Network growth model with intrinsic vertex fitness

I. E. Smolyarenko, K. Hoppe, and G. J. Rodgers

Department of Mathematical Sciences, Brunel University, Uxbridge, Middlesex UB8 3PH, United Kingdom

We study a class of network growth models with attachment rules governed by intrinsic node fitness. Both the individual node degree distribution and the degree correlation properties of the network are obtained as functions of the network growth rules. We also find analytical solutions to the inverse, *design*, problems of matching the growth rules to the required (e.g., power-law) node degree distribution and more generally to the required degree correlation function. We find that the design problems do not always have solutions. Among the specific conditions on the existence of solutions to the design problems is the requirement that the node degree distribution has to be broader than a certain threshold and the fact that factorizability of the correlation functions requires singular distributions of the node fitnesses. More generally, the restrictions on the input distributions and correlations that ensure solvability of the design problems are expressed in terms of the analytical properties of their generating functions.

I. INTRODUCTION

Numerous problems arising in a variety of social and natural sciences (biology, sociology, information systems, etc.) can be cast in terms of assemblies of interacting agents. Network theory, built on the basis of graph theory, became a powerful tool for the study of such systems in the past dozen years [1–3]. In the language of network theory, agents are represented as nodes of a graph, while edges encode interactions.

Many of the experimentally observed networks exhibit common topological features [2]. In particular, low average path lengths (small-world property) and power-law degree distributions characterize the class of networks known as scale-free networks [4]. Most of the theoretical effort is geared towards modeling and explaining these topological features of the networks.

The two broad classes of network models are static models, where the set of nodes is fixed, and dynamic network growth models. In the former case, edges are distributed among a given set of nodes according to a (typically probabilistic) rule. In effect, the model is a fixed-size (symmetric if the graph is unoriented) random adjacency matrix. In the latter case, the network is modeled as a stochastic process with both nodes and edges added sequentially according to a set of microscopic rules specific to a given model.

The simplest static model is the classical Erdős-Rényi (ER) random graph model [5], where a set of n edges is distributed among N nodes, usually with $n \sim O(N)$. Every pair of nodes receives an edge with equal probability $2n/N(N-1)$ and the resulting degree distribution $p(k)$ is Poisson in the $N \to \infty$ limit, decaying considerably faster with large k than most empirically observed networks.

A paradigmatic example of a dynamical network growth model is the Barabási-Albert (BA) preferential attachment network [6], which successfully reproduces key empirical observations: the small-world features and the power-law decay of degree distribution $p(k) \sim k^{-\alpha}$, with the exponent $\alpha = 3$. Despite its enormous success in modeling the basic features of empirically observed networks, the BA model possesses several limitations. In particular, the preferential attachment rule implies perfect knowledge on the part of the newly joining node about the current degrees of every node of the growing network, which may not be possible in some real-life situations. Possibly reflective of this limitation is the fact that empirical studies of growing networks tend to show many instances where the growth mechanism is inconsistent with preferential attachment, yet power-law degree distributions still emerge in the macroscopic limit [7].

The models that avoid the need to assume perfect knowledge about network topology are sometimes known as hidden variable models [8–12]. The linking rules for edges in such models are governed by some measure of the intrinsic quality of the nodes, commonly called *fitness*. The argument behind this type of model can be illustrated using the example of business relationships between companies: An entrepreneur is unlikely to know all buyer-supplier relationships or other contractual relations of the prospective business partner, but information on its general attractiveness (reputation, market capitalization, gross sales, etc.) may be readily available and moreover easily compared to that of other companies. Hence the perfect information aspect is still retained, but a different type of information is required compared to the preferential attachment model.

A static model of this type has been introduced in Ref. [8]. Every node i in the network is assigned a fitness value x_i, drawn from a probability distribution $\rho(x)$. An element ij of the adjacency matrix is equal to 1 with probability which depends on the fitnesses x_i and x_j of the nodes i and j via some linking function $f(x,y)$. It can be viewed as a generalization of the ER model, which in this language corresponds to all nodes having identically equal fitness.

Several variations of this model have been studied. A solution geared towards producing broad power-law distributions (which treats the distribution of degrees for nodes of the same fitness as negligibly narrow compared to the global degree distribution) has been found in Ref. [9], where it has been shown that for a given $\rho(x)$ there always exists an $f(x,y)$ such that the global degree distribution is asymptotically governed by a power law and vice versa. Another approach towards fitness-based static networks that exhibit power-law node degree distributions can also be found in [13].

Since most real-life networks feature growth in the number of nodes with time, static models have inherently restricted applicability as a modeling tool. Dynamic growth models incorporating fitness-based growth rules have been considered in Refs. [10–12]. Some of these models are based on mixed rules incorporating both the fitness and the instantaneous degree of a node, hence still requiring perfect information about current network topology. Pure fitness-based growth rules were investigated using a mixture of analytical and numerical techniques in [12], where numerical evidence was found tying power-law degree distributions to a similarly broad distribution of node fitnesses.

Dynamic fitness-based network models are the subject of the present paper. We obtain the full analytical description of the degree distribution and degree correlation for a class of fitness-based growth models. We show that the resulting degree distribution is parametrized by a single function, the distribution of node linking propensities.

The proposed scheme provides both the solution to the problem of determining the degree distribution given the growth rules and the closed form solution to the design problem of finding the growth rules that would model a network given the degree distribution and correlation properties. We show that this class of network growth models has an upper bound on how fast the degree distribution can decay at large k: not faster than $e^{k \ln(1-q/2)}$, where $0 < q \leqslant 1$ is the asymptotic ratio of the number of nodes to the number of edges. In the class of models considered below, the networks are sparse, so this ratio is $O(1)$. In accordance with previous findings for models with preferential attachment rules [4], the qualitative explanation for this bound on the degree distribution is the fact that a growing network is comprised of nodes that have different ages and therefore different amounts of time to acquire their neighborhoods. We also show that for the same reason, node degree correlations in growing networks of this type are almost always present and possess distinctive properties.

II. MODEL

Following the standard convention, the growth model is formulated in discrete time. At every time step t, a new edge is added to the network in one of the following two ways.

(i) With probability q, a new node is created and endowed with fitness x drawn from a fixed distribution $\rho(x)$. It is joined to the network by adding an edge connecting the new node to a target node inside the network. The probability to choose a particular target node i for this connection is $f(x,x_i)/\sum_{j=1}^{N(t)} f(x,x_j)$, where $f(x,y)$ is the linking function that depends on two fitness variables, x_i is the fitness of the target node, and $N(t)$ is the total number of nodes in the network at time t.

(ii) With probability $1 - q$, a new edge is added connecting two existing nodes in the network. A particular pair of nodes ij is chosen with the probability $f(x_i,x_j)/\sum_{m<n}^{N(t)} f(x_m,x_n)$.

We consider only the case of finite $q \leqslant 1$, hence the total number of edges at any time is $O(N(t))$ and the probability of adding a second edge connecting an already connected pair of nodes is $O(1/N(t))$, which is negligible in the leading approximation. We also do not consider oriented graphs, hence the linking function $f(x,y)$ is assumed to be a symmetric

function of its arguments. When $q = 1$, however, the growth rules explicitly differentiate between new and existing nodes, therefore this symmetry assumption is not essential in the formalism below and may be relaxed.

Consider now an existing node with fitness x. The probability that it receives a link at time t to a node whose fitness lies in the interval $[y, y + dy)$ is

$$
\gamma(x,y,t)dy
$$
$$
= \left[q \frac{f(x,y)\rho(y)}{\sum_{j=1}^{N(t)} f(x_j,y)} + (1-q) \frac{f(x,y)\sigma(y,t)}{\sum_{i<j}^{N(t)} f(x_i,x_j)} \right] dy, \quad (1)
$$

where $\sigma(x,t) = \sum_{i=1}^{N(t)} \delta(x - x_j)$ is the instantaneous node fitness density and the sums run over all nodes in existence at time t. The first term accounts for the connections to new nodes and the second term gives the contribution of the rewiring process of adding edges between existing nodes.

In the long-time limit $t \to \infty$, the random variable $N(t)$ is normally distributed with $\mathbb{E}[N(t)] = qt$ and variance equal to $q(1-q)t$. All considerations below will be made in the leading order in $1/t$, therefore it is sufficient to set $N(t) \to qt$. It is also simple to show that $\sigma(x,t) \to qt\rho(x)$ and therefore sums over nodes can be evaluated in the continuous limit as follows:

$$
\sum_{j=1}^{N(t)} f(x,x_j) = \int f(x,\xi)\sigma(\xi,t)d\xi \to qt \int f(x,\xi)\rho(\xi)d\xi,
$$
$$
(2a)
$$

$$
\sum_{i<j}^{N(t)} f(x_i,x_j) \to \frac{q^2 t^2}{2} \int f(\xi,\eta)\rho(\xi)\rho(\eta)d\xi d\eta. \quad (2b)
$$

The range of integration over the fitness variable is a matter of convention since in practice fitness serves as a quantitative proxy for some, maybe not even easily quantifiable, qualitative property of a node. We assume, however, that there is no ambiguity in relative ranking of nodes with respect to their fitness. In other words, given a group of nodes, it is always possible to rank order them by their fitness. For this assumption to hold it is sufficient to ensure that pairwise ordering of nodes by fitness is a transitive binary operation.

