Size-dependent degree distribution of a scale-free growing network

S. N. Dorogovtsev,[1,2,*] J. F. F. Mendes,[1,†] and A. N. Samukhin[2,‡]
[1]*Departamento de Física and Centro de Física do Porto, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre 687, 4169-007 Porto, Portugal*
[2]*A.F. Ioffe Physico-Technical Institute, 194021 St. Petersburg, Russia*
(Received 29 January 2001; published 21 May 2001)

We propose the simplest model of scale-free growing networks and obtain the exact form of its degree distribution for any size of the network (degree is a number of connections of a node). We demonstrate that a trace of initial conditions — a hump near cutoff of the degree distribution at $k_{cut} \sim t^{\beta}$ — may be found for any network size. Here $\beta = 1/(\gamma-1)$, where γ is the exponent of the degree distribution of the network. These size effects implement a natural boundary for the observation of the scale-free networks.

DOI: 10.1103/PhysRevE.63.062101 PACS number(s): 05.10.−a, 05.40.−a, 05.50.+q, 87.18.Sn

Significant progress was made recently in the field of evolving networks [1–7]. It was observed that a number of growing networks in nature (World Wide Web, Internet, collaboration nets, some networks in biology, etc.) are *scale-free*, i.e., their degree distribution is of a power-law form (degree is a number of connections of a node [8]). Moreover, it was found that at least many of the natural networks must have degree distributions with long tails, otherwise growing networks are not resilient enough to random breakdowns [9–11]. Infinite networks with the degree distribution exponent $\gamma \leqslant 3$ do not decay for *any* concentration (less than one) of randomly removed nodes or links [10].

The proposed mechanism of self-organization of networks into scale-free structures, the preferential linking, is quite natural [12]. New links of the growing networks are preferentially attached to nodes that already have many connections (degree k). In fact, it is the realization of a general principle — *popularity is attractive*. Phenomena of such kind were first considered by Simon long ago [13,14]. Several types of preferential linking were proposed [12,15–19] that provide a variety of the γ exponent values between 2 and infinity.

One should emphasize that only a few scale-free networks are known yet. The range of the values of degree, in which the power-law behavior can be observed, is usually too narrow for a precise measurement of the exponent γ. Why are so few scale-free networks observed? Why are the values of γ for all of them only between 2 and 3? (Note that not any network has to be resilient, e.g., neither nodes nor links of collaboration networks are removable by definition [21].) Here, we discuss these questions.

In previous papers, degree distributions $P(k,t)$ of scale-free networks were calculated only in the ''thermodynamic limit,'' i.e., in the limit of the large system size t, which also plays the role of time, if one node is added at each increment of time. In this case, the distribution is stationary, and is of the form $P(k) \sim k^{-\gamma}$ in all ranges of large enough k, $k \gg 1$. Nevertheless, real networks are finite. The evolution of $P(k,t)$ to the stationary distribution turns to be nontrivial. Using a simple exactly solvable model, we demonstrate below that, for finite networks, the power-law region of degree distribution has the cutoff at $k_{cut} \sim t^{\beta}$, where $\beta = 1/(\gamma-1)$. We show that the trace of the initial conditions, i.e., of the initial configuration of the network—the hump at $k_h \sim k_{cut} \sim t^{\beta}$—may be observed at *any* size of the network. This cutoff in degree distribution allows observation of the power-law dependence only for very large networks. For large values of γ, the power-law dependence is practically unobservable.

Let us introduce the model of the scale-free growing network with *undirected* links (see Fig. 1). Initially ($t=2$), there are three nodes, $s=0,1,2$, each with degree 2.

(i) At each increment of time, a new node is added.

(ii) It is connected to both ends of a randomly chosen link by two undirected links.

