

Stochastically evolving networksDerek Y. C. Chan,^{*} Barry D. Hughes,[†] and Alex S. Leong*Particulate Fluids Processing Centre, Department of Mathematics and Statistics, The University of Melbourne, Victoria 3010, Australia*William J. Reed[‡]*Department of Mathematics and Statistics, University of Victoria, Victoria, British Columbia, Canada V8W 3P4*

(Received 4 June 2003; published 31 December 2003)

We discuss a class of models for the evolution of networks in which new nodes are recruited into the network at random times, and links between existing nodes that are not yet directly connected may also form at random times. The class contains both models that produce “small-world” networks and less tightly linked models. We produce both trees, appropriate in certain biological applications, and networks in which closed loops can appear, which model communication networks and networks of human sexual interactions. One of our models is closely related to random recursive trees, and some exact results known in that context can be exploited. The other models are more subtle and difficult to analyze. Our analysis includes a number of exact results for moments, correlations, and distributions of coordination number and network size. We report simulations and also discuss some mean-field approximations. If the system has evolved for a long time and the state of a random node (which thus has a random age) is observed, power-law distributions for properties of the system arise in some of these models.

DOI: 10.1103/PhysRevE.68.066124

PACS number(s): 89.75.Fb, 89.75.Da, 89.75.Hc

I. INTRODUCTION

The rapidly increasing volume of theoretical work on models of growing networks parallels the continuing explosive growth of the communication networks that the models purport to describe [1]. Evolving network models also have important applications in biology and social phenomena [2]. The reviews of Albert and Barabási [3] and Dorogovtsev and Mendes [4] collect many applications of evolving network models, and outline the principal approaches to date in the formulation and analysis of such models. There is particular interest in “scale-free” networks for which many properties have power-law distributions, and in networks that possess the “small-world” property that internode distances are typically small.

In this paper, we formulate a class of network models as well-defined stochastically evolving systems, and we are able to exhibit a number of exact results on the time evolution of the number of nodes in the network, the coordination number [5] of a chosen node, and other observable network properties. In addition, computer simulations and mean-field-type arguments [6] are used to study interesting quantities that do not seem easily captured by analytic arguments. Some of the results of Secs. II and III have been previously obtained in analogous discrete-step models [7], but the continuous-time models discussed here are especially suited to the analysis by techniques with which physicists may be more familiar and these techniques are found useful for the more subtle models of Sec. IV.

The models are discussed in order of increasing complex-

ity. In Sec. II, we consider a growing tree, based on the birth process of Yule. We recover the limiting coordination-number distribution 2^{-k} ($k \in \mathbb{N}$) long known as a rigorous result for the “random recursive tree” discrete-time analog [8], and exploiting the relation between Yule trees and random recursive trees [9–11] leads to additional exact results for Yule trees.

Yule trees lack the small-world property of many observed networks. In Sec. III, we discuss a stochastic model in the spirit of the discrete-time models of Szymański [12,13] and Albert and Barabási [3], introduced by Reed and Hughes [14], in which nodes of high coordination number in a growing tree are favored to recruit new nodes into the network, and an asymptotic power-law distribution of coordination number (long known [12] and periodically rediscovered in the discrete context) is rigorously derived.

In Sec. IV, we consider the extensions of the models of Secs. II and III to incorporate cross-linking. Here the analysis becomes more difficult, but some analytic results are still available. Mean-field arguments suggest that for one of the cross-linked models, the mean coordination number is asymptotically proportional to the network size, while for the other model mean coordination number grows as the square root of network size. The first of these mean-field results is also established rigorously, while the second is shown in excellent agreement with simulation.

Several different types of observation of evolving networks need to be carefully distinguished. One may consider the number of nodes in the network, or the coordination number of an individual *fixed* node in the network, as a function of time, and many interesting observable properties exhibit exponential growth or decay. For the second type of observation, a *random node* is selected. The distribution of the coordination number of this node gives the marginal distribution of the coordination number over all nodes in such

^{*}Email address: D.Chan@ms.unimelb.edu.au

[†]Email address: B.Hughes@ms.unimelb.edu.au

[‡]Email address: reed@math.uvic.ca

networks. As the age of the randomly chosen node is itself random, this has an influence on the coordination number of the node and, for some of the models, power-law distributions for the coordination number result. For the third type of observation, the whole network is inspected at a *random time* with a given probability density function, and the variation of node age affects the properties of the system. This models the structure of a network at the instant before its death due to a single catastrophic random disaster.

Where we find power laws, these result from the competition between exponential growth of the structural properties, and exponential decay of the waiting-time density for the random observation or for the probability density function for node age, paralleling similar observations of two of the authors in a number of other contexts in taxonomy of living [15] and extinct [16] genera, gene and protein family size distributions [17], surname distributions [18], and other areas [19], and in a similar spirit to an explanation proposed by Fermi for the energy spectrum of cosmic rays [20]. Some remarks on possible applications of the network growth models in the present paper will be found in Sec. V.

The distribution of shortest path lengths between nodes in random networks is of considerable interest, and we include some mean-field predictions for our two treelike networks and simulation results for all networks we have studied. A related quantity is the “Wiener index” [21], which is the sum of the shortest connecting path lengths for all pairs of nodes in the network, and we include some results for this quantity.

We adopt the following notational conventions: random variables are capitalized, with their generic values denoted by corresponding lowercase letters; angle brackets denote expectation; and if a subscripted quantity is summed over one of its subscripts, that subscript is replaced by a bullet, e.g., $a_{l,\bullet,n} = \sum_m a_{l,m,n}$. In a number of figures, we display predicted distributions of integer-valued random variables. For clarity in the comparison between predictions and simulation results, the predicted distributions are usually interpolated to produce continuous curves. These interpolations are made by analytic continuations of the formulas for the predicted distributions, for example, replacing $k!$ by $\Gamma(k+1)$, and all special function calculations are performed using MATHEMATICA [22]. Standard deviations of simulation data are always computed as the standard deviation of the empirical distribution, which differs only slightly for the sample sizes considered from the standard deviations that would be obtained from unbiased estimates of the variance.

II. YULE TREES

Suppose there are n nodes in the network, with varying coordination numbers [5]. We allow any node to connect to an external isolated node, thereby bringing it into the network. For the present, we make this phenomenon independent of the current coordination number of the node, and we assume that the probability of a given node bringing a new node into the network in the time interval $(t, t+h]$ is $\lambda h + o(h)$ as $h \rightarrow 0$. This process is simply a standard linear birth process (Yule process), so that we shall call the resulting time-evolving trees “Yule trees,” but we shall ask some

questions about the trees that have no natural analog in the original birth process.

Since we shall answer some questions about this process using simulation, we observe that apart from the efficient choice of a data structure to store information of interest (see Appendix A), simulation is straightforward. The waiting-time density for a given node next to give birth is $\lambda e^{-\lambda t}$. Each time any node gives birth, we draw exponentially distributed random numbers to determine when that node next gives birth, and when its newly created offspring next gives birth. We can therefore easily create the complete time history of a realization of the process, and collect data as a function of time, or as a function of the size (number of nodes) in the network.

Denote the number of nodes in the network at time t by $N(t)$ and the number connected to a specific node (call it node $*$) by $K(t)$. Let $p_{k,n}(t) = \Pr\{K(t)=k, N(t)=n\}$, with $p_{k,n}(t)=0$ for $n \leq 0$. Then

$$p_{k,n}(t+h) = p_{k-1,n-1}(t)\lambda h + p_{k,n-1}(t)\lambda h(n-2) + p_{k,n}(t)[1 - \lambda hn] + o(h). \quad (1)$$

The first three terms on the right-hand side account for one birth at node $*$, one birth at some node other than node $*$, and no births at all in the time interval $(t, t+h]$, respectively. All other events have probabilities that are $o(h)$ as $h \rightarrow 0$ and so are negligible compared to the three terms exhibited. Subtracting $p_{k,n}(t)$ from both sides of Eq. (1), dividing by h , and taking the limit $h \rightarrow 0$, we obtain the differential-difference equation

$$\frac{d}{dt}p_{k,n} = \lambda p_{k-1,n-1} + \lambda(n-2)p_{k,n-1} - \lambda n p_{k,n}. \quad (2)$$

The equivalent partial differential equation for the generating function $\mathcal{P}(\kappa, \zeta, t) = \sum_{k=0}^{\infty} \sum_{n=1}^{\infty} p_{k,n}(t) \kappa^k \zeta^n$ is solved in Appendix B 1, and from that solution we are able to show that $K(t)$ and $N(t)$ are correlated for all finite times t , but that the correlation decays to zero as $t \rightarrow \infty$. We are also able to obtain exact formulas for the conditional mean and variance of $K(t)$ for a given network size, from which follow the asymptotic formulas

$$\langle K(t) | N(t) = n \rangle = \ln n + k_0 - \psi(n_0) + O(n^{-1}), \quad (3)$$

$$\text{var}\{K(t) | N(t) = n\} = \ln n - \psi(n_0) - \psi'(n_0) + O(n^{-1}), \quad (4)$$

where the digamma function $\psi(z) = \Gamma'(z)/\Gamma(z)$ is the logarithmic derivative of the usual Γ function.

We are able obtain more directly several quantities of interest that also come from the full analysis of Appendix B 1. If we sum over k we find that the marginal distribution of $N(t)$, $p_{\bullet,n}(t) = \sum_{k=0}^{\infty} p_{k,n}(t) = \Pr\{N(t) = n\}$, satisfies the evolution equation

$$\frac{d}{dt}p_{\bullet,n} = \lambda(n-1)p_{\bullet,n-1} - \lambda n p_{\bullet,n} \quad (5)$$

of the standard Yule process [23] with parameter λ . It follows from standard results (an easy proof uses generating functions) that $N(t)$ has a negative binomial distribution with parameters n_0 and $e^{-\lambda t}$:

$$p_{\bullet,n}(t) = \frac{(n-1)! e^{-n_0 \lambda t} (1 - e^{-\lambda t})^{n-n_0}}{(n_0-1)! (n-n_0)!}, \quad n \geq n_0; \quad (6)$$

$$\langle N(t) \rangle = n_0 e^{\lambda t}, \quad \text{var}\{N(t)\} = n_0 e^{\lambda t} (e^{\lambda t} - 1). \quad (7)$$

In particular, with $n_0=1$ we obtain the geometric distribution

$$p_{\bullet,n}(t) = e^{-\lambda t} (1 - e^{-\lambda t})^{n-1} \quad \text{for } n = 1, 2, \dots \quad (8)$$

We record for later use the result that for $n_0=1$,

$$\left\langle \frac{1}{N(t)} \right\rangle = \frac{e^{-\lambda t}}{1 - e^{-\lambda t}} \sum_{n=1}^{\infty} \frac{(1 - e^{-\lambda t})^n}{n} = \frac{\lambda t e^{-\lambda t}}{1 - e^{-\lambda t}}. \quad (9)$$

If we sum over n in Eq. (2), we find that the marginal distribution $p_{k,\bullet}(t) = \sum_{n=1}^{\infty} p_{k,n}(t) = \text{Pr}\{K(t) = k\}$ of $K(t)$ satisfies the evolution equation

$$\frac{d}{dt} p_{k,\bullet} = \lambda p_{k-1,\bullet} - \lambda p_{k,\bullet}. \quad (10)$$

This equation governs the Poisson process, and so has solution corresponding to the initial condition $K(0)=0$ given by $p_{k,\bullet}(t) = e^{-\lambda t} (\lambda t)^k / k!$ for $k=0, 1, 2, \dots$. This is the solution appropriate for the unique initial node of the system, which we shall call the ‘‘primal node.’’ For any other node, joined to the network at time t_* , we have $K(t_*)=1$, and we find that

$$p_{k,\bullet}(t) = e^{-\lambda(t-t_*)} [\lambda(t-t_*)]^{k-1} / (k-1)! \quad (11)$$

for $k \in \mathbb{N}$ and $t \geq t_*$.

Feigin [24] has shown that for a small number of stochastic processes that create elements at random times and possess the ‘‘order statistics property,’’ the lifetime distribution of a random element can be simply calculated. In particular, for the Yule process that underlies the present model, the lifetime probability density function for a node (other than the primal node from which the entire network was created) is given by

$$f(\tau) = \frac{\lambda e^{-\lambda \tau}}{1 - e^{-\lambda t}}, \quad 0 \leq \tau \leq t, \quad (12)$$

if the current age of the whole system is t . Hence the coordination-number distribution of a randomly chosen node is given by

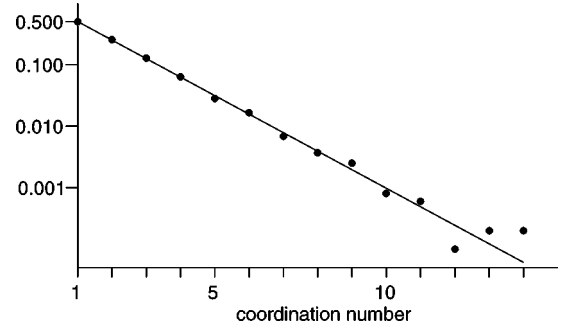


FIG. 1. For Yule trees (Sec. II) we compare in a log-linear plot the limiting coordination-number distribution 2^{-k} (solid line) with the distribution obtained from one realization of a tree of 10 000 nodes obtained at $\lambda t \approx 12.35$.

$$\text{Pr}\{K(t) = k(\text{random node})\} = \frac{e^{-\lambda t} (\lambda t)^k}{k!} \left\langle \frac{1}{N(t)} \right\rangle + \left\langle 1 - \frac{1}{N(t)} \right\rangle \int_0^t \frac{\lambda (\lambda \tau)^{k-1} e^{-2\lambda \tau} d\tau}{(1 - e^{-\lambda \tau}) (k-1)!}. \quad (13)$$

Using Eq. (9) we deduce that

$$\lim_{t \rightarrow \infty} \text{Pr}\{K(t) = k(\text{random node})\} = 2^{-k}. \quad (14)$$

In Fig. 1 we compare the limiting coordination-number distribution 2^{-k} with one realization of a network of 10 000 nodes. The fit is excellent except for the largest coordination numbers encountered. For the limiting distribution the mean coordination number is exactly 2, and half of all nodes have a single link. In the discrete-time context, the limiting distribution (14) was proved in 1987 by Szymański [8], who also notes that it was known earlier, and has been independently obtained in a nonrigorous manner more recently by Krapivsky *et al.* [25].

In the present model, each node has a unique path back to the primal node from which the entire tree has grown. We may think of all nodes with the same distance back to the primal node as lying on the same ring when the tree is drawn so that its nodes lie on concentric rings, and we refer to the number of links between a given node and the primal node as the ‘‘ring number’’ r . We shall derive a mean-field estimate (that is, an estimate neglecting fluctuations) [6] of the number of nodes with ring number r at time t , and we denote this estimate by $n(r, t)$. We have, in a mean-field treatment,

$$\frac{\partial}{\partial t} n(r, t) = \lambda n(r-1, t), \quad r \geq 1, \quad (15)$$

with $n(r, 0) = 0$ for $r \geq 1$, and $n(0, t) = 1$ for all t . The obvious generating function solution shows that

$$n(r, t) = (\lambda t)^r / r!, \quad (16)$$

and the total number of nodes present at time t is predicted to be $n_t = \sum_r (\lambda t)^r / r! = e^{\lambda t}$. This agrees with the exact result that $\langle N(t) \rangle = e^{\lambda t}$ noted above. We therefore predict that for a randomly chosen node,

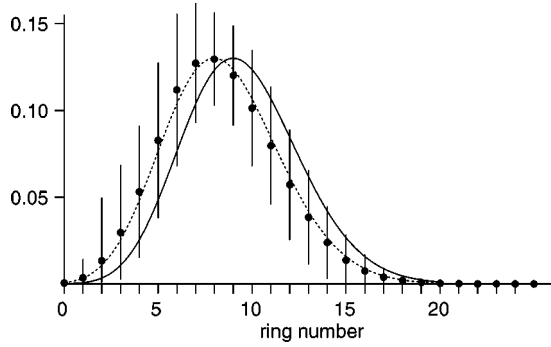


FIG. 2. For Yule trees (Sec. II), we compare the ring-number distribution predicted by the mean-field theory (17), shown as the solid line, with simulations, showing mean \pm one standard deviation. The data correspond to 100 realizations stopped at $\lambda t = 9.5$, when the mean network size is approximately 11 201 nodes (standard deviation $\approx 13\,061$). The dotted line is the exact ring-number distribution (23).

$$\Pr\{\text{ring number} = r\} = e^{-\lambda t} (\lambda t)^r / r!, \quad (17)$$

a Poisson distribution with mean λt . In Fig. 2 we compare simulation results with prediction (17). We find that for a fixed time t , the realization-to-realization fluctuations in the ring-number distribution are large and, in particular, the effects of those realizations in which the number of nodes generated is comparatively small spoils the performance of the mean-field theory.

