Discrete Applied Mathematics 8 (1984) 295–305
North-Holland

RECOGNIZING GRAPHS WITH FIXED INTERVAL NUMBER IS NP-COMPLETE

Douglas B. WEST
University of Illinois, Urbana, IL 61801, USA

David B. SHMOYS*
University of California, Berkeley, CA 94720, USA

Received 20 August 1982
Revised 1 June 1983

A *t-interval representation* of a graph expresses it as the intersection graph of a family of subsets of the real line. Each vertex is assigned a set consisting of at most t disjoint closed intervals, in such a way that vertices are adjacent if and only if some interval for one intersects some interval for the other. The interval number $i(G)$ of a graph G is the smallest number t such that G has a t-representation. We prove that, for any *fixed* value of t with $t \geq 2$, determining whether $i(G) \leq t$ is NP-complete.

1. Introduction

The notion of the interval number of a simple undirected graph (henceforth 'graph') generalizes the concept of interval graphs. A graph on n vertices is an *interval graph* if it is the intersection graph of a collection of finite intervals on the real line. In other words, each vertex is assigned a real interval such that two vertices of G are adjacent if and only if the corresponding intervals intersect.

To generalize this concept, we may assign more than one interval to each vertex. That is, given an integer t, we say that G has a *t-interval representation* (or simply a *t-representation*) if to each vertex of G we can assign at most t disjoint closed intervals such that two vertices v and w are adjacent in G if and only if some interval for v intersects some interval for w. Without loss of generality we assume that different intervals do not share endpoints. The interval number of G, denoted $i(G)$, is the smallest integer t such that G has a t-interval representation. Thus the interval graphs are the graphs G with $i(G) \leq 1$.

Several authors have remarked that interval graphs and interval numbers are useful in scheduling and allocation problems. Here we describe an application where 'multiple-interval' graphs may be more appropriate than interval graphs. Benzer [1] used interval graphs to discuss the linear arrangement of nucleotides in genes. Many genes are encoded in a single molecule of DNA and each gene contains information

*This work was supported by an NSF Fellowship, the Eugene C. Gee and Mona Fay Gee Scholarship, NSF Grant No. MCS78-20054 and DARPA Grant No. N00039-82-C-0235.

for the assembly of a specific protein. DNA is a linear molecule composed of a sequence of sub-units, each of which contains one of four possible nucleotides. The specific sequence of these nucleotides encodes the genetic information. For many years biologists thought that the nucleotides encoding a single gene appear in an unbroken sequence, covering a single 'interval' on the chain of nucleotides. Recently, Chambon [3] and others have shown that many genes are not represented as single unbroken sequences but rather as a collection of unbroken sequences on the DNA strand. This is analogous to a family of intervals on the real line. Thus, interval numbers may be a useful tool for studying the structure of genes; if an upper bound can be given on the number of intervals used for each gene, then finding interval numbers can be used to test gene compositions.

Bounds on the interval number in terms of other parameters of the graph appear in [5], [6], [7], [8], and [11]. Several of those papers pose the related problems of characterizing graphs with interval number t (even for $t=2$) and finding an efficient algorithm for computing the interval number. In this paper, we show that for any fixed value of t with $t \geq 2$, determining whether $i(G) \leq t$ is NP-complete. To do this, we reduce the problem of recognizing triangle-free 3-regular graphs having a Hamiltonian circuit to the problem of recognizing graphs having a 2-representation. Then we reduce the $(t-1)$-representation problem to the t-representation problem. (The t-representation problem is similar in some respects to problems proved NP-complete in [9] and [10]. However, there does not seem to be any immediate relationship among these problems.)

To put this result in perspective, we note that there exists a linear time algorithm for the case $t=1$, which is the problem of interval graph recognition [2]. Our result then implies that, for the class of recognition problems parametrized by a fixed value of the interval number, the boundary between P and NP-complete has been completely determined. This NP-completeness result also implies that a 'nice' characterization of t-interval graphs is unlikely to exist for any $t \geq 2$.

2. A preliminary reduction

We wish to reduce from the Hamiltonian circuit problem in 3-regular triangle-free graphs to the 2-representation problem. It is well known that testing for a Hamiltonian circuit in a 3-regular graph (also called a *cubic* graph) is NP-complete, as listed in [4]. To show that this remains true when triangles are forbidden from the graph, we provide a preliminary reduction.

Hamiltonian Circuit in Cubic Graph (3HC)

Instance: A graph $G(V,E)$ in which every vertex has three neighbors.

