DIMACS Technical Report 98-12
May 1998

Tagged Probe Interval Graphs

by

Li Sheng[1]
Rutgers Center for Operations Research
Rutgers University
New Brunswick, New Jersey 08903

Chi Wang[2]
Department of Mathematics
University of Louisville
Louisville, KY 40290

Peisen Zhang
Department of Genetics and Development
Columbia University
New York, NY 10032

[1]Research supported by Rutgers University NSF grant with grant number SES-9211492.
[2]Research supported by University of Louisville IRIG grant.

DIMACS is a partnership of Rutgers University, Princeton University, AT&T Labs-Research, Bell Labs, Bellcore and NEC Research Institute.

DIMACS is an NSF Science and Technology Center, funded under contract STC–91–19999; and also receives support from the New Jersey Commission on Science and Technology.

ABSTRACT

A generalization of interval graph is introduced for cosmid contig mapping of DNA. A graph is a tagged probe interval graph if its vertex set can be partitioned into two subsets of probes and nonprobes, and a closed interval can be assigned to each vertex such that two vertices are adjacent if and only if at least one of them is a probe and one end of its corresponding interval is contained in the interval corresponding to the other vertex. We show that tagged probe interval graphs are weakly triangulated graphs, hence are perfect graphs. For a tagged probe interval graph with a given partition, we give a chordal completion that is consistent to any tagged interval completions with respect to the same vertex partition. Forbidden induced subgraph lists are given for trees with or without a given vertex partition. A heuristic that construct map candidates is given for cosmid contig mapping.

1 Cosmid Contig Mapping and Tagged Probe Interval Graphs

All graphs in this paper are simple graphs without loops or parallel edges. Suppose $\mathcal{F}$ is a family of subsets of a set S. A graph $G = (V, E)$ is the *intersection graph* of $\mathcal{F}$ if there is a bijection $\Phi : V \to \mathcal{F}$ such that for any $u, v \in V$, $uv \in E$ if and only if $\Phi(u) \cap \Phi(v) \neq \emptyset$. A graph $G = (V, E)$ is an *interval graph* if it is the intersection graphs of a family of intervals. Interval graphs have been used in physical mapping of DNA for reconstruction of relative positions of DNA fragments. In this paper, we introduce a generalization of interval graph for physical mapping of DNA. Structural properties related to algorithmic development are studied and used for constructing heuristic.

In molecular biology, to replicate and study a certain contiguous stretch of DNA, DNA is cut into small fragments. Those fragments are replicated into clones. The goal of physical mapping is to reconstruct the relative positions of clones along original DNA. There are various techniques to determine if two clones overlap. Most of them involve obtaining fingerprints for clones, and determine that two clones overlap if their fingerprints are sufficiently similar [2, 3, 6, 12]. Overlap relation of clones can be represented by an interval graph, where the vertices correspond to clones and two vertices are adjacent if their corresponding clones overlap. When complete overlap information is available and correct, there are efficient graph theory algorithms that produce a list of map candidates based on the interval graph model. On the other hand, reliable and complete overlap information is very costly and practically not available. Therefore, map construction could be more efficient if it can be done by using only a subset of overlap information.

Different from fingerprinting, cosmid contig mapping generates overlap information by hybridization [4, 15]. In this process, a set of clones is placed on a filter for colony hybridization. The filter is probed with clones that have been radioactively labeled at their ends. We detect that two clones overlap if one of them is labeled and at least one end of this labeled clone is contained in the other clone. Interval graph model is not applicable for cosmid contig mapping. First, if only a subset of clones is selected to be labeled, we do not have complete overlap information among clones. Second, if an unlabeled clone is completely contained inside a labeled clone, their overlap cannot be detected.

To model cosmid contig mapping and develop mathematical algorithms to construct maps, we introduce a generalization of interval graphs that capture the property of overlap information from cosmid contig mapping. Let $G = (V, E)$ be a graph whose vertex set is V and edge set is E . Let $V = \{P, N\}$ be a partition of V. We call G a *tagged probe interval graph with respect to* P if there is a set of closed intervals $I = \{I_v | v \in V\}$ such that vertices u and v are adjacent in G if and only if the following hold

(1) $I_u \cap I_v \neq \emptyset$;

(2) $\{u, v\} \cap P \neq \emptyset$;

(3) $\{l_u, r_u\} \cap I_v \neq \emptyset$ if $u \in P$ and $v \in N$, where $I_u = [l_u, r_u]$.

We call vertices in P *probes* and call vertices in N *nonprobes.* I is called a *tagged (probe interval) representation* of G with respect to the given partition. We call a graph $G = (V, E)$

a *tagged probe interval graph* if there is a partition $V = \{P, N\}$ such that G is a tagged probe interval graph with respect to P. When such a partition $V = \{P, N\}$ is given, we denote the graph by $G = (V, P, E)$ to emphasize the given partition. If I is a tagged representation of $G = (V, P, E)$, then the intersection graph of I is a *tagged interval completion of* G. We will denote $I_u = [l_u, r_u]$. We say interval I_v is *properly contained* in interval I_u if $I_v \subset I_u$ with $l_u \neq l_v$ and $r_u \neq r_v$.

