Vertex intrinsic fitness: How to produce arbitrary scale-free networks

Vito D. P. Servedio and Guido Caldarelli
INFM UdR Roma1 and Dipartimento di Fisica, Università "La Sapienza," Piazzale A. Moro 2, I-00185 Roma, Italy

Paolo Buttà
Dipartimento di Matematica, Università di Roma "La Sapienza," Piazzale A. Moro 2, I-00185 Roma, Italy
(Received 29 September 2003; revised manuscript received 27 April 2004; published 23 November 2004)

We study a recent model of random networks based on the presence of an intrinsic character of the vertices called fitness. The vertex fitnesses are drawn from a given probability distribution density. The edges between pairs of vertices are drawn according to a linking probability function depending on the fitnesses of the two vertices involved. We study here different choices for the probability distribution densities and the linking functions. We find that, irrespective of the particular choices, the generation of scale-free networks is straightforward. We then derive the general conditions under which scale-free behavior appears. This model could then represent a possible explanation for the ubiquity and robustness of such structures.

DOI: 10.1103/PhysRevE.70.056126 PACS number(s): 89.75.Hc, 84.35.+i, 89.75.Da

In the last few years, much attention has been focused on the study of complex networks. A network is a mathematical object consisting of a collection of vertices (nodes) connected by edges (links) [1,2]. Networks arise in many areas of science: biology [3–5], social sciences [6–8], Internet [9–11], WWW [12], etc., where vertices and links can be, for example, proteins and their mutual interaction, individuals and sexual relationship [13], and computers and cable connections. Very interestingly, the same nontrivial statistical properties appear ubiquitously in all the above situations. A more traditional view, indeed, is represented by the binomial model inspired to the random graph model of Erdős-Rényi [14]. Here, each vertex has the same probability to connect to any other, resulting in a network with vertex degree, i.e., the number of edges connected to each vertex, distributed according to a binomial probability distribution. This is not the case of the above real data, where instead, the structure is self-similar, resulting in a scale-free (SF) probability distribution for the degree. More specifically, the degree k of the vertices is distributed according to a power law $P(k) \propto k^{\alpha}$ with usually $-3 < \alpha < -2$.

In order to explain the occurrence of SF networks, the ingredients of growth preferential attachment have been introduced [15]. The network increases the number of vertices with time; the newcomers tend to be connected with old vertices with a large degree. This means that in the network, one needs the knowledge of the degree value of all vertices in order to decide with whom to link. This is certainly a rather strange assumption in a variety of different situations. In fact, in some cases, we have the same SF properties without either growth of the system or a preferential attachment mechanism. As an example, the finite set of protein interactions in a cell forms a self-similar network. This is done without growth of the system size and ignoring their reciprocal degree. Possibly, some external influence on intrinsic properties such as chemical affinity is instead driving the phenomenon. Another important example is the sexual interaction network [13]. Here it is evident that the knowledge of the number of previous partners (if possible to achieve) could result in the opposite effect of preferential attachment. Rather, the driving force is the "beauty" of people involved, of which the number of partners is an effective measure.

To take into account this new mechanism, the varying fitness model has been introduced by Caldarelli *et al.* [16]. In this model, considering, e.g., only undirected graphs, one extracts a real non-negative variable x (the hidden variable) for each vertex of the graph from a probability distribution density $\rho(x)$. This variable x is the *fitness* of the vertex. Links between vertices are successively formed with a probability function $f(x,y)$, a symmetric function of its arguments.

A static simplified form of the vertex hidden variable model has been considered for one particular case by Goh *et al.* [17], while Bianconi *et al.* [18] introduced a fitness mechanism coupled to the preferential attachment. In the paper of Caldarelli *et al.* [16], the onset of SF behavior is instead directly related only to the fitness presence of any kind. This SF behavior is checked for two different fitness probability distribution densities. In this manuscript, on the contrary, we present an exhaustive study on the general conditions needed in order to produce a SF network with the vertex hidden variable model. Finally, we apply this study to obtain the analytic expressions connecting the fitness distribution density $\rho(x)$ and probability function $f(x,y)$ in three special cases.

