2013 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE)

Scale-Free Property of Copying Model

Rongrong Dai, Zhenting Hou
School of Mathematics
Central South University
Changsha, China
RongrongDAI@csu.edu.cn

***Abstract*—We study and generalize the copying model of Kumar et al. [Stochastic models for the web graph, FOCS 2000]. The basic idea is that a new web is often made by copying an old one, and then changing some of the links. Our main contributions are as follows: we show that (1) the copying model of Kumar et al. does generate a power-law degree distribution as stated in Kumar et al. [Stochastic models for the web graph, FOCS 2000]; (2) We generalize the Kumar et al. copying process. We prove that this generalized model has a power-law degree distribution and give the exact solution from the perspective of Markov chain.**

Keywords-copying model; scale-free; Markov chain

I. INTRODUCTION

Many natural and man-made complex systems can be represented in terms of networks, in which the vertices stand for the elementary units, while the edges picture the interactions between them [1].

The classical random graph models studied by Erdos and Renyi which is a static model with a Poisson degree distribution [2]. However, it is well known that many "real-world" complex systems such as neuronal networks, social networks, and scientific-collaboration networks show a power-law degree distribution which is absent in Erdos and Renyi graphs [3-5]. For seek to understand the growth dynamics, there are a number of recently developed alternative random graph models which do generate power-law degree distributions. For example, in the preferential attachment model (BA model, Barabasi and Albert) the probability that an existing vertex is connected to the newly created vertex increases with the degree of the vertex [6].

The preferential attachment model was proposed as a random graph model for the web by Barabasi and Albert [6]. And it is perhaps the most basic 'scale-free' random graph models. Barabasi and Albert proved that the degree sequence of such graphs does follow a power-law distribution. Attention has also been given to models where the attractiveness of vertices fades over time [7]. More recently Cooper and Frieze gave a general analysis of random graph processes revealing that many graphs generated by preferential attachment exhibit power-law degree distributions [8]. This analysis obtained graphs with a power-law parameter larger than 2 but smaller than 3 by using a graph generation model that allows edge insertion between existing vertices. The insight behind these concepts is the realization of two facts: Firstly, most complex networks are the result of a growth process, in which new vertices are added in time to the system. Secondly, new edges are not placed at random, but preferentially connect to vertices which already have a larger degree. It turns out that these two ingredients are able to reproduce power-law degree distributions [6]. Moreover, it has been shown that not all sorts of preferential attachment are able to generate a power-law degree distribution, but only those in which new edges attach to vertices with a probability strictly linear in their degree [9].

Networks such as the World Wide Web do not grow in an orderly manner. For example, the Web is created by millions of users and lacks an engineered architecture. Although such networks are complex in structure, their large size is a simplifying feature, and for infinitely large networks we can provides analytical predictions for basic network characteristics. The World Wide Web, are directed graphs, i.e., their vertices are connected by directed edges. Some alternative mechanisms, such as the copying model implicitly define a linear preferential attachment dynamics. Around the same time as the BA model, Kumar, Raghavan, Rajagopalan, Sivakumar, Tomkins and Upfal gave rather different model to explain the observed power laws in the web graph, as well as explaining the high number of dense bipartite sub-graphs found in the web graph [10]. The basic idea is that a new web page is often made by copying an old one, and then changing some of the links. Kumar et al. suggested the following random graph process as a model [10].

The copying model of Kumar et al. is parameterized by a copy factor $\alpha \in [0,1)$ and a constant out-degree $m \geq 1$. At each time step, one vertex u with m out-links is added [10]. We begin by choosing a 'prototype' vertex p uniformly at random from V_t (the old vertices). The i^{th} out-link of u is then chosen as follows: with probability α, the destination is chosen uniformly at random from V_t, and with the remaining probability $1-\alpha$, the out-link is taken to be the i^{th} out-link of p. Thus, the prototype is chosen once in advance. The m out-links are chosen by α-biased independent coin flips, either randomly from V_t, or by copying the corresponding out-link of the prototype.

The intuition behind this model is the following. When an author decides to create a new web page, the author is likely to have some topic in mind. The choice of prototype represents the choice of topic-larger topics are more likely to be chosen. The Bernoulli copying events reflect the following intuition: a new viewpoint about the topic will probably link to many pages 'within' the topic, but will also probably introduce a new spin on the topic, linking to some new pages whose connection to the topic was previously unrecognized.