Under this assumption, the choice of label x for fitness is arbitrary and in the following we will use a universal label u defined for any initial choice of labeling x via the cumulative fitness distribution function $u(x) = \int^x \rho(\xi)d\xi$, $u(x_{\min}) = 0$, $u(x_{\max}) = 1$. The rank ordering assumption is sufficient to ensure that $u(x)$ is strictly monotonically increasing, hence there exists a unique inverse $x(u)$. We therefore redefine the linking function in terms of these universal ranking variables:

$$
\widetilde{f}(u,v)dudv = f(x(u),y(v))\rho(x(u))\rho(y(v))dxdy. \quad (3)
$$

Consequently, Eq. (1) takes the form

$$
\widetilde{\gamma}(u,v,t)
$$
$$
= \frac{1}{t} \left[\frac{\widetilde{f}(u,v)}{\int_0^1 \widetilde{f}(w,v)dw} + \frac{2(1-q)}{q} \frac{\widetilde{f}(u,v)}{\int_0^1 \widetilde{f}(w,w')dwdw'} \right],
$$
$$
(4)
$$

where now $\widetilde{\gamma}(u,v,t) = \widetilde{\gamma}(u,v)/t$ is the probability density that a node with ranking u receives at time t a connection to a node whose ranking is v.

The node degree distribution will be shown below to be controlled by the cumulative linking propensity $\lambda(u,t) = \int_0^1 \widetilde{\gamma}(u,v,t)dv$, which gives the overall probability for a node with ranking u to receive a new connection at time t. Since $\widetilde{\gamma}(u,v,t)$ is proportional to $1/t$ in the $t \to \infty$ limit, we define $\lambda(u) = t\lambda(u,t)$,

$$\lambda(u) = \int_0^1 \frac{\widetilde{f}(v,u)}{\int_0^1 \widetilde{f}(v,w)dw}dv + \frac{2(1-q)}{q}\frac{\int_0^1 \widetilde{f}(v,u)dv}{\int_0^1 \widetilde{f}(w,w')dwdw'}.$$

(5)

Integrating this definition over u results in the following normalization condition:

$$\int_0^1 \lambda(u)du = 1 + 2(1-q)/q.$$

(6)

The factor of 2 on the right-hand side (rhs) above arises because in the sum over all nodes of the probabilities to acquire a new connection, a link added to a pair of existing nodes is counted twice. We note that Eqs. (4) and (5) are valid to the leading order in $1/t$ in the asymptotic regime when the expected network size is large, $t \gg 1/q$.

In the description of the dynamics of the degree for an individual node the distinction between forming connections to new versus existing nodes is irrelevant and the q dependence of the normalization of $\lambda(u)$ is the only remnant of the presence of both joining and rewiring mechanisms. Other topological properties of the network, however, depend on the details of the growth rules. For example, networks with $q = 1$ are trees and degree correlation properties of such tree networks are sensitive to the full structure of the linking function f.

As will be shown in the next section, the global degree distribution p_k, given the linking propensity function $\lambda(u)$, is stationary in the long-time limit and is given by

$$p_k = \int_0^1 \frac{du}{1+\lambda(u)}\left(\frac{\lambda(u)}{1+\lambda(u)}\right)^{k-1}.$$

(7)

Before presenting the derivation of Eq. (7), we would like to discuss some of its key properties. The integrand in Eq. (7) gives the degree distribution of nodes with a given ranking u. It has the standard form of exponential distribution with the parameter $[1+\lambda(u)]^{-1}$ and therefore the mean degree of nodes with a given ranking u is

$$\mathbb{E}[k|u] \equiv \bar{k}(u) = 1 + \lambda(u).$$

(8)

It is easy to check that the global average node degree is given by

$$\bar{k} \equiv \mathbb{E}[k] = \int_0^1 du[1 + \lambda(u)] = \frac{2}{q},$$

(9)

where the normalization property (6) has been used. This is consistent with the model growth rules: The average degree is twice the ratio of the number of edges to the number of nodes, hence in the $t \to \infty$ limit, $2t/qt = 2/q$. We thus recover the standard result [4] that purely random growing trees having $\lambda(u) = \text{const} = 1$ are characterized by the exponential degree
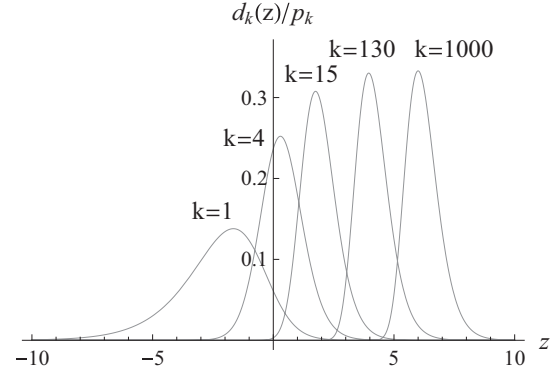

FIG. 1. Behavior of the integrand $d_k(z)$ in Eq. (11) for several values of k, normalized to p_k, for a power-law model $p_k = 1/k^\alpha \zeta(\alpha)$. The corresponding $\tilde{\rho}(\lambda)$ is given below in Eq. (21).

distribution. This conclusion is now also shown to hold for a constant $\lambda(u) = 2/q - 1$ when the rewiring process (ii) is included.

An alternative labeling of nodes is by their linking propensity λ. To avoid trivial technical issues, we will make a natural assumption that $\lambda(u)$ is a monotonic function and define the effective density $\widetilde{\rho}(\lambda) = 1/(d\lambda/du)$. The degree distribution is alternatively written as

$$p_k = \int_0^\infty \frac{\widetilde{\rho}(\lambda)d\lambda}{1+\lambda}\left(\frac{\lambda}{1+\lambda}\right)^{k-1},$$

(10)

which is better suited to analyzing the large-k behavior. In any specific model the range of possible values of λ may be smaller than the whole $\mathbb{R}_+$, however, this is easily taken into account by assuming a finite support for $\widetilde{\rho}$. Changing the variables to $\lambda = e^z$, we obtain

$$p_k = \int_{-\infty}^\infty \frac{dz\widetilde{\rho}(e^z)e^z}{1+e^z}\left(\frac{e^z}{1+e^z}\right)^{k-1} = \int_{-\infty}^\infty dz d_k(z). \quad (11)$$

Since $\widetilde{\rho}(\lambda)$ is a normalizable density, it must decay at large λ faster than $1/\lambda$, hence faster than e^{-z} as $z \to \infty$. Therefore, the integrand in p_k has the form of the product of a bounded function $e^z\widetilde{\rho}(e^z)$, with a function that is peaked at $z \sim \ln k$ and decays exponentially away from this value. The behavior of the integrand is shown in Fig. 1 for several characteristic values of k.

The width of the peak is of order 1 in the z variable. It therefore follows that if $\widetilde{\rho}$ is peaked at a much narrower scale, the distribution of degrees gets broadened to at least the shape of this peak. In the opposite limit of a power-law decay at large λ, $\widetilde{\rho}(\lambda) \sim C_\alpha \lambda^{-\alpha}$, where C_α is the normalization constant, the dominant contribution to the integral is proportional to $\widetilde{\rho}(z \sim \ln k)$. Hence the power-law behavior with the same exponent is inherited by p_k in the large-k limit, $p_k \approx C_\alpha\Gamma(\alpha)k^{-\alpha}$. The latter result has been observed numerically in [12]. This behavior will be confirmed and further investigated in the next section using the exact inversion of Eq. (10).

III. DYNAMICS OF NODE DEGREES

Since the degree of a given node at time t depends on the amount of time it had to accumulate its neighborhood, node

degree dynamics is restored from $p_k(t|u,\tau)$, the conditional probability that a node with ranking u that joined the network at time τ has degree k at time t. The evolution of $p_k(t|u,\tau)$ is governed in the large-t regime $t \gg 1/q$ by the following rate equation (see also [11,12]):

$$p_k(t+1|u,\tau) = p_{k-1}(t|u,\tau)\lambda(u,t) + p_k(t|u,\tau)[1 - \lambda(u,t)]. \tag{12}$$

We define the corresponding generating function as

$$\widetilde{G}(s,t|u,\tau) = \sum_{k=1}^{\infty} s^k p_k(t|u,\tau), \tag{13}$$

taking into account the fact that the growth rules ensure that the network is connected at all times, hence every node has at least one edge emanating from it so that $p_0(t|u,\tau) = 0$. Using the long-time asymptotics $\lambda(u,t) = \lambda(u)/t$ and approximating discrete time differences by derivatives, we find the following equation on $\widetilde{G}(s,t|u,\tau)$:

$$t\frac{\partial \widetilde{G}(s,t|u,\tau)}{\partial t} = \widetilde{G}(s,t|u,\tau)\lambda(u)(s-1). \tag{14}$$

The boundary condition is obtained from the observation that at the moment of joining the network, each node has a single connection to its parent, therefore $\lim_{t \to \tau^+} p_k(t|u,\tau) = \delta_{k1}$ and $\lim_{t \to \tau^+} \widetilde{G}(s,t|u,\tau) = s$. The equation is solved by

$$\widetilde{G}(s,t|u,\tau) = s\left(\frac{t}{\tau}\right)^{(s-1)\lambda(u)}. \tag{15}$$

The global degree distribution is obtained by averaging over the ranking of the node joining at time τ and over the times τ of joining the network. Since the incoming nodes are drawn from a time-independent distribution, the order of averaging can be reversed. Introducing $\zeta = \tau/t$, we have

$$G(s) = \int_0^1 du \int_0^t \frac{d\tau}{t}\widetilde{G}(s,t|u,\tau) = \int_0^1 du \int_0^1 d\zeta s\zeta^{(1-s)\lambda(u)}$$
$$= \int_0^1 \frac{s\,du}{(1-s)\lambda(u)+1}. \tag{16}$$

The t dependence of the global distribution function is suppressed in the notation since the result is t independent, showing that the global degree distribution of the network is stationary in the long-time limit. Expanding $G(s) = \sum_{k=1}^{\infty} p_k s^k$ and reading off the coefficients of s^k recovers the global degree distribution given in Eq. (7).