Question: Does G have a Hamiltonian circuit?

Hamiltonian Circuit in Triangle-free Cubic Graph (Δ3HC)

Instance: A graph $G=(V,E)$ in which every vertex has three neighbors but no three vertices form a cycle.

Question: Does G have a Hamiltonian circuit?

Lemma 1. *Δ3HC is NP-complete.*

Proof. We reduce 3HC to Δ3HC. Given an instance G of 3HC, we construct in polynomial time a triangle-free cubic graph G' such that G' has a Hamiltonian circuit if and only if G has a Hamiltonian circuit.

Begin by examining all the triples of vertices in the graph to determine which ones induce triangles. This can be done in $O(n^3)$-time. By looking at the remaining neighbors of the vertices in each triangle, we can determine whether or not the triangle shares an edge with another triangle. If so, we call such a triangle matched; otherwise it is 'isolated'. If G is 3-regular and $G \neq K_4$, then every triangle must be matched or isolated.

To form G', replace each isolated triangle by bisecting its edges with new vertices and adding a fourth new vertex joined to the new vertices on the three old edges. For each matched pair of triangles, replace one of the two triangles as above. These local replacements are shown in Fig. 1. No new triangles are created, so G' is triangle-free.

To show that G has a Hamiltonian circuit if and only if G' also does, note that any Hamiltonian circuit that enters one of the components shown in either graph must visit all vertices in that component before leaving the component, since it has at most three connections to the rest of the graph. The heavy lines in Fig. 1 indicate the only way (up to symmetry) that a Hamiltonian circuit can contain all of the vertices in one of these components. By making the indicated substitutions we can transform a Hamiltonian circuit in either graph into a Hamiltonian circuit in the other. $\square$

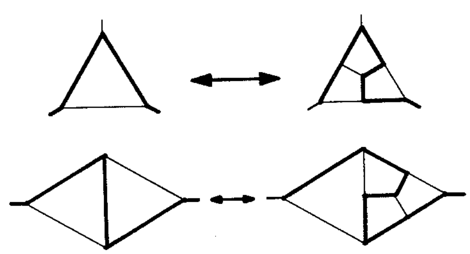

Fig. 1. Local replacement for triangle-free cubic graphs.

3. Useful facts about interval numbers

Complete bipartite graphs play a significant role in the subsequent reductions. We refer to the maximal independent sets of a complete bipartite graph as the *parts* of the graph. To facilitate discussion of interval representations, we let $f(v)$ denote the union of the intervals assigned to v. A t-representation of a graph G is *displayed* if each set $f(v)$ contains an open interval disjoint from all other sets $f(w)$. We say that a t-representation is of depth r if no point on the real line is contained in intervals for $r+1$ of the vertices.

Trotter and Harary [12] showed that $i(K_{m,n}) = t(m,n)$, where $t(m,n) = \lceil (mn+1)/(m+n) \rceil$. In fact, they showed that $K_{m,n}$ has a $t(m,n)$-representation that is displayed. To obtain an important fact about the representations of certain bipartite graphs, we give the argument for the lower bound. Since the depth is 2, when we read the representation from left to right we obtain at most one new edge at the left endpoint of each interval. No new edge arises if the left endpoint of this interval is contained in no other interval, such as with the leftmost interval in the representation. Since there are at most t intervals per vertex, this yields an upper bound on the number of edges that can be represented: $|E| \le (m+n)t-1$. Hence $i(K_{m,n}) \ge (mn+1)/(m+n)$. If $(mn+1)/(m+n)$ is an integer t, then we say that $K_{m,n}$ is *t-tight*. To represent all the edges in a t-tight $K_{m,n}$, the argument above implies that an intersection of intervals must occur at the left endpoint of *every* interval after the first. This means that their union is a single closed interval. We say that such a set of intervals 'appears contiguously'.

Lemma 2. *The graph $K_{t^2+t-1,t+1}$ is t-tight. If $K_{t^2+t-1,t+1}$ is an induced subgraph of a graph G, then in any t-representation of G the intervals for vertices of $K_{t^2+t-1,t+1}$ appear contiguously. Furthermore, if u and v are any specified vertices from opposite parts of $K_{m,n}$, then $K_{m,n}$ has a displayed $t(m,n)$-representation in which u and v are assigned the leftmost and rightmost intervals in the representation, respectively.*

Proof. Since $[(t+1)(t^2+t-1)+1]/(t^2+2t) = t$, $K_{t^2+t-1,t+1}$ is t-tight. In any t-representation of a graph G, the intervals corresponding to the vertices of an induced subgraph of G must give a t-representation of the subgraph, so the intervals for $K_{t^2+t-1,t+1}$ must appear contiguously.

