Exact solutions for models of evolving networks with addition and deletion of nodes

Cristopher Moore,[1,2,3] Gourab Ghoshal,[4,5] and M. E. J. Newman[3,4]
[1]*Department of Computer Science and Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico 87131, USA*
[2]*Santa Fe Institute, Santa Fe, New Mexico 87501, USA*
[3]*Center for the Study of Complex Systems, University of Michigan, Ann Arbor, Michigan 48109, USA*
[4]*Department of Physics, University of Michigan, Ann Arbor, Michigan 48109, USA*
[5]*Michigan Center for Theoretical Physics, University of Michigan, Ann Arbor, Michigan, 48109, USA*
(Received 11 April 2006; published 28 September 2006)

There has been considerable recent interest in the properties of networks, such as citation networks and the worldwide web, that grow by the addition of vertices, and a number of simple solvable models of network growth have been studied. In the real world, however, many networks, including the web, not only add vertices but also lose them. Here we formulate models of the time evolution of such networks and give exact solutions for a number of cases of particular interest. For the case of net growth and so-called preferential attachment—in which newly appearing vertices attach to previously existing ones in proportion to vertex degree—we show that the resulting networks have power-law degree distributions, but with an exponent that diverges as the growth rate vanishes. We conjecture that the low exponent values observed in real-world networks are thus the result of vigorous growth in which the rate of addition of vertices far exceeds the rate of removal. Were growth to slow in the future—for instance, in a more mature future version of the web—we would expect to see exponents increase, potentially without bound.

DOI: 10.1103/PhysRevE.74.036121 PACS number(s): 89.75.Hc, 89.20.Hh, 05.10.Gg, 05.65.+b

I. INTRODUCTION

The study of networks has attracted a substantial amount of attention from the physics community in the last few years [1–3], in part because of networks' broad utility as representations of real-world complex systems and in part because of the demonstrable successes of physics techniques in shedding light on network phenomena. One topic that has been the subject of a particularly large volume of work is growing networks, such as citation networks [4,5] and the worldwide web [6,7]. Perhaps the best-known body of work on this topic is that dealing with "preferential attachment" models [8,9], in which vertices are added to a network with edges that attach to preexisting vertices with probabilities depending on those vertices' degrees. When the attachment probability is precisely linear in the degree of the target vertex the resulting degree sequence for the network follows a Yule distribution in the limit of large network size, meaning it has a power-law tail [8–12]. This case is of special interest because both citation networks and the worldwide web are observed to have degree distributions that approximately follow power laws.

The preferential attachment model may be quite a good model for citation networks, which is one of the cases for which it was originally proposed [8,10]. For other networks, however, and especially for the worldwide web, it is, as many authors have pointed out, necessarily incomplete [13–17]. On the web there are clearly other processes taking place in addition to the deposition of vertices and edges. In particular, it is a matter of common experience that vertices (i.e., web pages) are often removed from the web and with them the links that they had to other pages. Models of this process have been touched upon occasionally in the literature [18–20], and the evidence suggests that in some cases vertex deletion affects the crucial power-law behavior of the degree distribution, while in other cases it does not.

In this paper, we study the general process in which a network grows (or, potentially, shrinks) by the constant addition and removal of vertices and edges. We show that a class of such processes can be solved exactly for the degree distributions they generate by solving differential equations governing the probability generating functions for those distributions. In particular, we give solutions for three example problems of this type, having uniform or preferential attachment and having stationary size or net growth. The case of uniform attachment and stationary size is of interest as a possible model for the structure of peer-to-peer filesharing networks, while the preferential-attachment stationary-size case displays a nontrivial stretched exponential form in the tail of the degree distribution. Our solution of the preferential attachment case with net growth confirms earlier results indicating that this process generates a power-law distribution, although the exponent of the power law diverges as the growth rate tends to zero, giving degree distributions that are numerically indistinguishable from exponential for small growth rates. This suggests that the clear power law seen in the real worldwide web is a signature of a network whose rate of vertex accrual far outstrips the rate at which vertices are removed. The relative rates of addition and removal could, however, change as the web matures, possibly leading to a loss of power-law behavior at some point in the future.

II. THE MODEL

Consider a network that evolves by the addition and removal of vertices. In each unit of time, we add a single vertex to the network and remove r vertices. When a vertex is removed, so too are all the edges incident on that vertex, which means that the degrees of the vertices at the other ends of those edges will decrease. Noninteger values of r are permitted and are interpreted in the usual stochastic fashion. (For example, values $r<1$ can be interpreted as the probability per unit time that a vertex is removed.) The value $r=1$ corresponds to a network of fixed size in which there is vertex turnover but no growth; values $r<1$ correspond to growing networks. In principle one could also look at values $r>1$, which correspond to shrinking networks, and the methods derived here are applicable to that case. However, we are not aware of any real-world examples of shrinking networks in which the asymptotic degree distribution is of interest, so we will not pursue the shrinking case here.

We make two further assumptions, which have also been made by most previous authors in studying these types of systems: (1) that all vertices added have the same initial degree, which we denote c; (2) that the vertices removed are selected uniformly at random from the set of all extant vertices. Note, however, that we will not assume that the network is uncorrelated (i.e., that it is a random multigraph conditioned on its degree distribution as in the so-called configuration model). In general the networks we consider will have correlations among the degrees of their vertices but our solutions will nonetheless be exact.

