

Distribution of Nodes of a Tree by Degree*

HWA SUNG NA AND ANATOL RAPOPORT

Mental Health Research Institute

The University of Michigan

Ann Arbor, Michigan

Communicated by Richard Bellman

ABSTRACT

Two types of random trees, "static" and "growing," are studied. The "growing" type of trees is constructed with an ordering bias imposed on its population of N nodes. Exact formulas for the expected number of nodes of degree i are derived in each case. Asymptotic behavior of such formulas is determined for large values of N .

INTRODUCTION

Many situations are representable structurally by graphs; all situations, in fact, where the objects of interest are a set of elements and a binary relation which does or does not obtain between each pair of elements in the set. Well-known examples are sociograms, communication nets, acquaintance nets, and so on.

A question of interest is how the structure of these systems came about. One possible history is the prior existence of a population of elements (points), each of which eventually establishes connections with other elements, chosen by some probabilistic rule from the population. The resulting graph would be a directed one. Or else *pairs* of elements might be selected by a probabilistic rule (say, when they meet or collide), and a symmetric relation is thereby established between them, resulting in an undirected graph. Still another way is by the addition of new elements to

* The research work on which this paper is based was supported by the National Science Foundation, Grant GS-1735, for the period June 1967-June 1969.

the population, whereby the new element establishes a (symmetric) relation with one or more of the elements already present. Depending on the method of generation and the probabilistic rules, the resulting graph will have different statistical properties.

Is it possible to infer the method of generation by examining the statistical properties of the graph? One way of approaching this problem is to assume a method of generation, and to infer the statistical properties induced by it. In what follows, we examine one statistical property of a special type of graph (a tree) under two different assumptions of how the tree "came about."

One such tree is a "growing tree." As an example, consider the following experiment on word association, illustrated in Figs. 1a and 1b.

A subject is given a vocabulary of n words. He is told to select one word arbitrarily, then look through the vocabulary and select a word "most closely associated" with it, which he indicates by joining the second word, w_2 , to the first, w_1 (Fig. 1a):

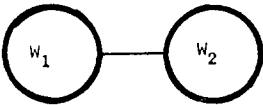


FIG. 1a

Next, he looks through the list to find a word associated *either* with w_1 or with w_2 , which he then joins to the one or to the other. As he continues in this fashion, the tree "grows," as in Fig. 1b.

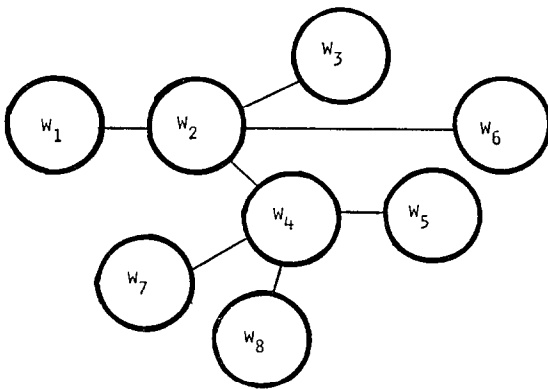


FIG. 1b

Under the null hypothesis, the selection of the word to which to join each new word is random. This method of generation will give rise to a certain distribution of node degrees (derived below). Systematic departures of this distribution would indicate the operation of *biases*. Hypotheses concerning the nature of these biases (e.g., unequal “penumbras of association” of words, or the effect of the order of selection) amount to hypotheses about the word association process. If the corresponding distributions of node degrees can be derived from the assumed biased process, the underlying hypotheses can be put to a test. This method of generation corresponds to what we call below a “growing tree.”

In contrast, in a “static tree” the number of elements, n , is initially given, and pairs of elements are selected at random to be joined by links. If the resulting graph is to be a tree, there must be no cycles. The mode of generating such a tree might be the following. Initially, single elements move about at random. Each collision forms a couple. A collision of a couple with a single element forms a triple, a collision of an s -tuple with a t -tuple forms an $(s + t)$ -tuple, and so on. At each collision a link is established between an element of one X -tuple and an element of another, the links being rigid so that the elements of the same k -tuple cannot collide. The process goes on until the entire set of n elements has been joined into an n -tuple. The null hypothesis states that the probability of collision of any two X -tuples is the same as what it would be if all the elements were “free.”

Let $G(N, h)$ denote a graph of N nodes and h links joining the N nodes. A tree is a special type of connected graph $G(N, N - 1)$, namely

$$h = N - 1. \quad (1)$$

We refer to a tree with N labeled points as a tree of order N .

The distribution of nodes of a tree, T , by degree (defined below) depends on the type of trees of which T is a member. Two types of trees are considered: static trees and growing trees. Differences in results are compared and examined.

