MATHEMATICS

THE EXPECTED NODE-INDEPENDENCE NUMBER OF RANDOM TREES

BY

A. MEIR AND J. W. MOON

(Communicated by Prof. N. G. DE BRUIJN at the meeting of February 24, 1973)

§ 1. Summary

We shall derive a formula for the expected value $\mu(n)$ of the node-independence number of a random tree with n labelled nodes and we shall determine the asymptotic behaviour of $\mu(n)$ as n tends to infinity.

§ 2. Introduction

A subset S of nodes of a graph G is *independent* if no two nodes of S are joined by an edge in G. The *node-independence number* of G is the number $N(G)$ of nodes in any largest independent subset of nodes of G. There are n^{n-2} trees T with n labelled nodes; let $\mu(n)$ denote the expected value of $N(T)$ over the set of such trees. (For definitions not given here see [1] or [2].) Our main object is to show that

$$\mu(n) = \sum_{k=1}^{n} \binom{n}{k} \left(\frac{-k}{n}\right)^{k-1},$$

for $n = 1, 2, \ldots$, and that

$$\mu(n)/n \to \varrho,$$

as $n \to \infty$, where $\varrho = .5671 \ldots$ is the unique solution of the equation $xe^x = 1$. We also give an estimate for the variance of $N(T)$ which is used to show that

$$Pr\left\{\left|\frac{N(T)}{n} - \varrho\right| < \varepsilon\right\} \to 1$$

as $n \to \infty$ for any fixed positive ε.

§ 3. Some lemmas

Let T denote a tree that is rooted at some node r. If every set of $N(T)$ independent nodes of T contains r we say T is a type I tree; if at least one set of $N(T)$ independent nodes of T does not contain r we say T is a type II tree. If we remove the root r of T we obtain a (possibly empty) collection of rooted trees $U_1, \ldots, U_j$ whose roots were originally joined to r. The proof of the following lemma is a straightforward exercise.

LEMMA 1. If each of the rooted trees $U_1, \ldots, U_j$ is a type II tree, then T is a type I tree and

$$N(T) = 1 + \sum_{i=1}^{j} N(U_i); \tag{1}$$

if at least one of the rooted trees $U_1, \ldots, U_j$ is a type I tree, then T is a type II tree and

$$N(T) = \sum_{i=1}^{j} N(U_i). \tag{2}$$

Let $y_{k,n}$ denote the number of rooted trees T with n labelled nodes such that $N(T)=k$; let $g_{k,n}$ and $f_{k,n}$ denote the number of these trees that are of types I and II, respectively. Consider the generating functions

$$Y = Y(z, x) = \sum_{n=1}^{\infty} \left(\sum_{k=1}^{n} y_{k,n} z^k \right) \frac{x^n}{n!},$$

$$G = G(z, x) = \sum_{n=1}^{\infty} \left(\sum_{k=1}^{n} g_{k,n} z^k \right) \frac{x^n}{n!},$$

and

$$F = F(z, x) = \sum_{n=1}^{\infty} \left(\sum_{k=1}^{n} f_{k,n} z^k \right) \frac{x^n}{n!}.$$

Since $g_{k,n} + f_{k,n} = y_{k,n}$ and $\sum_k y_{k,n} = n^{n-1}$, the number of rooted trees with n labelled nodes, it follows that these series converge for $|z| \leqslant 1$ and $|x| \leqslant e^{-1}$ and that

$$Y = G + F. \tag{3}$$

LEMMA 2. The functions G, F and Y satisfy the relations

$$G = zxe^F, \tag{4}$$

$$F = x(e^G - 1)e^F, \tag{5}$$

and

$$Y = xe^Y + (z-1)xe^F. \tag{6}$$

PROOF. Suppose the rooted tree T is a type I tree; then each of the rooted trees $U_1, \ldots, U_j$ obtained by removing the root of T is a type II tree by Lemma 1. The generating function for families of j rooted type II trees is $F^j/j!$ for $j = 0, 1, \ldots$. It follows, therefore, that

$$G = zx\{1 + F + F^2/2! + \ldots\} = zxe^F;$$

the factor x is present to account for the root of T and the factor z is present because of equation (1).

If T is a type II tree then at least one of the rooted trees $U_1, \ldots, U_j$ must be a type I tree; the generating function for non-empty families of type I trees is

$$G + G^2/2! + \ldots = e^G - 1.$$

There may or there may not be some type II trees among the trees $U_1, \ldots, U_j$ this time; the generating function for (possibly empty) families of type II trees is e^F, as before. These observations imply that

$$F = x(e^G - 1)e^F;$$

the factor x is again present to account for the root of T, but because of equation (2) the factor z is not included here. Since $Y = G + F$ it now follows from equations (4) and (5) that

$$Y = xze^F + xe^{G+F} - xe^F = xe^Y + (z-1)xe^F,$$

as required.

