The Square of a Tree

By IAN C. ROSS and FRANK HARARY*

(Manuscript received June 4, 1959)

The adjacency matrix of a graph of n points is the square matrix of order n, in which the i,j element is one if and only if the ith point and the jth point are adjacent, or $i = j$; and is zero otherwise. Let A be the adjacency matrix of graph G considered as a boolean matrix so that $1 + 1 = 1$. Then G^2, the square of G, is the graph whose adjacency matrix is A^2. We obtain a necessary and sufficient condition for a graph to be the square of a tree by providing an algorithm for determining a tree that is the square root of any graph known to be the square of some tree. This algorithm cannot be carried through when a graph is not the square of a tree. It is shown that, if a graph is the square of a tree, then it has a unique tree square root. The method utilizes a previous result for determining all the cliques in a given graph, where a clique is a maximal complete subgraph. This result was obtained while attempting the more general problem of characterizing boolean matrices having a square root, or, in general, an nth root.

I. INTRODUCTION

The correspondence between graphs, matrices and relations is well known and presents an interesting field of investigation. With any graph G there is associated a symmetric square matrix of 0's and 1's, called its "adjacency matrix". Forming the boolean square of this matrix, we may call the corresponding graph "the square of G". It is an open problem (communicated to us by N. J. Fine) to characterize those graphs that have at least one square root graph, and in general those graphs that have an nth root for any positive integer n. This paper presents a partial solution to the general problem by characterizing those graphs with a tree for a square root. It is also shown that, if the graphs obtained on squaring two trees are isomorphic, then the trees themselves are isomorphic.

In the course of the development of this paper, an algorithm is given

* Princeton University and Institute for Advanced Study (presently at University of Michigan); work supported by a grant from the Office of Naval Research to the Princeton University Logistics Project.

for constructing a tree from a given graph known (or found by the characterization) to be the square of a tree. This is based on a previous method for determining all the cliques in any graph.[1]

A graph G consists of a finite collection of *points* together with certain *lines* joining pairs of distinct points. Two points of G are *adjacent* if they are joined by a line. A *complete graph* is one in which every two distinct points are adjacent. A *clique* is a maximal complete subgraph of G containing at least three points. A point of G is *cliqual* if it is in at least one clique; it is *unicliqual* if it is in exactly one clique and *multicliqual* if it is in more than one clique. Two points are *cocliqual* if there is a clique containing both of them. The *neighborhood* of a point b of G consists of b together with all points adjacent to b. A point is *neighborly* if it is cocliqual with each point in its neighborhood. Thus, every neighborly point is cliqual.

A *path* is a collection of lines of the form $b_1b_2, b_2b_3, \cdots, b_{t-1}b_t$, where these t points are distinct. A *cycle* consists of a path together with the line b_tb_1 joining its end points. A graph is *connected* if there is a path between any two points. A *tree* is a connected graph with no cycles. An *endpoint* of a graph is one which is incident to exactly one line. It is well known[2] that every tree containing more than one point has an endpoint. A *next-point* is one which is adjacent to an endpoint. An *articulation point* or *cut point* of a connected graph is one whose removal results in a disconnected graph. A *block* is a connected graph with no articulation points.

Let the points of G be $b_1, \cdots, b_p$. The *adjacency matrix* of G is the matrix $A = (a_{ij})$, where $a_{ij} = 1$ if points b_i and b_j are adjacent and $a_{ij} = 0$ otherwise, except that we take (arbitrarily) the diagonal of A to contain only 1's. By the *boolean product* of two matrices of 0's and 1's is meant their ordinary product with the stipulation that $1 + 1 = 1$. The *square of G*, written G^2, is that graph whose adjacency matrix is A^2 (all matrix products will be regarded as boolean).

II. LEMMAS

It is convenient to derive the following sequence of lemmas. In this section, we have as the hypothesis that G is a given graph known to be the square of some tree T. The tree T itself is not given; only the graph G is available and it is assumed that G has at least three points.

Lemma 1: G is complete if and only if T has exactly one next-point.

Proof: If T has exactly one next-point b, then the neighborhood of b contains all the points of T. Therefore G has exactly one clique and is complete. If G is complete, then T must have a unique next-point. For if T has two distinct next-points, then there are two distinct points of G that are endpoints of T and are not adjacent in G.

Lemma 2: Every point of G is neighborly and G is connected.

Proof: Since $G = T^2$, every point of G lies on a triangle and so is cliqual. Further, every point is neighborly since every line of G is contained in a triangle. Obviously, the square of a connected graph contains that graph and is defined on the same set of points; G is therefore connected.

