Studia Scientiarum Mathematicarum Hungarica 5 (1970) 5—10

ON THE NUMBER OF ENDPOINTS OF A k-TREE

by

A. RÉNYI

§ 1. Definitions

A k-tree ($k=1, 2, 3, \ldots$) is the natural generalization to k dimensions of an ordinary (i.e. one-dimensional) tree (see [1], [2], [3], [4]). A k-tree can be considered either as a k-dimensional simplicial complex with certain properties, or as a graph. We shall take the second point of view.

A k-tree can be most conveniently defined inductively as follows: A k-tree of order $k+1$ is a complete $(k+1)$-graph. A k-tree $t^{(k)}_{(n+1)}$ of order $n+1$ ($n \geqq k+1$) is obtained by choosing an arbitrary k-tree $t^{(k)}_n$ of order n, and adding a new vertex, joining it to k such points of $t^{(k)}_n$ which form a complete graph in $t^{(k)}_n$, i.e. to the points of a $(k-1)$-cell of $t^{(k)}_n$. Thus a k-tree of order n contains n points (vertices), $n-k$ k-cells (i.e. complete subgraphs of order $k+1$) and $k(n-k)+1$ $(k-1)$-cells (i.e. complete subgraphs of order k).

A point of a k-tree of order n ($\geqq k+1$) is called an *endpoint* if it belongs to a single k-cell of the k-tree.

It is easy to see by induction that the number of endpoints of a k-tree of order $n \geqq k+1$ is at least 2 and at most $n-k$.

As a matter of fact two endpoints can not belong to the same $(k-1)$-cell if $n \geqq k+2$; thus if we take a k-tree $t^{(k)}_n$ of order $n \geqq k+2$ and form a k-tree $t^{(k)}_{n+1}$ of order $n+1$ by adding to $t^{(k)}_n$ a new point (joining it with the points of a $(k-1)$-cell of $t^{(k)}_n$) then the new point will be an endpoint, of $t^{(k)}_{n+1}$ and among the endpoints of $t^{(k)}_n$ at most one will not be an endpoint of $t^{(k)}_{n+1}$, thus the number of endpoints is either unchanged, or is increased by one; as a k-tree of order $k+2$ consists of two k-cells which have a $(k-1)$-cell in common and thus it contains exactly 2 endpoints, it follows that any k-tree of order $n \geqq k+2$ contains at least 2 and at most $n-k$ endpoints.

In this paper we consider *labeled*, more exactly: point-labeled k-trees, i.e. such k-trees of order n, the points of which are labeled by the numbers $1, 2, \ldots, n$. It has been shown (see [2] and [3]) that denoting by $T_k(n)$ the total number of labeled k-trees of order n, one has

$$T_k(n) = \binom{n}{k}[k(n-k)+1]^{n-k-2} \qquad (k=1, 2, \ldots). \tag{1}$$

The aim of the present paper is to determine the number $T_k(n, r)$ of labeled k-trees of order n having r endpoints ($2 \leqq r \leqq n-k$). The corresponding problem or 1-trees has been solved in [5].

§ 2. Exact formulae

We prove first the following recursion formula

$$T_k(n+1, r)r = (n+1)[(k(n-k-r+1)+1)T_k(n, r-1)+rkT_k(n, r)] \tag{2}$$

where $2 \leqq r \leqq n+1-k$, $n \geqq k+2$.

To prove (2) let us mention that if we take any k-tree $t_n^{(k)}$ of order n having r endpoints, and join a new point to one of the $(k-1)$-cells of $t_n^{(k)}$ containing an endpoint of $t_n^{(k)}$, we get a k-tree $t_{n+1}^{(k)}$ of order $n+1$ with r endpoints, because one endpoint of $t_n^{(k)}$ disappears and one new endpoint is created. On the other hand if we take a k-tree $t_n^{(k)}$ of order n having $r-1$ endpoints and join a new point to one of the $(k-1)$-cells not containing any of the endpoints of $t_n^{(k)}$, we get a k-tree $t_{n+1}^{(k)}$ of order $n+1$ having r endpoints, because all the $r-1$ endpoints of $t_n^{(k)}$ will be endpoints of $t_{n+1}^{(k)}$, and the new point will also be an endpoint of $t_{n+1}^{(k)}$. If we do this with all k-trees of order n labeled by the numbers $1, 2, \ldots, n+1$ except j (where $j = 1, 2, \ldots, n+1$), we get all possible k-trees of order $n+1$ having r endpoints, each exactly r times; taking into account that each endpoint of a k-tree of order $n \geqq k+2$ belongs to k of its $k(n-k)+1$ $(k-1)$-cells, (2) follows.

