Rate equation approach for correlations in growing network models

Alain Barrat[1] and Romualdo Pastor-Satorras[2]
[1]*Laboratoire de Physique Théorique (UMR du CNRS 8627), Bâtiment 210, Université de Paris-Sud 91405 Orsay, France*
[2]*Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord B4, 08034 Barcelona, Spain*
(Received 25 October 2004; published 22 March 2005)

We propose a rate equation approach to compute two vertex correlations in scale-free growing network models based on the preferential attachment mechanism. The formalism, based on previous work of Szabó *et al.* [Phys. Rev. E. **67**, 056102 (2002)] for the clustering spectrum, measuring three vertex correlations, is based on a rate equation in the continuous degree and time approximation for the average degree of the nearest neighbors of vertices of degree k, with an appropriate boundary condition. We study the properties of both two and three vertex correlations for linear preferential attachment models, and also for a model yielding a large clustering coefficient. The expressions obtained are checked by means of extensive numerical simulations. The rate equation proposed can be generalized to more sophisticated growing network models, and also extended to deal with related correlation measures. As an example, we consider the case of a recently proposed model of weighted networks, for which we are able to compute a weighted two vertex correlation function, taking into account the strength of the interactions between connected vertices.

DOI: 10.1103/PhysRevE.71.036127 PACS number(s): 89.75.Hc, 87.23.Ge, 05.70.Ln

I. INTRODUCTION

Many natural and manmade complex systems can be fruitfully represented and studied in terms of networks or graphs [1], in which the vertices stand for the elementary units that compose the system, while the edges picture the interactions or relations between pairs of units. This topological representation has found many applications in fields as diverse as the internet [2], the worldwide web [3], biological interacting networks [4–6], or social systems [7], leading to the development of a new branch of statistical mechanics, the modern theory of complex networks [8,9].

The empirical study of real complex networks, promoted by the recent accessibility to computers powerful enough to deal with very large databases, has uncovered the presence of some typical characteristics. The three most relevant of these are the following. (i) The small-world property [10], defined by an average shortest path length—average distance between any pair of vertices—increasing very slowly (logarithmically or more slowly [11]) with the network size N. (ii) The presence of a large transitivity [7], which implies that two neighbors of a given vertex are also connected to each other with large probability. Transitivity can be quantitatively measured by means of the clustering coefficient c_i of vertex i [10], defined as the ratio between the number of edges m_i existing betwen the k_i neighbors of i, and its maximum possible value, i.e., $c_i=2m_i/[k_i(k_i-1)]$. The average clustering coefficient, defined as $C=\Sigma_i c_i/N$, usually takes quite large values in real complex networks. (iii) A scale-free behavior for the degree distribution $P(k)$ [8,9], defined as the probability that a vertex is connected to k other vertices (has degree k), that shows a power-law behavior

$$P(k) \sim k^{-\gamma}, \tag{1}$$

where γ is a characteristic degree exponent, usually in the range $2<\gamma<3$. A major role is especially played by the scale-free nature of many real complex networks, which implies a large connectivity heterogeneity, at the basis of the peculiar behavior shown by dynamical processes taking place on top of these networks, such as the resilience to random damage [12–14], the spreading of infectious agents [15–18], or diffusion-annihilation processes [19,20].

From a theoretical point of view, the empirical research has inspired the proposal of new network models, aimed at reproducing and explaining the properties exhibited by complex networks. In this respect, many efforts have been devoted to develop models capable of accounting for a scale-free degree distribution. Classical network modeling was previously based on the Erdös-Renyi random graph model [21,22], which is a static model (i.e., defined for a fixed number of vertices N) yielding small-world networks with a Poisson degree distribution. A change of perspective in network modeling took place after the introduction of the preferential attachment paradigm first proposed by Barabási and Albert (BA) [23]. The insight behind this concept is the realization of two facts. First, most complex networks are the result of a growth process, in which new vertices are added in time to the system. Second, new edges are not placed at random, but tend to connect to vertices that already have a large degree. It turns out that these two ingredients are able to reproduce scale-free degree distributions with a tunable degree exponent [23,24]. Moreover, it has been shown that not all sorts of preferential attachment are able to generate a power-law degree distribution, but only those in which new edges attach to vertices with a probability strictly linear in their degree [25], and that some alternative mechanisms, such as the copying model [26], implicitly define a linear preferential attachment dynamics.

While a proper characterization and understanding of the origin of the scale-free degree distribution displayed by most real complex networks is a fundamental task, it has been recently realized that this property does not provide a suffient characterization of natural networks. In fact, these systems seem to exhibit also ubiquitous degree correlations, which translate into the fact that the degrees of the vertices at the

end points of any given edge are not independent [27–30]. Two vertex degree correlation can be conveniently measured by means of the conditional probability $P(k'|k)$ that a vertex of degree k is connected to a vertex of degree k'. For uncorrelated networks, in which this conditional probability is independent of k, it can be estimated as the probability that any edge end points to a vertex of degree k', which is simply given by $P_{\rm nc}(k'|k)=k'P(k')/\langle k\rangle$ [31]. The empirical evaluation of $P(k'|k)$ in real networks is usually a cumbersome task, restricted by finite size data, yielding noisy results. For this reason, it is more practical to analyze instead the average degree of the nearest neighbors of the vertices of degree k, which is formally defined as [27]

$$\overline{k}_{\rm NN}(k) = \sum_{k'} k' P(k'|k). \tag{2}$$

For uncorrelated networks, in which $P(k'|k)$ does not depend on k, we have

$$\overline{k}^{\rm nc}_{\rm NN}(k) = \frac{\langle k^2\rangle}{\langle k\rangle}, \tag{3}$$

independent of k. Thus, a $\overline{k}_{\rm NN}(k)$ function with an explicit dependence on the degree signals the presence of two vertex degree correlations in the network. When $\overline{k}_{\rm NN}(k)$ is an increasing function of k, the network shows *assortative mixing* [29] (vertices of large degree connected preferably with vertices of large degree, and vice versa). Negative correlations (low degree vertices connected preferably with large degree vertices), on the other hand, give rise to *disassortative mixing*, detected by a decreasing $\overline{k}_{\rm NN}(k)$ function.

Analogously to two vertex correlations, correlations implying three vertices can be mesured by means of the probability $P(k',k''|k)$ that a vertex of degree k is simultaneously connected to vertices of degree k' and k''. Again, the difficulties in directly estimating this conditional probability can be overcome by analyzing the clustering coefficient. The average clustering coefficient of the vertices of degree k (the clustering spectrum), $\overline{c}(k)$ [28,32], can be formally computed as the probability that a vertex of degree k is connected to vertices of degree k' and k'', and that those two vertices are at the same time joined by an edge, averaged over all the possible values of k' and k'' [31]. Thus, we can write $\overline{c}(k)$ as a function of the three vertex correlations as

$$\overline{c}(k) = \sum_{k',k''} P(k',k''|k) p^k_{k',k''}, \tag{4}$$

where $p^k_{k',k''}$ is the probability that vertices k' and k'' are connected, provided that they have a common neighbor k.[1] From this expression, the average clustering coefficient can be computed as

$$C = \sum_k P(k)\overline{c}(k). \tag{5}$$

For uncorrelated networks, we have that $P_{\rm nc}(k',k''|k) = P_{\rm nc}(k'|k)P_{\rm nc}(k''|k)$ [31], and $p^k_{k',k''}=(k'-1)(k''-1)/\langle k\rangle N$ [33]. Therefore we obtain

$$\overline{c}_{\rm nc}(k) = \frac{(\langle k^2\rangle - \langle k\rangle)^2}{\langle k\rangle^3 N}. \tag{6}$$

That is, $\overline{c}_{\rm nc}(k)$ is independent of k and equal to the average clustering coefficient C [33]. A functional dependence of $\overline{c}(k)$ on the degree can thus be attributed to the presence of a structure in the three vertex correlations. In particular, for scale-free networks it has been observed that in many instances, the clustering spectrum exhibits also a power-law behavior $\overline{c}(k)\sim k^{-\alpha}$. A value of α close to 1 has been empirically observed in several natural networks, and analytically found in some growing network models [32,34,35]. These findings have led to propose the clustering spectrum $\overline{c}(k)$ as a tool to measure hierarchical organization and modularity in complex networks [32].

The presence of correlations is thus a very relevant issue in order to understand and classify complex networks, especially in view of the important consequences that they can have on dynamical processes taking place on the topology defined by the networks [36–38]. While there are quite a few empirical results for real networks, the situation is not the same for network models, and therefore there is no consensus regarding the origin of assortative and dissasortative mixing, and its relation to the power law behavior of the clustering spectrum $\overline{c}(k)$. In fact, most works devoted to analytical calculations of correlations in complex network models have been performed only for particular cases [31,32,34,35,39,40]. In this respect, a noteworthy development is the rate equation formalism proposed by Szabó *et al.* in Ref. [34] (see also [41]) to compute $\overline{c}(k)$ in growing network models with preferential attachment. However, to our knowledge, no such formalism has been developed to deal with two vertex correlations, as given by the $\overline{k}_{\rm NN}(k)$ function.