The number of links associated with a node in a graph is *the degree* of the node in that graph. Since here we are concerned with undirected graphs, no distinction will be made between in- and out-degrees.

If a specific tree is given, then the degrees of each of its nodes are fixed. However, if a set, S , of trees is given (or, when only the rules for constructing such graphs are specified), one can only speak of the “expected” degrees of the nodes of any arbitrary member T in the set. That

is, how many nodes of T from S are expected to be of degree i , with $i = 1, 2, \dots, N - 1$?

A. Static trees

Consider a set of N labeled points with $N - 1$ links joining them. From such a set a collection U_1 of N^{N-2} distinct trees, all of order N , can be formed. Trees of this type are called "static" as the number of their nodes is arbitrarily given [1]. When such a tree is arbitrarily chosen from U_1 , how many nodes of this tree can be expected to be of degree i ($i = 1, 2, \dots, N - 1$)?

Let $f(i, N)$ denote the expected number of nodes of degree i ($i = 1, 2, \dots, N - 1$) in a tree $T(N)$ of order N .

It has been shown by Rényi [4] that for $i = 1$,

$$f(1, N) = N \left[1 - \frac{1}{N} \right]^{N-2} \quad (2)$$

and

$$f(1, N) \approx \frac{N}{e} \quad (3)$$

for large values of N .

To determine the number of nodes of higher degrees in a static tree, we shall first explain a technique originally developed by Prüfer [3]. This technique establishes a one-to-one correspondence between a tree $T(N)$ of order N and a unique vector R_{N-2} of $N - 2$ components such that each node of degree i ($i = 1, \dots, N - 1$) is associated with $i - 1$ components of the vector. The following example describes the technique briefly.

Consider a tree $T(7)$ of seven labeled points, $P_1, P_2, P_3, P_4, P_5, P_6$, and P_7 , as shown in Fig. 2a.

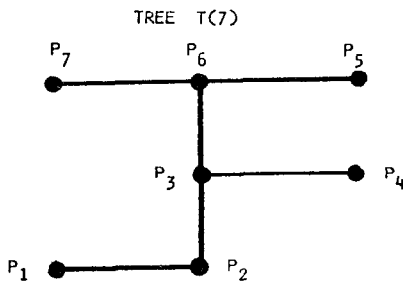


FIG. 2a

Procedure 1. Let $E(N)$ be the set of nodes in the tree $T(N)$, with degree 1; for example,

$$E(7) = \{P_1, P_4, P_5, P_7\}.$$

Procedure 2. Remove from the tree $T(N)$ the node P_{i_0} in $E(N)$ with the least index. (In the present example, Fig. 2a, remove P_1 .)

Procedure 3. Remove now the link $\overline{P_{i_0}P_{i_1}}$ associated with the node which has just been removed. (In this example, it is $\overline{P_1P_2}$.)

Since the node removed is of degree 1, there is exactly one link associated with it. Hence, when this link is removed in Procedure 3, P_{i_1} is the only node in tree $T(N)$ to be affected. The degree of P_{i_1} is reduced by one.

Procedure 4. The index i_1 of the node affected by the process of removal serves as the first component of vector R_{N-2} . (For the present example, $i_1 = 2$.)

Procedure 5. The resulting tree then has only $N - 1$ nodes, which we designate $T(N - 1)$.

Procedure 6. To obtain the second and subsequent components of R_{N-2} , repeat Procedures 1 through 5 on $T(N - 1)$.

Procedure 7. Stop the process when there is only one link (and two nodes) left in the graph.

Since only $N - 2$ nodes, and $N - 2$ links, are removed from $T(N)$, and each removal produces one component for the vector R_{N-2} , R_{N-2} has exactly $N - 2$ components. For the tree in Fig. 2a, its corresponding vector R_5 is

$$R_5 = \{2, 3, 3, 6, 6\}. \quad (4)$$

Trees $T(6)$, $T(5)$, $T(4)$, $T(3)$, and $T(2)$, resulting from the successive removal of nodes and links from $T(7)$, are shown in Fig. 2b.

Suppose node P_{i_1} is of degree i in $T(N)$. The removal of a link associated with P_{i_1} , as in Procedure 3, reduces its degree from i to $i - 1$ in $T(N - 1)$. Thus it takes $i - 1$ removals to reduce node P_{i_1} to a node of degree 1 in a certain reduced tree $T(K)$ from $T(N)$. Therefore the index i_1 will appear exactly $i - 1$ times in the components of R_{N-2} .

Now, when a vector R_{N-2} of $N - 2$ components is given, the procedure of *constructing* a tree $T(N)$ from R_{N-2} is rather obvious: it can be done by reversing the above described process. Hence it is established that the correspondence between $T(N)$ and R_{N-2} is one-to-one.

