Review of Scale Free Networks and its variances

Aoxi Li*
MIT Department of Physics
(Dated: May 17, 2011)

Networks describe a wide range of complex systems in nature and society. Among the most well-known examples are the cell, networks of chemicals, the Internet and networks of computers. Traditionally, these system have been modeled as random graphs. However, some detailed properties, including the degree distribution and clustering effect cannot be entirely explained merely using the traditional model. more recently, the underlying topology of these networks has been increasingly studied. This article reviews Barabási-Albert model and its variances. The insight of the scale-free property is demonstrated in this article.

1. INTRODUCTION

Network has become a valuable approach in explaining biological systems. As early as 1980s, researchers started viewing various biological systems as networks. The interactions of protein-protein and DNA-protein are good examples which demonstrate similar characteristics to networks. Researchers can best represent and analyze the essence of biology's layers computable networks, e.g. for protein-protein interactions, each protein can be modeled as a node of a network and the interactions in between can be modeled as edges.

Not only network method can be applied in biology, it can also be used to analyze various social phenomena, e.g. the Internet, a network of routers and computers, the network co-authorship, which make network a worthnoting mathematical model to study on.

Networks can have sophisticated structures built up by nodes and edges to model the complex systems. A network can be represented by nodes $i = 1, 2, \cdots, N$ and up to $N(N-1)/2$ edges.

2. RANDOM GRAPH THEORY

Several random graph theories were proposed to make the properties of network solvable. Two prominent models, Erdös- Rényi Model and Barabási-Albert Model , will be subsequently introduced in this paper.

2.1. Erdös- Rényi Model

Paul Erdös and Alfréd Rényi model is the simplest model which used a probabilistic method to generate random graphs. This model provides a tunable expected edge density of graphs. In the model, a graph consists of N nodes, with any pair connected independently with probability p. Thus, an expected number of $pN(N-1)/2$ edges are connected between these nodes.

random graphs have several important features. For example, we can obtain the most direct feature of Erdös-Rényi Model, the expected number of subgraphs of n nodes and l links is

$$N(n,l) = \binom{N}{n} p^l \times \frac{n!}{\text{symmetry factors}} \tag{1}$$

2.2. Degree Distribution, Diameter and Clustering Coefficient

When investigating the properties of a random graph, degree distribution, diameter and clustering coefficient are three characteristic criteria.

- **Degree distribution** In a random graph with independent connection probability p, the degree k_i of a node i follows a binomial distribution with parameters $N-1$ and p:

  $$P(k_i = k) = C_{N-1}^{k} p^k (1-p)^{N-1-k} \tag{2}$$

  The expected degree of each node is thus to be $\langle k \rangle = pN$.

- **Connectedness** In random graph theory, there are two major ways to describe the connectedness of a graph, namely diameter and average distance.

  Diameter of a graph is defined as the maximal distance between any pair of its nodes. This property has been studies widely. [**?**]cl)

  In Erdös- Rényi Model, the diameter of a graph with parameters p and N,

  $$d = \frac{\ln N}{\ln pN} = \frac{\ln N}{\ln \langle k \rangle} \tag{3}$$

  There is another way to characterize the spread of a random graph, average path length. In In Erdös-Rényi Model,

  $$l \sim \frac{\ln N}{\ln \langle k \rangle} \tag{4}$$

*Electronic address: aoxion@mit.edu; URL: http://aoxion.mit.edu/

- **Clustering Coefficient** Complex network also exhibit a large degree of clustering. Considering a node in a random graph and its neighbors, the probability that two of these neighbors are connected may vary in different models and this probability describes how clustering the network is.

  In Erdös- Rényi Model,

$$C = p = \frac{\langle k \rangle}{N} \tag{5}$$

  since all connections are independent.

