MULTITRACK INTERVAL GRAPHS

András Gyárfás[1]

Computer and Automation Institute, Hungarian Academy of Sciences
Department of Mathematics, University of Memphis

Douglas West[2]

Department of Mathematics, University of Illinois

Abstract. A *d-track interval* is a union of d intervals, one each from d parallel lines. The intersection graphs of d-track intervals are the unions of d interval graphs. The *multitrack interval number* or simply *track number* of a graph G is the minimum number of interval graphs whose union is G. The track number for $K_{m,n}$ is determined by proving that the arboricity of $K_{m,n}$ equals its "caterpillar arboricity". Recognition of graphs with track number 2 is shown to be NP-complete.

1. Introduction

Combinatorial properties of interval systems were investigated as early as the 1930's by Tibor Gallai. His two beautiful unpublished remarks are now phrased as saying that interval graphs and their complements are perfect graphs. In 1968, Gallai suggested considering more general set systems consisting of unions of d intervals, one each from d parallel lines. Such set systems have been called *separated d-intervals*, since the d parallel lines can be viewed as d disjoint host intervals on a single line. Here we propose the term *d-track intervals*, or *multitrack intervals* if d is not specified. This distinguishes between these set systems and collections of *d-intervals* or *multi-intervals*, which are arbitrary unions of d intervals on a single line. A collection of d-track intervals is a special case of a collection of d-intervals.

Combinatorial properties of d-track intervals and d-intervals were studied in [6]. There the transversal number τ is bounded by a function of the matching number ν for families of d-intervals. The best bound is established for small values of d and ν. It seems very difficult to find good bounds on τ in terms of ν. Recent

1 Supported by OTKA grant 7309.
2 Supported by NSA/MSP Grant MDA904-93-H-3040.

developments on this question include a promising approach by G. Tardos [14], who applied algebraic topology to prove that $\tau \leq 2\nu$ for every family of 2-track intervals. Tardos thinks there is no elementary proof of this for $\nu \geq 2$ (for $\nu = 1$ a simple proof works [6], even for more general set systems).

The intersection graph of a family of d-intervals is a *d-interval graph.* In addition to [6], the study of d-interval graphs was begun also by Trotter and Harary [16] and by Griggs and West [5]. The smallest d for which G is a d-interval graph is the *interval number* $i(G)$. In this language, the result of [6] states that the clique cover number is bounded by a function of the independence number for graphs of fixed interval number.

In [9], Kumar and Deo proposed the study of intersection graphs of d-track intervals. We call the intersection graph of a family of d-track intervals a *d-track interval graph* or simply a *d-track graph.* The interval graphs are precisely the 1-track graphs and also the 1-interval graphs. The result of Tardos cited above says that the clique-cover number is at most twice the independence number for every 2-track graph. As observed in [9], 2-track graphs may be used to model similarity in DNA sequences [7].

The *multitrack interval number* or *track number* of G (written $t(G)$) is the smallest d such that G is a d-track graph. As observed in [9], $t(G)$ is also the smallest d such that G can be expressed as the union of d interval graphs. Kumar and Deo [9] called $t(G)$ the *dimensionality* of G; we feel that "track number" better suggests the meaning of the parameter and its potential applications and avoids potential confusion with many other notions of dimension of graphs in the literature ([2] discusses dimension parameters involving intersection graphs).

We clarify some remarks in [9] about related problems. The *boxicity* of a graph G is the minimum number of interval graphs whose intersection is G; since the complement of an interval graph need not be an interval graph, boxicity of G does not generally equal $t(\bar{G})$. Roberts [10] proved that the maximum boxicity of an n-vertex graph is $\lfloor n/2 \rfloor$. Trotter [15] characterized the graphs with boxicity $\lfloor n/2 \rfloor$. The observation in [9] that the boxicity of every graph is at least its interval number is false; every complete bipartite graph has boxicity at most 2, while complete bipartite graphs have arbitrarily large interval number.

For the interval number, Griggs [4] proved $i(G) \leq \lceil (n+1)/4 \rceil$ for n-vertex graphs, and Andreae [1] proved Trotter's conjecture that when n is a multiple of 4 the only graph achieving this is $K_{n/2,n/2}$. The upper bound of $i(G) \leq \lceil (\Delta+1)/2 \rceil$ for graphs with maximum degree Δ, achieved by triangle-free regular graphs, was proved in [5], with a later compact presentation in [17]. The maximum of $i(G)$ for graphs with e edges is unknown; [13] presents an upper bound of $1 + \lceil \sqrt{e/2} \rceil$, but the conjectured optimum for $e \geq 4$ is $\lceil (1+\sqrt{e})/2 \rceil$, achieved by $K_{\sqrt{e},\sqrt{e}}$.