For the second part of the lemma, take a displayed t-representation of $K_{m,n}$, as guaranteed by the Trotter–Harary construction. Note that, for either of the parts in $K_{m,n}$, the leftmost of the intervals assigned to vertices in that part can be extended leftward to become the leftmost interval in the representation. Therefore, by permuting vertex labels we can have the leftmost interval assigned to u. We can do the same to assign the rightmost interval to any vertex v belonging to the other part. □

In fact, the lemma remains true if we drop the requirement that u and v belong to different parts in the bipartite graph. However, this is harder to prove, and our reductions require only the version stated above. We will make critical use of the graphs $K_{t^2+t-1,t+1}$, such as $K_{5,3}$, to ensure that representations can be constructed only in certain ways.

4. The main reduction

Consider the following decision problem.

t-Interval Representation (tREP)
Instance: A graph $G=(V,E)$.
Question: Does G have a representation as the intersection graph of sets consisting of at most t intervals on the real line?

Theorem 1. 2REP *is NP-complete.*

Proof. We prove this by reducing Δ3HC to 2REP. Let G be a triangle-free cubic graph. We will construct a supergraph G' of G such that G' has a 2-representation if and only if G is Hamiltonian.

First we note that it is easy to construct a 2-representation of a cubic Hamiltonian graph G. The edges of G can be partitioned into a Hamiltonian cycle and a complete matching. Suppose the Hamiltonian cycle is $v_0, v_1, \ldots, v_n, v_0$. The Hamiltonian path from v_0 to v_n can be represented contiguously, using one interval per vertex, such that intervals for consecutive vertices in the path overlap. Complete the Hamiltonian cycle by using a second interval for v_0. Suppose that (v_0, v_j) is the edge incident to v_0 in the complete matching. Extend the second interval for v_0 to overlap a second interval for v_j. For any vertex whose matched edge has not been represented, only one interval has been assigned. Therefore, we represent these remaining edges by pairs of intervals that intersect each other. This representation, which we call the *H-representation* of G, is shown in Fig. 2. Note that a cubic graph has an H-representation (for some labeling of the vertices) if and only if it is Hamiltonian.

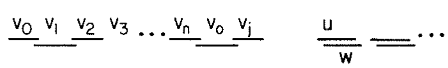

Fig. 2. The H-representation of a cubic Hamiltonian graph.

For an arbitrary cubic graph, many different 2-representations may be possible. We want to restrict the representation of G so that it must be an H-representation. The first step is to consider only triangle-free graphs, which restricts the depth to two. In addition, we add 'gadgets' to G so that any H-representation remains feasible, but other possible representations are eliminated. More precisely, if G has an H-representation, then the supergraph G' will have a 2-representation of G, and G' will have no 2-representation in which the intervals for G appear in any other way.

Roughly speaking, there are two steps to this restriction. First, we want to add gadgets to separate the intervals for vertices of G into two sets, each containing one of the intervals for each vertex. We call these the 'inside' and 'outside' intervals, for reasons that will soon be apparent. Second, we construct the gadgets so that the outside intervals can intersect at most one other interval for a vertex in G. This will force the matching of an H-representation to occur among the outside intervals and a Hamiltonian cycle to occur among the inside intervals. The reader may wish to refer to Fig. 3 as we describe the gadgets to be added.

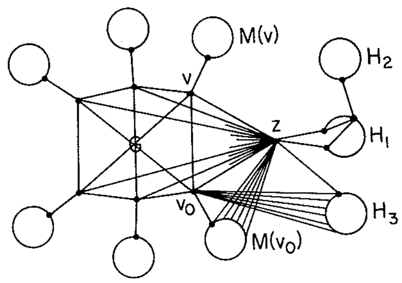

Fig. 3. The transformed graph G'.

The idea behind the gadgets added for the first step is straightforward. Add a vertex z joined to every vertex of G. This forces every vertex of G to have an interval intersecting an interval for z. These we call the 'inside' intervals. Actually, to enforce the condition that all 2-representations of G' contain an H-representation for G, we want all of these 'inside' intervals to lie within a single interval for z. So, we attach additional gadgets to z to insure that only one of the intervals for z contains any of the 'inside' intervals, and to control what happens at the endpoints of the intervals in $f(z)$. The details of this will be described more explicitly below.

