On the Steiner distance of trees from certain families

L.H. Clark
Southern Illinois University at Carbondale
Carbondale, Illinois 62901-4408 U.S.A

A. Meir
York University
N. York, Ontario M3J 1P3 Canada

J.W. Moon
University of Alberta
Edmonton, Alberta T6G 2G1 Canada

Abstract

The Steiner distance $d_G(S)$ of a subset S of nodes of a connected graph G is the minimum number of edges in a connected subgraph of G that contains S. We consider the behaviour of the expected value $\mu_k(n)$ of $d_T(S)$ and the probability $p_k(n,m)$ that $d_T(S) = m-1$ over all subsets S of k nodes of trees T of order n from certain families of trees.

1. Introduction

The *Steiner distance* $d_G(S)$ of a subset S of nodes of a connected graph G is the minimum number of edges in a connected subgraph of G that contains S. For $k \geq 2$, the *total Steiner k-distance* $D_k(G)$ of a connected graph G is the sum of $d_G(S)$ over all subsets S of k nodes of G. The total Steiner 2-distance of a graph, also known as the Wiener Index or the total distance of a graph, has been rather extensively studied (see, e.g., the references in [2] or [3]). The *average Steiner k-distance* $\mu_k(G)$ of a connected graph G of order n is given by $\mu_k(G) = D_k(G)/\binom{n}{k}$.

Dankelmann, Oellerman, and Swart [2] showed that if G is a connected graph of order n and $2 \leq k \leq n$, then

$$k-1 \leq \mu_k(G) \leq (n+1)(k-1)/(k+1)$$

with equality on the left if and only if G is $(n-k+1)$-connected or $n=k$ and equality on the right if and only if G is a path or $n=k$. They also showed that if T is a tree of order n and $2 \leq k \leq n$, then

$$k(n-1)/n \leq \mu_k(T) \leq (n+1)(k-1)/(k+1)$$

with equality on the left if and only if T is a star or $n = k$ and equality on the right if and only if T is a path or $n = k$.

Our main objects here are to consider (a) the expected value $\mu_k(n)$ of $d_T(S)$ and (b) the probability $p_k(n,m)$ that $d_T(S) = m - 1$ over all subsets S of k nodes of trees T of order n in certain families of trees. In §2 we describe the families we shall be considering, the simply generated families, and we state some technical results we shall need later. In §3 we derive a formula for $\mu_k(n)$ that we use in §4 to determine the asymptotic behaviour of $\mu_k(n)$ for $2 \le k \le n$.

Then in §5 we derive, in effect, an expression for the generating function for the probabilities $p_k(n,m)$. We use this expression in §6 to determine the limiting behaviour of $p_k(n,m)$ for fixed k as $n \to \infty$, assuming that $m = O(n^{2/3})$; we also use it to determine the behaviour of the second moment of $d_T(S)$ for fixed k. We illustrate some of these results for three particular families in §7; we conclude with an example that shows that the behaviour of $\mu_k(n)$ for non-simply generated families can be quite different from its behaviour for simply generated families.

2. Preliminaries

We recall that *ordered* trees are (finite) rooted trees with an ordering specified for the branches incident with each node as one proceeds away from the root (see [6; p. 306]). Given a sequence $\Gamma = \{\varphi_0, \varphi_1, \ldots\}$ of non-negative numbers, with $\varphi_0 = 1$, we define $\mathcal{F} = \mathcal{F}_\Gamma$ to be the set of weighted ordered trees such that each ordered tree T is assigned the *weight*

$$\omega(T) = \prod_i \varphi_i^{N_i(T)},$$

where $N_i(T)$ denotes the number of nodes of T of *out-degree* i (i.e. incident with i edges leading away from the root). We call such a family a *simply generated* family of trees (see, e.g., [8] or [12]).

Let $\mathcal{F}_n$ denote the subset of trees of $\mathcal{F}_\Gamma$ that have n nodes and let $y_n = \sum_{T \in \mathcal{F}_n} \omega(T)$; we refer to y_n as the number of (weighted) trees in $\mathcal{F}_n$. It is not difficult to see that the generating function $Y = \sum_1^\infty y_n x^n$ of the simply generated family $\mathcal{F}$ satisfies the relation

$$Y = x\Phi(Y), \tag{2.1}$$

where $\Phi(t) = 1 + \sum_1^\infty \varphi_i t^i$. Three familiar examples of simply generated families are the ordinary ordered trees for which $\Phi(t) = (1-t)^{-1}$ and $y_n = \binom{2n-1}{n-1}/n$; the rooted labelled trees for which $\Phi(t) = e^t$ and $y_n = n^{n-1}/n!$; and the binary trees for which $\Phi(t) = 1 + t^2$ and $y_{2m-1} = \binom{2m-2}{m-1}/m$ and $y_{2m} = 0$.

We shall assume henceforth that $\mathcal{F}$ is some given simply generated family of trees. And when deriving general results of an asymptotic nature we shall assume that the function $\Phi(t)$ appearing in relation (2.1) is analytic in the disk $|t| < R \le \infty$

and that

$$\varphi_i \geq 0 \quad \text{for} \quad i \geq 1 \quad \text{and} \quad \varphi_i > 0 \quad \text{for some} \quad i \geq 1;$$

$$\gcd\{i : i \geq 1 \quad \text{and} \quad \varphi_i > 0\} = 1; \text{ and}$$

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

It follows from these assumptions (cf. [8; p. 1000] or [12; p. 32]) that τ is unique and that $Y(x)$ is analytic in the disk $|x| \leq \rho = \tau/\Phi(\tau)$ except at $x = \rho$; furthermore, $Y(x)$ has an expansion in the neighbourhood of ρ of the form

$$Y(x) = \tau - b(\rho - x)^{1/2} - b_2(\rho - x) - \cdots, \tag{2.2}$$

where $b = \Phi(\tau)\big(2/\tau\Phi''(\tau)\big)^{1/2}$. Hence, by Darboux's theorem (cf. [13, p. 150]),

$$y_n = c\rho^{-n}n^{-3/2} \cdot \big(1 + O(n^{-1})\big) \tag{2.3}$$

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

We shall also need the following results later.

Lemma 1. *Let $F(t)$ and $G(t)$ be functions that are analytic in the disk $|t| < R_1$ and whose power series have non-negative coefficients (not all zero); let $0 < \tau < R_1$. If $m = O(n^{2/3})$ as $n \to \infty$, then*

$$\begin{aligned} &[x^n] : \big\{G\big(Y(x)\big)\big(F\big(Y(x)\big)^m\big) \cdot xY'/Y\big\} \\ &\qquad = G(\tau)F^m(\tau)(2\pi An)^{-1/2}\rho^{-n}e^{-(Bm)^2/2An} \\ &\qquad\quad \times \big(1 + O(1/n) + O(m/n) + O(m^3/n^2)\big), \end{aligned} \tag{2.4}$$

where $A = \tau^2\Phi''(\tau)/\Phi(\tau)$ and $B = \tau F'(\tau)/F(\tau)$.