Since every d-track representation is a d-interval representation, $t(G) \geq i(G)$ for every graph G. We expect most extremal results on interval number (including those listed above) to hold also for track number. The interval number of a tree is at most 2 [16], and [9] observes that this bound holds also for track number. The proof that $i(G) \leq 3$ for planar graphs [11] shows also that $i(G) \leq 2$ for outerplanar graphs; do these bounds also hold for track number? We also ask whether there exist graphs with track number d that are not the union of d pairwise edge-disjoint interval graphs.

For complete bipartite graphs, Trotter and Harary [16] proved that $i(K_{m,n}) = \lceil (mn+1)/(m+n) \rceil$. Kumar and Deo [9] observed the analogous lower bound $t(K_{m,n}) \geq \lceil mn/(m+n-1) \rceil$ and proved that equality holds when $m = n$ or when $m > 3n$. In this paper, we prove that equality holds in this formula for all m, n. This result has independent interest in the context of arboricity. The arboricity of a graph is the minimum number of forests needed to cover its edges. It is well known that the arboricity of $K_{m,n}$ is $\lceil mn/(m+n-1) \rceil$, and our result says that this can be achieved using only forests of caterpillars, which are the trees that are interval graphs (a *caterpillar* is a tree in which all edges are incident to a single path called the *spine* of the caterpillar).

Concerning recognition of graphs with fixed track number, Kumar and Deo wrote "it is likely that the recognition problem for this class of graphs can be solved in polynomial time." We prove that this is not true; recognizing graphs with track number 2 is NP-complete. Our proof is a modification of the proof of Shmoys and West [12] for the NP-completeness of recognizing $i(G) = 2$. As in [12], this reduction generates graphs that have triangles but no larger cliques. This suggests several open questions for refining the boundary between P and NP (posed for interval number in [12]). Does recognition of graphs with track number or interval number 2 remain NP-hard when the input is restricted to triangle-free graphs? Does either remain NP-hard when restricted to planar graphs? We note also that recognition of graphs with boxicity 2 is NP-complete [8].

2. Complete Bipartite Graphs

The lower bound $t(K_{n,m}) \geq \lceil mn/(m+n-1) \rceil$, proved in [9], follows from the remark that every triangle-free interval graph is a forest. It is well known (see [16], for example) that a forest is an interval graph if and only if its components are caterpillars. Hence equality in the formula follows from partitioning the edges of $K_{m,n}$ into $\lceil mn/(m+n-1) \rceil$ such forests.

Theorem 1. *$K_{n,m}$ can be decomposed into $p(m,n) = \lceil \frac{mn}{m+n-1} \rceil$ pairwise edge-disjoint forests of caterpillars.*

Proof. We may assume that $m \geq n$. Observe that if n is odd, then $p(n+1,n) = \lceil \frac{(n+1)n}{2n} \rceil = \frac{n+1}{2} = p(n,n)$. Similarly, if n is even, then $p(n+2,n) = \lceil \frac{(n+2)n}{2n+1} \rceil = n/2+1 = p(n,n)$. If H is an induced subgraph of G, then a decomposition of H into d caterpillar-forests can be obtained from such a decomposition of G by deleting the deleted vertices. It thus suffices to prove the claim when $m \geq 2\lceil (n+1)/2 \rceil > n$.

Observe that $mn = (m + n - 1)p - q$, where $0 \le q < m + n - 1$. We can write this also as $m(n - p) = (n - 1)p - q$. Since $m > n$, we have $p > n - p$ (i.e., $p > n/2$).

Let the partite sets be $A = \{a_1, \dots, a_n\}$ and $B = \mathbf{Z}_m$. We define p paths of length $2n - 1$, each starting and ending in A and visiting $n - 1$ elements of B. In each path, the vertices of A appear in the order $a_1, \dots, a_n$. In the ith path, the vertices of B appear in the order $x_i, x_i+1, \dots, x_i+(n-1)$. We will choose the values $\{x_i\}$ so that the paths are edge-disjoint and so that each element of B appears on at least $n - p$ paths. We prove first that this suffices.

Each vertex of B on a path has two neighbors in the path; because the paths are edge-disjoint, these neighbors are distinct. In addition to these, each $x \in B$ appearing in s paths needs $n-2s$ additional neighbors. Having appeared in s paths, x can be added to $p - s$ more interval subgraphs. Since $p - s \ge n - 2s$ if and only if $s \ge n - p$, it suffices to gain one additional neighbor from each additional track. Each vertex of A has a "displayed" interval in each track, and hence the desired neighbors may be obtained arbitrarily by adding x as a leaf in the caterpillar forests in which it does not yet appear.

