The Eighth International Symposium on Operations Research and Its Applications (ISORA'09)
Zhangjiajie, China, September 20–22, 2009

Degree Distribution Analysis of a Random Graph Process Based on Markov Chains*

Li Tan[1] Zhenting Hou[1,†] Xiangxing Kong[1]
Qinggui Zhao[2]

[1]School of Mathematics, Central South University, Changsha, Hunan, 410075, China.
[2]School of Mathematics, Chongqing University of Arts and Sciences, Chongqing, 402160, China.

Abstract In this paper, we investigate a growing random graph process, which allows loops and multiple edges. Based on the first-passage probability of Markov chains, we provide a rigorous and straightforward proof to the existence of the degree distribution of a random graph process, and finally show the precise expression of degree distribution. Numerical simulations reveal that our theoretical derivations are consistent with simulation values.

Keywords Random graph; Degree distribution; Markov chains; First-passage probability

1 Introduction

Recently, there has been considerable interest in the growing properties of human interaction networks such as the World Wide Web(WWW), the citation networks and the social networks. These networks have different physical structures, with different definitions for their nodes and links. However they appear to display considerable topological similarity, having degree distributions behave as power laws.

Barabasi and Albert[1-3] found that for many real-world networks, e.g., WWW, the fraction of vertices with degree k is proportional over a large range to a power-law tail, i.e. $P(k) \sim k^{-\gamma}$, where γ is a constant independent of the size of the network. For purpose of opening up mechanism producing scale-free property, they proposed the well known BA model and summarized the reasons: growth and preferential attachment.

Because of the limitation of BA model and imprecision of mean field method, scholars make their efforts to put forward different techniques to solve it. For example, Krapivsky et al.[4] let $N_k(t)$ be the number of vertices in network with degree k at time t, and they used the rate equations to show degree distribution for the BA model with $m = 1$. Dorogovtsev et al.[5] considered $k_i(t)$, the degree of vertex added at time i evolved at time t. They used $P(k,i,t)$ to express the probability of vertex i having k edges at time t, and let the average degree $P(k,t) = \frac{1}{t}\sum_{i=1}^{t} P(k,i,t)$ to be the network degree at time t. They presented the attraction model, where each new vertex has an initial attraction a, and got

*Supported by the National Natural Science Foundation of China (Grant No. 10671212, 90820302)
†Email:zthou@csu.edu.cn

stationary distribution: $P(k) = \frac{2m(m+1)}{k(k+1)(k+2)}$ with $a = m$. In fact, the definitions of [4] and [5] are equivalent, since with properties of expectation, we have $E(N_k(t)) = \sum_{i=1}^{t} P(k,i,t)$.

Bollobás et al.[6] presented a modified BA model, also called LCD model. They used n-pairing method to study $E(\sharp_m^n(d))$, the average number of vertices with indegree d and applied martingale inequality[7] to show $\frac{\sharp_m^n(d)}{n} \sim d^{-3}$. Buckley and Osthus[8] generalized the work of Bollobás[6] to incorporate random selections. Jordan[9] used a slightly different mathematically precise version of [6] to obtain the degree distribution of [5]. They also generalized the model with randomization.

In this paper,we study the LCD model proposed by Bollobás et al.[6] again from a new perspective. Based on the first-passage probability of Markov chains, we propose a new approach to provide a rigorous proof to the existence of the degree distribution of this model, and we also prove that the degree distribution obeys power-law form. Moreover, numerical simulations reveal that our theoretical derivations are consistent with simulation values. Because of allowing loops and multiple edges, this paper also gives more detailed results similar to the special case of Cooper and Frieze[10].

