

Markov chain-based numerical method for degree distributions of growing networks

Dinghua Shi and Qinghua Chen*

Department of Mathematics, Shanghai University, Shanghai 200436, China

Liming Liu†

Department of Industrial Engineering and Engineering Management, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong

(Received 18 June 2004; revised manuscript received 13 December 2004; published 25 March 2005)

In this paper, we establish a relation between growing networks and Markov chains, and propose a computational approach for network degree distributions. Using the Barabási-Albert model as an example, we first show that the degree evolution of a node in a growing network follows a nonhomogeneous Markov chain. Exploring the special structure of these Markov chains, we develop an efficient algorithm to compute the degree distribution numerically with a computation complexity of $O(t^2)$, where t is the number of time steps. We use three examples to demonstrate the computation procedure and compare the results with those from existing methods.

DOI: 10.1103/PhysRevE.71.036140

PACS number(s): 89.75.Hc, 84.35.+i, 64.60.Fr, 87.23.Ge

I. INTRODUCTION

Complex networks describe a wide range of practical systems of high technological, biological, and social importance [1,2]. For example, the Internet, the World Wide Web (WWW), biological cells, and communities of scientists can all be described as complex networks.

Erdős and Rényi [3] started the early studies of complex networks as random graphs in the 1960s. Many years later, Watts and Strogatz's construction of a small-world network model in 1998 [4] represents an interesting development for the study of complex networks in that it was motivated by observations of real system behaviors (e.g., Milgram's six degrees of separation [5]). A common feature of the random graph and small-world networks is that the degree distributions (the probability of finding a node with k connections) are Poissonian. However, empirical evidence from some growing networks, such as the Internet and WWW, show a fundamentally different picture, i.e., the tail of the degree distribution follows a power law. This led to the introduction in 1999 of scale-free networks by Albert, Jeong, and Barabási, in their pioneering work [6–8], and the start of a new phase in the study of complex networks. Recent studies [9–20] are characterized by empirical observations of scale-free behaviors in various practical systems and investigations of the formation mechanisms of scale-free networks. A number of important properties in scale-free networks have been identified, such as the small-world character, the emergence of hubs, and robustness and fragility. These properties show that scale-free networks can play an important role in the understanding of many complex and important systems.

Two general features can be observed in many real-world networks: successive additions of new nodes and certain

preferences in linking to existing nodes. Barabási, Albert, and Jeong proposed two mechanisms to characterize the evolution of a scale-free network [7,8]. One is the growth mechanism: starting from m_0 nodes, the network grows at a constant speed, i.e., adding one node at each time step and connecting to $m(m \leq m_0)$ existing nodes; and preferential attachment: the chance that an existing node receives a connection from a new node is proportional to the number of connections it already has. The authors show that, under these two mechanisms, a network evolves into a stationary scale-free state. Its degree distribution follows a power law with the degree exponent $\gamma = 2.9 \pm 0.1$ from simulation analysis and $\gamma = 3$ from the analytical result [8]. These results are significant for complex networks and the two mechanisms form the first model, referred to as the Barabási-Albert (BA) model, by which large networks can self-organize into a stationary scale-free state. Empirical evidence shows that in many networks, the number of edges grows faster than the number of nodes [10]. This led to the investigation of m -varying BA models, such as that of Dorogovtsev and Mendes [20].

Our research is mainly motivated by the following observation. While analytical solutions of the degree distribution for some simple models of growing networks, such as the BA model, can usually be obtained, one has to resort to simulation for solutions when the mechanisms in the model become more complex. This may inhibit the further development of the theory on complex networks. In this paper, we propose an alternative approach. We first find that the degree evolution of a growing network can be characterized by a sequence of nonhomogeneous Markov chains. By carefully analyzing the structure of these Markov chains, we can then develop an efficient numerical method to compute the network degree distribution. To show the feasibility and efficiency of our numerical method, we compute the degree distributions of growing networks under the basic BA model and two of its variants.

We organize the paper as follows. In Sec. II, we first review some of the existing methods for network degree dis-

*Also at College of Mathematics and Computer Science, Fujian Normal University, Fuzhou 350007, China.