In this paper we revise the formalism proposed in Ref. [34] for computing the clustering spectrum in growing network models with preferential attachment. Reconsidering the mean field rate equation in the continuous degree approximation for the $\overline{c}(k)$ presented in [34], we are able to provide a general expression for the boundary condition of this rate equation, which was neglected in the original treatment and which can have as a matter of fact relevant effects in the final solution, as we will show below. Inspired by this result we also propose a different rate equation for two vertex correlations, as measured by the $\overline{k}_{\rm NN}(k)$ function, and work out the correponding boundary condition. We remark that both equations are valid in general for the so-called citation networks [9], in which neither edge or vertex removal nor edge rewiring is allowed. Also, and due to the fact that the equations are formulated in the continuous degree approximation, we expect them to provide accurate results only in the limit of large k, especially in the case of scale-free networks. The general formalism is presented in Sec. II. The rate equations

[1] Note that the probability $p^k_{k',k''}$ can depend on the degree k of the common vertex.

obtained can be easily solved for growing networks with linear preferential attachement (LPA) [24], as shown in Sec. III. In particular, we are able to obtain expressions for the dependence of the correlations on the degree k and the system size N, for both the dissasortative and assortative regimes of the model, which are in very good agreement with numerical simulations and previous scaling arguments [42]. LPA models generate networks with a vanishing average clustering coefficient C. In order to assess the effects of a nonzero clustering, we study in Sec. IV a growing model presenting a large final clustering coefficient [43], which we are able to compute with very good accuracy. The results obtained are qualitatively similar to those shown by the Holme-Kim model [34,44]. The rate equation proposed for two vertex correlations can be easily generalized to deal with more involved situations. As an example of its flexibility, we examine in Sec. V a recently proposed model for the evolution of weighted complex networks [45]. In this case, we extend our formalism to compute a function estimating weighted two vertex correlations, in which the actual strength of the interactions between neighboring vertices is taken into account. Our results allow us to discuss the scaling form of two and three vertex correlation functions, and signal the possible relations that can be established between them.

II. RATE EQUATIONS FOR CORRELATIONS IN GROWING NETWORKS

Let us consider the class of growing network models in which, at each time step, a new vertex with m edges is added to the network. For the vertex introduced at time t, each of its emanating edges is connected to an existing vertex introduced at time s ($s<t$) with a connection probability $\Pi_s(\{k\},t)$, which is assumed to depend only on the degrees of the existing vertices at time t, $\{k\}=\{k_1(t),\ldots,k_{t-1}(t)\}$. Time runs from 1 to N (the final network size), and since for each new vertex m edges are added, the average degree is fixed and given by $\langle k\rangle=2m$. In the continuous k and t approximation [42], the average degree that the vertex s (i.e., the vertex introduced at time s) has at time t ($t>s$) can be computed from the rate equation

$$\frac{dk_s(t)}{dt} = m\Pi_s(\{k\},t), \tag{7}$$

with the boundary condition $k_s(s)=m$ (initially all vertices have m connections). From $k_s(t)$, the degree distribution can be obtained as

$$P(k,t) = -\frac{1}{t}\left(\frac{\partial k_s(t)}{\partial s}\right)^{-1}\Bigg|_{s=s(k,t)}, \tag{8}$$

where $s(k,t)$ is the solution of the implicit equation $k=k_s(t)$.

For this class of networks it is possible to obtain a rate equation for the clustering spectrum. Following Ref. [34], we recall that the clustering coefficient $c_s(t)$ of vertex s at time t is defined as the ratio between the number of edges between the neighbors of s and its maximum possible value. Then, if $M_s(t)$ is the number of connections between the neighbors of s at time t, we have that

$$c_s(t) = \frac{2M_s(t)}{k_s(t)[k_s(t)-1]}. \tag{9}$$

During the growth of the network, $M_s(t)$ can only increase by the simultaneous addition of an edge to s and one of its neighbors. Therefore, in the continuous k approximation, we can write down the following rate equation [34]:

$$\frac{dM_s(t)}{dt} = m(m-1)\Pi_s(\{k\},t)\sum_{j\in\mathcal{V}(s)}\Pi_j(\{k\},t), \tag{10}$$

where $\mathcal{V}(s)$ is the set of neighbors of vertex s. In order to solve this equation we must provide additionally a boundary condition. To do so, we observe that $M_s(s)$ is the number of triangles created by the introduction of vertex s. Therefore

$$M_s(s) = \frac{m(m-1)}{2}\sum_{j,n=1}^{s}\Pi_j(\{k\},s)\Pi_n(\{k\},s)\Pi_{j,n}, \tag{11}$$

that is, it is proportional to the probability that s is connected to vertices j and n, times the probability $\Pi_{j,n}$ that j and n are linked, averaged over all vertices j and n existing in the network at time s. The probability $\Pi_{j,n}$ is given by

$$\Pi_{j,n} = \Theta(j-n)m\Pi_n(\{k\},j) + \Theta(n-j)m\Pi_j(\{k\},n), \tag{12}$$

where $\Theta(x)$ is the Heaviside step function. Solving the equation for $M_s(t)$ with the boundary condition Eq. (11), we can obtain the clustering $c_s(t)$ from Eq. (9). Then, since in growing network models in the continuous k approximation the degree at time t is uniquely determined by the introduction time s, from $c_s(t)$ we can directly obtain the clustering spectrum $\overline{c}(k)$ as a function of k and the largest time $t=N$.

In the case of the two vertex correlation function $\overline{k}_{\mathrm{NN}}(k)$, we can proceed along similar lines. Let us define $R_s(t)$ as the sum of the degrees of the neighbors of vertex s, evaluated at time t. That is,

$$R_s(t) = \sum_{j\in\mathcal{V}(s)} k_j(t). \tag{13}$$

The average degree of the neighbors of vertex s, $\overline{k}_{\mathrm{NN}}(s)$, is then given by $\overline{k}_{\mathrm{NN}}(s)=R_s(t)/k_s(t)$. During the growth of the network, $R_s(t)$ can only increase by the adjunction of a new vertex connected either directly to s, or to a neighbor of s. In the first case $R_s(t)$ increases by an amount m (the degree of the newly linked vertex), while in the second case it increases by one unit. Therefore, in the continuous k approximation, we can write down the following rate equation:

$$\frac{dR_s(t)}{dt} = m[m\Pi_s(\{k\},t)] + m\sum_{j\in\mathcal{V}(s)}\Pi_j(\{k\},t). \tag{14}$$

In order to obtain the boundary condition for this equation,

we observe that, at time s, the new vertex s connects to an old vertex of degree $k_j(s)$ with probability $\Pi_j(\{k\},s)$, and that this vertex gains a new connection in the process. Therefore,

$$R_s(s) = m\sum_{j=1}^{s} \Pi_j(\{k\},s)[k_j(s)+1]. \tag{15}$$

From the solution of this rate equation, we can obtain $\overline{k}_{\rm NN}(s)$ and from it the two vertex correlation function by the functional dependence of s on k and $t=N$.

We must note that Eqs. (10)–(12), (14), and (15) are valid only for the so-called citation networks, in which neither edge removal nor rewiring [46] is allowed, since these two processes can induce nonlocal variations in the conectivity of the nearest neighbors.

III. LINEAR PREFERENTIAL ATTACHMENT MODELS

As an example of the application of the rate equations presented in the previous section, we consider the general LPA model proposed in Ref. [24], for which the rate equations for $R_s(t)$ and $M_s(t)$ can be closed and solved analytically. For general LPA, the connection probability takes the form

$$\Pi_s(\{k\},t) = \frac{b_1 k_s(t) + b_2}{\sum_j [b_1 k_j(t) + b_2]}, \tag{16}$$

where b_i are real constants. Since, for each new vertex, m edges are added to the network, the normalization constant in Eq. (16) takes the form $\Sigma_j[b_1k_j(t)+b_2]=(2mb_1+b_2)t$. Thus, the model depends only on the tuning parameter $a=b_2/b_1$, taking values in the interval $a\in\,]-m,\infty[$ [since the minimum degree is m, a cannot be lower than $-m$ in order for $\Pi_s(\{k\},t)$ to remain positive]. Thus, the connection probability for the LPA model reads

$$\Pi_s(\{k\},t) = \frac{k_s(t)+a}{(2m+a)t}. \tag{17}$$

Solving the rate equation for the degrees Eq. (7), we obtain

$$k_s(t) = (m+a)\left(\frac{t}{s}\right)^{\beta} - a, \quad \beta = \frac{m}{2m+a}. \tag{18}$$

Therefore, this model yields networks with a power-law degree distribution of the form

$$P(k) \sim k^{-\gamma}, \quad \gamma = 3 + a/m. \tag{19}$$

For $a>0$, we obtain a degree exponent $\gamma>3$, which corresponds to finite degree fluctuations in the thermodynamic limit. The case $a=0$ recovers the original BA model with $\gamma=3$ [23]. Finally, values $-m<a<0$ yield scale-free networks with a tunable degree exponent, in the range $\gamma\in\,]2,3[$.

