Vol 13 No 4, April 2004
1009-1963/2004/13(04)/0423-05

Chinese Physics

© 2004 Chin. Phys. Soc.
and IOP Publishing Ltd

Mirror nodes in growing random networks

Dai Shuo(戴 硕)† and Guo Yun-Jun(郭云均)

Department of Physics, Nanjing University, Nanjing 210093, *China*

(Received 16 July 2003; revised manuscript received 9 September 2003)

In this article, we consider mirror nodes, which are widely used to reduce local burden, and present two models that are practical for the Internet and probably some other networks. One model introduces an upper limit to the number of links that a node can have, beyond which the node will share the total links with a newly introduced one. Similarly, in the second model, if the number of links exceeds a limit, a new node will be introduced. Unlike the first model, the new node shares with the old one the chance of receiving new links but not the existing links. These models are analytically treated, and from the degree distribution, we can see that the number of nodes with medium links (half the upper limit) increases at the expense of the loss of highly connected nodes. By reducing the burden, this may improve the robustness of the networks.

Keywords: Internet, complex network systems, statistical mechanics
PACC: 0250, 0240

1. Introduction

In recent years, the Internet has played an increasingly important role in all branches of society, and how it may work efficiently becomes a central problem. The same problem also exists in various other working systems that can be described as complex networks,[1–3] such as epidemiology, ecology and gene regulation. The efforts on modelling complex networks have just begun and interesting features have been only partially revealed (see Refs.[1–3] for a review). In this field, one of the most important models is the Barabasi–Albert (BA) scale-free network model,[4] which considers network growth and preferential attachment. In this model, at each time step, a node is introduced and it is attached to an existing node with a probability proportional to the degree of the node (here degree is defined as the number of links of a node). Later, Krapivsky and Redner retained this basic idea and in the so-called growing random networks they considered more general situations (for details see below and Ref.[5]).

In the growing random network model, the rate equation can be written as

$$\frac{\mathrm{d}N_k}{\mathrm{d}t} = A^{-1}(A_{k-1}N_{k-1} - A_kN_k) + \delta_{k1}, \qquad (1)$$

where N_k is the number of nodes with degree k, and $A^{-1}A_k$ is the probability that a newly introduced node is linked to an existing node that already has k links. Assuming N_k and $A(t)$ go linearly with time, i.e. $N_k = n_k \cdot t$ and $A(t) = \mu \cdot t$, one can obtain the distribution of n_k.[5] This model, along with the simpler BA model, characterizes a great variety of networks in reality, including the Internet,[6–10] as mentioned at the beginning.

As the burden of the Internet grows heavier, in order to settle the consequent problems, new methods and techniques appear, which create new features that call for corresponding modifications to the former models.

One result of the heavy burden is that too many links may, as one can imagine, make a server unstable and may possibly lead to system downfall. Actually, a particular method is widely applied: i.e., people often use many servers for a single popular website simultaneously to maintain stability.

On the other hand, as the network grows larger, it may become more difficult for a newly introduced node to contact with a popular one that has many nodes due to the increasing distance. Therefore one may try to introduce new nodes as the mirror nodes of the original one, and this can certainly shorten the average distance of visit. In fact, the Internet has various mirror sites, which can be modelled as mirror nodes.

Considering the two improvements described above, we present two types of model, which are

†E-mail:daishuo2000@yahoo.com

based on growing random networks.[2] One is called the group server model and the other is called the mirror node model. The assumptions are each based on different prevailing techniques that are widely adopted today. In the following, we can see that the group-servers model gives the effect of diverting clusters and the mirror-nodes model gives the effect somewhat like 'satellite towns'.

2. Description of the models

2.1. Model I (group servers)

Just as in the growing random network model (the GN model),[5] a new node is introduced at each time step and is connected with one of the existing nodes in the network. For a randomly picked existing node with k links, the probability that a newly introduced node is linked to it is defined as A_k/A, where A is a normalization factor. In the GN model, A_k is proportional to k^γ. If $\gamma > 0$ (and this is the case that is of practical interest), the highly connected nodes will have a larger probability to receive new links. In the GN model, there is no upper limit of the links that a node can ultimately have, which means that for any node, the number of links it has may grow infinitely. But here, based on the considerations stated in Section 1, we try to keep it limited. The specific value of this limit may vary among different nodes. For example, for some nodes, the upper limit number may be $2C$; for other nodes, the upper limit number may be $2D$. When the number of links of a certain node (called H_1) exceeds its limit, we should introduce another node (called H_2), and each of the two nodes has half of the total links, that is C or D. When receiving new links, the two nodes equally share a probability that is obtained by viewing the two nodes as a single one in the context of the network. In this way, we can build a network whose nodes have an upper limit of connectivity. As is mentioned above, in reality the specific value of this limit may vary among different nodes. However, for the Internet, the technique is approximately on the same level all over the world. The limit of connectivity cannot exceed the capacity of the system. On the other hand, if this limit is too small, it will cause a waste of resource. Thus, we may suppose that there is a single-peak distribution of the limit of connectivity and use the average value instead of considering all the possibilities.

