

Network Topologies: Analysis And Simulations

MARJAN STERJEV and LJUPCO KOCAREV
Institute for Nonlinear Science
University of California
San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0402

USA

Abstract:-In this paper we present some of the network models and topologies that have been defined in the past few years as a result of the community efforts to model real world networks. For each presented model we provide computer calculated network measures of interest and restate some conclusions that separate one network model from another.

Key-Words:- network, topology

1 Introduction

Graphs and their properties have been studied for a long time in the classic graph theory. In the past few years there was a lot of work done in order to provide models for real world networks. At the current technology level an experimental data is also available making the model verification easy and accurate. In the network research community there are already accepted and well established network models.

The rest of the paper is outlined as follows. Section 2 presents network measures of interest. Sections 3, 4 and 5 present the **Random**, **Small- World** and **Scale-Free** network models. For each model we provide computer simulation results. Section 6 concludes.

2 Network Representation and Measures

Network is usually represented as a graph $G = (N, E)$ where N is a set of **nodes** and E is a set of **edges**. Let n represents the number of nodes in N . E contains from 0 up to $\binom{n}{2} = \frac{n(n-1)}{2}$ edges.

A node **degree** is the number of edges connected to that node. In the case of directed edges someone can differentiate between **out degree** and **in degree**. The **Laplacian** matrix $L(G)$ of a graph G , where G is an undirected, unweighted graph without graph loops or multiple edges from one node to another, is $n \times n$ symmetric matrix with one row and column for each node. The matrix elements are defined as follows:

$$L_{ij}(G) = \begin{cases} \text{node degree} & \text{if } i = j \\ -1 & \text{if } i \neq j \exists \text{ edge } (i, j) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

For example, a ring lattice of 4 elements can be represented by the following Laplacian matrix:

$$L(G) = \begin{bmatrix} 2 & -1 & 0 & -1 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ -1 & 0 & -1 & 2 \end{bmatrix} \quad (2)$$

The eigenvalues of $L(G)$ are nonnegative and ordered: $0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$. Connected component is a subset of mutually reachable nodes. The number of connected components of G is equal to the number of λ_i equal to 0. In particular, $\lambda_2 \neq 0$ if and only if G is connected, i.e. there is only one connected component containing all network nodes.

Network measures of interest in our observations are:

1. Average distance between two nodes;
2. Clustering coefficient;
3. Degree distribution.

The **distance** between two nodes is defined as the number of edges along the shortest path connecting them. In our computer simulations we use the iterative **Bellman-Ford** algorithm for the shortest

distance computation. If the value of λ_2 is 0, the graph is disconnected and we take -1 as the average distance between nodes.

We use the following definition of the **clustering coefficient**. Let the node i has k_i edges that connect it to the k_i neighbors. The maximum number of edges that can exist among these k_i nodes is $\binom{k_i}{2} = \frac{k_i(k_i - 1)}{2}$. If E_i represents the number of edges that actually exist among these k_i nodes, clustering coefficient for the node i is defined as the following quotient:

$$C_i = \frac{2E_i}{k_i(k_i - 1)} \quad (3)$$

The clustering coefficient C for the whole graph is the average of all C_i 's.

The **degree distribution** $P(k)$ is defined as the probability that a randomly chosen node has exactly k edges.

The following text presents 3 network models that have been differentiated in the contemporary network topology research: **Random**, **Small-World** and **Scale-Free** network model. For each model we provide network measures of interest collected through the computer calculations and simulations.

3 Random Model

The **Random model** is the oldest network model among the models presented in this paper. It is based on the random graph theory introduced by *Paul Erdős* and *Alfréd Rényi* ([1],[2]). Their work was followed by *Béla Bollobás* [3], today's one of the leading scientists on this field. Many complex networks with unknown organizing principles appear random and are investigated using random graph theory.

Erdős and *Rényi* have defined two models. The first one is defined as a graph with N nodes and n edges, where the edges are randomly chosen from the $\frac{N(N-1)}{2}$ possible ones. In the second model two nodes are connected with probability p . These two models become equivalent in the limit $N \rightarrow \infty$. Random graph theory is most concerned by the question at what probability p a specific graph property appears. For example at what connection

probability p a graph with N nodes becomes fully connected? Are there triangles of connected nodes? The greatest discovery in the **Random** model research is that many properties of these graphs appear very suddenly, i.e. below some probability threshold the property doesn't exist, but above that threshold almost every random graph with the same number of nodes has that property.