Lemma 3: Every clique of G is the neighborhood of a nonendpoint of T, and conversely.

Proof: We first remark that the neighborhood of an endpoint can be a clique only in a graph of two points that we exclude. Obviously, the neighborhood of any nonendpoint b of T generates a complete subgraph of G. It remains to show that this subgraph is maximal. This is immediate, since any point not in the neighborhood of b cannot be adjacent in G to all points of the neighborhood.

To prove the converse, let C be a clique of G. If an endpoint b of T is in C, then the clique C must consist of the neighborhood of the next-point of b. For any point of G not in this neighborhood cannot be adjacent to b in G. If C does not contain an endpoint of T, let c be any point of C. If C is the neighborhood of c, there is nothing to prove. Otherwise, there is a point d of C not in the neighborhood of c. In this case, both points c and d must have some common adjacent point e. Thus C is the neighborhood of e.

Lemma 4: If G has more than one clique, then b is a unicliqual point of G if and only if b is an endpoint of T. Alternatively, we may say that if G is not complete then b is a multicliqual point of G if and only if it is a nonendpoint of T.

Proof: If b is an endpoint of T, then the only clique containing b is given by the neighborhood of the next-point of b. If b is unicliqual in G, then it cannot be adjacent to two distinct points of T. For b would then belong to more than one clique, since by hypothesis G contains more than one clique.

Lemma 5: Two nonendpoints of T are adjacent in T if and only if their neighborhoods are cliques of G that intersect in exactly two points.

Proof: If two nonendpoints of T are adjacent in T, then they are both contained in the cliques in G given by their neighborhoods in T in accordance with Lemma 3. These two nonendpoints cannot both be adjacent in T to a common third point since T has no cycles. Hence the intersection of these two neighborhoods contains exactly these two points.

Conversely, if two distinct cliques of G intersect in exactly two points, then the corresponding nonendpoints in T (in accordance with Lemma 3) are adjacent in T. For if these two points are not adjacent, their neighborhoods in T can intersect in at most one point since T is a tree and has no cycles.

Lemma 6: G is a block.

Proof: Lemma 2 shows that G is connected. Points that are adjacent in T are also adjacent in G. Two points both adjacent to any given point in T are adjacent to each other in G. Therefore the removal of any point of G cannot disconnect it.

III. CHARACTERIZATION

We now state conditions that we shall see characterize every graph that is the square of a tree and demonstrate that, if $G = T^2$, then G satisfies these conditions. In the next section we shall present an algorithm for finding a tree root of such a graph, and shall show that the algorithm can be applied to graphs that meet the conditions of the present section. Therefore, a graph will be shown to meet these conditions if and only if it has a square root that is a tree.

The characterization has two cases. In Case 1, $G = K_p$, the complete graph of p points; in case 2, $G \neq K_p$.

Case 1

Consider the tree consisting of one point joined with all the others. When this tree is squared, the result is the complete graph. We illustrate with Fig. 1, in which $T^2 = K_5$.

Case 2

Consider the following five conditions for a graph $G \neq K_p$:

i. Every point of G is neighborly and G is connected.

ii. If two cliques meet at only one point b, then there is a third clique with which they share b and exactly one other point.

Fig. 1 — Graphs for Case 1.

iii. There is a one-to-one correspondence between the cliques and the multicliqual points b of G such that the clique $C(b)$ corresponding to b contains exactly as many multicliqual points as the number of cliques which include b.

iv. No two cliques intersect in more than two points.

v. The number of pairs of cliques that meet in two points is one less than the number of cliques.

We now prove that if there is a tree such that $T^2 = G$, then G will either be complete or will satisfy the above five conditions. We have already settled Case 1 in Lemma 1 and Fig. 1. We now turn to Case 2, where $G \neq K_p$.

The necessity of the first four of these conditions for a graph $G \neq K_p$ that is the square of a tree follows readily from the lemmas. Condition i is a restatement of Lemma 2. Condition ii follows from Lemmas 3 and 5, for, if $C_1 \cap C_2 = \{b\}$, then neither C_1 nor C_2 can be the clique associated with point b in accordance with Lemma 3. Hence, there is a third clique C_3 corresponding to b, and C_3 has the relationship stated in ii with C_1 and C_2 by Lemma 5. Condition iii is a combination of Lemmas 3 and 4, while iv is implied by Lemma 5. Condition v follows immediately from the well-known theorem that for trees the number of points exceeds the number of lines by one (see p. 51, satz 9 of Ref. 2).

