The power of choice in growing trees

R.M. D'Souza[1,2,a], P.L. Krapivsky[3], and C. Moore[2,4]

[1] Department of Mechanical and Aeronautical Engineering, University of California, Davis, CA 95616, USA
[2] The Santa Fe Institute, Santa Fe, NM 87501, USA
[3] Department of Physics, Boston University, Boston, MA 02215, USA
[4] Computer Science Department, University of New Mexico, Albuquerque, NM 87131, USA

Abstract. The "power of choice" has been shown to radically alter the behavior of a number of randomized algorithms. Here we explore the effects of choice on models of random tree growth. In our models each new node has k randomly chosen contacts, where $k > 1$ is a constant. It then attaches to whichever one of these contacts is most desirable in some sense, such as its distance from the root or its degree. Even when the new node has just two choices, i.e., when $k = 2$, the resulting tree can be very different from a random graph or tree. For instance, if the new node attaches to the contact which is closest to the root of the tree, the distribution of depths changes from Poisson to a traveling wave solution. If the new node attaches to the contact with the smallest degree, the degree distribution is closer to uniform than in a random graph, so that with high probability there are no nodes in the tree with degree greater than $O(\log \log N)$. Finally, if the new node attaches to the contact with the largest degree, we find that the degree distribution is a power law with exponent -1 up to degrees roughly equal to k, with an exponential cutoff beyond that; thus, in this case, we need $k \gg 1$ to see a power law over a wide range of degrees.

PACS. 89.75.Hc Networks and genealogical trees – 02.50.Ey Stochastic processes – 05.40.-a Fluctuation phenomena, random processes, noise, and Brownian motion

1 Formulation of the model

Over the past decade, the "power of choice" has emerged as a theme in research on optimization and randomized algorithms [1–4]. Consider a random decision process. Typically at each step of the process a decision is reached by choosing one outcome at random and accepting this choice. Now, rather then one random alternative being presented at each decision point, let a small set of randomly generated alternatives be presented, and let the best one be selected. It has been shown that with as few as two alternatives at each decision point, the resulting properties of the process can be radically altered. This was first explored in the context of load-balancing the allocation of jobs arriving at random times to a batch of processors. With as few as two choices, the maximum load on any one processor drops dramatically from $O(\log N)$ to $O(\log \log N)$. Increasing the number of choices beyond two only improves this by a constant factor, illustrating the "power of two choices".

Here we explore the effect of choice on growth of random trees. Perhaps the simplest way to build a growing random tree is to attach each new node to an existing node which is chosen uniformly at random. This process generates *random recursive trees* which have been studied

in great detail (see e.g. [5–8] and references therein). Here we discuss a simple generalization: for each new node we choose $k > 1$ existing 'contact' nodes uniformly at random, select the 'best' one according to some definition, and connect the new node to it. This creates a random tree [9] whose statistics may be very different from those of a random recursive tree.

We have to define, of course, the 'quality' of the node so that we can choose the best one. One natural definition of quality in a tree is distance to the root — the closer to the root, the better, so that the new node attaches to whichever one of its contacts is closest to the root (and, if more than one contact has this smallest distance, we choose one of them randomly). This could correspond, for instance, to someone joining a hierarchical organization, and choosing to become a daughter node of whichever one of their k contacts is highest up in the hierarchy.

Another natural definition is to measure quality by degree of the contact node: for instance, to attach the new node to the contact node with highest (or lowest) degree, again breaking ties randomly. Note that this is very different from the preferential attachment process [10], where the contact is selected from the entire graph with probability proportional to its degree. This latter process requires complete knowledge of the degree of all existing nodes. In contrast, our model assumes that the new node possesses only a small amount of local information, namely,

[a] e-mail: `raissa@cse.ucdavis.edu`

the degrees of a small number of potential contacts. This brings us to another motivation for this work: the desire to understand the effects of limited, local information on network growth.

For the smallest-depth model, we find a marked difference in behavior for $k \geq 2$ versus $k = 1$. The measure of interest in this case is the depth distribution (the fraction of nodes at each depth j). For $k = 1$, i.e., a random recursive tree, this distribution is Poisson. For $k \geq 2$, the same Poisson distribution is observed for distances close to the root, however for larger distances the depth distribution obeys a traveling wave solution. We also consider using *maximal* depth, rather than minimal depth, as the contact node selection criterion and find a similar traveling-wave solution.

For the highest-degree model, we find that the degree distribution decays exponentially for degree $i > k$. For $i < k$ the degree distribution exhibits power-law like behavior, thus in order to observe a power law for any substantial regime requires $k \gg 1$. In other words, a large amount of (overhead/knowledge of the system) is required to achieve a power law distribution.

Finally, in analogy to the above-referenced works on load balancing, the lowest-degree model achieves a degree distribution which is very close to uniform, in which the maximum degree in the entire graph is $O(\log \log N)$ as opposed to the maximum degree in a Poisson distribution, which is roughly $O(\log N)$. Thus, it is possible to generate a nearly-regular random tree (i.e., all vertices have nearly the same degree) without global coordination.

2 Smallest depth

Let N be the total number of nodes and $D_j(N)$ be the number of nodes at distance j from the root. By definition, $D_0(N) \equiv 1$, since the root is distance 0 from itself. Thus $D_0(N)$ is a deterministic quantity, while $D_j(N)$ with $1 \leq j < N$ are random variables. We shall focus on their averages $Q_j(N) \equiv \langle D_j(N) \rangle$. An average value provides a good description of a random variable when it is large and hence fluctuations are relatively small; we will see that this is indeed correct for $D_1(N)$.

To set the stage we begin in Section 2.1 with the simpler case of random recursive trees, for which everything is already known (see e.g. [11]). We then consider the influence of 2 or more choices in Section 2.2.

