

On the proper interval completion problem within some chordal subclasses

François Dross^a, Claire Hilaire^b, Ivo Koch^c, Valeria Leoni^d, Nina Pardal^{e,f,*},
María Inés Lopez Pujato^d, Vinicius Fernandes dos Santos^g

^a Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France

^b FAMNIT, University of Primorska, Koper, Slovenia

^c University of General Sarmiento, Argentina

^d National University of Rosario and CONICET, Argentina

^e ICC-CONICET, Argentina

^f University of Sheffield, UK

^g Computer Science Department, Federal University of Minas Gerais, Brazil

ARTICLE INFO

Article history:

Received 22 October 2021

Received in revised form 17 September 2024

Accepted 24 September 2024

Available online 7 October 2024

Keywords:

Proper interval completion

Split graph

Threshold graph

Quasi-threshold graph

Caterpillar

ABSTRACT

Given a property (graph class) Π , a graph G , and an integer k , the Π -completion problem consists of deciding whether we can turn G into a graph with the property Π by adding at most k edges to G . The Π -completion problem is known to be NP-hard for general graphs when Π is the property of being a proper interval graph (PIG). In this work, we study the PIG-completion problem within different subclasses of chordal graphs. We show that the problem remains NP-complete even when restricted to split graphs. We then turn our attention to positive results and present polynomial time algorithms to solve the PIG-completion problem when the input is restricted to caterpillar and threshold graphs. We also present an efficient algorithm for the minimum co-bipartite-completion for quasi-threshold graphs, which provides a lower bound for the PIG-completion problem within this graph class.

© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Graph modification problems can be used to address many fundamental problems, not only in graph theory itself, but also to model a large number of practical applications in several different fields. Some of those fields include molecular biology, computational algebra, and more generally, areas that involve modelling based on graphs where the missing edges are due to a lack of data, for example in data clustering problems [8,14]. In many of these applications, an edge modification of the graph that models the experimental data corresponds to correcting errors and inconsistencies in the data.

Given a graph G and a graph property Π , a *graph modification problem* consists of studying how to add or delete the minimum number of vertices or edges from G in order to obtain a graph that satisfies the property Π . In this article, the property Π will represent a graph class, such that the graph resulting from the modification belongs to this class. We focus our attention on one of the four basic graph modification problems: the Π -completion problem.

* Corresponding author.

E-mail addresses: n.pardal@conicet.gov.ar, n.pardal@sheffield.ac.uk (N. Pardal).

Given a property Π , a Π -completion of a graph $G = (V, E)$ is a supergraph $H = (V, E \cup F)$ such that H belongs to Π and $E \cap F = \emptyset$. The edges in F are referred to as *fill edges*. A Π -completion $H = (V, E \cup F)$ of G is *minimum* if, for any set of fill edges F' such that $H' = (V, E \cup F')$ belongs to Π , it holds that $|F'| \geq |F|$. In this case, $|F|$ is called the Π -completion number of G . The *minimum Π -completion problem* consists in finding the Π -completion number of a graph G . The associated decision problem—the Π -completion problem—consists in deciding, for a given integer k , if G has a Π -completion with at most k fill edges. Throughout this work, a Π -completion will always be a minimum one unless otherwise stated.

The Π -completion problem from an arbitrary graph is known to be NP-complete when Π is the class of chordal, interval, or proper interval graphs [7,8,12,20]. Furthermore, when Π is the class of interval graphs, it was shown that the problem remains NP-complete on line graphs [7], and also on co-bipartite graphs [21]. According to Peng et al. [19], the problem can be solved in $O(n)$ -time for trees, following from the results in [6]. In the case of chordal-completion and proper interval-completion (PIG-completion from now on), its study from the viewpoint of parameterized complexity was initiated by Kaplan et al. in 1999 [11]. In 2015, Bliznets et al. presented the first subexponential parameterized algorithm for PIG-completion that finds a solution in $k^{O(k^{2/3})} + O(nm(kn + m))$ -time [1].

A celebrated result by Courcelle et al. states that each graph property that is expressible in $MSOL_1$ (resp. $MSOL_2$) can be solved in polynomial time for graphs with bounded cliquewidth (resp. treewidth) [4]. Note that this result is mainly of theoretical interest and does not lead to practical algorithms. Since the problem of finding a PIG-completion with at most k edges can be expressed in $MSOL_2$ for fixed k , this motivates our search of efficient algorithms for subclasses of chordal graphs with bounded treewidth.

