Addition-Deletion Networks

E. Ben-Naim[1,*] and P. L. Krapivsky[2,†]

[1] *Theoretical Division and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*

[2] *Department of Physics, Boston University, Boston, Massachusetts 02215, USA*

We study structural properties of growing networks where both addition and deletion of nodes are possible. Our model network evolves via two independent processes. With rate r, a node is added to the system and this node links to a randomly selected existing node. With rate 1, a randomly selected node is deleted, and its parent node inherits the links of its immediate descendants. We show that the in-component size distribution decays algebraically, $c_k \sim k^{-\beta}$, as $k \to \infty$. The exponent $\beta = 2 + (r-1)^{-1}$ varies continuously with the addition rate r. Structural properties of the network including the height distribution, the diameter of the network, the average distance between two nodes, and the fraction of dangling nodes are also obtained analytically. Interestingly, the deletion process leads to a giant hub, a single node with a macroscopic degree whereas all other nodes have a microscopic degree.

PACS numbers: 89.75,Hc, 05.40.-a,05.20.Dd, 02.50.Ey

I. INTRODUCTION

Idealized models of random networks provide an important tool for characterizing real-world networks such as communication, infrastructure, and social networks [1–3]. For example, the emergence of hubs in heterogeneous networks can be reproduced through a "rich-gets-richer" mechanism. This phenomenon arises in networks that grow via preferential attachment [4–6].

Most network growth models are based upon sequential addition of nodes and subsequent attachment to existing nodes according to a prescribed mechanism. This framework does not allow for removal of nodes. In reality, there are two types of networks. One type of networks including for example citation networks may only expand, while another type of networks such as friendship networks may either expand or contract. Indeed, individuals may join or leave a social group [7, 8]. The same is true for technological networks such as the world-wide-web because websites may disappear.

There is substantial theoretical understanding of the structure of strictly expanding networks, but much less is known about networks that may also contract [7–13]. In the latter case, it is typically impossible to construct closed equations for quantities such as the degree distribution because removal of nodes generates memory or correlations.

In this paper, we introduce a simple network growth process with addition and deletion of nodes and show that it is possible to obtain closed equations for several structural properties of the network. In our growing network model, nodes may be added or deleted. When a node is added to the network, it is attached to a randomly selected existing node. When a node is deleted, its daughters are attached to its parent. This process is relevant for phylogenetic trees, namely trees that document the ancestries of distinct biological species [14–16] because in the process of evolution, mutation events generate new nodes while extinction events result in deletion of existing nodes [17–21].

We first investigate the in-component size distribution and show that this quantity obeys a closed evolution equation. We analytically obtain many characteristics of this distribution including the generating function, the moments, and the tail behavior. The in-component size distribution decays algebraically and the characteristic exponent varies continuously with the addition rate. Such behavior is in contrast with preferential attachment networks where this exponent is fixed.

Next, we obtain closed equations for the height distribution that characterizes the distance between a node and the root. This distribution obeys Poisson statistics. Other structural characteristics such as the diameter of the network and the degree of the most connected node follow directly from this quantity. In general, deletion leads to a condensation phenomena where a single node is connected to a finite fraction of all nodes, while the rest of the nodes have only a finite degree. We stress that condensation does not occur in the absence of deletion.

The rest of this paper is organized as follows. The addition-deletion process is introduced in section II. The in-component size distribution and related quantities such as the fraction of dangling nodes are discussed in section III. Next, in section IV, the height distribution and related quantities including the diameter are described. The degree distribution is analyzed using an approximate theory and numerical simulations in section V. Section VI contains a brief summary. Finally, appendices A and B contain several technical derivations.

*Electronic address: ebn@lanl.gov
†Electronic address: paulk@bu.edu

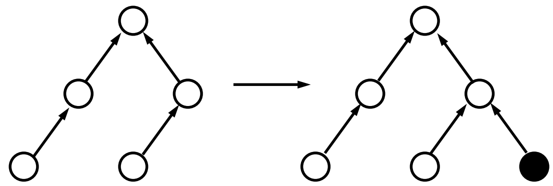

FIG. 1: Illustration of the node addition process.

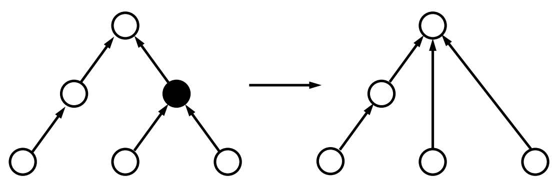

FIG. 2: Illustration of the node deletion process.

II. THE ADDITION-DELETION PROCESS

In the addition-deletion process, initially there is one seed node, and then, the network grows via addition of nodes and shrinks via deletion of nodes as follows.

1. **The addition process.** With rate r, a new node is added and it links to a randomly selected node (Fig. 1). This is an egalitarian attachment process as the new node is equally likely to link to any one of the existing nodes.