A single node has from 0 up to $N-1$ edges. In a **Random** network model having connection probability p , the node i has degree k with probability following binomial distribution:

$$P(k_i = k) = C_{N-1}^k p^k (1-p)^{N-1-k} \quad (4)$$

We can assume that the equation (4) holds for the whole graph. In the limit $N \rightarrow \infty$ equation (4) becomes Poisson distribution with the mean value pN :

$$P(k) = e^{-pN} \frac{(pN)^k}{k!} \quad (5)$$

Another important result from the random graph theory is that if $E(k) = pN \geq \ln(N)$ almost every graph is connected. In that case the average distance between nodes has the following approximation:

$$l_{random} \sim \frac{\ln(N)}{\ln(E(k))} = \frac{\ln(N)}{\ln(pN)} \quad (6)$$

We can think of the equation (6) this way. Every node has in average pN outgoing edges. After l steps, the spanning tree containing $(pN)^l$ nodes has reached every node in the graph, i.e. $(pN)^l = N$. The average clustering coefficient in the **Random** network model is given as:

$$C_{random} = p \quad (7)$$

The equation (7) has also intuitive background because the quotient between the number of existing edges and number of all possible edges is always p for every possible node's neighborhood.

Fig.1 shows the simulation results for the **Random** network model with $N=1024$ nodes. The network contains undirected edges and does not contain loops or multiple edges between nodes. We vary the connection probability in the interval $[0,1]$. In

Fig.1-a we plot the average distance between nodes. It can be seen that bellow $p=0.01$ the average distance is -1 , i.e. graph is disconnected. In Fig.1-b we plot the average clustering coefficient that is a straight line with the slope value equal to 1. In Fig.1-c we plot the node's degree distribution. It can be seen that the curve resembles the Poisson distribution.

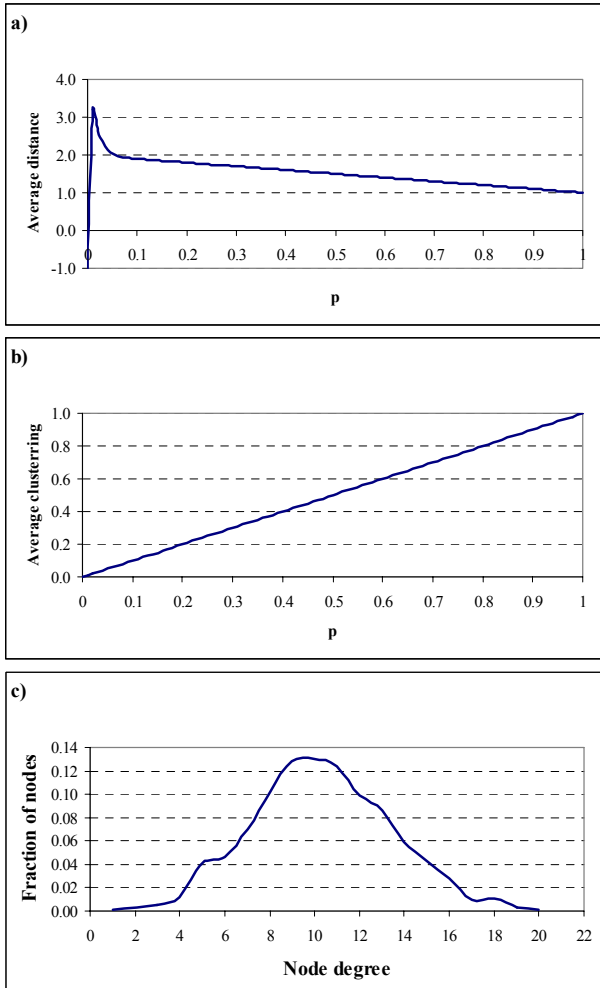


Fig.1. Random network model with $N=1024$ nodes
 a) Average distance b) Average clustering c) Degree distribution

We also created network visualization tool. Fig.2 is a visualization of the **Random** network model with the connection probability equal to 0.01. The nodes with a higher degree are positioned closer to the center.

