GENERALIZATIONS OF THE CONSECUTIVE ONES PROPERTY AND RELATED NP-COMPLETE PROBLEMS

Witold Lipski, Jr.*

Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801, USA

Abstract

A (0,1)-matrix A is said to have the consecutive ones property if its rows can be permuted so that the 1's appear consequtively in each column. We present four NP-complete problems connected with some generalizations of this notion. These problems concern decomposing the columns of a matrix into two subsets having the consecutive ones property, decomposing the rows into three subsets having the consecutive ones property, finding a subset of rows of maximal size having the consecutive ones property, and finding a permutation of the rows such that the 1's in any column are contained in a set of k consecutive rows, for a fixed "buffer size" k.

Key Words: Computational Complexity, NP-complete Problems, Matrices with the Consecutive Ones Property, Linear Families of Sets, File Organization, Consecutive Retrieval Property, Storage Space Minimization.

---

*On leave as a Fulbright Scholar from Institute of Computer Science, Polish Academy of Sciences, P. O. Box 22, 00-309 Warsaw PKiN, Poland.

1. INTRODUCTION

A (0,1)-matrix is said to have the <u>consecutive ones property</u> if its rows can be permuted so that the 1's appear consecutively in each column. For instance, the matrix

$$A = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

has this property, since the following permutation of rows brings together all ones in any column:

$$\begin{array}{c} 2 \\ 5 \\ 3 \\ 1 \\ 4 \end{array} \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

The matrix

$$B = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

(and, of course, any matrix containing B as a submatrix) is an example of a matrix without the consecutive ones property.

Matrices with the consecutive ones property have been introduced by Fulkerson and Gross [5] as a tool to investigate interval graphs. They received considerable attention when it turned out they could be applied to the design of efficient file organizations (see Ghosh [8], Lipski [12]). Still other applications, such as the so-called sequence dating in archeology, are discussed in the book of Roberts [17].

Let $\mathcal{M}$ be a family of subsets of a finite set X. $\mathcal{M}$ is said to be <u>linear</u> (see [12]) if there is a sequence

(1) $$x_1, x_2, \ldots, x_n$$

of elements of X such that

(i) every $x \in X$ occurs exactly once in the sequence, and

(ii) every $M \in \mathcal{M}$ appears as a <u>segment</u>, i.e. a set of $|M|$ consecutive terms of the sequence ($|M|$ denotes the cardinality of M).

It is obvious that a family $\mathcal{M} = \{M_1, \ldots, M_m\}$ of subsets of $X = \{x_1, \ldots, x_n\}$ is linear iff its <u>incidence matrix</u> $A = [a_{ij}]$, defined by

$$a_{ij} = \begin{cases} 1 & \text{if } x_i \in M_j \\ 0 & \text{if } x_i \notin M_j \end{cases}$$

has the consecutive ones property.

Several characterizations of linear families (or, equivalently, of matrices with the consecutive ones property) are known, see e.g. Tucker [19], Lipski [12], Nakano [15]. Efficient polynomial time algorithms to test for the linearity of a family of sets (and finding a suitable sequence (1), if there is one) have also been proposed, see e.g. Fulkerson and Gross [5], Lipski [12] and Booth and Lueker [3]. The algorithm of Booth and Lueker runs in linear time, more exactly, it requires $\mathcal{O}(m + n + f)$ steps when applied to a family $\mathcal{M} = \{M_1, \ldots, M_m\}$ of subsets of a set X, with $|X| = n$, $\Sigma_{i=1}^{m} |M_i| = f$.

Since the generalizations of the consecutive ones property we are going to consider are mostly connected with file organization problems, it is useful to describe this connection in some detail. We may treat X as the set of <u>records</u> in a file, and identify any $M \in \mathcal{M}$ with a <u>query</u>, more exactly, with the set of records relevant to this query. If $\mathcal{M}$ is linear then we can arrange the records without duplications in a linear storage in such a way that the response to each query can be retrieved as a set of consecutive records; consequently, our arrangement minimizes both the storage space and access time. Unfortunately, in most practical situations the class of linear families turns out to be too narrow to provide a basis for an efficient file organization. This fact rises the need to extend the class of linear families at a cost of decreasing the efficiency of the corresponding file organization.

One possibility is to relax condition (i) in the definition of a linear family, i.e., to allow repetitions in sequence (1). The problem of minimizing the number of repetitions can be formulated as a yes-no problem in the following way:

PROBLEM 1.

