Available online at www.sciencedirect.com
ScienceDirect
Communications in Nonlinear Science and Numerical Simulation 13 (2008) 1405–1410
Communications in Nonlinear Science and Numerical Simulation
www.elsevier.com/locate/cnsns

The average path length of scale free networks

Fei Chen *, Zengqiang Chen, Xiufeng Wang, Zhuzhi Yuan

Department of Automation, Nankai University, Tianjin, 300071, PR China

Received 23 March 2006; received in revised form 1 December 2006; accepted 5 December 2006
Available online 22 December 2006

Abstract

In this paper, the exact solution of average path length in Barabási–Albert model is given. The average path length is an important property of networks and attracts much attention in many areas. The Barabási–Albert model, also called scale free model, is a popular model used in modeling real systems. Hence it is valuable for us to examine the average path length of scale free model. There are two answers, regarding the exact solution for the average path length of scale free networks, already provided by Newman and Bollobas respectively. As Newman proposed, the average path length grows as $\log(n)$ with the network size n. However, Bollobas suggested that while it was true when $m = 1$, the answer changed to $\log(n)/\log(\log(n))$ when $m > 1$. In this paper, as we propose, the exact solution of average path length of BA model should approach $\log(n)/\log(\log(n))$ regardless the value of m. Finally, the simulation is presented to show the validity of our result.

PACS: 89.75.−k; 89.20.Hh; 05.65.+b

Keywords: Complex networks; BA model; Power-law; Average path length; Scale free

1. Introduction

When exploring a large system, people used to divide it into small parts, and then examine each part respectively. However, many proofs show that the relation between parts is at least as import as the part itself. Nevertheless, the structures of these systems are unknown which make the relation unobserved. The desire to understand these systems has boosted the study for complex networks.

There are many models which are used to simulate the structures of large systems. Some of them are worth noticing: the first model of complex networks was proposed by Erdös and Rényi which is called Erdös–Rényi (ER) model [1]. In this model, the probability of the appearance of each arc is fixed, denoted by p. In 1998, Watts and Strogatz proposed a simple model used to generate the small-world effect of real systems (Fig. 1) [2]. However, until now, the most popular model is the Barabási–Albert model [3,4], in short BA model, which reproduces power-law distributions regarding the degree of nodes.

* Corresponding author.
E-mail address: chernf@gmail.com (F. Chen).

1007-5704/$ - see front matter
doi:10.1016/j.cnsns.2006.12.003

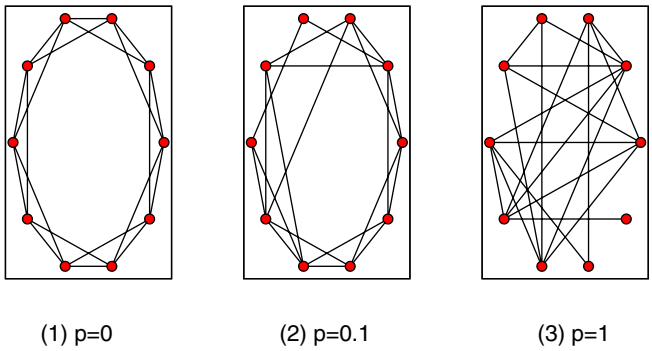

Fig. 1. The evolution process of small-world model. Three figures correspond to three different values of rewiring probability p, which is 0, 0.1 and 1 respectively.

The observation from real world networks shows that the degrees of most of the nodes are very low while some others are high which means that the degree distributions of these networks follow power-law distributions. The examples include the WWW, the Internet, metabolic network, etc. Moreover, the systems observed in real world are almost dynamic, which means the nodes are evolving. Barabási and Albert propose a tractable model to reproduce such features, called BA model in short [3,4]. The model is described as follows:

(1) Starting with m_0 nodes. At each time step, we add a node with $m(m \leqslant m_0)$ edges connected to the nodes already in the network.
(2) When selecting the edges, the probability of a node being selected is proportional to the degree of that node.

Fig. 2 shows a BA model with 100 nodes and the parameter $m = 1$.

