THE STRONG CONVERGENCE OF MAXIMAL DEGREES IN UNIFORM RANDOM RECURSIVE TREES AND DAGS

Luc Devroye
and
Jiang Lu
School of Computer Science
and
Department of Mathematics and Statistics
McGill University

ABSTRACT. We show that the maximal degree in a uniform random recursive tree is almost surely $(1+o(1))\log_2 n$. A random directed acyclic graph on n nodes is defined by by connecting the i-th node for each $i > m$ with r of its predecessors uniformly and at random. The maximal degree is shown to be almost surely $(1+o(1))\log_{1+1/r} n$.

KEYWORDS AND PHRASES. Uniform random recursive tree. Directed acyclic graph. Probabilistic analysis. Random tree. Strong convergence. Maximal degree. Second moment method.

Authors' address: School of Computer Science, McGill University, 3480 University Street, Montreal, Canada H3A 2K6. The authors' research was sponsored by NSERC Grant A3456 and FCAR Grant 90-ER-0291.

1. Introduction.

A uniform random recursive tree (or URRT) on n nodes is a tree recursively constructed by letting the i-th node pick its parent uniformly and at random from among the first $i-1$ nodes. A uniform random recursive dag (or URRD) on n nodes starts this process only at node $m+1$, so that the first m nodes are roots. Furthermore, the i-th node picks r nodes uniformly from among the first $i-1$ nodes to be its "parents", thus creating a directed acyclic graph. The object of this note is to study the maximal degree in these structures, as well as in certain subtrees rooted at the k-th node.

Na and Rapoport (1970), Moon (1974), Gastwirth (1977), Meir and Moon (1978), Najock and Heyde (1982), Dondajewski and Szymański (1982), Gastwirth and Bhattacharya (1984), Devroye (1987, 1988), Szymański (1987, 1990), Mahmoud (1992), Mahmoud and Smythe (1991), and Pittel (1994), have studied the URRT in some detail. A URRT of course is just a URRD with $m = 1$. Dags model expression trees in which the symbols are the roots and the mathematical operators correspond to internal nodes. They also model PERT networks, and represent partial orders in general. In the latter context, random dags, different from the ones studied here, were suggested by Winkler (1985) and Bollobás and Brightwell (1991). Key properties are obtained in these papers and by Bollobás and Winkler (1988) and Frieze (1991). In our model, both the number of roots (m) and the number of parents (r) can be controlled to yield a rich family of random dags.

The degree of a node is the number of children. In a URRD, the number of parents is m for all non-root nodes, and 0 for the roots. In a URRT, the degree sequence has been studied by Na and Rapoport (1970), Gastwirth and Bhattacharya (1984), and Mahmoud (1992). Let M_n be the maximal degree in a URRT or URRD. Szymański (1987, 1990) showed that for the URRT,

$$\limsup \frac{\mathbf{E}M_n}{\log_2 n} \leq 1.$$

Let T denote a URRD or URRT, and T_k be the subtree rooted at the k-th node. Quantities studied in this note include M_n, the maximal degree in T, and $M_{n,k}$, the maximal degree in T_k. Both are shown to be about a constant times $\log n$. As the URRT model is a natural asymptotic growth model, the strong convergence of various random parameters does matter. For that reason, we will spend some time on almost sure convergence results. Although some results for the URRT follow from those for the URRD, the proofs for the URRT are more transparent, and will be given separately.

THEOREM 1. *In a* URRT, $M_n/\log_2 n \to 1$ *almost surely and* $\lim_{n\to\infty} \mathbf{E}M_n/\log_2 n = 1$. *For a fixed integer* k, $M_{n,k}/\log_2 n \to 1$ *almost surely.*

THEOREM 2. *In a* URRD, *for fixed integers* r *and* m *with* $1 \le r \le m$,

$$\lim_{n\to\infty} \frac{M_n}{\log_{1+1/r} n} = 1 \text{ almost surely.}$$

REMARK. If we define a random graph on n nodes by choosing $n-1$ edges at random from the $n(n-1)/2$ possible edges, then the maximal degree is in probability asymptotic to $\log n/\log\log n$. This follows from arguments used in hashing (Gonnet, 1981; Devroye, 1985); see also Bollobás (1985, p. 72). Moon (1968) established a similar result for uniform labeled trees, in which each of the n^{n-2} trees is equally likely. Szymański (1987) showed that for a nonuniform random recursive tree in which a node is selected with probability proportional to its degree, $\mathbf{E}M_n \ge \mathbf{E}D_1 \sim 2\sqrt{n/\pi}$. In all these examples, the difference with the URRT is quite remarkable.

