

Chapter 3

Interval Graphs

An *interval graph* is defined to be any graph that is isomorphic to the intersection graph of a family of finite closed intervals of the real line, with each vertex v corresponding to a closed interval J_v ; the family of intervals is called an *interval representation* for the interval graph.

Interval graphs were first studied in [Hajós, 1957]. The standard references are section 3.4 of [Roberts, 1976] and Chapter 8 of [Golumbic, 1980].

Example 3.1 The graph shown on the left in Figure 3.1 could have the representation $J_a = [1, 4]$, $J_b = [1, 1]$ (a single-point closed interval), $J_c = [1, 2]$, $J_d = [2, 3]$, $J_e = [3, 4]$, and $J_f = [4, 4]$. If you are squeamish about length-zero intervals, you could of course use $J_b = [1, 1.1]$ and $J_f = [3.1, 4]$ instead. You could also use all open intervals, instead of closed.

3.1 Definitions and Characterizations

Much as in Chapter 2 for subtrees of a tree, it is easy to see that we can equivalently define interval graphs using subpaths of a path and so talk about *path representations*. Since subpaths of a path satisfy the Helly condition, Lemma 1.11 can be used to show that every interval graph has a *clique path representation*, or *clique path* (paralleling Theorem 2.1 for trees).

Example 3.1 (continued) The graph shown in Figure 3.1 has the clique path P shown there in which P_a has length three, P_b and P_f have length zero, and P_c , P_d , and P_e have length one.

Since paths are trees, interval graphs are chordal graphs, and so the cycle C_4 is a cheap example of a graph that is not an interval graph. A significantly different example would result from adding an edge dg to the graph

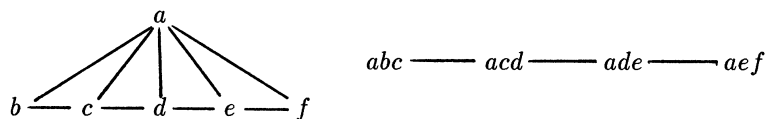


Figure 3.1: An interval graph with a path representation.

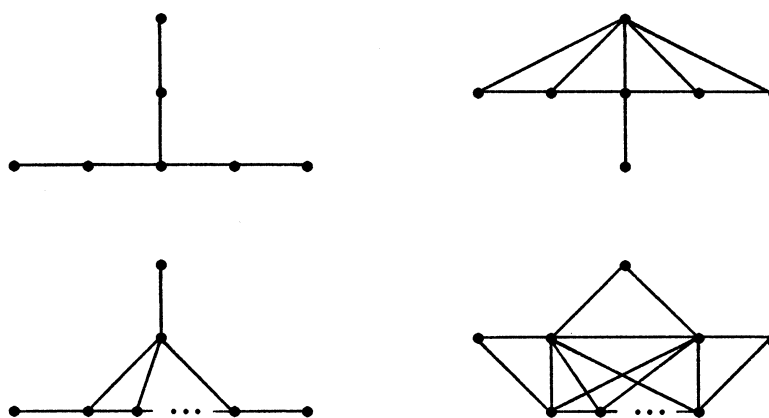


Figure 3.2: Some graphs that are not interval graphs; each of the lower two has order at least six.

on the left in Figure 3.1, producing the upper-right graph in Figure 3.2. (It might be instructive to attempt to find an interval representation or a path representation for that graph right now, instead of waiting for the characterizations just ahead in this chapter that show why it is impossible.) Indeed, none of the graphs shown in Figure 3.2 is an interval graph, while the graph in Figure 2.2 is an interval graph.

The following theorem consolidates the connection between interval and chordal graphs, extending a well-known theorem from [Fulkerson & Gross, 1965].

Theorem 3.1 *A connected graph G is an interval graph if and only if some maximum spanning tree of the weighted clique graph $K^w(G)$ is a clique path for G . Moreover, this is equivalent to every maximum spanning tree of the weighted clique graph $K^w(G)$ that is a path being a clique path for G .*

Proof. This is immediate from Theorem 2.3. \square

Corollary 3.2 (Fulkerson & Gross) *A graph is an interval graph if and only if the edge clique cover of all maxcliques can be arranged into a clique path representation.* \square

It is impractical to look for clique path representations by using Kruskal's algorithm to find all the maximum spanning trees and then trying to cull the nonpaths (indeed, being able to do that would be tantamount to solving the NP-hard problem of recognizing graphs that have hamiltonian paths). Yet interval graphs can be recognized efficiently. [Booth & Leuker, 1976] contains the classical, linear-time recognition algorithm, using the influential "PQ-tree" data structure that was introduced for that purpose; see also section 8.3 of [Golumbic, 1980]. [Simon, 1991], [Hsu & Ma, 1991], [Hsu, 1993], and [Corneil, Olariu, & Stewart, 1998] contain more recent recognition algorithms.

Theorem 3.1 relates to two prescient applications: Benzer's 1959 study of the fine structure of the gene, and Petrie's late nineteenth century work with archaeological seriation (see [Roberts, 1976] or [Golumbic, 1980]). Interval graphs *could* have been used in these contexts, but working with the appropriate incidence matrices as in Corollary 3.9 was quite sufficient. Section 3.4 contains "real" applications of interval graphs.

Three vertices form an *asteroidal triple* in a graph G if, for each two, there exists a path containing those two but no neighbor of the third. For instance, the three vertices of degree one in the upper-left graph in Figure 3.2 form an asteroidal triple. Notice that no two vertices of an asteroidal triple can be adjacent. (Section 7.6 will discuss those graphs that have no asteroidal triples.)

Exercise 3.1 Show that each of the graphs in Figure 3.2 has a unique asteroidal triple.

Theorem 3.3 is from [Lekkerkerker & Boland, 1962]. Recall from Chapter 2 that a graph is chordal if and only if contains no cycle C_k having $k \geq 4$ as an induced subgraph.