To insure that each vertex of G has another interval outside $f(z)$, we need only add, for each vertex $v \in G$, some structure $M(v)$ containing a vertex adjacent to v but not to z. We choose a structure $M(v)$ that accomplishes the second step mentioned above. In particular, for each $v \in V(G)$ let $M(v)=K_{5,3}$, and join v to one vertex of $M(v)$. No other edges join $M(v)$ to any other vertices in G'. Note that $K_{5,3}$ is 2-tight, so if the G' being constructed has 2-representation, then the intervals of each $M(v)$ appear contiguously. The only other interval that can intersect these is the outside interval for v. If it lies wholly within $f(M(v))$, it cannot be used for anything else. So, using Lemma 2, we may assume that the outside interval for v overlaps $f(M(v))$ at one of its endpoints.

To complete the construction of G', we describe the gadgets mentioned earlier that are used to pin down the ways in which the intervals for z can appear. Add three more copies of $K_{5,3}$ to the graph, called H_1, H_2, and H_3. Join z to two vertices in the same part of H_1, and join one vertex from the other part of H_1 to a vertex in H_2. Also, place one edge from z to a vertex of H_3. We have described all of the vertices of G', and this completes the construction except for a few more edges needed to maintain feasibility of the H-representations, as will be seen shortly.

To understand the role of the H_i, consider the intervals representing the subgraph induced by z and the vertices in the H_i. A 2-representation of G' must contain a 2-representation of this induced subgraph. The H_i are 2-tight, so the intervals in their representations appear contiguously. Therefore, the edge between H_1 and H_2 can be represented only by an intersection between end-intervals in the representations of H_1 and H_2. Assume that the intervals for H_2 appear to the left of those for H_1.

Since z neighbors only two independent vertices in H_1, and since each vertex of H_1 has already been assigned two intervals, it follows that two intervals for z must be used to represent its edges to H_1. Since one endpoint of $f(H_1)$ is already covered by $f(H_2)$, one of the intervals in $f(z)$ must be swallowed in a displayed portion of its neighbor in H_1. Since z has other neighbors in G' which do not neighbor anything in H_2, the other interval for z must overlap the other end of $f(H_1)$ and be displayed. By Lemma 2, this can be arranged, since the neighbors of z in H_1 are in the opposite part from the vertex with a neighbor in H_2. Finally, H_3 is also 2-tight, and its edge to z can only be represented by having the right end of the displayed interval for z overlap the left endpoint of $f(H_3)$.

Consider the graph G' constructed from a Hamiltonian graph G. The reader should refer to Fig. 4, where the 2-representation of G' obtained from the H-representation of G is illustrated. Now, let us check the feasibility of the H-representation shown, given these gadgets. The configuration of intervals representing z and the H_i can be inserted to enclose the Hamiltonian cycle part of the H-representation – almost. We want the second interval for v_0 to intersect the second interval for its matched neighbor v_j, but the latter is the outside interval for v_j. We add edges from v_0 to each vertex in H_3, so that this interval for v_0 can (and must) pass outside $f(z)$ and past $f(H_3)$ to intersect the outside interval for v_j. Now, $M(v_0)$ can only be added to the H-representation by attaching its (contiguous) representation to the interval for v_0 at the other end of the cycle, which is within $f(z)$. To enable this, we add edges from z to each vertex of $M(v_0)$. Finally, the other $M(v)$'s can be added to the representation before and after the matched pairs of intervals.

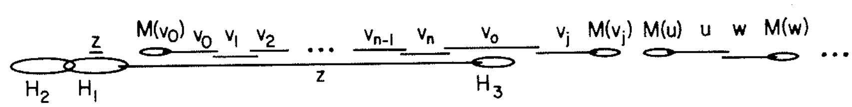

Fig. 4. The 2-representation of G'.

Summarizing the representation, from left to right, we have: (1) $f(H_2)$, (2) $f(H_1)$ containing an interval for z, (3) the displayed interval for z containing $f(M(v_0))$ and the Hamiltonian circuit from v_0 to v_0, and (4) $f(H_3)$. The second interval for v_0 extends throughout $f(H_3)$ and intersects the outside interval for its non-circuit neighbor. Finally, the remaining $M(v_i), v_i, v_j, M(v_j)$ pairs complete the representation.

