Scale-Free Networks from Varying Vertex Intrinsic Fitness

G. Caldarelli,[1] A. Capocci,[2] P. De Los Rios,[3,4] and M. A. Muñoz[5]

[1]*INFM UdR ROMA1 Dipartimento Fisica, Università di Roma "La Sapienza," Piazzale Aldo Moro 2 00185, Roma, Italy*
[2]*Département de Physique, Université de Fribourg-Pérolles, CH-1700 Fribourg, Switzerland*
[3]*Institut de Physique Théorique, Université de Lausanne, CH-1004 Lausanne, Switzerland*
[4]*INFM UdR Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy*
[5]*Instituto de Física Teórica y Computacional Carlos I, Universidad de Granada, Facultad de Ciencias, 18071-Granada, Spain*

(Received 15 July 2002; published 3 December 2002)

A new mechanism leading to scale-free networks is proposed in this Letter. It is shown that, in many cases of interest, the connectivity power-law behavior is neither related to dynamical properties nor to preferential attachment. Assigning a quenched fitness value x_i to every vertex, and drawing links among vertices with a probability depending on the fitnesses of the two involved sites, gives rise to what we call a good-get-richer mechanism, in which sites with larger fitness are more likely to become hubs (i.e., to be highly connected).

DOI: 10.1103/PhysRevLett.89.258702 PACS numbers: 89.75.Hc, 89.75.Da

Complex networks [1–3] are attracting much interest as one of the most promising tools to describe a large variety of social [4], biological [5,6], and technological systems, as the Internet [7–10] or the World Wide Web (WWW) [11]. Networks are abstract mathematical objects composed by vertices (sites) connected by arcs (links). In the aforementioned examples, vertices can represent people, proteins, species, routers, or html documents, while arcs correspond to acquaintances, physical interactions, predation relationships, cable connections, or hyperlinks, respectively. In recent developments, scale-free (SF) networks have emerged in many different contexts, as the WWW, the Internet, Email and scientific-citation networks, protein and gene interaction networks, etc., and have become paradigmatic [2,3]. In all these examples, the degree k of a vertex, i.e., the number of arcs linking it to other vertices, is power-law distributed, $P(k) \sim k^{-\gamma}$. SF networks also present the so-called *small-world* phenomenon [12], that is, by a few selected jumps (that can be either short or long range) it is possible to reach very different regions of the system and apparently distant environments.

To understand how SF networks arise, the concepts of *growing networks* and of *preferential attachment* have been introduced [2]. In particular, in the best known SF network model, i.e., the Barabási-Albert (BA) one, networks grow at a constant rate (modeling the fact that new Web pages are continuously created, new proteins emerge by mutation, and so forth), and new vertices are attached to older ones with a probability $p(k)$ which is a (linearly [13]) growing function of the number of preexisting links, k, at every site. In this way, a highly connected vertex is more likely to receive further links from newly arriving vertices: This is the so-called "rich get richer" rule.

In some other, recently proposed, models of protein interaction networks [14,15] and of the WWW, new vertices (proteins and Web pages, respectively) are added by copying (replicating) existing vertices, borrowing some of their links, and adding some new others. It has been shown that this mechanism leads also to an *effective* preferential attachment mechanism.

Yet, although in some contexts preferential attachment can be a very reasonable assumption, in many others it is certainly not. In particular, in some situations, the information about the degree of each and every single vertex is not available to newly added sites, neither in a direct nor in an effective way. Instead, it is reasonable that two vertices are connected when the link creates a mutual benefit (here we restrict ourselves to bidirectional links) depending on some of their intrinsic properties (authoritativeness, friendship, social success, scientific relevance, interaction strength, etc.). Therefore, it is reasonable to expect that for some of these systems the $P(k)$ scale-free behavior (when existing) has an origin unrelated to preferential attachment.

In order to explore this simple idea, we propose the following network-building algorithm: (i) Start by creating a total (large) number N of vertices. At every vertex i a fitness x_i, which is a real number measuring its importance or rank, is assigned. Fitnesses are random numbers taken from a given probability distribution $\rho(x)$. (ii) For every couple of vertices, i, j, a link is drawn with a probability $f(x_i, x_j)$ (f a symmetric function of its arguments) depending on the "importance" of both vertices, i.e., on x_i, x_j. Some remarks are in order before proceeding further.

