## 3. The analytical result

In the following, we will give and solve the rate equation of $T_s(t)$ ,the triangle number of nodes. From the model,the triangle number $T_s(t)$ of node $s$ increases in three case:

(a)$s$ is linked by the first link of new node,then the number will increase $m-1$;

(b) if one of it'neighbors $j$ get the first link,then whenever it receive an edge ,the triangle number increases 1;

(c) if(b)and if $I$ which is another neighbor $j$ receives a new link and simultaneously $i$ is one neighbor of $s$ then the triangle number of $s$ will increase.

So we obtain the rate equation

$$\frac{dT_s(t)}{dt} = (m-1)\frac{k_s}{\sum_j k_j} + \sum_{j\in(s,j)} \frac{k_j}{\sum_j k_j}\frac{m-1}{k_j s} + \sum_{j\in(s,j)} \frac{k_j}{\sum_j k_j}\frac{m-1}{k_j}\frac{2T_s}{k_s}\frac{m-2}{k_j} \tag{3}$$

We can rewrite (3) by a simple form

$$\frac{dT_s(t)}{dt} = 2(m-1)\frac{k_s}{\sum_j k_j} + \frac{2(m-1)(m-2)}{\sum_j k_j}\frac{2T_s}{k_s}\sum_{j\in(s,j)}\frac{1}{k_j}$$

Now,in the sense of average and using (2) we get

$$\sum_{j\in(s,j)} \frac{1}{k_j} = \frac{k_s}{t}\int_0^t \frac{1}{k_j} dj = \frac{2}{3m} k_s$$

So we get the simpler form of (3)

$$\frac{dT_s(t)}{dt} = 2(m-1)\frac{k_s}{\sum_j k_j} + \frac{2(m-1)(m-2)}{\sum_j k_j}\frac{2}{3m}T_s$$

By(2), we get

$$\frac{dT_s(t)}{dt} = (m-1)\frac{1}{t^{1/2}s^{1/2}} + \frac{2}{3}\frac{(m-1)(m-2)}{m^2}\frac{T_s}{t} \qquad (3^{'})$$

The general solution of $(3^{'})$is

$$T_s(t) = \frac{m-1}{\frac{1}{2} - \frac{2}{3}\frac{(m-1)(m-2)}{m^2}}(\frac{t}{s})^{\frac{1}{2}} + f(s)(\frac{t}{s})^{\frac{2}{3}\frac{(m-1)(m-2)}{m^2}} \qquad (4)$$

where $f(s)$ is the function only dependent on *s*. For the convenience of the following analysis,we can use the equivalent form of (4)

$$T_s(t) = \frac{m-1}{\frac{1}{2} - \frac{2}{3}\frac{(m-1)(m-2)}{m^2}}((\frac{t}{s})^{\frac{1}{2}} - 1) + T_s(s)(\frac{t}{s})^{\frac{2}{3}\frac{(m-1)(m-2)}{m^2}} \qquad (4')$$

where $T_s(s)$ is the very initial condition of $T_s(t)$ .

To get the exact solution of (3),we now try to find the form of $T_s(s)$. We perform integration on

*s* from 0 to *t* respectively to the left and right of (3 ). Then we get

$$2T_t = \int_0^t \frac{dT_s(t)}{dt}ds = 2(m-1) + \int_0^t \frac{2}{3}\frac{(m-1)(m-2)}{m^2}\frac{T_s(t)}{t}ds \qquad (5)$$

Making difference on both sides of (5 ) and using the (5) .

$$\frac{dT_t(t)}{dt} = \frac{(m-1)(m-2)}{m^2}(-\frac{1}{t^2}\int_0^t T_s(s)ds + \frac{1}{t}T_t(t)) = \frac{(m-1)(m-2)}{m^2}(-\frac{1}{t^2}t\frac{T_t(t)-(m-1)}{\frac{(m-1)(m-2)}{m^2}} + \frac{1}{t}T_t(t)) = (\frac{(m-1)(m-2)}{m^2} - 1)\frac{T_t(t)}{t} + \frac{m-1}{t}$$

Then we get the general solution of $T_t(t)$

$$T_t(t) = g(m)t^{\frac{(m-1)(m-2)}{m^2}-1} + \frac{m-1}{1-\frac{(m-1)(m-2)}{m^2}} \qquad (6)$$

where g(m) is the constant when m is given.And from the initial condition $T_m(m) = \frac{m(m-1)}{2}$, we get

$$g(m) = \frac{(m^2 - \frac{(m-1)(m-2)}{m^2})(m-1)(m-2)}{6m-4}$$

Now,we obtain the exact expression of $T_s(t)$

$$T_s(t) = \frac{m-1}{\frac{1}{2} - \frac{2}{3}\frac{(m-1)(m-2)}{m^2}}((\frac{t}{s})^{\frac{1}{2}} - 1) + (g(m)s^{\frac{(m-1)(m-2)}{m^2}-1} + \frac{m-1}{1-\frac{(m-1)(m-2)}{m^2}}(\frac{t}{s})^{\frac{2}{3}\frac{(m-1)(m-2)}{m^2}} \qquad (7)$$

Then,for node *s*,we obtain

$$C_s(t)=\frac{T_s(t)}{\frac{k_s(t)(k_s(t)-1)}{2}}=\frac{\frac{m-1}{\frac{1}{2}-\frac{2}{3}\frac{(m-1)(m-2)}{m^2}}((\frac{t}{s})^{\frac{1}{2}}-1)}{\frac{m^2\frac{t}{s}-m(\frac{t}{s})^{\frac{1}{2}}}{2}}+\frac{(g(m)s^{\frac{(m-1)(m-2)}{m^2}-1}+\frac{m-1}{1-\frac{(m-1)(m-2)}{m^2}})(\frac{t}{s})^{\frac{2}{3}\frac{(m-1)(m-2)}{m^2}}}{\frac{m^2\frac{t}{s}-m(\frac{t}{s})^{\frac{1}{2}}}{2}}$$