2. Proof of Theorem 1.

Let R_j be the node among $\{1,\ldots,j-1\}$ that is the parent of node j. Let D_i be the degree of node i in the URRT. We have

$$D_i = \sum_{j>i} I_{[R_j=i]}, i \ge 1.$$

By definition, $R_2,\ldots,R_n$ are independent. Also, for $j \ge 1$,

$$\mathbf{P}\{R_j = i\} = \frac{1}{j-1},\ 1 \le i < j.$$

Denoting $H_n = \sum_{j=1}^n (1/j)$, we have

$$\mathbf{E}D_i = H_{n-1} - H_i,\ 1 \le i \le n,$$

and

$$\mathrm{Var}\{D_i\} = \sum_{j=i+1}^{n} \frac{1}{j-1} \times \left(1 - \frac{1}{j-1}\right) \le \mathbf{E}D_i.$$

By Chebyshev's inequality, we conclude that for all $\epsilon > 0$,

$$\lim_{n\to\infty} \sup_{i \le \omega_n} \mathbf{P}\left\{\left|\frac{D_i}{\mathbf{E}D_i} - 1\right| > \epsilon\right\} = 0,$$

where $0 < \omega_n = o(n)$. The first few degrees are about $\log n$. This establishes that $\liminf \mathbf{E}M_n/\log n \ge 1$, which unfortunately, is not tight. We will be able to obtain Theorem 1 via the following device.

Lemma 1. *Let a_n be a sequence of positive numbers, and let A_{ni} be the event $[D_i \geq a_n]$. Then*

$$\sum_{i=1}^{n} \mathbf{P}\{A_{ni}\} \geq \mathbf{P}\left\{\cup_{i=1}^{n} A_{ni}\right\} \geq \frac{\sum_{i=1}^{n} \mathbf{P}\{A_{ni}\}}{1+\sum_{i=1}^{n} \mathbf{P}\{A_{ni}\}}.$$

Also,

$$\lim_{n\to\infty} \mathbf{P}\left\{M_n \geq a_n\right\} = \begin{cases} 0 & \textit{if} \lim_{n\to\infty} \sum_{i=1}^{n} \mathbf{P}\left\{A_{ni}\right\} = 0 \\ 1 & \textit{if} \lim_{n\to\infty} \sum_{i=1}^{n} \mathbf{P}\left\{A_{ni}\right\} = \infty \ . \end{cases}$$

Proof. The first chain of inequalities is obtained via Bonferroni's bound and by means of the Chung-Erdös inequality (Chung and Erdös, 1952; see also Dawson and Sankoff, 1967),

$$\mathbf{P}\left\{\cup_{i=1}^{n} A_{ni}\right\} \geq \frac{\left(\sum_{i=1}^{n} \mathbf{P}\left\{A_{ni}\right\}\right)^2}{\sum_{i=1}^{n} \mathbf{P}\left\{A_{ni}\right\} + \sum_{i\neq j} \mathbf{P}\left\{A_{ni}A_{nj}\right\}}.$$

As we will show further on,

$$\mathbf{P}\left\{A_{ni}A_{nj}\right\} \leq \mathbf{P}\left\{A_{ni}\right\} \mathbf{P}\left\{A_{nj}\right\},$$

for all $i \neq j$, so that

$$\begin{aligned} \mathbf{P}\left\{\cup_{i=1}^{n} A_{ni}\right\} &\geq \frac{\left(\sum_{i=1}^{n} \mathbf{P}\left\{A_{ni}\right\}\right)^2}{\sum_{i=1}^{n} \mathbf{P}\left\{A_{ni}\right\} + \left(\sum_{i=1}^{n} \mathbf{P}\left\{A_{ni}\right\}\right)^2 - \sum_{i=1}^{n} \left(\mathbf{P}\left\{A_{ni}\right\}\right)^2} \\ &\geq \frac{\sum_{i=1}^{n} \mathbf{P}\left\{A_{ni}\right\}}{1+\sum_{i=1}^{n} \mathbf{P}\left\{A_{ni}\right\}}. \end{aligned}$$

To show the former inequality, we note that the multinomial distribution is negative orthant dependent (see Joag-Dev and Proschan, 1983, for definitions and a discussion). In particular, if $(X_1, \ldots, X_k)$ is a multinomial random vector, then we have for all $x_1, \ldots, x_k$,

$$\mathbf{P}\left\{\cap_{i=1}^{k} [X_i \geq x_i]\right\} \leq \prod_{i=1}^{k} \mathbf{P}\left\{X_i \geq x_i\right\}.$$

Now, assume that $i < j$. Then, with $T = \sum_{k=i+1}^{j} I_{[R_k=i]}$, $V = \sum_{k=j+1}^{n} I_{[R_k=i]}$, $W = \sum_{k=j+1}^{n} I_{[R_k=j]}$, and a_n integer-valued, we have

$$\begin{aligned} \mathbf{P}\left\{A_{ni}A_{nj}\right\} &= \mathbf{P}\left\{T+V \geq a_n, W \geq a_n\right\} \\ &\leq \mathbf{P}\left\{T+V \geq a_n\right\} \mathbf{P}\left\{W \geq a_n\right\} \\ &= \mathbf{P}\left\{A_{ni}\right\} \mathbf{P}\left\{A_{nj}\right\}, \end{aligned}$$

where we used the fact that T is independent of (V, W), and that $(V, W, n-j-V-W)$ is a multinomial random vector. This establishes the first chain of inequalities. The second half of Lemma 1 follows from the first part and the fact that $[M_n \geq a_n] \equiv \cup_{i=1}^{n} A_{ni}$. □

We first show the first statement of Theorem 1 with "almost surely" replaced by "in probability". Clearly, we need only verify the limits of the sums shown in Lemma 1, first for $a_n = \lceil c \log n \rceil$ with $c > 1/\log 2$, and later with $c < 1/\log 2$. In the second part, we obtain strong convergence. In the third part, we deal with $M_{n,k}$.

