ELSEVIER

Physics Letters A 303 (2002) 337–344

PHYSICS LETTERS A

www.elsevier.com/locate/pla

Connectivity distribution and attack tolerance of general networks with both preferential and random attachments

Zonghua Liu [a,*], Ying-Cheng Lai [a,b], Nong Ye [c], Partha Dasgupta [d]

[a] *Department of Mathematics, Center for Systems Science and Engineering Research, Arizona State University, Tempe, AZ 85287, USA*
[b] *Department of Electrical Engineering and Physics, Arizona State University, Tempe, AZ 85287, USA*
[c] *Department of Industrial Engineering, Arizona State University, Tempe, AZ 85287, USA*
[d] *Department of Computer Science and Engineering, Arizona State University, Tempe, AZ 85287, USA*

Received 17 August 2002; received in revised form 17 August 2002; accepted 23 September 2002

Communicated by A.R. Bishop

Abstract

A general class of growing networks is constructed with both preferential and random attachments, which includes random and scale-free networks as limiting cases when a physical parameter is tuned. Formulas are derived characterizing the evolution and distribution of the connectivity, which are verified by numerical computations. Study of the effect of random failures and intentional attacks on the performance of network suggests that general networks which are neither completely random nor scale-free are desirable.

PACS: 89.75.Hc; 84.35.+i; 02.50.Cw; 05.40.-a

Since the ground-breaking papers by Barabási and Albert on scale-free networks [1] and by Watts and Strogatz on small-world networks [2], the interest on large, growing, and complex networks has soared [3]. Consider a network consisting of N (large) nodes, each having a number K of links to other nodes in the network. Since the network is large, K can be regarded as a random variable with a probability distribution $P(k)$. A realistic network is necessarily sparse, that is: $\langle K\rangle \ll N$, where $\langle K\rangle$ is the expectation of K. Barabási and Albert discovered [1,3] that many networks in nature appear to exhibit the scale-free feature in that the distribution $P(k)$ exhibits a power-law behavior in a range of k-values over several orders of magnitudes: $P(k) \sim k^{-\gamma}$. The mechanisms leading to the power-law distribution are argued to be growth and preferential attachment [1,4], where the former means that the size of the network keeps increasing with time and the latter underlies that the relative probability for an already heavily connected node to get new links is proportionally large. Growth and preferential attachments appear to be the fundamental organizing principle of the many complex networks. The small-world concept, on the other hand, describes the fact that the average path between any two nodes in a large network can be relatively short. While this concept has

[*] Corresponding author.
E-mail address: zhliu@chaos4.la.asu.edu (Z. Liu).

0375-9601/02/$ – see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0375-9601(02)01317-8

been known in social science in manifestations such as the “six degrees of separation” [5], Watts and Strogatz found that many realistic networks exhibit the small-world feature [2]. The small-world concept appears to be universal for large, sparse networks, regardless of whether they have an underlying organizing structure. In fact, the pioneering study on random graphs by Erdős and Rényi already indicated that the typical distance between any two nodes scales logarithmically with the number of nodes [6] and, many apparently scale-free networks are small-world, too [3].

A scale-free network, by its definition, permits a high degree of organization such as the existence of a set of nodes with great numbers of links. The scale-free situation is, however, idealized, for which the distribution $P(k)$ is strictly power-law. Indeed, the illuminating scale-free model proposed by Barabási et al. [1,4] predicts a universal power-law scaling behavior with exponent $\gamma = 3$. A random network, on the other hand, is characterized by the lack of any apparent structures and by an exponential distribution of $P(k)$. Complex networks in nature should fall somewhere in between these two extremes. Although numerical fittings of the distributions $P(k)$ for many realistic networks suggest a power-law behavior, there are also examples where the distribution is neither power-law nor exponential, such as the scientific collaboration network [7]. While the basic dynamical mechanisms for the scale-free property are relatively clear (i.e., growth and preferential attachment) [4] and there are many recent papers proposing various models for scale-free networks [8–19], the question of how a general network develops distribution $P(k)$ with a mixture of power-law and exponential behaviors remains interesting. It is desirable to know what *microscopic* dynamical mechanisms can produce such a mixed behavior in large, growing networks and, more importantly, whether there exists a parameter with a clear physical meaning which *controls* the relative weights between the power-law and exponential behaviors.

