ON THE MAXIMUM DEGREE IN A RANDOM TREE

J. W. Moon

1. INTRODUCTION

A *tree* is a connected graph that has no cycles (see, for example, Ore [7] as a general reference on graph theory). The *degree* of a node x in a tree is the number d(x) of edges joining x to other nodes. Since any tree T_n with n labelled nodes has exactly n - 1 edges, it follows that the average value of d(x), as x runs over the nodes of T_n, is 2(1 - 1/n). Let $D = D(T_n)$ denote the maximum degree of nodes in the tree T_n, that is, let $D(T_n) = \max\{d(x): x \in T_n\}$. Our object here is to derive, by elementary and crude arguments, an asymptotic formula for the average value of $D(T_n)$ over the set of the n^{n-2} trees T_n with n labelled points.

2. PRELIMINARY RESULTS

We first list some results that we shall use later. (In what follows, n and k will always denote integers such that $1 \leq k \leq n - 1$.)

LEMMA 1. *If the integers* d(i) (i = 1, 2, ⋯, n) *form a decomposition of* 2(n - 1), *then there exist*

$$\binom{n-2}{d(1)-1, \cdots, d(n)-1}$$

trees T_n *with* n *labelled nodes, the* i^{th} *node having degree* d(i).

This has been proved by Moon [5], [6] and Riordan [8].

LEMMA 2. *There are* $\binom{n-2}{k-1}(n-1)^{n-k-1}$ *trees* T_n *with* n *labelled nodes in which* d(x) = k *for each node* x.

This was first proved by Clarke [1]; it follows easily from Lemma 1.

LEMMA 3. *If* $k = \left[\frac{(1+\varepsilon)\log n}{\log\log n}\right]$, *then* $\frac{n}{k!} < n^{-\varepsilon+o(1)}$ *as* $n \to \infty$, *for any positive constant* ε.

LEMMA 4. *If* $k = \left[\frac{(1-\varepsilon)\log n}{\log\log n}\right]$, *then* $\frac{n}{k!} > n^{\varepsilon+o(1)}$ *as* $n \to \infty$, *for any positive constant* ε.

LEMMA 5. *If* $k = [\log n]$, *then* $\frac{n}{k!} < n^2 \log n / n^{\log\log n}$ *for all sufficiently large values of* n.

Received July 15, 1968.
This paper was completed during the Summer Research Institute of the Canadian Mathematical Congress at Queen's University.

The derivations of these three lemmas make use of the inequalities $(k/e)^k < k! < k^k$ and $t < -\log(1 - t) < t/(1 - t)$, where $0 < t < 1$; we omit the details, since they are straightforward.

3. AN UPPER BOUND FOR D

Suppose we pick a tree T_n at random from the set of the n^{n-2} trees with n labelled nodes and consider the degree of an arbitrary node x. It follows from Lemma 2 that

$$\Pr\{d(x) = k\} = \binom{n-2}{k-1} \frac{(n-1)^{n-k-1}}{n^{n-2}} = \frac{(1 - 1/n)^n}{(k-1)!} \cdot \frac{n^2}{(n-1)^2} \cdot \frac{(n-2)_{k-1}}{(n-1)^{k-1}}$$

$$< \frac{e^{-1}}{(k-1)!}, \quad \text{if } k \geq 3.$$

Therefore,

$$\Pr\{d(x) > k\} < e^{-1}\left(\frac{1}{k!} + \frac{1}{(k+1)!} + \cdots\right) < \frac{e^{-1}}{k!}\left(1 + \frac{1}{k+1} + \frac{1}{(k+1)^2} + \cdots\right)$$

$$= \frac{e^{-1}}{k!}\left(1 + \frac{1}{k}\right) < \frac{1}{k!}, \quad \text{if } k \geq 2.$$

The following result now follows from Boole's inequality

$$\Pr\left\{\bigcup E_i\right\} \leq \sum \Pr\{E_i\}$$

(the case $k = 1$ is obvious).

THEOREM 1.

$$\Pr\{D > k\} \leq \frac{n}{k!}.$$

COROLLARY 1. *If* ε *denotes a positive constant, then*

$$D \leq (1 + \varepsilon)\frac{\log n}{\log \log n}$$

for almost all trees T_n, *that is, for all but a fraction that tends to zero as* n *tends to infinity.*

This corollary follows from Theorem 1 and Lemma 3.