This can be proved by a straightforward extension of an argument based on the saddle-point method used to prove a related result in [10]; see also [4], [8; pp. 1002-1004], and [9]. (Note that if $m = 0$ and $G(Y) = Y$, then relation (2.4) is equivalent to (2.3).)

Lemma 2. *Suppose that $f \in C^1(0,1)$ and that*

$$\sup_{0<t<1} |f'(t) \cdot t^\alpha(1-t)^\alpha| = M < \infty,$$

where $1 < \alpha < 2$. Then, for $n = 2, 3, \ldots,$

$$\left|n^{-1} \cdot \sum_{j=1}^{n-1} f(j/n) - \int_0^1 f(t)dt\right| \leq HMn^{\alpha-2} + |f(1/2)| \cdot n^{-1},$$

where $H = H(\alpha)$ is independent of f.

The proof of this lemma is based upon the Mean Value Theorem and the assumed estimate for $f'(t)$, appropriately applied to subintervals of length $1/n$ of the interval $(0,1)$; we shall omit the details.

3. A formula for $\mu_k(n)$

Let $\mu_k(n)$ denote the expected value of the Steiner distance $d_T(S)$ where the expectation is taken over all subsets S of k nodes of all trees T in $\mathcal{F}_n$. The weights of the trees are taken into account (here and elsewhere), so that each tree T in $\mathcal{F}_n$ has probability $\omega(T)/y_n$, whence,

$$\mu_k(n)y_n = \binom{n}{k}^{-1} \cdot \sum_{T\in\mathcal{F}_n} \omega(T)\cdot D_k(T).$$

In this section we derive a formula for $\mu_k(n)$ in terms of the numbers $y_1, y_2, \ldots$ and $u_1, u_2, \ldots$, where u_j denotes the number of (weighted) trees $T \in \mathcal{F}_{j+1}$ with a distinguished terminal node (of out-degree zero); that is, $u_0 = 1$ and

$$u_j = \sum_{T\in\mathcal{F}_{j+1}} \omega(T)\cdot N_0(T)$$

for $j \geq 1$, where (as before) $N_0(T)$ denotes the number of nodes of out-degree zero in the tree T.

Proposition 1. *Let $2 \leq k \leq n$; then*

$$\mu_k(n)\binom{n}{k}y_n + \sum_{j=1}^{n-1}\left\{\binom{j}{k} + \binom{n-j}{k}\right\}u_j y_{n-j} = (n-1)\binom{n}{k}y_n, \tag{3.1}$$

where

$$\sum_{j=1}^{n-1} u_j y_{n-j} = (n-1)y_n. \tag{3.2}$$

Proof of (3.2). Suppose $1 \leq j \leq n-1$ and consider one of the u_j trees $T_{j+1} \in \mathcal{F}_{j+1}$ with root node r and distinguished terminal node v. If we identify the node v with the root node of one of the y_{n-j} trees $T_{n-j} \in \mathcal{F}_{n-j}$, we obtain a tree $T_n \in \mathcal{F}_n$ rooted at node r and with a distinguished edge, namely, the edge incident with v in the tree T_{j+1}; note that $\omega(T_{j+1})\cdot\omega(T_{n-j}) = \omega(T_n)$. When we carry out this construction in all possible ways, then each tree $T_n \in \mathcal{F}_n$ is obtained $n-1$ times (and with the proper weight). This implies relation (3.2).

We note, for later use, that (3.2) implies that

$$U(x) = 1 + \sum_{1}^{\infty} u_j x^j = xY'/Y; \tag{3.3}$$

and this implies, upon appealing to relation (2.2) and Darboux's theorem or to relation (2.3), that

$$u_n = \tau^{-1} n y_n + O(y_n) \tag{3.4}$$

as $n \to \infty$. (A weaker form of this relation, without the error term, was proved in [7; p. 164].)

Proof of (3.1). Consider the collection of partially coloured trees T_n in which one edge and the members of a k-set S of nodes of T_n have been given a special colour. The right hand side of relation (3.1) counts the total number of objects in this collection (taking the weights of the trees into account, as usual).

The coloured edge wv of any such partially coloured tree T_n partitions the nodes of T_n into two subsets, A and B, where we may suppose that the root node of T_n and node w belong to A and that node v belongs to B. If the coloured edge does not belong to the subtree determined by the set S of k coloured nodes, then either $S \subseteq A$ or $S \subseteq B$. It follows from this observation and the reasoning used in the derivation of relation (3.2) that the sum in the left hand side of relation (3.1) counts the objects in the collection in which the coloured edge does not belong to the subtree determined by the k coloured nodes.

Finally, it follows from the definition of $\mu_k(n)$ that the first term in relation (3.1) counts the objects in the collection in which the coloured edge does belong to the subtree determined by the k coloured nodes. Relation (3.1) must therefore hold, since the two sides count the same thing.

4. The behaviour of $\mu_k(n)$

We now use the preceding results to determine the asymptotic behaviour of $\mu_k(n)$ as $n \to \infty$; we found it necessary to treat different ranges of values of k separately.

Theorem 1. *If k is any fixed positive integer, then*

$$\mu_k(n) = (k-1)4^{1-k} \binom{2k-2}{k-1} (2\pi n/A)^{1/2} + O(k) \tag{4.1}$$

as $n \to \infty$, where $A = \tau^2 \Phi''(\tau)/\Phi(\tau)$.

If $k = o(n)$ as $k,\ n \to \infty$, then

$$\mu_k(n) = (2kn/A)^{1/2} + O\big((n/k)^{1/2}\big) + O(k). \tag{4.2}$$

If $k/n \to \beta$ as k, $n \to \infty$, where $0 < \beta < 1$, then

$$\mu_k(n) = n\big(1 - \tau^{-1}Y(\rho(1-\beta))\big) + o(n). \tag{4.3}$$

If $d := n - k = o(n)$ as k, $n \to \infty$, then

$$\mu_k(n) = n - 1 - \rho\tau^{-1}d + O(d^2/n). \tag{4.4}$$

Proof of (4.1) and (4.2). We assume for the time being only that $k = o(n)$ as $n \to \infty$. It follows from relations (3.1) and (3.2) that

$$\mu_k(n)(n)_k y_n = \sum_{j=1}^{n-1} \{(n)_k - (j)_k - (n-j)_k\} u_j y_{n-j}.$$

Now $u_j = \tau^{-1} j y_j + O(y_j)$ by (3.4) and

$$\sum_{j=1}^{n-1} y_j y_{n-j} = O(y_n) \tag{4.5}$$

by [8; Lemma 3.1(iii)]; therefore,

$$\mu_k(n)(n)_k y_n = \tau^{-1} \sum_{j=1}^{n-1} \{(n)_k - (j)_k - (n-j)_k\} \cdot j y_j y_{n-j} + O\big((n)_k y_n\big).$$