Let p_k be the fraction of vertices in the network at a given time that have degree k. By definition, p_k has the normalization

$$\sum_{k=0}^{\infty} p_k = 1. \tag{1}$$

Our primary goal in this paper will be to evaluate exactly the degree distribution p_k for various cases of interest.

Although the form of p_k is, as we will see, highly nontrivial in most cases, the mean degree of a vertex, $\langle k\rangle = \sum_{k=0}^{\infty} k p_k$, is easily derived in terms of the parameters r and c. The mean number of vertices added to the network per unit time is $1-r$. The mean number of edges removed when a randomly chosen vertex is removed from the network is by definition $\langle k\rangle$. Thus the mean number of edges added to the network per unit time is $c-r\langle k\rangle$. For a graph of m edges and n vertices, the mean degree is $\langle k\rangle=2m/n$. After time t we have $n=(1-r)t$ and, assuming that $\langle k\rangle$ has an asymptotically constant value, $m=(c-r\langle k\rangle)t$. Thus

$$\langle k\rangle = 2\frac{c-r\langle k\rangle}{1-r} \tag{2}$$

or, rearranging,

$$\langle k\rangle = \frac{2c}{1+r}. \tag{3}$$

In the special case $r=1$ of a constant-size network, this gives $\langle k\rangle=c$, which is clearly the correct answer.

We must also consider how an added vertex chooses the c other vertices to which it attaches. Let us define the attachment kernel π_k to be n times the probability that a given edge of a newly added vertex attaches to a given preexisting vertex of degree k. The factor of n here is convenient, since it means that the total probability that the given edge attaches to any vertex of degree k is simply $\pi_k p_k$. Since each edge must attach to a vertex of *some* degree, this also immediately implies that the correct normalization for π_k is

$$\sum_{k=0}^{\infty} \pi_k p_k = 1. \tag{4}$$

For the particular case of $\pi_k \propto k$ and $r<1$, which we consider in Sec. III C, models similar to ours have been studied previously by Sarshar and Roychowdhury [18], Chung and Lu [19], and Cooper, Frieze, and Vera [20]. While these authors did not seek an exact solution, our results on the power-law tail of the degree distribution in this case coincide with theirs.

A. Rate equation

Given these definitions, the evolution of the degree distribution is governed by a rate equation as follows. If there are a total of n vertices in the network at a given time, then the number of vertices with degree k is np_k. One unit of time later this number is $(n+1-r)p'_k$, where p'_k is the new value of p_k. Then

$$(n+1-r)p'_k = np_k + \delta_{kc} + c\pi_{k-1}p_{k-1} - c\pi_k p_k + r(k+1)p_{k+1} - rkp_k - rp_k. \tag{5}$$

The term δ_{kc} in Eq. (5) represents the addition of a vertex of degree c to the network. The terms $c\pi_{k-1}p_{k-1}$ and $-c\pi_k p_k$ describe the flow of vertices from degree $k-1$ to k and from k to $k+1$ as they gain extra edges when newly added vertices attach to them. The terms $(k+1)p_{k+1}$ and $-kp_k$ describe the flow from degree $k+1$ to k and from k to $k-1$ as vertices lose edges when one of their neighbors is removed from the network. And the term $-rp_k$ represents the removal with probability r of a vertex with degree k. Contributions from processes in which a vertex gains or loses two or more edges in a single unit of time vanish in the limit of large n and have been neglected.

We will be interested in the asymptotic form of p_k in the limit of large times for a given π_k. Setting $p'_k=p_k$ in Eq. (5) gives

$$\delta_{kc} + c\pi_{k-1}p_{k-1} - c\pi_k p_k + r(k+1)p_{k+1} - rkp_k - p_k = 0. \tag{6}$$

We can write the solution to Eq. (6) in terms of generating functions as follows. Let us define

$$f(z) = \sum_{k=0}^{\infty} \pi_k p_k z^k, \tag{7}$$

$$g(z) = \sum_{k=0}^{\infty} p_k z^k. \tag{8}$$

Then, upon multiplying both sides of Eq. (6) by z^k and summing over k (with the convention that $p_{-1}=0$), we derive a differential equation for $g(z)$ thus:

$$r(1-z)\frac{dg}{dz} - g(z) - c(1-z)f(z) + z^c = 0. \tag{9}$$

Note also that we can easily generalize our model to the case where the degrees of the vertices added are not all identical but are instead drawn at random from some distribution r_k. In that case, we simply replace δ_{kc} in Eq. (6) with r_k and z^c in Eq. (9) with the generating function $h(z)=\Sigma_k r_k z^k$.

In the following sections we solve Eq. (9) for a number of different choices of the attachment kernel π_k. Note that, since the definitions of both $f(z)$ and $g(z)$ incorporate the unknown distribution p_k, we must in general solve implicitly for $g(z)$ in terms of $f(z)$. In all of the cases of interest to us here, however, it turns out to be straightforward to derive an explicit equation for $g(z)$ as a special case of Eq. (9).