From the mean-field prediction (17), we can infer a corresponding mean-field prediction for the ring number at a given network size by writing $\lambda t = \ln n_t$. We arrive at the prediction that

$$\Pr\{\text{ring number} = r\} = (\ln n_t)^r / (n_t r!), \quad (18)$$

with the mean ring number being given by $\ln n_t$. We have compared prediction (18) with statistics gathered for networks of fixed size in Fig. 3. The fluctuations are greatly reduced, and the mean-field calculation closely tracks the mean found in simulations. Note however that Eq. (4) shows

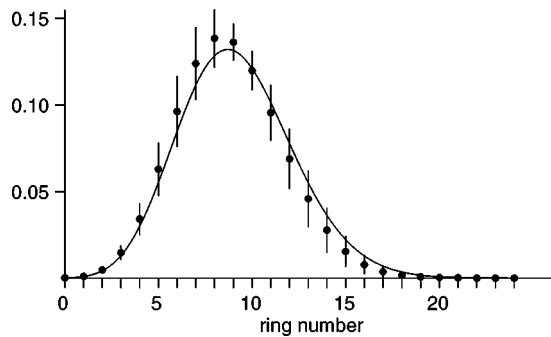


FIG. 3. For Yule trees (Sec. II), we compare the ring-number distribution predicted by the mean-field theory (18), shown as the solid line, with simulations, showing mean \pm one standard deviation. The data correspond to 100 realizations, all stopped when the network size is exactly 10 000 nodes (λt mean value ≈ 10.03 , with standard deviation ≈ 1.30).

that even after conditioning on a fixed large network size, there can be large fluctuations in the coordination numbers of older nodes (corresponding to $k_0 = 1$ and n_0 small).

We can extend the mean-field theory to include the coordination number of nodes. If we let $n(k, r, t)$ denote the mean-field estimate of the number of nodes with coordination number k and ring number r , we have

$$\begin{aligned} \frac{d}{dt} n(k, r, t) &\approx \lambda n(k-1, r, t) - \lambda n(k, r, t) \\ &+ \delta_{k,1} \sum_{k' \geq 1} \lambda n(k', r-1, t). \end{aligned} \quad (19)$$

We solve this equation using generating functions in Appendix C 1 and deduce that

$$\lim_{t \rightarrow \infty} \frac{\sum_{r=0}^{\infty} n(k, r, t)}{\sum_{k=0}^{\infty} \sum_{r=0}^{\infty} n(k, r, t)} = 2^{-k}, \quad (20)$$

so the mean-field theory produces the correct limiting coordination-number distribution (14).

If we suppress the explicit time dependence and study the structure of the tree that grows solely in terms of the number of nodes present at a given time (that is, we condition on $N_t = n_t$), then the problem reduces to that of the “random recursive tree,” for which several exact results are known. From work of Szymanski [8] it is known that for a randomly chosen node the ring-number distribution has the generating function

$$\sum_{r=0}^{\infty} \Pr\{\text{ring number} = r | N_t = n_t\} \rho^r = \frac{\Gamma(n_t + \rho)}{\Gamma(n_t + 1) \Gamma(\rho + 1)} \quad (21)$$

and $\Pr\{\text{ring number} = r | N_t = n_t\}$ can be expressed in terms of Stirling numbers. The mean ring number, found by differentiation of Eq. (21), is

$$\psi(n_t + 1) - \psi(2) = \ln(n_t + 1) + 1 - \gamma + O(n_t^{-1}) \sim \ln n_t,$$

where ψ is the digamma function as before and γ is Euler’s constant, so that the error in the mean-field prediction of the mean ring number at given network size n_t is $O(1)$ as $n_t \rightarrow \infty$.

For the case of a time-evolving tree network grown from a single initial node, we know the distribution of the number of nodes N_t from Eq. (8). Thus we can determine the generating function for the ring-number distribution at time t :

$$\begin{aligned}
& \sum_{r=0}^{\infty} \Pr\{\text{ring number}=r\} \rho^r \\
&= \sum_{n=1}^{\infty} \frac{\Gamma(n+\rho)}{\Gamma(n+1)\Gamma(\rho+1)} e^{-\lambda t} (1-e^{-\lambda t})^{n-1} \\
&= \frac{e^{-\lambda t}}{1-e^{-\lambda t}} \sum_{n=0}^{\infty} \frac{\Gamma(n+\rho+1)}{n!\Gamma(\rho+1)} \int_0^{1-e^{-\lambda t}} z^n dz \\
&= \frac{e^{-\lambda t}}{1-e^{-\lambda t}} \int_0^{1-e^{-\lambda t}} (1-z)^{-1-\rho} dz \\
&= \frac{e^{\rho\lambda t}-1}{\rho(e^{\lambda t}-1)} = \sum_{r=0}^{\infty} \frac{(\lambda t)^{r+1} \rho^r}{(e^{\lambda t}-1)(r+1)!} \quad (22)
\end{aligned}$$

and so we have the exact ring-number distribution

$$\Pr\{\text{ring number}=r\} = \frac{(\lambda t)^{r+1}}{(e^{\lambda t}-1)(r+1)!}. \quad (23)$$

The exact mean ring number, found by differentiation of Eq. (22), is $\lambda t(1-e^{-\lambda t})^{-1}-1 \sim \lambda t$ as $t \rightarrow \infty$. Mean-field theory predicted the mean ring number λt for all $t > 0$. We have plotted the exact distribution (23) as the dotted line in Fig. 2. It may be emphasized that this exact distribution is calculated over all tree realizations. It reproduces well the simulation averages, but the large standard deviations of the simulation data reflect the fact that this exact distribution may poorly capture the relative abundance of ring numbers in particular realizations at a given time of the time-evolving random tree. The covariance of the numbers of nodes for two given ring numbers is known for the random recursive tree [11], so that exact results for the covariance of the ring-number distribution for our time-evolving problem should also be able to be deduced, though we do not pursue this here.

In the mean-field treatment, $n(1,t) \rightarrow \infty$ as $t \rightarrow \infty$, that is, the number of different branches of the tree that join at the primal node diverges. Mean-field theory cannot resolve the sizes of these branches. It is tempting (though ill-advised, as discussed below) to propose on symmetry grounds that the branches are of comparable size, and if this is so, the probability that the (unique) shortest path joining an arbitrary pair of nodes with ring numbers r and s does not pass through the primal node decays to zero, and we may estimate the probability distribution function for the shortest path by assuming that the nodes inhabit different branches of the tree that join at the primal node. Hence

$$\begin{aligned}
& \Pr\{\text{path length}=l\} \\
& \approx \sum_{r=0}^l \Pr\{\text{ring number}=l-r\} \Pr\{\text{ring number}=r\} \quad (24)
\end{aligned}$$

and evaluating the sum on the right-hand side we find that in the simplest mean-field treatment,

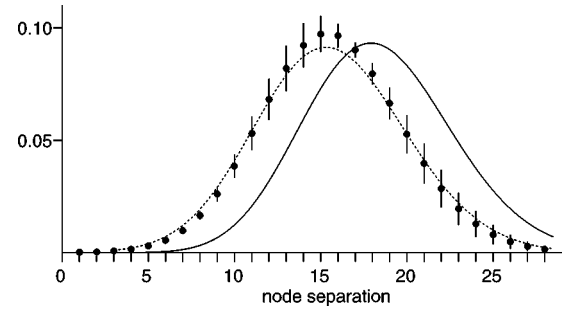


FIG. 4. For Yule trees (Sec. II), we compare the internode distance distribution predicted by mean-field arguments with simulations (showing mean \pm one standard deviation). The data correspond to 100 realizations, all stopped when the network size is exactly 10 000 nodes (λt mean value ≈ 10.03 , with standard deviation ≈ 1.30). The solid curve is the simple mean-field theory (26). The dotted curve is the improved mean-field theory (30).

$$\Pr\{\text{path length}=l\} \approx e^{-2\lambda t} (2\lambda t)^l / l!. \quad (25)$$

For an estimate written in terms of the number of nodes in the network rather than time, we set $n = e^{\lambda t}$ for the number of nodes, and deduce that

$$\Pr\{\text{path length}=l\} \approx (2 \ln n_t)^l / (n_t^2 l!). \quad (26)$$

When prediction (25) is compared against simulations at fixed time, it performs quite poorly. The fixed size prediction (26) does somewhat better, but is still disappointing (see Fig. 4, where the solid curve is the simple mean-field theory).

The reason for the failure of the mean-field theory to capture the path-length distribution comes from the typical asymmetry of individual realizations of the tree. Each pair of distinct nodes is joined by a unique path, and if we regard the nodes themselves as belonging to the path, there is a well-defined smallest ring number encountered on the path. In Table I, we show estimates from simulation of the probability ϕ_r that two randomly chosen distinct nodes have r for the smallest ring number encountered on the path that joins them. The data suggest that in an appropriate limit of long time or large network size, $\phi_r \rightarrow 2^{-r-1}$, but we have not been able to prove this. The probability that the path linking two arbitrarily chosen nodes passes through the primal node is around 0.5, whereas in the mean-field argument this probability should converge to 1 as time progresses.

We shall give a simple argument to demonstrate the asymmetry. Consider that the model of this section started not only with the primal node but with the primal node O and its first daughter A present at time $t=0$. At this instant, because of the lack of memory inherent in the exponential waiting-time density for births, the nodes O and A become equivalent and independent and their number of offspring will in each case be $\approx e^{\lambda t}$ as $t \rightarrow \infty$, so the first daughter has as descendants half the nodes of the tree. This conclusion can also be obtained by a more formal rigorous argument, based on generating functions [26]. This shows that individual realizations of the process typically exhibit significant asymmetry when

TABLE I. For Yule trees (Sec. II), we show estimates of the probability ϕ_r that the smallest ring number encountered on the unique path joining two randomly chosen distinct nodes is r . The results (experimental mean $\bar{\phi}_r$ and standard deviation) are obtained from 100 simulations of a network of 10 000 nodes.

| r | $\bar{\phi}_r$ | Standard deviation | $2^{r+1}\bar{\phi}_r$ |
|-----|------------------------|-----------------------|-----------------------|
| 0 | 5.03×10^{-1} | 2.03×10^{-1} | 1.006 |
| 1 | 2.46×10^{-1} | 1.47×10^{-1} | 0.983 |
| 2 | 1.23×10^{-1} | 8.10×10^{-2} | 0.982 |
| 3 | 6.57×10^{-2} | 5.40×10^{-2} | 1.051 |
| 4 | 3.23×10^{-2} | 2.68×10^{-2} | 1.035 |
| 5 | 1.47×10^{-2} | 1.30×10^{-2} | 0.943 |
| 6 | 8.11×10^{-3} | 7.31×10^{-3} | 1.038 |
| 7 | 3.76×10^{-3} | 3.28×10^{-3} | 0.963 |
| 8 | 1.94×10^{-3} | 1.82×10^{-3} | 0.992 |
| 9 | 9.69×10^{-4} | 8.31×10^{-4} | 0.992 |
| 10 | 4.45×10^{-4} | 3.78×10^{-4} | 0.911 |
| 11 | 2.06×10^{-4} | 1.65×10^{-4} | 0.842 |
| 12 | 1.01×10^{-4} | 1.17×10^{-4} | 0.825 |
| 13 | 4.56×10^{-5} | 5.40×10^{-5} | 0.748 |
| 14 | 1.89×10^{-5} | 2.58×10^{-5} | 0.619 |
| 15 | 7.67×10^{-6} | 1.13×10^{-5} | 0.502 |
| 16 | 3.87×10^{-6} | 9.01×10^{-6} | 0.507 |
| 17 | 1.35×10^{-6} | 3.24×10^{-6} | 0.355 |
| 18 | 4.83×10^{-7} | 1.51×10^{-6} | 0.253 |
| 19 | 1.50×10^{-7} | 4.98×10^{-7} | 0.157 |
| 20 | 5.16×10^{-8} | 2.21×10^{-7} | 0.108 |
| 21 | 1.92×10^{-8} | 1.00×10^{-7} | 0.081 |
| 22 | 6.20×10^{-9} | 3.73×10^{-8} | 0.052 |
| 23 | 8.00×10^{-10} | 6.27×10^{-9} | 0.013 |

viewed from the primal node, and this is why the mean-field prediction performs so poorly in comparison to simulations in Fig. 4.

We have been able to derive an improved mean-field theory. We consider network of size 2^m and we write $\varphi_m(l)$ for the probability that for an arbitrarily chosen pair of nodes the separation is l . The corresponding mean-field prediction is, from Eq. (26),

$$\varphi_m^0(l) \approx (m \ln 4)^l / (2^{2m} l!). \quad (27)$$

Assuming equivalence of the independent branches from O and A, we write

$$\varphi_{m+1}(l) = \frac{1}{2} [\varphi_m^0(l) + \varphi_m(l)]. \quad (28)$$

The two terms on the right correspond to cases where the nodes are on separate branches from the primal node [so each branch has approximately 2^m nodes—this is why we use $\varphi_m^0(l)$ rather than $\varphi_{m+1}^0(l)$], or on the same branch. The initial condition for the difference equation is $\varphi_1(l) = \delta_{l,1}$, and it is easy to deduce the exact solution

$$\varphi_m(l) = \frac{1}{2^m} \left(\sum_{k=1}^{m-1} 2^k \varphi_k^0(l) + 2 \delta_{l,1} \right). \quad (29)$$

This leads to a prediction of the internode distance distribution, but it is inconveniently restricted to networks of size 2^m . We recall that Lerch's transcendent $\Phi(z, s, a)$ and the polylogarithm $\text{Li}_m(z)$ are defined [27] by

$$\Phi(z, s, a) = \sum_{k=0}^{\infty} \frac{z^k}{(a+k)^s}, \quad \text{Li}_m(z) = \sum_{k=1}^{\infty} \frac{z^k}{k^m},$$

respectively. We now see that

$$\begin{aligned} \sum_{k=1}^{m-1} 2^k \varphi_m^0(l) &= \frac{(\ln 4)^l}{l!} \sum_{k=1}^{m-1} \frac{k^l}{2^k} \\ &= \frac{(\ln 4)^l}{l!} \left[\text{Li}_{-l} \left(\frac{1}{2} \right) - \Phi \left(\frac{1}{2}, -l, m \right) \right]. \end{aligned}$$

Since Lerch's transcendent $\Phi(z, s, a)$ is defined for all positive values of a , we are able to deduce a predicted internode distribution for arbitrary network size n by writing $m = \log_2 n$, giving

$$\begin{aligned} \text{Pr}\{\text{path length} = l\} &\approx \frac{2 \delta_{l,1}}{n} + \frac{(\ln 4)^l}{l! n} \\ &\times \left[\text{Li}_{-l} \left(\frac{1}{2} \right) - \Phi \left(\frac{1}{2}, -l, \log_2 n \right) \right]. \end{aligned} \quad (30)$$

This improved internode distance distribution is shown as the dotted curve in Fig. 4. The major failure of the simpler mean-field theory has been effectively repaired.

The average distance between pairs of nodes in a random recursive tree has been discussed by Moon [9] and Dobrow [28]. The related problem of the determination of the Wiener index $W(n)$, which is the sum over all node pairs of the internode distances, has been addressed by Neininger [29]. Where $H_n = \sum_{k=1}^n k^{-1}$ is the n th harmonic number, it is known [29] that

$$\langle W_n \rangle = n^2 H_n - 2n^2 + n H_n. \quad (31)$$

As $H_n = \ln n + \gamma + o(1)$ for $n \rightarrow \infty$, where γ is Euler's constant, we have the exact asymptotic form

$$\langle W(n) \rangle = n^2 (\ln n + \gamma - 2) + o(n^2). \quad (32)$$

We can use our naive mean-field theory to estimate the Wiener index

$$W(n) \approx \frac{1}{2} \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} (r+s) \frac{(\lambda t)^r}{r!} \frac{(\lambda t)^s}{s!} = \lambda t e^{2\lambda t}. \quad (33)$$

As noted above, in the mean-field approximation the total number of nodes is $n = e^{\lambda t}$, so we predict that

$$W(n) \approx n^2 \ln n. \quad (34)$$

The mean-field prediction thus agrees with the exact result to leading order, but the slow growth of the logarithm ensures that for quite large values of n the mean-field esti-

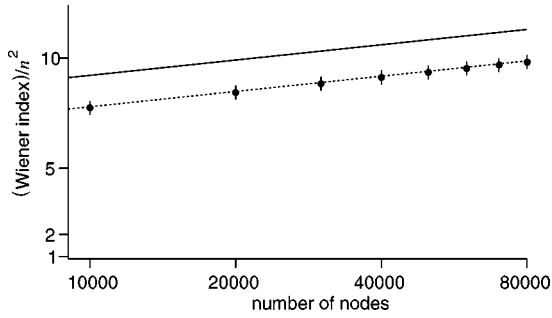


FIG. 5. For Yule trees (Sec. II), we compare the Wiener index predicted by Eq. (34) with simulations (ten realizations). We show the mean Wiener index $\langle W(n) \rangle$ from simulation \pm one standard deviation, scaled against n^2 . The solid curve is the mean-field estimate $W(n)/n^2 \approx \ln n$, which is asymptotically correct as $n \rightarrow \infty$, but performs poorly for the range of n values for which we have simulation data. The dotted line is the improved approximation $\langle W(n) \rangle/n^2 \approx \ln n + \gamma - 2$ based on the exact asymptotic form of $\langle W(n) \rangle$. Both approximations perform much better than the rigorous lower bound $W(n)/n^2 = (1 - n^{-2}) \sim 1$ and the rigorous upper bound $W(n)/n^2 = n(1 - n^{-2})/6 \sim n/6$ known for general trees.

mate differs significantly from the exact result. In Fig. 5 we compare the mean-field prediction $n^{-2}\langle W(n) \rangle \approx \ln n$ with the better approximation $n^{-2}\langle W(n) \rangle \approx (\ln n + \gamma - 2)$ based on the exact asymptotic form and simulation data. It is known in general that for an arbitrary tree of n nodes, the Wiener index $W(n)$ satisfies the inequality $(n-1)^2 \leq W(n) \leq n(n^2-1)/6$, with equality if and only if the tree is either a star or a linear chain, respectively [30], but neither of these bounds is usefully close to the exact result for $\langle W(n) \rangle$. One could derive an improved mean-field estimate of the expected Wiener index using Eq. (30), but as we have the exact value already, we have not pursued this.

