Manuscript Number:

Title: Algorithms for induced biclique optimization problems

Article Type: Regular Submission

Keywords: induced biclique; induced balanced biclique; polygon-circle graph; interval-filament graph; Protein-Protein-Interaction

Corresponding Author: Dr. Fanica Gavril, PH.D.

Corresponding Author’s Institution: Technion - Israel Institute of Technology

First Author: Fanica Gavril, PH.D.

Order of Authors: Fanica Gavril, PH.D.

Abstract: We present polynomial time algorithms for induced biclique optimization problems in the following families of graphs: polygon-circle graphs, 4-hole-free graphs, complements of interval-filament graphs and complements of subtree-filament graphs. Such problems are to find maximum: induced bicliques, induced balanced bicliques and induced edge bicliques. These problems have applications for biclique clustering of proteins by PPI criteria, of documents, and of web pages.
Algorithms for induced biclique optimization problems

Fanica Gavril

Computer Science Dept., Technion, Haifa 32000, Israel, gavril@cs.technion.ac.il

ABSTRACT: We present polynomial time algorithms for induced biclique optimization problems in the following families of graphs: polygon-circle graphs, 4-hole-free graphs, complements of interval-filament graphs and complements of subtree-filament graphs. Such problems are to find maximum: induced bicliques, induced balanced bicliques and induced edge bicliques. These problems have applications for biclique clustering of proteins by PPI criteria, of documents, and of web pages.

KEY WORDS: induced biclique, induced balanced biclique, polygon-circle graph, interval-filament graph, Protein-Protein-Interaction

1. Introduction

We consider only finite graphs $G(V,E)$ with no parallel edges and no self-loops, where $V$ is the set of vertices and $E$ the set of edges. For $U \subseteq V$, $G(U)$ is the subgraph induced by $U$. We denote $N(v)=\{u | u$ adjacent to $v\}$ and $N[v]=N(v) \cup \{v\}$. A graph is bipartite if its vertex set has a partition $U,W$ such that its edges are only between vertices of $U$ and vertices of $W$; when every two vertices $u \in U, v \in W$ are adjacent, the graph is complete bipartite. The complement of a complete bipartite graph is a biclique; a biclique is composed by two cliques with no interconnecting edges.

Two sets intersect if they have a non-empty intersection and they overlap if they intersect and none is contained into the other. A graph $G$ is an intersection graph (overlap graph) of a family $S$ of subsets of a set if there is a one-to-one correspondence between the vertices of $G$ and the subsets in $S$ such that two vertices are adjacent iff their corresponding subsets in $S$ intersect (overlap, respectively) [20]. Intersection graphs of intervals on a line, subtrees on a tree and arcs on a circle are called interval, chordal [7] and circular-arc graphs [20], respectively. Overlap graphs of intervals on a line and subtrees on a tree [6,4] are called circle graphs and subtree overlap graphs, respectively. A transitively orientable graph is called a comparability graph [20]. Polygon-circle graphs [13] are intersection graphs of families of convex polygons inscribed in a circle.
Gavril [8,9] introduced the interval-filament graphs which are defined as follows: In a Euclidean plane $PL$, consider a line $L$ defined by $y=0$, drawn from left to right. On $L$, consider a family of closed intervals $I$. For an interval $[l,r] \in I$, we define an interval-curve $c$ in $PL$ as a continuous function $c:[l,r] \rightarrow \mathbb{R}^+$ having $c(l)=c(r)=0$. Consider a family $ff$ of interval-curves fulfilling that $\cup_{c \in ff} c$ is a continuous curve: $a=\cup_{c \in ff} c$ is called an interval-filament. Two interval-filaments $a_1$, $a_2$ intersect if $a_1 \cap a_2 \neq \emptyset$. An interval-filament is delimited in $PL$ by its two extreme endpoints in $L$, hence, if two intervals are disjoint, their interval-filaments do not intersect. Clearly, the union of two intersecting interval-filaments is an interval-filament. The intersection graph of a family of interval-filaments is an interval-filament graph. For a vertex $v$, we denote by $i(v)$ the interval on $L$ delimiting the interval-filament corresponding to $v$. Gavril [8,9,10] similarly defined the subtree-filament graphs proved later to be the subtree overlap graphs [4].