We will choose the p values $\{x_i\}$ so that $x_{i+1} - x_i$ equals $\lfloor m/p \rfloor$ or $\lceil m/p \rceil$ for all p. The resulting paths are edge-disjoint if and only if $m/p \ge 2$. For fixed n, the smallest m we consider is $2\lceil (n+1)/2 \rceil$. We have already computed that $p = m/2$ when $m = 2\lceil (n+1)/2 \rceil$. We also must deal with n odd and $m = n+2$ by augmenting m; in this case $p(n+3, n) = (n+3)/2 = p(n+2, n)$. An additional check shows that $p(m, n) < m/2$ when $m = n+4$ and n is odd, and when $m = n+3$ and n is even. For larger m, we compute $\frac{(m+1)n}{m+n} < \frac{mn}{m+n-1}(\frac{m+1}{m}) = \frac{mn}{m+n-1} + \frac{n}{m+n-1} < \frac{mn}{m+n-1} + \frac{1}{2}$. Hence $p(m + 2, n) \le p(m, n) + 1$, and we may assume that $m/p \ge 2$ for all cases that need to be considered.

It remains to determine how to arrange the $\lfloor m/p \rfloor$'s and $\lceil m/p \rceil$'s (differences between consecutive x_i's) so that each element of B appears in at least $n-p$ paths. We do this by ensuring that no $n - p$ consecutive differences sum to more than $n - 1$. This ensures that when an element is just about to appear for the first time in a path, it appears in the next $n - p$ paths. We have p of these "windows" of $n - p$ consecutive differences. Since each unit contributes to $n - p$ windows, the average window contains $m(n - p)/p \le n - 1$ units, which is the desired bound. Hence it suffices to show that we can place m units into a cyclic arrangement of p buckets so that the populations of distinct buckets differ by at most 1 and the total populations of distinct windows of $n - p$ consecutive buckets differ by at most 1.

We prove that this holds for arbitrary positive integers m, p, s, where s is the number of buckets in each window and $s < p$. Index the buckets modulo p (starting with 1). If s and p are relatively prime, place the m units in the buckets $\{is: 1 \le i \le m\}$. Each bucket has $\lfloor m/p \rfloor$ or $\lceil m/p \rceil$ units, since no bucket repeats until all have been visited. The first unit augments the first s windows, and each successive unit augments the next s of the less populous windows. After the moment when all windows reach the same population, the next window augmented is the first. Explicitly, writing ms as $ap + b$ with $0 \le b < p$, the population of the first b

windows is $\lceil m/p \rceil$, and the population of the others is $\lfloor m/p \rfloor$. When s and p have a common divisor, the same technique is applied in each congruence class of buckets modulo that divisor successively, filling each class with one unit per bucket before continuing to the next. □

3. Complexity

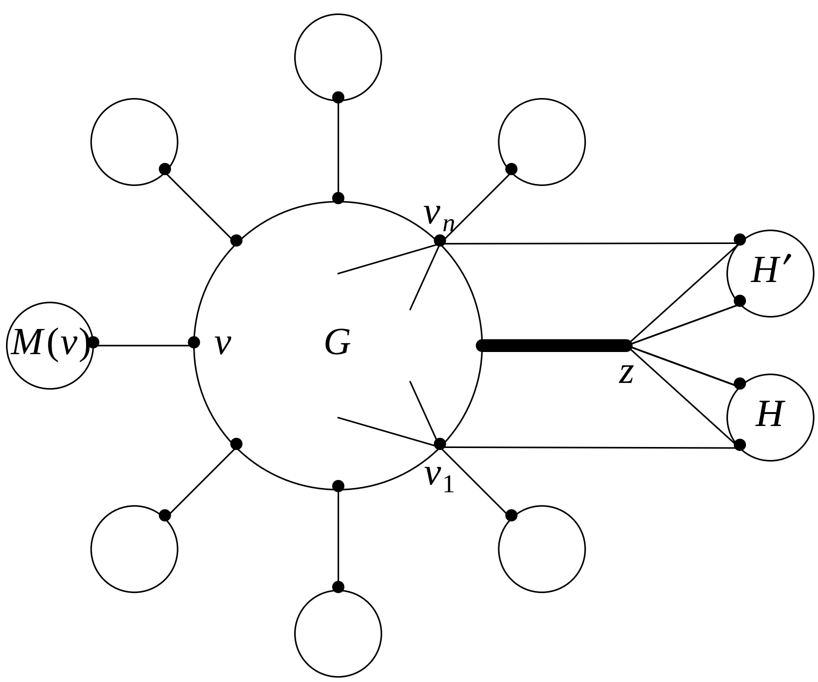
In this section we prove that recognition of graphs with track number 2 is NP-complete. It is easy to verify a d-track representation, so the problem is in NP. For NP-hardness, the problem we transform is that of checking for a Hamiltonian cycle through a specified edge of a triangle-free 3-regular graph. The NP-completeness of the Hamiltonian cycle problem for 3-regular graphs appears in [3]. The reduction to forbidding triangles appears in [12], though this is likely not its first appearance. Specifying an edge of the desired cycle does not make the problem easier, because one could run $|E(G)|$ instances of that problem to determine whether G has a Hamiltonian cycle. Finally, asking for a Hamiltonian cycle through the edge $v_1 v_n$ in a (3-regular triangle-free) graph is equivalent to deleting the edge and asking for a Hamiltonian v_1, v_n-path. We will show that this can be tested by testing whether an appropriate supergraph has a 2-track representation.