The expected Wiener index as a function of time can be calculated exactly, using Eqs. (8) and (31). We find after a little algebra that

$$\begin{aligned} \langle W(t) \rangle &= \sum_{n=1}^{\infty} e^{-\lambda t} (1 - e^{-\lambda t})^{n-1} (n^2 H_n - 2n^2 + n H_n) \\ &= 2(\lambda t + 1) e^{2\lambda t} - e^{\lambda t}. \end{aligned} \quad (35)$$

Note that although the naive mean-field theory produced the correct asymptotic form for the average Wiener index as a function of the number of nodes present, it is more seriously deficient for the time evolution.

To conclude our discussion of this simplest model of a growing network, note that if the growing network is killed at a random time T after the creation of the primal node, where T has probability density function $\Psi(t)$, then from Eq. (8) the network size N^\dagger at death has the probability distribution

$$\Pr\{N^\dagger = n\} = \int_0^\infty \Psi(t) e^{-\lambda t} (1 - e^{-\lambda t})^{n-1} dt$$

for $n = 1, 2, 3, \dots$. In particular, for death at an exponentially distributed time, with $\Psi(t) = \nu e^{-\nu t}$, we find (cf. Refs. [15,19])

$$\Pr\{N^\dagger = n\} = \frac{\nu \Gamma(\nu/\lambda + 1) \Gamma(n)}{\lambda \Gamma(n + \nu/\lambda + 1)} \sim \frac{\text{const}}{n^{\nu/\lambda + 1}}$$

as $n \rightarrow \infty$. We can also calculate the ring-number distribution for a randomly killed tree from the exact distribution (23). Let R^\dagger denote the ring number of a randomly chosen node in a randomly killed tree. Then

$$\Pr\{R^\dagger = r\} = \int_0^\infty \frac{\Psi(t) (\lambda t)^{r+1} dt}{(e^{\lambda t} - 1)(r+1)!}.$$

For death at an exponentially distributed time, with $\Psi(t) = \nu e^{-\nu t}$, the integral can be evaluated exactly in terms of the Hurwitz zeta function [31]

$$\zeta(s, a) = \sum_{n=0}^{\infty} \frac{1}{(n+a)^s} = \frac{1}{\Gamma(s)} \int_0^\infty \frac{x^{s-1} e^{-(1-a)x} dx}{e^x - 1},$$

and we find that

$$\Pr\{R^\dagger = r\} = \frac{\nu}{\lambda} \zeta\left(r + 2, \frac{\nu}{\lambda} + 1\right) \sim \frac{\nu/\lambda}{(1 + \nu/\lambda)^{r+2}}$$

as $r \rightarrow \infty$.

III. REED-HUGHES TREES

In 1985, Szymański [12] considered a discrete-time randomly growing tree for which the probability of selection of a node as the next node to give birth is proportional to the node's coordination number. In Theorem 5 of his paper he proves a result which implies, but is slightly stronger than, the statement that the probability that a randomly chosen node has coordination number k converges as the tree size grows to $4/[k(k+1)(k+2)]$. Coordination-number dependent growth probabilities, or mechanisms equivalent to this differently expressed, have been made popular by the work of Albert and Barabási [32] and Dorogovtsev and Mendes [33] and co-authors and generalized in various ways, but the paper of Szymański [12] has received little recognition.

We shall consider a continuous-time tree model that favors birth from nodes of high coordination number and leads to scale-free networks. In this specific form the model is that introduced by Reed and Hughes [34]. It is essentially a continuous-time analog of Szymański's problem, but also (as in our discussion of Yule trees) the Reed-Hughes formulation addresses the joint time evolution of the coordination number of a tagged node and the overall network size.

For a node in the network, with coordination number k_i at time t , suppose that the probability of it bringing a new node into the network in the time interval $(t, t+h]$ is $\lambda k_i h + o(h)$ as $h \rightarrow 0$. Thus the well-connected nodes are much more likely to establish a new connection than the less well-connected nodes. As in Sec. II, we denote the number of

nodes in the network at time t by $N(t)$ and the number connected to a specific node (node $*$, say) by $K(t)$. Let $p_{k,n}(t) = \Pr\{K(t)=k, N(t)=n\}$. By similar arguments to those used to derive Eq. (1), we find that

$$p_{k,n}(t+h) = p_{k-1,n-1}(t)\lambda(k-1)h + p_{k,n-1}(t)\lambda h \sum_{i \neq *} k_i + p_{k,n}(t) \left(1 - \lambda h \sum_i k_i \right) + o(h). \quad (36)$$

When $N(t)=n$, we know that $\sum_{i=1}^n k_i = 2n-2$, since the system evolves from one in which $\sum_{i=1}^n k_i = 2$ when $n=2$, and the addition of each new node to the network increases the sum over coordination numbers of all nodes by 2. In the limit $h \rightarrow 0$, the recurrence relation (36) yields the differential-difference equation

$$\frac{d}{dt} p_{k,n} = \lambda(k-1)p_{k-1,n-1} + \lambda(2n-4-k)p_{k,n-1} - \lambda(2n-2)p_{k,n}. \quad (37)$$

We shall extract some results of interest and compare them with simulations. If a node currently has coordination number k , then its waiting-time density for the next birth, that is, the creation of its next link, is $\lambda k e^{-\lambda k t}$. In all other respects, the simulation of this model is identical to the Yule tree model of Sec. II.

Summing Eq. (37) over n (from 1 to ∞) yields a differential-difference equation for the marginal distribution $p_{k,\bullet}(t) = \sum_n p_{k,n}(t)$ of $K(t)$:

$$\frac{d}{dt} p_{k,\bullet} = \lambda(k-1)p_{k-1,\bullet} - \lambda k p_{k,\bullet}. \quad (38)$$

This is the equation of a Yule process with parameter λ [cf. Sec. II, where we found that for the simpler model of that section, it is $N(t)$ that evolves as a Yule process]. Thus $K(t)$ has a negative binomial distribution with parameters $k_0 = K(0)$ and $e^{-\lambda t}$. In particular, with $k_0 = 1$ this reduces to a geometric distribution with

$$p_{k,\bullet}(t) = e^{-\lambda t} (1 - e^{-\lambda t})^{k-1} \quad \text{for } k=1,2,\dots \quad (39)$$

Similarly by summing Eq. (37) from k from 1 to ∞ , we obtain an evolution equation for the marginal distribution $p_{\bullet,n}(t) = \sum_k p_{k,n}(t)$ of $N(t)$:

$$\frac{d}{dt} p_{\bullet,n} = 2\lambda(n-2)p_{\bullet,n-1} - 2\lambda(n-1)p_{\bullet,n}. \quad (40)$$

It is easily shown by generating functions that the solution of the system corresponding to the initial condition $N(0)=2$ is

$$p_{\bullet,n}(t) = e^{-2\lambda t} (1 - e^{-2\lambda t})^{n-2} \quad (41)$$

and consequently that for all $t > 0$,

$$\langle N(t) \rangle = e^{2\lambda t} + 1, \quad \text{var}\{N(t)\} = e^{4\lambda t} - e^{2\lambda t}, \quad (42)$$

while $\langle 1/N(t) \rangle \sim 2\lambda t e^{-2\lambda t}$ as $t \rightarrow \infty$.

Equation (40) is the evolution equation of a nonhomogeneous birth process with

$$\Pr\{\text{birth in } (t, t+h] | N(t)=n\} = 2\lambda(n-1). \quad (43)$$

From this it follows that the number of new nodes, $U(t) = N(t) - n_0$, connected in $(0, t]$ is a birth process with immigration and has a negative binomial distribution. Such a process is an ‘‘order statistic’’ process [24], which means that the times of births since the start of the process have the same joint distribution as those of the order statistics of a sample of independent, identically distributed random variables: in this case of random variables with a truncated exponential distribution, that is, with probability density function

$$f(\tau) = \frac{2\lambda e^{-2\lambda\tau}}{1 - e^{-2\lambda t}}, \quad 0 < \tau < t, \quad (44)$$

where t is the elapsed time since the founding of the network. By similar arguments to those in Sec. II, we find that the limiting distribution of the coordination number of an arbitrary node is

$$\begin{aligned} \lim_{t \rightarrow \infty} \Pr\{K(t) = k(\text{random node})\} &= \int_0^\infty 2\lambda e^{-2\lambda\tau} e^{-\lambda\tau} (1 - e^{-\lambda\tau})^{k-1} d\tau \\ &= \int_0^1 2x^2(1-x)^{k-1} dx \\ &= \frac{2\Gamma(3)\Gamma(k)}{\Gamma(k+3)} = \frac{4}{k(k+1)(k+2)} \sim \frac{4}{k^3}, \end{aligned} \quad (45)$$

an asymptotic power-law distribution, as found in the discrete-time analog of the Reed-Hughes tree by Szymański [12] and subsequent authors [35]. We compare limit (45) with a simulation in Fig. 6. Note that

$$\lim_{t \rightarrow \infty} \langle K(t) | \text{random node} \rangle = \sum_{k=1}^\infty \frac{4}{(k+1)(k+2)} = 2.$$

We have managed to produce these results without having to determine the joint distribution $p_{k,n}(t)$. In Appendix B 2 we obtain the exact solution for the generating function for $p_{k,n}(t)$, and we show that the correlation coefficient for the number of nodes present and the coordination number of a specified starting node does not decay to zero in the long-time limit. This is another way in which the variable birth-rate model differs from the constant birth-rate Yule tree model of Sec. II.

By generalizing the ideas of Sec. II, we can derive a mean-field prediction of the ring-number distribution. If we denote the number of nodes with ring number r and coordination number k at time t by $n(k, r, t)$, we have

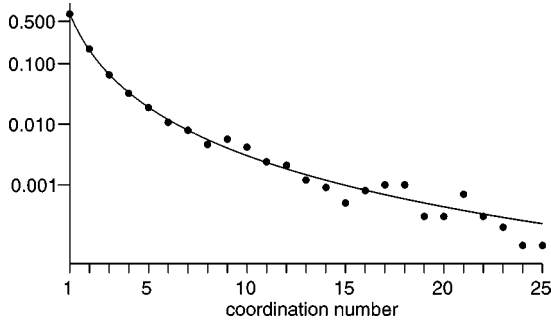


FIG. 6. For the Reed-Hughes tree model of Sec. III, where growth rates depend on coordination number, we compare in a log-linear plot the limiting coordination-number distribution $4[k(k+1)(k+2)]$ predicted by Eq. (45), shown as the curve, with the distribution obtained from one realization of a tree of 10 000 nodes ($\lambda t \approx 7.97$). In this realization all coordination numbers k with $1 \leq k \leq 27$ arose, but beyond there gaps appear. There were unique nodes with coordination numbers 220 and 209; the next largest coordination number was 76.

$$\begin{aligned} \frac{d}{dt}n(k,r,t) &\approx \lambda(k-1)n(k-1,r,t) - \lambda kn(k,r,t) \\ &+ \delta_{k,1} \sum_{k' \geq 1} \lambda k' n(k',r-1,t). \end{aligned} \quad (46)$$

We seek the solution, given that we start with one link and two nodes: $n(k,r,0) = (\delta_{r,0} + \delta_{r,1})\delta_{k,1}$. In Appendix C 2 we show that

$$\begin{aligned} \sum_{k=0}^{\infty} \sum_{r=0}^{\infty} n(k,r,t) \kappa^k \rho^r &= \int_0^t \frac{\lambda \rho (1 + \rho) \kappa e^{\lambda(1+\rho)\tau} d\tau}{\kappa + e^{\lambda(t-\tau)}(1-\kappa)} \\ &+ \frac{(1+\rho)\kappa}{\kappa + e^{\lambda t}(1-\kappa)}. \end{aligned} \quad (47)$$

Observe that in the mean-field theory the total number of nodes present at time t is given by

$$n_t = \sum_k \sum_r n(k,r,t) = e^{2\lambda t} + 1. \quad (48)$$

Recalling the exact result (42) that $\langle N(t) \rangle = e^{2\lambda t} + 1$, we see that the mean-field theory correctly produces one exact result for all $t \geq 0$: $n_t = \langle N(t) \rangle$.

Let us evaluate the mean-field prediction of the coordination-number distribution. Setting $\rho = 1$ in solution (47) for the generating function and then expanding in powers of κ , we find after a little algebra that

$$\begin{aligned} \sum_r n(k,r,t) &= 2e^{-\lambda t}(1 - e^{-\lambda t})^{k-1} \\ &+ e^{2\lambda t} \int_0^t 2\lambda e^{-3\lambda\tau}(1 - e^{-\lambda\tau})^{k-1} d\tau. \end{aligned} \quad (49)$$

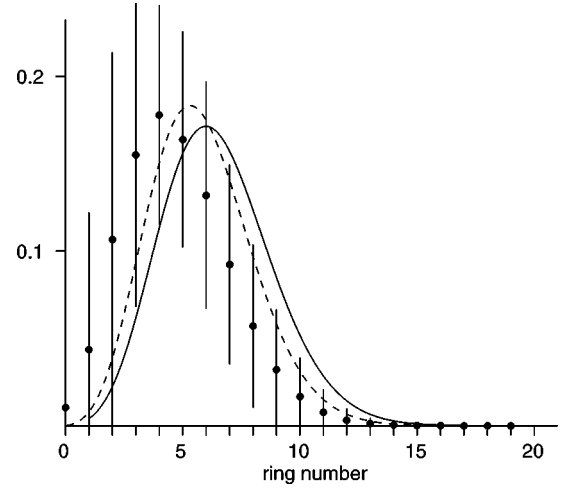


FIG. 7. For the Reed-Hughes tree model of Sec. III, we show the ring-number distribution obtained from 100 realizations at a fixed stopping time $\lambda t = 5.5$, corresponding to a mean network size $\approx 15\,400$ and standard deviation $\approx 29\,500$. The plotted data show the mean \pm one standard deviation. The continuous curve shows the mean-field prediction (51), while the broken curve corresponds to mean-field theory with the time adjusted so that the mean-field network size corresponds to the experimental mean network size.

Thus the mean-field coordination-number distribution $\sum_r n(k,r,t)/n_t$ converges at $t \rightarrow \infty$ to the same limit (45) as the earlier exact calculation.

Although we do not have a rigorous result for the distribution over ring numbers, we can easily extract a mean-field prediction. If we set $\kappa = 1$ in Eq. (47) and extract the coefficient of ρ^r we deduce that

$$\sum_k n(k,r,t) = e^{\lambda t} (\lambda t)^{r-1} / (r-1)!, \quad r \geq 1, \quad (50)$$

while $\sum_k n(k,0,t) = 1$. We now have mean-field predictions of the ring-number distribution for an arbitrarily chosen node:

$$\Pr\{\text{ring number} = r\} = \frac{e^{\lambda t} (\lambda t)^{r-1}}{(e^{2\lambda t} + 1)(r-1)!}, \quad r \geq 1; \quad (51)$$

$$\Pr\{\text{ring number} = 0\} = 1/(e^{2\lambda t} + 1). \quad (52)$$

The predicted mean ring number is $(\lambda t + 1)(1 + e^{-2\lambda t})^{-1}$. The mean-field prediction (51) is compared to simulations in Fig. 7. As in the Yule tree model of Sec. II, for a system inspected at a modest fixed time, the simulation results have the ring-number mode (most probable value) clearly less than the mean-field prediction (shown as a solid line). Figure 7 illustrates the effect of network size fluctuations in modest simulations. For the 100 realizations at fixed time $t = 5.5$ with $\lambda = 1$, the experimental mean network size is 15 416.8, but with standard deviation $\approx 29\,500$. The exact mean network size should be $e^{11} + 1 = 59\,875.1 \dots$. If instead we use the mean-field distribution with time t given by $t = \ln(n_t - 1)/(2\lambda)$, where n_t is taken as the experimental mean

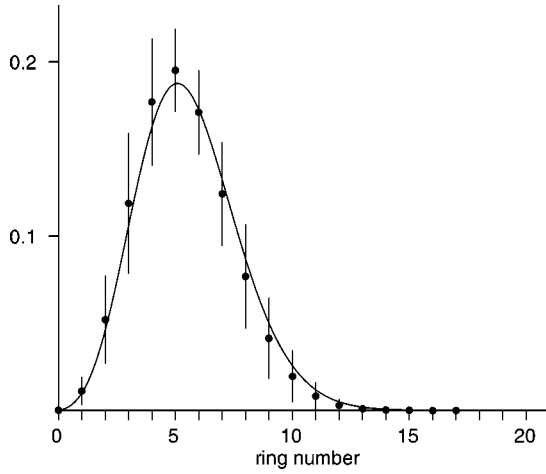


FIG. 8. For the Reed-Hughes tree model of Sec. III, we show the ring-number distribution obtained from 100 realizations at a fixed network size of 10 000 nodes, corresponding to a mean elapsed time $\lambda t \approx 6.13$ (standard deviation ≈ 1.21). The plotted data show the mean \pm one standard deviation. The continuous curve shows the mean-field prediction (53).

network size, we obtain the mean-field prediction shown as a broken line in Fig. 7, which is significantly better.

