Statistics of Changes in Lead Node in Connectivity-Driven Networks

P. L. Krapivsky* and S. Redner†

Center for BioDynamics, Center for Polymer Studies, and Department of Physics, Boston University, Boston, Massachusetts, 02215

(Received 17 July 2002; published 3 December 2002)

We study statistical properties of the highest degree, or most popular, nodes in growing networks. We show that the number of lead changes increases logarithmically with network size N, independent of the details of the growth mechanism. The probability that the first node retains the lead approaches a finite constant for popularity-driven growth, and decays as $N^{-\phi}(\ln N)^{-1/2}$, with $\phi = 0.08607\ldots$, for growth with no popularity bias.

DOI: 10.1103/PhysRevLett.89.258703 PACS numbers: 89.75.Hc, 02.50.Cw, 05.40.–a, 87.23.Ge

Extremes are vitally important in science and engineering. These quantities are used to determine the likelihood of a rare event, such as the probability of failure of a space shuttle launch or of a dam in flood conditions. The theory of extreme statistics provides a powerful tool to understand such real-world situations [1]. Extremes are also irresistible in everyday life—we are naturally drawn to compilations of various pinnacles of human endeavor, such as, for example, lists of the most beautiful people, the richest people, the most-cited scientists, athletic records, etc. [2].

This social perspective about extremes raises new questions for which much less is known compared to the magnitude of the extreme value itself [3]. For example, how does the identity of the leader—the individual who possesses the extreme value of a particular attribute—change as a function of time? What is the rate at which lead changes occur? What is the probability that a leader retains the lead as a function of time?

We address these questions within the framework of growing networks, where the relevant quantity is the node degree—the number of links that join to each node. We view the degree as quantifying the popularity (or wealth) of the node, and the leader is the node with the highest degree. We focus on generic network models with preferential attachment to already popular nodes [4–8], and networks with random attachment. The former describe, for example, the distributions of biological genera, word frequencies, publications, urban populations, and income [4,5], and contemporary applications to collaboration networks and the World Wide Web have been developed [9]. It has been posited that a hallmark of such systems is "the rich get richer"—that is, more popular nodes tend to remain so [5,6]. Our basic goal is to examine the consequences of preferential and random attachment mechanisms in growing networks and to test whether the adage of the rich get richer really does apply.

The network grows by adding nodes, each of which links to a preexisting node with an attachment rate A_k that depends only on the degree k of the target node. We choose $A_k = k + \lambda$ with $\lambda > -1$ [4,5]. For such networks, the degree distribution has an asymptotic power-law tail, $N_k \sim N/k^{3+\lambda}$, where N_k is the number of nodes of degree k and N is the total number of nodes [5,7]. For $\lambda \to \infty$ the growth mechanism reduces to random attachment and the degree distribution is exponential.

Identity of the leader.—We characterize the identity of each node by its index J. A node of index J is the Jth one introduced into the network. To start with an unambiguous leader, the initial system contains $N = 3$ nodes, with the initial leader having degree 2 (and index 1) and the other two nodes having degree 1. A new leader arises when its degree exceeds that of the current leader. For the linear attachment rate, $A_k = k$, the average index of the leader $J_{\rm lead}(N)$ saturates to a finite value of approximately 3.4 as $N \to \infty$ (Fig. 1). With probability ≈ 0.9, the leader is from among the ten earliest nodes, while the probability that the leader is not among the 30 earliest nodes is less than 0.01. Thus only the very earliest nodes have appreciable probabilities to be the leader; the rich really do get richer. Similarly in the general case of $A_k = k + \lambda$,

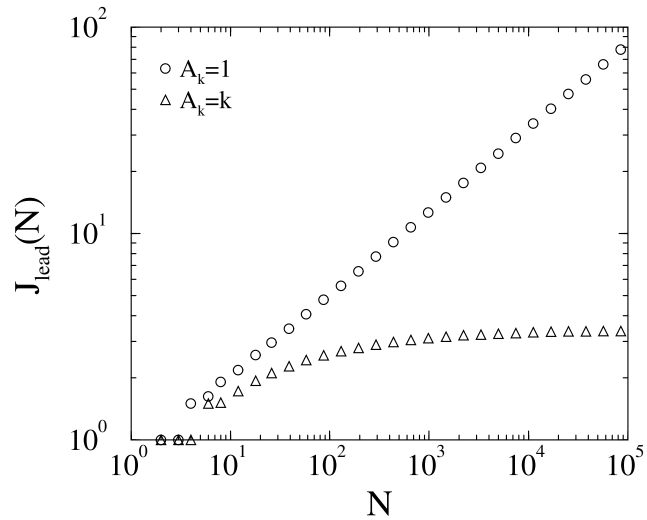

FIG. 1. Average index of the leader $J_{\rm lead}(N)$ as a function of the total number of nodes N for 10^5 realizations of a growing network of 10^5 nodes. Shown are the cases of attachment rates $A_k = 1$ and $A_k = k$.

the average index of the leader also saturates to a finite value that is a continuously increasing function of λ.

