

Anomalous scaling in redirection networks

Harrison Hartle ^{1,*}, P. L. Krapivsky ^{2,1,†}, S. Redner ^{1,‡} and Yuanzhao Zhang ^{1,§}

¹*Santa Fe Institute, Santa Fe, New Mexico 87501, USA*

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



(Received 1 April 2026; accepted 4 September 2026; published 25 September 2026)

In networks that grow by isotropic redirection (IR), a new node selects an initial target node uniformly at random and attaches to a randomly chosen neighbor of the target. The emerging networks exhibit leaf proliferation, in which the number of nonleaves scales sublinearly as N^μ and the degree distribution has an algebraic tail with exponent $1 + \mu$. To understand these mysterious properties, we introduce a class of models with redirection to leaves whenever possible. The resulting networks exhibit qualitatively similar phenomenology to IR networks, but avoid the inherent nonlocality of the IR growth rule. These networks admit an analytical description of the leaf degree distribution, from which we extract the exponent μ .

DOI: [10.1103/PhysRevE.114.034314](https://doi.org/10.1103/PhysRevE.114.034314)

I. INTRODUCTION

In the rich zoo of complex networks, two models stand out for their simplicity and richness: random recursive trees (RRTs) [1–5] and preferential attachment (PA) networks [6–11]. In RRTs, new nodes sequentially join the network by attaching to preexisting nodes uniformly at random. In PA networks, new nodes attach to preexisting nodes with a probability that is an increasing function of the degree of the target node. The case where the attachment probability increases linearly with the degree of the target is especially popular because this rule leads to scale-free networks where the degree distribution decays algebraically with a nonuniversal exponent.

An unexpected feature of PA networks is that it is not necessary to know the entire set of node degrees to determine where a new node should attach. The mechanism of redirection [12–17] provides an algorithmically efficient way to generate preferential attachment by a simple and local deformation of the RRT growth rule. This redirection algorithm gives PA networks that grow by both linear [13] and sublinear [18] preferential attachment. Thus, a purely local growth rule implicitly encodes the global information that is ostensibly needed to construct PA networks.

Redirection has even more intriguing consequences for undirected trees grown by the following rule:

- (1) Select a preexisting node uniformly at random.
- (2) Attach to a random neighbor of the selected node.

This rule is known as isotropic redirection (IR); see Fig. 1. Redirection-based mechanisms are akin to several real-world network formation processes; for instance, in citation graphs, an author may cite a paper found in the bibliography of another paper [19]; in social networks, a significant tie may

form via a weak-tie intermediary acquaintance [20]. Hereinafter, we assume that each new node attaches only to a single preexisting node, so that the resulting network is a tree. This isotropic redirection rule leads to highly modular trees in which almost all nodes are leaves (nodes of degree 1), while the number $\mathcal{C}$ of ‘core’ nodes (nonleaves) scales sublinearly with the total number of nodes N : $\mathcal{C} = O(N^\mu)$, with $\mu \approx 0.566$ [15]. Moreover, the degree distribution of the core nodes is extraordinarily heavy-tailed:

$$N_k \sim \frac{N^\mu}{k^{1+\mu}}, \quad (1)$$

where N_k is the number of nodes of degree $k \geq 2$. The exponent value $\mu \approx 0.566$ seems paradoxical because a power-law distribution with exponent $1 + \mu < 2$ has a diverging first moment, whereas any tree obeys the structural constraint $\sum_{k \geq 1} k N_k = 2(N - 1)$ [1], implying a constant average degree. This apparent contradiction is resolved by the fact that the number of core nodes of any degree scales sublinearly with N ; that is, $N_k \sim N^\mu$, with exponent shifted by one from the degree exponent. Since the number of leaves N_1 is $N_1 = N - \mathcal{C} = N - O(N^\mu)$, we have

$$\sum_{k=1}^N k N_k = N_1 + \sum_{k=2}^N k N_k \sim N + \sum_{k=2}^N \frac{N^\mu}{k^\mu} \sim N,$$

so that (1) is indeed consistent. This behavior of the degree distribution dramatically contrasts with typical sparse networks in which the degree distribution is extensive, $N_k \sim N/k^\nu$ with $\nu > 2$.

In spite of the algorithmic simplicity of the IR growth rule, analytical progress has been stymied by its nonlocal nature. Namely, the probability w_i to attach to a given node i depends on the degrees of all its neighbors:

$$w_i \propto \sum_{j \sim i} \frac{1}{k_j}, \quad (2)$$

where k_j is the degree of node j , and $j \sim i$ denotes the neighbors j of node i . Because of this nonlocality, determining the

*Contact author: hhartle@santafe.edu

†Contact author: pkrapivsky@gmail.com

‡Contact author: redner@santafe.edu

§Contact author: yuanzhao.zhang.1@gmail.com

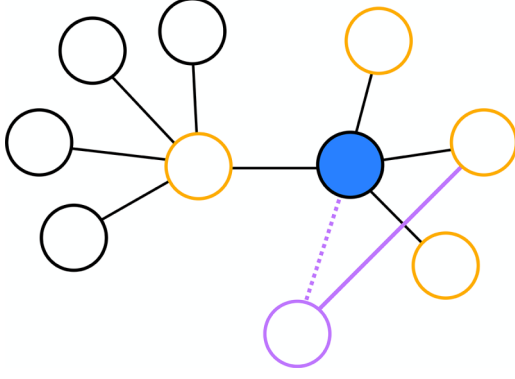


FIG. 1. Illustration of isotropic redirection (IR). A new node (purple circle) selects a preexisting target node (blue) uniformly at random (dashed line). The new node attaches (solid line) to one of the neighbors of the selected node (orange circles) uniformly at random.

degree distribution requires the joint degree distribution $N_{k,k'}$, which, in turn, requires higher-order degree correlations, etc. What is known rigorously about this model is that the fraction of leaves approaches 1 as $N \rightarrow \infty$ [21] and that $\mu \leq 0.9$ [22].

In this work, we avoid the complications caused by the nonlocality of the IR growth rule by introducing closely related growing network models that turn out to have qualitatively similar phenomenology to the IR network. In particular, we consider models in which redirection from nonleaves is only to leaves, hence avoiding any considerations of degree correlations. These models are analytically tractable via approaches recently developed in the context of growing unlabeled trees [23], leaf-based growth models [24], and the PA tree [25]. An essential aspect of these approaches is to consider the leaf degree of nodes, i.e., the number of leaves that a given node is attached to [26,27]. In particular, we derive and solve recurrences for how the distribution of leaf degree evolves under network growth, akin to conventional considerations of the degree distribution.

In Sec. II, we introduce models of growing trees with leaf-based redirection mechanisms. In Sec. III, we analyze these models, including analytical determination of the exponent μ that controls the sublinear core growth and degree distribution tail. In Appendix A, we estimate the asymptotic behavior of the core-size distribution by calculating the probability of star graphs and line graphs. In Appendix B, we discuss an extreme version of one of our network models that exhibits quite peculiar phenomenology, at the boundary between anomalous and nonanomalous scaling. In Appendix C, we present a toy two-population model that mimics some of the essential features of the growing network models, providing further insight into the absence of self-averaging in the network models.

II. MODELS

To define our models, it is useful to partition the network as shown in Fig. 2. Leaves (green) with degree 1 form the network periphery. Attached to leaves are rank-1 nodes (blue); by definition, rank-1 nodes are a distance of 1 from the periphery. We define the nucleus of the network (red outline) as nodes that are more than a unit distance from the periphery;

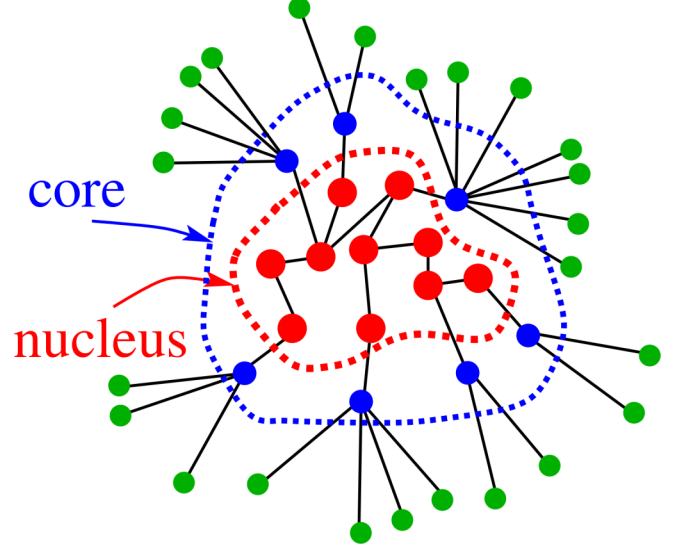


FIG. 2. A tree of 41 nodes that consists of 24 leaves (green), seven rank-1 nodes (blue), and $N = 10$ rank >1 nodes (red) that comprise the nucleus. The nucleus plus rank-1 nodes form the core of $C = 17$ nodes (within the blue outline).

these are nodes with rank >1 , also known as protected nodes [28,29]. The nucleus and rank-1 nodes together comprise the network core (blue outline). In terms of this picture, we define the following network growth models.

Direct attachment to the nucleus (DAN):

- (1) If a leaf is selected, attach to its rank-1 neighbor.
- (2) If a rank-1 node is selected, attach to its leaf.
- (3) If a nucleus node is selected, attach to this node.

Prohibited attachment to the nucleus (PAN):

- (1) If a leaf is selected, attach to its rank-1 neighbor.
- (2) If a rank-1 node is selected, attach to its leaf.
- (3) If a nucleus node is selected, reject this event.

Both processes resemble the IR growth rule, with two modifications: In rule 2, redirection from rank-1 nodes is always to leaves in both the DAN and PAN models, whereas in IR, redirection is isotropic among all neighboring leaves and core nodes; in rule 3, the attachment does not involve redirection. In both the DAN and PAN models, if a nucleus node is initially selected, redirection to a neighboring node is not allowed; instead, it undergoes direct attachment (DAN) or rejects the event (PAN). In the IR model, redirection from a core node can be to another core node; this is the process that leads to higher-body correlations in the underlying rate equations for the IR model. Because core-to-core redirection is forbidden in both the DAN and PAN models, they turn out to be analytically tractable while still exhibiting qualitatively similar behavior to the IR model.

