J. Austral. Math. Soc. (*Series A*) 31 (1981), 429–438

TERMINAL PATH NUMBERS FOR CERTAIN FAMILIES OF TREES

A. MEIR and J. W. MOON

(Received 20 June 1980)

Communicated by W. D. Wallis

Abstract

We determine the limiting distribution of the distance from the root of a tree to any nearest endnode of the tree (other than the root) for certain families of rooted trees.

1980 *Mathematics subject classification* (*Amer. Math. Soc.*): 05 C 05.

1. Introduction

The *terminal path number* $\delta(T_n)$ of a rooted tree T_n with n nodes is defined as follows (for definitions not given here see Harary and Palmer (1973) or Moon (1970)): if $n = 1$ then $\delta(T_n) = 0$; otherwise, $\delta(T_n)$ is the distance from the root of T_n to any nearest endnode of T_n (other than the root if it is an endnode). If $\mathfrak{F}$ denotes some family of rooted trees, let $p(n, k)$ denote the probability that $\delta(T_n) > k$ where the probability is taken over the trees T_n in $\mathfrak{F}$. Grimmett (1980) has derived a recurrence relation for the numbers

$$p_k = \lim_{n\to\infty} p(n, k)$$

when $\mathfrak{F}$ is the family of rooted labelled trees. His argument is based on a probabilistic representation of these trees in terms of a branching process satisfying certain conditions. Our object here is to consider this problem for a fairly general class of families of rooted trees by a direct combinatorial argument. Our main results are given in Section 3 after some preliminaries in Section 2; some numerical results are given in Section 4 for labelled trees, plane trees, and binary trees.

https://doi.org/10.1017/S1446788700024228 Published online by Cambridge University Press

2. Preliminaries

Let y_n denote the number of trees T_n in a given family $\mathcal{F}$ of rooted trees; if there are weights associated with trees in $\mathcal{F}$, then each tree is counted according to its weight in these definitions. We shall assume that $\mathcal{F}$ is a *simply generated family*, that is, that the generating function

$$y = y(x) = \sum_1^\infty y_n x^n$$

satisfies a relation of the type

$$y = x\varphi(y), \tag{1}$$

where

$$\varphi(y) = 1 + c_1 y + c_2 y^2 + \cdots$$

is a power series in y with non-negative coefficients. This implies that the trivial tree T_1 is in $\mathcal{F}$ and that any non-trivial tree T_n in $\mathcal{F}$ can be constructed by joining the roots of an ordered collection of smaller trees in $\mathcal{F}$-called the *branches* of T_n-to a new node which serves as the root of T_n. The factor x in equation (1) takes the new root node into account, and the coefficients c_i determine weights associated with the trees in $\mathcal{F}$. For further elaboration of this last point see Meir and Moon (1978); see also Otter (1949).

If $0 \leqslant k \leqslant n - 1$ let y_{nk} denote the number of trees T_n in $\mathcal{F}$ such that $\delta(T_n) \geqslant k$. We now give a recurrence relation for the generating functions

$$G_k(x) = \sum_{n=k+1}^\infty y_{nk} x^n.$$

THEOREM 1. *If* $y = x\varphi(y)$, *then* $G_0(x) = y(x)$ *and*

$$G_{k+1}(x) = x\varphi(G_k(x)) - x \tag{2}$$

for $k = 0, 1, \ldots$.

PROOF. It is easy to see that $\delta(T_n) \geqslant k + 1$ for a non-trivial rooted tree T_n if and only if $\delta(B) \geqslant k$ for each branch B of T_n. Relation (2) follows immediately from this observation and assumption (1); the term $-x$ excludes the trivial tree T_1 that has no branches.

We shall use the following results in the next section.

