ON THE MAXIMUM OUT-DEGREE IN RANDOM TREES

A. MEIR AND J.W. MOON

Department of Mathematics
University of Alberta
Edmonton, Alberta, Canada T6G 2G1

1. Introduction.

Let $\mathcal{F}$ denote a simply generated family of rooted trees whose generating function $y = \sum_1^\infty y_n x^n$ satisfies a relation of the form $y = x\Phi(y)$ where $\Phi(t) = 1 + \sum_1^\infty c_m t^m$. If some mild conditions are satisfied, then $y_n \sim c(\Phi(\tau)/\tau)^n \cdot n^{-3/2}$ where $\tau\Phi'(\tau) = \Phi(\tau)$. We say that a node v in a rooted tree T_n has *out-degree* k if v is incident with k edges that lead away from the root of T_n. Our object here is to investigate the behaviour of the maximum out-degree $\Delta = \Delta(T_n)$ of trees T_n in $\mathcal{F}$. After describing briefly in §2 the simply generated families of trees we shall be considering, we obtain bounds in §3 for $Pr\{\Delta < k\}$ in terms of the function $r_k(\tau) = \sum_k^\infty c_m \tau^m$. Then in §4 we obtain certain inequalities involving the functions $r_k(\tau)$, assuming henceforth that the coefficients are reasonably well-behaved. Our main result is in §5 where we show that if $D(n) = \max\{k : nr_k(\tau) \geq 1\}$ then $Pr\{(1-\epsilon)D(n) < \Delta(T_n) < (1+\epsilon)D(n)\} \to 1$ as $n \to \infty$. We consider the problem of estimating $D(n)$ in §6; we find that $D(n) \sim \log n/\log(R/\tau)$ if $\Phi(t)$ has a finite radius of convergence R while $D(n) = o(\log n)$ if $\Phi(t)$ is an entire function. Finally, in §7 we consider some particular families of trees. For example, if $\Phi(t) = (1-t)^{-1}$ and $\mathcal{F}$ is the family of plane trees then $D(n) = 1 + [\log_2 n]$; and if $\Phi(t) = e^t$ and $\mathcal{F}$ is the family of rooted labelled trees then $D(n) \sim \log n/\log\log n$. (We remark that the problem of determining the behaviour of $\Delta(T_n)$ for this last family was considered earlier in [6].)

Australasian Journal of Combinatorics 2(1990) pp 147-156

2. Simply generated families.

We recall that *plane* trees - or *ordered* trees, as they are called by some authors [2; p. 306] - are rooted trees with an ordering specified for the branches incident with each node. To each such tree T_n we assign a non-negative *weight* $\omega(T_n)$ satisfying the following condition: there exists a sequence of non-negative constants $c_0(=1), c_1, c_2, \ldots$ such that

$$\omega(T_n) = \prod_0^\infty c_m{}^{d_m(T_n)} \tag{2.1}$$

for every plane tree T_n, where $d_m(T_n)$ denotes the number of nodes of out-degree m in T_n. The collection of plane trees with such an assignment of weights will be called a *simply generated family*, henceforth denoted by $\mathcal{F}$. Let y_n denote the number of trees T_n in the family $\mathcal{F}$ where the weights are taken into account (both here and elsewhere); that is

$$y_n = \Sigma\, \omega(T_n) \tag{2.2}$$

where the sum is over all plane trees with n nodes. It is not difficult to see (*cf.* [3; p. 999] or [9; p. 24]) that if $\mathcal{F}$ is a simply generated family then its generating function $y = \sum_1^\infty y_n x^n$ satisfies the relation

$$y = x\Phi(y) \tag{2.3}$$

where $\Phi(t) = 1 + \sum_1^\infty c_m t^m$.