We have described all of the vertices and edges added to G to obtain G', and in the process we have shown that if G is Hamiltonian, then it has an H-representation and G' has a corresponding 2-representation. Note that we have distinguished a vertex v_0 by the edges added in the previous paragraph. Since a Hamiltonian circuit can 'start' at any point, it does not matter which vertex we call v_0. If G has n vertices, then it has $3n/2$ edges (being 3-regular), and G' has $9n+25$ vertices and $37n/2+65$ edges, so the transformation can be performed in polynomial time.

To complete the proof, we need only show that if G' has a 2-representation, then it contains an H-representation of G, and thus G has a Hamiltonian circuit. In motivating the construction we have already shown the required structure of the 2-representation for the subgraph induced on z and all the H_i, as well as the structure for the subgraph induced on just v and $M(v)$ for all vertices v. Since z does not neighbor the vertices of $M(v)$, and since $v \neq v_0$ does not neighbor any vertices in H_i, the representations of each of the subgraphs on v and $M(v)$ must lie outside $f(z)$, for $v \neq v_0$. Since each vertex of G neighbors z, each $f(v)$ must have an interval inside the displayed portion of $f(z)$. For $v \neq v_0$, we have forced an outside interval and an inside interval, but for v_0 we still have a 'free' interval.

Now, how can the edges of the original graph G appear in the representation? First, we show that for any $v \in V(G)$, the interval for v that intersects $f(M(v))$ cannot be used to represent more than one of the edges of G. First consider $v \neq v_0$. No interval for v can entirely contain the outside interval for any $u \notin \{v, v_0\}$, because this would produce an intersection between $f(v)$ and $f(M(u))$. Therefore, any intersection between $f(u)$ and the outside interval for $f(v)$ must use the 'free' endpoint of that interval. Since G has no triangles, it follows that only one of the edges incident to v in G can be represented using its outside interval, unless that interval entirely contains an interval for v_0. Suppose the outside interval for v entirely contains an interval for v_0. Then the interval for v_0 that intersects $f(M(v_0))$ must intersect intervals for the two other neighbors of v_0 in G. Showing this is impossible will complete all parts of the claim. Arguing as before, this interval for v_0 must entirely contain the inside interval for some vertex $w \in V(G)$. But then the outside interval for w must intersect outside intervals for two vertices of G other than v_0, which is impossible.

Now, consider the interval for $v \neq v_0$ inside $f(z)$. We claim it can intersect at most two other intervals for vertices of G. Suppose not. Since G has no triangles, one of these three intervals must be interely contained in $f(v)$; suppose it belongs to $f(w)$. Then the other interval for w must intersect $f(M(w))$, and it must be used to represent two edges of G. By the fact proved in the preceding paragraph, this is impossible.

We conclude that, for every $v \neq v_0$, the inside interval intersects intervals for two neighbors in G, and the outside interval intersects an interval for one neighbor in G. This means that the inside intervals group into paths. The only place such a path can end is at an interval for v_0, and there are only two intervals for v_0, so all the inside intervals must belong to a single path with intervals for v_0 at the ends. The intersection graph for these intervals is a Hamiltonian circuit in the original graph G. At this point, having assumed a 2-representation for G', we have forced a Hamiltonian circuit in G. This completes the proof. □

4. Further NP-completeness results and additional problems

The result of Theorem 1 holds equally well for $t>2$; 2-representations merely provide a convenient place to get started.

Theorem 2. *For any integer $t \geq 2$, tREP is NP-complete.*

Proof. Given Theorem 1, we need only reduce $(t-1)$-REP to tREP. Let G be any graph. We will construct a supergraph G' of G such that G has a $(t-1)$-representation if and only if G' has a t-representation. For each $v \in G$, create three copies of $K_{t^2+t-1,t+1}$. Call them $H_i(v)$, $i=1,2,3$. Join v to one vertex of $H_2(v)$. Add two more edges, one joining $H_2(v)$ to each of $H_1(v)$ and $H_3(v)$, using vertices from the two different parts of H_2 (see Fig. 5). If G has n vertices and m edges, then G' has $n(1+3t^2+6t)$ vertices and $m+3nt^2(t+2)$ edges, so this is a polynomial transformation.

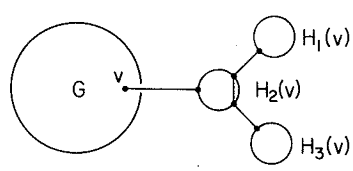

Fig. 5. The transformed graph G'.

