Sublinear but Never Superlinear Preferential Attachment by Local Network Growth

Alan Gabel and S. Redner

Center for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215, USA

Abstract.

We investigate a class of network growth rules that are based on a redirection algorithm wherein new nodes are added to a network by linking to a randomly chosen target node with some probability $1-r$ or linking to the parent node of the target node with probability r. For fixed $0 < r < 1$, the redirection algorithm is equivalent to linear preferential attachment. We show that when r is a decaying function of the degree of the parent of the initial target, the redirection algorithm produces sublinear preferential attachment network growth. We also argue that no local redirection algorithm can produce superlinear preferential attachment.

PACS numbers: 89.75.Fb, 89.75.Hc, 02.50.Cw, 05.40.-a

1. Introduction

A popular and highly successful theory to account for the growth of complex networks is preferential attachment [1, 2, 3, 4, 5]. Here new nodes are added sequentially to a network and each links to existing nodes according to an attachment rate A_k that is an increasing function of the degree k of the 'target' node to which linking occurs. Many real world networks appear to evolve according to this simple dynamics [6, 7, 8, 9, 10].

Preferential attachment naturally divide into three classes: sublinear, linear, and superlinear, in which the attachment rate grows with the degree of the target node as $A_k \sim k^\gamma$ with $0 < \gamma < 1$, $\gamma = 1$, and $\gamma > 1$, respectively [11, 12]. Each class produces networks with qualitatively different properties. Sublinear preferential attachment leads to networks with a universal stretched-exponential degree distribution. Linear preferential attachment leads to scale-free networks, but with a fragile power-law degree distribution. Here, fragile denotes that the exponent of the degree distribution depends sensitively on microscopic details of the growth mechanism. Superlinear preferential attachment networks are singular in character, as they contain one highly-connected "hub" node whose degree is of the order of the total number of nodes in the network [13, 14, 15, 16].

A basic feature of the preferential attachment growth rule is that each new node must 'know' the degree distribution of the entire network, as this global information is exploited to determine the identity of the target node. However, since real-world networks are typically large, it is unreasonable to expect that any new node has such global knowledge. A possible resolution is to use a redirection rule in which a new node only knows about some local portion of the network and attaches to a node in this region. As we will discuss, redirection generates linear preferential attachment growth by suitably-defined local growth rules [11, 17, 18]. An attractive feature of redirection is that this algorithm leads to extremely efficient simulations of network growth [19, 20, 21].

In this work, we introduce an extension of the redirection algorithm that produce networks that grow according to *sublinear* preferential attachment. In classic redirection, each new node attaches to either: (i) a randomly chosen target node with probability $1 - r$, where r is a fixed number strictly between 0 and 1, or (ii) the parent of the target, with probability r. We will show that this same redirection algorithm leads to sublinear preferential attachment growth when r is a suitably-chosen decreasing function of the degree of the parent node. Thus sublinear preferential attachment can also be achieved from a local growth rule. Furthermore, we will argue that no local redirection algorithm can produce superlinear preferential attachment. Our claim may help explain why linear and sublinear preferential attachment networks are found ubiquitously in empirical studies [6, 7, 22, 23, 24, 25], while evidence for superlinear attachment networks are relatively scarce [13, 26]: most real-world networks truly *do* grow according to local rules and thus cannot be governed by superlinear preferential attachment.

2. Preferential Attachment

We begin by summarizing basic aspects of preferential attachment and then describe how to achieve this type of network growth by redirection. Let N be the total number of nodes in a growing network and N_k be the number of nodes of degree k. For simplicity we consider directed tree-like networks, in which each new node has a single outgoing link. Thus a node of degree k will have a single parent (the node it first attached to upon entering the network) and $k-1$ children. We assume the network begins with a single node, which is thus its own parent. More general initial conditions give the same results when number of network nodes is sufficiently large. The network grows by introducing new nodes into the network sequentially. Each new node attached to a pre-existing node of degree k with attachment rate A_k.