PROOF OF THE WEAK CONVERGENCE: UPPER BOUND. We will begin by showing

$$\lim_{n\to\infty} \sum_{i=1}^{n} \mathbf{P}\left\{D_i \geq a_n\right\} = 0,$$

where $a_n = \lceil c \log n \rceil$ with $c > 1/\log 2$. We first obtain the following exponential inequality:

$$\mathbf{P}\left\{D_i \geq a_n\right\} \leq e^{c \log n (1-u+\log(u))},$$

where $u = (H_{n-1} - H_i)/(c \log n)$. Indeed, for $t > 0$, by Markov's inequality,

$$\begin{aligned}
\mathbf{P}\left\{D_i \geq a_n\right\} &\leq \mathbf{E}\left\{e^{tD_i - ta_n}\right\} \\
&\leq \mathbf{E}\left\{e^{tD_i - tc\log n}\right\} \\
&\leq e^{-tc\log n} \prod_{j>i} \left(1 - \frac{1}{j-1} + \frac{e^t}{j-1}\right) \text{ (by independence)} \\
&\leq e^{-tc\log n + \sum_{j>i}(e^t-1)/(j-1)} \\
&= e^{-tc\log n + (e^t-1)(H_{n-1}-H_i)} \ (i \geq 2).
\end{aligned}$$

The last upper bound is minimal when t is the solution of $e^t = c\log n/(H_{n-1} - H_i)$. Resubstitution yields the bound. Since u is a decreasing function of i, we note that

$$\begin{aligned}
\sum_{i<n^\epsilon} \mathbf{P}\left\{D_i \geq a_n\right\} &\leq n^\epsilon \mathbf{P}\left\{D_1 \geq a_n\right\} \\
&\leq n^\epsilon e^{c\log n\,(1+o(1))\left(1-\frac{1}{c}+\log\left(\frac{1}{c}\right)\right)} \\
&= e^{-\log n\,(o(1)+1-c+c\log c-\epsilon)}
\end{aligned}$$

and this tends to zero for $\epsilon < 1 - c + c\log(c)$, so the inequality can be satisfied for all ϵ small enough. Next, we have

$$\begin{aligned}
\sum_{i>n^{1-\epsilon}} \mathbf{P}\left\{D_i \geq a_n\right\} &\leq n\mathbf{P}\left\{D_{\lceil n^{1-\epsilon}\rceil} \geq a_n\right\} \\
&\leq ne^{c\log n\,(1+o(1))(1-\epsilon/c+\log(\epsilon/c))} \\
&\leq e^{-\log n\,(1+o(1))(-c+\epsilon-c\log(\epsilon/c)-1)} \ ,
\end{aligned}$$

which tends to 0 provided that $\epsilon < c/e^2$—again, this is satisfied for all ϵ small enough. The first part of the theorem follows if we can show that for all $\epsilon > 0$ small enough we have

$$\lim_{n\to\infty} \sum_{n^\epsilon \leq i \leq n^{1-\epsilon}} \mathbf{P}\left\{D_i \geq a_n\right\} = 0.$$

If we set $i = n^\beta$, with $\beta \in [\epsilon, 1-\epsilon]$, then $u = (1-\beta)/c + o(1)$, where the $o(1)$ term does not depend upon β. Also, for $n^\epsilon \le y, z \le n^{1-\epsilon}$, $|y-z| \le 1$, we have $|\phi(u(y)) - \phi(u(z))| \le 1/\epsilon n^\epsilon$, where $\phi(s) \stackrel{\text{def}}{=} 1 - s + \log(s)$. Thus, by the uniformity of this estimate,

$$\sum_{n^\epsilon \le i \le n^{1-\epsilon}} \mathbf{P}\left\{D_i \ge a_n\right\} \le o(1) + (1+o(1)) \int_{n^\epsilon}^{n^{1-\epsilon}} e^{c \log n \left(1 - \frac{1}{c}\left(1 - \frac{\log x}{\log n}\right) + \log \frac{1}{c}\left(1 - \frac{\log x}{\log n}\right)\right)} \, dx.$$

By the transform $v = 1 - \log x / \log n$, we can rewrite the last integral as

$$\begin{aligned}
\int_\epsilon^{1-\epsilon} e^{c \log n (1 - v/c + \log(v/c))} n^{1-v} \log n \, dv &= n^{1+c-c\log c} \log n \int_\epsilon^{1-\epsilon} n^{c \log v - 2v} \, dv \\
&= \frac{n^{1+c-c\log c} \log n}{(2 \log n)^{1+c\log n}} \int_{2\epsilon \log n}^{2(1-\epsilon)\log n} w^{c \log n} e^{-w} dw \\
&\le \frac{n^{1+c-c\log c} \log n}{(2 \log n)^{1+c\log n}} \Gamma(1 + c \log n) \\
&\sim n^{1 - c \log 2} \sqrt{\pi c \log n / 2}\ ,
\end{aligned}$$

where we used the gamma integral and Stirling's formula. The last expression tends to zero with n.

