Complex growing networks with intrinsic vertex fitness

C. Bedogne' and G. J. Rodgers
Department of Mathematical Sciences, Brunel University, Uxbridge, Middlesex UB8 3PH, United Kingdom
(Received 3 April 2006; revised manuscript received 21 July 2006; published 23 October 2006)

One of the major questions in complex network research is to identify the range of mechanisms by which a complex network can self organize into a scale-free state. In this paper we investigate the interplay between a fitness linking mechanism and both random and preferential attachment. In our models, each vertex is assigned a fitness x, drawn from a probability distribution $\rho(x)$. In Model A, at each time step a vertex is added and joined to an existing vertex, selected at random, with probability p and an edge is introduced between vertices with fitnesses x and y, with a rate $f(x,y)$, with probability $1-p$. Model B differs from Model A in that, with probability p, edges are added with preferential attachment rather than randomly. The analysis of Model A shows that, for every fixed fitness x, the network's degree distribution decays exponentially. In Model B we recover instead a power-law degree distribution whose exponent depends only on p, and we show how this result can be generalized. The properties of a number of particular networks are examined.

DOI: 10.1103/PhysRevE.74.046115 PACS number(s): 89.75.Hc, 84.35.+i, 89.75.Da

I. INTRODUCTION

Complex networks are normally modeled using graph theory in which evolving or growing sets of vertices are connected by edges. The vertices can represent the individuals of the system under consideration, the edges then symbolize relations between them. Complex networks are ubiquitous in physics, biology, sociology, and computer science. Despite the difference in the interpretation of the vertices and edges, and the networks' diversity of function, complex networks display considerable statistical and topological similarities.

The analysis of complex networks was initially approached through random graph theory [1], where the vertices are connected at random. These models result in exponential degree distributions (the degree of a vertex is the number of edges connecting it to other vertices). Thus the structure of these random graphs is uniform with most vertices having approximately the same degree and only few vertices of high degree.

It soon became clear, however, that this model, usually referred to as the Erdös-Renyi model [1,2], cannot explain the behavior of the many networks that display power-law rather than exponential degree distributions. These networks are called *scale-free*, because of the self-similarity property of the power law. Power laws decay very slowly, in contrast to the Poisson degree distribution of random graphs, and consequently scale-free networks have a number of hubs which are often the most important vertices. Hubs can connect apparently very distant environments and play a crucial role in the spreading of infective agents (such as human or computer viruses [3,4]) or the stability of network structures under external attack [5].

The widespread occurrence of scale-free topologies demonstrates that networks cannot be fundamentally random objects. The topological similarities between structures as diverse as the world-wide web, the web of human sexual contacts, and some metabolic networks suggest that networks cannot be understood only in terms of properties of their *parts*, and that one needs to identify the self-organizing principles controlling the evolution of the network *as a whole*.

Through a variety of different analytical approaches, usually borrowing tools from statistical physics [1,6–9], it has been shown that if the following two conditions hold: (i) the network grows in time by addition of new vertices at a constant rate; and (ii) the newly introduced vertices connect preferentially to highly connected vertices, we recover the power-law degree distribution displayed by many real-world networks. The simplest analytical model, usually referred to as the Albert-Barabási model [1], assumes growth at a constant rate and a *linear* form of preferential attachment [1,6], yielding a power law with exponent 3. While only a fraction of real-world networks have been fully characterized empirically, data mining usually recovers power laws with exponents ranging between 2 and 3.5. Thus growth and preferential attachment seem to be the principal mechanisms for scale-free degree distributions.

However, it has also been argued [10–12] that preferential attachment is not always a natural hypothesis since a new vertex, informally speaking, needs to know the global degree distribution of the network in order to decide with whom to link. It is obvious that in many contexts, and especially when the networks involved are very large, this information will not be available.

The varying fitness model [11,12] was thus proposed as an alternative mechanism for recovering scale-free topologies. A real non-negative variable x, drawn from a probability distribution $\rho(x)$, is assigned to each vertex of the network, representing the *fitness*, or ability to compete for new edges of the vertex. In real world networks the fitness will be related to intrinsic qualities of the vertex, such as rank, beauty, or wealth. Edges between vertices successively form according to a linking function $f(x,y)$, designed to model the idea of mutual benefit in a competitive environment.