The DAN and PAN models can be subsumed into a general class of models that we term variable attachment to the nucleus $\text{VAN}(\omega)$. Here, nucleus nodes are assigned a weight ω , and all other nodes have a weight of 1. For leaves and rank-1 nodes, the redirection mechanism is exactly the same as in the DAN and PAN models. If a nucleus node is selected as the target, which happens with probability $\omega/(\omega N + N - N)$, where N is the number of nodes in the nucleus, the new node

TABLE I. Basic network observables.

N	total number of network nodes
$\mathcal{C}$	number of nodes in the core of the network
$\mathcal{N}$	number of nodes in the nucleus of the network
N_k	number of nodes of degree k
M_ℓ	number of nodes of leaf degree ℓ

attaches to this target. The DAN model corresponds to $\omega = 1$, and the PAN model is recovered when $\omega = 0$. In the extreme case of $\omega = \infty$, each nucleus node that arises is immediately attached to in the subsequent growth step, becoming rank-1.

Table I summarizes the basic network observables of the models defined in this work. These include the aforementioned N , $\mathcal{C}$, $\mathcal{N}$, and N_k , but also M_ℓ : the number of nodes of leaf degree ℓ .

Networks that are generated by the IR, DAN, and PAN mechanisms are indeed similar to one another in appearance and in structural details (Fig. 3). They also have striking qualitative similarities to some real networks, e.g., the transcriptional regulatory networks observed in Ref. [30]. Networks from these models exhibit the following essential features:

(1) Leaf proliferation: Almost all nodes are leaves, while the remaining core of the network grows as

$$\mathcal{C} = N - N_1 \sim N^\mu, \quad (3)$$

with $\mu < 1$.

(2) Non-self-averaging of $\mathcal{C}$: The size of the core is broadly distributed. For $N, \mathcal{C} \rightarrow \infty$, with $z = \mathcal{C}/\langle \mathcal{C} \rangle$ finite, the probability distribution of core size $P_N(\mathcal{C})$ approaches the scaling form

$$P_N(\mathcal{C}) = \langle \mathcal{C} \rangle^{-1} \mathcal{P}(z), \quad z = \mathcal{C}/\langle \mathcal{C} \rangle. \quad (4)$$

Similar non-self-averaging holds for the variables N_k ($k > 1$), M_ℓ ($\ell > 0$), and $\mathcal{N}$.

(3) Algebraic decay of N_k and M_ℓ : The average values of N_k and M_ℓ satisfy

$$\langle N_k \rangle \sim N^\mu k^{-1-\mu}, \quad k \gg 1,$$

$$\langle M_\ell \rangle \sim N^\mu \ell^{-1-\mu}, \quad \ell \gg 1, \quad (5)$$

for the same exponent $\mu < 1$ as in (3). The equality of the degree and leaf degree tail exponents is a consequence of the leaf proliferation phenomenon.

Our main results determine the exponent μ : for the DAN model, we show that μ is the root of the transcendental equation

$$1 - \mu^2 = \int_0^1 dx x^\mu e^{x-1}, \quad (6a)$$

while for the PAN model, we show that μ is instead determined by

$$1 - \mu = \int_0^1 dx x^\mu e^{x-1}. \quad (6b)$$

The numerically exact values for the exponent μ are obtainable by the solution of (6). To six decimal places, the values are $\mu \approx 0.771\,192$ (DAN) and $\mu \approx 0.549\,735$ (PAN). Additionally, we determine the tail behavior of the scaled distribution of core size (4), namely,

$$\mathcal{P}(z) \sim z^{-1+1/\mu} \quad z \rightarrow 0, \quad (7a)$$

while for $z \rightarrow \infty$

$$-\ln \mathcal{P}(z) \sim \begin{cases} z^{1/(1-\mu)} \ln z & \text{(DAN)} \\ z^{1/(1-\mu)} & \text{(PAN)}, \end{cases} \quad (7b)$$

as derived in Appendix A. Furthermore, this work clarifies the nature of self-averaging in these models: ratios of the form $\mathcal{N}/\mathcal{C}$, $M_\ell/\mathcal{C}$ ($\ell > 0$), and $N_k/\mathcal{C}$ ($k > 1$) are asymptotically self-averaging—approaching nonrandom limiting values, denoted q_0 , q_ℓ , and c_k , respectively—despite the absolute quantities in the numerators and denominators being non-self-averaging.

III. THEORETICAL ANALYSIS

In this section, we establish recurrences for the leaf degree distributions of the models introduced in Sec. II, resulting in the implicit formulas (6) for the anomalous exponent μ that quantifies the core size growth and the degree distribution tail. The variables of interest are $\mathcal{N}$ (the number of nodes

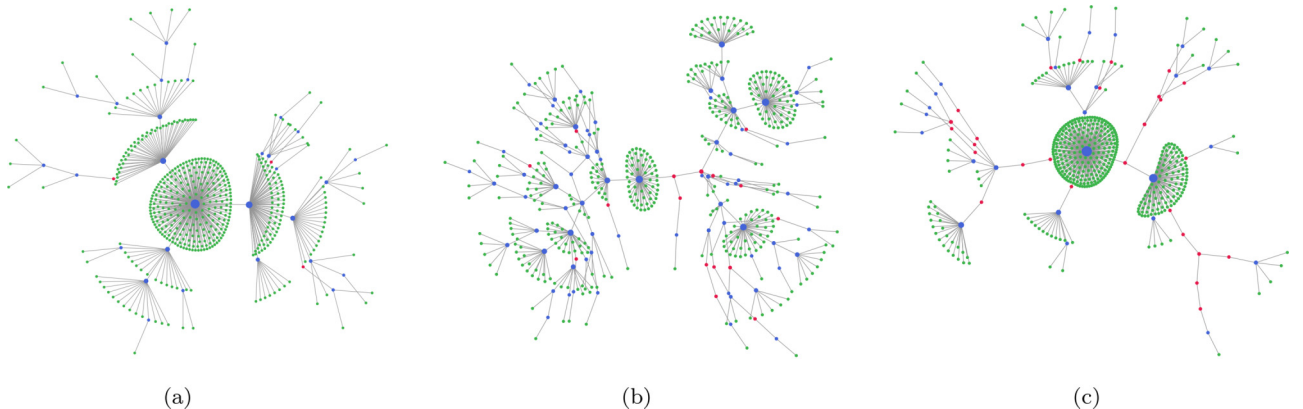


FIG. 3. Random trees grown by: (a) the IR model, (b) the DAN model, and (c) the PAN model. Node colors are as in Fig. 2.

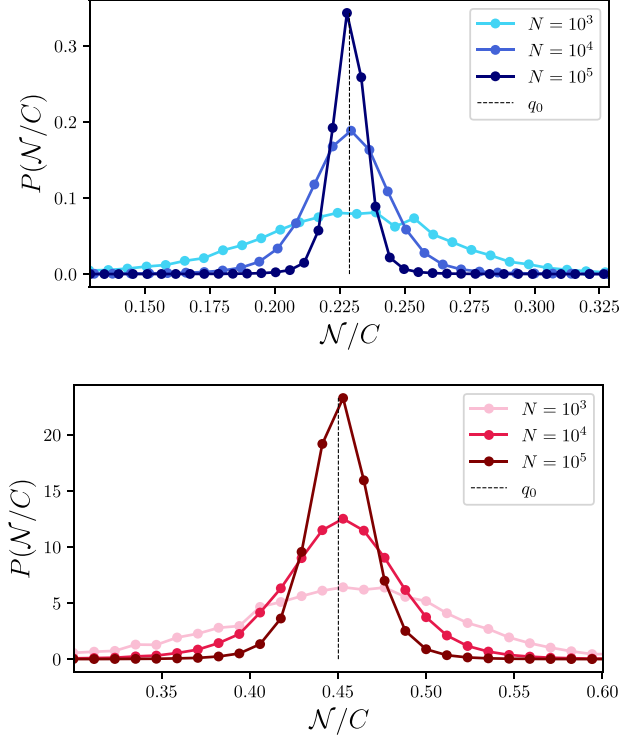


FIG. 4. Illustration of self-averaging for the nucleus-to-core ratio. The scaled probability distribution $P(N/C)$ for the ratio of nucleus to core size for a given network realization N/C as a function of the number of nodes N . The distribution converges to a delta function that is located at the deterministic value q_0 . Shown are data for the DAN model (upper panel) and the PAN model (lower panel).

in the nucleus), $N_1 = N - C$ (the number of leaves), and M_ℓ (the number of nodes with leaf degree ℓ , for $\ell = 1, 2, 3, \dots$). Nodes with leaf degree zero consist of nodes in the nucleus and the leaves themselves, so that $M_0 = N_1 + \mathcal{N}$. Thus, the quantities $(M_0, N_1, \mathcal{N})$ are not all independent, and we can choose any pair of them to characterize the network. We choose $(N_1, \mathcal{N})$, together with M_ℓ for $\ell > 0$, as the fundamental variables in what follows. It will also be useful to exploit the sum rules obeyed by the leaf degree distribution:

$$\sum_{\ell \geq 0} M_\ell = N, \quad \sum_{\ell \geq 1} \ell M_\ell = N_1. \quad (8)$$

The first equation states that the number of nodes with any leaf degree must equal the total number of nodes; the second equation expresses the total number of leaves through the leaf degree distribution.

The variables M_ℓ , N_k , $\mathcal{N}$, and C remain asymptotically non-self-averaging, namely, relative fluctuations between different network realizations do not vanish in the $N \rightarrow \infty$ limit. On the other hand, we find that ratios of the above basic random quantities are asymptotically self-averaging. We analyzed some representative such quantities, particularly $\mathcal{N}/C$, M_ℓ/C with small $\ell > 0$, and N_k/C with small $k > 1$. Figure 4 illustrates the asymptotic self-averaging of the ratio of nucleus size $\mathcal{N}$ to core size C for both the DAN and PAN models. The high correlation of M_ℓ with C results in ratios which are

asymptotically deterministic:

$$\frac{\mathcal{N}}{C} \rightarrow q_0, \quad \frac{M_\ell}{C} \rightarrow q_\ell. \quad (9)$$

This notion of asymptotic self-averaging is the basis of the following analysis. We also observe convergence of the degree distribution within the core of the form $N_k/C \rightarrow c_k$ for $k > 1$, but our analysis pertains to the leaf degree distribution.

