Long and short paths in uniform random recursive dags

Luc Devroye

School of Computer Science
McGill University
3450 University Street
Montreal H3A 2K6
Canada
luc@cs.mcgill.ca

Svante Janson

Matematiska Institutionen
Uppsala Universitet
Box 480, 751 06 Uppsala
Sweden
Svante.Janson@math.uu.se

May 28, 2009

Abstract. In a uniform random recursive k-dag, there is a root, 0, and each node in turn, from 1 to n, chooses k uniform random parents from among the nodes of smaller index. If S_n is the shortest path distance from node n to the root, then we determine the constant σ such that $S_n/\log n \to \sigma$ in probability as $n \to \infty$. We also show that $\max_{1 \le i \le n} S_i/\log n \to \sigma$ in probability.

Keywords and phrases. Uniform random recursive dag. Randomly generated circuit. Random web model. Longest paths. Probabilistic analysis of algorithms. Branching process.

CR Categories: F.2.0, G.3.0, G.2.2

2000 Mathematics Subject Classifications: 60C05, 60F99, 68R01.

The first author's research was sponsored by NSERC Grant A3456. The research was mostly done at the Institute Mittag-Leffler during the programme "Discrete Probability" held in 2009.

1. Introduction.

A uniform random k-dag is an infinite directed graph defined as follows. For each of the integers $1, 2, \ldots$, we pick a random set of k parents with replacement uniformly from among the smaller non-negative integers. This defines an infinite directed acyclic graph (or, dag) with one root (0), and can be viewed as a (too) simplistic model of the web, a random recursive circuit (Diaz, Sperna, Spirakis, Toran and Tsukiji, 1994, and Tsukiji and Xhafa, 1996), and a generalization of the URRT (uniform random recursive tree), which is obtained for $k = 1$. All the asymptotic results in the paper remain valid when parents are selected without replacement.

The uniform random k-dag restricted to vertices $0, 1, \ldots, n$, is denoted by $U_{k,n}$ or simply U_n. Indeed, we will take $k = 2$ in the main part of the paper, and point out the obvious modifications needed when $k > 2$ as we proceed. The infinite dag is denoted by U_∞.

From a given node n, let $\mathcal{P}_n$ be the collection of paths from node n to the origin. The length of path $p \in \mathcal{P}_n$ is $L(p)$. One can consider various path lengths:

$$S_n = \min_{p \in \mathcal{P}_n} L(p) \ , \ R_n^- = L(P_n^-) \ , \ R_n = L(P_n) \ , \ R_n^+ = L(P_n^+) \ , \ L_n = \max_{p \in \mathcal{P}_n} L(p),$$

where S, R and L are mnemonics for shortest, random, and longest, and P_n^-, P_n and P_n^+ are the paths in $\mathcal{P}_n$, where we follow the parent with the smallest index, the first parent and the parent with the largest index, respectively. We can regard R_n^- and R_n^+ as greedy approximations of S_n and L_n respectively. Note that, at least in a stochastic sense,

$$S_n \le R_n^- \le R_n \le R_n^+ \le L_n.$$

The length of the longest path is relevant for the time to compute the value of node n in a random recursive circuit, when nodes know their value only when all parents know their value. However, there are situations in which node values are determined as soon as one parent or a subset of parents know their value—they are called self-time circuits by Codenotti, Gemmell and Simon (1995). For the one-parent case, this leads naturally to the study of S_n. In networks, in general, shortest paths have been of interest almost since they were conceived (Prim, 1957; Dijkstra, 1959).

It is of interest to study the extreme behavior, as measured by

$$\max_{1 \le \ell \le n} S_\ell \ , \ \max_{1 \le \ell \le n} R_\ell^- \ , \ \max_{1 \le \ell \le n} R_\ell \ , \ \max_{1 \le \ell \le n} R_\ell^+ \ , \ \max_{1 \le \ell \le n} L_\ell.$$

If we replace max by min in these definitions, we obtain the constant 1, and it is therefore more meaningful to ask for the exteme minimal behavior as defined by

$$\min_{n/2 \le \ell \le n} S_\ell \ , \ \min_{n/2 \le \ell \le n} R_\ell^- \ , \ \min_{n/2 \le \ell \le n} R_\ell \ , \ \min_{n/2 \le \ell \le n} R_\ell^+ \ , \ \min_{n/2 \le \ell \le n} L_\ell.$$

So, in all, there are fifteen parameters that could be studied.

We take this opportunity to introduce the label process, which will be referred to throughout the paper. The label of each parent of n is distributed as $\lfloor nU \rfloor$, with U uniform $[0,1]$. An ℓ-th generation ancestor has a label distributed like

$$\lfloor \cdots \lfloor \lfloor nU_1 \rfloor U_2 \rfloor \cdots U_\ell \rfloor \in [nU_1U_2 \cdots U_\ell - \ell, nU_1U_2 \cdots U_\ell],$$

where the U_i's are i.i.d. uniform $[0,1]$ random variables.