In cosmid contig mapping, radioactively labeled clones correspond to vertices in P and the other clones correspond to vertices in $N = V - P$. The task in cosmid contig mapping can now be interpreted as to test if a graph with a given partition is a tagged probe interval graph and to construct a tagged representation for G when G is a tagged probe interval graph.

2 Preliminaries

Let $G = (V, E)$ be a graph. For a subset $S \subseteq V$, denote $G(S)$ the subgraph of G induced by S. Denote $N(x) = \{u \in V | ux \in E\}$ the *neighborhood* of $x \in V$ and denote $N[x] = N(x) \cup \{x\}$ the *closed neighborhood of* x. If $G = (V, E)$ is not an interval graph Booth76 $G' = (V, E \cup E')$ is an interval graph, then G' is an *interval completion* of G.

We first remark that it is possible that a graph can be a tagged probe interval graph for different partitions of probes and nonprobes. In practice, a partition, i.e., pre-selected probes and nonprobes, is often given by biologists. In the remainder of this paper, most of our results is given for graphs with given partitions.

Tagged probe interval graphs introduced here can be viewed as a refinement of probe interval graphs. [15] and [14]. A graph $G = (V, E)$ is a *probe interval graph* if the vertex set V can be partitioned into subsets P and N, and we can assign every vertex $v \in V$ an interval I_v such that $uv \in E$ if and only if $I_u \cap I_v \neq \emptyset$ and $\{u, v\} \cap P \neq \emptyset$, where I is called a *probe (interval) representation of* G with respect to P. Probe interval graphs were also introduced for the cosmid contig mapping. If we assume that probes are not only labeled on their two ends, Booth76 also on their interior points, then probe interval graphs can be used to represent overlap information obtained.

An interval graph is both a tagged probe interval graph and a probe interval graph with respect to $P = V$. A tagged probe interval graph or a probe interval graph may not be an interval graph. For example, it is well known that chordless cycle C_4 of length four is not an interval graph. However, C_4 is a tagged probe interval graph as well as a probe interval graph.

On the other hand, there is no containment relation between tagged probe interval graphs and probe interval graphs. There are graphs that are probe interval graphs Booth76 not tagged probe interval graphs with respect to any possible partition. There are also graphs that are tagged probe interval graphs Booth76 not probe interval graphs with respect to any possible partition.

The following lemma summarizes some observations which will be used in the remainder of this paper.

Lemma 2.1 *Suppose $G = (V, P, E)$ is a tagged probe interval graph and $I = \{I_v | v \in V\}$ is a representation of G with respect to the given partition.*

(1) $G(P)$ is an interval graph. In particular, if $P = V$ then G is an interval graph;

(2) $G(V - P)$ is an independent set;

(3) If $H \subseteq V$, then $G(H)$ is also a tagged probe interval graph;

(4) If $u \in P$, $v \in N$ and $uv \notin E$, then I_v not properly contained in I_u implies that $I_u \cap I_v = \emptyset$;

(5) If $|\{u, v\} \cap P| = 1$ and $uv \notin E$ imply $I_u \cap I_v = \emptyset$ for any $u, v \in V$, then G is a probe interval graph with respect to P and I is a probe representation of G with respect to P.

Proof. (1) - (4) follow directly from the definition. Condition (5) implies that $uv \in E$ if and only if at least one of u, v is a probe and $I_u \cap I_v \neq \emptyset$. Hence G is a probe interval graph and the lemma follows. □

Suppose $G = (V, P, E)$ is a tagged probe interval graph. We call two nonadjacent vertices v_1, v_2 an *enhanceable pair* if at least one of v_1, v_2 is in N and there are two nonadjacent vertices $u_1, u_2 \in P$ such that $u_i v_j \in E\,(i, j = 1, 2)$. The *enhancement* of G is obtained by adding an edge between every enhanceable pair. Denote the enhancement of G by $G^* = (V, P, E \cup E^*)$ where E^* contains all new edges between enhanceable pairs. Edges in E^* are called *enhanced edges*. It is easy to see that if $G = (V, P, E)$ is a tagged probe interval graph with $I = \{I_v | v \in V\}$ a tagged representation, and $G^* = (V, P, E \cup E^*)$ is the enhancement of G , then $uv \in E \cup E^*$ implies that $I_u \cap I_v \neq \emptyset$.