Theorem 3.3 (Lekkerkerker & Boland) *A graph is an interval graph if and only if it is chordal and has no asteroidal triple.*

Proof. First suppose G is an interval graph with a clique path P . As we have already observed, P is also a clique tree and so G is chordal by

Theorem 2.1. Suppose u, v, w are pairwise nonadjacent and, without loss of generality, that P_v is in between P_u and P_w along P . Since P is a clique path, every path from u to w in G will have to contain a neighbor of v , and so u, v, w cannot be an asteroidal triple.

Conversely, suppose G is a chordal graph and, among all clique trees for G , that T has a minimum number of leaves and that number is at least three; we show that G must then contain an asteroidal triple. Suppose Q_1, Q_2 , and Q_3 are three different leaves of T and let, respectively, Q'_1, Q'_2 , and Q'_3 (not necessarily distinct) be their unique neighbors in T .

For each $i = 1, 2, 3$, choose $v_i \in V(Q_i)$ such that $v_i \notin V(Q'_i)$. We show that $\{v_1, v_2, v_3\}$ is an asteroidal triple. Suppose rather, without loss of generality, arguing toward a contradiction, that every path in G connecting v_1 and v_3 contains a neighbor of v_2 . Not every edge of the path $T(Q_1, Q_3)$ in T can contain a nonneighbor of v_2 , since otherwise those nonneighbors could be used to induce a v_1 -to- v_3 path in G that contained no neighbor of v_2 . Therefore, the path $T(Q_1, Q_3)$ in T would contain some edge Q^*Q^{**} with $Q^* \cap Q^{**}$ consisting entirely of neighbors of v_2 , making $Q^* \cap Q^{**} \subseteq Q_2 \cap Q'_2$. Without loss of generality, suppose Q^* is closer to Q_2 in T than is Q^{**} . Then replacing edge Q^*Q^{**} with a new edge $Q^{**}Q_2$ would create a clique tree for G that has one fewer leaf than T , which is a contradiction. \square

Using Theorem 3.3, [Lekkerkerker & Boland, 1962] proves that a graph is an interval graph if and only if it is chordal and contains none of the graphs in Figure 3.2 as an induced subgraph. [Harary & Kabell, 1984] contains a similar characterization of “infinite-interval graphs” in which the intervals are taken to be one- or two-way infinite intervals of the real line.

Exercise 3.2 Let G be a split graph. Show that G is an interval graph if and only if G contains none of the graphs in Figure 3.3 as an induced subgraph.

Before stating the next characterization of interval graphs, we need to review some terminology and results about directed graphs (digraphs). A digraph D is defined to have a vertex set $V(D)$ and a set $A(D)$ of arcs, where $vw \in A(D)$ denotes an arc from vertex v to vertex w . We assume that there are no multiple arcs (meaning that there are never two arcs from v to w , although it is possible to have both $vw, wv \in A(D)$) and, in this chapter, no loops (meaning no $vv \in A(D)$). A digraph is *transitive* if, for $u, v, w \in V(D)$, $uv, vw \in A(D)$ with $u \neq w$ implies that $uw \in A(D)$. Given any graph G , an *orientation* of G is a digraph formed by specifying a direction for each edge

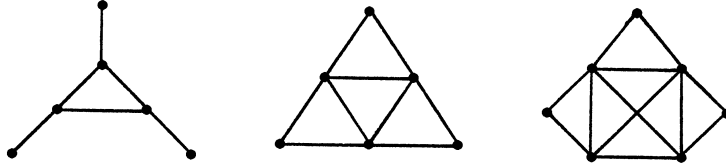


Figure 3.3: *Three split graphs that are not interval graphs.*

of G , producing an *oriented graph*. The orientation is a *transitive orientation* if the oriented graph is transitive.

Exercise 3.3 Show that the cycle C_4 has a transitive orientation but that C_5 does not.

A *directed hamiltonian path* of a digraph is a directed path that includes every vertex. A *tournament* is an orientation D of a complete graph; thus $u, v \in V(D)$ and $u \neq v$ imply that either $uv \in A(D)$ or $vu \in A(D)$ but not both. The following is from [Rédei, 1934].

Lemma 3.4 (Rédei) *Every tournament has a directed hamiltonian path.*

Proof. Suppose D is a tournament. We argue by induction on $n = |V(D)|$, with the result trivial for $n \leq 2$. Suppose $n > 2$ and $v \in V(D)$. By induction hypothesis, the tournament $D - v$ has a directed hamiltonian path v_1, v_2, \dots, v_{n-1} . If $vv_1 \in A(D)$, then v, v_1, \dots, v_{n-1} is a directed hamiltonian path in D . Otherwise, $v_1v \in A(D)$ and we choose i to be the largest integer for which $v_iv \in A(D)$. If $i = n$, then v_1, \dots, v_{n-1}, v is a directed hamiltonian path in D . If $i < n$, then $vv_{i+1} \in A(D)$, making $v_1, \dots, v_i, v, v_{i+1}, \dots, v_{n-1}$ a directed hamiltonian path in D . \square

Exercise 3.4 Show that every transitive tournament of order n has a unique vertex of each possible out-degree $0, \dots, n - 1$ and that taking these in order determines a directed hamiltonian path.

Exercise 3.5 Show that every transitive tournament has a *unique* directed hamiltonian path.

Recall that, for a graph G , the *complement* of G , denoted \overline{G} , is the graph having $V(\overline{G}) = V(G)$ where, for any distinct vertices u and v of G , $uv \in E(\overline{G})$ if and only if $uv \notin E(G)$.

We are finally ready to state and prove the characterization from [Gilmore & Hoffman, 1964].

Theorem 3.5 (Gilmore & Hoffman) *A graph is an interval graph if and only if it does not contain C_4 as an induced subgraph and its complement has a transitive orientation.*

Proof. First suppose G is an interval graph with clique path P laid out horizontally. Since G must be chordal, it cannot contain an induced C_4 .

Form an oriented graph \vec{G} by putting $uv \in A(\vec{G})$ if and only if P_u is totally to the left of P_v in P (i.e., every vertex of P_u is to the left of every vertex of P_v), noting that $uv \in A(\vec{G})$ implies $P_u \cap P_v = \emptyset$ and so $uv \notin E(G)$. It is easy to see that this is a transitive orientation of \vec{G} .

