Structure of Growing Networks with Preferential Linking

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 10 April 2000)

The model of growing networks with the preferential attachment of new links is generalized to include initial attractiveness of sites. We find the exact form of the stationary distribution of the number of incoming links of sites in the limit of long times, $P(q)$, and the long-time limit of the average connectivity $\overline{q}(s,t)$ of a site s at time t (one site is added per unit of time). At long times, $P(q) \sim q^{-\gamma}$ at $q \to \infty$ and $\overline{q}(s,t) \sim (s/t)^{-\beta}$ at $s/t \to 0$, where the exponent γ varies from 2 to ∞ depending on the initial attractiveness of sites. We show that the relation $\beta(\gamma - 1) = 1$ between the exponents is universal.

PACS numbers: 84.35.+i, 05.40.–a, 05.50.+q, 87.18.Sn

It was observed recently that the distributions of several quantities in various growing networks have a power-law form. This scaling behavior was observed in the World Wide Web, in neural and social networks, in nets of citations of scientific papers, etc.; see [1–14], and references therein. These observations challenge us to find the general reasons of such behavior. It is only recently that scientists became aware of the ever increasing impact of various evolving networks on everyone's life. Earlier studies [15–20] concentrated on simple random networks, and it was recently discovered that many complex networks are hierarchically organized [5,6,21].

Mostly, the interest is concentrated on the distribution of shortest paths between the different sites of a network [1] and on the distribution of the number of connections with a site [2–6]. The second quantity is obviously simpler to obtain than the first one but even for it, in the case of the networks with scaling behavior, no exact results are known.

The only known mechanism of self-organization of a growing network into a free-scale structure is preferential linking [7–9], i.e., new links are preferentially attached to sites with high numbers of connections. A simple model of a growing network with preferential linking was proposed by Barabási and Albert [7] (BA model). At each time step a new site is added. It connects with old sites by a fixed number of links. The probability of an old site to get a new link is proportional to the total number of connections with this site. It was found in [7,8] that the distribution of the number of links has a power-law form at long times. The value of the corresponding scaling exponent, γ, obtained using a mean-field approach, equals 3. This value is close to that one observed in the network of citations [3], but other examples of evolving networks show different values of γ. Introduction of the aging of sites changes γ [22] and may even break the scaling behavior [13,22].

In the present Letter, we generalize the BA model and find the exact form of the distribution of *incoming* links of sites in the limit of large sizes of the growing network. We derive a scaling relation connecting the scaling exponent of the distribution of incoming links and the exponent of the temporal behavior of the average connectivity, and demonstrate that it applies for a large class of evolving networks.

The model.—At each time step a new site appears (see Fig. 1). Simultaneously, m new *directed* links coming out from nonspecified sites are introduced. Let the connectivity q_s be the number of incoming links to a site s, i.e., to a site added at time s. The new links are distributed between sites according to the following rule. The probability that a new link points to a given site s is proportional to the following characteristic of the site:

$$A_s = A + q_s \,, \tag{1}$$

thereafter called its *attractiveness*. All sites are born with some initial attractiveness $A \geq 0$, but afterwards it increases because of the q_s term. The introduced parameter A, the initial attractiveness, governs the probability for "young" sites to get new links.

We emphasize that we do not specify sites from which the new links come out. They may come out from the new site, from old sites, or even from outside of the network. Our results do not depend on that. Therefore, the model describes also the particular case when every new site is the source of all the m new links like in the BA model. In this case, every site has m outgoing links and the total

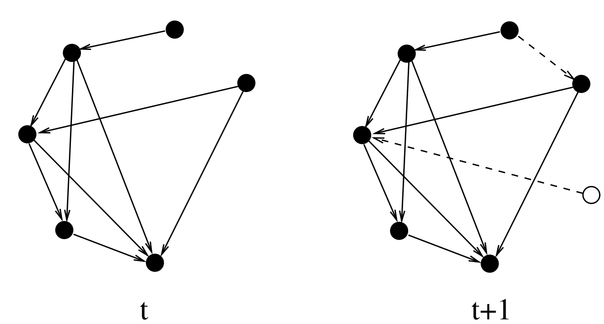

FIG. 1. Illustration of the growing network under consideration. Each instant a new site (open circle) and m (here, $m = 2$) new directed links (dashed arrows) are added. These links are distributed between the sites according to the rule introduced in the text.