Theorem 2.2 *Suppose $G = (V, P, E)$ is a tagged probe interval graph. If every nonprobe of G is incident to an enhanced edge, then $G' = \left(V, P, E \cup E'\right)$ is a probe interval graph where E' contains all enhanced edges that has a probe as one end.*

Proof. Let $I = \{I_u | u \in V\}$ be a tagged representation of G and let $u \in N$ be a nonprobe. Since u is incident to an enhanced edge, there are $x, y \in P$ such that $xy \notin E, xu \in E, yu \in E$. Therefore $I_x \cap I_y = \emptyset$, $I_u \cap I_x \neq \emptyset$, and $I_u \cap I_y \neq \emptyset$. If a probe $w \in P$ is such that $I_u \cap I_w \neq \emptyset$ and I_u does not contain any end point of I_w, then I_u is properly contained in I_w and $uw \notin E$. It follows that $I_w \cap I_x \neq \emptyset, I_w \cap I_y \neq \emptyset$ and $uw \in E'$. Therefore for any $u, v \in V$, $uv \in E \cup E'$ if and only if $I_u \cap I_v \neq \emptyset$ and $\{u, v\} \cap P \neq \emptyset$. By definition, G' is a probe interval graph. □

To end this section, we present a general frame of classification of graphs that has tagged probe interval graphs and probe interval graphs as special cases. A graph $G = (V, E)$ is a *split graph* if there is a partition $V = \{V_1, V_2\}$ such that $G(V_1)$ is a clique and $G(V_2)$ is an independent set. We call a graph $G = (V, E)$ a *property-Q split graph* with respect to V' for some subset $V' \subseteq V$ if $G(V')$ is a graph of property Q and $G(V - V')$ is an independent set. Various classes of graphs in the literature can be considered as special cases of this generalization of split graphs. For example, *apex graphs*, i.e., planar graphs with an extra vertex are special cases of *planar split graphs* according to our definition here. Graphs that can be partitioned into a split graph and an independent set (or a clique) can be viewed as *split split graphs*, which are studied in [9]. In the same spirit, we call a graph $G = (V, E)$ an *interval split graph* if there is a partition of $V = \{V_1, V_2\}$ such that $G(V_1)$ is an interval graph and $G(V_2)$ is an independent set. Tagged probe interval graphs and probe interval graphs are special cases of interval split graphs. In the remainder of this paper, we call a graph $G = (V, P, E)$ an

interval split graph with respect to P when a partition $V = \{P, N\}$ is given and $G(P)$ is an interval graph and $G(N)$ is an independent set.

3 Perfectness of Tagged Probe Interval Graphs

In this section, we show that tagged probe interval graphs are perfect graphs. Given graph $G = (V, E)$, the *clique number* $\omega(G)$ is the number of vertices in a maximum complete subgraph of G. The *chromatic number* $\chi(G)$ *of* G is the minimum number of independent sets needed to cover all vertices of G. G is *perfect* if $\omega(H) = \chi(H)$ for all induced subgraphs H of G. Perfect graphs are important in combinatorial optimization and algorithmic theory. Interval graphs are well known perfect graphs. A graph is *weakly triangulated* if both G and G^c, the complement of G, do not have any induced subgraph isomorphic to a chordless cycle of at least five vertices. Hayword [8] showed that weakly triangulated graphs are prefect graphs. The following was shown in [11].

Theorem 3.1 *[11] Probe interval graphs are weakly triangulated.*

In the following, we show that, similar to probe interval graphs, tagged probe interval graphs are weakly triangulated graphs.

Lemma 3.2 *A chordless cycle C_k $(k \geq 5)$ is not a tagged probe interval graph.*

Proof. Let $C_k = (V, E)$ $(k \geq 5)$ be a chordless cycle. Suppose on the contrary that $\{I_v | v \in V\}$ is a tagged representation of C_k $(k \geq 5)$ with respect to some $P \subseteq V$. Then $V \neq P$ since C_k is not an interval graph. Suppose u, v are a pair of none adjacent vertices such that $u \in P$ and $v \in N$. By (2) of Lemma 2.1, the two vertices adjacent to v in C_k, say x and y, must be probes. Since $k \geq 5$ and $uv \notin E$, it follows that $\{x, y\} \nsubseteq N[u]$ and $u \neq x, y$. By (1) of Lemma 2.1, either $I_u \cap I_x = \emptyset$ or $I_u \cap I_y = \emptyset$. Therefore I_v cannot be properly contained in I_u. Then $uv \notin E$ implies that $I_u \cap I_v = \emptyset$. By (5) of Lemma 2.1 C_k would be a probe interval graph, contradicting Theorem 3.1. □

Lemma 3.3 *The complement of any chordless cycle C_k $(k \geq 5)$ is not a tagged probe interval graph.*

