Growing network with heritable connectivity of nodes

S.N. Dorogovtsev\textsuperscript{1,2,*}, I.F.F. Mendes\textsuperscript{1,†}, and A.N. Samukhin\textsuperscript{2,‡}

\textsuperscript{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

\textsuperscript{2} A.F. Ioffe Physico-Technical Institute, 194021 St. Petersburg, Russia

We propose a model of a growing network, in which preferential linking is combined with partial inheritance of connectivity (number of incoming links) of individual nodes by new ones. The non-trivial version of this model is solved exactly in the limit of a large network size. We demonstrate, that the connectivity distribution depends on the network size, \( t \), in a \textit{multifractal} fashion. When the size of the network tends to infinity, the distribution behaves as \( \sim \exp(-\gamma \ln t) \), where \( \gamma = \sqrt{2} \). For the finite-size network, this behavior is observed for \( 1 < q \lesssim \exp(\ln^{1/2} t) \) but the multifractality is determined by the far wider part, \( 1 < q \lesssim \sqrt{t} \), of the distribution function.

Hierarchically organized networks play an outstanding role in Nature. The well known examples are the World Wide Web, scientific citations of papers, networks of people’s personal relations, neural networks, etc. (see, e.g., \cite{1–8}). Most of these networks belong to the class of \textit{small world ones}, i.e., their diameters are proportional to the logarithm of the network sizes \cite{1}. An important characteristic of a network is its connectivity distribution function, where the connectivity is the number of connections of a node (vertex degree, in the language of the graph theory). Properties of networks depend dramatically on the form of this distribution. It has been realized recently, that growing networks with power-law connectivity distributions, called \textit{scale-free networks}, are of special importance \cite{2}. In particular, they are resilient to random breakdowns \cite{3,11}. A natural way to obtain a scale-free growing network is provided by the mechanism of \textit{preferential linking} \cite{3}. This principle is similar to the one introduced in the well known Simon’s model \cite{12}, used to explain power-law distributions in various social and economic systems. In fact, all these models belong to the class of stochastic multiplicative processes \cite{3}. New links are attached preferentially to nodes of a growing network with a high number of connections. Several types of preferential linking were proposed, which produced the \( \gamma \) exponent of the connectivity distribution, \( \Pi(q) \sim q^{-\gamma} \) in the range \( (2, \infty) \) \cite{3,14,16}. Recently a model with \( \gamma < 2 \) was also proposed \cite{10}.

What are the scenarios of the evolution of networks produced by preferential linking? What are the arising connectivity distributions? These questions were considered in several recently published papers \cite{14,16}. However, in all the previous models, where the idea of preferential linking was used, it was assumed, that new nodes appear with the same properties, independent of the state of network at this moment, i.e., all nodes are born equal. One may say, that new nodes are created by some invariable external source. Here we put forward a different concept: \textit{new nodes are born with random properties, which reflects the state of the network at the moment of birth.}

In this respect, one can say, that they are created by the network itself. Of course, various realizations of this idea are possible. For example, in \cite{15}, an evolutionary model of such a type was proposed, but without preferential linking. No power law distributions were found there.

Such inheritance of properties seems to be a rather usual feature of networks. Let us discuss briefly, how it can arise in, e.g., networks of scientific citations. In these networks, new nodes (scientific papers) arise not in empty air. Each of them has its predecessors (e.g., some of the previous papers of the same author or some papers on the same topic, etc.), and inherits a part of their attractiveness. Roughly speaking, each paper is condemned to popularity or oblivion already at the instant of its birth, and its future depends on its direct predecessors. Similar inheritance may be essential in collaboration networks, in networks of relations between firms in economy, etc. Of course, a particular mechanism of the inheritance depends on a particular network, but the idea of the inheritance is natural.