2.1 Random recursive trees and depth

The quantity D_j grows each time a node at distance $j - 1$ is selected as the contact node. The average depth distribution thus satisfies the master equation [12]

$$Q_j(N + 1) = Q_j(N) + \frac{1}{N} Q_{j-1}(N). \qquad (1)$$

This equation is *exact* and it applies even for $j = 0$ if we set $Q_{-1}(N) \equiv 0$. Using the recursive nature of (1), we first solve for $Q_1(N)$, then $Q_2(N)$, etc. This gives

$$Q_j(N + 1) = \sum_{1 \leq m_1 < \cdots < m_j \leq N} \frac{1}{m_1 \times \ldots \times m_j}. \qquad (2)$$

Equivalently, we can recast the j-fold sums into simple sums, although the results look less neat. For example,

$$Q_1(N) = H_{N-1} \qquad (3a)$$

$$Q_2(N) = \frac{1}{2} \left[(H_{N-1})^2 - H_{N-1}^{(2)} \right] \qquad (3b)$$

where $H_N^{(p)} = \sum_{1 \leq n \leq N} n^{-p}$ are harmonic numbers. The asymptotic behaviors of $H_N \equiv H_N^{(1)}$, $H_N^{(2)}$, and other harmonic numbers are well-known [13], and the resulting asymptotics of the depth distribution are

$$Q_1(N + 1) = \ln N + \gamma + \frac{1}{2N} - \frac{1}{12N^2} + \cdots$$

$$Q_2(N + 1) = \frac{1}{2} (\ln N)^2 + \gamma \ln N + \frac{1}{2} \left[\gamma^2 - \frac{\pi^2}{6} \right] + \ldots,$$

where $\gamma \approx 0.577$ is the Euler-Mascheroni constant. Analogous results hold for $Q_j(N)$ for larger j.

If we merely want to establish the leading asymptotic behavior, we can replace the summation in (2) by integration. This leads to the simple result

$$Q_j(N) \to \frac{(\ln N)^j}{j!} \qquad (4)$$

showing that in the limit $N \to \infty$, the depth distribution is Poisson with mean $\ln N$. Alternatively, we can derive (4) within a continuum approach by replacing finite differences by derivatives in the $N \to \infty$ limit of (1). This procedure recasts discrete master equations into differential equations

$$\frac{dQ_j}{dN} = \frac{1}{N} Q_{j-1}. \qquad (5)$$

Solving (5) one recovers (4).

The normalization requirement $\sum_{j \geq 0} D_j(N) = N$ implies the sum rule for the averages

$$\sum_{j \geq 0} Q_j(N) = N. \qquad (6)$$

The continuum approximation (4) agrees with the sum rule (2) implying that it well approximates the depth distribution in the entire range. We therefore use it to find the depth of the recursive random tree. The depth is defined as the maximal $j_{\max}$. The criterion $Q_{j_{\max}} = 1$ leads to an estimate [11]

$$j_{\max} = e \ln N. \qquad (7)$$

It is possible to derive this result within the exact (discrete) approach and to determine the fluctuations of $j_{\max}$. However, for our purposes (7) is sufficient.

2.2 The model with k = 2 choices

Now suppose the new node has $k = 2$ choices. In this case, we have $D_j(N+1) = D_j(N) + 1$ if the two contact nodes have minimum depth $j-1$, or equivalently, if both of them have depth at least $j-1$, but if they do not both have depth greater than $j-1$. The probability of this is

$$N^{-2}\left[\left(\sum_{i \geq j-1} D_i\right)^2 - \left(\sum_{i \geq j} D_i\right)^2\right] =$$
$$N^{-2}\left(D_{j-1}^2 + 2D_{j-1}\sum_{i \geq j} D_i\right). \quad (8)$$

This leads to the exact recurrence

$$Q_j(N+1) = Q_j(N) + N^{-2}\left\langle D_{j-1}^2 + 2D_{j-1}\sum_{i \geq j} D_i. \right\rangle.$$

Unfortunately, this is not very helpful since the average of the product of random quantities differs from the product of their averages, viz. $\langle D_i D_j \rangle \neq \langle D_i \rangle \langle D_j \rangle$. One can, of course, write down an exact recurrence for $\langle D_i D_j \rangle$, but this involves third order moments $\langle D_i D_j D_k \rangle$, and so on. Thus the hierarchical nature of the governing equations does not allow us to obtain complete and rigorous results as is possible for the case $k = 1$.

The cases of $j = 1, 2$ are exceptional and one can determine Q_1 and Q_2 analytically. For $j = 1$ the analysis is especially simple since $D_0 = 1$, $\sum_{i \geq 1} D_i = N - 1$, and the growth rate (8) simplifies to $[1 + 2(N-1)]/N^2$. Therefore the average number of the neighbors of the root grows according to an exact and closed recurrence

$$Q_1(N+1) = Q_1(N) + \frac{2N-1}{N^2}. \quad (9)$$

Solving (9) subject to $Q_1(1) = 0$ yields

$$Q_1(N) = \sum_{n=1}^{N-1} \frac{2n-1}{n^2} = 2H_{N-1} - H_{N-1}^{(2)}. \quad (10)$$

Similarly for $j = 2$ we use relation $\sum_{i \geq 2} D_i = N - 1 - D_1$ and obtain

$$Q_2(N+1) = Q_2(N) + 2\frac{N-1}{N^2} Q_1(N) - \frac{\langle D_1^2(N) \rangle}{N^2}. \quad (11)$$