<u>Given</u>: A finite set X, a family $\mathcal{M}$ of subsets of X, and a nonnegative integer k.

<u>Question</u>: Does there exist a sequence of elements of X of length k such that every $M \in \mathcal{M}$ appears as a set of $|M|$ consecutive terms of the sequence?

Alternatively, we can drop condition (ii), i.e., allow each $M \in \mathcal{M}$ to consist of several disjoint segments, called blocks. In such a case we are led to

PROBLEM 2.

Given: A finite set X, a family $\mathcal{M}$ of subsets of X, and a nonnegative integer k.

Question: Does there exist a sequence of elements of X, without repetitions, such that the total number of blocks corresponding to all $M \in \mathcal{M}$ does not exceed k?

Unfortunately, there is a strong evidence that no efficient algorithms to solve Problems 1, 2 exist, since both problems have been proved NP-complete by Kou [10] (see Aho et al. [1] for a discussion of NP-complete problems). Further examples of combinatorial problems related to the consecutive ones property include the following two problems proved NP-complete by Booth [2]:

PROBLEM 3.

Given: A (0,1)-matrix A and a nonnegative integer k.

Question: Is it possible to transform A into a matrix with the consecutive ones property by replacing at most k 0's by 1's?

PROBLEM 4.

Given: A (0,1)-matrix A and a nonnegative integer k.

Question: Does there exist a subset of k columns of A which defines a submatrix with the consecutive ones property?

Still another problem is obtained if we consider the "two-dimensional" (2D) file organization on drum-type storage proposed by Ghosh [9]. A family $\mathcal{M}$ of subsets of a finite set X is said to be <u>2D linear</u> if there is a positive integer k and a sequence $X_1,\ldots,X_k$ of pairwise disjoint sets with $X_1 \cup \ldots \cup X_k = X$ such that

(i) $|M \cap X_i| \leq 1$ for all $M \in \mathcal{M}$, $1 \leq i \leq k$, and

(ii) every $M \in \mathcal{M}$ is contained in the union $X_i \cup X_{i+1} \cup \ldots \cup X_{i+|M|-1}$ for a suitable $i \leq k$.

(Each of the sets X_i corresponds to the set of records accessible to a set of recording heads of a drum-type storage at an instant of time; only one head is activated at any instant of time.)

PROBLEM 5.

<u>Given</u>: A family $\mathcal{M}$ of subsets of a finite set X.

<u>Question</u>: Is $\mathcal{M}$ 2D linear?

This problem has also been proved NP-complete (see Lipski [13]).

In this paper we give four other problems related to the consecutive ones property. Three of them are listed below:

PROBLEM 6.

<u>Given</u>: A family $\mathcal{M}$ of subsets of a finite set X.

<u>Question</u>: Can $\mathcal{M}$ be partitioned into two linear families?

For any family $\mathcal{M}$ of subsets of X, and any $Y \subseteq X$ we shall denote

$$\mathcal{M}|_Y = \{M \cap Y : M \in \mathcal{M}\}.$$

It is easy to see that if $\mathcal{M}$ is linear then $\mathcal{M}|_Y$ is also linear, for any $Y \subseteq X$.

PROBLEM 7.

Given: A family $\mathcal{M}$ of subsets of a finite set X.

Question: Does there exist a partition $X = X_1 \cup X_2 \cup X_3$ such that the families $\mathcal{M}|_{X_i}$, $i = 1,2,3$ are linear?

PROBLEM 8.

Given: A family $\mathcal{M}$ of subsets of a finite set X and a nonnegative integer k.

Question: Does there exist a subset $Y \subseteq X$ such that $|Y| \geq k$ and $\mathcal{M}|_Y$ is linear?

The last problem we are going to consider concerns a file organization technique proposed recently by Tanaka et al. [18]. A family $\mathcal{M}$ of subsets of X will be called quasi-linear with buffer size k, if there exists a sequence of elements of X such that

(i) every $x \in X$ occurs exactly once in the sequence, and

(ii) every $M \in \mathcal{M}$ is contained in a set of k consecutive terms of the sequence.

PROBLEM 9.

Given: A family $\mathcal{M}$ of subsets of a finite set X and a positive integer k.

Question: Is $\mathcal{M}$ quasi-linear with upper size k?

2. PROOFS OF THE RESULTS

It is clear that each of Problems 6-9 can be solved in polynomial time by a usual "guess-and-check" non-deterministic algorithm. In order to prove that our problems are NP-complete, it suffices to show a (deterministic) polynomial time transformation from a known NP-complete problem to each of them (see Aho et al. [1]).