†Electronic address: liulim@ust.hk

tributions. We then show, using the BA model, how the degree evolution of a node in a growing network can be represented as a Markov chain. Exploring the special structure of the transition probability matrices of these Markov chains, we derive a formula by which the network degree distribution can be efficiently computed. In Sec. III, we present the numerical results for the BA model, comparing them with the analytical results, and discuss issues related to the implementation of the computation procedure. In Sec. IV, we construct the Markov chains for two m -varying BA models and provide the numerical results of the network degree distributions. We conclude the paper and point out some future research opportunities in Sec. V.

II. DEGREE EVOLUTION AND MARKOV CHAINS

With the preferential attachment mechanism of the BA model, the probability that node i added at time step i receives a connection from an upcoming new node is proportional to its own degree k_i [7], i.e.,

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}. \quad (1)$$

Assuming continuity of $k_i(t)$ and treating $\Pi(k_i)$ as its rate of growth, the degree of node i at time step t then satisfies the following dynamic equation [7,8]:

$$\frac{\partial k_i}{\partial t} = m\Pi(k_i) = m \frac{k_i}{\sum_j k_j} = \frac{k_i}{2t}. \quad (2)$$

Under the initial condition that $k_i(t_i) = m$, the solution of Eq. (2) leads to

$$k_i(t) = m \left(\frac{t}{t_i} \right)^\beta, \quad \beta = \frac{1}{2}, \quad (3)$$

where t_i is the time when node i joins the network, and to the degree distribution

$$P(k) \sim 2m^2 k^{-\gamma}, \quad \gamma = 3. \quad (4)$$

Here, β is called the dynamic exponent while γ is the degree exponent.

The above simple analytical method is often referred to as the continuum (mean field) theory. Similar power-law results for the degree distribution of the BA model are also obtained using different analytical methods by other authors. For example, with the master-equation approach [14], Dorogovtsev, Mendes, and Samukhin treat the degree $k_i(t)$ of a node i at a fixed time t as a random variable. Thus its probability $P(k, t_i, t)$ satisfies the following relation:

$$P(k, t_i, t+1) = \frac{k-1}{2t} P(k-1, t_i, t) + \left(1 - \frac{k}{2t} \right) P(k, t_i, t). \quad (5)$$

Let

$$P(k, t) = \frac{\sum_{t_i} P(k, t_i, t)}{t}. \quad (6)$$

Assuming that the limit $P(k) = \lim_{t \rightarrow \infty} P(k, t)$ exists and $\lim_{t \rightarrow \infty} t[P(k, t+1) - P(k, t)] = 0$ (an additional condition), the degree distribution $P(k)$ satisfies the equation

$$2P(k) - 2\delta_{k,m} = (k-1)P(k-1) - kP(k), \quad (7)$$

with the solution

$$P(k) = \frac{2m(m+1)}{k(k+1)(k+2)}. \quad (8)$$

Krapivsky, Redner, and Leyvraz's rate-equation approach [15] focuses on the number $N_k(t)$ of nodes with k edges at time t . For the BA model, $N_k(t)$ is shown to satisfy

$$\frac{dN_k}{dt} = m \frac{(k-1)N_{k-1}(t) - kN_k(t)}{\sum_k kN_k(t)} + \delta_{k,m}. \quad (9)$$

Asymptotically, $N_k(t) = tP(k)$ and $\sum_k kN_k(t) = 2mt$, which leads to Eq. (7).

While the above methods handle well simple models, such as the BA model, they may not render analytical solutions for more complicated models. In this case, one can usually use simulation. While simulation is widely applicable, it can be quite time consuming and may not be flexible enough for in-depth analysis of network behaviors. Here, we propose a different approach to capture the network dynamics. We use the BA model to present the idea and the procedure of our approach.