Proof. Let $C_k^c = (V, E)$ be the complement of C_k. Since $C_5^c = C_5$, C_5^c is not a tagged probe interval graph with respect to any $P \subseteq V$ by Lemma 3.2. For $k \geq 6$, assume on the contrary that $I = \{I_v | v \in V\}$ is a tagged representation of C_k^c with respect to some $P \subseteq V$. Since C_k^c has induced subgraphs isomorphic to C_4, by (1) of Lemma 2.1, $P \neq V$. Similar to the proof of Lemma 3.2, let u, v be a pair of nonadjacent vertices of C_k^c such that $u \in P$ and $v \in N$. In the representation I_v cannot be properly contained in I_u for otherwise any vertex adjacent to v is a probe and hence adjacent to u by (1) of Lemma 2.1. It follows from $uv \notin E$ that $I_u \cap I_v = \emptyset$. By (5) of Lemma 2.1, C_k^c $(k \geq 6)$ would be a probe interval graph, contradicting Theorem 3.1. □

Theorem 3.4 *Tagged probe interval graphs are weakly triangulated.*

Proof. By (3) of Lemma 2.1 an induced subgraph of a tagged probe interval graph is also a tagged probe interval graph. Therefore by Lemmas 3.2 and 3.3, a tagged probe interval graph cannot contain any induced subgraph isomorphic to C_k or C_k^c $(k \geq 5)$. □

Corollary 3.5 *Tagged probe interval graphs are perfect.*

4 Chordal Completion of Tagged Probe Interval Graphs

A graph is *chordal* or *triangulated* if it contains no induced subgraph isomorphic to C_k for any $k \geq 4$. If $G = (V, E)$ is not a chordal graph Booth76 $H = (V, E \cup E')$ is a chordal graph, then H is a *chordal completion* of G. If H has minimum number of edges among all chordal completion of G, then H is called a *minimum chordal completion* of G. In this section, for a tagged probe interval graph G with a given partition, we show that the enhancement of G is a chordal completion of G that is contained in any tagged interval completion of G with respect to the given partition.

Theorem 4.1 *Suppose that $G^* = (V, P, E \cup E^*)$ is the enhancement of a tagged probe interval graph $G = (V, P, E)$. Then G^* is a subgraph of any tagged interval completion of G with respect to P.*

Proof. Suppose that G_I is a tagged interval completion of G. Let $I = \{I_u | u \in V\}$ be an interval representation of G_I. Clearly every edge of G is an edge of G_I. We only need to show that every enhanced edge is also an edge of G_I. Suppose u_1, u_2 is an enhanceable pair of G. Then there are probes v_1, v_2 such that $v_1 v_2 \notin E$ and $u_i v_j \in E$ ($i = 1, 2$ and $j = 1, 2$). Hence $I_{v_1} \cap I_{v_2} = \emptyset$, $I_{u_i} \cap I_{v_j} \neq \emptyset$ ($i = 1, 2$ and $j = 1, 2$). It follows that $I_{u_1} \cap I_{u_2} \neq \emptyset$. Therefore $u_1 u_2$ is an edge of G_I. □

Lemma 4.2 *Let $G = (V, P, E)$ be a tagged probe interval graph with a tagged representation $I = \{I_v | v \in V\}$.*

(1) Suppose $u_1, u_2, w \in V$ are such that $I_{u_1} \subseteq I_w$ or $I_{u_2} \subseteq I_w$. If $u_1 u_2$ is an enhanced edge, then either $u_i w \in E$ or $u_i w$ is an enhanced edge for $i = 1, 2$.

(2) *Suppose* $I_{u_1},\ I_{u_2},\ I_p,\ I_{v_1}, I_{v_2} \in I$ *are such that* $\emptyset \neq I_{u_1} \cap I_{u_2} \subseteq I_{v_1} \cap I_{v_2}$. *Then* $I_p \cap\ I_{u_i} \neq \emptyset$ *for* $i = 1, 2$ *implies that* $I_p \cap\ I_{v_i} \neq \emptyset$ *for* $i = 1, 2$.

Proof. If u_1, u_2 is an enhanceable pair, there exist two nonadjacent probes p_1, p_2 such that $I_{u_i} \cap I_{p_j} \neq \emptyset\,(i = 1, 2$ and $j = 1, 2)$. If $I_{u_1} \subseteq I_w$ or $I_{u_2} \subseteq I_w$ then $I_w \cap I_{p_j} \neq \emptyset\,(j = 1, 2)$. Therefore (1) follows.

For (2), since intervals satisfy Helly property, $I_p \cap I_{v_1} \cap I_{v_2} \supseteq I_p \cap\ I_{u_1} \cap\ I_{u_2} \neq \emptyset$. Therefore, $I_p \cap\ I_{v_i} \neq \emptyset$ and the lemma follows. $\Box$

Theorem 4.3 *Let* $G = (V, P, E)$ *be a tagged probe interval graph and let* E^* *be the set of enhanced edges. Then* $G^* = (V, P, E \cup E^*)$ *is a chordal graph.*

Proof. Let $I = \{I_v | v \in V\}$ be a tagged representation of G. Let C be a cycle of G^* with vertices $v_1, \cdots, v_k$ $(k \geq 4)$ and edges $v_k v_1$ and $v_i v_{i+1}\,(i = 1, ..., k-1)$. Let intervals corresponding to v_i be $I_i = [l_i, r_i]$ $(i = 1, ..., k)$. Without loss of generality, assume that $r_k = max\,\{r_j | 1 \leq j \leq k\}$ and that $r_k \geq r_{k-1} \geq r_1$. Suppose on the contrary that C has no chords in G^*. Then $v_1 v_{k-1} \notin E$, $v_{k-2} v_k \notin E$ and $v_2 v_k \notin E$.