We now turn to the design problem: Given a global degree distribution p_k, the task is to determine the growth rules required to produce it, i.e., to solve Eq. (10) with respect to $\widetilde{\rho}(\lambda)$. Rewriting it in terms of the generating function, one obtains the following singular integral equation:

$$G(s) = s\int_0^{\infty} \frac{\widetilde{\rho}(\lambda)d\lambda}{(1-s)\lambda+1}. \tag{17}$$

Up to a change of variables, the rhs is the Stieltjes transform of the density $\widetilde{\rho}(\lambda)$ [14].

The general constraints on the admissible global distributions are evident from the analytical properties of the rhs of Eq. (17). Its analytical continuation outside the $|s| \leqslant 1$ circle has a cut that extends along the real axis over the

support of $\widetilde{\rho}(\lambda)$ mapped by $1 + 1/\lambda$. Also, since the average of λ weighted by $\widetilde{\rho}(\lambda)$ is $2/q - 1$, the lower edge of the cut must be at or below $2/(2-q)$. Consequently, this bounds the radius of convergence of the series defining $G(s)$ and therefore gives the bound on the rate of decay of p_k quoted in the Introduction. It is interesting to note that there does not exist a fitness-based growth mechanism within this model that would produce the degree distribution of the static ER network since the generating function of the Poisson distribution is entire.

Consider now a given global distribution whose generating function $G(s)$ has an analytical continuation that satisfies the constraints elucidated above. Except for the degenerate case of constant λ, corresponding to pure exponential decay of p_k, the generating function must have a cut along at least some portion of the ray $[1,\infty)$. The solution of the singular integral equation on $\widetilde{\rho}(\lambda)$ is obtained by evaluating the jump of the (analytically continued) $G(s)$ across this cut [14]:

$$\widetilde{\rho}(\lambda) = \frac{1}{\pi}\frac{1}{1+\lambda}\text{Im}\,G\left(1 + \frac{1}{\lambda} + i0\right). \tag{18}$$

Equation (18) provides a general solution to the design problem of finding the linking propensity that generates a preset degree distribution. In terms of the original labeling x, the linking propensity $\lambda(x)$ satisfies $\widetilde{\rho}(\lambda)(d\lambda/dx) = \rho(x)$, hence it is given by substituting Eq. (18) into the implicit relation

$$\int_0^{\lambda(x)} \widetilde{\rho}(\lambda)d\lambda = u(x). \tag{19}$$

We now investigate Eq. (18) in application to several empirically interesting cases. In the case of a pure power law $p_k = k^{-\alpha}/\zeta(\alpha)$, we have

$$G(s) = \frac{1}{\zeta(\alpha)}\text{Li}_\alpha(s) = \frac{s}{\zeta(\alpha)\Gamma(\alpha)}\int_0^{\infty} \frac{t^{\alpha-1}dt}{e^t - s} \tag{20}$$

and applying Eq. (18) we find

$$\widetilde{\rho}(\lambda) = \frac{1}{1+\lambda}\frac{1+1/\lambda}{\zeta(\alpha)\Gamma(\alpha)}\int_0^{\infty} dt\,t^{\alpha-1}\delta(e^t - 1 - 1/\lambda)$$
$$= \frac{1}{\zeta(\alpha)\Gamma(\alpha)}\frac{\ln^{\alpha-1}(1+1/\lambda)}{1+\lambda}. \tag{21}$$

Consistent with the qualitative observations earlier, $\rho(\lambda)$ decays asymptotically as $\lambda^{-\alpha}$.

As noted above, the growth rules fix the average node degree to be $2/q$, therefore, the exponent α cannot be chosen arbitrarily: $2/q = \bar{k} = \zeta(\alpha)^{-1}\sum_{k=1}^{\infty}kk^{-\alpha} = \zeta(\alpha-1)/\zeta(\alpha)$. Since $0 < q \leqslant 1$, this relation restricts α to the $(2,2.478\ldots]$ range, where $\alpha(\bar{k}=2) \approx 2.478$ and $\alpha(\bar{k} \to \infty) \to 2$. Curiously, this range is close to the range of α in empirically observed networks [2]. However, the origin of the constraint is the assumption of a pure power law at small k, which is very unlikely to be universal.

For example, the degree distribution in the BA model is $4/k(k+1)(k+2)$ [15], decaying asymptotically with $\alpha = 3$

and having $\bar{k} = 2$. The generating function in this case is

$$G(s) = \frac{2}{s^2}\left\{\frac{3}{2}s^2 - s - (1-s)^2 \ln(1-s)\right\} \quad (22)$$

and the application of Eq. (18) gives

$$\widetilde{\rho}(\lambda) = \frac{2}{(1+\lambda)^3} \quad (23)$$

for the effective density that would mimic the BA degree distribution using fitness-based growth rules. It exhibits, as expected, the matching $1/\lambda^3$ decay at large λ. Inverting the $\widetilde{\rho}(\lambda)d\lambda = du$ relation gives $\lambda_{BA}(u) = 1/\sqrt{1-u} - 1$.

To illustrate the fact that generically the degree distribution is broader than the distribution of λ, as the last example we consider the intermediate case of the stretched exponential degree distribution. Such distributions have been observed empirically [16] and they arise theoretically in sublinear preferential attachment models [15,17]. Setting $p_k = Ak^{n/2}a^{\sqrt{k}}$, where A is the appropriate normalization constant, we find

$$G(s) = A\frac{\partial^n}{\partial(\ln a)^n}\frac{s\ln a}{2\sqrt{\pi}}\int_0^{\infty}\frac{t^{-3/2}e^{-\ln^2 a/4t}dt}{s - e^t} \quad (24)$$

and correspondingly

$$\widetilde{\rho}(\lambda) = A\frac{\partial^n}{\partial(\ln a)^n}\frac{\ln a}{2\sqrt{\pi}(1+\lambda)}\ln^{-3/2}\left(1 + \frac{1}{\lambda}\right)$$
$$\times \exp\left\{-\frac{\ln^2 a}{4\ln(1+1/\lambda)}\right\}. \quad (25)$$

The asymptotic behavior at large λ is exponential, $\widetilde{\rho}(\lambda) \sim \lambda^{n+1/2}e^{-\lambda \ln^2 a/4}$, considerably narrower than the postulated stretched exponential degree distribution.

It should be noted that Eq. (17) is supplemented by restrictions on $\widetilde{\rho}(\lambda)$ stemming from its probabilistic origin. To begin with, $\widetilde{\rho}(\lambda) \geqslant 0$ for all non-negative λ. Therefore, the existence of a formal inverse given by Eq. (18) does not guarantee that there actually exists a growth model that reproduces a given degree distribution. An instructive counterexample is given by the following generating function:

$$G(s) = -\frac{12}{s^2} + \frac{20}{s} - 7 - \frac{2(6-s)(1-s)^2}{s^3}\ln(1-s)$$
$$= 4\sum_{k=1}^{\infty}\frac{5k-3}{k(k+1)(k+2)(k+3)}s^k. \quad (26)$$

The degree distribution corresponding to Eq. (26) behaves as k^{-3} at large k and is characterized by $G'(1) = \bar{k} = 4$, so a putative growth model would have $q = 1/2$. However, applying Eq. (18) gives $\widetilde{\rho}(\lambda) = (5\lambda - 1)/(\lambda + 1)^4$, which obviously changes sign at $\lambda = 1/5$. Hence a growth model satisfying the set of rules stated at the beginning of Sec. II that would reproduce the degree distribution (26) does not exist.

Further, both the normalizing integral $1 = \int_0^{\infty}\widetilde{\rho}(\lambda)d\lambda$ and the relation (6), rewritten in terms of $\widetilde{\rho}$,

$$\int_0^{\infty}\lambda\widetilde{\rho}(\lambda)d\lambda = 1 + 2(1-q)/q, \quad (27)$$

impose conditions on $\widetilde{\rho}(\lambda)$ to ensure convergence of these integrals. Therefore, the behavior of $G(s)$ as $s \to \infty$ is severely restricted by the properties of the integral in Eq. (17). If

$\int_0^{\infty}d\lambda\widetilde{\rho}(\lambda)/\lambda$ is convergent at the lower limit [it automatically converges at the upper limit since $\widetilde{\rho}(\lambda)$ has to be normalizable and the extra power of $1/\lambda$ improves convergence], we have $G(s) \to \text{const}$ as $s \to \infty$. If the integral is divergent, convergence of the normalizing integral for $\widetilde{\rho}(\lambda)$ at the lower limit bounds its rate of growth as $\lambda \to 0$ to $1/\lambda$ with logarithmic corrections and hence $G(s)$ cannot grow faster than linear with the corresponding logarithmic corrections. For example, this immediately excludes distributions whose generating functions have polynomial terms corresponding to extra spikes at one or more special values of k.

We now turn to the question of matching the growth rules to the required $\widetilde{\rho}(\lambda)$ [or, equivalently, $\lambda(u)$]. Obviously the rules are massively nonunique, as one effectively imposes a single constraint on a function of two variables. Nevertheless, resolving this constraint is not entirely trivial, as one has to simultaneously ensure that $\widetilde{f}(u,v)$ is a symmetric and positive operator.