Consider the degree $K_i(t)$ of node i at time t . Based on the preferential attachment mechanism of the BA model, the sequence $\{K_i(t), t=i, i+1, \dots\}$ is a stochastic process with the state space $\Omega = \{m, m+1, \dots\}$ (noting that the original nodes are not considered). Here and below, we use the upper case K to emphasize the fact that the degree sequence is a stochastic process. The attachment mechanism also indicates that the future evolution of the process is independent of the past history, given its current state; but it is time dependent. This shows that the process $\{K_i(t)\}$ is in fact a nonhomogeneous Markov chain [21], with time-dependent transition probabilities

$$p_{kj}(t+1) = P\{K_i(t+1) = j | K_i(t) = k\} = \begin{cases} 1 - k/2t, & j = k, \\ k/2t, & j = k+1 \\ 0 & \text{otherwise,} \end{cases} \quad (10)$$

for $k = m, \dots, m+t-i$, and for $k > m+t-i$, we set

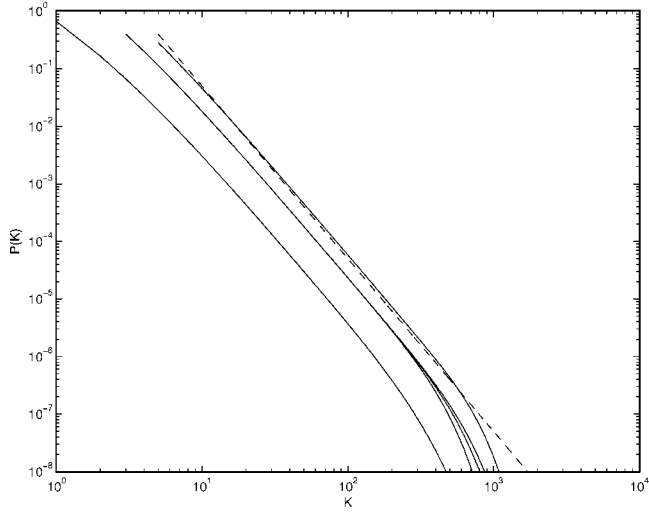


FIG. 1. The complete degree distributions of the BA model with different m and t . From left to right by the numerical method: (1) $m=1$ and $S=1$ with $t=150\,000$; (2) $m=3$ and $S=2$ with $t=100\,000$, $150\,000$, and $200\,000$, respectively; and (3) $m=5$ and $S=3$ and $t=150\,000$. The analytical results from formula (4) is the dashed line, with $m=5$.

$$\vec{F}^{(S,T)}(t+1) = (\{\cdots[\vec{e}_1 \mathbf{P}_S(S+1) + \vec{e}_1] \mathbf{P}_S(S+2) + \cdots\} + \vec{e}_1) \times \mathbf{P}_S(T+1) \cdots \mathbf{P}_S(t+1). \quad (20)$$

With Eq. (20), the computation of $\vec{F}^{(S,t)}(t+1)$ becomes very easy. We start from the innermost bracket. After one multiplication and one summation, we obtain a row vector whose first two elements are nonzero. The second round of multiplication and summation leads to a row vector with the first three elements being nonzero, and so on and so forth. The final result is a row vector with the first $t-S+1$ elements being nonzero. An efficient algorithm with a complexity of $O(t^2)$ can be developed to implement formula (20) numerically.

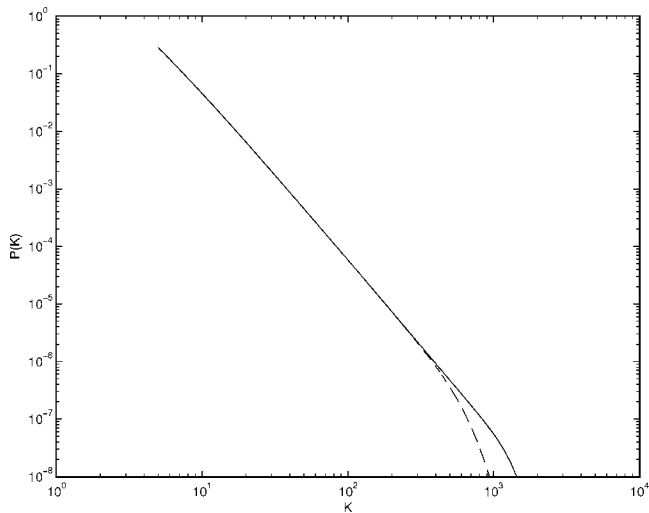


FIG. 2. The degree distributions of the BA model with $m=5$ for $S=3$ and 13 .