Theorem 1: Problem 6 is NP-complete.

Proof: We shall show that Undirected Hamiltonian Path With Degree At Most 3 (UHP3) transforms to Problem 6. We recall the formulation of UHP3: Given an undirected graph G with the degree of each vertex less or equal to 3, decide whether G contains a Hamiltonian path. This problem was proved NP-complete by Garey et al. [7] (actually, a circuit version is considered in [7], but the result can easily be extended to the path version).

Let $G = \langle V,E \rangle$ be an undirected graph with the set of vertices V, the set of edges E, such that the degree $d(v)$ of any vertex $v \in X$ is at most 3. (An edge $e \in E$ joining vertices $u,v \in V$ is identified with the 2-element set $\{u,v\}$.) We extend our graph by adding $4-d(v)$ pending edges incident with vertex v, for any $v \in V$. In the resulting graph, $G^* = \langle V^*,E^* \rangle$, every "old" vertex $v \in V$ has degree 4, and every "new" vertex $v \in V^*-V$ has degree 1. Define $\mathcal{M} = E^* \cup \{V\}$.

Claim: G contains a Hamiltonian path iff $\mathcal{M}$ can be partitioned into two linear subfamilies.

Suppose first that G contains a Hamiltonian path $v_1,v_2,\ldots v_n$ ($n = |V|$, $v_i \neq v_j$ for $i \neq j$, and $\{v_i,v_{i+1}\} \in E$ for $1 \leq i < n$). Let $v_0,v_{n+1} \in V^*-V$ be adjacent to v_1 and v_n, respectively. Define

$$\mathcal{M}_1 = \{\{v_i, v_{i+1}\}: \ 0 \leq i \leq n\} \cup \{V\},$$
$$\mathcal{M}_2 = \mathcal{M} - \mathcal{M}_1.$$

It is easy to see that $\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2$ is a partition of $\mathcal{M}$ into two linear families. The linearity of $\mathcal{M}_1$ is obvious, and the linearity of $\mathcal{M}_2$ follows from the fact that the graph $\langle V^*, \mathcal{M}_2\rangle$ is a collection of vertex-disjoint paths of length 2 or 3 (see Fig. 1).

Conversely, suppose that there exists a partition $\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2$, where $\mathcal{M}_1$, $\mathcal{M}_2$ are linear. Without loss of generality we may assume that $V \in \mathcal{M}_1$. Every $v \in V$ is incident with four different edges in E^*, and it is clear that two of them must be in $\mathcal{M}_1$ and the other two in $\mathcal{M}_2$. Any sequence of elements of V^* which realizes the linearity of $\mathcal{M}_1$ contains a subsequence of consecutive terms $v_1, \ldots v_n$ such that $\{v_1, \ldots, v_n\} = V$. Consequently, $\{v_i, v_{i+1}\} \in \mathcal{M}_1$, $1 \leq i < n$, which means that G contains a Hamiltonian path.

It is obvious that the family $\mathcal{M}$ can be constructed in time bounded by a polynomial in the size of G. This completes the proof. □

We note that a similar result can be proved for partitioning $\mathcal{M}$ linear subfamilies, for any fixed $k \geq 2$. The only change in the proof is that we add $2k-d(v)$ (rather than $4-d(v)$) pending edges incident with any $v \in V$.

By the theorem just proved, we should not expect efficient algorithms for partitioning a given family of sets into the minimal possible number of linear subfamilies. This does not mean that the technique of decomposing into linear subfamilies cannot be of practical value. A very simple situation arises if we restrict any subfamily in the decomposition to consist of at most two subsets (of course, <u>any</u> family consisting of two subsets is linear). From the file organization point of view (storage space minimization) it is desirable to find a decomposition $\mathcal{M} = \{M_1, M_2\} \cup \{M_3, M_4\} \cup \ldots \cup \{M_{n-1}, M_n\}$ such that