The evolution of N_k with the addition of each node is thus governed by [16]

$$\frac{dN_k}{dN} = \frac{A_{k-1}N_{k-1} - A_k N_k}{A} + \delta_{k,1}\,, \tag{1}$$

where $A = \sum_j A_j N_j$ is the total attachment rate. The first term $A_{k-1}N_{k-1}/A$ gives the probability that the new node attaches to a pre-existing node of degree $(k-1)$; this connection converts the node to have degree k, thereby increasing N_k by 1. Similarly, the term $A_k N_k/A$ corresponds to the probability that the new node connects to a node of degree k, thereby decreasing N_k by 1. Finally, the term $\delta_{k,1}$ arises because every new node has degree 1 and so increases N_1 by 1.

In sublinear and linear preferential attachment, it can be shown [11] that the total attachment rate scales as $A = \mu N$, where μ is a constant. In contrast, for superlinear attachment, the total rate scales as $A \sim N^\gamma$ because $A = \sum_j A_j N_j$ is dominated by the term in the sum that is associated with the hub, whose degree is of order N. The formal solution for N_k can be readily obtained by writing $N_k = N n_k$ and using $A = \mu N$ in (1) to give [11]

$$n_k = \frac{\mu}{A_k} \prod_{j=1}^{k} \left(1 + \frac{\mu}{A_j}\right)^{-1}. \tag{2}$$

For sublinear preferential attachment, the asymptotic behavior may be found by writing the above product as the exponential of the sum of a logarithm, converting the sum to an integral, expanding the logarithm in inverse powers of k^γ, and then performing the integrals. These manipulations lead to a degree distribution that asymptotically has the stretched exponential form

$$n_k \sim k^{-\gamma} \exp\left[-\frac{\mu}{1-\gamma}\, k^{1-\gamma}\right]. \tag{3}$$

For linear preferential attachment, and more generally for shifted linear attachment, where $A_k = k + \lambda$, the asymptotic degree distribution has the *non-universal* power-law form $n_k \sim k^{-3-\lambda}$n [11]. Here λ must satisfy the constraint $\lambda > -1$; otherwise, network evolution is pathological because it is not possible to attach to nodes of degree 1.

3. Generalized Redirection Algorithm

We now discuss how a suitably-defined redirection algorithm leads to a complex network whose growth is governed by *sublinear* preferential attachment. We assume that the network starts in a configuration in which each node has one parent. Let us first review the classic redirection algorithm that leads to shifted linear preferential attachment. The steps to add a new node to the network are as follows:

(i) Randomly select an existing node as the target.

(ii) A new node attaches to the target with probability $1 - r$, with $0 < r < 1$.

(iii) With probability r, the new node attaches to the parent of target.

This growth rule is local because each new node only needs to know about a single randomly-chosen node and its immediate environment, rather than the entire network structure. This redirection algorithm can be straightforwardly extended to allow the new node to make connections to multiple nodes in the network [27]. The surprising aspect of this innocuous redirection mechanism is that it leads precisely to linear preferential attachment for the case $r = \frac{1}{2}$ [11], a growth rule that ostensibly requires knowing the degrees of all nodes in the network.

We now generalize the redirection algorithm to allow $r = r(a, b)$ to be a function of the degrees of the target and ancestor nodes, a and b, respectively. To show how sublinear preferential attachment can be achieved from this still-local information, let us define f_k as the total probability that an incoming link is redirected *from* a randomly-selected target node of degree k to the parent of the target. Similarly, we define t_k as the total probability that an incoming link is redirected *to* a parent node of degree k after the incoming node initially selected one of the daughter nodes of this parent. Formally, these probabilities are defined in terms of the redirection probabilities by

$$f_k = \sum_{b\geq 1} \frac{r(k,b)N(k,b)}{N_k}\,, \qquad\qquad t_k = \sum_{a\geq 1} \frac{r(a,k)N(a,k)}{(k-1)N_k}\,, \tag{4}$$

where $N_k = \sum_{b\geq 1} N(k,b)$ and $N(a,b)$ is the correlation function that specifies the number of nodes of degree a that have a parent of degree b. Thus f_k is the mean redirection probability averaged over all N_k possible target nodes of degree k. Likewise, since each node of degree k has $k-1$ children, there are $(k-1)N_k$ possible target nodes whose redirection probabilities are averaged to give t_k.

