A New Network Model Based on Predictor

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

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

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

Abstract

Due to the rapid development of network technology, traditional network models are difficult to characterize the topology of current networks. The theory of complex networks provides new ideas for the study of the topology of real networks, while there is considerable interest in the link prediction of complex networks. In this paper we give a model of complex networks with predictor, and prove that the new network model has a stationary degree distribution. In addition, empirical research shows that the model has a high clustering coefficient when the size of the network tends to infinity.

Keywords

complex networks,predictor, stationary distribution of the degree.

1. Introduction

As an important tool of many fields, networks are attracting more and more attentions. It is showed that the topological structures will directly lead to different network performances[1-4]..

In 1999, Barabási and Albert [5-6]pointed out the importance topological properties of a real network: the degree distribution of network obeys power-law distribution, i.e. $P(k) \propto k^{-\gamma}$,where γ is a parameter of power-law distribution, regardless of the size of the networks. And the network with such a feature is called scale-free network. The scale-free networks has different degree distribution characteristics with other random networks: it's the vast majority of nodes connected to only a small number of other nodes, but there are some nodes are connected with the large number of nodes. So we can not use the random network model to describe the topology of the Internet, because its node degree distribution does not obey Poisson distribution. Therefore, the researchers believe that there must be some organization in complex networks of their own laws, and it is different from the principle of random networks[7], and certainly not in line with the nature of the rules networks[8].

Many social, biological and information systems can be well described as a network in which the nodes represent individual,and links represent the relationship or interactions between the nodes. Therefore, the study of complex networks has become the focus of many branches of science. Link prediction[9] was proposed as an important link between network structure and network evolution,. Link prediction algorithm can also predict the evolution of the network links that may arise. For example, on online social networks, it can help users find new friends, there by increasing their loyalty to the site.

In this paper we study some stationary properties of the complex networks. Firstly, we proposed a Scale-free model of complex networks, and its initial condition is similar with the BA model. We proved that this model has a stationary degree distribution in line with the actual characteristics of the real network. Furthermore, we introduced the predictors to improve the model with the idea of link prediction, it also proves that the new model has a stationary degree distribution. Then the network is simulated by computer programming and we verifies the corresponding properties. In addition, network simulation shows that our model has a larger clustering coefficient . After the study, we will verify the other features of networks one by one.

2. Scale-free model with its degree distribution

We consider a growth model connected by undirected edges, which satisfies the following conditions:

(1) Assuming it is a complete graph of m nodes at the initial moment;

(2)A new node with m edges is added to the network per unit time: the first edge of the new node connects to the existing node with preference probability, i.e. the probability of a new node connects the node i is proportional to the degree of node i k_i, that is $\frac{k_i}{\sum_j k_j}$. the remaining m-1 edges of the new node connects to the neighbor nodes of node i equiprobably.

As can be seen from the model, at time t, a node s may be connected by new node in two ways: (a) be connected to the first edge of the new node by the probability of $\frac{k_s}{\sum_j k_j}$; (b) if the neighbor node of node s has been connected to the first edge of a new node, then node s is connected to the new node by the probability of $\frac{m-1}{k_j}$.

So we can draw on differential equations with $k_s(t)$, such as equation (2-1):

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

in which $j\in(s,j)$ represent all the neighbors of node j. Known from the initial conditions, node s is added to the network by $k_s(s)=m$ at the initial time s, and at the time of t, the total degrees of nodes in the network is $\sum_j k_j = 2mt+m(m-1)$. So

$$\frac{dk_s(t)}{dt}=\frac{k_s(t-1)}{2\mathrm{t}+m-1} \tag{2-2}$$

Then, according to the initial conditions,

$$k_s(t)=m(\frac{2t+m-1}{2s+m-1})^{1/2} \tag{2-3}$$

By (2-3), we can see that: if t is bigger, and s is smaller, then the degree of the node is greater. And it is precisely a good explanation of "Matthew Effect", which is the phenomenon of rich get richer in network science. However, random network can not explain. (2-3) also shows that the probability of the degree of a node that less than a fixed value k is

$$P(k_s(t)<k)=P(s>\frac{(2t+m-1)m^2}{2\mathrm{k}^2}+\frac{1-m}{2}) \tag{2-4}$$

Assume that new node is added to the network by equal time intervals, so

$$P(s)=\frac{1}{m+t} \tag{2-5}$$

From (2-4) and (2-5), we obtain

$$\begin{aligned}P(s>\frac{(2t+m-1)m^2}{2k^2}+\frac{1-m}{2}) &= 1-P(s\le\frac{(2t+m-1)m^2}{2k^2}+\frac{1-m}{2})\\ &=1-(\frac{(2t+m-1)m^2}{2k^2}+\frac{1-m}{2})P(s)\\ &=1-(\frac{(2t+m-1)m^2}{2k^2}+\frac{1-m}{2})\frac{1}{m+t}\end{aligned} \tag{2-6}$$