The family of subtree-filament graphs contains the family of interval-filament graphs, which in turn contains the polygon-circle graphs and the cocomparability graphs. The family of polygon-circle graphs contains the circular-arc graphs, the circle graphs and the chordal graphs. The families of interval and subtree-filament graphs have polynomial time algorithms for many problems: for example maximum weight cliques, maximum independent sets [8], maximum weight induced matchings [2] and maximum induced split subgraphs [10]. On the other hand, Pergel [16] proved that the recognition problem for polygon-circle graphs and interval-filament graphs is NP-complete.

In the present paper we consider the existence of polytime algorithms for the following vertex induced subgraph problems:

**Biclique optimization problems:** Given a graph $G(V,E)$, find a vertex induced biclique $B(C_1,C_2)$ which maximizes a given function $f(|C_1|,|C_2|)$. For example, for maximum biclique maximize $|C_1|+|C_2|$, for maximum balanced biclique maximize $k=|C_1|=|C_2|$ and for maximum edge biclique maximize $|C_1| \cdot |C_2|$. Additional functions to maximize are $\min(|C_1|,|C_2|)$, and $\min(|C_1|,|C_2|)$ when $|C_1|+|C_2|$ is maximum. The algorithms allow also the enumeration of the solutions to the problems. All these problems are NP-complete for general graphs and all except the maximum biclique problem are NP-complete for cocomparability graphs [11,17], and thus for cocomparability graphs.

**Complete bipartite optimization problems** for a graph $G(V,E)$ are exactly the biclique optimization problems for the complement of $G$; these problems are trivial for planar graphs and 4-hole-free graphs since they can contain only induced complete bipartite subgraphs of limited size.
In the present paper, we present polytime algorithms to solve biclique optimization problems in the following families of graphs: polygon-circle graphs, 4-hole-free graphs, complements of interval-filament graphs, complements of subtree-filament graphs, and some subfamilies of interval-filament and subtree-filament graphs.

Yannakakis [22] gave a simple polytime algorithm for the maximum biclique problem in cobipartite graphs: consider its bipartite complement, interchange edges and missing edges, and in the new graph find a maximum independent set by the usual max-flow polytime algorithm. This algorithm can be extended to the family, discussed in [19], of hereditary graphs with a polynomial number of maximal cliques: to find a maximum induced biclique, apply the algorithm in [22] to the cobipartite subgraphs induced by every two maximal cliques. This family contains the families of 4-hole-free graphs, circular-arc graphs, chordal graphs, planar graphs and bounded boxicity graphs.

The biclique optimization problems have applications for pairs of entities related by interaction or containment and represented as bipartite graphs. For example, the two sets of entities can be documents and words, images and features, genes and species (taxa), diseases and genes, proteins and motifs, and proteins and domains. The problems have applications also when there is a single set of entities related by some property, to be clustered in bicliques by some strongly connected vs. non-connected or similarity vs. dissimilarity criteria. The most important application is in Protein-Protein-Interaction (PPI) problems for co-clustering proteins using the lock-and-key criteria [15], the complementary domains criteria [21], the domain-domain interaction criteria [12] and interacting motifs criteria [14]. But it has applications in many other domains, such as clustering documents with overlap content and web pages with overlap content [18].

2. Algorithms for bicliques

In this Section we describe algorithms for biclique optimization problems in families of graphs $G(V,E)$ which have a polytime algorithm for maximum clique, and which have a polynomial family $F$ of pairs of subgraphs $<G(U),G(W)>$, with no interconnecting edges, such that every biclique $B(C1,C2)$ fulfils $C1\subseteq U$ and $C2\subseteq W$ for some pair $<G(U),G(W)> \in F$. The algorithm works as follows:

Separation Algorithm: We consider every pair of subgraphs $<G(U),G(W)>$ in $F$, and find maximum cliques in $G(U)$, $G(W)$; these cliques define a biclique, since there are no edges between $U$ and $W$. Also, every biclique can be obtained in this way. The solutions
to biclique optimization problems are within these pairs of maximum cliques. For maximum biclique we take the pair $C_1, C_2$ which maximizes $|C_1|+|C_2|$ and by taking the one that also has maximum $\min(|C_1|,|C_2|)$ we obtain a solution to $\min(|C_1|,|C_2|)$ when $|C_1|+|C_2|$ is maximum. For maximum edge biclique we take the pair $C_1, C_2$ which maximizes $|C_1||C_2|$. For maximizing $\min(|C_1|,|C_2|)$ we take the pair $C_1, C_2$ which maximizes $\min(|C_1|,|C_2|)$ and for maximum balanced biclique we reduce the greater clique to the size $k=\min(|C_1|,|C_2|)$. The enumeration of the solutions is obtained by enumerating all maximal cliques in $G(U), G(W)$. □

The Separation Algorithm works not only for bicliques, but for any pair of induced structures with no interconnecting edges, like: two connected subgraphs, two subtrees, two split graphs, two holes, a clique and a hole, etc..