The aim of this Letter is to present a model to explain the general scaling behavior, i.e., mixture of power-law and exponential distributions, in a natural manner. Specifically, we identify a physical parameter that balances the power-law and exponential contributions to $P(k)$. To be as realistic as possible, we also consider the effect of temporal fluctuations in the number of links a new node can have [20]. In addition, we address the practically important issues of robustness and security, that is, how the performance of our general network is affected by random failures and intentional attacks.

The starting point of most existing models on growing networks is the quantity Π_i, the probability that a new node to be connected to node i already in the network. Starting with a small number (m_0) of nodes, a new node with m links (edges) is added at each time step with links determined by the probability Π_i. The original model by Barabási et al. [1,4] assumes a linear dependence of Π_i on k_i, the number of already existing links for node i. Generalization to algebraic dependence in the form of $\Pi_i \sim k_i^\alpha$ has been considered by Krapivsky et al. [8] which, for $\alpha = 1$, reduces to the model by Barabási et al. While such pre-determined dependence of Π_i on k_i is necessary for the network to exhibit the scale-free feature, there can also be random factors affecting this dependence. In particular, as new nodes are added to the network, although there is a tendency for an already heavily linked node to acquire more connections, the new connections can also be random. For instance, in the internet, convenience is a key factor determining how new nodes are added, besides popularity of certain hosts. Our idea is thus that, generally, a realistic network grows in time according to an attachment rule that is neither completely preferential nor completely random. In terms of the quantity Π_i, it should contain both a deterministic component reflecting preferential attachment, and a random component as well. In particular, we assume

$$\Pi_i = \frac{(1-p)k_i + p}{\sum_j [(1-p)k_j + p]}, \tag{1}$$

where $0 \leqslant p \leqslant 1$ is a parameter characterizing the relative weights between the deterministic and random contributions to Π_i, and the summation is over the whole network at a given time. Clearly, p is the probability that a new node is randomly connected to the existing node i and $(1-p)$ is the probability that the new node is preferentially attached to i. The model reduces to that by Barabási et al. for $p = 0$ and it becomes a completely random network for $p = 1$. It can thus generate any network structure from scale-free to random. One result of this Letter is that the model (1) yields, for $0 \leqslant p \leqslant 1$, a scaling that generally lies between the two extremes of power-law and exponential distributions, as follows,

$$P(k) \sim \left(\frac{\frac{k}{m}+b}{1+b}\right)^{-\gamma}, \tag{2}$$

where the scaling exponent γ is

$$\gamma = 3 + b, \quad \text{where } b = \frac{p}{m(1-p)}. \tag{3}$$

Clearly, the power-law scaling for scale-free networks is recovered for $p = 0$ and the distribution becomes exponential $P(k) \sim e^{-k/m}$ for $p \to 1$.

To derive the scaling law for $P(k)$, it is necessary to obtain the temporal evolution of the connectivity of a given node. We use the mean-field approach [4]. Under the approximation that k_i is continuous, the probability Π_i is in fact the continuous rate of change of k_i,

$$\frac{\partial k_i}{\partial t} = m\Pi_i(k_i) = \frac{m[(1-p)k_i + p]}{\sum_j [(1-p)k_j + p]}. \tag{4}$$

Because of the growing nature of the network, the summation in Eq. (4) increases with time t:

$$\sum_j \big[(1-p)k_j + p\big] = 2m(1-p)t + pt.$$

This, together with the initial condition $k_i(t_i) = m$, yields the solution to Eq. (4),

$$k_i(t) = \left(m + \frac{p}{1-p}\right)\left(\frac{t}{t_i}\right)^{\frac{m(1-p)}{2m+(1-2m)p}} - \frac{p}{1-p} \sim t^{\beta} \quad \text{for } t \text{ large and } 0 \leqslant p < 1, \tag{5}$$