2 Model Description

As Bollobás[6] stated, start with the case $m = 1$. Consider a fixed sequence of vertices $v_1, v_2, \cdots$. Inductively define a random graph process $(G_1^t)_{t\geq 0}$ so that G_1^t is a directed graph on $\{v_i : 1 \leq i \leq t\}$, as follows. Start with G_1^0 the graph with no vertices, or with G_1^1 the graph with one vertex and one loop. Given G_1^{t-1}, form G_1^t by adding the vertex v_t together with a single edge directed from v_t to v_i, where i is chosen randomly with

$$P\{i = s\} = \begin{cases} \frac{dG_1^{t-1}(v_s)}{2t-1} & 1 \leq s \leq t-1, \\ \frac{1}{2t-1} & s = t. \end{cases} \tag{1}$$

For $m > 1$ add m edges from v_t one at a time. Define the process $(G_m^t)_{t\geq 0}$ by running the process (G_1^t) on a sequence $v_1', v_2', \cdots$; the graph G_m^t is formed from G_1^{mt} by identifying the vertices $v_1', v_2', \cdots, v_m'$ to form v_1, identifying $v_{m+1}', v_{m+2}', \cdots, v_{2m}'$ to form v_2, and so on.

3 Degree distribution analysis with $m = 1$

Theorem 1.
For $m = 1$, the steady-state degree distribution of the LCD model exists, and is given by

$$P(k) = \frac{2(1+1)}{k(k+1)(k+2)} \sim 4k^{-3} > 0. \tag{2}$$

Instead of studying martingale $X_t = E[\sharp_1^n(k)|G_1^t)$ (where $0 \leq t \leq n$), following Dorogovtsev et al.[5], consider $k_i(t)$ as a random variable, and let $P(k,i,t) = P\{k_i(t) = k\}$ be the probability of vertex i having k edges at time t, also take the average degree $P(k,t) = \frac{1}{t}\sum_{i=1}^{t} P(k,i,t)$ to be the definition of the network degree at time t. Thus $k_i(t)$

is a nonhomogeneous Markov chain[11]. The state-transition probability of this Markov chain is given by:

$$P\{k_i(t+1)=l|k_i(t)=k\}=\begin{cases}1-a_{k,t} & l=k,\\ a_{k,t} & l=k+1,\\ 0 & otherwise.\end{cases} \tag{3}$$

where $a_{k,t}=\frac{k}{2t+1}$, $k=1,2,\cdots,t+2-i$, and $i=1,2,\cdots$.

Obviously, to the LCD model, $P(k,t,t)=\delta_{k,1}(1-a_{1,t-1})+\delta_{k,2}a_{1,t-1}$, where δ is the Kronecker-Delta function.

Consider the first-passage probability $f(k,i,t)$ of the Markov chain $k_i(t)$, $f(k,i,t)=P\{k_i(t)=k,k_i(l)\neq k,l=1,2,\cdots,t-1\}$. Based on the concept and techniques of Markov chains, the relationship between the first-passage probability and the probability of vertex degrees are as follows:

Lemma 1.
For $m=1$, when $k>1,i\neq s$, we have

$$f(k,i,s)=P(k-1,i,s-1)a_{k-1,s-1}, \tag{4}$$

$$P(k,i,t)=(1-a_{1,i-1})f(k,i,i+k-1)\prod_{j=i+k}^{t}(1-a_{k,j-1})+\sum_{s=i+k-2}^{t}f(k,i,s)\prod_{j=s+1}^{t}(1-a_{k,j-1}). \tag{5}$$

Proof. From the construction of LCD model, first passage to degree k at time s means the degree reaches $k-1$ at time $s-1$, and will get a degree at time s. This is what Eq.(4) means.