III. SOLUTIONS FOR SPECIFIC CASES

In this section we study three specific examples of the class of models defined in the preceding section: namely, linear preferential attachment models ($\pi_k \propto k$) for both growing and fixed-size networks and uniform attachment (π_k=constant) for fixed size. As we will see, each of these cases turns out to have interesting features.

A. Uniform attachment and constant size

For the first of our example models we study the case where the size of the network is constant ($r=1$) and in which each vertex added chooses the c others to which it attaches uniformly at random. This means that π_k is constant, independent of k, and combining Eqs. (1) and (4), we immediately see that the correct normalization for the attachment kernel is $\pi_k=1$ for all k. Then we have $\pi_k p_k = p_k$ so that $f(z)=g(z)$ in Eq. (9), which gives

$$\left(c + \frac{1}{1-z}\right)g(z) - \frac{dg}{dz} = \frac{z^c}{1-z}. \tag{10}$$

Noting that $(1-z)e^{-cz}$ is an integrating factor and that $g(z)$ must obey the boundary condition $g(1)=1$, we readily determine that

$$\begin{aligned} g(z) &= \frac{e^{cz}}{1-z}\int_z^1 t^c e^{-ct}dt \\ &= \frac{e^{cz}}{1-z}c^{-(c+1)}[\Gamma(c+1,cz) - \Gamma(c+1,c)], \end{aligned} \tag{11}$$

where

$$\Gamma(c+1,x) = \int_x^\infty t^c e^{-t}dt \tag{12}$$

is the incomplete Γ function.

One can easily check that this gives a mean degree $g'(1)=c$, as it must, and that the variance of the degree $g''(1)+g'(1)-c^2$ is equal to $\frac{2}{3}c$, indicating a tightly peaked degree distribution when c is large.

To obtain an explicit expression for the degree distribution, we make use of

$$\Gamma(c+1,x) = \Gamma(c+1)e^{-x}\sum_{m=0}^{c}\frac{x^m}{m!}, \tag{13}$$

$$e^x = \sum_{m=0}^{\infty}\frac{x^m}{m!}, \tag{14}$$

$$(1-z)^{-1} = \sum_{k=0}^{\infty} z^k, \tag{15}$$

to write

$$\begin{aligned} g(z) = c^{-(c+1)}\sum_{k=0}^{\infty} z^k \Bigg[&\Gamma(c+1)\sum_{m=0}^{c}\frac{(cz)^m}{m!} \\ &- \Gamma(c+1,c)\sum_{m=0}^{\infty}\frac{(cz)^m}{m!}\Bigg]. \end{aligned} \tag{16}$$

The z dependence in the first term of this expression can be rewritten

$$\begin{aligned} \sum_{k=0}^{\infty} z^k \sum_{m=0}^{c}\frac{(cz)^m}{m!} &= \sum_{m=0}^{c}\sum_{k=m}^{\infty} z^k \frac{c^m}{m!} \\ &= \sum_{k=0}^{\infty} z^k \sum_{m=0}^{\min(k,c)}\frac{c^m}{m!} = e^c \sum_{k=0}^{\infty} z^k \frac{\Gamma(\min(k,c)+1,c)}{\Gamma(\min(k,c)+1)}, \end{aligned} \tag{17}$$

where $\min(k,c)$ denotes the smaller of k and c and we have again employed Eq. (13). A similar sequence of manipulations leads to an expression for the second term also, thus:

$$\sum_{k=0}^{\infty} z^k \sum_{m=0}^{\infty}\frac{(cz)^m}{m!} = e^c \sum_{k=0}^{\infty} z^k \frac{\Gamma(k+1,c)}{\Gamma(k+1)}. \tag{18}$$

Combining these identities with Eq. (16), it is then a simple matter to read off the term in $g(z)$ involving z^k, which is by definition our p_k. We find two separate expressions for the cases of k above and below c:

$$p_k = \frac{e^c}{c^{c+1}}[\Gamma(c+1) - \Gamma(c+1,c)]\frac{\Gamma(k+1,c)}{\Gamma(k+1)}, \quad \text{for } k < c, \tag{19}$$

and

$$p_k = \frac{e^c}{c^{c+1}}\Gamma(c+1,c)\left[1 - \frac{\Gamma(k+1,c)}{\Gamma(k+1)}\right], \quad \text{for } k \geqslant c. \tag{20}$$

Note that the quantity $\Gamma(k+1,c)/\Gamma(k+1)$ appearing in both these expressions is the probability that a Poisson-distributed variable with mean c is less than or equal to k. Thus the degree distribution has a tail that decays as the cumulative distribution of such a Poisson variable, implying that it falls off rapidly. To see this more explicitly, we note that for fixed c and $k \gg c$