To obtain a closed recurrence for Q_2 we need to determine $\langle D_1^2(N) \rangle$, the average of the square of the number of neighbors of the root. Then (8) leads to

$$D_1(N+1) = \begin{cases} D_1(N) + 1 & \text{prob} \quad N^{-2}(2N-1) \\ D_1(N) & \text{prob} \quad 1 - N^{-2}(2N-1). \end{cases}$$

Squaring this equation and averaging we obtain

$$\langle D_1^2(N+1) \rangle = \left(1 - \frac{2N-1}{N^2}\right) \langle D_1^2(N) \rangle$$
$$+ \frac{2N-1}{N^2}[\langle D_1^2(N) \rangle + 2Q_1(N) + 1]$$
$$= \langle D_1^2(N) \rangle + 2\frac{2N-1}{N^2} Q_1(N) + \frac{2N-1}{N^2}.$$

Rather than directly solving this recurrence, we can use it together with (9) to establish a simpler recurrence for the variance $V_1(N) = \langle D_1^2(N) \rangle - \langle D_1(N) \rangle^2$. We find

$$V_1(N+1) = V_1(N) + \frac{2N-1}{N^2} - \left(\frac{2N-1}{N^2}\right)^2 \quad (12)$$

which is readily solved to give

$$V_1(N+1) = 2H_N - 5H_N^{(2)} + 4H_N^{(3)} - H_N^{(4)}.$$

Thus $\langle D_1^2 \rangle \neq \langle D_1 \rangle^2$, yet the variance is asymptotically $2 \ln N$ and therefore fluctuations of the random variable D_1 are indeed small compared to its average which grows as $2 \ln N$, see (10).

We determined $\langle D_1^2(N) \rangle = V_1(N) + Q_1^2(N)$ and therefore Q_2 satisfies a closed solvable recurrence (11). The solution reads

$$Q_2(N) = \frac{1}{2}[Q_1(N)]^2 - \frac{1}{2}\sum_{n=1}^{N-1}\left(\frac{2n-1}{n^2}\right)^2$$
$$- \sum_{n=1}^{N-1} \frac{V_1(n) + [Q_1(n)]^2 + Q_1(n)}{n^2}.$$

For $j \geq 3$, the problem becomes genuinely hierarchical and intractable. If we are seeking only the leading behavior, however, we can proceed. When $N \gg 1$ and j is sufficiently small, namely such that $\sum_{i \leq j} Q_i \ll N$, we can replace the sum $\sum_{i \geq j} D_i$ by N and the growth rate (8) by $2D_{j-1}/N$. Thus we arrive at a set of differential equations

$$\frac{dQ_j}{dN} = 2\frac{Q_{j-1}}{N}. \quad (13)$$

Solving these equations we obtain

$$Q_j(N) = \frac{(2 \ln N)^j}{j!}. \quad (14)$$

We check the validity of this approximation by substituting it back into our assumption $\sum_{i \leq j} Q_i \ll N$ which we used in the derivation of (13). This suggests that (14) holds when $j < v \ln N$ (i.e., for small distances from the root) where v is the smallest positive root of

$$v \ln\left(\frac{2e}{v}\right) = 1. \quad (15)$$

We can write v in terms of Lambert's function $W(x)$, defined as the root of $We^W = x$:

$$v = -1/W_{-1}(-1/2e) \quad (16)$$

where W_{-1} denotes the -1st branch of the Lambert function. Numerically, $v = 0.373365...$

When $j \geq v \ln N$ we cannot use (14). However, as long as Q_j is much larger than 1, let us assume that the fluctuations in D_j are small. In that case we can replace averages $\langle D_j D_k \rangle$ by $Q_j Q_k$, and in this regime we obtain

$$\frac{dQ_j}{dN} = N^{-2} \left(Q_{j-1}^2 + 2Q_{j-1} \sum_{k \geq j} Q_k \right). \qquad (17)$$

It is convenient to introduce the cumulative variable

$$q_j = \frac{1}{N} \sum_{i \geq j} Q_i \qquad (18)$$

that is, the average fraction of nodes whose depth is at least j. Summing (17) over all $i \geq j$ we arrive at a neat recurrence

$$\frac{d}{dN} N q_j = q_{j-1}^2. \qquad (19)$$

The form of this equation suggests the introduction of a new 'time' variable

$$t = \ln N. \qquad (20)$$

This transformation recasts (19) into

$$\frac{dq_j}{dt} = -q_j + q_{j-1}^2 \qquad (21)$$

which should be solved subject to the step function initial condition: $q_j(0) = 1$ for $j \leq 0$ and $q_j(0) = 0$ for $j > 0$.

Equation (21) has appeared in various contexts (see e.g. [14]) and while it is unsolvable, an asymptotic behavior of its solution is understood. In the long time limit, the solution approaches a 'traveling wave' form,

$$q_j(t) \to q(j - vt). \qquad (22)$$

Plugging (22) into (21) one finds that $q(x)$ satisfies

$$v \frac{dq}{dx} = q(x) - q(x-1)^2. \qquad (23)$$

The boundary conditions are

$$q(-\infty) = 1, \quad q(+\infty) = 0. \qquad (24)$$

The boundary-value problem (23)–(24) is still intractable analytically. However, the velocity v can be determined even without a complete solution for $q(x)$. The method relies on the analysis of the tail region $x \to -\infty$. One notices that (23) admits an exponential solution in this region,

$$1 - q(x) \propto e^{\lambda x} \quad \text{as} \quad x \to -\infty. \qquad (25)$$

Plugging this into (23) shows that the velocity v is related to λ via the dispersion relation [14]

$$v = \frac{1 - 2e^{-\lambda}}{\lambda}. \qquad (26)$$

The maximum of $v = v(\lambda)$ is given by (15) and it occurs at the largest positive root λ of the transcendental equation $2(1 + \lambda) = e^\lambda$. This is

$$\lambda = -1 - W_{-1}(-1/2e) \qquad (27)$$

or numerically, $\lambda = 1.67835...$ Comparing with (16), we see that λ and v are related as follows,

$$\lambda = -1 + 1/v. \qquad (28)$$

Strictly speaking, one can only assert that velocity does not exceed the maximum of (26). However, the so-called *selection principle* tells us that this extremal value is realized for any initial conditions which vanish sufficiently rapidly at infinity. The selection principle has been rigorously proven for a few nonlinear parabolic partial differential equations. Yet heuristic arguments and numerical evidence indicate that the its range of applicability is much broader. This is reviewed in [15] in the context of partial differential equations and in [16] in the context of difference equations.