So, the degree distribution $P(k)$ is

$$P(k)=\frac{\partial P(k_s(t)<k)}{\partial k}=\frac{(2t+m-1)m^2}{m+t}\frac{1}{k^3} \tag{2-7}$$

When the size of the network tends to infinity, i.e. $t\to\infty$, the model has a stable distribution of the degree. This is consistent with the results in the literature [9].

3. New Scale-free model added predictor

Since we need to consider much influence between the existing nodes and the new nodes during the process of the formation of the Scale-free model, whether the link between the nodes is also influenced by other factors. Here, we introduce the predictive factor to consider its impact on network model. So this new model would meet the following conditions:

(1) It is a complete graph of m nodes at the initial moment;

(2)A new node with m edges is added to the network per unit time: the first edge of the new node connects to the existing node with preference probability, meanwhile it is influenced by predictors α, i.e. the probability of a new node connects the node i is relate to α and proportional to the degree of node i ($k_i+\alpha$), that is $\frac{k_i+\alpha}{\sum_j(k_j+\alpha)}$. the remaining m-1 edges of the new node connects to the neighbor nodes of node i equiprobably.

As can be seen from the model, a node s may be connected by new node in two ways: (a) be connected to the first edge of the new node by the probability of $\frac{k_s+\alpha}{\sum_j(k_j+\alpha)}$; (b) if node j, the neighbor node of node s, has been connected to the first edge of a new node, then node s is connected to the new node by the probability of $\frac{m-1}{k_j}$.

So we can draw on differential equations with $k_s(t)$, such as equation (3-1):

$$\frac{dk_s}{dt}=\frac{k_s+\alpha}{\sum_j(k_j+\alpha)}+\sum_{j\in(s,j)}\frac{k_j}{\sum_j(k_j+\alpha)}\frac{m-1}{k_j} \tag{3-1}$$

where $j \in (s, j)$ represent all the neighbors of node j. According to the initial conditions, node s is added to the network by $k_s(s) = m$ at the initial time s, and at the time of t, the total degrees of nodes in the network is $\sum_j k_j = 2mt + m(m-1)$. So

$$\frac{dk_s}{dt} = \frac{\mathrm{m}k_s + \alpha}{2m\mathrm{t} + m(m-1) + (m+t)\alpha} \tag{3-2}$$

Solving the differential equations (3-2),

$$\frac{1}{m}\ln(mk_s + \alpha) = \frac{1}{2m+\alpha}\ln[(2m+\alpha)t + m(m-1+\alpha)] + C \tag{3-3}$$

where C is a constant, depending on the initial conditions, substitute $k_s(s) = m$ into the equation (3-3),

$$C = \ln \frac{(m^2+\alpha)^{\frac{1}{m}}}{[(2m+\alpha)s + m(m-1+\alpha)]^{\frac{1}{2m+\alpha}}} \tag{3-4}$$

Therefore, substitute the (3-4) into (3-3), we obtain:

$$k_s(t) = \frac{m^2+\alpha}{m}\left[\frac{(2\mathrm{m}+\alpha)t + m(m-1+\alpha)}{(2\mathrm{m}+\alpha)s + m(m-1+\alpha)}\right]^{\frac{m}{2m+\alpha}} - \frac{\alpha}{m} \tag{3-5}$$

It is similar to the result of the model without adding the predictor, (3-5) can also explain the phenomenon of rich get richer precisely. However, random networks and rule networks can not explain.