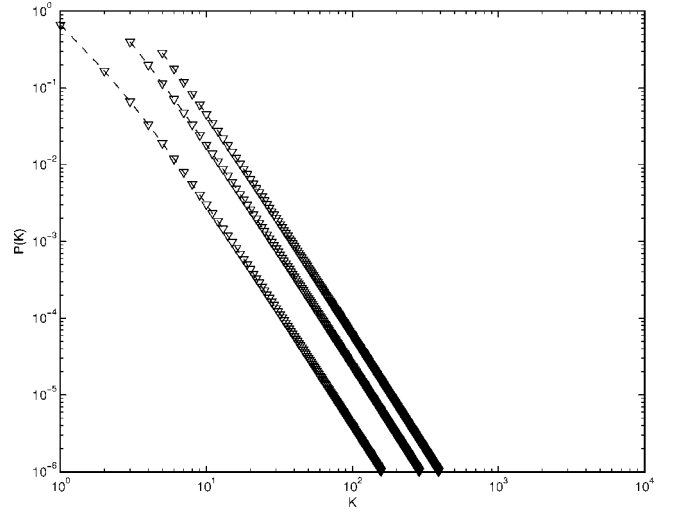


FIG. 3. The degree distributions of the BA model for $m=1$, 3 , and 5 from left to right.

III. COMPUTATION RESULTS AND DISCUSSION

In this section, we discuss the numerical results for the BA model.

Figure 1 shows the complete degree distributions of the BA model on a log-log scale, with different computation times t . We observe that each solid curve can be divided into three different sections: a straight middle section, clearly exhibiting the power-law property; a head section and an end section that bend inward, i.e., very small and very large degrees with probabilities smaller than the power law dictates. The numerical result matches the corresponding analytical result (the dashed line) almost exactly except at the end of the curve.

Why are the probabilities computed here smaller than the power law would give at the end of the curves? By the BA model, the degrees of the initial nodes are not considered in the computation of the network degree distribution, but they are likely to be among the group of nodes with the largest degrees, i.e., with degrees close to the end of the curve. With the numerical method (or simulation for this matter), we have to stop the computation at a finite t , and thus the proportion of the ignored initial nodes among all the nodes with largest degrees can be large enough to lower the degree probabilities of the end section. The sooner the computation stops, the earlier the curve bends. This can be seen from the three curves of case (2) with different stopping times t . With

TABLE I. Degree exponent and coefficients of the BA model.

Parameter m	Time t	Exponent γ	Coefficient c
1	150000	2.960830	3.147515
3	100000	2.989636	21.79266
3	150000	2.990032	21.89667
3	200000	2.980275	21.01711
5	150000	2.978894	52.58430

the analytic method, we do not observe this effect (e.g., from the dashed line) because it gives the degree distribution as t approaches infinity.

Now consider the bending at the head of the curves. Formula (4) is obtained by the continuum theory. Strictly speaking, it is a density (integrating from m to ∞ yields 1). Using it to approximate the discrete probability for large k is sufficiently accurate. For small degrees, it tends to overestimate the probability. This can be seen from formula (8) which is obtained with a discrete approach. When k is small, the probabilities obtained from Eq. (8) are smaller than the corresponding power-law probabilities. So formula (8) can be used to estimate small degree probabilities. The Markov chain method is a discrete approach and can also take care of the small degrees.