In the DAN and PAN models, and more generally, the $\text{VAN}(\omega)$ model class, we will see that the limiting ratios $\{q_\ell\}_{\ell \geq 0}$ satisfy closed and solvable recurrences. In contrast, for the IR model, the fractions q_ℓ do not satisfy a closed recurrence. Moreover, the fractions $c_k = N_k/C$ do not satisfy closed recurrences for any of these models. This simplification in the DAN and PAN models highlights the importance of focusing on the leaf degree as a fundamental dynamical variable. We note that closed two-index recurrences for the joint distribution of degree and leaf degree [25] (within the core) are obtainable for the DAN, PAN, and $\text{VAN}(\omega)$ model, but appear not to be for IR.

A. DAN model

For the DAN model, the quantities $(N_1, \mathcal{N})$ evolve according to

$$\frac{dN_1}{dN} = \frac{N_1 + \mathcal{N}}{N}, \quad \frac{d\mathcal{N}}{dN} = \frac{M_1 - \mathcal{N}}{N}. \quad (10)$$

For ease of notation in (10) and all similar formulas, we denote $\frac{dF}{dN} := F(N+1) - F(N)$. That is, the derivative notation signifies the change in a quantity $F = F(N)$ when one new node is added to the network. The equation for dN_1/dN expresses that the number of leaves increases by 1 whenever a leaf or nucleus node is initially selected. In the former case, the new node attaches to a rank-1 node; in the latter, direct attachment to the nucleus node occurs. In both cases, the number of leaves increases by 1. The equation for $d\mathcal{N}/dN$ expresses that whenever a new node directly attaches to a nucleus node, the latter becomes a rank-1 node and is removed from the nucleus. When a new node first selects a rank-1 node with a single leaf, redirection to the leaf occurs. In this case, the number of nucleus nodes increases by 1.

The number of nodes with leaf degree equal to 1, M_1 , grows with N according to

$$\frac{dM_1}{dN} = \frac{N - N_1 - 2M_1 + M_2}{N}. \quad (11)$$

Four distinct processes contribute to dM_1/dN , as illustrated in Fig. 5. (a) A rank-1 node with leaf degree ≥ 3 is selected and redirection to a leaf occurs; here, M_1 increases by 1. (b) A rank-1 node with leaf degree 2 is selected and redirection to the leaf occurs; here, M_1 increases by 2. (c) A leaf that is attached to a rank-1 node is selected, and redirection to the rank-1 node occurs; here, M_1 decreases by 1. (d) A nucleus node is selected, and attachment to this node occurs; M_1 increases by 1. Assembling these processes, using the sum rule $\sum_\ell M_\ell = N$, as well as $M_0 = \mathcal{N} + N_1$, then gives the right-hand side of Eq. (11).

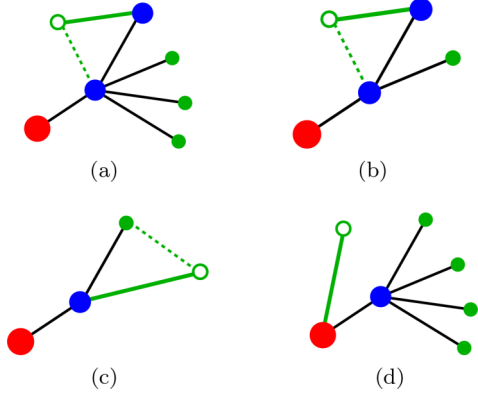


FIG. 5. The four processes that contribute to the change in M_1 . The new node, necessarily a leaf, is indicated by the open green circle. The dashed green line indicates the initially selected node, and the solid green line indicates the attachment event. The blue node is rank-1, and the red node is protected (other neighbors not displayed). In (a) and (b), the new attachment turns a leaf into a rank-1 node.

Finally, the leaf degree distribution M_ℓ with $\ell \geq 2$ satisfy the infinite system of equations

$$\frac{dM_\ell}{dN} = \frac{(\ell-1)M_{\ell-1} + M_{\ell+1} - (\ell+1)M_\ell}{N}. \quad (12)$$

The three processes that contribute to dM_ℓ/dN are the following: (a) Select one of the $\ell-1$ leaves neighboring a rank-1 node with leaf degree $\ell-1$ and attach to that rank-1 node; in this case, $M_\ell \rightarrow M_\ell + 1$. (b) Select a rank-1 node of leaf degree $\ell+1$ and attach to one of the leaves of this node; again, $M_\ell \rightarrow M_\ell + 1$. (c) Either select one of the ℓ leaves whose rank-1 neighbor has leaf degree ℓ and attach to the rank-1 node, or select a rank-1 node of leaf degree ℓ and attach to one of its leaves; the result is $M_\ell \rightarrow M_\ell - 1$.

A linear ansatz of the form $M_\ell \simeq m_\ell N$ results in the trivial condition $m_\ell = 0$ for all $\ell > 0$, and $m_0 = 1$, indicative of leaf proliferation. Instead, a sublinear growth ansatz is taken; Fig. 6 confirms the sublinear growth of the average size of the core, $\langle C \rangle \sim N^\mu$, with numerically estimated exponent μ close to the values we analytically derive in what follows. Introducing the sublinear scaling ansatz and limiting fractions q_ℓ of (9) into (12), the system reduces to recurrence

$$\mu q_\ell = (\ell-1)q_{\ell-1} + q_{\ell+1} - (\ell+1)q_\ell, \quad \ell \geq 2, \quad (13)$$

while Eqs. (10) and (11) become

$$\begin{aligned} q_0 &= 1 - \mu, & q_1 &= 1 - \mu^2, \\ \mu q_1 &= 1 - 2q_1 + q_2. \end{aligned} \quad (14)$$

Substituting (14) into (13), one can express q_ℓ as polynomials in μ of degree $\ell+1$. For example, the first few of these polynomials are

$$\begin{aligned} q_2 &= 1 + \mu - 2\mu^2 - \mu^3, \\ q_3 &= 2 + 4\mu - 4\mu^2 - 5\mu^3 - \mu^4, \\ q_4 &= 6 + 16\mu - 8\mu^2 - 22\mu^3 - 9\mu^4 - \mu^5, \end{aligned}$$

etc. In general, solutions to (13) grow factorially with ℓ . Indeed, Eq. (13) becomes $q_{\ell+1} = (\ell + \mu + 1)q_\ell$ to leading

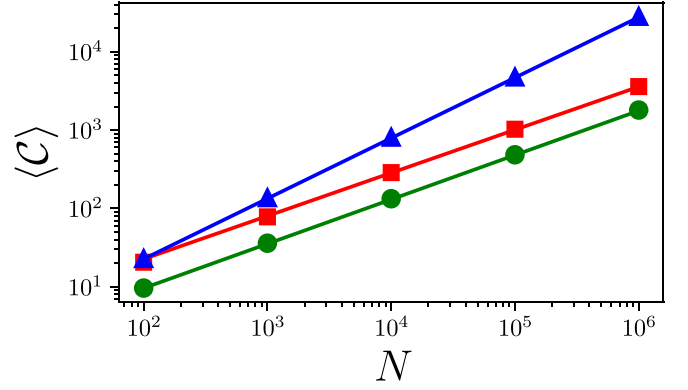


FIG. 6. The average number of core nodes $\langle C \rangle$ as a function of the number of network nodes from $N = 10^2$ to $N = 10^6$ for the DAN ($\blacktriangle$) and PAN ($\blacksquare$) models, as well as for the IR model ($\bullet$). Each point represents the mean of C over 2400 independent realizations. A non-linear least-squares fit provides accurate estimates for the exponents: $\mu \approx 0.569$ for the IR model, and $\mu \approx 0.773$ and 0.545 for DAN and PAN, respectively, aligning approximately with the analytically computed values.

order. This fact suggests that we seek a solution of the form $q_\ell = Q_\ell \Gamma(\ell + \mu + 1)$. Substituting this ansatz into Eq. (13), we find a simple recurrence for Q_ℓ that has an asymptotic solution $Q_\ell \simeq C(\mu)/(\ell + \mu)$. Hence the growth of q_ℓ is indeed factorial: $q_\ell \simeq C(\mu)\Gamma(\ell + \mu)$. For specific values of μ , the amplitude $C(\mu)$ vanishes, and a second linearly independent solution of the linear three-term recurrence (13) determines the asymptotic. The asymptotic of this minimal at infinity solution turns out to be algebraic, so it can be extracted using a continuum framework where (13) simplifies to $\frac{d}{d\ell}(\ell q_\ell) = -\mu q_\ell$, leading to $q_\ell \sim \ell^{-1-\mu}$ for $\ell \gg 1$. Figure 7 shows the tail behavior of q_ℓ , corroborating the predicted algebraic decay.