If we replace j by $n - j$ in the third sum on the right-hand side and then combine it with the second sum, we find that

$$\mu_k(n)(n)_k y_n = \tau^{-1} \sum_{j=1}^{n-1} \{j(n)_k - n(j)_k\} \cdot y_j y_{n-j} + O\big((n)_k y_n\big),$$

or

$$\mu_k(n) = \tau^{-1} \sum_{j=1}^{n-1} \{1 - (j-1)_{k-1}/(n-1)_{k-1}\} \cdot j y_j y_{n-j}/y_n + O(1). \tag{4.6}$$

We now observe that

$$j y_j y_{n-j}/y_n = c j^{-1/2}(1 - j/n)^{-3/2} \cdot \big\{1 + O\big(j^{-1}(1-j/n)^{-1}\big)\big\}, \tag{4.7}$$

by (2.3). Furthermore, if $1 \le k \le j \le n-1$, then

$$\begin{aligned} (j/n)^k \ge (j)_k/(n)_k &\ge \big((j-k)/(n-k)\big)^k \\ &= (j/n)^k \cdot \{1 - k(n-j)/j(n-k)\}^k \\ &\ge (j/n)^k \cdot \{1 - k^2(n-j)/j(n-k)\}; \end{aligned}$$

consequently,

$$(j)_k/(n)_k = (j/n)^k \cdot \{1 + O(k^2 j^{-1}(1 - j/n))\} \tag{4.8}$$

since we are assuming that $k = o(n)$; we note that this relation still holds if $j < k$. Hence, relations (4.6) - (4.8) imply that

$$\mu_k(n) = c\tau^{-1} \sum_{j=1}^{n-1} \{1 - (j/n)^{k-1}\} \cdot j^{-1/2}(1 - j/n)^{-3/2} + A_n + B_n + O(1), \tag{4.9}$$

where

$$A_n = O(1) \cdot \sum_{j=1}^{n-1} \{1 - (j/n)^{k-1}\} \cdot j^{-3/2}(1 - j/n)^{-5/2}$$

and

$$B_n = O(k^2) \cdot \sum_{j=1}^{n-1} (j/n)^{k-1} j^{-3/2}(1 - j/n)^{-1/2}.$$

It is not difficult to see that $1 - t^{k-1} \leq k(1 - t)$ if $0 < t < 1$; hence

$$\begin{aligned} A_n &= O(kn^{3/2}) \cdot \sum_{j=1}^{n-1} j^{-3/2}(n - j)^{-3/2} \\ &= O(k), \end{aligned}$$

where we have appealed to [8; Lemma 3.1(iii)] at the last step. Furthermore,

$$B_n = O(k^2 n^{3/2-k}) \cdot \sum_{j=1}^{n-1} j^{k-5/2} = O(k).$$

Thus it follows from (4.9) and these estimates that

$$\mu_k(n) = c\tau^{-1} n^{-1/2} \sum_{j=1}^{n-1} f(j/n) + O(k),$$

where

$$f(t) = (1 - t^{k-1}) t^{-1/2} (1 - t)^{-3/2}.$$

We may apply Lemma 2 to this sum with $\alpha = 3/2$. It can be shown that $M \leq k - 1$, say, and that $f(1/2) \leq 4$ for all k; consequently,

$$\mu_k(n) = c\tau^{-1} n^{1/2} \cdot I_k + O(k), \tag{4.10}$$

where

$$I_k = \int_0^1 (1 - t^{k-1}) t^{-1/2} (1 - t)^{-3/2} dt.$$

One way of evaluating this integral is to integrate by parts, observing that

$$\frac{d}{dt}\left(2t^{1/2}(1-t)^{-1/2}\right) = t^{-1/2}(1-t)^{-3/2},$$

and then make the substitution $t = \sin^2 u$. Hence,

$$I_k = (1-t^{k-1})\cdot 2t^{1/2}(1-t)^{-1/2}\Big|_0^1 + 2(k-1)\int_0^1 t^{k-3/2}(1-t)^{-1/2}\,dt$$

$$(4.11) \qquad = 4(k-1)\int_0^{\pi/2} \sin^{2k-2} u\, du$$

$$= 2\pi(k-1)4^{1-k}\binom{2k-2}{k-1}.$$

Now

$$(4.12) \qquad 2\pi c\tau^{-1} = (2\pi/A)^{1/2},$$

by the definitions of c and A following (2.3) and (2.4). Thus we may conclude from (4.10) - (4.12) that

$$\mu_k(n) = (k-1)4^{1-k}\binom{2k-2}{k-1}(2\pi n/A)^{1/2} + O(k)$$

provided that $k = o(n)$. This is the required conclusion (4.1) when k is fixed; if $k \to \infty$ but $k = o(n)$, then conclusion (4.2) follows upon applying Stirling's formula.

We remark that the result

$$\mu_2(n) = \frac{1}{2}(2\pi n/A)^{1/2} + O(1)$$

was proved earlier in [3] by a different argument. It is not difficult to see that $\mu_3(T_n) = 3/2\,\mu_2(T_n)$ for any tree T_n with $n \geq 3$ nodes. This implies that

$$\mu_3(n) = \frac{3}{2}\mu_2(n) = \frac{3}{4}(2\pi n/A)^{1/2} + O(1),$$

in accordance with (4.1) when $k = 3$.

Proof of (4.3) and (4.4). It follows from (3.1) and (3.4) that

$$\mu_k(n)(n)_k y_n = (n-1)(n)_k y_n - \sum_{j=1}^{n-1}\{(j)_k + (n-j)_k\}u_j y_{n-j}$$

$$= (n-1)(n)_k y_n - \sum_{j=1}^{n-k}(n-j)_k\{u_j y_{n-j} + u_{n-j} y_j\}$$

$$= (n-1)(n)_k y_n - \left(\tau^{-1}n + O(1)\right)\cdot\sum_{j=1}^{n-k}(n-j)_k y_j y_{n-j}$$

or that

$$\mu_k(n) = n - 1 - \left(\tau^{-1}n + O(1)\right) \cdot \sum_{j=1}^{n-k} f(n,k;j)y_j, \tag{4.13}$$

where

$$f(n,k;j) = \left((n-j)_k/(n)_k\right) \cdot (y_{n-j}/y_n).$$

We now assume that $k/n \to \beta$ as $k, n \to \infty$, where $0 < \beta < 1$. Then, using the facts that $(n-j)_k/(n)_k = (n-k)_j/(n)_j$ and that $y_{n-j}/y_n \to \rho^j$, in view of (2.3), we find that

$$f(n,k;j) \to (1-\beta)^j \rho^j$$

for every fixed value of j, as $k, n \to \infty$. Also, clearly,

$$f(n,k;j) \leq (1 - j/n)^k \cdot (y_{n-j}/y_n);$$

and it follows from (2.3) that $y_{n-j}/y_n \leq K(1-j/n)^{-3/2}\rho^j$ for all n and j, where K is a suitable constant. Hence,

$$f(n,k;j) \leq K(1-j/n)^{k-3/2}\rho^j \leq K\rho^j$$

for all n, k, and j, since we may assume that $k \geq 2$. Now $\sum y_j \rho^j$ converges, so we may apply Tannery's theorem to the sum in the right-hand side of (4.13) and conclude that

$$\sum_{j=1}^{n-k} f(n,k;j)y_j \to \sum_{j=1}^{\infty} (1-\beta)^j \rho^j y_j = Y\left((1-\beta)\rho\right)$$

as k, $n \to \infty$. This and (4.13) imply conclusion (4.3).

