

## The Research of Complex Networks and Its Modeling

Wen-cheng JIAO, Shuai HAN, Pei-zhang CUI, Xiao-hui Li

Dept. of Information Engineering, Ordnance Engineering College, Shijiazhuang Hebei 050003, China

jaowc@163.com

**Keywords:** complex networks, model

**Abstract.** The measurement of complex networks is based on graph theory. Firstly, the basic character of complex networks was analyzed. Then complex networks models such as ER model, small-world model, BA scale-free model were researched. Its construct rule, evolution process and dynamics characteristic were also analyzed.

### Introduction

The purpose of modeling is to simulate the real system as a network, through the network to study the complexity of the system. The measurement of the complex network is based on the graph theory, and the dynamic topological properties of the nodes and edges are calculated. Therefore, for the modeling of complex systems, it is necessary to have a basic knowledge of the dynamical properties of the graph theory and the system, for common complex network need to master its model and basic characteristics.

### Basic Properties of Complex Networks

**Graph Representation of Networks.** Graph is a data structure consisting of point sets and edge sets denoted by  $G = (V, E)$ . A concrete network can be abstracted as a graph composed of nodes and edges. Graphs are usually represented by arrays, adjacency lists, adjacency matrices, crisscross lists, and multiple linked lists. If there is an edge between node  $i$  and node  $j$ , and  $(i, j)$  and  $(j, i)$  correspond to the same edge, the network is undirected graph, otherwise it is directed graph. The connections between nodes are shown in Equation (1).

$$A_{ij} = \begin{cases} W_{ij} & \text{If there is an edge between node } i \text{ and node } j \\ 0 & \text{If there is no edge between node } i \text{ and node } j \end{cases} \quad (1)$$

$W_{ij}$  denotes the connection strength between nodes which is the weight of the edge. According to the weight of the edge, it can be divided into weighted network and unweighted network. Matrix is symmetric when a graph is an undirected graph otherwise asymmetric. Sequence of node and edge between two vertices is known as the path. The number of edges is called path length. If any vertices in a graph are connected, the graph is a connected graph.

**Degree, Degree Distribution and Correlation.** The degree of node  $i$  refers to the number of other nodes connected to the node in network. In real network, the degree of node  $i$  means the importance of node  $i$  to other nodes or the proportion of node  $i$  occupied in the network. It also describes the degree of its influence and location in network. The mean of network nodes average degree refers to the average of all nodes' degree in a whole network. On the basis of node degree, the degree distribution of nodes is described by distribution function  $P(k)$ , which shows the probability that the degree of a node is chosen as  $k$  randomly. And it also shows the probability of  $k$ -connected edges in the real network which is shown in Equation (2).

$$P(k) = \sum_{k=k'}^{\infty} P(k') \quad (2)$$

Degree distribution is mainly used to measure the characteristics of each node in network. Degree and degree distribution is an important scale usually used to distinguish the type of network. Correlation is used to describe the relationship between nodes.

**Average Path Length and Validity.** In complex networks, the distance between two nodes is defined as the number of edges of the shortest path between two nodes  $i$  and  $j$ , which is used to measure the shortest distance between nodes in a network. And the maximum value of distance between any node  $i$  and node  $j$  is called the diameter of a network. The average path length  $l$  of a network is the average length of the shortest distance between any two nodes in a network. As shown in Equation 3.

$$L = \frac{2}{n(n+1)} \sum_{i \geq j} d_{ij} \cdot \quad (3)$$

$n$  is the number of nodes in a network. Validity is inversely related to path length and it can be used to estimate the size of topological distance numerically. Average path length is significant in network applications and represents the efficiency of network nodes and networks.

