Topology of Evolving Networks: Local Events and Universality

Réka Albert and Albert-László Barabási*

Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556
(Received 7 December 1999)

Networks grow and evolve by local events, such as the addition of new nodes and links, or rewiring of links from one node to another. We show that depending on the frequency of these processes two topologically different networks can emerge, the connectivity distribution following either a generalized power law or an exponential. We propose a continuum theory that predicts these two regimes as well as the scaling function and the exponents, in good agreement with numerical results. Finally, we use the obtained predictions to fit the connectivity distribution of the network describing the professional links between movie actors.

The complexity of numerous social, biological, or communication systems is rooted in the rather interwoven web defined by the system's components and their interactions. For example, cell functioning is guaranteed by a complex metabolic network, whose nodes are substrates and enzymes, and edges represent chemical interactions [1]. Similarly, the society is characterized by a huge social network whose nodes are individuals or organizations, connected by social interactions [2], but equally complex networks appear in the business world, where nodes are companies and edges represent diverse trade relationships, or describe the world-wide web (www), whose nodes are html documents connected by links pointing from one page to another [3–5].

The study of random networks has been dominated by the model of Erdős and Rényi (ER) [6] that views the network as a set of nodes which are connected pairwise with equal probability. Recently Watts and Strogatz (WS) [7] found that local clustering is an important characteristic of networks, offering the first indication that real networks could be more complex than predicted by the ER model. A common feature of the ER and WS models is that the probability $P(k)$ that a node in the network is connected to k other nodes is bounded, decaying exponentially for large k. In contrast, measurements on the www connectivity [3–5], actor networks [7,8], citation network of scientists [9], and recent results on the Internet topology [10] found that, independently of the nature of the system and the identity of its constituents, $P(k)$ decays as a power law, following $P(k) \sim k^{-\gamma}$. These results offered the first evidence that some large networks can self-organize into a scale-free state, a feature unexpected by previous network models [6,7]. The origin of this scale-free behavior has been traced back to two mechanisms that are present in many systems, and have a strong impact on the final topology [8]. First, networks develop by the addition of new nodes that are connected to those already present in the system. Second, there is a higher probability that a new node is linked to a node that already has a large number of connections. These two ingredients led to the formulation of the scale-free model that generates a network for which

$P(k)$ follows a power law with $\gamma = 3$. While this model correctly predicts the emergence of power-law scaling, the agreement between the measured and predicted exponents is less than satisfactory: For real systems, γ is scattered between 2.1 and 3 [8], raising important questions about the universality of network formation.

In this paper we introduce an extended model of network evolution that gives a more realistic description of the local processes, incorporating the addition of new nodes, new links, and the rewiring of links. We show that, depending on the relative frequency of these local processes, networks can develop two fundamentally different topologies. In the first regime $P(k)$ has a power-law tail, but the exponent γ depends continuously on the frequency of the local events. In the second regime $P(k)$ approaches an exponential. We derive a phase diagram that predicts the transition between these two topologically distinct regimes, and we use the obtained prediction to fit the connectivity distribution of the actor network.

Extended model.—We start with m_0 isolated nodes, and at each time step we perform one of the following three operations (see Fig. 1a).

(i) With probability p we add $m(m \leq m_0)$ new links: For this we randomly select a node as the starting point of the new link, describing, for example, that a web developer decides to add a new hyperlink to a page. The other end of the link is selected with probability

$$\Pi(k_i) = \frac{k_i + 1}{\sum_j (k_j + 1)}, \qquad (1)$$

incorporating the fact that new links preferentially point to popular nodes, with a high number of connections [8]. This process is repeated m times.

(ii) With probability q we rewire m links: For this we randomly select a node i and a link l_{ij} connected to it. Next we remove this link and replace it with a new link $l_{ij'}$ that connects i with node j' chosen with probability $\Pi(k_j')$ given by (1). This process is repeated m times.