A. Two vertex degree correlations

The rate equation for $R_s(t)$ takes in this case the form

$$\frac{dR_s(t)}{dt} = m^2\frac{k_s(t)+a}{(2m+a)t} + \sum_{j\in\mathcal{V}(s)} m\frac{k_j(t)+a}{(2m+a)t}$$
$$= \beta\frac{(m+a)k_s(t)+am}{t} + \beta\frac{R_s(t)}{t}, \tag{20}$$

where we have used the definition of $R_s(t)$, Eq. (13). The general solution of the previous equation is

$$R_s(t) = A_0(s)t^{\beta} + \beta(m+a)^2\left(\frac{t}{s}\right)^{\beta}\ln t + a^2 \tag{21}$$

where $A_0(s)$ is given by the boundary condition $R_s(s)$. From Eq. (15), we have that

$$R_s(s) = m\sum_{j=1}^{s}\frac{a+(a+1)k_j(s)+k_j^2(s)}{(2m+a)s}$$
$$= \beta a + 2m\beta(a+1) + \frac{\beta}{s}\sum_{j=1}^{s}k_j^2(s). \tag{22}$$

Plugging $k_j(s)=(m+a)(s/j)^{\beta}-a$ into $R_s(s)$ results in

$$R_s(s) = m(1-a) + \beta(m+a)^2 s^{2\beta-1}\sum_{j=1}^{s} j^{-2\beta}. \tag{23}$$

In order to estimate the behavior of the previous expression, we have to distinguish the different cases corresponding to the possible values of β (namely, a).

(i) $-m<a<0$ *(i.e.,* $\beta>1/2$, $\gamma<3$*).* In this case, for large s, $\Sigma_{j=1}^{s}j^{-2\beta}\simeq\zeta(2\beta)$, where $\zeta(x)$ is the Riemann zeta function, and thus, at leading order,

$$R_s(s) \simeq \beta\zeta(2\beta)(m+a)^2 s^{2\beta-1}. \tag{24}$$

The determination of the integration constant $A_0(s)$ from Eq. (21) yields then

$$R_s(t) \simeq \beta\zeta(2\beta)(m+a)^2 t^{\beta}s^{\beta-1} + \beta(m+a)^2\left(\frac{t}{s}\right)^{\beta}\ln\left(\frac{t}{s}\right), \tag{25}$$

where terms independent of t and s and terms going to zero in the large t or s limit have been neglected. From the definition of $\overline{k}_{\rm NN}(s)$, and substituting s as a function of k and $t=N$ (the network final size) in the limit of large k and N we obtain the following expression for the average degree of the neighbors of the vertices of degree k:

$$\overline{k}_{\rm NN}(k,N) \simeq \beta\zeta(2\beta)(m+a)^{3-1/\beta}N^{2\beta-1}k^{-2+1/\beta}$$
$$+ (m+a)\ln\left(\frac{k}{m+a}\right). \tag{26}$$

From this expression, we conclude that the LPA with $a<0$ yields in the large N limit networks with disassortative two vertex correlations, characterized by a power-law decay $\overline{k}_{\rm NN}(k,N)\sim N^{2\beta-1}k^{-2+1/\beta}$. This exponent was previously ob-

tained by scaling arguments in Ref. [42]. The dependence of the prefactor on N implies that $\overline{k}_{\mathrm{NN}}(k,N)$ diverges in the thermodynamic limit $N\to\infty$, in agreement with the theoretical arguments provided in Ref. [47]. For finite N, however, the logarithmic term with constant prefactor can induce corrections to the power-law scaling. Since $2\beta-1$ is at most 1, the growth of $\overline{k}_{\mathrm{NN}}(k,N)$ is not very steep with N and these corrections are observable in numerical simulations, as we will see below in this section.

(ii) $a=0$ (i.e., $\beta=1/2$, $\gamma=3$). For this value of β Eq. (23) is dominated by a logarithmic divergence $\Sigma_{j=1}^{s} j^{-1}\simeq \ln s$, yielding

$$R_s(s)\simeq \frac{m^2}{2}\ln s. \tag{27}$$

From here, we obtain

$$R_s(t)\simeq \frac{m^2}{2}\sqrt{\frac{t}{s}}\ln t, \tag{28}$$

and finally

$$\overline{k}_{\mathrm{NN}}(k,N)\simeq \frac{m}{2}\ln N. \tag{29}$$

That is, two vertex correlations are independent of the degree and grow with the system size as $\ln N$, in agreement with the behavior expected for an uncorrelated scale-free network with degree exponent $\gamma=3$, Eq. (3). Numerical simulations of the BA model [28] show actually a very weak dependence on k in the $\overline{k}_{\mathrm{NN}}(k,N)$ function, compatible nevertheless with the behavior given by our rate equation approach in the large k limit. This k dependence, evidentiated at small values of the degree, cannot be detected within our approach, since it has been formulated in the continuous k approximation.

(iii) $a>0$ (i.e., $\beta<1/2$, $\gamma>3$). In this situation, the summation in Eq. (23), $\Sigma_{j=1}^{s} j^{-2\beta}\simeq s^{1-2\beta}/(1-2\beta)$, is divergent, and therefore $R_s(s)$ becomes independent of s. This leads to

$$R_s(t)\simeq \beta(m+a)^2\left(\frac{t}{s}\right)^{\beta}\ln\left(\frac{t}{s}\right) + \left[m(1-a)+\frac{\beta(m+a)^2}{1-2\beta}-a^2\right]\left(\frac{t}{s}\right)^{\beta}, \tag{30}$$

and finally the dominant behavior for the correlation function is

$$\overline{k}_{\mathrm{NN}}(k,N)\simeq (m+a)\ln\left(\frac{k}{m+a}\right). \tag{31}$$

In this case, $\overline{k}_{\mathrm{NN}}(k,N)$ is independent of the network size, and increases logarithmically with k: For $\gamma>3$, LPA yields networks with weak assortative mixing.

B. Three vertex correlations

In order to estimate three vertex degree correlations by means of the clustering spectrum $\overline{c}(k)$, we start from the rate equation Eq. (10), which for the LPA model takes the form

$$\frac{dM_s(t)}{dt}=m(m-1)\frac{k_s(t)+a}{(2m+a)t}\sum_{j\in\mathcal{V}(s)}\frac{k_j(t)+a}{(2m+a)t} = m(m-1)\frac{k_s(t)+a}{(2m+a)^2t^2}[R_s(t)+ak_s(t)]. \tag{32}$$

The boundary condition $M_s(s)$ can be written as

$$\begin{aligned} M_s(s) &= \frac{m(m-1)}{2}\sum_{j,n}\Pi_j(\{k\},s)\Pi_n(\{k\},s)\Pi_{j,n} \\ &= \frac{\beta^2(m-1)(m+a)^3}{2(2m+a)}s^{2\beta-2} \times\left\{\sum_{n=1}^{s} n^{-2\beta}\sum_{j=n+1}^{s} j^{-1}+\sum_{j=1}^{s} j^{-2\beta}\sum_{n=j+1}^{s} n^{-1}\right\} \\ &= \frac{\beta^2(m-1)(m+a)^3}{2(2m+a)}s^{2\beta-2}\times 2\sum_{n=1}^{s} n^{-2\beta}\sum_{j=n+1}^{s} j^{-1}. \end{aligned} \tag{33}$$

In order to solve Eq. (32), we approximate $k_s(t)$ and $R_s(t)$ by their dominant terms for large t and s, as computed above for the different possible values of a.

(i) $-m<a<0$. In this case we have

$$k_s(t)\simeq (m+a)\left(\frac{t}{s}\right)^{\beta},\quad R_s(t)\simeq \beta\zeta(2\beta)(m+a)^2t^{\beta}s^{\beta-1}. \tag{34}$$

Introducing this expression into Eq. (32), we obtain at leading order

$$M_s(t)\simeq \beta^2\frac{(m-1)(m+a)^3\zeta(2\beta)}{(2\beta-1)(2m+a)}(t^{2\beta-1}-s^{2\beta-1})s^{-1}+M_s(s). \tag{35}$$

In order to compute $M_s(s)$, we observe that the double summation in Eq. (33) takes the form at large s

$$\mathcal{S}=\sum_{n=1}^{s} n^{-2\beta}\sum_{j=n+1}^{s} j^{-1}\simeq\sum_{n=1}^{s} n^{-2\beta}(\ln s-\ln n)\simeq\zeta(2\beta)\ln s, \tag{36}$$

since $\Sigma_{n=1}^{\infty} n^{-2\beta}\ln n$ is convergent for $\beta>1/2$. Thus we obtain

$$M_s(t)\simeq \beta^2\frac{(m-1)(m+a)^3\zeta(2\beta)}{(2\beta-1)(2m+a)}(t^{2\beta-1}-s^{2\beta-1})s^{-1} + \beta^2\frac{(m-1)(m+a)^3}{2m+a}\zeta(2\beta)s^{2\beta-2}\ln s, \tag{37}$$

and from here the expression for the three vertex correlation function follows:

$$\overline{c}(k,N) \simeq \frac{2\beta^2(m-1)(m+a)^{3-1/\beta}\zeta(2\beta)}{(2\beta-1)(2m+a)} N^{2\beta-2}k^{-2+1/\beta} + \frac{2\beta^2\zeta(2\beta)(m-1)(m+a)^{5-2/\beta}}{2m+a}(\ln N)N^{2\beta-2}k^{-4+2/\beta}. \tag{38}$$

To understand the asymptotic behavior of $\overline{c}(k,N)$, two limits have to be taken, corresponding to large N and large k: (1) At fixed and large N, the leading behavior at large k is $\overline{c}(k,N) \sim N^{2\beta-2}k^{-2+1/\beta}$; (2) at fixed $k \lesssim (\ln N)^{\beta/(2\beta-1)}$ and large N, the leading behavior is instead $\overline{c}(k,N) \sim N^{2\beta-2} \ln N k^{-4+2/\beta}$. Therefore, in the numerical simulations we should expect to observe a crossover between these two scaling regimes.