Thus there is a sharp front at depth $j_{\text{front}} \approx vt = v \ln N$ to leading order, where the depth of most nodes in the tree is concentrated. Furthermore, the width of this front remains finite even in the limit $N \to \infty$. It is also possible to compute the sub-leading correction to the position of the front [14], giving an improved estimate of its location:

$$j_{\text{front}} \approx v \ln N + \frac{3}{2\lambda} \ln \ln N. \qquad (29)$$

To estimate the *maximum* depth j_{max}, it is necessary to bound the tail of $q(x)$ in the positive direction $x \to +\infty$. To do this, note that by definition $q(x)$ is monotonically decreasing, and by (23) this implies that

$$q(x) \leq q(x-1)^2$$

and therefore that this tail is doubly exponential,

$$q(x) \propto e^{-A \cdot 2^x} \qquad (30)$$

for some constant $A > 0$. Setting $q(x) = 1/N$ then gives the estimate

$$j_{\text{max}} \approx j_{\text{front}} + \frac{\ln \ln N}{\ln 2} \qquad (31)$$

minus a constant $C = \ln A / \ln 2$. As shown in Figure 1, (29) and (31) are indeed excellent estimates of the average and maximum depth respectively.

2.3 The effect of choice

At first sight, it seems that having two choices instead of one does not qualitatively affect the outcome, since the depth distributions (4) and (14) both seem Poissonian, and both have typical depth $O(\log N)$. This is, however, an illusion. First of all, the distribution (4) for random recursive trees is indeed Poissonian while (14) is valid only for $j < v \ln N$. Secondly, while both types of trees have depth $O(\log N)$, choice causes the depth to be much more
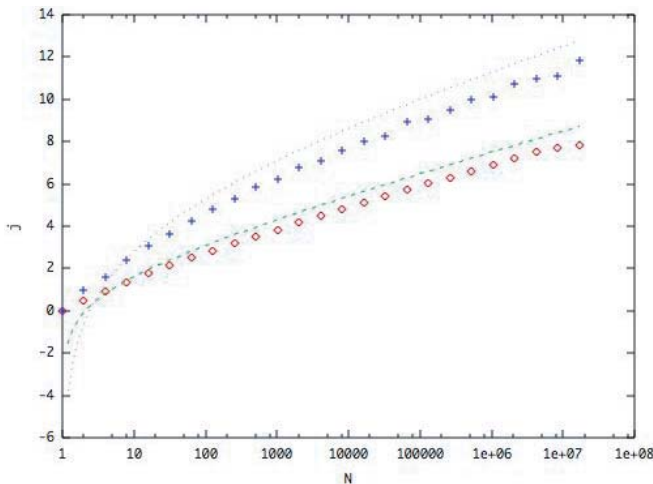
Fig. 1. The average depth (circles) and maximum depth (crosses) of a tree with $k = 2$, averaged over 10^3 independent trials for each value of N, and (dashed) the expressions (29) and (31) for j_{front} and j_{max} respectively.

concentrated. This is easiest to see if we consider the cumulative depth distribution (18). For random recursive trees, $q_j(t)$ is asymptotically

$$q_j(t) = \frac{1}{2} \operatorname{erfc}\left(\frac{j - t}{\sqrt{2t}}\right) \tag{32}$$

where $\operatorname{erfc}(z)$ is the error function

$$\operatorname{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_z^\infty d\eta \, e^{-\eta^2}. \tag{33}$$

Thus

$$q_j(t) = \begin{cases} 1 & j - t \ll -\sqrt{t} \\ 0 & j - t \gg +\sqrt{t}. \end{cases}$$

The boundary layer where q changes from one to zero is not a true front as its width grows with 'time' as $\sqrt{t} \sim \sqrt{\ln N}$.

On the other hand, for the model with choice the cumulative depth distribution has a traveling wave shape with a front of *constant* width. Thus

$$q_j(t) = \begin{cases} 1 & j - j_{\text{front}} \ll -1 \\ 0 & j - j_{\text{front}} \gg +1. \end{cases}$$

2.4 Multiple choices

What if the new node has more than two choices? The cases with $k \geq 3$ (with k constant) are morally similar to the $k = 2$ case: the cumulative depth distribution obeys the differential equation

$$\frac{dq_j}{dt} = -q_j + q_{j-1}^k. \tag{34}$$

Transforming this to $q_j(t) = q(j - vt)$ as before, we obtain

$$v\frac{dq}{dx} = q(x) - q(x - 1)^k. \tag{35}$$

The solution is again a traveling wave, whose velocity v depends on k. Assuming the selection principle, v is the smallest positive root of

$$v \ln\left(\frac{ke}{v}\right) = 1 \tag{36}$$

which can be written in terms of Lambert's function as

$$v = -1/W_{-1}(-1/ke). \tag{37}$$

Asymptotically, as k grows we have

$$v \approx \frac{1}{\ln ke + \ln\ln ke} = \frac{1}{\ln k}\left(1 - O\left(\frac{\ln\ln k}{\ln k}\right)\right). \tag{38}$$