2. **The deletion process.** With rate 1, a randomly selected node (together with its outgoing link) is deleted. The parent of the deleted node inherits all the incoming links of the deleted node (Fig. 2). This inheritance mechanism preserves ancestral relations as is the case in phylogenetic trees: when a species goes extinct, its immediate descendants are linked to its immediate ancestor.

We stress that the addition process and the deletion process are completely independent of each other and that in the deletion process all links are removed with an equal probability.

This addition-deletion process preserves the complexity of the graph since nodes and links are added and deleted in pairs. Thus, starting with a single seed node and no links, the network has a tree structure. The seed node can not be deleted (or the ancestry would be destroyed) and therefore, the seed is the root of the tree.

The average total number of nodes at time t, $N(t)$, evolves according to $dN/dt = r-1$. Therefore, this quantity grows linearly with time and since $N(0) = 1$,

$$N(t) = 1 + (r-1)t. \tag{1}$$

The total number of links, $L(t)$, follows from $L = N - 1$ and therefore, $L(t) = (r-1)t$. We restrict our attention to the case of growing networks, $r > 1$. In this case, the size of the network increases indefinitely with time and fluctuations are negligible in the long time limit. The total number of nodes grows diffusively and is subject to large fluctuations in the limit $r \to 1$.

III. THE IN-COMPONENT SIZE DISTRIBUTION

The in-component of a node $\mathbf{n}$ is the set of all nodes which are connected to $\mathbf{n}$ via a path of directed links. In the context of a phylogenetic tree, the in-component is simply the set of all descendants. Let C_k be the average number of nodes with an in-component of size k. In our definition, the in-component includes the node itself, and therefore $k \geq 1$. The in-component size distribution evolves according to

$$\frac{dC_k}{dt} = r\delta_{k,1} + \frac{r}{N}\left[(k-1)C_{k-1} - kC_k\right] + \frac{k}{L}\left[C_{k+1} - C_k\right]. \tag{2}$$

The first three terms correspond to changes caused by node addition. New nodes generate a new in-component of size one and thus, the first term. In-components grow by one when a node links to any of the nodes in that in-component and the next two terms account for this augmentation. The corresponding attachment rate is the addition rate r normalized by the total number of nodes N. The last two terms correspond to node deletion. Deletion of any one of the k links in an in-component of size $k+1$ generates an in-component of size k and hence the gain term. Similarly, deletion of any one of the k nodes in an in-component of size k decreases the in-component size by one and hence the loss term. All nodes except for the root may be deleted and accordingly, the deletion rate is normalized by the total number of links L. By summing equations (2) one can verify that the total number of nodes $N = 1 + \sum_{k\geq 1} C_k$ satisfies $dN/dt = r - 1$. We stress that Eq. (2) is exact.

We are interested in the long time asymptotics where the difference between the total number of nodes and the total number of links is negligible. Therefore, we replace L with N in (2). We also make the transformation

$$C_k = Nc_k \tag{3}$$

where c_k is the fraction of nodes with in-component degree equal to k and $\sum_{k\geq 1} c_k = 1$. We seek stationary solutions ($dc_k/dt = 0$) for which the distribution function c_k satisfies the difference equation

$$(r-1)c_k = r\left[\delta_{k,1} + (k-1)c_{k-1} - kc_k\right] + k(c_{k+1} - c_k). \tag{4}$$

The tail of this distribution can be obtained using a continuum approximation, that is, through replacement of differences with derivatives: $c_{k+1} - c_k \to dc/dk$, etc. This change transforms the difference equation (4) into the differential equation

$$r\frac{d}{dk}(k\,c) - k\frac{dc}{dk} + (r-1)c = 0. \tag{5}$$

The solution of this equi-dimensional equation is a power-law

$$c_k \simeq B\,k^{-\beta}\,, \qquad \beta = 2 + \frac{1}{r-1}, \tag{6}$$

as $k \to \infty$. Thus, the in-component size distribution has a power-law tail. The exponent $2 \le \beta < \infty$ varies continuously with the addition rate. The characteristic exponent is minimal, $\beta = 2$, for the well-understood case of random recursive trees ($r \to \infty$) where deletion is irrelevant [22–27]. The exponent diverges in the more interesting limit of marginally growing networks ($r \to 1$) where deletion is as strong as addition.

Addition-deletion networks differ from strictly growing networks where the exponent characterizing the in-component size distribution is independent of the details of the attachment mechanism, $\beta = 2$ [26]. Thus the deletion process qualitatively changes the network structure.

The moments $M_n = \sum_k k^n C_k$ provide complementary information about the in-component size distribution. Of course, $M_0 = N - 1$. The first moment satisfies

$$\frac{dM_1}{dN} = \frac{r-2}{r-1}\,\frac{M_1}{N} + \frac{r+1}{r-1}, \tag{7}$$

as seen by summing (2). Equation (7) describes the evolution as a function of the total number of nodes rather than time. Since we are interested in the asymptotic behavior, we again replace L with N in (2). Solving Eq. (7), we find that the first moment is proportional to the number of nodes, $M_1 = (1 + r)N$. Hence the average in-component size is linear in the addition rate, $\langle k\rangle = 1 + r$.