Conversely, suppose G contains no induced C_4 and that \vec{G} has a transitive orientation \vec{G} . We form a digraph D whose vertices are precisely the maxcliques of G , with arcs as follows: For every two maxcliques Q, Q' of G pick $v \in Q$ and $v' \in Q'$ such that $vv' \notin E(G)$, and then put $QQ' \in A(D)$ if and only if $vv' \in A(\vec{G})$. Of course we must show that D really is well defined. Arguing toward a contradiction, suppose that $u, v \in Q$ and $u', v' \in Q'$ where $uu', vv' \notin E(G)$ and $uu', v'v \in A(\vec{G})$. Observe that either $uv' \notin E(G)$ or $u'v \notin E(G)$, since the cycle u, v, u', v', u cannot be induced in G . Without loss of generality, we suppose that $uv' \notin E(G)$. Thus either $uv' \in A(\vec{G})$ or $v'u \in A(\vec{G})$. If $uv' \in A(\vec{G})$, then $uv', v'v \in A(\vec{G})$ forces $uv \in A(\vec{G})$, since \vec{G} is transitively oriented. But then $uv \in E(\vec{G})$, contradicting that u and v are in a common maxclique of G . A similar contradiction occurs if $v'u \in A(\vec{G})$. Thus D is well defined.

It is easy to check that D is transitive since \vec{G} is transitive, so D is a transitive tournament. By Lemma 3.4, D has a directed hamiltonian path $P : Q_1, \dots, Q_m$. We now show that P is a clique path for G . Suppose that $v \in V(G)$ is in two nonadjacent vertices Q, Q'' of P yet, arguing toward a contradiction, that $v \notin Q'$ for some $Q' \in V(P(Q, Q''))$. Without loss of generality, we can assume that $QQ'' \in A(D)$. Pick $u, w \in Q'$ such that $u \notin Q$ and $uv \notin E(G)$, while $w \notin Q''$ and $wv \notin E(G)$. Then $vu, wv \in A(\vec{G})$, so $u \neq w$ and, by transitivity, $wu \in A(\vec{G})$. But then $uw \notin E(G)$, contradicting that $u, w \in Q'$. \square

Example 3.2 Figure 3.4 shows one possible transitive orientation of \vec{G} , where G is the graph in Figure 3.1. Check that the only other transitive orientation is the reverse of this one. The corresponding digraph D used in the preceding proof is also shown. Notice that the directed hamiltonian path in D corresponds to the clique path in Figure 3.1.

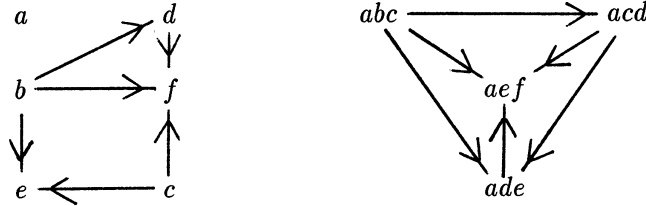


Figure 3.4: A transitively oriented digraph to illustrate the proof of Theorem 3.5.

[Opsut & Roberts, 1981] shows that the intersection number of an interval graph equals the number of maxcliques minus the number of isolated vertices.

3.2 Interval Hypergraphs

Continuing the discussion of tree hypergraphs in section 2.3, a hypergraph (X, \mathcal{E}) is an *interval hypergraph* if there is a path P with $X = V(P)$ such that, for each $S_i \in \mathcal{E}$, there is a subpath P_i of P with $V(P_i) = S_i$.

Exercise 3.6 Show that the hypergraph given in Exercise 2.17 is an interval hypergraph.

Precisely as in the easy direction of Theorem 2.7, every interval hypergraph must be a Helly hypergraph with an interval line graph. But, unlike what happened for tree hypergraphs, the following exercise shows that the converse fails.

Exercise 3.7 Show that the hypergraph $(\{a, b, c, d\}, \mathcal{E})$ with $\mathcal{E} = \{\{a, c\}, \{b, c\}, \{c, d\}\}$ is a Helly hypergraph and its line graph $\Omega(\mathcal{E})$ is an interval graph, but no path P exists as required in the definition of an interval hypergraph.

A *path in a hypergraph* (X, \mathcal{E}) is a sequence $v_0, S_1, v_1, S_2, \dots, S_k, v_k$ where S_1, \dots, S_k are distinct edges, v_0, \dots, v_k are distinct vertices, and each $v_{i-1}v_i \in S_i$; such a path is said to *join* the vertices v_0 and v_k . The hypergraph (X, \mathcal{E}) is *connected* if every two vertices are joined by a path. A vertex v is said to *lie between* vertices u and w in a hypergraph if every path in the hypergraph that joins u and w contains an edge that contains v . [Duchet, 1978, 1984]

contain proofs of the following characterization that is in the spirit of the Fulkerson–Gross result in Corollary 3.2.

Theorem 3.6 (Duchet) *A connected hypergraph is an interval hypergraph if and only if, for every three vertices, one of them lies between the other two.*

Proof. First, suppose (X, \mathcal{E}) is any connected hypergraph and there is a path P with $X = V(P)$ for which each $S_i \in \mathcal{E}$ corresponds to a subpath P_i of P such that $V(P_i) = S_i$. Then a vertex y lies between vertices x and z along P if and only if y lies between x and z in the hypergraph.

Conversely, suppose that (X, \mathcal{E}) satisfies the condition in the theorem—so for every three vertices, one of them lies between the other two. Choose a hypergraph (X, \mathcal{E}^*) with $\mathcal{E} \subseteq \mathcal{E}^*$ such that, among hypergraphs satisfying the condition in the theorem, \mathcal{E}^* is maximal. Let \mathcal{E}' be the set of minimal edges S of \mathcal{E}^* for which $|S| \geq 2$.

