PHYSICAL REVIEW E, VOLUME 65, 017102

Scale-free growing networks imply linear preferential attachment

Kasper Astrup Eriksen* and Michael Hörnquist†
NORDITA, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark
(Received 22 August 2001; published 17 December 2001)

It has been recognized for some time that a network grown by the addition of nodes with linear preferential attachment will possess a scale-free distribution of connectivities. Here we prove by some analytical arguments that the linearity is a necessary component to obtain this kind of distribution. However, the preferential linking rate does not necessarily apply to single nodes, but to groups of nodes of the same connectivity. We also point out that for a time-varying mean connectivity the linking rate will deviate from a linear expression by an extra asymptotically logarithmic term.

DOI: 10.1103/PhysRevE.65.017102 PACS number(s): 89.75.Hc, 02.50.Ey, 05.65.+b, 89.75.Da

In naturally occurring networks, distributions of connectivities that are approximately power laws seem to be ubiquitous. For example, it has been shown that systems as disparate as the number of hyperlinks pointing to a particular web page [1], collaborations of movie actors [2], citation patterns of scientific publications [3], metabolic networks [4], protein-protein interaction network [5], and human sexual contacts [6] all possess this kind of distribution.

For people interested in a particular system, it is indeed a nice observation to make that the distribution of connectivities is approximately a power law, but soon afterward questions about the microscopic details of its emergence start to arise. In response to such questions, various specific microscopic dynamical models that give rise to a power-law distribution have been investigated [2,7,8]. These works have in common that the networks are grown by addition of new nodes and that each such new node preferentially attaches to other nodes with a high number of connections. A scale-free distribution appeared in these investigations only when the probability of attachment was proportional, at least asymptotically, to the number of links already attached to the specific node—so-called linear preferential attachment. It has also been shown that at least three real-world networks—a science citation network, the internet, and a science collaboration network—grow by linear preferential linking [9]. Here we go the other way, and prove that this *linear* form of the connection kernel is not only sufficient, but also necessary, for a scale-free distribution to appear in a growing network. We remark that this result was indicated for a *homogeneous* connection kernel already in [2] and proved in [8]. Here, however, we make no *a priori* assumptions on the functional form of the attachment probability.

In spite of the fact that we focus only on the node degree distribution, leaving aside all other aspects of the network topology, such as the average path length, diameter, clustering, etc., we will in this Brief Report borrow our notation from the network community. However, with a change of terminology, the finding can be applied also to other types of growth processes. In particular, let the distribution function $N(k,t)$ be the number of nodes with k links (connectivity k) at time t. We prove that if the distribution function $N(k,t)$ of a system satisfies the following conditions: (1) the distribution is stationary and scale-free, at least asymptotically, i.e., $N(k,t)\propto k^{-\gamma}$ $(k\to\infty)$, (2) the exponent $\gamma>2$, and (3) the constant of proportionality $A(t)=N(k,t)/k^{-\gamma}$ grows as a function of time, then the underlying dynamics is *necessarily*, on average, governed by a growth process with asymptotically linear preferential attachment.

Before we proceed to the actual proof, let us elaborate a bit more on some of the above statements. The prerequisite $\gamma>2$ is included in order to make all the entering sums convergent and thus relatively independent of an upper cutoff. Fortunately, most real-world distributions have $2<\gamma<3$. The last condition is crucial, because otherwise the present setting will be a special case of the theory of stochastic multiplicative processes, where power-law distributions are known to occur by a different mechanism [10].

Originally, the term ''linear preferential attachment'' was coined for a pure growth model where one link at a time is added to an existing node with k connections with a probability that is proportional to k. This kind of model had in fact already been studied in 1955 by Simon [7]. Furthermore, Simon showed that the important entity was the *average* rate at which a link is added to one of the nodes with k links [11]. Let us introduce the rank

$$r(k,t)=1+\sum_{i=k+1}^{\infty}N(i,t) \tag{1}$$

and the attachment rate as the generalized current

$$J(k,t)=r(k,t+1)-r(k,t). \tag{2}$$

The average attachment rate to a single node with k links can now be expressed as [12]

$$j(k,t)=\frac{\langle J(k,t)\rangle}{\langle N(k,t)\rangle}, \tag{3}$$

and the linear preferential attachment property is formulated as $j\propto k$.

*Email address: kasper@nordita.dk
†Permanent address: Applied Mathematical Physics, Department of Technology and Science, Linköping University, Campus Norrköping, S-601 74 Norrköping, Sweden. Email address: micho@itn.liu.se

We now generalize to a general situation where it is microscopically also possible to remove link ends. As an example, consider the situation where at time t two links are added to node A, which already has nine links. At time $t+1$, the links are removed again. Obviously, nothing has happened, and consequently the attachment rate we refer to above is the net attachment rate. Secondly, the step of adding two links at a time can be viewed as two separate addition steps: first a link is added to a node (A) with nine links and then a link is added to a node with ten links. The correct way to define the attachment rate, say, for nodes with ten links, is thus in terms of the generalized current $J(10,t)$, i.e., the number of nodes that in one step get more than 10 links.

