JOURNAL OF RESEARCH of the National Bureau of Standards — B. Mathematical Sciences
Vol. 78B, No. 4, October–December 1974

Maximizing the Number of Spanning Trees in a Graph With *n* Nodes and *m* Edges*

D. R. Shier

Institute for Basic Standards, National Bureau of Standards, Washington, D.C. 20234

(September 4, 1974)

The problem considered is that of determining, among all graphs on n nodes and m edges, those having the maximum number of spanning trees. The possible candidate graphs can be obtained by deleting some number k of edges from a complete n-node graph. For $k \leqslant n/2$, it is shown that the maximum occurs when the k edges are mutually nonadjacent.

Key words: Combinatorial analysis; enumeration; graphs; maximization; spanning trees; trees

Consider the class of undirected graphs having n nodes and m edges. The problem to be addressed here is that of finding specific configurations of m edges on the given n nodes so that the resulting graph will contain the largest number of spanning trees. In particular, an explicit solution to this problem will be exhibited for graphs which have "enough" edges.

To be specific, let the set E of k edges be deleted from the complete graph K_n on n nodes: K_n has an edge between every pair of distinct nodes and thus contains $n(n-1)/2$ edges. For the case when $k \leqslant n/2$, we will demonstrate that the number of spanning trees $T(n, E)$ in the resulting graph is maximized by choosing the k deleted edges to be mutually nonadjacent. The (apparently more complicated) cases with $k > n/2$ await resolution.

Let P_k denote a set of k nonadjacent ("parallel") edges in K_n, where $k \leqslant n/2$. We will show that $|E| = k$ implies $T(n, E) \leqslant T(n, P_k)$. First the case when the edge set E is disconnected will be disposed of. This will be done in the context of the inductive hypothesis that if $i < k$ and $|S| = i$, then $T(n, S) \leqslant T(n, P_i)$ — whether or not S is connected. Suppose $|E| = k$ and that E can be decomposed into connected edge sets $C_1, C_2, \ldots, C_p$ with $p \geqslant 2$. Certainly, when $k=2$ the only disconnected set E possible consists of two nonadjacent edges, so that the inductive hypothesis holds. More generally, if $|E| = k$ then $|C_1| < k$ and $n \geqslant 2k > 2|C_1|$, whence $0 \leqslant T(n, C_1) \leqslant T(n, P_\alpha)$, $\alpha = |C_1|$, by the inductive hypothesis. Similarly, we have $0 \leqslant T(n, \bar{C}_1) \leqslant T(n, P_\beta)$, $\beta = |\bar{C}_1|$, where $\bar{C}_1 = C_2 \cup \ldots \cup C_p$ in the decomposition $E = C_1 \cup C_2 \cup \ldots \cup C_p$. From [2, p. 106] it is known that

$$T(n, P_i) = n^{n-2}\left(1 - \frac{2}{n}\right)^i,$$

whereupon

$$T(n, C_1)T(n, \bar{C}_1) \leqslant n^{n-2}n^{n-2}\left(1 - \frac{2}{n}\right)^k.$$

Also, the following relation [2, p. 106] obtains for disjoint edge sets:

$$T(n, C_1)T(n, \bar{C}_1) = n^{n-2}T(n, C_1 \cup \bar{C}_1),$$

AMS Subject Classification: 05C05

*This work was done while the author was a National Academy of Sciences-National Research Council Postdoctoral Research Associate at the National Bureau of Standards, Washington, D.C. 20234.

from which it is immediate that

$$T(n, E) \leqslant n^{n-2}\left(1-\frac{2}{n}\right)^k = T(n, P_k).$$

Subsequently, then, it will be assumed that E is a connected set of k edges on r nodes. It is only necessary to consider the case when $r > 2$ and show that $T(n, E) \leqslant T(n, P_k)$; if $r=2$ then $k=1$ and this relation is immediately satisfied with equality. The following result is most helpful in establishing the relation for $r > 2$; it involves a symmetric function ϕ for which the number of arguments, as well as their values, will be considered variable.

LEMMA: *Suppose that* $\phi(\mathbf{d}) = \phi(d_1, \ldots, d_r) \equiv \prod_{i=1}^{r}\left(1-\frac{d_i}{n}\right)$ *with* $\sum_{i=1}^{r} d_i = 2k, n \geqslant 2k, k+1 \geqslant r > 2$ *and all* d_i *positive integers. Then* $\phi(\mathbf{d})$ *is maximized over all* $\mathbf{d} \neq (1, 1, 2, 2, \ldots, 2)$ *by* $\mathbf{d}_0 = (2, 2, \ldots, 2)$.

PROOF: Given a sequence $\mathbf{d}$, execute the following procedure.

PROCEDURE REDUCE.