Fortunately, the explicit solutions to the generic three-term recurrences of the form $q_{\ell+1} + a_\ell q_\ell + b_\ell q_{\ell-1} = 0$ that are minimal at infinity are known [31]. These solutions exhibit a remarkable representation in terms of the continued fractions [31]. In particular, for the generic three-term recurrence, the ratio q_2/q_1 is

$$\frac{q_2}{q_1} = -\frac{b_2}{a_2 - \frac{b_3}{a_3 - \frac{b_4}{a_4 - \dots}}}. \quad (15)$$

Recalling that $q_1 = 1 - \mu^2$ and $q_2 = 1 + \mu - 2\mu^2 - \mu^3$ in the DAN model, we find that the ratio q_2/q_1 is

$$\frac{q_2}{q_1} = 2 + \mu - \frac{1}{1 - \mu^2}.$$

Equating this result to (15), which we specialize to $a_\ell = -(\ell + 1 + \mu)$ and $b_\ell = \ell - 1$ as per the three-term recurrence (13), we obtain

$$1 - \mu^2 = F(\mu), \quad (16)$$

where we introduce the shorthand notation

$$\frac{1}{F(\mu)} = 2 + \mu - \frac{1}{3 + \mu - \frac{2}{4 + \mu - \frac{3}{5 + \mu - \dots}}}. \quad (17)$$

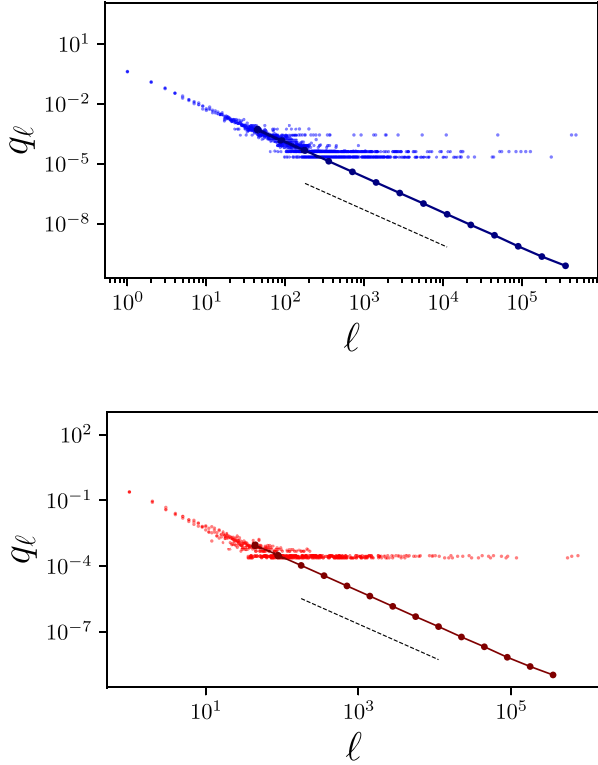


FIG. 7. Leaf degree distributions of the DAN model (upper panel) and PAN model (lower panel). The colored dots are the raw data (values of M_ℓ/C) for three graphs of size $N = 10^6$; the dotted curve represents log-binned data from 4800 graphs of size $N = 10^6$; the black dashed lines indicate pure power-law decay with slopes predicted by our theory: $1 + \mu_{\text{DAN}} \approx 1.771$ and $1 + \mu_{\text{PAN}} \approx 1.549$, respectively.

With the help of the handbook [32], we can express the continued fraction in terms of an integral representation. We set $a = 1 + \mu$ and $z = -1$ in identity (12.6.30) of [32], and then use the integral representation of the incomplete gamma function $\gamma(a, z) = \gamma(1 + \mu, -1)$ in the identity (12.6.30) to give

$$F(\mu) = \int_0^1 dx x^\mu e^{x-1}.$$

Combining this with (16), we obtain the transcendental equation for μ

$$1 - \mu^2 = \int_0^1 dx x^\mu e^{x-1}. \quad (18)$$

The numerically exact value of the exponent μ to six decimal places is $\mu \approx 0.771\,192$.

B. PAN model

For the PAN model, the quantities M_ℓ with $\ell \geq 2$ satisfy equations similar to Eqs. (12), except that the denominator N should now be replaced by $N - \mathcal{N}$. The quantity $\mathcal{N}$ is subleading in the $N \rightarrow \infty$ limit, so Eqs. (13) remain valid. The primary differences with the DAN model are Eqs. (10)

and (11) for N_1 , $\mathcal{N}$, and M_1 . First, we note that in each node addition, the target node is selected uniformly at random among all $N - \mathcal{N}$ non-nucleus nodes. By enumerating all the ways that N_1 and $\mathcal{N}$ can change when a new node joins the network, we arrive at the rate equations

$$\frac{dN_1}{dN} = \frac{N_1}{N - \mathcal{N}}, \quad \frac{d\mathcal{N}}{dN} = \frac{M_1}{N - \mathcal{N}}. \quad (19)$$

In the equation for dN_1/dN , the number of leaves increases by 1 whenever the new node first selects a leaf and then attaches to its rank-1 neighbor. In the equation for $d\mathcal{N}/dN$, the size of the nucleus increases by 1 whenever the new node selects a rank-1 node of leaf degree 1 and then attaches to its leaf.

We determine the rate equation for M_1 by the same detailed bookkeeping of the processes that contributed to the change in M_1 in the DAN model (Sec. III A). We thereby obtain

$$\frac{dM_1}{dN} = \frac{N - \mathcal{N} - N_1 - 2M_1 + M_2}{N - \mathcal{N}}. \quad (20)$$

This is nearly the same as Eq. (11) for M_1 in the DAN model, except there is now an extra loss term $-\mathcal{N}$. This term reflects the fact that attachment to a nucleus node is not allowed [the event in Fig. 5(d) is forbidden]. The corresponding equation for M_ℓ with $\ell \geq 2$ is

$$\frac{dM_\ell}{dN} = \frac{(\ell - 1)M_{\ell-1} + M_{\ell+1} - (\ell + 1)M_\ell}{N - \mathcal{N}}. \quad (21)$$

One may again demonstrate directly that a linear scaling ansatz yields the trivial outcome $M_\ell/N \rightarrow 0$ for $\ell > 1$, indicating sublinear growth, alongside $N_1/N \rightarrow 1$ indicating leaf proliferation; we proceed again with the ansatz $N_1 = N - O(N^\mu)$ and $M_\ell \sim N^\mu q_\ell$. From Eqs. (19) and (20), the analogs of Eqs. (14) are

$$\begin{aligned} q_0 &= 1 - \mu, \\ q_1 &= \mu - \mu^2, \\ \mu q_1 &= 1 - q_0 - 2q_1 + q_2. \end{aligned} \quad (22)$$

Therefore, $q_2 = \mu - \mu^2 - \mu^3$, and now using (22), the ratio q_2/q_1 is

$$\frac{q_2}{q_1} = 2 + \mu - \frac{1}{1 - \mu}.$$

Following the same steps as those used for the DAN model, we obtain the transcendental equation for μ ,

$$1 - \mu = F(\mu) = \int_0^1 dx x^\mu e^{x-1}. \quad (23)$$

The numerically exact value of the exponent μ , up to six decimal places, is $\mu \approx 0.549\,735$.

C. VAN(ω) model

For the VAN(ω) model, the growth of the number of leaves, the nucleus, and the number of nodes of each leaf degree value $\ell > 0$ are accounted for by the rate equations,

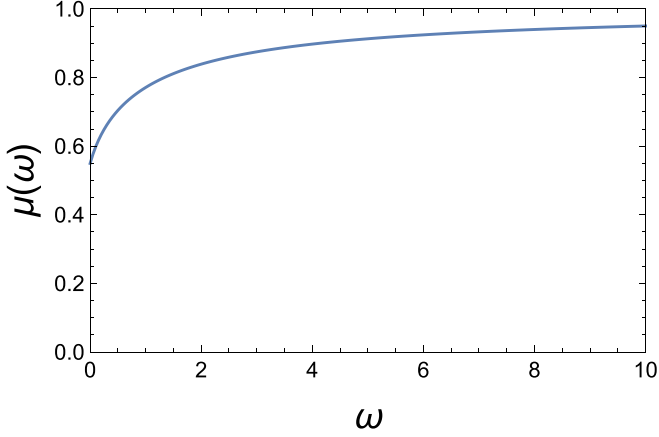


FIG. 8. The scaling exponent $\mu(\omega)$ of the $\text{VAN}(\omega)$ model. It monotonically increases from $\mu(0) = \mu_{\text{PAN}} \approx 0.549$ to $\mu(\infty) = 1$ with asymptotic behavior (27).

$$\begin{aligned} \frac{dN_1}{dN} &= \frac{N_1 + \omega\mathcal{N}}{N + (\omega - 1)\mathcal{N}}, \\ \frac{d\mathcal{N}}{dN} &= \frac{M_1 - \omega\mathcal{N}}{N + (\omega - 1)\mathcal{N}}, \\ \frac{dM_1}{dN} &= \frac{N + (\omega - 1)\mathcal{N} - N_1 - 2M_1 + M_2}{N + (\omega - 1)\mathcal{N}}, \\ \frac{dM_\ell}{dN} &= \frac{(\ell - 1)M_{\ell-1} + M_{\ell+1} - (\ell + 1)M_\ell}{N + (\omega - 1)\mathcal{N}}. \end{aligned} \quad (24)$$

The same approach as in the DAN and PAN models yields the same universal result $q_0 = 1 - \mu$ for the fraction of nucleus nodes in the core, while the ratios q_1/q_0 and q_2/q_1 are given by

$$\frac{q_1}{q_0} = \mu + \omega, \quad \frac{q_2}{q_1} = 2 + \mu - \frac{\omega + (1 - \omega)\mu}{(1 - \mu)(\mu + \omega)}. \quad (25)$$

Finally, the exponent μ is determined from the root of the transcendental equation

$$\frac{(1 - \mu)(\mu + \omega)}{\omega + (1 - \omega)\mu} = \int_0^1 dx x^\mu e^{x-1}. \quad (26)$$

The exponent $\mu = \mu(\omega)$ is a monotonic function of ω that increases from $\mu(0) = \mu_{\text{PAN}} \approx 0.549$ to $\mu(\infty) = 1$. The asymptotic behavior of the exponent in the $\omega \rightarrow \infty$ limit,

$$1 - \mu(\omega) \simeq \frac{1}{(e - 1)\omega}, \quad (27)$$

follows from (26); Fig. 8 displays $\mu(\omega)$.

D. $\text{VAN}(\infty)$ model

The limiting model $\text{VAN}(\infty)$ exhibits interesting marginal behavior. In this model, any protected node arising is immediately attached in the following growth step. The exponent $\mu(\infty) = 1$ leads to a quadratically decaying leaf degree tail, $q_\ell \sim \ell^{-2}$. The core grows only logarithmically slower than N :

$$\mathcal{C} \sim \frac{N}{\ln N}, \quad (28)$$

illustrating that $\text{VAN}(\infty)$ resides at the boundary between anomalous and nonanomalous scaling.