In the present Letter, we propose a model of a growing network with \textit{directed} links, in which preferential linking is combined with partial inheritance of connectivity of individual nodes by new ones. It is a simple example, demonstrating features of the network growth with such a combination of factors. The model is solved exactly in the limit of large network sizes. This means that we do not pass to the continuous limit of connectivity but use the \textit{discrete} \( q \) version of the model. Such an approach lets us to study features of the connectivity distribution which are invisible within continuous approximation.

In the model, during the network growth, many nodes appear and stay without any incoming connections. After exclusion of these nodes from the statistics, the average connectivity (here it is a number of \textit{incoming} links) of the remained part of the network grows by a power law. The connectivity distribution becomes a \textit{multifractal} one. The body of the distribution \( (q \lesssim \exp(\ln^{1/2} t)) \), which includes most of the nodes, is of the form, \( \Pi(q) \sim q^{-\gamma} \ln t \), where \( \gamma = \sqrt{2} < 2 \). The first moment of connectivity
is divergent, that means, that large fraction of links is concentrated “in hands” of a small fraction of the nodes. Moreover, all the moments (except zero-th one) are determined by the behavior of the partition function in the region, \( \exp[\ln^{1/2} t] \lesssim q \ll \sqrt{t} \). Here the function is size-dependent but independent of the initial conditions. Its form ensures that the moments scale with the network size with exponents, nonlinearly dependent on their order. For \( q \gtrsim \sqrt{t} \), the connectivity distribution is exponential.

**The model.**—At each increment of time, \( dt \), a new node is added with probability \( dt \). At time \( t \), the system consists of \( N(t) \) nodes, \( N(t) = 1 + O(t^{1/2}) \). We assume that a new node is born with a random number of incoming links, \( qN(t) \), which is distributed according to some time-dependent distribution function, \( \Pi_i(t, qN) \). At the same time, a link between sites is created with probability \( m dt \). Thus, the connectivity distribution function of connectivity \( \Pi(t, q) \) is defined as:

\[
\Pi_i(t, q) = \left\langle \frac{1}{N} \sum_{j=1}^{N} \delta[q_j(t) - q] \right\rangle - \frac{1}{t} \left\langle \sum_{j=1}^{N} \delta[q_j(t) - q] \right\rangle .
\]

(1)

Here, the average is over realizations of the process. Its infinitesimal variation, \( d\Pi = d_\alpha \Pi + d_t \Pi \), consists of two parts: \( d_\alpha \Pi = [(dt/t) \Pi_i(t, q) - \Pi(t, q)] \) arising from addition of new nodes and \( d_t \Pi = [m dt/tq(t)] [q - q \Pi(t, q)] \) originating from creation of new links. Therefore, we get the following master equation:

\[
t \frac{\partial \Pi(t, q)}{\partial t} + \Pi(t, q) + \frac{m}{q(t)} [q \Pi(t, q) - (q - 1) \Pi(t, q - 1)] \\
= \Pi_i(t, q) .
\]

(2)

This equation is similar to the one used in [4], apart the term \( \delta(q - q_0) \) on the rhs replaced by \( \Pi_i(t, q) \). Unlike the original Barabási-Albert’s model [5], a new node is appeared not with predefined initial connectivity \( q_i = \text{const} \), but \( q_i \) is assumed to be a random number, with a time-dependent probability distribution function \( \Pi_i(t, q) \).

We assume that \( \Pi_i \) is some functional of the network properties, in particular, on the connectivity distribution function, \( \Pi \). In this respect, new sites are not created by some external source, but are born by the network itself. We propose the following inheritance rule for new nodes. We assume, that every new node is born by some randomly chosen old one. At the moment of birth it “inherits” (copies), in average, a part \( c, 0 < c < 1 \), of its parent’s connectivity. More precisely, with probability \( c \), every of \( k \) incoming links of a parent creates a link, pointing at its heir. The parameter \( c \) is, in turn, assumed to be a random number, distributed with probability density \( h(c) \). Hence,

\[
\Pi_i(t, q) = \int_0^1 dc \frac{h(c)}{c} \sum_{k=q}^{\infty} \frac{k}{c} (1 - c)^{k-q} \Pi(t, k) .
\]

(3)

From Eqs. (2) and (3) one can easily obtain:

\[
\int_0^1 dc \frac{h(c)}{c} \frac{\partial \Phi(t, y)}{\partial y} = 0 .
\]

(4)

The initial condition is \( \Phi(t_0, y) = \Phi_0(y) \), \( t_0 \gg 1 \) (this equation is valid for \( t \gg 1 \)). After rescaling of the size variable \( t \rightarrow t/t_0 \) the initial condition becomes \( \Phi(1, y) = \Phi_0(y) \).