1. Let $d_i =$ a smallest element of $\mathbf{d}$,
   $d_j =$ a largest element of $\mathbf{d}$.
2. If $d_j - d_i > 1$ then $d_i := d_i + 1$, $d_j := d_j - 1$ and go to Step 1. Otherwise, terminate.

Note that this procedure is always finite and does not change the values for r and $\sum_{i=1}^{r} d_i$. Moreover, if the procedure yields as output the vector $\mathbf{d}^*$ for the input sequence $\mathbf{d}$, then $\phi(\mathbf{d}) \leqslant \phi(\mathbf{d}^*)$. In fact

$$\left(1-\frac{d_i}{n}\right)\left(1-\frac{d_j}{n}\right) \leqslant \left(1-\frac{d_i+1}{n}\right)\left(1-\frac{d_j-1}{n}\right)$$

$$\Leftrightarrow d_i d_j \leqslant (d_i+1)(d_j-1)$$
$$\Leftrightarrow 1 \leqslant d_j - d_i.$$

Thus at each step of REDUCE the value of ϕ cannot decrease, and so $\phi(\mathbf{d}) \leqslant \phi(\mathbf{d}^*)$. Upon termination of this procedure, one of the following final forms for $\mathbf{d}^*$ will be obtained (to within permutation):

(i) $(1, 1, \ldots, 1)$
(ii) $(s, s, \ldots, s), s > 1$
(iii) $(s, \ldots, s, s+1, \ldots, s+1), s > 1$
(iv) $(1, \ldots, 1, 2, \ldots, 2)$.

In Case (i), $1+1+\ldots+1 = r = 2k$. However, $r \leqslant k+1$ implies $2k \leqslant k+1$ so that $k=1$. This case cannot then occur since it is assumed that $k+1 > 2$.

In Case (ii), consider the situation when s is even. The following relation will allow a further simplification of $\mathbf{d}^*$. Namely, if $d_i \geqslant 2$ then

$$\left(1-\frac{d_i}{n}\right) \leqslant \left(1-\frac{2}{n}\right)\left(1-\frac{d_i-2}{n}\right). \tag{1}$$

(The above inequality is valid since $2(d_i - 2) \geqslant 0$.) Therefore, by replacing each s (if $s > 2$) with $(2, s-2)$, the value of ϕ cannot decrease and so eventually $\phi(s, \ldots, s) \leqslant \phi(2, \ldots, 2)$. If s is odd

$(s \geq 3)$, then the simplification $s := (2, s-2)$ eventually results in $\mathbf{d}_1 = (2, \ldots, 2, 3, \ldots, 3)$, where the number t of such 3's is necessarily even since $\Sigma\, d_i = 2k$. For each pair (3, 3) the following inequality holds:

$$\left(1-\frac{3}{n}\right)\left(1-\frac{3}{n}\right) \leq \left(1-\frac{2}{n}\right)^3, \tag{2}$$

because $n \geq 2k > 2$ implies $9n \leq 12n - 8$. Together with (1) the above inequality shows that $\phi(\mathbf{d}^*) \leq \phi(\mathbf{d}_1) \leq \phi(2, 2, \ldots, 2)$. In any event, then, $\phi(\mathbf{d}) \leq \phi(\mathbf{d}^*) \leq \phi(\mathbf{d}_0)$.

Suppose that s is odd in Case (iii), so that there are an even number of these s values in $(s, \ldots, s, s+1, \ldots, s+1)$. As in Case (ii) the simplifications implied by (1) and (2) yield $\phi(\mathbf{d}^*) \leq \phi(2, 2, \ldots, 2)$. A similar argument applies when s is even.

In Case (iv), suppose that the number of 1's is t, so the number of 2's is $r - t$. Now,

$$2r - 2 \leq 2k = \sum_{i=1}^{r} d_i = t + 2(r-t) = 2r - t,$$

whence $t \leq 2$. Since t must be even, either $t = 0$ or $t = 2$. In the former situation we have the vector $(2, 2, \ldots, 2)$ which is already subsumed under Case (ii); in the latter, we have $\mathbf{d}^* = (1, 1, 2, 2, \ldots, 2)$. Note that unless $\mathbf{d} = \mathbf{d}^*$ this possibility must have arisen from applying Step 2 of REDUCE to the vector $\mathbf{d}_2 = (1, 1, 1, 2, 2, \ldots, 2, 3)$, which in turn arose from zero or more applications of Step 2 starting with the original $\mathbf{d}$. Thus $\phi(\mathbf{d}) \leq \phi(\mathbf{d}_2)$.