The varying fitness model shows that scale-free behavior does not necessarily imply an underlying preferential attachment mechanism. Indeed, given any fitness probability distribution ρ, it is always possible to determine a linking function $f(x,y)$ such that the resulting degree distribution decays as a power law with a given real exponent [12].

While the varying fitness model interprets network topology exclusively in terms of fitnesses, attempts [13,14] have been made to explain network evolution on the basis of a preferential attachment mechanism which depends on the fitnesses, arguing that fitter vertices might overcome highly connected but less fit ones. It has been shown [13,14] that the introduction of a random additive fitness (or quenched disorder) to each vertex leads to a power law whose exponent depends on the average fitness of the network. The introduction of both a random additive fitness x and a multiplicative fitness y leads instead to multiscaling, that is, to a power law with a fitness dependent exponent.

Following this approach, we use rate equations [6–8] to investigate models where a fitness linking process is combined with random and preferential attachment (rather than *coupled* to preferential attachment as in Refs. [13,14]). Since the self-organizing principles underlying network evolution might be system-dependent and by no means unique, we wish to analyze the interplay and relative strength of the different processes likely to shape any network's structure.

In Sec. II, generalizing growing Erdös-Renyi graphs [1,2], we analyze a network built by adding the fitness linking process to random attachment. In Sec. III, generalizing the Albert-Barabási model [1], we analyze a model that incorporates preferential rather than random attachment. In the last section we summarize our results and draw conclusions.

II. MODEL A

Following Refs. [11,12] we introduce a distribution density function $\rho(x)$ and a linking probability function $f(x,y)$. We then introduce a network built as follows. At each time step

(a) with probability p, $0<p\leq 1$, a new vertex is introduced that then connects at random to one of the earlier vertices;

(b) or with probability $1-p$ an edge is introduced between vertices of fitnesses x,y with rate $f(x,y)$.

The average number of vertices with fitness x and degree k, $N_k(x,t)$, evolves as

$$\frac{\partial N_k(x,t)}{\partial t} = \frac{p}{N(t)}(N_{k-1}(x,t)-N_k(x,t)) + p\delta_{k1}\rho(x) + 2(1-p) \times \frac{w(x,t)}{\sum_{k=1}^{\infty}\int_0^{\infty} w(x,t)N_k(x,t)dx}(N_{k-1}(x,t)-N_k(x,t)), \tag{1}$$

where $N(t)$ is the total number of vertices at the time t and

$$w(x,t) = \sum_{k=1}^{\infty}\int_0^{\infty} f(x,y)N_k(y,t)dy. \tag{2}$$

The first term on the right hand side of Eq. (1) represents the change in the average number of vertices with fitness x and degree k due to process (a). The second term accounts for the continuous introduction, with probability p, of new vertices with fitness drawn from the probability distribution $\rho(x)$. The third term represents the change in the average number of vertices with fitness x and degree k due to process (b). We define also

$$N(x,t) = \sum_{k=1}^{\infty} N_k(x,t), \tag{3}$$

and express the total number of vertices at the time t as

$$N(t) = \int_0^{\infty} N(x,t)dx. \tag{4}$$

Summing Eq. (1) over k we obtain

$$\frac{\partial N(x,t)}{\partial t} = p\rho(x) \tag{5}$$

which, given our initial conditions (no vertex is present in the network at the time $t=0$), yields

$$N(x,t) = p\rho(x)t \tag{6}$$

in the large t limit. Integrating Eq. (6) we find as expected $N(t)=pt$, and Eq. (6) reads now $N(x,t)=\rho(x)N(t)$. Equation (6) allows us to express the integrals in the third term of Eq. (1) in terms of f and ρ:

$$\frac{w(x,t)}{\sum_{k=1}^{\infty} w(x,t)N_k(x,t)dx} = \frac{1}{pt}\frac{\int_0^{\infty} f(x,y)\rho(y)dy}{\int_0^{\infty}\int_0^{\infty} f(x,y)\rho(x)\rho(y)dxdy}. \tag{7}$$

We introduce the function

$$D(x) = \frac{\int_0^{\infty} f(x,y)\rho(y)dy}{\int_0^{\infty}\int_0^{\infty} f(x,y)\rho(x)\rho(y)dxdy}, \tag{8}$$

which can be interpreted physically (in the long time limit) as the mean degree of a vertex of fitness x divided by the total degree of the network, referring of course only to the degree accumulated by process (b).