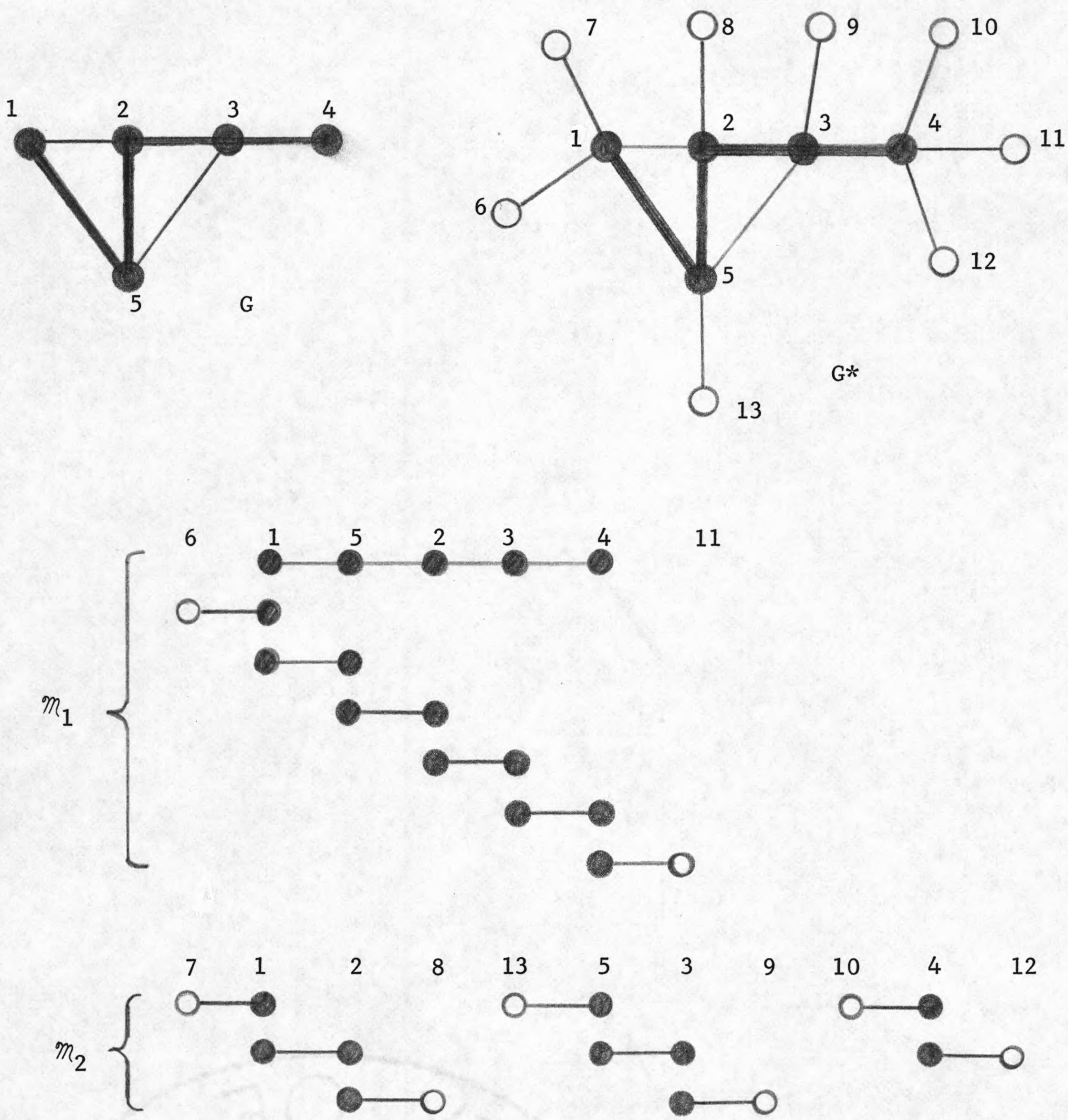

Fig. 1. A graph G with a Hamiltonian path, and the corresponding partition of $\mathfrak{M}$ into linear families.

$\Sigma_{k=1}^{n/2}|M_{2k-1} \cup M_{2k}| = \Sigma_{i=1}^{n}|M_i| - \Sigma_{k=1}^{n/2}|M_{2k-1} \cap M_{2k}|$ is minimal, i.e. $\Sigma_{k=1}^{n/2}|M_{2k-1} \cap M_{2k}|$ is maximal, see Lipski and Marek [14] (we assume that n is even, since we may extend $\mathcal{M}$ by adding a dummy subset $M_{n+1} = \emptyset$). An optimal decomposition of this type can be obtained in the following way. We construct a complete graph whose vertices correspond to (and are identified with) subsets in $\mathcal{M}$, and we associate the weight equal to $|M \cap N|$ with any edge $\{M,N\}$ of this graph. It is easy to see that any optimal solution corresponds to a maximum weighted matching in our graph. Hence we may use the (polynomial time) maximum weighted matching algorithm of Edmonds [4] (see also Lawler [11]).