However,

$$\phi(\mathbf{d}_2) = \left(1-\frac{1}{n}\right)^3\left(1-\frac{3}{n}\right)\left(1-\frac{2}{n}\right)^{k-3}$$

$$\phi(\mathbf{d}_0) = \left(1-\frac{2}{n}\right)^k$$

so that

$$\phi(\mathbf{d}_2) \leq \phi(\mathbf{d}_0) \Leftrightarrow \left(1-\frac{1}{n}\right)^3\left(1-\frac{3}{n}\right) \leq \left(1-\frac{2}{n}\right)^3$$

$$\Leftrightarrow 8n + 3 \leq 10n.$$

Since $n \geq 2k > 2$ this latter relation certainly holds and so $\phi(\mathbf{d}_2) \leq \phi(\mathbf{d}_0)$. It therefore follows that if $\mathbf{d} \neq (1, 1, 2, 2, \ldots, 2)$ then $\phi(\mathbf{d}) \leq \phi(\mathbf{d}_0)$ in Case (iv).

What is concluded from these four cases is that $\phi(\mathbf{d}) \leq \phi(\mathbf{d}_0)$ for all $\mathbf{d} \neq (1, 1, 2, 2, \ldots, 2)$, which is precisely the result to be established in the Lemma. It should be noted that the degrees d_i of a connected graph with $r > 2$ nodes and k edges satisfy the hypotheses of the Lemma if $n \geq 2k$.

Now we are in a position to prove the main result. Let $\{1, 2, \ldots, n\}$ be the nodes of K_n, and let $\{1, 2, \ldots, r\}$ be the nodes of a connected set E of k edges, with $r > 2$ and $n \geq 2k$. Also, let d_i denote the degree in E of node i. There are two cases to consider.

(i) E forms a chain. Then from [2, p. 108]

$$T(n, E) = n^{n-2} \sum_{j=0}^{k} \binom{2k-j+1}{j} \left(-\frac{1}{n}\right)^j$$

$$\leq n^{n-2}\left(1-\frac{2}{n}\right)^k,$$

using a combinatorial inequality proved in [4]. Thus $T(n, E) \leqslant T(n, P_k)$ and so the desired relation is obtained.

(ii) E does not form a chain. Then the degree sequence $\mathbf{d} = (d_1, \ldots, d_r) \neq (1, 1, 2, 2, \ldots, 2)$. Since E is a connected set of edges, the hypotheses of the above lemma are satisfied and so it is concluded that $\phi(\mathbf{d}) \leqslant \phi(\mathbf{d}_0) = \left(1 - \frac{2}{n}\right)^k$. Moreover, an expression for $T(n, E)$ is given [6] by $T(n, E) = n^{n-2} \det A$, where $A = (a_{ij})$ is the $r \times r$ matrix defined by

$$a_{ij} = \begin{cases} 1 - \dfrac{d_i}{n} & \text{if } i = j \\ \dfrac{1}{n} & \text{if } i \neq j,\ (i, j) \epsilon E \\ 0 & \text{otherwise.} \end{cases}$$

Now since E has k edges, $d_i \leqslant k \leqslant n/2$ or $1 - \frac{d_i}{n} \geqslant \frac{1}{2}$; thus $a_{ii} \geqslant \sum_{j \neq i} |a_{ij}|$ and accordingly matrix A is positive semidefinite. Also, for some i, $\frac{d_i}{n} < \frac{1}{2}$ since otherwise

$$\frac{nr}{2} = \sum_{i=1}^{r} d_i = 2k \leqslant n$$

or $r \leqslant 2$, a case which has been previously excluded. Inasmuch as A is symmetric and for some i, $a_{ii} > \frac{1}{2} > \sum_{j \neq i} |a_{ij}|$, it follows [5] that A is nonsingular and hence positive definite. Furthermore, using the Hadamard bound on det A [1, p. 129], we obtain

$$T(n, E) = n^{n-2} \det A \leqslant n^{n-2} \phi(\mathbf{d}) \leqslant n^{n-2} (1 - \frac{2}{n})^k = T(n, P_k).$$

Thus in either case $T(n, E) \leqslant T(n, P_k)$ and so the number of spanning trees is maximized by choosing the k edges to be mutually nonadjacent. Of course, in general there are numerous ways of choosing k parallel edges to delete from K_n, all of which result in the same (maximum) number of spanning trees. The restriction that $n \geqslant 2k$ is certainly a natural one, since it ensures the possibility of being able to select k parallel edges from the n-node complete graph. Finding the maximum (or minimum) number of spanning trees when $k > n/2$ appears to be an open problem [3].

References

[1] Bellman, R., Introduction to Matrix Analysis (McGraw-Hill, New York, 1970).
[2] Berge, C., Principles of Combinatorics (Academic Press, New York, 1971).
[3] Sedláček, J., On the number of spanning trees of finite graphs, Časopis Pěst. Mat. **94**, 217–221 (1969).
[4] Shier, D., Two combinatorial inequalities, submitted for publication.
[5] Taussky, O., A recurring theorem on determinants, Amer. Math. Monthly **56**, 672–676 (1949).
[6] Temperley, H. N. V., On the mutual cancellation of cluster integrals in Mayer's fugacity series, Proc. Phys. Soc. (London) **83**, 3–16 (1964).

(Paper 78B4–413)