First consider the case that $I_{k-1} \subseteq I_k$. Note that $I_{k-2} \cap I_k \supseteq I_{k-2} \cap I_{k-1} \neq \emptyset$. Since $r_k \geq r_{k-2}$, we have $v_{k-2} \in N$, for otherwise $v_{k-2} v_k$ would be a chord by $I_{k-2} \cap I_k \neq \emptyset$. If $v_{k-1} \in N$, then $v_{k-1} v_{k-2} \in E^*$ implies $v_{k-2} v_k \in E^*$ by (1) of Lemma 4.2, a contradiction. Therefore $v_{k-1} \in P$.

If $I_{k-2} \subseteq I_k$, then $v_{k-3} \in P$ for otherwise we would have that $v_{k-3} v_{k-2} \in E^*$, and so $v_{k-2} v_k \in E^*$ by (1) of Lemma 4.2. Now since $I_{k-3} \cap I_k \neq \emptyset$ and $r_{k-3} \leq r_k, v_{k-3} v_k \in E$. So $k = 4$ and $v_{k-3} = v_1$. Hence, v_{k-2} and v_k are each adjacent to two nonadjacent probes v_1 and v_{k-1}, which implies that $v_{k-2} v_k \in E^*$. This contradicts our assumption on C, so $I_{k-2} \nsubseteq I_k$ and $l_k \in I_{k-2}$. Hence, $v_k \in N$ for otherwise we would have $v_{k-2} v_k \in E$ since $I_{k-2} \cap I_k \neq \emptyset$.

Suppose $I_1 \cap I_{k-1} \neq \emptyset$. Then $v_1 \in N$ and $I_1 \subset I_{k-1}$ for otherwise $v_1 v_{k-1} \in E$. Hence, $v_1 v_k \in E^*$, which implies that $v_1 v_{k-1} \in E^*$ by (1) of Lemma 4.2, a contradiction. Therefore $I_1 \cap I_{k-1} = \emptyset$, and so $r_1 < l_{k-1}$.

Note that since $I_{k-2} \nsubseteq I_k$, $I_{k-2} \cap I_{k-1} \neq \emptyset$ and $r_1 < l_{k-1}$, we have $r_1 \in I_{k-2}$ and $\emptyset \neq I_1 \cap I_k \subseteq I_{k-2} \cap I_k$. If $v_1 \in P$, then $v_{k-2} v_1 \in E$ since $r_1 \in I_{k-2}$, and so $k = 4$. Since v_{k-2} and v_k are each adjacent to two nonadjacent probes v_1 and v_{k-1} , so $v_{k-2} v_k \in E^*$, a contradiction. If $v_1 \in N$, then $v_1 v_k \in E^*$ would lead to $v_{k-2} v_k \in E^*$ by (2) of Lemma 4.2, again a contradiction.

Combining above we have that if $I_{k-1} \subseteq I_k$ then C has a chord.

Next consider the case that $I_{k-1} \nsubseteq I_k$. $I_{k-1} \cap I_1 \neq \emptyset$ since $r_1 \leq r_{k-1} \leq r_k$ and $I_1 \cap I_k \neq \emptyset$. If $I_1 \subseteq I_{k-1}$, then $v_1 \in N$ since $v_1 v_{k-1} \notin E$. It follows that $v_k \in P$ and $v_2 \in P$ for otherwise we would have $v_1 v_{k-1} \in E^*$ by (1) of Lemma 4.2. On the other hand, since $I_2 \cap I_{k-1} \neq \emptyset$, it follows from $I_k \cap I_{k-1} \neq \emptyset$ and v_2, v_k are two nonadjacent probes that $v_2 v_{k-1} \in E$. Thus v_1 and v_{k-1} are each adjacent to two nonadjacent probes v_2 and v_k, and so $v_1 v_{k-1} \in E^*$, a contradiction. Therefore, $I_1 \nsubseteq I_{k-1}$.

Similarly, $I_{k-1} \nsubseteq I_1$. Hence $v_1, v_{k-1} \in N$ and $I_{k-1} - I_k \subseteq I_1 - I_k$. Now $v_k \in P$ for otherwise since $\emptyset \neq I_1 \cap I_k \subseteq I_{k-1} \cap I_k$, $v_1 v_k \in E^*$ would imply that $v_1 v_{k-1} \in E^*$ by (2) of Lemma 4.2.