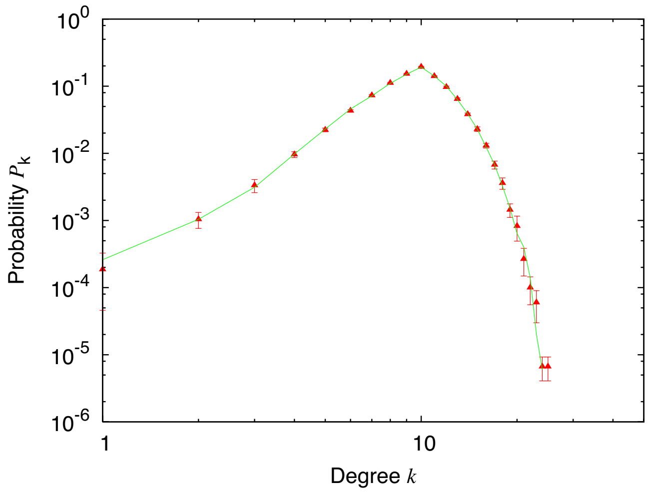

FIG. 1. (Color online) The degree distribution of our model for the case of uniform attachment (π_k=constant) with fixed size $n=50000$ and $c=10$. The points represent data from numerical simulations and the solid line is the analytic solution.

$$p_k = \frac{\Gamma(c+1,c)}{c^{c+1}} \sum_{m=k+1}^{\infty} \frac{c^m}{m!} \simeq \frac{\Gamma(c+1,c)}{\Gamma(k+2)} c^{k-c}, \tag{21}$$

since the sum is strongly dominated in this limit by its first term. Applying Stirling's approximation, $\Gamma(x) \simeq (x/e)^x \sqrt{2\pi/x}$, this gives

$$p_k \simeq \frac{\Gamma(c+1,c)}{c^c} k^{-3/2} e^k \left(\frac{c}{k}\right)^k, \tag{22}$$

which decays substantially faster asymptotically than any exponential.

As a check on these calculations, we have performed extensive computer simulations of the model. In Fig. 1 we show results for the case $c=10$, along with the exact solution from Eqs. (19) and (20). As the figure shows, the agreement between the two is excellent.

Before moving on to other issues, we note a different and particularly simple case of a growing network with uniform attachment, the case in which the vertices added have a Poisson degree distribution $c^k e^{-c}/k!$ with mean c. In that case the factor of z^c in Eq. (10) is replaced with the generating function $h(z)$ for the Poisson distribution:

$$h(z) = \sum_{k=0}^{\infty} \frac{c^k e^{-c}}{k!} z^k = e^{c(z-1)}, \tag{23}$$

and the solution, Eq. (11), becomes

$$g(z) = \frac{e^{cz}}{1-z} \int_z^1 h(t) e^{-ct} dt = e^{c(z-1)}, \tag{24}$$

which is itself the generating function for a Poisson distribution. Thus we see particularly clearly in this case that the equilibrium degree distribution in the steady-state uniform attachment network is sharply peaked with a Poisson tail. In fact, the network in this case is simply an uncorrelated random graph of the type famously studied by Erdős and Rényi [21]. It is straightforward to see that if one starts with such a graph and randomly adds and removes vertices with Poisson-distributed degrees, the graph remains an uncorrelated random graph with the same degree distribution, and hence this distribution is necessarily the fixed point of the evolution process, as the solution above demonstrates.

B. Preferential attachment and constant size

Our next example adds an extra degree of complexity to the picture: we consider vertices that attach to others in proportion to their degree, the so-called "preferential attachment" mechanism [9]. This implies that our attachment kernel π_k is linear in the degree: $\pi_k = Ak$ for some constant A. The normalization requirement (4) then implies that

$$\sum_{k=0}^{\infty} \pi_k p_k = A \sum_{k=0}^{\infty} k p_k = A\langle k \rangle = 1, \tag{25}$$

and hence $A = 1/\langle k \rangle$. For the moment, let us continue to focus on the case $r=1$ of constant network size, in which case $\langle k \rangle = c$ [Eq. (3)] and

$$\pi_k = \frac{k}{c}. \tag{26}$$

Then

$$f(z) = \frac{1}{c} \sum_{k=0}^{\infty} k p_k z^k = \frac{z}{c} g'(z), \tag{27}$$

and Eq. (9) becomes

$$\frac{g(z)}{(1-z)^2} - \frac{dg}{dz} = \frac{z^c}{(1-z)^2}. \tag{28}$$

The appropriate integrating factor in this case is $e^{-1/(1-z)}$, which, in conjunction with the boundary condition $g(1)=1$, gives

$$g(z) = e^{1/(1-z)} \int_z^1 \frac{t^c}{(1-t)^2} e^{-1/(1-t)} dt. \tag{29}$$

Changing the variable of integration to $y = 1/(1-t)$ this expression can be written

$$\begin{aligned} g(z) &= e^{1/(1-z)} \int_{1/(1-z)}^{\infty} \left(1 - \frac{1}{y}\right)^c e^{-y} dy \\ &= e^{1/(1-z)} \sum_{s=0}^{c} (-1)^s \binom{c}{s} \int_{1/(1-z)}^{\infty} \frac{e^{-y}}{y^s} dy \\ &= 1 + e^{1/(1-z)} \sum_{s=1}^{c} (-1)^s \binom{c}{s} \Gamma\left(1-s, \frac{1}{1-z}\right). \end{aligned} \tag{30}$$

where $\Gamma(1-s,x) = \int_x^{\infty} e^{-y} y^{-s} dy$ is again the incomplete Γ function, here appearing with a negative first argument.

