Accelerating, hyperaccelerating, and decelerating networks

M. J. Gagen* and J. S. Mattick

ARC Special Research Centre for Functional and Applied Genomics, Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland 4072, Australia
(Received 21 February 2005; published 20 July 2005)

Many growing networks possess accelerating statistics where the number of links added with each new node is an increasing function of network size so the total number of links increases faster than linearly with network size. In particular, biological networks can display a quadratic growth in regulator number with genome size even while remaining sparsely connected. These features are mutually incompatible in standard treatments of network theory which typically require that every new network node possesses at least one connection. To model sparsely connected networks, we generalize existing approaches and add each new node with a probabilistic number of links to generate either accelerating, hyperaccelerating, or even decelerating network statistics in different regimes. Under preferential attachment for example, slowly accelerating networks display stationary scale-free statistics relatively independent of network size while more rapidly accelerating networks display a transition from scale-free to exponential statistics with network growth. Such transitions explain, for instance, the evolutionary record of single-celled organisms which display strict size and complexity limits.

I. INTRODUCTION

The rapidly expanding field of network analysis, reviewed in [1–3], has provided examples of networks exhibiting "accelerating" network growth, where total link number grows faster than linearly with network size [4]. For instance, the internet and worldwide web appear to grow by adding links more quickly than sites [5,6] though the relative change over time is small [7]. Similarly, both the number of links per substrate in the metabolic networks of organisms [8] and the average number of links per scientist in collaboration networks increase linearly over time [9–13], while also languages appear to evolve via accelerated growth [14]. These studies have motivated general examinations of accelerating networks [15], including accelerating citation networks [16], duplication models exhibiting linear and superlinear growth in link number with network size [17–19], and stochastic accelerating networks where the number of new links added is probabilistic (though integral and greater than 1) causing network transitions from scale-free to exponential statistics [13,15,20–23].

These previous studies of deterministic or stochastic networks have typically considered networks where every node has at least one connection, which constrains the rate of acceleration that can be considered. In particular, if each new node added to a network of N nodes is accompanied by $N^\alpha \geq 1$ new links with acceleration parameter $\alpha \geq 0$, then ensuring that the network is less than fully connected constrains acceleration parameters to the range $0 \leq \alpha < 1$ [6]. Equivalent limits were considered in Refs. [15,17–19]. However, real networks can contain a majority of nodes that are entirely unconnected—many computers are unconnected or only intermittently connected to the internet and many

people are unconnected nodes in social or sexual contact networks. Modeling networks where a substantial proportion of the nodes are unconnected requires a probabilistic approach in which the number of links added with each new node scales as pN^α for some probability parameter $0 \leq p \leq 1$. The introduction of a probabilistic envelope lifts constraints on the acceleration parameter allowing networks displaying deceleration $\alpha < 0$, no acceleration $\alpha = 0$, acceleration $0 \leq \alpha < 1$, and hyperacceleration $\alpha \geq 1$ (while, as we will show, imposing additional constraints on the ability of the network to grow).

Hyperaccelerating probabilistic networks with $p \ll 1$ and $\alpha \geq 1$ are not merely of theoretical interest, and appear for instance in prokaryote gene regulatory networks where many gene nodes are unregulated and merely constitutively or stochastically expressed [24–26]. Prokaryote gene regulatory networks are sparsely connected ($p \ll 1$) and display hyperacceleration $\alpha = 1$ as established by independent comparative genomics analyses [27–31]. This is likely due to their reliance on sequence homology interactions between protein transcription factors and specific promoter binding sequence sites [31]. In these regulatory networks, outbound regulatory links are preferentially attached to existing regulator nodes as gene duplication events contribute to up to 75% of each new gene [32–38] and regulatory transcription factor [39,40]. In contrast, inbound links to regulated nodes are randomly formed due to the random drift of gene promotor sequences, although their subsequent fixation is determined by selection. The high acceleration parameter $\alpha = 1$ ensures that these prokaryote gene regulatory mechanisms are size and complexity constrained at predicted limits which closely match genomic size limits observed in the evolutionary record [24–26,31]. Accelerating networks are more prevalent and important in society and in biology than is commonly realized. In fact, any network in which the dynamical state of any node depends on the immediately preceding state of

*Electronic address: m.gagen@imb.uq.edu.au

(ideally) every other node is accelerating and will perhaps display either structural transitions from randomly connected, to scale-free statistics, to densely connected, and perhaps finally to fully connected statistics, or explicit size and complexity constraints [41].

This paper generalizes previous analyses to consider probabilistic accelerating networks in decelerating regimes with $\alpha < 0$, accelerating regimes with $0 \leq \alpha < 1$, and hyperaccelerating regimes where $\alpha \geq 1$. We will show that hyperaccelerating networks under preferential attachment of new connections to existing nodes can display transitions in their degree distribution statistics (measuring the connectivity, i.e., the number of connections at each node) from scale-free to exponential statistics dependent on the acceleration parameter α, the probability parameter p, and the network size N. These transitions occur at the point where the accelerating growth in connection number can no longer be sustained, introducing exponential cutoffs in the node degree distributions. These cutoffs can be generated by different mechanisms and appear in growing networks subject to strong aging effects [42], in finite-sized networks [43], with the discarding of nodes [44], and in information filtered networks [45,46].

Throughout, the degree of node n_i under the continuum approximation is k_{ij} at network size j. In existing treatments, the k_{ij} degree distribution suffices as every node has at least one connection. In contrast, we here consider sparsely connected networks where many nodes are initially unconnected, and it is sometimes necessary to consider the average degree distribution of only connected nodes, denoted $k_c(j, N)$. The node degree distribution k_{ij}, as usual, allows calculating the link distribution $P(k, N)$ specifying the probability that a node has degree k in a network of size N. However, for rapidly accelerating networks, it is possible that all nodes have fractional degree less than unity, $P(k, N) \ll 1$, which of course is unobservable as nodes must have an integer number of connections. This requires the specific calculation of a discrete degree distribution P_k specifying the probability that a network node has integer connectivity k given the continuous degree distribution $P(k, N)$.

To complete our examination of network transitions in accelerating, hyperaccelerating, and decelerating probabilistic networks, we first define our growing network models using undirected links in Sec. II. This definition allows an immediate rough quantification of the probability of an accelerating network forming a single giant connected component or undergoing a transition to a fully connected regime with network growth. In Sec. III we examine probabilistic networks growing through the preferential attachment of new links to existing connected network nodes, while in Sec. IV we examine growing probabilistic networks featuring the random attachment of new links to all existing nodes.

II. PROBABILISTIC ACCELERATING NETWORKS

In this section, we introduce notation describing our growing networks allowing the location of growth transitions. We consider networks growing through the sequential addition of numbered nodes n_i for $1 \leq i \leq N$ where at network

size j, node n_i $(1 \leq i \leq j)$ has k_{ij} undirected links. The addition of node n_j and its k_{jj} links will increase the probable number of links attached to existing nodes n_i for $1 \leq i \leq (j-1)$ so $k_{ij} \geq k_{ii}$. We assume new links are added only with each arriving node (so no new links are added between established nodes), and that the average number of new links attached to node n_i on its entry to the network is a function of network size,

$$k_{ii} = pi^\alpha, \tag{1}$$

where the probability constant p satisfies $0 \leq p \leq 1$. This average is then a decreasing function of network size for $\alpha < 0$, constant for $\alpha = 0$, and an increasing function of network size for $\alpha > 0$. The total expected number of links L in a probabilistic network of large size N is then

$$L = \sum_{i=1}^{N} pi^\alpha \approx \int_1^N pi^\alpha \, di$$

$$\approx \begin{cases} \dfrac{-p}{1+\alpha}(1 - N^{1+\alpha}) \approx -\dfrac{p}{1+\alpha}, & \alpha < -1, \\[2mm] p \ln N, & \alpha = -1, \\[2mm] \dfrac{p}{1+\alpha}(N^{1+\alpha} - 1) \approx \dfrac{pN^{1+\alpha}}{1+\alpha}, & \alpha > -1. \end{cases} \tag{2}$$

Consequently, the expected number of connections is $C = 2L$, and the average degree of each node is $\langle k \rangle = C/N$. In later numerical simulations, we compare the statistics of networks of different sizes, which is possible provided they possess the same average degree at some maximum simulated network size, denoted N_c. We achieve this for accelerating networks with $\alpha \geq 0$ by choosing different probability parameters p for each different sized network according to

$$p = \frac{(1+\alpha)\langle k \rangle}{2N_c^\alpha}. \tag{3}$$

This allows us to later consider different networks with identical average degree $\langle k \rangle = 0.4$, where networks have maximum simulated size of either $N_c = 1000$ nodes connected by $L = 200$ links with $p = 0.2(1+\alpha)10^{-3\alpha}$, or $N_c = 10\,000$ nodes connected by $L = 2000$ links with $p = 0.2(1+\alpha)10^{-4\alpha}$.