Recall that the initial nodes are ignored in the BA model in computing degree distributions. When m is greater than 1, we also need to take out additional nodes for the transition probability matrices to be properly defined. For example, the first node is ignored when $m=3$ and the first and second nodes are ignored when $m=5$. In the algorithm, we eliminate these nodes by starting the computation from time step S . Generally speaking, S mainly affects the end section of the distribution in the numerical method. In Fig. 2, the solid curve corresponds to the case with $S=3$ while the dashed curve corresponds to the case with $S=13$, and both have the same $m=5$ and the same stopping time $t=150\,000$. We observe that the two curves overlap all the way down to a point around $P(k) \approx 10^{-6}$, where both curves start bending inward, but the curve for $S=13$ obviously bends more. In other words, the exact value of S mainly affects how much the end section bends.

Summarizing the above analysis, we conclude that the head and middle sections of the curves computed by the Markov chain method capture the degree distribution of the BA model accurately. For the end section, the numerical method underestimates the degree probabilities because we can only compute up to a finite time. Fortunately, what we really want to know is whether the power law holds for a particular network (model) and what is the exact γ value. These can be determined by the middle section of the curves. If the curves show a power-law behavior, the network is scale-free. Otherwise, it is not. Either way, our method can help to determine the pattern as long as the degree evolution can be modeled by this type of Markov chain. In general, if the computation time is between 10^5 and 10^6 , the choice of S can be between 20 and 30, and the section of the curve with probabilities between 10^{-3} and 10^{-6} should be sufficient to reflect the asymptotic behavior of the degree distribution. Therefore, in Figs. 3–5 below, we will provide the degree distribution in the probability range from 10^0 to 10^{-6} .

We now compare the numerical results of the BA model with the analytic results from formula (8).

The dashed curves for the analytical results and the triangle curves for the corresponding numerical results in Fig. 3 overlap almost completely. It is clear that the numerical results are extremely close to the analytical results. The triangle curve in the middle is in fact the overlapping of three curves corresponding to three different t values. This shows that the degree distribution is stationary. Furthermore, except

for the small degrees, the three triangle curves are basically parallel, indicating that the degree exponent of the BA model is independent of m .

In Table I, the exponent γ and the coefficient c are fitted for different m and t using the least squares method. We note that for the degree exponent in Table I, we do not include the data for small degrees in the calculation. Similarly, they are not included for the exponents in Tables III and V below. We observe that the degree exponent is independent of m and the value matches those of simulation and the analytical solution with the mean field method. The coefficient of degree distribution c is between $2m^2$ and $2m(m+1)$, again matching the theoretical value from the master-equation approach. Furthermore, results for $m=3$ show that the coefficient c is independent of t , i.e., the network is stationary.

IV. m -VARYING BA MODELS

Section III shows that the Markov chain method provides accurate results with a moderate computation time. In this section, we show that this method can be efficiently applied to more complex models. We will construct the Markov chains and compute the degree distributions of two cases of the BA model with m -varying functions, which are of interests as empirical evidences have shown that the number of edges grows faster than the number of nodes in many networks [10].

A. Power function

Let mt^θ be the number of new links added at time step t , $0 \leq \theta < 1$, i.e., node t will link itself to mt^θ different existing nodes. Everything else remains the same as in the BA model.

We note that after t time steps, the network has $N=t+m_0$ nodes and approximately $\int_0^t mx^\theta dx$ links. The total number of degrees in the system at time t is

$$\sum_j k_j \approx 2 \int_0^t mx^\theta dx = \frac{2m}{\theta+1} t^{\theta+1}. \quad (21)$$

Assuming continuity, $k_i(t)$ then satisfies the following dynamic equation:

$$\frac{\partial k_i}{\partial t} = mt^\theta \Pi(k_i) = mt^\theta \frac{k_i}{\sum_j k_j} = \frac{(\theta+1)k_i}{2t}. \quad (22)$$

Under the initial condition $k_i(t_i) = mt_i^\theta$, where t_i is the time when node i joins the network, we solve this equation and obtain

$$k_i(t) = mt_i^\theta \left(\frac{t}{t_i} \right)^{(1+\theta)/2} = mt_i^\theta \left(\frac{t}{t_i} \right)^\beta, \quad \beta = \frac{1-\theta}{2}. \quad (23)$$