where

$$\beta = m(1-p)/\big[2m + (1-2m)p\big].$$

Note that $\beta = 0.5$ for $p = 0$, $\beta = m/(2m+1)$ for $p = 0.5$, and for $p \to 1$, Eq. (5) gives $k_i(t) \sim m \ln t$. These are consistent with the results in Ref. [4]. Using Eq. (5), we can write down the probability that a node has a connectivity $k_i(t)$ smaller than k, $\Phi\{k_i(t) < k\}$, as follows,

$$\Phi\big\{k_i(t) < k\big\} = \Phi\left\{t_i > t\left[\frac{m+(1-m)p}{k+(1-k)p}\right]^{2+\frac{p}{m(1-p)}}\right\}. \tag{6}$$

Suppose nodes are added at equal time intervals to the system, the probability density of t_i is then $\phi_i(t_i) = 1/(m_0 + t)$. Substituting this into Eq. (6), we obtain

$$\begin{aligned} &\Phi\left\{t_i > t\left[\frac{m+(1-m)p}{k+(1-k)p}\right]^{2+\frac{p}{m(1-p)}}\right\} \\ &\quad = 1 - \Phi\left\{t_i \leqslant t\left[\frac{m+(1-m)p}{k+(1-k)p}\right]^{2+\frac{p}{m(1-p)}}\right\} \\ &\quad \sim 1 - \frac{t}{m_0+t}\left[\frac{m+(1-m)p}{k+(1-k)p}\right]^{2+\frac{p}{m(1-p)}}. \end{aligned} \tag{7}$$

The probability distribution $P(k)$ can then be obtained by the partial derivative $P(k) = \partial\Phi\{k_i(t) < k\}/\partial k$, which gives the general scaling law (2).

The scaling results (2) and (3) are obtained under the assumption that m, the number of links added to the network at each time step, is a constant. In a realistic situation, m can fluctuate with time. To model this, we choose a constant $\overline{m}$ and assume that $m(t)$ can vary in the range $[1, 2\overline{m} - 1]$. Specifically, we write $m(t) = \overline{m}[1 + \xi(t)]$, where $\xi(t)$ is a discrete random variable uniformly distributed in the range $[(1-\overline{m})/\overline{m}, (\overline{m}-1)/\overline{m}]$ with zero average. A similar derivation yields

$$\begin{aligned} &k_i(t) \sim t^{\beta(t)}, \\ &P(k) \sim \left(\frac{\frac{k}{m}+b}{1+b}\right)^{-\gamma(t)}, \end{aligned} \tag{8}$$

where the scaling exponents $\beta(t)$ and $\gamma(t)$ now have an explicit dependence on time and they are given by:

$$\beta(t) = \big[f(t)(1-p)\big]\big/\big[2\overline{m} + (1-2\overline{m})p\big]$$

and $\gamma(t) = 1 + 1/\beta(t)$ where $f(t)$ is a random function defined by:

$$f(t)\ln(t) = \int_{t_i}^{t} \frac{\overline{m}[1+\xi(t)]}{t}\, dt.$$

As $t \to \infty$, $f(t) \to \overline{m}$. We see that with the number of new links fluctuating, $P(k)$ has the same scaling as that in the case of no fluctuation but the scaling exponent changes with time. On average, the scaling exponent is the same as that for the case where $m = \overline{m}$.

We now present numerical support for the scaling results (2), (3), (5), and (8). We start with $m_0 = 3$ nodes. At each time step, a new node with either $m = \overline{m} = 3$ [case (a)] or $m(t) = \overline{m}[1 + \xi(t)]$ [case (b)] links is added to the network. The number of nodes at time step t is then $N(t) = m_0 + t \sim t$. Fig. 1(a) and (b)