If $mn/(m+n-1)$ is an integer d, then expressing $K_{m,n}$ as a d-track interval graph requires decomposing $K_{m,n}$ into d caterpillars. In particular, the interval representation of the interval graph in each track occupies a contiguous portion of the line, with no gaps. Furthermore, symmetry allows us to name any desired pair of distinct vertices as the vertices whose intervals are leftmost and rightmost in the representation in the first track.

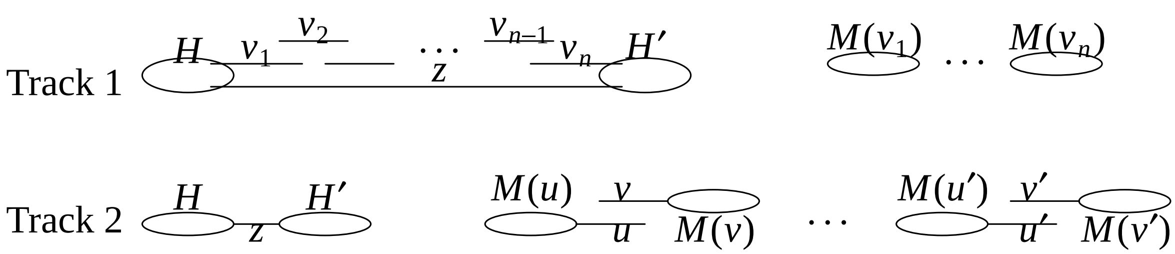
Theorem 2. *Recognizing 2-track interval graphs is NP-complete*

Proof. Suppose G is the graph obtained by deleting the edge $v_1 v_n$ from a triangle-free 3-regular graph. If G has a Hamiltonian v_1, v_n-path P, then deleting it leaves a 1-factor. We obtain a 2-track representation of G by representing P in one track and the 1-factor $G - E(P)$ in the other track. We construct a supergraph G' of G such that G' has a 2-track representation if and only if G has a Hamiltonian v_1, v_n-path. We will construct G' by adding gadgets so that in every 2-track representation of G', the induced subgraph G must be represented as described above.

For each $v \in V(G)$, let $M(v)$ be a copy of $K_{4,3}$; these are vertex disjoint. Make v adjacent to one vertex of $M(v)$. Let H, H' be two additional copies of $K_{4,3}$. Add one vertex z adjacent to $V(G)$ and to two vertices of degree 3 in each of H, H'. Finally, make v_1 adjacent to one neighbor of z in H, and make v_n adjacent to one neighbor of z in H'. This completes the construction of G' from G, illustrated below. Observe that $G' - z$ is triangle-free. The number of vertices and edges we have ended is linear in the number of vertices of G, so this is a polynomial transformation.



If G has a Hamiltonian v_1, v_n-path, then we can use it to form a 2-track representation of G' as illustrated below. In particular, we may partition H and H' into two spanning paths that together have all the 3-valent vertices as endpoints, and then z can be given the desired neighbors using these paths as the caterpillars for H and H'.



Conversely, suppose that G' has a 2-track representation f. The 2-track representation of each induced subgraph isomorphic to $K_{4,3}$ expresses it as an edge-disjoint union of two spanning caterpillars. The representation for each caterpillar occupies a contiguous portion of its track, as indicated in the illustrations above by small ellipses.

We let $f(v)$ denote the union of the intervals for v in the two tracks, and $f(S) = \cup_{v \in S} f(v)$ for $S \subseteq V(G)$. When an interval for z overlaps an interval for one of its neighbors x in H, it cannot intersect any other interval for H in that track, because z is not adjacent to any neighbor of x in H, and the caterpillar occupies a contiguous portion of the track. The same applies also for H', and z has four neighbors in $H \cup H'$. Hence in each track the interval for z contains the interval between $f(H)$ and $f(H')$.