The governing recurrence for $\ell \geq 2$ arises by taking $\mu = 1$ in (13),

$$0 = (\ell - 1)q_{\ell-1} - (\ell + 2)q_\ell + q_{\ell+1}. \quad (29)$$

We derive an exact integral expression for q_ℓ , namely,

$$q_\ell = \frac{1}{e - 1} \int_0^1 dt e^{1-t} (1 - t)t^{\ell-1}. \quad (30)$$

Details of the $\text{VAN}(\infty)$ are provided in Appendix B.

IV. DISCUSSION

We introduced a class of network growth models based on redirection in which redirection from nonleaves to other nonleaves is forbidden. The resulting networks exhibit leaf proliferation and anomalously broad degree and leaf degree distributions—bearing qualitative similarity to isotropic redirection (IR)—yet these models proved more tractable analytically than IR. The IR growth rule is exceedingly simple: pick a node uniformly at random and attach to a random neighbor of this selected node. Despite this simplicity, the IR network has resisted analytical understanding because the growth rule is nonlocal in character. The models we introduced retain the spirit and phenomenology of the IR network, but the growth rule remains local in character so that the resulting network is analytically tractable. In particular, this locality arises due to redirection from nonleaf nodes always being to leaves; for nucleus nodes, no such redirection is possible, so we considered the case of direct attachment to the nucleus (DAN) and prohibited attachment to the nucleus (PAN). We also introduced a class of models, $\text{VAN}(\omega)$, by assigning a weight ω to nucleus attachment, which includes the DAN and PAN models as special cases, and has an interesting limiting case $\text{VAN}(\infty)$.

Our main results include (a) analytical predictions of the exponent $\mu < 1$ that controls the growth rate of the core: $\mathcal{C} \sim N^\mu$, (b) the ultrabroad leaf degree distribution: $M_\ell \sim N^\mu \ell^{-(1+\mu)}$, and (c) the tail behavior of the rescaled core-size distribution $\mathcal{P}(z)$ where $z = \mathcal{C}/\langle \mathcal{C} \rangle$. We especially emphasize result (d) that the ratios $M_\ell/\mathcal{C}$, $N_k/\mathcal{C}$, and $\mathcal{N}/\mathcal{C}$ are self-averaging, despite the numerators and denominators being individually non-self-averaging. More generally, our work has provided models and a tractable analytical framework for exploring anomalous scaling and leaf proliferation in growing random trees.

The networks arising in our models qualitatively resemble a variety of real-world networks, not limited to trees. Sparsity and degree heterogeneity, for instance, are commonplace characteristics of network data [5,11], but moreover, several studies have reported degree distributions with estimated power-law exponents below two [14,33]. Such exponents also appear in modeling studies beyond isotropic redirection [34–36], but are primarily associated with diverging average degree [37] rather than leaf proliferation. Phenomena similar to leaf proliferation have been shown to occur in other growth models beyond trees [38].

A class of related growth models is based on duplication mechanisms [39–42]. For instance, in the partial duplication

model (PDM) [43], a random target node is selected, and each neighbor of that target is connected to with probability q . Duplication has been used to model the evolution of protein interaction networks, with duplicate proteins having similar functional roles [44]. The PDM is more tractable than IR in that the attachment probability is degree-proportionate as opposed to being dependent on all neighbor degrees as in Eq. (2). We note that in the very rare event that the PDM produces a tree, requiring at every growth step that exactly one link forms, the resulting tree is statistically identical to a sample from the IR model. A variety of other duplication and copying mechanisms have also been considered, e.g., attachment to the target node itself in addition to its neighbors [16,45,46], or statistical mimicry of links in an underlying hidden network [47].

We have not explicitly computed c_k , the degree distribution within the core, which does not satisfy a closed recurrence akin to that for q_ℓ . However, by leaf proliferation, the degree distribution is also a power law with an identical exponent, i.e., $c_k \sim k^{-1-\mu}$. Despite c_k not satisfying a closed recurrence in $\text{VAN}(\omega)$, the joint distribution of degree and leaf degree within the core $c_{k,\ell}$ does, providing a path to compute c_k . Similarly, the joint distribution of degree and leaf degree was used to obtain the leaf degree distribution in the PA tree [25], where it does not satisfy a closed recurrence. Numerous additional questions are of interest in these models—for instance, a more detailed characterization of the nucleus, such as its distribution of degrees $c_{k,0}$ or the number of components it has, which can exceed one due to formation of protected nodes located away from the existing nucleus, or by fragmentation in the event of direct attachment to the nucleus. In addition, it would be of interest to obtain concentration and normality results, akin to those previously established for RRTs and PA [48,49], but in the context of the unusual form of self-averaging arising in these models.

DATA AVAILABILITY

There are no publicly available research data or software supporting this manuscript. Requests for further information or data should be sent to the authors.

APPENDIX A: STAR AND LINE GRAPHS AND THE SCALED CORE-SIZE DISTRIBUTION $\mathcal{P}(z)$

We confirmed that the probability distribution of the core size approaches the scaling form

$$P_N(C) = \langle C \rangle^{-1} \mathcal{P}(C/\langle C \rangle). \quad (\text{A1})$$

The scaled distribution $\mathcal{P}(C/\langle C \rangle)$ is asymptotically stationary and broad, namely, it does not approach a delta function that would indicate asymptotic self-averaging; see Fig. 9. To obtain the small- z behavior (7a) and large- z behavior (7b) of $\mathcal{P}(z)$, we analyze the extremal probabilities of star graphs and line graphs.

A star graph contains a single hub while all other nodes are leaves. For the RRT, the natural initial condition is $N = 3$, which is necessarily a star. Thus, $S_3 = 1$. The probability S_{N+1} of creating a star of $N + 1$ nodes from a star of N nodes

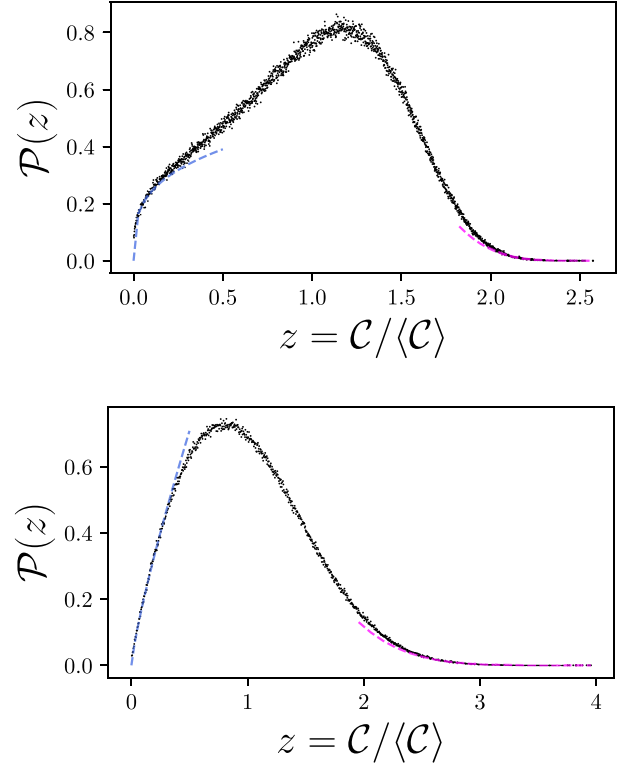


FIG. 9. Probability density of rescaled core size $\mathcal{P}(z)$ with $z = C/\langle C \rangle$, for graphs of size $N = 10^4$ from the DAN model (upper panel) and PAN model (lower panel), with data displayed from 2.4×10^6 independent realizations. Blue and pink dashed curves represent the asymptotics (7) with roughly estimated parameters.

satisfies $S_{N+1} = S_N/N$, from which

$$S_N = \frac{2}{(N-1)!}. \quad (\text{A2})$$

Similarly, for PA networks, the probability of creating a star of $N + 1$ nodes from a star of N nodes satisfies $S_{N+1} = S_N/2$, which yields

$$S_N = \frac{1}{2^{N-3}}. \quad (\text{A3})$$

On the other hand, for the DAN, PAN, and $\text{VAN}(\omega)$ models with arbitrary $\omega \geq 0$, as well as for the IR network, the probability of generating a star graph satisfies the recurrence $S_{N+1} = \frac{N-1}{N} S_N$, from which

$$S_N = \frac{2}{N-1}. \quad (\text{A4})$$

We now use this simple result to determine the small- z tail of the core-size distribution. Specializing (4) to $C = 1$ gives $P_N(1) = \langle C \rangle^{-1} \mathcal{P}(\langle C \rangle^{-1})$, which we equate to $S_N = P_N(1)$ given by (A4). Using $\langle C \rangle = AN^\mu$ gives

$$\mathcal{P}(z) \simeq 2A^{1/\mu} z^{-1+1/\mu}, \quad (\text{A5})$$

thus establishing (7a).

In contrast to the classic models of growing trees, the probability of creating a star in the DAN, PAN, and IR

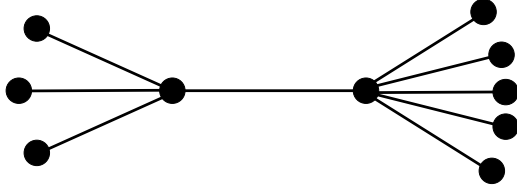


FIG. 10. The two-hub (3,5) graph.

networks decays very slowly with N . This observation provides a clue that typical network realizations grown by these rules should have non-negligible portions that are star-like, as seen in Fig. 3.