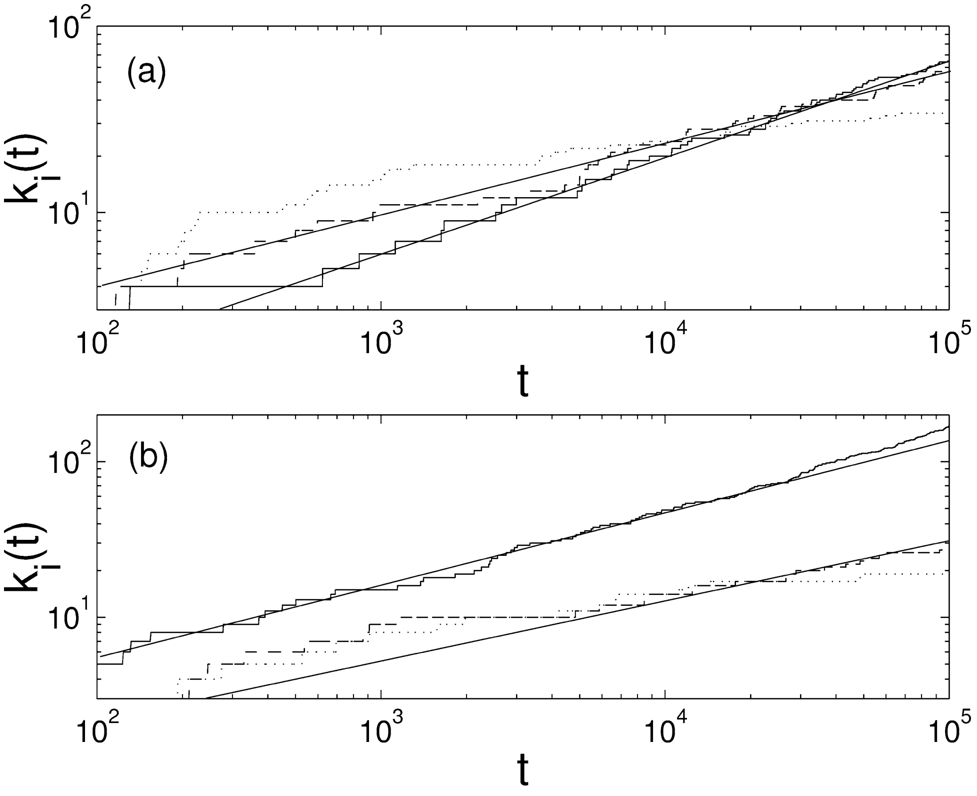

Fig. 1. Evolution of the connectivity $k_i(t)$ of a typical node in the network. The node is added to the system at $t = 95$. Cases (a) and (b) are for $\overline{m} = \text{constant} = 3$ and $m(t) = \overline{m}[1 + \xi(t)]$, respectively, where $\xi(t)$ is a discrete random variable. The solid, dashed, and dotted curves are for $p = 0$, $p = 0.5$, and $p = 1.0$, respectively.

show the scaling of $k_i(t)$ with t on a logarithmic scale for cases (a) and (b), respectively, where the solid, dashed, and dotted curves are for $p = 0$, $p = 0.5$, and $p = 1.0$, respectively. For $p = 0$ and $p = 0.5$ in case (a), there is a robust power-law scaling with the exponents $\beta \approx 0.5$ and $\beta \approx 0.42$, which agree very well with the predicted slopes

$$\beta = m(1 - p)/\bigl[2m + (1 - 2m)p\bigr].$$

When $m(t)$ fluctuates, as in case (b), there still appears to be a power-law scaling behavior for $p = 0$ and $p = 0.5$ with slopes similar to those in the case where m is constant, for large t. In fact, for p not too close to 1 in case (b), the power-law scaling behavior is always observed, but the scaling exponent exhibits small fluctuations about that for the constant m case, as predicted. For $p = 1$ in both cases (a) and (b), the scaling appears to be $K_i(t) \sim \ln t$, as predicted.

Fig. 2(a) and (b) show, on a logarithmic scale, the scaling behavior of the connectivity distribution $P(k)$ for $m = \text{constant}$ and $m(t) = \overline{m}[1+\xi(t)]$, respectively, where the open circles, stars, and squares denote $p = 0$, $p = 0.5$, and $p = 1$, respectively. The scaling is clearly power-law for $p = 0$, with the predicted slope $\gamma = 3$. For $p = 1$, the scaling is exponential, as can be seen by the corresponding curves on a semi-logarithmic scale in the insets. For $0 < p < 1$, the scaling lies somewhere in between the power-law and exponential behaviors. Comparing (a) with (b) suggests that fluctuations in $m(t)$ cause a plateau for small k region in the scaling of $P(k)$. These results thus indicate that our model can generate realistically observable scaling behaviors ranging from purely power-law to purely exponential, with the variation of a single control parameter p.