We remark that it can be shown, by a more refined argument, that if $2 \leq k \leq n-1$ then

$$\mu_k(n) = n - 1 - \tau^{-1}nY\left((1-k/n)\rho\right) + O\left((n/k)^{1/2}\right). \tag{4.14}$$

In particular, if $k = \beta n + o(n)$, where $0 < \beta < 1$, then (4.14) implies that

$$\mu_k(n) = n - \tau^{-1}nY\left((1-\beta)\rho\right) + O(k - \beta n) + O(1),$$

a slightly stronger version of conclusion (4.3). Conclusion (4.2) can also be deduced from (4.14).

It remains to establish conclusion (4.4). If $d = n - k$, then

$$(n-j)_k/(n)_k = (n-k)_j/(n)_j \leq (d/n)^j$$

for $1 \leq j \leq n - k$. Furthermore,

$$y_{n-1}y_1/y_n = \rho + O(1/n)$$

by (2.3). Therefore,

$$\sum_{j=1}^{n-k} f(n,k;j)y_j = dy_{n-1}y_1/ny_n + O(d^2/n^2)\cdot\sum_{j=2}^{d} y_{n-j}y_j/y_n$$

$$= d\rho/n + O(d^2/n^2),$$

where we have used relation (4.5) again at the last step. Thus it follows from (4.13) that if $d = n-k = o(n)$, then

$$\mu_k(n) = n-1-\left(\tau^{-1}n+O(1)\right)\cdot\left(d\rho/n+O(d^2/n^2)\right)$$

$$= n-1-\tau^{-1}\rho d + O(d^2/n),$$

as required.

5. A probability generating function

Suppose that $2 \le k \le m \le n$ and let $s_{k,m}(T)$ denote the number of subsets S of k nodes of the tree $T \in \mathcal{F}_n$ such that $d_T(S) = m-1$. In this section we derive an expression for the generating function of the numbers

$$N_k(n,m) = \sum_{T\in\mathcal{F}_n} \omega(T)\cdot s_{k,m}(T).$$

For expository purposes we shall regard the coefficients φ_i in the function $\Phi(t)$ as formal variables for the time being; in particular, we shall let φ_0 denote the (formal) weight factor associated with each terminal node of trees in $\mathcal{F}$. We first consider the case when $n = m$.

Proposition 2. *Let* $\Phi(t) = \varphi_0+\varphi_1 t+\varphi_2 t^2+\cdots$ *and* $Q(t) = \varphi_2 t + \varphi_3 t^2 + \cdots$. *Then*

$$(5.1)\quad N_k(m,m) = k^{-1}[t^{k-1}]\Phi^k(t)\sum_{h=0}^{k-1}\binom{h+k-1}{k-1}\binom{m-2}{h+k-2}Q^h(t)\varphi_1^{m-k-h}.$$

Proof. We recall that if $Y = \Sigma y_n x^n$ satisfies the relation $Y = x\Phi(Y)$, then it follows from Lagrange's inversion formula that

$$y_m = \sum_{T\in\mathcal{F}_m}\omega(T) = m^{-1}[t^{m-1}]\Phi^m(t).$$

Let y_{mp} denote the number of trees $T_m \in \mathcal{F}_m$ with p terminal nodes and let y^*_{mp} denote the number of these trees in which the root has out-degree one (where, as usual, the weights of the trees are taken into account). It is not difficult to see that

$$(5.2)\qquad y_{mp} = m^{-1}[t^{m-1}]\binom{m}{p}\varphi_0^p\left(\Phi(t)-\varphi_0\right)^{m-p}$$

and

$$(5.3)\qquad y^*_{mp} = \varphi_1 y_{m-1,p} = (m-1)^{-1}[t^{m-1}]\binom{m-1}{p}\varphi_0^p\big(\Phi(t)-\varphi_0\big)^{m-1-p}\cdot\varphi_1 t.$$

If a tree T_m has p terminal nodes then, clearly,

$$s_{k,m}(T_m) = \binom{m-p}{k-p} \quad \text{or} \quad \binom{m-p-1}{k-p-1}$$

according as the out-degree of the root of T_m is not or is equal to one. It therefore follows, appealing to relations (5.2) and (5.3), the binomial theorem, and the appropriate definitions, that

(5.4)

$$\begin{aligned}
N_k(m,m) &= \sum_p\left\{\binom{m-p}{k-p}(y_{mp}-y^*_{mp}) + \binom{m-p-1}{k-p-1}y^*_{mp}\right\}\\
&= \sum_p\left\{\binom{m-p}{k-p}y_{mp} - \binom{m-p-1}{k-p}y^*_{mp}\right\}\\
&= [t^{m-1}]\Big\{m^{-1}\sum_p\binom{m}{p}\binom{m-p}{k-p}\varphi_0^p\big(\Phi(t)-\varphi_0\big)^{m-p}\\
&\quad - (m-1)^{-1}\sum_p\binom{m-1}{p}\binom{m-p-1}{k-p}\varphi_0^p\big(\Phi(t)-\varphi_0\big)^{m-1-p}\cdot\varphi_1 t\Big\}\\
&= k^{-1}[t^{m-1}]\Big\{\binom{m-1}{k-1}\Phi^k(t)\big(\Phi(t)-\varphi_0\big)^{m-k}\\
&\quad - \binom{m-2}{k-1}\Phi^k(t)\big(\Phi(t)-\varphi_0\big)^{m-k-1}\cdot\varphi_1 t\Big\}\\
&= k^{-1}[t^{k-1}]\Big\{\binom{m-1}{k-1}\Phi^k(t)\big(\varphi_1+Q(t)\big)^{m-k}\\
&\quad - \binom{m-2}{k-1}\Phi^k(t)\big(\varphi_1+Q(t)\big)^{m-k-1}\cdot\varphi_1\Big\}.
\end{aligned}$$

Now

$$(5.5)\qquad \big(\varphi_1+Q(t)\big)^{m-k} = \sum_h\binom{m-k}{h}Q^h(t)\varphi_1^{m-k-h},$$

$$(5.6)\qquad \big(\varphi_1+Q(t)\big)^{m-k-1}\cdot\varphi_1 = \sum_h\binom{m-k-1}{h}Q^h(t)\varphi_1^{m-k-h},$$

and

$$(5.7)\quad \binom{m-1}{k-1}\binom{m-k}{h} - \binom{m-2}{k-1}\binom{m-k-1}{h} = \binom{h+k-1}{k-1}\binom{m-2}{h+k-2}.$$

We now substitute expansions (5.5) and (5.6) into the last line of expression (5.4); when we simplify, using identity (5.7), and note that $[t^{k-1}]\Phi^k(t)Q^h(t) = 0$ if $h \geq k$, we obtain the required relation (5.1).