If we set $y = y(x) = Y(1, x)$, $g = g(x) = G(1, x)$ and $f = f(x) = F(1, x)$, then

$$y = \sum_{n=1}^{\infty} n^{n-1} \frac{x^n}{n!} \tag{7}$$

and

$$y = xe^y, \tag{8}$$

appealing to the definition of Y and relation (6). (Relation (8) is well known; see [2; p. 26].) If follows from (4) that

$$g = xe^f \tag{9}$$

and since $y = f + g$, by (3), it follows from (8) and (9) that

$$y = xe^{f+g} = ge^g. \tag{10}$$

Relations (7)–(10) are valid for $|x| \leqslant e^{-1}$.

LEMMA 3. Let

$$M(x) = \sum_{n=1}^{\infty} \mu(n) n^{n-1} \frac{x^n}{n!};$$

then

$$M(x) = \frac{g(x)}{1 - y(x)}$$

for $|x| < e^{-1}$.

PROOF. Since $\mu(n) = n^{1-n} \sum_k k y_{k,n}$ it follows that $M(x) = (\partial Y/\partial z)_{z=1}$. If we differentiate both sides of equation (6) with respect to z, set $z = 1$, and appeal to equations (8) and (9), we obtain the required result.

§ 4. A FORMULA FOR $\mu(n)$

THEOREM 1. Let $n=1, 2, \ldots$; then

$$\mu(n)=\sum_{k=1}^{n}\binom{n}{k}\left(\frac{-k}{n}\right)^{k-1}.$$

PROOF. Since $y=ge^g$ for $|x|\leqslant e^{-1}$ and $y=0(x)$ as $x\to 0$, the first form of Lagrange's inversion formula (see [3; p. 125, exercise 206]) implies that

$$g=\sum_{k=1}^{\infty}(-k)^{k-1}\frac{y^k}{k!}$$

for sufficiently small values of x. Furthermore, since $y=xe^y$ for $|x|\leqslant e^{-1}$ the second form of Lagrange's formula (see [3; p. 125, exercise 207]) implies that

$$\frac{y^k}{1-xe^y}=\sum_{n=k}^{\infty}n^{n-k}\frac{x^n}{(n-k)!} \tag{11}$$

for $k=0, 1, \ldots$ and sufficiently small values of x. Hence,

$$M(x)=\frac{g}{1-y}=\frac{g}{1-xe^y}=\sum_{k=1}^{\infty}\frac{(-k)^{k-1}}{k!}\sum_{n=k}^{\infty}n^{n-k}\frac{x^n}{(n-k)!}$$
$$=\sum_{n=1}^{\infty}n^{n-1}\frac{x^n}{n!}\sum_{k=1}^{n}\binom{n}{k}\left(\frac{-k}{n}\right)^{k-1}.$$

This suffices to complete the proof of the theorem.

§ 5. THE ASYMPTOTIC BEHAVIOUR OF $\mu(n)$

THEOREM 2. Let ϱ denote the unique solution of the equation $xe^x=1$ so that $\varrho=.5671\ldots$; then

$$\mu(n)=\varrho n+0(n^{1/2}),$$

as $n\to\infty$.

PROOF. Let a_n denote the fraction of rooted trees with n labelled nodes that are type I trees so that

$$g(x)=\sum_{n=1}^{\infty}a_n n^{n-1}\frac{x^n}{n!};$$

clearly, $0\leqslant a_n\leqslant 1$. It follows from (7) that $y(e^{-1})$ is finite and from the relation $y=xe^y$ that $y(e^{-1})=1$; furthermore, since $y=ge^g$ it follows that $g(e^{-1})=\varrho$. Thus if we set $\alpha_n=a_n n^{n-1}/n!$, then

$$\sum_{n=1}^{\infty}\alpha_n e^{-n}=\varrho \tag{12}$$

and there is an absolute constant c_1 such that

$$\alpha_n e^{-n} < c_1 n^{-3/2} \tag{13}$$

for $n=1, 2, \dots$.

If we set $k=0$ in equation (11) we find that

$$(1-y(x))^{-1} = \sum_{n=0}^{\infty} n^n \frac{x^n}{n!}$$

where we adopt the convention that $0^0=1$. (This expansion also follows from the fact that $(1-y)^{-1}=1+xy'$ which may be established by differentiating the relation $y=xe^y$.) If we substitute these expansions for $g(x)$ and $(1-y(x))^{-1}$ in the formula for $M(x)$ given in Lemma 3 and then equate coefficients of x^n, we find that

$$\frac{\mu(n)}{n} = \sum_{k=1}^{n} \alpha_k e^{-k} \beta_{k,n} \tag{14}$$

where

$$\beta_{k,n} = \frac{n!}{n^n} \frac{(n-k)^{n-k}}{(n-k)!} e^k.$$

(Notice that $\beta_{k,n} > 1$ when $1 \leqslant k \leqslant n$ so

$$\frac{\mu(n)}{n} > \sum_{k=1}^{n} \alpha_k e^{-k}$$

for all n.)