Because link formation is probabilistic, we must specify a probability distribution for the formation of links with each node added. The total number of links attached to node n_i on entry to the network, denoted j say, ranges between 0 and i, so $0 \leq j \leq i$ with average pi^α. As each link either forms or does not form, we model the link formation process using a binomial distribution. For the binomial distribution, the average $k_{ii} = pi^\alpha$ equates to the product of the maximum number of links i and the link formation probability. Conversely, the size-dependent link formation probability at network size i equates to $k_{ii}/i = pi^{\alpha-1}$. Consequently, the probability distribution that node n_i forms exactly j links on entry to the network is

$$\mathcal{P}(j,i) = \binom{i}{j}(pi^{\alpha-1})^j[1 - pi^{\alpha-1}]^{i-j}. \qquad (4)$$

That is, when $\alpha=0$ the latest node n_i can potentially form links with every one of the i existing nodes (counting itself) with a network-size-dependent probability p/i to give an average number of new links of p connections with each new node. When $\alpha=1$, the latest node n_i can potentially form links with every other node with constant probability p to give an average number of new links of pi with each new node, and so on. The connection probability is constrained to be less than unity, $pi^{\alpha-1} \leq 1$, or equivalently, the average number of links added to node n_i is required to be less than i, $pi^{\alpha} \leq i$. For constant p, hyperaccelerating networks with $\alpha>1$ will eventually violate these constraints with network growth which effectively imposes size and structural constraints on these networks.

There are two types of network degree transitions which have been of interest in the literature. The first occurs when a growing network changes from sparse to dense connectivity as this change necessarily implies that the final degree statistics will become exponential in nature. In fact, any network of N nodes must be fully connected if it has $N^2/2$ undirected links assuming that no nodes are connected by more than one link. In this case, the degree distributions are δ-function distributed, and are thus in the exponential class of distributions. That is, while a sparsely connected network can possess either scale-free or exponential statistics, a densely connected network must exhibit exponential statistics (as multiple connections are not allowed between any two given nodes). This transition occurs for undirected links when approximately $L>N^2/2$. The second transition of interest occurs when the connected nodes coalesce to form a single giant island of interconnected nodes. To borrow results from random graph theory [47] (which have been roughly validated for accelerating biological networks [24–26]), the giant island is expected to form when approximately $L>N/2$. We now investigate the dependence of these transitions on the acceleration parameter α.

It is evident that decelerating networks with $\alpha<0$ are very sparsely connected so neither of the dense connectivity or giant island transitions will occur. Hereinafter, we only consider accelerating networks with $\alpha \geq 0$. For such networks, the dense connectivity and giant island transitions will occur, respectively, at the points

$$L > \frac{N^2}{2} \Rightarrow N^{\alpha-1} > \frac{1+\alpha}{2p} = \frac{N_c^{\alpha}}{\langle k \rangle},$$

$$L > \frac{N}{2} \Rightarrow N^{\alpha} > \frac{1+\alpha}{2p} = \frac{N_c^{\alpha}}{\langle k \rangle}. \qquad (5)$$

This establishes that, for instance, nonaccelerating networks with $\alpha=0$ can never undergo a transition to a densely connected regime for any choice of the probability parameter p for large network sizes N, though these networks can form giant islands with network growth as long as $p>1/2$. This inability to become densely connected remains for all accelerating regimes $\alpha<1$, but dense connectivity can emerge in

the hyperacceleration regime. At this point with $\alpha=1$, growing networks can undergo a transition to the densely connected regime when the probability parameter is large, $p\approx1$. For yet higher acceleration rates $\alpha>1$, all growing networks undergo a transition to the dense connectivity regime and display single giant islands, and these transitions occur at network sizes determined by the probability parameter p.

In the following sections, we determine the degree statistics for accelerating and hyperaccelerating networks under both preferential and random attachment of new links to established nodes.

III. PREFERENTIALLY ATTACHED ACCELERATING NETWORKS

We now turn to consider accelerating and hyperaccelerating networks ($\alpha \geq 0$) where new links are principally preferentially attached to existing connected nodes. The final generated link distribution will depend on the balance of different countervailing trends, namely, the respective weights given to preferential versus random attachment, and the magnitude of the acceleration parameter which influences both the rate at which links are attached to newer nodes and the rate at which older nodes accumulate links. We now examine these trends in detail.

We now calculate the degree distribution under preferential attachment. On entry into the network, node n_i establishes an average of $k_{ii}=pi^{\alpha}$ links with existing nodes $n_j \in \{n_1,\dots,n_i\}$. We suppose that these links either are preferentially attached to node n_j with probability proportional to n_j's current degree k_{ji}, or are randomly attached to n_j with probability proportional to β, a random connection parameter. Using the continuous approximation [4,48,49], the rate of growth in link number for node n_j is

$$\frac{\partial k_{ji}}{\partial i} = k_{ii}\frac{(k_{ji}+\beta)}{\displaystyle\int_0^i (k_{ji}+\beta)dj}. \qquad (6)$$

Here, the rate of growth in the node degree k_{ji} is determined by the number of new links added with node n_i (i.e., k_{ii}) which can be either preferentially attached to the existing nodes n_j with probability proportional to that node current degree (i.e., k_{ji}), or randomly attached if $\beta>0$ allowing even initially unconnected nodes to establish connections. (The case $\beta=0$ ensures that a node that receives zero links on entry to the network remains unconnected for all time.)

The denominator of Eq. (6) is a probability weighting to ensure normalization and is roughly equal to the total number of connections (C) for all nodes at network size i. Following [1], we can evaluate the denominator using the identity

$$\frac{\partial}{\partial i}\int_0^i k_{ji}dj = \int_0^i \frac{\partial k_{ji}}{\partial i}dj + k_{ii}, \qquad (7)$$

which can be evaluated using Eq. (6) to give

$$\frac{\partial}{\partial i}\int_0^i k_{ji}\,dj = 2k_{ii}. \tag{8}$$

Noting $k_{ii} \approx pi^\alpha$, this can be integrated to determine the denominator of Eq. (6) to be

$$\int_0^i (k_{ji}+\beta)dj = \frac{2p}{\alpha+1}i^{\alpha+1}+\beta i. \tag{9}$$

The first term on the right-hand side is the total number of connections at network size i and is in agreement with Eq. (2). Substituting this relation back into Eq. (6) gives

$$\frac{\partial k_{ji}}{\partial i} = \frac{pi^{\alpha-1}(k_{ji}+\beta)}{[2p/(\alpha+1)i^\alpha+\beta]}. \tag{10}$$

The degree statistics can readily be obtained by integrating Eq. (10) with the initial conditions $k_{jj} \approx pj^\alpha$ at time j to obtain

$$k_{jN}=\begin{cases} (p+\beta)\left(\dfrac{N}{j}\right)^{p/(2p+\beta)}-\beta, & \alpha=0,\\[2em] (pj^\alpha+\beta)\left(\dfrac{\dfrac{2p}{\alpha+1}N^\alpha+\beta}{\dfrac{2p}{\alpha+1}j^\alpha+\beta}\right)^{(\alpha+1)/2\alpha}-\beta, & \alpha>0. \end{cases} \tag{11}$$

Differentiation of these curves with respect to network size N satisfies Eq. (6), while integrating these link numbers over all mode numbers j (numerically for $\alpha>0$) gives the required total number of links as in Eq. (2). Further, differentiating the linkage distribution k_{jN} with respect to the age index j establishes that average node degree is monotonically decreasing with increasing node age j for $0 \leq \alpha<1$, flat for $\alpha=1$, and monotonically increasing for $\alpha>1$. In other words, irrespective of the choice of the β parameter, slowly accelerating networks with $0 \leq \alpha<1$ have the majority of their connections in the oldest portions of their networks, while hyperaccelerating networks with $\alpha>1$ place the majority of the connections in the youngest portions of their networks. The transition point occurs at $\alpha=1$ when links are uniformly distributed over the network irrespective of node age.

The further analysis of this age-dependent degree distribution is heavily dependent on the setting of the random choice parameter β. Self-evidently, setting β small ensures that preferential attachment processes dominate while the choice of large β ensures that random attachment dominates. Appendix A uses a Taylor series expansion to demonstrate that preferential attachment dominates hyperaccelerating networks only when β is so small as to be essentially zero, and hereinafter, we analyze preferential attachment using the setting $\beta=0$.

We are now in a position to simulate growing networks under pure preferential attachment with no random contribution $\beta=0$. For convenience, we note that the choice $\beta=0$ sets the degree versus age distribution to be
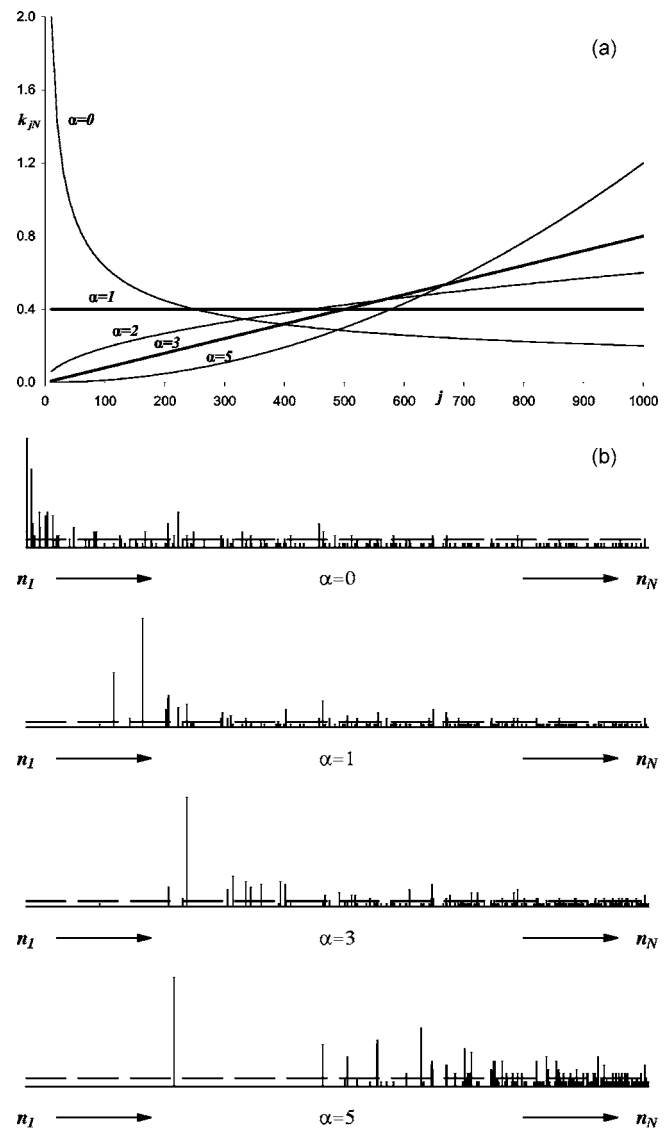