If $I_{k-2} \cap I_k \neq \emptyset$, then $v_{k-2} \in N$ and $I_{k-2} \subset I_k$ since $v_{k-2}v_k \notin E$. This is impossible since by (1) of Lemma 4.2, $v_{k-2}v_{k-1} \in E^*$ would imply $v_{k-2}v_k \in E^*$. Therefore $I_{k-2} \cap I_k = \emptyset$. It follows that $I_{k-2} \cap I_1 = I_{k-2} \cap (I_1 - I_k) \supseteq I_{k-2} \cap (I_{k-1} - I_k) = I_{k-2} \cap I_{k-1} \neq \emptyset$.

If $v_{k-2} \in P$, then since $I_1 \cap I_{k-2} \neq \emptyset$, it follows from $I_1 \cap I_k \neq \emptyset$ and v_{k-2}, v_k are two nonadjacent probes that $v_1 v_{k-2} \in E$. Thus v_1 and v_{k-1} are each adjacent to two nonadjacent probes v_{k-2} and v_k, and so $v_1 v_{k-1} \in E^*$, a contradiction. Hence $v_{k-2} \in N$. Then, since $I_{k-2} \cap I_1 \supseteq I_{k-2} \cap I_{k-1} \neq \emptyset$, $v_{k-2}v_{k-1} \in E^*$ would imply that $v_1 v_{k-1} \in E^*$ by (2) of Lemma 4.2, again a contradiction. Therefore C must have a chord in G^*. □

Notice that the enhancement G^* of a tagged probe interval graph G may not always be a minimum chordal completion of G. Furthermore, G^* may not be necessarily an interval graph yet.

5 Characterization of Cycle-Free Tagged Probe Interval Graphs

Since tagged probe interval graphs are closed for the operation of taking induced subgraphs, tagged probe interval graphs can be characterized by a list of forbidden induced subgraphs. Similarly, tagged probe interval graphs can also be characterized by a list of forbidden induced subgraphs where every graph in the list has prescribed partition of probes and nonprobes. In the following, we present such lists for trees.

Given a graph $G = (V, E)$, a subset $\{v_1, v_2, v_3\} \subseteq V$ is an *asteroidal triple* if there are paths P_{ij} from v_i to v_j such that $P_{ij} \cap N(v_k) = \emptyset$ for $i, j, k = 1, 2, 3$. The following is well-known:

Theorem 5.1 *[10] A chordal graph is an interval graph if and only if it has no asteroidal triple.*

Lemma 5.2 *Let $G = (V, P, E)$ be an interval split graph. If G contains an asteroid triple of three probes, then G is not a tagged probe interval graph with respect to P.*

Proof. Let x, y, z be an asteroid triple of G such that $\{x, y, z\} \subseteq P$. Let P_{xy} be a path between x and y such that $P_{xy} \cap N(z) = \emptyset$. Let P_{xz}, P_{yz} be similarly defined. Suppose on the contrary that $G = (V, P, E)$ is a tagged probe interval graph with a representation $I = \{I_v | v \in V\}$. Let $H = (V, P, E_H)$ be the intersection graph of I . Since H is an interval graph, $\{x, y, z\}$ is not an asteroidal triple of H. Without loss of generality, suppose $uz \in E_H$ for some vertices u on the path P_{xy} . Thus $u \in N$ and $I_u \subset I_z$ for otherwise $uz \in E$. Note that since x, y are probes, $u \neq x, y$, let w be a vertex on P_{xy} next to u. Then $w \in P$ and $I_w \cap I_u \neq \emptyset$. Hence $I_w \cap I_z \neq \emptyset$, and so $wz \in E$, which contradicts that z is not adjacent to any vertex on P_{xy} in G. □

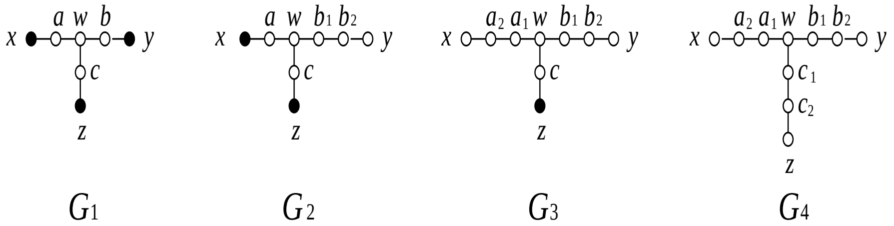

Figure 1: The forbidden subgraphs for trees to be TPIG

Lemma 5.3 *Let G_1, G_2, G_3 and G_4 be graphs in the Figure 1 with darkened vertices being probes and circled vertices being either probes or nonprobes. Then G_i is not a tagged probe interval graph for $i = 1, \ldots, 4$.*

Proof. Since the three probes x, y, z form an asteroid triple in G_1, by Lemma 5.2, G_1 is not a tagged probe interval graph.

If G_2 is a tagged probe interval graph, then b_2 must be a nonprobe, for otherwise, G_2 would have an induced subgraph isomorphic to G_1. Then y must be a probe and probes x, y, z form an asteroid triple of G_2, by Lemma 5.2, G_2 is not a tagged probe interval graph, a contradiction.