The simplest example is to assume factorizable $\widetilde{f}(u,v) = g(u)g(v)$. Equation (5) then implies that up to an arbitrary constant factor $g(u) = \lambda(u)$ for any q and consequently $\widetilde{\gamma}(u,v) = \lambda(u)$ for $q = 1$.

Another possible simple choice is the additive ansatz $\widetilde{f}(u,v) = g(u) + g(v)$. Curiously, however, it is not always possible to satisfy the positivity requirement on $\widetilde{f}$ with this choice. Since the relation between $\lambda(u)$ and $\widetilde{f}(u,v)$ is invariant under arbitrary rescaling of $\widetilde{f}$, we can set $\int_0^1 dvg(v) = 1$. In the $q = 1$ case we then have

$$\lambda(u) = \int_0^1 dv\frac{g(u) + g(v)}{1 + g(v)} \quad (28)$$

and hence $g(u) = 1 + (\lambda(u) - 1)/H$, where $H = \int_0^1 dv/[1 + g(v)]$. We therefore obtain a self-consistency condition

$$1 = \int\frac{\widetilde{\rho}(\lambda)d\lambda}{2H - 1 + \lambda}. \quad (29)$$

If λ ranges from 0 (as it does in the examples considered above), the positivity requirement implies $H \geqslant 1$ and then the rhs of Eq. (29) is bounded by

$$\int\frac{\widetilde{\rho}(\lambda)d\lambda}{2H - 1 + \lambda} \leqslant \int\frac{\widetilde{\rho}(\lambda)d\lambda}{1 + \lambda} < \int\widetilde{\rho}(\lambda)d\lambda = 1, \quad (30)$$

making the self-consistency condition impossible to satisfy. Additive solutions are possible, however, if the support of $\widetilde{\rho}$ is limited to some range $\lambda \geqslant \lambda_{\min}$. The sufficient condition for this is that

$$\int_{\lambda_{\min}}^{\infty}\frac{d\lambda\widetilde{\rho}(\lambda)}{\lambda + 1 - 2\lambda_{\min}} \geqslant 1.$$

More generally, $\widetilde{f}(u,v)$ can be parametrized as $g(u)K(u,v)g(v)$, where $K(u,v)$ is an arbitrary symmetric positive operator and $g(u)$ is adjusted to fit the required $\lambda(u)$. Explicitly, the condition on $g(u)$ reads

$$\lambda(u) = g(u)\int_0^1 dvK(u,v)\frac{1}{\int_0^1 dwK(v,w)g(w)}. \quad (31)$$

This is a nonlinear integral equation that can be solved by iterations. Starting from an arbitrary positive $g_0(u)$, we have
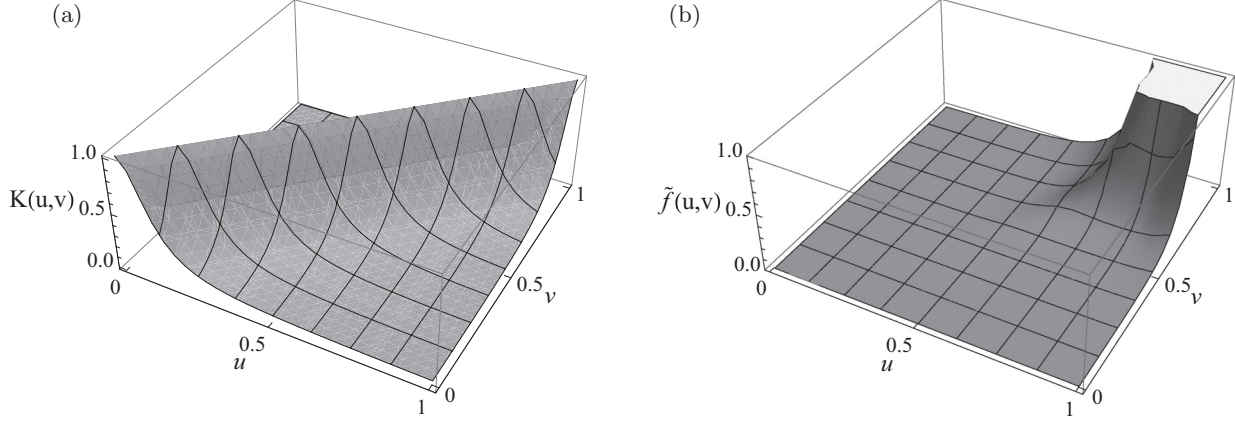
FIG. 2. Generating kernel and the linking function in the assortative case: (a) example of a generating kernel favoring assortative linking $K(u,v) = 1/[1 + 10(u - v)^2]$ and (b) linking function $\widetilde{f}(u,v)$ generated using Eq. (32) from the generating kernel in (a).

for the $(n + 1)$st iteration $g_{n+1}(u)$ the following recursion:

$$g_{n+1}(u) = \frac{\lambda(u)}{\int_0^1 dv\, K(u,v)(\int_0^1 dw\, K(v,w) g_n(w))^{-1}}. \quad (32)$$

Convergence properties of this iteration procedure require a separate investigation, however, numerical studies show that it converges fast for all reasonably smooth choices of the generating kernel K. Figures 2 and 3 show two examples of the initial generating kernel K and the corresponding linking functions $\widetilde{f}(u,v)$ obtained via iterations of Eq. (32), with $\lambda(u)$ given by inversion of Eq. (23) to match the degree distribution of the BA model. The first pair of surfaces models assortative linking between nodes with similar ranking, while the second pair corresponds to the choice of strongly disassortative behavior of the generating kernel K, both matched to produce the degree distribution of the BA model. It is seen that the linking function automatically favors connections to high-ranking nodes in order to ensure the power-law degree distribution.

IV. CORRELATION PROPERTIES OF THE NETWORK

Since each node of the network can be characterized by both its intrinsic fitness and its degree, there are two basic types of correlation function that can be defined. The simpler of those is the edge density

$$C(u,v,t) = \mathbb{E}\left[\frac{1}{2t} \sum_{i,j=1}^{N(t)} C_{ij}(t)\delta(u - u_i)\delta(v - u_j)\right]. \quad (33)$$

Here $C_{ij}(t)$ is the instantaneous adjacency matrix and the sum is normalized to the total number of edges, equal to t for any value of q. The factor of 2 accounts for double counting the edges in the unordered double sum over nodes. At the next time step, the change in C is governed by the addition to the adjacency matrix and the change in the normalization coefficient. Hence, to the leading order in $1/t$

$$\begin{aligned} C(u,v,t + 1) &= \left(1 - \frac{1}{t}\right)C(u,v,t) \\ &+ \frac{q}{2t}\left\{\frac{\widetilde{f}(u,v)}{\int \widetilde{f}(w,v)dw} + \frac{\widetilde{f}(u,v)}{\int \widetilde{f}(u,w)dw}\right\} \\ &+ \frac{1-q}{t}\frac{\widetilde{f}(u,v)}{\int \widetilde{f}(w,w')dwdw'}. \end{aligned} \quad (34)$$
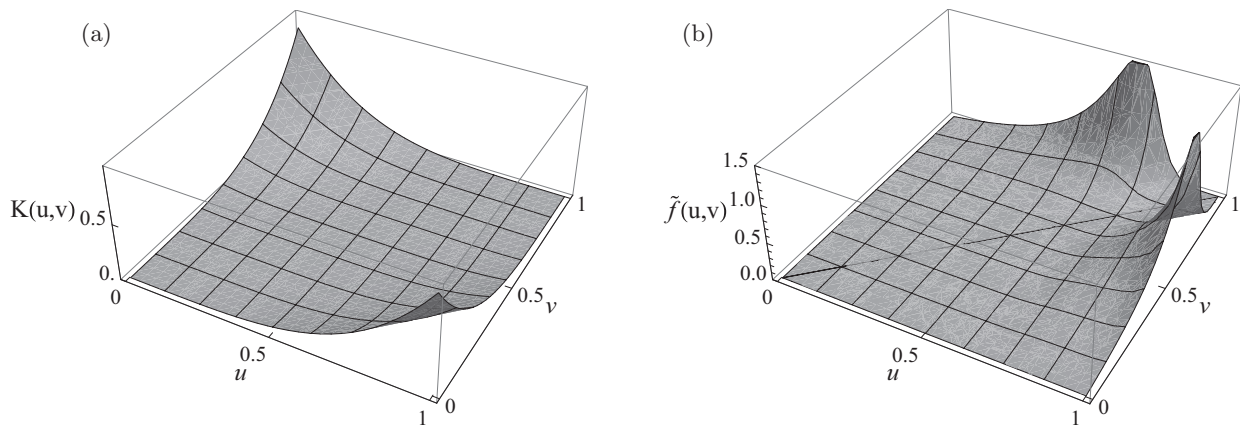

FIG. 3. Generating kernel and the linking function in the disassortative case: (a) example of a generating kernel favoring disassortative linking $K(u,v) = (u - v)^4$ and (b) linking function $\widetilde{f}(u,v)$ generated using Eq. (32) from the generating kernel in (a).