<u>Theorem 2</u>: Problem 7 is NP-complete.

<u>Proof</u>: The NP-complete problem which we shall transform to Problem 7 is 3-colorability (see Garey <u>et al</u>. [7]). We recall the formulation of 3-colorability: Given an undirected graph G, decide whether its vertices can be colored by using three colors in such a way that adjacent vertices are always assigned different colors. Let $G = \langle V,E \rangle$ be an undirected graph. We split every vertex $v \in V$ into 5 copies $v^{(1)},\dots,v^{(5)}$ and we join $u^{(i)}$ with $v^{(j)}$ iff u and v are joined in G. It is clear that the resulting graph $G^* = \langle V^*,E^* \rangle$, where

$$E^* = \{\{u^{(i)},v^{(j)}\}: \{u,v\} \in G \wedge 1 \leq i,j \leq 5\},$$

is 3-colorable iff G is 3-colorable. Now we put $X = V^*$, $\mathcal{M} = E^*$.

<u>Claim</u>: G is 3-colorable iff there is a partition $X = X_1 \cup X_2 \cup X_3$ such that $\mathcal{M}|_{X_i}$, $i = 1,2,3$ are linear.

Assume first that there exists a proper coloration of the vertices of G by three colors. For any $v \in V$, extend the color of v to all copies $v^{(1)},\dots,v^{(5)}$, and define X_i to be the set of vertices of ith color in V^*, $i = 1,2,3$. Then every $\mathcal{M}|_{X_i}$, $i = 1,2,3$ is trivially linear, since it is composed of subsets of cardinality at most 1.

Conversely, suppose that there exists the required partition $X = X_1 \cup X_2 \cup X_3$, and assign the ith color to all vertices of G* which are in X_i, $i = 1,2,3$. Let $A = \{u^{(1)},\ldots,u^{(5)}\}$, $B = \{v^{(1)},\ldots,v^{(5)}\}$, where $\{u,v\} \in E$. We shall prove that A contains a "large" monochromatic subset, consisting of at least three vertices. Indeed, if no such subset exists then we may assume without loss of generality that two vertices in A are assigned the first color, two vertices the second color, and one vertex the third color. The distribution of colors in B must also be of the type $2+2+1$, since otherwise one of the families $\mathcal{M}|X_i$ would contain a subfamily of the form $\{\{a,b\}, \{a,c\}, \{a,d\}\}$ $(a \in A,\ b,c,d \in B)$, which is evidently nonlinear. It follows that there are four vertices of the same color, $a,b \in A$ and $c,d \in B$, and consequently one of the families $\mathcal{M}|X_i$ contains $\{\{a,c\}, \{c,b\}, \{b,d\}, \{d,a\}\}$, a nonlinear subfamily. This contradiction proves the existence of a (trivially unique) large monochromatic subset of A, and more generally, of any subset $\{v^{(1)},\ldots,v^{(5)}\}$, $v \in V$ such that v is not isolated in G. Of course, if $\{u,v\} \in E$ then the large monochromatic subsets of $\{u^{(1)},\ldots,u^{(5)}\}$ and $\{v^{(1)},\ldots,v^{(5)}\}$ are of different colors. Now assign to every non-isolated $v \in V$ the color of the large monochromatic subset of $\{v^{(1)},\ldots,v^{(5)}\}$, and the first color to all isolated vertices. This defines a proper 3-coloring of G, and the proof of the claim is completed.

Now the theorem follows from the evident fact that the transformation which constructs $\mathcal{M}$ for a given G can be carried out in polynomial time. □

An interesting open problem is whether a similar result holds for partitioning X into two, rather than three, subsets.

<u>Theorem 3</u>: Problem 8 is NP-complete.

<u>Proof:</u> We shall transform Vertex Cover (see [1]), to Problem 8. First let us recall the formulation of Vertex Cover: Given an undirected graph

$G = \langle V,E\rangle$ and a nonnegative integer k, decide whether there exists a set S (called a vertex cover) of at most k vertices with the property that every edge of G is incident with some vertex in S.