Because of allowing loops, the degree of a vertex can be 1 or 2 at the first step. If the degree is 1, the earliest time for the degree of vertex i to reach k is at step $i+k-1$, and the latest time to do so is at step t. Or else, the earliest time for the degree of vertex i to reach k is at step $i+k-2$. After the degree of this vertex becomes k, it will not increase any more. Thus Eq.(5) is established. □

Lemma 2.
For $m=1$, $\lim_{t\to\infty}P(1,t)$ exists, and

$$P(1)\triangleq\lim_{t\to\infty}P(1,t)=\frac{2}{3}>0. \tag{6}$$

Proof. From the construction of LCD model or with Eq.(3), it follows that

$$P(1,i,t+1)=(1-a_{1,t})P(1,i,t).$$

Summing i from 1 to $t+1$ on both sides, since $P(1,t+1,t+1)=1-a_{1,t}$, we have

$$P(1,t+1)=(1-a_{1,t})\frac{t}{t+1}P(1,t)+\frac{1}{t+1}(1-a_{1,t}).$$

With initial condition $P(1,1)=P(1,1,1)=0$ and Stolz theorem[12], Eq.(6) is obtained. □

Lemma 3.
For $m = 1$, when $k > 1$, if $\lim_{t\to\infty} P(k-1,t)$ exists, then $\lim_{t\to\infty} P(k,t)$ also exists and

$$P(k) \triangleq \lim_{t\to\infty} P(k,t) = \frac{k-1}{k+2} P(k-1) > 0. \tag{7}$$

Proof. Observe that $P(k,i,t) = 0$, when $i > t+2-k$. Since in this case even if the degree of vertex i increases by 1 each time, it can't reach degree k at t. Then, it follows from Lemma 1 that

$$\begin{aligned} P(k,t) &= \frac{1}{t}\sum_{i=1}^{t} P(k,i,t) = \frac{1}{t}\sum_{i=1}^{t+2-k} P(k,i,t) \\ &= P(t) + \frac{1}{t}\prod_{j=k}^{t}(1-a_{k,j-1}) \times \{P(k-1,k-2)\frac{(k-1)(k-2)}{2k-3} \\ &\quad + \sum_{l=k}^{t} P(k-1,l-1)\frac{(k-1)(l-1)}{2l-1}\prod_{h=k}^{l}(1-a_{k,h-1})^{-1}\}. \end{aligned}$$

where $P(t) = \frac{1}{t}\sum_{i=1}^{t+2-k}[(1-a_{1,i-1})P(k-1,i,i+k-2)a_{k-1,i+k-2}\prod_{j=i+k}^{t}(1-a_{k,j-1})]$. However, $P(t) < \frac{1}{t}\sum_{i=1}^{t+2-k}\frac{k-1}{2(i+k)-3} < (k-1)\frac{\ln t+\beta}{t}$, where β is the Euler constant. Take limits on both sides, it is easy to get $\lim_{t\to\infty} P(t) = 0$. Use Stolz theorem like Lemma 2, we have $\lim_{t\to\infty} P(k,t) = \frac{k-1}{k+2}P(k-1) > 0$. Therefore, $\lim_{t\to\infty} P(k,t)$ also exists and Eq.(7) is thus proved. □

Proof of Theorem 1
By mathematical induction, if follows from Lemmas 2 and 3 that the steady-state degree distribution of the LCD model with $m = 1$ exists. Then, solving Eq.(7) iteratively, we obtain $P(k) = \frac{k-1}{k+2}P(k-1) = \frac{2(1+1)}{k(k+1)(k+2)} \sim 4k^{-3} > 0$.□

4 Degree distribution analysis with $m \geq 2$

Theorem 2.
For $m \geq 2$, the steady-state degree distribution of the LCD model exists, and is given by

$$P(k) = \frac{2m(m+1)}{k(k+1)(k+2)} \sim 2m^2k^{-3} > 0. \tag{8}$$

For $m \geq 2$, according to the description of LCD model, m steps of graph G_1^{mn} generates one vertex in graph G_m^n. Note capital letters as the time and vertices in graph G_m^n, while lower case as the time and vertices in graph G_1^{mn}. From the step $(I-1)m+1$ to $(I-1)m+m$ in graph G_1^{mn} is the adding of the Ith vertex in graph G_m^n. Denote $t = (I-1)m+r$, where $I = 1,2,\cdots,\lceil \frac{t}{m} \rceil$; $r = 1,2,\cdots,m$. Consider $k_I(t)$, the degree of vertex I at time $t \geq (I-1)m+1$. For variable t, $k_I(t)$ is a nonhomogeneous Markov chain.