If we eliminate time in favor of network size n_t , the mean-field prediction becomes

$$\Pr\{\text{ring number} = r\} = \frac{(n_t - 1)^{1/2} [\ln(n_t - 1)]^{r-1}}{2^{r-1} (r-1)! n_t} \quad (53)$$

for $r \geq 1$, with $\Pr\{\text{ring number} = 0\} = n_t^{-1}$. Prediction (53) is compared to simulation data for a fixed network size in Fig. 8. We find here very good agreement between mean-field theory and simulations.

As in Sec. II, we may attempt to predict the asymptotic distribution of internode distances, using the observation that the statistics are dominated by node pairs for which the connecting path passes through one of the two nodes initially present, so that Eq. (24) applies. Using the mean-field ring-number distribution (50) we predict that for $l \geq 2$,

$$\Pr\{\text{path length} = l\} = \frac{e^{2\lambda t} (2\lambda t)^{l-2}}{(e^{2\lambda t} + 1)^2 (l-2)!} + \frac{2e^{\lambda t} (\lambda t)^{l-1}}{(e^{2\lambda t} + 1)^2 (l-1)!},$$

so that $\Pr\{\text{path length} = l\} \sim e^{-2\lambda t} (2\lambda t)^{l-2} / (l-2)!$, while

$$\Pr\{\text{path length} = 0\} = (e^{2\lambda t} + 1)^{-2} \sim e^{-4\lambda t},$$

$$\Pr\{\text{path length} = 1\} = 2e^{\lambda t} (e^{2\lambda t} + 1)^{-2} \sim 2e^{-3\lambda t}.$$

Eliminating time in favor of $n = e^{2\lambda t} + 1$, we infer a corresponding mean-field path length distribution for a given network size n :

$$\Pr\{\text{path length} = 0\} = n^{-2}, \quad (54)$$

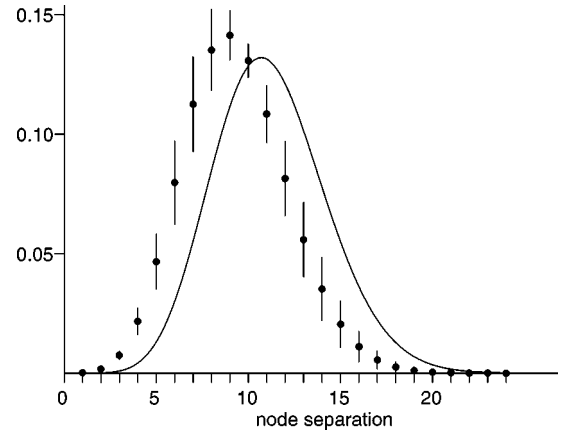


FIG. 9. For the Reed-Hughes tree model of Sec. III, we compare the internode distance distribution predicted by Eq. (56) with simulations for 100 realizations of fixed network size 10 000 (mean running time $\lambda t = 6.1282$, standard deviation 1.2097).

$$\Pr\{\text{path length} = 1\} = 2(n-1)^{1/2} n^{-2}, \quad (55)$$

and

$$\Pr\{\text{path length} = l\} = \frac{(n-1) [\ln(n-1)]^{l-2}}{n^2 (l-2)!} + \frac{2(n-1)^{1/2} \left[\frac{1}{2} \ln(n-1) \right]^{l-1}}{n^2 (l-1)!} \quad (l \geq 2). \quad (56)$$

The mean-field prediction of the path length distribution at constant time performs badly (as was the case with the analogous prediction for Yule trees in Sec. II). Prediction (56) is a little better, but still disappointing (Fig. 9).

As in Sec. II, the disappointing quality of the mean-field predictions of the distance distribution in comparison to simulations in Fig. 9 is due to the asymmetry of typical realizations of the tree when viewed from the primal node, and the argument to demonstrate this phenomenon is similar to that in Sec. II. One may attempt to derive an improved mean-field theory for the present model using the approach based on Eq. (28), which was successful for Yule trees in Sec. II. However, the natural interpolation from networks of size 2^m to general sizes seems very difficult, and we have not pursued this here.

As for the Yule tree model of Sec. II, we may study the probability ϕ_r that the smallest ring number encountered on the unique path joining two randomly chosen distinct nodes is r . Some simulation data is reported in Table II. The data are tolerably well approximated by

$$\phi_r^0 = \begin{cases} 1/2, & r = 0 \\ 3^{-r}, & r > 0, \end{cases} \quad (57)$$

for which we now offer a partial explanation. In some respects the designation of one of the two nodes initially present as the origin (node O, with ring number 0) and the

TABLE II. For the Reed-Hughes tree model of Sec. III, we show estimates of the probability ϕ_r that the smallest ring number encountered on the unique path joining two randomly chosen distinct nodes is r . The results (experimental mean $\bar{\phi}_r$ and standard deviation) are obtained from 100 simulations of a network of 10 000 nodes. The last column compares the experimental mean with the empirical relation (57).

| r | $\bar{\phi}_r$ | Standard deviation | $\bar{\phi}_r/\phi_r^0$ |
|-----|------------------------|-----------------------|-------------------------|
| 0 | 5.10×10^{-1} | 2.98×10^{-1} | 1.021 |
| 1 | 3.40×10^{-1} | 2.44×10^{-1} | 1.020 |
| 2 | 9.77×10^{-2} | 1.06×10^{-1} | 0.879 |
| 3 | 3.59×10^{-2} | 5.35×10^{-2} | 0.969 |
| 4 | 1.07×10^{-2} | 1.77×10^{-2} | 0.866 |
| 5 | 3.82×10^{-3} | 7.26×10^{-3} | 0.928 |
| 6 | 1.15×10^{-3} | 1.69×10^{-3} | 0.841 |
| 7 | 3.69×10^{-4} | 5.39×10^{-4} | 0.808 |
| 8 | 1.24×10^{-4} | 2.01×10^{-4} | 0.816 |
| 9 | 3.45×10^{-5} | 5.53×10^{-5} | 0.679 |
| 10 | 9.97×10^{-6} | 2.18×10^{-5} | 0.589 |
| 11 | 2.41×10^{-6} | 5.38×10^{-6} | 0.426 |
| 12 | 5.80×10^{-7} | 1.12×10^{-6} | 0.308 |
| 13 | 1.31×10^{-7} | 3.21×10^{-7} | 0.208 |
| 14 | 1.80×10^{-8} | 5.68×10^{-8} | 0.086 |
| 15 | 3.60×10^{-9} | 1.58×10^{-8} | 0.052 |
| 16 | 4.00×10^{-10} | 2.80×10^{-9} | 0.017 |

other as offspring (node A, having ring number 1) is a little artificial. Let us instead consider as our measure of remoteness the shortest distance (measured in units of link length) from whichever of nodes O or A is closest, and write $\hat{\phi}_r$ for the probability that for the unique path joining any two arbitrary nodes, the smallest remoteness is r . For paths that are closer to O than to A, this value of r is the same as the ring number, but for paths that are closer to A, it is one less than the ring number. If we assume that the two branches of the tree are statistically equivalent, which is appropriate as at time $t=0$ the initial states of nodes O and A are identical, we see that

$$\hat{\phi}_0 = \phi_0^0 + \frac{\phi_1^0}{2}; \quad \hat{\phi}_r = \frac{\phi_r^0 + \phi_{r+1}^0}{2} \quad (r \geq 1). \quad (58)$$

The empirical approximation (57) takes the simple form

$$\hat{\phi}_r = \frac{2}{3^{r+1}} \quad (r \geq 0). \quad (59)$$

It would be interesting to have a result of this type established rigorously.

From the mean-field distance distribution it is easy to estimate the Wiener index $W(n)$ and we predict that

$$W(n) \sim \frac{1}{2} n^2 \ln n, \quad (60)$$

cf. Fig. 10. The performance of the mean-field estimate for the Wiener index is much better for the Reed-Hughes tree

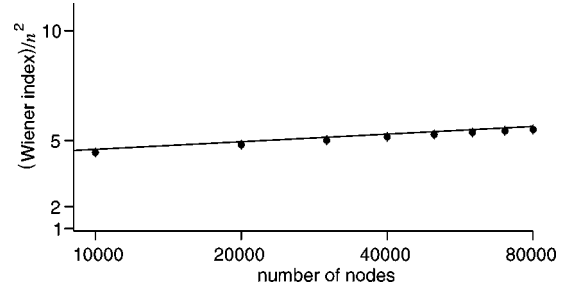


FIG. 10. For the Reed-Hughes tree model of Sec. III, we compare the Wiener index predicted by Eq. (60) with simulations (ten realizations with $\lambda=1$). We show the mean Wiener index $W(n) \pm$ one standard deviation, scaled against n^2 . The solid curve is the mean-field estimate $W(n)/n^2 \approx (1/2) \ln n$ which performs much better than the rigorous lower bound $W(n)/n^2 = (1 - n^{-2}) \sim 1$ and the rigorous upper bound $W(n)/n^2 = n(1 - n^{-2})/6 \sim n/6$.

model than for the Yule tree model of Sec. II, and better than one might expect, given the inadequacy of the mean-field distance distribution prediction manifest in Fig. 9.

If the growing network is killed at a random time T after the creation of the first two nodes, where T has probability density function $\Psi(t)$, then from Eq. (41) the network size N^\dagger at death has the probability distribution

$$\Pr\{N^\dagger = n\} = \int_0^\infty \Psi(t) e^{-2\lambda t} (1 - e^{-2\lambda t})^{n-2} dt$$

for $n=2,3,4, \dots$. In particular, for death at an exponentially distributed time, with $\Psi(t) = \nu e^{-\nu t}$, we find

$$\Pr\{N^\dagger = n\} = \frac{\nu \Gamma(\nu/(2\lambda) + 1) \Gamma(n-1)}{2\lambda \Gamma(n + \nu/(2\lambda))} \sim \frac{\text{const}}{n^{\nu/(2\lambda) + 1}}$$

as $n \rightarrow \infty$.

A natural generalization of the continuous-time model discussed above would replace the birth rate $\lambda K_i(t)$ at node i by a more general function of $K_i(t)$, and there has been work in this direction for discrete-time models. For example, Krapivsky *et al.* [25] consider growth nodes selected with probabilities proportional to $K_i(t)^\gamma$. When γ is neither 0 or 1, the sum rules that enable us to focus on the coordination-number distribution of a tagged node in the Yule tree and Reed-Hughes tree models, respectively, are lost, and only mean-field and simulation approaches seem profitable [36].

IV. CROSS-LINKED NETWORKS

The models of Secs. II (Yule trees) and III (Reed-Hughes trees) can be extended to allow for cross-linking in the network structure. It is simplest to explain the simulation algorithm first, as this raises a subtle issue.

A. A class of models for simulation

We associate with each node $*$ a waiting-time density $\Psi_*(t)$ for it next to create a new link to a node not yet part of the network. We consider two cases: *constant fertility*, $\Psi_*(t) = \lambda e^{-\lambda t}$ (cf. Sec. II) and *variable fertility*, $\Psi_*(t)$

$=k\lambda e^{-k\lambda t}$ (cf. Sec. III). In the variable fertility case, k is the number of nodes to which node $*$ has given birth. The value of k is updated each time a birth occurs and as the exponential distribution has no memory, we draw a new random time for the next birth from the same mother. Thus far, the constant and variable fertility cases are the same as the Yule tree model of Sec. II and the Reed-Hughes tree model of Sec. III, respectively.

To escape the strict tree structure, we also introduce a random waiting time for the creation of the next link to an existing node to which node $*$ has not been previously linked. This link can be described as ‘‘marriage.’’ If the number of eligible suitors (marriage partners) for node $*$ at time t is denoted by s , we associate a waiting-time density $\phi_*(t) = \mu s e^{-\mu s t}$ for node $*$ to initiate its next marriage. However, this waiting-time density only applies until either (a) a node other than $*$ gives birth, or (b) some node other than $*$ initiates a marriage to node $*$, or (c) a pair of nodes not including $*$ marry. As these events all change the eligible suitor set for some nodes, continual resetting of the waiting-time densities for marriages for all nodes must be accommodated. Although this may be seen as a practical nuisance in simulation, it enables us to switch off attempted marriages for node $*$ for appropriate time intervals when there are no suitors available, but resume seeking marriages when new eligible partners are created.

It remains to define precisely what is meant by an eligible suitor, and we distinguish between the following two quite different cases.

(1) *Polygamy*: an eligible suitor for node $*$ is any node not yet linked to it as parent or offspring or by marriage.

(2) *Monogamy*: an eligible suitor for node $*$ is any node not yet linked to it as parent or offspring that has not yet been married, so that in this case, no node ever contracts more than one marriage.

We shall only develop theory and report simulations for polygamous cases, in which extensive cross-linking can occur. The monogamous case, or a variant in which each node is allowed to acquire a limited number of marriage partners, produces a more sparsely linked network.

B. Transitions due to marriage

Let $N(t)$, $K_j(t)$, and $S_j(t)$, respectively, denote the number of nodes in the network, the coordination number of node j of the network, and the number of eligible suitors of node j at time t . Note that

$$S_j(t) = N(t) - 1 - K_j(t), \quad (61)$$

so that although we use $S_j(t)$ in the derivation of our evolution equations, it can be eliminated in the later stages of the analysis. For a specified node $*$ of the network, we shall for brevity replace $K_*(t)$ and $S_*(t)$ with $K(t)$ and $S(t)$, respectively. The set of suitor nodes eligible to marry node $*$ at time t will be denoted by $\sigma(*)$. In addition, let $M(t)$ denote the total number of links in the network, so that

$$\sum_j K_j(t) = K(t) + \sum_{j \neq *} K_j(t) = 2M(t). \quad (62)$$

We shall calculate the probability of marriages occurring in the time interval $(t, t+h)$. For brevity, all quantities that are $o(h)$ are suppressed in the analysis. Consider first a marriage involving node $*$. This marriage may be initiated by node $*$ with probability $\mu S(t)h$, or initiated by another node $j \in \sigma(*)$. In the latter case, node j initiates marriages with probability $\mu S_j(t)h$, but only one out of its $S_j(t)$ choices of partner selects node $*$, so its probability of initiating a marriage to node $*$ is μh . Thus

$$\begin{aligned} \Pr\{\text{node } * \text{ marries}\} &= \mu S(t)h + \sum_{j \in \sigma(*)} \mu h = 2\mu S(t)h \\ &= 2\mu [N(t) - 1 - K(t)]h. \end{aligned}$$

A marriage not involving node $*$ may be initiated by a node eligible to marry node $*$, or a node not eligible to marry node $*$, and thus

$$\begin{aligned} \Pr\{\text{marriage not involving } *\} &= \sum_{j \in \sigma(*)} \mu (S_j(t) - 1)h + \sum_{j \notin \sigma(*), j \neq *} \mu S_j(t)h \\ &= \sum_j \mu S_j(t)h - 2\mu S(t)h \\ &= \mu \{N(t)[N(t) - 1] - 2M(t) - 2[N(t) - 1 - K(t)]\}h, \end{aligned}$$

where we have used Eqs. (61) and (62). Considering the complement of the events just considered, we deduce that

$$\Pr\{\text{no marriages}\} = 1 - \mu \{N(t)[N(t) - 1] - 2M(t)\}h.$$

We briefly consider a system in which there are no births, with links only created via marriage. A state of the system in which $K(t) = k, M(t) = m, N(t) = n$, and $S(t) = s$ will be denoted by (k, m, n, s) for brevity, and recalling that $S(t)$ is determined by $N(t)$ and $K(t)$ we write

$$\Pr\{K(t) = k, M(t) = m, N(t) = n\} = p_{k,m,n}(t).$$

This probability is zero if any of k, m, n is negative, or if $k \geq n$. Ignoring all events with probability $o(h)$, the system can be found in state (k, m, n, s) at time $t+h$ in only the following three ways.

(i) Node $*$ marries between times t and $t+h$, corresponding to $(k-1, m-1, n, s+1) \rightarrow (k, m, n, s)$, with probability $2\mu [n-1-(k-1)]h p_{k-1, m-1, n}(t)$.

(ii) A marriage that does not include node $*$ occurs, so that $(k, m-1, n, s) \rightarrow (k, m, n, s)$ with probability $\mu [n(n-1) - 2(m-1) - 2(n-1-k)]h p_{k, m-1, n}(t)$.

(iii) No marriages occur, so that $(k, m, n, s) \rightarrow (k, m, n, s)$ with probability $\{1 - \mu [n(n-1) - 2m]h\} p_{k, m, n}(t)$. We readily deduce the evolution equation

$$\begin{aligned} \frac{d}{dt} p_{k,m,n} = & 2\mu(n-k)p_{k-1,m-1,n} \\ & + \mu[n(n-1) - 2(m-1) - 2(n-1-k)]p_{k,m-1,n} \\ & - \mu[n(n-1) - 2m]p_{k,m,n}. \end{aligned} \quad (63)$$

Note that n is constant in the process, as there are no births. If we sum over m we find that

$$\frac{d}{dt} p_{k,\bullet,n} = 2\mu(n-k)p_{k-1,\bullet,n} - 2\mu(n-1-k)p_{k,\bullet,n}$$

and hence the generating function

$$\mathcal{P}_n(\kappa, t) = \sum_{k=0}^{\infty} p_{k,\bullet,n}(t) \kappa^k$$

satisfies the partial differential equation

$$\frac{\partial \mathcal{P}_n}{\partial t} - 2\mu\kappa(1-\kappa) \frac{\partial \mathcal{P}_n}{\partial \kappa} = -2\mu(n-1)(1-\kappa)\mathcal{P}_n. \quad (64)$$

One may verify directly that the solution of this equation, given that $K(0) = k_0 \leq n-1$, is

$$\mathcal{P}_n(\kappa, t) = \kappa^{k_0} [e^{-2\mu t} + \kappa(1 - e^{-2\mu t})]^{n-1-k_0}, \quad (65)$$

from which it follows that for $k_0 \leq k \leq n-1$,

$$p_{k,\bullet,n}(t) = \binom{n-1-k_0}{k-k_0} e^{-2\mu t(n-1-k)} (1 - e^{-2\mu t})^{k-k_0}. \quad (66)$$

The model evolves sensibly and converges to the fully married or fully cross-linked state $K(t) = N(t) - 1$ in the sense that $\Pr\{K(t) = n-1\} \rightarrow 1$ as $t \rightarrow \infty$.