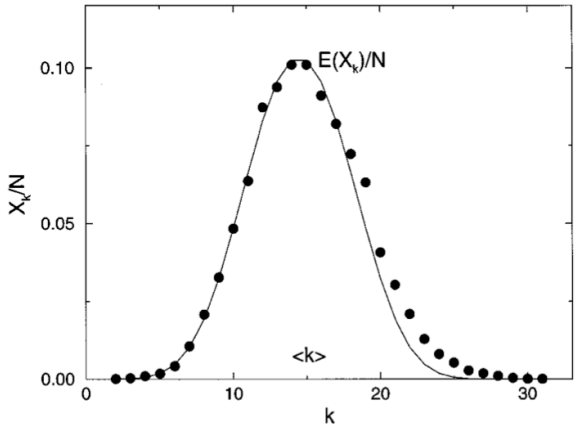

FIG. 1: This is a diagram demonstrating the probability distribution. It follows a binomial distribution and has an expenentially decaying tail.

3. SCALE-FREE NETWORKS

In Erdös- Rényi Model, the clustering coefficient is fixed to be $C = p$ which means there is no clustering effect at all. However, in most real networks the clustering coefficient is typically much larger than it is in a comparable independent random network.

Researches on degree distribution of real networks have demonstrated most real networks follow a power-law degree distribution,

$$P(k) \sim k^{-\gamma}. \tag{6}$$

[1] Such networks are called scale free.

These discoveries have initiated a revival of network modeling, resulting in the introduction of Barabási-Albert Model and similar class of models.

3.1. Barabási-Albert Model

The empirical result shows that many large networks are scale free, in other words, their degree distribution follows a power law for large k, which Erdös- Rényi Model cannot produces.

Barabási and Albert [2] addressed that the origin of the power-law degree distribution observed in networks is rooted in the *growing* nature and *preferential attachment* of networks. For example, a web page will more likely include hyperlinks to popular websites which already have high indegrees.

These two ingredients, growth and preferential attachment, inspired the introduction of the Barabási-Albert model, which led to a network with a power-law degree distribution for the first time. The algorithm of Barabási-Albert model is as following:

- *Growth* Starting with a small number (m_0) of nodes, at every time step and add one new node with $m(\leq m_0)$ edges that link the new node to m different nodes already present in the system.
- *Preferential attachment* When choosing the nodes to which the new node connects, we assume that the probability Π that a new node will be connected to node i depends on the degree k_i of node i, such as a linear relation,

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j} \tag{7}$$

3.2. Theoretical Approach

Barabási and Albert proposed an approach to solve the model by exploiting the continuum theory [3] Assuming that the degree k_i of a given node i is a continuous real variable, the rate at which k_i changes can be written as

$$\begin{aligned} \frac{\partial k_i}{\partial t} &= m\Pi(k_i) \\ &= m\frac{k_i}{\sum_{j=1}^{N-1} k_j} \\ &= \frac{k_i}{2t} \end{aligned} \tag{8}$$

The solution of Eq. 8 with the initial condition to be every node i at its introduction to the network has $k_i(t_i) = m$ is

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{\beta} \text{ where } \beta = \frac{1}{2} \tag{9}$$

Using Eq. 9, one can derive that

$$P[k_i(t) < k] = P\left(t_i > \frac{m^{1/\beta}t}{k^{1/\beta}}\right). \tag{10}$$

Assuming that we add the nodes at nodes at equal time intervals, the t_i values have a constant probability density

$$P(t_i) = \frac{1}{m_0 + t} \tag{11}$$

Network	Size	$\langle k\rangle$	γ_{out}	γ_{in}	l	l_{rand}	C	C_{rand}
WWW, site level	153127	35.21		1.94	3.1	3.35	0.1078	0.00023
Internet,domain	3015-4389	3.42-3.76	2.1-2,2	2.1-2.2	3.7-3.76	6.36-6.18	0.18-0.3	0.001
Movie actors	225226	61	2.3	2.3	3.65	2.99	0.79	0.00027
Math. co-authorship	70975	3.9	2.5	2.5	9.5	8.2	0.59	5.4×10^{-5}
SPIRES co-authorship	56627	173	1.2	1.2	4.0	2.12	0.726	0.003
Words, synonyms	22311	13.48	2.8	2.8	4.5	3.84	0.7	0.0006
Words, co-occurrence	460902	70.13	2.7	2.7	2.67	3.03	0.437	0.0001
Phone call	53×10^6	3.16	2.1	2.1				
Citation	783339	8.57		3				