A useful identify for the case $s \geqslant 1$ can be derived by integrating by parts thus:

$$\Gamma(-s,x) = \frac{1}{s}\left[\frac{e^{-x}}{x^s} - \Gamma(1-s,x)\right]. \tag{31}$$

Iterating this expression then gives

$$\Gamma(1-s,x) = -\frac{(-1)^s}{(s-1)!}\left[\Gamma(0,x) + e^{-x}\sum_{m=1}^{s-1}\frac{(-1)^m(m-1)!}{x^m}\right]. \tag{32}$$

$[\Gamma(0,x)=\int_x^\infty(e^{-y}/y)dy$ is also known as the exponential integral function $-\mathrm{Ei}(-x)$.] Applying this identity to Eq. (30) gives

$$\begin{aligned} g(z) &= 1 - \sum_{s=1}^{c}\binom{c}{s}\frac{1}{(s-1)!} \\ &\quad\times\left[e^{1/(1-z)}\Gamma\left(0,\frac{1}{1-z}\right) + \sum_{m=1}^{s-1}(-1)^m(m-1)!(1-z)^m\right] \\ &= q(z) - A_c e^{1/(1-z)}\Gamma\left(0,\frac{1}{1-z}\right), \end{aligned} \tag{33}$$

where $q(z)$ is a polynomial of degree $c-1$ and

$$A_c = \sum_{s=1}^{c}\binom{c}{s}\bigg/(s-1)!$$

depends only on c. For $k\geq c$, then, the degree distribution p_k is given by the coefficients of z^k in $-A_c e^{1/(1-z)}\Gamma(0,1/(1-z))$. We determine these coefficients as follows. Changing the variable of integration to $x=y-z/(1-z)$, we find

$$-e^{1/(1-z)}\Gamma\left(0,\frac{1}{1-z}\right) = -e\int_1^\infty \frac{e^{-x}}{x+z/(1-z)}dx. \tag{34}$$

Then we expand the integrand to get

$$\frac{1}{x+z/(1-z)} = \frac{1}{x} - \sum_{k=1}^{\infty}\left(1-\frac{1}{x}\right)^{k-1}\frac{z^k}{x^2}. \tag{35}$$

Commuting the sum and the integral, we obtain

$$-e^{1/(1-z)}\Gamma\left(0,\frac{1}{1-z}\right) = \sum_{k=0}^{\infty}a_k z^k, \tag{36}$$

where

$$a_0 = -e\int_1^\infty \frac{e^{-x}}{x}dx = -e\Gamma(0,1), \tag{37}$$

and for $k\geq 1$,

$$a_k = e\int_1^\infty\left(1-\frac{1}{x}\right)^{k-1}\frac{e^{-x}}{x^2}dx. \tag{38}$$

Integrating by parts, we obtain a slightly simpler expression

$$a_k = \frac{e}{k}\int_1^\infty\left(1-\frac{1}{x}\right)^{k}e^{-x}dx. \tag{39}$$

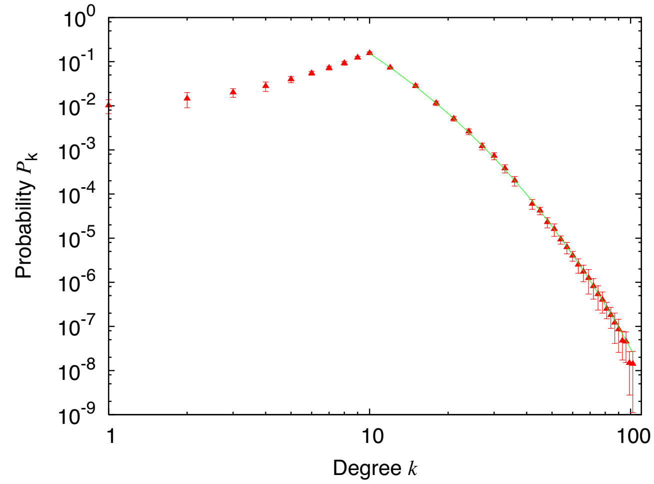

FIG. 2. (Color online) The degree distribution for our model in the case of fixed size $n=50\,000$ and $c=10$ with linear preferential attachment. The points represent data from our numerical simulations, and the solid line is the analytic solution for $k\geq c$. Note that the tail of the distribution does not follow a power law as in growing networks with preferential attachment, but instead decays faster than a power law, as a stretched exponential.

While the coefficients a_k can be expressed exactly using hypergeometric functions, a more informative approach is to employ a saddle-point expansion. The integrand of Eq. (39) is unimodal in the interval between 1 and ∞ and peaks at $x=\frac{1}{2}(1+\sqrt{4k+1})\simeq\sqrt{k}$. Approximating the integrand as a Gaussian around this point, we obtain as $k\to\infty$,

$$a_k \simeq \sqrt{\pi e}k^{-3/4}e^{-2\sqrt{k}} \tag{40}$$

and $p_k=A_c a_k$ for $k\geq c$ as stated above.

Figure 2 shows the form of this solution for the case $c=10$. Also shown in the figure are results from computer simulations of the model on systems of size $n=50000$ with $c=10$, which agree well with the analytic results. The appearance of the stretched exponential in Eq. (40) is worthy of note. We are aware of only a few cases of graphs with stretched exponential degree distributions that have been discussed previously—for instance, in growing networks with sublinear preferential attachment [22] as well as in empirical network data [23].