As for the BA model, it turns out that the degree distributions exhibit a finite-size scaling behavior. In fact, since the prototype vertices are selected randomly, the probability that a Web page with incoming degree k will receive a new hyperlink is proportional to $(1-\alpha)k+m\alpha$, indicating that the copying mechanism effectively amounts to a linear preferential attachment. Note also that the mixture of uniform and preferential attachment is easily seen to be equivalent to the preferential attachment with constant initial attractiveness [11]. Kumar et al. prove that the expectation of the incoming degree distribution is [10]

$$P\left(k_{in}\right) = k_{in}^{-\frac{2-\alpha}{1-\alpha}} \tag{1}$$

Thus $P\left(k_{in}\right)$ follows a power-law with an exponent that varies between 2 and ∞ .These models were designed explicitly to model the World Wide Web. Indeed, they show that their model has a large number of complete bipartite sub-graphs, as has been observed in the WWW graph, whereas several other models, do not.

But in Lemma 3 as stated in Kumar et al., one claim which the lemma follows from has a little mistake in the estimate [10]:

$$Q_1 = \sum_{k=1}^{t-1} \prod_{i=1}^{k} (1-\alpha/l)\theta \cdot 2, \ with\ 0 < \alpha < 1 \tag{2}$$

and

$$\mathrm{Q}_1 \geq \frac{t-1}{1+\alpha} - \alpha^2 \ln t \tag{3}$$

Is it in evidence?

$$\begin{aligned} &1-\alpha \geq \frac{Q_1}{t-1} \geq \frac{1}{1+\alpha} - \alpha^2 \frac{\ln(\mathrm{t}-1)}{t-1} \\ &\Rightarrow 1 \leq (1+\alpha)\frac{\ln(\mathrm{t}-1)}{t-1} \end{aligned} \tag{4}$$

Contradiction.

So, we give the exact solution of the degree distribution from the perspective of Markov chain by using the formalism based on previous work of Hou et al. [12].

II. Degree Distribution of Copying Model

Let $k_i^+(\mathrm{n})$ and $k_i^-(\mathrm{n})$ denote the random variables as the number of vertices with in-degree and out-degree of the vertex which inter the network at time i . Obviously, $k_i^-(\mathrm{n}) = \mathrm{m}$.

Let $P(k,i,n) = \mathrm{P}(k_i^+(\mathrm{n}) = \mathrm{k})$ denote that probability that the degree of the vertex which inter the network at time i arrive k in the graph at time t .

$$P(k,n) = \frac{1}{n}\sum_{i=1}^{n} P(k,i,n) \tag{5}$$

If $\lim_{n\to\infty} P(k,n)$ exists, then $P(\mathrm{k}) = \lim_{n\to\infty} P(k,n)$.

The model indicates that the evolution of $k_i^+(t)$ is independent of the network before t-1 when given its current state. So $\{k_i^+(t)\}$ is a Markov chain. We obtain the following transition probability

$$\begin{aligned} &\mathrm{P}\{k_i^+(t+1) = 1 \mid k_i^+(t) = \mathrm{k}\} = \\ &\begin{cases} \frac{m\alpha}{t} + (1-\alpha)\frac{k}{t} & l = k+1, \\ 1 - \left\{\frac{m\alpha}{t} + (1-\alpha)\frac{k}{t}\right\} & l = k, \\ 0 & otherwise. \end{cases} \end{aligned} \tag{6}$$

Let the probability arriving for the first time $f(k,i,s) = P\{k_i^+(s) = k, k_i^+ < \mathrm{k}, \mathrm{l} = 1,2,\ldots,\mathrm{s}-1\}$ denote the probability that the degree of the vertex which inter the network at time i first arrive k at time s .

Lemma 1: when $k>0$,

$$f(k,i,s) = P(\mathrm{k}-1,\mathrm{i},\mathrm{s}-1)(\frac{m\alpha}{s-1} + \frac{k-1}{s-1}) \tag{7}$$

Lemma 2: $\lim_{n\to\infty} P(0,n)$ exists and

$$P(0) = \lim_{n\to\infty} P(0,n) = \frac{1}{1+m\alpha} \tag{8}$$

Proof: From the copying model, the only way that a vertex without in-link chosen by new vertex is uniformly at random, so

$$P(0,i,n+1) = (1-\frac{m\alpha}{n})P(0,i,n) \tag{9}$$

And the new coming vertex is always arriving in the network with 0 in-link

$$P(0,n+1,n+1) = 1 \tag{10}$$

We obtain

$$P(0,n+1) = \frac{1}{n+1}\sum_{i=1}^{n+1} P(0,\mathrm{i},n) = (1-\frac{m\alpha}{n})\frac{n}{n+1}P(0,n) + \frac{1}{n+1} \tag{11}$$