Another direction on which current research on this topic is focused is in finding and characterizing minimal completions for input graph classes for which the minimum version is hard in the most efficient possible way from a computational point of view [5,10,15].

Throughout this work, we consider the target class Π to be the subclass of interval graphs given by proper interval graphs. The most well known motivation for the PIG-completion problem comes from molecular biology. In [18], Benzer first gave strong evidence that the collection of DNA composing a bacterial gene was linear. This linear structure could be represented as overlapping intervals on the real line, and therefore as an interval graph. In order to study various properties of a certain DNA sequence, the original piece of DNA is fragmented into smaller pieces which are then cloned many times. When all the clones have the same size, the resulting graph should not only be interval, but proper interval. Deciding whether two clones should overlap or not is the critical part. However, there might be some false positive or false negatives due to erroneous interpretation of some data. Thus, correcting the model to get rid of inconsistencies is equivalent to removing or adding as few edges as possible to the graph so that it becomes interval.

This work is organized as follows. We start by proving in Section 3 that the PIG-completion problem remains hard even if the input graph G is split. Since split graphs are in particular chordal, this result implies that the problem remains hard when G is a chordal graph, which leads us to study proper subclasses of chordal graphs where the PIG-completion problem might be tractable. More precisely, in Section 4 we first give an efficient algorithm for the PIG-completion problem on threshold graphs. We finish the section by showing an efficient dynamic programming algorithm for the co-bipartite-completion problem on quasi-threshold graphs following a brief discussion on the difficulties of generalizing the previous result to this superclass of threshold graphs. In Section 5, we show an efficient algorithm for a very sparse class of graphs, a subclass of both interval graphs and trees called caterpillars. We conclude the paper with some final remarks and possible future directions in Section 6.

2. Definitions

We give in this section the basic definitions and fix the notation that will be used throughout this work. All graphs in this paper are undirected and simple. Let G be a graph, and let $V(G)$ and $E(G)$ denote its vertex and edge sets, respectively. We denote by n the number of vertices and by m , the number of edges. Whenever it is clear from the context, we simply write V and E and denote $G = (V, E)$. For basic definitions not included here, we refer the reader to [3].

Given a graph G and $S \subseteq V$, the *subgraph of G induced by S* , denoted by $G[S]$, is the graph with vertex set S and such that two vertices of S are adjacent if and only if they are adjacent in G . When G' and $G[S]$ are isomorphic for some $S \subseteq V$, with a slight abuse of terminology we simply say that G' is an *induced subgraph* of G . For any family $\mathcal{F}$ of graphs, we say that G is $\mathcal{F}$ -free if G does not contain any graph $F \in \mathcal{F}$ as an induced subgraph. If a graph G is $\mathcal{F}$ -free, then the graphs in $\mathcal{F}$ are called the *forbidden induced subgraphs* of G .

A *clique* in a graph G is a complete induced subgraph of G . Also, we will often use this term for the vertex set that induces the clique.

A graph is *chordal* if it does not contain an induced cycle of size 4 or more.

A graph G is an *interval graph* if it admits an intersection model consisting of intervals on the real line, that is, a family $\mathbf{I}$ of intervals on the real line and a mapping from the set of vertices of G to the intervals of $\mathbf{I}$ such that two vertices are adjacent in G if and only if the corresponding intervals intersect. Notice that the class of interval graphs is a subclass of the class of chordal graphs.

A *proper interval graph* is an interval graph that admits a *proper interval model*, that is, an intersection model in which no interval is properly contained in any other.

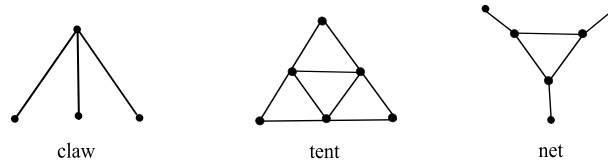


Fig. 1. Some of the forbidden induced subgraphs for proper interval graphs.

A *unit interval graph* is an interval graph that has an interval representation in which each interval has unit length. Every proper interval graph is a unit interval graph, and vice versa [17].