Thus, Eq. (7) takes the form

$$\frac{w(x,t)}{\sum_{k=1}^{\infty} w(x,t)N_k(x,t)dx} = \frac{D(x)}{N(t)}. \tag{9}$$

We proceed now to solve the rate equations for Model A following the methods in Refs. [6–8]. Since $N_k(x,t)$ grows linearly with time, we introduce functions $c_k(x)$ such that

$$N_k(x,t) = tc_k(x), \tag{10}$$

and notice that Eq. (10) becomes *exact* in the long time limit.

Substituting Eqs. (9) and (10) into Eq. (1) gives recursive equations for the functions $c_k(x)$;

$$c_k(x) = \frac{H(x)}{1+H(x)}c_{k-1}(x) + \frac{p\delta_{k1}\rho(x)}{1+H(x)}, \tag{11}$$

where the function H is defined by

$$H(x) = 1 + 2\frac{1-p}{p}D(x). \tag{12}$$

Equation (11) can be easily solved recursively and yields

$$c_k(x) = \frac{H(x)^{k-1}}{(1+H(x))^k}p\rho(x). \tag{13}$$

Defining now

$$F(x) = \int_0^\infty f(x,y)\rho(y)dy \tag{14}$$

and

$$\Lambda = \int_0^\infty F(x)\rho(x)dx, \tag{15}$$

we can express Eq. (13) in the form

$$c_k(x) = \frac{p}{2^k}\frac{\left(1+2\frac{1-p}{p\Lambda}F(x)\right)^{k-1}}{\left(1+\frac{1-p}{p\Lambda}F(x)\right)^k}\rho(x). \tag{16}$$

Hence

$$c_k(x) = (T(x))^k A(x), \tag{17}$$

where

$$T(x) = \frac{1}{2}\left[1 + \frac{\frac{1-p}{p\Lambda}F(x)}{1+\frac{1-p}{p\Lambda}F(x)}\right] \tag{18}$$

and

$$A(x) = \frac{p\rho(x)}{1+2\frac{1-p}{p\Lambda}F(x)}. \tag{19}$$

Equation (18) shows that the inequality $\frac{1}{2} \leqslant T(x) < 1$ holds. There exists then a function $A = A(x)$ and a function $B = B(x)$ such that

$$c_k(x) = A(x)\frac{1}{B(x)^k}, \tag{20}$$

with $1 < B(x) \leqslant 2$. Thus *for every fixed* x, the network's degree distribution decays exponentially. The function $c_k(x)$ can be equivalently expressed in the form

$$c_k(x) = \frac{p}{2^{k-1}}E(x)^{k-1}\rho(x) - \frac{p}{2^k}E(x)^k\rho(x), \tag{21}$$

where $E(x) = 2T(x)$.

The global degree distribution is given by $c_k = \int_0^\infty c_k(x)dx$, and can thus be expressed in the form $c_k = (P/2^{k-1})a_{k-1} - (p/2^k)a_k$, where $a_k = \int_0^\infty E(x)^k\rho(x)dx$. Notice that the coefficients $\{a_k,\ k=1,2,\ldots\}$ satisfy the inequalities $a_{k-1} < a_k$, $a_k \leqslant 2a_{k-1}$. Notice also that Eqs. (6) and (10) imply that $\Sigma_{k=1}^\infty c_k = p$. It follows, as expected, $\lim_{k\to\infty} c_k = 0$.