In our approach, we now follow the above technique and derive the

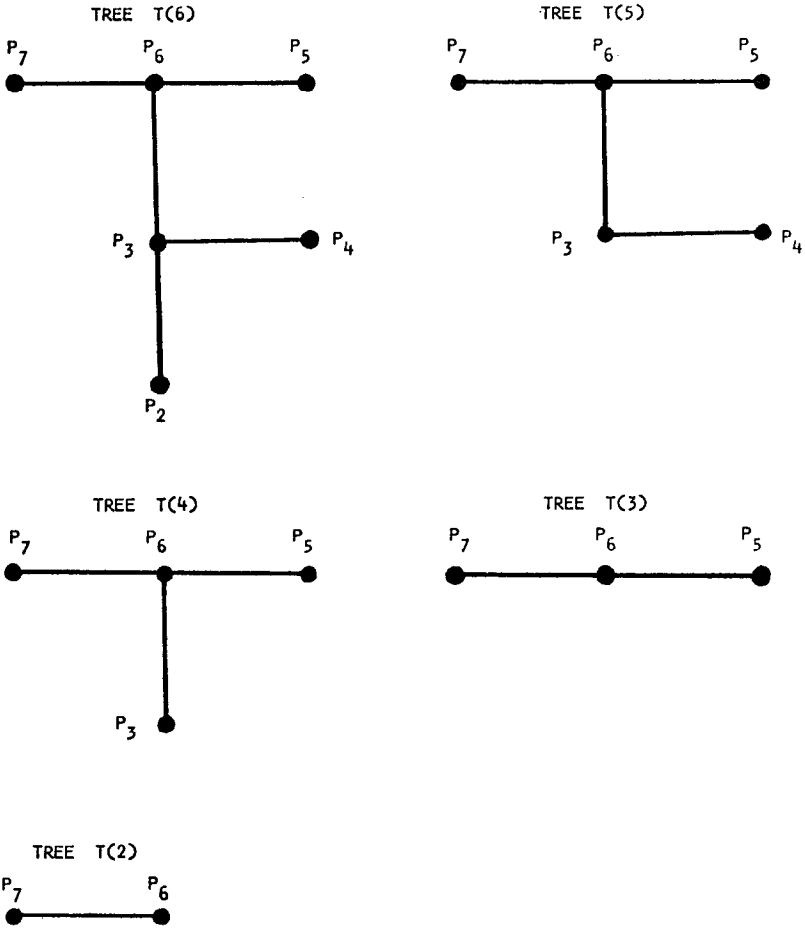


FIG. 2b

formula for $f(i, N), i = 2, 3, \dots, N - 1$, by making use of the coefficients of Bell's polynomials.

Let \mathbf{k} be a vector of $N - 1$ components, $(k_1, k_2, \dots, k_{N-1})$, such that

$$k_i \geq 0, \quad \text{for all } i, \tag{5}$$

and

$$\sum_{i=1}^{N-1} k_i = N. \tag{6}$$

$B_{k_1}(\mathbf{k})$ will be the coefficient associated with the term $g(\mathbf{k})$ in Bell's

polynomial, and

$$\begin{aligned}
 g(\mathbf{k}) &= g_1^{k_2} g_2^{k_3} \cdots g_{N-2}^{k_{N-1}} \\
 &= \prod_{i=1}^{N-2} g_i^{k_{i+1}}.
 \end{aligned}
 \tag{7}$$

The Bell polynomial of $(N - 2)$ th order can now be written as follows:

$$Y_{N-2} = \sum_{k_1=2}^{N-1} f_{k_1}(S_{k_1}),
 \tag{8}$$

where

$$S_{k_1} = \sum_{\mathbf{k} \in K_1} B_{k_1}(\mathbf{k})g(\mathbf{k})
 \tag{9}$$

and K_1 stands for the set of vectors \mathbf{k} such that all \mathbf{k} have the same value k_1 for the first component. (A short table of some Bell polynomials and the recurrence formula for them have been given, e.g., by Riordan [5].)

Note that if we interpret the vector \mathbf{k} as an unrestricted partition of N into $N - 1$ parts, then Bell's number $B_{k_1}(\mathbf{k})$ gives the number of ways of distributing $N - 2$ distinguishable things into $N - k_1$ classes such that precisely k_i classes have $i - 1$ things each, and

$$\sum_{i=1}^{N-2} i \cdot k_{i+1} = N - 2.
 \tag{10}$$

Now, consider a tree $T_{k_1}(N)$ with precisely k_i nodes of degree i , $i = 1, 2, \dots, N - 1$. For convenience, nodes of degree 1 will also be referred to as end points of a tree. And we note that

$$2 \leq k_1 \leq N - 1.
 \tag{11}$$

Since the index r of a node P_r of degree i will appear exactly $i - 1$ times as components of R_{N-2} , the $N - 2$ components of the vector R corresponding to $T_{k_1}(N)$ will not involve the k_1 indices of its end points. ($T_{k_1}(N)$ denotes a tree of order N with exactly k_1 end points.)