Consider an instance of Vertex Cover, i.e. an undirected graph $G = \langle V,E\rangle$, and an integer $k \geq 0$. Let $V = \{v_1,v_3,v_5,\ldots,v_{2n-1}\}$, and let $X = \{v_1,v_2,v_3,\ldots,v_{2n-1}\}$, where $v_2,v_4,\ldots,v_{2n-2}$ are some new elements not in V. Define

$$S_i = \{v_1,v_2,\ldots,v_i\}, \quad 1 \leq i \leq 2n-2,$$

$$\mathcal{A} = \{S_i : 1 \leq i \leq 2n-2\} \cup \{X-S_i : 1 \leq i \leq 2n-2\},$$

$$\mathcal{M} = \mathcal{A} \cup E .$$

<u>Claim</u>: There is a vertex cover of cardinality at most k in G iff there is a subset $Y \subseteq X$ such that $|Y| \geq 2n-1-k$ and $\mathcal{M}_{|Y}$ is linear.

Assume that S is a vertex cover in G, $|S| \geq k$, and put $Y = X-S$. Then $|Y| \geq 2n-1-k$ and $\mathcal{M}_{|Y}$ is linear. Indeed, $|M \cap Y| \leq 1$ for any $M \in E$, and $\mathcal{A}_{|Y}$ is obviously linear, since $\mathcal{A}$ is linear. Notice that $\mathcal{A}$ keeps a fixed ordering (up to reversal) on the elements of X, and $\mathcal{A}_{|Y}$ keeps the same ordering on the elements of Y. Suppose now that there is a subset $Y \subseteq X$ with $|Y| \geq 2n-1-k$ such that $\mathcal{M}_{|Y}$ is linear. Assume that $v_i,v_j \in Y$, $i < j$, and $\{v_i,v_j\} \in E$. Then it is easily seen that $v_p \notin Y$ for $i < p < j$, and that we may replace v_i in Y by v_{i+1}: $\mathcal{M}_{|(Y-\{v_i\}) \cup \{v_{i+1}\}}$ is linear. By applying these replacements repeatedly, we obtain a subset $Y' \subseteq X$ such that $|Y'| = |Y| \geq 2n-1-k$, $\mathcal{M}_{|Y}$ is linear, and Y' does not contain any edge of G. Define $S = (X-Y') \cap V$. Then S is a vertex cover in G, and $|S| \leq k$. This proves the claim, and the whole theorem, since our transformation can evidently be done in polynomial time. $\square$

Before we prove that Problem 9 is NP-complete, we shall need some definitions. Let $G = \langle V,E\rangle$ be an undirected graph, and let f be a sequence $v_1,\ldots,v_n$ containing each vertex $v \in V$ exactly once. We call every such sequence a <u>layout</u> of G, and we define its <u>bandwidth</u> as follows

$$\text{bandwidth }(f) = \{|i-j| : \{v_i,v_j\} \in E\} .$$

The bandwidth of G is defined to be

$$\text{Bandwidth }(G) = \min\{\text{bandwidth }(f)\text{: f is a layout of G}\}.$$

Let A be the vertex adjacency matrix of G. It is easily seen that Bandwidth (G) $\leq$ k iff there is a permutation matrix P such that all ones in P^TAP lie within the "band" composed of the diagonal, the first k super-diagonals, and the first k subdiagonals. Let Bandwidth be the following problem: Given an undirected graph G and a nonnegative integer k, decide whether Bandwidth (G) $\leq$ k. This problem has been proved NP-complete by Papadimitriou [16] (see Garey *et al*. [6] for related results).

<u>Theorem 4</u>: Problem 9 is NP-complete.

<u>Proof</u>: It is a rather trivial fact that Bandwidth transforms to Problem 9. Indeed, f is a layout of G = $\langle V,E \rangle$ with bandwidth (f) $\leq$ k iff f realizes the quasi-linearity with buffer size k + 1 of E. □

We also remark that it is a trivial consequence of a non-trivial algorithm of Garey *et al*. [6], that testing for the quasi-linearity with buffer size 3 (and finding a suitable sequence, if there is one) can be done in linear time. Indeed, if $|M| > 3$ for some $M \in \mathcal{M}$ then clearly $\mathcal{M}$ is not quasi-linear with buffer size 3. If $|M| \leq 3$ for all $M \in \mathcal{M}$ then we replace every 3-element subset $\{a,b,c\} \in \mathcal{M}$ by three subsets $\{a,b\}$, $\{b,c\}$, $\{a,c\}$, and we delete all subsets of cardinality less than 2. The resulting family E determines the set of edges of some graph G such that Bandwidth (G) $\leq$ 2 iff $\mathcal{M}$ is quasi-linear with buffer size 3. Now it is sufficient to apply the linear time algorithm of Garey *et al*. to test whether Bandwidth (G) $\leq$ 2.