Suppose a and b are distinct vertices in $S \in \mathcal{E}'$. If $S \neq \{a, b\}$, then $\{a, b\} \notin \mathcal{E}^*$ and so the hypergraph $(X, \mathcal{E}^* \cup \{\{a, b\}\})$ will not satisfy the condition in the theorem— X contains x, y, z and $(X, \mathcal{E}^* \cup \{\{a, b\}\})$ contains minimal-length paths

$$\begin{aligned} x, E_1, \dots, E_p, y \text{ with } z \notin E_1 \cup \dots \cup E_p, \\ y, E'_1, \dots, E'_q, z \text{ with } x \notin E'_1 \cup \dots \cup E'_q, \\ x, E''_1, \dots, E''_r, z \text{ with } y \notin E''_1 \cup \dots \cup E''_r, \end{aligned}$$

where $\{a, b\} \in \{E_1, \dots, E_p, E'_1, \dots, E'_q, E''_1, \dots, E''_r\}$. But if each occurrence of $\{a, b\}$ among the E_i 's, E'_i 's, and E''_i 's is replaced by $S \in \mathcal{E}^*$, then by the condition in the theorem one of x, y, z will be between the other two; without loss of generality, say that y is between x and z . That means that $\{a, b\} = E''_i$ for some $1 \leq i \leq r$, and so $a \neq y \neq b$ and $y \in S$. Without loss of generality, using the minimality of the path $x, E''_1, \dots, E''_r, z$, we can suppose that $a \in E''_{i-1}$ (or, possibly, $a = x$). By the assumed maximality of \mathcal{E}^* , we can assume that $A = E''_1 \cup \dots \cup E''_{i-1} \in \mathcal{E}^*$ (or, if $a = x$, that $A = \{a\} \in \mathcal{E}^*$). Thus there exists $A \in \mathcal{E}^*$ such that $a, x \in A$ and $b, y \notin A$. Similarly, there exists $B \in \mathcal{E}^*$ such that $b, z \in B$ and $a, y \notin B$. Again using the assumed maximality of \mathcal{E}^* , we can assume that $S \setminus A, S \setminus B \in \mathcal{E}^*$, and so a and b are connected by the path $a, S \setminus B, y, S \setminus A, b$ in (X, \mathcal{E}^*) .

Thus we have shown that every pair of vertices in an edge $S \in \mathcal{E}^*$ are linked by a path in (X, \mathcal{E}^*) whose edges are subsets of S , and that every edge $S \in \mathcal{E}'$ has cardinality two, since otherwise $S \setminus B \in \mathcal{E}^*$ would contradict S 's assumed minimality. Therefore, (X, \mathcal{E}') is a graph. The assumed maximality of \mathcal{E}^* implies that $X \in \mathcal{E}^*$, so (X, \mathcal{E}') is connected, and the condition

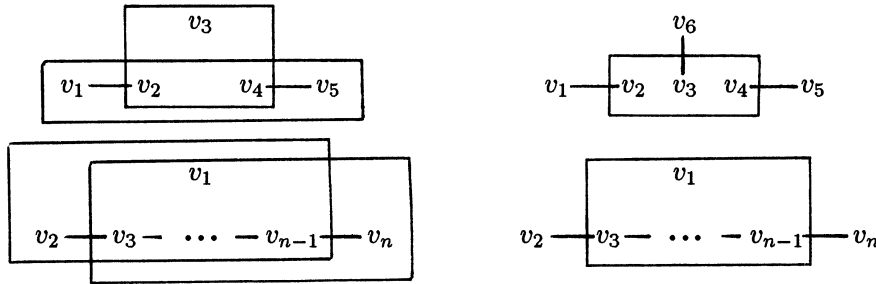


Figure 3.5: *The four noncycle forbidden subhypergraphs for an interval hypergraph.*

in the theorem implies that (X, \mathcal{E}') is a path. Finally, every edge $S \in \mathcal{E}^*$ can be seen to be a connected subset of (X, \mathcal{E}') by an inductive argument on $|S|$. \square

[Tucker, 1972] proves that a hypergraph is an interval hypergraph if and only if it contains none of the following five induced subhypergraphs $(\{v_1, \dots, v_n\}, \mathcal{E})$ (see Figure 3.5 for the last four):

- (1) $n \geq 3$ and $\mathcal{E} = \{\{v_1, v_2\}, \{v_2, v_3\}, \dots, \{v_n, v_1\}\}$.
- (2) $n = 5$ and $\mathcal{E} = \{\{v_1, v_2\}, \{v_2, v_3, v_4\}, \{v_4, v_5\}, \{v_1, v_2, v_4, v_5\}\}$.
- (3) $n = 6$ and $\mathcal{E} = \{\{v_1, v_2\}, \{v_2, v_3, v_4\}, \{v_4, v_5\}, \{v_3, v_6\}\}$.
- (4) $n \geq 4$ and $\mathcal{E} = \{\{v_2, v_3\}, \dots, \{v_{n-1}, v_n\}, \{v_1, v_2, \dots, v_{n-1}\}, \{v_1, v_3, \dots, v_n\}\}$.
- (5) $n \geq 4$ and $\mathcal{E} = \{\{v_2, v_3\}, \dots, \{v_{n-1}, v_n\}, \{v_1, v_3, \dots, v_{n-1}\}\}$.

[Trotter & Moore, 1976] gives a shorter proof, and [Duchet, 1984] contains a short proof using Theorem 3.6.

See [Lehel, 1983] and [Duchet, 1984, 1995] and their references for more on various sorts of representation of hypergraphs by intervals.