FIG. 1. Snapshots of the nonstationary statistics of preferentially attached growing networks all possessing an average number of links per node $\langle k \rangle=0.4$ when at network size $N=1000$. (a) The average degree distribution k_{jN} as a function of node age j for various values of the acceleration parameter α. The degree distribution is monotonically decreasing for $0 \leq \alpha<1$, flat for $\alpha=1$, and monotonically increasing for $\alpha>1$. (b) Example simulated networks for various acceleration parameters α with nodes listed sequentially left to right from n_1 to n_N with the degree of each node indicated by vertical lines. The horizontal dashed line indicates an average degree of 2.

$$k_{jN}=pN^{(\alpha+1)/2}j^{(\alpha-1)/2}, \tag{12}$$

for arbitrary α. The choice $\alpha=0$ duplicates existing results found for growing networks which add a constant number of links with each new node subject to preferential attachment [2], while additionally, choosing $p=1$ duplicates the deterministic results of Ref. [15] in the regimes $0 \leq \alpha<1$ considered in that paper. This purely preferentially attached distribution is shown in Fig. 1 along with example simulation runs at different acceleration parameters α. (The floating free end

of the first link to the first connected node is randomly attached to any of the existing nodes as they all possess zero links reducing preferential attachment to random attachment.) Part (a) of this figure shows the average degree k_{jN} of node n_j which is monotonically decreasing with respect to node age for $0 \leq \alpha < 1$ and monotonically increasing when $\alpha > 1$. These different trends depend on whether the accumulative effects of preferential attachment outweighs the accelerating number of added links or not. The setting $\alpha = 1$ ensures the average number of links per node is independent of node number as preferential attachment exactly cancels the bias in initial link number towards later nodes. Part (b) of this figure shows the actual degree of node n_j (rather than the average degree) in example simulated networks, making it evident that actual node degree is generally a monotonically decreasing function of node age. Why this is so is discussed later.

Observed statistics normally lack age or node number information, and we now follow standard practice in calculating the age-independent degree distribution. The k_{jN} distribution contains information about both node degree and node age and so approximates network statistics (simulated or observed) when all of this information is available. However, it is usually the case that node age information is unavailable necessitating calculation of degree distributions that are not conditioned on node age. This effectively requires binning together all nodes irrespective of their age to obtain a final link distribution. Following Ref. [24], the usual continuum approach [4,48,49] must be modified when nonmonotonically decreasing degree statistics are present. In particular, the final degree distribution is obtained via

$$P(k,N) = \frac{1}{N} \int_0^N dj \; \delta(k - k_{jN}) = \pm \frac{1}{N} \left(\frac{\partial k_{jN}}{\partial j} \right)^{-1} \quad \text{at } j = j(k,N),$$

(13)

where $j(k,N)$ is the solution of the equation $k = k_{jN}$. The top line is used when all nodes possess the same average link number while the second line is applicable with the plus (negative) sign when the average numbers of links per node is monotonically increasing (decreasing) with node number. Nonmonotonic cases require alternative approaches.

We now use Eq. (12) to calculate the age-independent final link distribution $P(k,N)$ relevant when age information is unavailable. In the case $\alpha = 1$ we have $j \neq j(k)$, meaning the degree k is independent of the node age j, so the δ function of Eq. (13) immediately integrates to give $\int_0^N dj \; \delta(k - pN)/N = \delta(k - pN)$. For other cases, we have the constraint

$$j = p^{2/(1-\alpha)} N^{(1+\alpha)/(1-\alpha)} k^{2/(\alpha-1)}. \quad (14)$$

The age constraint $1 \leq j \leq N$ translates into the degree constraints $k \in [pN^\alpha, pN^{(\alpha+1)/2}]$ for $0 \leq \alpha < 1$, and $k \in [pN^{(\alpha+1)/2}, pN^\alpha]$ for $\alpha > 1$. Via Eq. (13), the final continuous connection distribution under preferential attachment is then
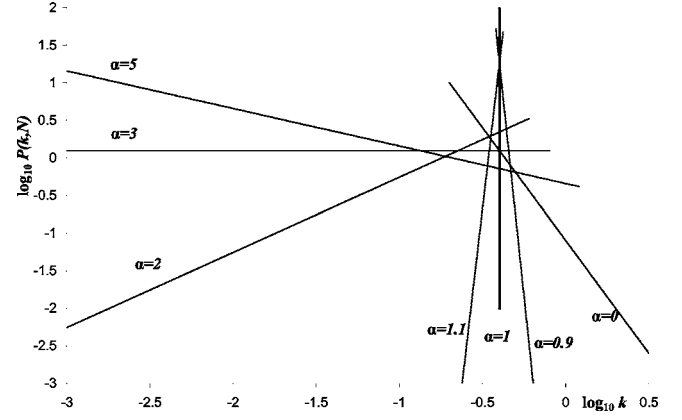

FIG. 2. The age-independent degree distribution $P(k,N)$ as a function of acceleration parameter α. The logarithmic plot means that straight lines indicate power-law distributions. The degree distributions are monotonically decreasing for $0 \leq \alpha < 1$ over the range $k \in [pN^\alpha, \infty)$, a δ function for $\alpha = 1$, and over the range $k \in [0, pN^\alpha]$, are monotonically increasing for $1 < \alpha < 3$, flat for $\alpha = 3$, and monotonically decreasing for $\alpha > 3$. Networks in this figure possess an average degree per node of $\langle k \rangle = 0.4$ at a network size of $N = 10\,000$.

$$P(k,N) = \begin{cases} \dfrac{2}{1-\alpha}(pN^\alpha)^{2/(1-\alpha)}\dfrac{1}{k^{(3-\alpha)/(1-\alpha)}}, & 0 \leq \alpha < 1, \\[2mm] \delta(k - pN), & \alpha = 1, \\[2mm] \dfrac{2}{\alpha-1}\dfrac{k^{(3-\alpha)/(\alpha-1)}}{(pN^\alpha)^{2/(\alpha-1)}}, & \alpha > 1. \end{cases}$$

(15)

Each of these separate distributions is normalized over $k \in [k_0, k_\infty]$ via $\int_{k_0}^{k_\infty} P(k,N) = 1$ where $k \in [pN^\alpha, \infty)$ for $0 \leq \alpha < 1$, $k \in [0,N]$ for $\alpha = 1$, and $k \in [0, pN^\alpha]$ for $\alpha > 1$. The average degree per node $\langle k \rangle = \int_{k_0}^{k_\infty} k P(k,N) = C/N$ as required over these same ranges. Again, the choice $p = 1$ and $0 \leq \alpha < 1$ duplicates the deterministic results of Ref. [15].

The resulting age-independent degree distributions $P(k,N)$ are shown in Fig. 2 and are monotonically decreasing for acceleration parameters $0 \leq \alpha < 1$, monotonically increasing for $1 < \alpha < 3$, uniformly flat for $\alpha = 3$, and monotonically decreasing for $\alpha > 3$. In all cases except $\alpha = 1$, the age-independent distribution $P(k,N)$ is a power law proportional to $k^{-\gamma}$ with $\gamma = (\alpha - 3)/(\alpha - 1)$. The exponent here can take the values $\gamma \in [3, \infty)$ for $\alpha \in [0,1)$ and $\gamma \in (-\infty, 1)$ for $\alpha \in (1, \infty)$, so no value of the acceleration parameter permits exponents in the range $\gamma \in [1,3]$.