In the continuous time approximation, this takes the form

$$\frac{\partial}{\partial \ln t} C(u,v,t) = -C(u,v,t) + D(u,v), \qquad (35)$$

where

$$D(u,v) = \frac{q}{2}\left\{\frac{\widetilde{f}(u,v)}{\int \widetilde{f}(w,v)dw} + \frac{\widetilde{f}(u,v)}{\int \widetilde{f}(u,w)dw}\right\}$$
$$+ \frac{(1-q)\widetilde{f}(u,v)}{\int \widetilde{f}(w,w')dwdw'}. \qquad (36)$$

The contribution of the initial conditions decays as $1/t$ and the stationary correlation function is $C(u,v) = D(u,v)$. In the factorizable case $\widetilde{f}(u,v) = \lambda(u)\lambda(v)$ this simplifies to

$$C(u,v) = \frac{q^2}{2(2-q)}[\lambda(u) + \lambda(v)] + \frac{q^2(1-q)}{(2-q)^2}\lambda(u)\lambda(v). \quad (37)$$

In the examples considered above, such as $\lambda_{BA}(u)$, the function $\lambda(u)$ is an increasing function of ranking. Further, this property is automatically enforced if $\lambda(u)$ is obtained by solving a design problem, as evident from Eq. (19). In the corresponding factorizable models, according to Eq. (37), highly ranked nodes are much more likely to be connected to each other than mixed or low-ranked pairs. However, more general behavior is possible if $\widetilde{f}$ is nonfactorizable.

The second class of correlation functions is the degree correlation function $P_{kk'}$, which we define as the joint probability to find a node of degree k connected to a descendant of degree k'. Leaving the investigation of the topology of fitness-based networks with loops to future work, below we consider only the case $q = 1$, when the network is a random tree.

Consider a node with ranking u created at time τ. To illustrate the method, we first rederive constructively the probability $p_k(t)$ that this node has degree k at time t, which is given by the shifted Poisson distribution according to the series expansion of Eq. (15). Summing over all possible times $t_i, \tau < t_i \leqslant t$, when $k - 1$ descendant nodes could link to the original node, we obtain

$$p_k(t|u,\tau) = \frac{1}{(k-1)!} \sum_{\{t_i\}_{i=1}^{k-1}} \prod_{i=1}^{k-1} \frac{\lambda(u)}{t_i} \prod_{\substack{j=\tau+1\\j\neq t_i}}^{t} \left(1 - \frac{\lambda(u)}{j}\right),$$
$$(38)$$

where each factor gives either the probability of linking at time t_i, $\lambda(u)/t_i$, or the probability $(1 - \frac{\lambda(u)}{j})$ that no linking occurs at times j other than any of the t_i. The factorial in the prefactor compensates for equivalent permutations. Evaluating the sum in the continuous limit gives

$$p_k(t|u,\tau) = \frac{1}{(k-1)!}\left(\int_\tau^t \frac{\lambda(u)d\xi}{\xi}\right)^{k-1} \exp\left\{-\int_\tau^t \frac{\lambda(u)d\xi'}{\xi'}\right\}$$
$$= \frac{1}{(k-1)!}\left(\lambda(u)\ln\frac{t}{\tau}\right)^{k-1} e^{-\lambda(u)\ln t/\tau}, \qquad (39)$$

as expected from Eq. (15). The degree correlation function is now readily constructed by noting that any one of the $k - 1$ descendants has degree k' at time t with probability $p_{k'}(t|v,\tau')$, where v is the ranking of the descendant and τ' is its time of joining, which happens with probability $\widetilde{\gamma}(u,v)/\tau'$:

$$P_{kk'} = \frac{1}{(k-1)!}(k-1)\int_0^1 du \int_0^t \frac{d\tau}{t}\left(\lambda(u)\ln\frac{t}{\tau}\right)^{k-2}$$
$$\times e^{-\lambda(u)\ln t/\tau}\int_0^1 dv \int_\tau^t \frac{\widetilde{\gamma}(u,v)d\tau'}{\tau'} p_{k'}(t|v,\tau'). \quad (40)$$

The normalization of this expression is easily checked: $\sum_{k=2}^\infty P_{kk'} = p_{k'}$ and $\sum_{k'=1}^\infty P_{kk'} = (k-1)p_k$. The asymmetry of the normalization is due to the distinction between parent and descendant nodes and the factor $k - 1$ comes from the fact that the parent node of the node u is not included in the count of its neighbors with degree k' in the definition of $P_{kk'}$.

Substituting the explicit expression for $p_{k'}(t|v,\tau')$ and changing integration variables to $z = [1 + \lambda(u)] \ln t/\tau$, $\zeta = \lambda(v) \ln t/\tau'$, we find

$$P_{kk'} = \int_0^1 du \int_0^1 dv \frac{\widetilde{\gamma}(u,v)}{\lambda(u)\lambda(v)} \frac{1}{(k-2)!(k'-1)!}$$
$$\times \left(\frac{\lambda(u)}{1+\lambda(u)}\right)^{k-1} \int_0^\infty dz z^{k-2} e^{-z} \int_0^{z\lambda(v)/[1+\lambda(u)]}$$
$$\times d\zeta \zeta^{k'-1} e^{-\zeta}. \qquad (41)$$

A further change of variables $\zeta = \xi z/[1 + \lambda(u)]$ allows us to evaluate the integral over z, leading to

$$P_{kk'} = \Gamma_{kk'}\int_0^1 du \int_0^1 dv \frac{\widetilde{\gamma}(u,v)}{\lambda(u)\lambda(v)}[\lambda(u)]^{k-1}$$
$$\times \int_0^{\lambda(v)} \frac{\xi^{k'-1}d\xi}{[1+\lambda(u)+\xi]^{k'+k-1}}, \qquad (42)$$

where $\Gamma_{kk'} = (k' + k - 2)!/(k-2)!(k'-1)!$. An example of the correlation function $P_{kk'}$ in the case of factorized $\widetilde{f}(u,v) = \lambda(u)\lambda(v)$ for the faux-BA model is shown in Fig. 4 together with the data from direct numerical simulation of such networks.

We now analyze this expression in the asymptotic regime $k \gg 1$, $k' \gg 1$, with finite $\eta = (k'-1)/k \approx k'/k$. The integral over ξ can be evaluated in the saddle-point approximation. The "action" $\mathcal{F}(\xi) = -(k'-1)\ln\xi + (k'+k-1)\ln(1+\lambda(u)+\xi)$ is extremal at $\xi_0 = [1+\lambda(u)]\eta$, with $\mathcal{F}(\xi_0) = k\ln[1+\lambda(u)] + k\{(1+\eta)\ln(1+\eta) - \eta\ln\eta\}$. In the same large-$k$ approximation one has

$$\Gamma_{kk'} = \exp\{k[(1+\eta)\ln(1+\eta) - \eta\ln\eta]\}\sqrt{\frac{k}{2\pi\eta(1+\eta)}},$$

and taking into account the cutoff of the ξ integral at $\lambda(v)$, we find

$$P_{kk'} = \frac{1}{2}\int_0^\infty \widetilde{\rho}(\lambda)d\lambda \int_0^\infty \widetilde{\rho}(\lambda')d\lambda' \frac{\widetilde{\gamma}(\lambda,\lambda')}{\lambda\lambda'} e^{-(k-1)\ln(1+1/\lambda)}$$
$$\times \left\{\text{erf}\left(\sqrt{\frac{k}{2\eta(1+\eta)}}\left[\frac{\lambda'}{\lambda} - \eta\right]\right) + \text{erf}\left(\sqrt{\frac{k\eta}{2(1+\eta)}}\right)\right\},$$
$$(43)$$

where it was also convenient to change to integrating over $\lambda = \lambda(u)$ and $\lambda' = \lambda(v)$ with the corresponding densities. The Gaussian approximation in the integral over ξ would of course break down when the saddle point ξ_0 moves far to the right
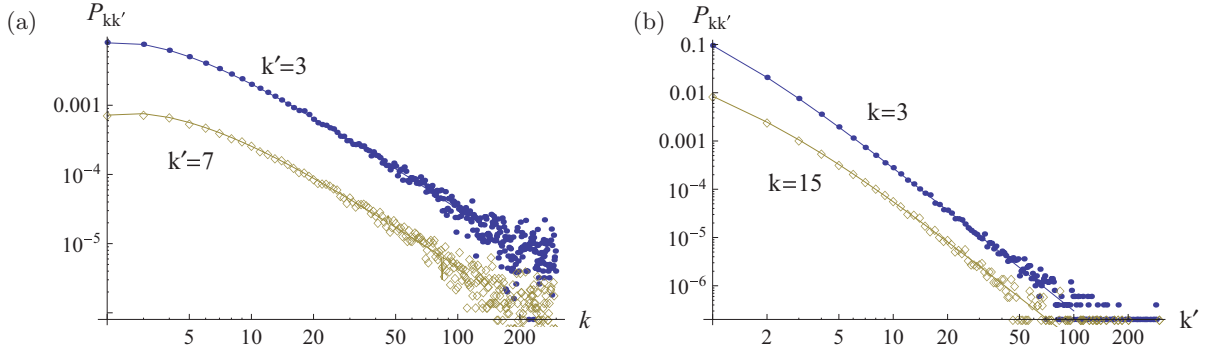
FIG. 4. (Color online) Node degree correlations, theory vs numerical simulation (50 runs, 100 000 time steps each), in the faux-BA factorized model: (a) $P_{kk'}$ as a function of k with fixed k' and (b) $P_{kk'}$ as a function of k' with fixed k.

of λ', however, this regime can be ignored as its contribution to the integrals over λ and λ' is exponentially suppressed. The factor $e^{-(k-1)\ln(1+1/\lambda)}$ in the large-k regime pushes the main region of integration over λ to $\lambda \sim k$ and the error functions restrict $\mu = \lambda'/\lambda$ to be greater than η (which is assumed to be finite). The combination of the error functions in the large curly brackets thus effectively behaves as a step function $\theta(\mu - \eta)$ smeared on the scale of $1/\sqrt{k}$.