4. A LOWER BOUND FOR D

If $t(n, k)$ denotes the number of trees T_n for which $D(T_n) \leq k$, then it follows from Lemma 1 that

$$t(n, k) = (n-2)! \times \text{the coefficient of } z^{n-2} \text{ in } \left\{1 + z + \frac{z^2}{2!} + \cdots + \frac{z^{k-1}}{(k-1)!}\right\}^n.$$

Hence,

$$t(n, k) < (n-2)! \left\{1 + 1 + \frac{1}{2!} + \cdots + \frac{1}{(k-1)!}\right\}^n < (n-2)! \left\{e - \frac{1}{k!}\right\}^n$$

$$< cn^{n-3/2} \left\{1 - \frac{1}{e \cdot k!}\right\}^n < cn^{n-3/2} \cdot \exp(-n/e \cdot k!),$$

for some constant c $(c < 2e)$, by Stirling's formula (see, for example, Feller [2, p. 52]). If we divide both sides of this inequality by n^{n-2}, the total number of trees T_n, we obtain the following result.

THEOREM 2.

$$\Pr\{D \leq k\} < cn^{1/2} \exp(-n/e \cdot k!).$$

COROLLARY 2. *If* ε *denotes a positive constant, then*

$$D > (1 - \varepsilon) \frac{\log n}{\log \log n}$$

for almost all trees T_n.

This corollary follows from Theorem 2 and Lemma 4.

5. THE EXPECTED VALUE OF D

Let ε be any positive constant; if $E(D)$ denotes the expected value of $D(T_n)$, taken over the set of the n^{n-2} trees T_n, then

$$E(D) = \sum_{k=1}^{n-1} k \Pr\{D = k\}.$$

Hence,

$$E(D) \leq k_1 \Pr\{D \leq k_1\} + k_2 \Pr\{D > k_1\} + (n-1) \Pr\{D > k_2\},$$

where

$$k_1 = \left[(1 + \varepsilon) \frac{\log n}{\log \log n} \right] \quad \text{and} \quad k_2 = [\log n];$$

therefore, it follows from Theorem 1 and Lemmas 3 and 5 that

$$E(D) \leq (1 + \varepsilon) \frac{\log n}{\log \log n} + (\log n) n^{-\varepsilon + o(1)} + \frac{n^3 \log n}{n^{\log \log n}}$$

$$= (1 + \varepsilon + o(1)) \frac{\log n}{\log \log n}, \quad \text{as } n \to \infty.$$

Also,

$$E(D) \geq (1 - \varepsilon) \frac{\log n}{\log \log n} \Pr\left\{D > (1 - \varepsilon) \frac{\log n}{\log \log n}\right\}$$

$$\geq (1 - \varepsilon - o(1)) \frac{\log n}{\log \log n}, \quad \text{as } n \to \infty,$$

by Corollary 2. Since ε can be arbitrarily small, the two preceding inequalities imply the following result.

THEOREM 3.

$$E(D) \sim \frac{\log n}{\log \log n} \quad \textit{as} \ n \to \infty .$$

By using sharper estimates for $n/k!$ in Theorems 1 and 2, one can show that

$$(1 - \varepsilon)\frac{\log n}{\log \log n} \cdot \frac{\log \log \log n}{\log \log n} < D(T_n) - \frac{\log n}{\log \log n} < (1 + \varepsilon)\frac{\log n}{\log \log n} \cdot \frac{\log \log \log n}{\log \log n},$$

for almost all trees T_n and each positive constant ε.

In view of Lemma 1, the problem we have considered here is but a special case of the following problem (see [3] and [4]). If m different balls are randomly distributed among n different boxes, what is the expected number of balls in the box containing the maximum number of balls? We leave it to the reader to decide what restrictions on the relative sizes of m and n are necessary for the preceding argument to work in the more general case.

REFERENCES

1. L. E. Clarke, *On Cayley's formula for counting trees*. J. London Math. Soc. 33 (1958), 471-474.

2. W. Feller, *An introduction to probability theory and its applications*. Vol. I, Second Edition. John Wiley and Sons, Inc., New York, 1957.

3. I. J. Good, *Saddle-point methods for the multinomial distribution*. Ann. Math. Statist. 28 (1957), 861-881.

4. R. E. Greenwood and M. O. Glasgow, *Distribution of maximum and minimum frequencies in a sample drawn from a multinomial distribution*. Ann. Math. Statist. 21 (1950), 416-424.

5. J. W. Moon, *The second moment of the complexity of a graph*. Mathematika 11 (1964), 95-98.

6. ———, *Enumerating labelled trees*. Graph Theory and Theoretical Physics (F. Harary, Editor), Academic Press, London, 1967, 261-272.

7. O. Ore, *Theory of graphs*. Amer. Math. Soc. Colloquium Publications 38, Providence, R.I., 1962.

8. J. Riordan, *The enumeration of labeled trees by degrees*. Bull. Amer. Math. Soc. 72 (1966), 110-112.

University of Alberta
Edmonton, Alberta, Canada