INCIDENCE MATRICES WITH THE CONSECUTIVE 1's PROPERTY

BY D. R. FULKERSON AND O. A. GROSS[1]

Communicated by R. C. Buck, May 13, 1964

1. **Introduction.** Let $A=(a_{ij})$ be an m-by-n matrix whose entries a_{ij} are all either 0 or 1. For certain applications, one of which will be discussed below, it is of interest to know whether there is an m-by-m permutation matrix P such that the 1's in each column of PA occur in consecutive positions. In this note we state certain results that have relevance for this problem. Proofs of these, together with an efficient computational method for deciding the question in any given case, will be published elsewhere.

The problem posed above includes that of determining whether a given finite undirected graph is an interval graph. The study of interval graphs [2], [3], [4], [5] was stimulated in part by an application concerning the fine structure of genes. A basic genetic problem, discussed in [1], is to decide whether or not the sub-elements of genes are linked together in a linear order. A way of approaching this problem is also described in [1]. Briefly, it is as follows. For certain microorganisms, there are a standard form and mutants, the latter arising from the standard form by alteration of some connected portion of the genetic structure. Experiments can be devised for determining whether or not the blemished parts of two mutant genes intersect or not. Thus the mathematical problem becomes: Given a large number of mutants together with intersection data on pairs of mutants, to decide whether this information is compatible with a linear model of the gene. If one represents the intersection data by a graph (two mutants, i.e., vertices, being joined by an edge if their blemished portions intersect), the problem is to decide whether this graph is an interval graph.

2. **A basic theorem.** We say that a (0, 1)-matrix A has the *consecutive 1's property* (for columns) if there is a permutation matrix P such that the 1's in each column of PA occur consecutively. The first question that naturally arises is how much information about A is needed to decide whether it has the property or not. Do we need to

[1] This research is sponsored by the United States Air Force under Project RAND—contract No. AF 49(638)-700 monitored by the Directorate of Development Planning, Deputy Chief of Staff, Research and Development, Hq USAF. Views or conclusions contained in this Memorandum should not be interpreted as representing the official opinion or policy of the United States Air Force.

know A itself, or will something less suffice? Theorem 2.1 below provides a partial answer to this question; it shows that a knowledge of the matrix A^TA is enough. Here A^T denotes the transpose of A.

THEOREM 2.1. *Let A and B be* (0, 1)*-matrices satisfying*

$$A^TA = B^TB. \tag{2.1}$$

Then either both A and B have the consecutive 1*'s property or neither does. Moreover, if A and B have the same number of rows and A has the property, then there is a permutation P such that $B = PA$.*

The first part of Theorem 2.1 follows easily from the second. The second assertion can be proved by induction on the number of columns of A.

In view of Theorem 2.1, it would be interesting to know conditions on A^TA in order that A have the consecutive 1's property. Later on we shall state a theorem which reduces this question to the consideration of (0, 1)-matrices having connected "overlap graphs." For such matrices, there is a simple construction for testing the property, but we do not know explicit necessary and sufficient conditions.

3. The overlap graph and component graph. Let a and b be (0, 1)-vectors having m components. Their inner product $a \cdot b$ satisfies

$$0 \leqq a\cdot b \leqq \min(a\cdot a,\ b\cdot b). \tag{3.1}$$

If strict inequality holds throughout (3.1), we say that a and b ***overlap***. We also say that a *contains* b if

$$a\cdot b = b\cdot b. \tag{3.2}$$

Now let A be an m-by-n (0, 1)-matrix having column vectors a_j, $j = 1, 2, \cdots, n$. It is convenient, and presents no loss of generality in studying the consecutive 1's property, to assume that $a_j \neq 0$, $j = 1, 2, \cdots, n$, and that $a_i \neq a_j$ for $i \neq j$. We refer to such an A as *proper*.

There are various graphs one can associate with a (0, 1)-matrix A that are meaningful insofar as the consecutive 1's property is concerned. We describe two such graphs, one being an undirected graph, the other a directed graph. The first of these is obtained from A by taking vertices $x_1, x_2, \cdots, x_n$ corresponding to the columns $a_1, a_2, \cdots, a_n$ of A, and putting in undirected edges (x_i, x_j) corresponding to overlapping column vectors a_i and a_j. We call this the *overlap graph* of A and denote it by $\mathcal{G} = \mathcal{G}(A)$. The overlap graph of A splits up into connected components $\mathcal{G}_1, \mathcal{G}_2, \cdots, \mathcal{G}_p$, and this decomposition yields a corresponding partition of A into m-rowed submatrices $A_1, A_2, \cdots,$

A_p. We now form a second (directed) graph by taking vertices $X_1, X_2, \cdots, X_p$ corresponding to these submatrices, and putting in an edge $[X_i, X_j]$ directed from X_i to X_j, if there is a column vector a of A_i and a column vector b of A_j such that a contains b. We call this directed graph the *component graph* of A and denote it by $\mathfrak{D}=\mathfrak{D}(A)$.