Finally, we remark that most of the NP-completeness results described in this paper can be extended to cyclic (rather than linear) arrangements of the underlying set X, and even to more general structures considered in [12]. We leave it to the reader.

REFERENCES

[1] A. V. Aho, J. E. Hopcroft, and J. D. Ullman, Design and Analysis of Computer Algorithms, Addison-Wesley, Reading, MA, 1974.

[2] K. S. Booth, PQ-tree Algorithms, Ph.D. Dissertation, Dept. of Electrical Engineering and Computer Science, University of California, Berkeley, CA, 1975.

[3] K. S. Booth and G. S. Lueker, Testing for the consecutive ones property, interval graphs, and graph planarity using PQ-tree algorithms, J. Comput. System Sci., 13 (1976), pp. 335-379.

[4] J. Edmonds, Maximum matching and a polyhedron with 0,1 vertices, J. Res. NBS, 69B (1965), pp. 125-130.

[5] D. R. Fulkerson and O. A. Gross, Incidence matrices and interval graphs, Pacific J. Math., 15 (1965), pp. 835-855.

[6] M. R. Garey, R. L. Graham, D. S. Johnson and D. E. Knuth, Complexity results for bandwidth minimization, SIAM J. Appl. Math., 34 (1978), pp. 477-495.

[7] M. R. Garey, D. S. Johnson and L. J. Stockmeyer, Some simplified NP-complete graph problems, Theoret. Comput. Sci., 1 (1976), pp. 237-276.

[8] S. P. Ghosh, File organization: the consecutive retrieval property, Comm. ACM, 9 (1972), pp. 802-808.

[9] S. P. Ghosh, File organization: consecutive storage of relevant records on a drum-type storage, Information and Control, 25 (1974), pp. 145-165.

[10] L. T. Kou, Polynomial complete consecutive information retrieval problems, SIAM J. Comput., 6 (1977), pp. 67-75.

[11] E. L. Lawler, Combinational Optimization: Networks and Matroids, Holt, Rinehart and Winston, New York, NY, 1976.

[12] W. Lipski, Information storage and retrieval - mathematical foundations II (Combinatorial problems), Theoret. Comput. Sci., 3 (1976), pp. 183-211.

[13] W. Lipski, One more polynomial complete consecutive retrival problem, Information Processing Lett., 6 (1977), pp. 91-93.

[14] W. Lipski and W. Marek, File organization, an application of graph theory, Proc. 2nd Internat. Colloquium on Automata Languages and Programming, Saarbrucken, 1974, J. Loeckx, ed., Lecture Notes in Computer Science, 14, Springer-Verlay, Berlin, 1974, pp. 270-273.

[15] T. Nakano, A chracterization of intervals; the consecutive (one's or retrieval) property, Comment Math. Univ. Sancti Pauli, 22 (1973), pp. 49-53.

[16] C. H. Papadimitriou, The NP-completeness of the bandwidth minimization problem, Computing, 16 (1976), pp. 263-270.

[17] F. S. Roberts, Discrete Mathematical Models, with Applications to Social, Biological and Environmental Problems, Prentice-Hall, Englewood Cliffs, NJ, 1976.

[18] K. Tanaka, Y. Kambayashi, and S. Yajima, Organization of quasi-consecutive retrieval files, Proc. Languages and Automata Symp., July 21-23, 1977, Nagoya, Japan.

[19] A. C. Tucker, A structure theorem for the consecutive 1's property, J. Combinatorial Theory, 12B (1972), pp. 153-162.