We now consider in more detail the factorized case $\widetilde{f}(u,v) = \lambda(u)\lambda(v)$, which gives $\widetilde{\gamma}(u,v) = \lambda(u)$. In the main case of interest, when the degree distribution is asymptotically a power law, the structure of the integrals over λ and λ' picks out the asymptotic regimes of $\widetilde{\rho}(\lambda) \sim C_\alpha/\lambda^\alpha$ for both variables λ and λ'. Integrating first over μ and approximating $e^{-(k-1)\ln(1+1/\lambda)} \approx e^{-k/\lambda}$, we find

$$P_{kk'} = \frac{C_\alpha^2}{\alpha\eta^\alpha} \int_0^\infty d\lambda\, \lambda^{-2\alpha} e^{-k/\lambda} = \frac{C_\alpha^2 \Gamma(2\alpha-1)}{\alpha\eta^\alpha k^{2\alpha-1}}. \quad (44)$$

The joint degree distribution of neighboring nodes completely factorizes in this regime, $P_{kk'} \propto k p_k p_{k'}$. Curiously, however, the factorized correlation function shows an enhancement by a factor of $\Gamma(2\alpha-1)/\alpha\Gamma^2(\alpha)$ compared to the product $k p_k p_{k'}$, so links between high-degree nodes are more favored by a finite factor than what would be expected if degree distributions of neighboring nodes were completely independent. The enhancement factor tends to 1 in the limiting normalizable case $\alpha \to 2$ and is equal to 2 if $\alpha = 3$. This has to be contrasted with the behavior of the actual BA model in the same regime [Eq. (44) of Ref. [15]], where the asymptotic expression also factorizes into power-law functions of k and k', but the powers are different from 2 and 3, respectively.

In the complementary limiting case of large k and finite k', again assuming factorization of $\widetilde{f}$ and the asymptotic power-law behavior of $\widetilde{\rho}(\lambda)$, it is convenient to integrate first over λ by exploiting the suppression of small-λ contributions by $(1+1/\lambda)^{-(k-1)}$. This gives

$$P_{kk'} = \frac{C_\alpha \Gamma(k'+\alpha-1)}{k^{\alpha-1}(k'-1)!} \int_0^\infty \frac{\widetilde{\rho}(\lambda')d\lambda'}{\lambda'} \int_0^{\lambda'} \frac{x^{k'-1}dx}{(1+x)^{k'+\alpha-1}}$$

$$= \frac{C_\alpha \Gamma(k'+\alpha-1)}{k^{\alpha-1}(k'-1)!} \int_0^\infty d\lambda' \frac{\widetilde{\rho}(\lambda')(\lambda')^{k'-1}}{k'}$$

$$\times {}_2F_1(k'+\alpha-1,k';k'+1;-\lambda'), \quad (45)$$

where we have used $\Gamma_{kk'} \sim k^{k'}/(k'-1)!$ and ${}_2F_1$ is the hypergeometric function. Alternatively, introducing $I(\lambda) = \int_\lambda^\infty d\mu\,\widetilde{\rho}(\mu)/\mu$, the correlation function can be written as (see Appendix A)

$$P_{kk'} = \frac{C_\alpha \Gamma(k'+\alpha-1)}{k^{\alpha-1}(k'-1)!} \int_0^\infty \frac{d\lambda'(\lambda')^{k'-1}I(\lambda')}{(1+\lambda')^{k'+\alpha-1}}. \quad (46)$$

Although this expression is again factorized, it now shows suppression of the joint probability compared to the uncorrelated product $k p_k p_{k'}$ by a numerical factor. The actual suppression factor $h(k') = P_{kk'}/k p_k p_{k'}$ is sensitive to the details of $\widetilde{\rho}(\lambda)$ at small λ. For example, in the faux-BA case $\widetilde{\rho}(\lambda) = 2/(1+\lambda)^3$, explicit evaluation gives for the suppression factor $h(1) = 7/16$, $h(2) = 11/20$, $h(3) = 5/8$, and $h(4) = 19/28$.

It is seen from Eq. (43) that the degree correlation function is also sensitive to the details of the linking rules. If factorization of $\widetilde{f}$ is not assumed, the leading approximation to $P_{kk'}$ as $k \to \infty$, $k'/k = \eta$, is

$$P_{kk'} = \int_0^\infty d\lambda\, d\mu\, \widetilde{\rho}(\lambda)\widetilde{\rho}(\lambda\mu) \frac{\widetilde{\gamma}(u(\lambda),u(\lambda\mu))}{\lambda\mu} e^{-k/\lambda}\theta(\mu-\eta). \quad (47)$$

This expression can be rewritten to emphasize the scaling behavior in the case of power-law degree distributions. Defining $\widetilde{\gamma}(u(\lambda),u(\lambda')) = \lambda\delta(\lambda,\lambda')$ (with $\delta = 1$ in the factorized case) and further introducing a scaling function $\Delta_\alpha(\lambda,\lambda')$ via

$$\int_{\lambda'}^\infty \delta(\lambda,l) \frac{dl}{l^{\alpha+1}} = \frac{1}{\alpha}(\lambda')^{-\alpha}\Delta_\alpha(\lambda,\lambda'), \quad (48)$$

we now use the asymptotic form of the density $\widetilde{\rho}$ and switch the remaining integration variable to $x = k/\lambda$ to obtain

$$P_{kk'} = \frac{C_\alpha^2}{\eta^\alpha k^{2\alpha-1}} \int_0^\infty dx\, x^{2\alpha-2} e^{-x}\Delta_\alpha\left(\frac{k}{x},\frac{k\eta}{x}\right). \quad (49)$$

Since there is an infinite number of degrees of freedom in choosing the growth rules for a given $\widetilde{\rho}$, the function Δ_α can enforce a variety of correlation behaviors for a given degree distribution. The two surfaces in Fig. 5 show the ratio of the node degree correlation function to the uncorrelated baseline $(k-1)p_k p_{k'}$ for the two choices of the linking function corresponding to Figs. 2 and 3. These examples very starkly illustrate the fact that, despite producing identical node degree distributions (the distribution of the BA model
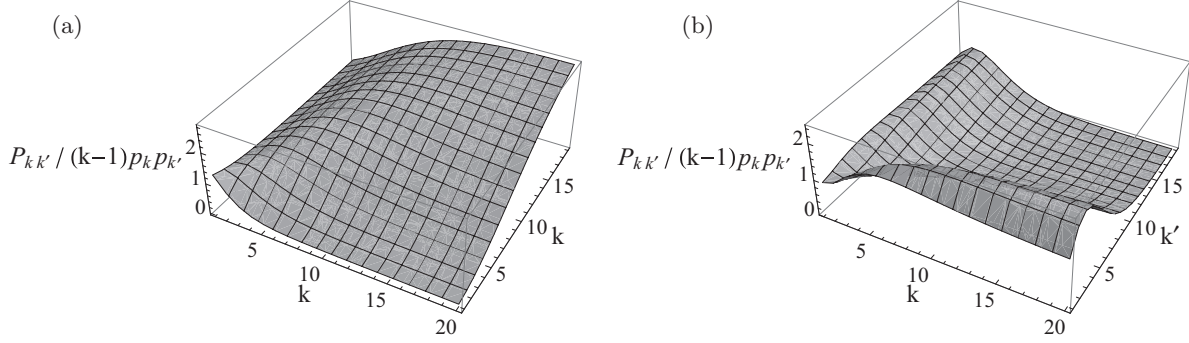
FIG. 5. Ratio of $P_{kk'}$ to the uncorrelated baseline $(k-1)p_k p_{k'}$ for the case when linking rules favor connections between nodes with (a) similar rankings and (b) different rankings.

in both cases), these two choices of the linking function generate very different types of correlation behavior. The rules that favor connections between nodes of similar degrees induce a considerably enhanced probability of connections between nodes of high degree compared to the uncorrelated null hypothesis and vice versa, the case where connections between nodes of similar rankings are disfavored results in a network in which connections between high-degree hubs are suppressed. Further study is needed to determine whether this difference qualitatively affects the behavior of the average path lengths.

The wide latitude in choosing the linking function producing the same degree distribution, as discussed above, provides the motivation to investigate a more general design problem: to match a given function $P_{kk'}$ to the network growth rules. We define the corresponding generating function as

$$G(s,\sigma) = \sum_{k=2}^{\infty} \sum_{k'=1}^{\infty} P_{kk'} s^{k-2} \sigma^{k'-1}. \tag{50}$$

Substituting the explicit form of $P_{kk'}$ from Eq. (42) we obtain, after evaluating the sums,

$$G(s,\sigma) = \int_0^{\infty} d\lambda \int_0^{\infty} d\lambda' \frac{\gamma_{\text{eff}}(\lambda,\lambda')}{(1+z\lambda)(1+z\lambda+w\lambda')}, \tag{51}$$

where $\gamma_{\text{eff}}(\lambda,\lambda') = \widetilde{\gamma}(u(\lambda),u(\lambda'))\widetilde{\rho}(\lambda)\widetilde{\rho}(\lambda')$, $z = 1 - s$, and $w = 1 - \sigma$.