Hence the degree distribution at time t is

$$P(k, t) \sim \frac{2}{1-\theta} m^{2/(1-\theta)} t^\gamma k^{-\gamma}, \quad (24)$$

with

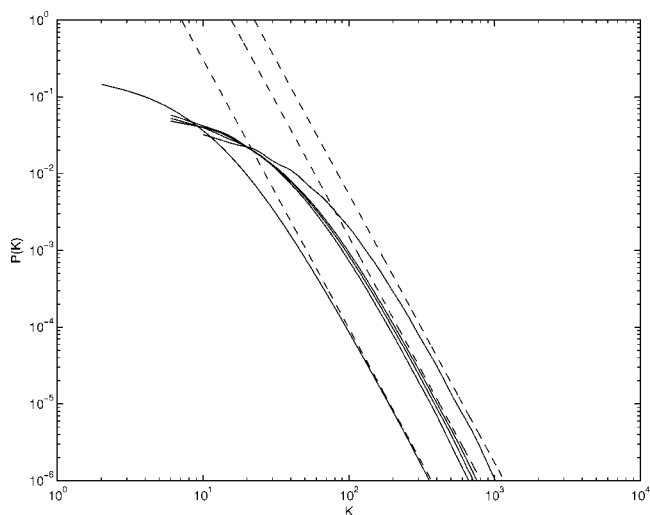


FIG. 4. The degree distributions of the power function case. Five solid curves for numerical results from left to right for (1) $m=1$ with $t=150\,000$; (2) $m=3$ with $t=100\,000$, $150\,000$, and $200\,000$; and (3) $m=5$ with $t=150\,000$. Three dashed lines for the corresponding analytic results from left to right for: $m=1$, $t=150\,000$; $m=3$, $t=200\,000$; and $m=5$, $t=150\,000$.

sults match very well in term of the degree exponent. Since (24) is obtained using the mean field method, it does not characterize the small degree accurately as discussed in Sec. III. This can be seen in Fig. 4 in which the solid curves bend obviously on the top while the dashed lines are completely straight. We also observe that when $m=5$, there is a gap between the lower part of the solid curve and the dashed line. This can be attributed to the difference between the continuum approach for (24) and the discrete approach for the numerical computation. The continuum approach may overestimate the number of links while the discrete approach tends to underestimate the number of links, in particular when m is large. For example, when $m=5$, the number of links added between times $t=32$ and 242 (see Table II) is $5t^{0.2}$ for the mean field method, but for our numerical method, it remains a constant $5\lceil t^{0.2} \rceil = 10$.

B. Logarithmic function

Let $m \ln t$ be the number of new links in time step t for $t \geq 2$.

TABLE III. Numerical results of the power function case.

Parameter m	Time t	Exponent γ	Coefficient c
1	150000	3.502938	891.641
3	100000	3.499978	8213.46
3	150000	3.502746	10920.8
3	200000	3.496971	12300.2
5	150000	3.503176	37303.5

We note that after t time steps, the network has $N=t+m_0$ nodes and approximately $\int_0^t m \ln x \, dx$ links. Then, the number of total degrees of the system at time t is

$$\sum_j k_j \approx 2 \int_0^t m \ln x \, dx = 2mt(\ln t - 1). \quad (31)$$

The average system degree is $\bar{k} = 2m(\ln t - 1)$, i.e., it follows a logarithmic law. Due to its complexity, there has been no analytical results for the degree distribution for this m -varying BA model.

We now construct the Markov chains for the degree sequence $\{K_i(t), t=i, i+1, \dots\}$. The state space is $\Omega_i = \{m_i, m_i + 1, \dots\}$, where $m_i = m \lceil \ln i \rceil$ (or $m_i = \lceil m \ln i \rceil$). At time t , the probability that an existing node i will connect with the new node is given by

$$m \ln t \frac{k_i}{\sum_j k_j} \approx \frac{k_i \ln t}{2t(\ln t - 1)}. \quad (32)$$

Hence, the one-step transition probabilities are

$$p_{kj}(t+1) = P\{K_i(t+1) = j | K_i(t) = k\} = \begin{cases} 1 - \frac{k \ln t}{2t(\ln t - 1)}, & j = k, \\ \frac{k \ln t}{2t(\ln t - 1)}, & j = k + 1, \\ 0, & \text{otherwise} \end{cases} \quad (33)$$

for $k = m_i, \dots, m_i + t - i$. Again, since $P\{K_i(t) = k\} = 0$, we set, for $k > m_i + t - i$,

$$p_{kj}(t+1) = \begin{cases} 1 & j = k, \\ 0 & j \neq k. \end{cases} \quad (34)$$

TABLE IV. Intervals of constant m_i for the logarithmic function case.