The average path length of the network is defined as the average distance between all pair of its nodes [5]. It is relevant in many areas such as: circuit design [6], epidemic spreading [7], etc [8–12]. The average path length of the WWW has been studied by Réka Albert [13] indicating that the web forms a small world. The analytical result of average path length of networks was first studied by Newman et al. [14] using generating function, which suggested that the average path length should approach $\log(n)$. But the solution proposed by Bollobas and Riordan [15] suggested that it was true only for $m = 1$. When $m > 1$, the answer changed to $\log(n)/\log(\log(n))$. In this paper, we provide an exact solution for the average path length of BA Model which says that the average path length always approaches $\log(n)/\log(\log(n))$ regardless of m. This gives people another choice when predicting the average path length of a network and the method we used is easy to understand.

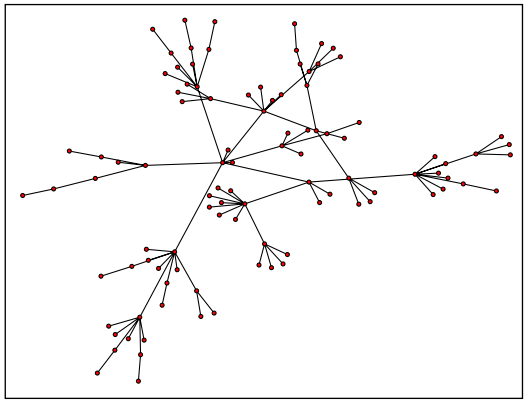

Fig. 2. An example of BA model with 100 nodes and m is set to 1.

Moreover, we give some simulations which are lacking in the previous papers [14,15] to make our result convictive.

The rest of the paper is organized as follows: in Section 2, the exact solution of average path length for BA model is presented. Section 3 is devoted to the simulation of our result. Finally, in Section 4, some conclusion are drawn to complete our discussion.

2. Theoretical analysis of average path length

Let d_i denotes the degree of a given node i. The change of d_i is described by the following equation, where t_i denotes the adding time of node i [16]:

$$d_i = m\left(\frac{t}{t_i}\right)^{\frac{1}{2}} \tag{1}$$

So we get

$$t_i = \frac{nm^2}{d_i^2} \tag{2}$$

Since $t_i \geqslant 1$ and the nodes are born with degree m, hence

$$m \leqslant d_i \leqslant m\sqrt{n} \tag{3}$$

The degree distribution of BA model follows a power law.

$$p(k) = 2m^2k^{-r} \tag{4}$$

where r is the exponent of the power law and $r = 3$ in Eq. (4).

The distribution $p(k)$ denotes the probability of the degree of a randomly selected node is k. The expectation of the degree is easily be obtained from Eq. (4).

$$\langle k\rangle = 2m \tag{5}$$

Suppose now we randomly select a node, then the number of its first neighbors, which means the number of nodes that directly connect to this node, follows a power-law distribution $p(k)$ as described by Eq. (4). However what is the distribution of its second neighbors. Since the probability that the outgoing edges of the first neighbors connect to the original vertex or its immediate neighbors goes as n^{-1} where n denotes the number of nodes in this network, it can be neglected [1] when the networks size is large. Moreover, we notice a fact that the number of second neighbors equals the sum of degrees of all its first neighbors, each minus one to allow for the original vertex. The degree distribution of each of these first neighbors is the same and follows:

$$q(k) = \frac{kp(k)}{\int kp(k)} \tag{6}$$

Solving Eq. (6), we obtain

$$q(k) = \frac{mk^{-2}}{1-\frac{1}{\sqrt{n}}} \tag{7}$$

Then the expectation of the degree of each of the first neighbors is

$$\langle k_q\rangle = \frac{m\log(n)}{2(1-\frac{1}{\sqrt{n}})} \tag{8}$$

Now we can use the Wald Theorem to make an estimate for the number of second neighbors.