All observed nodes have only a discrete number of connections, requiring calculation of a discrete degree distribution from the above continuous degree distribution. We consider how this is accomplished now. The continuous degree distributions $P(k,N)$ are typically defined over very limited ranges $k \in [k_0, k_\infty]$ so it can often be the case that every node of the network is predicted to have fractional degree less than unity. Of course, it is impossible to observe nodes with fractional degrees—every node either has zero connections or one connection, or two connections, and so on. When most
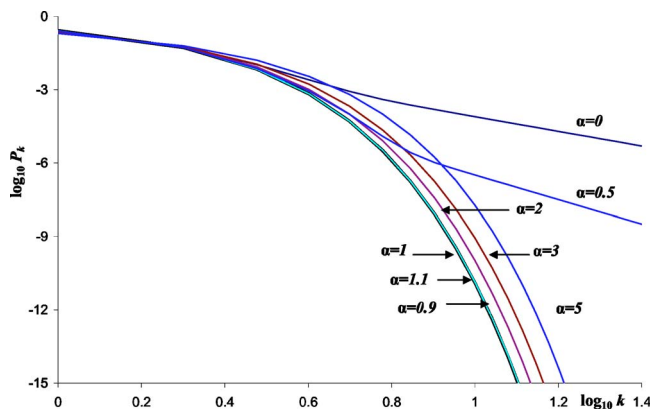
FIG. 3. (Color online) The age-independent discrete degree distribution P_k as a function of acceleration parameter α under preferential attachment. It is evident that while the unobservable continuous distributions $P(k,N)$ appear to be distinguishable, the observable discrete distributions are virtually indistinguishable at low degree numbers.

connected nodes have fractional degree, it is necessary to convert the unobservable continuous distribution $P(k,N)$ to an observable discrete distribution, which we denote P_k, giving the proportion of network nodes possessing an integral number of k links for $k \in [0,N]$. For networks with degree distributions which are long tailed, the continuous distribution $P(k,N)$ closely approximates the discrete distribution P_k and this step is often not required. In Appendix B, we detail how to calculate the observed integral degree distribution P_k generated from a continuous power-law degree distribution

$P(k,N)$ defined over a finite range $k \in [k_0,k_\infty]$. It is generally the case that finite ranges impose exponential cutoffs on scale-free continuous distributions $P(k,N)$ to give discrete distribution P_k possessing exponential statistics. A number of alternative mechanisms can impose an exponential cutoff on a power-law degree distribution [42–45], and we now add hyperacceleration to this list.

Appendix B specifies a numerical algorithm for calculating the observable discrete distribution P_k defined by the unobservable continuous distribution $P(k,N)$. The calculated discrete distributions equivalent to the continuous distributions of Fig. 2 appear in Fig. 3 showing that low acceleration networks with most nodes possessing on average more than one connection are observed to have power-law discrete degree distributions. In contrast, high acceleration networks have most nodes possessing on average only a fractional degree show an observed discrete degree distributions which is exponential in nature. In fact, the generated discrete distributions are virtually indistinguishable over wide ranges of the acceleration parameter in marked contrast to the very different shapes of the continuous distributions $P(k,N)$ at different acceleration parameters. This is due to the exponential cutoffs imposed when the continuous distributions are defined only over finite ranges. All of the calculations involved in the numerical algorithm of Appendix B can in principle be done by hand as we now illustrate using distributions $P(k,N)$ where the boundary point b lies within the range of support $k_0 < b < k_\infty$ for low acceleration regimes $0 \leq \alpha < 1$, but outside the range of support $k_0 < k_\infty < b$ for high acceleration regimes $\alpha \geq 1$. Using Appendix B, this choice gives

$$P_k^+ = \begin{cases} \left[\dfrac{pN^\alpha}{k-1/2} \right]^{2/(1-\alpha)} - \left[\dfrac{pN^\alpha}{k+1/2} \right]^{2/(1-\alpha)}, & 0 < \alpha < 1, \ k \in [2,\infty), \\ 0, & \alpha = 1, \ k \in [0,N], \\ 0, & \alpha > 1. \end{cases} \tag{16}$$

All of these distributions then have nonzero support to the left of the boundary b and so can be partitioned to give NP^- nodes labeled $0 \leq j \leq (NP^- - 1)$ each with average degree $k_j = p_j N$ equal to

$$k_j = \begin{cases} pN^\alpha \left(1 - \dfrac{j}{N} \right)^{(\alpha-1)/2}, & 0 < \alpha < 1, \\ pN, & \alpha = 1, \\ pN^{\alpha/2} j^{(\alpha-1)/2}, & \alpha > 1. \end{cases} \tag{17}$$

These values then feed directly into Eq. (B11) to give the total discrete distribution P_k. To further illustrate this last stage of the calculation, consider the case $\alpha = 1$ where we have $k_j = p_j N = pN$ and $p_j = p$ for all nodes $0 \leq j \leq N-1 \approx N$, and noting the absence of any contribution from the right of b so $P^+ = 0$, we have

$$P_k = P_k^- = \binom{N}{k} p^k (1-p)^{N-k}, \quad \alpha = 1, \tag{18}$$

for $k \in [0,N]$.

We now turn to consider the degree distribution of only that subset of nodes with at least one connection. As mentioned above, Fig. 1 makes it apparent that the actual degree of connected nodes generally decreases with node age even though average node degree can be monotonically increasing or decreasing. The reason is that the k_{jN} curves average connectivity over all nodes rather than only over connected nodes. (Recall that $\beta = 0$ means that nodes that are unconnected on entry to the network forever remain unconnected.) The probability that node n_j is connected is straightforwardly $\mathcal{P}_c(j) = 1 - \mathcal{P}(0,j)$ using Eq. (4) or
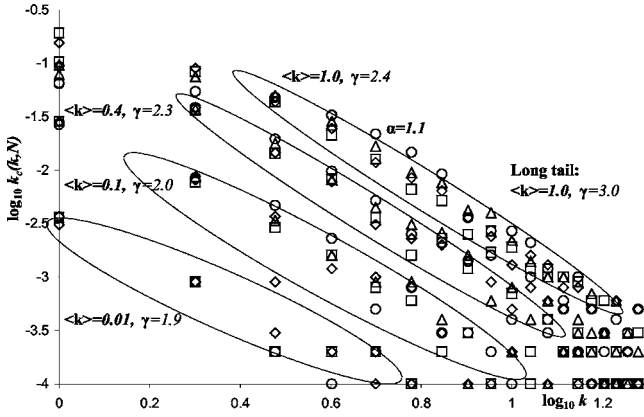
FIG. 4. The proportional degree distribution $k_c(k,N)$ for connected nodes under preferential attachment for various values of the average network degree $\langle k \rangle$ at a snapshot network size of $N = 10\,000$. The curves naturally group together according to average network degree, and are characterized by a power law $k^{-\gamma}$ with slope γ as shown. Further network growth would eventually convert these scale-free distributions to exponential distributions. For average degree $\langle k \rangle = 0.01$, we show curves for $(\alpha, p) = (0, 0.005), (1, 10^{-6}), (3, 2 \times 10^{-14}), (5, 3 \times 10^{-22})$; for $\langle k \rangle = 0.1$, $(\alpha, p) = (0, 0.05), (1, 10^{-5}), (3, 2 \times 10^{-13}), (5, 3 \times 10^{-21})$; for $\langle k \rangle = 0.4$, $(\alpha, p) = (0, 0.2), (1, 4 \times 10^{-5}), (3, 8 \times 10^{-13}), (5, 10^{-20})$; for $\langle k \rangle = 1.0$, $(\alpha, p) = (0, 0.5), (1, 10^{-4}), (3, 2 \times 10^{-12}), (5, 3 \times 10^{-20})$.

$$\mathcal{P}_c(j) = 1 - (1 - pj^{\alpha-1})^j \approx pj^{\alpha}. \qquad (19)$$

Hence, the average link number per node at node n_j [Eq. (12)] equates to the product of the average number of links per connected node at node n_j, denoted $k_c(j,N)$, and the probability that node n_j is connected, or $\mathcal{P}_c(j)$. By definition then, we have

$$k_{jN} = k_c(j,N)\mathcal{P}_c(j), \qquad (20)$$

giving

$$k_c(j,N) = \frac{pN^{(\alpha+1)/2}j^{(\alpha-1)/2}}{1 - (1 - pj^{\alpha-1})^j} \approx \left(\frac{N}{j}\right)^{(\alpha+1)/2}. \qquad (21)$$

Consequently, the average number of connections of connected nodes $k_c(j,N)$ is always a monotonically decreasing function of node age j. (An equivalent derivation appeared in Ref. [24].)

It is difficult to further analyze the degree distribution for actually connected nodes as the distribution of Eq. (21) is only incompletely sampled—node n_j is connected only with probability $1 - \mathcal{P}(0,j)$. Figure 4 shows age-independent degree statistics for simulated networks at a snapshot size of $N = 10\,000$ nodes and with different acceleration and probability parameters (α, p) chosen to generate networks over a range of average degrees $0.01 \leq \langle k \rangle \leq 1.0$. To do this, we choose different acceleration parameters in the set $\alpha \in \{0, 1, 3, 5\}$ with probability parameters set according to Eq. (3). The resulting slopes of the power-law degree distributions $k^{-\gamma}$ averaged over the corresponding acceleration parameters varies over the range $1.9 \leq \gamma \leq 3.0$. This regime was forbidden prior to taking account of the exponential cutoffs

imposed by the finite ranges. [The choice $(\alpha, p) = (1, 10^{-5})$ generating $\gamma \approx 2$ is roughly equivalent to the accelerating simulations of prokaryote gene regulatory networks [24].] Additional network growth would see these scale-free distributions undergoing a full transition to exponential statistics.