(ii) $a=0$. We now have

$$k_s(t) \simeq m\sqrt{\frac{t}{s}}, \quad R_s(t) \simeq \frac{m^2}{2}\sqrt{\frac{t}{s}}\ln t, \tag{39}$$

which yields

$$M_s(t) \simeq \frac{m^2(m-1)}{16s}(\ln^2 t - \ln^2 s) + M_s(s). \tag{40}$$

Since $\beta=1/2$, $M_s(s)$, as given by Eq. (33), can be easily shown to be

$$M_s(s) = \frac{m^2(m-1)}{16s}\ln^2 s, \tag{41}$$

and we obtain

$$M_s(t) \simeq \frac{m^2(m-1)}{16s}\ln^2 t, \tag{42}$$

which results in a clustering coefficient at large N

$$\overline{c}(k,N) \simeq \frac{m-1}{8}\frac{\ln^2 N}{N}. \tag{43}$$

We recover the well-known result for the BA model that the clustering spectrum is constant, and scaling as $(\ln^2 N)/N$, as observed in Ref. [34]. It is worth noting that the computation of the boundary condition (41) is essential in recovering this result. Interestingly, from Eq. (42) we can also compute the total number of triangles in BA networks as a function of the network size, $T(N)$, i.e.,

$$T(N) \simeq \frac{1}{3}\int_1^N M_s(N)ds \simeq \frac{m^2(m-1)}{48}\ln^3 N, \tag{44}$$

where the factor $1/3$ comes from the fact that in the computation each triangle is seen once by each of its three vertices. This value can be compared to the exact result obtained in Ref. [48], namely,

$$T(N) \simeq \frac{m(m-1)(m+1)}{48}\ln^3 N. \tag{45}$$

We can see here that, even though our rate equation approach captures the correct scaling with N, it underestimates the value of the numerical prefactor, due to the continuous degree approximation.

(iii) $a>0$. For this range of values of a we have

$$k_s(t) \simeq (m+a)\left(\frac{t}{s}\right)^{\beta}, \quad R_s(t) \simeq \beta(m+a)^2\left(\frac{t}{s}\right)^{\beta}\ln\left(\frac{t}{s}\right), \tag{46}$$

yielding

$$M_s(t) \simeq \beta^2\frac{(m-1)(m+a)^3}{(2m+a)(1-2\beta)}s^{-2\beta}\left\{-t^{2\beta-1}\ln\left(\frac{t}{s}\right) + \frac{s^{2\beta-1}-t^{2\beta-1}}{1-2\beta}\right\} + M_s(s). \tag{47}$$

For the evaluation of $M_s(s)$, we observe that the double summation $\mathcal{S}$ defined in Eq. (36) shows now a power-law divergence $\mathcal{S} \simeq s^{1-2\beta}/(1-2\beta)^2$. Thus we have

$$M_s(t) \simeq \beta^2\frac{(m-1)(m+a)^3}{(2m+a)(1-2\beta)}s^{-2\beta}\left\{-t^{2\beta-1}\ln\left(\frac{t}{s}\right) + \frac{2s^{2\beta-1}-t^{2\beta-1}}{1-2\beta}\right\}, \tag{48}$$

yielding a three vertex correlation function

$$\overline{c}(k,N) \simeq \frac{4\beta^2(m-1)(m+a)^{3-1/\beta}}{(2m+a)(1-2\beta)^2}N^{-1}k^{-2+1/\beta}. \tag{49}$$

Therefore, for $a>0$ (i.e., $\beta<1/2$), we obtain that the average clustering of the vertices of degree k is a growing function of k, scaling as $\overline{c}(k,N) \sim N^{-1}k^{-2+1/\beta}$. Since by definition the clustering must be smaller than 1, this growing behavior must be restricted to degree values $k \lesssim N^{\beta/(1-2\beta)}$.

C. Computer simulations

In order to check the analytical results obtained in this section, we have performed extensive numerical simulations of the LPA model. Simulations were performed for system sizes ranging from $N=10^3$ to 10^6, averaging over 100 network samples for each value of N and a. We focus in particular in the ranges $a<0$ and $a>0$, which have not been previously explored (for numerical data corresponding to $a=0$, the BA model, see Refs. [28,49]).

In Figs. 1 and 2 we explore the behavior of networks generated for $a<0$. We consider first the average degree of the nearest neighbors $\overline{k}_{\mathrm{NN}}(k,N)$. Figure 1(a) corresponds to $m=4$, $a=-2$, values that yield $\beta=2/3$ and $\gamma=2.5$, while Fig. 1(b) plots data for $m=4$, $a=-3$, corresponding to $\beta=4/5$ and $\gamma=2.25$. The dashed lines represent the power-law behavior $k^{-2+1/\beta}$ expected analytically. We observe that, as the size of the network increases, the data follow the predicted scaling $\overline{k}_{\mathrm{NN}}(k,N) \sim N^{2\beta-1}k^{-2+1/\beta}$ on larger and larger ranges. Nevertheless, the logarithmic corrections present in Eq. (26) are clearly visible from the large k deviations shown by the data (middle plots in Fig. 1). The logarithmic correction can, in fact, be taken into account if one rescales $\overline{k}_{\mathrm{NN}}(k,N)$ appropriately; namely, if we define

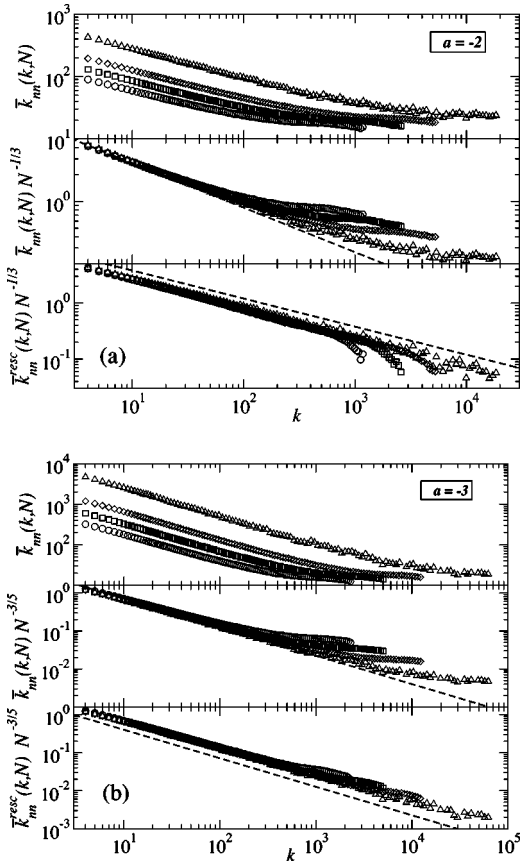

FIG. 1. Average degree for the nearest neighbors of the vertices of degree k, $\overline{k}_{\rm NN}(k,N)$, for the LPA model for $m=4$, with $a=-2$ (a) and -3 (b). Symbols correspond to the different system sizes $N=10^4$ (○), 3×10^4 (□), 10^5 (◇), and 10^6 (△). Top plots: Raw data. Middle plots: Data rescaled by the size prefactor $N^{1-2\beta}$. Bottom plots: Data rescaled by the size prefactor with logarithmic corrections. The dashed lines represent a power-law decay with exponent $-2+1/\beta$.

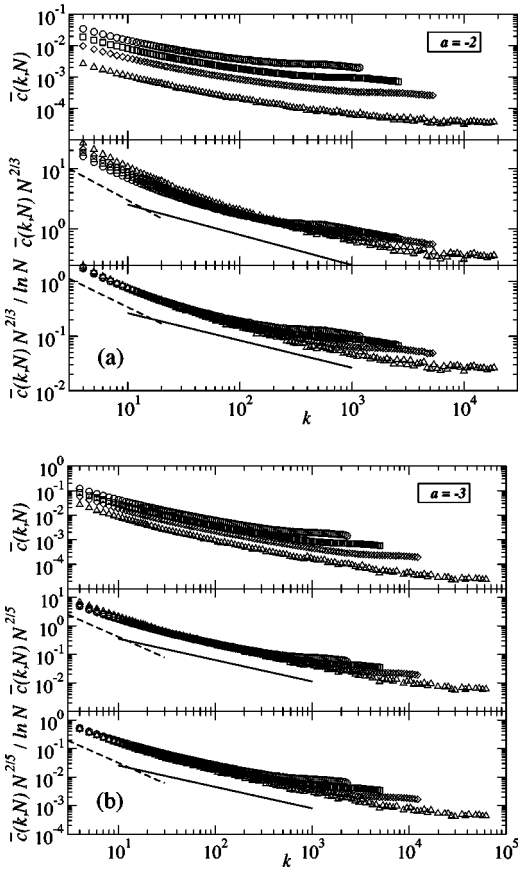

FIG. 2. Clustering spectrum $\overline{c}(k,N)$ for the LPA model for $m=4$, with $a=-2$ (a) and -3 (b). Symbols correspond to the different system sizes $N=10^4$ (○), $N=3\times10^4$ (□), $N=10^5$ (◇), and $N=10^6$ (△). Top plots: Raw data. Middle plots: Data rescaled by the size prefactor $N^{2-2\beta}$, corresponding to large k. Bottom plots: Data rescaled by the size prefactor $N^{2-2\beta}/\ln(N)$, corresponding to small k. The full and dashed lines represent power-law decays with exponent $-2+1/\beta$ and $-4+2/\beta$, respectively.

$$\overline{k}^{\rm resc}_{\rm NN}(k,N) = \overline{k}_{\rm NN}(k,N) - (m+a)\ln\left(\frac{k}{m+a}\right), \qquad (50)$$

then, from Eq. (26), we expect

$$\overline{k}^{\rm resc}_{\rm NN}(k,N)N^{1-2\beta} \sim k^{-2+1/\beta}. \qquad (51)$$

In the bottom plots of Fig. 1 we draw the rescaled average degree of the nearest neighbors with logarithmic corrections. The collapse of the data is indeed surprisingly good, given the numerous approximations and leading order cancellations made in our calculations. The remaining discrepancy at very large k is presumably due to the subdominant terms we have neglected.