3.3 Proper Interval Graphs

A *proper interval graph* is the intersection graph of a family of closed intervals of the real line, none of which is *properly* contained in another. This is equivalent to being the intersection graph of a family of subpaths of a path, none of which is a proper subpath of another; such a path is called a *proper path representation* of the graph. For instance, Figure 3.6 shows a proper

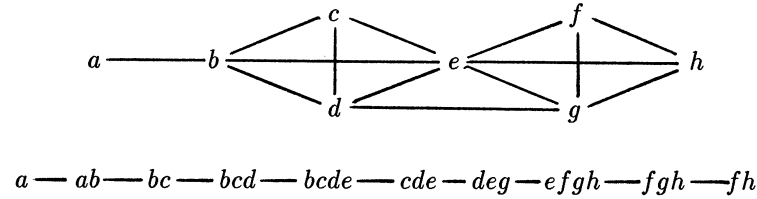


Figure 3.6: A proper interval graph with a proper path representation.

interval graph and a proper path representation for it. (Notice that we cannot have the vertices of the path just be the maxcliques, as was true for clique paths.) Proper interval graphs were introduced in [Roberts, 1969a] as “indifference graphs,” for reasons we discuss in subsection 3.4.2; see also section 8.5 of [Golumbic, 1980]. They also were introduced in epidemiology as “time graphs”; see [Hedman, 1984].

Exercise 3.8 Show that a proper interval graph cannot contain $K_{1,3}$ (the graph on the left in Figure 3.7) as an induced subgraph, and so that the graph in Figure 3.1 is not a proper interval graph.

Exercise 3.9 Show that the clique graph of an interval graph must be a proper interval graph.

Exercise 3.10 Show that every chordal graph has a “proper tree representation,” meaning a tree representation T in which no T_v is properly contained in a T_w .

Theorem 3.7 (Roberts) *A graph is a proper interval graph if and only if it is an interval graph that does not contain an induced subgraph isomorphic to $K_{1,3}$.*

Proof. Exercise 3.8 gives the implication one way. For the converse, suppose G has a clique path P and contains no induced subgraph isomorphic to $K_{1,3}$. Suppose the subpath P_v is properly contained in P_w . We first show that there cannot be vertices Q^ℓ, Q^r in P_w with P_v in between such that there are $x \in Q^\ell \setminus \{Q : Q \in P_v\}$ and $y \in Q^r \setminus \{Q : Q \in P_v\}$, since otherwise v, w, x, y would induce a $K_{1,3}$ centered at w . Thus we can assume that P_v and P_w share one common end-vertex. Without loss of generality, assume

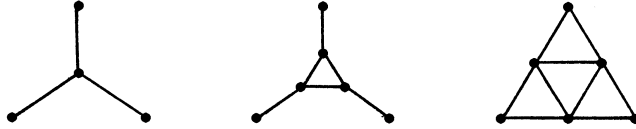


Figure 3.7: *Three graphs that are not proper interval graphs.*

Q is their common left end-vertex. Now modify P by inserting a new vertex Q' just to the left of Q with $Q' = Q \setminus \{u : Q \text{ is a left end-vertex of } P_u \text{ and } P_v \text{ is properly contained in } P_u\}$. After this modification, P is still a path representation of G but P_v is no longer contained in P_w (or in any of the P_u 's in the definition of Q'). Repeating this lengthens P into a proper path representation of G . \square

Exercise 3.11 Use the construction in the proof of Theorem 3.7 to make a clique path for the graph in Figure 3.6 into a proper path representation.

Exercise 3.12 (Roberts) Show that a graph is a proper interval graph if and only if it contains no cycle of length greater than or equal to four and contains none of the graphs of Figure 3.7 as induced subgraphs.

[Wegner, 1967] and [Roberts, 1969a] define a *unit interval graph* to be the intersection graph of a family of closed intervals of the real line, all of which have the same length (which is often taken to be one). This is equivalent to being the intersection graph of a family of subpaths of a path, all of which have the same length; such a path is called a *unit path representation* of the graph. Since every unit path representation is a proper path representation, every unit interval graph is a proper interval graph. But the following theorem shows that much more is true.

Theorem 3.8 (Roberts) *A graph is a proper interval graph if and only if it is a unit interval graph.*

Proof. As we observed, the implication one way is immediate. To prove the converse, suppose G is a proper interval graph. Arguing inductively, suppose that, for every proper subgraph G' of G , every proper path representation of G' can be made into a unit path representation of G' by

simply inserting duplicates of some of the vertices into the path. Suppose P is a proper path representation for G and pick a $v \in V(G)$ that occurs in an end-vertex of P . Obtain a proper path representation P^+ from P by inserting duplicate vertices into P so that removing all occurrences of v from P^+ would leave a unit path representation of the subgraph induced by $V(G) \setminus \{v\}$. We can assume that v still occurs in an end-vertex of P^+ . Let $k = |V(P_w^+)|$ for each $w \neq v$. Then $k \geq |V(P_v^+)|$, since P^+ is still a proper path representation, and so adding $k - |V(P_v^+)|$ new vertices, each equal to $\{v\}$, to the end-vertex that contains v will produce a unit path representation of G . \square

[Jackowski, 1992] defines an *astral triple* in a graph as three vertices such that, for each two, there exists a path containing those two but not the third vertex that does not have two consecutive vertices that are neighbors of the third. For instance, the three vertices of degree one in the graph on the left in Figure 3.7 form an astral triple (but not an asteroidal triple). Paralleling the characterization of interval graphs in Theorem 3.3, Jackowski proves that a graph is a proper interval graph if and only if it contains no astral triple.

Exercise 3.13 (Jackowski) Show that every nonchordal graph contains an astral triple of vertices.

For any graph G with vertices indexed by $\{1, \dots, n\}$ and maxcliques indexed by $\{1, \dots, m\}$, define the *maxclique-vertex matrix* $M(G)$ to be the $m \times n$ matrix with entry $m_{ij} = 1$ if the i th maxclique contains the j th vertex, and $m_{ij} = 0$ otherwise. For instance, the graph G in Figure 3.6 has the maxclique-vertex matrix

$$M(G) = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix},$$

where the columns correspond to the vertices in alphabetical order and the rows correspond to the maxcliques in the order $ab, deg, bcde, efgh$.

A matrix has the *consecutive ones property for columns* if its rows can be permuted so as to make all the 1 entries in each column consecutive. The *consecutive ones property for rows* is defined similarly. For instance, you can show that the above matrix has the consecutive ones property for columns by interchanging the second and third rows; it also has the consecutive ones property for rows. The following merely rephrases Corollary 3.2.