We can always assume that all the endpoints of the intervals in a (unit) interval model are pairwise distinct (so in particular that no interval is reduced to a single point, no two intervals are identical, and no two intervals intersect in exactly one point) [9].

Three nonadjacent vertices of a graph form an *AT (asteroidal triple)* if every two of them are connected by a path avoiding the neighbourhood of the third. Interval graphs are precisely those chordal graphs that are also AT-free [13] and proper interval graphs are precisely those chordal graphs that are also {claw, tent, net}-free [16,17] (see Fig. 1).

A graph $G = (C \cup I, E)$ is a *split graph* if its vertex set can be partitioned into a set C of pairwise adjacent vertices and set I of pairwise nonadjacent vertices.

A *threshold graph* is a split graph in which any two nonadjacent vertices satisfy that the neighbourhood of one is contained in the neighbourhood of the other. Equivalently, G is a threshold graph if it can be constructed from the empty graph by repeatedly adding either an *isolated vertex* (nonadjacent to every other vertex) or a *dominating vertex* (adjacent to every other vertex). Let the ordering of $V(G)$ according to this construction procedure be the *threshold ordering*. Threshold graphs are characterised precisely as the $\{2K_2, C_4, P_4\}$ -free graphs.

Quasi-threshold graphs, also called *trivially perfect graphs*, are the $\{P_4, C_4\}$ -free graphs. A connected quasi-threshold graph $G = (V, E_G)$ admits a rooted tree $T = (V, E_T)$ on the same vertex set V , rooted on a vertex r , such that $uv \in E_G$ if and only if there is a path in T starting in r containing both u and v .

A graph G is a *caterpillar* if G is a tree in which the removal of all the pendant vertices (i.e., the *leaves*) results in a path (i.e. the *spine* or *central path*).

A graph is *perfect* if, for every induced subgraph, the size of a largest independent set equals the smallest number of cliques needed to cover the subgraph.

A family $\mathcal{S}$ of nonempty sets has the *Helly property* if every subfamily of $\mathcal{S}$ of pairwise intersecting sets has a nonempty intersection. This property is also known as the *2-Helly property*. For example, any family of pairwise-intersecting intervals in the real line has the 2-Helly property.

3. PIG-completion within split graphs

We now devote our attention to the complexity of the PIG-completion problem when the input graph belongs to the class of split graphs. We start by citing some useful results and stating a few lemmas that characterise PIG-completions when the split partition fulfils certain properties. These results will be useful in Section 3.1, where we give a reduction to the PIG-completion problem.

Recall the following result:

Theorem 3.1 (Peng et al. [19]). *The threshold-completion problem is NP-complete on split graphs.*

Let us consider a connected split graph $G = (C \cup I, E)$, such that C is maximum in the sense that there is no vertex in I adjacent to every vertex of C , and let $H = (C \cup I, E \cup F)$ be a PIG-completion of G .

Lemma 3.2. *There exist both a partition $\{C', I'\}$ of the vertex set of H , where $I' \subseteq I$ and C' is a clique with $C' \supseteq C$, and a partition $\{I_l, I_r\}$ of I' such that I_l and I_r are both cliques in H .*

Proof. Let $\{I_u\}_{u \in V(H)}$ be a unit interval model for the graph H , which we know it exists since H is proper interval. It follows from the 2-Helly property of intervals and the fact that C is a clique in H that there exists a real point p in $\bigcap_{u \in C} I_u$. For each vertex $v \in I$, if either the right endpoint of I_v is to the left of $p - 1$, or the left endpoint of I_v is to the right of $p + 1$, then I_v does not intersect any interval in $\{I_u : u \in C\}$. We may conclude that v is pairwise nonadjacent to every vertex of C in H and thus v is an isolated vertex in G , which leads to a contradiction given that G is connected. Let us now consider $I_l = \{u \in I : p - 1 \in I_u\}$ and $I_r = \{u \in I : p + 1 \in I_u\}$, which are both cliques in H and let $I' = I_l \cup I_r$. Notice that $\{I_l, I_r\}$ is indeed a partition of I' since no unit interval can contain the points $p - 1$ and $p + 1$ at the same time. Moreover, if we define $C' = C \cup I \setminus (I_l \cup I_r)$, then it is clear from our previous remark that, for every vertex $u \in I \setminus (I_l \cup I_r)$, the unit interval I_u is fully contained in the interval $[p - 1, p + 1]$ and thus it contains the point p . Furthermore, p is also contained in the interval corresponding to every vertex in C , and therefore C' is a clique in H as well. ■

Lemma 3.3. Let $\{C', I'\}$ be a partition of the vertex set of H and $\{I_l, I_r\}$ be a partition of I' into cliques as given by the previous lemma. Then, $H_l = (C' \cup I_l, E_l)$ and $H_r = (C' \cup I_r, E_r)$ are threshold graphs, where $E_* = E(H[C' \cup I_*]) \setminus F_{I'}$ for $* \in \{l, r\}$ and $F_{I'} = \{uv \in F : u, v \in I'\}$.