We now turn to addressing the effect of random failure and intentional attack on general networks, which

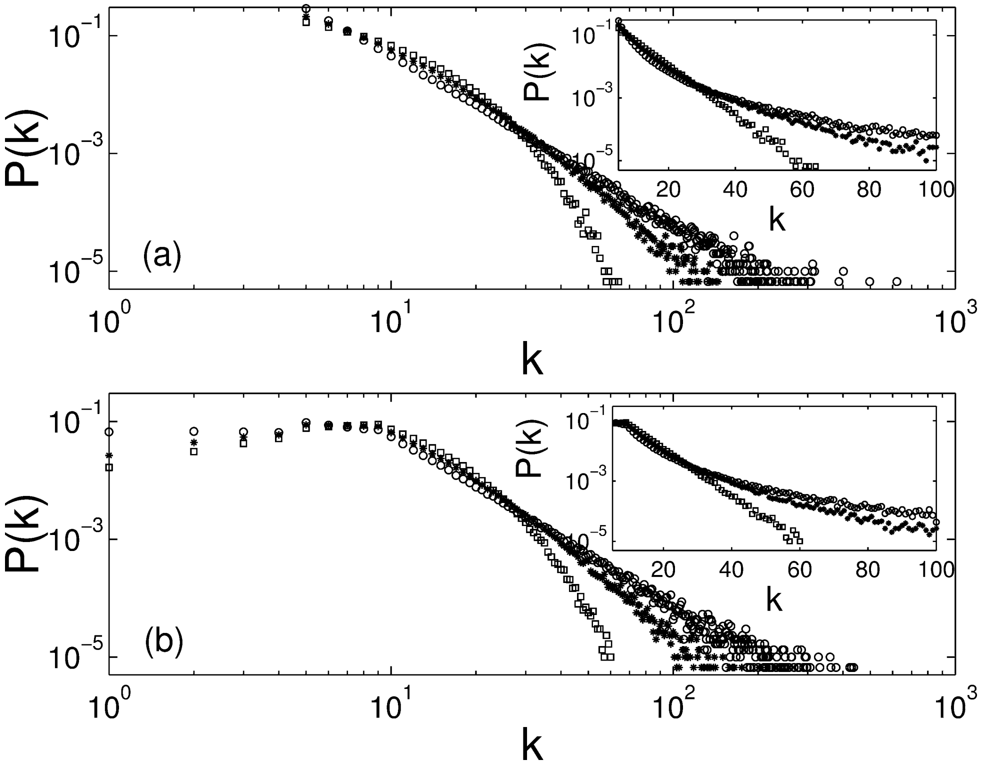

Fig. 2. The connectivity distribution $P(k)$ for (a) $m =$ constant and (b) $m(t) = \overline{m}[1 + \xi(t)]$, where the open circles, stars, and squares denote $p = 0$, $p = 0.5$, and $p = 1$, respectively. For $p = 0$, the scaling is clearly power-law, with deviations from it as p is increased from zero. The scaling for $p = 1$ is exponential, as shown in the insets. Fluctuations in $m(t)$ cause a plateau for small k region in the scaling of $P(k)$ [case (b)].