It is worth pointing out that the small- z tail of $\mathcal{P}(z)$ depends on the initial condition. To see how this dependence arises, consider the growth of a network starting with the linear graph L_4 with 4 nodes and core size $\mathcal{C} = 2$ instead of the linear graph L_3 , which was our canonical initial condition. Since the size of the core can only increase, the minimal core size remains equal to 2 as the network grows. A graph with core size 2 is necessarily a two-hub tree (Fig. 10). We denote by $H_{m,n}$ the probability of generating a two-hub (m, n) graph, that is, a tree with m leaves attached to one hub, n leaves attached to the other hub, and a single link between the hubs. For the DAN, PAN, and VAN(ω) models, the recurrence that determines the evolution of these probabilities when a new node attaches to the network is

$$H_{m,n} = \frac{m-1}{N-1} H_{m-1,n} + \frac{n-1}{N-1} H_{m,n-1}. \quad (\text{A6})$$

Solving the first few of (A6) subject to the initial condition $H_{1,1} = 1$ corresponding to the linear graph L_4 gives

$$\begin{aligned} H_{2,1} &= H_{1,2} = \frac{1}{4}, \\ H_{3,1} &= H_{2,2} = H_{1,3} = \frac{1}{10}, \\ H_{4,1} &= H_{3,2} = H_{2,3} = H_{1,4} = \frac{1}{20}. \end{aligned} \quad (\text{A7})$$

These probabilities clearly depend only on the total number of nodes: $H_{m,n} = H_N$ for all $N = m + n + 2$. This observation is readily proved by induction. The recurrence (A6) then simplifies to $H_N = \frac{N-4}{N-1} H_{N-1}$, which we solve subject to the initial condition $H_4 = 1$ to give

$$\begin{aligned} H_N &= \frac{6}{(N-3)(N-2)(N-1)} \\ &= \frac{6}{(m+n-1)(m+n)(m+n+1)}, \end{aligned} \quad (\text{A8})$$

for $N \geq 4$. The probability to build a two-hub tree of size N is then

$$P_N(2) = \sum_{m+n+2=N} H_{m,n} = \frac{6}{(N-2)(N-1)}. \quad (\text{A9})$$

Specializing the core-size distribution in (4) to $\mathcal{C} = 2$, the minimal size of the core when the network starts with the graph L_4 is $P_N(2) = \langle \mathcal{C} \rangle^{-1} \mathcal{P}(2/\langle \mathcal{C} \rangle)$. We now equate this to $P_N(2)$ in (A9) to give

$$\mathcal{P}(z) \simeq 12 \left(\frac{A}{2} \right)^{2/\mu} z^{-1+2/\mu}. \quad (\text{A10})$$

The essential point is that $\mathcal{P}(z)$ vanishes as $z^{-1+2/\mu}$ when network growth starts with the graph L_4 , rather than $\mathcal{P}(z) \simeq z^{-1+1/\mu}$ when growth begins with L_3 —the small- z tail of $\mathcal{P}(z)$ depends on the initial condition.

As a complement to the above presentation, we now focus on the large- z tail of $\mathcal{P}(z)$. We probe this limiting behavior by computing the probability Q_N to create networks with the minimal number of leaves—namely, the linear graph of length N , L_N . We consider only the canonical initial condition where the network is L_3 so that $Q_3 = 1$. By straightforward enumeration, we obtain $Q_4 = \frac{1}{3}$ and $Q_5 = \frac{1}{6}$ for all of our redirection models.

For the PAN model, the probabilities Q_N satisfy recurrence $Q_N = \frac{1}{2} Q_{N-1}$, from which

$$Q_N = \frac{2^{-(N-4)}}{3}. \quad (\text{A11})$$

We now substitute $Q_N = P_N(N-2) \simeq \langle \mathcal{C} \rangle^{-1} \mathcal{P}(N/\langle \mathcal{C} \rangle)$ and $\langle \mathcal{C} \rangle = AN^\mu$ in the above equation and eliminate N in favor of z . This gives the large- z asymptotic

$$\ln \mathcal{P}(z) \simeq -\ln 2 (Az)^{1/(1-\mu)}, \quad (\text{A12})$$

up to the unknown amplitude A in $\mathcal{C} \simeq AN^\mu$, as first quoted in Eq. (7b).

For the VAN(ω) model with $\omega > 0$, the recurrence satisfied by the probability Q_N to build the line graph L_N is $Q_{N+1} = \frac{2}{4+\omega(N-4)} Q_N$, from which

$$Q_N = \frac{1}{6} \left(\frac{2}{\omega} \right)^{N-5} \frac{\Gamma(1 + \frac{4}{\omega})}{\Gamma(N-4 + \frac{4}{\omega})}, \quad (\text{A13})$$

for $N \geq 5$. From this expression, we again obtain the same large- z asymptotic behavior for $\mathcal{P}(z)$ as for DAN, namely, $\ln \mathcal{P}(z) \sim -z^{\frac{1}{1-\mu}} \ln z$, for any $\omega > 0$.

APPENDIX B: VAN(∞) MODEL

The VAN(∞) model has unusual dynamics, residing at the boundary between linear and sublinear scaling. Since nodes in the nucleus have an infinite weight, whenever a nucleus node is created, it is immediately converted to a rank-1 node in the next attachment event. Thus, the nucleus is either empty or consists of a single node. The limiting exponent for VAN(∞) is $\mu(\infty) = 1$ with asymptotic approach (27) as $\omega \rightarrow \infty$, from which we anticipate tail behavior $q_\ell \sim \ell^{-2}$. Equivalently, the leaf degree distribution is expected to scale as $M_\ell \sim \mathcal{C}/\ell^2$. We now recall the identity

$$\sum_{\ell=1}^{N-1} \ell M_\ell = N_1. \quad (\text{B1})$$

The sum on the left-hand side of Eq. (B1) is dominated by the tail, so that

$$\sum_{\ell=1}^{N-1} \ell M_\ell \sim \mathcal{C} \sum_{\ell=1}^{N-1} \ell^{-1} \simeq \mathcal{C} \ln N. \quad (\text{B2})$$

Substituting this result, together with $N_1 = N - \mathcal{C}$, into Eq. (B1), we obtain

$$\mathcal{C} \sim \frac{N}{\ln N}. \quad (\text{B3})$$

Thus, the size of the core that is generated by $\text{VAN}(\infty)$ grows only logarithmically slower than N . Proceeding in the same heuristic manner, we deduce q_ℓ for all $\ell \geq 2$. Specializing (13) to $\mu = 1$ gives

$$(\ell + 2)q_\ell = (\ell - 1)q_{\ell-1} + q_{\ell+1}. \quad (\text{B4})$$

The minimal at infinity solution to this recurrence can be expressed via the generalized hypergeometric function

$$q_\ell = A\ell^{-2}F[1, \ell - 1; \ell + 1, \ell + 1; 1]. \quad (\text{B5})$$

Higher-order hypergeometric functions can be expressed through integrals over the lower-order ones. Specifically, we use the identity

$$F[a, c; b, d; 1] = \int_0^1 dt \frac{t^{c-1}(1-t)^{d-c-1}\Gamma(d)}{\Gamma(c)\Gamma(d-c)} F[a; b; t],$$

and recast Eq. (B5) into

$$q_\ell = A \frac{\ell - 1}{\ell} \int_0^1 dt t^{\ell-2}(1-t) F[1; \ell + 1; t], \quad (\text{B6})$$

where $F[1; \ell + 1; t]$ is the confluent hypergeometric function. Computing the integral, we obtain

$$q_\ell = q_1[1 - e(\ell - 1)\gamma(\ell, 1)], \quad (\text{B7})$$

where $\gamma(\ell, 1) = \int_0^1 du u^{\ell-1} e^{-u}$ is the incomplete gamma function. The amplitude q_1 is computed by normalization:

$$\frac{1}{q_1} = e \sum_{\ell=1}^{\infty} [\gamma(\ell, 1) - \gamma(\ell + 1, 1)] = e\gamma(1, 1) = e - 1. \quad (\text{B8})$$

The solution (B7) has integral representation

$$q_\ell = \frac{1}{e-1} \int_0^1 dt e^{1-t} (1-t)t^{\ell-1}. \quad (\text{B9})$$

Change of variable $t \rightarrow (1-t)\ell$ in (B9) yields the asymptotic $q_\ell \simeq q_1 \ell^{-2}$. Exact expressions for q_ℓ for any $\ell \geq 1$ can be extracted; for instance,

$$q_2 = \frac{3-e}{e-1}, \quad q_3 = \frac{11-4e}{e-1}, \quad q_4 = \frac{49-18e}{e-1},$$

suggesting the general form

$$q_\ell = \frac{A_\ell - B_\ell e}{e-1}, \quad (\text{B10a})$$

with integer A_ℓ and B_ℓ for all $\ell \geq 2$. The general expressions for A_ℓ and B_ℓ are

$$A_\ell = \sum_{j=0}^{\ell-1} (j+1)! \binom{\ell-1}{j}, \quad B_\ell = \ell! - (\ell-1)!. \quad (\text{B10b})$$

The exact representation (B10) is compact, but, in contrast to (B7) or (B9), not useful for extracting the asymptotic behavior. The scaled core-size distribution and leaf degree distribution for $\text{VAN}(\infty)$ are displayed in Fig. 11.

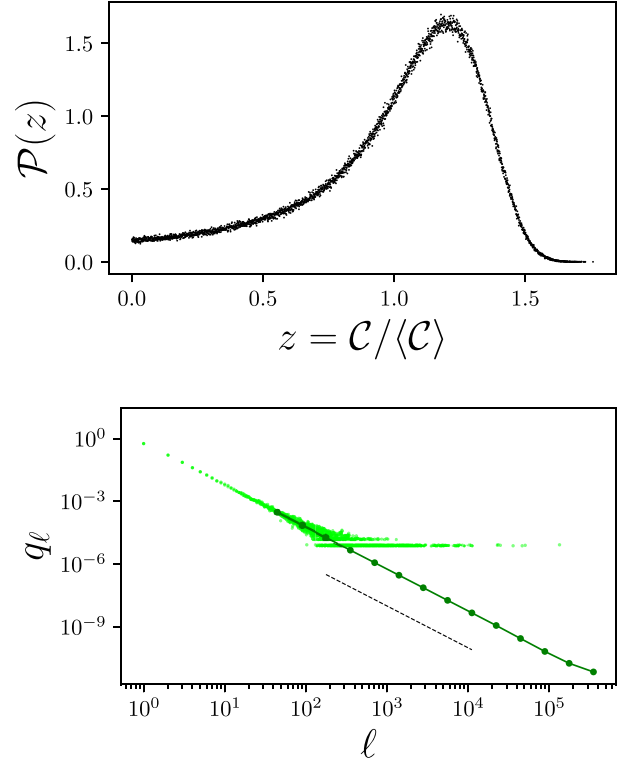


FIG. 11. Scaled core size distribution $\mathcal{P}(z)$ and leaf degree distribution $q_\ell \sim \ell^{-2}$ in the $\text{VAN}(\infty)$ model. Data for the core size distribution are from 2.4×10^6 realizations at size $N = 10^4$, and for the leaf degree distribution are from 4800 realizations at size $N = 10^6$, with ratios $M_\ell/\mathcal{C}$ for three graphs shown as green dots; the dashed line represents pure power-law decay of the form ℓ^{-2} .