As far as we know, it is the simplest model of a scale-free network. The preferential linking arises in it not because of some special rule including a function of degree as in [12] but naturally. Indeed, in the model that we consider, the probability that a node has the randomly chosen link attached to it is equal to the degree k of the node divided by the total

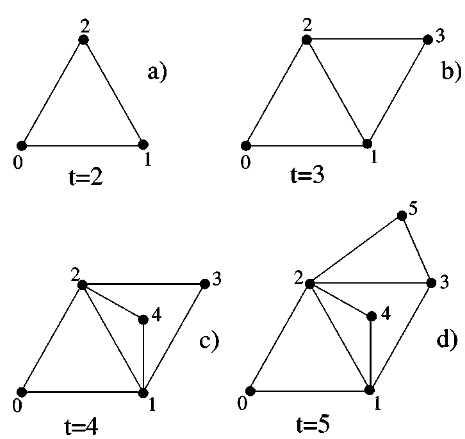

FIG. 1. Illustration of the simplest model of scale-free growing networks. In the initial configuration $t=2$ three sites are present, $s=0,1,2$ (a). At each increment of time, a new node with two links is added. These links are attached to the ends of a randomly chosen link of the network.

*Email address: sdorogov@fc.up.pt
†Email address: jfmendes@fc.up.pt
‡Email address: alnis@samaln.ioffe.rssi.ru

number of links, $2t-1$. Therefore, the evolution of the network is described by the following master equation,

$$p(k,s,t+1)=\frac{k-1}{2t-1}p(k-1,s,t)+\frac{2t-1-k}{2t-1}p(k,s,t), \tag{1}$$

with the initial condition, $p(k,s=\{0,1,2\},t=2)=\delta_{k,2}$. Also, $p(k,t,t)=\delta_{k,2}$. Here, $p(k,s,t)$ is the probability that the site $0\leqslant s\leqslant t$ has k connections at time t. Note that this master equation and all the following ones are exact for all $t\geqslant 2$. Equation (1) has the form similar to that of the Barabási-Albert model [12]. Therefore, one may expect that the scaling exponents of these models have to coincide. One should emphasize that, here, unlike the Simon's model, a number of nodes (one) added at each time step is fixed.

From Eq. (1), we can obtain several useful exact relations for our model. In particular, from Eq. (1), one may find the equation for the average degree of an individual node, $\overline{k}(s,t)\equiv\Sigma_{k=2}^{t-s+2}kp(k,s,t)$:

$$\overline{k}(s,t+1)=\frac{2t}{2t-1}\overline{k}(s,t), \quad \overline{k}(t,t)=2. \tag{2}$$

One can obtain easily its solution,

$$\overline{k}(s,t)=2^{t-s+1}\frac{(t-1)!}{(s-1)!}\frac{(2s-3)!!}{(2t-3)!!}\overset{s,t\gg 1}{\cong}2\sqrt{\frac{t}{s}}. \tag{3}$$

Here, $s\geqslant 2$ and $\overline{k}(0,t)=\overline{k}(1,t)=\overline{k}(2,t)$. Hence, the scaling exponent β, defined through the relation, $\overline{k}(s,t)\propto(s/t)^{-\beta}$, equals 1/2 like for the Barabási-Albert model.

Also, one can find the average number $\overline{b}(s,s')$ of links between the sites s and s' for any $s<s'\leqslant t$, $0<\overline{b}(s,s')\leqslant 1$. In fact, $\overline{b}(s,s')$ is the average of the element of the adjacency matrix [8] over all possible realizations of the growth. The equation for this quantity is

$$\overline{b}(s,s'+1)=\frac{1}{2t-1}\left[\sum_{u=0}^{s-1}\overline{b}(u,s)+\sum_{u=s+1}^{s'}\overline{b}(s,u)\right]. \tag{4}$$

Its exact solution for $s<s'$ is of the form,

$$\overline{b}(s,s')=2^{s'-s}\frac{(s'-2)!}{(s-1)!}\frac{(2s-3)!!}{(2s'-3)!!}\overset{s,s'\gg 1}{\cong}\frac{1}{\sqrt{ss'}}, \tag{5}$$

and $\overline{b}(0,s')=\overline{b}(1,s')=\overline{b}(2,s')=1$. Expression (5) demonstrates distribution of links in the network.

We found exactly the degree distribution of the oldest nodes, $p(k,0,t)=p(k,1,t)=p(k,2,t)$,

$$p(k,2,t)=\frac{(k-1)}{2^{t-k}(t-k)!}\frac{(2t-k-2)!}{(2t-3)!!}\overset{t\gg k^2}{\cong}\frac{(k-1)}{2t}. \tag{6}$$

This relation turns to be useful for finding the total degree distribution. Also, one may obtain the relation, $p(2,s,t)=(2s-3)/(2t-3)$. The scaling form of $p(k,s,t)$ for $k,s,t\gg 1$ and $k\sqrt{s/t}$ fixed is obtained using the Z transform. The scaling relation is of the form,

$$p(k,s,t)=\sqrt{\frac{s}{t}}\left(k\sqrt{\frac{s}{t}}\right)\exp\left(-k\sqrt{\frac{s}{t}}\right). \tag{7}$$

This is a particular case of the corresponding scaling relations for the scale-free networks [17].

