THE RECONSTRUCTION OF A TREE FROM ITS MAXIMAL SUBTREES

FRANK HARARY AND ED PALMER

One of the many interesting conjectures proposed by S. M. Ulam in **(5)** can be stated as follows:

If G and H are two graphs with p points v_i and u_i respectively ($p \geqslant 3$) such that for all i, $G - v_i$ is isomorphic with $H - u_i$, then G and H are themselves isomorphic.

P. J. Kelly **(3)** has shown this to be true for trees. The conjecture is, of course, not true for $p = 2$, but Kelly has verified by exhaustion that it holds for all of the other graphs with at most six points. Harary and Palmer **(2)** found the same to be true of the seven-point graphs.

In **(1)** Harary reformulated the conjecture as a problem of reconstructing G from its subgraphs $G - v_i$ and derived several of the invariants of G from the collection $G - v_i$.

The purpose of this paper is to show that any tree T can be reconstructed from fewer than all the subgraphs $T - v_i$, namely from only its maximal (proper) subtrees. This is an improvement of Kelly's result because the latter depends on all of the subgraphs (both the trees and the other forests) obtained by the deletion of a point. Definitions that do not appear here, as well as basic theorems on trees, may be found in König **(4)**. In particular, a *forest* is a graph with no cycles; a *tree* is a connected graph with no cycles.

Let $u_1, \ldots, u_p$ be the points of a connected graph G and let $G_i = G - u_i$ be the graph obtained by deleting the point u_i from G (as well as all the lines incident with u_i). Henceforth we consider only graphs with more than two points.

If T is any tree, then every subgraph $T - v_i$ is a forest. Since we have assumed that $p \geqslant 3$, at least one of these forests is not a tree. Now G itself is a cycle of length p if and only if each G_i is a path of length $p - 2$. Hence the question of whether or not G is a tree is easily determined from the collection G_i because G is a tree if and only if G is not a cycle and each G_i is a forest.

The following ten theorems show that every tree can be reconstructed from those subgraphs (obtained by removing one point at a time) that are themselves trees. Let $v_1, \ldots, v_n$ be the points of degree one (end points) in the tree T and let $T_i = T - v_i$ be the subtree obtained by deleting v_i from T. Recall

Received June 22, 1965. Work supported in part by the U.S. Air Force Office of Scientific Research under Grant AF-AFOSR-754-65.

that every tree is either centred or bicentred; see **(4**, p. 65**)**. The first two theorems are simple but useful.

THEOREM 1. *If T is centred, at most two of the subtrees T_i are bicentred.*

Proof. Let T be centred at c and, without loss of generality, let $P(v_1, v_2)$ be a path in T whose length is the diameter of T. Then c is a point of $P(v_1, v_2)$ and for each $i \geqslant 3$, T_i is centred at c. Obviously if T_1 or T_2 is centred, c is the centre. And if T_1 or T_2 is bicentred, c is one of the centres.

COROLLARY 1.1. *If at least three of the T_i are bicentred, T is bicentred.*

This corollary is merely the contrapositive of Theorem 1. Similar reasoning yields the next result.

THEOREM 2. *If T is bicentred, at most two of the T_i are centred.*

COROLLARY 2.1. *If at least three of the T_i are centred, T is centred.*

COROLLARY 2.2. *If T has at least five end points, then T is centred or bicentred according as the majority of the T_i are centred or bicentred.*

When the number of end points of T is less than five, the question of whether or not T is centred is not necessarily determined by the number of subtrees T_i which are centred. But the main result is still settled more easily here than in the general case. Clearly T must be homeomorphic to one of the trees shown in Figure 1 according as T has two, three, or four end points. By examining all of the possibilities for the subtrees T_i, we obtain Theorem 3. The exhaustive details are omitted.

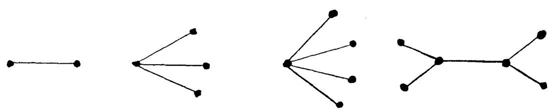

FIGURE 1.

THEOREM 3. *If T has less than five end points, then T is determined by the T_i.*

The next theorem shows how to determine the degree of the centre c of T from the degree of c in the subtrees T_i. Note that when T is centred, the radius of T is the same as the radius of any subtree T_i that is centred. Recall that n denotes the number of end points of T. We call a tree a *star* if all of its points except one are end points.