**Multifractality.**—The connectivity distribution in our model is of a multifractal type. Indeed, the moments of the distribution may be expressed as:

\[
\mathbb{E}[q^n] = \mathbb{E}[\Pi(t, q) q^n] = \mathbb{E}[\Pi(t, q)] \mathbb{E}[q^n] .
\]

One can easily derive the equation below, that their dependence on network size \( t \) is \( \mathbb{E}[q^n] \sim A_{nn} t^n + A_{n,n-1} t^{n-1} + \ldots \), and \( \tau(n) = (1 - c) n + \mathbb{E}[q] - 1 \), \( \mathbb{E}[q] = \sum_{q=1}^{\infty} q \Pi(t, q) \). The coefficients \( A_{nn} \) depend on the initial distribution. At long times \( t \gg \tau(n) \), since \( \tau(n) \) is a growing function for \( n > 1 \). When \( \tau(n) \) is not a linear function of \( n \), this type of size-dependent distribution is called multifractal [6]. (For a pure fractal distribution it would be \( M_n(t) \sim t^{\tau(n)} \), where \( D_f \) is the dimension of a fractal.) This function, \( \tau(n) \), encodes the set of generalized dimensions of the multifractal, \( D_f(n) = \tau(n) / n \).

Also, one may describe a multifractal distribution as a statistical mixture of fractal ones with “dimensionalities”, \( \alpha = dr/dn \). These fractal distributions have supports (sets of sites with non-zero connectivity) with dimensionalities, \( \alpha = f(\alpha) \). Where \( \alpha, c = \text{Legendre transform of } \tau(n), \tau(n) + f(\alpha) = 0 \). In other words, every fractal distribution enters with a statistical weight, scaling with network size as \( t^{f(\alpha)} \).

For example, if the distribution of \( c \) is uniform, \( h(c) = \Theta(c) \Theta(1 - c) \), we obtain for our network: \( \tau(n) = n/2 - n/(n + 1) \); \( D_f(n) = n/[2(n + 1)] \), \( \alpha = 1/2 - 1/(n + 1)^2 \) and \( f(\alpha) = (1 - \sqrt{1/2 - \alpha^2}) \) with \( -\infty < \alpha < 1/2 \).
its “succession rights”. Here, we choose this rule as: $h(c) = \Theta(c\Theta(1-c)$, i.e., $c$ is assumed to be uniformly distributed within the interval $(0,1)$. In this nontrivial and important case we are able to find the exact solution of the problem. Indeed, for the homogeneous $h(c)$, after application of the operator $\partial_y(y)$, Eq. (3), may be reduced to a linear partial differential equation. This equation, after the Mellin’s transformation with respect to time, $\psi(y, q) = \int_1^\infty dt t^{q-1}\Phi(t, y)$, takes the form,

$$y^2(1-y) \partial_y^2 \psi + y(2y - 3q) \partial_y \psi + 2\eta \psi = -2\partial_y(y\Phi_0),$$

(5)

where $\Phi_0(y)$ is the initial distribution in the $z$-representation. In the following we shall consider the case $\Phi_0(y) = (1-y)^k$, that is $\Pi(1, q) = \delta(q-k)$.