We now proceed to the proof. From Eqs. (1), (2), and (3) it follows immediately that the average attachment rate can be written

$$j(k,t)=\frac{1}{\langle N(k,t)\rangle}\sum_{i=k+1}^{\infty}[\langle N(i,t+1)\rangle-\langle N(i,t)\rangle]. \quad (4)$$

With the scale-free distribution $\langle N(k,t)\rangle = A(t)k^{-\gamma}$, these sums can for $k\to\infty$ be replaced by integrals [13], which yields the closed form

$$j(k,t)\approx\frac{A(t+1)-A(t)}{A(t)}\frac{k}{\gamma-1}\propto k, \quad k\to\infty. \quad (5)$$

This result means that as soon as we have a growing system where the distribution of connectivities is given asymptotically by a power-law distribution, there has to be a mechanism of preferential attachment to existing nodes with a probability that on average is asymptotically proportional to k.

Let us emphasize that our derivation of Eq. (4) nowhere used the first requirement about a stationary scale-free distribution, which means that it is general. For instance, one can show by straightforward calculations that for a power-law distribution of connectivities, where the exponent changes with time (for instance, due to a change in the average connectivity, as measured for the World Wide Web [14]), the lowest order correction to the linear preferential attachment is a term proportional to $k\log k$ asymptotically in k [15].

To summarize, we have shown that, for all distributions that asymptotically resemble power laws, it is necessary for the net rate of attachment of new nodes to one of the nodes with k links to be asymptotically proportional to k, i.e., for the so-called linear preferential attachment rule to necessarily apply. Furthermore, our proof is not limited to *pure* growth models, but also covers reorganization and loss of internal links. It is thus possible to rephrase the question of why power laws are ubiquitous in real life in terms of why linear preferential attachment is so common. If attachment is viewed as a stochastic process, then, given no other information, the best guess at the intrinsic attractiveness of a node is simply proportional to the number of connections the node already possesses. This then leads to linear preferential attachment, hinting at an explanation similar in spirit to the random matrix distributions observed in many other systems.

Finally, in order to avoid any misunderstanding, we stress that of course there are many other mechanisms for obtaining scale-free distributions in other cases, e.g., self-organized criticality [16] and multiplicative random processes with drift and limits [10]. However, these cases do not satisfy our requirements above, and hence are not affected by the present statement on the origin of power laws.

The authors would like to thank Kim Sneppen, Jesper Borg, Ingve Simonsen, and Svante Jonsell for careful reading of the manuscript and many valuable suggestions. M.H. acknowledges financial support from STINT, The Swedish Foundation for International Cooperation in Research and Higher Education.

[1] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **401**, 130 (1999).

[2] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).

[3] S. Redner, Eur. Phys. J. B **4**, 131 (1998).

[4] H. Jeong *et al.*, Nature (London) **407**, 651 (2000).

[5] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai, Nature (London) **411**, 41 (2001).

[6] F. Liljeros *et al.*, Nature (London) **411**, 907 (2001).

[7] H. A. Simon, Biometrika **42**, 425 (1955).

[8] P. L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000); P. L. Krapivsky and S. Redner, Phys. Rev. E **63**, 066123 (2001).

[9] M. E. J. Newman, Phys. Rev. E **64**, 025102(R) (2001); H. Jeong, Z. Néda, and A.-L. Barabási, e-print cond-mat/0104131.

[10] M. Levy and S. Solomon, Int. J. Mod. Phys. C **7**, 595 (1996); D. Sornette and R. Cont, J. Phys. I **7**, 431 (1997).

[11] This average is over all nodes with k links, so the situation is covered where, for instance, half the nodes are dead, in the sense that it is impossible to link to them, while the remaining nodes have twice as big a probability of being picked.

[12] The average taken depends on the system considered. For a computer simulation, it is the ensemble average over many independent runs, while for a real system, where one has only a single copy of the network, it is the average over a series of time steps around t, or an average over nodes with approximately the same number of connections k. Further, one might instead have used $\langle J(k,t)/N(k,t)\rangle$ as the definition of $j(k,t)$. In the limit of infinitely large networks, the two definitions coalesce and the choice in the main text is easier to treat analytically.

[13] With the Yule distribution $N(k,t)=A(t)B(k,\gamma)$ the sums can be calculated exactly, with a final expression that is identical to Eq. (5). Here $B(k,\gamma)=\Gamma(k)\Gamma(\gamma)/\Gamma(k+\gamma)$ is Euler's Beta function, and Γ is the Gamma function,

[14] A. Broder *et al.*, Comput. Netw. **33**, 309 (2000).

[15] The average attachment rate will in this case take the form

$$j(k,t)=\frac{k}{\gamma-1}\left(\frac{A(t+1)-A(t)}{A(t)}-\frac{d\gamma}{dt}\sum_{i=0}^{k}\frac{1}{\gamma-1+i}\right).$$

[16] P. Bak, C. Tang, and K. Wiesenfeld, Phys. Rev. Lett. **59**, 381 (1987).