We now consider the general case.

Theorem 2. *Let*

$$P = P(x,t) = x\Phi\big(Y(x)+t\big) = x\Phi(Y) + x\Phi'(Y)t + x\Phi''(Y)t^2/2! + \cdots$$

and

$$Q = Q(x,t) = t^{-1}\{x\Phi\big(Y(x)+t\big) - x\Phi\big(Y(x)\big) - x\Phi'\big(Y(x)\big)t\}$$
$$= x\Phi''(Y)t/2! + x\Phi'''(Y)t^2/3! + \cdots .$$

Then

(5.8)

$$N_k(n,m) = k^{-1}[t^{k-1}x^n]P^k$$
$$\times \sum_{h=0}^{k-1} \binom{h+k-1}{k-1}\binom{m-2}{h+k-2} Q^h \left(x\Phi'(Y)\right)^{m-k-h} \cdot xY'/Y.$$

Proof. We begin by observing that

$$N_k(n,m) = \sum_{T_n} \omega(T_n)\cdot s_{k,m}(T_n)$$
$$= \sum_{T_n}\Big\{\sum_{T_m\subseteq T_n} s_{k,m}(T_m)\Big\}\cdot\omega(T_n)$$
$$= \sum_{T_m}\Big\{\sum_{T_n:T_m\subseteq T_n} \omega(T_n)\Big\}\cdot s_{k,m}(T_m).$$

We already have an expression for

$$N_k(m,m) = \sum_{T_m} \omega(T_m)\cdot s_{k,m}(T_m)$$

in Proposition 2. We shall make certain modifications to the expression for $N_k(m,m)$ that will have the effect, for each tree T_m, of replacing the weight factor $\omega(T_m)$ by the sum $\sum' \omega(T_n)$ of the weights of all trees T_n that contain T_m as a subtree.

Consider any term in expression (5.1) for $N_k(m,m)$ that is associated with some tree T_m. Such a term, apart from numerical factors, is a product $\varphi_{d_1}\cdots\varphi_{d_m}$ of m factors; and each factor φ_d corresponds to a node v of out-degree d in T_m. Suppose that in a tree T_n that contains T_m as a subtree there are e additional edges attached to node v (and leading away from the root of T_m). These additional e edges could be attached to v amongst the d original edges leading away from v in $(d+1)(d+2)\cdots(d+e)/e! = (d+e)_d/d!$ ways; and the generating function that enumerates the (weights of) the branches determined by these additional e edges is simply $Y^e(x)$. The weight factor φ_d originally associated with node v in T_m should now be replaced by the weight factor φ_{d+e}. Consequently, to obtain the generating function for all the trees obtained by attaching additional branches to the nodes of T_m, each factor φ_d that appears in expression (5.1) should be replaced by

$$x\left\{\binom{d}{d}\varphi_d + \binom{d+1}{d}\varphi_{d+1}Y(x) + \binom{d+2}{d}\varphi_{d+2}Y^2(x) + \cdots\right\} = x\Phi^{(d)}(Y)/d!\,;$$

the additional x factors are to take the nodes v of T_m into account. This may be accomplished formally by making the following replacements in the right hand side of (5.1):

(i) the function $\Phi(t) = \sum_d \varphi_d t^d$ is replaced by $x\sum_d \Phi^{(d)}(Y)t^d/d! = P(x,t)$;
(ii) the function $Q(t) = \sum_{d\ge 2}\varphi_d t^{d-1}$ is replaced by $x\sum_{d\ge 2}\Phi^{(d)}(Y)t^{d-1}/d! = Q(x,t)$;
(iii) the factor φ_1 is replaced by $x\Phi'(Y)$.

We are almost, but not quite finished. If a tree T_n contains the tree T_m, let T_{n-j} denote the subtree of T_n determined by all nodes z such that the path from z to the root of T_n contains the root of T_m; and let T_{j+1} denote the subtree of T_n determined by the root T_m and any nodes not in T_{n-j}. In the last paragraph we described how to obtain the generating function for the trees T_{n-j}, in effect. Now the tree T_{j+1} is either the trivial tree consisting of a single node (if the roots of T_m and T_n coincide) or it is a non-trivial tree rooted at the root of the tree T_n with a distinguished terminal node, namely, the root of T_m. The number of such trees is u_j. So once we have made the replacements to expression (5.1) described in (i) - (iii) above, we should multiply by $U(x) = xY'/Y$, in view of (3.3), to take into account the possibilities for the trees T_{j+1}. The coefficient of $t^{k-1}x^n$ in the resulting expression equals $N_k(n,m)$. This completes the proof of the theorem.

We remark, by way of illustration, that when $k=2$ relation (5.8) implies that

$$N_2(n,m) = [x^n]\left\{Y\big(x\Phi'(Y)\big)^{m-1} + (m-2)Y^2\big(x\Phi''(Y)/2\big)\big(x\Phi'(Y)\big)^{m-3}\right\}\cdot xY'/Y.$$

In this case the subtree T_m considered above is simply a path of length $m-1$. The node of T_m that is closest to the root of a tree T_n containing T_m is either one of the end-nodes of the path T_m or one of the $m-2$ interior nodes of the path. The two terms in the expression for $N_2(n,m)$ correspond to these two alternatives. The factors Y, $x\Phi'(Y)$, and $x\Phi''(Y)/2$ correspond to the nodes of out-degree $0, 1$, and 2 in the path T_m.

6. A limiting distribution

Suppose that $2 \le k \le m \le n$ and let $p_k(n, m)$ denote the probability that $d_T(S) = m - 1$ over all subsets S of k nodes of all trees $T \in \mathcal{F}_n$, i.e.,

$$p_k(n, m) = N_k(n, m) \Big/ \binom{n}{k} y_n.$$

We now determine the limiting behaviour of $p_k(n, m)$ for fixed values of k if $m = O(n^{2/3})$ as $m, n \to \infty$.

Theorem 3. *Let k be any fixed integer, $k \ge 2$. If $m = O(n^{2/3})$ as $m, n \to \infty$, then*

$$p_k(n, m) = \frac{A}{(k-2)!} \left(\frac{Am^2}{2n} \right)^{k-2} \frac{m}{n} e^{-Am^2/2n} \cdot \left(1 + O(1/m) + O(m/n) + O(m^3/n^2)\right),$$

where $A = \tau^2 \Phi''(\tau)/\Phi(\tau)$.