Theorem 1. *Let X_1, X_2, X_3, ... be a sequence of random variables with the same distribution, and let T be a stopping time for them, which has finite expectation. Define*

$$S = X_1 + X_2 + \cdots + X_T$$

Then

$$E(S) = E(X_1)E(T)$$

We notice that the number of terms added to S is itself a random variable, but this is well defined when T is almost surely finite, and we're even assuming finite expectation. Using the Wald theorem, the expectation of the number of second neighbors is obtained:

$$E(N_2) = E(N_1)\langle k_q - 1\rangle \tag{9}$$

Where N_1, N_2 denote the number of first neighbors and second neighbors respectively. Since $E(N_1) = \langle k\rangle$,

$$E(N_2) = \langle k\rangle\langle k_q - 1\rangle \tag{10}$$

Eq. (10) can be generalized into nth neighbors as

$$E(N_n) = \langle k\rangle\langle k_q - 1\rangle^{n-1} \tag{11}$$

Since the number of all its neighbors up to the average path length l, equals the number of nodes in this network,

$$1 + \sum_{i=1}^{l} E(N_i) = n \tag{12}$$

where n is the size of network. Combining with Eq. (11), we obtain

$$1 + \sum_{i=1}^{l} \langle k\rangle\langle k_q - 1\rangle^{n-1} = n \tag{13}$$

Solving Eq. (13),

$$l = \log_{\frac{m\log(n)}{2\left(1-\frac{1}{\sqrt{n}}\right)}-1}\left(2-\frac{m\log(n)}{2\left(1-\frac{1}{\sqrt{n}}\right)}\right)^{1-\frac{n-1}{2m}} \tag{14}$$

As $n \to \infty$, Eq. (14) reduces to

$$l \approx \frac{\log(n)}{\log(\log(n))} \tag{15}$$

3. Simulation

In the following we give the simulation of the average path length of different BA models. The algorithm used to find the nearest distance between any pair of nodes is Dijkstra algorithm. The time complexity for obtaining a average path length using Dijkstra algorithm is $O(n^3)$. Due to this fact, the number of nodes used in our simulation is no more than 10,000.

Fig. 3 shows the curve of average path length versus number of nodes when $m = 5$ as described above. The number of nodes is from 100 to 10,000. The circles correspond to the real data, the squares denote our theoretical result and the diamonds represent the curve of $\log(n)$. It can be easily deduced from the slopes of these curves that our result is in good agreement with the simulation and is much better than $\log(n)$. Moreover, as the network size grows our result will be much better. Furthermore, when the network is sparse, the assumption we made before becomes more accurate, therefore we would expect our analytical result becomes more precise.

Fig. 4 shows the curve of average path length versus number of nodes when $m = 3$ as described above. The symbols have the same meaning as before. It is evident that our result is in very good agreement with the

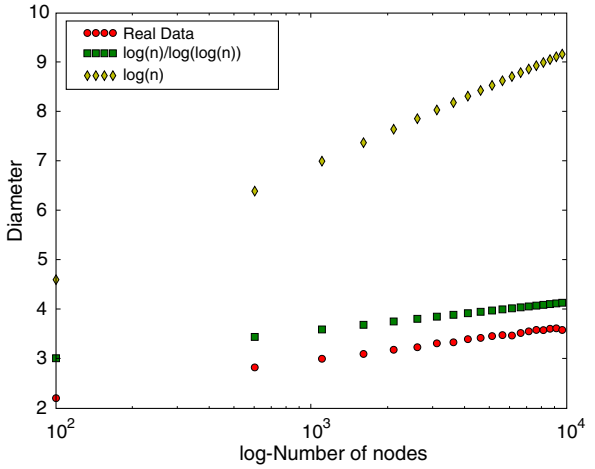

Fig. 3. The average path length versus the number of nodes when $m = 5$. The number of nodes is from 100 to 10,000. The circles correspond to the real data, the squares denote our theoretical result and the diamonds represent the analytical result provided by Newman. It is evident that our result is much better than the result provided by Newman et al. and is in good agreement with the simulation result.