means, respectively, random removal of fractions of nodes and targeted destruction of certain nodes, most likely those heavily connected ones, in the network. The recent study by Albert et al. [21] indicates that growing networks with exponentially distributed connectivity are robust against both intentional attacks and random removals of a relatively small fraction of nodes. Scale-free networks, on the other hand, appear to be robust against random failures but are more sensitive to intentional attacks. These results are intuitively understandable as, for instance, there exists a few heavily connected nodes in a scale-free network. Attack on even a few of these nodes would immediately cripple the network. Albert et al. suggested [21] using the concept of *diameter* [22] to quantify the performance of the network under random failures or attacks. Roughly speaking, the diameter of a network is the average number of links between any two nodes in the network, which is the same concept as the average shortest path in the small-world characterization of networks [2]. Computation of the diameter requires searching through all pairs of nodes in the network, which is numerically intensive when the size of the network is large. We have thus developed the following simple method to compute the diameter D of a large network, which is numerically efficient. Specifically, at a given time, we randomly choose n nodes from the network such that increasing n does not result in appreciable variations in D. Each node is regarded as a “center” and the number 1 is assigned to nodes that are directly connected to it. These are the first nearest neighbors. Nodes that require two links to reach the center are assigned the number 2 and they are the second nearest neighbors, and so on. This process continues until every node that is linked to the center, directly or indirectly, is assigned a number. The average value of all these numbers gives the distance d_1 from the center node to an arbitrary node. Choosing

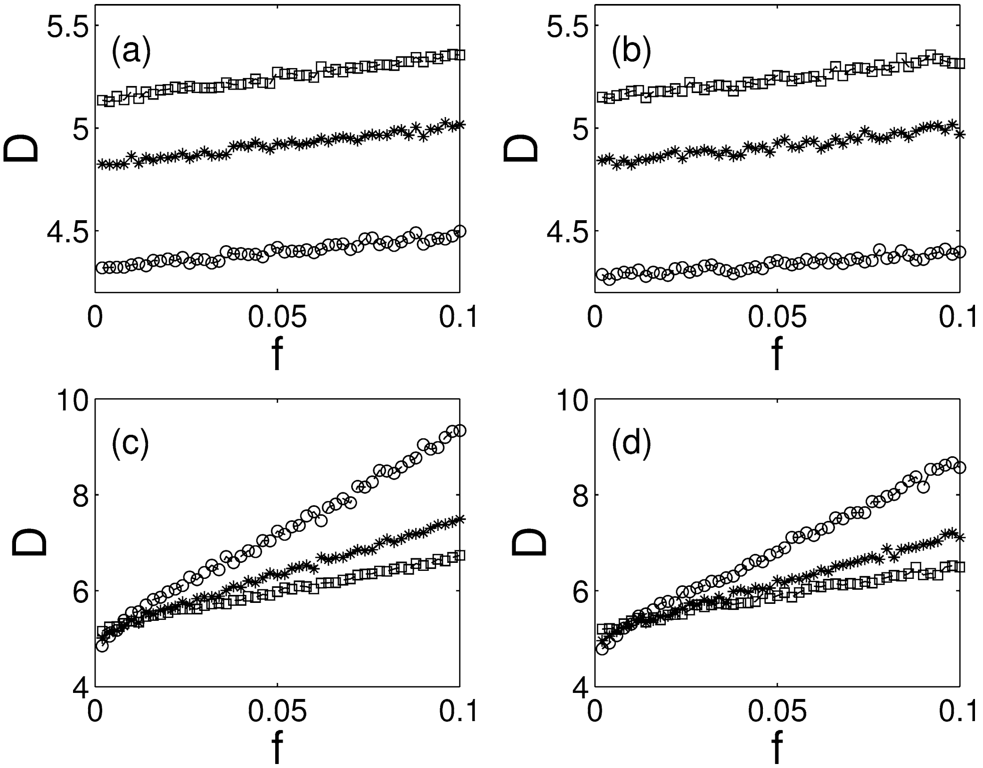

Fig. 3. Changes in the diameter D as a function of the fraction f of removed nodes: (a) random failure with constant m, (b) random failure with fluctuating $m(t)$, (c) intentional attack with constant m, and (d) intentional attack with fluctuating $m(t)$. The open circles, stars, and open squares denote the $p = 0$, $p = 0.5$, and $p = 1$ cases, respectively. The general network with $p = 0.5$ appears to be a good trade-off between scale-free and random networks against both random failures and intentional attacks.

the center node in turn yields an additional $(n - 1)$ distances d_i $(i = 2, \dots, n)$. The diameter D is taken to be the average value of the n distances. We find numerically that the method yields essentially the same result when different schemes for neighbor assignments are used, indicating that the computed value of the diameter reflects correctly the situation which it aims to describe, i.e., the average path length between two arbitrary nodes in the network.