For random attachment ($A_k = 1$), the average index of the leader grows algebraically, $J_{\text{lead}}(N) \sim N^{\psi}$ with $\psi \approx 0.41$ (Fig. 1). The leader is still an early node (since $\psi < 1$), but not necessarily one of the earliest. For example, simulation indicates that, for $N = 10^5$, a node with index greater than 100 has a probability of approximately 10^{-2} of being the leader. Thus, in random attachment, the order of node creation plays a significant but not deterministic role in the identity of the leader node.

The identity of the leader can be determined from the joint index-degree distribution. Let $C_k(J, N)$ be the average number of nodes of index J and degree k. As shown in [7], for constant attachment rate, this joint distribution has the Poisson form,

$$C_k(J, N) = \frac{J}{N} \frac{|\ln(J/N)|^{k-1}}{(k-1)!}. \tag{1}$$

From this, the average index of a node of degree k is

$$J_k(N) = \frac{\sum_{1 \le J \le N} J\, C_k(J, N)}{\sum_{1 \le J \le N} C_k(J, N)} = N\left(\frac{2}{3}\right)^k, \tag{2}$$

implying $J_{\text{lead}}(N) = N(2/3)^{k_{\max}}$. We estimate the maximum degree from the extreme value criterion $\sum_{k \ge k_{\max}} N_k(N) \approx 1$. Using $N_k(N) = N/2^k$ [7], we find $2^{k_{\max}} \approx N$, or $k_{\max} \sim \ln N/\ln 2$. Therefore

$$J_{\text{lead}}(N) \propto N^{\psi}, \quad \text{with } \psi = 2 - \frac{\ln 3}{\ln 2} \approx 0.415\,037,$$

in excellent agreement with our numerical results.

For the linear attachment rate, the joint index-degree distribution is [7]

$$C_k(J, N) = \sqrt{\frac{J}{N}}\left\{1 - \sqrt{\frac{J}{N}}\right\}^{k-1}, \tag{3}$$

from which the average index of a node of degree k is $J_k(N) = 12N/[(k+3)(k+4)]$. Since $N_k(N) \sim 4N/k^3$ for the linear attachment rate [6,7], the extreme statistics criterion $\sum_{k \ge k_{\max}} N_k(N) \approx 1$ gives $k_{\max} \approx \sqrt{N}$. Therefore $J_{\text{lead}}(N) \sim 12N/k_{\max}^2 = \mathcal{O}(1)$ indeed saturates to a finite value. A similar result holds in the general case $A_k = k + \lambda$. Thus the leader is one of the first few nodes in the network.

Lead changes.—We find that the average number of lead changes $L(N)$ grows logarithmically in N for both the attachment rates $A_k = 1$ and $A_k = k$ (Fig. 2). There is, however, a significant difference in the distribution of the number of lead changes, $P(L)$, at fixed N. For $A_k = 1$, this distribution is sharply localized, with the average value $L \approx 5.609$ in a network of $N = 10^5$ nodes, while the maximum number of lead changes in 10^5 realizations was 16. On the other hand, for $A_k = k$, $P(L)$ has a significant tail and the maximum number of lead changes is 63. This longer tail in $P(L)$ for linear attachment stems from repeated lead changes among the two leading nodes. Even though the distribution is visually broader, the average number of lead changes, $L \approx 5.096$, is less than that for $A_k = 1$. Related to lead changes is the number of *distinct* nodes that enjoy the lead over the history of the network. Simulations indicate that this quantity also grows logarithmically in N.