The fitness model can be easily generalized in order to have more than one fitness variable per vertex [19]. In the following, we consider a single real variable x per vertex, with $x \geqslant 0$. As a probability distribution density function, ρ satisfies $\{\rho(x) \geqslant 0 \,|\, \int_0^\infty \rho(z)dz = 1\}$, while the linking probability $0 \leqslant f(x,y) \leqslant 1$. We define the primitive function of $\rho(x)$, the probability distribution $R(x) = \int_0^x \rho(z)dz$. Indicating the number of vertices in the graph with N, one has the vertex degree

$$k(x) = N \int_0^\infty f(x,z)\rho(z)dz. \qquad (1)$$

Other quantities of interest are the average nearest-neighbor connectivity (vertex degree correlation),

 ©2004 The American Physical Society

$$K_{\mathrm{nn}}(x) = \frac{N}{k(x)}\int_0^\infty f(x,z)k(z)\rho(z)dz, \qquad (2)$$

expressing the average degree of vertices that are nearest neighbors of vertices with fitness x, and the clustering coefficient (vertex transitivity),

$$C(x) = N^2\frac{\int_0^\infty \int_0^\infty f(x,y)f(y,z)f(z,x)\rho(y)\rho(z)dydz}{k(x)^2}, \qquad (3)$$

that counts the fraction of nearest neighbors of vertices with fitness x that are also nearest neighbors each other. Equations (1), (2), and (3) are valid asymptotically when N approaches infinity. Equations (2) and (3) were first derived in Ref. [20], and expressed in a different form.

If $k(x)$ is an invertible and increasing function of x, then the probability distribution $P(k)$ is given by

$$P(k) = \rho(x(k))x'(k) \qquad (4)$$

or, as a function of x,

$$P(k(x)) = \frac{\rho(x)}{k'(x)}. \qquad (5)$$

Since the degree probability is power-law distributed in most of the physical situations, we impose in Eq. (5) $P(k)=ck^{\alpha}$ with $\alpha \in \mathbb{R}$. The constant c is fixed by the normalization condition $\int_{k_0}^{k_\infty} P(k)dk=1$,

$$c = \begin{cases} \dfrac{\alpha+1}{k_\infty^{\alpha+1}-k_0^{\alpha+1}}, & \text{if } \alpha \neq -1, \\ \left(\log\dfrac{k_\infty}{k_0}\right)^{-1}, & \text{if } \alpha = -1, \end{cases} \qquad (6)$$

with $k_s=\lim_{x\to s}k(x)$. Note that, according to Eq. (1), $k_0=\beta N$ and $k_\infty=\gamma N$ for some $0<\beta<\gamma\leqslant 1$, so that $c \propto N^{-(\alpha+1)}$. Equation (5) becomes

$$ck'(x)(k(x))^{\alpha} = \rho(x). \qquad (7)$$

By integrating Eq. (7) from 0 to x, we get the following nonlinear integral equation:

$$k(x) = \begin{cases} \left(k_0^{\alpha+1} + \dfrac{\alpha+1}{c}R(x)\right)^{1/(\alpha+1)}, & \text{if } \alpha \neq -1, \\ k_0 e^{R(x)/c}, & \text{if } \alpha = -1, \end{cases} \qquad (8)$$

with $k(x)$ given by Eq. (1).

By multiplying both sides of Eq. (8) by $\rho(x)$ and integrating from 0 to ∞, we get an analytic expression for the average vertex degree $\langle k\rangle$. This expression can be used to write k_0 as a function of $\langle k\rangle$, so that the final expressions do depend on the physical quantity $\langle k\rangle$ only. For this purpose, the integral on the right-hand side is simply solved using the relation $\rho(x)dx=dR(x)$.

In the following, we show an application of the model in three special cases of interest, comparing the analytic results with numerical simulations. It has to be noticed that once N is fixed, in order to compute the quantities $P(k)$, $K_{\mathrm{nn}}(k)$, $C(k)$, from ensemble statistics, we need to perform two different average procedures. First, we should extract an $\{x_i\}_{i=1,\ldots,N}$ configuration with the distribution density $\rho(x)$ and keep it fixed, while creating ensemble elements using the linking probability $f(x,y)$ and averaging at the end. Secondly, we should repeat the above procedure a sufficient number of times. We assume that for large enough N and ensemble elements, the procedure of first averaging with respect to the f can be skipped.

Here we focus on two different problems: first, there is what we call a direct problem, in which one assigns a distribution density function $\rho(x)$ and tries to find the linking probability function $f(x,y)$; secondly, there is what we call an inverse problem, in which one assigns the linking probability function and tries to determine the fitness probability distribution density $\rho(x)$. The inverse problem is by far more complex and interesting than the direct one. For instance, in the case of a protein SF network by assuming a reasonable linking function, we can retrieve the probability density distribution of fitness (e.g., some basic property of the macromolecules).