LEMMA 1. *Suppose*

$$\varphi(t) = 1 + c_1 t + c_2 t^2 + \cdots$$

is a regular function of t *when* $|t| < R \leqslant +\infty$ *and let*

$$y = y(x) = x + y_2 x^2 + y_3 x^3 + \cdots$$

denote the solution of $y(x) = x\varphi(y(x))$ *in a neighbourhood of* $x = 0$. *If*

(i) $c_1 > 0$ *and* $c_j > 0$ *for some* $j \geqslant 2$,

(ii) $c_i \geqslant 0$ *for* $i \geqslant 2$, *and*

(iii) $\tau\varphi'(\tau) = \varphi(\tau)$ *for some* τ, *where* $0 < \tau < R$, *then*

$$y_n \sim c\rho^{-n} n^{-3/2} \tag{3}$$

as $n \to \infty$, *where* $\rho = \tau/\varphi(\tau)$ *and* $c = \{\varphi(\tau)/(2\pi\varphi''(\tau))\}^{1/2}$. *Furthermore, if* $x\varphi'(y(x)) = \sum_1^\infty d_n x^n$, *then*

$$d_n \sim c\varphi''(\tau)\rho^{-n+1} n^{-3/2} \tag{4}$$

as $n \to \infty$.

LEMMA 2. *Let* $A(x) = \sum_0^\infty a_n x^n$, $B(x) = \sum_0^\infty b_n x^n$, *and* $A(x)B(x) = \sum_0^\infty c_n x^n$, *and suppose there exist positive constants* a *and* ρ *such that*

$$a_n \sim a\rho^{-n} n^{-1/2} \quad \text{and} \quad b_n = O(\rho^{-n} n^{-3/2})$$

as $n \to \infty$. *If* $B(\rho) \neq 0$, *then*

$$c_n \sim B(\rho) a_n$$

as $n \to \infty$.

Relation (3) was proved in Meir and Moon (1978) and the proof of relation (4) is essentially the same except that Darboux's theorem is applied to expansion (3.5) in that paper instead of to expansion (3.3); we remark that a result closely related to relation (3) was proved earlier in Otter (1949). Lemma 2 was proved in Meir and Moon (1977).

3. Main results

We now determine the limiting behaviour of the probability $p(n, k) = y_{nk}/y_n$ that $\delta(T_n) \geqslant k$ for a tree T_n chosen from the simply generated family $\mathfrak{F}$. We assume throughout this section that the function φ that appears in relation (1) satisfies the hypothesis of Lemma 1.

https://doi.org/10.1017/S1446788700024228 Published online by Cambridge University Press

THEOREM 2. *If k is any fixed non-negative integer, then*

$$p_k = \lim_{n\to\infty} p(n, k) = \gamma_1\gamma_2 \cdots \gamma_k,$$

where

$$\gamma_j = \rho\varphi'(G_{j-1}(\rho))$$

for $j \geq 1$ *and an empty product is interpreted as one.*

PROOF. In what follows we let $\mathcal{C}_n\{f(x)\}$ denote the coefficient of x^n in any power series $f(x)$.

If we differentiate both sides of equation (2) with respect to x, multiply throughout by x, and then simplify slightly, we find that

$$xG'_{k+1}(x) = x\varphi'(G_k(x)) \cdot xG'_k(x) + G_{k+1}(x) \tag{5}$$

for $k \geq 0$. We may assume as our induction hypothesis that

$$\mathcal{C}_n\{xG'_k(x)\} = ny_{nk} = np(n, k)y_n \sim cp_k\rho^{-n}n^{-1/2}$$

as $n \to \infty$, in view of relation (3). Furthermore, it follows from relation (4) that

$$\mathcal{C}_n\{x\varphi'(G_k(x))\} \leq \mathcal{C}_n\{x\varphi'(y)\} = O(\rho^{-n}n^{-3/2}).$$

Therefore, since

$$\mathcal{C}_n\{G_{k+1}(x)\} \leq y_n = O(\rho^{-n}n^{-3/2}),$$

it follows from (5) and Lemma 2 that

$$np(n, k+1)y_n = \mathcal{C}_n\{xG'_{k+1}(x)\} \sim \rho\varphi'(G_k(\rho)) \cdot np(n, k)y_n,$$

or that

$$p(n, k+1) \to \gamma_{k+1}p_k = p_{k+1}$$

as $n \to \infty$. This suffices to complete the proof of the theorem.

COROLLARY 2.1. *There exists a constant A that depends on* $\mathcal{F}$ *such that*

$$p_k(\rho c_1)^{-k} \to A$$

as $k \to \infty$.