We shall assume henceforth that $\mathcal{F}$ is some given simply generated family such that the function $\Phi(t)$ appearing in (2.3) is regular when $|t| < R \leq \infty$. We further assume that

(2.4) $\quad c_m \geq 0 \quad \text{for} \quad m \geq 1,$

(2.5) $\quad \gcd\{m : c_m > 0\} = 1, \quad$ and

(2.6) $\quad \tau\Phi'(\tau) = \Phi(\tau) \quad \text{for some } \tau, \quad \text{where} \quad 0 < \tau < R.$

It follows from these assumptions (see [8; p. 216], [3; p. 999], or [9; p. 32]) that $y(x)$ is regular when $|x| \leq \rho$, $x \neq \rho$, where $\rho = \tau/\Phi(\tau)$; moreover, $y(\rho) = \tau$ and

$$y_n \sim c(\Phi(\tau)/\tau)^n \cdot n^{-3/2} \tag{2.7}$$

as $n \to \infty$, where $c = (\Phi(\tau)/2\pi\Phi''(\tau))^{1/2}$.

3. Bounds for $Pr\{\Delta(T_n) < k\}$.

Let $r_k(t) = \sum_k^{\infty} c_m t^m$ for $k = 0, 1, \ldots$. We now establish bounds for $Pr\{\Delta(T_n) < k\}$, where $\Delta(T_n)$ denotes the maximum out-degree of nodes in the tree T_n, assuming that any given tree T_n is selected from the trees in $\mathcal{F}$ with n nodes with probability $\omega(T_n)/y_n$.

LEMMA 1. *Let $A = 2\tau/c$ and $B = 1/\Phi(\tau)$. Then*

$$Pr\{\Delta(T_n) < k\} < An^{1/2} \cdot e^{-Bnr_k(\tau)} \tag{3.1}$$

for $k = 1, 2, \ldots$ and all sufficiently large values of n.

PROOF: Let $\bar{y}_k = \bar{y}_k(x) = \Sigma\ \bar{y}_{kn} x^n$ where $\bar{y}_{kn}$ denotes the number of trees T_n in $\mathcal{F}$ such that $\Delta(T_n) < k$. Then it is not difficult to see that $\bar{y}_k$ satisfies the relation $\bar{y}_k = x\Phi_k(\bar{y}_k)$ where $\Phi_k(t) = \Phi(t) - r_k(t)$. Thus it follows from Lagrange's inversion formula (see, *e.g.*, [1; 148]) that

$$\bar{y}_{kn} = n^{-1} \cdot C_{n-1}\left\{\left(\Phi_k(t)\right)^n\right\}$$

where $C_m\{f(t)\}$ denotes the coefficient of t^m in the power series expansion of $f(t)$. Therefore,

$$\begin{aligned}\bar{y}_{kn}\tau^{n-1} &\le n^{-1} \cdot \left(\Phi_k(\tau)\right)^n \\ &= n^{-1} \cdot \{\Phi(\tau) - r_k(\tau)\}^n \\ &\le n^{-1}\Phi^n(\tau) \cdot e^{-Bnr_k(\tau)}.\end{aligned}$$

This implies inequality (3.1) since $Pr\{\Delta(T_n) < k\} = \bar{y}_{kn}/y_n$ and $y_n \sim c(\Phi(\tau)/\tau)^n \cdot n^{-3/2}$.

Notice that if $\Phi(t)$ is a polynomial of exact degree h, then it follows from Lemma 1 that $Pr\{\Delta(T_n) = h\} \to 1$ as $n \to \infty$. So we will assume henceforth that $c_m > 0$ for infinitely many m, i.e., that $r_k(\tau) > 0$ for all k.

We shall need two known results in proving the next inequality. Firstly, let $e_k(n)$ denote the expected number of nodes of out-degree at least k in a tree T_n in $\mathcal{F}$, where the expectation is taken over all trees T_n in $\mathcal{F}$; then it follows from [4; Theorem 4] (see also [9; p. 26]) that

$$\sum_1^{\infty} e_k(n) y_n x^n = x^2 r_k(y(x)) y'(x)/y(x) \tag{3.2}$$

for $k = 0, 1, \ldots$. Secondly, it was shown in [5] (see also [7; p. 305]) that for each family $\mathcal{F}$ there exists a positive constant Q such that if $y = \tau e^{i\theta}$, where $|\theta| \leq \pi$, then

$$|\Phi(y)| \leq \Phi(\tau) e^{-Q\theta^2}. \tag{3.3}$$

LEMMA 2. *Let* $K = \rho(c^2 \pi Q)^{-1/2}$. *Then*

$$Pr\{\Delta(T_n) \geq k\} \leq K n r_k(\tau) \tag{3.4}$$

for $k = 0, 1, \ldots$ *and all sufficiently large values of* n.

