Probe Interval Graph and Its Applications to Physical Mapping of DNA

Peisen Zhang
pz6@columbia.edu
Columbia Genome Center, Columbia University Russ Berrie medical science Pavilion
New York, NY 10032, USA

Keywords: interval graph, triangulated graph, physical mapping

1 Introduction

A new class of graph called the *probe interval graph* has been introduced. It is an extension of the interval graph. It requires only partial overlap information. An *enhanced probe interval graph* has been derived from the probe interval graph. Its fundamental properties (such as that the enhanced probe interval graph is triangulated) are investigated. The structure of the probe interval graph is characterized. A special kind of probe interval graph, the so-called STS-probe interval graph, is presented. The probe interval graph has been used in the human genome project (Zhang *et al.* [1994]). The probe interval graph is a new class of graphs. Only very fundamental properties have been studied in this work. Some studies about this graph have been published (McMorris *et al.* [1999]).

Definition 1 *Let V is a finite set of intervals of a real line. Then $G(V, E)$ denotes the graph whose vertex set is V and whose edge set is E. $(v_1, v_2) \in E$, if and only if $v_1 \cap v_2 \neq \emptyset$. The graph $G(V, E)$ is an interval graph (see Golumbic 1980).*

Definition 2 *Let V be a vertex set of undirected graph, and P be a subset of V. Then $G(V, P, E)$ denotes a probe graph if its edge set $E \subset (P \times V)$. The elements of P are called probes.*

Definition 3 *Let V be a finite set of intervals of a real line, and P be a subset of V. Then a probe graph $G(V, P, E)$ whose vertex set is V, whose probe set is P, and whose edge set is E, will be called a probe interval graph, if its vertex set is $E = \{(v_i, v_j) \mid v_i \cap v_j \neq \emptyset$ and $v_i \in P$ or $v_j \in P\}$ (Zhang et al. 1994). Note that $G(V, V, E) = G(V, E)$, hence the probe interval graph is an extension of the interval graph.*

Definition 4 *Let $G(V, P, E)$ be a probe graph, if two non-overlapping probes v_1 and v_2 overlap two non-probe vertices t_1 and t_2, then a new edge (t_1, t_2) can be introduced as an enhanced edge. The probe graph $G(V, P, E)$ with all enhanced edges will be called the enhanced probe graph of $G(V, P, E)$. It is easy to see that for a probe interval graph $G(V, P, E)$, the enhanced edges represent the overlaps among non-probe intervals. The probe interval graph $G(V, P, E)$ with all enhanced edges will be called the enhanced probe interval graph of $G(V, P, E)$*

An undirected graph G is called triangulated if every cycle of length greater than 3 possesses a chord, that is, an edge joining two nonconsecutive vertices of the cycle. The interval graph is triangulated (see Golumbic 1980).

Theorem 1 *The Enhanced Probe Interval Graph is triangulated.*

Definition 5 *Given a probe interval graph $G(V, P, E)$, if $E \subset (P \times V \backslash P)$, we will call G an STS-probe interval graph.*

2 The Structure of the probe interval graph

Given an enhanced probe graph $G(V, P, E)$, we define the following:

Quasi-clique QC is a set of vertices including probes and non-probes. In QC, there is a non-empty subset $C \supset (\mathrm{QC} \cap P)$ that is a clique and every non-probe is adjacent to every probe in QC. The maximal cliques in QC will be called the ***cores*** of QC.

Quasi-maximal clique QM is a quasi-clique and at least one of its cores is a maximal clique in G.

Complete set of quasi-maximal cliques is the set of quasi-maximal cliques such that every maximal clique of G is a subset of one and only one quasi-maximal clique of that set.

Theorem 2 *A probe graph $G(V, P, E)$ where $P \subset V$ and $E \subset (V \times P)$ is a probe interval graph if and only if there is a linearly ordered complete set of quasi-maximal cliques for its enhanced graph of $G(V, P, E)$ such that, for each vertex v (probe or non-probe), the quasi-maximal cliques containing v occur consecutively.*

The STS-probe interval graph is a special class of probe interval graph. Its probes do not overlap. The STS-probe interval graph has a simple structure. Every quasi-maximal clique only includes one probe. We can deal with an STS-probe interval graph as an interval graph. We can use the MPQ-tree algorithm to get an MPQ-tree representation with all probes as leaves in a linear time (Korte and Möring 1989).

3 Discussion

In comparison with the interval graph, the probe interval graph is a more powerful and more flexible tool for the assembly of contigs in physical mapping of DNA. Any number of probes can be used to generate a probe interval graph. But in order to generate an interval graph, every member in the library of DNA fragments should be used as a probe. Further, the probe interval graph can be used in *real time*. After each set of hybridization data is generated, a probe interval graph will be formulated according to the accumulated data and a corresponding map will be generated. The map will provide useful feedback information to the next step. But for the interval graph, a map can not be obtained until all data have been collected.

References

[1] Benzer, S., On the topology of the genetic fine structure, *Proc. Nat. Acad. Sci. USA*, 45:1607–1620, 1959.

[2] Golumbic, M. C., Algorithmic Graph Theory and the Perfect Graph *Academic Press*, Orlando, Fla., 1980.

[3] Golumbic, M. C. and Shamir R., Complexity and algorithms for reasoning about time: a graph-theoretic approach, *JACM*, 40:1108–1133, 1993.

[4] Korte, N., and Möring, R. H., An incremental linear time algorithm for recognizing interval graph, *SIAM J. Comput.*, 18:68–81, 1989.

[5] McMorris, F. R., Wang, C., and Zhang, P., On probe interval graphs, *Discrete Applied Math*, 89:287–295, 1999.

[6] Watson, J. D., Gilman, Michael, Witkowski, Jan, and Zoller Mark, Recombinant DNA Second Edition, *W. H. Freeman and Company*, New York, 1992.

[7] Zhang, P., Schon, Eric A., Fischer, S. G., Cayanis, E., Weiss, J., Kistler S., and Bourne, P. E., An algorithm based on graph theory for the assembly of contigs in physical mapping of DNA, *CABIOS*, 10:309–317, 1994.

[8] Zhang, P., Ye, X., Liao, l., Russo, J., and Fischer, S. G., Integrated mapping package – a physical mapping software tool Kit, *Genomics*, 55:78–87, 1999.