Random Network: 1024 nodes

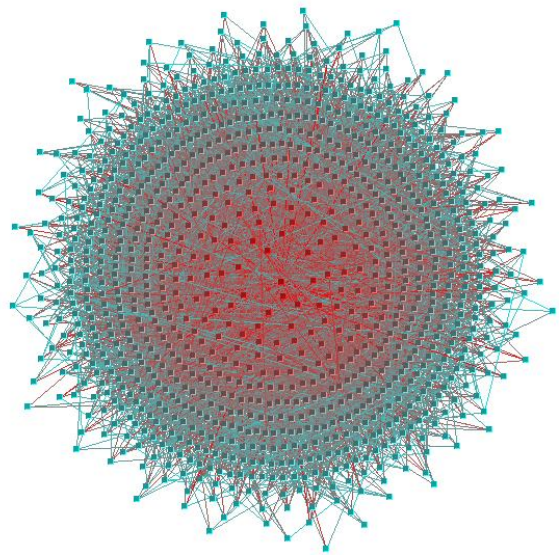


Fig.2. Visualization of the Random network model with $N=1024$ nodes and $p=0.01$

4 Small-World Model

The idea behind the **Small-World** network model comes from the social systems and the relationships therein. Most of the people in the social networks are friends with their immediate neighbors imposing that way very large short distance clustering. However, some people have friends who are a long way away (old relationship or acquaintance). This relationship "shortcuts" make the average distance between people relatively small. In a folk wisdom this is known as "six degrees of separation".

The **Small-World** network model was first introduced by *Watts* and *Strogats* [4]. The model was extended and mathematically analyzed by *Watts* and *Newman* [5],[6]. They model the network as an ordered lattice with clustering coefficient which is network size (N) independent. Clustering coefficient in a lattice depends only on the coordination number K (the number of neighbors the node is connected to). A ring lattice with N nodes and coordination number K has $\frac{NK}{2}$ edges. In such lattice *Watts* and *Strogats* randomly rewire each edge with a small probability p leading to a small number of $p \frac{NK}{2}$ rewired edges. Instead of edge rewiring, *Watts* and *Newman* in [5] add $p \frac{NK}{2}$ new edges without removing the old ones. This small change makes the model more realistic (there is no loops or isolated

components) and it is easier to analyze. Fig.3 shows a ring lattice with $N=20$ nodes, coordination number $K=4$ and 3 shortcuts added.

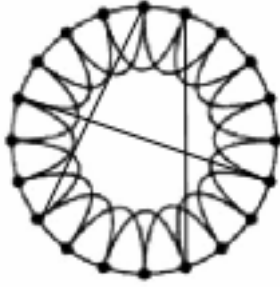


Fig.3. Ring lattice with 3 shortcuts

A ring lattice without shortcuts has clustering coefficient given as [7]:

$$C_{lattice} = \frac{3(K - 2)}{4(K - 1)} \quad (8)$$

and average distance between nodes given as:

$$l_{lattice} = \frac{N}{2K} \quad (9)$$

The **Small-World** network model interpolates between an ordered lattice ($p=0$) and **Random** network model ($p=1$). *Watts* and *Strogatz* has first noticed that there is a region of small p with large and almost unchanged clustering coefficient (property of an ordered lattice) and small average distance that scales as logarithm of N (property of the **Random** network model). *Watts* and *Newman* in [5], using recombination group principles, have derived the following equation for the average distance between nodes in one dimensional lattice:

$$l_{sw} = \frac{N}{K} f(pKN) \quad (10)$$

$$f(x) \sim \begin{cases} \text{constant} & \text{for } x \ll 1 \\ \frac{\log(x)}{x} & \text{for } x \gg 1 \end{cases}$$

Fig.4 shows the simulation results for the **Small-World** model with $N=1024$ nodes. The network contains undirected edges and does not contain loops or multiple edges between nodes. We vary the number of shortcuts in the interval $[0,100]$. In Fig.4-a we plot the average distance between nodes. It can be seen that the average distance in the case of 100 shortcuts is greatly reduced compared with an

ordered lattice. In Fig.4-b we plot the average clustering which is almost unchanged.

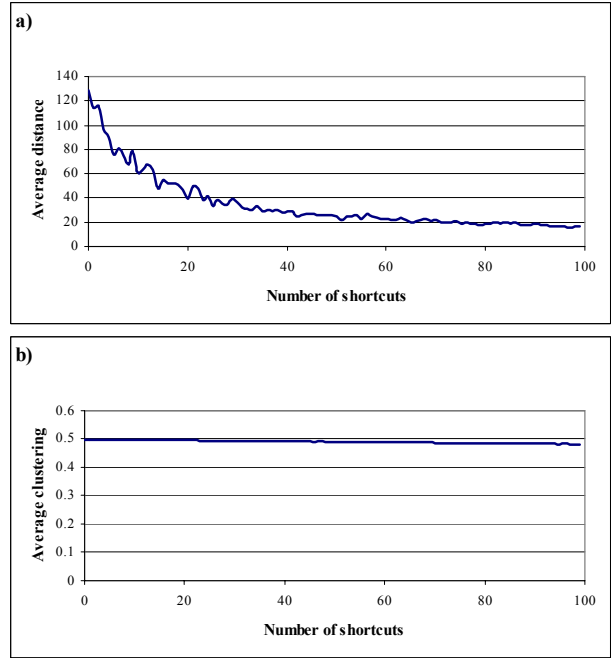


Fig.4. Small-World network model with $N=1024$ nodes a) Average distance b) Average clustering

