JOURNAL OF RESEARCH of the National Bureau of Standards — B. Mathematical Sciences
Volume 76B, Nos. 1 and 2, January-June 1972

Three Results for Trees, Using Mathematical Induction

W. A. Horn

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

(December 12, 1971)

The paper illustrates the use of mathematical induction to extend results which are true for a line segment to trees. Three separate theorems are stated and proved, each of which has some importance in its own right.

Key words: Helly's theorem; mathematical induction; minimum-length coverings; node-covering; trees.

1. Introduction

This paper presents three theorems each of which extends to trees a result that is known or obvious for a line segment. While the theorems are of interest in themselves, it is the purpose of the paper to use them as examples of how the inductive process may be effective in proving theorems for trees or for graphs with treelike structures. Other examples of the application of induction to trees include the proof that a tree with n nodes has $n-1$ arcs (cf. [5],[1] p. 35), or a paper such as [4].

The definition of a tree which will be used throughout is the following:

A *tree* is a connected graph containing no cycles.

2. A Minimal Node-Cover Problem for Trees

This section deals with the following problem: Given a (finite) tree T and a collection $\{T_i\}_1^m$ of subtrees, find a set of nodes P in T of least cardinality, such that

$$P \cap T_i \neq \phi$$

for all i. This problem has been considered for more complex graphs than trees (cf. [1], for example), but the following recursion algorithm, applicable to trees, is particularly simple to use.

ALGORITHM. Consider any "tip" node (node of order 1), $x \epsilon T$. Let A_x consist of x and its associated arc.

1. If there exists some $T_j = \{x\}$, take x to be in P and derive the remaining members of P as the set of nodes which solve the problem for the tree $T' = T - A_x$, with the collection of subtrees

$$\{T_i \colon i = 1,2, \ldots, m, x \notin T_i\}.$$

2. If no $T_i = \{x\}$, find P as a solution to the problem on the tree T' with the collection of subtrees $\{T_i'\}_1^m$, where

$$T_i' = T_i - A_x$$

for each i.

AMS Subject Classification: 0540.

[1] Figures in brackets indicate the literature references at the end of this paper.

THEOREM 2.1. *The above algorithm produces a minimal set* P.

PROOF. Consider first case 1, where some $T_j = \{x\}$. Let P' be some solution to the reduced problem stated in the algorithm. If P is any solution to the original problem, then clearly $x \epsilon P$, while $P - \{x\}$ meets at least every T_i not containing x. Thus by the optimality of P and P', $P - \{x\}$ and P' both have the same cardinality. Clearly $\{x\} \cup P'$ is a feasible solution for the original problem.

In the second case, if $x \epsilon T_i$ for some i, then each such T_i must contain some element besides x. Since each such T_i is connected, it must therefore contain y, the node at the other end of the arc containing x, since y cuts A_x from the rest of the tree T.

Now let P' be some optimal solution set of nodes for the reduced problem of case 2. Since $T_i' \cap P' \neq \phi$ for each i, $T_i \cap P' \neq \phi$ also, so that P' is feasible for the original problem. Let P be any optimal node set for the original problem. If $x \notin P$, then P is a feasible solution for the reduced problem, so that $|P| \geqslant |P'|$, proving the optimality of P'. On the other hand, if $x \epsilon P$, consider the new node set

$$P_1 = P - \{x\} \cup \{y\}.$$

Since $P \cap T_i \neq \phi$ for each i, it follows that $P_1 \cap T_i \neq \phi$ for each i, since if $x \epsilon T_i$ then $y \epsilon T_i$ for any i. But again P_1 is a feasible solution for the reduced problem, so that

$$|P| = |P_1| \geqslant |P'|,$$

proving the optimality of P'.

3. Minimum-Length Coverings by Intersecting Subgraphs

This section deals with an extension of a problem previously considered by the author [2].

Let G be an undirected, connected graph whose edges have nonnegative lengths. If A and B are disjoint subgraphs of G, we shall use the notation $P(A, B)$ to refer to a (any) path in G with one endpoint in A, the other in B, and no other point in $A \cup B$.

