An Improved B-A Model for Scale-free Network Evolution Model

Hongyan Wei¹, Jiangong Wang²* and Tianqi Wang²

¹Beijing Institute of Information Technology, Beijing, 100094, China
²Troop 31401 PLA, Jinan, Shandong, 250014, China.

*Corresponding author’s e-mail: hw12xian@hotmail.com

Abstract. Scale free network can describe complex network very well. Barabasi and Albert proposed B-A model, which reveals the nature of many complex phenomena in the real world. The generated network degree distribution satisfies power-law distribution. In this paper, two shortcomings of the model are found through careful study, and an improved scale-free network evolution model is proposed. The experimental data show that the improved model has significant scale-free characteristics and can better reflect some characteristics of complex networks.

1. Introduction

The study of complex networks has a long history. It began with the famous mathematician Euler’s study of Konigsberg’s seven bridge problem, which opened a new chapter in the study of complex networks with graph theory as a tool. Then the random graph theory established by Erdös and Renyi, a Hungarian mathematician, started the systematic study of complex network theory. Since then, the random graph theory has been the basic theory of complex network research. Their E-R model is widely used as a complex network model. In 1998, Watts and Strogatz proposed the famous small world model. In 1999, Barabási, Albert et al. Studied a world wide network subnet with 325729 nodes, and found that although the world wide network has a short average path similar to the small world, its degree distribution is not a symmetric Poisson distribution like the random network and the small world network, but a power-law distribution. Based on this finding, they proposed a scale-free model (B-A model).

The B-A model inevitably has some obvious limitations compared with the real network. For example, the actual network often has some non-power law characteristics, such as exponential truncation, small variable saturation and so on. Because of the limitation of B-A model, there have been a lot of extended model research based on BA model. Many important extension models are made by modifying the Ba and the preferred connection mode of the model. Bianconi and Barabási (2001) introduced the concept of fitness, and proposed fitness model[3], in which the probability of priority connection is directly proportional to the product of node’s degree and fitness [4]. Li Xiang and Chen Guanrong (2003) proposed the local world evolutionary network model, which is different from the BA model in that the newly added nodes are not selected from all local nodes, but from the local world according to the priority connection principle [5]. Tan Jinsong and he Zheng (2009) built five cluster self-organization models based on different connection modes between enterprises and conducted simulation research [6]. In this paper, the shortcomings of B-A model are analyzed, and the improved B-A model is proposed. The advantages of the improved model are verified by simulation.
2. B-A model and the shortcomings

Scale free model (B-A), as a model of complex network, has been more and more widely recognized by people, and has been applied to many scientific fields, such as biology, communication technology, social science, economics and so on. Metabolic network, actor network, author network of scientific and technological literature, web network, Internet routing network and other complex systems can be abstracted into complex network models. A large number of studies show that they all have scale-free characteristics—the degree distribution of network nodes meets the power-law distribution.

\[ p(k) = C_k k^{-\gamma} \]  

in the formula, \( C_k \) is a constant, and \( \gamma \) is degree distribution index.

Barabasi and Albert proposed scale-free model by studying the real world network. The scale-free network model has the following two characteristics: growth characteristics, and optimal connection characteristics.

The network generation process is as follows.

Step 1: growth. Start with a small number of nodes \( m_0 \), add a new node with \( m(\leq m_0) \) edges at each time interval, and connect to the nodes that already exist in the system.

Step 2: select the best connection. When selecting a node connected to a new node, assuming that the probability of the new node connected to the node i depends on the connectivity \( k_i \) of the node:

\[ \prod(k_i) = \frac{k_i}{\sum k_j} \]  

B-A model reveals the essence of many complex phenomena in the real world. The degree distribution of the generated network satisfies the power-law distribution, and the degree distribution index \( \gamma = 2.9 \pm 0.1 \), while the degree distribution index of the real network is generally \( 2.1 < \gamma < 4 \). This model is consistent with the real network, but it also has some limitations. Firstly, for the initial network, there is no definite limit, only \( m_0 \) isolated nodes. However, the real network always has an initial basic network and continues to expand on the basis of the basic network to form a larger network; secondly, each time interval adds a new node with \( m(\leq m_0) \) edges, which is not consistent with the real world network. The number of edges of new nodes should not be fixed. Each new node joining the network should establish a connection with the network nodes according to its own needs, which should not be fixed.

In view of the shortcomings of B-A model, this paper proposes an improved B-A model.

3. Improved B-A network model

There are two shortcomings of B-A: the first one is that \( m_0 \) isolated nodes are used as the initial network; the second is that the number of sides added at each time is fixed. This is not consistent with the real world network generation process. In view of these two shortcomings, this paper proposes the improvement of B-A model, first of all, change the initial network of the model, and take two connected nodes as the initial network of the network; secondly, add nodes every time, nodes and each existing node have the possibility of establishing a connection. The probability is determined by the degree of the existing nodes, and the larger the degree, the greater the probability of establishing the connection. According to these two principles, the specific algorithm of model generation is as follows

Step 1: is to select two connected points as the starting network.

Step 2: add new nodes at each time, and connect the new nodes with the existing nodes according to the probability \( (\prod) \)

\[ \prod(k_i) = \text{prob} \times \frac{k_i}{(t+2)} \]  

In formula (3), \( \text{prob} \) is a constant (usually \( \text{prob} \leq 0.6 \)), \( k_i \) is connectivity, \( t \) is the current time, \( t+2 \) is the number of nodes in the current network.
Step 3: whether to join the network when the node is disconnected. If you join, the number of nodes in the network will increase by 1. Otherwise, the node will be abandoned.