Equation (51) can be interpreted as the integral equation on γ_{eff} if $G(s,\sigma)$ is a given function for a network whose structure the growth rules have to match. In the model considered so far this makes the problem overdetermined: If the linking function $\widetilde{f}(u,v)$ is required to be a symmetric function of its arguments, the corresponding function $\widetilde{\gamma}(u,v)$ defines a (continuous state space) reversible Markov chain. Indeed, the definition (4) implies $\int \widetilde{\gamma}(u,v)du = 1$, hence the stochasticity property is satisfied; thus denoting $\int \widetilde{f}(u,v)dv = I(u)$ we have

$$\widetilde{\gamma}(u,v)I(v) = \widetilde{f}(u,v) = \widetilde{f}(v,u) = \widetilde{\gamma}(v,u)I(u), \tag{52}$$

which is the definition of reversibility. In a generic case, even assuming that other conditions on the solvability of Eq. (51) are satisfied (see below), the symmetry in Eq. (52) has no enforcing mechanism. Nevertheless, this design problem can be given a meaning by relaxing the symmetry requirement imposed on $\widetilde{f}$: as noted in Sec. II, in the $q = 1$ case considered here a

model in which $\widetilde{f}$ is not symmetric can be treated within the same formalism. Note that in such an expanded model, the task of determining $\widetilde{f}$ once $\widetilde{\gamma}$ is known is reduced to determining the stationary distribution $I(v)$ in the corresponding Markov chain.

Performing an analytical continuation of $G(s,\sigma)$, originally given by the sum in Eq. (50) with a finite radius of convergence in s, to $|s| > 1$ and interpreting z as the variable of the Laplace transform, the corresponding inverse transform turns the equation into a double Laplace transform with respect to the variables $1/\lambda$ and λ'/λ. These Laplace transforms are straightforwardly inverted to give

$$\gamma_{\text{eff}}(\lambda,\lambda') = \frac{1}{\lambda} \int_{d-i\infty}^{d+i\infty} \frac{da}{2\pi i} e^{a/\lambda} \int_{c-i\infty}^{c+i\infty} \frac{dw}{2\pi i} e^{wa\lambda'/\lambda}$$
$$\times \frac{\partial}{\partial w} \left[w \int_{-\infty}^{\infty} \frac{d\zeta}{2\pi} \mathcal{G}(i\zeta,w)e^{i\zeta a} \right], \tag{53}$$

where the real constants d and c are chosen in the usual way so as to leave all the singularities to the left of the corresponding contours and $\mathcal{G}(z,w) = G(1-z,1-w)$. It is important that the analytical continuations are performed in the order implied by the sequence of integrations: a and w are real and positive in the integral over $\zeta = -iz$, after which an analytical continuation over w is performed with a still real and positive, and an analytical continuation in a is performed in the outermost integral.

Similarly to the discussion in Sec. III, the existence of the formal inverse in Eq. (53) requires that the analytical properties in z and w implied by the integral in the rhs of Eq. (51) match those of $\mathcal{G}(z,w)$. Further, the formal inverse does not yet guarantee that the corresponding growth model exists: The result of integration in Eq. (53) has to be a non-negative function for all values of its arguments. Note also that setting either σ or s to one and integrating over the corresponding λ variable shows that Eq. (51) contains Eq. (10). Hence the solvability of Eq. (10) is a necessary condition on the solvability of Eq. (51).

It was shown earlier that a factorizable linking function leads to a certain residual amount of correlations in the degree distributions of neighboring nodes. The qualitative explanation for the persistence of correlations is the presence of a mixture of different ages of nodes in the network. Armed with Eq. (53), we are now in a position to investigate whether correlations

between degrees of neighboring nodes can ever be completely designed away.

The absence of correlations is encoded in a fully factorized correlation function $P_{kk'} = (k-1)p_k p_{k'}$. The generating function is therefore $G(s,\sigma) = (d/ds)[G(s)/s]G(\sigma)/\sigma$ and assuming the existence of the corresponding growth process, we have

$$\mathcal{G}(z,w) = \int_0^\infty \frac{d\lambda\lambda\widetilde{\rho}(\lambda)}{(1+z\lambda)^2} \int_0^\infty \frac{d\lambda'\widetilde{\rho}(\lambda')}{1+w\lambda'}. \qquad (54)$$

Applying now Eq. (53), a lengthy but straightforward calculation gives

$$\gamma_{\text{eff}}(\lambda,\lambda') = \left(2\lambda^2\frac{\partial}{\partial\lambda} + \lambda^3\frac{\partial^2}{\partial\lambda^2}\right)\int_\lambda^\infty \frac{dx}{x}\widetilde{\rho}(x)\widetilde{\rho}\left(\frac{x\lambda'}{x-\lambda}\right). \qquad (55)$$

A remarkable feature of this expression is that it is not guaranteed to be non-negative. In other words, degree correlations can indeed be designed away (allowing for nonsymmetric linking), but only for a subset of densities $\widetilde{\rho}$. To classify the behavior of Eq. (55), it is convenient to rewrite it in terms of new variables $l = 1/\lambda$ and $\mu = \lambda'/\lambda$. It now takes the form

$$\gamma_{\text{eff}}(1/l,\mu/l) = l\frac{\partial^2}{\partial l^2}\int_0^1 \frac{dy}{y}R\left(\frac{l}{\mu}(1-y)\right)R(ly), \qquad (56)$$

where $R(x) = \widetilde{\rho}(1/x)$.

The behavior of $R(x)$ is governed by the convergence requirements on the integrals involving $\widetilde{\rho}(\lambda)$, as discussed in Sec. III: It must decay at least as x^3 as $x \to 0$ and is either constant (possibly 0) as $x \to \infty$ or diverges slower than x. As a consequence, in the first case, $R(x)$ necessarily becomes concave at sufficiently large values of l and the second derivative in Eq. (56) turns negative, invalidating the solution as a putative linking probability density.

If $R(x)$ is divergent at infinity, one has to distinguish between divergence slower than $x^{1/2}$, in which case the integral grows slower than l and the second derivative is negative at least in some region of small μ and sufficiently large l. In the remaining case of singular behavior as $x^{1/2}$ or faster, it is possible to ensure that the the second derivative in Eq. (56) never turns negative. Such a divergence is a necessary condition for the existence of a growth model with factorizable degree correlations. It is not a sufficient condition since this divergence can still coexist with nonmonotonic behavior, which would turn the second derivative negative. It should be remarked that the faux-BA density $\widetilde{\rho}(\lambda) = 2/(1+\lambda)^3$ is nonsingular at $\lambda \to 0$ and therefore does not allow for correlationless networks.

The qualitative picture that therefore emerges is that degree correlations can only be designed away if there is an accumulation of nodes with low linking propensity, so new nodes mostly connect to a few well-separated hubs, while the majority of nodes never acquire many connections. This is consistent with the previously remarked fact of correlations being an (almost) unavoidable feature of growing networks: Beyond a certain threshold, descendant nodes acquiring descendants of their own necessarily generate correlations between degrees of connected nodes. The particular threshold requiring a square root or faster divergence of the density of linking propensities

is a surprising technical result, which does not seem to have an immediate qualitative explanation.

As the second application of this formalism, we find the linking probability $\gamma_{\text{eff}}(\lambda,\lambda')$, which reproduces the correlation function of the BA model. The details of this somewhat technical exercise are given in Appendix B. The result, Eq. (B7), is a non-negative function, albeit not corresponding to a reversible Markov chain. Hence a static fitness growth model indeed exists such that it reproduces precisely both the degree distribution and the degree correlations of the standard preferential attachment model, although it requires a nonsymmetric linking function $\widetilde{f}$.

V. DISCUSSION

The key finding of this study is the fact that design problems of matching the growth rules to the required statistics of node degrees can be solved in a broad, but still circumscribed, class of models. The implications of this fact are twofold. On the one hand, the restrictions on the solvability of the design problems can be used as diagnostic tools: Given an empirical network, it may be possible to exclude the hidden variable growth process if the corresponding design problem cannot be solved based on the known degree statistics.

On the other hand, the capacity of the fitness models to mimic the node degree statistics of the preferential attachment model means that one should be careful in inferring preferential attachment behavior from empirical networks. Consider a network whose growth process (for simplicity without rewiring) can be observed in time. Without knowing whether a hidden variable mechanism operates, one may assume an attachment kernel $\Pi(k)$, i.e., the probability that an incoming node joins a particular node i with degree k_i is $\Pi(k_i)/t$, where $\Pi(k)$ is assumed to be normalized so that $\sum_j \Pi(k_j) = t$. [The standard linear preferential attachment model has $\Pi(k) = k/\bar{k}$.] Hence the probability that the incoming node chooses a node of degree k to connect to is $\Pi(k)p_k$, where p_k is the degree distribution, assumed to be stationary. One may use this correspondence to extract the attachment kernel by observing which existing nodes the incoming nodes choose to join [13].

Assume now that actually a fitness-based growth process is in operation, with each node possessing a hidden fitness, or ranking, label. The probability that the incoming node chooses a node of ranking u is $\lambda(u)$ and using Eqs. (7) and (10) we find the probability that the chosen node has degree k to be

$$\int du\lambda(u)p_k(t|u) = \int_0^\infty d\lambda\widetilde{\rho}(\lambda)\left(\frac{\lambda}{1+\lambda}\right)^k. \qquad (57)$$

If $\widetilde{\rho}(\lambda) \sim \lambda^{-\alpha}$, the integral can be estimated as proportional to $k^{-\alpha+1}$. Therefore, recalling that a power-law tail in $\widetilde{\rho}$ reproduces itself as the corresponding power law in p_k, we see that the integral behaves as kp_k, leading one to conclude that linear preferential attachment is in operation. Remarkably, this behavior is obtained for any α, not just $\alpha = 3$ of the orthodox preferential attachment model.

In conclusion, we mention a few outstanding questions that may warrant further study. In connection with the iteration algorithm outlined at the end of Sec. III, it would be useful to investigate under what conditions on the generating kernel K the corresponding iteration operator is contracting so that the

convergence of iterations would follow from the Banach fixed point theorem.