APPENDIX C: TWO-POPULATION MODEL

We have presented a self-contained theory for the phenomena of leaf proliferation, the ultrabroad leaf degree distribution, and the vanishingly small relative size of the core, in the family of random growing trees $\text{VAN}(\omega)$, emphasizing $\omega = 0$, $\omega = 1$, and $\omega = \infty$. To provide further intuition for these unusual properties, we analyze a toy model that sheds light on how leaf proliferation and a vanishingly small relative size of the core can arise. Our discussion here serves as a natural counterpoint to that given in Appendix A about the relative abundance of star graphs and reinforces the message that redirection models naturally lead to leaf proliferation and a vanishingly small core.

In our toy model, we partition the network into two populations, leaf nodes and core nodes, and disregard all other details, such as the degree of the core nodes. The number of leaf nodes and core nodes are N_1 and $\mathcal{C}$, respectively, with $N_1 + \mathcal{C} = N$. Because these two populations are not independent, it suffices to consider only the dynamics of the core nodes. When a new node is added to the network, only two outcomes can occur:

- (1) A core node is added with probability $\mu\mathcal{C}(N)/N$.
- (2) Otherwise, a leaf node is added.

These two steps are meant to mimic the isotropic redirection growth rule in which a core node is initially selected with probability $\mathcal{C}/N$, and then in a fraction μ of these selections attachment to a leaf occurs, which increases $\mathcal{C}$ by 1. In terms of the unknown parameter μ , the number $\mathcal{C}(N)$ of core nodes changes according to the stochastic equation

$$\mathcal{C}(N+1) = \begin{cases} \mathcal{C}(N) + 1 & \text{prob. } \frac{\mu \mathcal{C}(N)}{N} \\ \mathcal{C}(N) & \text{prob. } 1 - \frac{\mu \mathcal{C}(N)}{N} \end{cases}, \quad (\text{C1})$$

for $N \geq 3$. Starting with this exact equation, we now compute moments of $\mathcal{C}$ and determine the probability distribution of $\mathcal{C}$.

1. The average and the variance

The number $\mathcal{C}(N)$ of core nodes is a random quantity for all $N \geq 4$. Averaging (C1) gives

$$\langle \mathcal{C}(N+1) \rangle = \left[1 + \frac{\mu}{N} \right] \langle \mathcal{C}(N) \rangle. \quad (\text{C2})$$

Iterating this recurrence and using the initial condition $\mathcal{C}(3) = 1$, we obtain

$$\langle \mathcal{C}(N) \rangle = A \frac{\Gamma(N+\mu)}{\Gamma(N)} \quad A = \frac{2}{\Gamma(3+\mu)}. \quad (\text{C3})$$

From the asymptotics of the gamma function, we have

$$\langle \mathcal{C}(N) \rangle = A N^\mu - \frac{\mu(1-\mu)A}{2} N^{\mu-1} + \dots \quad (\text{C4})$$

This two-population toy model thus predicts sublinear growth of $\mathcal{C}(N)$ on N , but also gives a simple mechanism, embodied by the process in Eq. (C1), for this anomalously slow growth.

In the same way as for the average, the recurrence for the second moment is

$$\langle \mathcal{C}^2(N+1) \rangle = \left[1 + \frac{2\mu}{N} \right] \langle \mathcal{C}^2(N) \rangle + \mu A \frac{\Gamma(N+\mu)}{\Gamma(N+1)},$$

from which

$$\langle \mathcal{C}^2(N) \rangle = \frac{4}{\Gamma(3+2\mu)} \frac{\Gamma(N+2\mu)}{\Gamma(N)} - A \frac{\Gamma(N+\mu)}{\Gamma(N)}. \quad (\text{C5})$$

The variance $\sigma^2(N) = \langle \mathcal{C}^2(N) \rangle - \langle \mathcal{C}(N) \rangle^2$ thus has the asymptotic behavior

$$\sigma^2(N) = \left[\frac{4}{\Gamma(3+2\mu)} - A^2 \right] N^{2\mu} - A N^\mu + \dots \quad (\text{C6})$$

The crucial implication is that the standard deviation is of the same order as the average. Thus, different realizations of the growth process lead to differences in $\mathcal{C}(N)$ that are of the order of $\mathcal{C}(N)$ itself. That is, there is no asymptotic self-averaging in the number of core nodes, as we have observed in numerical simulations of networks grown by the DAN and PAN rules.

2. The probability distribution $P_N(\mathcal{C})$

The leading asymptotic behaviors of the average and variance, Eqs. (C4) and (C6), imply that the random variable $\mathcal{C}(N)$ remains broadly distributed as $N \rightarrow \infty$. We therefore expect that the probability distribution $P_N(\mathcal{C}) := \text{Prob}[\mathcal{C}(N) = \mathcal{C}]$ will converge to the scaling form

$$P_N(\mathcal{C}) \simeq N^{-\mu} \mathcal{P}(z), \quad z = \mathcal{C}/N^\mu, \quad (\text{C7})$$

in the limit $C, N \rightarrow \infty$, with $z = C/N^\mu$ finite. The lack of asymptotic self-averaging implies that the scaled distribution $\mathcal{P}(z)$ approaches a fixed finite-width shape.

Following the same reasoning as that used for the first two moments of $\mathcal{C}$, the probability distribution satisfies

$$P_{N+1}(\mathcal{C}) = \left[1 - \frac{\mu \mathcal{C}}{N} \right] P_N(\mathcal{C}) + \frac{\mu(\mathcal{C}-1)}{N} P_N(\mathcal{C}-1). \quad (\text{C8})$$

Let us first derive the asymptotic behaviors of $\mathcal{P}(z)$:

$$\mathcal{P}(0) = \frac{2}{\Gamma(3-\mu)}, \quad (\text{C9a})$$

$$\ln \mathcal{P}(z) \sim -z^{1/(1-\mu)} \quad (z \rightarrow \infty). \quad (\text{C9b})$$

To prove (C9a), we determine $P_N(1)$, i.e., the probability of generating the star graph. Using (C1), we obtain

$$P_N(1) = \prod_{k=3}^{N-1} \left[1 - \frac{\mu}{k} \right] = \frac{2}{\Gamma(N)} \frac{\Gamma(N-\mu)}{\Gamma(3-\mu)}. \quad (\text{C10})$$

For $N \rightarrow \infty$, $P_N(1) \simeq \frac{2}{\Gamma(3-\mu)} N^{-\mu}$, which is consistent with (C7) if $\mathcal{P}(0)$ is given by (C9a).

Let us now determine the probability to generate the opposite extreme of a line graph, namely, $P_N(N-2)$. Again using (C1), we obtain

$$P_N(N-2) = \prod_{k=3}^{N-1} \frac{\mu(k-2)}{k} = \frac{2\mu^{N-3}}{(N-1)(N-2)}. \quad (\text{C11})$$

Therefore, $\ln P_N(N-2) \simeq \ln \mathcal{P}(N^{1-\mu}) \simeq N \ln \mu$. This leading behavior is consistent with (C9b), and suggests that $\ln \mathcal{P}(z) \simeq z^{1/(1-\mu)} \ln \mu$ as $z \rightarrow \infty$.

For the extreme case of the maximum possible value $\mu = 1$, rather special behavior arises. In this case, the extreme events of $\mathcal{C} = 1$ and $\mathcal{C} = N-2$ occur with probabilities

$$P_N(1) = \frac{2}{N-1} \quad \text{and} \quad P_N(N-2) = \frac{2}{(N-1)(N-2)}.$$

A natural guess for $P_N(\mathcal{C})$ that interpolates between these two extreme values is

$$P_N(\mathcal{C}) = \frac{2}{N-1} \left[1 - \frac{\mathcal{C}-1}{N-2} \right]. \quad (\text{C12})$$

One can verify that Eq. (C12) is indeed the solution by substituting this form into (C8) with $\mu = 1$. Solvability of the stochastic process (C1) with $\mu = 1$ is not surprising, as it reduces to the Pólya urn process [50].

The exact probability distribution (C12) can be rewritten in the scaling form (C7) with $\mu = 1$, thereby justifying the emergence of scaling. The scaled distribution reads

$$\mathcal{P}(z) = \begin{cases} 2(1-z) & 0 < z < 1 \\ 0 & z \geq 1 \end{cases}. \quad (\text{C13})$$

The scaled distribution depends on the initial condition. For our general models, we quantitatively exemplified this phenomenon by computing the behavior of $\mathcal{P}(z)$ in the $z \rightarrow 0$ limit for the default initial condition L_3 and the initial condition L_4 . (The linear graph L_4 emerges from L_3 with probability

$\frac{1}{3}$ for the VAN(ω) models independently of ω .) The scaled distribution vanishes as $z^{\frac{1}{\mu}-1}$ for the initial condition L_3 and as $z^{\frac{2}{\mu}-1}$ for the initial condition L_4 ; cf. (7a) and (A10). For the two-population model with $\mu = 1$, the initial condition L_4 corresponds to $C = 2$ when $N = 4$, and the scaled form of the solution of the Pólya urn process with this initial condition reads

$$\mathcal{P}(z) = \begin{cases} 6z(1-z) & 0 < z < 1 \\ 0 & z \geq 1 \end{cases}. \quad (\text{C14})$$

We finally derive the exact solution in the $0 < \mu < 1$ range. We already know $P_N(1)$, Eq. (C10). Specializing (C8) to $C = 2$ and using (C10), we obtain an inhomogeneous recurrence

$$P_{N+1}(2) = \left[1 - \frac{2\mu}{N}\right]P_N(2) + \frac{2\mu\Gamma(N-\mu)}{\Gamma(N+1)\Gamma(3-\mu)}, \quad (\text{C15})$$

which is solved (subject to $P_3(2) = 0$) to yield

$$P_N(2) = \frac{2}{\Gamma(N)} \left[\frac{\Gamma(N-\mu)}{\Gamma(3-\mu)} - \frac{\Gamma(N-2\mu)}{\Gamma(3-2\mu)} \right]. \quad (\text{C16})$$