Corollary 3.9 *A graph G is an interval graph if and only if $M(G)$ has the consecutive ones property for columns.* \square

The following result seems to have been first stated in this form in [Deogun & Gopalakrishnan, to appear], in connection with the application in section 3.4.3, although all the pieces were certainly in [Roberts, 1968]. Section 7.1 contains more about variations of the consecutive ones properties.

Theorem 3.10 *A graph G is a proper interval graph if and only if $M(G)$ has the consecutive ones property for both rows and columns.*

Proof. First suppose G has a proper path representation P . Corollary 3.9 shows that $M(G)$ has the consecutive ones property for columns. Let v_1, \dots, v_n be the vertices in the order of their leftmost appearance along P . Suppose $i < j < k$ and $v_i v_k \in E(G)$. Then P_{v_i} (the subpath of P corresponding to v_i) will intersect P_{v_k} and so also P_{v_j} , ensuring that $v_i v_j \in E(G)$. Because P_{v_j} cannot be properly contained in P_{v_i} , path P_{v_j} will have to intersect P_{v_k} , ensuring that $v_j v_k \in E(G)$. Thus, each maxclique of G will correspond to consecutive vertices in the ordering v_1, \dots, v_n , and this means that $M(G)$ has the consecutive ones property for rows.

Conversely, suppose $M(G)$ has the consecutive ones property for both columns and rows. The former of these implies that G is an interval by Corollary 3.9. The latter implies that G does not contain $K_{1,3}$ as an induced subgraph, and so G is a proper interval graph by Exercise 3.7. \square

Sections 3.4.2, 7.1, and 7.2 contain other characterizations of proper interval graphs, and [Gutierrez & Oubiña, 1996] contains various order-theoretic characterizations. [Gutierrez & Oubiña, 1995] shows that every proper interval graph satisfies

$$|V(G)| \geq 2c(G) - c(K(G)),$$

where $K(\cdot)$ is the clique graph operator from section 1.4 and $c(\cdot)$ counts the number of maxcliques and then investigates the graphs for which equality holds.

See [Corneil, Kim, Natarajan, Olariu, & Sprague, 1995], [Hell & Huang, 1995], and [de Figueiredo, Meidanis, & de Mello, 1995] for recognition algorithms for proper interval graphs, and [Hell & Huang, 1995] and [Deng, Hell, & Huang, 1996] for representation algorithms.

[Leibowitz, Assman, & Peck, 1982] generalizes the notion of a unit interval graph by defining the “interval count” of an interval graph to be the

minimum number of different lengths of intervals needed in an interval representation. [Skrien, 1984] characterizes graphs that have interval count two, where one of the allowed lengths is zero—in other words, the intersection graphs of points and unit intervals; section 5.2 will show that “threshold” graphs are of this type.

[Pe’er & Shamir, 1995] investigates a host of restrictions on interval graphs, including bounding the maximum lengths of intervals.

3.4 Some Applications of Interval Graphs

Each of the following subsections is merely a brief sketch of one application of interval graphs. As we did in section 2.4, we have selected applications that make essential use of the intersection definition of interval graphs, rather than other important applications that involve interval graphs. One example of the latter involves on-line coloring algorithms: a graph is presented one vertex at a time, along with its neighbors among earlier vertices, and the graph is to be properly colored with as few colors as possible. This is a highly practical problem in many contexts, dynamic storage problems for one. Papers such as [Gyárfás & Lehel, 1988], [Ślusarek, 1989], [Kierstead, 1991] and [Kierstead & Qin, 1995] contain results for interval graphs; the first of these also studies proper interval graphs, while [Ślusarek, 1995] and [Marathe, Hunt, & Ravi, 1996] study circular-arc graphs (section 7.1).

3.4.1 Applications to Biology

Probably the first paper on interval graphs was an application in biology. Although stated in terms of incidence matrices rather than graphs, the question in [Benzer, 1959] was whether certain fragment overlap data on the DNA making up a bacterial gene was consistent with the gene having a linear structure—in other words, whether the graph constructed from the data was an interval graph. Of course today we know that the gene is indeed a linear arrangement and, as mentioned in section 2.4.1, DNA strands are sequences (words) built from the four letter alphabet $\{A, C, G, T\}$.

One of the problems involving DNA is to try to assemble subsequences involving possible overlaps into longer sequences. Certainly one would expect interval graphs and their variants to be useful, and indeed this has been the case. [Jungeck, Dick, & Dick, 1982] is a very readable introductory paper. See [Fellows, Hallett, & Wareham, 1993], [Goldberg, Golumbic, Kaplan, & Shamir, 1995], and [Nicholson, 1995] for more recent views. [Mirkin

& Rodin, 1984] and [Waterman, 1995] are two excellent books with in-depth treatment.

We now briefly present some oversimplified background relevant to what is called *physical mapping* of DNA. Sequence fragments called *clones* are obtained from an unknown DNA sequence and form a “clone library.” Experiments are then carried out that can decide if a very short molecule, called a *probe*, overlaps each clone. The goal is to reconstruct the placement of the clones along the DNA sequence, the sequence having been destroyed during the construction of the clone library. If only some of the clones are used as probes, then the overlap information is not available between clones that are not probes. [Zhang, to appear] introduced the following generalization of interval graphs to deal with this situation.

A graph G is a *probe interval graph* if $V(G)$ can be partitioned into subsets P and N (corresponding to the probes and nonprobes) and each $v \in V(G)$ can be assigned to an interval I_v such that $uv \in E(G)$ if and only if both $I_u \cap I_v \neq \emptyset$ and at least one of u and v is in P . Interval graphs are simply probe interval graphs with $N = \emptyset$.

Exercise 3.14 Show that, although C_4 and the graph in the middle of Figure 3.7 are not interval graphs, they are both probe interval graphs.

Results on probe interval graphs and their variants can be found in [Zhang, to appear], [McMorris, Wang, & Zhang, to appear], [Wan, Lee, Wang, & Zhang, to appear] and [Sheng, Wang, & Zhang, to appear]; also see [Atkins & Middendorf, 1996]. The following result considerably restricts the possible structure of probe interval graphs; the graphs described therein are the *weakly chordal graphs*, that are discussed further in section 7.3.