THEOREM 4. *Let T be a tree with at least three end points which is centred at c. Let r be the radius of T. Let m be the maximum degree of c in those subtrees T_i that are centred. Then* $\deg c = m$ *in T unless each centred subtree T_i is isomorphic to a tree T' constructed by identifying the centres of a star having $n - 3$ end points and a path of length $2r$. In this case* $\deg c = m + 1$.

Proof. Without loss of generality we can let $P(v_1, v_2)$ be a path in T whose length is $2r$, the diameter of T. Since T has at least three end points, we can let v_3 be one such that the distance d from v_3 to c is maximal among all the end points v_i with $i > 2$. If $d > 1$, then $\deg c = m$. If $d = 1$, then all of the end points of T except possibly v_1 and v_2 are at distance one from c. Hence T_i is isomorphic to T' for $i > 2$. If the radius $r = 1$, then T_1 and T_2 are also isomorphic to T'. Obviously $\deg c = m + 1$ in this case.

The next three theorems show how to reconstruct T from the subtrees T_i when T is centred. It is convenient to consider separately the cases when T has none, one, or two subtrees T_i that are bicentred.

For the reconstruction, we need the following definition. A *branch* (B, v) of a tree at the point v is a maximal connected subtree containing v and exactly one point of the tree which is adjacent with v.

THEOREM 5. *Let T be a tree with at least five end points such that exactly two of the subtrees T_i are bicentred. Then T is determined by the T_i.*

Proof. Since T has at least five end points, at least three of the subtrees T_i are centred. Therefore Corollary 2.1 implies that T is centred, say at c. Without loss of generality we can let T_1 and T_2 be the two bicentred subtrees of the hypothesis. Therefore each T_i for $i \geqslant 3$ is centred. In the proof of Theorem 1, it is shown that this centre is c for every T_i. Using Theorem 4, we can determine the degree of c in T from the degree of c in the subtrees T_i. Now the proof is in two parts according as the degree of c is two or greater than two.

Part 1. The degree of c is two.

Since there are at least four end points in each T_i, we have the following two cases:

Case 1.1. For each $i \geqslant 3$, T_i has a branch at c with only one end point in T_i.

Case 1.2. For some $i \geqslant 3$, T_i has more than one end point in each of its two branches at c.

In Case 1.1, let r be the radius of T. Since T has at least five end points, one of the branches of T at c, say (B_1, c), must be a path of length r. For each $i \geqslant 3$ let w_i be the point of T_i with $\deg w_i > 2$ in T_i and minimum distance $d(c, w_i)$. Let $d = \min d(c, w_i)$ for $i \geqslant 3$. Then d is the distance in T from c to the nearest point of degree greater than two. We can assume that T_1 has centres c_1 and c_2 and a point u of degree greater than two with $d(u, c_2) = d - 1$. Therefore T_1 must have been obtained by deleting the point of degree one that is in (B_1, c). Hence the branch of T_1 at c_1 which contains c_2 is the other branch of T, say (B_2, c).

In Case 1.2, for each $i \geqslant 3$ let the two branches of T_i at c be denoted by (B_i^1, c) and (B_i^2, c). Let $p_1 + 1$ be the maximum of the number of points in any of these branches and let p_2 be the minimum. Then if (B_1, c) and (B_2, c) again denote the two branches of T at c, we can assume that (B_1, c) contains

https://doi.org/10.4153/CJM-1966-079-8 Published online by Cambridge University Press

$p_1 + 1$ points and (B_2, c) contains $p_2 + 1$ points. For each $i \geqslant 3$, the branches of T_i at c have either p_1 and $p_2 + 1$ points or $p_1 + 1$ and p_2 points, and so $p_1 \geqslant p_2$.

If $p_1 > p_2 + 1$, then any branch of T_i at c which contains $p_1 + 1$ points is also a branch of T at c. The other branch of T at c is obtained by taking a branch some T_j at c which contains $p_2 + 1$ points.

If $p_1 < p_2 + 1$, then $p_1 + 1 = p_2 + 1$ because $p_1 \geqslant p_2$. Hence for each $i \geqslant 3$ we can assume that (B_i^1, c) has $p_1 + 1$ points and (B_i^2, c) has p_1 points. If the branches (B_i^1, c) are all isomorphic, then each branch of T at c is (B_i^1, c). Otherwise if (B_j^1, c) and (B_k^1, c) are not isomorphic, then (B_j^1, c) and (B_k^1, c) are the branches of T at c.

