RECONSTRUCTION OF TREES

BENNET MANVEL

Every tree T determines a set of distinct maximal proper subtrees $T_i = T - v_i$, which are obtained by the deletion of an endpoint of T. In this paper we prove that a tree is almost always uniquely determined by this set of its subtrees, and point out two interesting consequences of this result.

In [**5**], Ulam proposed the following conjecture, which we state in a slightly stronger form due to Harary [**1**].

ULAM'S CONJECTURE. *A graph G with at least three points is uniquely determined up to isomorphism by the subgraphs $G_i = G - v_i$.*

Kelly [**4**] proved the conjecture for trees and Harary and Palmer [**3**] showed that not all of the G_i are needed in that case by proving Corollary 1 below. If we remove from the list of subgraphs G_i of a graph G all but one graph of each isomorphism type, we obtain a set of G_i which are distinct up to isomorphism. The following conjecture involving this set is due to Harary. It seems considerably stronger than Ulam's conjecture, and in fact many people would say that it is extremely unlikely. There is, however, no counter-example with seven or fewer points.

HARARY'S CONJECTURE. *A graph G with $p \geqq 4$ points can be reconstructed uniquely from its set of non-isomorphic subgraphs $G_i = G - v_i$.*

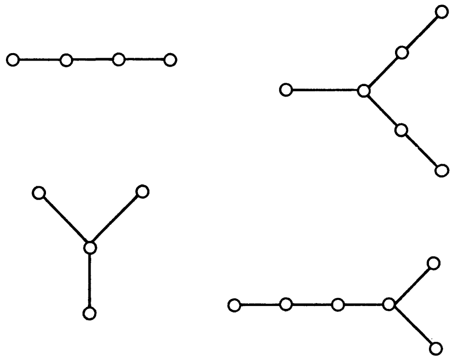
Note that we know here only what graphs are in the list of subgraphs G_i, not how many times each occurs there. The following theorem shows that usually not even that much information is necessary for reconstructing a tree. The maximal proper subtrees of a tree T are just those which result from the deletion of an endpoint of T.

THEOREM 1. *A tree T is determined by its set of non-isomorphic maximal proper subtrees, except in the two cases illustrated in Figure* 1.

Proof. We call the set of all central points of a tree T its *centre*. If v_i is an endpoint of T and the tree $T_i = T - v_i$ has the same centre as T, then, following Kelly, we call v_i a *non-essential point* or *n.-e. point*. An endpoint whose removal changes the centre of T is naturally called an *essential* (*e.*) *point*, and any point at maximal distance from the centre of T is *radial*. The *branches of a centred tree* are the branches of its central point, rooted at that point, and the *branches of a bicentred tree* are those rooted branches of either of its central

Received September 6, 1968 and in revised form, December 13, 1968. This research was supported by grant AFOSR 68–1515 from the U.S. Air Force Office of Scientific Research.

points which do not contain the axis joining those two points. A branch is *radial* if it contains a radial point. Any further questions about our terminology may be settled by consulting [**2**].



(a) (b)

FIGURE 1

We proceed to find the degree sequence of T. If T has a point of degree 2 adjacent to an endpoint, then the degree sequence of any T_i with a maximal number of endpoints can be augmented to that of T by adding a 2. If T has no such point, then if we select from the subtrees a tree T_j with a point of degree d adjacent to a peripheral endpoint, where d is the smallest degree of any point adjacent to such an endpoint of a T_i, the degree sequence of T will be that of T_j with d replaced by $d + 1$ and a 1 added. In order to find the degree sequence of G, therefore, we need merely decide whether or not T has a point of degree 2 adjacent to an endpoint. To do that we note that if some subtree has more endpoints than another, then there is such a point, and if all subtrees have the same number of endpoints, then either no endpoint is adjacent to a point of degree 2 or they all are. The latter possibility holds if and only if in every subtree at most one endpoint is not adjacent to a point of degree 2, with the exception of the star of Figure 1(a). We may therefore assume, with that exception, that we can find the degree sequence of T.

Since we have the degree sequence, we may assume in what follows that T is not a path. Thus T has some n.-e. points, whose deletion results in T_i having the same centre as T. We call such a T_i, which is just a T_i with maximum diameter, an *n.-e.* T_i, and an *essential* T_i is defined and recognized in an analogous way. From any n.-e. T_i the radius and the central or bicentral nature of T can be deduced.

Case 1. *T is bicentral.* The two components which result when the axis is removed from T will be called the *halves* of T and will be denoted by H_1 and H_2. Suppose first that only one half of T contains an n.-e. point of T. This is the case if and only if all n.-e. T_i have at least one half which is a path, all

https://doi.org/10.4153/CJM-1970-007-4 Published online by Cambridge University Press

(at least one) essential T_i have a radial path as one branch, and, if those two conditions hold for a tree T with exactly one essential and one n.-e. T_i, then T does not have two points of degree 3.