If $T_{k_1}^*(N)$ is another tree having exactly the same set, E , of k_1 end points as $T_{k_1}(N)$, then the $N - 2$ components of the vector R^* , corresponding to $T_{k_1}^*(N)$, will use the same $N - k_1$ indices as the vector R . However, the distribution of the $N - 2$ components over these $N - k_1$ indices will be different for the two vectors R and R^* . Thus every distinct distribution of the $N - 2$ components over the same $N - k_1$ indices represents a different member of the collection of trees, U_1 , having the same set E of k_1 end points as $T_{k_1}(N)$. There are $A(k_1)$ such members, including $T_{k_1}(N)$ itself, where

$$A(k_1) = \sum_{\mathbf{k} \in K_1} B_{k_1}(\mathbf{k}) \cdot (N - k_1)!.
 \tag{12}$$

Equation (12) will be self-evident if we consider $B_{k_1}(\mathbf{k}) \cdot (N - k_1)!$ as the number of different vectors R_{N-2} with k_i indices appearing $i - 1$ times each as its components, $i = 2, \dots, N - 1$, and

$$\sum_{i=2}^{N-1} k_i = N - k_1. \tag{13}$$

Now let $d(i, N)$ denote the total number of nodes of degree i in the whole collection U_1 of N^{N-2} trees of order N . We have

$$d(i, N) = \sum_{k_1=2}^{N-1} \binom{N}{k_1} \cdot \sum_{\mathbf{k} \in K_1} k_i \cdot B_{k_1}(\mathbf{k}) \cdot (N - k_1)!. \tag{14}$$

Equation (14) is so because: For a given partition \mathbf{k} of N into $N - 1$ parts, there are $B_{k_1}(\mathbf{k}) \cdot (N - k_1)!$ trees each having exactly k_i nodes of degree i . This gives

$$k_i \cdot (B_{k_1}(\mathbf{k}) \cdot (N - k_1)!)$$

nodes of degree i . Next, for a given set E of k_1 end points, there are $|K_1|$ different partitions \mathbf{k} . Hence we have

$$\sum_{\mathbf{k} \in K_1} k_i \cdot B_{k_1}(\mathbf{k}) \cdot (N - k_1)!$$

nodes of degree i for each set E of k_1 end points. As all the points are labeled, there are

$$\binom{N}{k_1} = \frac{N!}{k_1! (N - k_1)!}$$

different sets of E . Finally, the number of end points for a tree may vary from two to $N - 1$; thus we have to sum over all possible values of k_1 .

Since there are N^{N-2} trees in the collection U_1 , and $d(i, N)$ nodes of degree i in U_1 , a randomly chosen tree from U_1 can be expected to have $f(i, N)$ nodes of degree i , where

$$\begin{aligned} f(i, N) &= \frac{d(i, N)}{N^{N-2}} \\ &= \frac{1}{N^{N-2}} \sum_{k_1=2}^{N-1} \binom{N}{k_1} \cdot \left[\sum_{\mathbf{k} \in K_1} k_i \cdot B_{k_1}(\mathbf{k}) \cdot (N - k_1)! \right] \\ &\text{for } i = 2, 3, \dots, N - 1. \end{aligned} \tag{15}$$

It has been shown by Meir and Moon [2] that when N becomes very large,

$$f(i, N) \approx \frac{N}{e} \frac{1}{(i - 1)!}. \tag{16}$$

B. Growing trees

In “growing” trees, the nodes are not only labeled but also ordered. That is, in the process of constructing these trees, the entire set of n labeled points is not presented at once, as in the case of “static” trees. Instead, these nodes are introduced one at a time. Thus they are arranged in an ordered sequence: namely, the 1st node, the 2nd node, . . . , the n th node. This ordering among nodes can be considered as a *bias* introduced onto (or imposed upon) the population of nodes. The effect of this bias is certain to be reflected in the statistics of the distribution studied.

To derive the distribution of nodes by degrees in this case, it is necessary first to describe in further detail the process of constructing such a “growing” tree.

Assume that a tree $T(n)$ of order n has already been constructed. The $(n + 1)$ th node, P_{n+1} , is then added to the graph $T(n)$ together with a new link L_{n+1} joining P_{n+1} to $T(n)$. L_{n+1} is so placed that P_{n+1} may be joined to any one of the n nodes of $T(n)$ with equal probability. The resulting graph now has $n + 1$ nodes and n links (with $n - 1$ links from the tree $T(n)$ and a new link, L_{n+1}). Thus it is again a tree, but now of order $n + 1$. Hence we have constructed a tree $T(n + 1)$. Repeating the same procedure, trees of any order can be constructed.