Let $\{S_i\}_1^n$ be a collection of connected nonempty subgraphs of G with the property that if $S_i \cap S_{i+1} = \phi$, for any i, then there exists a unique path $P(S_i, S_{i+1})$. A collection $\{T_i\}_1^n$ of connected subgraphs of G will be called *feasible*, relative to $\{S_i\}_1^n$, if

$$S_i \subset T_i \qquad i = 1,2,\ldots,n, \tag{3.1}$$

$$T_i \cap T_{i+1} \neq \phi \qquad i = 1,2,\ldots,n-1. \tag{3.2}$$

We treat the problem of choosing among feasible $\{T_i\}_1^n$ so as to

$$\text{minimize} \sum_{i=1}^{n} |T_i|, \tag{3.3}$$

where $|T_i|$ is the sum of the lengths of all edges in T_i. A simple algorithm solves this problem. (The problem considered in [2] was the same as that treated here, except that compact intervals on the real line replaced the subgraphs T_i, above.)

ALGORITHM.

1. Let $T_1 = S_1$.
2. Having determined T_k, find T_{k+1} as follows.

   (a) If $S_{k+1} \cap T_k \neq \phi$, set $T_{k+1} = S_{k+1}$.

   (b) If $S_{k+1} \cap T_k = \phi$, let $P(S_{k+1}, T_k)$ be any path between S_{k+1} and T_k. Set $T_{k+1} = S_{k+1} \cup P(S_{k+1}, T_k)$.

THEOREM 3.1. *The algorithm gives a minimal value for* $\Sigma|T_i|$.

PROOF. The proof is by induction on the number n of subgraphs S_i. Clearly the algorithm works for one subgraph. Assuming it works for $n-1$ subgraphs, we prove its validity for n subgraphs.

First we show that there exists an optimal feasible solution $\{T_i\}_1{}^n$, which minimizes $\Sigma|T_i|$, for which $T_1=S_1$. For if $\{T_i'\}_1{}^n$ is optimal, but $T_1' \neq S_1$, let $P(S_1,T_2')$ be any shortest path from S_1 to T_2'. Define a new collection of subgraphs

$$T_1=S_1, \tag{3.4}$$

$$T_2=T_2' \cup P(S_1,T_2'), \tag{3.5}$$

$$T_i=T_i' \qquad (3 \leq i \leq n). \tag{3.6}$$

Then it is clear that $S_i \subset T_i$, $1 \leq i \leq n$, and that T_2 is connected, since $P(S_1,T_2')$ is connected and intersects T_2'. Also, $T_i \cap T_{i+1} \neq \phi$, for all i, since $T_1 \cap T_2$ contains the common point on $P(S_1,T_2')$, and $T_i \cap T_{i+1} \subset T_i' \cap T'_{i+1}$ for $2 \leq i \leq n$. Next we note that

$$|P(S_1,T_2')| \leq |T_1' - S_1|, \tag{3.7}$$

for the connectivity of T_1', together with the fact that $T_1' \cap T_2' \neq \phi$, imply that there exists a path from S_1 to T_2' which is contained in T_1' but does not contain any arcs of S_1. Thus we have from (3.4) and (3.5),

$$\begin{aligned} |T_1|+|T_2| &\leq |S_1|+|T_2'|+|P(S_1,T_2')| \\ &\leq |S_1|+|T_2'|+|T_1'-S_1| \\ &=|T_1'|+|T_2'|, \end{aligned} \tag{3.8}$$

so that

$$\sum_{i=1}^{n} |T_i| \leq \sum_{i=1}^{n} |T_i'| \tag{3.9}$$

and the collection $\{T_i\}$ is optimal.

Now let $\{T_i\}_1{}^n$ represent any optimal collection for which $T_1=S_1$, and let $\{T_i{}^*\}_1{}^n$ be a collection obtained from using the algorithm. We first note that $T_2{}^* \subset T_2$. For if $S_1 \cap S_2 \neq \phi$, this is clear, since $T_2{}^*=S_2$. If $S_1 \cap S_2=\phi$, and $P(S_1,S_2)$ is the unique path from S_1 to S_2 in G, then $P(S_1,S_2) \subset T_2$, since if this were not the case then T_2 would be connected and contain S_2, while intersecting S_1, so that there would exist another path in T_2 from S_2 to S_1, contradicting the uniqueness of $P(S_1,S_2)$. Since $T_2{}^*=S_2 \cup P(S_1,S_2)$, by definition, $T_2{}^* \subset T_2$.

Thus the members of the collection $\{T_i\}_2{}^n$ cover the respective members of the collection $\{T_2{}^*, S_3, S_4, \ldots, S_n\}$ in G. If there exists a unique path in G between T_2^* and S_3 (or if $T_2^* \cap S_3 \neq \phi$), then the induction hypothesis applies to the set $\{T_i{}^*\}_2{}^n$, which was obtained by using the algorithm on $\{T_2^*, S_3, S_4, \ldots, S_n\}$, a set of $n-1$ elements, so that we have

$$\sum_{i=2}^{n} | T_i{}^* | \leq \sum_{i=2}^{n} | T_i |. \tag{3.10}$$

Since $|T_1|=|T_1{}^*|$, the proof is complete once it is shown that any path from $T_2{}^*$ to S_3 in G is unique.