PROOF: We first observe that $Pr\{\Delta(T_n) \geq k\} \leq e_k(n)$, by Boole's inequality. Hence it follows from (3.2) and Cauchy's theorem that

$$Pr\{\Delta(T_n) \geq k\} y_n \leq (2\pi i)^{-1} \int r_k(y(x)) x^{1-n} (y'(x)/y(x)) dx$$

where the integration is along a small circle around $x = 0$. Since $y'(0) = 1$ we may change variables by letting $x = y/\Phi(y)$; consequently

$$Pr\{\Delta(T_n) \geq k\} y_n \leq (2\pi i)^{-1} \int r_k(y) \Phi^{n-1}(y) \cdot y^{-n} dy$$

where now the integration can be taken around the circle $|y| = \tau$. But if $y = \tau e^{i\theta}$, where $|\theta| \leq \pi$, then $|r_k(y)| \leq r_k(\tau)$. Therefore, taking absolute values and applying (3.3), we find that

$$\begin{aligned} Pr\{\Delta(T_n) \geq k\} y_n &\leq (2\pi)^{-1} r_k(\tau) \cdot (\Phi(\tau)/\tau)^{n-1} \int_{-\pi}^{\pi} e^{-(n-1)Q\theta^2} d\theta \\ &\leq (4\pi(n-1)Q)^{-1/2} \cdot (\Phi(\tau)/\tau)^{n-1} \cdot r_k(\tau). \end{aligned}$$

This implies inequality (3.4) since $y_n \sim c(\Phi(\tau)/\tau)^n \cdot n^{-3/2}$.

4. Some inequalities for $r_k(\tau)$.

In the last section we obtained bounds for $Pr\{\Delta(T_n) < k\}$ in terms of the function $r_k(\tau) = \sum_k^\infty c_m \tau^m$. We need some information about the behaviour of $r_k = r_k(\tau)$ as $k \to \infty$ in order to exploit these bounds. In the following two lemmas we consider separately the cases when R, the radius of convergence of the function $\Phi(t)$, is finite or is infinite.

LEMMA 3. *Suppose that $R < \infty$ and that*

$$c_m{}^{1/m} \to R^{-1} \tag{4.1}$$

as $m \to \infty$. Then for every $\delta > 0$ there exists a number J such that if $j \geq J$ and $\ell \geq (1+2\delta)j$, then

$$r_\ell < r_j{}^{1+\delta}. \tag{4.2}$$

PROOF: Let ϵ denote a positive constant such that

$$\epsilon < \frac{1}{2}(R - \tau) \tag{4.3}$$

and

$$\{(R+\epsilon)/(R-\epsilon)\}^{1+\delta} \cdot \{\tau/(R-\epsilon)\}^{\delta} < 1. \tag{4.4}$$

It follows from assumption (4.1) that there exists an integer N such that if $m \geq N$, then

$$(R+\epsilon)^{-m} < c_m < (R-\epsilon)^{-m}.$$

Consequently, if $k \geq N$ then

$$\begin{aligned} (\tau/(R+\epsilon))^k < r_k &< \sum_k^\infty (\tau/(R-\epsilon))^m \\ &< U \cdot (\tau/(R-\epsilon))^k \end{aligned} \tag{4.5}$$

where $U = 2R/(R-\tau)$, using (4.3) at the last step.