(i) The concept of "vertex-importance" or fitness has been already introduced successfully in the field of complex networks, but as an additional ingredient on top of the BA network [16]. Contrarily, here we put the emphasis on fitness itself, by eliminating the preferential attachment rule. (ii) A trivial realization of the above rules is the standard Erdős-Rényi model [17], where $f(x_i, x_j)$ is constant and equal to p for all vertex couples.

This particular choice does not produce SF networks, but in what follows we will show that other realizations of the general rules do so. (iii) The model, as defined, is static, but it can straightforwardly be considered a dynamic one by adding new vertices at every time step and linking them to the existing ones according to the above attaching rule. (iv) It is also easy to generalize the model to include asymmetric or directed links. (v) A somehow similar static model was studied by Goh *et al.* [18].

A general expression for $P(k)$ can be easily derived. Indeed, the mean degree of a vertex of fitness x is simply

$$k(x) = N \int_0^\infty f(x, y)\rho(y)dy = NF(x) \tag{1}$$

[with $x_i \in (0, \infty)$]. Assuming $F(x)$ to be a monotonous function of x, and for large enough N, we have the simple relation,

$$P(k) = \rho\left[F^{-1}\left(\frac{k}{N}\right)\right]\frac{d}{dk}F^{-1}\left(\frac{k}{N}\right). \tag{2}$$

For finite values of N corrections to this equation emerge [19]. As a particular example, consider $f(x_i, x_j) = (x_i x_j)/x_M^2$, where x_M is the largest value of x in the network. Then

$$k(x) = \frac{Nx}{x_M^2}\int_0^\infty y\rho(y)dy = N\frac{\langle x\rangle x}{x_M^2}, \tag{3}$$

and we have the simple relation,

$$P(k) = \frac{x_M^2}{N\langle x\rangle}\rho\left(\frac{x_M^2}{N\langle x\rangle}k\right). \tag{4}$$

A particularly simple realization of the model emerges if we consider power-law distributed fitnesses. This choice can be naturally justified by arguing that power laws appear rather generically in many contexts when one ranks, for example, people according to their incomes or cities according to their population, etc. This is the so-called Zipf law which establishes that the rank $R(x)$ behaves as $R(x) \propto x^{-\alpha}$ in a quite universal fashion [20]. The reason for the ubiquitous presence of the Zipf law yields on the multiplicative nature of the intrinsic fluctuations which generically leads to flat distributions in logarithmic space and, consequently, to power laws [20].

Clearly, if $\rho(x) \sim x^{-\beta}$ (Zipf's behavior, with $\beta = 1 + 1/\alpha$ [20]) then, using Eq. (4), also the degree distribution $P(k)$ is a power law and the network shows SF behavior. In Fig. 1, we show the degree distributions from simulations with $\beta = 2.5, 3, 4$ (corresponding to Zipf exponents $\alpha = 2/3, 1/2, 1/3$); the asymptotic behavior is, in all cases, well described by Eq. (4). This result is hardly surprising: From SF fitnesses we generate SF networks, but still it provides a new generic path to SF networks and takes into account the widespread occurrence of the Zipf's behavior in nature. In order to extend this result and check whether SF networks can be generated even when $\rho(x)$ is not SF itself, we consider an exponential distribution of fitnesses, $\rho(x) = e^{-x}$ (representing a random, Poisson distribution)

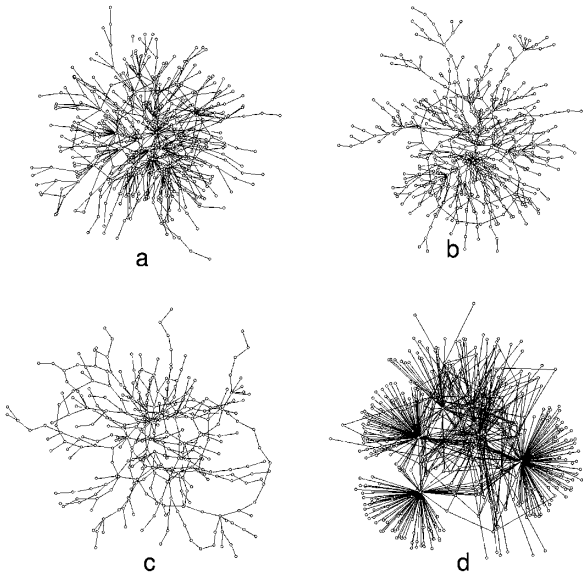

FIG. 1. Graph representations of four networks produced, respectively, with (a) $\beta = 2.5$, (b) $\beta = 3.0$, (c) $\beta = 4.0$, and (d) $\rho(x) = e^{-x}$ with a threshold rule. Graphs have been produced with the Pajek software.