THE PARAMETER R_n. It is clear that R_n is just the distance from node n in a URRT to its root. In particular, R_n and its minimal and maximal versions do not depend upon k. We dispense immediately with R_n and its extensions because of well-known results on the URRT obtained via the study of branching random walks by Devroye (1987) and the equivalence between R_n and the number of records in an i.i.d. sequence of continuous random variables (see, e.g., Rényi (1962), Pyke (1965), Glick (1975) or Devroye (1988)). Only the minimal parameter for R_n requires a gentle intervention. We know that

$$\frac{R_n}{\log n} \to 1 \text{ in probability,}$$

for example. Furthermore,

$$\frac{R_n - \log n}{\sqrt{\log n}} \xrightarrow{\mathcal{L}} \mathcal{N}$$

where $\mathcal{N}$ is a standard normal random variable, and $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution. Furthermore, an explicit tail bound on R_n will be needed further on in the paper. The maximal value of $R_\ell, 1 \le \ell \le n$, follows immediately from either Devroye (1987) or Pittel (1994). We summarize:

THEOREM 1. *We have*

$$\frac{R_n}{\log n} \to 1 \text{ in probability,}$$

$$\frac{\max_{1\le\ell\le n} R_\ell}{\log n} \to e \text{ in probability,}$$

and

$$\lim_{n\to\infty} \mathbb{P}\left\{ \min_{n/2\le\ell\le n} R_\ell \le 2 \right\} = 1.$$

Finally, for $t \ge \log n$ integer,

$$\mathbb{P}\{R_n > t\} \le \exp\left(t - \log n - t\log(t/\log n)\right).$$

PROOF. An outline of proof is needed for the third part and the explicit bound in part four. Let us count the number of nodes of index in $[1, n/2]$ that connect directly to the root. This number is

$$Z = \sum_{\ell=1}^{n/2} \xi_{1/\ell},$$

where ξ_p is Bernoulli (p). Let A be the event that no node of index in $(n/2, n]$ connects to a node counted in Z. This probability is smaller than

$$\begin{aligned}\mathbb{E}\left\{\left(1-\frac{Z}{n}\right)^{n/2}\right\} &\le \mathbb{E}\left\{e^{-Z/2}\right\} \\ &\le \prod_{\ell=1}^{n/2}\left(1-1/\ell+1/(\sqrt{e}\ell)\right) \\ &\le \exp\left(-\sum_{\ell=1}^{n/2}\frac{1-1/\sqrt{e}}{\ell}\right) \\ &\le (\lfloor n/2\rfloor)^{1-1/\sqrt{e}}\,.\end{aligned}$$

If the complement of A holds, then clearly, $\min_{n/2\le\ell\le n} R_\ell \le 2$, and thus, we have shown the third part of Theorem 1. Turning to part four, note that $R_n \le \min\{t : nU_1\cdots U_t < 1\}$, and thus that

$$\mathbb{P}\{R_n > t\} \le \inf_{\lambda>0}\mathbb{E}\left\{(nU_1\cdots U_t)^\lambda\right\} = \inf_{\lambda>0} n^\lambda(\lambda+1)^{-t} = \exp\left(t-\log n - t\log(t/\log n)\right). \ \square$$

Conjecture. *For all fifteen parameters, generically denoted by X_n, there exist finite constants $x = x(k) \ge 0$ such that*

$$\frac{X_n}{\log n} \to x \text{ in probability.}$$

Remark. The limits in the conjecture are denoted by σ, ρ^-, ρ, ρ^+ and λ for S_n, R_n^-, R_n, R_n^+ and L_n, respectively. For the minimal and maximal versions of these parameters, we will use the subscripts min and max, respectively, as in $\rho^+_{\min}$ and $\sigma_{\max}$, for example.

Let us briefly survey what is known and provide conjectures in the other cases.

The parameter L_n. Tsukiji and Xhafa (1996) showed that $\lambda_{\max} = ke$. The Chernoff large deviation bound shows that λ is at most the largest solution x of

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

and thus $\lambda < \lambda_{\max}$. We believe that λ is indeed given by (1) based on arguments not unlike the proof of Theorem 2 below. We have no guess at this point about the value of $\lambda_{\min}$.

The parameter R_n^+. In the label process, the parent's index is approximately distributed as $n\max(U_1,\ldots,U_k)$, where the U_i's are i.i.d. uniform $[0,1]$ random variables. If U, as elsewhere in this paper, is uniform $[0,1]$, then the parent's index is thus roughly like $nU^{1/k}$. By renewal theory, this implies that

$$\frac{R_n^+}{\log n} \to k \stackrel{\text{def}}{=} \rho^+ \text{ in probability.}$$

Chernoff's large deviation bound show that $\rho^+_{\max}$ is at most the unique solution x of (2) that is above k:

$$\left(\frac{ke}{x}\right)^x e^{1-k} = 1. \tag{2}$$

We believe that the solution of (2) yields $\rho^+_{\max}$. Applying Chernoff to the other tail shows that $\rho^+_{\min}$ is at least the other solution of (2), as (2) has two solutions, one below k and one above k. Furthermore, we believe that the two solutions of (2) yield $\rho^+_{\min}$ and $\rho^+_{\max}$.

For $k = 2$, the parameter R^+_n is intimately linked to the random binary search tree, which can be grown incrementally by a well-known process described as follows: given an n-node random binary search tree, sample one of its $n+1$ external nodes uniformly at random, replace it by node $n+1$, and continue. The parent of that node is either its neighbor (in the total ordering) to the left or its neighbor to the right, and in fact, it is the neighbor added last to the tree. But the labels (times of insertion) of the neighbors are uniformly drawn without replacement from $\{1, \ldots, n\}$, and are thus roughly distributed as nU, so that the parent of $n+1$ is roughly distributed as $n\sqrt{U}$, because the maximum of two i.i.d. uniform $[0, 1]$ random variables is distributed as $\sqrt{U}$. With this in mind, $\max_{1\le \ell \le n} R^+_\ell$ is the height of the random binary search tree, R^+_n is the depth (distance to the root) of the node of label n (the n-th node inserted), and $\min_{n/2\le \ell\le n} R^+_\ell$ is very roughly the shortest distance from leaf to root, or fill-up level. These quantities behave in probability as described above, as shown by Devroye (1986, 1987), and this explains the values $\rho^+_{\max} = 4.31107\ldots$, $\rho^+ = 2$ and $\rho^+_{\min} = 0.3733\ldots$.