Now let ℓ and j be such that $\ell \geq (1+2\delta)j$. Then, appealing to (4.5) twice, we find that

$$\begin{aligned} r_\ell &< U \cdot (\tau/(R-\epsilon))^\ell \\ &= U \cdot \left(\frac{R+\epsilon}{R-\epsilon}\right)^{(1+\delta)j} \cdot \left(\frac{\tau}{R-\epsilon}\right)^{\ell-(1+\delta)j} \cdot \left(\frac{\tau}{R+\epsilon}\right)^{(1+\delta)j} \\ &< U \cdot \left\{\left(\frac{R+\epsilon}{R-\epsilon}\right)^{1+\delta} \cdot \left(\frac{\tau}{R-\epsilon}\right)^{\delta}\right\}^j \cdot r_j{}^{1+\delta} \end{aligned}$$

if $j \geq N$. Condition (4.4) implies that the coefficient of $r_j{}^{1+\delta}$ in this last expression is less than one when $j \geq M$ for sufficiently large M. Hence $r_\ell < r_j{}^{1+\delta}$ when $j \geq J = \max\{M, N\}$, as required.

LEMMA 4. *Suppose that* $R = \infty$ *so that*

$$c_m{}^{1/m} \to 0 \tag{4.6}$$

as $m \to \infty$. *In addition, suppose that*

$$c_{k+1}{}^{1/(k+1)} \leq c_k{}^{1/k} \tag{4.7}$$

when $k \geq N$. *Then for every* $\delta > 0$ *there exists a number* L *such that if* $j \geq L$ *and* $\ell \geq (1+2\delta)j$, *then*

$$r_\ell < r_j{}^{1+\delta}. \tag{4.8}$$

PROOF: Choose L so that $L \geq N$, $c_j{}^{1/j}\tau < 1/2$, and $(c_j{}^{1/j}\tau)^\delta < 1/2$ for $j \geq L$. Then it follows from our assumptions that

$$\begin{aligned} c_{\ell+\nu}\tau^{\ell+\nu} &\leq (c_\ell \tau^\ell)^{(\ell+\nu)/\ell} \\ &= c_\ell\tau^\ell \cdot (c_\ell{}^{1/\ell}\tau)^\nu \leq c_\ell\tau^\ell(1/2)^\nu \end{aligned}$$

for $\nu = 0, 1, \ldots$. Hence $r_\ell \leq 2c_\ell\tau^\ell$.

On the other hand, it also follows from our assumptions that

$$\begin{aligned} c_\ell\tau^\ell &\leq (c_j{}^{1/j}\tau)^\ell = (c_j{}^{1/j}\tau)^{\ell-\delta j} \cdot (c_j{}^{1/j}\tau)^{\delta j} \\ &< (c_j{}^{1/j}\tau)^{(1+\delta)j} \cdot (1/2)^j = (c_j\tau^j)^{1+\delta} \cdot (1/2)^j < r_j{}^{1+\delta} \cdot (1/2)^j. \end{aligned}$$

Hence, $r_\ell \le 2c_\ell\tau^\ell < r_j{}^{1+\delta} \cdot (1/2)^{j-1} \le r_j{}^{1+\delta}$, as required.

We remark that it can be shown that conclusion (4.8) also holds if condition (4.7) is replaced by either of the following conditions: there exist positive numbers N and H such that $\max\{c_h^{1/h} : k \le h \le 2k\} \le H c_k{}^{1/k}$ for all $k \ge N$ or $c_m > 0$ for all m and c_{k+1}/c_k decreases to 0 as $k \to \infty$.

5. Main result.

Let $D(n) = \max\{k : nr_k(\tau) \ge 1\}$. We now show that the distribution of $\Delta(T_n)$ is concentrated around $D(n)$ if the coefficients c_m are reasonably well-behaved.

THEOREM 1. *Suppose the coefficients c_m satisfy the conditions of Lemma 3 or of Lemma 4. Then for every $\epsilon > 0$*

$$Pr\{(1-\epsilon)D(n) < \Delta(T_n) < (1+\epsilon)D(n)\} \to 1$$

as $n \to \infty$.