Similarly, the second moment obeys

$$\frac{dM_2}{dN} = \frac{2r-3}{r-1}\,\frac{M_2}{N} + \frac{r+3}{r-1}\,\frac{M_1}{N} + 1. \tag{8}$$

There are three types of asymptotic behavior

$$M_2 \simeq \begin{cases} \frac{(r^2+5r+2)}{2-r} N & 1 < r < 2; \\ 16\,N \ln N & r = 2; \\ \kappa N^{\frac{r-2}{r-1}} & r > 2. \end{cases} \tag{9}$$

The proportionality constant κ depends on the initial conditions. These asymptotic behaviors are consistent with the algebraic decay (6): the second moment grows linearly when $\beta > 3$, it increases super-linearly when $\beta < 3$, and it acquires a logarithmic correction in the marginal case $\beta = 3$.

The generating function

$$F(x) = \sum_{k=1}^{\infty} c_k x^k \tag{10}$$

contains complete information about the in-component size distribution. Normalization implies $F(1) = 1$ and since $c_0 = 0$, then $F(0) = 0$. Multiplying Eq. (4) by x^k

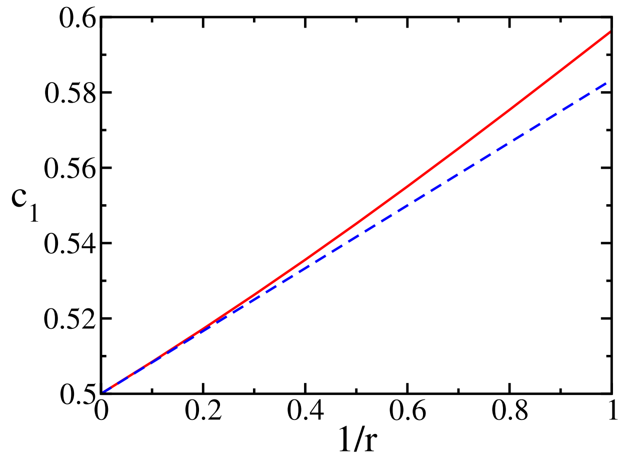

FIG. 3: The fraction of dangling nodes c_1 versus inverse of the addition rate $1/r$ (solid line). Also shown is the large-r asymptotic behavior $c_1 = \frac{1}{2} + \frac{1}{12r}$ (dashed line).

and summing over k, we arrive at a first order differential equation for the generating function

$$(x-1)(rx-1)F'(x) + \left(1 - r - x^{-1}\right) F(x) + rx = 0 \tag{11}$$

where the prime denotes differentiation with respect to x. The homogeneous part of equation (11) admits the solution $x(1-x)^{\beta-1}|rx-1|^{-\beta}$. We make the transformation $F(x) = x(1-x)^{\beta-1}(rx-1)^{-\beta}\,a(x)$ when $1/r < x < 1$, and then the auxiliary function $a(x)$ satisfies $a' = r\,(1-x)^{-\beta}(rx-1)^{\beta-1}$. Integration of this equation yields the generating function

$$F(x) = r\,\frac{x(1-x)^{\beta-1}}{(rx-1)^{\beta}} \int_{1/r}^{x} dy\, \frac{(ry-1)^{\beta-1}}{(1-y)^{\beta}} \tag{12}$$

in the range $1/r < x < 1$. The lower limit of integration was chosen to assure that $F(1/r)$ is finite. One can verify that $\lim_{x\to 1} F(x) = 1$ in agreement with the normalization requirement. A similar calculation gives the generating function in the range $0 < x < 1/r$

$$F(x) = r\,\frac{x(1-x)^{\beta-1}}{(1-rx)^{\beta}} \int_{x}^{1/r} dy\, \frac{(1-ry)^{\beta-1}}{(1-y)^{\beta}}. \tag{13}$$

The large-k behavior is encoded in the $x \to 1$ behavior of the generating function. First, we confirm the power-law tail (6) by using (12). Second, we obtain the corresponding proportionality constant

$$B = \Gamma(\beta)(\beta-1)^{\beta}. \tag{14}$$

The appendix details the analysis leading to this result.

Conversely, the small-k behavior is reflected by the $x \to 0$ behavior of the generating function. The fraction of dangling nodes, that is, nodes with no incoming links, is given by $c_1 = \lim_{x\to 0} F'(x)$ and by differentiation of the generating function (13) we readily find

$$c_1 = \int_0^1 dy\,(1-y)^{\beta-1}(1-y/r)^{-\beta}. \tag{15}$$

The integral can be expressed via the hypergeometric function, $c_1 = \beta^{-1} F(\beta, 1; \beta+1; r^{-1})$, or alternatively as the power series $c_1 = \sum_{k\geq 0}(k+\beta)^{-1} r^{-k}$. The quantity c_1 is a monotonically decreasing function of the addition rate r and is roughly linear in $1/r$ (figure 3). It is maximal for marginally growing networks (see subsection A below),

$$\lim_{r\to 1} c_1 = e\, E_1(1) = 0.596347\ldots, \tag{16}$$

where $E_1(x) = \int_x^\infty dz z^{-1} e^{-z}$ is the exponential integral [28]. The fraction of dangling nodes is minimal for random recursive trees, $\lim_{r\to\infty} c_1 = 1/2$. In general, over half of the nodes are dangling nodes.