In Secs. IV C and IV D we add this marriage or cross-linking mechanism to the models with birth rates of Secs. II and III, respectively. We note that the special case of Eq. (66) for the initial condition $k_0 = 0$ is produced exactly by the following mean-field argument. Let the number of nodes with coordination number k be denoted by $n(k, t)$. For each such node, the number of eligible suitors is $n_0 - 1 - k$. (Later, when we allow the number of nodes in the network to change, n_0 will be replaced by the current number of nodes.) We take the evolution equation

$$\begin{aligned} \frac{\partial}{\partial t} n(k, t) = & 2\mu[n_0 - 1 - (k-1)]n(k-1, t) \\ & - 2\mu(n_0 - 1 - k)n(k, t), \end{aligned} \quad (67)$$

with initial condition $n(k, s, 0) = n_0 \delta_{k,0}$. The coefficient 2μ appears in the evolution equation because a node may marry either at its own initiative or at the initiative of an eligible suitor (cf. the earlier exact analysis). A standard generating function solution (see Appendix C 3) shows that

$$\frac{n(k, t)}{n_0} = \binom{n_0-1}{k} (1 - e^{-2\mu t})^k e^{-2\mu t(n_0-1-k)}.$$

This result establishes the appropriateness of the mean-field treatment of the marriage contribution in our subsequent approximate analysis of models in which both birth and marriage occur. Indeed, for this special model with marriage alone, the mean-field evolution equation (67) is equivalent to the exact evolution equation (64), with $p_{k,\bullet,n}(t)$ corresponding to $n(k, t)/n_0$.

C. Constant fertility polygamy

We retain the notation of Sec. IV B and seek an evolution equation for the distribution $p_{k,m,n}(t)$ of a system with constant birth rates (cf. the Yule trees in Sec. II) and the marriage mechanism of Sec. IV B. In the infinitesimal interval $(t, t+h]$ the following four transitions that add one link to the network can occur to produce the state (k, m, n) ; all other events involving multiple births and/or marriages have probability $o(h)$, and all $o(h)$ terms are suppressed in the analysis for brevity as usual.

(i) Node * gives birth:

$$(k-1, m-1, n-1, s) \rightarrow (k, m, n, s),$$

with probability $\lambda h p_{k-1, m-1, n-1}(t)$.

(ii) A node other than * gives birth:

$$(k, m-1, n-1, s-1) \rightarrow (k, m, n, s),$$

with probability $\lambda(n-2)h p_{k, m-1, n-1}(t)$.

(iii) Node * marries:

$$(k-1, m-1, n, s+1) \rightarrow (k, m, n, s),$$

with probability $2\mu(n-k)h p_{k-1, m-1, n}(t)$.

(iv) A marriage that does not involve * occurs:

$$(k, m-1, n, s) \rightarrow (k, m, n, s),$$

with probability

$$\mu[n(n-1) - 2(m-1) - 2(n-1-k)]h p_{k, m-1, n}(t).$$

The probability that no event of the types (i)–(iv) occurs is $\{1 - \lambda n h - \mu[n(n-1) - 2m]h\} p_{k, m, n}(t)$. We now readily deduce that

$$\begin{aligned} \frac{d}{dt} p_{k,m,n} = & \lambda p_{k-1, m-1, n-1} + \lambda(n-2)p_{k, m-1, n-1} \\ & + 2\mu(n-k)p_{k-1, m-1, n} \\ & + \mu[n(n-1) - 2(m-1) - 2(n-1-k)]p_{k, m-1, n} \\ & - \{\lambda n + \mu[n(n-1) - 2m]\} p_{k, m, n}. \end{aligned} \quad (68)$$

As the random variable $M(t)$ is not of primary importance, we sum out over m to obtain an evolution equation for the joint distribution $p_{k,\bullet,n}(t)$ of $K(t)$ and $N(t)$:

$$\begin{aligned} \frac{d}{dt} p_{k,\bullet,n} &= \lambda p_{k-1,\bullet,n-1} + \lambda(n-2)p_{k,\bullet,n-1} \\ &+ 2\mu(n-k)p_{k-1,\bullet,n} - \{2\mu(n-k-1) + \lambda n\} p_{k,\bullet,n}. \end{aligned} \quad (69)$$

Note that if we also sum out over k , we obtain an evolution equation for the marginal distribution $p_{\bullet,\bullet,n} = \Pr\{N(t) = n\}$, namely,

$$\frac{d}{dt} p_{\bullet,\bullet,n} = \lambda(n-1)p_{\bullet,\bullet,n-1} - \lambda n p_{\bullet,\bullet,n}, \quad (70)$$

so that $N(t)$ is a Yule process and the node ages therefore have the order statistic property as in Sec. II. However, if instead we attempt to sum out over n in Eq. (69) we do not obtain an evolution equation for the marginal distribution of $K(t)$ alone.

We introduce the generating function

$$\mathcal{P}(\kappa, \zeta, t) = \sum_{k=0}^{\infty} \sum_{n=1}^{\infty} p_{k,\bullet,n}(t) \kappa^k \zeta^n \quad (71)$$

and obtain the partial differential equation

$$\begin{aligned} \frac{\partial \mathcal{P}}{\partial t} + \{\lambda \zeta(1-\zeta) + 2\mu \zeta(1-\kappa)\} \frac{\partial \mathcal{P}}{\partial \zeta} - 2\mu \kappa(1-\kappa) \frac{\partial \mathcal{P}}{\partial \kappa} \\ = (\lambda \zeta - 2\mu)(\kappa - 1) \mathcal{P}. \end{aligned} \quad (72)$$

Although it is possible to solve this equation using similar methods to those employed in Appendix B for the tree models of Secs. II and III and for the model with marriage alone in Sec. IV B, the resulting solution is expressed in terms of a relatively intractable integral, which we refrain from recording here.

If we only wish to extract the means of $N(t)$ and $K(t)$ we can do so more directly by differentiating Eq. (72) with respect to ζ or κ and setting $\kappa = \zeta = 1$, from which we find that

$$\frac{\partial}{\partial t} \langle N(t) \rangle - \lambda \langle N(t) \rangle = 0, \quad (73)$$

$$\frac{\partial}{\partial t} \langle K(t) \rangle + 2\mu \langle K(t) \rangle = \lambda - 2\mu + 2\mu \langle N(t) \rangle. \quad (74)$$

With the initial conditions $N(0) = n_0$ and $K(0) = k_0$, we find that

$$\langle N(t) \rangle = n_0 e^{\lambda t}, \quad (75)$$

$$\begin{aligned} \langle K(t) \rangle &= k_0 e^{-2\mu t} + \frac{\lambda - 2\mu}{2\mu} (1 - e^{-2\mu t}) \\ &+ \frac{2\mu n_0}{\lambda + 2\mu} (e^{\lambda t} - e^{-2\mu t}). \end{aligned} \quad (76)$$

If we want the mean coordination number for an arbitrary node (other than the primal node), we have to average the latter equation appropriately to take account of the time the

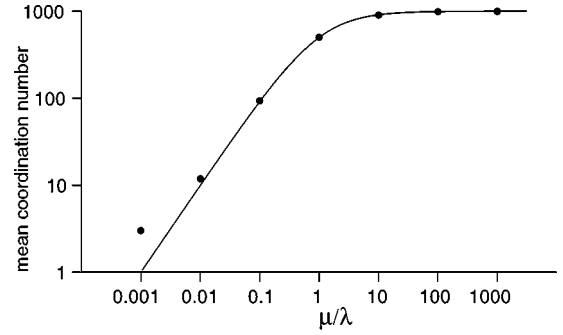


FIG. 11. The mean coordination number of the constant fertility polygamy case of Sec. IV C. We show results obtained from 20 realizations of a network of 1000 nodes for each μ/λ value. The curve is the approximation $1000 \mu/(\lambda + \mu)$ based on Eq. (79).

node has been in existence, which has probability density function $\lambda e^{-\lambda\tau}/(1 - e^{-\lambda t})$ if the chosen node was created at time $t - \tau$. Then the initial network size n_0 has to be taken as the random variable $N(t - \tau)$ for a process started at time 0 with one node initially present, and so has expectation $e^{\lambda(t-\tau)}$. We thus have for a random node other than the primal node

$$\begin{aligned} (1 - e^{-\lambda t}) \langle K(t) | \text{random nonprimal node} \rangle \\ = \int_0^t \lambda e^{-\lambda\tau} \left\{ \frac{\lambda}{2\mu} - 1 + \left(2 - \frac{\lambda}{2\mu} \right) e^{-2\mu\tau} \right\} d\tau \\ + \int_0^t \lambda e^{-\lambda\tau} e^{\lambda(t-\tau)} \frac{2\mu}{\lambda + 2\mu} (e^{\lambda\tau} - e^{-2\mu\tau}) d\tau \\ = \left(\frac{\lambda}{2\mu} - 1 \right) (1 - e^{-\lambda t}) + \frac{\lambda(4\mu - \lambda)}{2\mu(\lambda + 2\mu)} (1 - e^{-(\lambda + 2\mu)t}) \\ + \frac{2\mu e^{\lambda t}}{\lambda + 2\mu} \left\{ 1 - e^{-\lambda t} - \frac{\lambda}{2(\lambda + \mu)} (1 - e^{-2(\lambda + \mu)t}) \right\} \end{aligned} \quad (77)$$

and so as $t \rightarrow \infty$,

$$\langle K(t) | \text{random nonprimal node} \rangle = \frac{\mu e^{\lambda t}}{\lambda + \mu} + O(1). \quad (78)$$

That is, for a process started at time 0 from a single primal node,

$$\langle K(t) | \text{random nonprimal node} \rangle \sim \frac{\mu}{\lambda + \mu} \langle N(t) \rangle. \quad (79)$$

In Fig. 11 we show estimates of the mean coordination number as a function of μ/λ (20 realizations of a network of 1000 nodes for each μ/λ value shown), together with the approximation $1000\mu/(\lambda + \mu)$ based on Eq. (79). The fit is excellent for $\mu/\lambda > 0.01$. For a given network size, the difference between the mean-field prediction and simulations is to be expected for sufficiently small μ/λ , since we know from Sec. II that for $\mu = 0$, the mean coordination number of a randomly

TABLE III. We show the results of ten simulations of mean coordination number scaled against network size for constant birth-rate polygamy at three fixed times ($t=4$, $t=6$, and $t=8$) for $\lambda=1$ and three values of μ ($\mu=0.1$, 1, or 10), together with the corresponding prediction $\mu/(\lambda+\mu)$ for this ratio.

| | Time | Mean | Standard deviation | Predicted | |
|-------------------|-------|--------|--------------------|-----------|---------|
| $\mu/\lambda=0.1$ | $t=4$ | 0.1895 | 0.1037 | 0.0909 | (1/11) |
| | $t=6$ | 0.1094 | 0.0276 | | |
| | $t=8$ | 0.0931 | 0.0038 | | |
| $\mu/\lambda=1$ | $t=4$ | 0.5113 | 0.0736 | 0.5000 | (1/2) |
| | $t=6$ | 0.5107 | 0.0353 | | |
| | $t=8$ | 0.5039 | 0.0289 | | |
| $\mu/\lambda=10$ | $t=4$ | 0.7803 | 0.1570 | 0.9091 | (10/11) |
| | $t=6$ | 0.8851 | 0.0626 | | |
| | $t=8$ | 0.9056 | 0.0076 | | |

chosen node converges to 2 as $n \rightarrow \infty$. Some simulations at constant time are reported in Table III.

Consider now the mean-field approximation for this model. We denote by $n(k,t)$ the number of nodes with coordination number k at time t and we write $n(t) = \sum_k n(k,t)$ for the total number of nodes in the network at time t . The evolution equation for $n(k,t)$ is obtained by combining the growth terms of the mean-field analysis of Sec. II, with the terms from the mean-field treatment of network evolution due to marriage in Sec. IV B:

$$\begin{aligned} \frac{\partial}{\partial t} n(k,t) = & \lambda n(k-1,t) - \lambda n(k,t) + \lambda \delta_{k,1} n(t) \\ & + 2\mu[n(t) - 1 - (k-1)]n(k-1,t) \\ & - 2\mu[n(t) - 1 - k]n(k,t). \end{aligned} \quad (80)$$

In the third term we have noted that each birth creates a node able to marry all but one of the nodes currently present. The equation is slightly in error, since we treat $n(t)$ as an integer variable in setting up the evolution equation, but subsequently treat $n(t)$ and $n(k,t)$ as continuously varying functions of time. We shall consider here only the initial condition $n(k,0) = \delta_{k,0}$ appropriate to a network grown from a single initial node. If we sum Eq. (80) over k we obtain $n'(t) = \lambda n(t)$, leading to the appropriate mean-field prediction that $n(t) = e^{\lambda t}$, which we now use. The solution of the mean-field Eq. (80) using generating functions is discussed in Appendix C 4. We show there in particular that as $t \rightarrow \infty$, the mean-field model predicts that the mean coordination number is

$$\frac{1}{n(t)} \sum_k k n(k,t) \sim \frac{\mu n(t)}{\lambda + \mu}, \quad (81)$$

which agrees with the rigorous asymptotic value for $\langle K | \text{random node} \rangle$ given by Eq. (79).

A sample of simulation data for networks inspected at fixed times is given in Table IV. These data illustrate the

TABLE IV. Mean coordination number and internode distance for constant birth-rate polygamy at three fixed times ($t=4$, $t=6$, and $t=8$) for $\lambda=1$ and three values of μ ($\mu=0.1$, 1, or 10).

| Case $\lambda=1, \mu=0.1$ | | Mean | Standard deviation |
|---|-------|----------|--------------------|
| Number of nodes generated | $t=4$ | 25.4 | 13.7 |
| | $t=6$ | 181.2 | 104.4 |
| | $t=8$ | 1346.2 | 806.8 |
| $\langle K(t) \text{random node} \rangle$ | $t=4$ | 3.641 | 1.514 |
| | $t=6$ | 18.065 | 9.900 |
| | $t=8$ | 123.043 | 72.112 |
| Mean internode distance | $t=4$ | 2.376 | 0.527 |
| | $t=6$ | 2.273 | 0.223 |
| | $t=8$ | 2.030 | 0.095 |
| Case $\lambda=1, \mu=1$ | | Mean | Standard deviation |
| Number of nodes generated | $t=4$ | 26.7 | 18.4 |
| | $t=6$ | 192.2 | 137.6 |
| | $t=8$ | 1398.0 | 1041.1 |
| $\langle K(t) \text{random node} \rangle$ | $t=4$ | 13.010 | 8.185 |
| | $t=6$ | 94.446 | 65.942 |
| | $t=8$ | 700.192 | 522.527 |
| Mean internode distance | $t=4$ | 1.446 | 0.133 |
| | $t=6$ | 1.494 | 0.037 |
| | $t=8$ | 1.499 | 0.026 |
| Case $\lambda=1, \mu=10$ | | Mean | Standard deviation |
| Number of nodes generated | $t=4$ | 28.6 | 19.1 |
| | $t=6$ | 227.1 | 163.0 |
| | $t=8$ | 1680.0 | 1215.6 |
| $\langle K(t) \text{random node} \rangle$ | $t=4$ | 24.689 | 17.927 |
| | $t=6$ | 205.851 | 148.890 |
| | $t=8$ | 1524.762 | 1103.106 |
| Mean internode distance | $t=4$ | 1.160 | 0.102 |
| | $t=6$ | 1.104 | 0.061 |
| | $t=8$ | 1.092 | 0.007 |

significant fluctuations in network size and coordination number, but show that at fixed sufficiently large times the fluctuations in internode distance are small. For fixed size data, see Table V. We have not been able to find a simple prediction of the mean internode distance.

D. Variable fertility polygamy

We continue the notation of Sec. IV B and seek an evolution equation for the distribution $p_{k,m,n}(t)$ of a system with variable birth rates as in Sec. III and the marriage mechanism of Sec. IV B. In the infinitesimal interval $(t, t+h]$ the following four transitions that add one link to the network can occur to produce the state (k,m,n) ; all other events involv-

TABLE V. Mean coordination number and internode distance for constant birth-rate polygamy (20 realizations of a network of 1000 nodes).

| μ/λ | Coordination number | | Internode path length | |
|---------------|---------------------|--------------------|-----------------------|--------------------|
| | Mean | Standard deviation | Mean | Standard deviation |
| 0.001 | 3.012 | 0.047 | 6.0522 | 0.0891 |
| 0.01 | 11.84 | 0.41 | 3.1541 | 0.0362 |
| 0.1 | 93.18 | 2.49 | 2.0072 | 0.0112 |
| 1 | 499.4 | 12.9 | 1.5033 | 0.0133 |
| 10 | 904.7 | 5.4 | 1.0947 | 0.0054 |
| 100 | 987.3 | 1.9 | 1.0118 | 0.0019 |
| 1000 | 996.2 | 0.9 | 1.0028 | 0.0009 |

ing multiple births and/or marriages, or have probability $o(h)$ and all $o(h)$ terms are suppressed in the analysis for brevity as usual.