Note that, from the above description, two assumptions have been made; namely,

(i) there is a constant probability, α , that the newly introduced node is of degree 1, namely $\alpha = 1$;

(ii) the probability that P_{n+1} will be joined to a node of degree i in $T(n)$ is $f(i, n)/n$, where $f(i, n)$ is the number of expected nodes of degree i in $T(n)$.

These two assumptions describe a stochastic process. Thus trees of the “growing type” are constructed according to a stochastic process.

Furthermore, the process permits node P_n to have $n - 1$ different choices to be joined to $T(n - 1)$. Thus, from a particular realization of $T(n - 1)$, it is possible to construct $n - 1$ different trees of order n . It follows that there are altogether $(n - 1)!$ trees of order n . They form the collection $U_2(n)$.

In what follows, the distribution of nodes by degree for members of U_2 will be derived by first setting up a pair of difference equations, and then finding the asymptotic behavior of the distribution when N becomes very large.

1. The Difference Equations

When node P_{N+1} is added to the graph $T(N)$ to form a new tree $T(N+1)$, it is joined to only one of the N nodes, say V_1 . If node V_1 is originally of degree i in $T(N)$, it becomes a node of degree $i+1$ in $T(N+1)$. Whereas the respective degrees of all the other $N-1$ nodes, V_2, V_3, \dots, V_N , will remain the same in both $T(N)$ and $T(N+1)$; and P_{N+1} is of degree 1 in $T(N+1)$. Thus the number $f(i+1, N+1)$ of the new tree $T(N+1)$ is one *more* than $f(i+1, N)$ of $T(N)$. At the same time, $f(i, N+1)$ is one *less* than $f(i, N)$ except when $i=1$. Furthermore, $f(1, N+1)$ is one more than $f(1, N)$ if $i \neq 1$. For the rest of values $f(j, N)$ and $f(j, N+1)$, they will be the same for all j except for $j=1$ and $j=i$.

In the special case where V_1 is originally of degree 1 in $T(N)$, we have the following situation. Node V_1 becomes of degree 2 in $T(N+1)$. Hence $f(2, N+1)$ is increased by one from $f(2, N)$. However, the number $f(1, N+1)$ now will be the same as $f(1, N)$ because of the addition of P_{N+1} , a degree-one node, to compensate for node V_1 . All the other values $f(j, N+1)$, $j \neq 2$, also remain the same as $f(j, N)$.

Next, the problem is to determine the chance of P_{N+1} being connected to a node of degree exactly i in $T(N)$. According to the equiprobable assumption in laying down L_{N+1} , the probability for V_1 to be of degree i is simply

$$\frac{f(i, N)}{N}.$$

The above argument leads to the two difference equations (17) and (18), which give the expected number of nodes of degree i , $i=1, 2, \dots, N$, in any of a "growing"-type tree of order $N+1$.

$$f(1, N+1) = \left[\frac{f(1, N)}{N} \right] \cdot f(1, N) + \left[1 - \frac{f(1, N)}{N} \right] \cdot [f(1, N) + 1]; \quad (17)$$

$$\begin{aligned} f(i, N+1) &= \left[\frac{f(i, N)}{N} \right] \cdot [f(i, N) - 1] + \left[\frac{f(i-1, N)}{N} \right] \cdot [f(i, N) + 1] \\ &\quad + \left[1 - \frac{f(i, N)}{N} - \frac{f(i-1, N)}{N} \right] \cdot f(i, N) \\ &\text{for } i = 2, 3, \dots, N \end{aligned} \quad (18)$$

The first term on the right-hand side of Eq. (17) takes care of the case where V_1 is a degree-one node in $T(N)$. The second term considers cases where V_1 is of higher degrees in $T(N)$. In Eq. (18), for a particular value i under consideration, the first term on the right-hand side represents situations where V_1 is of degree i in $T(N)$; the second term applies for cases where V_1 is of degree $i - 1$ in $T(N)$; and the last term takes care of every other value of the degree of V_1 in $T(N)$.