Proof of the weak convergence: lower bound. Assume next that $c < 1/\log 2$. Define $b_n = b = \lfloor \delta \log n \rfloor$, where $\delta \in (0, 1/\log 2 - c)$ is an arbitrary small positive number. Furthermore, $\epsilon > 0$ is a fixed small number such that $2\epsilon < c + \delta < 2(1-\epsilon)$. We consider an integer i with $n^\epsilon \le i \le n^{1-\epsilon}$, and assume that n is so large that $b(i-1) > 4$ for all such i. Let $B_2, B_3, \ldots$ be independent Bernoulli random variables with $\mathbf{E} B_j = 1/(j-1)$, and let $P_2, P_3, \ldots,$ be independent Poisson random variables with $\mathbf{E} P_j = 1/(j-1)$. Set $a = a_n = \lceil c \log n \rceil$. In trying to derive a lower bound for $\mathbf{P}\{D_i \ge a\}$, we cannot use the standard bounds on the closeness of the distribution of a sum of Bernoulli random variables to a Poisson random variable (see Le Cam (1960); see Deheuvels and Pfeifer (1986) for the most recent references), as these bounds are too large. However, in a rich enough probability space, there exists an embedding such that

$$\begin{aligned}
D_i = \sum_{j > i} I_{[R_j = i]} &= \sum_{j>i} B_j \\
&\ge \sum_{j>i} P_j I_{[P_j \le 1]} \\
&= \sum_{j>i} P_j - \sum_{j>i} P_j I_{[P_j > 1]} \\
&\stackrel{\text{def}}{=} W_i - Z_i.
\end{aligned}$$

To see this, let U be uniform $[0,1]$, and define B_j and P_j both by the probability integral transform. Thus, $B_j = 0$ if and only if $U \leq 1 - \mu$. This implies that $U \leq e^{-\mu}$ and thus that $P_j = 0$. If $B_j = 1$, we trivially have $B_j \geq P_j I_{[P_j \leq 1]}$. Repeat this for all j. Note that W_i is Poisson with parameter $\mu \stackrel{\text{def}}{=} \sum_{j>i} 1/(j-1) = H_{n-1} - H_i$. From the trivial inequality

$$\mathbf{P}\left\{D_i \geq a\right\} \geq \mathbf{P}\left\{W_i \geq a+b\right\} - \mathbf{P}\left\{Z_i \geq b\right\} \ ,$$

we note that it suffices to show that

$$\lim_{n\to\infty} \sum_{n^\epsilon \leq i \leq n^{1-\epsilon}} \mathbf{P}\left\{W_i \geq a+b\right\} = \infty \ , \tag{1}$$

and

$$\lim_{n\to\infty} \sum_{n^\epsilon \leq i \leq n^{1-\epsilon}} \mathbf{P}\left\{Z_i \geq b\right\} = 0 \ , \tag{2}$$

to conclude by Lemma 1 that $\lim \mathbf{P}\{M_n \geq a_n\} = 1$.

VERIFICATION OF (1). If V_i is Poisson with parameter $\log((n-1)/(i-1))$, then uniformly over our i,

$$\begin{aligned}
\mathbf{P}\left\{W_i \geq a+b\right\} &\geq \mathbf{P}\left\{V_i \geq a+b\right\} \\
&\geq \frac{1}{(a+b)!}\log^{a+b}\left(\frac{n-1}{i-1}\right)\frac{i-1}{n-1} \\
&= \frac{\log^{a+b}(n-1)}{(a+b)!}\frac{i-1}{n-1}\left(1-\frac{\log(i-1)}{\log(n-1)}\right)^{a+b} \\
&= \frac{(1-o(1))\log^{a+b}(n)}{(a+b)!}\frac{i}{n}\left(1-\frac{\log(i)}{\log(n)}\right)^{a+b} .
\end{aligned}$$

Thus,

$$\begin{aligned}
\sum_{n^\epsilon \leq i \leq n^{1-\epsilon}} \mathbf{P}\left\{W_i \geq a+b\right\} &\geq \frac{(1-o(1))\log^{a+b}(n)}{(a+b)!}\int_{n^\epsilon}^{n^{1-\epsilon}} \frac{x}{n}\left(1-\frac{\log(x)}{\log(n)}\right)^{a+b} dx \\
&= \frac{(1-o(1))\log^{a+b}(n)}{(a+b)!}\int_{\epsilon\log n}^{(1-\epsilon)\log n} n\left(\frac{w}{\log(n)}\right)^{a+b} e^{-2w} dw \\
&\sim \frac{\log^{a+b}(n)}{(a+b)!} n \frac{(a+b)!}{2^{a+b+1}\log^{a+b}(n)} \\
&= \frac{n}{2^{a+b+1}}.
\end{aligned}$$

In this chain, we used the weak law of large numbers for the gamma distribution (the probability that a gamma $(a+b+1)$ random variable takes values in $[2\epsilon\log n, 2(1-\epsilon)\log n]$ tends to one when $2\epsilon < c+\delta < 2(1-\epsilon)$). It should be noted that the lower bound tends to ∞ when $c+\delta < 1/\log 2$.