(i) Node * gives birth:

$$(k-1, m-1, n-1, s) \rightarrow (k, m, n, s),$$

with probability $\lambda(k-1)hp_{k-1, m-1, n-1}$.

(ii) A node other than * gives birth:

$$(k, m-1, n-1, s-1) \rightarrow (k, m, n, s),$$

with probability

$$\sum_{j \neq *} \lambda k_j h p_{k, m-1, n-1} = \lambda [2(m-1) - k] h p_{k, m-1, n-1}.$$

(iii) Node * marries, as in Secs. IV B and IV C.

(iv) A marriage that does not involve * occurs as in Secs. IV B and IV C.

The probability that no event of types (i)–(iv) occurs is $\{1 - 2\lambda mh - \mu[n(n-1) - 2m]h\}p_{k, m, n}(t)$.

We again adopt the convention that $p_{k, m, n} = 0$ if $k \leq 0$, or $m < 0$, $n < 0$, or $k \geq n$. Taking account of all the events discussed above, we obtain the evolution equation

$$\begin{aligned} \frac{d}{dt} p_{k, m, n} = & \lambda(k-1)p_{k-1, m-1, n-1} + \lambda[2(m-1) - k] \\ & \times p_{k, m-1, n-1} + 2\mu(n-k)p_{k-1, m-1, n} \\ & + \mu[n(n-1) - 2(m-1) - 2(n-1-k)] \\ & \times p_{k, m-1, n} - \{2\lambda m + \mu[n(n-1) - 2m]\} p_{k, m, n}. \end{aligned} \quad (82)$$

If we introduce the generating function

$$\mathcal{P}(\kappa, \xi, \zeta, t) = \sum_k \sum_m \sum_n p_{k, m, n}(t) \kappa^k \xi^m \zeta^n, \quad (83)$$

we find that

$$\begin{aligned} \frac{\partial \mathcal{P}}{\partial t} + \mu \zeta^2 (1 - \xi) \frac{\partial^2 \mathcal{P}}{\partial \zeta^2} + \kappa (1 - \kappa) \xi (\lambda \zeta - 2\mu) \frac{\partial \mathcal{P}}{\partial \kappa} \\ + [2\lambda \xi (1 - \xi \zeta) - 2\mu \xi (1 - \xi)] \frac{\partial \mathcal{P}}{\partial \xi} + 2\mu \xi \zeta (1 - \kappa) \frac{\partial \mathcal{P}}{\partial \zeta} \\ = 2\mu \xi (1 - \kappa) \mathcal{P}. \end{aligned} \quad (84)$$

The second-order ζ derivative in this equation makes the analysis somewhat more difficult than that encountered in the previous models. It is possible to obtain a slightly simpler equation for $\mathcal{Q}(\xi, \zeta, t) = \mathcal{P}(1, \xi, \zeta, t)$, but even this equation does not appear to permit any significant analytical progress towards the calculation of probability distributions of interest, so we confine our attention to an analysis of moments.

Before attempting an exact analysis of moments, we turn to mean-field arguments for guidance, knowing these arguments have given correct results for the models studied earlier in this paper. The mean-field evolution equation for $n(k, t)$, the number of nodes with coordination number k , is

$$\begin{aligned} \frac{\partial}{\partial t} n(k, t) = & \lambda(k-1)n(k-1, t) - k\lambda n(k, t) \\ & + \lambda \delta_{k,1} \sum_{k'} k' n(k', t) + 2\mu[n(t) - 1 - (k-1)] \\ & \times n(k-1, t) - 2\mu[n(t) - 1 - k]n(k, t). \end{aligned} \quad (85)$$

This is a direct combination of the mean-field treatments of the model of Sec. III and of marriage in Sec. IV B. The initial condition is $n(k, 0) = 2\delta_{k,1}$. We show in Appendix C 5 that Eq. (85) implies that the number of nodes $n(t)$ and the mean coordination number $c(t)$ are related by

$$c(t) = \frac{n'(t)}{\lambda n(t)}, \quad (86)$$

while

$$nc \frac{dc}{dn} - 2 \left(1 - \frac{\mu}{\lambda}\right) c + c^2 = \frac{2\mu(n-1)}{\lambda}. \quad (87)$$

We have not found a closed-form solution of this differential equation except for $\mu/\lambda = 1$ (see below), but it is easily verified that the asymptotic solution for large n is

$$c = \left(\frac{4\mu n}{3\lambda}\right)^{1/2} + \frac{4}{5} \left(1 - \frac{\mu}{\lambda}\right) + O(n^{-1/2}). \quad (88)$$

In Fig. 12 we show estimates of the mean coordination number for a network of 1000 nodes as a function of μ/λ . Simulations (20 realizations for each μ/λ value) are shown as disks. The straight line corresponds to the first term in the asymptotic expansion (88). We note that for a fixed large n , expansion (88) must fail in the limits $\mu/\lambda \rightarrow 0$ and $\mu/\lambda \rightarrow \infty$, since we know from our earlier analyzes that $c(t) \rightarrow 2$ and $c(t) = n(t) - 1 + \dots$ in the cases $\mu = 0$ and $\lambda = 0$, respectively. However, we find that a numerical solution [37] of the differential Eq. (87) (with the appropriate initial condition

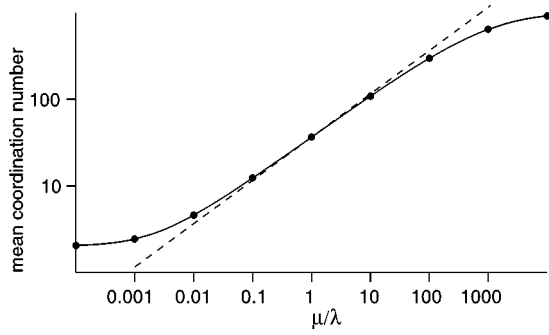


FIG. 12. The mean coordination number of the variable fertility polygamy case of Sec. IV D. Results obtained from 20 realizations of a network of $n(t) = 1000$ nodes for each μ/λ value are shown as disks. The dashed line corresponds to the one-term mean-field approximation $\sqrt{4\mu n(t)/(3\lambda)}$, and the continuous curve to the numerical solutions of the differential equation (87) for the mean-field treatment. At the resolution of the figure, the latter curve is indistinguishable from the simulation results.

$c = 1$ for $n = 2$) gives mean-field predictions (shown by the continuous curve in Fig. 12) that closely coincide with the simulation results for nine decades of μ/λ . The quality of the mean-field approximation can also be seen from the data given in Table VI.

In the special case $\mu/\lambda = 1$, the differential equation (87) is easily shown to have the exact solution

$$c^2 = \frac{4n}{3} - 2 + \frac{4}{3n^2} \quad (89)$$

and so

$$\frac{n'(t)}{\lambda n(t)} = \left(\frac{4}{3}\right)^{1/2} \left(n - \frac{3}{2} + \frac{1}{n^2}\right)^{1/2}, \quad (90)$$

that is

$$\lambda \frac{dt}{dn} = \frac{\sqrt{3}}{2\lambda} \left(n^3 - \frac{3}{2}n^2 + 1\right)^{-1/2}. \quad (91)$$

TABLE VI. Mean coordination number and mean internode distances for variable birth-rate polygamy for a network of 1000 nodes. Simulation results for $0.0001 \leq \mu/\lambda \leq 10000$ are averages over 20 realizations. We also give mean-field predictions. The asymptotic expansion (88), taken to one or two terms, is only useful for $0.01 \leq \mu/\lambda \leq 100$, and indeed the two-term expansion becomes negative for large μ/λ . However, the predictions based on numerical solution of Eq. (87) agree well with simulations over the entire range and show the correct transition to known long-time limits (2 for $\mu/\lambda \rightarrow 0$ and 999 for $\mu/\lambda \rightarrow \infty$).

| | μ/λ | 0.0001 | 0.001 | 0.01 | 0.1 | 1 | 10 | 100 | 1000 | 10000 |
|---|--------------------------------|--------|-------|-------|-------|-------|-------|-------|--------|--------|
| Mean-field approximation for mean coordination number | One term from expansion (88) | 0.365 | 1.155 | 3.651 | 11.55 | 36.51 | 115.5 | 365.1 | 1154.7 | 3651.5 |
| | Two terms from expansion (88) | 1.165 | 1.954 | 4.443 | 12.27 | 36.51 | 108.3 | 285.9 | 355.5 | (<0) |
| | Numerical solution of Eq. (87) | 2.047 | 2.432 | 4.589 | 12.29 | 36.49 | 108.5 | 297.0 | 640.7 | 917.9 |
| Computer simulation for coordination number | Mean | 2.046 | 2.429 | 4.595 | 12.36 | 36.64 | 108.5 | 298.0 | 642.2 | 913.0 |
| | Standard deviation | 0.010 | 0.019 | 0.088 | 0.19 | 0.84 | 2.2 | 7.1 | 9.8 | 7.2 |
| Computer simulation for internode distance | Mean | 6.642 | 5.710 | 4.256 | 3.086 | 2.388 | 1.986 | 1.714 | 1.359 | 1.086 |
| | Standard deviation | 0.402 | 0.187 | 0.031 | 0.014 | 0.018 | 0.011 | 0.009 | 0.010 | 0.007 |

Integrating this from $n = 2$ to $n = \infty$, we deduce that in mean-field theory for $\lambda = \mu$, the network size becomes infinite at the finite time

$$t_c = \frac{\sqrt{3}}{2\lambda} \int_2^\infty \frac{dn}{\sqrt{n^3 - (3/2)n^2 + 1}} \approx \frac{1.44}{\lambda}. \quad (92)$$

Moreover, we have for large $n(t)$,

$$t_c - t \sim \frac{\sqrt{3}}{2\lambda} \int_{n(t)}^\infty \frac{dz}{z^{3/2}} = \frac{\sqrt{3}}{\lambda \sqrt{n(t)}}. \quad (93)$$

Thus

$$n(t) \sim \frac{3}{[\lambda(t_c - t)]^2} \quad \text{as } t \rightarrow t_c. \quad (94)$$

One may argue from an asymptotic analysis of the differential Eq. (87) for general λ/μ that in the mean-field treatment divergence of the network size occurs at some finite time $t_c(\lambda, \mu)$ for $0 < \mu/\lambda < \infty$, but we have not pursued accurate estimates of $t_c(\lambda, \mu)$ other than the special value $t_c(\lambda, \lambda)$ found above.

Guided by the mean-field analysis, we see what we can establish by exact arguments. Rigorous relations involving averages of $K(t)$, $M(t)$, and $N(t)$ can be found by differentiating Eq. (84) with respect to κ , ξ , or ζ and setting $\kappa = \xi = \zeta = 1$, and we find that

$$\frac{d}{dt} \langle K(t) \rangle - (\lambda - 2\mu) \langle K(t) \rangle = 2\mu [\langle N(t) \rangle - 1], \quad (95)$$

$$\frac{d}{dt} \langle M(t) \rangle - 2(\lambda - \mu) \langle M(t) \rangle = \mu \langle N(t) [N(t) - 1] \rangle, \quad (96)$$

$$\frac{d}{dt} \langle N(t) \rangle = 2\lambda \langle M(t) \rangle. \quad (97)$$

As these equations do not decouple, we cannot extract exact formulas for the moments. However, if we eliminate $\langle M(t) \rangle$, we deduce that

$$\begin{aligned} & \frac{d^2}{dt^2} \langle N(t) \rangle - 2(\lambda - \mu) \frac{d}{dt} \langle N(t) \rangle + 2\lambda\mu \langle N(t) \rangle \\ &= 2\lambda\mu \langle N(t)^2 \rangle \\ &= 2\lambda\mu [\langle N(t) \rangle^2 + \text{var}\{N(t)\}] \geq 2\lambda\mu \langle N(t) \rangle^2. \end{aligned}$$

If we assume that there exist constants t_c , G , and γ with $\langle N(t) \rangle \sim G(t_c - t)^{-\gamma}$ as $t \rightarrow t_c$, we see from inequality (98) that for t sufficiently close to t_c we have

$$G\gamma(\gamma + 1)(t_c - t)^{-\gamma-2} \geq 2\lambda\mu G^2(t_c - t)^{-2\gamma},$$

so that $\gamma \leq 2$ and if $\gamma = 2$, then $G \leq 3/(\lambda\mu)$. Further, if $\gamma > 1$, it follows from Eq. (95) that for any fixed node,

$$\langle K(t) \rangle \sim 2\mu G(\gamma - 1)^{-1} (t_c - t)^{1-\gamma} \quad (98)$$

and thus as $\langle N(t) \rangle \rightarrow \infty$ we have

$$\langle K(t) \rangle \sim \frac{2\mu G^{1/\gamma}}{\gamma - 1} \langle N(t) \rangle^{1-1/\gamma}. \quad (99)$$

The mean-field calculation for $\lambda = \mu$ suggests that $\gamma = 2$. If one makes this stronger assumption, it follows that $\langle K(t) \rangle \sim H \langle N(t) \rangle^{1/2}$, with the constant $H \leq 2\sqrt{3\mu/\lambda}$, for any fixed node.

V. DISCUSSION

In this paper we have considered four models for the stochastic growth of networks. All four models are based on well-defined time-evolving random processes, and we have been able to study properties of the resulting networks both as functions of time and as functions of network size, using exact analysis, mean-field approximations and simulations.

The first two models produce trees, that is, random networks without cross-links. The ‘‘Yule tree’’ model of Sec. II assigns constant birth rates to nodes, while the ‘‘Reed-Hughes tree’’ model of Sec. III has coordination-number dependent birth rates. In both cases, the mean coordination number in the network converges to 2 in the long-time limit, but the coordination-number distributions differ greatly. In the constant birth-rate Yule tree model, the coordination-number distribution converges to 2^{-k} , $k = 1, 2, 3, \dots$ [Eq. (14)]. For the variable birth-rate Reed-Hughes tree model, the coordination-number distribution converges to the heavy-tailed distribution $4/[k(k+1)(k+2)]$, $k = 1, 2, 3, \dots$ [Eq. (45)]. These limiting distributions have been derived rigorously here (they have been previously proven [8,12] and rediscovered [25] for the discrete analog) and are also identical with the limiting distributions in a mean-field treatment. The striking difference between the two models reflects the fact that when a random node is examined, its coordination-number distribution depends on how long the node has been present in the network. The interplay of exponential growth of a system and random lifetime of elements in the system being able to produce heavy tails is in line with general observations of two of the authors [19] that find applications in a wide variety of contexts [15–18], and embody a perspective that goes back at least to Fermi in 1949 [20], but may

have a much longer history, and should perhaps be better known.

For the tree networks we have made some progress in studying the metrical structure of the network. The exact results we give for Yule trees build on known results for random recursive trees, a discrete-time model that underlies Yule trees. The apparent absence of analogous discrete-time results applicable to Reed-Hughes trees makes analysis of this problem somewhat harder.

For both tree models we have discussed, mean-field treatments give predictions of the distribution of the ring number (the distance of nodes from the initial node) that agree well with simulations for a given large network size. Realization-to-realization temporal fluctuations make the mean-field predictions much less accurate for predicting the distribution for a given time. For Yule trees, building on results for random recursive trees, we have derived the exact distribution for the ring number at fixed time.

In contrast to these successes of mean-field theory, the simple mean-field approach gives disappointing predictions of the internode distance distribution, which can be understood by noting (as we have shown) that the treelike structures have a strong statistical anisotropy. For Yule trees, an improved mean-field theory that addresses this anisotropy enables us to find a much better approximation to the internode distance distribution

One quantitative measure of the metrical structure of the tree is the Wiener index, a sum over all internode distances. For Yule trees the mean Wiener index for fixed network size is known from earlier work on random recursive trees. We have deduced from this the mean Wiener index at fixed time. For both tree models, our mean-field analysis suggests that the Wiener index grows asymptotically as a multiple of $\langle n \rangle^2 \ln \langle n \rangle$, where n is the network size. This prediction agrees to leading order with the exact result for the mean Wiener index for Yule trees. One may conjecture that with high probability in any given realization ω of either process, the Wiener index $W_i(\omega)$ will have the asymptotic behavior $W_i(\omega) \sim C(\omega) N(t)^2 \ln N(t)$ as $t \rightarrow \infty$, where $C(\omega)$ has a well-defined distribution.

Concerning applications of our treelike models, the Yule tree (constant birth rate) model of Sec. II can be viewed as the classical linear birth process, and therefore inherits all applications of that process, although the questions we have asked differ from those normally asked in many of the standard applications. The Yule tree does not exhibit scale-free [3] behavior. In contrast, the Reed-Hughes tree (variable birth rate) model of Sec. III does produce scale-free networks.