Proof. Suppose that $0 \le h \le k-1$ and let the function $G_h(Y) = G_h\big(Y(x)\big)$ be defined as follows:

$$\begin{aligned} x^{k+h} G_h(Y) &= [t^{k-1}]\{P^k Q^h\} \\ &= x^{k+h} [t^{k-1}] \Big(\sum_{j \ge 0} \Phi^{(j)}(Y)\, t^j / j! \Big)^k \Big(\sum_{j \ge 2} \Phi^{(j)}(Y) t^{j-1} / j! \Big)^h. \end{aligned}$$

For example,

$$G_{k-1}(Y) = \big(\Phi(Y)\big)^k \big(\Phi''(Y)/2\big)^{k-1} \tag{6.1}$$

and

$$\begin{aligned} G_{k-2}(Y) &= k\big(\Phi(Y)\big)^{k-1} \Phi'(Y) \big(\Phi''(Y)/2\big)^{k-2} \\ &\quad + (k-2)\big(\Phi(Y)\big)^k \big(\Phi''(Y)/2\big)^{k-3} \big(\Phi'''(Y)/3!\big). \end{aligned}$$

Then formula (5.8) of Theorem 2 can be rewritten as

$$p_k(n, m) \binom{n}{k} y_n = \sum_{h=0}^{k-1} a_h(m) f_h(n, m), \tag{6.2}$$

where

$$a_h(m) = \frac{h + k - 1}{m(m-1)} \binom{m}{k} \binom{m-k}{h} \tag{6.3}$$

and

(6.4) $$f_h(n,m) = [x^{n-m}]\Big\{G_h(Y)\big(\Phi'(Y)\big)^{m-k-h}\cdot xY'/Y\Big\}$$

for $0 \le h \le k-1$.

We evaluate $f_h(n,m)$ by appealing to relation (2.4) with $G(Y) = G_h(Y)$, $F(Y) = \Phi'(Y)$, and with n and m replaced by $n-m$ and $m-k-h$, respectively. To simplify the resulting expression we note that

$$(m-k-h)^2/(n-m) = m^2/n + O(m/n) + O(m^3/n^2),$$

$$(n-m)^{-1/2} = n^{-1/2}\cdot\big(1+O(m/n)\big),$$

$$B = \tau F'(\tau)/F(\tau) = \tau\Phi''(\tau)/\Phi'(\tau)$$

$$= \tau^2\Phi''(\tau)/\Phi(\tau) = A,$$

and that

$$\rho\Phi'(\tau) = \tau\Phi'(\tau)/\Phi(\tau) = 1.$$

In this way we find that if $0 \le h \le k-1$, then

(6.5) $$f_h(n,m) = G_h(\tau)\rho^{-n+k+h}(2\pi An)^{-1/2}e^{-Am^2/2n}\cdot\big(1+O(m/n)+O(m^3/n^2)\big).$$

In particular,

$$G_{k-1}(\tau) = \big(\Phi(\tau)\big)^k\big(\Phi''(\tau)/2\big)^{k-1} = \tau(A/2)^{k-1}\rho^{1-2k},$$

from (6.1), so

(6.6)

$$f_{k-1}(n,m) = \tau(A/2)^{k-1}(2\pi An)^{-1/2}\rho^{-n}e^{-Am^2/2n}\cdot\big(1+O(m/n)+O(m^3/n^2)\big)$$

$$= (A/2)^{k-1}\cdot ny_ne^{-Am^2/2n}\cdot\big(1+O(m/n)+O(m^3/n^2)\big),$$

appealing to (2.3) at the last step.

We now observe that the coefficient $a_{k-1}(m)$ dominates the remaining coefficients that appear on the right-hand side of (6.2). For,

(6.7) $$a_h(m)/a_{k-1}(m) = \frac{h+k-1}{2(k-1)}\cdot\frac{(k-1)_{k-1-h}}{(m-k-h)_{k-1-h}}$$

$$\le \big((k-1)/(m-k-h)\big)^{k-1-h} = O\big((1/m)^{k-1-h}\big)$$

if $0 \le h \le k-1$. Moreover,

$$(6.8)\qquad \begin{aligned} a_{k-1}(m) &= \frac{2(k-1)}{m(m-1)} \binom{m}{k} \binom{m-k}{k-1} \\ &= \frac{2}{k!(k-2)!}\, m^{2k-3} \cdot (1 + O(1/m)) \end{aligned}$$

from (6.3). When we combine relations (6.2) and (6.5) - (6.8) we find that

$$\begin{aligned} p_k(n,m) &= \big(1 + O(1/m)\big) \cdot a_{k-1}(m) f_{k-1}(n,m) \Big/ \binom{n}{k} y_n \\ &= \frac{A}{(k-2)!} \left(\frac{Am^2}{2n}\right)^{k-2} \frac{m}{n}\, e^{-Am^2/2n} \cdot \big(1 + O(1/m) + O(m/n) + O(m^3/n^2)\big), \end{aligned}$$

as required.

Corollary 3.1. *Let k be any fixed integer, $k \ge 2$. If λ is any positive constant, then*

$$\sum_{m \le (2\lambda n/A)^{1/2}} p_k(n,m) = \frac{1}{(k-2)!} \int_0^\lambda u^{k-2} e^{-u} du + O(n^{-1/2})$$

as $n \to \infty$.

This result follows readily from Theorem 3.

Let $\nu_k(n)$ denote the expected value of $d_T(S) \cdot \big(d_T(S)+1\big)$ over all subsets S of k nodes of all trees $T \in \mathcal{F}_n$. It is possible to derive a rather complicated formula for $\nu_k(n)$ analogous to formula (3.1) for $\mu_k(n)$; but we were not able to deduce the asymptotic behaviour of $\nu_k(n)$ from this formula. However, Theorem 2 can be used to determine the behaviour of both $\mu_k(n)$ and $\nu_k(n)$ when k is fixed. We shall give the argument only for $\nu_k(n)$ since the argument for $\mu_k(n)$ is very similar (and would not add anything to the results of Theorem 1).

Theorem 4. *If k is any fixed positive integer, then*

$$\nu_k(n) = 2(k-1)n/A + O(n^{1/2})$$

as $n \to \infty$, where $A = \tau^2 \Phi''(\tau)/\Phi(\tau)$.

Proof. For any fixed value of k, consider the generating function

$$V(x) = \sum_n \nu_k(n) \binom{n}{k} y_n x^n.$$

Let

$$(6.9)\qquad b_h(x) = (h+k-1) \binom{h+k}{k} [t^{k-1}]\{P^k(x,t) Q^h(x,t)\}$$

for $0 \le h \le k-1$, where $P(x,t)$ and $Q(x,t)$ are as defined in Theorem 2. We assert that

(6.10) $$V(x) = \sum_{h=0}^{k-1} b_h(x)(xY'/Y)^{h+k+2}.$$

This expression follows upon multiplying both sides of equation (5.8) by $m(m-1)x^n$ and then summing over m and n; to simplify the right-hand side of the resulting expression, we interchange the order of summation and appeal to Newton's binomial theorem and the identity $\left(1 - x\Phi'(Y)\right)^{-1} = xY'/Y$. We will determine the asymptotic behaviour of $\nu_k(n)$ by applying Darboux's theorem to the function $V(x)$.