Proof. If H_l is not threshold, then there are vertices $v, w \in I_l$ (nonadjacent in H_l) and $x, y \in C'$ such that $vx, wy \in E_l$ and $vy, wx \notin E_l$. Notice that v and w are adjacent in H since they both lie in I_l , which is a clique in H . Therefore, we find an induced C_4 in H , which results in a contradiction since H is a proper interval graph. ■

Finally, we will need the following property of proper interval graphs. Let G be an interval graph, and let $\mathbf{I}_G$ be an interval model of G . We say that a clique C of G is a *first clique* of G in $\mathbf{I}_G$ if there is a real point p such that p intersects every interval corresponding to the vertices of C , and every other interval of $\mathbf{I}_G$ lies strictly on the same side of p . Notice that such a clique always exists, since it can be found by restricting to the rightmost interval of the model.

Lemma 3.4. Let $G = (V, E)$ and $G' = (V', E')$ be two vertex-disjoint proper interval graphs, with respective proper interval models $\mathbf{I}_G$ and $\mathbf{I}_{G'}$, and let C (resp. C') be a first clique of G in $\mathbf{I}_G$ (resp. G' in $\mathbf{I}_{G'}$). Then, the graph $G'' = (V \cup V', E \cup E' \cup F)$ is also a proper interval graph, where F consists of all the possible edges between C and C' .

Proof. Let $\mathbf{I}_G$ be a proper interval model of G . Recall that we can assume that the endpoints of intervals in any (unit) interval model are distinct. Up to inverting right and left in $\mathbf{I}_G$, we can assume that C is on the right side of the model. Thus, G admits a proper interval model $\mathbf{I}_G = \{[a_u, b_u]\}_{u \in V} = \{[a_u, b_u]\}_{u \in V}$, with $a_u < b_u$, and there is a real point p intersecting every interval of C such that all the other intervals lie strictly to the left of p .

Let $k \geq 1$ be the size of C , let n be the size of V with $n \geq k$, and let $V = \{u_1, \dots, u_n\}$ be an ordering of the vertices such that $C = \{u_1, \dots, u_k\}$ and $b_{u_{k+1}} < \dots < b_{u_n} < p \leq b_{u_1} < \dots < b_{u_k}$. For each $i \in \{1, \dots, n\}$, let $c_{u_i} = p + 1 + \frac{i}{k}$ if $i \leq k$, and $c_{u_i} = b_{u_i}$ for the remaining vertices. Observe that $p < p + 1 < c_{u_1} < \dots < c_{u_k} = p + 2$. Moreover, for each $i \in \{1, \dots, n\}$, it follows that $a_{u_i} < p$, and also $c_{u_i} < p$ for every $i > k$.

Consider now the interval model given by $\{[a_u, c_u]\}_{u \in V}$. Note that, since we have not modified the ordering of the endpoints of the intervals, this is also a proper interval representation of G where every interval that does not represent a vertex of C lies strictly on the left of p , and all the remaining intervals are to the left of $p + 2$.

Consider now the proper interval model of G' given by $\mathbf{I}_{G'} = \{[a_v, b_v]\}_{v \in V'} = \{[a_v, b_v]\}_{v \in V'}$, and let p' be a real point intersecting C' such that every other interval lies strictly to the right of p' . Upon shifting all those intervals, we can assume that $p' = p + 2$.