THE PARAMETER R^-_n. Arguing as above, the parent's index is approximately distributed as $n\min(U_1, \ldots, U_k)$. By a property of the uniform (or exponential) distribution, using a sequence of i.i.d. exponential random variables $E_1, E_2, \ldots$, we have this distributional identity:

$$n\min(U_1,\ldots,U_k) \stackrel{\mathcal{L}}{=} nU_1U_2^{1/2}\cdots U_k^{1/k} \stackrel{\mathcal{L}}{=} \exp\left(\log n - \sum_{j=1}^{k}\frac{E_j}{j}\right).$$

Renewal theory easily gives the law of large numbers and central limit theorem for R^-_n. For example,

$$\frac{R^-_n}{\log n} \to \frac{1}{H_k} \stackrel{\text{def}}{=} \rho^- \text{ in probability,}$$

where $H_k = \sum_{j=1}^k (1/j)$ is the k-th harmonic number. Using large deviation bounds similar to the ones used below in showing part of Theorem 2, one gets that

$$\lim_{n\to\infty} \mathbb{P}\left\{\max_{1\le\ell\le n} R^-_\ell \ge (x+\epsilon)\log n\right\} = 0$$

for all $\epsilon > 0$, where x is the solution greater than $1/H_k$ of

$$1 + f(x) = x\sum_{j=1}^{k}\log\left(1 + f(x)/j\right),$$

and $f(x) > 0$ is implicitly defined by

$$\sum_{j=1}^{k}\frac{1}{j+f(x)} = \frac{1}{x}, \qquad x > 1/H_k.$$

These equations follow from the obvious Chernoff bound. We conjecture that $\rho^-_{\max}$ equals this upper bound, but a rigorous proof that $\rho^-_{\max}$ is indeed as described above is not given in this paper.

THE PARAMETER S_n. The most important parameter for computer scientists and combinatorialists is the one in which graph distances are defined by shortest paths, and this leads to the study of S_n. That was the original motivation of the paper, and we will settle first order asymptotics in this paper. Theorem 1 implies, for example, that with probability tending to one,

$$\min_{n/2 \le \ell \le n} S_\ell \le 2.$$

So we turn to σ and $\sigma_{\max}$:

THEOREM 2. *Assume $k \ge 2$. Then $\sigma = \sigma_{\max}$, where σ is given by the solution $x \in (0,1)$ of*

$$\varphi(x) \stackrel{\text{def}}{=} \left(\frac{ke}{x}\right)^x e^{-1} = 1. \tag{3}$$

[Note that φ is indeed an increasing function on $(0,1)$.]

Observe that Theorem 2 does not extend to $k = 1$, because in that case, $S_n \equiv R_n \equiv L_n$, and similarly for the maximal versions of these parameters, in view of the equivalence with the URRT. Thus, $S_n/\log n \to 1$ and $\max_{1\le\ell\le n} S_\ell/\log n \to e$ in probability.

The following is a table of constants in the Conjecture for $k = 2$. The constants involving σ (top row) are obtained in this paper, while those involving ρ (third row) are covered by Theorem 1. The constants ρ^- and ρ^+ follow from ordinary renewal theory. The zeroes in the table follow from Theorem 1. Finally, $\lambda_{\max}$ is due to Tsukiji and Xhafa (1996). There are thus four conjectured constants, which happen to be one-sided bounds ($\rho^-_{\max}$, $\rho^+_{\min}$, $\rho^+_{\max}$, λ), and one unknown constant, $\lambda_{\min}$.

$\sigma_{\min}\ \sigma\ \sigma_{\max}$	0	0.3733...	0.3733...
$\rho^-_{\min}\ \rho^-\ \rho^-_{\max}$	0	0.6666... (= 2/3)	1.6737...
$\rho_{\min}\ \rho\ \rho_{\max}$	0	1	2.7182... (= e)
$\rho^+_{\min}\ \rho^+\ \rho^+_{\max}$	0.3733...	2	4.3110...
$\lambda_{\min}\ \lambda\ \lambda_{\max}$	?	4.3110...	5.4365... (= $2e$)

The following is a table of σ, ρ^- and $\rho^-_{\max}$ for different numbers of parents k.

k	σ	ρ^-	$\rho^-_{\max}$
2	0.3733...	0.6666...	1.6737...
3	0.3040...	0.5454...	1.3025...
4	0.2708...	0.48	1.1060...
5	0.2503...	0.4379...	0.9818...
6	0.2361...	0.4081...	0.8951...
7	0.2254...	0.3856...	0.8305...
8	0.2170...	0.3679...	0.7800...
9	0.2102...	0.3534...	0.7393...
10	0.2045...	0.3414...	0.7057...
11	0.1996...	0.3311...	0.6773...
12	0.1954...	0.3222...	0.6531...
13	0.1916...	0.3144...	0.6318...
14	0.1883...	0.3075...	0.6132...
15	0.1854...	0.3013...	0.5966...
16	0.1827...	0.2957...	0.5816...
17	0.1802...	0.2907...	0.5683...
18	0.1780...	0.2861...	0.5560...
19	0.1760...	0.2818...	0.5448...
20	0.1740...	0.2779...	0.5346...
21	0.1723...	0.2743...	0.5251...
22	0.1706...	0.2709...	0.5164...
23	0.1691...	0.2677...	0.5083...
24	0.1676...	0.2648...	0.5007...
25	0.1663...	0.2620...	0.4936...
26	0.1650...	0.2594...	0.4868...
27	0.1638...	0.2569...	0.4805...
28	0.1626...	0.2546...	0.4747...
29	0.1615...	0.2524...	0.4690...
30	0.1604...	0.2503...	0.4638...
35	0.1559...	0.2411...	0.4409...
40	0.1521...	0.2337...	0.4225...
45	0.1490...	0.2275...	0.4074...
50	0.1463...	0.2222...	0.3946...