In terms of these probabilities f_k and t_k, the master equation that governs the evolution of N_k is

$$\frac{dN_k}{dN} = \frac{(1-f_{k-1})N_{k-1} - (1-f_k)N_k}{N} + \frac{(k-2)t_{k-1}N_{k-1} - (k-1)t_kN_k}{N} + \delta_{k,1}. \tag{5}$$

The first ratio corresponds to instances of the growth process for which the incoming node actually attaches to the initial target node. For example, the term $(1-f_k)N_k/N$ gives the probability that one of the N_k target nodes of degree k is randomly selected and that the link from the new node is *not* redirected away from this target. Similarly, the second ratio corresponds to instances in which the link to the target node *is* redirected to the parent. For example, the term $(k-1)t_kN_k/N$ gives the probability that one of the $(k-1)N_k$ children of nodes of degree k is chosen as the target and that the new node *is* redirected. Lastly, the term $\delta_{k,1}$ accounts for the newly-added node of degree 1.

By simple rearrangement, we can express the master equation (5) in the generic form of (1), with attachment rate given by

$$\frac{A_k}{A} = \frac{(k-1)t_k + 1 - f_k}{N}\,. \tag{6}$$

As a simple check of this expression, note that when the redirection probability is constant, we have $f_k = t_k = r$ and the expected linear dependence $A_k \sim k$ is recovered.

4. Sublinear Preferential Attachment

The asymptotic behavior of the attachment rate in (6) is $A_k \sim k\,t_k$. Thus a redirection probability $r(a,b)$ for which t_k is a decreasing function of k will asymptotically correspond to sublinear preferential attachment. Let us therefore choose $r(a,b) = b^{\gamma-1}$, with $0 < \gamma < 1$. Because r depends only on the degree of the parent node, Eq. (4) reduces to $t_k = k^{\gamma-1}$. Using this form of t_k in Eq. (6) yields

$$\frac{A_k}{A} = \frac{k^\gamma - k^{\gamma-1} + 1 - f_k}{N} \tag{7}$$

whose leading behavior is indeed sublinear preferential attachment: $A_k \sim k^\gamma$. Because f_k is a bounded probability, it is a subdominant contribution to A_k.

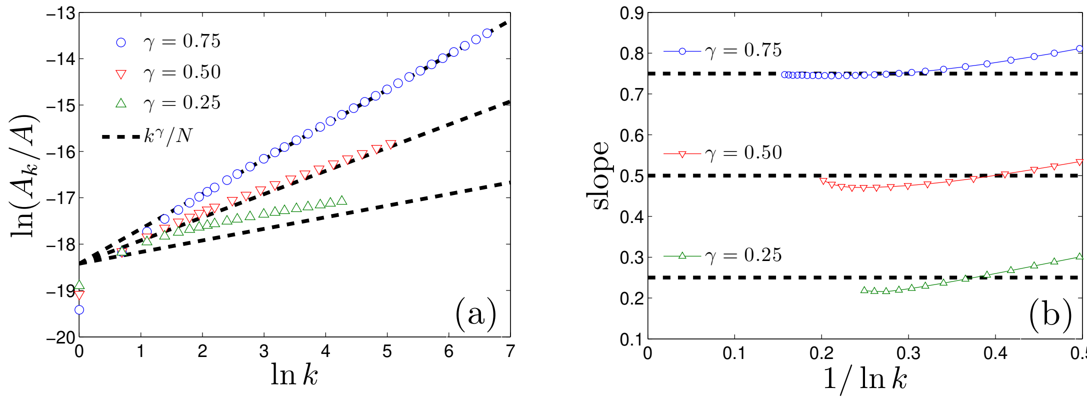

Figure 1. (a) Simulated attachment probabilities A_k/A versus k from generalized nearest-neighbor redirection. (b) Local slopes of the successive data points in (a) versus $1/\ln k$ for $\gamma = 0.75$ ($\circ$), $\gamma = 0.5$ (∇), and $\gamma = 0.25$ (Δ). The dashed curve represents $A_k/A = k^\gamma/N$.