We apply now the Separation Algorithm to various families of graphs.

**Polygon-circle graphs.** Let $G(V,E)$ be a polygon-circle graph, with a representation by convex polygons in a circle $CR$. Consider two disjoint vertex subsets $U,W$, such that the subgraphs $G(U), G(W)$ are connected. As proved by Cameron [2], the union of the polygons corresponding to the vertices in $U$ and $W$, are contained in two convex polygons $P_U, P_W$, defined by their points on $CR$. Cameron [2] also proved that $P_U, P_W$ intersect iff there are two intersecting polygons corresponding to vertices $u \in U, v \in W$. This implies that if there are no adjacent vertices $u \in U, v \in W$, then $P_U, P_W$ are disjoint and there is a chord in $CR$ separating them.

In the representation of $G$ by polygons on a circle $CR$, we consider every two arcs with no polygon points, on $CR$, a chord connecting two points $X,Y$ in the arcs, and the two vertex sets $U_{XY}, W_{XY}$ corresponding to the polygons on either side of the chord $XY$. Let $F$ be the family of these pairs $<G(U_{XY}), G(W_{XY})>$. In the polygon-circle subgraphs $G(U_{XY}), G(W_{XY})$, we find maximum cliques, by the algorithm in [8]. Then, to solve biclique optimization problems we apply the Separation Algorithm to the family $F$.

**Extensions of polygon-circle graphs.** We can also solve biclique optimization problems in the family of graphs $G(V,E)$ fulfilling that for every two non-adjacent vertices $u,v$ the subgraph $G(N[u,v],E)$ is a polygon-circle graph: Every biclique of $G$ contains two non-adjacent vertices $u,v$, and is contained in $G(N[u,v],E)$. Therefore, for every pair of non-adjacent vertices $u,v$, we apply the Separation Algorithm to the polygon-circle graph $G(N[u,v],E)$.  

Consider a family of graphs $G(V,E)$ obtained from a family of graphs $G(V,E \cup E2)$, with a polynomial number of maximal cliques, by deleting the $E2$ edges, and fulfilling that for every two maximal cliques $V_X, V_Y$ of $G(V,E \cup E2)$ the subgraph $G(V_X \cup V_Y, E)$ is a polygon-circle graph. Every clique $C$ of $G(V,E)$ is a clique of $G(V,E \cup E2)$ and is contained in a maximal clique $V_X$ of $G(V,E \cup E2)$. Hence $C$ is contained in the subgraph $G(V_X,E)$. Therefore, we can solve biclique optimization problems by considering every two maximal cliques $V_X, V_Y$ of $G(V,E \cup E2)$ and applying the Separation Algorithm to the polygon-circle graphs $G(V_X \cup V_Y, E)$. Such families are, for example, the interval-filament and the subtree-filament graphs [8] fulfilling also that for every two points $X,Y$ on $L$ or $T$, the subgraph $G(V_X \cup V_Y, E)$ is a polygon-circle graph, where $V_X=\{v \mid X \in i(v)\}$, $V_Y=\{v \mid Y \in i(v)\}$.

Unfortunately, the recognition problem for polygon-circle graphs is NP-complete [16]. In addition:

**Lemma 1.** The recognition problem for polygon-circle graphs is reducible to recognizing whether for two non-adjacent vertices $u,v$ the subgraph $G(N[u,v], E)$ is a polygon-circle graph; therefore this problem is NP-complete also.

**Proof:** To an input for the first problem we add a new vertex $x$ adjacent only to $u$ and a new vertex $y$ adjacent to all other vertices of $G$. A polygon-circle representation for the second problem gives a polygon-circle representation to the first, and a polygon-circle representation to the first problem gives a polygon-circle representation to the second, by representing $x$ as a small arc containing only a point of $u$ on $CR$, and representing $y$ by a disjoint arc covering the rest of $CR$. □

But, the recognition problem for families of graphs with more restricted conditions is polynomial; for example when we request the subgraphs $G(N[u,v], E)$ and $G(V_X \cup V_Y, E)$ to be circular-arc graphs or circle graphs, which have polytime recognition algorithms. Such a subfamily of polygon-circle graphs is characterized in the following Theorem:

**Theorem 2.** Consider a polygon-circle graph $G(V,E)$ fulfilling that for every vertex $v$, the subgraph $G(V-N[v], E)$ is connected. Then $G(V,E)$ is a circular-arc graph.