2. The shortest path length S_n.

We will establish Theorem 2 in two parts. First we show that for all $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\{S_n \le (1-\epsilon)\sigma \log n\} = 0, \tag{4}$$

and then that

$$\lim_{n\to\infty} \mathbb{P}\left\{\max_{1\le \ell\le n} S_\ell \ge (1+\epsilon)\sigma\log n\right\} = 0. \tag{5}$$

We only consider the case $k = 2$ since the case $k > 2$ follows quite easily.

Lemma 1. *Let G_a be gamma(a), with $a \ge 1$. Then*

$$\frac{\mathbb{P}\{G_a \ge x\}}{\frac{x^{a-1}e^{-x}}{\Gamma(a)}} \le \frac{1}{1-\frac{a-1}{x}}, x > a-1,$$

and

$$\frac{\mathbb{P}\{G_a \le x\}}{\frac{x^{a-1}e^{-x}}{\Gamma(a)}} \le \frac{1}{\frac{a-1}{x}-1}, x < a-1.$$

Proof. The gamma density is $f(y) = y^{a-1}e^{-y}/\Gamma(a)$. It is log-concave for $a \ge 1$, and thus, a first-term Taylor series bound yields the inequality

$$f(y) \le f(x)e^{(y-x)(\log f)'(x)} = f(x)e^{(y-x)((a-1)/x-1)}.$$

Integrating the upper bound out over $[x,\infty)$ or $(-\infty, x]$ then immediately yields the results. $\square$

From node n, we can consider the index of the first of the 2^ℓ ℓ-th level ancestors, which is distributed as

$$\lfloor \cdots \lfloor\lfloor nU_1\rfloor U_2\rfloor \cdots U_\ell\rfloor \ge nU_1U_2\cdots U_\ell - \ell \stackrel{\mathcal{L}}{=} n\exp(-G_\ell) - \ell,$$

where $\stackrel{\mathcal{L}}{=}$ denotes equality in distribution, and G_ℓ is gamma(ℓ). If these indices are $I_1, \ldots, I_{2^\ell}$, then we have

$$\begin{aligned}
\mathbb{P}\{S_n \le \ell\} &= \mathbb{P}\left\{\min_{1\le i\le 2^\ell} I_i = 0\right\} \\
&\le 2^\ell \mathbb{P}\{I_1 = 0\} \\
&\le 2^\ell \mathbb{P}\{n\exp(-G_\ell) - \ell \le 0\} \\
&= 2^\ell \mathbb{P}\{G_\ell \ge \log(n/\ell)\} \\
&\le \frac{2^\ell(\log(n/\ell))^{\ell-1}e^{-\log(n/\ell)}}{\Gamma(\ell)\left(1-\frac{\ell-1}{\log(n/\ell)}\right)} \qquad (\text{if } \log(n/\ell) \ge \ell-1) \\
&\le \frac{\ell^{3/2}(2\log(n))^\ell e^{-\log(n)}}{(\ell/e)^\ell\left(1-\frac{\ell-1}{\log(n/\ell)}\right)}.
\end{aligned}$$

Set $\ell = \lfloor t \log n \rfloor$ for $t \in (0,1)$, and note that the upper bound is

$$\Theta\left(\log^{3/2}(n)\right) \times (\varphi(t))^{\log n},$$

where $\varphi(t) = (2e/t)^t/e$ is as in (3). We have $\varphi(\sigma) = 1$ for $\sigma = 0.3733\ldots$. Thus, we have shown (4): for all $\epsilon > 0$,

$$\mathbb{P}\{S_n \le (\sigma - \epsilon)\log n\} = o(1).$$

Although we will not need it directly, we will also deal with the upper bound on S_n. This can be done in a number of ways, but the shortest route is perhaps via the great-grandparent strategy that jumps ℓ generations at a time, where ℓ now is a large but fixed integer. We denote this by ℓ-GGP. We associate with each node n two independent uniform $[0,1]$ integers U and V and let the parent labels be $\lfloor nU \rfloor$ and $\lfloor nV \rfloor$. Let A_n be the event that any of the 2^ℓ ancestors of node n conicide. It is clear that $\mathbb{P}\{A_n\} \to 0$ as $n \to \infty$. As an ancestor label is described by

$$\lfloor \cdots \lfloor \lfloor nU_1 \rfloor U_2 \rfloor \cdots U_\ell \rfloor \le nU_1U_2\cdots U_\ell \overset{\mathcal{L}}{=} n\exp(-G_\ell),$$

we define

$$Z_\ell = \min_{p \in \mathcal{P}} \prod_{e \in p} U_e$$

where $\mathcal{P}$ is the collection of all paths of length ℓ above node n, and each $p \in \mathcal{P}$ consists of edges e that each have an independent uniform random variable associated with it. If $\epsilon > 0$ and n is greater than some n_ϵ, then the ℓ-GGP gives with probability greater than $1-\epsilon$ a node with label less than $Z_\ell n$. Define

$$Z_\ell^{(\epsilon)} = \min(Z_\ell, b)$$

where b is chosen such that $\mathbb{P}\{Z_\ell > b\} = \epsilon$. As long as the label stays above n_ϵ, one can dominate the labels in the ℓ-GGP by multiplying n with successive independent copies of $Z_\ell^{(\epsilon)}$. Let T_n be the number of steps until the label in ℓ-GGP reaches n_ϵ or less. Renewal theory shows that with probability tending to one,