2.2. Model II (mirror nodes)

In Model II, we define the upper limit number as M. Similarly, when the number of the links of one node (called B_1) exceeds the upper limit, a mirror node (called B_2) will be introduced. The links of the original node remain unchanged, and initially the mirror node has no links to outside nodes. Then (B_1+B_2) is viewed as a single one when we calculate the probability that it receives new links, and this probability is further equally shared by B_1 and B_2. So, for a newly introduced node, the chance to link with the original node and the mirror node is the same. Although there is a link between the original node and the mirror node, this link has no effect on the probability of receiving new links. It is simply used to connect the original node and the mirror node.

Models I and II are similar in that when the links of the original node exceeds a certain limit, a new node is introduced and it has an equal probability of receiving new links as the original one. The difference is that, in Model I, the original node H_1 has to share the total links with the new node H_2, but in Model II, the mirror node B_2 has only one 'link' to the original node B_1 without sharing any of the existing links of B_1.

3. The degree distribution of the two models

3.1. The degree distribution of Model I

In Model I, the group server model, when the links of one node reach the upper limit, the node will be separated, and its links are equally divided. As mentioned above, the sub-nodes are viewed as a whole when receiving new links. Thus, with some necessary modification, the results of the GN model are still applicable. We should divide the nodes with the links exceeding the upper limit until all of the sub-nodes have their connectivity within the assumed range, and then we can obtain the degree distribution.

We assume $A_k = k^\gamma$, and the upper limit number is $2C$. For the nodes with the number of links smaller than C, the distribution is the same as in the GN model. In the GN model we define $n_k = N_k/t$, where N_k denotes the number of nodes with degree k, and we have

$$n_k = \frac{\mu}{k} \prod_{i=1}^{k} \left(1 + \frac{\mu}{i}\right)^{-1}, \qquad (2)$$

where μ is a parameter related to γ, and $\mu = 2$ for $\gamma = 1$ (for details see Ref.[5]). Since in the GN model

the total number of nodes N equals time t, we have $P_k = n_k$, where P_k is the probability distribution that a randomly picked node has degree k. In the present model the total number of nodes is larger than t, but we still have $P_k \propto n_k$, and

$$P_k = \frac{1}{P}\frac{\mu}{k}\prod_{i=1}^{k}\left(1+\frac{\mu}{i}\right)^{-1}, \quad (\text{when } k < C), \qquad (3)$$

where $1/P = t/N$ is a normalization factor.

Now we turn to the nodes with the number of links no smaller than C, and here we have to take into consideration the sub-nodes of the nodes with their degrees exceeding the upper limit, $2C$. First we calculate the number of the nodes with degree C. Obviously n_c itself has a contribution to P_c. From n_{c+1} to n_{2c-1} there is no contribution, and what we need to know is how to separate nodes with links more than $2C$. There are two sources of contribution.

(1) When $k = 2C$ the node should be separated into two sub-nodes, each of which has C links; when $k = 4C$, the node should be separated into four sub-nodes, each with C links, etc; when $k = 2^nC$ the node should be separated into 2^n sub-nodes, each of which has C links.

There are other nodes that also have a contribution. These nodes have their number of links close to 2^nC, and they can be further divided into two groups.

(2) When $k = 2^nC + 1$, the node should be separated into 2^n sub-nodes, of which $2^n - 1$ nodes have C links, and the remaining one has $C+1$ links. When $k = 2^nC + 2$ the node should be separated into 2^n sub-nodes, of which $2^n - 2$ sub-nodes have C links, and the remaining ones have $C + 1$ links, etc. The final case is that, when $k = 2^nC + 2^n - 2$, the node should be separated into 2^n sub-nodes, of which only one sub-node has C links, and all of the remaining ones have $C + 1$ links.

(3) When $k = 2^nC - 1$, the node should be separated into $2^n - 1$ sub-nodes, of which $2^n - 2$ sub-nodes have C links, and the other one has $2C - 1$ links; when $k = 2^nC - 2$, the node should be separated into $2^n - 2$ sub-nodes, of which $2^n - 4$ have C links, and the remaining ones have $2C - 1$ links, etc. The final case is that, when $k = 2^nC - 2^{n-1} + 1$, the node should be separated into $2^{n-1} + 1$ sub-nodes, of which only two have C links, and all of the remaining ones have $2C - 1$ links.