A more precise estimate for j_{front} is again given by (29), with λ given by (28). For $j \ll j_{\text{front}}$, (4) and (14) generalize to

$$Q_j(N) = \frac{(k\ln N)^j}{j!}. \tag{39}$$

Finally, the tail of $q(x)$ is doubly exponential,

$$q(x) \approx e^{-Ak^x} \tag{40}$$

and the maximum depth is given by

$$j_{\text{max}} \approx j_{\text{front}} + \frac{\ln\ln N}{\ln k}. \tag{41}$$

3 Largest depth

We pause here to consider a model in which we reverse our definition of the 'better' node, and attach each new node to the contact node which is furthest from the root. If $k = 2$, then we have $D_j(N + 1) = D_j(N) + 1$ whenever the maximum depth of the two nodes is $j - 1$, and this occurs with probability

$$N^{-2}\left[\left(\sum_{i=0}^{j-1} D_i\right)^2 - \left(\sum_{i=0}^{j-2} D_i\right)^2\right] = \\ N^{-2}\left(D_{j-1}^2 + 2D_{j-1}\sum_{i=0}^{j-2} D_i\right). \tag{42}$$

For instance, the average number of the neighbors of the root grows according to

$$Q_1(N + 1) = Q_1(N) + \frac{1}{N^2} \tag{43}$$

and therefore

$$Q_1(N) = H_{N-1}^{(2)}. \tag{44}$$

Thus the average number of neighbors of the root does not diverge as in the smallest depth model, but instead approaches the constant $\zeta(2) = \pi^2/6$. Generally, the behavior of $Q_j(N)$ for small j is very different from (14), viz. for $j = O(1)$ the average number of nodes of depth j

remains finite in the $N \to \infty$ limit. Therefore in contrast with the smallest depth model, the quantities $D_j(\infty)$ are not self-averaging when $j = O(1)$ and their averages do not characterize them. Yet, the probability distribution

$$P(s) = \text{Prob}[D_1(\infty) = s] \tag{45}$$

can be determined. For instance, $D_1(2) = 1$ and the probability that the root still has one neighbor when the tree size reaches N is

$$\text{Prob}[D_1(N) = 1] = \prod_{n=2}^{N-1} \left(1 - \frac{1}{n^2}\right)$$

and therefore

$$P(1) = \prod_{n=2}^{\infty} \left(1 - \frac{1}{n^2}\right) = \frac{1}{2}. \tag{46}$$

Proceeding with this line of reasoning one obtains

$$P(s+1) = \frac{1}{2} \sum_{2 \leq n_1 < \cdots < n_s}^{\infty} \frac{1}{(n_1^2 - 1) \times \cdots \times (n_s^2 - 1)}$$

which can be expressed as a sum involving the zeta function at positive integers.

However, even though the D_j are not self-averaging, there are many similarities between the smallest depth model and this one. In particular, the cumulative depth distribution has a traveling wave shape (22). Indeed, after several mappings [16] the model becomes identical to one which has appeared in studies of collision processes in gases [17], fragmentation processes [18], and other problems [14]. If we define the cumulative variable as

$$q_j = \frac{1}{N} \sum_{i \geq j} Q_i, \tag{47}$$

then writing $q_j(t) = q(j - vt)$ gives (19), (21) and (23) again, but now with the boundary conditions

$$q(-\infty) = 0, \quad q(+\infty) = 1. \tag{48}$$

With these boundary conditions, (23) admits a solution whose tail in the positive direction is exponential,

$$1 - q(x) \propto e^{-\mu x} \quad \text{as} \quad x \to +\infty \tag{49}$$

and the dispersion relation is now

$$v = \frac{2e^\mu - 1}{\mu}. \tag{50}$$

The selection principle now suggests that v is the *minimum* of (50). This is the larger of the two real roots of the transcendental equation (15), which is $v = 4.31107...$ A more precise estimate of j_{front} is

$$j_{\text{front}} = v \ln N - \frac{3}{2\mu} \ln \ln N \tag{51}$$

where $\mu = 0.768039...$ is the larger root of $2(1 - \mu) = e^{-\mu}$.

More generally, for $k > 2$ the velocity v is the larger real root of (36), or

$$v = -1/W_1(-1/ke) \tag{52}$$

which, as k grows, approaches

$$v \approx ke - 1. \tag{53}$$

The position of the front is given by (51) with

$$\mu = 1 - 1/v. \tag{54}$$

Finally, since the tail of $q(x)$ in the positive direction is given by (49), setting $q_j = 1 - 1/N$ gives the following estimate of the maximum depth,

$$j_{\text{max}} = j_{\text{front}} + \frac{1}{\mu} \ln N. \tag{55}$$

Note that, unlike the minimum depth model, $j_{\text{max}} - j_{\text{front}}$ is $O(\log N)$ instead of $O(\log \log N)$, since the tail (49) is exponential rather than doubly exponential.