$$T_n \le \frac{(1+\epsilon)\log n}{\mathbb{E}\left\{-\log\left(Z_\ell^{(\epsilon)}\right)\right\}}.$$

because the ℓ-GGP takes ℓ steps at a time, and because a node with label n_ϵ is not further than n_ϵ away from the origin, we see that with probability tending to one,

$$\begin{aligned} S_n &\le n_\epsilon + \frac{\ell(1+\epsilon)\log n}{\mathbb{E}\left\{-\log\left(Z_\ell^{(\epsilon)}\right)\right\}} \\ &\le \frac{\ell(1+2\epsilon)\log n}{\mathbb{E}\left\{-\log\left(Z_\ell^{(\epsilon)}\right)\right\}}. \end{aligned}$$

Uniform integrability implies that

$$\lim_{\epsilon \downarrow 0} \mathbb{E}\left\{-\log\left(Z_\ell^{(\epsilon)}\right)\right\} = \mathbb{E}\left\{-\log\left(Z_\ell\right)\right\}.$$

Therefore, for any (new, fresh) $\epsilon > 0$ and $\ell \ge 1$, with probability going to one,

$$S_n \le \frac{\ell(1+\epsilon)\log n}{\mathbb{E}\left\{-\log\left(Z_\ell\right)\right\}}.$$

Observe that

$$\frac{-\log(Z_\ell)}{\ell} \stackrel{\mathcal{L}}{\leq} \frac{1}{\ell} \max_{p\in\mathcal{P}} \sum_{e\in p} E_u,$$

where the E_u are i.i.d. exponential random variables. From the theory of branching random walks, it is easy to verify (see, e.g., Biggins (1977), or Devroye (1986, 1987)) that, as $\ell \to \infty$,

$$\frac{1}{\ell} \max_{p\in\mathcal{P}} \sum_{e\in p} E_u \to \frac{1}{\sigma}$$

in probability. Thus,

$$\liminf_{\ell\to\infty} \frac{-\mathbb{E}\{\log(Z_\ell)\}}{\ell} \geq \frac{1}{\sigma},$$

and thus, by choosing ℓ large enough, we see that with probability tending to one,

$$S_n \leq (1+2\epsilon)\sigma \log n.$$

This concludes the proof of the first part of Theorem 2.

The next section requires an explicit rate of convergence. To this end, still restricting ourselves to $k=2$ only, let $Z_{\ell,1}^{(\epsilon)}, Z_{\ell,2}^{(\epsilon)}, \ldots$ be i.i.d. copies of $Z_\ell^{(\epsilon)}$, and note that,

$$\begin{aligned} T_n &\leq \min\left\{t : nZ_{\ell,1}^{(\epsilon)} \cdots Z_{\ell,t}^{(\epsilon)} < 1\right\} \\ &= \min\left\{t : \log\left(1/Z_{\ell,1}^{(\epsilon)}\right) + \cdots + \log\left(1/Z_{\ell,t}^{(\epsilon)}\right) > \log n\right\}. \end{aligned}$$

Set $\mu = \mathbb{E}\left\{\log\left(1/Z_\ell^{(\epsilon)}\right)\right\}$. Then, assuming $\delta^* \in (0,1/2)$ and $\delta \in (\delta^*, 2\delta^*)$ such that $m = (1/\mu+\delta)\log n$ is integer-valued,

$$\begin{aligned} \mathbb{P}\{T_n > m\} &\leq \mathbb{P}\left\{\log\left(1/Z_{\ell,1}^{(\epsilon)}\right) + \cdots + \log\left(1/Z_{\ell,m}^{(\epsilon)}\right) < \log n\right\} \\ &= \mathbb{P}\left\{\log\left(1/Z_{\ell,1}^{(\epsilon)}\right) + \cdots + \log\left(1/Z_{\ell,m}^{(\epsilon)}\right) - m\mu < -\delta\mu\log n\right\}. \end{aligned}$$

Let $p > 2$ be a fixed number. Rosenthal's inequality (Rosenthal, 1970, Fuk and Nagaev, 1971, see also Petrov, 1975) states that there is a constant C_p with the following property. If $\{X_n, n \geq 1\}$ is a sequence of centered and independent random variables, and if $Y_n = X_1 + \cdots + X_n$, and if $\mathbb{E}\{|X_n|^p\} < \infty$ for all n, then

$$\mathbb{E}\{|Y_n|^p\} \leq C_p \left(\sum_{j=1}^n \mathbb{E}\{|X_j|^p\} + (\mathbb{V}\{Y_n\})^{p/2}\right).$$

For i.i.d. random variables with $X_1 = X$, we have

$$\mathbb{E}\{|Y_n|^p\} \leq C_p \left(n\mathbb{E}\{|X|^p\} + n^{p/2}\left(\mathbb{E}\{X^2\}\right)^{p/2}\right) \leq 2C_p \max(n, n^{p/2})\mathbb{E}\{|X|^p\}.$$

Applied to our situation with $p=4$, using Markov's inequality, we have

$$\begin{aligned} \mathbb{P}\{T_n > m\} &\leq (\delta\mu\log n)^{-4}\mathbb{E}\left\{\left(\log\left(1/Z_{\ell,1}^{(\epsilon)}\right) + \cdots + \log\left(1/Z_{\ell,m}^{(\epsilon)}\right) - m\right)^4\right\} \\ &\leq 2C_4(\delta\mu\log n)^{-4} m^2 \mathbb{E}\left\{\left|\log\left(1/Z_\ell^{(\epsilon)}\right) - \mu\right|^4\right\} \\ &\leq C(\log n)^{-2}\delta^{*-4}, \end{aligned}$$

where C depends upon ϵ and ℓ only. The remainder of the argument involving an appropriate choice of ℓ remains valid, and we can conclude that for any $\epsilon > 0$,

$$\mathbb{P}\{S_n > (\sigma + \epsilon)\log n\} = O\left(1/\log^2 n\right), \tag{6}$$

with room to spare.