When $2^nC + 2^n \le k \le 2^{n+1}C - 2^n$, the nodes n_k have no contribution. So we get the general formula:

$$P_C = \frac{1}{P}\sum_{n=0}^{\infty}\left[\sum_{m=1}^{2^n-1}(2^n - 2m)n_{2^nC-m} + \sum_{m=0}^{2^n-1}(2^n - m)n_{2^nC+m}\right], \qquad (4)$$

where P_C is the probability that a randomly selected node has degree C, and $1/P$ is the normalization factor that also appears in Eq.(2).

When $k = a$ $(C + 1 \le a \le 2C - 1)$, we can follow the same procedure. When $k = 2^na$, the node can be separated into 2^n parts all with a links; when $k = 2^na \pm 1$, the node can be separated into 2^n sub-nodes, of which $2^n - 1$ have a links, and the remaining one has $a \pm 1$ links; finally, when $k = 2^nx \pm (2^n - 1)$, the node can be separated into 2^n parts, of which only one sub-node has a links, and all of the remaining ones have $a \pm 1$ links. When $2^nx + 2^n \le k \le 2^{n+1}x - 2^{n+1}$, the nodes n_k have no contribution to P_a. Based on the above analysis, we get the general formula

$$P_k = \frac{1}{P}\sum_{n=0}^{\infty}\sum_{m=1-2^n}^{2^n-1}(2^n - |m|)n_{2^nk-m}, \qquad (5)$$

where $2^nx + 2^n \le k \le 2^{n+1}x - 2^{n+1}$, P_k is the probability that a randomly selected node has degree k and $1/P$ is the normalization factor.

To illustrate the above analytical result, in Fig.1 we show the degree distribution P_k, with $\mu = 2$ and $2C = 30$. Besides the large-scale power-law decay, there is a clearly identifiable peak at $k = C + 1 = 16$. This indicates that the number of nodes with medium degree (approximately half of the limit connectivity) is increased, which is compensated by the loss of the high degree nodes.

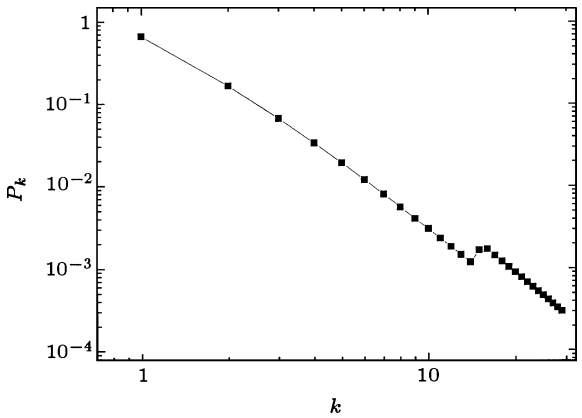


Fig.1. Degree distribution of Model I. The probability that a randomly selected node has degree k, P_k, is plotted as function of k, with $2C = 30$ and $\mu = 2$.

3.2. Degree distribution of Model II

In Model II, the mirror node Model, we define the upper limit number as M. When the degree of one node exceeds M, a first-order mirror node is created. Then the original node and the mirror node have the same chance to link to a newly introduced node. When the mirror node has M links (at that time, on average the original node has $2M$ links), a 'mirror's mirror', or a second-order mirror, node will be created. Thus, when the original node has qM links, the qth-order mirror nodes will be created. On average, the original node has M more links than its first-order mirror nodes, $2M$ more links than its second-order mirror nodes, and qM more links than its qth-order mirror nodes.

If we consider the original node with its mirror nodes as a whole, Model II can be also viewed as the GN model. We define A_k as the probability that a newly introduced node links to an existing node (containing the original node and all its mirror nodes) that already has k links. If we assume $A_k \propto k^\gamma$, the degree distribution $n_k = \dfrac{\mu}{A_k}\prod_{i=1}^{k}\left(1+\dfrac{\mu}{A_k}\right)^{-1}$ in the GN model is still applicable.