number of its connections equals $q_s + m$. This number coincides with the attractiveness of the site, Eq. (1), if one sets $A = m$. In this case, our rule for the distribution of new links among sites coincides with the corresponding rule of the BA model. Hence, the model that we consider here is equivalent to the BA model in the particular case of the initial attractiveness equal to m.

In fact, our model may be mapped to the following general problem. Each instant, m new particles (i.e., incoming links) have to be distributed between an *increasing* number (one per time step) of boxes (i.e., sites) acording to the introduced rule.

The master equation.—Let us derive the equation for the distribution $p(q,s,t)$ of the connectivity q of the site s. At time t $(t = 1,2,\ldots)$ the network consists of t sites connected by $m(t-1)$ directed links (it is convenient to assume that initially $(t = 1)$ we have one site with m incoming links—the resulting behavior at long times is independent of the initial condition). The total attractiveness of the network at time t is $A_\Sigma = (m + A)t = (1 + a)m$, where $a \equiv A/m$. Note that one may allow multiple links, i.e., the connectivity of a given site may increase simultaneously by more than one. That is not essential for long times, when probability to receive simultaneously more than one link of the m is vanishing. The probability that a new link is connected with the site s equals A_s/A_Σ. The probability for the site s to receive exactly l new links of the m injected is $\mathcal{P}_s^{(ml)} = \binom{m}{l}(A_s/A_\Sigma)^l(1 - A_s/A_\Sigma)^{m-l}$. Hence, the connectivity distribution of a site obeys the following master equation:

$$p(q,s,t+1) = \sum_{l=0}^{m} \mathcal{P}_s^{(ml)} p(q-l,s,t) = \sum_{l=0}^{m} \binom{m}{l} \left[\frac{q-l+am}{(1+a)mt}\right]^l \left[1 - \frac{q-l+am}{(1+a)mt}\right]^{m-l} p(q-l,s,t). \quad (2)$$

Equation (2) is supplied with the initial condition $p(q,s,s) = \delta(q)$, which means that sites are born with zero connectivity (i.e., without incoming links in our definition).

The connectivity distribution of the entire network is $P(q,t) = \sum_{u=1}^{t} p(q,u,t)/t$. Summing up Eq. (2) over s from 1 to t, one gets

$$(t+1)P(q,t+1) - p(q,t+1,t+1) = \left(t - \frac{q+am}{1+a}\right)P(q,t) + \frac{q-1+am}{1+a}P(q-1,t) + \mathcal{O}\left(\frac{P}{t}\right). \quad (3)$$

At long times, we obtain

$$(1+a)t\frac{\partial P}{\partial t}(q,t) + (1+a)P(q,t) + (q+am)P(q,t) - (q-1+am)P(q-1,t) = (1+a)\delta(q). \quad (4)$$

Finally, assuming that the limit $P(q) = P(q, t \to \infty)$ exists, we get the following equation for the stationary connectivity distribution:

$$(1+a)P(q) + (q+ma)P(q) - (q-1+ma)P(q-1) = (1+a)\delta(q). \quad (5)$$

The stationary distribution.—To solve Eq. (5) one may use the Z transform of the distribution function:

$$\Phi(z) = \sum_{q=0}^{\infty} P(q) z^q. \quad (6)$$

Then one gets from Eq. (5)

$$z(1-z)\frac{d\Phi}{dz} + ma(1-z)\Phi + (1+a)\Phi = 1 + a. \quad (7)$$

The solution of Eq. (7) that is analytic at $z = 0$ has the following form:

$$\Phi(z) = (1+a)z^{-1-(m+1)a}(1-z)^{1+a}\int_0^z dx\,\frac{x^{(m+1)a}}{(1-x)^{2+a}} = \frac{1+a}{1+(m+1)a}\,{}_2F_1[1, ma; 2+(m+1)a; z], \quad (8)$$

where ${}_2F_1[\,]$ is the hypergeometric function. Using its expansion [23] in z, we obtain, comparing with Eq. (6),

$$P(q) = (1+a)\frac{\Gamma[(m+1)a+1]}{\Gamma(ma)}\frac{\Gamma(q+ma)}{\Gamma[q+2+(m+1)a]}, \quad (9)$$

that is our main result (see Fig. 2). In particular, when $a = 1$, that corresponds to the case $A_s = m + q_s$, studied in [7,8], we get

$$P(q) = \frac{2m(m+1)}{(q+m)(q+m+1)(q+m+2)}. \quad (10)$$

This expression in the limit $q \to \infty$ approaches the corresponding result of [7,8] obtained in the frames of an approximate scheme, but the prefactors are different. In fact, the "mean field" approach, used in [8], is equivalent to the continuous-q approximation in our discrete-difference equations. Indeed, if we replace the finite difference with a derivative over q, we get the expression obtained in [7,8].

For $ma + q \gg 1$, the distribution function (9) takes the form

$$P(q) \cong (1 + a) \frac{\Gamma[(m + 1)a + 1]}{\Gamma(ma)} (q + ma)^{-(2+a)}. \tag{11}$$

Therefore, we find the scaling exponent γ of the distribution function:

$$\gamma = 2 + a = 2 + A/m, \tag{12}$$

where A is the initial attractiveness of a site.

The distribution $p(q, s, t)$.—Let us find the connectivity distribution $p(q, s, t)$ for the site s. At long times $t \gg 1$, keeping only two leading terms in $1/t$ in Eq. (2), one gets

$$p(q, s, t + 1) = \left[1 - \frac{q + am}{(1 + a)t}\right] p(q, s, t) + \frac{q - 1 + am}{(1 + a)t} p(q - 1, s, t) + \mathcal{O}\left(\frac{p}{t^2}\right). \tag{13}$$

Assuming that the scale of time variation is much larger than 1, we can replace the finite t difference with a derivative

$$(1 + a)t \frac{\partial p}{\partial t}(q, s, t) = (q - 1 + am)p(q - 1, s, t) - (q + am)p(q, s, t). \tag{14}$$

Finally, using the Z transform in the similar way as before, we obtain the solution of Eq. (14), i.e., the connectivity distribution of individual sites:

$$p(q, s, t) = \frac{\Gamma(am + q)}{\Gamma(am)q!} \left(\frac{s}{t}\right)^{am/(1+a)} \left[1 - \left(\frac{s}{t}\right)^{1/(1+a)}\right]^q. \tag{15}$$

Hence, this distribution has an exponential tail. Now one may get also the expression for the average connectivity of a given site:

$$\overline{q}(s, t) = \sum_{q=0}^{\infty} q p(q, s, t) = am\left[\left(\frac{s}{t}\right)^{-1/(1+a)} - 1\right]. \tag{16}$$

Thus, at a fixed time t the average connectivity of an old site $s \ll t$ depends upon its age as $\sim s^{-\beta}$, where the exponent $\beta = 1/(1 + a)$. Therefore, we have the following relation between the exponents of the considered network:

$$\beta(\gamma - 1) = 1, \tag{17}$$

that was previously obtained in the continuous approximation [22].

We can show that Eq. (17) is universal and may be obtained from the most general considerations. In fact, we assume only that the averaged connectivity $\overline{q}(s, t)$ and the connectivity distribution $P(q)$ show scaling behavior. Then, in the scaling region, the quantity of interest, i.e., the probability $p(q, s, t)$, has to be of the following form: $p(q, s, t) = (s/t)^{\Delta_1} f[q^{\Delta_2}(s/t)^{\Delta_3}]$. Obviously, one can set $\Delta_2 = 1$. $\Delta_1 = \Delta_3$ because of the normalization condition for $p(q, s, t)$ at a fixed s, $\sum_{q=0}^{\infty} p(q, s, t) = 1$. Then, the relation $\overline{q}(s, t) \propto (s/t)^{-\beta}$ leads to $\Delta_1 = \Delta_3 = \beta$ [we use the definition (16)], and finally, inserting $p(q, s, t)$ in such a form into the relation $P(q) \propto q^{-\gamma}$ at large q and t, one gets the relation (17).