When $t = (I-1)m + r, r = 1, 2, \cdots, m-1$, the state-transition probability of this Markov chain is given by:

$$P\{k_I(t+1) = l | k_I(t) = k\} = \begin{cases} 1 - b_{k+1,t} & l = k+1, \\ b_{k+1,t} & l = k+2, \\ 0 & otherwise. \end{cases} \tag{9}$$

where $b_{k+1,t} = \frac{k+1}{2t+1}$ and $k = 1, 2, \cdots, 2r$.

When $t \geq Im$, the state-transition probability of this Markov chain is given by:

$$P\{k_I(t+1) = l | k_I(t) = k\} = \begin{cases} 1 - c_{k,t} & l = k, \\ c_{k,t} & l = k+1, \\ 0 & otherwise. \end{cases} \tag{10}$$

where $c_{k,t} = \frac{k}{2t+1}$ and $k = m, m+1, \cdots, 2m+t-Im$.

Let $P(k,I,T) = P\{K_I(T) = k\}$ (or $\hat{P}(k,I,t) = P\{k_I(t) = k\}$) be the degree probability of vertex I at step T (or t). As above, take the average degree $P(k,T) = \frac{1}{T}\sum_{I=1}^{T} P(k,I,T)$ to be the definition of the network degree at time T.

Consider the first-passage probability of the Markov chain $k_I(t)$. $\hat{f}(k,I,t) = P\{k_I(t) = k, k_I(l) \neq k, l = 1, 2, \cdots, t-1\}$ and the first-passage probability of the Markov chain $k_I(T)$. $f(k,I,T) = P\{K_I(T) = k, K_I(L) \neq k, L = 1, 2, \cdots, T-1\}$. Based on the concept and techniques of Markov chains, we have:

Lemma 4.
For the first-passage probability, when $1 < I = S$, we have

$$f(k,I,I) = \begin{cases} \prod_{r=1}^{m} (1 - b_{r,(I-1)m+r-1}) & k = m, \\ \prod_{r=1}^{m} b_{2(r-1)+1,(I-1)m+r-1} & k = 2m, \\ \sum_{r=k-m}^{m} \hat{f}(k-(m-r), I, (I-1)m+r) \cdot & \\ \prod_{q=r+1}^{m} (1 - b_{k-(m-r)+(q-r),(I-1)m+q-1}) & m < k < 2m. \end{cases} \tag{11}$$

When $1 < I = S, m < k < 2m$, we have

$$\hat{f}(k-(m-r), I, (I-1)m+r) = \hat{P}(k-(m-r)-2, I, (I-1)m+r-1) b_{k-(m-r)-1,(I-1)m+r-1}. \tag{12}$$

When $1 \leq I < S, k > m$, we have

$$f(k,I,S) = \sum_{r=1}^{m} \hat{f}(k,I,(S-1)m+r) \prod_{u=r+1}^{m} (1 - c_{k,(S-1)m+u-1}), \tag{13}$$

$$\hat{f}(k,I,(S-1)m+r) = \hat{P}(k-1,I,(S-1)m+r-1) c_{k-1,(S-1)m+r-1}. \tag{14}$$

For the vertex degree probability, when $1 < I \leq T$, we have

$$P(k,I,T) = \begin{cases} \prod_{r=1}^{m}(1-c_{r,(I-1)m+r-1}) \prod_{A=I+1}^{T}\prod_{q=1}^{m}(1-c_{m,(A-1)m+q-1}) & k=m, \\ \sum_{S=I}^{T} f(k,I,S) \prod_{A=S+1}^{T}\prod_{r=1}^{m}(1-c_{k,(A-1)m+r-1}) & k>m. \end{cases} \tag{15}$$