To test the prediction that $A_k \sim k^\gamma$ with $\gamma < 1$, we simulated 100 network realizations that are grown to $N = 10^8$ nodes by generalized redirection. Once a network reaches 10^8 nodes, we measure the probability that attachment to a node of degree k actually occurs by systematically making 'test' attachments to each node of the network according to generalized redirection. The term test attachment means that the network is returned to its original state after each such event. We count all test events that ultimately lead to attachment to a node of degree k. Dividing this number of events by the total number of nodes N gives the attachment probability to nodes of degree k, $A_k N_k/A$.

Our simulations indeed show that the A_k grows sublinearly with k (Fig. 1). The agreement is best for γ less than, but close to 1, where the degree distribution is fairly broad. As γ decreases, the asymptotic behavior is contaminated by the appearance of progressively more slowly-decaying sub-asymptotic corrections terms in Eq. (6). We also compare the simulated degree distribution to the analytic result given in Eq. (2) with $A_k = k^\gamma$. To make this comparison, we fit the simulated distribution to

$$n_k = C\,\frac{\tilde{\mu}}{k^\gamma}\prod_{j=1}^{k}\left(1+\frac{\tilde{\mu}}{j^\gamma}\right)^{-1}, \tag{8}$$

where C and $\tilde{\mu}$ are fitting parameters. We introduce these parameters because our redirection algorithm gives the attachment rate $A_k \sim k^\gamma$ only asymptotically. Thus the degree distribution that arises from our generalized redirection algorithm should match the theoretical prediction (2) only as $k \to \infty$ (Fig. 2). While the best-fit value of $\tilde{\mu}$ does not obey the bound $\mu > 1$ for sublinear preferential attachment [11], the fitting parameters do not affect the nature of the dependence of n_k on k and γ.

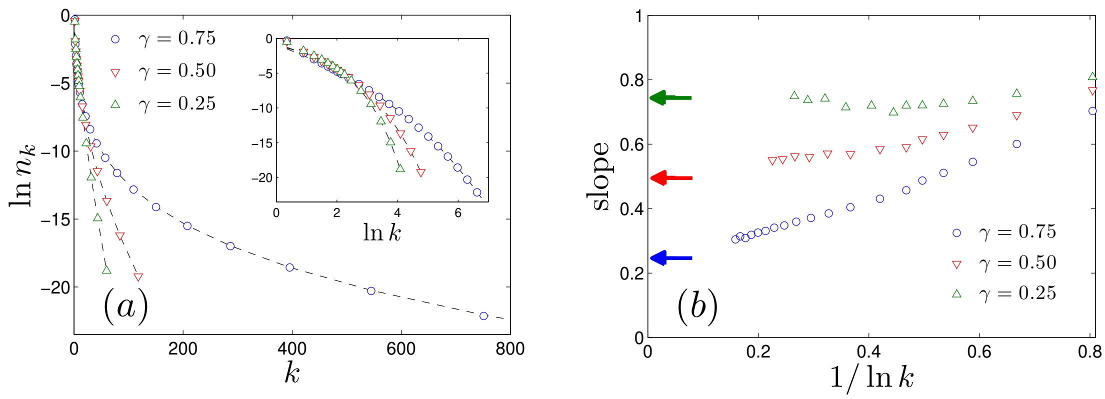

Figure 2. (a) Degree distribution n_k versus k for generalized redirection with $r(a,b) = b^{\gamma-1}$ for $\gamma = 0.75$ ($\circ$), $\gamma = 0.5$ ($\bigtriangledown$), and $\gamma = 0.25$ ($\bigtriangleup$). The data are accumulated in equal-size bins on a logarithmic scale. Dashed curves are least-squares fits based on Eq. (8) for $k \geq 4$ with parameters $(C, \tilde{\mu}) = (0.45, 0.99)$ for $\gamma = 0.75$, $(0.59, 0.90)$ for $\gamma = 0.5$, and $(0.59, 0.75)$ for $\gamma = 0.25$. Inset: the same data on a double logarithmic scale. (b) The local slopes from a plot of $\ln[\ln(k^\gamma n_k)]$ versus $\ln k$; the arrows show the expected asymptotic result from Eq. (3), $\ln k^\gamma n_k \sim -k^{1-\gamma}$.