C. Preferential attachment in a growing network

We now come to the third and most complex of our example networks, in which we combine preferential attachment with net growth of the network, $r<1$. (Logically, we should perhaps first solve the case of a growing network without preferential attachment, which in fact we have done. But the solution turns out to have no qualitatively new features to distinguish it from the constant size case and is mathematically tedious besides. Given the large amount of effort it requires and its modest rewards, therefore, we prefer to skip this case and move on to more fertile ground.)

As before, perfect linear preferential attachment implies $\pi_k=k/\langle k\rangle$ or

$$\pi_k = \frac{1}{2}(1+r)\frac{k}{c}, \tag{41}$$

where we have made use of Eq. (3). Then $f(z)=(1+r)zg'(z)/2c$ and Eq. (9) becomes

$$g(z) - (1-z)\left(r - \frac{1}{2}(1+r)z\right)\frac{dg}{dz} = z^c. \tag{42}$$

An integrating factor for the left-hand side in this case is $|(\alpha-z)/(1-z)|^{-2/(1-r)}$ where $\alpha=2r/(1+r)$. (Note that $\alpha<1$ when $r<1$.) Unfortunately, this integrating factor is nonanalytic at $z=\alpha$, which makes integrals traversing this point cumbersome. To circumvent this difficulty, we observe that the second term in Eq. (42) vanishes at $z=\alpha$, giving $g(\alpha)=\alpha^c$. This provides us with an alternative boundary condition on $g(z)$, allowing us to fix the integrating constant while only integrating up to $z=\alpha$. It is then straightforward to show that

$$g(z) = \frac{2}{1+r}\left(\frac{\alpha-z}{1-z}\right)^{-2/(1-r)}\int_z^\alpha \left(\frac{\alpha-t}{1-t}\right)^{2/(1-r)}\frac{t^c dt}{(1-t)(\alpha-t)}, \tag{43}$$

for $z\leqslant\alpha$. Since the degree distribution is entirely determined by the behavior of $g(z)$ at the origin, it is adequate to restrict our solution to this regime.

Changing variables to $u=(\alpha-t)/(1-\alpha)$, we find

$$g(z) = \frac{2}{1+r}\left(\frac{\alpha-z}{1-z}\right)^{1-\gamma}(1-\alpha)^{-1} \times\int_0^{(\alpha-z)/(1-\alpha)}\left(\frac{u}{1+u}\right)^{\gamma}[\alpha-(1-\alpha)u]^c\frac{du}{u^2}, \tag{44}$$

where $\gamma=(3-r)/(1-r)$. If we expand the last factor in the integrand, this becomes

$$g(z) = \frac{2}{1+r}\sum_{s=0}^{c}(-1)^s\binom{c}{s}(1-\alpha)^{s-1}\alpha^{c-s} \times\left(\frac{\alpha-z}{1-z}\right)^{1-\gamma}\int_0^{(\alpha-z)/(1-\alpha)}\frac{u^{s+\gamma-2}}{(1+u)^\gamma}du. \tag{45}$$

We observe the following useful identity:

$$\begin{aligned}\int_0^x \frac{u^\beta}{(1+u)^\gamma}du &= \int_0^x\left(\frac{u}{1+u}\right)^\beta(1+u)^{\beta-\gamma}du\\ &= \frac{x^\beta}{(\beta-\gamma+1)(1+x)^{\gamma-1}}\\ &\quad - \frac{\beta}{\beta-\gamma+1}\int_0^x\frac{u^{\beta-1}}{(1+u)^\gamma}du,\end{aligned} \tag{46}$$

where the second equality is derived via integration by parts. Setting $\beta=s+\gamma-2$ and $x=(\alpha-z)/(1-\alpha)$ and noting that the last integral has the same form as the first, we can employ this identity iteratively $s-1$ times to get

$$\begin{aligned}&\left(\frac{\alpha-z}{1-z}\right)^{1-\gamma}\int_0^{(\alpha-z)/(1-\alpha)}\frac{u^{s+\gamma-2}}{(1+u)^\gamma}du\\ &= (-1)^{s+1}\frac{\Gamma(s+\gamma-1)}{\Gamma(s)}\\ &\times\left[\frac{1}{\Gamma(\gamma)}\left(\frac{\alpha-z}{1-z}\right)^{1-\gamma}\int_0^{(\alpha-z)/(1-\alpha)}\frac{u^{\gamma-1}}{(1+u)^\gamma}du\right.\\ &\left.+\sum_{m=1}^{s-1}\frac{(-1)^m}{\Gamma(\gamma+m)}\left(\frac{\alpha-z}{1-z}\right)^m\right].\end{aligned} \tag{47}$$

The final sum can be evaluated in closed form in terms of the incomplete Γ function, but our primary focus here is on the preceding term. Substituting into Eq. (45), we see that $g(z)=q(z)+A_{c,r}h(z)$, where

$$h(z) = -\left(\frac{\alpha-z}{1-z}\right)^{1-\gamma}\int_0^{(\alpha-z)/(1-\alpha)}\frac{u^{\gamma-1}}{(1+u)^\gamma}du, \tag{48}$$

$$A_{c,r} = \frac{2}{1+r}\sum_{s=0}^{c}\binom{c}{s}(1-\alpha)^{s-1}\alpha^{c-s}\frac{\Gamma(\gamma+s-1)}{\Gamma(\gamma)\Gamma(s)}, \tag{49}$$

and $q(z)$ is a polynomial of order $c-1$ in z.