PROOF: We first show that

$$Pr\{\Delta(T_n) \ge (1+\epsilon)D(n)\} \to 0 \tag{5.1}$$

as $n \to \infty$. Let $h = D(n)$, $\ell = [(1+\epsilon)h]$, and $\delta = \epsilon/3$. Since we are assuming that $\varphi(t)$ is not a polynomial, it follows that $h \to \infty$ as $n \to \infty$. Thus we may suppose that n is large enough to ensure that $(1+2\delta)(h+1) < \ell$ and that, by Lemma 3 or Lemma 4, $r_\ell < (r_{h+1})^{1+\delta}$. But $r_{h+1} < n^{-1}$, by the definition of h, so $nr_\ell < n \cdot n^{-1-\delta} = n^{-\delta}$. Hence it follows from Lemma 2 that

$$\begin{aligned} Pr\{\Delta(T_n) \ge (1+\epsilon)h\} &\le Pr\{\Delta(T_n) \ge \ell\} \\ &\le Knr_\ell < Kn^{-\delta}, \end{aligned}$$

and this implies (5.1).

We now show that

$$Pr\{\Delta(T_n) \le (1-\epsilon)D(n)\} \to 0 \tag{5.2}$$

as $n \to \infty$. As before, we let $h = D(n)$ and $\delta = \epsilon/3$ but this time we let $j = [(1-\epsilon)h] + 1$. We may suppose that n is large enough to ensure that $(1+2\delta)j \le h$ and that, by Lemma 3 or Lemma 4, $r_h < {r_j}^{1+\delta}$. But $r_h \ge n^{-1}$, so

$$nr_j > n{r_h}^{1/(1+\delta)} \ge n \cdot n^{-1/(1+\delta)} = n^{\delta/(1+\delta)}.$$

Hence it follows from Lemma 1 that

$$\begin{aligned} Pr\{\Delta(T_n) \le (1-\epsilon)h\} &= Pr\{\Delta(T_n) < j\} \\ &< An^{1/2} \cdot e^{-Bnr_j} < An^{1/2} \cdot e^{-Bn^{\delta/(1+\delta)}}, \end{aligned}$$

and this implies (5.2) and completes the proof of the theorem.

6. The behaviour of $D(n)$.

We now consider the problem of estimating the function $D(n) = \max\{k : nr_k(\tau) \ge 1\}$. It turns out that the behaviour of $D(n)$ depends on whether R, the radius of convergence of $\varphi(t)$, is finite or infinite.

THEOREM 2. *Suppose that $R < \infty$ and that*

$$c_m{}^{1/m} \to R^{-1}$$

as $m \to \infty$. Then

$$D(n) \sim \log n / \log(R/\tau)$$

as $n \to \infty$.

PROOF: We saw in the proof of Lemma 3 that if ϵ is any sufficiently small positive constant, then there exists an integer N such that if $k \ge N$ then

$$(6.1) \qquad (\tau/(R+\epsilon))^k < r_k < U \cdot (\tau/(R-\epsilon))^k$$

where $U = 2R/(R-\tau)$. Now let $h = D(n)$; we may suppose that n is large enough to ensure that $h \ge N$. Then it follows from (6.1) and the definition of h that

$$U^{-1} \cdot ((R-\epsilon)/\tau)^h < {r_h}^{-1} \le n < {r_{h+1}}^{-1} < ((R+\epsilon)/\tau)^{h+1}.$$

Consequently

$$h \log((R-\epsilon)/\tau) - \log U < \log n < (h+1)\log((R+\epsilon)/\tau),$$

which implies the required result.

THEOREM 3. *Suppose that $R = \infty$ so that*

$$c_m{}^{1/m} \to 0$$

as $m \to \infty$. Then

$$D(n) = o(\log n)$$

as $n \to \infty$.

PROOF: For any small positive ϵ, let $\delta = \tau^{-1} \cdot e^{-1/\epsilon}$; we may suppose that $\epsilon < 1/\log 2$ so that $\delta\tau < 1/2$. There exists an integer N such that if $m \geq N$ then $c_m < \delta^m$; hence if $k \geq N$, then

$$r_k < \sum_k^{\infty} (\delta\tau)^m < 2(\delta\tau)^k.$$

Now let $h = D(n)$; we may suppose that n is large enough to ensure that $h \geq N$. Then, since $r_h \geq n^{-1}$, it follows that

$$n \geq r_h{}^{-1} > \frac{1}{2}(\delta\tau)^{-h} = \frac{1}{2} \cdot e^{h/\epsilon}.$$

Consequently, $h < \epsilon \log(2n)$ and this implies the required result.