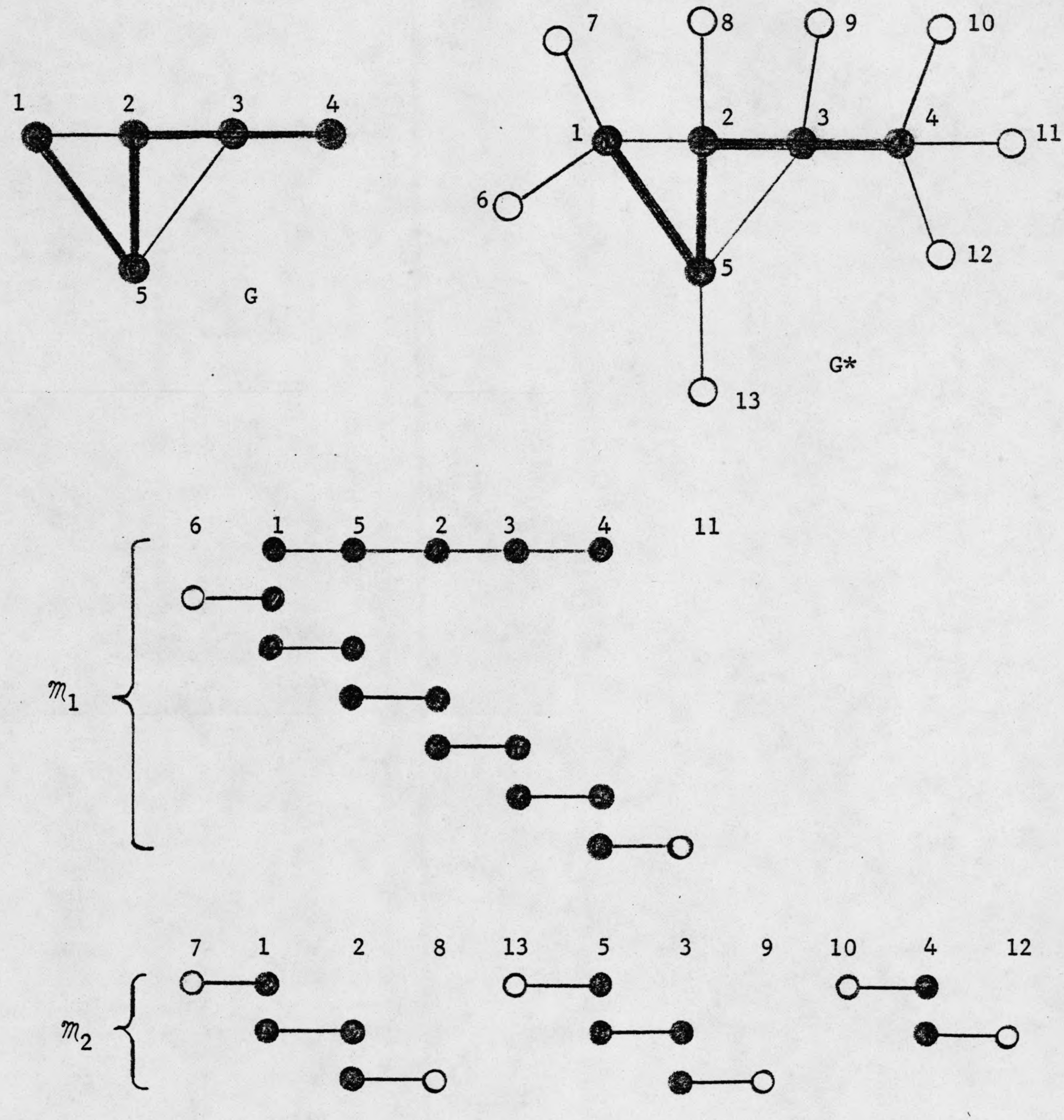

Fig. 1. A graph G with a Hamiltonian path, and the corresponding partition of $\mathcal{M}$ into linear families.

Matrices with the consecutive ones property have been introduced by Fulkerson and Gross [5] as a tool to investigate interval graphs. They received considerable attention when it turned out they could be applied to the design of efficient file organizations (see Ghosh [8], Lipski [12]). Still other applications, such as the so-called sequence dating in archeology, are discussed in the book of Roberts [17].

Let $\mathcal{M}$ be a family of subsets of a finite set X. $\mathcal{M}$ is said to be <u>linear</u> (see [12]) if there is a sequence

(1) $$x_1, x_2, \ldots, x_n$$

of elements of X such that

(i) every $x \in X$ occurs exactly once in the sequence, and

(ii) every $M \in \mathcal{M}$ appears as a <u>segment</u>, i.e. a set of $|M|$ consecutive terms of the sequence ($|M|$ denotes the cardinality of M).

It is obvious that a family $\mathcal{M} = \{M_1, \ldots, M_m\}$ of subsets of $X = \{x_1, \ldots, x_n\}$ is linear iff its <u>incidence matrix</u> $A = [a_{ij}]$, defined by

$$a_{ij} = \begin{cases} 1 & \text{if } x_i \in M_j \\ 0 & \text{if } x_i \notin M_j \end{cases}$$

has the consecutive ones property.

Several characterizations of linear families (or, equivalently, of matrices with the consecutive ones property) are known, see e.g. Tucker [19], Lipski [12], Nakano [15]. Efficient polynomial time algorithms to test for the linearity of a family of sets (and finding a suitable sequence (1), if there is one) have also been proposed, see