Let us put now

$$t_k(n, s) = \frac{T_k(n, n-s)}{\binom{n}{s}}. \tag{3}$$

It follows that

$$t_k(n+1, s) = [k(s-k)+1]\cdot t_k(n, s)+kst_k(n, s-1) \quad \text{for} \quad k \leqq s \leqq n-1, \tag{4}$$

where $t_k(k+1, k) = 0$.
Thus, putting

$$G_k(z, s) = \sum_{n=s+2}^{+\infty} t_k(n, s)z^{n-s-2} \qquad (s = k, k+1, \ldots) \tag{5}$$

we obtain

$$G_k(z, s) = \frac{ksG_k(z, s-1)}{1-z[k(s-k)+1]}. \tag{6}$$

As however $T_k(n, n-k) = \binom{n}{k}$ if $n \geqq k+2$, we have

$$G_k(z, k) = \frac{1}{1-z}. \tag{7}$$

Thus it follows from (6) that

$$G_k(z, s) = \frac{s!\, k^{s-k}}{k! \prod_{j=0}^{s-k} (1-z(jk+1))}. \tag{8}$$

It follows in particular for $k=1$ that

$$G_1(z,s)=\frac{s!}{\prod_{h=1}^{s}(1-hz)}. \tag{9}$$

Now it is known (see [5]) that for $k=1$

$$T_1(n,n-s)=S(n-2,s)s!\binom{n}{s} \tag{10}$$

where $S(m,s)$ are the Stirling numbers of the second type, defined by

$$y^m=\sum_{s=1}^{m}S(m,s)y(y-1)\dots(y-s+1). \tag{11}$$

From (10) we can deduce (9) directly as follows.

It follows from (11) that

$$\sum_{m=s}^{\infty}\frac{S(m,s)z^m}{m!}=\frac{(e^z-1)^s}{s!} \tag{12}$$

and thus for $|z|<\frac{1}{s}$

$$\sum_{m=s}^{\infty}S(m,s)z^m=\sum_{m=s}^{\infty}\frac{S(m,s)z^m}{m!}\int_0^{\infty}y^m e^{-y}\,dy=\int_0^{\infty}\frac{(e^{zy}-1)^s}{s!}e^{-y}\,dy. \tag{13}$$

It follows by partial integration that

$$\sum_{m=s}^{\infty}S(m,s)z^m=\frac{z^s}{\prod_{h=1}^{s}(1-hz)} \tag{14}$$

i. e.

$$G_1(z,s)=s!\sum_{m=s}^{\infty}S(m,s)z^{m-s}=\frac{s!}{\prod_{h=1}^{s}(1-hz)} \tag{15}$$

in accordance with (9).

To get an explicit expression for $T_k(n,r)$ we need the following identity

$$\frac{1}{\prod_{j=0}^{m}[1-z(jk+1)]}=\frac{1}{m!\,k^m}\sum_{j=0}^{m}\binom{m}{j}\frac{(-1)^{m-j}(jk+1)^m}{1-z(jk+1)} \tag{16}$$

which can be proved e.g. by elementary function theory.

It follows from (8) and (16)

$$G_k(z,s)=\binom{s}{k}\sum_{j=0}^{s-k}\binom{s-k}{j}\frac{(-1)^{s-k-j}(jk+1)^{s-k}}{1-z(jk+1)} \tag{17}$$

and thus

$$(18)\qquad t_k(n,s)=\frac{T_k(n,n-s)}{\binom{n}{s}}=\binom{s}{k}\sum_{j=0}^{s-k}\binom{s-k}{j}(-1)^{s-k-j}(jk+1)^{n-k-2}.$$

Thus we obtain finally

$$(19)\qquad T_k(n,n-s)=\binom{n}{s}\binom{s}{k}\sum_{j=0}^{s-k}\binom{s-k}{j}(-1)^{s-k-j}(jk+1)^{n-k-2}.$$

By adding the values of $T_k(n,n-s)$ for $s=k,k+1,\dots,n-2$ we must of course get the known formula (1) (see [2], [3]) for the total number $T_k(n)$ of k-trees of order n. As a matter of fact we get from (19)

$$(20)\qquad T_k(n)=\sum_{s=k}^{n-2}T_k(n,n-s)=$$

$$=\binom{n}{k}\sum_{j=0}^{n-k-2}\binom{n-k}{j}(jk+1)^{n-k-2}(n-k-j-1)(-1)^{n-k-j}.$$

Taking into account that

$$\sum_{j=0}^{n-k-2}\binom{n-k}{j}(jk+1)^{n-k-2}(n-k-j-1)=$$

$$=\frac{1}{k}\left[(k(n-k-1)+1)H^{(n-k-2)}(0)-H^{(n-k-1)}(0)\right]$$

where

$$H(x)=e^x[(e^{kx}-1)^{n-k}+(n-k)e^{k(m-k-1)x}-e^{k(n-k)x}]$$

and we have

$$H^{(n-k-2)}(0)=(n-k)[k(n-k-1)+1]^{n-k-2}-[k(n-k)+1]^{n-k-2}$$

and

$$H^{(n-k-1)}(0)=(n-k)[k(n-k-1)+1]^{n-k-1}-[k(n-k)+1]^{n-k-1}$$

and thus

$$\frac{1}{k}\left[(k(n-k-1)+1)H^{(n-k-2)}(0)-H^{(n-k-1)}(0)\right]=[k(n-k)+1]^{n-k-2}$$

it follows that (1) holds.