**Clustering Coefficient and Module.** The clustering coefficient  $C_i$  of any node  $i$  in a network is defined as the ratio of the number of edges actually connected to node  $i$  in a network to the total number of edges that may be connected to node  $i$ , as shown in Equation 4.

$$C_i = \frac{2 * e_i}{k_i * (k_i - 1)} \cdot \quad (4)$$

The clustering coefficient  $C$  of the whole network is the average of clustering coefficients of all nodes in a network. As shown in Equation 5.

$$C = \frac{1}{n} \sum_{i=1}^n c_i \cdot \quad (5)$$

$k_i$  is the degree of node  $i$  in a network and  $e_i$  is the number of edges actually existing in a network between node  $i$  and its adjacent nodes. The clustering coefficient of whole network is an important parameter to measure the degree of network collectivizing. It reflects the collectivizing ability of a network.

Module is connected with cluster. It mainly refers to fixed subgraph appeared with higher frequency in a network. It is also used to describe the close interrelationship between neighbor nodes. Therefore, clustering coefficient can also be equivalently defined by triangular model, as shown in equation 6.

$$C_i = \frac{\text{The number of triangles connected to point } i}{\text{The number of triples connected to point } i} \cdot \quad (6)$$

**Betweenness and Centrality.** Betweenness is divided into vertices and edges betweenness. The betweenness of node  $i$  is the probability of all the shortest paths in a network passing through node  $i$ . Edge betweenness is the probability of shortest path that passing through the edge. The concept of betweenness describes the role and importance of nodes and edges in whole network. The centrality of node  $i$  is similar to betweenness. Nodes with high betweenness or centrality are called Hub nodes.

### Several Complex Network Models

**Regular Network.** The most typical network is nearest neighbor network and global coupling network.

### (1) Nearest Neighbour Network

Each node in network is connected to its neighbour nodes which is often called a lattice. The nearest neighbour network contains  $N$  dots that form a ring. Each node is connected to its left and right  $K/2$  neighbour nodes.  $K$  is an even number. Nearest neighbour network is highly clustered, but do not has small-world effects.

### (2) Global coupling network

In this network, each node is connected to all other nodes. In all networks with same number of nodes, global coupling network has a minimum average path length  $L = 1$  and a maximum clustering coefficient  $C = 1$ . Although global coupling network model reflects the clustering properties and small-world effects of many practical networks, its significant limitation is that in a network with  $N$  nodes is common with  $N(N-1)/2$  edges. But in real network, edge is relatively sparse. A general network has the number of edges is the order of  $N$  rather than the order of  $N^2$ .

**ER Random Graph Network Model.** ER (Erdos-Renyi) random graph model is a kind of undirected non-weighted network model. ER model construction rules are as follows: Let total  $N$  nodes in a network, each pair of nodes connected with probability  $P$ . In ER model, the maximum number of connected edges is  $N(N-1)/2$ ; if nodes are connected by probability  $P$ , the number of actual connections is  $PN(N-1)/2$ ; the distribution of degrees satisfies the binomial distribution:

$$P(k) = \binom{N-1}{K} P^K (1-P)^{N-1-K} \quad (7)$$

The average degree of the entire network is  $\bar{k} = p(N-1)$ . When  $N$  is large, the distribution expression of Eq. (7) will take the form of Poisson distribution. The degree distribution of Poisson form is unique to ER model network and has become the biggest difference from other networks. So degree distribution of network is usually one of the criteria to distinguish different networks.

In ER model network, the average shortest path length is closely related to  $N$  and  $p$ :

$$\bar{l} \propto \ln(N)/\ln[pN] \quad (8)$$

In the case of small  $P$ , the entire network has a very small clustering. The majority of nodes isolated so that the diameter  $\bar{l} = 1$ ; when  $N$  and  $p$  is large enough, the ER model network will have large-scale connectivity groups. The analysis and study of percolation theory found that when the connection probability  $p$  exceeds the threshold  $P_c \cong 1/N$ , the network will turn from disconnected to connected in  $\bar{k}_c = 1$ .

The ER model has the following network dynamics properties:

- 1) The average shortest distance of network is small;
- 2) The average clustering coefficient of network is small;
- 3) The node degree of network presents Poisson distribution;
- 4) The ability of deliberate attack resistance is strong;
- 5) The network is less fault-tolerant.