PROOF. Let $\beta_k = G_k(\rho)$ for $k = 0, 1, \ldots$. Then $\beta_0 = y(\rho) = \tau$ by Lemma 1, and

$$\beta_{k+1} = g(\beta_k),$$

where

$$g(t) = \rho(\varphi(t) - 1)$$

in view of Theorem 1. If $0 < t \leq \tau$ then

$$g(t) = \rho(c_1t + c_2t^2 + \cdots) = t\rho(c_1 + c_2t + \cdots) < t\rho\varphi'(\tau) = t$$

https://doi.org/10.1017/S1446788700024228 Published online by Cambridge University Press

since $c_i > 0$ for some $i \geqslant 2$ and since $\rho\varphi'(\tau) = 1$ by Lemma 1. It follows, therefore, that the sequence $\beta_0, \beta_1, \ldots$ decreases monotonically to a limit which must necessarily be zero. This implies, since $0 < \rho c_1 < 1$, that

$$\beta_k(\rho c_1)^{-k} \to a$$

for some constant a as $k \to \infty$; see DeBruijn (1970), Section 8.3.

Now

$$\gamma_j = \rho\varphi'(\beta_{j-1}) = \rho c_1(1 + O(\beta_{j-1})) = \rho c_1(1 + O((c_1\rho)^{j-1}))$$

since $\varphi'(t)$ is regular when $|t| < R$ and $\beta_{j-1} \leqslant \tau < R$ when $j \geqslant 1$. Consequently,

$$p_k(\rho c_1)^{-k} = \prod_1^k (\gamma_j/\rho c_1) = \prod_1^k (1 + O((c_1\rho)^{j-1})) \to A$$

for some constant A as $k \to \infty$, since $\Sigma(c_1\rho)^{j-1}$ converges. This proves the required result.

Let $\mu(n)$ denote the expected value of $\delta(T_n)$ over all trees T_n in $\mathfrak{F}$. Before determining the limiting value of $\mu(n)$ we need to introduce some more terminology and results.

If the tree T_n is rooted at node r suppose we select an edge rs incident with node r; we next select an edge su incident with node s where $u \neq r$, and so on. Let $c(T_n)$ denote the number of edges selected before reaching an endnode of T_n (other than r if r is an endnode) where the process terminates. We assume that at each step the next edge chosen is chosen at random from the admissible edges; we define $c(T_n)$ to be zero when $n = 1$. If $0 \leqslant k \leqslant n - 1$, let $q(n, k)$ denote the probability that $c(T_n) \geqslant k$ where the probability is taken over all trees T_n in $\mathfrak{F}$. It follows from results in Meir and Moon (1975) that

(6) $$q_k = \lim_{n\to\infty} q(n, k) = (1 - \rho/\tau)^{k-1}(1 + (k - 1)\rho/\tau)$$

for each non-negative integer k, and that

(7) $$\lim_{n\to\infty} \sum_{k=1}^{n-1} q(n, k) = \sum_{k=1}^{\infty} q_k = 2\tau/\rho - 1.$$

COROLLARY 2.2. *$\mu(n) \to p_1 + p_2 + \cdots$ as $n \to \infty$.*

PROOF. If $\delta(T_n) \geqslant k$, then it must be that $c(T_n) \geqslant k$. It follows readily from this observation that

(8) $$p(n, k) \leqslant q(n, k)$$

for all n and k.

Let the functions $P_n(t)$, $P(t)$, $Q_n(t)$ and $Q(t)$, where $n = 1, 2, \ldots$ and t is non-negative real variable, be defined by the relations $P_n(t) = p(n, [t+1])$, $P(t) = p_{[t+1]}$, $Q_n(t) = q(n, [t+1])$, and $Q(t) = q_{[t+1]}$. In particular,

$$P_n(t) = Q_n(t) = 0$$

if $t \geqslant n - 1$. It follows from Corollary 2.1 and (6) that

$$P_n(t) \to P(t) \qquad \text{and} \qquad Q_n(t) \to Q(t)$$

for each fixed value of t as $n \to \infty$. Furthermore,

$$P_n(t) \leqslant Q_n(t)$$

for all n and t, by (8); and

$$\int Q_n(t) \to \int Q(t)$$

as $n \to \infty$, by (7), where the integrals are over the interval $[0, \infty)$.