According to (3-5), the probability of the degree of a node that less than a fixed value k is

$$P(k_s(t) < k) = P\left(s > \frac{[(2m+\alpha)t + m(m-1+\alpha)](m^2+\alpha)^{\frac{2m+\alpha}{m}}}{(2m+\alpha)(mk+\alpha)^{\frac{2m+\alpha}{m}}} - \frac{m(m-1+\alpha)}{2m+\alpha}\right) \tag{3-6}$$

And assume that new node is added to the network by equal time intervals, so

$$P(s) = \frac{1}{m+t} \tag{3-7}$$

From (3-6) and (3-7), we obtain

$$P\left(s > \frac{[(2m+\alpha)t+m(m-1+\alpha)](m^2+\alpha)^{\frac{2m+\alpha}{m}}}{(2m+\alpha)(mk+\alpha)^{\frac{2m+\alpha}{m}}} - \frac{m(m-1+\alpha)}{2m+\alpha}\right)$$

$$= 1 - P\left(s \le \frac{[(2m+\alpha)t+m(m-1+\alpha)](m^2+\alpha)^{\frac{2m+\alpha}{m}}}{(2m+\alpha)(mk+\alpha)^{\frac{2m+\alpha}{m}}} - \frac{m(m-1+\alpha)}{2m+\alpha}\right) \quad (3\text{-}8)$$

$$= 1 - \left\{\frac{[(2m+\alpha)t+m(m-1+\alpha)](m^2+\alpha)^{\frac{2m+\alpha}{m}}}{(2m+\alpha)(mk+\alpha)^{\frac{2m+\alpha}{m}}} - \frac{m(m-1+\alpha)}{2m+\alpha}\right\} P(s)$$

$$= 1 - \left\{\frac{[(2m+\alpha)t+m(m-1+\alpha)](m^2+\alpha)^{\frac{2m+\alpha}{m}}}{(2m+\alpha)(mk+\alpha)^{\frac{2m+\alpha}{m}}} - \frac{m(m-1+\alpha)}{2m+\alpha}\right\} \frac{1}{m+t}$$

So, the degree distribution $P(k)$ is

$$P(k) = \frac{\partial P(k_s(t) < k)}{\partial k} = \frac{[(2m+\alpha)t+m(m-1+\alpha)](m^2+\alpha)^{2+\frac{\alpha}{m}}}{m+t} \frac{1}{(mk+\alpha)^{3+\frac{\alpha}{m}}} \quad (3\text{-}9)$$

When the size of the network tends to infinity, i.e. $t \to \infty$, the degree distribution of the model is

$$P(k) = \frac{(2m+\alpha)(m^2+\alpha)^{2+\frac{\alpha}{m}}}{(mk+\alpha)^{3+\frac{\alpha}{m}}} \quad (3\text{-}10)$$

From (3-10), we can find that degree distribution of the network is approximately proportional to $\mathrm{k}^{-(3+\frac{\alpha}{m})}$, denoted

$$P(k) \sim \mathrm{k}^{-(3+\frac{\alpha}{m})} \quad (3\text{-}11)$$

It is a distribution with a power-law decay tail, so it also has a behavior of power-law. Similarly, the model has a stable distribution of the degree.

So, by deduction, we proved that the new Scale-free model after adding predictors also has a stable distribution of the degree.

4. Results analysis and discussion by computer simulation

To test the theoretical results, we simulate the evolution of the network by computer programming, then we calculate the degree distribution and clustering coefficient of the network, and drawing based on data obtained by simulation.

Firstly, we can plot the graphics of $P(k) - k$ according to the data obtained by programming. However, we can not judge whether it presents a distribution of power-law directly form the graphics. If the degree distribution satisfies $P(k) \propto k^{-\gamma}$, then taking the logarithm of its two sides by using the idea of replacing the curve by straight lines, we obtain:

$$\ln(P(k)) \propto -\gamma \ln(k) \quad (4\text{-}1)$$

So the graphics about $\ln(P(k))-\ln(k)$ would be a monotonically decreasing line based on the data obtained, which is easy to distinguish. Therefore we must plot the graphics of $\ln(P(k))-\ln(k)$. Finally, plot the graphics of $C(t)-t$, where $C(t)$ is the clustering coefficient of the network.

4.1 Graphics of degree distribution

Plot the graphics of $P(k)-k$ and $\ln(P(k))-\ln(k)$ according to the data obtained by programming.

Let m0 = 6, m = 6, α = -1, sumN = 10000, draw the graphics of degree distribution of the model, as shown in Fig. 4-1.