Both the Yule tree and the Reed-Hughes tree can be thought of as models for the spread of infectious diseases: the former for diseases such as influenza or SARS, and the latter for sexually transmitted diseases in which more promiscuous individuals (those with many previous contacts) are more likely to spread the disease to new uninfected partners. The Reed-Hughes tree may also be a reasonable model for food webs and networks of interacting proteins. For the

Reed-Hughes model to apply to a system, the exponents in the empirical distributions of connectivities should be close to 3. In fact estimates cited by Albert and Barabási [3] are 3.4 for a network of sexual partners, 2.4 for the protein network of the yeast *Saccharomyces cerevisiae*, and 1.05 and 1.13 for two food webs. The first two are not too far from the value of 3 derived in the model. Those for the food webs however are very different, but it should be borne in mind that the food webs are rather small, the largest having 186 nodes, so there is probably considerable sampling error in the estimates.

In Sec. IV we have addressed models in which links between nodes can be established not only by birth (which brings new nodes into the network) but also by a cross-linking process that we describe as marriage, which introduces a new characteristic cross-linking rate parameter μ . In the case $\mu=0$ the models of Sec. IV reduce to the tree-growing models of previous sections. Models with cross-linking are of more potential interest in communication network modeling, since existing tortuous connection paths may be supplanted as the network evolves by shorter paths.

We have studied the degenerate case in which the birth component of network evolution is disabled in Sec. IV B, and shown that a mean-field treatment produces coordination-number predictions that agree exactly with rigorous calculations. When constant birth rates and marriage are combined in Sec. IV C, we find that the problem remains sufficiently tractable such that we can prove rigorously that the mean coordination number of the network is asymptotically proportional to the mean network size in the long-time limit (this is also predicted by a mean-field analysis). There is scope for further work on the properties of this model, including the difficult problem of determining the distribution of internode distance.

The last specific model we have studied (Sec. IV D) combines coordination number dependent birth rates with marriage. The typical example of such a network is the World Wide Web. When new nodes are added to the web, they are more likely to have links to nodes, such as Google, Adobe, etc., that are already well connected than to weakly connected nodes, and nodes already present in the network are more likely to be found by and linked to already well-connected nodes. For the model of Sec. IV D few exact results seem to be available, but mean-field arguments and simulations predict that the mean coordination number scales as the square root of current network size (in excellent agreement with simulations), and also predict that the mean number of nodes present in the system diverges as $(t_c - t)^{-2}$ at a finite time t_c . Some rigorous bounds on exponents that may characterize divergence of the mean number of nodes and the mean coordination number of a fixed node are derived, but there are opportunities for further analysis. In particular, networks with directed links merit examination, and the techniques of the present paper may be useful in that context.

ACKNOWLEDGMENTS

This research was supported in part by the Particulate Fluids Processing Center (an ARC Special Research Center) at the University of Melbourne and by NSERC Discovery Grant 7252.

APPENDIX A: SIMULATION DETAILS

Simulation code was written in C. For brevity below, (fp) and (int) denote floating point and integer, respectively. For simulating *trees* (Secs. II and III), two lists are maintained. The first is a list of all *birth events*, both past and future. Each birth event is a C structure consisting of birth time (fp), identity number of mother (int), and ring number of mother (int). This list is sorted in order of increasing birth times, and the data structure used for the list is a binary search tree. Since we will often be inserting new items into the middle of a sorted list, the binary tree data structure should be much more efficient than either arrays or linked lists.

The second list, namely, the *node property list* is a list of nodes and certain information associated with each node. Each node entry is a C structure consisting of birth time (fp), ring number (int), identity number of mother (int), number of daughters (int), list of daughters (implemented as an array) consisting of daughter's identity number (int) and daughter's birthtime (fp), and *dlimit* (int)—a number used to check whether we have filled up the daughters array, to then allocate more memory. The node property list is also implemented as an array, and the subscript of the array is used as the identity number of the node. Identity numbers are ordered such that higher identity numbers are created at later times.

For trees, the calculation of the distance between any pair of nodes is straightforward, since there is a unique path joining the nodes. We first move back through the tree towards the primal node from the node with higher ring number until we have the same ring number as the other node. We then move from both nodes towards the primal node until we encounter a common node.

For simulating *networks with internal links*, three lists are maintained. The first is a list of all birth events, both past and future. The second is a list of *all birth and marriage events that have occurred*, but no future events. Each event is a C structure consisting of time of event (fp), type of event (b or m), *id1* (int), and *id2* (int). For birth events, *id1* gives birth to *id2*. For marriage events, *id1* chooses to link to *id2*. Both lists are sorted in order of increasing birth times. The list of birth events is implemented as a binary search tree for the same reasons as for tree networks. The list of all events that have occurred is implemented as an array. This list is not really required in practice, but is a useful check that the simulation is running correctly. The third list is a list of nodes. Each node entry is a C structure consisting of birth time (fp), number of partners (int), list of partners (implemented as an array) consisting of partner's identity number (int) and time the link was created with partner (fp) sorted in order of increasing link times (we do not differentiate between nodes being a mother, daughter, etc.), *plimit* (int), a number used to check whether we have filled up the partners array, to then allocate more memory. The node list is implemented as an array, and the subscript of the array is used as the identity number of the node, as for trees.

The algorithm used in determining the shortest distance

between any two nodes in a network with cross-links is the “breadth first search” algorithm [38]. Conceptually this is quite simple. To find the shortest distance $d(A,B)$ between two distinct nodes A and B, we pick one of the nodes, say A, and set all of A’s nearest neighbors to be distance 1 from A. Call this set of nearest neighbors \mathcal{S}_1 . If $B \in \mathcal{S}_1$, then $d(A,B)=1$ and the search ends. Otherwise, consider each node of \mathcal{S}_1 in turn, search its nearest neighbors, and for any of them that have not yet been assigned a distance from A, assign distance 2. The process continues until node B is assigned a distance. Since the network is connected, node B will always be found.

So far as the actual practical implementation of this algorithm goes, we control the order in which we process the different nodes with a first-in first-out queue data structure. The queue used is array based, though queues using linked lists are also possible [39]. When we first set all of A’s nearest neighbors to be distance 1 from A, we also put the identity numbers of these nearest neighbors into a queue. Now we get the first element from the queue, search among its nearest neighbors, and any of these nodes which have not already been processed or is already in the queue are inserted into the end of the queue. Then we get the second element in the queue and repeat the step above. We keep repeating this until either B is found (this always occurs for a connected network such as those considered in the present paper) or the queue is empty (for a disconnected network with A and B in disjoint components of the network).

When we wish to determine the distances between all pairs of nodes in the network, the simple approach of choosing each pair (A,B) of nodes in turn and implementing the above algorithm is inefficient. Instead it is better to consider each node A in turn, and use the basic breadth-first search algorithm to determine the distances of all nodes from A.

APPENDIX B: JOINT DISTRIBUTIONS

We outline here the solution for the joint distributions of coordination number and nodes for the models of randomly growing trees discussed in Secs. II and III. The method of characteristics is used to solve an appropriate first-order partial differential equation for the generating function

$$\mathcal{P}(\kappa, \zeta, t) = \sum_{k=0}^{\infty} \sum_{n=1}^{\infty} p_{k,n}(t) \kappa^k \zeta^n, \tag{B1}$$

where $p_{k,n}(t) = \text{Pr}\{K(t)=k, N(t)=n\}$. The same technique can be used for the exact analysis of model involving marriage without birth discussed in Sec. IV B, but as the details are practically identical to those for the mean-field analysis of that model in Appendix C 3, we do not give them here. The models with cross-linking produce harder partial differential equations. For the model of Sec. IV C, the partial differential equation for the generating function of

$$\text{Pr}\{K(t)=k, N(t)=n\} = p_{k,\bullet,n}(t) = \sum_m p_{k,m,n}(t)$$

can be solved in terms of an integral, but as useful exact results do not appear easily extracted from it, we do not give details here. The second of our models with cross-linking (Sec. IV D) produces a partial differential equation of second order for the generating function, and does not appear amenable to explicit solution.

1. The Yule tree model of Sec. II

From Eq. (2) we find that the partial differential equation to be solved for the generating function $\mathcal{P}(\kappa, \zeta, t)$ defined by Eq. (B1) is

$$\frac{\partial \mathcal{P}}{\partial t} + \lambda \zeta (1 - \zeta) \frac{\partial \mathcal{P}}{\partial \zeta} = \lambda (\kappa - 1) \zeta \mathcal{P}. \tag{B2}$$

We first solve the ordinary differential equation $d\zeta/dt = \lambda \zeta (1 - \zeta)$, which gives $e^{\lambda t} (1 - \zeta) / \zeta = \text{const}$ as the characteristic curves of Eq. (B2). We then write $\mathcal{P}(\kappa, \zeta, t) = \Phi(X, t)$, where $X = e^{\lambda t} (1 - \zeta) / \zeta$ and obtain the equation

$$\frac{1}{\Phi} \frac{\partial \Phi}{\partial t} = \frac{\lambda (\kappa - 1) e^{\lambda t}}{X + e^{\lambda t}},$$

giving $\Phi(X, t) = F(X) (X + e^{\lambda t})^{\kappa - 1}$. The function $F(X)$ is found from the initial condition $\mathcal{P}(\kappa, \zeta, 0) = \kappa^{k_0} \zeta^{n_0}$, corresponding to $K(0) = k_0$ and $N(0) = n_0$, and we deduce that

$$\mathcal{P}(\kappa, \zeta, t) = \frac{\kappa^{k_0} \zeta^{n_0} e^{-n_0 \lambda t}}{[1 - \zeta (1 - e^{-\lambda t})]^{n_0 + \kappa - 1}}. \tag{B3}$$

This reduces in the special cases $\kappa=1$ and $\zeta=1$ to simple single-variable generating functions from which the marginal distributions of $N(t)$ and $K(t)$, respectively, follow. We can also deduce by differentiating the generating function appropriately the moments of the distribution. The covariance is $\text{cov}\{K(t), N(t)\} = e^{\lambda t} - 1$, and the correlation coefficient is

$$\frac{\text{cov}\{K(t), N(t)\}}{\sqrt{\text{var}\{K(t)\} \text{var}\{N(t)\}}} = \frac{(1 - e^{-\lambda t})^{1/2}}{(n_0 \lambda t)^{1/2}},$$

so that $K(t)$ and $N(t)$ are asymptotically uncorrelated as $t \rightarrow \infty$.

The expansion of the generating function we have found in both variables is messy and unenlightening, but we observe that where $(\alpha)_n = \Gamma(n + \alpha) / \Gamma(\alpha)$, with Γ the usual Gamma function, we have

$$\sum_{k=k_0}^{\infty} p_{k,n}(t) \kappa^k = \frac{(n_0 + \kappa - 1)_{n - n_0} \kappa^{k_0}}{e^{n_0 \lambda t} (n - n_0)!} (1 - e^{-\lambda t})^{n - n_0},$$

while

$$\sum_{k=k_0}^{\infty} p_{k,n}(t) = p_{\bullet,n}(t) = \frac{(n_0)_{n - n_0}}{e^{n_0 \lambda t} (n - n_0)!} (1 - e^{-\lambda t})^{n - n_0}$$

and so the generating function for the conditional distribution of $K(t)$, given $N(t) = n$, has the form

$$\sum_{k=k_0}^{\infty} \Pr\{K(t)=k|N(t)=n\} \kappa^k = \frac{(n_0 + \kappa - 1)_{n-n_0} \kappa^{k_0}}{(n_0)_{n-n_0}}, \quad (\text{B4})$$

and this contains no explicit time dependence. Moreover, it is a polynomial in κ , and for values of n that are not extravagantly large, we can calculate the conditional probabilities using standard computer algebra packages.

We note, in particular, that on appropriately differentiating Eq. (B4) and setting $\kappa=1$ the conditional mean and variance of the coordination number can be found exactly in terms of the digamma function $\psi(z) = \Gamma'(z)/\Gamma(z)$ and its derivative, and the large- n asymptotic behavior extracted:

$$\langle K(t)|N(t)=n \rangle = k_0 + \psi(n) - \psi(n_0) = k_0 + \sum_{j=n_0}^{n-1} \frac{1}{j},$$

$$\text{var}\{K(t)|N(t)=n\} = \psi(n) - \psi(n_0) + \psi'(n) - \psi'(n_0)$$

and the asymptotic forms (3) and (4) follow.

2. The Reed-Hughes tree model of Sec. III

From Eq. (37) the partial differential equation for the generating function $\mathcal{P}(\kappa, \zeta, t)$ is

$$\frac{\partial \mathcal{P}}{\partial t} + \lambda \kappa (1 - \kappa) \zeta \frac{\partial \mathcal{P}}{\partial \kappa} + 2\lambda \zeta (1 - \zeta) \frac{\partial \mathcal{P}}{\partial \zeta} = 2\lambda (1 - \zeta) \mathcal{P}.$$

We construct two families of characteristics by solving the simultaneous equations

$$\frac{d\kappa}{dt} = \lambda \kappa (1 - \kappa) \zeta, \quad \frac{d\zeta}{dt} = 2\lambda \zeta (1 - \zeta),$$

and are thus led to write $\mathcal{P}(\kappa, \zeta, t) = \exp[\mathcal{Q}(X, Y, t)]$, where

$$X = \left(\frac{1 - \kappa}{\kappa} \right)^2 \frac{1}{1 - \zeta}, \quad Y = e^{2\lambda t} \left(\frac{1 - \zeta}{\zeta} \right).$$

This gives

$$\frac{\partial \mathcal{Q}}{\partial t} = 2\lambda \left(1 - \frac{e^{2\lambda t}}{Y + e^{2\lambda t}} \right),$$

so that $\mathcal{Q}(X, Y, t) = 2\lambda t - \ln(Y + e^{2\lambda t}) + f(X, Y)$, with the function $f(X, Y)$ to be determined from the initial condition, which we take to be $\mathcal{P}(\kappa, \zeta, 0) = \kappa \zeta^2$, so that the specified node whose coordination we study is here taken to be one of the two nodes initially present. After a little algebra we find that

$$\mathcal{P}(\kappa, \zeta, t) = \frac{\zeta^2 e^{-2\lambda t}}{1 - (1 - e^{-2\lambda t}) \zeta} \left(1 + \frac{(1 - \kappa)/\kappa}{\sqrt{1 - (1 - e^{-2\lambda t}) \zeta}} \right)^{-1}.$$

From derivatives of this solution we may show that the covariance is $\text{cov}\{K(t), N(t)\} = \frac{1}{2}(e^{3\lambda t} - e^{\lambda t})$. The correlation coefficient is

$$\frac{\text{cov}\{K(t), N(t)\}}{\sqrt{\text{var}\{K(t)\} \text{var}\{N(t)\}}} = \frac{(1 - e^{-2\lambda t})^{1/2}}{2(1 - e^{-\lambda t})^{1/2}},$$

and this converges to 1/2 in the long-time limit, so that the coordination number of a given one of the two starting nodes retains its correlation with the number of nodes present for all times. Observe that the correlation coefficient does not converge to 1 as $t \rightarrow 0^+$; a similar result holds for the constant birth-rate model when the system is started with more than one node initially present.

APPENDIX C: MEAN-FIELD SOLUTIONS

1. The Yule tree model of Sec. II

If we define the generating function

$$\mathcal{N}(\kappa, \rho, t) = \sum_{k=0}^{\infty} \sum_{r=0}^{\infty} \kappa^k \rho^r n(k, r, t), \quad (\text{C1})$$

we find the evolution Eq. (19) corresponds to

$$\frac{\partial \mathcal{N}}{\partial t}(\kappa, \rho, t) + \lambda(1 - \kappa)\mathcal{N}(\kappa, \rho, t) = \lambda \kappa \rho \mathcal{N}(1, \rho, t), \quad (\text{C2})$$

which we solve with initial condition $n(k, r, 0) = \delta_{k,0} \delta_{r,0}$, so that $\mathcal{N}(\kappa, \rho, 0) = 1$. Write $\Phi(\rho, t) = \mathcal{N}(1, \rho, t)$, so that $\Phi(\rho, 0) = 1$, and set $\kappa=1$ in Eq. (C2) to deduce that

$$\frac{\partial}{\partial t} \Phi(\rho, t) = \lambda \rho \Phi(\rho, t)$$

so that $\Phi(\rho, t) = e^{\lambda \rho t}$ and the required solution of Eq. (C2) is easily shown to be

$$\mathcal{N}(\kappa, \rho, t) = \frac{(1 + \rho)(1 - \kappa)e^{-\lambda(1 - \kappa)t}}{1 + \rho - \kappa} + \frac{\kappa \rho}{1 + \rho - \kappa} e^{\lambda \rho t}.$$

Thus

$$\lim_{t \rightarrow \infty} \frac{\sum_{k=0}^{\infty} \sum_{r=0}^{\infty} n(k, r, t) \kappa^k}{\sum_{k=0}^{\infty} \sum_{r=0}^{\infty} n(k, r, t)} = \lim_{t \rightarrow \infty} \frac{\mathcal{N}(\kappa, 1, t)}{\mathcal{N}(1, 1, t)} = \frac{\kappa}{2 - \kappa} = \sum_{k=1}^{\infty} \frac{\kappa^k}{2^k}$$

and Eq. (20) follows.