In such a situation, that centred subtree with more points near its centre can be augmented to T by adding a point adjacent to the endpoint of its radial path.

If both halves of T have n.-e. points, we assume without loss of generality that $|V(H_1)| \geqq |V(H_2)|$. If all pairs of halves of n.-e. T_i have one half just one point larger than the other, then $|V(H_1)| = |V(H_2)|$, and H_1 and H_2 are just the larger halves of two pairs, chosen to be different if possible. Otherwise $|V(H_1)| > |V(H_2)|$ and we can easily determine the two halves from two T_i, one formed by deletion of an n.-e. point from H_1 and the other formed in the same way from H_2, unless it happens that $H_2 \cong H_1 - v$, for some n.-e. point v. In that case there will be some confusion about H_2, but there will be some T_i from an n.-e. point of H_1 which will have two isomorphic halves, and those must both be H_2, and our proof is complete.

Case 2. *T is central.* We need to find first whether or not just one branch of T contains an n.-e. point. If T has only one n.-e. point, then there is no problem, and if there are three or more such points, then they all lie on the same branch of T if and only if every n.-e. T_i has all its n.-e. points on one branch. Thus the only difficulty occurs when T has exactly two n.-e. points. We first investigate that case, using T_1 and T_2 for the two subtrees T_i with the same centre as T, and then explain how the number of n.-e. points of T can be found so that we know which case we are dealing with.

Since T has exactly two radial branches if it has exactly two n.-e. points, there are only three possibilities. First, if one of T_1 and T_2 has an n.-e. point on a radial branch and the other does not, then it is easy to see that one n.-e. point of T is on a radial branch and the other on a non-radial one. Second, if neither T_1 nor T_2 has an n.-e. point on a radial branch, then the two n.-e. points of T lie on one or two branches as T has maximum degree 3 or 4, since two n.-e. points lying on different branches must be the endpoints of non-radial paths leading from the centre of T, which must thus have degree 4. Finally, if both T_1 and T_2 show an n.-e. point on a radial branch, then if there is only one essential subtree T_i, both n.-e. points are on one branch if and only if one of the halves of that T_i is a path. If there are two essential subtrees T_i, then the two n.-e. points are on different branches of T if and only if either T_1 or T_2 has three branches or both halves of each essential T_i have at least one n.-e. point.

In order to use these arguments, we must find the number s of n.-e. points in T. If we let t be the maximum number of n.-e. points in any n.-e. T_i, then it is clear that s is either t, $t + 1$, or $t + 2$. Thus if t is at least 3, s is also, and we have all the information we need for the above argument. Furthermore, if t is 0, s must be 1 or 2 since we are assuming that we have eliminated paths. It will be 2 if and only if there is exactly one essential T_i and it has two endpoints at distance 2 from each other, with the exception of the last tree in Figure 1. It remains to decide for $t = 1$ whether s is 1, 2, or 3 and for $t = 2$ whether s is 2, 3, or 4.

If t is 1, we can generally distinguish $s = 1$ or 3 from $s = 2$ by the degree sequence of T, since we usually have in the latter case either two points of degree 3 or one of degree 4, which cannot happen if s is 1 or 3. If a tree T with only one point of degree 3 and none of degree 4 has two different essential T_i, then it must have only one n.-e. point since two of its three endpoints are essential. If it has only one T_i, then its three endpoints are similar, and therefore $s = 3$. Finally, if T has exactly one essential and one n.-e. T_i, we can have either of the two situations shown in Figure 2.

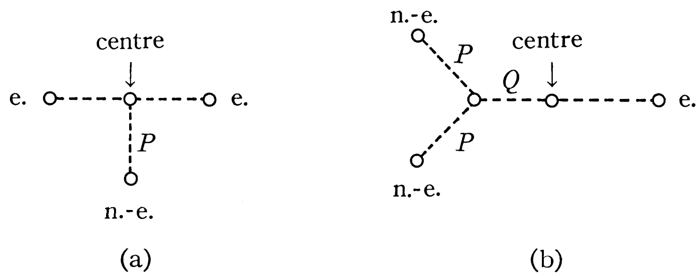


FIGURE 2

If P has length at least 2, then (a) and (b) can be distinguished since (a) has its n.-e. subtree with a centre of degree 3 and the centre of the n.-e. subtree of (b) must be of degree 2. If P has length 1, then, if T has at least seven points, Q must have length at least 2, and therefore we can distinguish (a) and (b) since the former has an essential T_i with bicentre of degree 3 and the latter does not. By examining the eight central trees with less than seven points, one can verify that the theorem holds for all such trees except those shown in Figure 1.