The in-component size distribution can be calculated recursively for small k by using (4), $c_2 = r(2c_1 - 1), c_3 = r\,[(6r+1)c_1 - (3r+1)]$, etc.

We now investigate the limits $r \to 1$ and $r \to \infty$ in detail.

A. Marginally Growing Networks

The recursion relation (4) becomes

$$c_{k+1} = \delta_{k,1} + (k-1)c_{k-1} - 2kc_k + (k+1)c_{k+1} \tag{17}$$

for marginally growing networks ($r \to 1$). Repeating the steps leading to (5) yields the second order linear differential equation

$$\frac{d^2}{dk^2}(kc) - \frac{dc}{dk} - c = 0. \tag{18}$$

The tail of the in-component size distribution has a stretched exponential tail

$$c_k \sim k^{-1/4}\, e^{-2\sqrt{k}}, \tag{19}$$

a result that can be obtained using the WKB technique. This sharp decay is consistent with the divergent β.

To find the generating function, we repeat the steps leading to (12). The governing equation (11) becomes $(x-1)^2 F' - x^{-1}F + x = 0$, and the solution of this equation is $F(x) = \frac{x}{1-x}\, e^{\frac{1}{1-x}} \int_0^{1-x} \frac{dy}{y} \exp\left(-\frac{1}{y}\right)$. The generating function can be expressed in terms of the exponential integral

$$F(x) = \frac{x}{1-x}\, e^{\frac{1}{1-x}}\, E_1\left(\frac{1}{1-x}\right), \tag{20}$$

from which the fraction of dangling nodes (16) follows.

B. Random Recursive Trees

The second limiting case, $r \to \infty$, corresponds to random recursive trees. Now, the in-component size distribution satisfies the recursion equation

$$(k+1)c_k = (k-1)c_{k-1} + \delta_{k,1}. \tag{21}$$

Unlike in the general case, this simpler equation can be solved recursively starting at $k=1$: $c_1 = \frac{1}{1\cdot 2}$, $c_2 = \frac{1}{2\cdot 3}$, $c_3 = \frac{1}{3\cdot 4}$, etc. We thus recover the well-known result [26]

$$c_k = \frac{1}{k(k+1)}. \tag{22}$$

Alternatively, one can obtain the distribution from the generating function $F(x) = 1 + \frac{1-x}{x}\,\ln(1-x)$.

IV. THE HEIGHT DISTRIBUTION

Another important structural characteristic is the height of a node, that is, the number of links separating a given node from the root [29, 30]. Let $H_k(t)$ be the average number of nodes with height k at time t. This quantity evolves according to

$$\frac{dH_k}{dt} = \frac{r}{N} H_{k-1} + \frac{k}{L}\,(H_{k+1} - H_k) \tag{23}$$

for $k \geq 1$. The initial condition is $H_k(0) = \delta_{k,0}$ and the boundary conditions are $H_0 = 1$ and $H_{-1} = 0$. The first term on the right-hand side accounts for gain due to node addition and the next two terms describe changes due to node deletion. The factor k reflects that deletion of a node of height k reduces the heights of all of its descendants by one. Again, the addition and the deletion rates are normalized by the total number of nodes and the total number of links, respectively. One can verify that the total number of nodes $N = \sum_k H_k$ satisfies $dN/dt = r - 1$.

Following the above analysis of the in-component size distribution, we study the long time asymptotic behavior of h_k, the fraction of nodes with height k, using the transformation

$$H_k = N h_k. \tag{24}$$

Normalization implies $\sum_k h_k = 1$ and the boundary condition is now $h_0 = 0$ since $H_0 = 1$. Let us substitute (24) into the rate equation (23). The distribution function h_k satisfies the recursion relation

$$r(h_k - h_{k-1}) + (h_{k+1} - h_k) = (k+1)h_{k+1} - kh_k. \tag{25}$$

This equation was obtained by adding H_{k+1} to both side of (23).