and $f(x_i, x_j) = \theta[x_i + x_j - z(N)]$, where $\theta(x)$ is the usual Heaviside step function. This represents processes where two vertices are linked only if the sum of their fitnesses is larger than a given *threshold* $z(N)$. Using these rules, we obtain analytically (and confirm in computer simulations) that $P(k) \sim k^{-2}$ [21]. This leads to the nontrivial result that *even non-scale-free fitness distributions can generate scale-free networks* (see Fig. 2). Also different implementations of the threshold rule, such as $f(x_i, x_j) = \theta[x_i^n + x_j^n - z(N)^n]$ (where n is an integer number) give rise to the same inverse square behavior (although, in some cases, with logarithmic corrections).

In a future publication we will explore, in a more systematic way, the necessary and sufficient conditions for the fitness distribution and attaching rule under which well-behaved SF networks are generated.

Let us stress that the model, as defined, has a diverging average connectivity in the large N limit, as can be easily inferred from Eq. (1); i.e., it is severely *accelerated* [22]. Nevertheless, we can introduce in a rather natural way an upper cutoff accounting for the fact that every site has limited information on the rest of the world and, therefore, connection is attempted with a finite number, m, of different sites. Alternatively, vertices can be linked with the above rule and, after that, links are kept with probability p (so that, for example, $pN = m$). By including this modification, the N factor in Eq. (1) is substituted by m, and the connectivity is finite in the thermodynamic limit. In order to generate different accelerated networks (with the averaged connectivity not reaching a stationary value but growing with N in different possible ways [22]), other selection rules can be easily implemented.

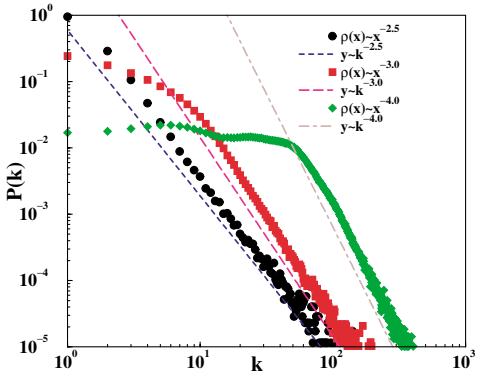

FIG. 2 (color online). Degree distribution for networks generated using $\rho(x) \sim x^{-\beta}$ with $\beta = 2.5$, 3, 4, and $f(x_i, x_j) = (x_i x_j)/x_M^2$; power laws with the corresponding analytical values are explicitly drawn in straight lines. Data obtained for networks with 10^4 vertices, averaged over 100 realizations.

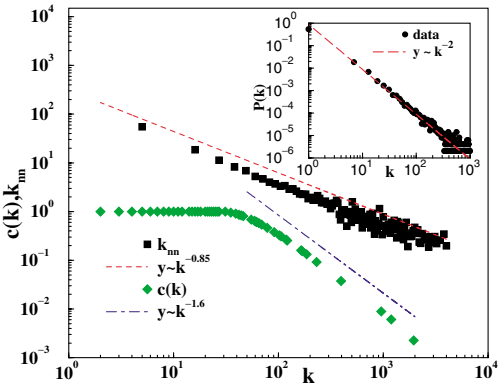

FIG. 3 (color online). Average neighbor connectivity k_{nn} against k, for networks generated using $\rho(x) \sim e^{-x}$ and a threshold rule. Results are averaged over 100 realizations of size 10^4. Inset: The degree distribution $P(k)$; the straight line corresponds to k^{-2}.