VERIFICATION OF (2). We will use Chernoff's exponential bounding technique to bound $\mathbf{P}\{Z_i \geq b\}$. First, observe that

$$\mathbf{E}Z_i = \sum_{j>i} (\mathbf{E}P_j - \mathbf{P}\{P_j = 1\}) = \sum_{j>i} \frac{1}{j-1} \left(1 - e^{-\frac{1}{j-1}}\right) \leq \sum_{j>i} \left(\frac{1}{j-1}\right)^2 \leq \frac{1}{i-1}.$$

Since $b(i-1) > 4$, we see that

$$\mathbf{P}\{Z_i \geq b\} \leq \mathbf{P}\{Z_i - \mathbf{E}Z_i \geq b/2\} \leq e^{-\frac{tb}{2}} \mathbf{E}\left\{e^{t(Z_i - \mathbf{E}Z_i)}\right\}.$$

We will abbreviate the Poisson parameter $1/(j-1)$ of P_j to μ. The last expectation may be written as a product over $j > i$ of expected values of the form

$$\begin{aligned}
&\mathbf{E}\left\{e^{t(P_j - I_{[P_j=1]} - \mathbf{E}P_j + \mathbf{P}\{P_j=1\})}\right\} \\
&\leq \mathbf{E}\left\{e^{t(P_j - I_{[P_j=1]})}\right\} \\
&= e^{-\mu}(1+\mu) + \sum_{k=2}^{\infty} \frac{\mu^k}{k!} e^{-\mu+tk} \\
&= e^{-\mu+\mu e^t}(1 + \mu(1-e^t)e^{-\mu e^t}) \\
&\leq e^{-\mu+\mu e^t}(1 + \mu(1-e^t)(1-\mu e^t)) \\
&\leq e^{\mu^2 e^{2t}} \quad (\text{use } 1 - u \leq e^{-u}) .
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathbf{P}\{Z_i \geq b\} &\leq e^{-\frac{tb}{2} + \sum_{j>i}(j-1)^{-2}e^{2t}} \\
&\leq e^{-\frac{tb}{2} + \frac{e^{2t}}{i-1}} \\
&= \left(\frac{4e}{b(i-1)}\right)^{b/4} \quad (\text{take the optimal value for } t,\ t = (1/2)\log(b(i-1)/4) \\
&\leq \left(\frac{4e}{b(n^\epsilon - 1)}\right)^{b/4} \\
&\leq \left(\frac{1}{2(n^\epsilon - 1)}\right)^{b/4} \quad (\text{if } b \geq 8e) \\
&\leq n^{-\epsilon b/4} \quad (\text{if } n \geq 2) .
\end{aligned}$$

Thus, when $n \geq 2$ is so large that $b \geq 8e$,

$$\sum_{n^\epsilon \leq i \leq n^{1-\epsilon}} \mathbf{P}\{Z_i \geq b\} \leq n^{1-\epsilon\delta \log n/4} .$$

This is $o(1)$, as required. The weak convergence of $M_n/\log_2 n$ to 1 thus follows. This implies that $\liminf_{n\to\infty} \mathbf{E}M_n/\log_2 n \geq 1$, which together with Szymanski's result (or directly via tail bounds for M_n given above) shows that $\mathbf{E}M_n/\log_2 n \to 1$ as $n \to \infty$.

Proof of the strong convergence. The monotonicity of both M_n and $\log n$ are used in the inequalities

$$\inf_{N\leq n} \frac{M_{2^n}}{\log 2^{n+1}} \leq \inf_{2^N\leq n} \frac{M_n}{\log n} \leq \sup_{2^N\leq n} \frac{M_n}{\log n} \leq \sup_{N\leq n} \frac{M_{2^{n+1}}}{\log 2^n} \ .$$

To show that

$$\limsup_{n\to\infty} \frac{M_n}{\log n} \leq \frac{1}{\log 2} \text{ almost surely },$$

we need only show

$$\limsup_{n\to\infty} \frac{M_{2^n}}{n\log 2} \leq \frac{1}{\log 2} \text{ almost surely }.$$

From the tail bounds derived above, for any $c > 1/\log 2$, there exists an $\varepsilon > 0$ small enough such that for all n large enough,

$$\mathbf{P}\left\{\frac{M_{2^n}}{n\log 2} > c\right\} \leq 2e^{-\frac{\varepsilon}{2}(n-1)\log 2} + \frac{\sqrt{2\pi c n\log 2}}{2} e^{(1-c\log 2)n\log 2} \ .$$

As this is clearly summable in n, we may apply the Borel-Cantelli lemma to conclude that $\limsup_{n\to\infty} M_n/\log n \leq 1/\log 2$ almost surely. From exponential bounds derived earlier and the association inequality of Lemma 1, for any $c < 1/\log 2$, there exists an $\varepsilon > 0$ small enough such that for all n large enough,

$$\begin{aligned}
\mathbf{P}\left\{\frac{M_{2^n}}{n\log 2} < c\right\} &= 1 - \mathbf{P}\left\{\frac{M_{2^n}}{n\log 2} \geq c\right\} \\
&\leq 1 - \frac{\sum_{i=1}^{2^n} \mathbf{P}\left\{D_i \geq cn\log 2\right\}}{1 + \sum_{i=1}^{2^n} \mathbf{P}\left\{D_i \geq cn\log 2\right\}} \\
&\leq \frac{1}{2^{\varepsilon n}+1} \ .
\end{aligned}$$

This too is a summable series. By the Borel-Cantelli lemma,

$$\liminf_{n\to\infty} \frac{M_n}{\log n} \geq \frac{1}{\log 2} \text{ almost surely.}$$