In this case, it is convenient to use the exponential generating function, defined via

$$G(z) = \sum_{k=0}^{\infty} h_k e^{kz}. \tag{26}$$

Multiplication of the recursion relation (25) by e^{kz} and summation over k leads to the first order ordinary differential equation

$$\frac{dG}{dz} = G(z)(1 + re^z). \tag{27}$$

The generating function is found by integration of this equation subject to the boundary condition $G(0) = 1$,

$$G(z) = e^{z+r(e^z-1)}. \tag{28}$$

Using the definition (26) we find that the height distribution function is Poissonian

$$h_k = e^{-r}\,\frac{r^{k-1}}{(k-1)!}. \tag{29}$$

The moments of the height distribution $m_n = \sum_k k^n h_k$ can be found by differentiation of the generating function (28). For instance, $m_1 = G'(0) = 1 + r$, and therefore the average node height is equal to the average in-component size, $\langle k\rangle = 1 + r$. Similarly, fluctuations in the height are obtained from the second moment $m_2 = c''(0) = r^2+3r+1$, and hence, the variance is

$$\langle k^2\rangle - \langle k\rangle^2 = r. \tag{30}$$

This variance is sufficient to characterize the behavior in the vicinity of the average, where the Poissonian height distribution becomes Gaussian.

We can immediately deduce a number of characteristics using the height distribution. There are H_k nodes at height k with incoming links from the H_{k+1} nodes at height $k+1$. Thus, the average in-degree I_k of nodes at height k is $I_k = H_{k+1}/H_k$. Using $H_0 = 1$ and (29) we find

$$I_k = \begin{cases} Ne^{-r} & k = 0;\\ r\,k^{-1} & k > 0.\end{cases} \tag{31}$$

The second result shows that the average degree is inversely proportional to the height.

The root, with its macroscopic degree, is the most connected node whereas all other nodes have a microscopic degree. Hence, the root is a giant hub. The degree of the root is maximal for marginally growing networks, $\lim_{r\to 1} I_0 = Ne^{-1}$, and it decreases indefinitely as r increases. Therefore, deletion must be responsible for this condensation phenomena. Indeed, since the attachment rate to each node is $1/N$, addition can be responsible for at most $\ln N$ of the Ne^{-r} connections as is the case for random recursive trees [24–26]. We conclude that emergence of the giant hub is due to deletion.

Consider now the highest node. The maximal height $k_{\rm max}$ is determined from the extreme statistics criterion $H_{k_{\rm max}} \sim 1$. Using the Poisson distribution (29) and the Stirling formula $k! \sim (k/e)^k$, the highest node grows very slowly with the network size, $k_{\rm max} \approx \ln N/\ln(\ln N)$. Remarkably, the maximal height is independent of the addition rate r.

Asymptotically, the diameter of the network D is twice the maximal height

$$D \approx 2\,\frac{\ln N}{\ln(\ln N)}. \tag{32}$$

For strictly growing random trees, the diameter exhibits a logarithmic growth independent of the attachment mechanism, $D \sim \ln N$ [26]. However, the diameter can be affected by the topology. If every new node links to m existing nodes, this strictly growing network is no longer a tree (for $m \geq 2$) and $D \sim \ln N/\ln\ln N$. This result [31] is again very robust, namely it holds for strictly growing networks with reasonable attachment mechanisms and for all $m \geq 2$. Equation (32) shows that deletion qualitatively affects the growth of the diameter. Surprisingly, despite their distinct topologies these strictly growing networks and the addition-deletion networks have similar diameters.

Finally, we address the average distance between two nodes, $\langle l\rangle$. Since the degree of the root is proportional to the total number of nodes N, a path connecting two randomly selected nodes almost surely includes the root. As a result, the average distance is asymptotically equal to twice the average height $\langle l\rangle = 2(1 + r)$.

V. THE DEGREE DISTRIBUTION

In contrast with the in-component size distribution and the height distribution, it is impossible to construct closed equations for the degree distribution and we perform an approximate analysis.

When a node with in-degree j is deleted, the in-degree of the parent i is augmented by $j-1$. This is simply the aggregation process [32]

$$(i,j) \xrightarrow{i/(NL)} i + j - 1. \tag{33}$$

The aggregation rate is proportional to the in-degree of the parent node because any of the daughter nodes can be deleted. The normalization factor NL reflects that the probability of picking the deleted node is inversely proportional to the number of nodes and the probability of picking the parent node is inversely proportional to the number of links.

Let $A_k(t)$ be the average number of nodes with in-degree k at time t (the out-degree is always one). The total number of nodes and the total number of links are given by $N = \sum_k A_k$ and $L = \sum_k kA_k$, respectively. Assuming that the degrees of neighboring nodes are completely uncorrelated, the quantity A_k obeys the nonlinear rate equation [33, 34]

$$\begin{aligned}\frac{dA_k}{dt} &= r\delta_{k,0} + \frac{r}{N}(A_{k-1} - A_k) \\ &+ \frac{1}{NL}\left[\sum_{i+j=k+1} iA_iA_j - (kN + L)A_k\right].\end{aligned} \tag{34}$$

The initial condition is $A_k(0) = \delta_{k,0}$ and the boundary condition is $A_{-1}(t) = 0$. The first three terms represent addition events. These terms are linear and thus, exact. The last three terms account for deletion