Theorem 3.11 (McMorris, Wang, & Zhang) *Neither a probe interval graph nor its complement can contain an induced cycle of length greater than or equal to five.*

Proof. Suppose G is a probe interval graph with respect to the partition $V(G) = P \cup N$, with I_v the interval assigned to each $v \in V(G)$. Let G^* be defined precisely the same as G except with $uv \in E(G^*)$ if and only if $I_u \cap I_v \neq \emptyset$ (*without* the addition $\{u, v\} \cap P \neq \emptyset$ assumption). Clearly G^* is an interval graph. Suppose C is an induced cycle of G of length at least four and $u, v \in P$ are adjacent along C . Since the only edges in $E(G^*) \setminus E(G)$ are between vertices of N , u and v have the same neighborhoods in G and G^* . But then some subset of the vertices of C would induce a chordless cycle

containing u and v of length at least four in G^* , contradicting that G^* is an interval (and so chordal) graph. Therefore, vertices of P and N alternate on C , and so G has no induced cycles of odd length greater than three.

Now suppose C is an induced cycle of G of even length at least six. Let $x, y, z \in P$ be nonadjacent vertices along C , and let paths P_1, P_2 , and P_3 be, respectively, the segments of C between x and y , between y and z , and between z and x ; thus z is not adjacent to any vertex on P_1 , x is not adjacent to any vertex on P_2 , and y is not adjacent to any vertex on P_3 . Since $x, y, z \in P$, they have no new neighbors in G^* and so form an asteroidal triple in G^* , again contradicting that G^* is an interval graph. Therefore, a probe interval graph has no induced cycles of length greater than or equal to five.

To show that G contains no complement of an induced cycle of length larger than four, first notice that the complement of an induced cycle of length five would also be an induced cycle of length five, which we now know is impossible in G . Therefore, we only need to show that G contains no complement of an induced cycle of length six or more. Suppose to the contrary that $\{v_1, \dots, v_n\}$ induces a complement of an induced cycle in G with $n \geq 6$, where $v_1v_n \notin E(G)$ and each $v_iv_{i+1} \notin E(G)$, while all other v_iv_j 's are in $E(G)$. Then v_1, v_4, v_2, v_5, v_1 will be an induced cycle in G , and so we can assume without loss of generality that $v_1, v_2 \in P$ and $v_4, v_5 \in N$. Therefore, $v_3v_5 \in E(G)$ implies that $v_3 \in P$ and $v_4v_6 \in E(G)$ implies that $v_6 \in P$. But then v_2, v_5, v_3, v_6, v_2 would be a length-four chordless cycle in G having three vertices from P , contradicting vertices from P and N alternating around induced cycles in G . \square

Exercise 3.15 (Zhang) Show that “enhancing” a probe interval graph by adding edges between pairs of nonprobes that have two nonadjacent probes as common neighbors produces a chordal graph.

Also related to DNA matters, [Bodlaender & de Fluiter, 1996] discusses a “chromatic interval completion problem,” paralleling the chromatic chordal completion problem—inserting edges so as to make an interval graph—in section 2.4.1. Going the other way, [Wang, 1994] discusses removing edges from a bipartite graph so as to leave an interval graph.

3.4.2 Applications to Psychology

While classical theories of measurement are based on the physical sciences, much work has also been done on notions of measurement that are more

suitable for the social sciences. Much of this work has been within the context of psychology. Our discussion below is based on [Roberts, 1976, 1978b]. [Roberts, 1979] is a related treatment of various aspects of “measurement theory.”

Suppose A is a set of alternatives, such as types of cars or food products, and a person has preferences among the elements of A . [Luce, 1956] motivates seeking a real-valued function f on A such that, for $a, b \in A$, preferring alternative a to alternative b implies $f(a) > f(b) + \delta$, where the positive constant δ represents a *threshold* or *just noticeable difference* between alternatives.

Define a binary relation R on a finite set A to be an *interval order* if it satisfies the two following axioms.

Axiom 1: For all $a \in A$, not aRa .

Axiom 2: For all $a, b, c, d \in A$, if aRb and cRd , then either aRd or cRb .

Exercise 3.16 Suppose f is a real-valued function defined on A , and aRb is defined to mean that $f(a) > f(b) + \delta$, where δ is a positive constant. Show that R is an interval order on A .

For any binary relation R on any finite set A , define the graph $G(R)$ to have $V(G(R)) = A$ with $ab \in E(G(R))$ if and only if neither aRb nor bRa ; edges thereby correspond to “indifference” with respect to R . (Warning: An “indifference graph” per se is defined somewhat differently and is equivalent to being a proper interval graph.) The following result is from [Fishburn, 1970a, 1970b].

Proposition 3.12 (Fishburn) *A binary relation R on a finite set A is an interval order on A if and only if R is transitive and $G(R)$ is an interval graph.*

Proof sketch. First suppose R satisfies Axioms 1 and 2; transitivity follows directly. Theorem 3.5 shows that $G = G(R)$ is an interval graph as follows: G cannot contain an induced cycle a, b, c, d, a since $ac, bd \notin E(G)$ would imply both $(aRc$ or $cRa)$ and $(bRd$ or $dRb)$, and each of the four possible cases would lead to a contradiction using Axiom 2; and R is itself a transitive orientation of \overline{G} .

Conversely, suppose R is transitive and $G = G(R)$ is an interval graph. Axiom 1 follows from G being loopless. Arguing toward a contradiction

with Axiom 2, suppose aRb and cRd , so $ab, cd \notin E(G)$, yet neither aRd nor cRb . Then by transitivity, neither dRa nor bRc , so $ad, bc \in E(G)$. Transitivity similarly shows $ac, bd \in E(G)$, producing an induced cycle a, d, b, c, a , contradicting Theorem 3.5. \square

Exercise 3.17 Discuss whether or not our proof of Proposition 3.12 actually shows something stronger: that a binary relation R on a finite set A is an interval order on A if and only if R is transitive and $G(R)$ contains no induced C_4 .