Consider how a t-representation of G' must be arranged. All the H_i are t-tight. Hence in any t-representation of G' the $H_i(v)$ appear contiguously for a fixed vertex v, using t intervals per vertex. The edges between the $H_i(v)$ can only be realized by overlapping the intervals at the ends of the $f(H_i(v))$. By Lemma 2, this can be done. Having done this, the endpoints of $f(H_2(v))$ are covered, and the edge from v to $H_2(v)$ must be represented by inserting an interval for v into the displayed portion of its neighbor in $H_2(v)$. This interval for v cannot intersect any other intervals, because v and its neighbor in $H_2(v)$ have no mutual neighbors.

Creating this part of the representation for each vertex $v \in G$ represents all of the new edges but none of the old edges and uses one interval for each $v \in G$. None of these intervals can intersect any other intervals in the representation. Therefore, G' has a t-representation if and only if the subgraph G has a $(t-1)$-representation. □

Another related result follows easily at this point. Recall that the depth of a representation is the maximum number of intervals in it that cover a single point on the line, and the depth-r interval number $i_r(G)$ is the smallest t such that the graph has a depth-r t-representation.

Depth-r t-Representation (t_rREP)
Instance: A graph $G=(V,E)$.
Question: Is $i_r(G) \leq t$?

Theorem 3. *For every $t \geq 2$ and $r \geq 3$, t_r-REP is NP-complete.*

Proof. Note that the 2-representation of G' constructed in the proof of Theorem 1 has depth 3, so for any $r \geq 3$ it is true that G has a Hamiltonian circuit if and only if G' has a 2-representation of depth (at most) r. Furthermore, note that in the proof of Theorem 2 the additional gadgets have t-representations of depth 2, so if the representation of G or G' is of depth at least 2, the representation of the other has the same depth. The theorem follows immediately from these two facts. □

Several interesting special cases remain. What happens for depth 2? In particular, any t-representation of a triangle-free graph must have depth 2. This leads us to ask whether there is an efficient algorithm to determine the interval number of a triangle-free graph. To reach NP-completeness we need not have anything larger than triangles, because the graph generated in the main transformation has no K_4.

Similarly, Scheinerman and West [11] proved that the interval number of a planar graph never exceeds 3. Given a planar representation of a graph, their proof gives an efficient algorithm to find a 3-representation, but it does not determine whether the interval number is 3 or smaller. Hence we ask, is there an efficient algorithm to determine the interval number of a planar graph?

Acknowledgement

We wish to thank our good friend Pattie Suyemoto for suggesting this collaboration.

References

[1] S. Benzer, On the topology of the genetic fine structure, Proc. Nat. Acad. Sci. 45 (1959) 1607–1620.
[2] K.S. Booth and G.S. Leuker, Testing for the consecutive ones property, interval graphs, and graph planarity testing using PQ-tree algorithms, J. Comput. System. Sci. 13 (1976) 335–379.
[3] P. Chambon, Split genes, Scientific American 244 (May 1981) 60–71.
[4] M.R. Garey and D.S. Johnson, Computers and Intractibility, A Guide to the Theory of NP-Completeness (Freeman, San Francisco, 1979).
[5] J.R. Griggs, Extremal values of the interval number of a graph, II, Discrete Math. 28 (1979) 37–47.
[6] J.R. Griggs and D.B. West, Extremal values of the interval number of a graph, SIAM J. Algebraic Discrete Methods 1 (1980) 1–7.
[7] L. Hopkins and W.T. Trotter Jr., A bound on the interval number of a complete mutipartite graph, in: G. Chartrand, Y. Alavi, D.L. Goldsmith, L. Lesniak-Foster and D.R. Lick, eds., The Theory and Applications of Graphs (Wiley, New York, 1981) 391–407.
[8] L. Hopkins, W.T. Trotter Jr. and D.B. West, The interval number of the complete multipartite graph, Discrete Appl. Math. 8 (1984) 163–187.
[9] L.T. Kou, Polynomial complete consecutive information retrieval problems, SIAM J. Comput. 6 (1977) 67–75.

[10] L.T. Kou, L.J. Stockmeyer and C.K. Wong, Covering edges by cliques with regard to keyword conflicts and intersection graphs, Comm. ACM 21 (1978) 135–138.

[11] E.R. Scheinerman and D.B. West, The interval number of a planar graph: three intervals suffice, J. Combin. Theory (B), to appear.

[12] W.T. Trotter Jr. and F. Harary, On double and multiple interval graphs, J. Graph Theory 3 (1979) 205–211.