In Fig. 2 we represent the clustering spectrum $\overline{c}(k,N)$ for the same parameters as before, i.e., $m=4$, $a=-2$ (a) and $m=4$, $a=-3$ (b). The top plots represent the corresponding nonrescaled raw data. According to the solution provided in Eq. (38), for small values of k an asymptotic scaling is expected of the form $\overline{c}(k,N)\sim N^{2\beta-2}\ln Nk^{-4+2/\beta}$. This behavior is approximately recovered in the bottom plots in Fig. 2 for both values of a, where we can see that the first points in the graphics for different values of N collapse onto the same curve, with approximately the predicted k dependence. For large values of k, on the other hand, we expect instead a scaling $\overline{c}(k,N)\sim N^{2\beta-2}k^{-2+1/\beta}$, which is again recovered in the middle plots of this figure, showing a better collapse in the intermediate range of k values. At very large k values, finally, the neglected logarithmic terms come into play, affecting the scaling of the data. It is important to notice the important role played by the boundary condition Eq. (11), which is responsible for the second term in Eq. (38), giving the correct scaling behavior for small k.

In Fig. 3 we finally explore the average degree of the nearest neighbors (a) and the clustering spectrum (b) for the LPA model with $a>0$. We focus in particular on the values $m=4$ and $a=2$ (top plots), yielding $\beta=2/5$, $\gamma=3.5$; $a=5$ (middle plots), with $\beta=4/13$, $\gamma=4.25$; and $a=10$ (bottom plots), which corresponds to $\beta=2/9$, $\gamma=5.5$. For the $\overline{k}_{\rm NN}(k,N)$ function our theoretical analysis predicts a function independent of the network size, and slowly (logarithmically) growing with the degree. These predictions are con-

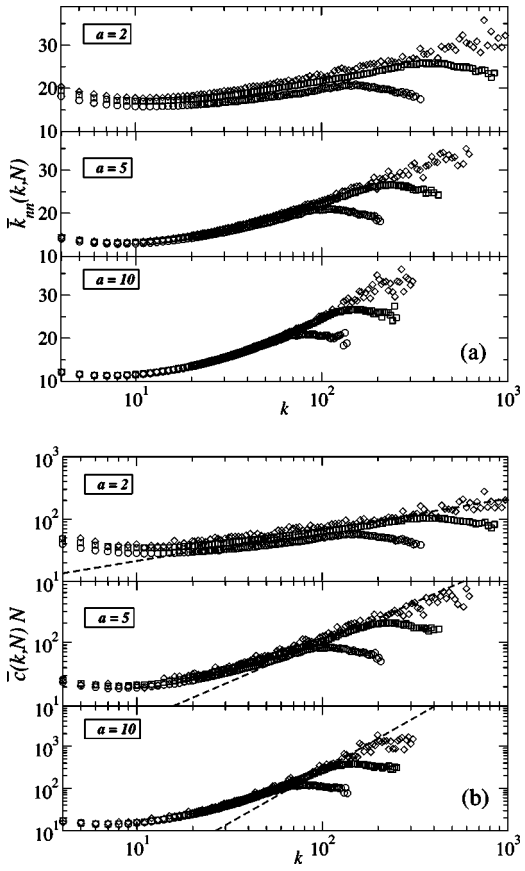

FIG. 3. Average degree for the nearest neighbors of the vertices of degree k, $\overline{k_{\mathrm{NN}}}(k,N)$ (a) and clustering spectrum $\overline{c}(k,N)$ (b) for the LPA with $m=4$ and positive values of a. Symbols correspond to the different system sizes $N=10^4$ (○), $N=10^5$ (□), and $N=10^6$ (◇). Top plots: $a=2$. Middle plots: $a=5$. Bottom plots: $a=10$. Data for $\overline{c}(k,N)$ have been rescaled by the theoretical size prefactor N. The dashed lines represent a power-law behavior with exponent $-2+1/\beta$.

firmed in Fig. 3(a). It is noteworthy that the theoretical prediction becomes more accurate for large a: While the collapse is quite good for $a \geqslant 5$, the graphs are a bit scattered for the smallest value of a considered. This fact is due to the slow convergence (as N grows) to the theoretical asymptotic form for small a. Analogously, the clustering spectrum shows the predicted scaling $\overline{c}(k,N) \sim N^{-1}k^{-2+1/\beta}$ [Fig. 3(b)]. The dependence on system size is correctly captured by our analysis for larger values of a. In this range, however, the power-law dependence on k seems to depart from the theoretical exponent $-2+1/\beta$. This apparent departure can be due to the limited range of degrees for such large values of the degree exponent (the degree range decreases for increasing a), as well as to the subdominant terms neglected in the asymptotic expression Eq. (49).

IV. GROWING NETWORKS WITH LARGE CLUSTERING

As we have seen in the previous section, the LPA model yields a clustering spectrum $\overline{c}(k)$ that, even if presenting a nontrivial scaling, vanishes in the thermodynamic limit, i.e., $\lim_{N\to\infty}\overline{c}(k,N)=0$. However, for many complex networks, such as the internet [2], we observe a function $\overline{c}(k)$ scaling with k, together with a finite clustering.

Several models have been proposed which reproduce this feature. In particular, Dorogovtsev, Mendes, and Samukhin (DMS) introduced in Ref. [43] a scale-free growing network with large clustering coefficient C. The model is defined as follows: At each time step, a vertex is added and connected to *the two extremities of a randomly chosen edge*, thus forming a triangle. The resulting network has a power-law degree distribution $P(k)\sim k^{-3}$, with $\langle k\rangle=4$, and since each new vertex induces the creation of at least one triangle, we expect this model to generate networks with finite clustering coefficient. We consider here a generalization of the DMS model, in which every new node is connected to the extremities of $m/2$ randomly chosen *edges*, where m is an even number. The original model corresponds thus to $m=2$, and this generalization allows one to tune the average degree, setting it to $\langle k\rangle=2m$.

It is important to notice that this model actually contains the LPA mechanism in a disguised form. Indeed, the probability to choose a vertex s is clearly proportional to the number of edges arriving at s, i.e., to its degree k_s. At time t there are mt edges so that $\Sigma_s k_s=2mt$ and the probability to choose s when choosing one edge is $k_s/(mt)$ [$\Sigma_s k_s/(mt)=2$ since one chooses indeed two vertices]. This process is repeated $m/2$ times and thus at each time step the probability to choose s is $k_s/(2t)$.

This shows that another way of formulating the random choice of an edge is in fact the following: a vertex s is chosen with the usual preferential attachment probability $k_s/(2mt)$, and then one of its neighbors is chosen at random, i.e., with probability $1/k_s$.

It is then clear that the rate equation for the degree is given by

$$\frac{dk_s(t)}{dt}=\frac{k_s(t)}{2t}, \tag{52}$$

leading to $k_s(t)=m(t/s)^{1/2}$ and to a scale-free degree distribution of the form $P(k)\approx 2m^2k^{-3}$.

We are now in position to write down the rate equations for the network correlations, taking into account that, each time a vertex is chosen, so is one of its neighbors.

A. Two vertex degree correlations

At each time step, $R_s(t)$ can increase either if the vertex s is chosen (and then R_s increases by $m+1$ because a neighbor of s is also chosen), or if a neighbor j is chosen together with a neighbor l of j, with $l\neq s$ (and then R_s increases by 1). Therefore, we have that

$$\frac{dR_s(t)}{dt}=(m+1)\frac{k_s(t)}{2t}+\sum_{j\in\mathcal{V}(s)}\frac{k_j(t)}{2t}\left(1-\frac{1}{k_j(t)}\right)$$
$$=\frac{mk_s(t)}{2t}+\frac{R_s(t)}{2t}. \tag{53}$$

This is exactly the same equation as for the LPA with $a=0$,

i.e., the BA model. Moreover, the boundary condition for $R_s(s)$ can be written as

$$R_s(s) = \frac{m}{2}\sum_{j=1}^{s} \frac{k_j(s)}{2ms}\left\{ k_j(s) + 1 + \sum_{l\in\mathcal{V}(j)} \frac{1}{k_j(s)}[k_l(s)+1]\right\}, \tag{54}$$

where, for each one of the $m/2$ edges chosen by s, the first term corresponds to the contribution of j [chosen with probability $k_j/(2ms)$], and the second term to the contribution of a neighbor l of j (chosen with probability $1/k_j$). This expression is easily reduced to

$$R_s(s) = \frac{1}{2}\sum_{j=1}^{s} \frac{k_j(s)[k_j(s)+1]}{s} \simeq \frac{m^2}{2}\ln s. \tag{55}$$

Once again we obtain the same result as for the LPA with $a=0$. The conclusion is that the $\overline{k}_{\mathrm{NN}}(k,N)$ function for the generalized DMS model is given by Eq. (29): the two vertex correlations are independent of the degree and grow with the network size as $\ln N$, in the same fashion as in the BA model.