3. The maximal shortest path length

The purpose of this section is to show (5). We let σ be as in the first part of the proof, and let $\epsilon > 0$ be arbitrary. Fix n large enough. From (6),

$$\mathbb{E}\left\{\left|\{j : n/2 \le j \le n, S_j > (\sigma+\epsilon)\log n\}\right|\right\} = O\left(\frac{n}{\log^2 n}\right),$$

and thus $\mathbb{P}\{A(n)\} = O\left(\frac{1}{\log^2 n}\right)$, where

$$A(n) \stackrel{\text{def}}{=} \left[\left|\{j : n/2 \le j \le n, S_j > (\sigma+\epsilon)\log n\}\right| > \frac{n}{4}\right].$$

If we take an incremental view of the process of adding edges, then a node with index in $[n, 2n]$ selects a parent of depth $\le (\sigma+\epsilon)\log n$ and index $\ge n/2$ with probability $\ge 1/8$ if $A(n)$ fails to hold. It is this observation that will allow us to uniformly bound all depths by something close to $(\sigma+\epsilon)\log n$.

Consider the indices in dyadic groups, $\{2^{r-1}+1, \ldots, 2^r\}$, $r \ge 1$. We recall from a comparison with the URRT, that $S_n \le R_n$ and thus that $\max_{1\le j\le n} S_j \le \max_{1\le j\le n} R_j$, and that (see Theorem 1)

$$\mathbb{P}\left\{\max_{1\le j\le n} R_j > 2e\log n\right\} \le n^{-2e\log(2)} < n^{-3}.$$

Thus, for $\gamma > 0$ small enough,

$$\mathbb{P}\left\{\max_{1\le j\le \lfloor n^\gamma\rfloor} S_j > (\sigma+\epsilon)\log n\right\} = O(n^{-3\gamma}) = o(1).$$

It remains to show that

$$\mathbb{P}\left\{\max_{n^\gamma\le j\le n} S_j > (\sigma+\epsilon)\log n\right\} = o(1).$$

Consider the event

$$B(r) = \bigcup_{r'\le s\le r} A(2^s),$$

where r' is the largest integer such that $2^{r'} < n^\gamma$. Clearly, $\mathbb{P}\{B(r)\} = O(1/r') = O(1/\log n)$. On the complement, $(B(r))^c$, intersected with $\left[\max_{1\le j\le\lfloor n^\gamma\rfloor} S_j \le (\sigma+\epsilon)\log(n)\right]$, we look at the process started at a node $m \le n$ and assume that its index m is in $\left\{2^r+1, \ldots, 2^{r+1}\right\}$. That process is looked at as a binary tree of consecutive parents, and will be cut off at height $h = \lfloor 10\log\log n\rfloor$. There may be duplicate parents (in which case the tree degenerates to a dag), so we need to be a bit careful. If any parent in the tree is selected with index $\le 2^{r'} < n^\gamma$, then $S_m \le (\sigma+\epsilon)\log n + h$, and thus, we can assume that in this "tree" any node j selects its parent uniformly in the range $(2^{r'}, j)$. At any stage, by our assumption, the probability of picking a parent i having $S_i \le (\sigma+\epsilon)\log n$ is at least $1/8$ (and this is why we needed the dyadic trick, so that we can make this statement regardless of the choice of i within the range $(2^{r'}, n]$).

We claim that this "tree" has at least 2^{h-1} leaves or reaches $[1, 2^{r'}]$ with overwhelming probability. To see this, note that a node j in it picks a node already selected with probability not exceeding $2^h/j$. But the index j is stochastically larger than

$$X_h \stackrel{\text{def}}{=} \lfloor \cdots \lfloor \lfloor mU_1 \rfloor U_2 \rfloor \cdots U_h \rfloor$$

by our remarks about the labeling process. The probability that there are in fact at least two such unwanted parent selections (but none of them less than n^γ) in that "tree" is not more than

$$2^{2h+2} \times \mathbb{E}^2 \left\{ \frac{2^h}{X_h} \mathbb{1}_{[X_h \geq n^\gamma]} \right\} \leq 2^{4h+2} \times \mathbb{E}^2 \left\{ \frac{1}{X_h} \mathbb{1}_{[X_h \geq n^\gamma]} \right\} \tag{7}$$

We have

$$\begin{aligned}
\mathbb{E}\left\{ X_h^{-1} \mathbb{1}_{[X_h \geq n^\gamma]} \right\} &= \int_0^\infty \mathbb{P}\{X_h^{-1} \mathbb{1}_{[X_h \geq n^\gamma]} > t\}\, dt \\
&= \int_0^{1/n^\gamma} \mathbb{P}\{X_h < 1/t\}\, dt \\
&\leq \int_0^{1/n^\gamma} \mathbb{P}\{mU_1 \cdots U_h < h + 1/t\}\, dt \\
&= \int_0^{1/n^\gamma} \mathbb{P}\{\log \frac{m}{h+1/t} < G_h\}\, dt \\
&= \int_0^{1/n^\gamma} \int_{\log_+ \frac{m}{h+1/t}}^\infty \frac{y^{h-1}e^{-y}}{\Gamma(h)}\, dy\, dt \\
&= \int_0^\infty \frac{y^{h-1}e^{-y}}{\Gamma(h)} \min\left(n^{-\gamma}, \frac{1}{(me^{-y} - h)_+} \right) dy \\
&\leq \int_0^{\log(m/2h)} \frac{2y^{h-1}}{\Gamma(h)m}\, dy + n^{-\gamma} \int_{\log(m/2h)}^\infty \frac{y^{h-1}e^{-y}}{\Gamma(h)}\, dy \\
&\leq \frac{2(\log(n))^h}{m\, h!} + \frac{n^{-\gamma}(\log(n))^{h-1}4h}{\Gamma(h)m} \quad \text{(for n large enough, by lemma 1)} \\
&= O\left(n^{o(1)}/m\right) = O\left(m^{-1+o(1)}\right).
\end{aligned}$$