Finally, we briefly consider the size of the largest connected island in the artificial case of purely preferentially attached networks where initially unconnected nodes remain always unconnected. In this particular case, preferential attachment ensures that every connected node belongs to the same interconnected largest island whose size then equates to the number of connected nodes. As a node is connected only if it gains some connections on entry to the network with probability $1 - \mathcal{P}(0,i)$ from Eq. (4), the size s of the largest island is

$$s = \sum_{i=1}^{N} [1 - (1 - pi^{\alpha-1})^i]. \qquad (22)$$

For small p and small network sizes N, this equates to the number of links $s \approx \sum_{i=1}^{N} pi^{\alpha} \approx L$ as each connected node likely has only one link in sparsely connected networks. In this regime, the largest island grows at an acceleration rate determined by the acceleration parameter α. However, outside this regime, the growth of the largest island eventually saturates when almost every new node is connected [i.e., when $1 - (1 - pi^{\alpha-1})^i \approx 1$], meaning the largest island grows proportionally to network size N. This occurs in the limit $pi^{\alpha-1} \rightarrow 1$ or N large giving $s \approx N$. Should a network absolutely need to maintain an accelerated growth in the size of its largest island, as conjectured for the regulatory gene network of prokaryotes due to competitive pressures, then the point at which the largest island makes a transition from accelerated to linear growth will represent an upper network size limit [24–26].

IV. RANDOMLY ATTACHED ACCELERATING NETWORKS

We now turn to consider accelerating and hyperaccelerating networks ($\alpha \geq 0$) where new links are randomly distributed over all existing nodes. The final generated link distribution will depend on the balance of two countervailing trends. The first sees newer nodes attracting more links due to acceleration which confers an increasing average number of links on younger nodes, while in contrast, older nodes have a longer time to acquire links from the newer nodes.

We now calculate the degree distribution under random attachment, where the rate of growth of the link number for node n_j when the network has grown to size i is given under the continuous approximation [4,48,49] as

$$\frac{\partial k_{ji}}{\partial i} = \frac{k_{ii}}{i} = pi^{\alpha-1}, \qquad (23)$$

where $k_{ii} = pi^{\alpha}$. Here, the rate of growth in the degree k_{ji} is determined by the number of new links added with node n_i (k_{ii}) which are equally proportioned over the i existing nodes. The resulting degree statistics can readily be obtained

through integrating this equation with the initial conditions $k_{jj} \approx pj^{\alpha}$ at time j to obtain

$$k_{jN} = \begin{cases} p\left[1 + \ln\dfrac{N}{j}\right], & \alpha = 0, \\ \dfrac{p}{\alpha}[N^{\alpha} + (\alpha - 1)j^{\alpha}], & \alpha > 0. \end{cases} \qquad (24)$$

These results satisfy Eq. (23), while integration of the link numbers over all node numbers j gives the required total number of links as in Eq. (2). [As established in Appendix A, these random attachment equations are exactly reproduced by applying a Taylor series expansion of the general equations Eq. (11) about the point $x = pN^{\alpha}/\beta \approx 0$ with the retention only of terms linear in x.] For interest, Appendix C provides the exact distribution for links under random attachment.

The resulting degree k_{jN} versus node age j distributions are shown in Fig. 5 along with example simulation runs at different acceleration parameters α. This figure makes clear that the degree distribution k_{jN} is monotonically decreasing with j for $0 \leq \alpha < 1$ so younger nodes are more heavily connected than older nodes on average, and monotonically increasing for $\alpha > 1$ so younger nodes are less heavily connected than older nodes on average. It is only on the boundary between these two regions at $\alpha = 1$ that the degree distribution is flat so average node degree is independent of node age.

As previously, observed statistics normally lack age or node number information, and we now follow standard practice in calculating the age independent degree distribution. The k_{jN} distribution contains information about both node degree and node age and so approximates network statistics (simulated or observed) when all of this information is available. We now calculate the age-independent final link distribution $P(k, N)$ as in Eq. (13). In the case $\alpha = 1$ we have $j \neq j(k)$, meaning the node degree k is independent of the node age j, so the δ function of Eq. (13) again integrates to give $\delta(k - pN)$. For other cases, we have

$$j = \begin{cases} Ne^{(1-k/p)}, & \alpha = 0, \\ \left[\dfrac{1}{1-\alpha}\left(N^{\alpha} - \dfrac{\alpha k}{p}\right)\right]^{1/\alpha}, & 0 < \alpha, \ \alpha \neq 1. \end{cases} \qquad (25)$$

The age constraint $1 \leq j \leq N$ translates into the degree constraints $k \in [p, p(1 + \ln N)]$ for $\alpha = 0$, $k \in [pN^{\alpha}, pN^{\alpha}/\alpha]$ for $0 < \alpha < 1$, $k = pN$ for $\alpha = 1$, and $k \in [pN^{\alpha}/\alpha, pN^{\alpha}]$ for $\alpha > 1$. Via Eq. (13), these relations determine the predicted age-independent final link distribution $P(k, N)$ under random attachment to be
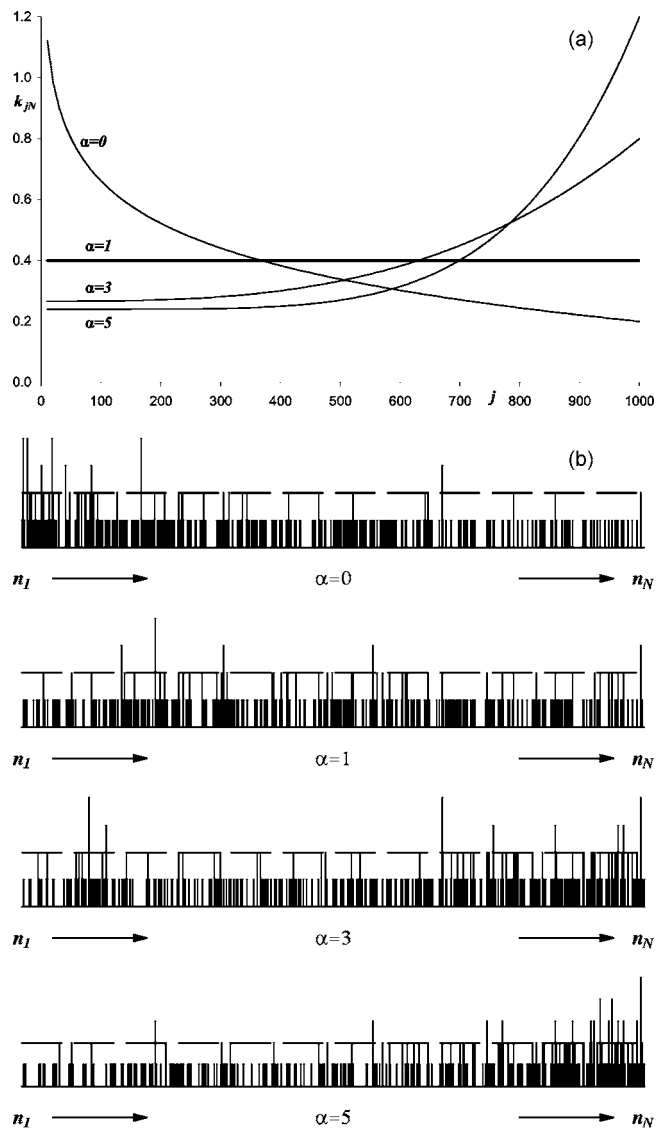

FIG. 5. (a) The average degree distribution k_{jN} under random attachment as a function of node age j for various values of the acceleration parameter α. The degree distribution is monotonically decreasing for $0 \leq \alpha < 1$, flat for $\alpha = 1$, and monotonically increasing for $\alpha > 1$. (b) Example simulated networks for various acceleration parameters α with nodes listed sequentially left to right from n_1 to n_N with the degree of each node indicated by vertical lines. The horizontal dashed line indicates an average degree of 2.

$$P(k, N) = \begin{cases} \dfrac{1}{p}e^{(1-k/p)}, & \alpha = 0, \\ \dfrac{1}{pN(1-\alpha)^{1/\alpha}}\left(N^{\alpha} - \dfrac{\alpha k}{p}\right)^{(1-\alpha)/\alpha}, & 0 \leq \alpha < 1, \\ \delta(k - pN), & \alpha = 1, \\ \dfrac{1}{pN(\alpha - 1)^{1/\alpha}}\dfrac{1}{(\alpha k/p - N^{\alpha})^{(\alpha-1)/\alpha}}, & \alpha > 1. \end{cases}$$
$$(26)$$

Each of these separate distributions is normalized over $k \in [k_0, k_{\infty})$ via $\int_{k_0}^{k_{\infty}} P(k, N) = 1$ where $k \in [p, \infty)$ for $\alpha = 0$,
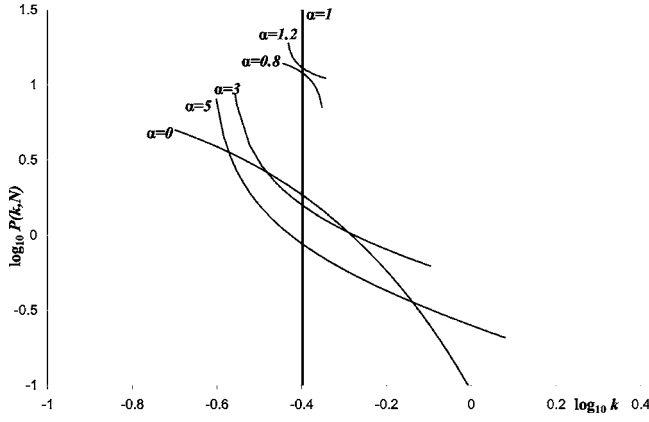
FIG. 6. The age-independent degree distribution $P(k,N)$ as a function of acceleration parameter α for randomly connected networks. The distributions are monotonically decreasing for $\alpha \neq 1$ over the respective ranges $k \in [p,\infty)$ for $\alpha=0$, $k \in [pN^\alpha, pN^\alpha/\alpha]$ for $0 < \alpha < 1$, and $k \in [pN^\alpha/\alpha, pN^\alpha]$ for $\alpha > 1$.