Equations (17) and (18) are difference equations of two variables, N and i . By collecting terms, they can be reduced to the following forms:

$$f(1, N + 1) = \frac{N - 1}{N} f(1, N) + 1; \tag{19}$$

$$f(i, N + 1) = \frac{N - 1}{N} f(i, N) + \frac{1}{N} f(i - 1, N),$$

$$\text{for } i = 2, 3, \dots, N. \tag{20}$$

The boundary, or initial, conditions for these equations are

$$f(1, 2) = 2 \tag{21}$$

$$f(i, N) = 0 \quad \text{for } N \leq i. \tag{22}$$

2. Solving for $f(1, N + 1)$ and $f(i, N + 1)$

To solve for Eq. (19) first we write

$$f(1, N) = \frac{N - 2}{N - 1} \cdot f(1, N - 1) + 1 \tag{23}$$

by reducing the argument in Eq. (18) by one. Now, if Eq. (23) is inserted into Eq. (19) it becomes

$$f(1, N + 1) = \frac{N - 2}{N} \cdot f(1, N - 1) + \frac{1}{N} [N + (N - 1)]. \tag{24}$$

Repeating the same procedure, Eq. (24) can, in general, be written as

$$f(1, N + 1) = \frac{N - k}{N} \cdot f(1, N - k + 1) + \frac{1}{N} \left[\sum_{l=0}^{k-1} (N - l) \right], \tag{25}$$

for some k . If we set

$$k = N - 1 \tag{26}$$

and insert Eq. (26) into Eq. (25), it becomes

$$f(1, N + 1) = \frac{1}{N} \cdot f(1, 2) + \frac{1}{N} \left[\sum_{l=0}^{N-2} (N - l) \right]. \tag{27}$$

If we apply the boundary condition in Eq. (21) to Eq. (27), using the well-known summation formula

$$\sum_{k=1}^m k = \frac{m(m+1)}{2}, \quad \text{for } m \geq 1, \tag{28}$$

Eq. (27) becomes

$$f(1, N+1) = \frac{2}{N} + \frac{1}{N} \left[\frac{N(N+1)}{2} - 1 \right], \tag{29}$$

or

$$f(1, N+1) = \frac{N+1}{2} + \frac{1}{N}, \quad N \geq 2. \tag{30}$$

Note that in this study we consider only trees of order 2 or more or both. For the least interesting case where $N = 1$, we simply have

$$f(1, 1) = 0. \tag{31}$$

To solve for Eq. (20) using the same method as for Eq. (19), we have

$$\begin{aligned} f(i, N+1) &= \left(\frac{N-1}{N} \right) \cdot \left[\frac{N-2}{N-1} \cdot f(i, N-1) \right. \\ &\quad \left. + \frac{1}{N-1} \cdot f(i-1, N-1) \right] + \frac{1}{N} \cdot f(i-1, N) \\ &= \frac{N-2}{N} \cdot f(i, N-1) + \frac{1}{N} [f(i-1, N) + f(i-1, N-1)] \\ &= \frac{N-k}{N} \cdot f(i, N-k+1) + \frac{1}{N} \left[\sum_{l=0}^{k-1} f(i-1, N-l) \right] \\ &= \frac{i-1}{N} \cdot f(i, i) + \frac{1}{N} \left[\sum_{K=i}^N f(i-1, K) \right]. \end{aligned} \tag{32}$$

If the boundary condition in Eq. (22) is applied to Eq. (32), it becomes

$$f(i, N+1) = \frac{1}{N} \sum_{K=i}^N f(i-1, K), \quad \text{for } i = 2, \dots, N. \tag{33}$$

In the case of $i = 2$, Eq. (33) gives

$$\begin{aligned} f(2, N+1) &= \frac{1}{N} \sum_{K=2}^N f(1, K) \\ &= \frac{1}{N} \left[\sum_{K=2}^N \left(\frac{K}{2} + \frac{1}{K-1} \right) \right] \\ &= \frac{N+1}{2^2} + S(2, N), \end{aligned} \tag{34}$$

where

$$\begin{aligned}
 S(2, N) &= \frac{1}{N} + \frac{1}{2N} + \cdots + \frac{1}{N(N-1)} - \frac{1}{2N} \\
 &= \sum_{k=1}^{N-1} \frac{1}{kN} - \frac{1}{2N}.
 \end{aligned}
 \tag{35}$$

Similarly, let $i = 3$; then, inserting Eq. (34) into Eq. (33), we obtain

$$\begin{aligned}
 f(3, N + 1) &= \frac{1}{N} \sum_{K=3}^N f(2, K) \\
 &= \frac{1}{N} \left[\sum_{K=3}^N \left(\frac{K}{2^2} + S(2, K - 1) \right) \right] \\
 &= \frac{N + 1}{2^3} + S(3, N),
 \end{aligned}
 \tag{36}$$

where

$$\begin{aligned}
 S(3, N) &= \sum_{K=3}^N \frac{1}{N} \cdot S(2, K - 1) - \frac{3}{4N} \\
 &= \frac{1}{N} \sum_{K=2}^{N-1} \sum_{\alpha=1}^{K-1} \frac{1}{\alpha K} - \frac{1}{2N} \sum_{K=2}^{N-1} \frac{1}{K} - \frac{3}{4N}.
 \end{aligned}
 \tag{37}$$