(iii) With probability $1 - p - q$ we add a new node: The new node has m new links that with probability $\Pi(k_i)$ are connected to nodes i already present in the system.
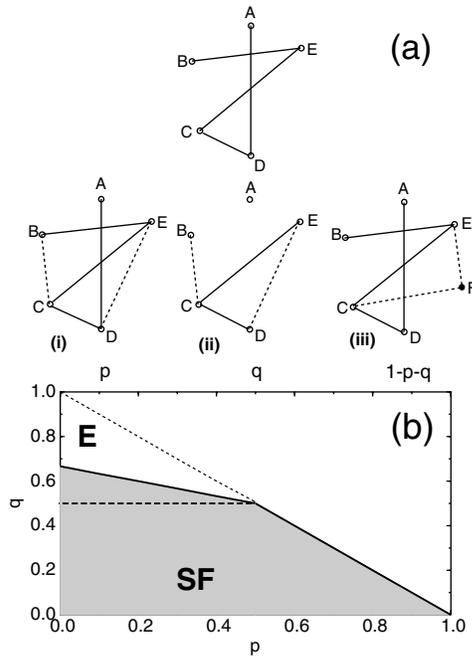
FIG. 1. (a) Illustration of the possible elementary processes in the extended model for $m_0 = 3$ and $m = 2$. At time $t = 2$ there are five nodes in the system connected by four links (upper panel). In the next time step, one of three possible events can take place: (i) Add $m = 2$ links with probability p. One end of the new link is selected randomly, and the other is selected using preferential attachment [Eq. (1)]. The new links are drawn with dashed lines. (ii) Rewire $m = 2$ links with probability $1 - q$. The AD link is disconnected from its A end and connected preferentially to the highly connected node E. (iii) Add a new node (F) and connect it with $m = 2$ links to the nodes in the system with probability $1 - p - q$, using preferential attachment. (b) Phase diagram for the extended model. The scale-free (SF) regime for $m = 1$ is shaded; the remaining $p + q < 1$ domain corresponds to the exponential (E) regime. The boundary between E and SF is shown as a dotted line when $m \to 0$, or as a dashed line when $m \to \infty$.

Since our goal is to investigate the generic mechanisms of network evolution, we use bidirectional links. However, our results can be easily generalized to directed networks as well. In the model, the probabilities p and q can be varied in the interval $0 \le p < 1$ and $0 \le q < 1 - p$. Note that we choose the probability $\Pi(k_i)$ to be proportional to $k_i + 1$, such that there is a nonzero probability that isolated nodes ($k_i = 0$) acquire new links. Finally, in the $p = q = 0$ limit the model reduces to the scale-free model investigated in Ref. [8].

Continuum theory.—In the model the probability that a node i increases its connectivity k_i depends only on k_i and quantities characterizing the whole network (the parameters p, q, m and the number of nodes and links). We assume that k_i changes continuously, and the probability $\Pi(k_i)$ can be interpreted as the rate at which k_i changes [11]. Consequently, the processes (i)–(iii) all contribute to k_i, each being incorporated in the continuum theory as follows.

(i) Addition of m new links with probability p:

$$\left(\frac{\partial k_i}{\partial t}\right)_{(i)} = pA\frac{1}{N} + pA\frac{k_i + 1}{\sum_j (k_j + 1)}, \qquad (2)$$

where N is the size of the system. The first term on the right-hand side corresponds to the random selection of one end of the new link, while the second term reflects the preferential attachment (1) used to select the other end of the link. Since the total change in connectivity after a step is $\Delta k = 2m$, we have $A = m$.

(ii) Rewiring of m links with probability q:

$$\left(\frac{\partial k_i}{\partial t}\right)_{(ii)} = -qB\frac{1}{N} + qB\frac{k_i + 1}{\sum_j (k_j + 1)}. \qquad (3)$$

The first term incorporates the decreasing connectivity of the node from which the link was removed, and the second term represents the increasing connectivity of the node that the link is reconnected to. The total connectivity does not change during the rewiring process, but B can be calculated by separating the two processes, obtaining $B = m$.

(iii) Addition of a new node with probability $1 - p - q$:

$$\left(\frac{\partial k_i}{\partial t}\right)_{(iii)} = (1 - p - q)C\frac{k_i + 1}{\sum_j (k_j + 1)}. \qquad (4)$$

The number of links connecting the new node to the nodes in the system is m, thus we have $C = m$.