The following theorem may be established in a straightforward manner.

THEOREM 3.1. *The component graph* $\mathfrak{D}(A)$ *of a proper* (0, 1)*-matrix* A *is acyclic and transitive.*

That is, $\mathfrak{D}(A)$ contains no directed cycles, and if $[X, Y]$ and $[Y, Z]$ are edges, then $[X, Z]$ is an edge. Thus $\mathfrak{D}(A)$ is the graph of a partial ordering. This partial ordering of components of $\mathcal{G}(A)$ is special in the sense that an element can have at most one immediate predecessor. Thus if we omit from $\mathfrak{D}$ every edge whose existence is implied by transitivity, the resulting graph is simply a collection of rooted trees.

The structure of the component graph $\mathfrak{D}(A)$ is useful in establishing the decomposition theorem of the next section.

4. A decomposition theorem. For an arbitrary (0, 1)-matrix A, we can rearrange columns and write

$$A = (A_1, A_2, \cdots, A_p), \tag{4.1}$$

where each submatrix A_k, $k=1, 2, \cdots, p$, corresponds to a component of the overlap graph $\mathcal{G}(A)$. We term (4.1) an *overlap decomposition* of A, and refer to the submatrices A_k as *components* of A. If A has just one component, we say that A is *connected.*

THEOREM 4.1. *A* (0, 1)*-matrix* A *has the consecutive* 1*'s property if and only if each of its components has the property.*

Necessity in Theorem 4.1 is of course trivial. Sufficiency can be established by induction on the number of components of a proper A. The induction step proceeds by deleting a component of A which corresponds to a minimal element in the partial ordering given by $\mathfrak{D}(A)$.

Theorem 4.1 effectively solves the problem posed in §1, since one can describe a very simple and efficient procedure for testing whether or not a connected matrix has the consecutive 1's property. Moreover, having arranged each individual component of a disconnected A so that its 1's appear consecutively in each column, the proof of Theorem 4.1 indicates how to fit these components together so as to yield a permuted form of A which has consecutive 1's in each column. The entire process is computationally efficient, requiring no more than $O(n^2)$ steps if A has n columns.

5. **Application to interval graphs.** A graph $\mathcal{G}$ (finite, undirected, without multiple edges or loops) is an interval graph provided $\mathcal{G}$ can be represented as the intersection graph of a set of intervals on the real line. The theorems and methods described in preceding sections can be applied to the problem of determining when a graph $\mathcal{G}$ is an interval graph by considering a certain incidence matrix which specifies $\mathcal{G}$. We term this incidence matrix the *dominant-clique-vs.-vertex* matrix, and define it as follows. First of all, a *clique* in $\mathcal{G}$ is a set of vertices, every two of which are joined by an edge. We may partially order the set of all cliques of $\mathcal{G}$ by inclusion. The maximal elements in this ordering will be termed *dominant cliques.* Since two vertices of $\mathcal{G}$ are joined by an edge if and only if they belong to some dominant clique, the dominant-clique-vs.-vertex incidence matrix characterizes $\mathcal{G}$.

THEOREM 5.1. *A graph $\mathcal{G}$ is an interval graph if and only if the dominant-clique-vs.-vertex incidence matrix of $\mathcal{G}$ has the consecutive* 1*'s property.*

We also note that an interval graph is necessarily a rigid-circuit graph [2], and that one can describe a simple method to test for the rigid-circuit property. (A graph is a rigid-circuit graph if every circuit with more than three vertices has a chord. The test is based on the known fact that such a graph always contains simplicial vertices, a simplicial vertex being one whose neighboring vertices are a clique [2], [3].) If the test succeeds, the method automatically generates all dominant cliques. Thus to discover if $\mathcal{G}$ is an interval graph, one can first apply an easy test for the rigid-circuit property, and then test the resulting dominant-clique-vs.-vertex incidence matrix for the consecutive 1's property.

REFERENCES

1. S. Benzer, *On the topology of the genetic fine structure*, Proc. Nat. Acad. Sci. U.S.A. **45** (1959), 1607.

2. G. A. Dirac, *On rigid circuit graphs*, Abh. Math. Sem. Univ. Hamburg **25** (1961), 72.

3. C. G. Lekkerkerker and J. Ch. Boland, *Representation of a finite graph by a set of intervals on the real line*, Fund. Math. **51** (1962), 45.

4. G. Hajos, *Über eine Art von Graphen*, Internat. Math. Nachr. **11** (1957), 65.

5. P. C. Gilmore and A. J. Hoffman, *A characterization of comparability graphs and of interval graphs*, Canad. J. Math. **16** (1964), 539–548.

RAND CORPORATION, SANTA MONICA, CALIFORNIA