FIG. 7. (Color online) The age-independent discrete degree distribution P_k as a function of acceleration parameter α for random attachment. It is evident that while the unobservable continuous distributions $P(k,N)$ are distinguishable, the observable discrete distributions are virtually indistinguishable. We also show the actual points for different network simulations of $N=10\,000$ nodes with average degree $\langle k \rangle = 0.4$ with $\alpha=0$ (triangles), 1 (diamonds), 3 (circles), and 5 (squares).

$k \in [pN^\alpha, pN^\alpha/\alpha]$ for $0 < \alpha < 1$, $k = pN$ for $\alpha = 1$, and $k \in [pN^\alpha/\alpha, pN^\alpha]$ for $\alpha > 1$. The average degree per node $\langle k \rangle = \int_{k_0}^{k_\infty} k P(k,N) = C/N$ as required over these same ranges. As shown in Fig. 6, the predicted age-independent continuous degree distribution $P(k,N)$ is monotonically decreasing as a function of k for both $\alpha < 1$ and $\alpha > 1$, and a δ function for $\alpha = 1$.

Again, all observed nodes have only a discrete number of connections, requiring that calculation of a discrete degree distribution from the above continuous degree distribution. The continuous degree distributions $P(k,N)$ are again typically defined over very limited ranges $k \in [k_0, k_\infty]$ for acceleration parameters greater than zero so many nodes possess fractional degree less than unity. This again requires that the continuous distribution $P(k,N)$ be converted into an equivalent discrete distribution P_k following the methods of Appendix B. Figure 7 shows the results of taking full account of the finite ranges of the continuous distributions $P(k,N)$ which
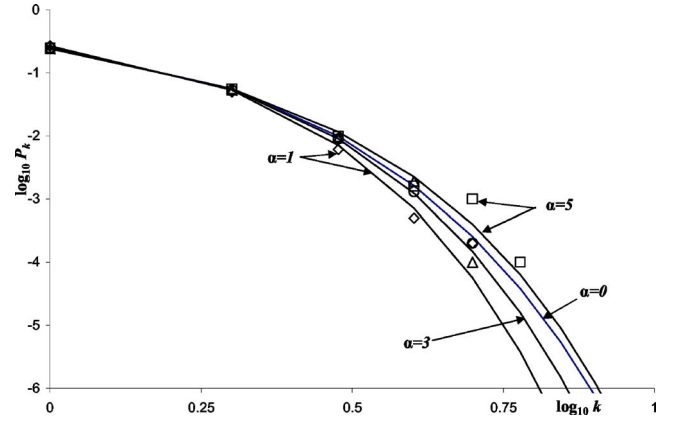
impose exponential cutoffs on the generated discrete distribution P_k. It is clear that the exponential cutoffs render the observable discrete distributions for different acceleration parameters essentially indistinguishable. This result is borne out by simulations of networks with $N=10\,000$ nodes with average degree $\langle k \rangle = 0.4$ for different acceleration parameters (also shown in Fig. 7).

Again, we illustrate the numerical algorithm of Appendix B via specific examples for randomly attached networks. Consider continuous distributions $P(k,N)$ where the boundary point b lies within the range of support for the distribution for all acceleration parameters except $\alpha=1$. That is, we assume $pN^\alpha < b < pN^\alpha/\alpha$ for $0 \leq \alpha < 1$ and $pN^\alpha/\alpha < b < pN^\alpha$ for $\alpha > 1$, while for $\alpha=1$ we assume $pN < b$. These illustrative choices then give

$$P_k^+ = \begin{cases} 2\sinh\left(\dfrac{1}{2p}\right)e^{(1-k/p)}, & \alpha = 0, \\[2ex] \left[\dfrac{1-(\alpha/pN^\alpha)(k-1/2)}{1-\alpha}\right]^{1/\alpha} - \left[\dfrac{1-(\alpha/pN^\alpha)(k+1/2)}{1-\alpha}\right]^{1/\alpha}, & 0 < \alpha < 1, \\[2ex] 0, & \alpha = 1, \\[2ex] \left[\dfrac{(\alpha/pN^\alpha)(k+1/2)-1}{\alpha-1}\right]^{1/\alpha} - \left[\dfrac{(\alpha/pN^\alpha)(k-1/2)-1}{\alpha-1}\right]^{1/\alpha}, & \alpha > 1. \end{cases} \tag{27}$$

In this example, the allowed ranges for the degree k are $k \in [2,\infty)$ for $\alpha=0$, $k \in [2, pN^\alpha/\alpha]$ for $0 < \alpha < 1$, and $k \in [2, pN^\alpha]$ for $\alpha > 1$. Our example distributions have nonzero support to the left of the boundary point b and so can be partitioned to give NP^- nodes labeled $0 \leq j \leq (NP^- - 1)$ each with average degree $k_j = p_j N$ equal to

$$
k_j = \begin{cases}
p\left[1 - \ln\left(1 - \dfrac{j}{N}\right)\right], & \alpha = 0, \\[2ex]
\dfrac{pN^\alpha}{\alpha}\left[1 - (1-\alpha)\left(1 - \dfrac{j}{N}\right)^\alpha\right], & 0 < \alpha < 1, \\[2ex]
pN, & \alpha = 1, \\[2ex]
\dfrac{pN^\alpha}{\alpha}\left[1 + (\alpha - 1)\left(\dfrac{j}{N}\right)^\alpha\right], & \alpha > 1.
\end{cases}
\tag{28}
$$

These values then feed directly into Eq. (B11) to give the total discrete distribution P_k. To illustrate this last stage of the calculation, consider the case $\alpha = 1$ where we have $k_j = p_j N = pN$ and $p_j = p$ for all nodes $0 \leq j \leq N-1 \approx N$, and noting the absence of any contribution from the right of b, so $P^+ = 0$, we have

$$
P_k = P_k^- = \binom{N}{k} p^k (1-p)^{N-k}, \quad \alpha = 1,
\tag{29}
$$

for $k \in [0, N]$.

We now turn to establish how the formation of a single giant connected component depends on the acceleration parameter α. The number of nodes in the largest island s will be a function of network size and will grow as new nodes and new connections are added to the network. The largest island will grow by one when the new node n_i forms one link with probability $\mathcal{P}(1, i)$ which attaches to the largest island with probability s/i, and will grow by the average size of all smaller islands $\bar{s}$ when node n_i forms two links with probability $\mathcal{P}(2, i)$ either of which connects to the largest island with probability s/i while the remaining link connects to an external island with probability $(i-s)/i$. Altogether, the rate of growth of the largest island varies as

$$
\frac{ds}{di} = \frac{s}{i}\mathcal{P}(1, i) + 2\bar{s}\frac{(i-s)s}{i^2}\mathcal{P}(2, i) \approx p i^{\alpha-1}s + p^2 \bar{s} i^{2\alpha-2}(i-s)s,
\tag{30}
$$

using Eq. (4) and noting the restriction $p i^{\alpha-1} \leq 1$. Numerical simulations (and some analytically tractable solutions) indicate that the first term here is negligible compared to the second, and hereinafter we ignore this first term. For initial conditions, we have $s(n_c) = 2$ at node n_c where the first link likely appears (see Appendix B). Figure 8 shows the predicted growth of the largest island compared to simulated networks for different acceleration parameters α, where all networks possess an average degree per node of $\langle k \rangle = 0.4$ at a maximum simulated network size of $N = 1000$, and we assume an average size of smaller islands of $\bar{s} = 7.5$. There is an evident close connection between theoretically predicted curves and observed statistics.

V. CONCLUSION

We have examined the network structural transitions displayed by accelerating and hyperaccelerating probabilistic networks motivated by the observation that important biological networks such as prokaryote gene regulatory networks are both hyperaccelerating and size limited due to network structural transitions. We examined accelerating growing networks of nodes connected by undirected links which were probabilistically added with each new node and either preferentially or randomly attached to existing nodes. The addition of a probabilistic envelope allowing the number of new links added with each new node to be an integer greater than or equal to zero allowed us to extend network theory to model sparsely connected networks where the majority of nodes are entirely unconnected (e.g., prokaryote gene regulatory networks). Consequently, the probabilistic envelope also allowed us to lift existing constraints on the modeling of accelerating networks, allowing us to treat decelerating, accelerating, and hyperaccelerating networks. These latter two classes of networks were shown to be subject to transitions in which either a single giant connected component forms or the network condenses into a fully connected state with exponential statistics. We were able to roughly locate these transitions as a function of the acceleration and probability parameters and the network size. Mean field approximations were compared to network simulations over a wide range of parameters and shown to be consistent.