The only remaining possibility has $p_1 = p_2 + 1$. For some $i \geqslant 3$ we can choose T_i so that (B_i^1, c) has $p_1 + 1$ points and (B_i^2, c) has $p_2 = p_1 - 1$ points. Then (B_i^1, c) is a branch of T at c. Let w be the point of (B_i^1, c) that has degree greater than two and minimum distance $d = d(c, w)$ from c. We can assume that T_1 has centres c_1 and c_2 and a point u of degree greater than two such that $d(c_1, u) = d$, that u is in the branch of T_1 at c_1 that does not contain c_2, and that there are p_1 points in this branch. Then T_1 must have been obtained from T by deleting a point from the larger branch of T at c and the other branch of T at c is the branch of T_1 at c_1 which contains c_2. Thus T is reconstructed, completing the proof of Part 1.

Part 2. The degree of c is greater than two.

As above let r be the radius of T. Then v_1 and v_2 are the only points of T at distance r from c and each occurs in a different branch of T at c. Let these two branches be (B_1, c) and (B_2, c). Let $(B_3, c), \ldots, (B_s, c)$ be the other branches of T at c.

For each $i \geqslant 3$ let (B_i^1, c) and (B_i^2, c) be the two branches of T_i at c that contain points at distance r from c, in accordance with the hypothesis of the theorem. Let M be the maximum for $i \geqslant 3$ of the number of points in $(B_i^1, c) \cup (B_i^2, c)$. Choose T_k with M points in $(B_k^1, c) \cup (B_k^2, c)$. Clearly T_k must have been obtained by deleting a point from some branch (B_i, c) of T with $i \geqslant 3$. Hence the branches (B_1, c) and (B_2, c) of T are precisely (B_k^1, c) and (B_k^2, c).

To determine the other branches of T we consider two cases:

Case 2.1. (B_1, c) and (B_2, c) are paths of length r.

Case 2.2. (B_1, c) has more than one end point in T.

For Case 2.1, we know that T_1 and T_2 are isomorphic. Let c_1 and c_2 be the centres of T_1. Suppose c_1 is the centre at which there is a branch in T_1 that is a path of length r. Then there is another branch at c_1 that is a path of length $r - 1$. The remaining branches of T_1 at c_1 can be taken as the other branches $(B_3, c), \ldots, (B_s, c)$ of T at c.

For Case 2.2, choose T_j with $j \geqslant 3$ so that there are $M - 1$ points in $(B_j^1, c) \cup (B_j^2, c)$. Then T_j must have been obtained by deleting a point from either (B_1, c) or (B_2, c). Let $(B_j^3, c), \ldots, (B_j^s, c)$ be the other branches of T_j at c. Then the branches $(B_3, c), \ldots, (B_s, c)$ of T at c are precisely $(B_j^3, c), \ldots, (B_j^s, c)$. Hence we know all the branches of T at c, so that T has been reconstructed. This completes the proof of Part 2 and of the theorem.

THEOREM 6. *Let T be a tree with at least four end points such that exactly one of the subtrees T_i is bicentred. Then T is determined by the T_i.*

Proof. Since T has at least four end points, at least three of the subtrees T_i are centred. Therefore Corollary 2.1 implies that T is centred, say at c. Without loss of generality we can let T_1 be the bicentred subtree of the hypothesis. Thus each T_i for $i \geqslant 2$ is centred and we know that this centre is c. Again using Theorem 4 to determine the degree of c in T from the degree of c in the subtrees T_i, we present the proof in two parts according as the degree of c is two or greater than two.

Part 1. The degree of c is two.

Let r be the radius of T. Since only T_1 is bicentred, we know that T has one branch at c, say (B_1, c), which contains exactly one point at distance r from c, namely v_1. The other branch, say (B_2, c), contains at least two such points.

Suppose T_1 has a branch at one centre, say c_1, which is a path of length r containing the other centre c_2. Then one branch of T at c is a path of length r and the other is the branch of T_1 at c_2 that contains c_1.