The distribution, obtained with such initial condition, is denoted as $\Pi(t; q, k)$, and its Mellin’s-time-transform

$$\phi_1(\zeta, k) = \frac{4k\Gamma(\zeta\Gamma(2+\zeta)/(1-\zeta)}{4\zeta + (1+\zeta)} \int_0^1 dz z^{2-2/(1+\zeta)} (1-z)\exp(-\frac{2(\zeta-1)}{\Gamma(2+\zeta)}) (1-z)^{k-1} F(1 - \frac{2}{1+\zeta}, 1 - \frac{2}{1+\zeta}; \eta + \frac{2}{1+\zeta}; z),$$

(7)

$$\phi_2(\zeta, q) = \frac{\Gamma(1+\zeta + \frac{2}{1+\zeta})}{\Gamma(2+\zeta) - \zeta} \sin \pi \zeta \int_0^\infty dy \sin^2(1+y)^{-q} F(1, 2+\zeta; 2+\zeta - \frac{2}{1+\zeta}; -y).$$

(8)

Here $F$ is the hypergeometric function, $2F1$.

Our main goal is to obtain $\Pi(t; q, k)$, at long $t$. One can calculate $\Pi$, the inverse Mellin’s-transform of $\Psi$, using the saddle point approximation replacing the integration variable $\eta$ with $\zeta(\eta) = \zeta$. At long $t$ the saddle point $\zeta_c$ is close to its position at $t = \infty$, $\zeta_c = \sqrt{2} - 1$, $\eta = \zeta(\zeta_c) = 0$. To avoid the situation when the integrand is zero in the saddle point, one may integrate the inverse Mellin’s-transform by parts. Then, in the saddle point approximation at $t = \infty$, we obtain the following relation,

$$\Pi(t; q, k) \approx \frac{t^{-3/2+\sqrt{2}}}{2^{1/4}\sqrt{\pi\ln^{1/2}t}} \frac{\partial \phi_1(\zeta, k) \partial \phi_2(\zeta, q)}{\partial \zeta} \mid_{\zeta = \sqrt{2} - 1},$$

(9)

for $q, k > 0$.

Note, however, that, if $t$ is finite, this expression becomes invalid for large enough $q$, since in this case, $\phi_2 \sim q^{-\zeta}$ varies essentially within the saddle point maximum. The region of validity of Eq. (10) may be estimated as $\ln q \ll \ln^{1/2}t$. At larger $q$, the saddle point position becomes $q$-dependent. For $1 \ll \ln q \lesssim (1/2)\ln t$, the saddle point position remains close enough to $\zeta_\infty$, and we obtain:

$$\Pi(t; q, k) \approx$$

---

as $\Psi(q; q, k)$. Eq. (11) may be reduced to an inhomogeneous hypergeometric one after the substitution, $\psi(y) = y^2 \chi(y)$, where $\zeta$ is one of the root of the characteristic equation: $\zeta^2 - (1 - 2q) \zeta + 2\eta = 0$. Here we present the result in terms of $\Psi(q; q, k)$, which is the $q$-th term of Taylor’s series of $\psi(y)$ around the point $y = 1$. After lengthy calculations, we obtain

$$\Psi(q; q, k) = -\phi_1(\zeta_1, k) \Phi_2(\zeta_1, q) = \delta(q - k),$$

(6)

where $\zeta_1$ is the root of the characteristic equation, which is positive for $\Re \eta < 0$, $\zeta_2$ is the other root, and functions $\phi_{1,2}$ may be expressed as:

$$d \frac{\ln(aq)}{(t \ln t)^{3/2}} \exp \left[\sqrt{2 \ln t \ln(t/q^2)} \frac{\partial \phi_1(\zeta, k)}{\partial \zeta} \right]_{\zeta = \sqrt{2} - 1},$$

(10)

where: $d = 0.174 \ldots$ and $a = 0.940 \ldots$

One can see from Eq. (10), that at long times, the fraction of zero-connectivity nodes is

$$\Pi(t; 0, k) \approx 1 + t^{-3/2+\sqrt{2}} \frac{\partial \phi_1(\zeta, k)}{\partial \zeta} \mid_{\zeta = \sqrt{2} - 1}. $$

(11)