The network that we consider here belongs to the class of scale-free growing networks (we use the classification of growing networks presented in [13]). For the known real networks of this class (see the most complete description [13]), no data for the variation of the average connectivity of a site with its age are available yet. These data are necessary to obtain the exponent β. It would be intriguing to study this quantity in the real scale-free networks and to check Eq. (17). One should note that, for the network with aging of sites, the relation (17) was already confirmed by the simulation [22].

The particular form of the scaling function $f(\xi)$, $\xi \equiv q(s/t)^{-\beta}$, depends on the specific model of the growing network. In the case under consideration, it follows from Eq. (15) that

$$f(\xi) = \frac{1}{\Gamma(am)} \xi^{am-1} \exp(-\xi) \tag{18}$$

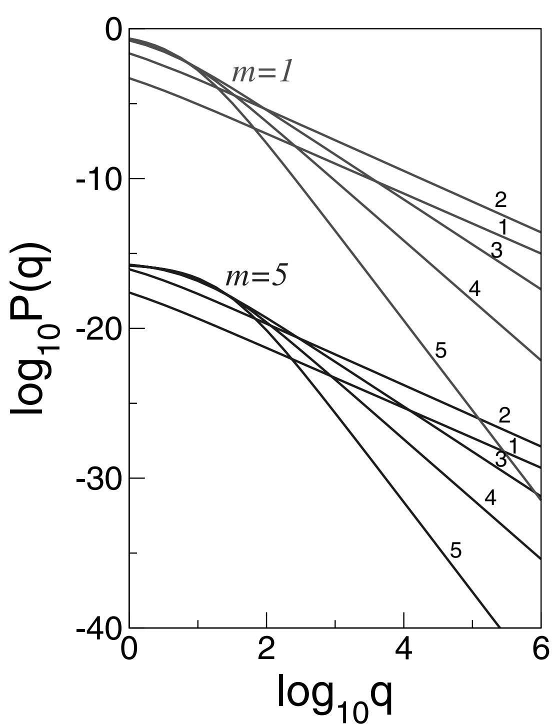

FIG. 2. Log-log plot of the distribution of the incoming links of sites for $m = 1$ and $m = 5$ (the curves for $m = 5$ are displaced down by 15). (1) $a = 0.001$, (2) $a = 0.05$, (3) $a = 1.0$ (BA model), (4) $a = 2.0$, (5) $a = 4.0$.

for $s/t \to 0$, $q \to \infty$ and the fixed $q(s/t)^{\beta}$. (Here, we use the asymptote: $\Gamma(am + q)/q! \to q^{am-1}$ at $q \to \infty$.)

Discussion.—In the limit of zero initial attractiveness of sites ($a = 0$) all the links lead to the first site since all others have no chance to get a new incoming link. In this case, Eqs. (12) and (17) give $\gamma = 2$ and $\beta = 1$. For $a = 1$, i.e., for the BA model, $\gamma = 3$ and $\beta = 1/2$ [7,8]. Finally, when $a \to \infty$, i.e., all sites have equal attractivity all the time, and the scaling breaks, one sees from Eqs. (12) and (17) that $\gamma \to \infty$ and $\beta \to 0$. In the last case, our system appears to be out of a class of networks with preferential linking. Note that the ranges of variation of γ, $2 < \gamma < \infty$, and β, $1 < \beta < 0$, are the same as for the scale-free network with aging of sites [22]. Note also that the observed value of the scaling exponent of the distribution of incoming links in the World Wide Web, 2.1 [5,9], is in this range.