Assuming $T_2{}^* \cap S_3=\phi$, let P and Q be two different paths from $T_2{}^*$ to S_3 in G. Since $S_2 \subset T_2{}^*$, there exist paths M and N (possibly the same path) from S_2 to the points (possibly the same point) where P and Q, respectively, meet $T_2{}^*$. But then $M \cup P$ and $N \cup Q$ are different paths from S_2 to S_3, contrary to the initial assumption.

COROLLARY 3.2. *The algorithm is valid when* G *is a tree and the* S_i *are subtrees of* G.

PROOF. We need only show that there exists a unique path between any two disjoint subtrees in a given tree. Suppose that there exist two distinct paths P and Q between subtrees A and B, having respective endpoints p_A and q_A in A and p_B and q_B in B. If $p_A = q_A$ and $p_B = q_B$, then $P = Q$, since there is a unique path between points in a tree. Therefore, assume $p_A \neq q_A$, and let R_A be a path between p_A and q_A in A. If $P \cap Q \neq \phi$, then clearly $P \cup Q \cup R_A$ contains a cycle. If $P \cap Q = \phi$, let R_B be a path between p_B and q_B in B. Then $P \cup Q \cup R_A \cup R_B$ is a cycle. Thus there exists a unique path between A and B.

4. Helly's Theorem for Trees

Helly's theorem [2] states, in one dimension, that a collection of closed connected segments on a line E^1 has a nonempty intersection if and only if every pair of segments has a nonempty intersection. In this section we extend the result to trees.

THEOREM 4.1. *A collection of subtrees of a tree has at least one common node if and only if every pair of subtrees has at least one common node.*

PROOF. The proof is by induction on the number of nodes. For a tree with 1 or 2 nodes the theorem is obvious.

Assume the theorem true for a tree with $n-1$ or fewer nodes and consider a tree T with n nodes. Let $S_1, S_2, \ldots, S_m$ be a collection of subtrees of T such that

$$S_i \cap S_j \neq \phi$$

for every pair (i, j).

Let x be any node of T which is also a cut point of T, and let $C_1, C_2, \ldots, C_k$ be the connected components of $T - \{x\}$. Let

$$A = \{x\} \cup C_1,$$

$$B = \{x\} \cup \bigcup_{i=2}^{k} C_i.$$

Then A and B are subtrees of T. Furthermore, either

$$S_i \cap A \neq \phi, \qquad \text{all } i,$$

or

$$S_i \cap B \neq \phi, \qquad \text{all } i.$$

For if this were not the case, then we would have some $S_i \subset A - \{x\}$ and some $S_j \subset B - \{x\}$, so that $S_i \cap S_j = \phi$.

Suppose, without prejudice to the argument, that

$$S_i \cap A \neq \phi, \qquad \text{all } i,$$

Let

$$S_i' = S_i \cap A$$

for each i. Then

$$S_i' \cap S_j' \neq \phi,$$

for every pair (i, j). For if $S_i' \cap S_j' = \phi$ for some (i, j), then since $S_i \cap S_j \neq \phi$ we must have a point $y \in B - \{x\}$ such that

$$y \in S_i \cap S_j.$$

But since x is a cut point and $S_i \cap A \neq \phi$, $S_j \cap A \neq \phi$, we must have $x \in S_i \cap S_j$, by the connectedness of S_i and S_j.

Thus $\{S_i'\}$ is a collection of pairwise intersecting subtrees of the finite tree A of order less than n. By the induction hypothesis, there exists a point

$$z \in \bigcap_{i=1}^{m} S_i' \subset \bigcap_{i=1}^{m} S_i,$$

proving the theorem.

5. References

[1] Edmonds, J., Covers and packings in a family of sets, AMS Bull. **68,** 494-499 (1962).

[2] Helly, E., Über Mengen konvexer Korper mit gemeinschaftlichen Punkten, Jber. Deut. Math. Verein. **32,** 175-176 (1923).

[3] Horn, W. A., Minimum-length covering by intersecting intervals, J. Res. Nat. Bur. Stand. (U.S.), **73B,** No. 1, 49-51 (1969).

[4] Horn, W. A., Optimal design of sorting networks, submitted for publication.

[5] Ore, Oystein, Graphs and Their Uses, (Random House, 1963) (paperback).

(Paper 76B1&2–359)