Finding the explicit analytical form of the infinitesimal sequence $\{c_k,\ k=1,2,\ldots\}$ is very hard, even for simple choices of f and ρ, as the coefficients $\{a_k,\ k=1,2,\ldots\}$ usually cannot be expressed in terms of elementary functions. Some particular solutions can, however, be obtained for c_k, for instance when $\rho(x)$ is a linear sum of delta functions, so that the system can be divided into interacting subpopulations. For instance, if $f(x,y) = xy$ and

$$\rho(x) = \tfrac{1}{2}[\delta(x-\beta) + \delta(x)], \tag{22}$$

where β is some positive constant, then

$$c_k = \frac{p}{2}\left[2^{-k} + \frac{p}{4-3p}\left(\frac{4-2p}{4-3p}\right)^{-k}\right]. \tag{23}$$

For large k, the degree distribution is dominated by the second term. Some solutions are also possible for continuous $\rho(x)$ and $f(x,y)$. For instance, when $\rho(x) = 1$, $0 < x < 1$ and $f(x,y) = (x-y)^2$, so that vertices with different fitnesses are more likely to be joined than those with similar fitnesses, it is a simple matter to show that

$$c_k \sim \frac{1}{(1+p)^k}. \tag{24}$$

The same result holds for any any α if $f(x,y) = (x^\alpha - y^\alpha)^2$.

We notice at last that, although is not always possible to express the functions analytically, it is nevertheless possible to find f and ρ such that the global degree is power-law distributed. Supposing for instance that $T(x) = e^{-x}$ and $A(x) = Dx^\alpha$, where $D > 0$, Eq. (17) yields $c_k = D\Gamma(\alpha+1)(1/k^{\alpha+1})$. We can easily see that ρ must now take the form $\rho(x) = (D/p)[x^\alpha/(e^x-1)]$. Thus we can choose D such that $\int_0^\infty \rho(x)dx = 1$. The linking function f will then be a (normalized) element of the linear space of the solutions of the integral equation,

$$\int_0^\infty f(x,y)\rho(y)dy = \frac{p}{2(1-p)}\frac{2-e^x}{e^x-1} \times \int_0^\infty\int_0^\infty f(x,y)\rho(x)\rho(y)dxdy. \tag{25}$$

It is obvious that this technique can be generalized. Supposing that T and A are chosen such that the integration of Eq. (17) yields a power law, without any loss in generality we can multiply the function A by a constant which will be used to normalize the probability distribution ρ. The linking function f will then be a solution of an integral equation of the same kind of the one above.

We end this section by observing that if we forget the random process (a) considering only the second and third terms of Eq. (1), we still recover an exponential degree distribution. With calculations similar to the ones above, we can easily obtain the recursive equations

$$c_k(x) = \frac{H(x)-1}{H(x)}c_{k-1}(x) + \frac{p\delta_{K1}\rho(x)}{H(x)}, \tag{26}$$

whose solution is

$$c_k(x) = \frac{p\rho(x)}{H(x)-1}\left(1-\frac{1}{H(x)}\right)^k. \tag{27}$$

Since $H(x)>1$, Eq. (27) has the form

$$c_k(x) = U(x)\frac{1}{V(x)^k} \tag{28}$$

where $V(x)>1$, showing that the network's degree distribution decays exponentially for every fixed x.

Notice that in this case, given *any* probability distribution ρ, it is always possible [12] to find a linking function $f(x,y)$ such that c_k scales as a power law with a fixed real exponent. On the other hand, if we choose U and V such that integration of Eq. (28) gives a power law, we can always determine f and ρ accordingly. The probability distribution $\rho(x)$ is uniquely determined by U, V and takes the form

$$\rho(x) = \frac{1}{p}\frac{U(x)}{V(x)-1}. \tag{29}$$

Since U can always be multiplied by an arbitrary constant, we can assume that

$$\frac{1}{p}\int_0^\infty \frac{U(x)}{V(x)-1}dx = 1. \tag{30}$$

The linking function obeys then the integral equation

$$\int_0^\infty f(x,y)\frac{U(y)}{V(y)-1}dy = \frac{1}{2(1-p)}\frac{1}{V(x)-1}\int_0^\infty\int_0^\infty \times f(x,y)\frac{U(x)}{V(x)-1}\frac{U(y)}{V(y)-1}dxdy. \tag{31}$$