It follows from Stirling's formula that there exist absolute constants c_2 and c_3 such that

$$\beta_{k,n} < c_2 n^{1/2} (n-k+1)^{-1/2} \tag{15}$$

when $1 \leqslant k \leqslant n$, and

$$|\beta_{k,n} - 1| < c_3 k n^{-1} \tag{16}$$

when $1 \leqslant k \leqslant \frac{1}{2}n$. Hence,

$$\sum_{k \leqslant \frac{1}{2}n} \alpha_k e^{-k} \beta_{k,n} = \sum_{k \leqslant \frac{1}{2}n} \alpha_k e^{-k} + 0(n^{-1}) \sum_{k \leqslant \frac{1}{2}n} k^{-1/2} = \varrho + 0(n^{-1/2}) \tag{17}$$

by (16), (12) and (13); furthermore,

$$\sum_{\frac{1}{2}n < k \leqslant n} \alpha_k e^{-k} \beta_{k,n} = 0(n^{-1}) \sum_{\frac{1}{2}n < k \leqslant n} (n-k+1)^{-1/2} = 0(n^{-1/2}) \tag{18}$$

by (13) and (15). It now follows from (14), (17) and (18) that

$$\mu(n)/n = \varrho + 0(n^{-1/2})$$

as required.

The entries in the following table were obtained using Theorem 1 and were verified by examining the diagrams of the trees with up to ten nodes given in [1; pp. 233–234].

TABLE: Values of $\mu(n)/n$

n	1	2	3	4	5	6	7	8	9	10
$\mu(n)/n$	1	.5	.6667	.5625	.6080	.5748	.5894	.5770	.5818	.5766

§ 6. THE VARIANCE OF $N(T)$

Let $\lambda(n)$ denote the second factorial moment of $N(T)$ over the set of the n^{n-2} trees T with n labelled nodes and let $\sigma^2(n)$ denote the variance of $N(T)$ over the same set. If we differentiate both sides of equation (6) with respect to z twice, set $z=1$, and appeal to the relations that have been established for Y, G and F, we find after some calculations that

$$\sum_{n=1}^{\infty} \lambda(n) n^{n-1} \frac{x^n}{n!} = \left(\frac{\partial^2 Y}{\partial z^2}\right)_{z=1}$$

$$= \frac{yg^2}{(1-y)^3} + \frac{2yg^2}{(1+g)\ (1-y)^2}$$

$$= (x^2 y'' + xy') g^2 + \frac{2g^2}{y(1+g)} (xy')^2.$$

Using this identity, equation (12), and Stirling's formula, we can show by an elementary but tedious argument that

$$\lambda(n) = \varrho^2 n^2 + 0(n^{3/2}) \tag{19}$$

as $n \to \infty$; we omit the details of his argument. Relation (19) and Theorem 2 imply that

$$\sigma^2(n) = \lambda(n) + \mu(n) - \mu^2(n) = 0(n^{3/2}). \tag{20}$$

The following result now follows immediately from Theorem 2, equation (20), and Chebyshev's inequality.

THEOREM 3. If $N(T)$ denotes the node-independence number of a tree T chosen at random from the set of the n^{n-2} trees with n labelled nodes, then

$$Pr\left\{\left|\frac{N(T)}{n} - \varrho\right| < \varepsilon\right\} \to 1$$

as $n \to \infty$ for any fixed position ε.

§ 7. INDEPENDENT EDGES IN A RANDOM TREE

A subset of edges of a graph G is *independent* if no two of the edges have a node in common. The *edge-independence number* of G is the number

$E(G)$ of edges in any largest independent subset of edges of G. If T is any tree with n nodes, then theorems of Gallai and König (see [1; pp. 95–96]) imply that

$$N(T)+E(T)=n.$$

The following result now follows from Theorems 1 and 2.

THEOREM 4. If $\nu(n)$ denotes the expected value of $E(T)$ over the set of the n^{n-2} trees T with n labelled nodes, then

$$\nu(n)=-\sum_{k=2}^{n}\binom{n}{k}\left(\frac{-k}{n}\right)^{k-1}$$

for $n=2, 3, \ldots,$ and

$$\nu(n)/n \to 1-\varrho=.4329\ldots$$

as $n \to \infty$.

§ 8. ACKNOWLEDGEMENTS

The authors are indebted to Professor N. G. de Bruijn for suggestions that led to shorter proofs of Theorems 1 and 2 than the ones they originally gave; in particular, the idea of expressing the node-independence number of T in terms of the node-independence numbers of the subtrees attached to the root of T and the results in § 3 are due to him.

The preparation of this paper was assisted by grants from the National Research Council of Canada.

University of Alberta

REFERENCES

1. HARARY, F., Graph Theory, Addison–Wesley, Reading, 1969.
2. MOON, J. W., Counting Labelled Trees, Canadian Mathematical Congress, Montreal, 1970.
3. PÓLYA, G. and G. SZEGÖ, Aufgaben und Lehrsätze aus der Analysis, I, Springer, Berlin, 1964.