The matter of interest is the total degree distribution, $P(k,t)\equiv\Sigma_{s=0}^{t}p(k,s,t)/(t+1)$. The equation for it can be derived from Eq. (1),

$$P(k,t)=\frac{t}{t+1}\left[\frac{k-1}{2t-3}P(k-1,t-1)+\left(1-\frac{k}{2t-3}\right)P(k,t-1)\right]+\frac{1}{t+1}\delta_{k,2} \tag{8}$$

with the initial condition $P(k,2)=\delta_{k,2}$.

The exact solution of Eq. (8) is

$$P(k,t)=\frac{24}{k(k+1)(k+2)}\frac{1}{(t+1)(2t-3)!!}\frac{(2t-k-2)!}{2^{t-k}(t-k)!}\times\left\{(t-k)\left[t+\frac{(k-2)(k+1)}{4}\right]+\frac{(k-1)k(k+1)(k+2)}{8}\right\}. \tag{9}$$

One may check Eq. (9) inserting it directly into Eq. (8). We obtained Eq. (9) using the distribution function $\tilde{P}(k,t)\equiv\Sigma_{s=3}^{t}p(k,s,t)/(t-2)$, which looks less cumbersome than $P(k,t)$ and may be found without great problems, and the expression for $p(k,2,t)$, Eq. (6).

From Eq. (8) with $t\to\infty$, it follows the equation for the stationary distribution $P(k)$,

$$(k-1)P(k-1)-(k+2)P(k)+2\delta_{k,2}=0, \tag{10}$$

where the solution is

$$P(k)=\frac{12}{k(k+1)(k+2)}. \tag{11}$$

Equation (11) is similar to the form of the stationary degree distribution found for the Barabási-Albert model [17,18]. One sees that $\gamma=3$.

Our aim is to find how the stationary distribution is reached. From Eq. (9), for $t\gg k\gg 1$, one gets

$$P(k,t)=P(k)\left[1+\frac{1}{4}\frac{k^2}{t}+\frac{1}{8}\left(\frac{k^2}{t}\right)^2\right]\exp\left\{-\frac{1}{4}\frac{k^2}{t}\right\}. \tag{12}$$

The factor $P(k,t)/P(k)\equiv g(k/\sqrt{t})$ depends only on the combination $k/\sqrt{t}$. Therefore, the peculiarities of the distribution induced by the size effects never disappear but only move with increasing time in the direction of large degree. The function $g(k/\sqrt{t})$ is close to 1 for $k<\sqrt{t}$, has a hump at k_{max}

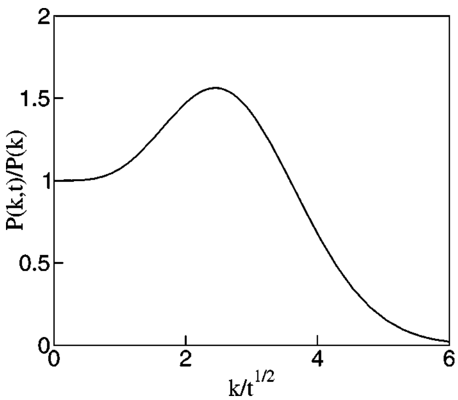

FIG. 2. Deviation of the degree distribution of the finite-size network from the stationary one, $P(k,t)/P(k,t\to\infty)$, vs $k/\sqrt{t}$. The form of the hump depends on the initial configuration.

between $\sqrt{t}$ and $4\sqrt{t}$ with a maximum at $k_{max}/\sqrt{t}=\sqrt{6}=2.449\ldots$, $g(k_{max}/\sqrt{t})=7\,e^{-3/2}=1.562\ldots$, and the cutoff at $k_{cut}\sim 4\sqrt{t}$ (see Fig. 2). Hence, the power-law behavior is observable only in a rather narrow region, $1\ll k\ll\sqrt{t}$.