We find, for our general network with initial conditions $m_0 = 3$ at large time, the diameter D is small and does not change appreciably for $0 \leqslant p \leqslant 1$. For instance, for $t = 10^5$ (so there are about 10^5 nodes in the network at this time), the scale-free network $(p = 0)$ has $D \approx 4.3$ while the completely random network $(p = 1)$ has $D \approx 5.1$, both are negligible comparing with the number of nodes in the network, suggesting that all networks resulted from our model are small-world and therefore are efficient in terms of communication or propagation of information within the network, regardless of whether there is an organized structure (the scale-free case) or there is a total lack of such a structure (the random case). Fig. 3(a) and (b) show the diameter of the network as a function of f, the fraction of randomly removed nodes, for cases of constant and fluctuating m, respectively, where the lower trace (open circles) is for the scale-free, the upper (squares) for random, and the middle (stars) for general networks. Randomly removing less than 10% of the nodes results in only an incremental increase in the diameter, suggesting that all networks are robust against random failures of a small fraction of nodes. The situation changes with intentional attacks, as shown in Fig. 3(c) and (d) for cases of constant and fluctuating m, respectively. We see that for both cases, the diameters of the scale-free networks (upper traces, open circles) increase more rapidly with f, comparing with the more general (middle traces, stars) and random (lower traces, squares) networks. The interesting observation is that with respect to the diameter measure, a general network (say with $p = 0.5$) is much more robust against intentional attacks as compared with a scale-free one, while at the same time, performs better under random failures as compared with a random network. Thus, the parameter p can be tuned to generate networks with any desired trade-off between the scale-free and random properties in consideration of security versus robustness against random failures.

We noticed that Ref. [9] gave a model which is mathematically equivalent to our model (1) but with different physical meaning. In Ref. [9], the following attachment rule is considered. At each time step, m new links are added to the network and, the probability that a new link is attached to node i is proportional to $A + q_i$, where $A \geqslant 0$ is a predefined constant and q_i is the number of incoming links to node i. If A is chosen to be m, then the quantity $A + q_i$ becomes k_i, the total number of links of node i. Our attachment rule Eq. (1), however, is fundamentally different from this rule. Firstly, Eq. (1) contains both a deterministic and a random components, the latter simulates realistic physical processes such as random rewiring, removal, and addition of new links. This random component is clearly important, but Ref. [9] ignores this completely. Secondly, the attachment probability in Ref. [9] is special in the sense that it is determined by the number of *incoming* links only, while our rule is more general because it involves both incoming and outgoing links. In many realistic networks such as those arising in epidemiology and biology, both incoming and outgoing links are important. Although in form, the attachment rule in Ref. [9] appears to be equivalent to our rule by the simple conversion $A = p/(1 - p)$, the meanings of these two rules and hence the consequences are qualitatively different, as explained. As we have demonstrated, because of the complication to include a random component, a mean-field treatment appears to be feasible and proper, yielding results that can be verified through numerical experiments.

We stress the following three differences between our work and Ref. [9]. Firstly, our main theoretical prediction is that for realistic networks, the connectivity distribution is typically a mixture of algebraic and exponential components, which has indeed been observed, while Ref. [9] assumes algebraic connectivity distribution and focuses on the variation of the algebraic scaling exponent. In this sense, the model and result in Ref. [9] are somewhat special. Secondly, we analyze the effect of fluctuating links, which is also an ingredient in many realistic networks. Thirdly, we consider the practically important issues of network security against random failures or intentional attacks. Despite the inclusion of these practical factors, we are still able to predict the connectivity distribution through a mean-field approach. The treatment in Ref. [9] is rigorous and elegant, but the situation considered there is special.

In conclusion, we have constructed a class of general growing networks based on the intuitive but realistic consideration that nodes are added to the network with both preferential and random attachments, taking into consideration that the number of new nodes can fluctuate with time. We have derived theoretical formulas for both the temporal evolution and distribution of the connectivity, and these are verified by extensive numerical computations. The effects of random failure and intentional attacks on the performance of the network are also addressed. From the engineering pointview of designing large, complex, growing networks, our recommendation is that both scale-free and random architectures should be avoided. Instead, one should consider general networks with approximately equal amounts of preferential and random factors.