We start with the special case of $f(x,y)=g(x)h(y)$ where both the direct and inverse problems can be analytically solved. Because of the symmetry of $f(x,y)$ with respect to its arguments, one has $g(x)\equiv h(x)$, so that $f(x,y)=g(x)g(y)$. Equation (1) becomes

$$k(x) = Ng(x)\int_0^\infty g(z)\rho(z)dz, \qquad (9)$$

which substituted into Eq. (8) gives equations in g and ρ. If one fixes a given function $\rho(x)$, the equations in $g(x)$ can be easily solved. Take for instance the second equation corresponding to $\alpha=-1$. One gets

$$Ng(x)\langle g\rangle = k_0 e^{R(x)/c}, \quad \langle g\rangle = \int_0^\infty g(z)\rho(z)dz.$$

By multiplying the left- and right-hand side by $\rho(x)$ and integrating from 0 to ∞, considering that $\rho(x)dx=dR(x)$, we get

$$\langle g\rangle = \sqrt{k_0 c(e^{1/c}-1)/N}.$$

Finally, after substituting the value of c taken from Eq. (6) with $\alpha=-1$, the solution reads

$$g(x) = \beta\sqrt{\frac{\log\gamma-\log\beta}{\gamma-\beta}}e^{R(x)\log(\gamma/\beta)}. \qquad (10)$$

The condition that $g(x)$ be a probability, i.e., $g(\infty)=\lim_{x\to\infty}g(x)\leqslant 1$, fixes the dependence between γ and β. With the choice $g(\infty)=1$, one ensures the broadest range of k such that $P(k)$ is a power law with the desired exponent. This procedure is applicable for any value of α. Equation (10) generates random networks with degree probability distribution $P(k)\propto 1/k$. In order to test the result, we take the choice reported in the caption of Fig. 1.

We conclude that *for any given $\rho(x)$ there exists a function $g(x)$ such that the network generated by $\rho(x)$ and $f(x,y)=g(x)g(y)$ is scale-free with an arbitrary real exponent.*

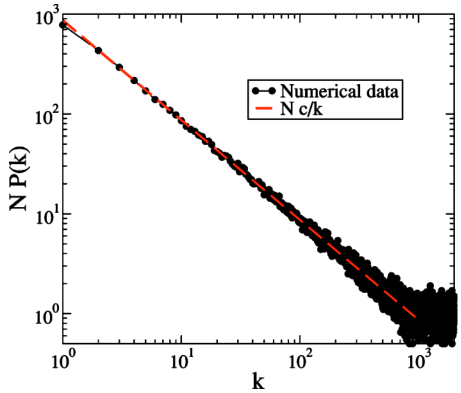

FIG. 1. (Color online) Vertex degree distribution generated by $f(x,y)=g(x)g(y)$, $\alpha=-1$, $\rho(x)=e^{-x}$, $N=10^4$, $k_0=0.1$ [resulting in $k_\infty \approx 1077$ by requiring $g(\infty)=1$]. The function $g(x)$ is given by Eq. (10). This figure is obtained averaging over 20 realizations.

In this case, both the average nearest-neighbor connectivity and clustering coefficient are constant [20]. Respectively,

$$K_{\mathrm{nn}} = N\langle g^2\rangle, \quad C = \frac{\langle g^2\rangle^2}{\langle g\rangle^2}, \tag{11}$$

as it can be derived from Eq. (2) and Eq. (3). This special case is in some sense close to the preferential attachment mechanism, in that vertices with a large fitness value are likely to be the most connected ones in the network. To large vertex degree values correspond large vertex fitness values, so that the preferential attachment rule is recovered in a more natural way without the necessity, from the newcoming nodes, of the knowledge of the whole set of vertex degrees.

The inverse problem for $f(x,y)=g(x)g(y)$ is solved by substituting Eq. (9) into Eq. (7),

$$\rho(x) = cg'(x)g(x)^{\alpha}(N\langle g\rangle)^{\alpha+1}.$$

Let us remark that the assumptions on $k(x)$ force $g(x)$ to be nondecreasing with $g(\infty) > g(0) > 0$.