**Proof.** Let $G(V,E)$ be a polygon-circle graph fulfilling that every $G(V-N[v], E)$ is connected. Consider a vertex $v$ and its polygon $P_v$. Since $G(V-N[v], E)$ is connected, the points on $CR$ of the union of the polygons $\cup_{u \in V-N[v]} P_u$ defines [2] a convex polygon $P_{V-N[v]}$ disjoint from $P_v$. For every vertex $v$, we delete from its polygon $P_v$ the chord facing $P_{V-N[v]}$ to obtain a polygonal line in $CR$. This does not change the intersection relationship since the intersection of two convex polygons in $CR$ involves the intersection of two pairs of chords,
one pair from each polygon. By the above process, two polygonal lines intersect iff their arcs on CR intersect. Therefore \( G(V,E) \) is a circular-arc graph. \( \square \)

Eschen and Spinrad [5] described a linear time recognition algorithm for circular-arc graphs. Therefore the polygon-circle graphs \( G(V,E) \) fulfilling that every subgraph \( G(V-N[v],E) \) is connected have a polytime algorithm for recognition. For example:

**Corollary 3.** The cobipartite polygon-circle graphs are circular-arc graphs, since every \( G(V-N[v],E) \) is connected.

**4-hole-free graphs.** Let us consider the well-known family of 4-hole-free graphs.

**Lemma 4.** For every two cliques \( C1,C2 \) in a 4-hole-free graph \( G(V,E) \), the subgraph \( G(C1 \cup C2,E) \) is an interval graph.

**Proof.** The subgraph \( G(C1 \cup C2,E) \) being cobipartite, has no holes with 5 or more vertices; \( G(C1 \cup C2,E) \) also is 4-hole-free and thus it is chordal. A cobipartite graph is a cocomparability graph. Therefore \( G(C1 \cup C2,E) \) being chordal and cocomparability, it is an interval graph [20]. \( \square \)

Since the 4-hole-free graphs have a polynomial number of maximal cliques, and the interval graphs are polygon-circle graphs, we can apply the Separation Algorithm to every \( G(C1 \cup C2,E) \), \( C1,C2 \) maximal cliques, to solve biclique optimization problems.

The more restricted family of even-hole-free graphs has been widely discussed in the literature. In [1] it is proved that they have an elimination scheme by cobipartite graphs and in [3] it is proved that they have an elimination scheme by chordal graphs. But by Lemma 4, a cobipartite graph with no 4-holes is an interval graph. Therefore

**Corollary 5.** An even-hole-free graph has an elimination scheme by interval graphs with minimum covering by cliques at most 2.

**A subfamily of interval-filament graphs.** Consider an interval-filament graph \( G(V,E) \): for a vertex \( v \) let \( a(v) \) denote its interval-filament and for a clique \( C \) let

\[
a(C) = \cup_{c \in C} a(v);
\]

clearly, there are two vertices \( x_C, y_C \) in \( C \) fulfilling that \( i(C) \subseteq i(x_C) \cup i(y_C) \).

We consider the subfamily of interval-filament graphs in which for every maximal clique \( C \) the vertices \( x_C, y_C \) fulfill that there are no edges between a vertex \( u \) having \( i(u) \subseteq (i(x_C) \cup i(y_C)) \) and a vertex \( v \) having \( i(x_C) \cup i(y_C) \subseteq (i(v)). \) For every two adjacent vertices \( x,y \), let \( U_{x,y} = \{ u \mid i(u) \subseteq (i(x) \cup i(y)) \} \), \( W_{x,y} = \{ w \mid i(x) \cup i(y) \subseteq (i(w)) \) or \( i(w) \cap (i(x) \cup i(y)) = \phi \}. \)