With Stolz Theorem [13], we have

$$P(0) = \lim_{n\to\infty} P(0,n) = \frac{1}{1+m\alpha} \tag{12}$$

Theorem 3: $k>0$, if $P(k-1)=\lim_{n\to\infty}P(k-1,n)$ exists and it is positive, then $P(k)=\lim_{n\to\infty}P(k,n),(\mathrm{k}=1,2,...)$ also exists, and

$$P(k)=\frac{k-\dfrac{1-2\alpha}{1-\alpha}}{k+\dfrac{1+\alpha}{1-\alpha}}P(k-1)\sim k^{-(1+\frac{1}{1-\alpha})} \tag{13}$$

Proof: When $k>0$, from the Lemma 1,

$$\begin{aligned} P(\mathrm{k},n) &= \frac{1}{n}\sum_{i=1}^{n}P(\mathrm{k},\mathrm{i},n) \\ &= \frac{1}{n}\sum_{i=1}^{n-k+1}\sum_{s=i+k-1}^{n}f(\mathrm{k},\mathrm{i},s)\prod_{j=s}^{n-1}(1-(\frac{m\alpha}{j}+(1-\alpha)\frac{k}{j})) \\ &= \frac{1}{n}\sum_{i=1}^{n-k+1}p(\mathrm{k-1},\mathrm{s-1})(\mathrm{s-1}) \qquad (14) \\ &\qquad (\frac{m\alpha}{s-1}+\frac{k-1}{s-1}\prod_{j=s}^{n-1}(1-(\frac{m\alpha}{j}+(1-\alpha)\frac{k}{j}))) \end{aligned}$$

with Stolz Theorem [13], we have

$$P(k)=\frac{k-\dfrac{1-2\alpha}{1-\alpha}}{k+\dfrac{1+\alpha}{1-\alpha}}P(k-1) \tag{15}$$

Thus, Theorem 3 is proved.

So, the power of in-degree distribution is $1+1/(1-\alpha)$. Clearly, the fact that the power $1+1/(1-\alpha)$ is regardless of the out-degree m is a drawback of the model. Moreover, it can easily be shown that all of edges of a resulting graph can be decomposed into m disjoint forests. Presumably, most massive real-world graphs with power-law degree distributions have a richer structure than this. As we will see, by considering a different evolution in our general model, our analysis does yield a degree distribution power-law dependent of m.

III. A Generalized Copying Model

Strongly, in the opinion of author, the prototype is among the new spin on the topic, but in fact, it is also the existing resource about the topic. So, we introduce a new copying model which the author consider the prototype firstly. As the copying model, The generalized copying model is also parameterized by a copy factor $\alpha\in(0,1)$ and a constant out-degree $m>1$. At each time step, one vertex u is added, and u is then given m out-links. To generate the out-links, we begin by choosing a 'prototype' vertex p uniformly at random from V_t (the old vertices). The i^{th} out-link of u is then chosen as follows:

(1) With probability β, the destination of the first out-link of u is prototype p, and with the remaining probability 1-β, is chosen uniformly at random from V_t.

(2) With probability α, the destination of the remaining $m-1$ out-links is chosen uniformly at random from V_t, and with the remaining probability $1-\alpha$ the out-link is taken to be the i^{th} out-link of p. The $m-1$ out-links are chosen by α-biased independent coin flips, either randomly from V_t, or by copying the corresponding out-link of the prototype.

The intuition behind this model is the following. When an author decides to create a new web page, the author is likely to have some topic in mind. The choice of prototype represents the choice of topic-larger topics are more likely to be chosen. The Bernoulli copying events reflect the following intuition: a new viewpoint about the topic will probably link to many pages 'within' the topic (i.e., pages already linked to by existing resource lists about the topic), but will also probably introduce a new spin on the topic, linking to some new pages whose connection to the topic was previously unrecognized.