Each $v \in V(G)$ is adjacent in G' to z and to a non-neighbor of z (in $M(v)$). Furthermore, v has non-neighbors in H and H'. Hence the interval for v that intersects $f(z)$ is contained in $f(z)$. Call this the "inside interval" for v, and call the interval for v that intersects an interval for a non-neighbor of z the "outside interval" for v.

Since $f(M(v))$ occupies one interval in each track, the outside interval for v generates no additional edges of G' unless its neighbor in $M(v)$ appears at an end of this interval for $M(v)$. Since $M(v)$ has no other neighbors, we may assume that each outside interval appears in this way. Now each outside interval intersects at most one other outside interval and no inside intervals. Since each vertex of G has three neighbors in $G \cup (H \cup H')$, each inside interval must intersect intervals assigned to two vertices in $G \cup (H \cup H')$. Since $G' - z$ is triangle-free, this implies that no inside interval is contained in other inside interval. Hence each inside interval intersects exactly two others, except that for its second "inside neighbor" the inside interval for v_1 intersects $f(H)$, and the inside interval for v_n intersects $f(H')$. Hence each outside interval intersects exactly one other outside interval.

At this point, we have proved that every 2-track representation for G' has the properties illustrated above. The outside intervals form a 1-factor of G. The inside intervals produce the remaining edges of G by forming a Hamiltonian path from v_1 to v_n. Hence we can test whether G has a Hamiltonian v_1, v_n-path by forming G' and testing whether it has a 2-track interval representation. □

We remark that [12] also reduced recognition of d-interval graphs to recognition of $d+1$-interval graphs, thereby proving that recognition of d-interval graphs is NP-complete for each fixed d at least 2. The corresponding reduction for d-track graphs does not seem as simple. We have not proved that recognition of d-track graphs is NP-hard for fixed d greater than 2, though we expect this to be true.

References

[1] T. Andreae, On an extremal problem concerning interval number of a graph, *Discrete Appl. Math.* 14(1986), 1–9.

[2] M. Cozzens and F.S. Roberts, On dimensional properties of graphs, *Graphs and Combinatorics* 5(1989), 29–46.

[3] M.R. Garey, D.S. Johnson, and R.E. Tarjan, The planar Hamiltonian circuit problem is NP-complete, *SIAM J. Comput.* 5(1976), 704-714.

[4] J.R. Griggs, Extremal values of the interval number of a graph II, *Discrete Math.* 28(1979), 37–47.

[5] J.R. Griggs and D.B. West, Extremal values of the interval number of a graph, *SIAM J. Algeb. Disc. Meth.* 1(1980), 1–7.

[6] A. Gyárfás and J. Lehel, A Helly type problem in trees, in *Combinatorial Theory and its Applications*, *Coll. Math. Soc. J'anos Bolyai* 4(1969), 571–584

[7] D. Joseph, J. Meidanis, and P. Tiwari, Determining DNA sequence similarity using maximum independent set algorithms for interval graphs, *Proc. SWAT* (1991), 326–337.

[8] J. Kratochvil, A special planar satisfiability problem and a consequence of its NP-completeness, preprint 1992.

[9] N. Kumar and N. Deo, Multidimensional interval graphs, *Congr. Numer.* 102 (1994), 45–56.

[10] F.S. Roberts, On the boxicity and cubicity of a graph, in *Recent Progress in Combinatorics* (W.T. Tutte, ed.) (Academic Press, 1969), 301–310.

[11] E.R. Scheinerman and D.B. West, The interval number of a planar graph—three intervals suffice, *J. Comb. Theory (B)* 35(1983), 224–239.

[12] D.B. Shmoys and D.B. West, Recognizing graphs with fixed interval number is NP-complete, *Discrete Appl. Math.* 8(1984), 295–305.

[13] J. Spinrad and G. Vijayan, and D.B. West, An improved edge bound on the interval number of a graph *J. Graph Theory* 11(1987), 447–449.

[14] G. Tardos, Transversals of 2-intervals, a topological approach, submitted.

[15] W.T. Trotter, A characterization of Roberts' inequality for boxicity, *Discrete Math.* 28(1979), 303-313.

[16] W.T. Trotter and F. Harary, On double and multiple interval graphs, *J. Graph Theory* 2(1978), 137–142.

[17] D.B. West, A short proof of the degree bound on interval number, *Discrete Math.* 73(1989), 309–310.