Proof of the last part of Theorem 1. Here we use a simple result on Pólya urns that may be found in Athreya and Ney (1972). It has been used in the analysis of uniform random recursive trees in other contexts by Mahmoud and Smythe (1991). As we grow our URRT and follow $|T_k|$, the size of the URRT, we observe that $|T_k|$, for $n \geq k$, follows a Pólya urn process: at $n = k$, the urn contains $k-1$ black balls and one white ball, the white ball corresponding to node k. Each subsequent node picks a previous node uniformly at random. Therefore, we may model this by picking a ball from the urn at random, and throwing two balls back of the same colour. It is known that the proportion of white balls tends almost surely to a beta $(1, k-1)$ random variable when $k > 1$:

$$\lim_{n\to\infty} \frac{|T_k|}{n} = Y \text{ almost surely .}$$

where Y is beta $(1, k-1)$. Since $0 < Y < 1$ with probability one, we have

$$\lim_{n\to\infty} \frac{\log |T_k|}{\log n} = 1 \text{ almost surely .}$$

The next observation is that T_k grows as a URRT in its own right. Hence, on $[|T_k| \to \infty]$, we have

$$\frac{M_{n,k}}{\log |T_k|} \to \frac{1}{\log 2} \text{ almost surely ,}$$

and thus,

$$\lim_{n\to\infty} \frac{M_{n,k}}{\log n} = \frac{1}{\log 2} \text{ almost surely .}$$

3. Proof of Theorem 2.

For $i > j$, the indicator of the presence of a directed edge between "child" i and "parent" j is denoted by R_{ij}. We define R_j as the vector $(R_{j1}, \ldots, R_{j,j-1})$. Clearly, $\{R_j \, , \, j > m\}$ are independent.

Lemma 2. *Let $a > 0$ be fixed. Define the degree of the i-th node,*

$$D_{ni} = \sum_{j=i+1}^{n} R_{ji} \ ,$$

and the event $A_{ni} = [D_{ni} \geq a]$, where $a > 0$ is a given number. Then, for $k > i > j > m$, R_{ki} and R_{kj} are negatively orthant dependent, i.e., for any increasing functions f and g,

$$\mathbf{E}\{f(R_{ki})g(R_{kj})\} \leq \mathbf{E}\{f(R_{ki})\}\mathbf{E}\{g(R_{kj})\} \ .$$

Furthermore, D_{ni} and D_{nj} are negatively orthant dependent. In particular, for any $a > 0$,

$$\mathbf{P}\{A_{ni}A_{nj}\} \leq \mathbf{P}\{A_{ni}\}\mathbf{P}\{A_{nj}\} \ .$$

PROOF. Clearly,

$$\begin{aligned}
&\mathbf{E}\{f(R_{ki})g(R_{kj})\} \\
&\quad = f(1)g(1)\frac{r(r-1)}{k(k-1)} + (f(1)g(0)+f(0)g(1))\frac{r(k-r)}{k(k-1)} + f(0)g(0)\frac{(k-r)(k-r-1)}{k(k-1)} \\
&\quad \le \left(\frac{r}{k}f(1)+\frac{k-r}{k}f(0)\right)\left(\frac{r}{k}g(1)+\frac{k-r}{k}g(0)\right) \\
&\quad = \mathbf{E}\{f(R_{ki})\}\mathbf{E}\{g(R_{kj})\}\ ,
\end{aligned}$$

because $0 \le (k-r)r(f(1)-f(0))(g(1)-g(0))$. Also, the pairs (R_{ki}, R_{kj}) for $k > i$ are independent. Hence, a straightforward argument shows that $\sum_{k>i} R_{ki}$ and $\sum_{k>i} R_{kj}$ are negatively orthant dependent. By conditioning, the addition of an independent component $\sum_{k=j+1}^{i} R_{kj}$ does not change matters, and we note that $D_{ni} = \sum_{k>i} R_{ki}$ and $D_{nj} = \sum_{k>j} R_{kj}$ are negatively orthant dependent as well. □

We note at this point that Lemma 1 remains formally valid if we define $A_{ni} = [D_{ni} \ge a_n]$. The proof of Theorem 2 proceeds along the lines of the proof of Theorem 1. Strong convergence is obtained from tail estimates that were used to prove weak convergence by applying the Borel-Cantelli lemma to the subsequence $\lfloor (1+1/r)^n \rfloor$, where $n \ge 1$, and employing simple monotonicity arguments. Thus, we only consider upper bounds for

$$\sum_{i=1}^{n} \mathbf{P}\left\{D_{ni} \ge a_n\right\}\ ,$$

when $a_n = (1+\epsilon)\log_{1+1/r} n$, and lower bounds for the same expression with $a_n = (1-\epsilon)\log_{1+1/r} n$. The upper bound should tend to zero, and the lower bound should tend to ∞ in order to be able to deduce that

$$\frac{M_n}{\log_{1+1/r} n} \to 1 \text{ in probability .}$$

AN UPPER BOUND. We proceed as in the proof of Theorem 1. Take $a_n = \lceil c \log n \rceil$ with $c > 1/\log(1+1/r)$, and define $u = r(H_{n-1} - H_i)/(c \log n)$. Note that $D_{ni} = \sum_{j>i} R_{ji}$, where R_{ji} is Bernoulli $(r/(j-1))$. Thus, for $i \geq m$, we have, along the lines of the proof of Theorem 1, for $t > 0$,