2. The Reed-Hughes tree model of Sec. III

If we again introduce the generating function (C1), we find that the evolution Eq. (46) and initial condition correspond to

$$\frac{\partial \mathcal{N}}{\partial t} + \lambda \kappa (1 - \kappa) \frac{\partial \mathcal{N}}{\partial \kappa} = \lambda \kappa \rho \frac{\partial \mathcal{N}}{\partial \kappa} \Big|_{\kappa=1}, \quad (\text{C3})$$

and $\mathcal{N}(\kappa, \rho, 0) = (1 + \rho)\kappa$, respectively. To solve the partial differential equation (C3), we use a variant of the method of characteristics. Suppressing the dependence on ρ in notation for brevity, we define

$$\Phi(t) = \left. \frac{\partial \mathcal{N}}{\partial \kappa} \right|_{\kappa=1} \quad (\text{C4})$$

so that the equation to be solved is

$$\frac{\partial \mathcal{N}}{\partial t} + \lambda \kappa (1 - \kappa) \frac{\partial \mathcal{N}}{\partial \kappa} = \lambda \kappa \rho \Phi(t). \quad (\text{C5})$$

Solving the equation $d\kappa/dt = \lambda \kappa (1 - \kappa)$ we find the characteristic curves to be given by $X = e^{\lambda t} (1 - \kappa) / \kappa$, and we seek the general solution of Eq. (C5) by writing $\mathcal{N} = Q(X, t)$, which gives

$$\frac{\partial Q}{\partial t} = \frac{\lambda \rho e^{\lambda t} \Phi(t)}{X + e^{\lambda t}}.$$

Integrating from 0 to t and writing $F(X) = Q(X, 0)$, we deduce that

$$\mathcal{N}(\kappa, \rho, t) = \int_0^t \frac{\lambda \rho e^{\lambda \tau} \Phi(\tau) d\tau}{X + e^{\lambda \tau}} + F(e^{\lambda t} (1 - \kappa) / \kappa).$$

From the initial condition $\mathcal{N}(\kappa, \rho, 0) = (1 + \rho)\kappa$ we have $F(\kappa^{-1} - 1) = (1 + \rho)\kappa$, whence $F(z) = (1 + \rho)/(1 + z)$. Replacing the variable X by the original variables, we have now shown that

$$\mathcal{N}(\kappa, \rho, t) = \int_0^t \frac{\lambda \rho \kappa \Phi(\tau) d\tau}{\kappa + e^{\lambda(t-\tau)}(1-\kappa)} + \frac{(1+\rho)\kappa}{\kappa + e^{\lambda t}(1-\kappa)}.$$

To determine the function $\Phi(t)$, we differentiate with respect to κ and set $\kappa=1$, giving

$$\Phi(t) = (1 + \rho)e^{\lambda t} + \lambda \rho e^{\lambda t} \int_0^t \Phi(\tau) e^{-\lambda \tau} d\tau.$$

If we write $\Phi(t) = e^{\lambda t} \chi(t)$ we obtain

$$\chi(t) = 1 + \rho + \lambda \rho \int_0^t \chi(\tau) d\tau,$$

leading to the differential equation $\chi'(t) = \lambda \rho \chi(t)$ and initial condition $\chi(0) = 1 + \rho$, so $\chi(t) = (1 + \rho)e^{\lambda \rho t}$, and it follows that

$$\mathcal{N}(\kappa, \rho, t) = \int_0^t \frac{\lambda \rho (1 + \rho) \kappa e^{\lambda(1+\rho)\tau} d\tau}{\kappa + e^{\lambda(t-\tau)}(1-\kappa)} + \frac{(1+\rho)\kappa}{\kappa + e^{\lambda t}(1-\kappa)},$$

which is Eq. (47).

3. Marriage alone

With $\mathcal{N}(\kappa, t) = \sum_{k=0}^{\infty} n(k, t) \kappa^k$, the evolution Eq. (67) becomes

$$\frac{\partial \mathcal{N}}{\partial t} - 2\mu \kappa (1 - \kappa) \frac{\partial \mathcal{N}}{\partial \kappa} + 2\mu (1 - \kappa) (n_0 - 1) \mathcal{N} = 0,$$

with the initial condition $n(k, 0) = n_0 \delta_{k,0}$ becoming $\mathcal{N}(\kappa, 0) = n_0$. Solving the equation $d\kappa/dt = -2\mu \kappa (1 - \kappa)$ we find the characteristic curves $X = e^{-2\mu t} (1 - \kappa) / \kappa$, and write $\mathcal{N}(\kappa, t) = Q(X, t)$. This gives

$$\frac{\partial Q}{\partial t} + 2\mu \left(1 - \frac{e^{-2\mu t}}{X + e^{-2\mu t}} \right) (n_0 - 1) Q = 0,$$

and so

$$Q(X, t) = Q(X, 0) e^{-2\mu(n_0-1)t/(X + e^{-2\mu t})n_0^{-1}}.$$

Fitting the initial condition, we find that

$$\begin{aligned} \mathcal{N}(\kappa, t) &= n_0 e^{-2\mu(n_0-1)t} \left(\frac{X+1}{X + e^{-2\mu t}} \right)^{n_0-1} \\ &= n_0 [e^{-2\mu t} (1 - \kappa) + \kappa]^{n_0-1} \end{aligned}$$

and using the binomial theorem to extract the coefficient of κ^k , we obtain the mean-field approximation

$$n(k, t) = n_0 \binom{n_0-1}{k} (1 - e^{-2\mu t})^k e^{-2\mu t(n_0-1-k)},$$

which agrees with an exact calculation of the coordination-number distribution in Sec. IV B. The general solution (66) of the exact evolution Eq. (64) for $p_{k,\bullet,n}(t)$ can be constructed by a slight extension of the preceding analysis.

4. Constant birth-rate polygamy

We introduce the generating function

$$\mathcal{N}(\kappa, t) = \sum_{k=0}^{\infty} n(k, t) \kappa^k,$$

so that the initial condition is $\mathcal{N}(\kappa, 0) = 1$. As we have already noted that $n(t) = e^{\lambda t}$ (both rigorously and in mean field), we are able to deduce from Eq. (80) that

$$\begin{aligned} \frac{\partial \mathcal{N}}{\partial t} - 2\mu \kappa (1 - \kappa) \frac{\partial \mathcal{N}}{\partial \kappa} + (1 - \kappa) [\lambda + 2\mu(e^{\lambda t} - 1)] \mathcal{N} \\ = \kappa \lambda e^{\lambda t}, \end{aligned} \quad (\text{C6})$$

with the initial condition $n(k, 0) = \delta_{k,0}$ becoming $\mathcal{N}(\kappa, 0) = 1$. The solution by characteristics is very similar to that in Appendix C 3, but the integrals that arise in the solution do not appear to be simply evaluable and we do not write them out. We settle for extracting the mean-field prediction of the mean coordination number. Write $m(t) = \sum_k k n(k, t) = \partial \mathcal{N} / \partial \kappa |_{\kappa=1}$. Then on differentiating Eq. (C6) with respect to κ and setting $\kappa=1$ we deduce that

$$m'(t) + 2\mu m(t) - \lambda e^{\lambda t} - 2\mu(e^{\lambda t} - 1)e^{\lambda t} = \lambda e^{\lambda t},$$

with $m(0)=0$. This differential equation is easily solved. We find that as $t \rightarrow \infty$, $m(t) \sim \mu e^{2\lambda t}/(\lambda + \mu)$, and so the mean-field prediction of the mean coordination number is $m(t)/n(t) \sim \mu n(t)/(\lambda + \mu)$.

5. Variable birth-rate polygamy

Introducing $\mathcal{N}(\kappa, t)$ as in Appendix C 4 and noting that the relevant initial condition is $\mathcal{N}(\kappa, 0) = 2\kappa$, we find the evolution equation

$$\frac{\partial \mathcal{N}}{\partial t} + (\lambda - 2\mu)\kappa(1 - \kappa) \frac{\partial \mathcal{N}}{\partial \kappa} + 2\mu(1 - \kappa) \times [\mathcal{N}|_{\kappa=1} - 1] \mathcal{N} = \lambda \kappa \frac{\partial \mathcal{N}}{\partial \kappa} \Big|_{\kappa=1}. \quad (\text{C7})$$

We write for brevity $n(t) = \mathcal{N}(1, t)$ for the total number of nodes present and $m(t) = \partial \mathcal{N} / \partial \kappa|_{\kappa=1}$. The mean coordination number is then

$$c(t) = \mathcal{N}(1, t)^{-1} \partial \mathcal{N} / \partial \kappa|_{\kappa=1} = m(t)/n(t).$$

Setting $\kappa=1$ in Eq. (C7) gives $n'(t) = \lambda m(t)$ and so $c(t) = n'(t)/[\lambda n(t)]$. Differentiating Eq. (C7) with respect to κ and then setting $\kappa=1$ gives

$$m'(t) - 2(\lambda - \mu)m(t) - 2\mu[n(t) - 1]n(t) = 0.$$

We may now eliminate $m(t)$ in favor of $n(t)$ to obtain the mean-field evolution equation for $n(t)$,

$$n''(t) - 2(\lambda - \mu)n'(t) - 2\lambda\mu[n(t) - 1]n(t) = 0, \quad (\text{C8})$$

to be solved subject to the initial conditions $n(0) = 2$, $n'(0) = 2\lambda$. The differential Eq. (C8) is autonomous, so we can obtain an associated first-order differential equation in the standard way. Write $v = n' = \lambda n c$ so that $n'' = v dv/dn = \lambda^2 n c d(nc)/dn$, giving an evolution Eq. (87) for c as a function of n .

-
- [1] B.A. Huberman and L.A. Adamic, *Nature (London)* **401**, 131 (1999); L.A. Adamic and B.A. Huberman, *Science* **287**, 2115 (2000).
- [2] M.E.J. Newman, D.J. Watts, and S.H. Strogatz, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2566 (2002).
- [3] R. Albert and A.L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
- [4] S.N. Dorogovtsev and J.F.F. Mendes, *Adv. Phys.* **51**, 1079 (2002).
- [5] By the ‘‘coordination number’’ of a node we mean the number of other nodes that are directly linked to it. In other contexts, this number is called the ‘‘valence,’’ the ‘‘connectivity,’’ or the ‘‘degree.’’
- [6] We use the term ‘‘mean field’’ fairly loosely to describe the replacement of stochastic evolution by deterministic evolution in a natural way. With this approach, we are able to obtain exact results in some cases, and excellent approximations in other cases. Where our mean-field arguments fail, we are able to devise appropriate correction procedures and an example of this is given in Sec. II. As our processes are intrinsically continuous-time ones, we do not have to deal with additional errors that arise from replacing differences with derivatives. Moreover, the same basic approach is used for all attributes of networks that we explore, so that for trees no *ad hoc* adjustments or parameter fits are needed to match predictions to simulation data. To illustrate this point, we note the interesting paper of G. Szabó, M. Alava, and J. Kertész, *Phys. Rev. E* **66**, 026101 (2002), where a tree model with coordination number varying deterministically with ring number is used to approximate scale-free networks, the degree distribution exponent -3 is lost, and the approximate theory derived has to be fitted to simulation data, whereas our approach has no adjustable parameters.
- [7] One may argue that the origins of discrete-time modeling of networks can be traced to a paper on copy and growth processes by H.A. Simon, *Biometrika* **42**, 425 (1955); and this connection has been noted by S. Bornholdt and H. Ebel, *Phys. Rev. E* **64**, 035104 (2001). However, random graph models much more in the spirit of recent interests can be fairly attributed to Szymański (Refs. [8] and [12]).
- [8] J. Szymański, in *Random Graphs '87 (Poznan 1987)*, edited by M. Karoński and A. Ruciński (Wiley, New York, 1990), pp. 313–324.
- [9] J.W. Moon, in *Combinatorics*, edited by T.P. McDonough and V.C. Mavron, London Math. Soc. Lecture Note Ser., Vol. 13 (Cambridge University Press, London, 1974), pp. 125–132.
- [10] A. Meir and J.W. Moon, *Can. J. Math.* **30**, 997 (1978).
- [11] R. van Hofstad, G. Hooghiemstra, and P. van Mieghem, *Random Struct. Algorithms* **20**, 519 (2002).
- [12] J. Szymański, in *Random Graphs '85*, edited by M. Karoński and Z. Palka, *Ann. Discr. Math.* Vol. 33 (Elsevier, Amsterdam, 1987), p. 297.
- [13] Szymański [12] also refers to earlier unpublished work of Dondajewski and Szymański.
- [14] W.J. Reed and B.D. Hughes, *Scientiae Mathematicae Japonicae Online* **8**, 329 (2003).
- [15] W.J. Reed and B.D. Hughes, *J. Theor. Biol.* **217**, 125 (2002).
- [16] B.D. Hughes and W.J. Reed, physics/0211090.
- [17] W. J. Reed and B.D. Hughes, *Math. Biosci.* (to be published).
- [18] W.J. Reed and B.D. Hughes, *Physica A* **319**, 579 (2003).
- [19] W.J. Reed and B.D. Hughes, *Phys. Rev. E* **66**, 067103 (2002).
- [20] E. Fermi, *Phys. Rev.* **75**, 1169 (1949); the authors are indebted to Professor J. I. Katz for this reference.
- [21] The Wiener index is not due to Norbert Wiener, but due to the chemist H. Wiener, *J. Am. Chem. Soc.* **69**, 17 (1947). For a collection of results and applications of the Wiener index, see A.A. Dobrynin, R. Entringer, and I. Gutman, *Acta Appl. Math.* **66**, 211 (2001). Some studies of the Wiener index of classes of random trees have been given recently by Neininger (Ref. [29]).
- [22] S. Wolfram, *Mathematica: A System for Doing Mathematics by Computer* (Addison-Wesley, Reading, MA, 1988). Computations were done using MATHEMATICA, Version 4.0 (Wolfram

- Research). An inaccuracy in the evaluation of the Hurwitz zeta function $\zeta(s, a)$ for some parameter values did not affect any of our reported results.
- [23] N.T.J. Bailey, *The Elements of Stochastic Processes* (Wiley, New York, 1964), p. 85.
- [24] P.D. Feigin, *J. Appl. Probab.* **16**, 297 (1979).
- [25] P.L. Krapivsky, S. Redner, and F. Leyvraz, *Phys. Rev. Lett.* **85**, 4629 (2000); P.L. Krapivsky, C.J. Rodgers, and S. Redner, *ibid.* **86**, 5401 (2001).
- [26] Briefly, let $A(t)$ be the number of nodes corresponding to the primal node's first-born daughter and all her offspring, and $B(t)$ the number of nodes corresponding to all other daughters of the primal node and their offspring. An evolution equation for $\Pr\{A(t)=a, B(t)=b\}$ can be set up and solved, and if we write $\Phi(t) = B(t)/N(t) = B(t)/[A(t) + N(t) + 1]$, we find that $\langle \Phi(t) \rangle \rightarrow 1/2$ and $\text{var}\{\Phi(t)\} \rightarrow 1/12$ as $t \rightarrow \infty$.
- [27] A. Erdélyi, W. Magnus, F. Oberhettinger, and F.G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 1, pp. 27–31.
- [28] R.P. Dobrow, *J. Appl. Probab.* **33**, 749 (1996).
- [29] R. Neininger, *Combinatorics, Probab. Comput.* **11**, 587 (2002).
- [30] R.C. Entringer, D.E. Jackson, and D.A. Snyder, *Czech. Math. J.* **26**, 283 (1976).
- [31] E.T. Whittaker and G.N. Watson, *A Course of Modern Analysis*, 4th ed. (Cambridge University Press, Cambridge, 1927), Chap. XIII.
- [32] See Ref. [3] and also A.L. Barabási and R. Albert, *Science* **286**, 509 (1999); A.L. Barabási, R. Albert, and H. Jeong, *Physica A* **272**, 173 (1999); **281**, 69 (2000); R. Albert and A.L. Barabási, *Phys. Rev. Lett.* **85**, 5234 (2000).
- [33] See Ref. [4] and S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin, *Phys. Rev. Lett.* **85**, 4633 (2000).
- [34] The model in Sec. III was briefly outlined by Reed and Hughes (Ref. [14]) but the simulation results and mean-field analysis given in Sec. III are new.
- [35] A second way of formulating a model for the evolution of a tree-structured network, which gives the same result as above, is closer to the more familiar model of the evolution of scale-free networks (cf. Ref. [3]). Rather than assuming a fixed number of new nodes is added to the network in every period, assume instead that new nodes are created in a Yule process at rate 2λ , and that any new node will connect to one of the existing nodes with probabilities proportional to the current coordination numbers of these existing nodes. After some steps this yields an evolution equation for $p_k(t)$ the same as Eq. (38) and in consequence a geometric distribution for $K(t)$, the same as Eq. (39). Since the Yule process is an order statistic process it follows that the time elapsed since the introduction of a specified node * will follow a truncated exponential distribution on $(0, \tau)$, again with parameter 2λ . Thus as $t \rightarrow \infty$ the distribution of the coordination number of a random node will follow the Yule distribution (45).
- [36] It is possible to derive a mean-field argument that suggests that $\langle K(t) \rangle \rightarrow 2$ whenever $\gamma \geq 1$, but we shall not give details here. Krapivsky *et al.* [25] obtain by nonrigorous means a number of interesting results on the behavior of observable quantities as γ changes.
- [37] The differential equation for c as a function of n was easily solved to an acceptable level of precision using the ODESOLVE routine in MUPAD (SciFace Software) and the solution was checked against the NDSOLVE package in MATHEMATICA 4.0 (Wolfram Research).
- [38] See, for example, G. Chartrand and O.R. Oellermann, *Applied and Algorithmic Graph Theory* (McGraw-Hill, New York, 1993); R. Sedgewick, *Algorithms in C*, 3rd ed. (Addison-Wesley, Reading, MA, 2002), pt. 5.
- [39] See R. Sedgewick, *Algorithms in C*, 3rd ed. (Reading, MA, Addison-Wesley, 1998), Pts. 1–4, for further details on both types of queues.