**Small-world Network Model.** Many real networks, especially social networks are clustered which leads to the study of small-world networks. The earliest small world network model is the network model (WS model) proposed by Watts and Strogatz in 1998. The model starts with a ring with  $N$  nodes. Each node on the ring has  $m$  edges connected to each side. Then randomly reconnect each edge with probability  $P$  (except self-join and re-edges), these reconnected edges are called long-range links, and the long-range connections greatly reduce the average path length of the network. It has little effect on the cluster coefficients of the network. The evolution of WS model is:

(1) Initialization: A neighboring node-coupled ring network with  $N$  nodes, each node  $i$  is connected to its  $K$  neighbors  $i \pm 1, i \pm 2, \dots, i \pm K/2$ .  $K$  is an even integer.

(2) Randomization: Randomly rewrite each edge of a network with probability  $p$ . That is to reconnect a ready edge to another fixed point with probability  $p$ , avoiding the situation of connecting to itself or with the existing edge.

Through the evolution of above process, the regular network will be a certain probability of short-range connection. Obviously, when  $p$  is very small, the properties of the whole network are very close to regular network. When  $p$  is large enough, the network is close to classical ER model network.

The small-world network constructed by WS model has the following network dynamics characteristics:

- 1) The average shortest distance of network is small;
- 2) The average clustering coefficient of network is high;
- 3) The node degree of network shows the distribution between impact function and Poisson function;
- 4) The ability of deliberate attack resistance is stronger than  $q$ ;
- 5) The network's fault tolerance is small;

Soon Newman and Watts improved WS model and proposed a variant model. It randomly select pairs of nodes as a long-range connection. Original lattice edge remain fixed. Later, Kasturirangan proposed an alternative model of WS model which also starts with a ring lattice. It adds nodes in the middle of lattice and randomly joins nodes on the lattice that act as "long-range connections" in WS model. In fact, if a new node is added in the middle of the grid and connect to enough grid edges, the network will show the character of small world.

**BA Scale-free Network Model.** The degree distribution of ER model and WS model is not consistent with many realistic networks. It is very difficult to describe these realistic networks. Barabasi and Albert had found that many complex networks have large scale of self-organizing feature by following the dynamic evolution process of World Wide Web. Node degree of most complex networks obeys the power-law distribution. Network with the power-law degree distribution is called scale-free network.

BA network model is the first stochastic scale-free network model. It is assumed that there are a few nodes in the system at the initial moment of BA model generation. A new node is added in each subsequent time interval and connect to a number of different nodes existed. When a node is selected to be connected to a new node in network, it is assumed that the probability of the selected node connecting to the new node is proportional to the degree of the selected node. This connection is called a preferred connection. BA network eventually evolved into a scale-invariant state. Node degree obey power-law distribution which index is equal to 3. The average path length of BA model is very small and the cluster coefficient is also small. But it is larger than the cluster coefficient of the same scale random graph. When network tends to infinity, the cluster coefficients of these two networks are almost zero. The realization of the algorithm is as follows:

(1) Growth characteristics: The entire network starts with an initial network with  $m_0$  nodes. It adds a new node to initial network after every interval time. And it generates  $m$  ( $m \leq m_0$ ) edges connected with nodes already exist in system.

(2) Preferred connection: When new node is added to network, it connects to nodes  $i$  with the probability  $\Pi(k_i)$ . The probability depends on the degree  $k_i$  of node  $i$ , that is:

$$\Pi(k_i) = k_i / \sum_j k_j. \quad (9)$$

The network will have  $N=m_0+t$  nodes, and  $mt$  new added edges after  $t$  time step operations. Depending on the choice of growth and preference, the network will eventually evolves into a state of constant scale: the degree distribution of network does not change with time  $t$ .