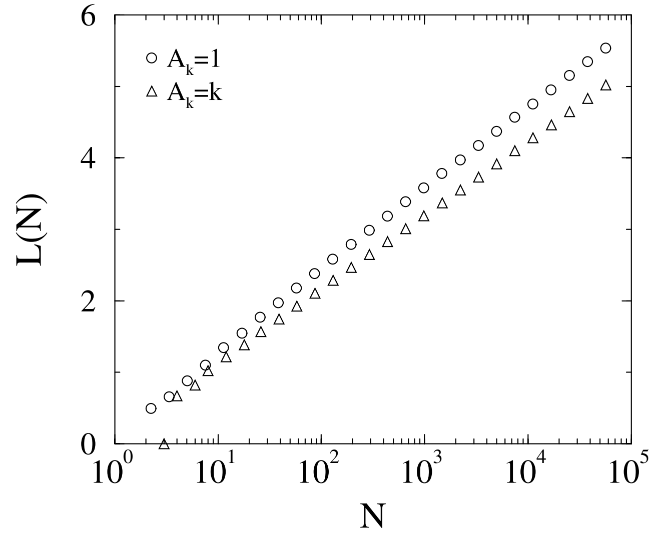

FIG. 2. Average number of lead changes $L(N)$ as a function of network size N for 10^5 realizations of the network for $A_k = 1$ and $A_k = k$.

This logarithmic behavior can be easily understood for the attachment rate $A_k = 1$. Here the number of lead changes cannot exceed the upper bound given by the maximal degree $k_{\max} \sim \ln N/\ln 2$. To establish the logarithmic growth in the general case we first note that when a new node is added, the lead changes if the leadership is currently shared between two (or more) nodes and the new node attaches to a coleader. The number of coleader nodes (with degree $k = k_{\max}$) is $N/k_{\max}^{3+\lambda}$, while the probability of attaching to a coleader is $k_{\max}/N$. Thus the average number of lead changes satisfies

$$\frac{d}{dN} L(N) \propto \frac{k_{\max}}{N} \frac{N}{k_{\max}^{3+\lambda}}. \tag{4}$$

Since the maximal degree $k_{\max}$ grows as $N^{1/(2+\lambda)}$, Eq. (4) reduces to $dL(N)/dN \propto N^{-1}$ and thus gives the logarithmic growth $L(N) \propto \ln N$. This argument can be adapted to networks with arbitrary attachment rates (except those growing faster than linearly with k [7]), and thus the growth law $L(N) \propto \ln N$ is universal. This universality is reminiscent of the radius of random networks which typically are proportional to $\ln N$, independent of their construction mechanism.

Fate of the first leader.—Figure 3 shows that the degree distribution of the first node depends on the initial conditions for the linear attachment rate; the same is true in

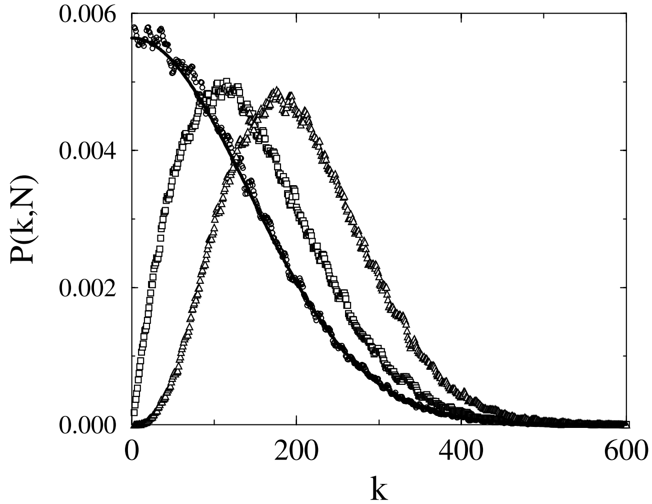

FIG. 3. Degree distribution of the first node for the dimer, trimer (2 links to the initial node), and pentamer (4 links) initial conditions based on 10^5 realizations of a network with 10^5 links. The curve is the prediction of Eq. (7).

the general case $A_k = k + \lambda$ while for $A_k = 1$ the initial condition is asymptotically irrelevant.

We can determine the degree distribution of the first node analytically for the constant and linear attachment rates. (A similar approach is given in Ref. [8]). Let $P(k, N)$ be the probability that the first node has degree k in a network of N links [10]. For $A_k = k$, this probability obeys the master equation

$$P(k, N + 1) = \frac{k - 1}{2N} P(k - 1, N) + \frac{2N - k}{2N} P(k, N). \tag{5}$$

The first term on the right accounts for the situation when the earliest node has degree $k - 1$. Then a new node attaches to it with probability $(k - 1)/2N$, thereby increasing the probability for the node to have degree k. Conversely, with probability $(2N - k)/2N$ a new node does not attach to the earliest node, thereby giving the second contribution to $P(k, N + 1)$.