To have a more extensive picture of the nature of the networks under consideration, we have studied the following topological properties [3], interest in which has been triggered by recent studies on the Internet structure [10,23]: (i) the average distance $\langle d \rangle$, measuring the average minimum number of arcs needed to connect two given sites; (ii) the *average neighbor connectivity* $k_{nn}(k)$, measuring the average degree of vertices neighbor of a k-degree vertex; (iii) the *clustering coefficient* $c(k)$ that measures the degree of interconnectivity of nearest neighbors of k-degree vertices. More specifically, the clustering-coefficient c_i of a vertex i, whose degree is k_i, is the ratio between the number of edges e_i in the subgraph identified by its neighbors and the maximum possible number of edges in the subgraph. That is $c_i = 2e_i/k_i(k_i - 1)$ [2]. $c(k)$ is obtained by averaging c_i for all vertices with fixed degree k; (iv) the probability distribution of the *betweenness*, b_i, defined as the total number of minimum paths between any couple of vertices in the network passing through vertex i [24]. This quantity gives a measure of the amount of traffic passing through a vertex. We studied, as in the aforementioned papers, both the probability distribution $P(b)$ and its first moment $\langle b \rangle / N$.

Computer simulations of our model show that networks with power-law distributed fitnesses, and different values of β, show nearly constant $k_{nn}(k)$ and $c(k)$, just as occurs for the original BA model [2]. The distribution of betweenness decays as a power law with an exponent $\gamma_b \approx 2.2$ for $\gamma = 2.5$ and $\gamma = 3$, and $\gamma_b \approx 2.6$ for $\gamma = 4$. This is in good agreement with that conjectured in Ref. [18]: All networks with $3 \geq \gamma > 2$ can be classified in only two groups according to the value of γ_b ($\gamma_b = 2$ and $\gamma_b = 2.2$, respectively), while for larger values of γ, larger nonuniversal values of γ_b are reported.

The exponential case behaves in a different way: For a network of size $N = 10^4$, $z = 10$, and $m = N$, we find $\langle d \rangle = 2$, $\langle c \rangle \simeq 0.1$, and $\langle b \rangle / N \simeq 0.1$, but a power-law behavior is found for the clustering magnitudes, i.e., $\langle k_{nn} \rangle \propto k^{-0.85}$ and $c(k) \propto k^{-1.6}$. The betweenness distribution, instead, shows an unexpected behavior, giving a power-law tail with an exponent $\gamma_b \approx 1.45$ (see Fig. 4). It is worth remarking that our model having $\gamma = 2$ is not included in the previously discussed classification of betweenness exponents [18] (Fig. 3).

Having explored the most basic properties of the model and some particular realizations, let us comment now on possible applications.

Email networks [25] are a good candidate to be represented by our model. In this case growth may occur, but agents (Email senders) do not have any access or knowledge of the degree of the receivers. Rather than preferential attachment there should be some intrinsic feature of the receiver playing a role in the phenomenon.

To further emphasize the utility of this new mechanism, let us mention the following possibility: One can imagine situations where a Poisson network is seen as SF just because the exploration method implicitly implements a probabilistic rule depending on the fitnesses (this applies, for example, when links are detected by "picking" them one by one, but not if the network is

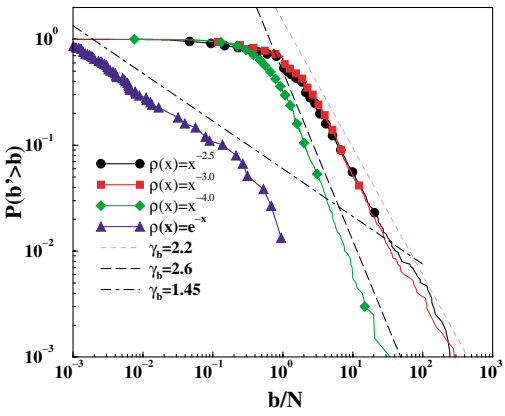

FIG. 4 (color online). Integrated betweenness distribution, $P(b' > b) \propto b^{1-\gamma_b}$, for different fitness distributions.

explored by crawling on it). Let us think, for example, of the case with threshold type of attaching rule. If only links with corresponding fitnesses over threshold are "seen" by the exploration method then, for example, an Erdős-Rényi network with exponentially distributed fitnesses can be seen as SF (with, obviously, a connectivity upper cutoff related to the maximum connectivity of the underlying network; in cases in which this connectivity is high, one can generate hubs in the "apparent" SF network). In particular, this scenario could be of relevance to protein networks. Let us argue why.