Otherwise each branch of T at c contains points v of degree one such that $T - v$ is centred. Let S be the set of all subtrees T_i with $i > 1$ such that T_i has a branch at c, say (B_i^1, c), that contains at least two points at distance r from c. For each T_i in S, let k_i be the number of points in (B_i^1, c) and let m_i be the number of points in (B_i^1, c) at distance r from c. Let k be the maximum of the k_i and let m be the maximum of the m_i. Let S' be the set of all T_i in S such that $k_i = k$ and $m_i = m$. Since each T_j in S' must have been obtained by deleting a point (B_1, c), the branch (B_2, c) of T is given by (B_j^1, c).

Let S'' be the set of all subtrees T_i with $i > 1$ that are not in S'. The branch (B_1, c) occurs as a branch of each T_i in S''. It can be identified because it is the only branch that is in each T_i of S'' and that has at least two end points exactly one of which is at distance r from c.

Part 2. The degree of c is greater than two.

As in Part 1 we can let (B_1, c) be the branch of T at c that contains exactly one point at distance r from c. Let (B_2, c) be the branch that contains at least two such points. Since at least one of these branches has more than one end point in T, the proof is the same as that of Theorem 5, Part 2, Case 2.2.

THEOREM 7. *Let T be a tree with at least three end points such that each subtree T_i is centred. Then T is determined by the T_i.*

Proof. Since the subtrees T_i of T are all centred, Corollary 2.1 implies that T is centred, say at c. Further the centre of each subtree T_i is also c.

Suppose the degree of c in T is not equal to the degree of c in some T_i. Then v_i must be an end point adjacent to c. Therefore T is easily obtained from this subtree T_i by adding a line at c.

Now we can assume that for each i the degree of c in T_i is the same as the degree of c in T. Let the branches of each T_i at c be $(B_i^1, c), \ldots, (B_i^s, c)$ where $s = \deg c$. Let $p_1 + 1$ be the maximum of the number of points in any branch among all the T_i.

Without loss of generality we can assume that T_1 has isomorphic branches $(B_1^1, c), \ldots, (B_1^m, c)$ each with $p_1 + 1$ points and that no T_i has more than m branches at c isomorphic with (B_1^1, c). Let S be the collection of all T_i with m branches at c, each isomorphic to (B_1^1, c). Let S' be the collection of subtrees T_i that are not in S. Then if T_i is in S', T_i has exactly $m - 1$ branches at c that are isomorphic to (B_1^1, c).

If $S' = \emptyset$, then $\deg c = m + 1$ and each branch of T at c is isomorphic to (B_1^1, c). If $S' \neq \emptyset$, then T has exactly m branches at c that are isomorphic to (B_1^1, c). Let the other branches of T at c be $(B_{m+1}, c), \ldots, (B_s, c)$.

Suppose (B_1^1, c) has n_1 end points in T_1, say $u_1, \ldots, u_{n_1}$. Then there is a subtree T' of T rooted at c such that exactly $m - 1$ of the branches of T' at c are isomorphic to (B_1^1, c) and such that the set S' of subtrees can be partitioned into m sets $S_1, \ldots, S_m$ subject to the following restrictions:

(1) Each set S_j consists of exactly n_1 subtrees $T_1^j, \ldots, T_{n_1}^j$ and for each i, T_i^j has at least one branch at c that is isomorphic to $(B_1^1 - u_i, c)$.

(2) The union of the other branches at c of T_i^j is a tree which when rooted at c is isomorphic to (T', c).

Let $(C_1, c), \ldots, (C_{s-m}, c)$ be the branches of (T', c) that are not isomorphic to (B_1^1, c). Then these can be taken as the other central branches $(B_{m+1}, c), \ldots, (B_s, c)$ of T. Hence we know all the central branches of T and so T is determined.

For centred trees, we have now shown that T may be reconstructed from its subtrees T_i.

The next three theorems show how to reconstruct T from the subtrees T_i when T is bicentred. When each subtree T_i is also bicentred, the proof is easy because we make use of the result for centred trees. Otherwise the proofs are much more involved and are similar to the proofs of Theorems 5 and 6.

THEOREM 8. *Let T be a tree with at least four end points such that each subtree T_i is bicentred. Then T is determined by the T_i.*

Proof. Since at least three of the T_i are bicentred, T is bicentred, say at c_1 and c_2. Let T' be the tree obtained from T by subdividing the central line $c_1 c_2$ introducing a new point, say c. Since each T_i is bicentred, with centres c_1 and c_2, we can subdivide the central line $c_1 c_2$ of T_i by inserting a new point, c, to obtain a new collection of trees T_i' centred at c. Now apply Theorem 3 or 7 to the T_i' to obtain T'. Then T is obtained from T' by replacing the c_1–c_2 path (of length 2) by a line.