Time i	21	55	149	404	1097	2981	8104	22027	59875	162755
$\lceil \ln i \rceil$	3	4	5	6	7	8	9	10	11	12

The use of Markov chains to model the degree evolution opens the door for the application of methodologies and results from a mature field to the exciting field of complex networks. For instance, can we model the joint degree evolution of two nodes by two-dimensional Markov chain and compute the joint degree distribution? Can we construct a different type of Markov chain to model the distance between two nodes so that the average path length can be computed? The fact that the evolution of a complex network can be modeled by Markov chains may indicate an important direction for us to investigate the underlying mechanisms and the essential properties of growing networks, utilizing the extensive understanding of the structural properties of

Markov chains through many years' theoretical development and applications to the study of many natural phenomena.

ACKNOWLEDGMENTS

We thank Wang Binbin and Zhou Huijie from Shanghai University for helping us with the numerical computation. This research is supported in part by the National Natural Science Foundation of China through Grant No. 70171059, by the 863 Project through Grant No. 2002AA234021, and by Hong Kong Research Grant Council through Grants No. HKUST6089/00E and No. HKUST6198/01E.

-
- [1] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
 - [2] S. H. Strogatz, *Nature (London)* **410**, 268 (2001).
 - [3] P. Erdős and A. Rényi, *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17 (1960).
 - [4] D. J. Watts and S. H. Strogatz, *Nature (London)* **393**, 440 (1998).
 - [5] S. Milgram, *Psychol. Today* **1**, 60 (1967).
 - [6] R. Albert, H. Jeong, and A.-L. Barabási, *Nature (London)* **401**, 130 (1999).
 - [7] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
 - [8] A.-L. Barabási, R. Albert, and H. Jeong, *Physica A* **272**, 173 (1999).
 - [9] R. Albert, H. Jeong, and A.-L. Barabási, *Nature (London)* **406**, 378 (2000).
 - [10] M. Faloutsos, P. Faloutsos, and C. Faloutsos, *Comput. Commun. Rev.* **29**, 251 (1999).
 - [11] R. Albert and A.-L. Barabási, *Phys. Rev. Lett.* **85**, 5234 (2000).
 - [12] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, *Nature (London)* **407**, 651 (2000).
 - [13] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai, *Nature (London)* **411**, 41 (2001).
 - [14] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, *Phys. Rev. Lett.* **85**, 4633 (2000).
 - [15] P. L. Krapivsky, S. Redner, and F. Leyvraz, *Phys. Rev. Lett.* **85**, 4629 (2000).
 - [16] S. N. Dorogovtsev and J. F. F. Mendes, *Phys. Rev. E* **62**, 1842 (2000).
 - [17] G. Bianconi and A.-L. Barabási, *Phys. Rev. Lett.* **86**, 5632 (2001).
 - [18] E. Ravasz and A.-L. Barabási, e-print cond-mat/0206130.
 - [19] Q. H. Chen and D. H. Shi, *Physica A* **335**, 240 (2004).
 - [20] S. N. Dorogovtsev and J. F. F. Mendes, *Phys. Rev. E* **63**, 056125 (2001).
 - [21] S. M. Ross, *Stochastic Processes* (John Wiley and Sons, New York, 1983).
 - [22] D. H. Shi, J. L. Guo, and L. Liu, *Matrix-Analytic Methods in Stochastic Models*, edited by S. R. Chakravarthy and A. S. Alfa (Marcel Dekker, New York, 1996), p. 207.