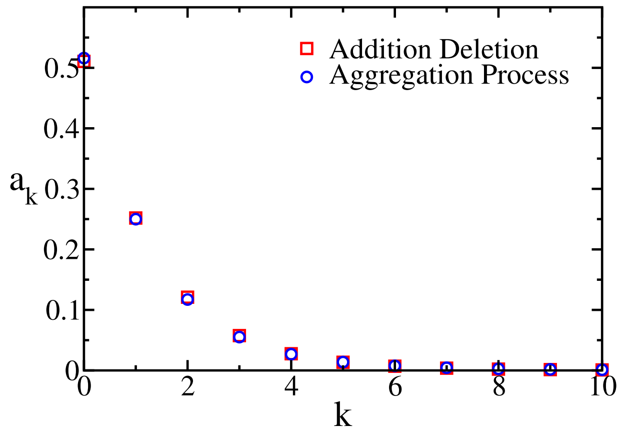

FIG. 4: The degree distribution for the aggregation process and for the addition-deletion network. Shown are results of one Monte Carlo simulation run with 10^7 nodes and $r = 4$. Roughly 90% of the nodes have a degree smaller or than $k = 10$.

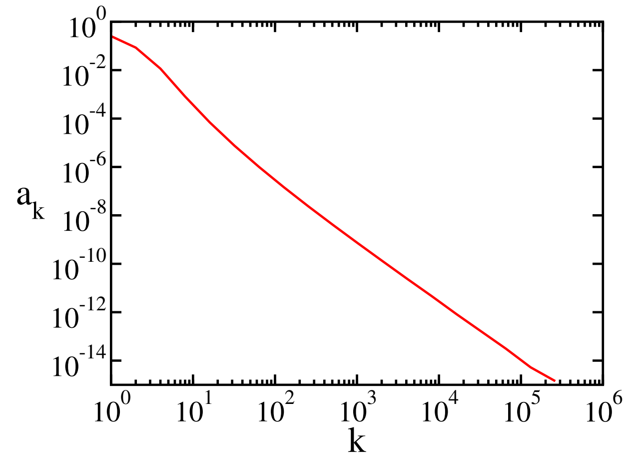

FIG. 5: The tail of the degree distribution for the addition-deletion network. The simulation results were obtained from 10^3 independent realizations in a network with 10^7 nodes and $r = 4$. The decay exponent is $\gamma = 2.45 \pm 0.05$

events. These terms are non-linear as they represent two node interactions. Moreover, the gain term has a convolution structure, reflecting the aggregation process (33). Of course, the aggregation process becomes exact in the limit $r \to \infty$ where deletion becomes irrelevant. By summation of the rate equation (34) we recover $dN/dt = dL/dt = r - 1$.

Following our above analysis, we study the fraction of nodes with in-degree k, a_k. Making the transformation $A_k = Na_k$, the degree distribution satisfies

$$(k + 2r)a_k = r\delta_{k,0} + ra_{k-1} + \sum_{i+j=k+1} ia_i a_j. \tag{35}$$

We analyze this hierarchy of equations using the generating function $U(x) = \sum_k a_k x^k$. The normalizations $\sum_k a_k = \sum_k ka_k = 1$ imply $U(1) = U'(1) = 1$. By multiplying the recursive equations (35) by x^k and summing over k, we find that the generating function obeys the nonlinear ordinary differential equation

$$(x - U)\frac{dU}{dx} = r(x - 2)U + r. \tag{36}$$

This equation is consistent with the normalization conditions $U(1) = U'(1) = 1$.

The moments of the degree distribution defined by $m_n = \sum_k a_k k^n$ are obtained by successive differentiation of the generating function. In particular, the high-order moments grow rapidly although they do remain finite (appendix B)

$$m_n \sim \left[\frac{r}{(r-1)^2}\right]^n (n!)^2, \tag{37}$$

as $n \to \infty$. Using these asymptotics we deduce (see Appendix B) that the tail of the degree-distribution decays as a stretched exponential

$$a_k \sim e^{-\lambda\sqrt{k}} \tag{38}$$

with $\lambda = [2(r-1)]/\sqrt{r}$.

How well does this approximation compare with addition-deletion networks? In addition-deletion networks, the root has a macroscopic degree. In other words, it is a condensate that contains a finite fraction of all nodes. In the aggregation process (33) all clusters are finite, as seen from $U'(1) = \sum_k ka_k = 1$ [34]. This is a consequence of the fact that all nodes are treated in the same way. Introducing a fixed node that can not be deleted should significantly improve the approximation.

We also examined the degree distribution using Monte Carlo simulations of the aggregation process (33) and the addition-deletion process. We find that when the degree is very small, the aggregation process approximates the addition-deletion process very well (Figure 4). For small degree this agreement holds independent of r. However, this approximation is poor when the degree is large.

For addition-deletion networks, the degree distribution follows a power law behavior (Figure 5)

$$a_k \sim k^{-\gamma} \tag{39}$$

as $k \to \infty$. Our approximate approach neglects degree-degree correlation. The fact that our approximate approach is accurate at small degrees but inaccurate at higher degrees indicates that degree-degree correlations are negligible for small degrees but substantial for large degrees. Indeed, highly connected nodes are long lived and thus, involve strong memory.