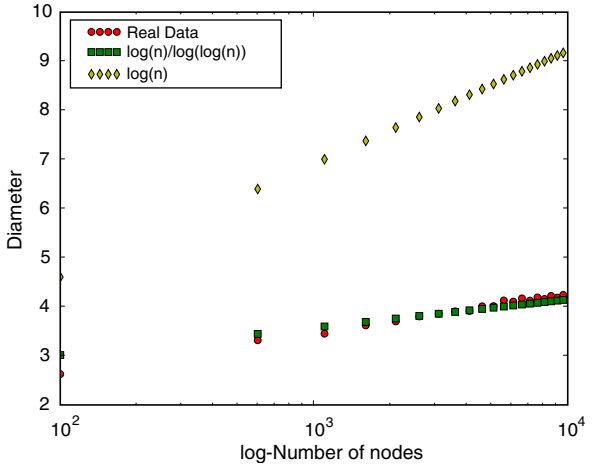

Fig. 4. The average path length versus the number of nodes when $m = 3$. The number of nodes is from 100 to 10,000. The circles correspond to the real data, the squares denote our theoretical result and the diamonds represent the analytical result provided by Newman et al. It is evident that our result is in very good agreement with the simulation result even when the number of nodes is not very large (no more than 10,000).

simulation and is better than the case $m = 5$ which is denser than $m = 3$. Moreover, since most of the real networks are sparse, our result may get a lot of applications in practice.

4. Conclusions

In this paper, we give the exact solution of average path length for BA model. The result shows that the average path length of BA model goes as $\frac{\log(n)}{\log(\log(n))}$ which is identical with the solution provided by Bollobas and Riordan [15] when $m > 1$. The present result provides another good choice when predicting the average path lengths of scale free networks. And we find that our result is much better than $\log(n)$. Furthermore, when the networks is sparse, our result is found to be in very good agreement with the simulation. Since most of the

real networks are sparse, our result can be applied in many areas for predicting the networks' average path lengths.

Acknowledgements

The author would like to thank the anonymous reviewer for valuable suggestions and comments.

This work was supported by the CNSF Grants No. 60574036, the Specialized Research Fund for the Doctoral Program of Higher Education of China Grant No. 20050055013, and the Program for New Century Excellent Talents of China (NCET).

References

[1] Erdös P, Rényi A. On random graphs I. Publ Math Debrecen 1959;6:290–7.
[2] Watts DJ, Strogatz SH. Collective dynamics of small world. Nature 1998;393:440–2.
[3] Barabasi AL, Albert R. Emergence of scaling in random network. Science 1999;286:509–12.
[4] Jeong H. Complex scale-free networks. Physica A 2003;321:226.
[5] Newman MEJ. The structure and function of complex network. IAM Rev 2003;45(2):167–256.
[6] Ferrer R, Janssen C, Sole RV. Topology of technology graphs: small world patterns in electronic circuits. Phys Rev E 2001;64:32767.
[7] Meyers LA, Pourbohloul B, Newman MEJ, Skowronski DM, Brunham RC. Network theory and SARS: predicting outbreak diversity. J Theor Biol 2005;232(1):71–81.
[8] Scott J. Social network analysis: a handbook. London: Sage Publications; 2000.
[9] Dorogovtsev SN, Goltsev AV, Mendes JFF. Pseudofractal scale-free web. Phys Rev E 2002;65:066122.
[10] Ravasz E, Somera AL, Mongru DA, Oltvai Z, Barabási AL. Hierarchical organization of modularity in metabolic networks. Science 2002;297:1551–5.
[11] Broder A, Kumar R, Maghoul F, Raghavan P, Rajagopalan S, Stata R, et al. Graph structure in the web: experiments and models. J Comput Networks 2000;33:309–20.
[12] Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási AL. The large-scale organization. of metabolic networks. Nature 2000;407:651–4.
[13] Albert R, Jeong H, Barabási A-L. Diameter of the World Wide Web. Nature 1999;401:130–1.
[14] Newman MEJ, Strogatz SH, Watts DJ. Random graphs with arbitrary degree distributions and their applications. Phys Rev E 2001;64:026118.
[15] Bollobas B, Riordan O. The diameter of a scale-free random graph. Combinatorica 2004;24(1).
[16] Albert R, Barabási A L. The statistical mechanics of complex network. Rev Mod Phys 2002;74:47–97.