The normalization condition $R(\infty)=1$ results in

$$\rho(x) = \begin{cases} \dfrac{\alpha+1}{g(\infty)^{\alpha+1} - g(0)^{\alpha+1}} g'(x)g(x)^{\alpha}, & \text{if } \alpha \neq -1, \\ \left(\log\dfrac{g(\infty)}{g(0)}\right)^{-1} g'(x)g(x)^{-1}, & \text{if } \alpha = -1. \end{cases} \tag{12}$$

The case $f(x,y)=f(x-y)$ is more complicated. In this case, both the nearest-neighbor connectivity and clustering coefficient depend on the fitness x and conversely on the degree k. We managed to solve this case in the particular case of an exponentially distributed fitness. We indicate with $F(x)$ the right-hand side of Eq. (8). Thus Eq. (8) becomes

$$\int_0^{\infty} f(x-u)\rho(u)du = F(x)/N.$$

By changing the integration variable into $z=x-u$, we get

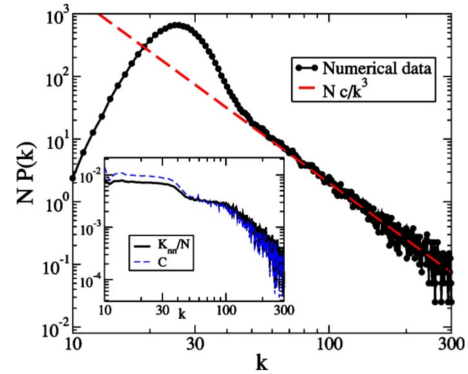

FIG. 2. (Color online) Degree distribution in the case $f(x,y) = f(x-y)$, $\alpha=-3$, $\rho(x)=e^{-x}$, $k_0=10$, $k_\infty=N=10^4$, $f(u)=[F(u) + F'(u)]/N$ with $F(x)$ given by the right-hand side of Eq. (8), averaged 40 times. The value of c is calculated from Eq. (6). The inset shows the vertex degree correlation and transitivity as functions of the vertex degree.

$$\int_{-\infty}^{x} \rho(x-z)f(z)dz = F(x)/N,$$

which in the special case $\rho(x)=e^{-x}$ becomes

$$\int_{-\infty}^{x} e^{z}f(z)dz = e^{x}F(x)/N.$$

By differentiating with respect to the variable x, we finally obtain

$$f(x,y) = \frac{F(x-y) + F'(x-y)}{N}. \tag{13}$$

In order to test the result, we take the function and parameter choice of the Fig. 2 caption.

The case $f(x,y)=f(x+y)$ is analogous. Again, we consider the special case $\rho(x)=e^{-x}$, getting now

$$f(x,y) = \frac{F(x+y) - F'(x+y)}{N}. \tag{14}$$

In both previous cases of Eqs. (13) and (14), the value of γ can be chosen independently from β since $\lim_{z\to\infty} f(z)=\gamma$, as it can be simply derived from Eq. (7), by which one finds that $\lim_{z\to\infty} F'(z)=0$. The best choice is of course $\gamma=1$.

The solution of Eq. (8) for $\alpha=-2$ obtained via Eq. (14) reads, recalling that $k_0=\beta N$, $k_\infty=\gamma N$, and using Eq. (6),

$$f(x,y) = \frac{\gamma}{[1 + (\gamma\beta^{-1} - 1)e^{-(x+y)}]^2}. \tag{15}$$

Through Eq. (15) we clarify the assumption made in the original paper by Caldarelli *et al.* [16], where $f(x,y)=\Theta(x+y-z)$ with $z=z(N)$. Note that now with the latter choice of $f(x,y)$ one gets $P(k)=Ne^{-z}k^{-2}$ that forces z to depend logarithmically upon N in order to get the correct normalization. The functional form of the $z(N)$ was already guessed numerically by Ref. [21]. To test the result, we take the parameters reported in the caption of Fig. 3. In these last two cases, both

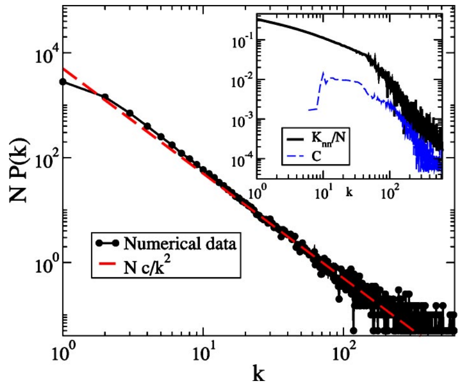

FIG. 3. (Color online) Degree distribution in the case $f(x,y) = f(x+y)$ and $\alpha=-2$, $\rho(x)=e^{-x}$, $k_0=0.5$, $k_\infty=N=10^4$, $f(v)$ from Eq. (15), averaged 20 times. The inset shows the vertex degree correlation and transitivity as functions of the vertex degree.

the nearest-neighbor connectivity and clustering coefficient show a nontrivial k dependence.