Proof. From Eq.(9), it is easy to obtain Eq.(11)-Eq.(12), substitute Eq.(9) for Eq.(10), we will get Eq.(13)-Eq.(15). □

Lemma 5.
For $m \geq 2$, $\lim_{T\to\infty} P(m,T)$ exists, and

$$P(m) \triangleq \lim_{T\to\infty} P(m,T) = \frac{2}{m+2} > 0. \tag{16}$$

Proof. For $1 \leq I \leq T$, from the construction of LCD model, it is easy to get that $P(m,1,T) = 0$. With Eq.(15) of Lemma 4 and Stolz theorem, Eq.(16) is proved like Lemma 2. □

Lemma 6.
For $m \geq 2$, when $k > m$, we have

$$\lim_{T\to\infty} \frac{1}{T} \sum_{I=1}^{T-1} f(k,I,I) \prod_{A=I+1}^{T}\prod_{r=1}^{m}(1-b_{k,(A-1)m+r-1}) = 0. \tag{17}$$

And if $\lim_{T\to\infty} P(k-1,T) = P(k-1)$ exists, then

$$\lim_{T\to\infty} \frac{1}{T} \sum_{I=1}^{T} \hat{P}(k-1,I,Tm+r-1) = P(k-1). \tag{18}$$

Proof. With Eq.(11) and Eq.(12)of Lemma 4, Eq.(17) is easy to be established. Eq.(18) is obtained with principle of adjacency. □

Lemma 7.
For $m \geq 2$, when $k > m$, if $\lim_{T\to\infty} P(k-1,T)$ exists, then $\lim_{T\to\infty} P(k,T)$ also exists and

$$P(k) \triangleq \lim_{T\to\infty} P(k,T) = \frac{k-1}{k+2} P(k-1) > 0 \tag{19}$$

Proof. It follows from Eq.(15) of Lemma 4 that

$$\begin{aligned} P(k,T) &= \frac{1}{T}P(k,T,T) + \frac{1}{T}\sum_{I=1}^{T-1} f(k,I,I) \prod_{A=I+1}^{T}\prod_{r=1}^{m}(1-c_{k,(A-1)m+r-1}) \\ &\quad + \frac{1}{T}\sum_{I=1}^{T-1}\sum_{S=I+1}^{T} f(k,I,S) \prod_{A=S+1}^{T}\prod_{r=1}^{m}(1-c_{k,(A-1)m+r-1}) \end{aligned}$$

Obviously, the limit of first item is also 0. With Lemma 6, the limit of second item is 0. We only need to consider the third item.

From Eq.(13) and Eq.(14), we obtain

$$
\begin{aligned}
&\frac{1}{T}\sum_{I=1}^{T-1}\sum_{S=I+1}^{T} f(k,I,S)\prod_{A=S+1}^{T}\prod_{r=1}^{m}(1-c_{k,(A-1)m+r-1})\\
=\;&\frac{1}{T}\prod_{A=3}^{T}\prod_{q=1}^{m}(1-c_{k,(A-1)m+q-1})\Big\{\\
&\sum_{r=1}^{m}\hat{P}(k-1,I,m+r-1)c_{k-1,m+r-1}\prod_{u=r+1}^{m}(1-c_{k,m+u-1})\\
&+\sum_{S=3}^{T}\sum_{r=1}^{m}\sum_{I=1}^{S-1}\hat{P}(k-1,I,(S-1)m+r-1)c_{k-1,(S-1)m+r-1}\prod_{u=r+1}^{m}(1-c_{k,(S-1)m+u-1})\\
&\prod_{B=3}^{S}\prod_{h=1}^{m}(1-c_{k,(B-1)m+h-1})^{-1}\Big\}.
\end{aligned}
$$

With Stolz theorem, we have $P(k) \triangleq \lim_{T\to\infty} P(k,T) = \frac{k-1}{k+2}P(k-1) > 0$. □