By adding the contribution of the three processes, we obtain

$$\frac{\partial k_i}{\partial t} = (p - q)m\frac{1}{N} + m\frac{k_i + 1}{\sum_j (k_j + 1)}. \qquad (5)$$

In (5) the system size N and the total number of links $\sum_j k_j$ vary with time as $N(t) = m_0 + (1 - p - q)t$ and $\sum_j k_j = (1 - q)2mt - m$, indicating that for large t we can neglect the constants m_0 and m compared to the terms linearly increasing with time. Using as an initial condition the connectivity of a node added at time t_i, $k_i(t_i) = m$, the solution of (5) for $k_i(t)$ has the form

$$k_i(t) = [A(p,q,m) + m + 1]\left(\frac{t}{t_i}\right)^{1/B(p,q,m)}$$
$$- A(p,q,m) - 1, \qquad (6)$$

where

$$A(p,q,m) = (p - q)\left(\frac{2m(1 - q)}{1 - p - q} + 1\right),$$
$$B(p,q,m) = \frac{2m(1 - q) + 1 - p - q}{m}. \qquad (7)$$

The probability that a node has a connectivity $k_i(t)$ smaller than k, $P[k_i(t) < k]$, can be written as $P[k_i(t) < k] = P[t_i > C(p,q,m)t]$, where

$$C(p,q,m) = \left(\frac{m + A(p,q,m) + 1}{k + A(p,q,m) + 1}\right)^{B(p,q,m)}. \qquad (8)$$

Since t_i must satisfy the condition $0 \le t_i \le t$, we can distinguish three cases.

(I) If $C(p,q,m) > 1$, then $P[k_i(t) < k] = 0$. Thus the condition that $P(k)$ is nonzero is that $k > m$.

(II) If $C(p,q,m)$ is not real, then $P[k_i(t) < k]$ is not well-defined. Thus to be able to calculate $P(k)$ we must have $\frac{m+A(p,q,m)+1}{k+A(p,q,m)+1} > 0$ for all $k > m$, satisfied if $A(p,q,m) + m + 1 > 0$.

(III) Finally, if $0 < C(p,q,m) < 1$, the connectivity distribution $P(k)$ can be determined analytically. Defining the unit of time in the model as one growth/rewire/new link attempt, the probability density of t_i is $P_i(t_i) = 1/(m_0 + t)$, thus

$$P[k_i(t) < k] = 1 - C(p,q,m)\,\frac{t}{m_0 + t} \qquad (9)$$

from which, using $P(k) = \frac{\partial P[k_i(t)<k]}{\partial k}$, we obtain

$$P(k) = \frac{t}{m_0 + t} D(p,q,m)$$
$$\times [k + A(p,q,m) + 1]^{-1-B(p,q,m)}, \quad (10)$$

where $D(p,q,m) = [m + A(p,q,m) + 1]^{B(p,q,m)} \times B(p,q,m)$.

Thus the connectivity distribution, the main result provided by the continuum theory, has a generalized power-law form

$$P(k) \propto [k + \kappa(p,q,m)]^{-\gamma(p,q,m)}, \qquad (11)$$

where $\kappa(p,q,m) = A(p,q,m) + 1$ and $\gamma(p,q,m) = B(p,q,m) + 1$.

Phase diagram and scaling.—Equations (10) and (11) are valid only when $A(p,q,m) + m + 1 > 0$ [see (I) and (II)], which, for fixed p and m, translates into $q < q_{max} = \min\{1 - p, (1 - p + m)/(1 + 2m)\}$. The (p,q) phase diagram, shown in Fig. 1b, indicates the existence of two regions in the parameter space: For $q < q_{max}$, $P(k)$ is given by (11), thus the connectivity distribution is scale-free. For $q > q_{max}$, however, Eq. (11) is not valid; the continuum theory fails to predict the behavior of the system. We demonstrate that, in this regime, $P(k)$ crosses over to an exponential. The boundary between the scale-free and exponential regimes depends on the parameter m, being a line with slope $-m/(1 + 2m)$ (Fig. 1b).