5. Unattainability of Superlinear Preferential Attachment

From Eq. (6), we see that A_k will be superlinear in k only if t_k can grow as a power law in k. Since t_k is a probability that must be less than one, any nearest-neighbor redirection algorithm cannot produce superlinear preferential attachment. In redirection, an incoming node can attach to an arbitrary node $\mathbf{x}$ either directly or by attaching to one of the $k-1$ children of $\mathbf{x}$ and then redirecting to $\mathbf{x}$. Thus there are a maximum k ways for the new node to attach to node $\mathbf{x}$. Redirection cannot provide any additional k-dependent amplification beyond the local environment size because the factor t_k in (6) is bounded as $k \to \infty$. Therefore the attachment rate to $\mathbf{x}$ cannot grow faster than linearly in its degree by nearest-neighbor redirection.

The inability of nearest-neighbor redirection to produce superlinear preferential attachment suggests attaching to more distant ancestors of a node. It is natural to consider grandparent redirection [28] in which the incoming node attaches to a randomly-selected target node with probability $1 - r(a,b,c) - s(a,b,c)$, to the parent of the target with probability $r(a,b,c)$, and to the grandparent of the target with probability $s(a,b,c)$, where a, b, and c are the degrees of the target, parent, and grandparent nodes, respectively.

Following the same steps as in our parent redirection algorithm, f_k and t_k are again the respective probabilities that an incoming link is: (a) redirected *from* a randomly selected target of degree k, and (b) *to* a parent node of degree k from an initially selected child node. We also introduce u_k as the probability a link is redirected to the *grandparent* of degree k from an initially selected grandchild of this node. These probabilities are formally defined as

$$
\begin{aligned}
f_k &= \sum_{b\geq 1}\sum_{c\geq 1} \frac{[r(k,b,c)+s(k,b,c)]\,N(k,b,c)}{N_k}\ , \\
t_k &= \sum_{a\geq 1}\sum_{c\geq 1} \frac{r(a,k,c)N(a,k,c)}{(k-1)N_k}\ , \\
u_k &= \sum_{a\geq 1}\sum_{b\geq 1} \frac{s(a,b,k)N(a,b,k)}{g_k N_k}\ ,
\end{aligned}
\tag{9}
$$

where the correlation function $N(a,b,c)$ is the number of nodes of degree a with a parent of degree b and a grandparent of degree c. As in nearest-neighbor redirection, f_k is the total redirection probability averaged over all N_k nodes of degree k and t_k is the parent redirection probability averaged over all $(k-1)N_k$ children whose parents have degree k. The function g_k is defined as the mean number of grandchildren of a node of degree k, and therefore, u_k is the grandparent redirection probability averaged over all $g_k N_k$ grandchildren of degree-k nodes.

Alternatively, we can define g_k based on c_k, the mean degree of a child of a degree k node. A node of degree k has $k-1$ children, and each child has $c_k - 1$ children, on average. Therefore, a node of degree k has an average of $g_k = (k-1)(c_k-1)$ grandchildren. Formally, g_k and c_k are given by

$$
g_k = \sum_{a\geq 1}\sum_{b\geq 1} N(a,b,k), \qquad\qquad c_k = \sum_{a\geq 1} aN(a,k). \tag{10}
$$

Using the identities $\sum_a N(a,b,c) = (b-1)N(b,c)$ and $\sum_a N(a,b) = (b-1)N_b$, we rearrange (10) to give the relation $g_k = (k-1)(c_k-1)$.