By(2)and(7),we obtain $\overline{C}_k(N)$, the average clustering coefficient of nodes whose degree is k.

$$\overline{C}_k(N)=\frac{\frac{m-1}{\frac{1}{2}-\frac{2}{3}\frac{(m-1)(m-2)}{m^2}}(\frac{k}{m}-1)}{\frac{k^2-k}{2}}+\frac{(g(m)\frac{k^2}{m^2N}^{1-\frac{(m-1)(m-2)}{m^2}}}{\frac{k^2-k}{2}}+\frac{\frac{m-1}{1-\frac{(m-1)(m-2)}{m^2}})(\frac{k^2}{m^2})^{\frac{2}{3}\frac{(m-1)(m-2)}{m^2}}}{\frac{k^2-k}{2}}$$

When $N\to\infty$,we obtain

$$\overline{C}(k)=\frac{\frac{m-1}{\frac{1}{2}-\frac{2}{3}\frac{(m-1)(m-2)}{m^2}}(\frac{k}{m}-1)}{\frac{k^2-k}{2}}+\frac{\frac{m-1}{1-\frac{(m-1)(m-2)}{m^2}})(\frac{k^2}{m^2})^{\frac{2}{3}\frac{(m-1)(m-2)}{m^2}}}{\frac{k^2-k}{2}} \qquad (8)$$

and as we see, at large *N*, $\overline{C}_k(N)$ is independent of *N*.

Further,we can calculate the average clustering coefficient of the whole network

$$C(m)=\int_m^{\infty}P(k)\overline{C}(k)dk$$

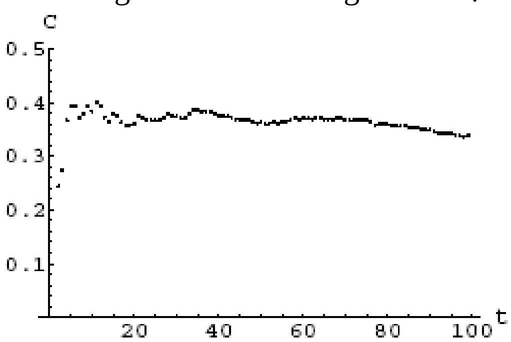
From (2)and(8), it is easy to see that for given *m* the average clustering coefficient *C* is a constant independent of *N*.

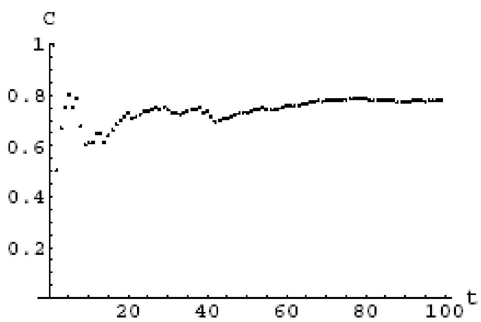
## 4. Computer simulations

In order to check our analytical result we have performed numerical simulations of our model. See the following figures.

Figure1: Clustering coefficient changes with $\sqrt{t}$ when $m0 = 6$,$m = 6$ and $t < 100000$

Figure2 :Clustering coefficient changes with $\sqrt{t}$ when $m0 = 2$,$m = 2$ and $t < 100000$




## 5. Conclusion

Both the theory and simulation results show that we have proposed a model that has the high clustering coefficient and power-law degree distribution simultaneously. Both the degree distribution and the triangle distribution depend on the network size and when the size of the network tends to infinity, the average clustering coefficient *C*(*N*) independent of the network size *N*.

## Acknowledgements

This work is supported by the Fundamental Research Funds for the Central Universities (14CX02177A).

## References

[1] Newman M E J. Complex systems: A survey [J]. Am. J. Phys. 79, 800-810, 2011.
[2] S.N.Dorogovtsev and J.F.F. Mendes.Evolution of networks.Adv.Phys.51, 1079-1187, 2002.
[3] M.E.J.Newman. The structure and function of complex networks.SIAM Review 45,167-256, 2003.
[4] Erdos P,Renyi A. On random graphs. Publications Mathematicae, 6, 290-297, 1959.
[5] M.Faloutsos,P. Faloutsos, and C.Faloutsos. On Power Law Relationships of the Internet Topology. Comput.Commun.Rev.29:251-262, 1999.
[6] A.L.Barabasi and R.Albert. Emergence of scaling in random networks. Science286, 509-512, 1999.
[7] Erzsebet Ravasz, Albert-Laszlo Barabasi. Hierarchical Organization in Complex Networks. Physical Review E67, 026112, 2003.
[8] Sara Nadiv Soffer and Alexei Vazquez. Clustering coefficient without degree correlations biases, Physical Review E71, 057101, 2005.
[9] Watts D.J, Strogatz S H. Collective dynamics ofsmall worldnetworks. Nature, 393, 440-442, 1998.
[10] R.Cohen and S.Havlin. Scale-Free Networks Are Ultrasmall. Phys.Rev.Lett.90:058701,2003.
[11] Janusz A.Holyst,Julian Sienkiewicz,Agata Fronczak,Piotr Fronczak and Krzysztof Suchecki. Universal scaling of distances in complex networks,Physical Review E 72, 026108,2005.
[12] R.Albert, H.Jeong, and A.-L.Barabasi. Diameter of the World Wide Web. Nature 401: 130 -131, 1999.