One may check that the form of the hump in Fig. 2 depends on the initial conditions. In our case, the evolution starts from the configuration shown in Fig. 1(a). If we start the growth from another configuration, the form would be different.

We have demonstrated above the size-dependence of degree distribution using the exactly solvable example. What are the general reasons of such behavior of scale-free networks?

Measuring of degree distributions is always impeded by the strong fluctuations at large k. The reason for such fluctuations is the poor statistics in this region. One can easily estimate the characteristic value k_f above which the fluctuations are strong. If $P(k)\sim k^{-\gamma}$, and $\gamma>2$, then $tk_f^{-\gamma}\sim 1$. Therefore, $k_f\sim t^{1/\gamma}$. One may improve the situation using the cumulative distributions, $P_{cum}(k)\equiv\int_k^\infty dkP(k)$, instead of $P(k)$. Also, in simulations, one may make a lot of runs to increase the statistics. Nevertheless, one cannot pass the cutoff k_{cut} that we discuss. One can estimate the cutoff: $tP_{cum}(k_{cut})\sim 1$, so $k_{cut}\sim t^{1/(\gamma-1)}$. This cutoff is the real barrier for the observation of the power-law dependence.

We have shown that the degree distribution of individual sites is an exponentially decreasing function at large k [see Eq. (7)]. For the scale-free networks, it can be written in the general scaling form [17], $p(k,s,t)=(s/t)^\beta f[k(s/t)^\beta]$, where $f(x)$ is a scaling function, and the relation [16,17] between the exponents β and γ is $\beta(\gamma-1)=1$. In the particular case of the proposed model, $f(x)=x\exp(-x)$. The exponent β also figures in the relation for the average degree, $\overline{k}(s,t)\propto(s/t)^{-\beta}$. It follows from the scaling form of $p(k,s,t)$ that the cutoff of the total distribution is determined by the degree distribution of the individual nodes with the smallest number s, i.e., by the oldest ones. Therefore, $k_{cut}(1/t)^\beta\sim$ const and $k_{cut}\sim t^\beta=t^{1/(\gamma-1)}$. For the considered model, $\beta=1/2$, see Eq. (7). The degree distributions of the oldest nodes (and their quantity) depend strongly on the initial conditions. Hence, the part of the total degree distribution near the cutoff depends strongly on this factor. The power-law dependence of the distribution can be observed only if it exists for at least 2 or 3 decades of the degree. For this, the networks have to be large, $t>10^{2.5(\gamma-1)}$. But there are only a few large networks in nature! For large γ, one practically has no chances to find the scale-free behavior.

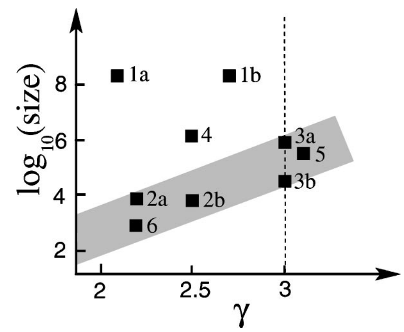

FIG. 3. Log-linear plot of the size vs the γ exponent value for the networks reported as having power-law degree distributions. The line $\log_{10}t\sim 2.5(\gamma-1)$ is the finite-size boundary for the observation of the power-law degree distributions. The dashed line $\gamma=3$ is the resilience boundary. This boundary is important for those growing networks that have to be stable to random breakdowns. The points: $1a$ and $1b$ are obtained for incoming and outgoing links of the pages of the World Wide Web [4,7] (also, $\gamma_{in}=2.1$ and $\gamma_{out}=2.45$ were obtained from the complete map of the nd.edu domain of the web, 325,729 nodes [12], $\gamma_{in}=1.94$ was obtained for the domain level of the web in spring 1997 [22]), $2a$ is for outgoing links for the inter-domain structure of the Internet and $2b$ is for outgoing links for the Internet at the router level [3], $3a$ and $3b$ are for citations of the ISI data base and Phys. Rev. D [1] (also, estimations of Ref. [18] lead to $\gamma=2.5$ or, alternatively, even to some possibility that these networks are not scale-free), 4 is for the collaboration network of MEDLINE [21], 5 is for the collaboration network of movie actors [20], (also, $\gamma=2.3$ was obtained for this network in [6]) 6 is for incoming and outgoing links of the networks of the metabolic reactions [23]. The precision of the upper points is about ± 0.1 and is much worse for points in the dashed region.