If G_3 is a tagged probe interval graph, we may assume that a_2, b_2 are nonprobes for otherwise G_3 would have an induced subgraph isomorphic to G_2. Then x, y are both probes and the three probes x, y, z form an asteroid triple in G_3, contradicting Lemma 5.2.

If G_4 is a tagged probe interval graph, we may assume that a_2, b_2, c_2 are all nonprobes for otherwise G_4 would have an induced subgraph isomorphic to G_3. Then x, y, z must be all probes and they form an asteroid triple in G_4, contradicting Lemma 5.2. $\square$

The *length* of a path is the number of edges on the path. The distance between two vertices u, v, denoted $d(u, v)$, is the length of any shortest path joining u, v. A pair (G, u), consisting of a graph G and a specified vertex u, is called a *rooted graph* with the *root* u. For a rooted tree (T, r), the root r is the *ancestor* of all the other vertices in T; and the *height* of (T, r), denoted by $h(T, r)$, is the largest distance between r and a vertex of T. Let $e = uv$ be an edge of T. We refer to u as the *parent* of v or v as a *child* of u if $d(r, u) < d(r, v)$. Vertices that have the same parent are called *siblings*.

Let $P = x_1 x_2 \ldots x_m$ be a path of m vertices in a tree $T = (V, E)$. Deleting edges of P from T, the remaining graph is partitioned into m rooted subtrees $(T_1, x_1), (T_2, x_2), \ldots, (T_m, x_m)$. Denote V_i the vertex set of (T_i, x_i) $(i = 1, 2, \ldots, m)$. Then $G(\bigcup_{t=i}^{j} V_i)$ is a tree for any $(1 \leq i < j \leq m)$.

Lemma 5.4 *Let $T = (V, E)$ be a tree and $P \subseteq V$, $N = V - P$. Suppose $P^* = x_1 x_2 \ldots x_{m+1}$ is a longest path of T and (T_i, x_i) $(i = 1, ..., m+1)$ are rooted trees with vertex set V_i defined above. If $T(N)$ is an independent set, and T has no induced subgraph isomorphic to graph G_4 in Figure 1, then*

(1) $h(T_i, x_i) \leq 2$ *for* $i = 1, ..., m+1$.

(2) *If* $\{x \in T_i | d(x, x_i) = 2\} \bigcap P = \emptyset$ *for* $i = 1, ..., m+1$, *then* $T = (V, P, E)$ *is a tagged probe interval graph with respect to* P.

Proof. (1) is trivial. To prove (2), we construct a tagged representation of T. First, let $I_i = [l_i, r_i]\,(i = 1, ..., m+1)$ be such that $l_i < r_{i-1} < l_{i+1} < r_i$. Then $I_i^* = (r_{i-1}, l_{i+1}) \subset I_i$ and $I_i^* \cap I_j = \emptyset$ for any $j \neq i$. By (1), x_i only has children or grandchildren in (T_i, x_i) for every i. For each nonprobe child of x_i, let $[l_i, d_i]$ be its corresponding interval where $l_i < d_i < r_{i-1}$. We can assign a set of disjoint subintervals of I_i^* corresponding to probe children of x_i. We assign grandchildren of x_i the same interval as its parent. Then it is easy to check that this is a tagged representation of T. □

Theorem 5.5 *Let* $T = (V, P, E)$ *be a tree and* $P \subseteq V$, $N = V - P$. *Then* T *is a tagged probe interval graph with respect to* P *if and only if* $T(N)$ *is an independent set and* T *has no induced subgraph isomorphic to graph* G_1, G_2, G_3 *and* G_4 *in Figure 1.*

Proof. The necessity follows from Lemma 5.3. To prove the sufficiency, by Lemma 5.4, we only need to show that T has a longest path P^* satisfying the condition in part (2) of Lemma 5.4. It is easy to see that if T has no induced subgraph isomorphic to G_1, G_2 or G_3, then such a longest path P^* exits in T. □

Theorem 5.6 *Let* $T = (V, E)$ *be a tree. Then* T *is a tagged probe interval graph if and only if* T *has no induced subgraph isomorphic to graph* G_4 *in Figure 1.*

Proof. The necessity follows from Lemma 5.3. To prove the sufficiency, assume now that tree T has no induced subgraph isomorphic to graph G_4 in Figure 1. Let $P^* = x_1 x_2 \ldots x_{m+1}$ be a longest path in T. Let T_i be defined the same as before. If $m < 4$ then T is an interval graph. So T is a tagged probe interval graph for $P = V$. Assume now that $m \geq 4$. Since T has no induced subgraph isomorphic to G_4 in Figure 1, by Lemma 5.4, $h(T_i, x_i) \leq 2$. For $i = 1, ..., m+1$, vertex x_i has only children or grandchildren in the rooted tree (T_i, x_i). For $i = 1, ..., m+1$, let all the grandchildren, if any, of x_i be nonprobes, and let the rest vertices of T_i be probes. Then $T(N)$ is an independent set. By Lemma 5.4, $T = (V, P, E)$ is a tagged probe interval graph with respect to the choosing partition of probes and nonprobes. □

6 A Heuristic for Tagged Probe Interval Graphs

In this section we give a heuristic that constructs map candidates based on tagged probe interval graphs for cosmid contig mapping. Algorithmic terminologies not defined here are referred to [7, 1].