4 Highest degree

We now consider a model in which quality is measured not by depth, but by the degree of the contact node — the higher the degree, the better. As we will show below, in this case the degree distribution exhibits a power law up to degree $j \sim k$, beyond which it decays exponentially. Therefore, in this model we need a large number of choices, $k \gg 1$, in order to observe a power law over a wide range of degrees.
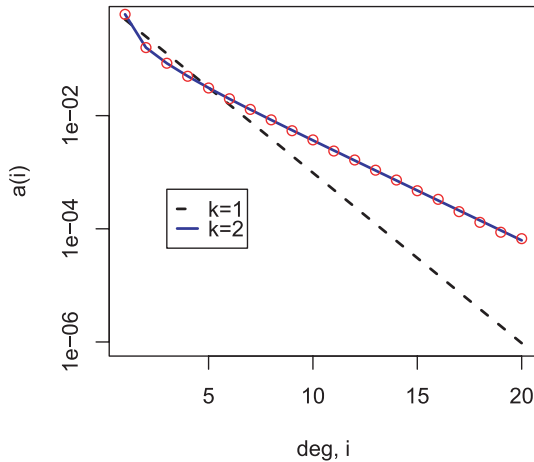
4.1 Recurrence for the degree distribution

We start by writing a master equation for the degree distribution of the tree. We add one node at each step, so at time t there are t nodes in the tree. Let $N_i(t)$ be the number of nodes which have degree i at time t, and let $C_i(t) = \sum_{j=1}^{i} N_j(t)$ be the corresponding total number of nodes of degree i or less at time t. Normalizing these numbers, let $a_i(t) = N_i(t)/t$ be the fraction of nodes which have degree i, and let $c_i(t) = \sum_{j=1}^{i} a_j(t) = C_i(t)/t$ be the corresponding cumulative distribution.

At each iteration, we choose k contact nodes at random from the t existing nodes, and connect the new node to the contact node of highest degree, with ties broken randomly. The evolution of the expected cumulative degree distribution can can be written as, for all $i \geq 1$,

$$C_i(t+1) = C_i(t) + 1 - \left(c_i(t)^k - c_{i-1}(t)^k\right), \tag{56}$$

since C_i increases by 1 for each new node added, and decreases precisely when the new node connects to a node of degree i. This latter event occurs when all k nodes have

Fig. 2. The degree distribution, a_i, for the highest degree model, for both $k = 1$ and $k = 2$. The points are data from numerical simulation of the model with $k = 2$.

Fig. 3. Numerical simulation results for $k = 16$. Note that for $i < k$ we observe $a_i \sim i^{-1.5}$, while for $i > k$ we observe $i \sim e^{-i/k}$.

degree i or less, but not all have degree $i-1$ or less. Writing $c_i(t) = C_i(t)/t$ and making the assumption that a steady-state limit exists, we obtain the recurrence

$$c_i = 1 - (c_i^k - c_{i-1}^k). \qquad (57)$$

We note that in the case $k = 1$, where there is no choice, the solution to (57) is simply

$$c_i = 1 - 2^{-i} \text{ and } a_i = 2^{-i} \qquad (58)$$

which is the degree distribution of a random recursive tree.
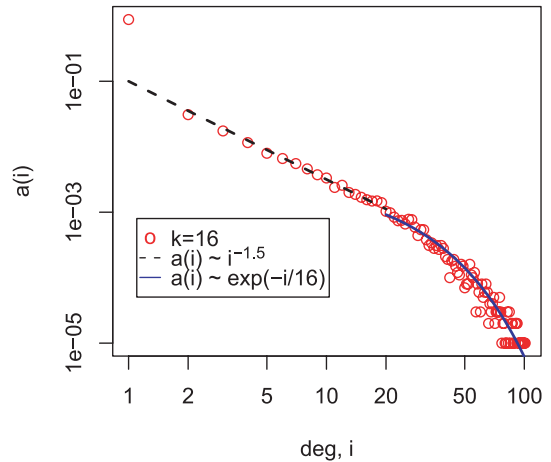
4.2 The model with k ≥ 2 choices

We are particularly interested in the behavior for small k. Recall that the "power of choice" comes from situations where results vary dramatically if $k = 2$ rather than $k = 1$. For $k \geq 2$ we can solve (57) analytically only in the regime $i \gg 1$ as discussed in detail below. Yet, for $k = 2$, equation (57) is very easy to solve numerically as it reduces to the quadratic equation:

$$c_i^2 + c_i - (1 + c_{i-1}^2) = 0. \qquad (59)$$

Figure 2 is a plot of the degree distribution, a_i, for $k = 1$ and $k = 2$. Recall $a_i = c_i - c_{i-1}$. The data points are from a numerical simulation with $k = 2$, grown to size 1×10^6 nodes. Note the excellent agreement. Though the distribution for $k = 2$ decays less slowly than $k = 1$ both exhibit exponential decay, thus the nature of the solution is not altered with such minor amounts of choice.

From numerical simulation with $k \geq 2$ we find different behaviors for $i > k$ than for $i < k$ (see Fig. 3). For degree $i > k$ we observe $a_i \sim \exp(-i/k)$. For $i < k$ we observe what appears to be a power law in that regime, $a_i \sim k^{-\gamma}$, with $\gamma \approx 1.5$. The largest k we simulated was $k = 32$, hence the "power law" regime is quite small. Rather than computer simulation, we can look at the asymptotic limits of (57) and arrive at these similar results in the limit $i \gg 1$ and $k \gg 1$. Note, the asymptotic limit will give $\gamma = 1$, and we can attribute the difference with numerical results to finite size effects in simulation.

4.3 Asymptotic limits

In the asymptotic regime $i \gg 1$ we write $c_i = 1 - \epsilon_i$ and assume that $\epsilon_i \ll 1$. To first order, $c_i^k = 1 - k\epsilon_i$. Simplifying (57), we find $(k + 1)\epsilon_i = k\epsilon_{i-1}$ and therefore

$$1 - c_i = A_k \left(\frac{k}{k+1} \right)^i \quad \text{when} \quad i \gg 1, \qquad (60)$$

where A_k is a constant depending on k. We argue below that

$$A_k \sim k^{-1} \quad \text{as} \quad k \to \infty \qquad (61)$$

In the rest of this section we always assume that $k \gg 1$. Let us start with nodes of degree one (which are often called 'leaves'). In this case we have $c_1 = a_1$ and equation (57) reduces to

$$a_1 = 1 - a_1^k. \qquad (62)$$

Writing

$$a_1 = 1 - \frac{W}{k} \qquad (63)$$

and assuming that $W \ll k$ yields $a_1^k = e^{-W}$. This allows us to recast (62) into

$$We^W = k \qquad (64)$$

so W is Lambert's function $W(k)$. For large k, we have $W(k) \approx \ln k$, justifying our assumption that $W \ll k$. Thus almost all nodes are leaves: the fraction of nodes whose degree exceeds one is $1 - a_1 = W(k)/k \approx (\ln k)/k$.