Let $k' \geq 1$ be the size of C' , let n' be the size of V' with $n' \geq k'$, and let $V' = \{v_1, \dots, v_{n'}\}$ be an ordering of the vertices such that $C' = \{v_1, \dots, v_{k'}\}$ and $a_{v_{k'+1}} < \dots < a_{v_1} \leq p' < a_{v_{k'+1}} < \dots < a_{v_{n'}}$. For each $i \in \{1, \dots, n'\}$, let $c_{v_i} = p + 1 - \frac{i}{k'}$ if $i \leq k'$, and let $c_v = a_v$ for every other vertex. Observe that $p = c_{v_{k'}} < \dots < c_{v_1} < p + 1 < p + 2$. Furthermore, for each $i \in \{1, \dots, n'\}$ it follows that $b_{v_i} > p + 2$, and also $c_{v_i} > p + 2$ for every $i > k'$.

Consider the interval model given by $\{[c_v, b_v]\}_{v \in V'}$. Once more, since we have not modified the ordering of the endpoints of the intervals, this yields a proper interval representation of G' such that all the intervals that do not represent a vertex of C' lie strictly to the right of $p + 2$, and all the intervals of C' lie to the right of p .

Finally, consider the set of intervals $\{[a_u, c_u]\}_{u \in V} \cup \{[c_v, b_v]\}_{v \in V'}$ and let us see that no interval is contained in another. Suppose that there are two intervals $[a, b]$ and $[a', b']$, representing u and u' respectively, such that $a < a' < b' < b$. By construction, one of them lies in V and the other in V' . Suppose that $u \in V$ and $u' \in V'$. Hence, $b \leq p + 2$ and $p + 2 < b'$, which contradicts the fact that $b' < b$. We reach an analogous contradiction if $u' \in V$ and $u \in V'$. Therefore, $\mathbf{I} = \{[a_u, c_u]\}_{u \in V} \cup \{[c_v, b_v]\}_{v \in V'}$ is a proper interval model of some graph. Let $G'' = (V'', E'')$ be the proper interval graph corresponding to $\mathbf{I}$. By construction, it follows that $V \cup V' = V''$ and $E \cup E' \subseteq E''$. For every $u \in C$ and $v \in C'$, their corresponding intervals intersect in $p + 1$, hence G'' contains all the edges between C and C' . Let F be that set of edges. Let $u \in V \setminus C$, let $v \in V'$, and let $[a, b], [a', b']$ be their respective intervals. By construction, $a < b < p \leq a' < b'$ and thus u and v must be nonadjacent. Similarly, if $u \in V$ and $v \in V' \setminus C$, then $a < b \leq p + 2 < a' < b'$. Therefore, the only edges between V and V' are precisely those in F . ■

3.1. NP-completeness

We are now ready to prove that obtaining a PIG-completion is still NP-complete when the input graph is split. In order to do this, we strongly rely on the previous lemmas and Theorem 3.1.

Theorem 3.5. The PIG-completion problem is NP-complete on split graphs.

Proof. Given a completion of a split graph, it can be checked in polynomial time if this is in fact a PIG-completion, hence the problem is in NP.

We give a reduction from threshold-completion on split graphs (see Fig. 2 for a schema of the gadget used for the reduction). Let (G, ℓ) be an instance of threshold-completion on split graphs, where $G = (C \cup I, E)$ is a (connected, for simplicity) split graph on n vertices.

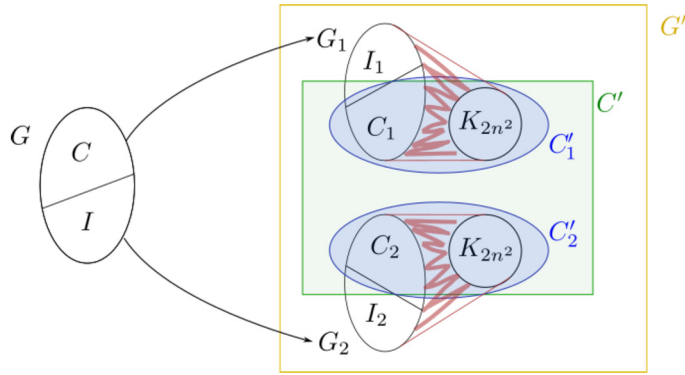


Fig. 2. A schema of the gadget used for the reduction of Theorem 3.5.