B. Three vertex correlations

In order to write down the rate equation for $M_s(t)$, we have to take into account that, at each time step, $m/2$ triangles are formed by the choice of $m/2$ edges, and that, moreover, additional triangles may be formed for $m>2$ by choosing two different edges with a common vertex. At each time step, the increase in $M_s(t)$ is thus given by two terms. The first one comes from choosing the vertex s with probability $k_s(t)/(2t)$. In this case, M_s increases by 1, since one of the neighbors of s is also chosen. The second contribution comes from the following situation: one of the edges chosen is $s-l$, and another one is $j-l'$, with $j\in\mathcal{V}(s)$, $j\neq l$, and $l'\neq s$.

The resulting rate equation reads

$$\begin{aligned}\frac{dM_s(t)}{dt} &= \frac{k_s(t)}{2t} + \frac{m}{2}\left(\frac{m}{2}-1\right)\frac{k_s(t)}{mt}\sum_{j\in\mathcal{V}(s)}\frac{k_j(t)}{mt}\left(1-\frac{1}{k_s(t)}\right)\\ &= \frac{k_s(t)}{2t} + \frac{m-2}{4mt^2}[k_s(t)-1]R_s(t).\end{aligned} \tag{56}$$

We use $k_s \simeq m\sqrt{t/s}$ and $R_s \simeq m^2 \ln t\sqrt{t/s}/2$ to obtain

$$\frac{dM_s(t)}{dt} \simeq \frac{m}{2\sqrt{ts}} + \frac{m^2(m-2)\ln t}{8st}, \tag{57}$$

whose solution reads

$$M_s(t) \simeq m\left(\sqrt{\frac{t}{s}}-1\right) + \frac{m^2(m-2)}{16s}(\ln^2 t - \ln^2 s) + M_s(s). \tag{58}$$

The boundary condition is again given by two contributions. First, $m/2$ triangles are created by attaching s to $m/2$ edges. The second contribution is given by

$$\frac{m}{2}\left(\frac{m}{2}-1\right)\frac{1}{2}\sum_{j=1}^{s}\sum_{l\in\mathcal{V}(j)}\frac{k_j(s)}{ms}\left(1-\frac{1}{k_j(s)}\right)\frac{k_l(s)}{ms}, \tag{59}$$

i.e., the sum over all vertices j of the probability that, among the $m/2$ edges chosen by the new node s, one has j for extremity, and another one has a neighbor l of j for extremity [the factor $1/2$ is due to the double counting of the links (j,l) and (l,j)]. This yields

$$M_s(s) = \frac{m}{2} + \frac{m-2}{8ms^2}\sum_{j=1}^{s} R_j(s)[k_j(s)-1] \simeq \frac{m}{2} + \frac{m^2(m-2)\ln^2 s}{16s} \tag{60}$$

and finally

$$M_s(t) \simeq m\sqrt{\frac{t}{s}} - \frac{m}{2} + \frac{m^2(m-2)}{16s}\ln^2 t. \tag{61}$$

The clustering spectrum can therefore be written as

$$\overline{c}(k,N) \simeq \frac{2k-m}{k(k-1)} + \frac{m-2}{8N}\ln^2 N. \tag{62}$$

The clustering spectrum is now finite in the infinite size limit,

$$\overline{c}(k) = \lim_{N\to\infty}\overline{c}(k,N) \simeq \frac{2k-m}{k(k-1)}. \tag{63}$$

It is interesting to see that, for the original model with $m=2$, the finite size corrections actually vanish and we obtain the result $\overline{c}(k,N)=2/k$, independent of N. This scaling is also similar to that obtained for the Holme-Kim model in [34]. The knowledge of the exact form of the degree distribution for $m=2$, $P(k)=12/[k(k+1)(k+2)]$ [43] allows us to obtain the average clustering coefficient $C(m=2)=2\pi^2-19 \approx 0.739$. More generally, for large m, approximating $P(k)$ by $2m^2/k^3$ and sums by integrals yields

$$\begin{aligned}C(m) &= \int_m^\infty P(k)\overline{c}(k)dk \simeq 2m^2 - 3m - 4/3\\ &\quad + 2m^2(2-m)\ln\left(\frac{m}{m-1}\right).\end{aligned} \tag{64}$$

C. Computer simulations

We have performed extensive numerical simulations of the generalized DMS model studied in this section. We focus on the clustering spectrum $\overline{c}(k,N)$ since the results for $\overline{k}_{\mathrm{NN}}(k)$ are expected to be equal to the case of the BA model. Figure 4(a) shows the excellent agreement between the predicted behavior Eq. (62) and the numerical data for various values of m and sizes ranging from $N=10^4$ to 10^6. As expected, no finite size corrections are present for $m=2$, while they are correctly described by the analytical approach for larger m. Moreover, the prediction for the average clustering coefficient $C(m)$, Eq. (64), is also shown to be in excellent agreement with numerical data, Fig. 4(b).

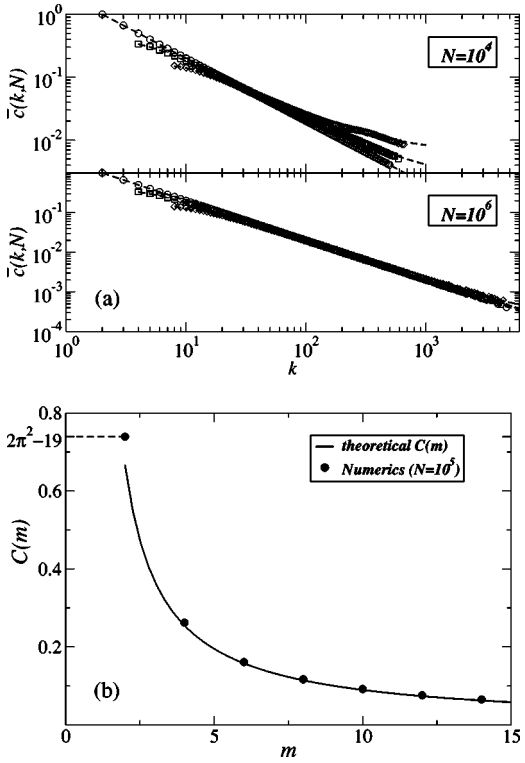

FIG. 4. (a) Clustering spectrum $\overline{c}(k,N)$ for the generalized DMS model. The top plot corresponds to a system size $N=10^4$; the bottom plot is for $N=10^6$. Symbols correspond to different values of the average degree $m=2$ (○), 4 (□), and 8 (◇). The dashed lines are the theoretical predictions given by Eq. (62). (b) Average clustering coefficient $C(m)$ as a function of m, for networks of size $N=10^5$. The full line represents the theoretical prediction Eq. (64). The dashed line marks the theoretical value for $m=2$, $C(2)=2\pi^2-19$.

V. WEIGHTED GROWING NETWORKS

In the previous sections we have applied the rate equation formalism to analyze two and three vertex correlations in standard models with either vanishing or constant clustering coefficient. The formalism for the two vertex correlations, however, is not limited to these particular cases, and can be easily extended to analyze more complex growing network models. As an example, in this section we will consider a recently proposed growing weighted network model [45]. Weighted networks [50] are a natural generalization of graphs in which a real quantity is assigned to each edge, representing the importance or weight w_{ij} of the interaction between the vertices i and j. Recently [51], it has been pointed out that real weighted networks present a complex architecture, characterized by broad distributions of weights, as well as nontrivial correlations between the values of the weights and the topological structure of the network.

Motivated by these findings, Ref. [45] proposed a dynamic growing weighted network model, in which new edges are attached to old vertices with a connection probability depending on the strength, or total weight, of the vertex. In order to define the model, let us consider a weighted network characterized by the elements w_{ij} defining the weight assigned to the edge connecting vertices i and j. We assume the elements w_{ij} to be symmetric, that is, $w_{ij}=w_{ji}$. Each vertex i is characterized by both its degree k_i and its strength σ_i, defined as

$$\sigma_i = \sum_{j \in \mathcal{V}(i)} w_{ij}. \tag{65}$$

For nonweighted networks, in which $w_{ij}=1$, we obviously recover $\sigma_i=k_i$. The model proposed in Ref. [45] considers a growing network in which at each time step, a new vertex is added to the system and connected with m edges to older vertices. The probability that the new vertex t is connected to s $(s<t)$ is given by the connection probability

$$\Pi_s(\{s\},t) = \frac{\sigma_s(t)}{\sum_j \sigma_j(t)}, \tag{66}$$

that is, linearly proportional to the strength of the old vertex s. Each new edge carries an initial weight $w_0=1$. Additionally, there is a dynamic rearrangement of the weights belonging to the edges of the receiving vertex: When the vertex s receives a new connection, the weight of its edges is increased by an amount

$$w_{sj} \to w_{sj} + \delta \frac{w_{sj}}{\sigma_s}, \quad j \in \mathcal{V}(s). \tag{67}$$

This rule implies that, for each new vertex added, the total strength of the network is increased by an amount $2m+2m\delta$; therefore, the normalization constant in Eq. (66) is $\Sigma_j \sigma_j(t)=2m(1+\delta)t$. It can be shown, within the continuous k approximation [45], that this model generates scale-free networks, characterized by the quantities

$$\sigma_s(t) = m\left(\frac{t}{s}\right)^{\beta}, \quad k_s(t) = \frac{\sigma_s(t)+2m\delta}{2\delta+1}, \quad P(k) \sim k^{-\gamma}, \tag{68}$$

with exponents

$$\beta = \frac{2\delta+1}{2\delta+2}, \quad \gamma = \frac{4\delta+3}{2\delta+1}. \tag{69}$$

Therefore, for $\delta>0$, this model yields power-law degree distributions with degree exponent $\gamma \in]2,3[$ and $\beta>1/2$. The case $\delta=0$ recovers the BA model.