Analyzing (57) for $i = 2, 3, \ldots$ one finds that the following ansatz is useful:

$$c_i = 1 - \frac{W - w_i}{k}. \qquad (65)$$

Plugging (65) into (57) we obtain

$$1 + e^{w_{i-1}} - e^{w_i} = W^{-1}w_i \qquad (66)$$

Since $W \to \infty$ as $k \to \infty$, equation (66) simplifies to

$$1 + e^{w_{i-1}} - e^{w_i} = 0 \qquad (67)$$

whose solution (satisfying $w_1 = 0$) is $w_i = \ln i$. Plugging this to (65) we find that $a_i = c_i - c_{i-1}$ is given by

$$a_i = k^{-1} \ln \left(\frac{i}{i-1} \right) \quad \text{when} \quad 2 \leq i \ll k. \qquad (68)$$

The upper bound $i \ll k$ is necessary since we can use (67) instead of (66) only when $w_i \ll W$ which is equivalent to $\ln i \ll \ln k$. Note that we can further simplify (68) when $i \gg 1$, viz.

$$a_i = \frac{1}{k} \cdot \frac{1}{i} \quad \text{when} \quad 1 \ll i \ll k. \qquad (69)$$

Thus up to a crossover at $i = k$, the degree distribution exhibits an algebraic behavior $a_i \sim i^{-1}$ with unusually small exponent.

The derivation of (60) actually holds when $i \gg k$. Using (60) we compute $a_i = c_i - c_{i-1}$ to give

$$a_i = k^{-1} A_k \left(\frac{k}{k+1} \right)^i \quad \text{when} \quad i \gg k. \qquad (70)$$

The regions of the validity of (69) and (70) do not formally overlap. It is reasonable to assume, however, that they remain qualitatively correct. Then from equation (69) we obtain $a_k \sim k^{-2}$ while equation (70) leads to $a_k \sim k^{-1} A_k$. Matching this values we confirm the announced asymptotic of the amplitude, equation (61). Furthermore, we find

$$a_i \sim \left(\frac{k}{k+1} \right)^i \approx e^{-i/k} \quad \text{when} \quad 1 \ll k \ll i. \qquad (71)$$

5 Lowest degree

There are situations where one wants to ensure that all nodes have low degree, for instance consider the case of load-balancing discussed in Section 1. Thus the final variant we consider is when an incoming node connects to the target node of *lowest* degree.

5.1 Recurrence for the degree distribution

As in Section 4, we begin by writing the master equation for the degree distribution of the tree. Again let $N_i(t)$ be the number of nodes which have degree i at time t, and now let $\overline{C}_i(t) = \sum_{j \geq i} N_j(t)$ be the corresponding total number of nodes of degree i or greater at time t. Normalizing, let $a_i(t) = N_i/t$ and let $\overline{c}_i(t) = \sum_{j \geq i} a_j(t) = \overline{C}_i(t)/t$ be the complementary cumulative distribution.

At each iteration, we choose k contact nodes at random from the t existing nodes, and connect the new node to the contact node of *lowest* degree, with ties broken randomly.

The evolution of the expected complementary cumulative degree distribution can can be written, for all $i > 1$, as

$$\overline{C}_i(t+1) = \overline{C}_i(t) + \left[\overline{c}_{i-1}(t)^k - \overline{c}_i(t)^k \right], \qquad (72)$$

since $\overline{C}_i$ increases precisely when the new node connects to a node of degree $i - 1$. This event occurs when all k nodes have degree $i - 1$ or greater, but not all have degree i or greater. Writing $\overline{c}_i(t) = \overline{C}_i(t)/t$ and making the assumption that a steady-state limit exists, we obtain the recurrence

$$\overline{c}_i = \overline{c}_{i-1}^k - \overline{c}_i^k. \qquad (73)$$

We note that in the case $k = 1$, where there is no choice, the solution to (73) is simply

$$\overline{c}_i = 2^{-(i-1)} \quad \text{and} \quad a_i = 2^{-i} \qquad (74)$$

which, as (58), is the degree distribution of a random recursive tree.

5.2 The model with k ≥ 2 choices

For $k = 2$, (73) is very easy to solve numerically as it reduces to the quadratic equation:

$$\overline{c}_i^2 + \overline{c}_i - \overline{c}_{i-1}^2 = 0. \qquad (75)$$

Figure 4 is a plot of the degree distribution, a_i, for $k = 1$ and $k = 2$. Recall here, $a_i = \overline{c}_i - \overline{c}_{i+1}$. The data points are from a numerical simulation with $k = 2$, grown to size 1×10^6 nodes. Note the excellent agreement. With minor choice, the degree distribution is radically altered.