During the evolution of network, the degree  $k_i$  of node  $i$  changes with time and the change of  $k_i$  is a continuous value. According to mathematical theory, it is known that the changing rate of node degree is proportional to the connection probability  $\Pi$ . And  $k_i$  satisfies the following dynamic equation:

$$\frac{\partial k_i}{\partial t} = m \prod (k_i) = m \frac{k_i}{\sum_{j=1}^{N-1} k_j} \quad (10)$$

The sum of denominator degrees includes all nodes except the newly added node. So its value is  $\sum_j k_j = 2mt - m$ , the equation (10) becomes:

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t} \quad (11)$$

The solution of (11) is:

$$k_i(t) = m \left( \frac{t}{t_i} \right)^\beta \quad (12)$$

$\beta=1/2$ ,  $k_i(t_i)=m$ . Equation (12) shows that degree of all nodes in network are of the same evolution law. They are subject to power-law distribution. We can get the probability that the degree of one node is less than  $k_i(t)$

$$P[k_i(t) < k] = P\left( t_i > \frac{\frac{1}{m^\beta t}}{K^\beta} \right) \quad (13)$$

If the interval of newly added nodes is equal, the probability density of  $t_i$  is evenly distributed.

$$p(t_i) = \frac{1}{m_0 + t} \quad (14)$$

So

$$P\left( t_i > \frac{\frac{1}{m^\beta t}}{K^\beta} \right) = 1 - \frac{\frac{1}{m^\beta t}}{K^{\frac{1}{\beta}(t+m_0)}} \quad (15)$$

The probability distribution of node degree of the whole network can be expressed as:

$$p(k) = \frac{\partial p[k_i(t) < k]}{\partial k} = \frac{2m^{\frac{1}{\beta}} t}{t+m_0} \frac{1}{k^{\frac{1}{\beta}+1}} \quad (16)$$

When time  $t$  tends to infinity and network is in steady state, the probability distribution of  $k$  tends to

$$p(k) \propto 2 m^{\frac{1}{\beta}} k^{-\gamma} \quad (17)$$

$\gamma = \frac{1}{\beta} + 1 = 3$ . And it is independent of  $m$ . The power exponent is the parameter describing the state of network. It is independent of the system size, the number of edges and the initial number of nodes  $m_0$ . It only depends on the network topology. So the network is called scale-free network.

The network constructed by BA scale-free network model has the following network dynamics properties:

- 1) The shortest average distance of the network is small;
- 2) The network clustering coefficient is small;
- 3) Degree of network nodes shows power - law distribution;
- 4) The fault-tolerant ability of network is large;
- 5) The ability of deliberate attack resistance is less.

**Evolutionary Network Model.** BA model provides a new perspective for people to study complex systems and opens up a new situation in complex network research. But compared with real network, BA model still has obvious flaw. Real network mostly has big cluster coefficient and degree distribution index is not only equal to 3. It is located between [2, 3]. A variety of subtle changes are likely to affect network topology during the evolution process of real network, The evolution of different real networks is very different subjected to factors such as cost, competition and other factors. It is necessary to establish a specific network model for different realistic networks. So network evolution has become one of the hotspots in recent complex network research. The evolution of networks is mainly based on the establishment of dynamic model. People Identify and capture the dynamic changing process of factors which can affect network topology structure. So people can achieve the purpose of understanding network topology.

## Conclusion

Complex networks have become a hotspot in academic research. It has potential and extensive application in many fields. Constructing evolution model of small-world network and scale-free network which conform to the nature of real system, not only has important theoretical significance but also has potential application value. It has extremely important effect for people to grasp network function and its dynamic characteristics.

## References

- [1] Newman M E J. Models of the small world. *Journal of Statistical Physics*, 2000, 101:819-841.
- [2] Strogatz S H. Exploring complex networks. *Nature*, 2001, 410: 268-276.
- [3] Xu T, Chen R, He Y, et al. Complex network properties of Chinese power grid. *International Journal of Modern Physics B*, 2004, 18(17-19):2599-2603.
- [4] Verspagen B, Duysters G. The small worlds of strategic technology alliances. *Technovation*, 2004, 24: 563-571.