Addition-deletion networks can be simulated using an efficient algorithm that is linear in the network size. Throughout the growth process we keep track of all nodes, whether they have been deleted (passive nodes) or not (active nodes). In an addition event, an active node is added and it links to a randomly selected active node. In a deletion event, an active node is marked passive. When the required network size is reached, passive nodes are removed in reverse chronological order to preserve the ancestral relations.

Simulations of the aggregation process are performed as follows. The system consists of aggregates, each with a given size. In an addition step, an aggregate of size zero is added and simultaneously, the size of another, randomly chosen aggregate increases by one. In a deletion step one aggregate is chosen with probability proportional to its size and it merges with another randomly chosen aggregate according to (33).

VI. SUMMARY

We have studied networks which undergo a biased growth — new nodes are added at a certain rate and old nodes are deleted at a smaller rate in such a way that the ancestry is preserved. This growing network model is relevant to biological and technological networks that preserve ancestral relations between nodes. Our main conclusion is that it is possible to characterize the structure of such networks analytically. For instance, the in-component size distribution has a power-law tail and the characteristic exponent varies continuously with the addition rate. We have also shown that the height distribution obeys Poisson statistics. Additional properties including the fraction of dangling nodes and the average degree as a function of height have been obtained as well.

Interestingly, the addition-deletion process results in a giant hub that is connected to a finite fraction of the nodes in the system. This fraction decreases to zero as deletion becomes weaker, showing that the deletion process causes this condensation phenomena.

We have also seen that the degree distribution evolves by an aggregation process since the parent node inherits all incoming links of a deleted node. Treating the underlying aggregation process as completely random is valid only at small degrees because degree-degree correlations become significant at large degrees. Constructing closed equations for this aggregation process from which the degree distribution can be calculated is an outstanding challenge.

Acknowledgments. We acknowledge financial support from DOE grant DE-AC52-06NA25396 and NSF grant CHE-0532969.

---

[1] D. J. Watts and S. H. Strogatz, Nature **393**, 440 (1998).
[2] R. Albert and A. L. Barabasi, Rev. Mod. Phys. **74**, 47 (2002).
[3] S. N. Dorogovtsev and J. F. F. Mendes, Adv. in Phys. **51**, 1079 (2002).
[4] A. L. Barabasi and R. Albert, Science **286**, 509 (1999).
[5] P. L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000).
[6] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000); Phys. Rev. E **63**, 062101 (2001).
[7] J. Davidsen, H. Ebel, and S. Bornholdt, Phys. Rev. Lett. **88**, 128701 (2002).
[8] J. M. Pacheco, A. Traulsen, and M. A. Nowak, Phys. Rev. Lett. **97**, 258103 (2006).
[9] Y. Moreno, J. B. Gómez, and A. F. Pacheco, Europhys. Lett. **58**, 630 (2002).
[10] F. Chung and L. Lu, Internet Mathematics **1**, 409 (2004).
[11] C. Cooper, A. Frieze, and J. Vera, Internet Mathematics **1**, 463 (2004).
[12] C. Moore, G. Ghosal, and M. E. J. Newman, Phys. Rev. E **74**, 036121 (2006).
[13] N. Farid and K. Christensen, New. J. Phys. **8**, 212 (2006).
[14] M. Kimura, *The natural theory of evolution* (Cambridge Press, Cambridge 1983).
[15] M. S. Waterman, *Introduction to Computational Biology: Maps, sequences and genomes* (Chapman & Hall, London, 1995).
[16] R. Durbin, S. Eddy, A. Krogh, and G. Mitchison, *Biological sequence analysis* (Cambridge University Press, Cambridge, 1998).
[17] M. A. Huynen and E. van Nimwegen, Mol. Biol. Evol. **15**, 583 (1998).
[18] E. V. Koonin, Y. Wold, and G. Karev, Nature **420**, 218 (2002).
[19] S. C. Manrubia and D. H. Zannette, J. Theor. Biol. **216**, 461 (2002).
[20] P. W. Messer, P. F. Arndt, and M. Lässig, Phys. Rev. Lett. **94**, 1318103 (2005).
[21] E. Ben-Naim and P. L. Krapivsky, J. Stat. Mech. P07001 (2006).
[22] H. M. Mahmoud, *Evolution of Random Search Trees* (John Wiley & Sons, New York, 1992).
[23] D. E. Knuth, *The Art of Computer Programming, vol. 3, Sorting and Searching* (Addison-Wesley, Reading, 1998).
[24] R. T. Smythe and H. Mahmoud, Theory Probab. Math. Statist. **51**, 1–27 (1995).
[25] M. Drmota and B. Gittenberger, Random Struct. Alg. **10**, 421–451 (1997); M. Drmota and H.-K. Hwang, Adv. Appl. Probab. **37**, 321–341 (2005).
[26] P. L. Krapivsky and S. Redner, Phys. Rev. E **63**, 066123 (2001).
[27] P. L. Krapivsky and S. Redner, Phys. Rev. Lett. **89**, 258703 (2002).
[28] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).
[29] E. Ben-Naim, P. L. Krapivsky, and S. N. Majumdar Phys. Rev. E **64**, 035101(R) (2001)
[30] R. Lambiotte and M. Ausloos, Europhys. Lett. **77**, 58002 (2007).
[31] B. Bollobás and O. Riordan, Combinatorica **24**, 5 (2004).
[32] F. Leyvraz, Phys. Rep. **383**, 95 (2003).
[33] M. von Smoluchowski, Z. Phys. **17**, 557 (1916).
[34] E. Ben-Naim and P. L. Krapivsky, Phys. Rev. E **75**, 011103 (2007).
[35] In general, $\ln a_k \propto -k^{1/\nu}$ when $U_n \sim (n!)^\nu$.