Similarly, if $t = 2$, we have $s = 4$ if and only if there is no essential T_i. Furthermore, we can usually distinguish $s = 2$ from $s = 3$ by the degree sequence of T, since in the latter case but not the former we may have $4, 3, 2, \ldots, 1$ or $3, 3, 3, 2, \ldots, 1$. The only difficulty occurs when we have a tree such as that of Figure 3(a) to which there is attached an n.-e. path at one of the points a, b, c (centre), or d. Call these trees T_a, T_b, T_c, and T_d, respectively. They each have three n.-e. points, one essential point, and a degree sequence beginning with either a single 4 and no 3 or just two 3s, and therefore may be confused with the trees of Figure 3(b) which have only two n.-e. points but otherwise seem to have the same properties, since their essential points are similar. Now T_c and T_d can be set apart from the others since their unique bicentral subtree has two halves, neither of which is a path, however T_a might be confused with T_3 or T_b with T_4. The trees T_3 and T_a are

https://doi.org/10.4153/CJM-1970-007-4 Published online by Cambridge University Press

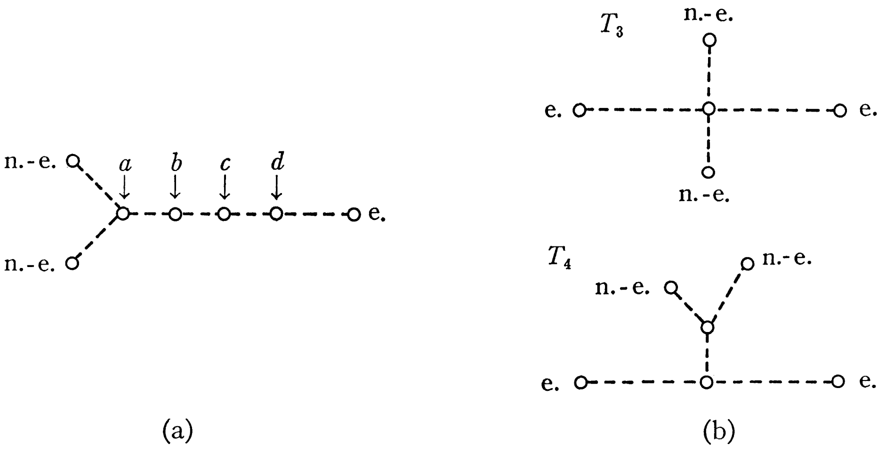


FIGURE 3

distinguished by observing that one has its point of maximum degree at the centre of its n.-e. subtrees and the other does not. To distinguish T_4 from T_b we examine their essential subtrees, and find that in general that of the former cannot have two radial points on the same branch while that of the latter must. The only exception is when the two n.-e. points of T_4 are similar, in which case T_4 will be distinguished by having only one n.-e. T_i, while T_b has two. Thus in every case we can find the number of n.-e. points of T.

Now if only one branch of T has an n.-e. point, we proceed in two cases as follows. All n.-e. points of T lie on a non-radial branch if and only if in all essential T_i one bicentre has degree 2 and the other has degree 3. In that case T is clearly obtained from any such T_i by adding v_i adjacent to the end of a radial path leading from the bicentre of degree 3. In the other case, T has just two branches, both radial, and all n.-e. points lie on one of them. To reconstruct T in this case we select that essential tree T_k which has a point of degree 3 or more as close as possible to a bicentre. Then T_k may be augmented to T by adding v_k adjacent to the endpoint of the unique (radial) path leading from a bicentre of degree 2.

If several branches of T have n.-e. points, then every branch of T will appear as a branch in some central T_i. Select a largest branch B containing an n.-e. point v. Then if T_i is a central subtree containing a minimum number of branches isomorphic to B and a maximum number isomorphic to $B - v$, T can be obtained from T_i by replacing one branch $B - v$ by B. Thus T can be reconstructed in every case, and the theorem is proved.

Since the pairs of trees shown in Figure 1 do not have common sets of T_i, the following result is immediate.

COROLLARY 1. *A tree T can be reconstructed uniquely from its set of non-isomorphic subforests $T_i = T - v_i$.*

https://doi.org/10.4153/CJM-1970-007-4 Published online by Cambridge University Press

Our proof may be shortened considerably if the multiplicities of the T_i are given, and is then a new proof of the following result, due to Harary and Palmer [**3**].

COROLLARY 2. *Any tree T can be reconstructed from its list of maximal proper subtrees.*

REFERENCES

1. F. Harary, *On the reconstruction of a graph from a collection of subgraphs*, pp. 47–52 in *Theory of graphs and its applications*, edited by M. Fiedler (Academic Press, New York, 1964).

2. ——— *Graph theory* (Addison-Wesley, Reading, Massachusetts 1969).

3. F. Harary and E. M. Palmer, *The reconstruction of a tree from its maximal subtrees*, Can. J. Math. *18* (1966), 803–810.

4. P. J. Kelly, *A congruence theorem for trees*, Pacific J. Math. *7* (1957), 961–968.

5. S. M. Ulam, *A collection of mathematical problems*, p. 29 (Wiley (Interscience), New York, 1960).

The University of Michigan,
Ann Arbor, Michigan