One can see from Eq. (11), that at long times, the fraction of zero-connectivity nodes tends to 1. These nodes do not have incoming links but only outgoing ones (remind, that connectivity, in the present paper, is defined as the number of incoming links). They are passive constituents of the network, i.e., their connectivity remains unchanged all the time. Although the fraction of active nodes (with non-zero connectivity) tends to zero as the network grows, their total number, increases with time as $t^{\sqrt{2}-1/2}/\ln^{3/2}t$. One can introduce the distribution function of active nodes, $\Pi_1(t; q, k)$. It follows from Eqs. (11) and (11), that in the large time limit, it tends to the distribution, $\Pi_1(q)$, which depends neither on time nor on the initial conditions:

$$\Pi_1(t; q, k) \equiv (1 - \delta_q) \frac{\Pi(t; q, k)}{1 - \Pi(t; 0, k)} \rightarrow$$

---

3
\[
\Pi_1(q) = \frac{d}{q^{2}} \ln(aq) .
\] (13)

A plot of \( \Pi_1(q) \) is presented at Fig. 1. One can obtain from Eqs. (12) and (8) the following asymptotic expression for large \( q \):

\[
\Pi_1(q) = \frac{1}{\Gamma(k) t^{3/2}} \left( \frac{q-k}{t} \right)^k \exp\left(-\frac{q}{\sqrt{t}}\right) .
\] (15)

Apart from the logarithmic factor, this is a power law distribution. The exponent is less than 2. Indeed, the first moment of \( \Pi_1(t, q) \) (average number of links per active node) diverges at \( t \to \infty \). At finite \( t \), Eq. (13) becomes invalid when \( \ln q \gtrsim \ln^{1/2} t \). One can replace Eq. (13) with more general expression, which follows from Eqs. (14) and (11):

\[
\Pi_1(t, q) = dt^{-3/2} \ln(aq) \exp \left( \frac{\sqrt{2} \ln t \ln (t/q^2)}{2} \right) .
\] (14)

Here, \( 1 \ll q \ll k\sqrt{t} \). Note, that the distribution in this region depends on the network size but not on the initial condition. Eq. (13) follows from Eq. (14) in the appropriate region of \( q \). Dependence on the initial connectivity distribution appears when \( q \gtrsim k\sqrt{t} \).

One can also consider the limit, \( q \to \infty \), keeping \( t \) fixed. This determines the form of the large-connectivity cut-off of the scaling distribution. For \( q \approx k\sqrt{t} \gg 1 \), the saddle point of the inverse Mellin’s transform integral is shifted to the region \( \eta < 0, |\eta| \gg 1 \). Evaluation of the integral yields

\[
\Pi(t; q, k) \approx \frac{1}{\Gamma(k) t^{3/2}} \left( \frac{q-k}{\sqrt{t}} \right)^k \exp\left(-\frac{q}{\sqrt{t}}\right) .
\] (15)

for \( q \gtrsim k\sqrt{t} \). The estimation of the total number of nodes for which the connectivity values are within this tail, \( t_0 \sum_{q=k\sqrt{t}}^{q_0} \Pi(t; q, k) \), gives a value of the order of \( t_0 \), the number of nodes in the initial state of the network. The characteristic scale of the \( q \) variation, \( \sqrt{t} = t^{1-\epsilon} \), can be obtained simply in the continuous \( q \) limit of Eqs. (13) and (11) where the natural variable arises, \( q / t^{1-\epsilon} = q / \sqrt{t} \).

Conclusions.—We have found the exact large-size solution of a growing network model, which exhibits power-law type behavior, \( \Pi(q) \sim q^{-\gamma} \ln q \), \( 1 < \gamma < 2 \). In our model, \( \gamma = \sqrt{2} \). With probability close to 1, a randomly chosen node has its connectivity within the power law dependence region. In a recent paper [9], a number of real networks with \( \gamma < 2 \) were presented. Note that the connectivity distributions with \( \gamma < 2 \) were obtained analytically and by simulation for networks with accelerating growth [10]. In such networks, the average connectivity grows with the network size. In the present case, the number of links per active site also grows with time in a power law manner, i.e., as \( t^{3/2-\sqrt{2}} \ln^{3/2} t \). We have found that the characteristic feature of evolving networks with inheritance is the multifractality of the connectivity distribution. In our network, this intriguing property originates from the wide region, \( \ln^{1/2} t \lesssim q \lesssim k\sqrt{t} \) of the connectivity distribution.