Since $A_{c,r}$ depends only on c and r and $q(z)$ has no terms in z of order z^c or higher, the degree distribution for $k\geqslant c$ is, to within a multiplicative constant, given by the coefficients in the expansion of $h(z)$ about zero. Making the change of variables

$$u = \frac{y}{(1-z)/(\alpha-z)-y}, \tag{50}$$

we find that

$$h(z) = -\int_0^1\frac{y^{\gamma-1}dy}{(1-z)/(\alpha-z)-y}, \tag{51}$$

and expanding the integrand in powers of z we obtain $h(z)=\sum_{k=0}^\infty a_k z^k$ with

$$\begin{aligned}a_k &= (1-\alpha)\int_0^1\frac{(1-y)^{k-1}}{(1-\alpha y)^{k+1}}y^{\gamma-1}dy\\ &= \frac{\gamma-1}{k}\int_0^1\left(\frac{1-y}{1-\alpha y}\right)^k y^{\gamma-2}dy,\end{aligned} \tag{52}$$

for $k\geqslant1$, where the second equality follows via an integration by parts.

As in the case of constant size, we can express these coefficients in closed form using special functions, but if we are primarily interested in the form of the tail of the degree distribution, then a more revealing approach is to make a further substitution $y=x/k$, giving

$$a_k = (\gamma-1)k^{-\gamma}\int_0^k\frac{(1-x/k)^k}{(1-\alpha x/k)^k}x^{\gamma-2}dx. \tag{53}$$

In the limit of large k this becomes

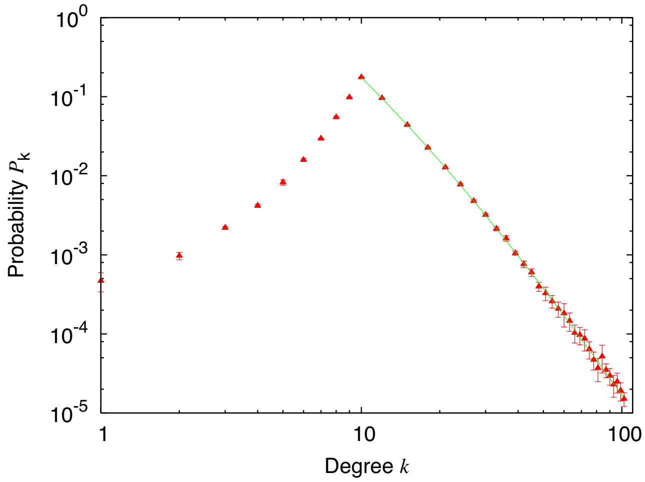

FIG. 3. (Color online) Degree distribution for a growing network with linear preferential attachment and $r=\frac{1}{2}$, $c=10$. The solid line represents the analytic solution, Eqs. (49) and (52), for $k \geqslant c$, and the points represent simulation results for systems with final size $n=100\,000$ vertices.

$$a_k \simeq (\gamma-1)k^{-\gamma}\int_0^\infty e^{-(1-\alpha)x}x^{\gamma-2}dx = \frac{\Gamma(\gamma)}{(1-\alpha)^{\gamma-1}}k^{-\gamma}, \quad (54)$$

and $p_k=A_{c,r}a_k$ for $k \geqslant c$ as stated above.

Thus the tail of the degree distribution follows a power law with exponent

$$\gamma = \frac{3-r}{1-r}. \quad (55)$$

Note that this exponent diverges as $r \to 1$ so that the power law becomes ever steeper as the growth rate slows, eventually assuming the stretched exponential form of Eq. (40)—steeper than any power law—in the limit $r=1$. In the limit $r \to 0$ we recover the established power-law behavior $a_k \sim k^{-3}$ for growing graphs with preferential attachment and no vertex removal [8–11].

In Fig. 3 we show the form of the degree distribution for this model for the case $r=\frac{1}{2}$, $c=10$, along with numerical results from simulations of the model on networks of (final) size $n=100\,000$ vertices. The power-law behavior is clearly visible on the logarithmic scales used as a straight line in the tail of the distribution. Once again the analytic solution and simulations are in excellent agreement.

We note that Sarshar and Roychowdhury [18] and, subsequently, Chung and Lu [19] and Cooper, Frieze, and Vera [20] independently demonstrated power-law behavior in the degree distribution of networks in the case $r<1$. Their results focus on the tail of the distribution rather than on exact solutions, but they find the same dependence of the exponent on the growth rate.

IV. DISCUSSION

In this paper we have studied models of the time evolution of networks in which, in addition to the widely considered case of addition of vertices, we also include vertex removal. We have given exact solutions for cases in which vertices are added and removed at the same rate, a potential model for steady-state networks such as peer-to-peer networks, and cases in which the rate of addition exceeds the rate of removal, which we regard as a simple model for the growth of, for example, the worldwide web.