We may apply Fatou's Lemma (see Royden (1968), p. 226) to the functions $Q_n(t) - P_n(t)$ and conclude that

$$\begin{aligned}\int (Q(t) - P(t)) &\leqslant \liminf_{n\to\infty} \int (Q_n(t) - P_n(t)) \\ &\leqslant \limsup_{n\to\infty} \int Q_n(t) - \limsup_{n\to\infty} \int P_n(t) \\ &= \int Q(t) - \limsup_{n\to\infty} \int P_n(t),\end{aligned}$$

whence,

$$\limsup_{n\to\infty} \int P_n(t) \leqslant \int P(t).$$

But if we apply Fatou's Lemma to the functions $P_n(t)$ we find that

$$\liminf_{n\to\infty} \int P_n(t) \geqslant \int P(t).$$

It follows, therefore, that

$$\mu(n) = \int P_n(t) \to \int P(t) = p_1 + p_2 + \cdots$$

as $n \to \infty$, as required.

4. Results for particular cases

We now consider the limiting values of $p(n, k)$, for fixed k, for the following families $\mathfrak{F}$ of simply generated trees: labelled trees for which $c_i = 1/i!$ for all i, plane trees for which $c_i = 1$ for all i, and binary trees for which $c_2 = 1$ and $c_i = 0$ otherwise (see Meir and Moon (1978)).

THEOREM 3. *If $\mathcal{F}$ denotes the family of rooted labelled trees, then*

$$p_k = e^{-k+\beta_0+\beta_1+\cdots+\beta_{k-1}},$$

where $\beta_0 = 1$ and

$$\beta_{k+1} = e^{\beta_k - 1} - e^{-1}$$

for $k = 0, 1, \ldots$.

PROOF. In this case

$$y = xe^y = \sum_1^\infty n^{n-1}\frac{x^n}{n!}$$

and $\varphi(t) = e^t$, so $\tau = 1$ and $\rho = e^{-1}$. The recurrence relation for the numbers $\beta_k = G_k(e^{-1})$ follows from Theorem 1; and the formula for p_k follows from Theorem 2 since

$$\gamma_{k+1} = \rho\varphi'(G_k(\rho)) = e^{\beta_k - 1}.$$

If we let $\alpha_k = \beta_k - 1$ then $\alpha_0 = 0$,

$$\alpha_{k+1} = \beta_{k+1} - 1 = e^{\beta_k} - e^{-1} - 1 = e^{\alpha_k - 1} - e^{-1} - 1,$$

and

$$p_k = e^{\alpha_0+\alpha_1+\cdots+\alpha_{k-1}}.$$

This expression for p_k was derived by Grimmett (1980) by a different argument.

Since $p_k = \gamma_1\gamma_2\cdots\gamma_k$ it is perhaps more convenient to work directly with the numbers

$$\gamma_{k+1} = \rho\varphi'(G_k(\rho)) = e^{\beta_k - 1} = \beta_{k+1} + e^{-1}.$$

Then $\gamma_1 = 1$ and

$$\gamma_{k+1} = \exp(\beta_k - 1) = \exp(\gamma_k - e^{-1} - 1) = \sigma e^{\gamma_k},$$

where $\sigma = \exp(-e^{-1} - 1)$ for $k \geqslant 1$. The numerical values of some of these numbers, truncated after the last digit displayed, are given in Table 1.

TABLE 1.
Data for Labelled Trees

k	β_k	$\beta_k e^k$	γ_k	p_k	$p_k e^k$	$\Sigma_1^k p_i$
1	.63212	1.71828	1.00000	1.00000	2.71828	1.00000
2	.32432	2.39642	.69220	.69220	5.11470	1.69220
3	.14093	2.83068	.50881	.35219	7.07411	2.04439
4	.05567	3.03986	.42355	.14917	8.14474	2.19357
5	.02106	3.12607	.38894	.05802	8.61108	2.25159
10	.00014	3.17854	.36802	.00040	8.90216	2.28596
15	9×10^{-7}	3.17890	.36788	2×10^{-6}	8.90418	2.28619
20	6×10^{-9}	3.17890	.36787	1×10^{-8}	8.90419	2.28619

https://doi.org/10.1017/S1446788700024228 Published online by Cambridge University Press

THEOREM 4. *If $\mathcal{F}$ denotes the family of plane trees, then*

$$p_k = 9 \cdot 4^{-k}(1 + 2 \cdot 4^{-k})^{-2}$$

for $k = 0, 1, \ldots$.