We obtain the following transition probability

$$\mathrm{P}\{k_i^+(t+1)=l\,|\,k_i^+(t)=\mathrm{k}\}= \begin{cases} \dfrac{\beta}{t}+\dfrac{1-\beta}{t}+\dfrac{(\mathrm{m}-1)\alpha}{t}+\left(1-\alpha\right)\dfrac{m-1}{m}\dfrac{k}{t} & l=k+1, \\ 1-(\dfrac{\beta}{t}+\dfrac{1-\beta}{t}+\dfrac{(\mathrm{m}-1)\alpha}{t}+\left(1-\alpha\right)\dfrac{m-1}{m}\dfrac{k}{t}) & l=k, \\ 0 & otherwise. \end{cases} \tag{16}$$

Lemma 4: when $k>0$,

$$f(k,i,s)=P(\mathrm{k-1},\mathrm{i},\mathrm{s-1})(\frac{\beta}{s}+\frac{1-\beta}{s-1}+\frac{(m-1)\alpha}{s-1}+(1-\alpha)\frac{m-1}{m}\frac{k}{s-1}) \tag{17}$$

Lemma 5: $P(0)=\lim_{n\to\infty}P(0,n)$ exists and

$$P(0)=\lim_{n\to\infty}P(0,n)=\frac{1}{2+(m-1)\alpha} \tag{18}$$

Theorem 6: $k>0$, if $P(\mathrm{k}-1)=\lim_{n\to\infty}P(k-1,n)$ exists and it is positive, then $P(\mathrm{k})=\lim_{n\to\infty}P(k,n),(k=1,2,...)$ also exists, and

$$P(k)=\frac{k-1}{k+\dfrac{m}{(m-1)(1-a)}}P(k-1)\sim k^{-(1+\frac{m}{(m-1)(1-a)})} \tag{19}$$

The proof of Lemma 5 and Theorem 6 are similar with Lemma 2 and Theorem 3.

So, the power of in-degree distribution is $1+m/(m-1)(1-\alpha)$ which varies between $1+1/(1-\alpha)$ (for $m\to\infty$) and $1+2/(1-\alpha)$ (for m=2). Obviously, the power $1+m/(m-1)(1-a)$ is independent of the parameter β. However, it successfully depends of m. Even for $\beta=0$, the generalized copying model is a different model with the Kumar et al. model.

IV. Conclusion

In this paper we propose a generalized coping model and obtain the exact solution of the degree distribution by abstract Markov chain. The degree distribution of the new copying model depends on the parameter *m* of the model.

References

[1] B. Bollobas, Modern Graph Theory. New York: Springer-Verlag, 1998.

[2] H. Wang and H. Pham, Reliability and Optimal Maintenance. Berlin: Springer, 2006.

[3] D. J. Watts and D. H. Strogatz, "Collective dynamics of small-world networks," Nature, vol. 393, no. 7, pp. 440-442, 1998.

[4] S. Wasserman and K. Faust, Social Network Analysis. Cambridge: Cambridge university Press, 1994.

[5] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, "Classes of small-world networks," PNAS, vol. 97, no. 21, pp. 11149-11152, 1990.

[6] A. L. Barabasi and R. A. Albert, "Emergence of scaling in random networks," Science, vol. 286, no. 7, pp. 509-512, 1999.

[7] S. N. Dorogovstev and J. F. F. Mendes, "Evolution of networks with aging of sites," Physical Review Letters, vol. 62, no.1, pp. 1842-1850, 2000.

[8] C. Cooper and A. Frieze, "A general model of web graphs," Random Structures Algorithms, vol. 22, no. 3, pp. 311-335, 2003.

[9] P. L. Krapivsky, S. Redner, and F. Leyvraz, "Connectivity of growing random networks," Physical Review Letters, vol. 85, no. 21, pp. 4629-4632, 2000.

[10] R. Kumar, "Stochastic models for the web graph," In Proceedings of the 41th IEEE Annual Symposium on Foundations of Computer Science, pp. 57-65, 2000.

[11] S. N. Dorogovtsev, J. E. F. Mendes, and A. N. Samukhin, "Structure of growing networks with preferential linking," Physical Review Letters, vol. 85, bo. 21, pp. 4633-4636, 2000.

[12] Z. T. Hou, X. X. Kong, D. H. Shi, and G. R. Chen, "Degree-distribution stability of growing networks, lecture notes of the institute for computer sciences," Social Informatics and Telecommunications Engineering, vol. 5, no. 1, pp. 1827-1837, 2009.

[13] O. Stolz, Vorlesungen Uber Allgemiene Arithmetic. Leipzig: Teubner, 1886.