One observation from cosmid contig mapping is that proper containment between DNA fragments is rare. When one clone is a subfragment of another, often they share one end. This fact suggests that tagged probe interval graphs from cosmid contig mapping often have representations of intervals in which if two intervals I_u and I_v are such that $I_u \subseteq I_v$ then

I_u and I_v share at least one end. It is known that intersection graphs of such families are precisely the so-called *proper interval graphs.* Roberts [13] showed that a graph is a proper interval graph if and only if there is a *compatible order* $\prec$ on its vertex set such that if $x \prec y \prec z \prec w$ and $xw \in E$, then $yz \in E$. The following heuristic is based on this characterization and Theorem 4.1.

Heuristic: Input a tagged probe interval graph $G = (V, P, E)$ that is obtained from a cosmid contig mapping. Output a linear order of vertices of G.

1. Construct the enhancement G^* of G.
2. Let x be an arbitrary probe in G^*. Use breadth-first search and probe first tie-breaking rule to construct a spanning tree T rooted at x such that S contains only edges that are visited from a labeled vertex to an unlabeled vertex during the search.
3. Let y be of G^* that has the largest distance from x in S. Use breadth-first search and probe first tie-breaking rule again to construct another spanning tree T rooted at y.
4. Find a longest path l of T. Label vertices of l increasingly while traverse on l from one end to the another end. Let V_T be the vertex set of l.
5. If $V_T = V$, then stop. Otherwise, assign a *neighbor* u of V_T, i.e., $u \notin V_T$ and $ux \in E$ for some $x \in V_T$, a label that is the average of labels of its neighbors in V_T and break tie arbitrarily. Add u to V_T. Repeat.

References

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

[2] A.V. Carrano, P.J. de Jong, E. Branscomb, T. Slezak and B.W. Watkins. Constructing chromosome- and region-specific cosmid maps of the human genome, *Genome* 31 (1989), 1059-1065.

[3] A. Coulson, J. Sulston, S. Brenner and J. Karn. Toward a physical map of the genome of the nematode, *Caenorhabditis Elegans. Proc. Natl. Acad. Sci. USA* 83 (1987), 7821-7825.

[4] S.G. Fischer, E. Cayanis, J. Russo, I. Sunjevaric, B. Boukhgalter, X.L. Li, P. Zhang, P. Rothstein, M.T. Yu, D. Warburton, I.S. Edelman and A. Efstratiadis. Assembly of ordered contigs from YAC-selected cosmids of human chromosome 13, *Genomics* 21 (1994), 525-537.

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

[6] R.M. Gemmill, J.F. Coyle-Morris, F.D. McPeek Jr., L.F. Wara-Uribe and F. Hecht. Construction of long-range restriction maps in human DNA using pulsed field gel electrophoresis, *Gene Anal. Technol.* 4 (1987), 119-131.

[7] M.C. Golumbic. *Algorithmic Graph Theory and Perfect Graphs.* Academic Press, New York, 1980.

[8] R.B. Hayward. Weakly triangulated graphs, *J. Comb. Theory (B)* 39 (1985), 200—209.

[9] A.E. Kézdy, H.S. Snevily and C. Wang. Partitioning permutations into increasing and decreasing subsequences, *J. Comb. Theory (A)* 73(1996), 353—359.

[10] Lekkerkerker and Boland. Representation of a finite graph by a set of intervals on the real line. *Fund. Math.* 51(1962), 45-64.

[11] F.R. McMorris, C. Wang and P. Zhang. On probe interval graphs, *manuscript*, 1996.

[12] M.V. Olson, E. Dutchik, M.Y. Graham, G.M. Brodeur, C. Helms, M. Frank, M. MacCollin, R. Acheinman and T. Frand. Random-clone strategy for genomic restriction mapping in yeast, *Proc. Natl. Acad. Sci. USA* 83 (1986), 7826-7830.

[13] F.S. Roberts. Indifference graphs, *Proof Techniques in Graph Theory,* (F. Harary ed.), 139-146, Academic Press, New York, 1969.

[14] P. Zhang. Probe interval graphs and its applications to physical mapping of DNA, *manuscript*, 1994.

[15] P. Zhang, E.A. Schon, S.F. Fischer, E. Cayanis, J. Weiss, S. Kistler and P.E. Bourne. An algorithm based on graph theory for the assembly of contigs in physical mapping of DNA, *CABIOS* 10 (1994), 309-317.