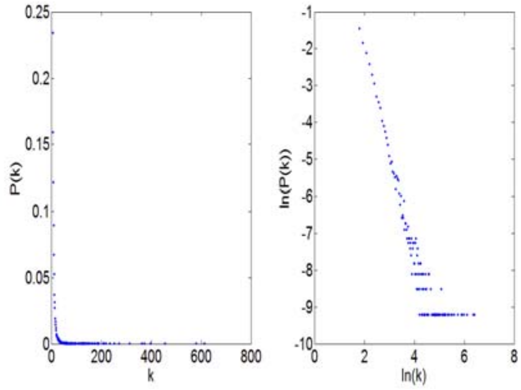

Fig. 4-1. degree distribution when m0 = 6, m = 6, α = -1, sumN = 10000

(2) Let m0 = 10, m = 10, α = -1, sumN = 10000, draw the graphics of degree distribution of the model, as shown in Fig. 4-2.

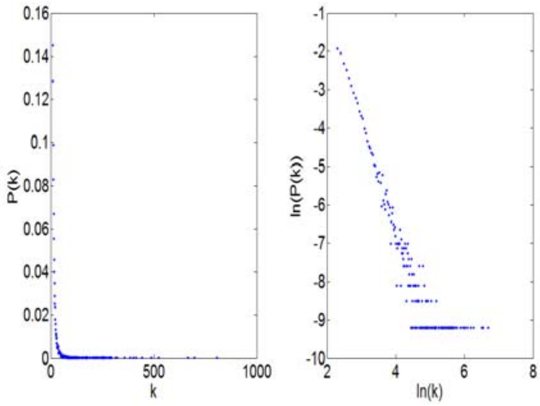

Fig. 4-2. degree distribution when m0 = 10, m = 10, α = -1, sumN = 10000

(3) Let m0 = 10, m = 10, α = -1, sumN = 50000, draw the graphics of degree distribution of the model, as shown in Fig. 4-3.

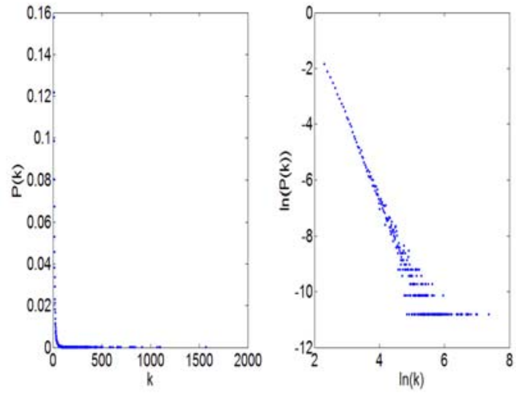

Fig. 4-3. degree distribution when m0 = 10, m = 10, α = -1, sumN = 50000

By comparing the above three figures, Fig. 4-1, Fig. 4-2 and Fig. 4-3 , we find that the degree distribution of the new model obeying the power law.

Comparing Fig. 4-1 and Fig. 4-2, we can find that when sumN is certain, m0 and m is different, the relationship between $P(k)$ and k is different, but they both present in the form of a power law.

Comparing Fig. 4-2 and Fig. 4-3, we can find that when m0 and m are certain, and sumNs are great but different, the relationship between $P(k)$ and k is basically the same. It provides further evidence of the model with stable distribution of degree. That is, when the size of network tends to infinity, the distribution of degree is only related with the degree k regardless of the size of network.

4.2 Graphics of clustering coefficient

Plot the graphics of $C(t)-t$ according to the data obtained by programming.

(1) Let m0 = 6, m = 6, α = -1, sumN = 10000, draw the graphics of clustering coefficient of the model, as shown in Fig. 4-4.

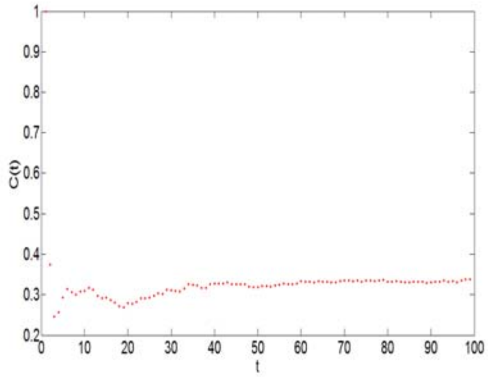

Fig. 4-4. clustering coefficient when m0 = 6, m = 6, α = -1, sumN = 10000

(2) Let m0 = 10, m = 10, α = -1, sumN = 10000, draw the graphics of clustering coefficient of the model, as shown in Fig. 4-5.