We recall that xY'/Y and, for any fixed j, the function $x\Phi^{(j)}(Y)$ are analytic in the disk $|x| \le \rho$ except at $x = \rho$ (cf. [8; p. 1000]); and, in view of (2.2), they have expansions in the neighbourhood of ρ of the form

(6.11) $$xY'/Y = (\rho b/2\tau)(\rho - x)^{-1/2} + a_1 + a_2(\rho - x)^{1/2} + \cdots$$

and

(6.12) $$x\Phi^{(j)}(Y) = \rho\Phi^{(j)}(\tau) + c_1(\rho - x)^{1/2} + c_2(\rho - x)^1 + \cdots .$$

Thus it follows from (2.2) and (6.9) - (6.11) that $V(x)$ is analytic in the disk $|x| \le \rho$ except at $x = \rho$ and that it has an expansion in the neighbourhood of ρ of the form

(6.13) $$V(x) = b_{k-1}(\rho) \cdot (\rho b/2\tau)^{2k+1}(\rho - x)^{-k-\frac{1}{2}} + v_1(\rho - x)^{-k} + \cdots .$$

If we now apply Darboux's theorem [13; p. 150] to $V(x)$, using expansion (6.13), we find that

(6.14)
$$\nu_k(n)\binom{n}{k}y_n = b_{k-1}(\rho)(\rho b/2\tau)^{2k+1} \cdot n^{k-\frac{1}{2}}\rho^{-n-k-\frac{1}{2}} \Big/ \Gamma(k + \frac{1}{2}) + O(n^{k-1}\rho^{-n}).$$

We observe that

$$(\rho b/2\tau)^2 = \rho^2\Phi^2(\tau)/2\tau^3\Phi''(\tau)$$
$$= \rho\Phi(\tau)/2\tau^2\Phi''(\tau) = \rho/2A,$$

by the definitions of b, ρ, and A given in §2; moreover,

$$b_{k-1}(x) = 2(k-1)\binom{2k-1}{k}Y^k\left(x\Phi''(Y)/2\right)^{k-1}.$$

Therefore,

$$b_{k-1}(\rho)(\rho b/2\tau)^{2k+1} = 4(k-1)\binom{2k-1}{k}\frac{\tau}{A}\left(\frac{\rho}{4}\right)^k\left(\frac{\rho}{2A}\right)^{1/2},$$

after some simplification. If we substitute this in the right-hand side of (6.14), recall that $\Gamma(k+1/2) = \pi^{1/2}(2k)!/4^k k!$, simplify, and appeal to (2.3), we find that

$$\begin{aligned}\nu_k(n)\binom{n}{k} y_n &= 4(k-1)\binom{2k-1}{k}\frac{k!}{(2k)!}\frac{\tau}{A}(2\pi A n)^{-1/2}\cdot n^k\rho^{-n} + O(n^{k-1}\rho^{-n})\\ &= 2(k-1)A^{-1}n^{k+1}y_n/k! + O(n^{k+\frac{1}{2}}y_n).\end{aligned}$$

This suffices to complete the proof of the theorem.

7. Special Cases

We now illustrate some of the preceding results for three particular families of simply generated trees. Let $\mathcal{F}$ denote the family of ordinary ordered trees (cf. [6; p. 112] or [12; p. 30]), for which $\Phi(t) = (1-t)^{-1}$ and

$$Y = \frac{1}{2}\{1-(1-4x)^{1/2}\} = \sum_1^{\infty}\binom{2n-2}{n-1}\frac{x^n}{n}. \tag{7.1}$$

In this case the expression in Theorem 2 can be simplified to yield the explicit formula

$$\begin{aligned} p_k(n,m) &= y_k\binom{m+k-3}{2k-3}\binom{2n-2}{n-m} \div \binom{n}{k} y_n\\ &= 2\binom{n-1}{k-2}\binom{n-k}{m-k} \div \binom{n+m-2}{n-k+1}.\end{aligned} \tag{7.2}$$

And it follows from (3.3) and (3.1) that $u_j = (2j-1)y_j$ for $j \geq 1$ and that

$$\begin{aligned}\mu_k(n)(n)_k y_n &= (n-1)(n)_k y_n - \sum_{j=1}^{n-1}\{(j)_k + (n-j)_k\}(2j-1)y_j y_{n-j}\\ &= (n-1)(n)_k y_n - 2(n-1)\sum_{j=k}^{n-1}(j)_k y_j y_{n-j}.\end{aligned}$$

When we divide by $(n-1)$ and express this last relation in terms of generating functions and simplify, making use of (7.1), we find that

$$\mu_k(n) = (k-1)\binom{2k-2}{k-1}4^{n-k} \div \binom{2n-2}{n-1}. \tag{7.3}$$

Hence

$$\mu_k(n) = (k-1)4^{1-k}\binom{2k-2}{k-1}\left(\pi(n-1)\right)^{1/2}\cdot\left(1+O(n^{-1})\right)$$

for any k, as $n \to \infty$, and

$$\mu_k(n) = \left((k-1)(n-1)\right)^{1/2}\cdot\left(1+O((n-k)/kn)\right)$$

as k, $n \to \infty$. (Formula (7.3) can also be deduced from formula (7.2).) It can also be shown that $\mu_k^{(2)}(n)$, the expected value of $\left(d_T(S)\right)^2$ over all subsets S of k nodes of the y_n ordered trees $T \in \mathcal{F}_n$ is given by the surprisingly simple formula

$$\mu_k^{(2)}(n) = (k-1)(n-1).$$

Now let $\mathcal{F}$ denote the family of rooted labelled trees, for which $\Phi(t) = e^t$ and $y_n = n^{n-1}/n!$ (cf. [5; p. 174] or [6; p. 392]). In this case it follows from (3.3) that $u_j = jy_j = j^j/j!$ for $j \geq 1$; and it is not difficult to see that the argument used to derive formula (4.13) leads here to the exact formula

$$\mu_k(n) = (n-1) - \sum_{j=1}^{n-k}\binom{n-k}{j}(j/n)^{j-1}(1-j/n)^{n-j-1}.$$

Note that this implies relations (4.3) and (4.4) in this case. Furthermore, it can also be shown, using a modified version of relation (5.8), that if $2 \leq k \leq n$ then

$$\mu_k(n) = \frac{n}{(n)_k}\sum_{j=k}^{n}(j)_k(n)_j n^{-j}\cdot[t^{j-2}]\{e^{kt}(e^t-1)^{j-k}\}$$

for the family of labelled trees. We remark that this expression can be used to give an alternate derivation of relation (4.1) in this case. In particular, it also implies that

$$\mu_2(n) = \left(n/(n)_2\right)\sum_{j=2}^{n}(j)_2(n)_j n^{-j}$$

and

$$\mu_3(n) = \left(n/2(n)_3\right)\sum_{j=3}^{n}(j+3)(j)_3(n)_j n^{-j} = 3/2\,\mu_2(n).$$