SND thanks PRAXIS XXI (Portugal) for a research grant PRAXIS XXI/BCC/16418/98. JFFM was partially supported by the project FCT-Sapiens 33141/99. We also thank A.V. Goltsev and B.N. Shalaev for many useful discussions. Note added.—After this manuscript had been prepared, a paper of Bianconi and Barabási [23] has appeared, in which the multiscale behavior and the connectivity distribution with logarithmic correction, \( \Pi(q) \sim q^{-\gamma} \ln q, \gamma > 2 \), was found for another model of growing networks.

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

[1] D.J. Watts and S.H. Strogatz, Nature (London) 393, 440 (1998).
[2] S. Redner, Eur. Phys. J. B 4, 131 (1998).
[3] B.A. Huberman, P.L.T. Pirolli, J.E. Pitkow and R.J. Lukose, Science 280, 95 (1998).
[4] R. Albert, H. Jeong and A.-L. Barabási, Nature 406, 130 (1999).
[5] A.-L. Barabási and R. Albert, Science 286, 509 (1999).
[6] L.A.N. Amaral, A. Scala, M. Barthelemy and H.E. Stanley, Proc. Natl. Acad. Sci. U.S.A. 97, 11149 (2000).
[7] M.E.J. Newman, cond-mat/0007214.
[8] G. Caldarelli, R. Marchetti, and L. Pietronero, cond-mat/0001758, to appear in Europhys. Lett.
[9] R. Albert, H. Jeong and A.-L. Barabási, Nature, 406, 378 (2000).
[10] D.S. Callaway, M.E.J. Newman, S.H. Strogatz and D.J. Watts, cond-mat/0007306.
[11] R. Cohen, K. Erez, D. ben-Avraham and S. Havlin, cond-mat/0007048.
[12] Simon H.A., Biometrika, 22, 425 (1955).
[13] S. Solomon and M. Levy, Int. Mod. Phys. C 7, 745 (1996); D. Sornette and R. Cont, J. Phys. I (Paris) 7, 431 (1997).
[14] H. Takayasu, A.-H. Sato, and M. Takayasu, Phys. Rev. Lett. 79, 966 (1997); M. Marsili, S. Maslov, and Y.-C. Zhang, Physica A 253, 403 (1998).
[15] S.N. Dorogovtsev, J.F.F. Mendes and A.N. Samukhin, Phys. Rev. Lett. 85, 20 November (2000); S.N. Dorogovtsev and J.F.F. Mendes, Phys. Rev. E 62, 1842 (2000); Europhys. Lett. 52, 33 (2000).
[16] P.L. Krapivsky, S. Redner and F. Leyvraz, Phys. Rev. Lett. 85, 20 November (2000).
[17] S.N. Dorogovtsev and J.F.F. Mendes, cond-mat/0000065.
[18] R. Albert and A.-L. Barabási, cond-mat/0005085.
FIG. 1. Log-log plot of the stationary distribution, $\Pi_1(q)$, of incoming links. The circles represent the exact result, Eq. (12). The asymptotic dependence, Eq. (13), is shown by the line.

[19] B.B. Mandelbrot, J. Fluid Mech., 62, 331 (1974)
[20] H.G.E. Hentschel and I. Procaccia, Physica D, 8, 835 (1983).
[21] T.C. Halsey, M.H. Jensen, L.P. Kadanoff, I. Procaccia and B.I. Shraiman, Phys. Rev. A, 33, 1141 (1986).
[22] G. Bianconi and A.-L. Barabási, cond-mat/0011029.