It is easy to see that, at the level of the mean field rate equations in the continuous k approximation, the weighted growing network model described above can be mapped into a growing network with LPA and negative parameter a given by

$$a = -\frac{2m\delta}{2\delta+1}. \tag{70}$$

Therefore, we expect to observe the two and three vertex correlation functions

$$\overline{k}_{\mathrm{NN}}(k,N) \simeq \frac{m\zeta(2\beta)}{2(1+\delta)}\left(\frac{m}{2\delta+1}\right)^{2-1/\beta} N^{2\beta-1}k^{-2+1/\beta} + \frac{m}{2\delta+1}\ln\left(\frac{2\delta+1}{m}k\right), \quad (71)$$

$$\overline{c}(k,N) \simeq \frac{(m-1)(2\delta+1)^2}{4\delta(\delta+1)^2}\left(\frac{m}{2\delta+1}\right)^{2-1/\beta} N^{2\beta-2}k^{-2+1/\beta} + \zeta(2\beta)\frac{(m-1)(2\delta+1)^2}{4(\delta+1)^3} \times\left(\frac{m}{2\delta+1}\right)^{4-2/\beta}(\ln N)N^{2\beta-2}k^{-4+1/\beta}. \quad (72)$$

A. Weighted two vertex correlations

The definition of the $\overline{k}_{\mathrm{NN}}(k)$ function we have computed above completely neglects the effect of the weights. Therefore, it provides a biased view of real correlations in the system (for example, two neighbors with the same degree but widely different weights give the same contribution). In order to take into account the effect of the weights associated with the edges, a different correlation measure has been proposed, the weighted average degree of the nearest neighbors $\overline{k}^{w}_{\mathrm{NN}}(k)$ [51], defined as follows:

$$\overline{k}^{w}_{\mathrm{NN}}(s) = \frac{1}{\sigma_s(t)}\sum_{j\in\mathcal{V}(s)} w_{sj}(t)k_j(t). \quad (73)$$

This definition implies that $\overline{k}^{w}_{\mathrm{NN}}(s) > \overline{k}_{\mathrm{NN}}(s)$ if the edges with largest weight point to the neighbors with largest degree, while $\overline{k}^{w}_{\mathrm{NN}}(s) < \overline{k}_{\mathrm{NN}}(s)$ in the opposite case. Therefore, $\overline{k}^{w}_{\mathrm{NN}}(s)$ measures the effective affinity to connect with large or small degee neighbors, according to the magnitude of the interaction weight. The weighted average degree of the nearest neighbors $\overline{k}^{w}_{\mathrm{NN}}(k)$, is defined as the average of $\overline{k}^{w}_{\mathrm{NN}}(s)$ for all the vertices with the same degree k.

We can study analytically the weighted two vertex correlations by seeking a rate equation for the quantity

$$Q_s(t) = \sum_{j\in\mathcal{V}(s)} w_{sj}(t)k_j(t). \quad (74)$$

According to the rules defining the model, at each time step $Q_s(t)$ can increase its value by two mechanisms: (1) If a new vertex is directly attached to s, $Q_s(t)$ increases by an amount $m+\delta Q_s/\sigma_s$; (2) if a new vertex is attached to a neighbor j of s, then $Q_s(t)$ increases by $w_{sj}+\delta w_{sj}/\sigma_j+\delta w_{sj}k_j/\sigma_j$. Therefore, the rate equation satisfied by $Q_s(t)$ is

$$\frac{dQ_s(t)}{dt} = m\Pi_s(\{\sigma\},t)\left(m+\frac{\delta}{\sigma_s(t)}Q_s(t)\right) + \sum_{j\in\mathcal{V}(s)} m\Pi_j(\{\sigma\},t) \times\left(w_{sj}+\delta\frac{w_{sj}}{\sigma_j(t)}+\delta w_{sj}\frac{k_j(t)}{\sigma_j(t)}\right), \quad (75)$$

which, in terms of $\sigma_s(t)$ and $Q_s(t)$, yields

$$\frac{dQ_s(t)}{dt} = \left(\beta+\frac{\delta}{1+\delta}\right)\frac{Q_s(t)}{t} + \frac{m+\delta-2m\delta}{2(1+\delta)}\frac{\sigma_s(t)}{t}. \quad (76)$$

Inserting the value of $\sigma_s(t)$ given by Eq. (68), the general solution of this equation is

$$Q_s(t) = A_0(s)t^{\beta+\delta/(1+\delta)} - \frac{m}{2\delta}(m+\delta-2m\delta)\left(\frac{t}{s}\right)^{\beta}. \quad (77)$$

Since all new edges have an initial weight $w_0=1$, the initial condition for $Q_s(t)$ coincides with that of $R_s(t)$. Solving for $A_0(s)$ from Eq. (24), substituting for the corresponding value of a given by Eq. (70), we finally obtain in the large k and N limit

$$\overline{k}^{w}_{\mathrm{NN}}(k,N) \simeq \frac{m\zeta(2\beta)}{2(1+\delta)(2\delta+1)}N^{2\beta-1}. \quad (78)$$

That is, in this model the weighted average degree of the nearest neighbors is independent of k, signaling the absence of two vertex weighted correlations, as indeed found numerically in Ref. [52]. There is, however, a scaling with the system size, given by the factor $N^{2\beta-1}$, which is the same as that found for the nonweighted correlations for the same value of γ.

B. Computer simulations

We have performed numerical simulations of the weighted growing network model described in Ref. [45], for sizes ranging from $N=10^3$ to 10^5, focusing on the behavior of the average degree of the nearest neighbors, for both its nonweighted and weighted versions. In Fig. 5 we plot the average degree of the nearest neighbors $\overline{k}_{\mathrm{NN}}(k,N)$ for $m=2$ and $\delta=2$ (a) which corresponds to a network with $\beta=5/6$, $\gamma=2.20$, and $\delta=5$ (b), that yields $\beta=11/12$, $\gamma=2.09$. As expected from the analytical analysis performed above, the obtained scaling is analogous to the LPA model: the numerical data follow the predicted form $\overline{k}_{\mathrm{NN}}(k,N)\sim N^{2\beta-1}k^{-2+1/\beta}$. The bottom plots highlight the presence of the logarithmic correction of Eq. (71), by plotting the rescaled function

$$\overline{k}^{\mathrm{resc}}_{\mathrm{NN}}(k,N) = \overline{k}_{\mathrm{NN}}(k,N) - \frac{m}{2\delta+1}\ln\left(\frac{2\delta+1}{m}\right). \quad (79)$$

In this case, it is noticeable that the rescaled $\overline{k}^{\mathrm{resc}}_{\mathrm{NN}}(k,N)$ function with logarithmic corrections yields a better data collapse than that shown by the LPA model. Even though both models are identical at the mean field level, the existing microscopic differences seem to yield smaller subleading corrections for the weighted growing network model.

For this same set of parameters, we have also evaluated the weighted average degree of the nearest neighbors, $\overline{k}^{w}_{\mathrm{NN}}(k,N)$, shown in the middle plot in Fig. 5(a) (filled symbols). We observe that the $\overline{k}^{w}_{\mathrm{NN}}(k,N)$ is indeed, as expected, independent of k, and scales with the system size with the predicted factor $N^{2\beta-1}$.

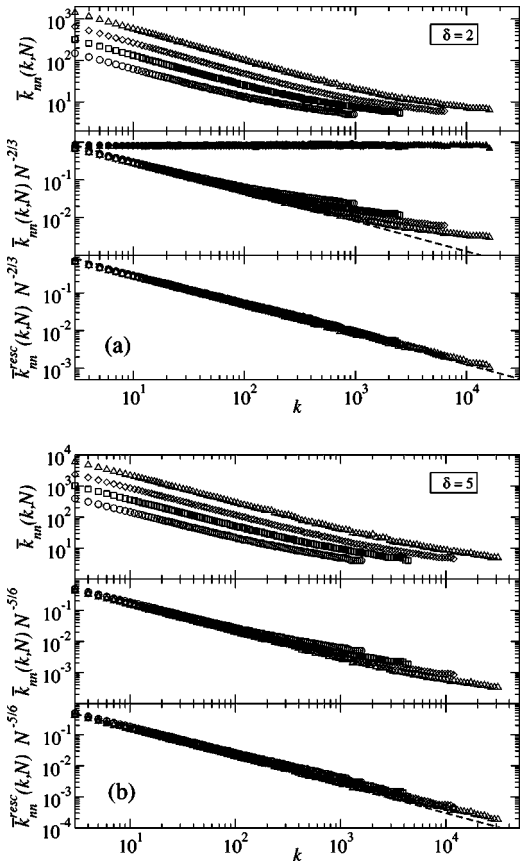

FIG. 5. Average degree for the nearest neighbors of the vertices of degree k, $\overline{k}_{\mathrm{NN}}(k,N)$, for the weighted growing model for $m=4$, with $\delta=2$ (a) and 5 (b). Symbols correspond to the different system sizes $N=3\times10^3$ (○), $N=10^4$ (□), $N=3\times10^4$ (◇), and $N=10^5$ (△). Top plot: Raw data. Middle plot: Data rescaled by the size prefactor $N^{1-2\beta}$. Bottom plot: Data rescaled by the size prefactor with logarithmic corrections. The dashed lines represent a power-law decay with exponent $-2+1/\beta$. The middle plots display also the values of $N^{1-2\beta}\overline{k}^{w}_{\mathrm{NN}}(k,N)$ (filled symbols), which collapse onto a horizontal line, in agreement with the analytical prediction Eq. (78).