Specializing (C8) to $C = 3$ and using (C16), we obtain an inhomogeneous recurrence

$$P_{N+1}(3) = \left[1 - \frac{3\mu}{N}\right]P_N(3) + \frac{4\mu}{\Gamma(N+1)} \left[\frac{\Gamma(N-\mu)}{\Gamma(3-\mu)} - \frac{\Gamma(N-2\mu)}{\Gamma(3-2\mu)} \right], \quad (\text{C17})$$

which is solved (subject to $P_4(3) = 0$) to yield

$$P_N(3) = \frac{2}{\Gamma(N)} \left[\frac{\Gamma(N-\mu)}{\Gamma(3-\mu)} - 2 \frac{\Gamma(N-2\mu)}{\Gamma(3-2\mu)} + \frac{\Gamma(N-3\mu)}{\Gamma(3-3\mu)} \right].$$

The exact results for $P_N(C)$ with $C = 1, 2, 3$ suggest the general solution

$$P_N(C) = \frac{2}{\Gamma(N)} \sum_{k=1}^C (-1)^{k-1} \binom{C-1}{k-1} \frac{\Gamma(N-k\mu)}{\Gamma(3-k\mu)}, \quad (\text{C18})$$

applicable to $C \geq 1$ and all $N \geq C+2$. The solution (C18) can be verified by direct substitution into (C8).

Unfortunately, it is difficult to extract useful information from the exact solution (C18). Since $\langle C \rangle \sim N^\mu$, the core is large, extracting the scaled distribution from (C18) requires simplifying the sum of a very large number of terms. As another example, we mention that extracting $P_N(N-2)$ from (C18) requires summing $N-2$ terms, while directly from the recurrence (C8) we deduced (C11).

-
- [1] M. Drmota, *Random Trees* (Springer, Vienna, 2009).
 - [2] P. Flajolet and R. Sedgewick, *Analytic Combinatorics* (Cambridge University Press, Cambridge, UK, 2009).
 - [3] P. L. Krapivsky, S. Redner, and E. Ben-Naim, *A Kinetic View of Statistical Physics* (Cambridge University Press, Cambridge, UK, 2010).
 - [4] A. Frieze and M. Karoński, *Introduction to Random Graphs* (Cambridge University Press, Cambridge, UK, 2016).
 - [5] R. van der Hofstad, *Random Graphs and Complex Networks* (Cambridge University Press, Cambridge, UK, 2016).
 - [6] H. A. Simon, On a class of skew distribution functions, *Biometrika* **42**, 425 (1955).
 - [7] A.-L. Barabási and R. Albert, Emergence of scaling in random networks, *Science* **286**, 509 (1999).
 - [8] P. L. Krapivsky, S. Redner, and F. Leyvraz, Connectivity of growing random networks, *Phys. Rev. Lett.* **85**, 4629 (2000).
 - [9] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Structure of growing networks with preferential linking, *Phys. Rev. Lett.* **85**, 4633 (2000).
 - [10] R. Albert and A.-L. Barabási, Statistical mechanics of complex networks, *Rev. Mod. Phys.* **74**, 47 (2002).
 - [11] M. Newman, *Networks: An Introduction* (Oxford University Press, New York, USA, 2010).
 - [12] J. M. Kleinberg, S. R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, The web as a graph: Measurements, models and methods, in *Computing and Combinatorics. COCOON 1999. Lecture Notes in Computer Science*, Vol. 1627, edited by T. Asano, H. Imai, D. T. Lee, S. Nakano, and T. Tokuyama (Springer, Berlin, 1999), pp. 1–18.
 - [13] P. L. Krapivsky and S. Redner, Organization of growing random networks, *Phys. Rev. E* **63**, 066123 (2001).
 - [14] A. Gabel, P. L. Krapivsky, and S. Redner, Highly dispersed networks generated by enhanced redirection, *J. Stat. Mech.* (2014) P04009.
 - [15] P. L. Krapivsky and S. Redner, Emergent network modularity, *J. Stat. Mech.* (2017) 073405.
 - [16] W. Levens, A. Szorkovszky, and D. J. T. Sumpter, Friend of a friend models of network growth, *R. Soc. Open Sci.* **9**, 221200 (2022).
 - [17] P. L. Krapivsky and S. Redner, The magic of networks grown by redirection, *Indian J. Phys.* **98**, 3777 (2024).
 - [18] A. Gabel and S. Redner, Sublinear but never superlinear preferential attachment by local network growth, *J. Stat. Mech.* (2013) P02043.
 - [19] S. Redner, How popular is your paper? An empirical study of the citation distribution, *Eur. Phys. J. B* **4**, 131 (1998).
 - [20] M. S. Granovetter, The strength of weak ties, *Am. J. Sociol.* **78**, 1360 (1973).
 - [21] C. Cannings and J. Jordan, Random walk attachment graphs, *Electron. Commun. Probab.* **18**, 1 (2013).
 - [22] L. Addario-Berry, S. Briend, L. Devroye, S. Donderwinkel, C. Kerriou, and G. Lugosi, Random friend trees, *Random Struct. Algorithms* **69**, e70082 (2026).
 - [23] H. Hartle, B. Klein, D. Krioukov, and P. L. Krapivsky, Growing unlabeled networks, *arXiv:2509.17200*.
 - [24] H. Hartle and P. L. Krapivsky, Statistics of leaves in growing random trees, *J. Stat. Mech.* (2026) 023210.
 - [25] H. Hartle and P. L. Krapivsky, Leaves of preferential attachment trees, *J. Phys. A: Math. Theor.* **59**, 165001 (2026).

- [26] A. Kaneko, Spanning trees with constraints on the leaf degree, *Discrete Appl. Math.* **115**, 73 (2001).
- [27] J. Lin, H. Liu, and L. You, Tight conditions for spanning trees with leaf degree at most k in graphs, *Discrete Appl. Math.* **374**, 97 (2025).
- [28] R. R. X. Du and H. Prodinger, Notes on protected nodes in digital search trees, *Appl. Math. Lett.* **25**, 1025 (2012).
- [29] H. M. Mahmoud and M. D. Ward, Asymptotic properties of protected nodes in random recursive trees, *J. Appl. Probab.* **52**, 290 (2015).
- [30] H.-W. Ma, J. Buer, and A.-P. Zeng, Hierarchical structure and modules in the *Escherichia coli* transcriptional regulatory network revealed by a new top-down approach, *BMC Bioinf.* **5**, 199 (2004).
- [31] W. Gautschi, Computational aspects of three-term recurrence relations, *SIAM Rev.* **9**, 24 (1967).
- [32] A. Cuyt, V. B. Petersen, B. Verdonk, H. Waadeland, and W. B. Jones, *Handbook of Continued Fractions for Special Functions* (Springer, Berlin, 2008).
- [33] Z. Li and X. Tang, A study on scale free social network evolution model with degree exponent < 2 , *J. Syst. Sci. Complex.* **33**, 87 (2020).
- [34] H. Seyed-allaei, G. Bianconi, and M. Marsili, Scale-free networks with an exponent less than two, *Phys. Rev. E* **73**, 046113 (2006).
- [35] G. Timár, S. N. Dorogovtsev, and J. F. F. Mendes, Scale-free networks with exponent one, *Phys. Rev. E* **94**, 022302 (2016).
- [36] A. Catanzaro, R. van der Hofstad, and D. Garlaschelli, Clustering without geometry in sparse networks with independent edges, *arXiv:2603.13159*.
- [37] O. T. Courtney and G. Bianconi, Dense power-law networks and simplicial complexes, *Phys. Rev. E* **97**, 052303 (2018).
- [38] P. Krapivsky and D. Krioukov, Scale-free networks as preasymptotic regimes of superlinear preferential attachment, *Phys. Rev. E* **78**, 026114 (2008).
- [39] G. Bebek, P. Berenbrink, C. Cooper, T. Friedetzky, J. Nadeau, and S. C. Sahinalp, The degree distribution of the generalized duplication model, *Theor. Comput. Sci.* **369**, 239 (2006).
- [40] C. Steinbock, O. Biham, and E. Katzav, Analytical results for the in-degree and out-degree distributions of directed random networks that grow by node duplication, *J. Stat. Mech.* (2019) 083403.
- [41] C. Steinbock, O. Biham, and E. Katzav, Analytical results for the distribution of shortest path lengths in directed random networks that grow by node duplication, *Eur. Phys. J. B* **92**, 130 (2019).
- [42] D. Borrelli, Duplication-divergence growing graph models, *arXiv:2506.15640*.
- [43] F. Hermann and P. Pfaffelhuber, Large-scale behavior of the partial duplication random graph, *ALEA, Lat. Am. J. Probab. Math. Stat.* **13**, 687 (2016).
- [44] I. Ispolatov, P. L. Krapivsky, and A. Yuryev, Duplication-divergence model of protein interaction network, *Phys. Rev. E* **71**, 061911 (2005).
- [45] G. Bianconi, R. K. Darst, J. Iacovacci, and S. Fortunato, Triadic closure as a basic generating mechanism of communities in complex networks, *Phys. Rev. E* **90**, 042806 (2014).
- [46] T. Y. Y. Lo, W. Levens, and D. J. T. Sumpter, Properties of the ‘friend of a friend’ model for network generation, *J. Complex Networks* **12**, cnae032 (2024).
- [47] M. Falkenberg, Heterogeneous node copying from hidden network structure, *Commun. Phys.* **4**, 200 (2021).
- [48] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády, The degree sequence of a scale-free random graph process, *Random Struct. Algorithm* **18**, 279 (2001).
- [49] S. Janson, Asymptotic degree distribution in random recursive trees, *Random Struct. Algorithms* **26**, 69 (2005).
- [50] H. M. Mahmoud, *Pólya Urn Models* (Chapman and Hall/CRC, New York, NY, 2008).