The solution to the master equation (5) for the "dimer" initial condition is [11]

$$P(k, N) = \frac{1}{2^{2N-k-1}} \frac{(2N - k - 1)!}{(N - k)!\,(N - 1)!}. \tag{6}$$

For $N \to \infty$, this simplifies to the Gaussian distribution

$$P(k, N) \sim \frac{1}{\sqrt{\pi N}} e^{-k^2/4N} \tag{7}$$

for finite values of the scaling variable $k/N^{1/2}$. Thus the typical degree of the first node is of the order of $N^{1/2}$; this is the same scaling behavior as the degree of the leader node. For the trimer initial condition (which we typically used in simulations) we obtained the degree distribution of the first node in the form of a series of ratios of gamma functions [11], in which $P(k, N)$ has an $e^{-k^2/4N}$ Gaussian tail, independent of the initial condition. The degree of the first node also approximates that of the leader node [3] more and more closely as the degree of the first node in the initial state is increased.

Although $P(k, N)$ contains all information about the degree of the first node, the behavior of its moments $\langle k^a \rangle_N = \sum k^a P(k, N)$ is simpler to appreciate. To determine the moments, it is more convenient to construct their governing recursion relations directly, rather than to calculate the moments from $P(k, N)$. Using Eq. (5), the average degree satisfies the recursion relation

$$\langle k \rangle_{N+1} = \langle k \rangle_N \left(1 + \frac{1}{2N}\right), \tag{8}$$

whose solution is

$$\langle k \rangle_N = \Lambda \frac{\Gamma(N + \frac{1}{2})}{\Gamma(\frac{1}{2})\Gamma(N)} \sim \frac{\Lambda}{\sqrt{\pi}} N^{1/2}. \tag{9}$$

The prefactor Λ depends on the initial condition, with $\Lambda = 2, 8/3, 16/5, \ldots$ for the dimer, trimer, tetramer, etc., initial conditions.

This multiplicative dependence on the initial condition means that the first few growth steps substantially affect the average degree of the first node. For example, for the dimer initial condition, the average degree of the first node is, asymptotically, $\langle k \rangle_N \sim 2\sqrt{N/\pi}$. However, if the second link attaches to the first node, an effective trimer initial condition arises and $\langle k \rangle_N \sim (8/3)\sqrt{N/\pi}$. Thus small initial perturbations lead to huge differences in the degree of the first node.

An intriguing manifestation of the rich get richer phenomenon is the behavior of the survival probability $S(N)$ that the first node leads throughout the growth up to size N (Fig. 4). For the linear attachment rate, $S(N)$ saturates to a finite nonzero value of approximately 0.277 as $N \to \infty$; saturation also occurs for the general attachment rate $A_k = k + \lambda$. Thus for these popularity-driven systems, the rich get richer holds in a strong form–the lead never changes with a positive probability.

For constant attachment rate, $S(N)$ decays to zero as $N \to \infty$, but asymptotic behavior is not apparent even when $N = 10^8$. A power-law $S(N) \propto N^{-\phi}$ is a reasonable fit, but the local exponent is still slowly decreasing at $N \approx 10^8$ where it has reached $\phi(N) \approx 0.18$. To understand the slow approach to asymptotic behavior, we study the degree distribution of the first node. This quantity satisfies the recursion relation

$$P(k, N) = \frac{1}{N} P(k - 1, N - 1) + \frac{N - 1}{N} P(k, N - 1), \tag{10}$$

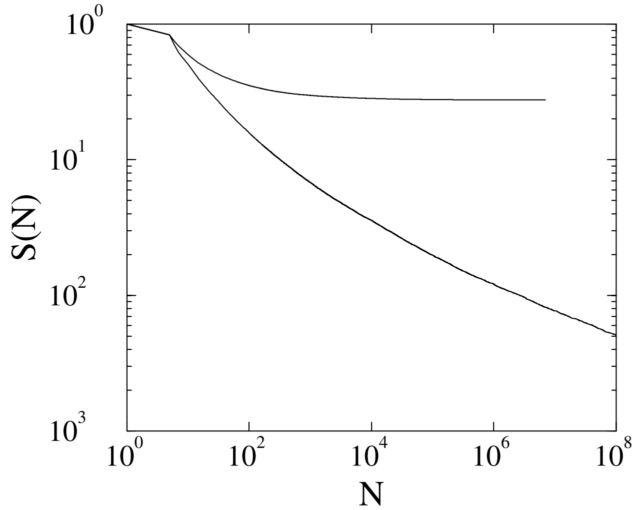

FIG. 4. Probability that the first node leads throughout the evolution for 10^5 realizations of up to size $N = 10^7$ for $A_k = k$ (upper), and up to $N = 10^8$ for $A_k = 1$ (lower).