Scale-free regime.—While a power-law tail is present in any point of this regime, the scaling is different from that predicted by the simpler scale-free model [8]. First, according to (10), for small k the probability saturates at $P(\kappa(p,q,m))$. This feature reproduces the results obtained for real networks: $P(k)$ obtained for the actor network or the citation network of scientists saturates at small k [8]. Second, the exponent $\gamma(p,q,m)$, characterizing the tail of $P(k)$ for $k \gg \kappa(p,q)$, changes continuously with p, q, and m, predicting a range of exponents between 2 and ∞. This allows us to account for the wide variations seen in real networks, for which γ varies from 2 to 3. To

demonstrate the predictive power of the continuum theory, we have studied the extended discrete model numerically, as defined in (i)–(iii). As Fig. 2a shows, Eq. (10) offers an excellent, parameter-free fit to the numerical results.

Exponential regime.—For $q > q_{max}$ the continuum theory is no longer valid, but we can investigate $P(k)$ by using numerical simulations. As Fig. 2b shows, in this regime, as $q \to 1$, $P(k)$ develops an exponential tail. This transition to an exponential can be understood in the terms of Model B discussed in Refs. [8,11], demonstrating that growth is an essential condition for power-law scaling. However, in the limit $q \to 1$ the growth is suppressed; the frequent rewiring process acts on a network with an almost constant number of nodes. The convergence of $P(k)$ to an exponential in this regime indicates that for this choice of parameters the model belongs to the class of networks defined by the ER [6] and WS [7] models [12].

Application to real networks.—To illustrate the predictive power of the obtained results, we chose to investigate the collaboration graph of movie actors [7]. In this network each actor is a node, and two actors are connected if they were cast in the same movie during their career. This network grows continuously by the addition of new actors (nodes). However, an important contribution to the network connectivity comes from another frequent event: A veteran actor plays in a new movie, establishing new internal links [process (i)] with actors (nodes) with whom he/she did not play before. Since links are never eliminated, rewiring is absent ($q = 0$); thus the evolution of this network can be described by a two parameter model (p,m): With probability p a new actor is introduced, playing with m other actors, and, with probability $1 - p$, m new internal links are generated by established actors
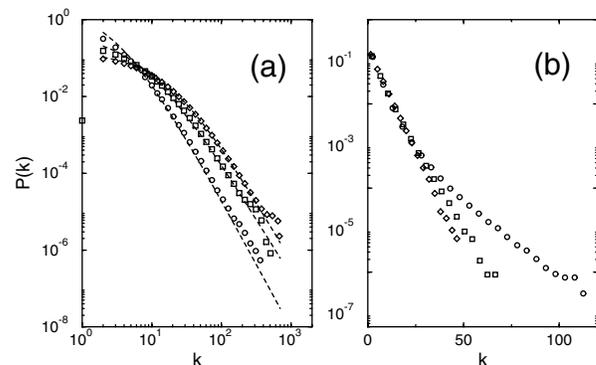