PROOF. In this case

$$y = x(1-y)^{-1} = \sum_{1}^{\infty} \binom{2n-2}{n-1} \frac{x^n}{n}$$

and $\varphi(t) = (1-t)^{-1}$, so $\tau = \frac{1}{2}$ and $\rho = \frac{1}{4}$. If $\beta_k = G_k(\frac{1}{4})$, then $\beta_0 = \frac{1}{2}$ and

$$\beta_k = \tfrac{1}{4}\beta_k(1-\beta_k)^{-1}$$

for $k = 0, 1, \ldots$, by Theorem 1, from which it follows readily by induction that

$$\beta_k = \tfrac{3}{2}4^{-k}(1 + 2 \cdot 4^{-k})^{-1}.$$

Now

$$\gamma_{k+1} = \rho\varphi'(G_k(\rho)) = \tfrac{1}{4}(1-\beta_k)^{-2} = (2\beta_{k+1}/\beta_k)^2,$$

whence

$$p_k = \gamma_1\gamma_2 \cdots \gamma_k = 4^{k+1}\beta_k^2 = 9 \cdot 4^{-k}(1 + 2 \cdot 4^{-k})^{-2}$$

as required.

It follows from Corollary 2.2 and Theorem 4 that

$$\mu(n) \to \sum_{1}^{\infty} p_k = 1.62297\ldots,$$

as $n \to \infty$, for the family of plane trees.

We adopt the notational convention that $(t)_0 = 1$ and $(t)_j = t(t-1) \cdots (t-j+1)$ for $j = 1, 2, \ldots$.

THEOREM 5. *If $\mathcal{F}$ denotes the family of binary trees, then*

$$p(2n+1, k) = 2^k(n)_{2^k-1}/(2n)_{2^k-1}.$$

PROOF. In this case $y = x(1 + y^2)$, whence

$$G_k(x) = x^{-1}(xy)^{2^k}$$

by Theorem 1. Now

$$y^m = \sum_{n=0}^{\infty} \frac{m}{n+m} \binom{2n+m-1}{n} x^{2n+m}$$

for $m = 1, 2, \ldots,$ by Lagrange's inversion formula. We find, therefore, that

$$G_k(x) = \sum_{n=2^k-1}^{\infty} \frac{2^k}{n+1}\binom{2n+1-2^k}{n} x^{2n+1}$$

for $k = 0, 1, \ldots$. Consequently,

$$p(2n+1, k) = y_{2n+1,k}/y_{2n+1} = 2^k\binom{2n+1-2^k}{n} \Big/ \binom{2n}{n}$$

and this reduces to the expression given above.

It follows readily from Theorem 5 that

$$\lim_{n\to\infty} p(2n+1, k) = 2^{k+1-2^k}$$

for each fixed integer k; hence

$$\lim_{n\to\infty} \mu(2n+1) = \sum_{1}^{\infty} 2^{k+1-2^k} = 1.56298\ldots,$$

by Tannery's theorem (see Bromwich (1931), p. 136).

5. Acknowledgements

We are indebted to Mr. W. Aiello for performing some numerical calculations for us. The preparation of this paper was assisted by grants from the Natural Sciences and Engineering Research Council of Canada.

References

T. Bromwich (1931), *An introduction to the theory of infinite series* (Macmillan, London).

N. G. De Bruijn (1970), *Asymptotic methods in analysis* (North-Holland, Amsterdam).

G. R. Grimmett (1980), 'Random labelled trees and their branching networks', *J. Austral. Math. Soc. Ser. A* **30**, 229–237.

F. Harary and E. Palmer (1973), *Graphical enumeration* (Academic Press, New York).

A. Meir and J. W. Moon (1973), 'Climbing certain types of rooted trees, I.', *Proc. Fifth British Comb. Conf.*, University of Aberdeen, edited by C. St. J. A. Nash-Williams and J. Sheehan, pp. 461–469. (Utilitas Mathematica Publishing Inc., Winnipeg).

A. Meir and J. W. Moon (1977), 'Packing and covering constants for certain families of trees, I', *J. Graph Theory* **1**, 157–174.

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

J. W. Moon (1970), *Counting labelled trees* (Canadian Mathematical Congress, Montreal).

R. Otter (1949), 'The multiplicative process', *Ann. Math. Statist.* **20**, 206–224.
H. L. Royden (1968), *Real analysis* (Macmillan, New York).

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