which reduces to the convection-diffusion equation

$$\left(\frac{\partial}{\partial \ln N} + \frac{\partial}{\partial k}\right)P = \frac{1}{2}\,\frac{\partial^2 P}{\partial k^2} \qquad (11)$$

in the continuum limit. The solution is a Gaussian

$$P(k, N) = \frac{1}{\sqrt{2\pi \ln N}}\,\exp\left\{-\frac{(k - \ln N)^2}{2\ln N}\right\}. \qquad (12)$$

Therefore the degree of the first node grows as $\ln N$, with fluctuations of the order of $\sqrt{\ln N}$. On the other hand, the maximal degree grows faster, as $v \ln N$ with $v = 1/\ln 2$, and its fluctuations are negligible.

We now estimate the large-N behavior of $S(N)$ as $\sum_{k \geq k_{\max}} P(k, N)$. This approximation gives

$$S(N) \propto \int_{v \ln N}^{\infty} \frac{dk}{\sqrt{\ln N}} \exp\left\{-\frac{(k - \ln N)^2}{2\ln N}\right\} \propto N^{-\phi}\,(\ln N)^{-1/2}, \qquad (13)$$

with $\phi = (v - 1)^2/2 \approx 0.097\,989\ldots$. The logarithmic factor leads to a very slow approach to asymptotic behavior.

The above estimate is based on the Gaussian approximate for $P(k, N)$ which is not accurate outside the scaling region, namely, for $k \gg \ln N + \sqrt{\ln N}$. However, we can determine $P(k, N)$ exactly because its defining recursion formula, Eq. (10), is closely related to that of the Stirling numbers of the first kind $\left[{N \atop k}\right]$ [12], and the solution for the dimer initial condition is $P(k, N) = \left[{N \atop k}\right]/N!$. The corresponding generating function is

$$S_N(x) = \sum_{k=1}^{N} P(k, N)\, x^k = \frac{x(x+1)\ldots(x+N-1)}{N!}.$$

Using the Cauchy theorem, we express $P(k, N)$ in terms of the contour integral $S_N(x)/x^{k+1}$. When $N \to \infty$, this contour integral is easily computed by applying the saddle point technique [11]. Finally, we arrive at Eq. (13) with the same logarithmic prefactor but with a slightly smaller *exact* exponent $\phi = 1 - v + v \ln v \approx 0.086\,07$.

In summary, lead changes are rare in popularity-driven network growth processes and leadership is restricted to the earliest nodes. With finite probability, the first node remains the leader throughout the evolution. For growth with no popularity bias, leadership is shared among a somewhat larger cadre of nodes. As a consequence the average index of the leader node grows as N^{ψ} with $\psi = 0.415\,037\ldots$. The possibility of sharing the lead among a larger subset of nodes gives a rich dynamics in which the probability that the first node retains the lead decays as $N^{-\phi}\,(\ln N)^{-1/2}$ with $\phi = 0.080\,67\ldots$.

We are grateful to NSF Grant No. DMR9978902 for partial financial support of this research.

*Electronic address: paulk@bu.edu
†Electronic address: redner@bu.edu

[1] J. Galambos, *The Asymptotic Theory of Extreme Order Statistics* (R. E. Krieger Publishing Co., Malabar, Florida, 1987).
[2] See, e.g., *The Guiness Book of World Records* (Bantam Books, New York, 2002).
[3] A. A. Moreira, J. S. de Andrade, Jr., and L. A. N. Amaral, cond-mat/0205411.
[4] G. U. Yule, Philos. Trans. R. Soc. London B **213**, 21 (1924); *The Statistical Study of Literary Vocabulary* (Cambridge University Press, Cambridge, 1944).
[5] H. A. Simon, Biometrica **42**, 425 (1955); Inf. Control **3**, 80 (1960).
[6] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[7] P. L. Krapivsky and S. Redner, Phys. Rev. E **63**, 066123 (2001).
[8] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. E **63**, 062101 (2001).
[9] See, e.g., S. H. Strogatz, Nature (London) **410**, 268 (2001).
[10] The normalized attachment probability is A_k/A, with $A = \sum A_j N_j$. For the linear attachment rate, A is twice the total number of links. Hence formulas become more neat if we denote by N the total number of links.
[11] P. L. Krapivsky and S. Redner (to be published).
[12] See R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science* (Addison-Wesley, Reading, MA, 1989).