Acknowledgements

This work was supported by AFOSR under Grant No. F49620-01-1-0317.

References

[1] A.-L. Barabási, R. Albert, Science 286 (1999) 509.
[2] D.J. Watts, S.H. Strogatz, Nature 393 (1998) 440.
[3] R. Albert, A.-L. Barabási, Rev. Mod. Phys. 74 (2002) 47.
[4] A.-L. Barabási, R. Albert, H. Jeong, Physica A 272 (1999) 173;
A.-L. Barabási, R. Albert, H. Jeong, Physica A 281 (2000) 69.
[5] S. Milgram, Psychol. Today 1 (1967) 60.
[6] P. Erdös, A. Rényi, Publ. Math. Inst. Hung. Acad. Sci. 5 (1960) 17;
B. Bollobaás, Random Graphs, Academic, London, 1985.
[7] M.E.J. Newman, Phys. Rev. E 64 (2001) 016132.
[8] P.L. Krapivsky, S. Redner, F. Leyvraz, Phys. Rev. Lett. 85 (2000) 4629;
Z. Liu, Y.-C. Lai, N. Ye, Phys. Rev. E 66 (2002) 036112.

[9] S.N. Dorogovtsev, J.F.F. Mendes, A.N. Samukhin, Phys. Rev. Lett. 85 (2000) 4633;
S.N. Dorogovtsev, J.F.F. Mendes, A.N. Samukhin, cond-mat/0009090;
S.N. Dorogovtsev, J.F.F. Mendes, A.N. Samukhin, cond-mat/0011077.

[10] R. Albert, A.-L. Barabási, Phys. Rev. Lett. 85 (2000) 5234.

[11] Some flavors of our general network model have also appeared in the model by Kumar et al., to explain the power-law scaling behavior of the World Wide Web. In their model, as a new node is added to the network, a prototype node is chosen randomly from the already existing nodes. With probability p, the ith edge of the new node is connected randomly to some node in the network, and with probability $(1-p)$ the edge is connected to the prototype node. A strictly power-law scaling behavior is obtained for the number of incoming links this way, with the exponent given by $(2-p)/(1-p)$. See, R. Kumar, P. Raghavan, S. Rajalopagan, D. Sivakumar, A.S. Tomkins, E. Upfal, in: Proceedings of the 19th Symposium on Principles of Database Systems, p. 1; Proceedings of the 41st IEEE Symposium on Foundations of Computer Science (IEEE Computing Soc.), p. 57.

[12] A. Vázquez, cond-mat/006132;
A. Vázquez, cond-mat/0105031.

[13] S.N. Dorogovtsev, J.F.F. Mendes, Phys. Rev. E 63 (2001) 025101;
S.N. Dorogovtsev, J.F.F. Mendes, Phys. Rev. E 63 (2001) 056125.

[14] A.-L. Barabási, H. Jeong, E. Ravasz, Z. Néda, A. Schubert, T. Vicsek, cond-mat/0104162.

[15] G. Bianconi, A.-L. Barabási, Europhys. Lett. 54 (2001) 436;
G. Bianconi, A.-L. Barabási, Phys. Rev. Lett. 86 (2001) 5632.

[16] G. Ergün, G.J. Rodgers, cond-mat/0103423.

[17] P.L. Krapivsky, S. Redner, Phys. Rev. E 63 (2001) 066123.

[18] P.L. Krapivsky, G.J. Rodgers, S. Redner, Phys. Rev. Lett. 86 (2001) 5401.

[19] B. Tadić, Physica A 293 (2001) 273;
B. Tadić, cond-mat/0104029.

[20] B.A. Huberman, L.A. Adamic, Nature 401 (1999) 131;
K.-I. Goh, B. Kahng, D. Kim, Phys. Rev. Lett. 88 (2002) 108701.

[21] R. Albert, H. Jeong, A.-L. Barabási, Nature 406 (2000) 378.

[22] R. Albert, H. Jeong, A.-L. Barabási, Nature 401 (1999) 130.