A Note on the Consecutive Ones Submatrix Problem

MohammadTaghi Hajiaghayi
mhajiaghayi@uwaterloo.ca
Yashar Ganjali
yganjali@uwaterloo.ca
Department of Computer Science
University of Waterloo *

Abstract

A binary matrix has the *Consecutive Ones Property (C1P) for columns* if there exists a permutation of its rows that leaves the 1's consecutive in every column. The problem of Consecutive Ones Property for a matrix is a special variant of Consecutive Ones Submatrix problem in which a positive integer K is given and we want to know if there exists a submatrix B of A consisting of K columns of A with C1P property. This paper presents an error in the proof of NP-completeness for this problem in the reference cited in text by Garey and Johnson [6].

keywords: Computational Complexity, Consecutive Ones Property, Consecutive Ones Submatrix.

1 Introduction

A binary matrix has the *Consecutive Ones Property (C1P) for columns* if there exists a permutation of its rows that places the 1's consecutive in every column. One can symmetrically define the equivalent property for rows. This problem of verifying this property has applications in computational biology [5], recognizing interval graphs [4] and file organization [7] in all of these the goal is to linearly arrange a set of objects with the constraint on the consecutiveness of objects in a class of given subsets.

The problem of Consecutive Ones Property for a matrix is introduced in [6] as a special variant of Consecutive Ones Submatrix problem (COS) in which for a matrix A and a positive integer K we want to know whether there exists a submatrix B consisting of K of the columns of A that satisfies C1P.

Fulkerson and Gross introduced C1P and gave the first polynomial time sequential algorithm for it [2]. The classic algorithm for solving C1P is a linear-time sequential algorithm of Booth and Lueker [3] based on PQ-trees which is quite complicated. Simpler algorithms were found by Meidanis et al., and Habib et al. [1, 9]. Nevertheless, the generalized COS problem is NP-complete. This NP-completeness result as well as the ones for many variants of this problem are used to show the NP-completeness of some other problems such as Interval Routing Skim Problems [8]. However, in the classic reference text of Garey and Johnson [6], the proof of NP-completeness for this problem is referenced to [10], which proves the NP-completeness of a

*The research was supported by the Natural Sciences and Engineering Research Council of Canada and Communication and Information Technology Ontario

slightly different problem rather than the original COS as defined above and in [6]. In these paper, we will present the error as well as another proof for this problem. We also consider the complexity of some special cases of this problem.

2 The error and the new proof

We consider the problem below presented in [6]:

Consecutive Ones Submatrix(COS)

INSTANCE: An $m \times n$ matrix A of 0's and 1's and a positive integer K.

QUESTION: Is there an $m \times K$ submatrix B of A such that has the "consecutive ones" property, i.e., the columns of B can be permuted such that in each row all the 1's occure consecutively?

In [6], this problem is referenced to [10] and has been stated that this problem has a transformation from the Hamiltonian Path problem. We discuss more about this below.

Theorem 4.24 from [10] says:

> **Theorem 4.24.** *Let M be an $m \times n$ matrix. Deciding if there exists any $m \times k$ submatrix having the consecutive ones property is NP-complete.*
>
> **Proof:** A simple reduction can be made from the Hamilton path problem. Given a graph $G = (V, E)$, let M be the incidence matrix. G has a Hamilton path if and only if M has an $n \times (n-1)$ submatrix with the consecutive ones property. The Hamilton path problem is known to be NP-complete. □

Below is Booth's note on the subject [11].

> The incidence matrix M for G is $n \times m$ (n is the number of vertices, m is the number of edges), where M has a row for each vertex and a column for each edge, with a 1 in the ij-th entry if and only if vertex i is incident with edge j. (Note that each column thus has exactly two ones). If there is an $n \times (n-1)$ submatrix M' with the consecutive ones property, then there is a permutation P such that PM' has consecutive ones. This permutation is the Hamilton path. To see this, notice that vertex i can only be in an edge with vertex $i-1$ or $i+1$ if the columns have consecutive ones, and since there are $n-1$ columns, it must be the case that vertex 1 connects to vertex 2, vertex 2 to vertex 3, etc. If this were not the case, the M' would correspond to the union of a set of disjoint paths in G, but any set of disjoint paths with more than one component has less than $n-1$ edges, which is a contradiction. So M' must be just a single path with all the vertices, which is the desired Hamilton path.

but the above theorem and the above proof are not the the proof of COS problem said in [6]. In fact, they are the proof of NP-completeness of the following variant of the problem:

Other variant of Consecutive Ones Sub-matrix(OCOS)

INSTANCE: An $m \times n$ matrix A of 0's and 1's and a positive integer K.

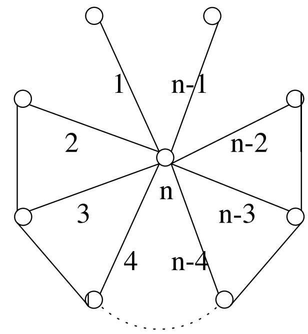

Figure 1: A counterexample

QUESTION: Is there an $m \times K$ submatrix B of A that has the "consecutive ones" property, i.e., the rows of B can be permuted such that in each column all the 1's occur consecutively?

Note that in the COS we permuted the columns after selecting them and we wanted to have the consecutive ones property for rows, but in OCOS we permute the rows after selecting the columns and we want to have consecutive ones for the columns.

These two problems are not equivalent if $k \neq n$ (if $k = n$, then one problem is the transpose of the other and the two are obviously symmetric). The cause of error is that in fact the original consecutive ones property was stated for columns (i.e., permuting the rows) and theorem 4.24 of [10] only talks about consecutive ones property (i.e., it does not say anything about the rows or the columns).