We may now write a master equation for the evolution of N_k in grandparent redirection:

$$
\begin{aligned}
\frac{dN_k}{dN} =& \frac{(1-f_{k-1})N_{k-1} - (1-f_k)N_k}{N} + \frac{(k-2)t_{k-1}N_{k-1} - (k-1)t_k N_k}{N} \\
&+ \frac{g_{k-1}u_{k-1}N_{k-1} - g_k u_k N_k}{N} + \delta_{k,1}\,.
\end{aligned}
\tag{11}
$$

The first two ratios are the same as in Eq. (5), and correspond to instances where an incoming node links to the target or to the parent of the target. The third ratio is specific to grandparent redirection and corresponds to the instances where a link is redirected to the grandparent of a target node. For example, $g_k u_k N_k/N$ is the probability that one of the $g_k N_k$ grandchildren of degree k nodes is initially targeted and the link is redirected to the degree-k grandparent.

Rearranging terms and eliminating g_k in favor of c_k, we may express the master equation in the canonical form of (1), with the attachment rate

$$
\frac{A_k}{A} = \frac{(k-1)(c_k-1)u_k + (k-1)t_k + 1 - f_k}{N}\ . \tag{12}
$$

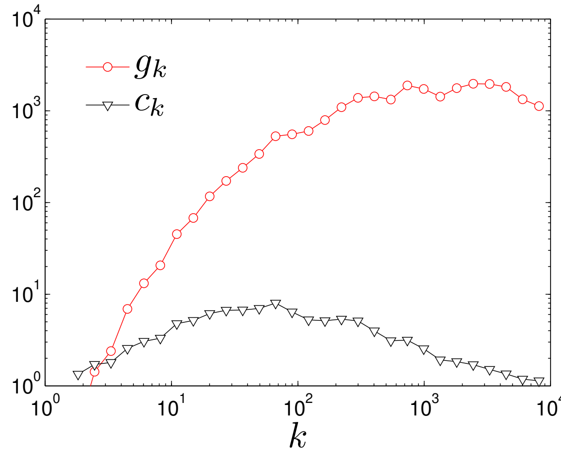

Figure 3. Dependence of g_k ($\circ$) and c_k ($\bigtriangledown$) versus k from direction simulations of superlinear preferential attachment growth for $\gamma = 1.3$. Data represents 10^3 network realizations of size $N = 10^4$. Qualitatively similar results are obtained for other values of $\gamma > 1$.

If c_k grew as a power law in k, superlinear preferential attachment would result. Direct simulations of superlinear preferential attachment networks of up to $N = 10^4$ nodes show instead that c_k asymptotically approaches 1 as $k \to N$ (Fig. 3). This behavior arises because of the shallowness of the underlying genealogical tree in this class of networks [13, 14, 16, 15]. Typically, such a network contains a single hub node whose degree is of order N that connects to order N 'leaf' nodes of degree 1. Moreover, the number of nodes with degrees larger than, but of the order of, one grows at most sublinearly with N [11]. Consequently, the mean degree of a child of the hub asymptotes to 1. Similarly, as their degree increases, non-hub nodes will be connected to a larger fraction of leaves. Thus c_k decreases with k for large k, so that the average number of grandchildren, $g_k = (k-1)(c_k-1)$, grows sublinearly with k (Fig. 3).

Grandparent redirection illustrates a basic shortcoming of any local growth rule in the quest to produce superlinear preferential attachment network growth. Consider a arbitrary node $\mathbf{x}$ in a network that grows by some local redirection attachment rule. The rate at which a new node attaches to $\mathbf{x}$ is limited by the size of its local environment, now defined as the set of nodes that, if initially targeted by an incoming node, can ultimately lead to attachment to $\mathbf{x}$. For nearest-neighbor redirection, the local environment of $\mathbf{x}$ consists of $\mathbf{x}$ and its children, and has size k if $\mathbf{x}$ has degree k. In grandparent redirection, the local environment of $\mathbf{x}$ consists of its children and grandchildren, so that its size equals $k + g_k$. Because g_k grows sublinearly in k, grandparent redirection cannot give superlinear preferential attachment.