For all $k \geq 2$ we can show the upper bound on the maximum degree is $O(\log \log N)$ using a method similar to that in [19]. From (73), for $i \geq 3$ we obtain the upper bound, $\overline{c}_i \leq \overline{c}_{i-1}^k$, and by recursion:

$$\overline{c}_i \leq \overline{c}_{i-1}^k \leq \overline{c}_2^K, \qquad (76)$$

where $K = k^{(i-2)}$. Since $\overline{c}_2 < 1$, $\overline{c}_i$ decreases doubly-exponentially. To find $i_{\max}$, the typical largest degree present after addition of N nodes, we set $\overline{c}_i = 1/N$. Solving this relation we find:

$$i_{\max} \leq \log_k \log_{1/\overline{c}_2} N = O(\log \log N). \qquad (77)$$

6 Discussion

We explore the "power of choice" in growth of random trees by introducing a minimalist generalization of random recursive trees. At each decision point $k > 1$ choices are presented and the most desirable one selected. If the criteria is to minimize or maximize depth, a small amount of choice has a dramatic effect. For $k = 1$ the depth distribution decays with a Poisson behavior. For $k \geq 2$ this Poisson decay is seen for distances close to the root, but for further distances, the depth distribution obeys a traveling wave behavior. If the criteria instead involves
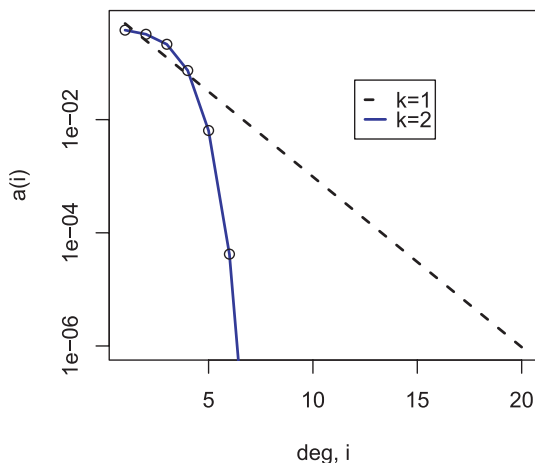
Fig. 4. The degree distribution, a_i, for the lowest degree model, for both $k = 1$ and $k = 2$. The points are data from numerical simulation of the model with $k = 2$.

node degree, we must distinguish the maximum degree model from the minimum degree one. For minimum degree, choice has a dramatic effect. Going from $k = 1$ to $k = 2$ the degree distribution changes from geometric decay to double-exponential decay (and hence the maximum degree observed in the tree changes from $O(\log N)$ to $O(\log \log N)$). In contrast, for maximum degree, a large number of choices, $k \gg 1$, must be allowed before a change from the $k = 1$ behavior is observed. The degree distribution decays exponentially for all small values of k. Once $k \gg 1$ a power law distribution results for nodes of degree $i < k$, while for nodes of degree $i > k$ the distribution decays exponentially.

We established many results about the depth distribution. Some of them are exact, others (namely the assumption that the maximum allowed value of velocity is realized, employed at the end of Sect. 2.2) utilize a selection principle which is not rigorously established for (21). There is no doubt of the validity of this principle in a broad range of contexts, and there is firm numerical support of all analytical results derived herein.

In this current work, our focus is on trees. An interesting avenue for future research is to explore the power of choice in network growth, by extending our analysis to situations where each incoming node connects to multiple parents. One challenge to analytic, and even numerical, exploration of such a model is that the depth of any older node can change with the addition of new nodes.

References

1. Y. Azar, A.Z. Broder, A.R. Karlin, E. Upfal, *Balanced allocations*, in *Proc. 26th ACM Symp. Theory of Computing* (1994) pp. 593–602
2. Y. Azar, A.Z. Broder, A.R. Karlin, E. Upfal, SIAM J. Comp. **29**, 180 (1999)
3. M. Adler, S. Chakarabarti, M. Mitzenmacher, L. Rasmussen, Rand. Struct. Alg. **13**, 159 (1998)
4. M. Mitzenmacher, E. Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis* (Cambridge University Press, New York, 2005)
5. R.T. Smythe, H. Mahmoud, Theory Probab. Math. Statist. **51**, 1 (1995)
6. M. Drmota, B. Gittenberger, Random Struct. Alg. **10**, 421 (1997); M. Drmota, H.-K. Hwang, Adv. Appl. Probab. **37**, 321 (2005)
7. P.L. Krapivsky, S. Redner, Phys. Rev. Lett. **89**, 258703 (2002); e-print `arXiv:cond-mat/0207370`
8. S. Janson, Alea Lat. Am. J. Probab. Math. Stat. **1**, 347 (2006); e-print `arXiv:math/0509471`
9. We always assume that the initial tree is a single node which is the root
10. A.L. Barabási, R. Albert, Science **286**, 509–512 (1999)
11. P.L. Krapivsky, S. Redner, Phys. Rev. E **63**, 066123 (2001); e-print `arXiv:cond-mat/0011094`
12. Equation (1) appears in many contexts, e.g. for networks with copying: P.L. Krapivsky, S. Redner, Phys. Rev. E **71**, 036118 (2005); e-print `arXiv:cond-mat/0410379`
13. R.L. Graham, D.E. Knuth, O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science* (Reading, Mass.: Addison-Wesley, 1989)
14. E. Ben-Naim, P.L. Krapivsky, S.N. Majumdar, Phys. Rev. E **64**, 035101 (2001); e-print `arXiv:cond-mat/0105`
15. W. van Saarloos, Phys. Rep. **386**, 29 (2003)
16. S.N. Majumdar, P.L. Krapivsky, Physica A **318**, 161 (2003); e-print `arXiv:cond-mat/0205581`
17. R. van Zon, H. van Beijeren, Ch. Dellago, Phys. Rev. Lett. **80**, 2035 (1998)
18. P.L. Krapivsky, S.N. Majumdar, Phys. Rev. Lett. **85**, 5492 (2000)
19. M. Mitzenmacher, IEEE Trans. Paral. Dist. Sys. **12**, 1094 (2001)