In general, Eq. (33) can be rewritten as

$$\begin{aligned}
 f(i, N + 1) &= \frac{1}{N} \sum_{K=i}^N \left[\frac{K}{2^{i-1}} + S(i - 1, K - 1) \right] \\
 &= \frac{N + 1}{2^i} + S(i - 1, N),
 \end{aligned}
 \tag{38}$$

where

$$S(i, N) = \frac{1}{N} \sum_{K=i}^N S(i - 1, K - 1) - \frac{i(i - 1)}{2^i \cdot N} \quad \text{for } i = 2, 3, \dots, N;
 \tag{39}$$

and

$$S(1, N) = \frac{1}{N}.
 \tag{40}$$

Note that $S(i, N)$ is not defined if $N < i$.

3. Asymptotic Behavior of $f(i, N)$

LEMMA 1.

$$(a) \quad S(1, N) = \frac{1}{2} \quad \text{for } N = 2; \quad (41)$$

$$(b) \quad S(1, N) \leq \frac{1}{3} \quad \text{for } N \geq 3; \quad (42)$$

$$(c) \quad S(2, N) \leq \frac{1}{3} \quad \text{for } N \geq 2; \quad (43)$$

$$(d) \quad S(i, N) < \frac{N - i + 1}{3N} \quad \text{for } i \geq 3 \text{ and } N \geq 3. \quad (44)$$

$$(e) \quad \lim_{N \rightarrow \infty} S(1, N) = 0 \quad (45)$$

$$(f) \quad \lim_{N \rightarrow \infty} S(2, N) = 0 \quad (46)$$

Proof. (a) By definition

$$S(1, 2) = \frac{1}{2}.$$

(b) By definition

$$S(1, N) = \frac{1}{N} \leq \frac{1}{3} \quad \text{for } N \geq 3.$$

(c)

$$S(2, 2) = \frac{1}{2} - \frac{1}{2 \cdot 2} = \frac{1}{4} < \frac{1}{3};$$

$$S(2, 3) = \frac{1}{3};$$

$$S(2, 4) = \frac{1}{4}(1 + \frac{1}{3}) = \frac{1}{3};$$

for all $N > 4$,

$$\begin{aligned} S(2, N) &= \sum_{k=1}^N \frac{1}{kN} - \frac{1}{2N} \\ &= \frac{1}{N} \left[1 + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{N-1} \right] \\ &< \frac{1}{N} \left[1 + \frac{1}{3} + \frac{1}{3} + \cdots + \frac{1}{3} \right] \\ &< \frac{N}{N} \cdot \frac{1}{3}. \end{aligned}$$

Therefore, $S(2, N) \leq \frac{1}{3}$ for all $N \geq 2$.

(d) Let $i = 3$,

$$\begin{aligned} S(3, N) &= \frac{1}{N} \sum_{k=2}^{N-1} S(2, k) - \frac{3}{4N} \\ &< \frac{1}{N} \sum_{k=2}^{N-1} \frac{1}{3} - \frac{3}{4N} \\ &< \frac{N-2}{N} \cdot \frac{1}{3} \quad \text{for all } N. \end{aligned}$$

Thus, for $i = 3$,

$$S(i, N) < \frac{N - i + 1}{3N} \text{ is true.}$$

Then, by induction,

$$\begin{aligned} S(i + 1, N) &= \frac{1}{N} \sum_{k=i}^{N-1} S(i, k) - \frac{i(i + 1)}{2^{i+1} \cdot N} \\ &< \frac{1}{N} \sum_{k=i}^{N-1} \frac{K - i + 1}{3K} \\ &< \frac{1}{N} \sum_{k=i}^{N-1} \frac{1}{3} \\ &< \frac{N - i + 1}{N3} \quad \text{for all } N. \end{aligned}$$

(e) $\lim_{N \rightarrow \infty} S(1, N) = \lim_{N \rightarrow \infty} \frac{1}{N} = 0.$

(f) $\lim_{N \rightarrow \infty} S(2, N) = \lim_{N \rightarrow \infty} \left\{ \frac{1}{N} \sum_{k=1}^{N-1} \frac{1}{k} - \frac{1}{2N} \right\}$
 $\cong \lim_{N \rightarrow \infty} \left\{ \frac{1}{N} \int_1^N \frac{dk}{k} - \frac{1}{2N} \right\}$
 $= \lim_{N \rightarrow \infty} \left(\frac{\ln N}{N} - \frac{1}{2N} \right) = 0.$