The present paper uses mean field network theory to model rapidly accelerating networks consisting of many un-
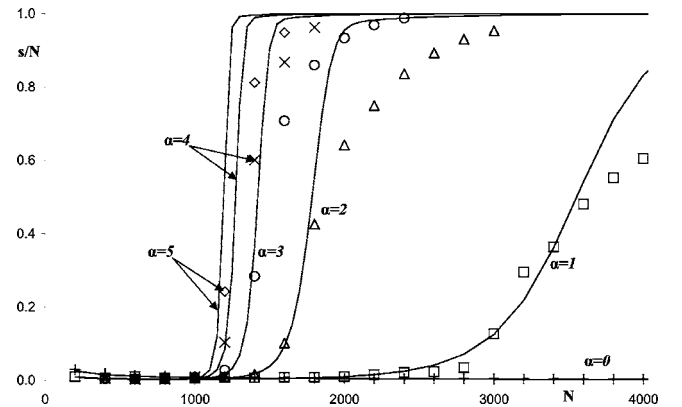

FIG. 8. The simulated and predicted size of the largest island s expressed as a proportion of growing network size for various acceleration parameters α. Each network has average degree $\langle k \rangle = 0.4$ at a maximum simulated network size $N_c = 1000$ nodes, and in every case, we assume that the average size of smaller islands is $\bar{s} = 7.5$.

connected nodes allowing the examination of the statistical transitions generated under accelerated growth. As such, these extended models will be useful in treating models of, for instance, accelerating biological regulatory networks mainly consisting of unconnected nodes and displaying network transitions which limit size and complexity. Such models are required to explain the observed evolutionary record of prokaryote gene regulatory networks.

APPENDIX A: RELATIVE WEIGHTS OF PREFERENTIAL AND RANDOM ATTACHMENT

In analyzing Eq. (11), it is first necessary to determine appropriate limits on β giving access to regimes where preferential attachment dominates random attachment for accelerating networks.

Appropriate limits on β can be obtained by performing a Taylor expansion of Eq. (11) about the point $x = pN^\alpha/\beta \approx 0$ while retaining only terms linear in x. Using $\alpha = 0$ in Eq. (11), we have

$$k_{jN} = \frac{p}{x}\left[(1+x)\left(\frac{N}{j}\right)^x - 1\right]. \tag{A1}$$

Noting $d/dx[(1+x)a^x] = a^x[1+(1+x)\ln a]$, the leading terms of a Taylor expansion immediately reproduce the $\alpha = 0$ result for the random attachment model of Eq. (24). Similarly, using $\alpha > 0$ in Eq. (11) gives

$$k_{jN} = \beta\left\{\left[1+x\left(\frac{j}{N}\right)^\alpha\right]\left[\frac{[2/(\alpha+1)]x^\alpha + 1}{[2/(\alpha+1)]x(j/N)^\alpha + 1}\right]^{(\alpha+1)/2\alpha} - 1\right\}. \tag{A2}$$

A straightforward differentiation then gives a Taylor expansion whose leading terms exactly equal the $\alpha > 0$ result for the random attachment model of Eq. (24). Hence, random attachment entirely dominates when $pN^\alpha/\beta \ll 1$ or equivalently when

$$\beta \gg \beta_{\text{random}} = pN^\alpha = \frac{1}{2}(1+\alpha)\langle k \rangle. \tag{A3}$$

For values of $\beta < \beta_{\text{random}}$, preferential attachment will influence the final distribution. For sufficiently small β, preferential attachment will dominate (rather than merely contribute) and this latter boundary can be located as follows. The first connected node, denoted n_c (i.e., the cth node), likely appears when the cumulative average number of added initial links sums roughly to unity,

$$\sum_{i=1}^{c} pi^\alpha \approx \frac{pc^{\alpha+1}}{\alpha+1} \approx 1. \tag{A4}$$

Now the floating end of the new link attached to node n_c can be either randomly attached to one of the nodes n_1 to n_c with probability proportional to $c\beta$, or preferentially attached to node n_c with probability proportional to $1+\beta \approx 1$ in the preferential attachment regime. For preferential attachment to dominate, we require $c\beta \ll 1$, or equivalently,
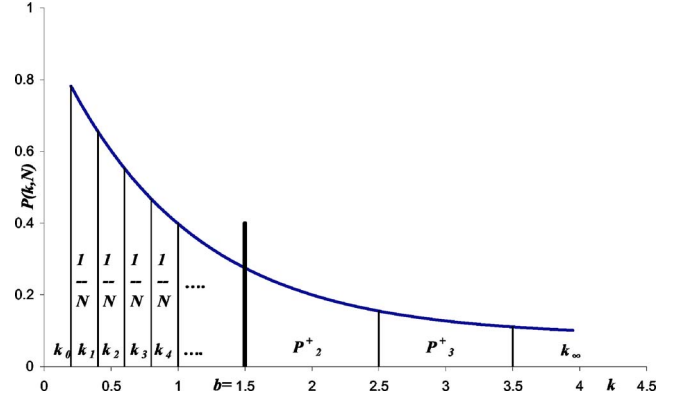

FIG. 9. (Color online) A continuous distribution $P(k,N)$ defined over the finite range $[k_0, k_\infty]$ generates an observable discrete distribution P_k for $k = 0, 1, 2, \dots$. The $P(k,N)$ distribution lying to the left of the boundary $b = 3/2$ represents network nodes with fractional average degree distributions, and is partitioned into regions each of area $1/N$ with boundaries $[k_j, k_{j+1})$ for $j \geq 0$. The $P(k,N)$ distribution lying to the right of the boundary $b = 3/2$ represents network nodes with integral average degree, and is partitioned into regions as shown.

$$\beta \ll \beta_{\text{pref}} = \frac{1}{c} = \left[\frac{p}{\alpha+1}\right]^{1/(\alpha+1)} = \left[\frac{\langle k \rangle}{2N_c^\alpha}\right]^{1/(\alpha+1)}. \tag{A5}$$

For the typical hyperaccelerating networks considered here, this last constraint ensures that preferential attachment dominates only when β is so small as to be essentially zero, and in this paper, we analyze preferential attachment using the setting $\beta = 0$.

APPENDIX B: EXPONENTIAL CUTOFFS IN HYPERACCELERATING NETWORKS

An arbitrary continuous network degree distribution $P(k,N)$ defined over a finite range $[k_0, k_\infty]$ is essentially unobservable, and in actuality, generates an equivalent discrete degree distribution P_k with $k \in [0, N]$ which can be compared to observed statistics. Here, we show how to calculate the observable discrete distribution from the theoretically predicted but unobservable continuous distribution.

The first step is to partition the continuous distribution $P(k,N)$ into two parts at an arbitrarily chosen boundary point b of order unity. We choose $b = 3/2$ as shown in Fig. 9. Then, that proportion of the continuous distribution $P(k,N)$ lying to the right of b can be considered to contribute to the long tail of the observed discrete distribution P_k in the normal way. That is, the proportion of the continuous distribution in the region $[k-1/2, k+1/2)$ for integral k equal to

$$P_k^+ = \int_{k-1/2}^{k+1/2} P(k,N)dk \tag{B1}$$

can be entirely assigned to the discrete distribution bin P_k. The total proportion of the continuous distribution $P(k,N)$ so assigned is

$$P^+ = \sum_{k=b+1/2}^{k_\infty} P_k^+ = \int_b^{k_\infty} P(k,N)dk, \qquad (B2)$$

and this proportion of the distribution describes the degree distribution of a total of NP^+ nodes. When this proportion is close to unity as usually applies for long tailed distributions, nothing further need be done and the distribution bins $P_k^+ = P_k$ then equate to the predicted discrete observable distribution.

However, when P^+ is significantly less than unity as applies for hyperaccelerating networks, it is necessary to assign the remaining proportion of the continuous distribution lying to the left of the chosen value b to the discrete distribution bins P_k. The proportion of the continuous distribution remaining to be assigned is

$$P^- = \int_{k_0}^b P(k,N)dk = 1 - P^+, \qquad (B3)$$

with this proportion of the distribution describing the degree distribution of a total of NP^- nodes. The best way to understand how this remaining assignment is done is through a successive partitioning of the usual normalization constraint,

$$1 = \int_{k_0}^{k_\infty} P(k,N)dk$$

$$= \int_{k_0}^b P(k,N)dk + \int_b^{k_\infty} P(k,N)dk$$

$$= \sum_{j=0}^{(NP^--1)} \int_{k_j}^{k_{j+1}} P(k,N)dk + P^+. \qquad (B4)$$

Here, the continuous distribution to the left of the point $k=b$ is partitioned into NP^- regions with respective boundaries $[k_j, k_{j+1})$ for $k_j \in [k_0, b]$ and $k_j < k_{j+1}$ while as noted above, that to the right of $k=b$ is partitioned into potentially an infinite number of regions with boundaries $[k-1/2, k+1/2)$ for integral $k>b$. The second contribution to the discrete bin P_k comes from the NP^- regions of the continuous distribution lying to the left of b. We would like each of these NP^- regions to correspond to a single node of the network with nonintegral average degree, and achieve this by choosing the regional boundaries k_j such that

$$\int_{k_0}^{k_j} P(k,N)dk = \frac{j}{N}, \quad j \in [0, NP^- - 1]. \qquad (B5)$$