$$\begin{aligned}
\mathbf{P}\left\{D_{ni} \geq a_n\right\} &\leq \mathbf{E}\left\{e^{tD_{ni} - ta_n}\right\} \text{ (by Markov's inequality)} \\
&\leq \mathbf{E}\left\{e^{tD_{ni} - tc\log n}\right\} \\
&\leq e^{-tc\log n} \prod_{j>i} \left(1 - \frac{r}{j-1} + \frac{re^t}{j-1}\right) \text{ (by independence)} \\
&\leq e^{-tc\log n + \sum_{j>i} r(e^t-1)/(j-1)} \\
&= e^{-tc\log n + r(e^t-1)(H_{n-1}-H_i)} \\
&= e^{-c\log n(t - u(e^t-1))} \text{ (definition of } u) \\
&= e^{c\log n(1-u+\log u)} \text{ (take } t = -\log u)
\end{aligned}$$

where we assume that n and i are such that $u \leq 1$ ($i > n^{1-c/r}$ will do; since $r\log(1+1/r) < 1$, this is satisfied for all i and all n large enough). We take $\epsilon > 0$ very small, and consider three ranges for our sum. Since u is a decreasing function of i, and since the value of u at $i = m$ is $r/c + O(1/\log n)$, we note that

$$\begin{aligned}
\sum_{m<i<n^\epsilon} \mathbf{P}\left\{D_{ni} \geq a_n\right\} &\leq n^\epsilon \mathbf{P}\left\{D_{nm} \geq a_n\right\} \\
&\leq n^\epsilon e^{c\log n(1-u+\log u)} \text{ (at } i = m) \\
&\leq n^\epsilon e^{c\log n(1-(r/c)+\log(r/c))+O(1)} \\
&\leq Cn^{\epsilon+c-r+c\log(r/c)} \text{ (for some constant } C) .
\end{aligned}$$

Note that the exponent of n is negative for all $\epsilon > 0$ small enough. Similarly,

$$\begin{aligned}
\sum_{i>n^{1-\epsilon}} \mathbf{P}\left\{D_{ni} \geq a_n\right\} &\leq n\mathbf{P}\left\{D_{n\lceil n^{1-\epsilon}\rceil} \geq a_n\right\} \\
&\leq ne^{c\log n(1-u+\log u)} \text{ (at } i = \lceil n^{1-\epsilon}\rceil) \\
&= ne^{O(1)+\log n(c-r\epsilon+c\log(r\epsilon/c))} \\
&\quad \text{(since the value of } u \text{ at } i = \lceil n^{1-\epsilon}\rceil \text{ is } r\epsilon/c + O(1/\log n)) \\
&\leq Cn^{1+c-r\epsilon+c\log(r\epsilon/c)} \text{ (for some constant } C) .
\end{aligned}$$

Once again, it is easy to verify that the exponent of n is negative: first of all, note that $c > 1/\log(1+1/r) > r$. Then, observe that the exponent is a unimodal function of c with a maximum at $c = r\epsilon$. Thus, if we replace c in the exponent by r, we necessarily obtain an upper bound, which is $1 + r - r\epsilon + r\log\epsilon < 0$ when $\epsilon < e^{-(1+1/r)}$. The sought upper bound follows if we can show that for all $\epsilon > 0$ small enough we have

$$\lim_{n\to\infty} \sum_{n^\epsilon \le i \le n^{1-\epsilon}} \mathbf{P}\left\{D_{ni} \ge a_n\right\} = 0.$$

If we set $i = n^\beta$ with $\beta \in [\epsilon, 1-\epsilon]$, then we note that $u = r(1-\beta)/c + o(1)$, where the $o(1)$ term does not depend upon β. Also, for $n^\epsilon \le y, z \le n^{1-\epsilon}$, $|y-z| \le 1$, we have $|\phi(u(y)) - \phi(u(z))| \le 1/\epsilon n^\epsilon$, where $\phi(s) \stackrel{\text{def}}{=} 1 - s + \log(s)$. Thus, by the uniformity of this estimate,

$$\sum_{n^\epsilon \le i \le n^{1-\epsilon}} \mathbf{P}\left\{D_{ni} \ge a_n\right\} \le o(1) + (1+o(1)) \int_{n^\epsilon}^{n^{1-\epsilon}} e^{c \log n \left(1 - \frac{r}{c}\left(1 - \frac{\log x}{\log n}\right) + \log \frac{r}{c}\left(1 - \frac{\log x}{\log n}\right)\right)}\, dx.$$

By the transform $v = 1 - \log x / \log n$, we can rewrite the last integral as

$$\begin{aligned}
&\int_\epsilon^{1-\epsilon} e^{c \log n (1 - vr/c + \log(vr/c))} n^{1-v} \log n\, dv \\
&\quad = n^{c\log r + 1 + c - c \log c} \log n \int_\epsilon^{1-\epsilon} v^{c \log n} e^{-(1+r) v \log n}\, dv \\
&\quad \le n^{c\log r + 1 + c - c \log c} \log n \int_0^{\infty} v^{c \log n} e^{-(1+r) v \log n}\, dv \\
&\quad = n^{c\log r + 1 + c - c \log c} \log n \times ((1+r)\log n)^{-c\log n - 1} \times \Gamma(c \log n + 1) \\
&\quad \sim \frac{n^{c \log(r/(1+r)) + 1} \sqrt{2\pi c \log n}}{1+r}
\end{aligned}$$

where we used the gamma integral and Stirling's formula. The last expression tends to zero with n at a polynomial rate.