To achieve superlinear preferential attachment, the size of the local environment of a node must grow superlinearly in the node degree. For a hub, this condition leads to the impossible situation that the local environment must have a size of order N^γ, which is larger than the entire network.

6. Conclusion

We presented a degree-dependent nearest-neighbor redirection algorithm to generate complex networks. We showed how this algorithm is equivalent to network growth by sublinear preferential attachment when the redirection probability is a suitably-chosen function of the degrees of the target and parent nodes. By exploiting only this local information in the immediate vicinity of the target node, we constructed sublinear preferential attachment networks extremely efficiently because just a few computer instructions are needed to create each new network node. We also argued that no local redirection rule can generate superlinear preferential attachment. The prevalence of linear and sublinear preferential attachment networks, along with the relative scarcity of superlinear preferential attachment networks, suggests that real-world networks should grow only according to local growth rules.

Acknowledgments

We gratefully acknowledge financial support from grant #FA9550-12-1-0391 from the U.S. Air Force Office of Scientific Research (AFOSR) and the Defense Advanced Research Projects Agency (DARPA).

References

[1] G. U. Yule, Phil. Trans. Roy. Soc. B **213**, 21 (1924).
[2] H. A. Simon, Biometrica **42**, 425 (1955).
[3] R. K. Merton, Science **159**, 56 (1968).
[4] Price, D. J. de S., Science **149**, 292 (1976).
[5] A. Barabási and R. Albert, Science **286**, 509 (1999).
[6] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[7] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, UK, 2003).
[8] M. E. J. Newman, A.-L. Barabási, and D. J. Watts, *The Structure and Dynamics of Networks* (Princeton University Press, Princeton, NJ, 2006).
[9] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, UK, 2008).
[10] M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2010).
[11] P. L. Krapivsky, S. Redner, F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000); P. L. Krapivsky and S. Redner, Phys. Rev. E **63**, 066123 (2001).
[12] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000).
[13] M. Gonen, S. Bar, and A. Wool, in *Passive and Active Network Measurement*, Lecture Notes in Computer Science vol. 3015, eds. C Barakat, and I. Pratt (Springer, Berlin, 2004).
[14] R. Oliverira and J. Spencer, *Internet Mathematics* **2**, 121 (2005).
[15] P. L. Krapivsky and D. Krioukov, Phys. Rev. E **78**, 026114 (2008).
[16] P. L. Krapivsky, S. Redner, and E. Ben-Naim, *A Kinetic View of Statistical Physics* (Cambridge University Press, Cambridge, 2010).

[17] J. Kleinberg, R. Kumar, P. Raphavan, S. Rajagopalan, and A. Tomkins, in: Lecture Notes in Computer Science, vol. 1627 (Springer-Verlag, Berlin, 1999).
[18] A. Vazquez, Phys. Rev. E **67**, 056104 (2003).
[19] P. L. Krapivsky and S. Redner, J. Phys. A **35**, 9517 (2002).
[20] B. Machta and J. Machta, Phys. Rev. E **71**, 026704 (2005).
[21] J. Chayes, S. Khanna, C. Borgs, M. Brautbar, and B. Lucier, ArXiv:1202.6033 (2012).
[22] Z. Neda, H. Jeong, and A. Barabási, Europhys. Lett. **61** 4, 567 (2003).
[23] M. E. J. Newman, S. Forrest, and J. Balthrop, Phys. Rev. Lett. **66**, 035101 (2002).
[24] M. Tomassini and L. Luthi, Physica A **385**, 750 (2007).
[25] L. E. C. Rocha, F. Liljeros, and P. Holme, Proc. Natl. Acad. Sci. **107** 5706 (2010).
[26] M. Golosovsky and S. Solomon, Phys. Rev. Lett. **109**, 098701 (2012).
[27] H. Rozenfeld and D. ben-Avraham, Phys. Rev. E **70**, 056107 (2004).
[28] Redirection to more distant ancestors was first considered in E. Ben-Naim and P. L. Krapivsky, J. Stat. Mech. P06004 (2010).