7. Special cases.

The plane trees and the rooted labelled trees illustrate the contrast in the behaviour of $D(n)$ when $R < \infty$ and when $R = \infty$. For the plane trees $\Phi(t) = (1-t)^{-1}$, so $R = 1$, $\tau = 1/2$, and $r_k(\tau) = (1/2)^{k-1}$; consequently, $D(n) = 1 + [\log_2 n]$ in accordance with Theorem 2. For the rooted labelled trees $\Phi(t) = e^t$ so $R = \infty$ and $\tau = 1$; it is not difficult to see that for this case $1/k! < r_k(\tau) < (1 + k^{-1})/k!$ from which it follows that $D(n) \sim \log n / \log\log n$ (see also [6]).

In general, when $R = \infty$ the behaviour of $D(n)$ depends very much on the rate at which the coefficients c_m of $\Phi(t)$ approach zero. For example, let $g(x)$ denote an increasing function of x such that $g(x) \to \infty$ and $g(x+1) - g(x) \to \infty$ as $x \to \infty$. Consider the function $\Phi(t) = 1 + \sum_1^{\infty} c_m t^m$ where $c_m = e^{-g(m)}$ for $m \geq 3$, $c_2 = 1 - \sum_3^{\infty} (m-1)c_m$, and $c_1 = 0$. Then $R = \infty, \tau = 1$, and $r_k(\tau) \sim c_k \sim$

$e^{-g(k)}$ as $k \to \infty$; consequently, $D(n) \sim g^{-1}(\log n)$ for this family, where g^{-1} denotes the inverse of the function g. In particular, if $g(x) = e^x$ then $D(n) \sim \log\log n$, if $g(x) = e^{e^x}$ then $D(n) \sim \log\log\log n$, and so on. And if, for example, $g(x) = x \log x$ then $D(n) \sim \log n / \log\log n$, and if $g(x) = x \log\log x$, then $D(n) \sim \log n / \log\log\log n$, and so on. Consequently, $D(n)$ can approach infinity arbitrarily slowly and Theorem 3 is, in a sense, best possible.

We remark in closing that it is not difficult to show that the expected value of $\Delta(T_n)$ is asymptotically equal to $D(n)$ as $n \to \infty$, assuming that $\Phi(t)$ satisfies the hypothesis of Lemma 3 or of Lemma 4.

The preparation of this paper was assisted by grants from the Natural Sciences and Engineering Research Council of Canada.

References

[1] L. Comtet, *Advanced Combinatorics*, Reidel, Dordrect, 1974.

[2] D.E. Knuth, *The Art of Computer Programming*, Vol. I, Addison-Wesley, Reading, 1973.

[3] A. Meir and J.W. Moon, On the altitude of nodes in random trees, *Can. J. Math.* **30** (1978) 997–1015.

[4] A. Meir and J.W. Moon, On nodes of out-degree one in random trees, *Colloq. Math. Soc. J. Bolyai* **52** (1987) 405–416.

[5] A. Meir and J.W. Moon, The asymptotic behaviour of coefficients of powers of certain generating functions, submitted for publication.

[6] J.W. Moon, On the maximum degree in a random tree, *Mich. Math. J.* **15** (1968) 429–432.

[7] A.M. Odlyzko and L.B. Richmond, On the unimodality of high convolutions of discrete distributions, *Ann. Probabl.* **13** (1985) 299–306.

[8] R. Otter, The multiplicative process, *Ann. Math. Satist.* 20 (1949) 206–224.

[9] J.-M. Steyaert and P. Flajolet, Patterns and pattern-matchings in trees: an analysis, *Inf. and Control* **58** (1983) 19–58.