We think that the fitness mechanism studied in this paper is responsible for the widespread occurrence of SF networks. Indeed, in many cases, such as Pareto's law [22] in economics or Zipf's law [23] in linguistic, or fractal growth [24], we find the presence of characteristic power-law distributions. Once those systems are represented by means of graphs, those power laws come back in the role of SF networks. In this respect then, we think that the proposed mechanism represents the connection between fractal growth theory and the study of SF networks. Moreover, a new class of phenomena where the various elements do not display any fractal (power-law) behavior results in the formation of SF networks. Indeed, with this new model of fitness growth, also Gaussian distributed sets of vertex fitnesses (probably because of the central limit theorem) may give rise to SF behavior when the linking probability function assumes the form stemming from the solution of Eq. (8). This is probably the origin of the fact that SF networks seem to be more general than fractal phenomena.

In conclusion, we present a general procedure to reproduce real SF networks with arbitrary vertex degree distribution densities. More specifically, we found that, given a fitness distribution density $\rho(x)$, it is always possible to find a symmetric linking probability function $f(x,y)$ such that the resulting random network is SF with a given real exponent. We give the recipe to find these linking functions, in three cases of interest. In order to allow the generation of networks even closer to the real data, it would be desirable to have control not only on the vertex degree distribution, but also on the vertex transitivity and vertex degree correlation, by solving simultaneously Eqs. (2), (3), and (5). As a first step, the compatibility of these three equations should be addressed, once the functions $P(k)$, $K_{\rm nn}(k)$, and $C(k)$ are given. The solution of this problem is certainly very hard and is left open for the future. The relative ease with which we obtain SF structures seems to be the key ingredient in order to explain the ubiquitous presence and robustness of the real data.

We acknowledge R. Pastor-Satorras, P. De Los Rios, D. Garlaschelli, F. Squartini, S. Millozzi, and S. Leonardi for useful discussions, and support from EU FET Open Project IST-2001-33555 COSIN (www.cosin.org). V.D.P.S. and G.C. thank IMFM PAIS PA_G02_4 for support.

[1] B. Bollobás, *Random Graphs*, 2nd ed. (Academic, New York, 2001).
[2] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks* (Oxford University Press, Oxford, 2003).
[3] J. Kim, P. L. Krapivsky, B. Kahng, and S. Redner, Phys. Rev. E **66**, 055101 (2002).
[4] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai, Nature (London) **411**, 41 (2001).
[5] D. Garlaschelli, G. Caldarelli, and L. Pietronero, Nature (London) **423**, 165 (2003)
[6] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11149 (2000).
[7] M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2001).
[8] M. E. J. Newman, Phys. Rev. E **67**, 026126 (2003)
[9] G. Caldarelli, R. Marchetti, and L. Pietronero, Europhys. Lett. **52**, 386 (2000).
[10] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[11] R. Pastor-Satorras, A. Vazquez, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).
[12] L. A. Adamic and B. A. Huberman, Science **287**, 2115 (2000).
[13] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Åberg, Nature (London) **411**, 907 (2001).
[14] P. Erdős and A. Rényi, Bull. Inst. Int. Stat. **38**, 343 (1961).
[15] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002), and references therein.
[16] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Muñoz, Phys. Rev. Lett. **89**, 258702 (2002).
[17] K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. **87**, 278701 (2001).
[18] G. Bianconi and A.-L. Barabási, Europhys. Lett. **54**, 436 (2001).
[19] M. Boguñá, R. Pastor-Satorras, A. Diaz-Guilera, and A. Arenas, e-print cond-mat/0309263.
[20] M. Boguñá and R. Pastor-Satorras, Phys. Rev. E **68**, 036112 (2003).
[21] L. Laura, S. Leonardi, and S. Millozzi (private communication). A suitable choice argued from previous computer simulations was $z(N) \propto \log N$, in accord with our findings.
[22] G. Sorel, *Review of Pareto's Course D'economie Politique*, Vol. 1, Le devenir social, 5-2 (1896); *Review of Pareto's Course D'economie Politique*, Vol. 2, 5-3 (1897).
[23] G. K. Zipf, *Human Behaviour and the Principle of Least Effort* (Addison-Wesley, Reading MA, 1949).
[24] B. Mandelbrot, *Les Objets Fractals: Forme, Hasard, et Dimension* (Flammarion, Paris, 1975).