III. MODEL B

We change now the model substituting preferential to random attachment. The rate equations now become

$$\frac{\partial N_k(x,t)}{\partial t} = \frac{p}{M(t)}[(k-1)N_{k-1}(x,t) - kN_k(x,t)] + p\delta_{k1}\rho(x) + 2(1-p)\frac{w(x,t)}{\sum_{k=1}^{\infty}\int_0^\infty w(x,t)N_k(x,t)dx} \times[N_{k-1}(x,t) - N_k(x,t)], \tag{32}$$

where the proper normalizing factor in the first term on the right hand side is now

$$M(t) = \int_0^\infty \sum_{k=1}^{\infty} kN_k(x,t)dx. \tag{33}$$

While the second and the third term on the right hand side of Eq. (32) are the same as in Model A, the first term on the right hand side now represents the change in the average number of vertices with fitness x and degree k due to preferential attachment.

Differentiating Eq. (33) gives

$$\frac{\partial M(t)}{\partial t} = 2, \tag{34}$$

which obviously yields

$$M(t) = 2t, \tag{35}$$

in the large t limit. Substituting now Eqs. (8), (10), and (35) into Eq. (32) gives recursive equations for the functions $c_k(x)$,

$$c_k(x) = \frac{H(x)-1+\frac{p}{2}(k-1)}{H(x)+\frac{pk}{2}}c_{k-1}(x) + \frac{p\delta_{k1}\rho(x)}{H(x)+\frac{pk}{2}}, \tag{36}$$

whose solution can be expressed in terms of gamma functions,

$$c_k(x) = 2\rho(x)\frac{\Gamma\left(k+2\frac{H(x)-1}{p}\right)\Gamma\left(1+2\frac{H(x)}{p}\right)}{\Gamma\left(1+2\frac{H(x)-1}{p}\right)\Gamma\left(k+1+\frac{2H(x)}{p}\right)}. \tag{37}$$

Equation (37) yields the asymptotic result [15]

$$c_k(x) \sim \frac{1}{k^{1+(2/p)}}. \tag{38}$$

For every x and for large k we have thus found a power law degree distribution, whose exponent $\gamma(p)=1+\frac{2}{p}$ only depends on p. When $p=1$ we recover $\gamma=3$ as expected for a random network growing preferentially.

Equation (38) shows that the interplay between process (a) and process (b) shifts the power law by $2\frac{1-p}{p}$ above the value 3 we would obtain when preferential attachment acts in isolation. We have also seen, on the other hand, that the degree distribution of the network described only by the second and third terms of Eq. (32) decays exponentially for every fixed x and, therefore, it cannot shift a power law by mere addition. We conclude that the (nonlinear) influence of process (b) on process (a), when evaluated on vertices of a given fitness x, turns out to *weaken* the effect of process (a). When $p\neq 1$, the power law exponent is larger then 3, and we would thus expect to find a smaller number of hubs than when preferential attachment acts in isolation.

Notice that when $p\to 0$ the preferential attachment mechanism still dominates the fitness linking process, in the sense that we still recover a power law distributed connectivity, but by choosing an appropriately small p we can make the power law decay arbitrarily fast.

There are very few $\rho(x)$ and $f(x,y)$ that allow an explicit solution for c_k. An exception is $f(x,y)=xy$ and $\rho(x)$ given by Eq. (22). Then we have an explicit solution:

$$c_k = \frac{\Gamma(k)\Gamma\left(\frac{2}{p}+1\right)}{\Gamma\left(k+\frac{2}{p}+1\right)} + \frac{\Gamma\left(k+8\frac{1-p}{p^2}\right)\Gamma\left(\frac{2}{p}+1+8\frac{1-p}{p^2}\right)}{\Gamma\left(k+1+\frac{2}{p}+8\frac{1-p}{p^2}\right)\Gamma\left(1+8\frac{1-p}{p^2}\right)}. \quad (39)$$

It is easy to see that both terms decay at the same rate for large k, independent of which value of x they correspond to, given by Eq. (38).

The three terms of Eq. (32) depend, respectively, on probabilities p, p, and $1-p$. We can easily generalize our results by replacing the triplet $(p,p,1-p)$ by a triplet of completely unrelated probability parameters (p,r,q).