THEOREM 9. *Let T be a tree with at least five end points such that exactly two of the subtrees T_i are centred. Then T is determined by the T_i.*

The details of the proof are rather involved and similar to those in the proof of Theorem 5, and are omitted here.

THEOREM 10. *Let T be a tree with at least four end points such that exactly one of the subtrees T_i is centred. Then T is determined by the T_i.*

The proof of this theorem is also omitted. The details are similar to those in the proof of Theorem 6.

The principal result now follows from a combination of Theorems 1 through 10.

MAIN THEOREM. *Every tree T is determined by its subtrees T_i.*

From Theorem 7 we obtain the corresponding result for rooted trees.

COROLLARY 7.1. *Let T be a rooted tree with root v. Let $v_1, \ldots, v_n$ be the points of T that have degree one with $v_i \neq v$ for all i. Then T is determined by the rooted subtrees $T_i = T - v_i$.*

Proof. As in the case of ordinary trees, the degree of v in T is easily determined from the degree of v in the rooted subtrees T_i. If the degree of v in T is greater than one, Theorem 7 provides a method for reconstructing T from the subtrees T_i. Otherwise we consider two cases:

Case 1. None of the T_i contain points of degree greater than two. Then T has at most two points of degree one besides v. In either case T is easily reconstructed.

Case 2. Some subtree T_i contains a point of degree greater than two. For each such T_i, let d_i be the distance in T_i from v to the nearest point of degree greater than two. Let d be the minimum of the d_i. Then d is the distance in T from v to the nearest point, say u, of degree greater than two. Let u_i be the point of T_i at distance d from v. Let T_i' be the tree rooted at u_i obtained from T_i by deleting the $v - u_i$ path from T_i. Let T' be the tree rooted at u which is obtained from T by deleting the $v - u$ path from T. Since the degree of u in T' is greater than or equal to two, Theorem 7 provides a method for reconstructing T' from the T_i'. Then T is obtained from T' by identifying one end of a path of length d with u and rooting the resulting tree at the other end of this path.

We next have the corresponding result for oriented trees. An *oriented graph* is obtained from a graph when each line is assigned a unique direction indicated by an arrow. One can show that Theorems 5 through 10 hold for oriented trees. The proofs need only to be modified slightly. Also it can be verified that any oriented tree T with three or four end points is determined by the subtrees T_i. Thus we have the following corollary of the Main Theorem.

COROLLARY 1. *If T is an oriented tree with at least three end points, then T is determined by the oriented subtrees T_i.*

Obviously no oriented tree with exactly three points is determined by the oriented subtrees T_i. Even when T is an oriented path with more than three points, T is not necessarily determined by the collection T_i. This is shown by the two oriented trees in Figure 2. Each of these trees has the same subtrees T_1 and T_2.

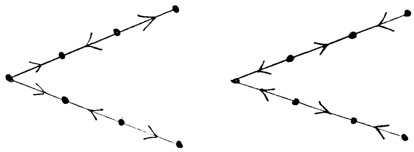

FIGURE 2

We conclude with another special case. A *signed tree* has the numbers $+1$ or -1 assigned to each of its lines. As in the case of oriented trees, Theorems 5 through 10 hold for signed trees. It can be verified that Theorem 3 also holds. Hence we have:

COROLLARY 2. *If T is a signed tree, T is determined by the collection T_i.*

Clearly one can derive the main theorem for various other species of trees by similar considerations.

REFERENCES

1. F. Harary, *On the reconstruction of a graph from a collection of subgraphs*, in M. Fiedler (ed.), *Theory of graphs and its applications* (Prague, 1964), pp. 47–52.
2. F. Harary and E. Palmer, *On similar points of a graph*, J. Math. Mech. *15* (1966), to appear.
3. P. J. Kelly, *A congruence theorem for trees*, Pacific J. Math., *7* (1957), 961–968.
4. D. König, *Theorie der endlichen und unendlichen Graphen* (Leipzig, 1936; reprinted New York, 1950).
5. S. M. Ulam, *A collection of mathematical problems* (New York, 1960), p. 29.

The University of Michigan,
Ann Arbor, Michigan