We find very different behaviors in these various cases. For a steady-state network in which newly added vertices attach to others at random we find a degree distribution, Eqs. (19) and (20), which is sharply peaked about its maximum and has a rapidly decaying (Poisson) tail. This distribution is quite unlike the right-skewed degree distributions found in many real-world networks, but as a possible form for a "designed" network such as a peer-to-peer network it might be preferable over skewed forms, being more homogeneous and hence distributing traffic more evenly.

If newly appearing vertices attach to others using a linear preferential attachment mechanism, whereby vertices gain new edges in proportion to the number they already possess, we find that the degree distribution becomes a stretched exponential, Eqs. (39) and (40), a substantially broader distribution than that of the random attachment case, though still more rapidly decaying than the power laws often seen in growing networks.

And in the case where the network shows net growth, adding vertices faster than it loses them, we find that the degree distribution follows a power law, Eqs. (52) and (54), with an exponent γ that assumes values in the range $3 \leqslant \gamma < \infty$, diverging as the growth rate tends to zero.

This last result is of interest for a number of reasons. First, it shows that power-law behavior can be rigorously established in networks that grow but also lose vertices. Most previous analytic models of network growth have focused solely on vertex addition. And while the real worldwide web and other networks appear to have degree distributions that closely follow power laws, these networks also clearly lose vertices as well as gaining them. The results presented here demonstrate that the widely studied mechanism of preferential attachment for generating power-law behavior also works in this regime.

On the other hand, the large values of the exponent γ generated by our model appear not to be in agreement with the behavior observed in real-world networks, most of which have exponents in the range from 2 to 3 [1–3]. There are well-known mechanisms that can reduce the exponent from 3 to values slightly lower—specifically the generalization of the preferential attachment model to the case of a directed network [8,11], which is in any case a more appropriate model for the worldwide web. In the limit of low growth rate, however, our model predicts a diverging exponent and, while the exact value may not be accurate because of a host of complicating factors, it seems likely that the divergence itself is a robust phenomenon; as other authors have commented, there are good reasons to believe that net growth is one of the fundamental requirements for the generation of power-law degree distributions by the kind of mechanisms considered here.

Thus the fact that we do not observe very large exponents in real networks appears to indicate that most networks are in a regime where growth dominates over vertex loss by a wide margin. It is possible, however, that this will not always be the case. The web, for example, has certainly being enjoying a period of very vigorous growth since its appearance in the early 1990s, but it could be that this is a sign primarily of its youth and that as the network matures its size will grow more slowly, the vertices added being more nearly balanced by those taken away. Were this to happen, we would expect to see the exponent of the degree distribution grow larger. A sufficiently large exponent would make the distribution indistinguishable experimentally from an exponential or stretched exponential distribution, although we do not realistically anticipate seeing behavior of this type any time in the near future.

ACKNOWLEDGMENTS

This work was funded in part by the National Science Foundation under Grant Nos. DMS-0405348 and PHY-0200909 and by the James S. McDonnell Foundation. C.M. thanks the University of Michigan and the Center for the Study of Complex Systems for their hospitality while part of this work was carried out and Tracy Conrad and Rosemary Moore for their support.

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[2] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).
[3] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[4] D. J. de S. Price, Science **149**, 510 (1965).
[5] S. Redner, Eur. Phys. J. B **4**, 131 (1998).
[6] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **401**, 130 (1999).
[7] J. M. Kleinberg, S. R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, in *Proceedings of the 5th Annual International Conference on Combinatorics and Computing*, edited by T. Asano, H. Imai, D. T. Lee, S.-I. Nakano, and T. Tokuyama, *Lecture Notes in Computer Science* (Springer, Berlin, 1999), Vol. 1627, pp. 1–18.
[8] D. J. de S. Price, J. Am. Soc. Inf. Sci. **27**, 292 (1976).
[9] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[10] P. L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000).
[11] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000).
[12] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády, Random Struct. Algorithms **18**, 279 (2001).
[13] S. N. Dorogovtsev and J. F. F. Mendes, Europhys. Lett. **52**, 33 (2000).
[14] R. Albert and A.-L. Barabási, Phys. Rev. Lett. **85**, 5234 (2000).
[15] P. L. Krapivsky and S. Redner, Comput. Netw. **39**, 261 (2002).
[16] B. Tadić, Physica A **314**, 278 (2002).
[17] A. Grönlund, K. Sneppen, and P. Minnhagen, Phys. Scr. **71**, 680 (2005).
[18] N. Sarshar and V. Roychowdhury, Phys. Rev. E **69**, 026101 (2004).
[19] F. Chung and L. Lu, Internet Mathematics **1**, 409 (2004).
[20] C. Cooper, A. Frieze, and J. Vera, Internet Mathematics **1**, 463 (2004).
[21] P. Erdős and A. Rényi, Publ. Math. Inst. Hung. Acad. Sci. **5**, 17 (1960).
[22] P. L. Krapivsky and S. Redner, Phys. Rev. E **63**, 066123 (2001).
[23] M. E. J. Newman, S. Forrest, and J. Balthrop, Phys. Rev. E **66**, 035101(R) (2002).