IV. ALGORITHM AND THEOREM

If the graph G is complete, we have already seen that there is a unique tree T such that $G = T^2$, and we have illustrated this with Fig. 1.

Now, let G be a graph that is not complete and satisfies the five conditions of Section III. We now show how to construct a unique tree T (up to isomorphism) that is a square root of G.

Step 1

Find all the cliques of G in accordance with the method of Ref. 1.

Step 2

Let the cliques of G be $C_1, \cdots, C_n$. Then $n > 1$, since G is not complete and condition i holds. Consider a collection of multicliqual points $b_1, \cdots, b_n$ corresponding to these cliques in accordance with condition iii. These are to be the nonendpoints of the tree T being constructed. Find all the pairwise intersections of the n cliques. By condition iv, no cliques meet in more than two points. We now form a graph S by joining the points b_i and b_j by a line if and only if the corresponding cliques C_i and C_j intersect in two points. By condition v, S is a tree.

Step 3

For each clique C_i of G, let n_i be the number of unicliqual points. Thus n_i is a nonnegative integer. To the tree S obtained in Step 2 attach n_i endpoints to b_i, obtaining a tree T. Since the points of G are either multicliqual or unicliqual and the points of T are either nonendpoints or endpoints, the number of points of G and T is equal.

It is clear that the tree T constructed by this algorithm is a tree square root of G. It also follows that the five conditions of the preceding section are sufficient for an incomplete graph to be the square of a tree. Further, the algorithm results in a unique tree T, up to isomorphism.

We have just proved the following theorem about the relationship between trees and their squares:

Theorem: Let T_1 and T_2 be trees. If T_1^2 and T_2^2 are isomorphic, then T_1 and T_2 are isomorphic.

V. EXAMPLE

We illustrate the algorithm with the graph G of Fig. 2, which meets the conditions for being the square of a tree. In this graph we indicate the points for convenience by the numerals 1, . . . , 8.

By the method of Ref. 1 we find all the cliques of G. There are four cliques whose composition is as follows:

$$C_1: \ 1256,$$
$$C_2: \ 1234,$$
$$C_3: \ 237,$$
$$C_4: \ 248.$$

In accordance with Step 2 of the algorithm, we introduce the points b_1, b_2, b_3 and b_4. Among the four cliques there are pairwise intersections between C_2 and each of the other cliques, but the other cliques have only single point intersections with each other. We therefore obtain the graph S shown in Fig. 3, which is a subtree of the tree T under construction.

We now proceed in accordance with Step 3. Let U_i be the set of unicliqual points of C_i; then we have:

$$U_1: \ 56,$$
$$U_2: \ —,$$
$$U_3: \ 7,$$
$$U_4: \ 8.$$

On joining each point b_i to the number of points in the set U_i we obtain the tree T shown in Fig. 4.

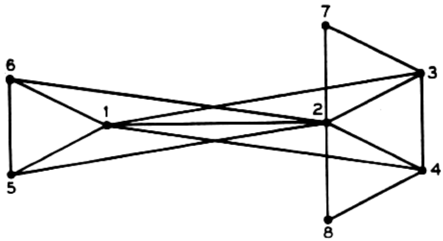

Fig. 2 — Graph G for the example.

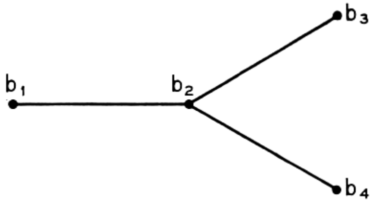

Fig. 3 — Graph S for the example.

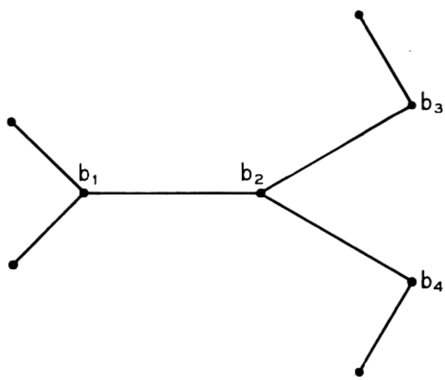

Fig. 4 — Graph T for the example.

REFERENCES

1. Harary, F. and Ross, I. C., A Procedure for Clique Detection Using the Group Matrix, Sociometry, **20**, 1957, p. 205.
2. König, D., *Theorie der endlichen und unendlichen Graphen*, Academische Verlagsgesellschaft, Leipzig, 1936.