First, we treat the nodes with the number of links smaller than M. We can not tell whether a node with k ($k < M$) links is an original node, or a mirror node. As is mentioned above, we consider the original node and all its mirror nodes as a whole when making new connections. So when the node is an original node, the probability that it has degree k, $p_{k,0}$ is proportional to n_k; when it is a first-order mirror node, approximately we have

$$p_{k,1} \propto n_{M+2k} + \frac{1}{2}n_{M+2k+1} + \frac{1}{2}n_{M+2k-1};$$

in general, when it is a qth-order mirror node, we can conclude that

$$p_{k,q} = \sum_{i=0}^{q-i}\frac{q-i}{q}\left(n_{\frac{q(q+1)}{2}M+qk+i} + n_{\frac{q(q+1)}{2}M+qk-i}\right). \quad (6)$$

By taking the sum of all these expressions, we get the degree distribution

$$P_k = \frac{1}{P}\left[n_k + \sum_{q=1}^{\infty}\sum_{i=0}^{q-1}\frac{q-i}{q} \times \left(n_{\frac{q(q+1)}{2}M+qk+i} + n_{\frac{q(q+1)}{2}M+qk-i}\right)\right], \quad (7)$$

where P_k is the probability that a randomly selected node has degree k ($k < M$) and $1/P$ is a normalization factor.

When the number of links is $jM + k$ ($j \neq 0$ and $k < M$), we can similarly obtain

$$P_{jM+k} = \frac{1}{P}\sum_{q=j}^{\infty}\sum_{i=0}^{q-1}\frac{q-i}{q} \times \left(n_{\frac{q(q+1)}{2}M+qk+i} + n_{\frac{q(q+1)}{2}M+qk-i}\right), \quad (8)$$

where P_{jM+k} is the probability that a randomly picked node has degree $jM + k$, and $1/P$ is the normalization factor. For brevity, we do not give the details of the calculation. Equations (7) and (8) are illustrated in Fig.2 in comparison with the degree distribution of the GN model, with $M = 30$ and $\mu = 2$. We can see that, while the power-law distribution is globally retained, there is a clearly identifiable peak at $k = M$.

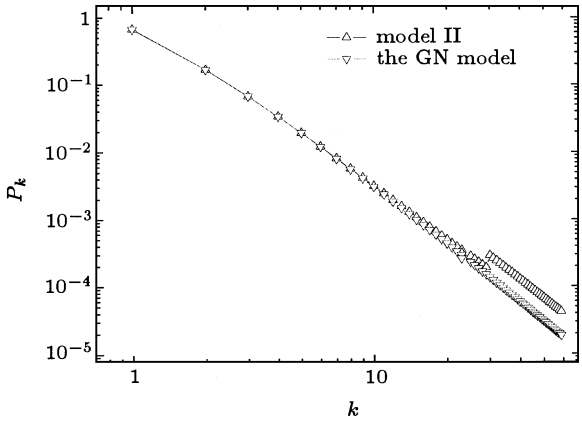


Fig.2. The degree distribution, P_k, of Model II is plotted as a function of k (up triangles), in comparison with that of the GN model (down triangles), with $M = 30$ and $\mu = 2$.

4. Summary and discussions

As mentioned in the Introduction, due to the problems caused by the increasingly heavy burden on the Internet and many other network systems, there are new methods and techniques that call for corresponding modelling works. In this article, these new features are characterized in two types of models, which are built by modifying the growing random network model.[5] The basic ideas of the two models are similar: i.e., when a node reaches the limit of connectivity, its burden is to some degree transferred to some other nodes. The difference lies in, when these other nodes are introduced, whether or not the existing links of the original node are shared. Of these two models,

one common feature is that, while the number of nodes with high connectivity is decreased, there are more nodes with medium links (approximately half of the connectivity limit), and there appears a clearly identifiable peak in the previous monotonically decreasing degree distribution.

In recent years, various complex network systems, especially the Internet, have been under both theoretical and empirical scrutiny, and the studies are fruitful. Yet, besides the large-scale degree distribution, there are still many aspects that we cannot explain and many new methods and techniques require theoretical description. We hope the burden-reducing features that are studied in this article may provide a useful reference to the technology developers and contribute to the general progress in the understanding of the Internet and other complex networks.

References

[1] Albert R and Barabasi A L 2002 *Rev. Mod. Phys.* **74** 47

[2] Dorogovtsev S N and Mendes J F F 2002 *Adv. Phys.* **51** 1079

[3] Zhu H, Wang X R and Zhu J Y 2003 *Physics* **32** 224 (in Chinese)

[4] Barabasi A L and Albert R 1999 *Science* **286** 509

[5] Krapivsky P L and Redner S 2001 *Phys. Rev.* E **63** 066123

[6] Pastor-Satorras R, Vazquez A and Vespignani A 2001 *Phys. Rev. Lett.* **87** 258701

[7] Cohen R, Erez K, ben-Avraham D and Havlin S 2001 *Phys. Rev. Lett.* **86** 3682

[8] Park J and Newman M E J 2003 *Phys. Rev.* E **68** 026112

[9] Vazquez A, Pastor-Satorras R and Vespignani A 2002 *Phys. Rev.* E **65** 066130

[10] Ebel H, Mielsch L I and Bornholdt S 2002 *Phys. Rev.* E **66** 035103