TABLE I: The general characteristics of several real networks. For each network we have indicated the number of nodes, the average degree $\langle k\rangle$, the average path length l, the clustering coefficient C, indegree and outdegree $(\gamma_{in},\gamma_{out})$ exponents when fitting $P(k)$ against $k^{-\gamma}$[2].

therefore, we can obtain

$$P\left(t_i > \frac{m^{1/\beta}t}{k^{1/\beta}}\right) = 1 - \frac{m^{1/\beta}t}{k^{1/\beta}(t+m_0)} \tag{12}$$

Further, we have

$$\begin{aligned} P(k) &= \frac{\partial P[k_i(t)<k]}{\partial k} \\ &= \frac{2m^{1/\beta}t}{m_0+t}\frac{1}{k^{1/\beta+1}} \\ &\sim 2m^{1/\beta}k^{-\gamma} \end{aligned} \tag{13}$$

where γ is fixed to be $\gamma = 1/\beta + 1 = 3$.

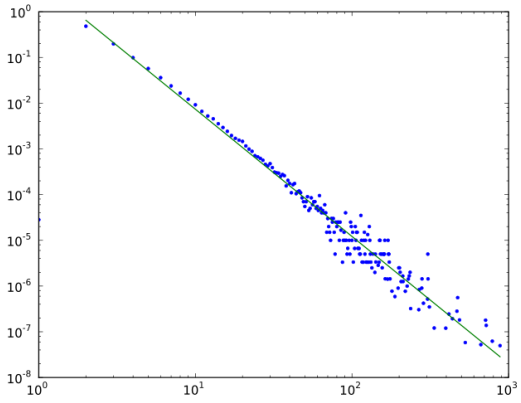

FIG. 2: This diagram shows the scale free property of Barabási-Albert model. $\ln P(k)$ and $\ln k$ demonstrate a good linear fit.

Barabási-Albert model also shows differences from Erdös-Rényi model in average path length and clustering coefficient. The average path length is $l \sim \ln N/\ln\ln N$ instead of $l \sim \ln N/\ln\langle k\rangle$, which means the network in scale-free model is more connected, while the clustering coefficient is $C \sim N^{-3/4}$ instead of $C \sim \langle k\rangle N^{-1}$, while means the network in scale-free model is also more clustered. For the simplicity, we will not discuss the detailed derivation here.

4. VARIATIONS OF BARABÁSI-ALBERT MODEL

The Barabási-Albert model discussed previously is only a minimal model that captures the mechanisms responsible for the power-law degree distribution. It is limited by the simplicity nature of the model itself: it predicts a power-law degree distribution with a fixed exponent $\gamma = 3$, while the exponents measured for real networks vary between 1 and 3 (See Table I). These discrepancies between the model and real networks bring up the question that if it is possible to change the scaling exponents. Here we'll discuss some variant model from the original Barabási-Albert model which can achieve different values of the exponent.

4.1. Change the preferential attachment function $\Pi(k)$

In the original Barabási-Albert model, we assumed a linear relation between the attachment probability $\Pi(k)$ and k.

There are two major approaches to modify the attachment probability $\Pi(k)$, nonlinear functions and initial attractiveness. Nonlinear functions expand the possible form of attachment functions to $\Pi(k) \sim k^\alpha$. Another general feature of $\Pi(k)$ in real networks is that $\Pi(k) \neq 0$, since in real networks, if there is a nonzero probability that a new node attaches to an isolated node. Thus the general form of $\Pi(k)$ has the form $\Pi(k) = A + k^\alpha$, where A is the initial attractiveness . In real networks, every node has a finite chance to be discovered and linked to , even if it has no edges to start with. Thus the parameter

A describes the likelihood that an isolated node will be discovered, such as a new article's being cited the first time.