Consider the graph G' defined as follows. Let $G_1 = (C_1 \cup I_1, E_1)$ and $G_2 = (C_2 \cup I_2, E_2)$ be two copies of G . For each $i \in \{1, 2\}$, we consider $G'_i = (C'_i \cup I_i, E'_i)$, the graph constructed from G_i by connecting $2n^2$ new vertices to all the vertices of G_i , where C'_i is the set formed by the union of C_i and the newly added vertices. Notice that G'_i is a split graph. We denote by V_i the vertex set of G_i and by V'_i the vertex set of G'_i , for each $i \in \{1, 2\}$. Finally, connect all the vertices of C'_1 and C'_2 into a clique C' . Let $G' = (V', E')$ be the resulting split graph on $2(2n^2 + n)$ vertices, where $V' = C' \cup I_1 \cup I_2$.

We show that G can be augmented to a threshold graph with at most ℓ fill edges if and only if G' can be augmented to a proper interval graph with at most $2k$ additional edges, where $k = \ell + \binom{|I|}{2}$.

First, suppose there is a minimum PIG-completion $H = (V', E' \cup F)$ of G' with $|F| \leq 2k$ edges. Notice that we may assume that $|F| = 2k$, since we can add additional edges if necessary and still keeping the property of being a PIG. One way to show this (assuming that all the endpoints in the interval model are distinct), is by shifting the leftmost interval to the right until it intersects a new interval. We will show that G can be augmented to a threshold graph with ℓ edges.

Observe that completing each G_i into a clique requires less than n^2 edges. Thus, a PIG-completion of G' would need less than $2n^2$ fill edges, which implies that $|F| < 2n^2$.

Let $\mathbf{I}_H = \{\mathbf{I}_u\}_{u \in V(H)}$ be a proper interval model for the graph H . By Lemmas 3.2 and 3.3, there is a partition C'', I_l, I_r of V' where $C' \subseteq C''$, $I' := I_l \cup I_r \subseteq I_1 \cup I_2$, and such that $H[I_l]$ and $H[I_r]$ are both cliques with fill edges in $F_{I'}$ (as defined in Lemma 3.3).

If there is a vertex $v \in C'' \setminus C'$, w.l.o.g. $v \in I_1$, then we need at least $2n^2$ fill edges to connect v to C'_2 in H , which contradicts $|F| < 2n^2$, thus $C'' = C'$.

Since C' induces a clique in H , it follows from the 2-Helly property of intervals that there is a real point $p \in \bigcap_{u \in C'} \mathbf{I}_u$. No other interval of the model $\mathbf{I}_H$ intersects the point p , otherwise this requires at least $2n^2$ fill edges, and thus each interval corresponding to the vertices in $I_1 \cup I_2$ lies either strictly to the left or to the right of p .

Claim 3.6. *There is no $v_1 \in I_1$ and $v_2 \in I_2$ whose corresponding intervals lie either both to the right of p or both to the left of p .*

Proof. Let $\mathbf{I}_1$ and $\mathbf{I}_2$ be the intervals corresponding to v_1 and v_2 , respectively. Suppose that $\mathbf{I}_1$ and $\mathbf{I}_2$ both lie to the left of p . Since $\{v_1\} \cup C'_1 \setminus C_1$ induces a clique in H , there is a real point p_1 intersecting all the intervals corresponding to this set, and the same holds for $\{v_2\} \cup C'_2 \setminus C_2$ and p_2 . Suppose w.l.o.g. that $p_1 \leq p_2 < p$. Then, all the intervals of $C'_1 \setminus C_1$ intersect p_2 , thus v_2 is adjacent to all the vertices of $C'_1 \setminus C_1$ introducing thus $2n^2$ fill edges, which results in a contradiction. ■

Therefore, we assume without loss of generality that, in the interval representation of H , the endpoints of the intervals corresponding to vertices of I_1 lie strictly to the left of p and the ones corresponding to vertices of I_2 lie strictly to the right of p . Since $\{I_l, I_r\}$ is a partition of $I_1 \cup I_2$ where each set induces a clique in H , then $I_l = I_1$ and $I_r = I_2$. Let us show that all the fill edges lie inside each G_i .

Claim 3.7. *For each $i \in \{1, 2\}$, let F_i be the set of fill edges inside $H[V_i]$. Then $F = F_1 \cup F_2$.*

Proof. Recall that for each $i \in \{1, 2\}$, all the possible edges between V_i and $V'_i \setminus V_i$ are already in E' , thus F_i is also the set of fill edges in $H[V'_i]$.