Equation (35) becomes now

$$M(t) = p + r + 2q, \quad (40)$$

and Eq. (36) takes the form

$$c_k(x) = \frac{W(x)-1+\alpha(k-1)}{W(x)+\alpha k} c_{k-1}(x) + \frac{r\delta_{k1}\rho(x)}{W(x)+\alpha k}, \quad (41)$$

where

$$W(x) = 1 + \frac{2q}{r} D(x) \quad (42)$$

and

$$\alpha = \frac{p}{p+r+2q}. \quad (43)$$

When $p\neq 0$ (the case $p=0$ can be treated analogously to the case discussed at the end of Sec. I) the solution of Eq. (41) is

$$c_k(x) = \frac{r\rho(x)}{\alpha} \frac{\Gamma\left(k+\frac{w(x)-1}{\alpha}\right)\Gamma\left(1+\frac{W(x)}{\alpha}\right)}{\Gamma\left(1+\frac{W(x)-1}{\alpha}\right)\Gamma\left(k+1+\frac{W(x)}{\alpha}\right)} \quad (44)$$

yielding the asymptotic result [15]

$$c_k(x) \sim \frac{1}{k^{1+(1/\alpha)}}. \quad (45)$$

We have still recovered, for every fixed x, a power law degree distribution. In this case we can easily see that the probabilities can be tuned to give any power-law exponent larger than 2.

IV. CONCLUSIONS

Using rate equations we have investigated the interplay between the fitness linking process introduced in the varying fitness model and the other two most important networks' self-organizing principles examined so far in the literature, that is random and preferential attachment.

In Sec. II (Model A), we have shown that the interplay between the fitness linking mechanism and random attachment results in an exponential degree distribution for any fixed fitness x. Although the analytical form of the exponential law depends explicitly on the choice of the probability distribution $\rho(x)$ and of the linking function $f(x,y)$, the exponential basis is always limited between 1 and 2. We also noticed that integrating this exponential distributions over x easily gives power laws, when f and ρ are chosen appropriately. We also examined special cases where the global degree is instead exponentially distributed. We have then shown that the fitness linking mechanism acting in isolation also yields an exponential degree distribution for every fixed x.

In Sec. III (Model B) we replaced random with preferential attachment showing that the degree distribution decays as a power law shifted by $2\frac{1-p}{p}$ above the value expected for a linear preferential attachment mechanism acting in isolation. We have interpreted this result as a weakening of the preferential attachment mechanism due to the fitness linking process. We observed that Model B can be generalized taking account of three completely unrelated probabilities and that the parameters thus introduced can be tuned to give any desired power law larger than 2.

ACKNOWLEDGMENT

We thank the European Union Marie Curie Program (NET-ACE project) for financial support.

[1] B. Bollobás, *Random Graphs* (Academic Press, London, 1985).
[2] R. Albert and A. L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[3] R. Pastor-Satorras and A. Vespignani, "Epidemics and immunization," in *Scale-Free Networks* (Wiley-VCH, Berlin, 2002).
[4] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[5] R. Albert, H. Jeong, and A. Barabási, Nature (London) **406**, 378 (2000).
[6] P. L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett.

85, 4629 (2000).
[7] P. L. Krapivsky and S. Redner, Phys. Rev. E **63**, 066123 (2001).
[8] P. L. Krapivsky, G. J. Rodgers, and S. Redner, Phys. Rev. Lett. **86**, 5401 (2002).
[9] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000).
[10] K. I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. **87**, 278701 (2001).
[11] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Muñoz, Phys. Rev. Lett. **89**, 258702 (2002).
[12] V. D. P. Servedio and G. Caldarelli, Phys. Rev. E **70**, 056126 (2004).
[13] G. Bianconi and A. L. Barabási, Europhys. Lett. **54**, 436 (2001).
[14] G. Ergün and G. J. Rodgers, Physica A **303**, 261 (2002).
[15] M. Abramovitch and I. A. Stegun, *Handbook of Mathematical Functions* (Dover Publications, New York, 1966).