We can argue that the introduction of the two effects can bring in different value of exponents. Exact solution for a class of growing network models has been given using master-equation approach [5]. For example, $\Pi(k_{in}) = A + k_{in}$. The calculation indicate that the degree distribution follows $P(k) \sim k^{-\gamma}$ where $\gamma = 2+A/m$. A more general model, Dorogovtsev-Mendes-Samukhin model, has been investigated and the model also gives a power-law degree distribution with $\gamma = 2+(n_r+n+A)/m$ where n_r edges are randomly distributed and initially every new node has degree of n.

4.2. Local Events

The Barabási-Albert model only considers network growth mechanism. However, for real networks, modification of existing nodes may also take place. A model that incorporates the rewiring was discussed in [2]. We perform one of following three operations:

1. With probability p add $m(m \leq m_0)$ new edges. One end of a new edge is selected randomly, the other with probability
$$\Pi(k_i) = \frac{k_i + 1}{\sum_j (k_j + 1)}. \quad (14)$$
2. With probabilility q we rewire m edges. We randomly select a node i and remove an edge l_{ij} connected to it and replace it with a new edge l'_{ij} that connect node i with node j' chosen with probability $\Pi(k'_j)$ given by Eq. 14.
3. With probability $1-p-q$, we add a new node and connect m edges into the graph using the original method.

Using continuum theory, the growth rate for degree of node i is given by

$$\frac{\partial k_i}{\partial t} = (p-q)m\frac{1}{N} + m\frac{k_i + 1}{\sum_j (k_j + 1)} \quad (15)$$

The solution is given as

$$k_i(t) = [A(p,q,m) + m + 1]\left(\frac{t}{t_i}\right)^{1/B(p,q,m)} - A(p,q,m) - 1 \quad (16)$$

where

$$A(p,q,m) = (p-q)\left(\frac{2m(1-q)}{1-p-q} + 1\right) \quad (17)$$

$$B(p,q,m) = \frac{2m(1-q)+1-p-q}{m}. \quad (18)$$

The corresponding degree distribution follows a generalized power-law form.

$$P(k) \propto [k + \kappa(p,q,m)]^{-\gamma(p,q,m)}, \quad (19)$$

where $\kappa(p,q,m) = A(p,q,m) + 1$ and $\gamma(p,q,m) = B(p,q,m) + 1$. When k gets large, the relation can be simply written as $P(k) \sim k^{-\gamma}$.

5. CONCLUSION

In this paper, we compared different models of networks, especially for the variances of Barabási-Albert model. We've demonstrated that Barabási-Albert model which include network growth and preferential attachment has provided us a better tool explaining the networks in real world. It naturally includes the scale-free property of real networks and has a degree distribution of $P(k) \sim k^{-\gamma}$.

Furthermore, since Barabási-Albert model can only have a fixed exponent, we investigated the variances of Barabási-Albert model, which take into consider different preferential attachment functions and removal of existing edges which have improved the flexibility of network models by making the exponent *gamma* tunable.

Recently, new tools are brought up as well, such as ideas from field theory (Burda et al., 2001) and quantum statistics (Bianconi, 2000). The theory of network will become more and more inspirational and influential not only in biology science, but also in social science and other fields.

6. ACKNOWLEDGEMENT

I would like to thank Prof. Mehran Kardar for introducing this interesting topic to us.

[1] Albert, R., and A.L. Barabási,Rev of modern physics, 74, 2002

[2] Albert, R., and A.L. Barabási, 2000, Phys. Rev. Lett. 85, 5234.

[3] Albert, R., H. Jeong, and A.L. Barabási, 2000, Nature (London) 406, 378; 2001, 409, 542(E).
[4] Erdös, P., and A. Rényi, 1959, Publ. Math. (Debrecen) 6, 290.
[5] Dorogovtsev, S. N., and J. F. F. Mendes, 2000a, Europhys. Lett. 50, 1.
[6] Chung, F., and L. Lu, 2001, Adv. Appl. Math. 26, 257.