Thus, our probability (7) is not larger than $O\left(m^{-2+o(1)}\right)$. If there is only one unwanted parent selection and we avoid indices below n^γ, and considering that the first parent selection at the root node is always good, we see that at least half of the 2^h potential leaves are in fact realized. Each of these leaves makes two independent parent selections. The probability that all these leaves avoid parents j with $S_j < (\sigma+\epsilon)\log n$ is at most $(7/8)^{2^{h-1}} = o(n^{-2})$. If there is a connection, however, to such a parent of low depth, then the root has shortest path length at most $h+1$ more than $(\sigma+\epsilon)\log n$. Hence, if $\mathcal{E}_m$ is the event $[S_m > (\sigma+\epsilon)\log n + h + 1]$, then

$$\mathbb{P}\left\{ \mathcal{E}_m \cap ((B(r))^c \cap \left[\max_{1 \leq j \leq \lfloor n^\gamma \rfloor} S_j \leq (\sigma+\epsilon)\log n \right] \right\} = O\left(m^{-2+o(1)}\right).$$

Thus

$$\begin{aligned}
&\mathbb{P}\left\{\max_{n^\gamma \le j \le n} S_j > (\sigma+\epsilon)\log n + h + 1\right\} \\
&= \mathbb{P}\left\{\cup_{m\ge n^\gamma}^n \mathcal{E}_m\right\} \\
&\le \mathbb{P}\left\{\max_{1\le j\le \lfloor n^\gamma \rfloor} S_j > (\sigma+\epsilon)\log n\right\} + \mathbb{P}\{B(r)\} + \sum_{m\ge n^\gamma}^{n} m^{-2+o(1)} \\
&= O(n^{-3\gamma}) + O(1/r') + n^{-\gamma+o(1)} \\
&= O(1/\log n).
\end{aligned}$$

This concludes the proof of the theorem. □

4. Bibliographic remarks and possible extensions.

The study of the URRT goes back as far as Na and Rapoport (1970) and Meir and Moon (1978). Single nonuniform parent selections have been considered as early as 1987 by Szymański. Szymański (1987) showed that if a parent is selected with probability proportional to its degree, then with high probability there is a node of degree $\Omega(\sqrt{n})$. This is nothing but the preferential attachment model of Barabasi and Albert (see Albert, Barabasi and Jeong, 1999, or Albert and Barabasi, 1999), which for a single parent is a special case of the linear recursive trees or PORT (plane-oriented recursive tree). For this model, the parameter R_n was studied by Mahmoud (1992a), and the height by Pittel (1994) and Biggins and Grey (1997), and in a rather general setting by Broutin and Devroye (2006): the height is in probability $(1.7956\ldots + o(1))\log n$. The profile (number of nodes at each depth level) was studied by Hwang (2005, 2007) and Sulzbach (2008).

One can ask the questions studied in the present paper for these more general models.

Various aspects of URRT's besides the depth and height have been studied by many researchers. These include the degrees of the nodes, the profile, sizes of certain subtrees of certain nodes, the number of leaves, and so forth. Surveys and references can be found in the book by Mahmoud (1992b) or the paper by Devroye (1998). Specific early papers include Timofeev (1984), Gastwirth (1997), Dondajewski and Szymański (1982), Mahmoud (1991), Mahmoud and Smythe (1991), Smythe and Mahmoud (1994), Szymański (1990), and the most recent contributions include Fuchs, Hwang and Neininger (2006), and Drmota, Janson and Neininger (2008). One may wonder how the profiles behave for uniform random k-dags.

4. References

R. Albert and A. Barabasi, 1999, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–512.

R. Albert, A. Barabasi, and H. Jeong, 1999, "Diameter of the World-Wide Web," *Nature*, vol. 401, p. 130.

J. L. Balcazar, J. Diaz, and J. Gabarro, 1995, *Structural Complexity I*, Springer-Verlag, Berlin.

J. D. Biggins, 1976, "The first and last-birth problems for a multitype age-dependent branching process," *Advances in Applied Probability*, vol. 8, pp. 446–459.

J. D. Biggins, 1977, "Chernoff's theorem in the branching random walk," *Journal of Applied Probability*, vol. 14, pp. 630–636.

J. D. Biggins and D. R. Grey, 1997, "A note on the growth of random trees," *Statistics and Probability letters*, vol. 32, pp. 339–342.

N. Broutin and L. Devroye, 2006, "Large deviations for the weighted height of an extended class of trees," *Algorithmica*, vol. 46, pp. 271–297.

N. Broutin, L. Devroye, and E. McLeish, 2008, "Weighted height of random trees," *Acta Informatica*, vol. 45, pp. 237–277.

B. Codenotti, P. Gemmell, and J. Simon, 1995, "Average circuit depth and average communication complexity," in: *Third European Symposium on Algorithms*, pp. 102–112. Springer-Verlag, Berlin.