It is interesting that at the first glance, it seems that the above proof also can be stated for COS but it fact this is not true. For a counter-example, consider figure 1. For the graph in this figure, if we select $n-1$ columns corresponding to $n-1$ edges numbered $1, \ldots, n-1$ which form a Star(S_n), from the incidence matrix, the obtained matrix has the consecutive ones property for the columns but, the graph does not have the Hamiltonian Path.

Now we present a new NP-completeness proof for this problem.

Theorem 1 *The COS problem is NP-complete.*

Proof: Obviously this problem is in NP. We reduce the *Undirected Hamiltonian Path with degree at most 3(UHP3)* to this problem. UHP3 is formulated as given an undirected graph G with degree of each vertex less or equal to 3, decide whether G contains a Hamiltonian Path (see [6] for the proof of NP-completeness).

Let $G(V, E)$ be the input graph with n vertices and m edges. Consider the graph G' extended from G by adding one pendant edge incident to each vertex v of G (the new adjacent vertex corresponding to v_i is v_i'). The new graph $G'(V', E')$ has $n' = 2n$ vertices and $m' = m + n$ edges. Now consider the incidence matrix M for G' (note that the incidence matrix M for G' is $n' \times m'$ where M has a row for each vertex and a column for each edge, with a 1 in the $M[i, j]$ entry if and only if vertex i is incident to edge j). Note that each column thus has exactly two ones and each row has at most 4 ones.

First suppose that G has a Hamiltonian path $v_1, v_2, \ldots, v_n$, then the columns corresponding to edges (v_1, v_1'), (v_1, v_2), (v_2, v_2'), (v_2, v_3), (v_3, v_3'), $\cdots$, (v_n, v_n') (in this order) form an $n' \times (n'-1)$ submatrix with consecutive ones property for the rows.

Conversely, we show that G contains a Hamiltonian Path if M has an $n' \times (n'-1)$ submatrix M' with the consecutive ones property for the rows. First note that no selection of K' columns of M (in fact, K' selection of edges of G') can admit a cycle in G' since otherwise these K' columns have at least one row which cannot have the consecutive ones property. So this selection corresponds to a forest in G' and because it has $n'-1$ columns (edges) it must be a tree that we call T. Now, let v be a vertex in the original vertex set V. We show that it cannot have more than two neighbors in T which are also in the original vertex set V. let v_1, v_2 and v_3 be the the first three vertices (corresponding to edges which are appear in a consecutive order) that are adjacent to v. v_2 has one pendant edge, but it cannot appear on either of its sides, because on one side of v_2 is the edge (v, v_1) and on the other side is (v, v_3).

So each original vertex has at most two adjacent original vertices. Hence, this tree corresponds to a Hamiltonian Path in G and the proof is complete. □

Note that in the above construction each column has at most two ones and each row has at most four ones. This gives us the following corollary.

Corollary 2 *The COS problem is NP-complete even for matrices with at most two ones in each column and at most four ones in each row.*

The following is another result in this regard.

Theorem 3 *For matrices which have at most two ones in each column and at most two ones in each row, the COS problem can be solved in polynomial time.*

Proof: Consider the graph G constructed from the input matrix A as follows: we associate one vertex corresponding to each row and put an edge between any two vertices v and u if and only if the two rows corresponding to v and u both have a 1 in some same column. The degree of each vertex in G is at most 2. So, its components are all cycles or paths. Obviously, we cannot select all columns corresponding to vertices of a cycle. Let the number of cycles in G be c' and the number of columns of the input matrix be c. It is easy to observe that the COS problem has solution if and only if $K \leq c - c'$. We note that all the above operations can be done in polynomial time. □

But, the complexity of the case in which we have at most two ones in each column and at most three ones in each row, is open.

Note that the variant of the COS in which we ask if B has the *circular ones* property for the rows, i.e. the columns of B can be permuted in such a way that in each row either all the 1's or all the 0's occur consecutively, is also NP-complete (its proof is similar to proof of theorem 1 with $K = n'$ instead of $K = n' - 1$). This variant is also introduced in [6] and it again suffers from the same error as mentioned above for the COS.

3 Conclusion

In this paper, we introduced an error in the proof of intractability of an NP-complete problem mentioned in text by Garey and Johnson [6] and provided a new proof for this problem. We also considered the complexity of some special cases of this problem.

Acknowledgments. We thank Anna Lubiw and Bardia Sadri for their good hints.

References

[1] J. Meidanis, O. Porto, G. P. Telles,"On the consecutive ones property", *Discrete Applied Mathematics 88(1998)*, 325-354.

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

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

[4] J. S. Deogun, K. Gopalakrishnan, "Consecutive retrieval property-revisited", *Information processing Letters 69(1999)*, 15-20.

[5] F. Annexstein, R. Swaminathan, "On testing consecutive-ones property in parallel", *Discrete Applied Mathematics 88(1998)*, 7-28.

[6] M. R. Garey, D. S. Johnson, *Computers and Intractability, A guide to the theory of NP-completeness*, W.H. Freeman and Company, 1979.

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

[8] M. Flammini, G. Gambosi, S. Salomone, "Boolean Routing", *Lecture Notes in Computer Science 725(1993).*

[9] M. Habib, R. McConnell, C. Paul, L. Viennot, "Lex-BFS and partition refinement, with applications to transitive orientation, interval graph recognition and consecutive ones testing", *Theoretical Computer Science 234(2000)*, 59- 84.

[10] K. S. Booth, "PQ-tree Algorithms", *Doctoral Thesis, Dept. of Electrical Engineering and Computer Science, University of California, Berkeley, CA.*

[11] Private Communication with Professor K.S. Booth in UBC.