Small-World Network: 1024 nodes

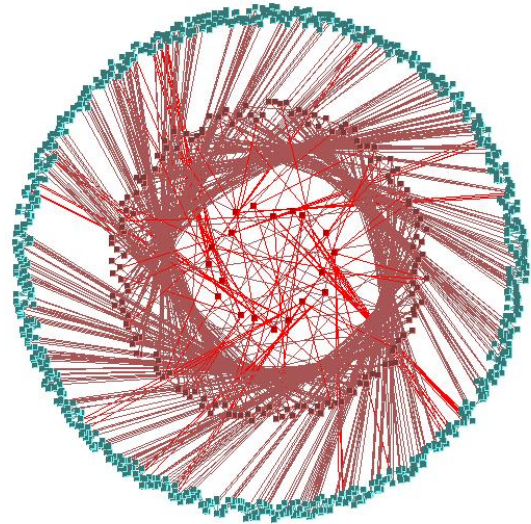


Fig.5. Visualization of the Small-World network model with $N=1024$ nodes, $K=4$ and 100 shortcuts

5 Scale-Free Model

The **Scale-Free** network model is an attempt to describe the results obtained by investigating data available from some real networks. Many systems

(networks) have the property that the degree distribution $P(k)$ follows a power law, i.e.

$P(k) \sim k^{-\gamma}$ and it is independent of the network size N (scale-free). For example, the analyzes of the movie actors collaboration graph [8] have shown that $P(k)$ follows a power law with $\gamma_{actor} = 2.3 \pm 1$. World Wide Web is also considered as a huge network where every HTML page represents a node and hyperlinks between pages represent edges. It has been shown [8] that WWW's degree distribution also follows a power law, having $\gamma_{www}^{in} = 2.1$ for incoming links and $\gamma_{www}^{out} = 2.45$ for outgoing links.

The first and most exploited **Scale-Free** network model is suggested by *Barabási* and *Albert* [8],[9]. They obtain the scale-free network properties using the following two aspects: **growth** and **preferential attachment**. The growth aspect states that the number of vertices increases during the time i.e. it is not fixed (the growth of the WWW). The preferential attachment aspect states that a new node is more probably attached to the nodes having higher degrees (*rich-becomes-richer* principle). For example, a newly introduced WWW page points more likely to some well known and established WWW sites.

The above two aspects are incorporated into the *Barabási -Albert* model (BA) as follows:

Growth: Starting with a small number of nodes (m_0), at every time step a new node is added with $m \leq m_0$ edges connecting it to the nodes already present in the system.

Preferential attachment: A new introduced node connects to the node i with probability proportional to the node's i degree k_i , i.e.:

$$P(i) = m \frac{k_i}{\sum_j k_j}. \quad (11)$$

Using the mean-field approach, *Barabási* and *Albert* have derived the following approximative degree distribution formula:

$$P(k) = \frac{2m^2 t}{m_0 + t} \frac{1}{k^3}. \quad (12)$$

In the above equation t is the number of executed time steps that correspond to network with $m_0 + t$ nodes and mt edges (plus the edges among the initial nodes). The equation (10) shows that the degree distribution follows a power law with $\gamma = 3$ (k^{-3}).

We take $m_0 = m = 2$ in the computer simulation. In order to avoid the possibility of having disconnected nodes we start with a network of $m_0 = 2$ connected nodes. At every time step we add a new node with $m = 2$ new edges. The initial state of two connected nodes doesn't matter in the limit of large N . It is worth mentioning that the BA model allows multiple edges between nodes, but doesn't allow loops. When we calculate the average clustering or average distance we count only one edge. The LCD model, introduced in [10], allows both multiple edges and loops.

Fig.6 plots the degree distribution of the BA network ($m_0 = m = 2$) at two growing levels: $N=1024$ nodes and $N=2048$ nodes. The red colored dotted lines plot a degree distribution according equation (10). It can be seen that in both cases the degree distribution in a large range follows a power law with $\gamma \sim 3$.