Let $\mathcal{F}$ now denote the family of binary trees (cf. [6; p. 389] or [12; p. 29]), for which $\Phi(t) = 1+t^2$ and

(7.4) $$Y = (1/2x)\{1-(1-4x^2)^{1/2}\} = \sum_{1}^{\infty}\binom{2m-2}{m-1}\frac{x^{2m-1}}{m}.$$

In this case it follows from (3.3) that $u_j = \frac{1}{2}(j+2)y_{j+1}$ when j is even (both sides equal zero when j is odd) and hence, by (3.1), that

$$\mu_k(n)(n)_k y_n + \frac{1}{2}\sum_{j=1}^{m-1}\{(j)_k + (n-j)_k\}(j+2)y_{j+1}y_{n-j} = (n-1)(n)_k y_n,$$

where we assume that n is odd (and that the sum is over even integers j). The function $\Phi(t) = 1 + t^2$ does not satisfy all the conditions stated in Section 1 that we have assumed were satisfied when deriving asymptotic results. Nevertheless, it can be shown that the conclusions of Theorems 1, 3, and 4 and Corollary 3.1 still hold for the family of binary trees provided that n is an odd integer. In particular, it can be shown, making use of relation (7.4), that if n is odd and $k/n \to \beta$ as k, $n \to \infty$, where $0 < \beta < 1$, then

$$\mu_k(n) = 2n\beta^{1/2}\left(\beta^{1/2} + (2-\beta)^{1/2}\right)^{-1} + o(n).$$

In the accompanying table we give the first four digits of some values of $1 - \tau^{-1}Y\big(\rho(1-\beta)\big)$ for these families of trees; these numbers give the limiting values of $\mu_k(n)/n$ if $k/n \to \beta$, in view of relation (4.3).

β	Ordered Trees	Labelled Trees	Binary Trees
.10	.3162	.3916	.3732
.25	.5000	.5801	.5485
.50	.7071	.7680	.7320
.75	.8660	.8981	.8729
.90	.9486	.9617	.9498

Table: Values of $1 - \tau^{-1}Y\big(\rho(1-\beta)\big)$

We conclude with an example that shows that the analogue of Theorem 1 does not necessarily hold for non-simply generated families of trees. We recall that a tree T_n with n labelled nodes, rooted at node 1, is a *recursive* tree if $n = 1$ or if $n \geq 2$ and T_n is obtained by joining the n-th node to a node of some recursive tree T_{n-1} (see, e.g., [11] or [8]). Let $\mu_k(n)$ now denote the expected value of $d_T(S)$ over all subsets S of k nodes of the $(n-1)!$ recursive trees T with n labelled nodes.

Theorem 5. *If* $2 \leq k \leq n$, *then*

$$\mu_k(n) = kn(n-k+1)^{-1} \cdot \sum_{j=k}^{n} j^{-1} - 1 - \varepsilon_k, \tag{7.5}$$

where $0 \leq \varepsilon_k \leq (n-k)/n(k-1)$.

Sketch of Proof. When we adapt the argument that led to relation (3.1) to the family of recursive trees, we find that

$$\mu_k(n)\binom{n}{k}(n-1)! + \sum_{j=1}^{n-1}\binom{n}{j+1}(j-1)!(n-j-1)!\left\{\binom{j}{k}+\binom{n-j}{k}\right\}$$

$$= (n-1)\binom{n}{k}(n-1)!$$

or that

$$(7.6) \qquad (\mu_k(n)+1)(n-1)_{k-1} = (n)_k - \sum_{j=1}^{n-1}\{(j)_k + (n-j)_k\}/j(j+1).$$

We now assert that

$$(7.7) \qquad (n)_k - \sum_{j=1}^{n-k}(n-j)_k/j(j+1) = k\sum_{j=1}^{n+1-k}(n-j)_{k-1}/j$$

$$= k(n)_{k-1}\cdot\sum_{j=k}^{n} j^{-1}.$$

The first equality follows from summation by parts and the second is an identity that can be proved by induction on k. Furthermore,

$$(7.8) \sum_{j=k}^{n-1}(j)_k/j(j+1) \le (n-2)/n\cdot\sum_{j=k}^{n-1}(j-2)_{k-2} = (n-2)(n-2)_{k-1}/n(k-1).$$

Conclusion (7.5) now follows from (7.6) - (7.8). We remark that the particular result

$$\mu_2(n) = 2(n+1)(n-1)^{-1}\cdot\sum_{j=2}^{n} j^{-1} - 2,$$

which follows readily from (7.6) and (7.7), was given earlier in [11].

Analyzing the behaviour of the second moment seems to be a more complicated problem. It can be shown, however, that if k is fixed then the variance of $d_T(S)$ over all subsets S of k nodes of the recursive trees T with n nodes equals $O(\log n)$. So it follows from Chebyshev's inequality that the Steiner distance of recursive trees T_n is concentrated around $k \log n$ when k is fixed and n is large.

Acknowledgement. The work of the second author was assisted by a Special General Research Grant from the University of Alberta.

References

[1] L.W. Beineke, O.R. Oellermann, and R.E. Pippert, On the Steiner median of a tree, *Disc. Appl. Math.*, **68** (1996), 249-258.

[2] P. Dankelmann, O.R. Oellermann, and H.C. Swart, The average Steiner distance of a graph, *J. Graph Th.*, **22** (1996), 15-22.

[3] R.C. Entringer, A. Meir, J.W. Moon and L.A. Székely, On the Wiener index of trees from certain families, *Australasian J. Comb.*, **10** (1994), 211-224.

[4] I.J. Good, Saddle-point methods for the multinomial distribution, *Ann. Math. Stat.*, **28** (1957), 861-888.

[5] I.P. Goulden and D.M. Jackson, *Combinatorial Enumeration*, Wiley, New York, 1983.

[6] D.E. Knuth, *The Art of Computer Programming*, Vol. **1**, Addison-Wesley, Reading, 1973.

[7] A. Meir and J.W. Moon, Packing and covering constants for certain families of trees, I, *J. Graph Th.*, **1** (1977), 157-174.

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

[9] A. Meir and J.W. Moon, The asymptotic behaviour of powers of certain generating functions, *Europ. J. Comb.*, **11** (1990), 581-587.

[10] A. Meir and J.W. Moon, On the bipartition numbers of random trees, II, *Ars Comb.* (to appear).

[11] J.W. Moon, The distance between nodes in recursive trees, *Proc. British Comb. Conf.* 1973 (Cambridge, 1974), 125-132.

[12] J.-M. Steyaert and P. Flajolet, Patterns and pattern-matching in trees: an analysis, *Information and Control*, **58** (1983), 19-58.

[13] H.S. Wilf, *Generatingfunctionology*, Academic Press, San Diego, 1990.

(Received 16/6/98)