VI. CONCLUSIONS

A complete theoretical characterization of a growing network model should imply not only the estimation of the corresponding degree distribution, but also an analytical study of the functional form of the correlations between the degrees of neighboring vertices. Capitalizing on the work of Szabó *et al.* [34,41], in this paper we have provided a formalism to compute two vertex correlations, expressed by means of the average degree of the nearest neighbors of the vertices of degree k, $\overline{k}_{\mathrm{NN}}(k)$, valid for growing network models generated by means of the preferential attachment mechanism and belonging to the class of so-called citation networks. The formalism is based on a rate equation in the continuous k approximation, together with the appropriate boundary condition, that can be easily solved in the case in which the preferential attachment is linear in the degree. Additionally, we have presented a more complete description of the rate equation determining the clustering spectrum $\overline{c}(k)$, by discussing the effects of boundary conditions. Applying this framework to several growing network models, we have obtained asymptotic expressions for the functions $\overline{k}_{\mathrm{NN}}(k,N)$ and $\overline{c}(k,N)$, showing both the degree dependence and the scaling with the system size, due to finite size effects. As a general result, we conclude that networks generated by LPA with degree exponent $\gamma<3$ exhibit the scaling behavior

$$\overline{k}_{\mathrm{NN}}(k,N) \sim N^{2\beta-1}k^{-2+1/\beta}, \tag{80}$$

previously obtained by means of scaling arguments [42], which is the signature of disassortative (negative) two vertex correlations. We have also been able to identify the presence of logarithmic corrections in models with LPA, which clearly appear in computer simulations of the model. For this LPA model, we also observe the presence of small assortative correlations for degree exponents $\gamma>3$, characterized by a logarithmic growth of the $\overline{k}_{\mathrm{NN}}(k,N)$ function, which is otherwise independent of the network size. The situation is more complex in what concerns the clustering spectrum $\overline{c}(k,N)$. For $\gamma>3$, we observe the presence of a crossover between two power-law decays in the degree, $\overline{c}(k)\sim k^{-\alpha}$, with $\alpha=-4+2/\beta$ for $k\lesssim(\ln N)^{\beta/(2\beta-1)}$, and $\alpha=-2+1/\beta$ in the asymptotic limit, while for $\gamma>3$ we obtain an increasing $\overline{c}(k,N)$ function, limited by an upper degree cutoff.

From these results we can conclude that the value $\alpha\simeq1$ observed in the literature [32,34] is not a generic feature of all scale-free networks [41]. However, we notice that LPA yields networks with a vanishing clustering coefficient. In order to assess the possible effects of this factor, we have considered the DMS model [43], which generates networks with a large value of C, as observed in real networks. In this case, we obtain a lack of two vertex correlations, while the clustering spectrum scales as $\overline{c}(k)\sim k^{-1}$. An analogous result is obtained for the similar Holme-Kim model [34,44].

As a final point, we have shown the flexibility of the rate equation approach to compute two vertex correlations by applying it to a recently proposed weighted growing network model, in which edges are further characterized by a distribution of weights that is dynamically coupled to the evolving topology of the network. For this model, we are able to extend our formalism to deal with weighted two vertex correlations, which measure the effect of the strength of the interactions between neighboring vertices.

The very good agreement shown between our analytical estimates and numerical simulations suggest that the method proposed in this paper to compute two vertex correlations is in general valid to characterize growing citation network models. An obvious improvement would be to extend it to deal with models in which vertex and edge removal, and edge rewiring, are allowed. This inclusion, however, would probably lead to quite complex nonlocal rate equation, whose solution would be much harder to tackle.

ACKNOWLEDGMENTS

We thank M. Alava, M. Boguñá, and A. Vespignani for helpful comments and discussions. This work has been partially supported by EC-FET Open Project No. IST-2001-

33555 and Contract No. 001907 (DELIS). R.P.-S. acknowledges financial support from the Ministerio de Ciencia y Tecnología (Spain), and from the Departament d'Universitats, Recerca i Societat de la Informació, Generalitat de Catalunya (Spain).

[1] B. Bollobás, *Modern Graph Theory* (Springer-Verlag, New York, 1998).
[2] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, Cambridge, U.K., 2004).
[3] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **401**, 130 (1999).
[4] A. Wagner, Mol. Biol. Evol. **18**, 1283 (2001).
[5] H. Jeong, S. S. Mason, A. L. Barabási, and Z. N. Oltvai, Nature (London) **411**, 41 (2001).
[6] S. Maslov and K. Sneppen, Science **296**, 910 (2002).
[7] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications* (Cambridge University Press, Cambridge, U.K., 1994).
[8] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[9] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).
[10] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[11] R. Cohen and S. Havlin, Phys. Rev. Lett. **90**, 058701 (2003).
[12] R. A. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **406**, 378 (2000).
[13] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. Lett. **85**, 5468 (2000).
[14] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000).
[15] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[16] R. Pastor-Satorras and A. Vespignani, Phys. Rev. E **63**, 066117 (2001).
[17] R. M. May and A. L. Lloyd, Phys. Rev. E **64**, 066112 (2001).
[18] M. Barthélemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. Lett. **92**, 178701 (2004).
[19] L. K. Gallos and P. Argyrakis, Phys. Rev. Lett. **92**, 138301 (2004).
[20] M. Catanzaro, M. Boguñá, and R. Pastor-Satorras, e-print cond-mat/0407447.
[21] P. Erdös and P. Rényi, Publ. Math. (Debrecen) **6**, 290 (1959).
[22] B. Bollobás, *Random Graphs* (Academic Press, London, 1985).
[23] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[24] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000).
[25] P. L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000).
[26] R. Kumar *et al.*, in *Proceedings of the 41st IEEE Symposium on Foundations of Computer Science (FOCS)* (IEEE, New York, 2000), pp. 57–65.
[27] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).
[28] A. Vázquez, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. E **65**, 066130 (2002).
[29] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
[30] M. E. J. Newman, Phys. Rev. E **67**, 026126 (2003).
[31] M. Boguñá and R. Pastor-Satorras, Phys. Rev. E **68**, 036112 (2003).
[32] E. Ravasz and A.-L. Barabási, Phys. Rev. E **67**, 026112 (2003).
[33] M. E. J. Newman, in *Handbook of Graphs and Networks: From the Genome to the Internet*, edited by S. Bornholdt and H. G. Schuster (Wiley-VCH, Berlin, 2003), pp. 35–68.
[34] G. Szabó, M. Alava, and J. Kertész, Phys. Rev. E **67**, 056102 (2003).
[35] A. Vázquez *et al.*, Phys. Rev. E **67**, 046111 (2003).
[36] M. Boguñá and R. Pastor-Satorras, Phys. Rev. E **66**, 047104 (2002).
[37] Y. Moreno and A. Vázquez, Eur. Phys. J. B **31**, 265 (2003).
[38] A. Vázquez and Y. Moreno, Phys. Rev. E **67**, 015101 (2003).
[39] P. L. Krapivsky and S. Redner, Phys. Rev. E **63**, 066123 (2001).
[40] P. L. Krapivsky and S. Redner, Comput. Netw. **39**, 261 (2002).
[41] G. Szabó, M. Alava, and J. Kertész, in *Complex Networks*, Lecture Notes in Physics Vol. 650, edited by E. Ben-Naim, H. Frauenfelder and Z. Toroczkai (Springer-Verlag, Berlin, 2004).
[42] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).
[43] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. E **63**, 062101 (2001).
[44] P. Holme and B. J. Kim, Phys. Rev. E **65**, 026107 (2002).
[45] A. Barrat, M. Barthélemy, and A. Vespignani, Phys. Rev. Lett. **92**, 228701 (2004).
[46] R. Albert and A.-L. Barabási, Phys. Rev. Lett. **85**, 5234 (2000).
[47] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. Lett. **90**, 028701 (2003).
[48] B. Bollobás and O. Riordan, in *Handbook of Graphs and Networks: From the Genome to the Internet* (Ref. [33]), pp. 1–34.
[49] K. Klemm and V. M. Eguíluz, Phys. Rev. E **65**, 057102 (2002).
[50] J. Clark and D. A. Holton, *A First Look at Graph Theory* (World Scientific, Singapore, 1990).
[51] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, Proc. Natl. Acad. Sci. U.S.A. **101**, 3747 (2004).
[52] A. Barrat, M. Barthélemy, and A. Vespignani, Phys. Rev. E **70**, 066149 (2004).