Since we already proved that there is no fill edge between I_1 and I_2 , it suffices to see that there is no fill edge between I_1 and C'_2 (resp. I_2 and C'_1). Toward a contradiction, suppose there is at least one of said fill edges. Let us construct a PIG H' on the same vertex set with edge set $E' \cup F_1 \cup F_2$.

The model given by $\{\mathbf{I}_u\}_{u \in V'_1}$ is a proper interval model of $H[V'_1]$. It follows from the above reasoning that C'_1 can be seen as a first clique in this model, since the point p lies exactly in the intervals that correspond to vertices of C' . Similarly, $\{\mathbf{I}_u\}_{u \in V'_2}$ is a proper interval model of $H[V'_2]$ and C'_2 is a first clique in this model.

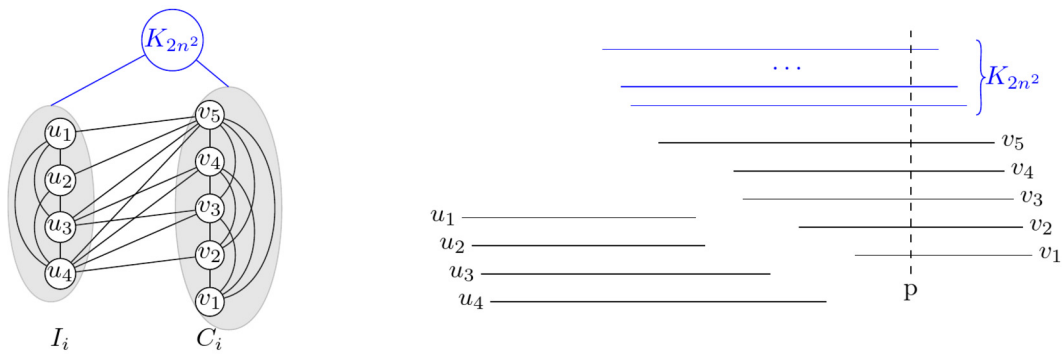


Fig. 3. Example of H_i'' and its proper interval model with $n = 9$.

It follows from Lemma 3.4 that there is a PIG $H' = (V'_1 \cup V'_2, E'_1 \cup F_1 \cup E'_2 \cup F_2 \cup F')$, where F' is the set of edges connecting all the vertices of C'_1 and C'_2 . Notice that $E'_1 \cup E'_2 \cup F' = E'$, hence H' is a PIG-completion of G' with edge set $E' \cup F_1 \cup F_2$. Thus, H' is a smaller PIG-completion than H , which contradicts the fact that H has minimum number of edges. ■

Assume without loss of generality that $|F_1| \leq |F_2|$. We will build a proper interval completion of G' with $2|F_1|$ fill edges, which will then prove that $|F_1| = |F_2|$ since H is a minimum PIG-completion of G' . To do this, consider any proper interval model of H , and keep only those images of vertices belonging to $I_1 \cup C'_1$. Now cut the intervals right after point p , keeping arbitrarily small parts to the right of p so that the intervals remain non-nested. By doing this we obtain a proper interval model of $H[I_1 \cup C'_1]$ where C'_1 is a first clique. Therefore, by considering twice the same graph in Lemma 3.4, we get a proper interval graph that is exactly two copies of $H[I_1 \cup C'_1]$ and has all the edges between the two copies of C'_1 . Note that since G_1 and G_2 are copies of the same graph, this proper interval graph is a PIG-completion of G' , with $2|F_1|$ fill edges, which proves that $|F_1| = |F_2| = k$.

Let $F'_1 = \{uv \in F : u, v \in I_1\}$, and let $H_1 = (V_1, E \cup F_1 \setminus F'_1)$. Note that since $H[I_1]$ is a clique and $G'[I_1]$ is an independent set, $|F'_1| = \binom{|I_1|}{2}$. In other words, all the possible edges between vertices of I_1 are fill edges. Therefore, H_1 has $\ell = k - \binom{|I_1|}{2}$ fill edges (as a completion of G_1). By Lemma 3.3, H_1 must be a threshold-completion of G_1 . Since G_1 is isomorphic to G , G has a threshold-completion with ℓ edges.

For the only if direction, suppose there is a minimum threshold-completion H of G with ℓ fill edges. We will construct a PIG-completion of G' with k' fill edges such that $k' = 2\left(\ell + \binom{|I|}{2}\right)$.