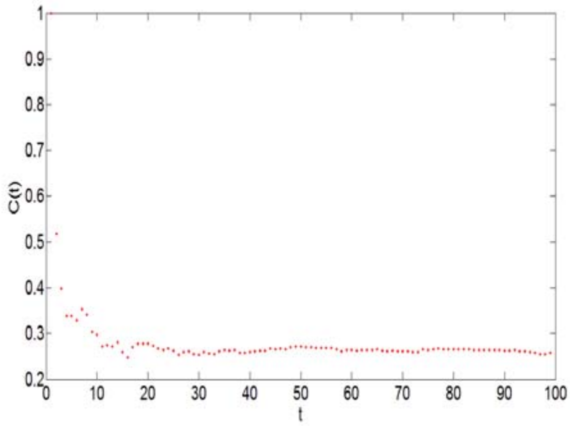

Fig. 4-5. clustering coefficient when m0 = 10, m = 10, α = -1, sumN = 10000

(3) Let m0 = 10, m = 10, α = -1, sumN = 50000, draw the graphics of clustering coefficient of the model, as shown in Fig. 4-6.

By comparing the above three figures, Fig. 4-4, Fig. 4-5 and Fig. 4-6, it is easy to see that when the size of the network tends to infinity, clustering coefficient tends to stable, and it is a constant which is independent of the size of network. This is also consistent with the conclusions of scholars. While in the random network, when the size of network tends to infinity, the clustering coefficient tends to 0. Thus, the model has a greater clustering coefficient.

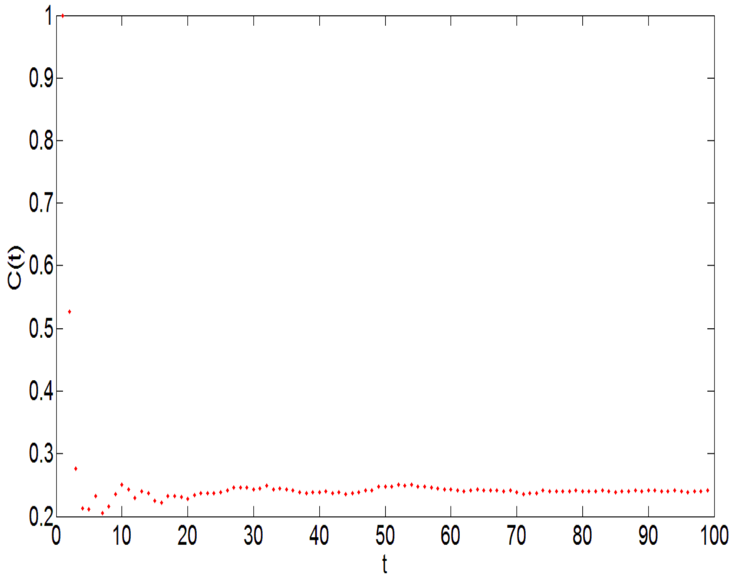

Fig. 4-6. clustering coefficient when m0 = 10, m = 10, α = -1, sumN = 50000

5. Conclusion

In this paper, we propose a Scale-free model in complex network, and verify it by both the theory method and computer simulation method. Then, we introduce the predictor, and make link prediction in the established model. The results show that the model we have established has a stationary distribution of degree, and the new model adding predictors also has the same properties. Meanwhile, the numerical experiments show that the clustering coefficient stabilizes, which is consistent with the topological properties of the real networks.

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(2011), 800-810.
[2] Barabási A L.Scale-free networks: a decade and beyond[J]. Science(2009), 325(5939): 412-413.
[3] Albert R, Barabási A L. Statistical mechanics of complex networks[J]. Reviews of Modern Physics(2002), 74(1): 47.
[4] Newman M E J. Models of the small world[J]. Journal of Statistical Physics(2000), 101(3-4): 819-841.
[5] Barabási A L, Albert R, Jeong H. Mean-field theory for Scale-free random networks [J]. Physica A: Statistical Mechanics and its Applications (1999), 272(1): 173-187.
[6] Barabási A L, Albert R. Emergence of scaling in random networks[J]. Science(1999), 286(5439): 509-512.
[7] Watts D J, Strogatz S H. Collective dynamics of ‘small-world’ networks [J]. Nature (1998), 393: 440-442.
[8] Newman M E J. Models of the small world [J]. Journal of Statistical Physics (2000), 101(3-4): 819-841.
[9] Lv L Y, Zhou T. Link prediction in complex networks: a survey [J]. Physica A: Statistical Mechanics and its Applications(2011), 390(6): 1150-1170.