Thus, we have

$$\begin{aligned} f(i, N + 1) &= \frac{N + 1}{2^i} + S(i, N) \\ &\leq \frac{N + 1}{2^i} + \frac{1}{3} \end{aligned} \tag{47}$$

for all i and N except when $i = 1$ and $N = 2$; in that case, we have

$$f(1, 3) = \frac{N + 1}{2^i} + \frac{1}{2} = 2. \tag{48}$$

Hence, for large values of N , the distribution is approximately geometric:

$$f(i, N + 1) \approx \frac{N + 1}{2^i}. \tag{49}$$

Furthermore, for finite values of N , the error introduced by using Eq. (49) instead of Eq. (30) and Eq. (38) is at most $\frac{1}{3}$ except when $N = 2$. For $N = 2$, the error becomes $\frac{1}{2}$ for $f(1, N + 1)$ and $\frac{1}{4}$ for $f(2, N + 1)$. However, this is the case of least interest in the present work, because, for $N = 2$, all the $f(i, N + 1)$'s with $i \geq 3$ are zero.

CONCLUSIONS

To summarize, we have established the asymptotic behavior of $f(i, N)$ as follows. For large values of N ,

$$f(i, N) = \begin{cases} \frac{1}{(i - 1)!} \cdot \frac{N}{e} & \text{for static trees,} \\ \frac{N}{2^i} & \text{for growing trees.} \end{cases}$$

In order to further illustrate its significance, a simple example will be added in which three labeled points, A , B , and C , are used to form trees.

In the case of growing trees, node B is joined to A first, and then connecting node C to the tree AB . Here only two trees, T_1 and T_2 , of order 3 may be obtained, as shown in Fig. 2c.

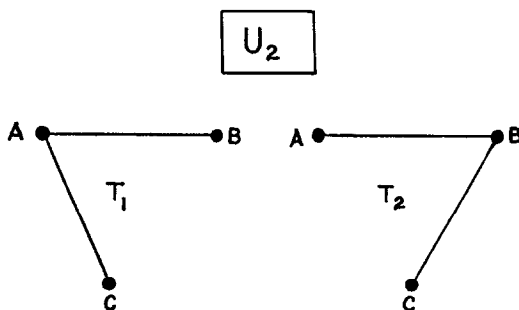


FIG. 2c

However, for the static type, there will be three trees: T_1 , T_2 , and T_3 , as shown in Fig. 2d.

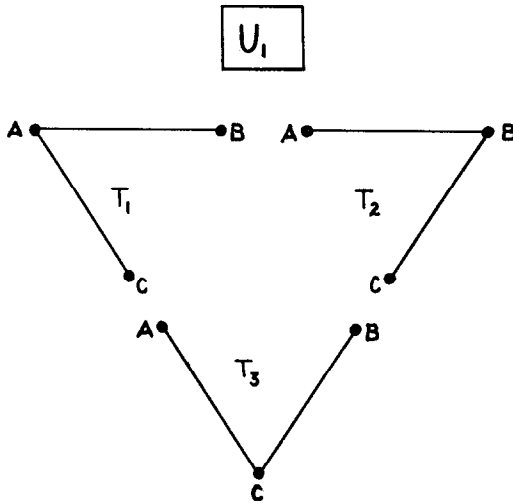


FIG. 2d

The above illustrations demonstrate how the introduction of ordered bias among the nodes operates to influence the distributions for the two types of trees. That is, as we have seen, when $i = 1$, we have

$$f(1, N) \cong \begin{cases} N/e & \text{for static trees;} \\ N/2 & \text{for growing trees.} \end{cases}$$

Thus it is shown that T_3 can never occur in the collection U_2 for the following reason: when the nodes are ordered, the links associated with them are ordered also. Thus, link \overline{AB} comes “before” link \overline{AC} or \overline{BC} . Consequently, \overline{AB} cannot be removed or replaced by either \overline{AC} or \overline{BC} . Yet this is precisely what is required to be able to construct T_3 .

REFERENCES

- 1 A. Cayley, A theorem on trees, *Quart. J. Math.* **23**(1889). 376–378; *Collected Papers*, Cambridge, **13**(1897), 26–28.
- 2 A. Meir and J. W. Moon, (1968). On nodes of degree two in random trees, *Mathematika* **15**(1968), 188–192.
- 3 A. Prüfer, Neuer Beweis eines Satzes über Permutationen, *Archiv für Mathematik Physik* **27**(1918), 122–142.
- 4 A. Rényi, Some remarks on the theory of trees, *Publ. Math. Inst. Hung. Acad. Sci.* **4**(1959), 73–85.
- 5 J. Riordan, *An introduction to combinatorial analysis*. Wiley, New York, 1958.

Mathematical Biosciences **6** (1970), 313–329