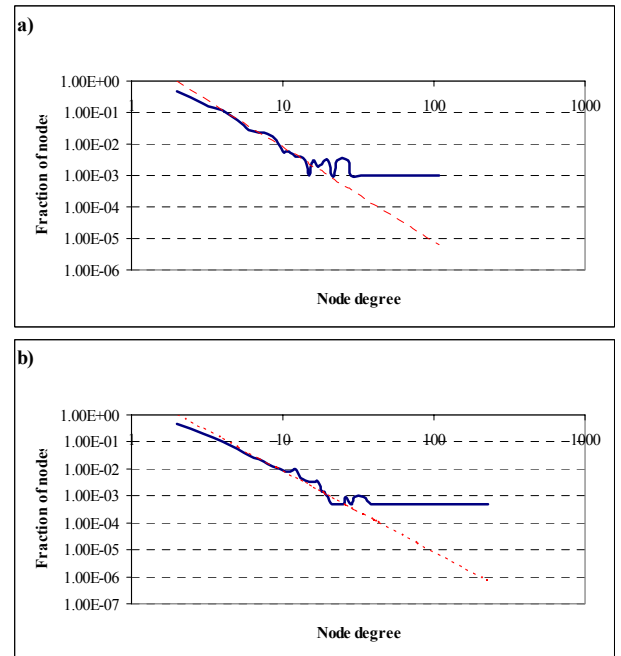


Fig.6. Degree distribution in the BA network model with $m_0 = m = 2$ a) $N=1024$ nodes b) $N=2048$ nodes

The calculated average distance in the BA network model with $N=1024$ nodes is 4.0. So, the BA network is highly connected graph. The calculated average clustering in the same network varies in the interval $[0.02,0.05]$ and it is greater than the clustering in a random network with the same number of nodes and edges.

Scale-Free Network: 1024 nodes

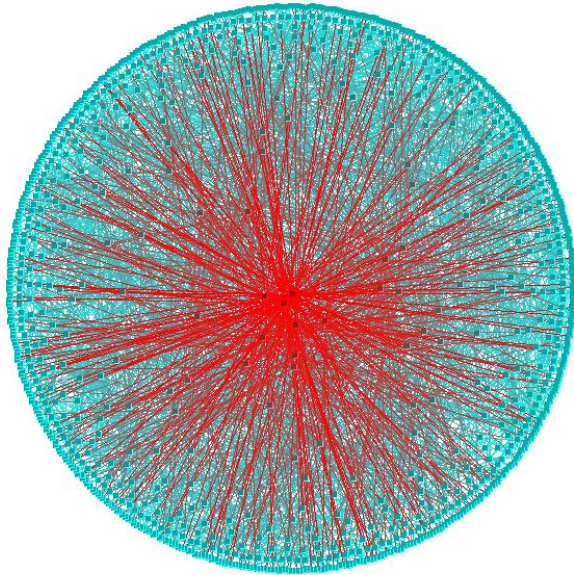


Fig.7. Visualization of the BA network model with $N=1024$ nodes and $m_0 = m = 2$

6. Conclusions

We presented three network models that make the basic network classification in the contemporary network topology research. However, variants of the models exist. Each of them is trying to incorporate additional network topology concepts. For example, an overview of the current scale-free network models is given in [11].

Network models and analysis therein are very important for predicting future network behaviour. Analyzing average distance and clustering may help in designing more effective and robust network services.

References

- [1] P.Erdős, A. Rényi, On random graphs I, *Publicationes Mathematicae Debrecen*, Vol. 5, 1959, pp. 290-297
- [2] P.Erdős, A. Rényi, On the evolution of random graphs, *Magyar Tud. Akad.Mat.Kutató Int.Közl.*, Vol. 5, 1960, pp. 17-61

- [3] B. Bollobás, *Random Graphs*, Cambridge University Press, 2001
- [4] D. Watts, S. Strogatz, Collective Dynamics of 'small-world' networks, *Nature*, Vol.4, 1998, pp. 440-442
- [5] M.Newman, D.Watts, Renormalization group analysis of the small-world network model, *Physics Letters*, A 263, 1999, pp.341-346
- [6] M.Newman, D.Watts, Scaling and percolation in the small-world network model, *Physics reviews*, E 60, 1999, pp. 7332-7342
- [7] R.Albert, A. Barabási, Statistical mechanics of complex networks, *Reviews of modern physics*, Vol. 74, 2002, pp. 47-97
- [8] R.Albert, A. Barabási, Emergence of Scaling in Random Networks, *Science*, Vol. 286, 1999, pp. 509-512
- [9] R.Albert, H. Jeong, A. Barabási, The diameter of the World-Wide Web, *Nature*, Vol. 401, 1999, pp. 130
- [10] B. Bollobás, O.Riordan, The diameter of a scale-free random graph, *Combinatorica*, 2003
- [11] B. Bollobás, *Mathematical results on scale-free random graphs*, 2003