Step 4: repeat steps 2 and 3 until the required number of network nodes is generated.

This algorithm takes two connected nodes as the initial network, which is in line with the real world network events, and also solves the problem of the uncertainty of the initial network in the B-A model. Based on the initial network, add nodes to expand the network. Each time a new node is added, the newly added node establishes the connection with the existing nodes in the network according to the probability \( \prod (k_i) = \text{prob} \times k_i / (t + 2) \). If no connection is established with the network, the node is automatically discarded until the required number of network nodes is generated. This kind of network expansion mode embodies the characteristics of priority connection, and also conforms to the objective reality. The number of connections established by the newly added nodes is determined by the probability, not by the fixed number of sides, which solves the second problem proposed in this paper for B-A model.

The simulation results show that the improved B-A network model has the following characteristics.

1. The generated network has scale-free characteristics, and the degree distribution follows power-law distribution.

2. The degree distribution index of the model is 2.5<\( \gamma <3.5 \).

3. The generated network is a sparse network with a certain amount or a small number of hub nodes. It is in line with many network forms in the real world.

4. Experimental data analysis

4.1 Model degree distribution test

There are two ways to test the degree distribution of the model: one is to generate a large network and take the node degree analysis in the middle; the other is to generate a large number of fixed node networks to consider the degree distribution jointly. Considering two reasons, this paper adopts the second method. The first reason is that the memory space of the computer is limited, and the large-scale network can’t be generated (the number of nodes exceeds 100000); the second reason is that the degree distribution tends to be stable, the linear effect of the simulation curve is good, and it is easy to observe.

100,000 networks are generated, and the number of nodes in each network is 500. According to statistics, the number of k nodes in all networks is \( n_k \), and the degree distribution is \( p(k) = n_k / 500\times10^4 \). Figure 1 shows the overall degree distribution of the network with 100,000 nodes of 500 in the double logarithmic coordinate. The X axis is the node degree, and the Y axis is the degree distribution \( p(k) \). It can be seen from the figure that the simulation curve shows good linearity when k is large. In other words, when the k value is large, the network degree distribution generated by the network generator obeys the power-law distribution.

![Figure 1: The overall degree distribution of a network](image-url)
In addition, the model in this paper is used to generate networks with 10000 and 20000 nodes respectively. The simulation results of their degree distribution are shown in Figure 2 (a, b). Figure 2 (c) shows the distribution of the node output of China Education Network (366442 nodes, 540775 sides). It can be observed from the simulation diagram that the tail of their degree distribution curve is power-law distribution.

![Figure 2](image_url)

Figure 2. simulation results of degree distribution

4.2 Estimate the degree distribution index range of the model

The degree distribution of scale-free network obeys power-law distribution, and its expression is \( p(k) = C k^{-\gamma} \). The degree distribution index \( \gamma \) is an important index to evaluate the network, which directly reflects some properties of the network.

1) when \( 0 \leq \gamma \leq 1 \), the network is a non-sparse network, that is, the number of edges is close to fully connected.

2) when \( 1 < \gamma \leq 2 \), the network is sparse, but there are many high node nodes (hub nodes).

3) when \( 2 < \gamma \leq 3 \), the network is sparse, and there are some high degree nodes (hub nodes).

4) when \( \gamma > 3 \), the network is sparse, and there are few high degree nodes (hub nodes) in the network.

| prob | \( \gamma \) |
|------|-------------|
| 0.10 | 3.08085     |
| 0.15 | 3.0271572   |
| 0.20 | 2.5708912   |
| 0.25 | 3.14372     |
| 0.30 | 2.552868    |
| 0.35 | 2.713744    |

![Figure 3](image_url)

Figure 3. calculation simulation curve

In this paper, the method of line fitting is used to estimate the degree distribution index of the network. Calculate the network degree distribution index and calculation square under different prob values. The average value of the degree distribution index of 1000 networks is calculated. The value of prob is 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6. Table 1 shows the results of calculation.
and simulation. Figure 3 is the calculation simulation curve, and the X axis is prob, the Y axis is value of average distribution index. Obviously, the degree distribution index of scale-free network generated by the model is up and down in 3.

5. conclusion
The improved B-A network model proposed in this paper, the degree distribution satisfies the power-law distribution, has a good scale-free characteristic. From the principle of network generation, it is more in line with the generation law of the real network, and can reflect the characteristics of the complex network well. It has a certain significance for the research of the complex network.

References
[1] Watts, D.J. and Strogatz, S.H. (1998) Collective dynamics of “small world” networks. Nature, 393, 440-442.
[2] Albert R, Barabási A L. Emergence of Scaling in Random Networks[J].Science, 1999,286.
[3] Bianconi G, Barabási A L. Bose-Einstein Condensation in Complex Network[J]. Phys. Lett., 2001,86.
[4] Li X, Chen G. A Local World Evolving Network Model[J]. Physica A, 2003,328.
[5] Newman M E J. The Structure and Function of Complex Networks[J]. Society for Industrial and Applied Mathematics Review, 2003,(45).
[6] Dorogovtsev S N, Mendes J F F, Samukhin A N. Structure of Growing Networks with Preferential Linking[J]. Phys. Rev. Lett., 2000, (85).
[7] Shi D H, Chen Q H, Liu L M. Markov Chain-based Numerical Method for Degree Distributions of Growing Networks[J]. Phy Rev. E, 2005, (71).