APPENDIX A: DERIVATION EQ. (14)

The tail of the in-component size distribution can be obtained from the $x \to 1$ asymptotics of the generating function $F(x)$. Let us consider for example the power-law tail $c_k \simeq Bk^{-2}$ as $k \to \infty$. In this case, the leading behavior of the derivative of the generating function is

$$\frac{dF(x)}{dx} = \sum_k k c_k x^{k-1} \simeq B \sum_k k^{-1} x^k = -B \ln(1-x)$$

in the limit $x \to 1$. Generally, the power-law decay (6) implies

$$\frac{d^{\beta-1}F(x)}{dx^{\beta-1}} = -B \ln(1-x) \tag{A1}$$

as $x \to 1$, for any positive integer β.

To evaluate this derivative, we rewrite the generating function (12) using the transformation $z = (\beta-1)(1-y)$

$$F(x) = r^\beta \, \frac{x(1-x)^{\beta-1}}{(rx-1)^\beta} \int_{(\beta-1)(1-x)}^1 dz \, \frac{(1-z)^{\beta-1}}{z^\beta}. \tag{A2}$$

For arbitrary positive integer β, the integral is evaluated by performing β successive integration by parts

$$F(x) = \frac{xP_{\beta-1}(x)}{(rx-1)^\beta} + (-r)^\beta \frac{x(1-x)^{\beta-1}}{(rx-1)^\beta} \ln(1-x) \tag{A3}$$

where $P_n(x)$ is a polynomial of degree n.

Differentiating the generating function (A3) $\beta-1$ times and noting that the second term dominates in the $x \to 1$ limit, we obtain

$$\frac{d^{\beta-1}F(x)}{dx^{\beta-1}} \simeq -(\beta-1)! \left(\frac{r}{r-1}\right)^\beta \ln(1-x). \tag{A4}$$

By comparing this expression with (A1) we obtain the coefficient $B = \Gamma(\beta)(\beta-1)^\beta$ for all integer $\beta \geq 2$. Finally, we use analytic continuation and extend this result to arbitrary β.

APPENDIX B: DERIVATION OF (37) AND (38)

We obtain the large-k behavior of the degree distribution using the $x \to 1$ behavior of the generating function $U(x)$. Differentiating equation (36) n times and setting $x = 1$ gives a recursion relation for the derivatives $U_n = d^nU(x)/dx^n|_{x=1}$

$$U_n = \frac{r\,n(n+r)}{(r-1)^2} U_{n-1} + \frac{1}{r-1} \sum_{m=2}^{n-3} \binom{n}{m} U_{n-m} U_{m+1}.$$

The first two coefficients are $U_0 = U_1 = 1$. Only the first term in the recursion equation is relevant asymptotically and therefore,

$$\frac{U_n}{U_{n-1}} \to \frac{r}{(r-1)^2}\, n^2 \tag{B1}$$

as $n \to \infty$. Eq. (37) is obtained by noting that the coefficients

$$U_n = \sum_{k \geq 0} k(k-1)\ldots(k-n+1)a_k$$

are asymptotically equivalent to the moments, $U_n \sim m_n$.

The large-n behavior of m_n is computed as follows:

$$\begin{aligned} m_n &\sim \int dk\, k^n a_k \\ &\sim \int dk \, \exp\left(n \ln k - \lambda\sqrt{k}\right) \\ &\sim \exp\left[2n \ln(2n/\lambda e)\right] \\ &\sim \left(\frac{2}{\lambda}\right)^{2n} (n!)^2. \end{aligned} \tag{B2}$$

In the first line we replaced the summation by integration, and in the second we used the presumed asymptotic $a_k \sim e^{-\lambda\sqrt{k}}$. The third line was obtained by using the steepest descent method — the function $f(k) = n \ln k - \lambda\sqrt{k}$ is maximal at $k_* = (2n/\lambda)^2$. The fourth line was obtained using the Stirling formula $n! \sim (n/e)^n$. Finally, by comparing equations (B2) and (37), we obtain the tail behavior (38) [35].