Let \( F \) be the family of these pairs \( \langle G(U_{x,y}), G(W_{x,y}) \rangle \). In the interval-filament subgraphs \( G(U_{x,y}), G(W_{x,y}) \), we can find maximum cliques, by the algorithm in [8]. Then, to solve biclique optimization problems we apply the Separation Algorithm to the family \( F \).
A subfamily of subtree-filament graphs. Consider a subtree-filament graph $G(V,E)$; as known [4], $G(V,E)$ has a representation as an overlap graph of subtrees $t(v)$ in a tree $T$. For every clique $C$ of $G$, denote the subtree $\cup_{v \in C} t(v)$ by $t(C)$. Since $G(V,E)$ is an overlap graph, there are no edges between a vertex $u$ having $t(u) \subseteq t(C)$ and a vertex $v$ having $t(C) \cap t(v) = \emptyset$. We consider the subfamily of subtree-filament graphs for which there exists a constant $k$ such that every maximal clique $C$ has a subset $A$ of at most $k$ vertices fulfilling $t(C) = \cup_{v \in A} t(v) = t(A)$. For every clique $A$ with at most $k$ vertices let $U_A = \{ u \mid t(u) \subseteq t(A) \}$ and $W_A = \{ w \mid t(A) \cap t(w) = \emptyset \}$. Let $F$ be the family of these pairs $<G(U_A), G(W_A)>$. In the subtree-filament subgraphs $G(U_A), G(W_A)$, we can find maximum cliques, by the algorithms in [8,4]. Then, to solve biclique optimization problems we apply the Separation Algorithm to the family $F$.

3. Algorithms for complete bipartite subgraphs

In this Section we describe optimization algorithms for complete bipartite subgraphs in interval-filament and subtree-filament graphs. Let $G(V,E)$ be an interval-filament graph with a representation on a line $L$. Consider two non-adjacent vertices $u,v$ having $i(u) \subset i(v)$ and let $U_{u,v} = \{ w \mid i(u) \subset i(w) \subset i(v), \ w \notin N(u) \cup N(v) \} \cup \{ u,v \}$.

Lemma 6. Every vertex $z \in N(u) \cap N(v)$ is adjacent to every vertex $w \in U_{u,v}$.

Proof. Assume that there are two non-adjacent vertices $z \in N(u) \cap N(v)$ and $w \in U_{u,v}$. If $i(w) \subset i(z)$ then $i(u) \subset i(w) \subset i(z)$ and if $i(z) \subset i(w)$ then $i(u) \subset i(z) \subset i(w)$. Also, if $i(w) \cap i(z) = \emptyset$, then $i(u) \cap i(z) = \emptyset$, since $i(u) \subset i(w)$. All three cases contradict the fact that $z$ is adjacent to both $u$ and $v$ but not to $w$. □

Let $B(IND1, IND2, E)$ be a complete bipartite subgraph of $G$. Assume that every two vertices $w,z \in IND1$ have $i(w) \subset i(z)$ and let $u,v$ be the vertices in $IND1$ with minimal and maximal intervals. Then, by Lemma 6, $IND1 \subseteq U_{u,v}$ and $IND2 \subseteq N(u) \cap N(v)$.

Assume now that $IND1$ has two vertices $w,z$ having $i(w) \cap i(z) = \emptyset$. Then, for every $u \in IND2$, $i(u)$ contains the interval between $i(w)$ and $i(z)$. Hence, every two vertices $u,v \in IND2$, being non-adjacent, fulfill $i(u) \subset i(v)$. Therefore, if $IND2$ has only one vertex $u$, then $IND1 \subseteq N(u)$, and if $IND2$ has more than one vertex, let $u,v$ be its vertices with minimal and maximal intervals. Then, by Lemma 6, $IND2 \subseteq U_{u,v}$ and $IND1 \subseteq N(u) \cap N(v)$.

We obtain every complete bipartite subgraph of $G$ by considering every vertex $w$ and maximum independent set $IND1$ in $G(N(w),E)$, and by considering every two non-
adjacent vertices \( u, v \) having \( i(u) \supseteq i(v) \) and finding maximum independent sets \( IND_1 \) in \( G(U_{u,v}, E) \) and \( IND_2 \) in \( G(N(u) \cap N(v), E) \), by the algorithm in [8]. Among all such pairs, we chose the solutions for the complete bipartite subgraphs optimization problems. Note that this is the Separation Algorithm on the complement graphs.

Lemma 6 and the above algorithms remain true for the subtree-filament graphs with "interval on a line" replaced by "subtree on a tree" and \( U_{u,v} = \{ w \mid t(u) \supseteq t(w) \supseteq t(v), \ w \notin N(u) \cup N(v) \cup \{u,v\} \} \). For two vertices \( w, z \in IND_1 \) with disjoint subtrees, the subtrees of the vertices in \( IND_2 \) are contained one into another since they contain the unique path between \( t(w) \) and \( t(z) \). Thus, we can solve by the above algorithms the complete bipartite subgraphs optimization problems in subtree-filament graphs.

References