L. Devroye, 1986, "A note on the height of binary search trees," *Journal of the ACM*, vol. 33, pp. 489–498.

L. Devroye, 1987, "Branching processes in the analysis of the heights of trees," *Acta Informatica*, vol. 24, pp. 277–298.

L. Devroye, 1988, "Applications of the theory of records in the study of random trees," *Acta Informatica*, vol. 26, pp. 123–130.

L. Devroye, 1998, "Branching processes and their applications in the analysis of tree structures and tree algorithms," in: *Probabilistic Methods for Algorithmic Discrete Mathematics*, edited by M. Habib, C. McDiarmid, J. Ramirez-Alfonsin and B. Reed, vol. 16, pp. 249–314. Springer-Verlag, Berlin.

L. Devroye, 1999, "Universal limit laws for depths in random trees," *SIAM Journal on Computing*, vol. 28, pp. 409–432.

J. Diaz, M. J. Serna, P. Spirakis, J. Toran, and T. Tsukiji, 1994, "On the expected depth of Boolean circuits," Technical Report LSI-94-7-R, Universitat Politecnica de Catalunya, Dep. LSI.

E. W. Dijkstra, 1959, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, pp. 269–271.

M. Dondajewski and J. Szymański, 1982, "On the distribution of vertex-degrees in a strata of a random recursive tree," *Bulletin de l'Académie Polonaise des Sciences, Série des Sciences Mathématiques*, vol. 30, pp. 205–209.

M. Drmota, S. Janson, and R. Neininger, 2008, "A functional limit theorem for the profile of search trees," *Annals of Applied Probability*, vol. 18, pp. 288–333.

M. Fuchs, H.-K. Hwang, and R. Neininger, 2006, "Profiles of random trees: Limit theorems for random recursive trees and binary search trees," *Algorithmica*, vol. 46, pp. 367–407.

D. K. Fuk and S. V. Nagaev, 1971, "Probability inequalities for sums of independent random variables," *Theory of Probability and its Applications*, vol. 16, pp. 643–660.

J. L. Gastwirth, 1977, "A probability model of a pyramid scheme," *The American Statistician*, vol. 31, pp. 79–82.

N. Glick, 1978, "Breaking records and breaking boards," *American Mathematical Monthly*, vol. 85, pp. 2–26.

H.-K. Hwang, 2005, "Profiles of random trees: plane-oriented recursive trees (Extended Abstract)," in: *International Conference on Analysis of Algorithms,DMTCS Proceedings AD*, pp. 193–200.

H.-K. Hwang, 2007, "Profiles of random trees: Plane-oriented recursive trees," *Random Structures and Algorithms*, vol. 30, pp. 380–413.

H. M. Mahmoud, 1991, "Limiting distributions for path lengths in recursive trees," *Probability in the Engineering and Informational Sciencies*, vol. 5, pp. 53–59.

H. Mahmoud, 1992a, "Distances in plane-oriented recursive trees," *Journal of Computers and Applications in Mathematics*, vol. 41, pp. 237–245.

H. Mahmoud, 1992b, *Evolution of Random Search Trees*, Wiley, New York.

H. M. Mahmoud and R. T. Smythe, 1991, "On the distribution of leaves in rooted subtrees of recursive trees," *Annals of Applied Probability*, vol. 1, pp. 406–418.

H. Mahmoud and B. Pittel, 1984, "On the most probable shape of a search tree grown from a random permutation," *SIAM Journal on Algebraic and Discrete Methods*, vol. 5, pp. 69–81.

H. S. Na and A. Rapoport, 1970, "Distribution of nodes of a tree by degree," *Mathematical Biosciences*, vol. 6, pp. 313–329.

V. V. Petrov, 1995, *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*, Clarendon Press, Oxford.

B. Pittel, 1984, "On growing random binary trees," *Journal of Mathematical Analysis and Applications*, vol. 103, pp. 461–480.

B. Pittel, 1994, "Note on the heights of random recursive trees and random m-ary search trees," *Random Structures and Algorithms*, vol. 5, pp. 337–347.

R. C. Prim, 1957, "Shortest connection networks and some generalizations," *BSTJ*, vol. 36, pp. 1389–1401.

R. Pyke, 1965, "Spacings," *Journal of the Royal Statistical Society Series B*, vol. 7, pp. 395–445.

A. Rényi, 1962, "Theorie des elements saillant d'une suite d'observations," in: *Colloquium on Combinatorial Methods in Probability Theory*, pp. 104–115. Mathematisk Institut, Aarhus Universitet, Denmark.

H. P. Rosenthal, 1970, “On the subspaces of L^p ($p > 2$) spanned by sequences of independent random variables,” *Israel Journal of Mathematics*, vol. 8, pp. 273–303.

R. T. Smythe and H. M. Mahmoud, 1994, “A survey of recursive trees,” *Teorya Imovirnostyta Mat. Stat. (in Ukrainian)*, vol. 51, pp. 1–29.

H. Sulzbach, 2008, “A functional limit law for the profile of plane-oriented recursive trees.,” in: *Fifth Colloquium on Mathematics and Computer Science, DMTCS Proceedings AI*, pp. 339–350.

J. Szymański, 1987, “On a nonuniform random recursive tree,” *Annals of Discrete Mathematics*, vol. 33, pp. 297–306.

J. Szymański, 1990, *On the maximum degree and height of a random recursive tree*, Wiley, New York.

E. A. Timofeev, 1984, “Random minimal trees,” *Theory of Probability and its Applications*, vol. 29, pp. 134–141.

T. Tsukiji and F. Xhafa, 1996, “On the depth of randomly generated circuits,” in: *Proceedings of Fourth European Symposium on Algorithms*.