Here, each region defines an area of exactly $1/N$ so j regions have a combined area of j/N (see Fig. 9). The average degree of the node corresponding to the jth region is approximately

$$\langle k_j \rangle = k_j \qquad (B6)$$

and is typically nonintegral. (Other alternatives such as $\langle k_j \rangle = (k_j + k_{j-1})/2$ might be chosen but the difference is negligible for large N.) As any given node can, in practice, only possess an integral number of connections, these nonintegral average node degrees require that the link numbers for these

nodes be assigned probabilistically over the range $k \in [0,N]$ so as to give the required average. This is achieved by assigning an integral number k_j of links to this region with $k_j \in [0,N]$ according to the probability distribution

$$\mathcal{P}_r(k_j) = \binom{N}{k_j} p_j^{k_j}(1-p_j)^{N-k_j}, \qquad (B7)$$

with probabilities p_j chosen to satisfy

$$p_j = \frac{\langle k_j \rangle}{N}, \qquad (B8)$$

to ensure that this distribution has average $\langle k_j \rangle = p_j N = \langle k_j \rangle$. We note that later results will not be overly dependent on whether the $\mathcal{P}_r$ distribution chosen here is binomial in form, or some other reasonable distribution. Summing over all the NP^- regions to the left of b, the proportion of nodes possessing an integral number of $k_j = k$ links is then

$$P_k^- = \frac{1}{N} \sum_{j=0}^{(NP^--1)} \binom{N}{k} p_j^k (1-p_j)^{N-k} \qquad (B9)$$

for $k \in [0,N]$ and zero otherwise. This is normalized according to $\Sigma_{k=0}^N P_k^- = P^-$. Consequently, the total proportion of the network possessing a discrete number of k links is expected to be

$$P_k = P_k^- + P_k^+. \qquad (B10)$$

As required, this distribution is normalized and has average $\langle k \rangle = L/N$. This gives the discrete probability that a node has k links for $k \in [0,\infty)$. It is also common to calculate the related probability that connected nodes have k links and this is given by $P_k' = P_k/(1-P_0)$ for $k \geq 1$.

In summary, for the general case of arbitrary continuous distributions $P(k,N)$ defined over the range $k \in [k_0, k_\infty]$, the predicted discrete distribution is

$$P_k = P_k^- + P_k^+ = \frac{1}{N} \sum_{j=0}^{NP^--1} \binom{N}{k} p_j^k (1-p_j)^{N-k} + \int_{k-1/2}^{k+1/2} P(k,N)dk, \qquad (B11)$$

where $P^- = \int_{k_0}^b P(k,N)dk$, $p_j = k_j/N$, k_j is chosen to satisfy $j/N = \int_{k_0}^{k_j} P(k,N)dk$, and the partition point b is arbitrarily chosen to be close to unity.

APPENDIX C: EXACT RANDOM ATTACHMENT LINKAGE DISTRIBUTION

We here present the exact distribution for links under random attachment. First, note that node n_i initially receives an average of $\langle j \rangle = pi^\alpha$ links where $j \in [0,i]$ with each link formed with probability $pi^{\alpha-1}$, and subsequently receives an average of $\langle j_i \rangle = pi^\alpha/i = pi^{\alpha-1}$ links from itself where $j_i \in [0,i]$ with each link formed with probability $pi^{\alpha-1}$. The arrival of node n_{i+1} is accompanied by an additional $p(i+1)^\alpha$ new links of which an average $\langle j_{i+1} \rangle = p(i+1)^\alpha/(i+1) = p(i+1)^{\alpha-1}$ attach to node n_i where $j_{i+1} \in [0,i+1]$ with each

link formed with probability $p(i+1)^{\alpha-2}$, and so on, until it receives an average of $\langle j_N \rangle = pN^{\alpha}/N = pN^{\alpha-1}$ links from node n_N with $j_N \in [0, N]$ with each link formed with probability $pN^{\alpha-2}$. The joint probability that node n_i obtains $j, j_i, \ldots, j_N$ connections is then

$$\mathcal{P}_i(j, j_i, j_{i+1}, \ldots, j_N) = \binom{i}{j} [pi^{\alpha-1}]^j [1 - pi^{\alpha-1}]^{i-j} \prod_{n=i}^{N} \binom{n}{j_n}$$
$$\times [pn^{\alpha-2}]^{j_n} [1 - pn^{\alpha-2}]^{n-j_n}. \qquad (C1)$$

The average number of inbound links for node n_i is then

$$\langle j + j_i + \cdots + j_N \rangle = pi^{\alpha} + \sum_{n=i}^{N} pn^{\alpha-1}, \qquad (C2)$$

which integrates to give Eq. (24) as found by the continuum approach. Unfortunately, further work with this more exact multivariate distribution is generally intractable.

[1] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).
[2] R. Albert and A. L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[3] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[4] S. N. Dorogovtsev and J. F. F. Mendes, Phys. Rev. E **63**, 056125 (2001).
[5] M. Faloutsos, P. Faloutsos, and C. Faloutsos, in *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, edited by L. Chapin, J. P. G. Sterbenz, G. Parulkar, and J. S. Turner (ACM Press, New York, 1999), pp. 251–262.
[6] S. N. Dorogovtsev and J. F. F. Mendes, Phys. Rev. E **63**, 025101(R) (2001).
[7] A. Vázquez, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. E **65**, 066130 (2002).
[8] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, Nature (London) **407**, 651 (2000).
[9] S. N. Dorogovtsev and J. F. F. Mendes, Europhys. Lett. **52**, 33 (2000).
[10] A. Vázquez, e-print cond-mat/0006132.
[11] M. E. J. Newman, Phys. Rev. E **64**, 016131 (2001); **64**, 016132 (2001).
[12] A. L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek, Physica A **311**, 590 (2002).
[13] A. Vázquez, Phys. Rev. E **67**, 056104 (2003).
[14] S. N. Dorogovtsev and J. F. F. Mendes, Proc. R. Soc. London, Ser. B **268**, 2603 (2001).
[15] P. Sen, Phys. Rev. E **69**, 046107 (2004).
[16] P. Sen, Physica A **346**, 139 (2005).
[17] I. Ispolatov, P. L. Krapivsky, and A. Yuryev, e-print q-bio.MN/0411052.
[18] P. L. Krapivsky and S. Redner, e-print cond-mat/0410379.
[19] I. Ispolatov, P. L. Krapivsky, I. Mazo, and A. Yuryev, e-print q-bio.MN/0502005.
[20] R. Albert and A. L. Barabási, Phys. Rev. Lett. **85**, 5234 (2000).
[21] P. L. Krapivsky, G. J. Rodgers, and S. Redner, Phys. Rev. Lett. **86**, 5401 (2001).
[22] Z. Liu, Y.-C. Lai, and N. Ye, Phys. Rev. E **66**, 036112 (2002).
[23] K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. **88**, 108701 (2002).

[24] M. J. Gagen and J. S. Mattick, e-print q-bio.MN/0312021.
[25] M. J. Gagen and J. S. Mattick, e-print q-bio.MN/0312022.
[26] M. J. Gagen and J. S. Mattick, Theory Biosci. **123**, 381 (2005).
[27] C. K. Stover *et al.*, Nature (London) **406**, 959 (2000).
[28] S. D. Bentley *et al.*, Nature (London) **417**, 141 (2002).
[29] E. van Nimwegen, Trends Genet. **19**, 479 (2003).
[30] I. Cases, V. de Lorenzo, and C. A. Ouzounis, Trends Microbiol. **11**, 248 (2003).
[31] L. J. Croft, M. J. Lercher, M. J. Gagen, and J. S. Mattick, e-print q-bio.MN/0311021.
[32] I. Yanai, C. J. Camacho, and C. DeLisi, Phys. Rev. Lett. **85**, 2641 (2000).
[33] J. Qian, N. M. Luscombe, and M. Gerstein, J. Mol. Biol. **313**, 673 (2001).
[34] A. Wagner, Proc. R. Soc. London, Ser. B **270**, 457 (2002).
[35] E. Eisenberg and E. Y. Levanon, Phys. Rev. Lett. **91**, 138701 (2003).
[36] R. Pastor Satorras, E. Smith, and R. Sole, J. Mol. Biol. **222**, 199 (2003).
[37] L. Giot *et al.*, Science **203**, 1727 (2003).
[38] V. Kunin and C. A. Ouzounis, Genome Res. **13**, 1589 (2003).
[39] M. Madan Babu and S. A. Teichmann, Nucleic Acids Res. **31**, 1234 (2003).
[40] A. Bhan, D. J. Galas, and T. G. Dewey, Bioinformatics **18**, 1486 (2002).
[41] J. S. Mattick and M. J. Gagen, Science **307**, 856 (2005).
[42] S. N. Dorogovtsev and J. F. F. Mendes, Phys. Rev. E **62**, 1842 (2000).
[43] A. A. Moreira, J. S. Andrade, Jr., and L. A. Nunes Amaral, Phys. Rev. Lett. **89**, 268703 (2002).
[44] T. Fenner, M. Levene, and G. Loizou, e-print cond-mat/0209463.
[45] S. Mossa, M. Barthélémy, H. E. Stanley, and L. A. Nunes Amaral, Phys. Rev. Lett. **88**, 138701 (2002).
[46] H. Štefančić and V. Zlatić, e-print cond-mat/0404495.
[47] P. Erdös and A. Renyi, Publ. Math. Inst. Hungarian Acad. Sci. **5**, 17 (1960).
[48] A. L. Barabási, R. Albert, and H. Jeong, Physica A **272**, 173 (1999).
[49] A. L. Barabási and R. Albert, Science **286**, 509 (1999).