Proof of Theorem 2 By mathematical induction, if follows from Lemmas 5 and 7 that the steady-state degree distribution of the LCD model with $m \geq 2$ exists. Then, solving Eq.(19) iteratively, we obtain $P(k) = \frac{k-1}{k+2}P(k-1) = \frac{2m(m+1)}{k(k+1)(k+2)} \sim 2m^2k^{-3} > 0$.□

5 Simulations

Section 3 and 4 shows that the degree distribution of the LCD model follows power-law with degree exponent 3. In this section, we performed numerical simulations to illustrate the schemes discussed above. The following figures are simulations and analytic results of LCD model with $m = 1,2,5$. Figure 1 is the case of $m = 1$, the simulation results fit with $P(k) \sim 4k^{-3}$ very well. Figure 2 is the case of $m > 1$, the simulation results fit well with $P(k) \sim 2m^2k^{-3}$ either. All numerical values are the average of 20 times.

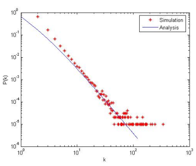

Figure 1: Degree distribution of the LCD model with $m = 1, t = 2000$. The slope of the solid line is $\gamma = 3$, which corresponds to Eq.(2) of theoretical derivation.

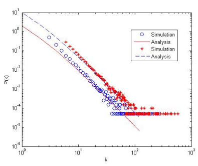

Figure 2: Degree distribution of the LCD model with $m \geq 1$. $\circ$ and $*$ are simulation results with $m = 2$, $t = 2000$ and $m = 5$, $t = 2000$ respectively. The solid line and dashed line correspond to Eq.(8) of theoretical derivation with $m = 2$ and $m = 5$ respectively. Both of their slope are $\gamma = 3$.

6 Conclusion

From the perspective of Markov chains, we considered the degree of vertex i at time t i.e., $k_i(t)$ as a Markov chain and with the technical of first-passage probability, we give a rigorous proof to the existence of the steady degree distribution and give the exact formula of degree distribution to the LCD model. In order to check the feasibility of the analytical results obtained in this paper, we compare the analytical results with the simulations, we can observe that there is an overall good agreement between the simulated data and the analytical results. The method proposed in this paper is of universal and can be used in other models, such as the attraction model[5] and models whether allowing multiple edges or not.

References

[1] Barabási A.L., Albert R., Emergence of scaling in random networks. Science. 286, 509-512(1999).

[2] Albert R., Jeong H., Barabási A.L., Diameter of the World-Wide web. Nature. 401, 130-131(1999).

[3] Barabási A.L., Albert R. and Jeong H., Mean-field theory for scale-free random networks. Physics A. 272, 173-187(1999).

[4] Krapivsky P.L., Redner S., Leyvraz F., Connectivity of growing random networks. Phys. Rev. Lett. 85, 4629-4632(2000).

[5] Dorogovtsev S.N., Mendes J.F.F. and Samukhin A.N., Structure of Growing Networks with Preferential Linking. Phys. Rev. Lett. 85, 4633-4636(2000).

[6] Bollobás B., Riordan O.M., Spencer J. and Tusnády G., The degree sequence of a scale-free random graph process. Rand. Struct. Algor. 18, 279-290(2001).

[7] Hoeffding W., Probability inequalities for sums of bounded random variables. J. Amer. Statist. Assoc. 58, 13-30(1963).

[8] Buckley P.G., Osthus D., Popularity based random graph models leading to a scale-free degree degree sequence. Discrete Mathematics. 282, 53-68(2004).

[9] Jordan J., The degree sequences and spectra of scale-free random graphs. Rand. Struct. Algor. 29, 226-242(2006).

[10] Cooper C. and Frieze A., A general model of web graphs. Rand. Struct. Algor. 22(3), 311-335(2003).

[11] Shi D.H., Chen Q.H., Liu L.M., Markov chain-based numerical method for degree distribution of growing networks. Phys. Rev. E. 71, 036140(2005).

[12] Stolz O., Vorlesungen uber allgemiene Arithmetic. Teubner, Leipzig(1886).