Define a binary relation R on a finite set A to be a *semiorder* if it satisfies Axioms 1 and 2 and also the following axiom.

Axiom 3: *For all $a, b, c, d \in A$, if aRb and bRc , then either aRd or dRc .*

Exercise 3.18 Suppose f, δ, A , and R are as in Exercise 3.16. Show that R is a semiorder of A .

We state the following result of [Roberts, 1969a, 1971], as stated in [Fishburn, 1985], without proof.

Proposition 3.13 (Roberts) *A binary relation R on a finite set A is a semiorder on A if and only if R is transitive and $G(R)$ is a proper interval graph.*

Again with [Luce, 1956] as motivation, [Roberts, 1971] defines a graph G to be *representable by just noticeable differences* if, for each $v \in V(G)$, there exists a real number r_v contained in a closed interval J_v of the real line such that $uv \in E(G)$ if and only if $r_u \in J_v$ (or, equivalently, $r_v \in J_u$). (Compare this with the concept of “catch graphs” in section 7.2.) While we state the following result without proof, Exercise 3.19 will be a simpler special case.

Proposition 3.14 (Roberts) *A graph is representable by just noticeable differences if and only if it is a proper interval graph.*

Exercise 3.19 (Roberts) Show that a graph G is a proper interval graph if and only if, for each $v \in V(G)$, there exists a real number r_v and a closed *unit* interval J_v centered at r_v such that $uv \in E(G)$ if and only if $r_u \in J_v$.

More general applications to psychology involve general seriation problems. For instance in developmental psychology, [Coombs & Smith, 1973] studies whether psychological “traits” could correspond to chronological intervals—interval graph models would clearly be useful here. [Hubert, 1974] surveys the role of interval and proper interval graphs in seriation problems in psychology. [Troxell, 1995] also considers proper interval graphs.

3.4.3 Applications to Computing

Applications of interval graphs tend to be examples of general “seriation” problems, determining whether certain data or objects are compatible with arrangement in a linear pattern. Such examples are often scheduling problems, with the linear dimension corresponding to time. A common example involves a graph having university courses as vertices, with two vertices adjacent if and only if the courses overlap in time of day and so cannot be assigned a common room. Such a graph will be an interval graph, and finding the minimum number of rooms needed corresponds to finding the graph’s chromatic number, a problem that is *much* easier for interval graphs than in general. (Making a hard—in this case NP-complete—problem tractable is an important role of interval graphs in computing, much as we mentioned for chordal graphs at the end of subsection 2.4.2; [Olariu, Schwing, & Zhang, 1995] is an up-to-date discussion.) [Kendall, 1969] contains another well-known “seriation in time” problem to which interval graphs are applicable, in this case to archaeology.

There are many applications to computing in which the seriation is not with respect to time. [Golumbic, 1984] gives one interesting example, and there are others in section 8.4 of [Golumbic, 1980]. Our discussion is in terms of the widely studied topic of consecutive retrieval file organization. The original idea appeared in [Ghosh, 1972], with [Eswaran, 1975] linking it to interval graphs. [Ghosh, Kambayashi, & Lipski, 1983] is a collection of articles on this subject, with [Lipski, 1983] listing almost 200 references on consecutive retrieval and interval graphs.

Suppose \mathcal{R} is a set of *records* (files) and \mathcal{Q} is a set of *queries*, each linked to a particular set of relevant records so that each query $Q_i \in \mathcal{Q}$ can be identified with a subset of \mathcal{R} . Such \mathcal{R} and \mathcal{Q} are said to satisfy the *consecutive retrieval property* if the records relevant to each query can be stored consecutively in linear storage without repeating records.

Example 3.3 Suppose A, B, C, D, E, F, G, H, I are nine records with $Q_1 = \{A, B, C\}$, $Q_2 = \{D, E, F, G\}$, $Q_3 = \{C, D, H, I\}$, $Q_4 = \{D, I\}$, and $Q_5 =$

$\{C, D, E, H, I\}$. Then one way to satisfy the consecutive retrieval property is shown by the linear arrangement

$$A - B - C - H - I - D - E - F - G.$$

Clearly \mathcal{R} and \mathcal{Q} satisfy the consecutive retrieval property if and only if $H = (\mathcal{R}, \mathcal{Q})$ is an interval hypergraph. As we observed in section 3.2, this means that if the records of \mathcal{R} can be arranged so as to satisfy the consecutive retrieval property with respect to the queries in \mathcal{Q} , then the line graph $\Omega(\mathcal{Q})$ of H must be an interval graph. But, as we also observed in section 3.2, the converse fails. This is a subtle, but important, point that can cause confusion in various seriation applications. The subtlety is shown by its incorrect inclusion as a theorem in [Ghosh, 1977] and its removal from the second edition, [Ghosh, 1986]. That $\Omega(\mathcal{Q})$ is an interval graph corresponds to \mathcal{Q} being representable by intervals, not to the arrangement of the members of \mathcal{R} . The hypergraphs of Example 2.6 and Exercise 3.7 can both be thought of as examples of \mathcal{R} and \mathcal{Q} where $\Omega(\mathcal{Q})$ is an interval graph, yet the records cannot be linearly arranged so as to satisfy the consecutive retrieval property. [Deogun & Gopalakrishnan, to appear] proves the following.

Proposition 3.15 (Deogun & Gopalakrishnan) *Suppose the hypergraph $H = (\mathcal{R}, \mathcal{Q})$ is such that there are not two $Q_i, Q_j \in \mathcal{Q}$ with $Q_i \subseteq Q_j$. Then the records in \mathcal{R} can be arranged so as to satisfy the consecutive retrieval property with respect to \mathcal{Q} if and only if the line graph of H is isomorphic to the clique graph of the line graph of the dual hypergraph H^* and this clique graph is a proper interval graph.*

The consecutive retrieval property has been generalized in various ways, many of them in [Ghosh, Kambayashi, & Lipski, 1983]. In particular, [Tanaka, 1983] investigates replacing linear storage with storage on trees, thereby replacing interval hypergraphs with tree hypergraphs.