We see that the approach of [7,8] based on the continuous-q approximation gives the proper values for the critical exponents (see also [22] for the network with aging of sites). Thus, this approximation is effective for calculation of the exponents of the scale-free networks.

A two-parameter fitting was proposed in [24] to describe the observed distribution for the citations of scientific papers [3]. One sees that the connectivity distribution of the considered growing networks is of quite different form. It seems that the difference occurs because we study the *growing* structure unlike the approach [24].

In conclusion, in the limit of long times (large sizes of the growing networks), we have found the exact solution of the master equation for the distribution of incoming links. This solution demonstrates the existence of the scaling region in a class of the growing networks with the preferential linking.

The considered growing networks are self-organized into free-scale structures. The input flow of the new links is distributed between the increasing number of sites. The scaling exponents are determined by the value of the initial attractiveness ascribed to every new site. Depending on this quantity, the scaling exponent γ of the connectivity distribution takes values from 2 to ∞. We have shown that the relation (17) between the scaling exponents γ and β is valid for a wide class of growing networks.

S. N. D. thanks PRAXIS XXI (Portugal) for Research Grant No. PRAXIS XXI/BCC/16418/98. J. F. F. M. was partially supported by FCT Sapiens Project No. 33141/99. We also thank E. J. S. Lage for reading the manuscript and A. V. Goltsev and Yu. G. Pogorelov for many useful discussions.

Note added.—After submission of this manuscript we learned of Krapivsky and co-workers' work [25] which overlaps some of our results.

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

[1] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[2] J. Laherrere and D. Sornette, Eur. Phys. J. B **2**, 525 (1998).
[3] S. Redner, Eur. Phys. J. B **4**, 131 (1998).
[4] B. A. Huberman, P. L. T. Pirolli, J. E. Pitkow, and R. J. Lukose, Science **280**, 95 (1998).
[5] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **401**, 130 (1999).
[6] B. A. Huberman and L. A. Adamic, Nature (London) **401**, 131 (1999).
[7] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[8] A.-L. Barabási, R. Albert, and H. Jeong, Physica (Amsterdam) **272A**, 173 (1999).
[9] R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, in *Proceedings of the 25th VLDB Conference, Edinburgh, 1999* (Morgan Kaufmann, Orlando, FL, 1999), pp. 639–650.
[10] M. Barthélémy and L. A. N. Amaral, Phys. Rev. Lett. **82**, 3180 (1999); Phys. Rev. Lett. **82**, 5180(E) (1999).
[11] M. E. J. Newman and D. J. Watts, Phys. Lett. A **263**, 341 (1999).
[12] A. Barrat and M. Weigt, Eur. Phys. J. B **13**, 547 (2000).
[13] L. A. N. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11 149 (2000).
[14] S. N. Dorogovtsev and J. F. F. Mendes, Europhys. Lett. **50**, 1 (2000).
[15] P. Erdös and A. Renýi, Publ. Math. Inst. Hung. Acad. Sci. **5**, 17 (1960).
[16] B. Bollobás, *Random Graphs* (Academic Press, London, 1985).
[17] G. Parisi, J. Phys. A **19**, L675 (1986).
[18] B. Derrida and Y. Pomeau, Europhys. Lett. **1**, 45 (1986).
[19] B. Derrida and D. Stauffer, Europhys. Lett. **2**, 739 (1987).
[20] R. Monasson and R. Zecchina, Phys. Rev. Lett. **75**, 2432 (1995).
[21] D. J. Watts, *Small Worlds* (Princeton University Press, Princeton, New Jersey, 1999).
[22] S. N. Dorogovtsev and J. F. F. Mendes, Phys. Rev. E **62**, 1842 (2000).
[23] H. Bateman and A. Erdélyi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 1.
[24] C. Tsallis and M. P. de Albuquerque, Eur. Phys. J. B **13**, 777 (2000).
[25] P. L. Krapivsky, S. Redner, and F. Leyvraz, preceding Letter, Phys. Rev. Lett. **85**, 4629 (2000).