Thus as a by-product we obtained another proof of the formula (1). Notice that for $k=1$ (1) reduces to CAYLEY's celebrated formula $T_1(n)=n^{n-2}$ for the total number of ordinary (one-dimensional) trees of order n. (For other proofs of this formula see [6] and [7].)

§ 3. The moments of the distribution of the number of endpoints

Let us first consider the mean value and the variance of the number of endpoints of a k-tree of order n. For the mean value

$$M_k(n) = \frac{1}{T_k(n)} \sum_{r=2}^{n-k} r T_k(n, r) \tag{21}$$

we get from the recursion formula (2) immediately

$$M_k(n) = \frac{n T_k(n-1)}{T_k(n)} [k(n-k-1)+1] \tag{22}$$

and thus

$$M_k(n) = \frac{n-k}{\left(1 + \frac{k}{k(n-k-1)+1}\right)^{n-k-2}}. \tag{23}$$

Thus we have

$$\lim_{n \to +\infty} \frac{M_k(n)}{n} = \frac{1}{e} \quad \text{for} \quad k = 1, 2, 3, \dots. \tag{24}$$

Similarly we get from (2) for the variance

$$D_k^2(n) = \frac{1}{T_k(n)} \sum_{r=2}^{n-k} [r - M_k(n)]^2 T_k(n, r) \tag{25}$$

$$D_k^2(n) = \frac{(n-k-1)(n-k)[(n-k-2)k+1]^{n-k-3}}{[k(n-k)+1]^{n-k-2}} + \tag{26}$$

$$+ \frac{(n-k)[k(n-k-1)+1]^{n-k-3}}{[k(n-k)+1]^{n-k-2}} - \frac{(n-k)^2[k(n-k-1)+1]^{2n-2k-6}}{[k(n-k)+1]^{2n-2k-4}}$$

and thus

$$\lim_{n \to +\infty} \frac{D_k^2(n)}{n} = \frac{1}{e}\left(1 - \frac{2}{e}\right) \qquad (k = 1, 2, \dots). \tag{2. 7}$$

It is remarkable that the asymptotic formulae for $M_k(n)$ and $D_k^2(n)$ do not depend on k, i.e. are the same as those obtained in [5] for $k=1$.

In [5] we have proved that if we choose at random one of the n^{n-2} labeled trees of order n, so that any one of these is chosen with the same probability, then denoting by v_n the number of endpoints of this random tree, the distribution of the random variable $\frac{v_n - M_1(n)}{D_1(n)}$ tends for $n \to +\infty$ to the standard normal distribution. By considering moments of every order one can prove that the same holds for every k, i.e. if we choose at random, with uniform distribution one of the $T_k(n)$ labeled k-trees of order n, and denote the number of its endpoints by $v_n^{(k)}$, then the distribution of $\frac{v_n^{(k)} - M_k(n)}{D_k(n)}$ tends for $n \to +\infty$ to the standard normal distribution.

REFERENCES

[1] HARARY, F. and PALMER, E. M.: On Acyclic simplicial complexes, *Mathematika* **15** (1968) 115—122.

[2] BEINEKE, L. W. and PIPPERT, R. E.: The number of labeled k-dimensional trees, *Journal of Combinatorial Theory* **6** (1969) 200—205.

[3] MOON, J. W.: The number of labeled k-trees, *Journal of Combinatorial Theory* **6** (1969) 196—199.

[4] PALMER, E. M.: On the number of labeled 2-trees, *Journal of Combinatorial Theory* **4** (1969) 206—207.

[5] RÉNYI, A.: Some remarks on the theory of trees, *Publ. Math. Inst. Hung. Acad. Sci.* **4** (1959) 73—85.

[6] MOON, J. W.: *Various proofs of Cayley's formula for counting trees*, Seminar on Graph Theory, ed. by F. Harary.

[7] RÉNYI, A.: On Cayley's polynomials for counting trees, *Proc. of the Calgary Int. Conference on Combinatorial Structures and their Applications* (in print).

Mathematical Institute of the Hungarian Academy of Sciences, Budapest

(Received August 1, 1969.)