A LOWER BOUND. We proceed once again as in the proof of Theorem 1. It suffices to note the few differences. Set $a = a_n = \lceil c \log n \rceil$ with $c < 1/\log(1 + 1/r)$, and set $b = \lfloor \delta \log n \rfloor$ where $\delta \in (0, 1/\log(1+1/r) - c)$. The independent Poisson random variables P_j have parameter $r/(j-1)$ instead of $1/(j-1)$. The definition of W_i and Z_i is as in Theorem 1. Using an argument as in Theorem 1, we may prove that

$$\sum_{n^\epsilon \le i \le n^{1-\epsilon}} \mathbf{P}\{W_i \ge a_n + b_n\} \ge n^\eta$$

for some $\eta > 0$ and all n large enough. By the Chernoff bounding method, it is also easy to show that

$$\mathbf{P}\{Z_i \ge b_n\} \le e^{-\frac{b_n}{4}\left(1 - \log \frac{b_n (i-1)}{4r^2}\right)}$$

for $i \ge n^\epsilon$ and n so large that $b_n > 4r^2/(i-1)$. Simple calculations then show that for some $\eta > 0$ and all n large enough,

$$\sum_{n^\epsilon \le i \le n^{1-\epsilon}} \mathbf{P}\{Z_i \ge b_n\} \le n^{-\eta}\ ,$$

as required.

Acknowledgment.

We thank the referees for their helpful reports.

References.

B. Bollobás, *Random Graphs*, Academic Press, New York, 1985.

B. Bollobás and G. Brightwell, "Box-spaces and random partial orders," *Transactions of the American Mathematical Society*, vol. 324, pp. 59–72, 1991.

B. Bollobás and P. M. Winkler, "The longest chain among random points in Euclidean space," *Proceedings of the American Mathematical Society*, vol. 103, pp. 347–353, 1988.

Y. S. Chow and H. Teicher, *Probability Theory*, Springer-Verlag, New York, N.Y., 1978.

K. L. Chung and P. Erdös, "On the application of the Borel-Cantelli lemma," *Transactions of the American Mathematical Society*, vol. 72, pp. 179–186, 1952.

D. A. Dawson and D. Sankoff, "An inequality for probabilities," *Proceedings of the American Mathematical Society*, vol. 66, pp. 504–507, 1967.

P. Deheuvels and D. Pfeifer, "A semigroup approach to Poisson approximation," *Annals of Probability*, vol. 14, pp. 663–676, 1986.

L. Devroye, "The expected length of the longest probe sequence when the distribution is not uniform," *Journal of Algorithms*, vol. 6, pp. 1–9, 1985.

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

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

M. Dondajewski and J. Szymański, "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, 1982.

A. Frieze, "On the length of the longest monotone subsequence in a random permutation," *Annals of Applied Probability*, vol. 1, pp. 301–305, 1991.

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

J. L. Gastwirth and P. K. Bhattacharya, "Two probability models of pyramid or chain letter schemes demonstrating that their promotional claims are unreliable," *Operations Research*, vol. 32, pp. 527–536, 1984.

G. H. Gonnet, "Expected length of the longest probe sequence in hash code searching," *Journal of the ACM*, vol. 28, pp. 289–304, 1981.

K. Joag-Dev and F. Proschan, "Negative association of random variables, with applications," *Annals of Statistics*, vol. 11, pp. 286–295, 1983.

D. E. Knuth and A. Schonhage, "The expected linearity of a simple equivalence algorithm," *Theoretical Computer Science*, vol. 6, pp. 281–315, 1978.

L. LeCam, "An approximation theorem for the Poisson binomial distribution," *Pacific Journal of Mathematics*, vol. 10, pp. 1181–1197, 1960.

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

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

A. Meir and J. W. Moon, "Path edge-covering constants for certain families of trees," *Utilitas Mathematica*, vol. 14, pp. 313–333, 1978.

J. W. Moon, "On the maximum degree in a random tree," *Michigan Mathematical Journal*, vol. 15, pp. 429–432, 1968.

J. W. Moon, "The distance between nodes in recursive trees," *London Mathematical Society Lecture Notes*, vol. 13, pp. 125–132, Cambridge University Press, London, 1974.

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

D. Najock and C. C. Heyde, "On the number of terminal vertices in certain random trees with an application to stemma construction in philology," *Journal of Applied Probability*, vol. 19, pp. 675–680, 1982.

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

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

J. Szymański, "On the maximum degree and the height of a random recursive tree," in: *Random Graphs 87*, ed. M. Karoński, J. Jaworski and A. Ruciński, pp. 313–324, John Wiley, Chichester, 1990.

P. M. Winkler, "Random orders," *Order*, vol. 1, pp. 317–335, 1985.

P. M. Winkler, "Connectedness and diameter for random orders of fixed dimension," *Order*, vol. 2, pp. 165–171, 1985.