In Fig. 3, in the log-linear scale, we present the sizes of all reported scale-free networks vs their γ exponent values. The plotted points are inside of the region restricted by the lines: $\gamma=2$, $\log_{10}t\sim 2.5(\gamma-1)$, and by the logarithm of the size of the largest scale-free network in nature—the World Wide Web—$\log_{10}t\sim 9$.

Our model demonstrates that the form of degree distribution is influenced by initial conditions even for large networks. Therefore, it is hard to obtain the values of the scaling exponents with high precision both from experimental data and simulations. One should note that including the aging of nodes, breaking of links, or disappearing of nodes suppresses the effect of the initial conditions and removes the hump (see the plots of the degree distributions in [16]).

In conclusion, using a simple model, we have described the size effect on degree distribution of scale-free growing networks. This cutoff and a trace of the initial conditions, a hump near the cutoff, impede observations of the power-law dependence even for large but finite networks. For large γ, such observations are impossible. If $\gamma\to 2$, then $k_{cut}\sim t$, so, in fact, the cutoff is absent.

The proposed model belongs to the class of the exactly solvable scale-free growing networks. One can consider another simple model. Instead of the connection of a new node with the ends of a randomly chosen link of the network, one may connect it each time with all three vertex nodes of a randomly chosen triangle of links. (Note that we forbid multiple links.) In other words, a new node connects with three random nearest neighbor nodes. Such a model has the same scaling exponents as the considered one.

Note added.—After this paper had been submitted, a work [24] with analytical expressions for size-dependent distributions of the Simon model was presented. Numerical calculation of these distributions was made in Ref. [25].

S.N.D. thanks PRAXIS XXI (Portugal) for a research Grant PRAXIS No. XXI/BCC/16418/98. J.F.F.M. was partially supported by Project No. POCTI/FIS/33141/99. We also thank A.-L. Barabási for providing us with the preprint [23] before publication.

[1] S. Redner, Eur. Phys. J. B **4**, 131 (1998).
[2] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **401**, 130 (1999).
[3] M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999).
[4] R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, in *Proceedings of the 25th International Conference on Very Large Data Bases (VLDB'99), Edinburgh, 1999* (Morgan Kaufmann, New York, 1999), pp. 639–650.
[5] B.A. Huberman and L.A. Adamic, Nature (London) **401**, 131 (1999).
[6] L.A.N. Amaral, A. Scala, M. Barthelemy, and H.E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11 149 (2000).
[7] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, in Proceedings of the 9th International World Wide Web, Amsterdam, 2000, (unpublished), pp. 309–320.
[8] B. Bollobás, *Random Graphs* (Academic Press, London, 1985).
[9] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **406**, 378 (2000).
[10] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000).
[11] D.S. Callaway, M.E.J. Newman, S.H. Strogatz, and D.J. Watts, Phys. Rev. Lett. **85**, 5468 (2000).
[12] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[13] H.A. Simon, Biometrika **42**, 425 (1955).
[14] H.A. Simon, *Models of Man* (Wiley, New York, 1957).
[15] A.-L. Barabási, R. Albert, and H. Jeong, Physica A **272**, 173 (1999).
[16] S.N. Dorogovtsev and J.F.F. Mendes, Phys. Rev. E **62**, 1842 (2000); Europhys. Lett. **52**, 33 (2000).
[17] S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000).
[18] P.L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000); P.L. Krapivsky and S. Redner, e-print cond-mat/0011094.
[19] S.N. Dorogovtsev and J.F.F. Mendes, Phys. Rev. E **63**, 056125 (2001).
[20] R. Albert and A.-L. Barabási, Phys. Rev. Lett. **85**, 5234 (2000).
[21] M.E.J. Newman, Proc. Natl. Acad. Sci. USA **98**, 404 (2001).
[22] L.A. Adamic and B.A. Huberman, Science **287**, 2115a (2000).
[23] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A-L. Barabási, Nature (London) **407**, 651 (2000).
[24] L. Kullmann and J. Kertész, e-print cond-mat/0012410.
[25] D.H. Zanette and S.C. Manrubia, e-print nlin.AO/0009046.