Another issue is the presence of complementary positivity requirements: Both $\widetilde{\rho}$ and the expansion coefficients p_k of $G(s)$ in Eq. (18) must be positive due to their probabilistic origins [another such dual pair is $\widetilde{\gamma}$ and $G(s,\sigma)$]. Formally, the requirement is imposed on the analytical structure of $G(s)$: Both the coefficients of the power-law expansion of $G(s)$ around 0 and the imaginary part of its jump across the cut extending from 1 to ∞ have to be positive. It may be interesting to explore whether the second requirement can be expressed as a condition on the coefficients p_k that does not amount to solving for $\widetilde{\rho}$.

As the final remark, the restrictions on solvability of design problems can be ultimately traced to the correlations between degrees of connected nodes stemming from the nature of the growth process. In a sense, the dynamics of node degrees may be viewed as a (discrete) flow of node degree density. In this language, the restrictions on allowed degree distributions that allow the existence of solutions to Eq. (18) [and similarly the correlation functions that admit solutions to Eq. (51)] can probably be understood as conservation laws.

APPENDIX A

Equation (46) is obtained from the first line of Eq. (45) as follows. As discussed in Sec. III, $\rho(\lambda)$ may be singular at 0, therefore, in order to integrate by parts over λ' the first line of Eq. (45), we define the function $I(\lambda) = \int_\lambda^\infty \widetilde{\rho}(\mu) d\mu/\mu$. Performing now the integration by parts, we obtain Eq. (46) plus the boundary terms

$$
\lim_{\lambda' \to 0} I(\lambda') \int_0^{\lambda'} \frac{x^{k'-1} dx}{(1+x)^{k'+\alpha-1}} - \lim_{\lambda' \to \infty} I(\lambda') \int_0^{\lambda'} \frac{x^{k'-1} dx}{(1+x)^{k'+\alpha-1}}.
\tag{A1}
$$

We now show that both of these contributions are equal to zero. Since the denominator in the integral over x is regular at 0, the first term is proportional to $\lim_{\lambda' \to 0}(\lambda')^{k'} I(\lambda')$. This expression is trivially equal to zero if a finite $\lim_{\lambda' \to 0} I(\lambda')$ exists. If the limit is infinite,

$$
\lim_{\lambda' \to 0}(\lambda')^{k'} I(\lambda') = \lim_{\lambda' \to 0} \frac{I'(\lambda')}{-k'/(\lambda')^{k'+1}} = \lim_{\lambda' \to 0}(\lambda')^{k'} \widetilde{\rho}(\lambda')/k' = 0,
\tag{A2}
$$

where the last equality follows from the fact that $\widetilde{\rho}(\lambda)$ has at most an integrable singularity at 0. As for the second term, the integral over x is finite for any k', while the definition of $I(\lambda')$ implies $\lim_{\lambda' \to \infty} I(\lambda') = 0$, hence their product is equal to zero.

APPENDIX B

The purpose of this appendix is to calculate the linking probability $\gamma_{\text{eff}}(\lambda,\lambda')$, which would generate a network whose degree correlations are identical to those of the classical preferential attachment (BA) model. The degree correlation

function of the BA model is calculated in Ref. [15] as

$$
P_{kk'} = \frac{4(k-1)}{k'(k'+1)(k'+k)(k'+k+1)(k'+k+2)}
$$
$$
+ \frac{12(k-1)}{k'(k'+k-1)(k'+k)(k'+k+1)(k'+k+2)}.
\tag{B1}
$$

Applying the definition in Eq. (50), the corresponding generating function is calculated to be

$$
G(s,\sigma) = \frac{3}{\sigma}\frac{\partial}{\partial s}\int_0^\sigma d\sigma' \frac{J(\sigma') - J(s)}{\sigma' - s} + \frac{\partial}{\partial s}\frac{s}{\sigma^2}\int_0^s
$$
$$
\times d\sigma'(\sigma - \sigma')\frac{\widetilde{G}(\sigma')/\sigma'^2 - \widetilde{G}(s)/s^2}{\sigma' - s},
\tag{B2}
$$

where $\widetilde{G}(s) = G(s) - \frac{2}{3}s$, $G(s)$ is the generating function of the degree distribution of the BA model given in Eq. (22), and

$$
J(s) = \int_0^s ds' \frac{\widetilde{G}(s')}{s'^2}.
\tag{B3}
$$

Given the structure of $G(s,\sigma)$, the inversion of the integral equation (51) is easier to perform by switching the order of integrations over z and w:

$$
\gamma_{\text{eff}}(\mu'/l', 1/l') = -l' \int_{d-i\infty}^{d+i\infty} \frac{da}{2\pi i} e^{al'} \int_{c-i\infty}^{c+i\infty} \frac{dz}{2\pi i} e^{za\mu'}
$$
$$
\times \int_{b-i\infty}^{b+i\infty} \frac{dw}{2\pi i} \mathcal{G}(z,w) a w e^{aw},
\tag{B4}
$$

where $l' = 1/\lambda'$, $\mu' = \lambda/\lambda'$, and all three contours lie to the right of the corresponding singularities of the integrands. The integral over w can now be evaluated by deforming the contour so that it is wrapped around the cut in the w plane extending from $w = 0$ to $-\infty$, which can be deduced from Eq. (B2). Evaluating the jump of $\mathcal{G}(z,w)$ across the cut using Eq. (B2), we have

$$
\frac{1}{\pi}\text{Im}\,\mathcal{G}(z, -\zeta + i0) = \frac{2\zeta}{(1+\zeta)^2}\int_0^\zeta dy \frac{dy}{(1+y)^2(z+y)^2},
\tag{B5}
$$

where z is assumed to be real and positive and $\zeta = -w$ is the coordinate along the cut. The analytical continuation over z is now trivial and after integration we obtain

$$
\gamma_{\text{eff}}(\mu'/l', 1/l')
$$
$$
= l' \int_{d-i\infty}^{d+i\infty} \frac{da}{2\pi i} e^{al'} a^2 \mu' \int_0^\infty d\zeta \frac{2\zeta^2}{(1+\zeta)^2} e^{-\zeta a}
$$
$$
\times \int_0^\zeta dy \frac{y^2}{(1+y)^2} e^{-ya\mu'}.
\tag{B6}
$$

The integration over a gives the second derivative of the δ function, which is easily integrated over to give

$$
\gamma_{\text{eff}}(\mu'/l', 1/l') = l' \frac{\partial^2}{\partial l'^2}\int_{l'/(\mu'+1)}^{l'} d\zeta \frac{2\zeta^2}{(1+\zeta)^2}\frac{(l'-\zeta)^2}{(l'+\mu'-\zeta)^2}.
\tag{B7}
$$

The remaining integral can be calculated in closed form

$$\gamma_{\text{eff}}(\lambda,\lambda') = \frac{4\lambda[\mathcal{P}_1(\lambda,\lambda') + \ln(1 + 1/\lambda')\mathcal{P}_2(\lambda,\lambda')]}{(\lambda' + 1)(\lambda' + \lambda + 1)^5}, \quad \text{(B8)}$$

where the polynomials $\mathcal{P}_1$ and $\mathcal{P}_2$ are

$$\mathcal{P}_1(\lambda,\lambda') = 2\lambda'^3 - 8\lambda\lambda'^2 - \lambda'^2 + 2(\lambda^2 - 3\lambda - 1)\lambda' + (\lambda + 1)^2 \quad \text{(B9)}$$

and

$$\mathcal{P}_2(\lambda,\lambda') = -2\lambda'^4 + 8\lambda'^3\lambda - 2\lambda'(\lambda^2 - 5\lambda - 3) \\ + 2\lambda'(2 + \lambda - \lambda^2). \quad \text{(B10)}$$

One can check that this linking probability is indeed nonnegative over the full range of λ and λ' and that the Kolmogorov condition for reversibility is not satisfied, hence the corresponding linking function $\widetilde{f}$ is not symmetric.

[1] M. E. Newman, SIAM Rev. **45**, 167 (2003).

[2] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).

[3] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Phys. Rep. **424**, 175 (2006).

[4] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).

[5] P. Erdős and A. Rényi, Mag. Tud. Akad. Mat. Kutató Int. Közl. **5**, 17 (1960).

[6] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).

[7] H. Jeong, Z. Néda, and A. Barabási, Europhys. Lett. **61**, 567 (2003).

[8] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Muñoz, Phys. Rev. Lett. **89**, 258702 (2002).

[9] V. D. P. Servedio, G. Caldarelli, and P. Butta, Phys. Rev. E **70**, 056126 (2004).

[10] G. Ergun and G. Rodgers, Physica A **303**, 261 (2002).

[11] C. Bedogne and G. J. Rodgers, Phys. Rev. E **74**, 046115 (2006).

[12] X.-J. Xu, L.-M. Zhang, and L.-J. Zhang, Int. J. Mod. Phys. C **21**, 129 (2010).

[13] S. Abe and S. Thurner, Phys. Rev. E **72**, 036102 (2005).

[14] A. D. Polyanin and A. V. Manzhirov, *Handbook of Integral Equations* (Chapman & Hall/CRC, Boca Raton, FL, 2008) .

[15] P. L. Krapivsky and S. Redner, Phys. Rev. E **63**, 066123 (2001).

[16] M. E. J. Newman, S. Forrest, and J. Balthrop, Phys. Rev. E **66**, 035101 (2002).

[17] C. Moore, G. Ghoshal, and M. E. J. Newman, Phys. Rev. E **74**, 036121 (2006).