FIG. 2. (a) Comparison between the numerical simulations and the prediction of the continuum theory in the scale-free regime. In the simulations we used $t = 100\,000$, $m_0 = m = 2$. Circles: $p = 0.3$, $q = 0$; squares: $p = 0.6$, $q = 0.1$; diamonds: $p = 0.8$, $q = 0$. The data were logarithmically binned. The parameter-free predictions of Eq. (10) are shown as dashed lines. (b) The numerically obtained $P(k)$ in the exponential regime, shown on a semilogarithmic plot, indicating the convergence of $P(k)$ to an exponential in the $q \to 1$ limit. Circles: $p = 0$, $q = 0.8$; squares: $p = 0$, $q = 0.95$; diamonds: $p = 0$, $q = 0.99$.
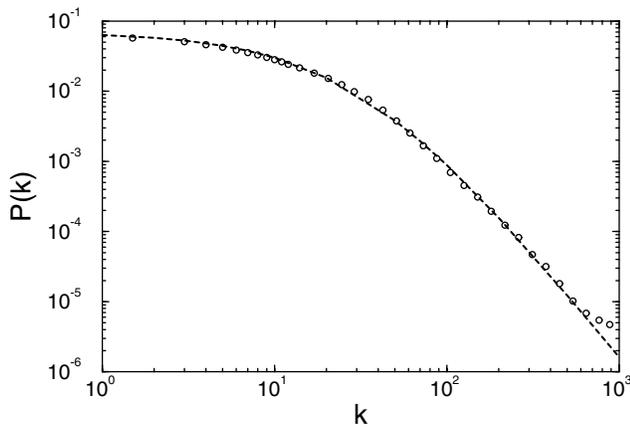
FIG. 3. The connectivity distribution of movie actors (circles) based on the Internet Movie Database containing 212 250 actors and 3 045 787 links (the same data was used in Refs. [8,11]). The data were logarithmically binned. The dashed line corresponds to the two parameter fit offered by Eq. (10) with $p = 0.937$ and $m = 1$. Note that TV films and series were excluded from the network because they lump together over many years actors that have not really played together on the same set.

playing in new movies. To obtain p and m, we fit the connectivity distribution $P(k)$ obtained for this network with Eq. (10), obtaining an excellent overlap for $p = 0.937$ and $m = 1$ (Fig. 3). The corresponding parameters of Eq. (10) are $\kappa(p,q) = 31.68$ and $\gamma(p,q,m) = 3.07$. This indicates that 93.7% of new links connect existing nodes, and only 6.3% of links come from new actors joining the movie industry. Naturally, in reality the number of actors cast in a movie varies from one movie to another. It can be shown, however, that fluctuations in m do not change the exponent γ [13]. Also, in more realistic models different parameters should be used for the number of links for new actors and for new internal links. However, such detailed modeling goes beyond the scope of this paper.

Discussion.—While in critical phenomena power-law scaling is typically associated with universality, implying that the exponents are independent of the microscopic details of the model, here we demonstrated that no such universality exists for scale-free networks, the scaling exponents depending continuously on the network's parameters. On the other hand, our results indicate the existence of a different criterion for universality based on the func-

tional form of $P(k)$: Our model predicts the existence of two regimes, the scale-free and the exponential regime. Some of the large networks investigated so far, such as the www or the actor network, are described by scale-free networks [3,5,8,10]. However, a number of fundamental network models [6,7] lead to $P(k)$ that decays exponentially, indicating the robustness of the exponential regime as well [12].

We wish to acknowledge useful discussions with Y. Tu, H. Jeong, and S. Yook, and thank B. Tjaden for providing the data on the actor collaborations. This work was partially supported by the NSF Career Award DMR-9710998 and PHY-9988674.

*Email address: alb@nd.edu
[1] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, Nature (London) **407**, 651 (2000).
[2] S. Wasserman and K. Faust, *Social Network Analysis* (Cambridge University Press, Cambridge, UK, 1994).
[3] Members of the Clever Project, Sci. Am. **280**, 54 (1999).
[4] R. Kumar, P. Raghavan, S. Rajalopagan, and A. Tomkins, in *Proceedings of the 25th International Conference on Very Large Databases* (Morgan Kaufmann Publishers, San Francisco, 1999), p. 639.
[5] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **401**, 130 (1999).
[6] P. Erdős and A. Rényi, Publ. Math. Inst. Hung. Acad. Sci. **5**, 17 (1960); B. Bollobás, *Random Graphs* (Academic Press, London, 1985).
[7] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[8] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[9] S. Redner, Eur. Phys. J. B **4**, 131 (1998).
[10] M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999).
[11] A.-L. Barabási, R. Albert, and H. Jeong, Physica (Amsterdam) **272A**, 173 (1999).
[12] For additional mechanisms leading to exponential networks, see Refs. [14–16].
[13] Y. Tu (private communication).
[14] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11 149 (2000).
[15] S. N. Dorogovtsev and J. F. F. Mendes, Phys. Rev. E **62**, 1842 (2000).
[16] P. L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000).