The way comprehensive protein networks have been obtained until now is through a bait-prey method, named "two-hybrid" method: Two proteins are hybridized with two fragments of a transcription factor (a protein that binding to a gene promotes its transcription into the corresponding RNA). The spliced promoter does not bind to the gene, transcription is inhibited, and the corresponding RNA is absent. Yet, if the two proteins interact, they bring together the two promoter fragments allowing it to bind to the gene and transcription to start. The presence of the corresponding RNA signals the interaction between the two proteins. We can imagine that the interaction strength between the two proteins has to be above a given threshold, else the typical promoter binding time will be too short for the RNA polymerase to bind to the gene and initiate transcription. In turn, it is reasonable to assume that the interaction strength is a function of some properties of the two proteins (such as, for example, their hydrophobicity, or their accessible surface area). This possibility has still to be checked through an analysis of the detailed physics behind the two-hybrid method.

In summary, we have presented an alternative model to justify the ubiquity of SF networks in nature. It is a natural generalization of the standard Erdős-Rényi. The main result is that emergence of SF properties is not necessarily linked to the ingredients of growth and preferential attachment. Instead, static structures characterized by quenched disorder (for different disorder distributions) and threshold phenomena may generate effects very similar to those measured in the real data. In particular, we recover the power-law behavior of degree, betweenness, and clustering-coefficient distributions. We believe that this model is particularly suitable for situations where the degree value of nodes is not publicly available.

We thank R. Pastor-Satorras for very useful comments and suggestions, and P. Hurtado for a critical reading of the manuscript. The authors thank FET Open Project IST-2001-33555 COSIN, P. D. L. R. thanks the OFES-Bern (Grant No. 02.0234), and M. A. M. acknowledges financial support from the MCYT (Grant No. BFM2001-2841).

[1] S. H. Strogatz, Nature (London) **410**, 268 (2001).
[2] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002), and references therein.
[3] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).
[4] M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2001).
[5] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai, Nature (London) **411**, 41 (2001).
[6] R. V. Solé and J. M. Montoya, Proc. R. Soc. London B **268**, 2039 (2001); J. Camacho *et al.*, Phys. Rev. Lett. **88**, 228102 (2002).
[7] M. Faloutsos, P. Faloutsos, and C. Faloutsos, *Proceedings of the ACM, SIGCOMM* [Comput. Commun. Rev. **29**, 251 (1999).
[8] G. Caldarelli, R. Marchetti, and L. Pietronero, Europhys. Lett. **52**, 386 (2000).
[9] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[10] R. Pastor-Satorras, A. Vazquez, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).
[11] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **401**, 130 (1999).
[12] D. Watts, *Small-Worlds* (Princeton University Press, Princeton, New Jersey, 1999).
[13] SF networks are generated only if $p(k) \sim k$; sublinear $p(k)$ produce skewed but not SF networks; P. L. Krapivsky and S. Redner, Phys. Rev. E **63**, 066123 (2001).
[14] A. Vazquez, A. Flammini, A. Maritan, and A. Vespignani, cond-mat/0108043.
[15] R. V. Solé, R. Pastor-Satorras, E. D. Smith, and T. Kepler, Adv. Compl. Systems **5**, 42 (2002).
[16] G. Bianconi and A.-L. Barabási, Europhys. Lett. **54**, 436 (2001).
[17] P. Erdős and A. Rényi, Bull. Inst. Int. Stat. **38**, 343 (1961).
[18] K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. **87**, 278701 (2001); K.-I. Goh *et al.*, cond-mat/0205232.
[19] P. L. Krapivsky and S. Redner, cond-mat/0207107.
[20] M. Marsili and Y.-C. Zhang, Phys. Rev. Lett. **80**, 2741 (1998);
[21] Being more precise, we obtain a power law plus a delta function of amplitude $\exp(-z)$ at $K = N$, corresponding to the fact that sites with x_i above threshold are connected to every other site. This is irrelevant for large enough z and N.
[22] S. N. Dorogovtsev and J. F. F. Mendes, in *Handbook of Graphs and Networks: From the Genome to the Internet*, edited by S. Bornholdt and H. G. Schuster (Wiley, Berlin, 2002), p. 320.
[23] A. Vazquez, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. E **65**, 066130 (2002); cond-mat/0206084.
[24] See M. E. J. Newman, Phys. Rev. E **64**, 016131 (2001); **64**, 016132 (2001), and references therein.
[25] H. Ebel, L.-I. Mielsch, and S. Bornholdt, Phys. Rev. E **66**, 035103 (2002).