Let F_i be the set of fill edges added to the vertices corresponding to each G_i to obtain a threshold graph H_i for each $i \in \{1, 2\}$, and let $H'_i = (C'_i \cup I_i, E'_i \cup F_i)$. Notice that H'_i is also a threshold-completion of G'_i for each $i \in \{1, 2\}$. Observe that we can consider the same partition of the vertices into a clique and an independent set for both G_i and H_i . Indeed, if a vertex $v \in I_i$ is in the clique of H_i , we can remove all the edges connecting v with vertices in the independent set. This way we could place v in the independent set of H_i instead. The same holds for the vertex partition of G'_i and H'_i . Consider F'_i to be the fill edges obtained by completing I_i into a clique for each $i \in \{1, 2\}$, and let $H''_i = (V'_i, E'_i \cup F_i \cup F'_i)$ for each $i \in \{1, 2\}$. This gives a total of $\ell + \binom{|I|}{2}$ fill edges.

Since each H'_i is a threshold graph, the neighbourhoods of the independent vertices are nested and hence we can consider an ordering of said vertices in terms of increasing containment of their neighbourhoods. Recall that it is possible to represent any clique with a proper interval model by overlapping the corresponding intervals such that each interval starts and ends in a different point. A proper interval model for H''_i is given as follows, for each $i \in \{1, 2\}$. Since I_i is a clique in H'_i , we can place the corresponding intervals such that they overlap. The same holds for the intervals corresponding to all the vertices in C'_i , and we can place the endpoints of the intervals corresponding to these vertices by following the ordering of the neighbourhoods of the vertices in I_i to do this. This way, we can place the endpoints of the intervals corresponding to the vertices in I_i according to the increasing ordering given by the neighbourhoods with regards to C'_i , and thus obtaining a proper interval model for each H''_i , as in Fig. 3. Observe that C'_i is a first clique in the described proper interval model of H''_i for each $i \in \{1, 2\}$. Finally, we obtain a PIG-completion of G' with the desired number of fill edges by applying Lemma 3.4. ■

4. An algorithm for PIG-completion on threshold graphs

In this section we present a simple polynomial-time algorithm for computing an optimal PIG-completion for a threshold graph G . To do this, we will show first that PIG-completion for threshold graphs is equivalent to co-bipartite-completion. This will enable us to give a procedure, based on the definition of threshold graphs, that iteratively places the vertices in one of the two cliques in an optimal way.

Let G be a threshold graph, and let $n = |V(G)|$. We will show how to compute a PIG-completion for G . Recall that, since G is a threshold graph, there is an ordering $v_1, v_2, \dots, v_n$ of the vertices of G such that for each $i \in \{2, 3, \dots, n\}$, v_i is either a dominating or an isolated vertex in the subgraph $G[\{v_1, v_2, \dots, v_i\}]$. In the former case, we say that v_i is a *dominating* vertex, and in the latter case, we say that v_i is an *isolated* vertex. If v_n is an isolated vertex in G , then a PIG-completion of G can be obtained by simply computing a PIG-completion of $G[\{v_1, v_2, \dots, v_{n-1}\}]$ and then adding v_n to the resulting graph. We assume v_n is a dominating vertex. Throughout this section, we will use indistinctly and interchangeably the terms minimum and optimal when referring to the best possible completion.

Since each interval contains at least one of the endpoints of the interval corresponding to the universal vertex, it follows that any PIG that contains a universal vertex is a co-bipartite graph. Hence, any PIG-completion of any graph that contains a universal vertex is also a co-bipartite-completion of it.

Lemma 4.1. *A graph H is a minimum PIG-completion of G if and only if it is a minimum co-bipartite-completion of G .*

Proof. Since G contains a universal vertex, any PIG-completion of G is also a co-bipartite-completion of G . Conversely, let H be a minimum co-bipartite-completion of G . Consider a bipartition of $V(H)$ into two sets A and B such that the subgraphs $H[A]$ and $H[B]$ are both cliques. H does not contain independent sets of size 3 or more, which in particular implies that H contains no claws or asteroidal triples as induced subgraphs. Towards a contradiction, suppose H contains an induced cycle C . Since H is co-bipartite, C has either 3 or 4 vertices. Suppose that C is a 4-cycle, and let v_1, v_2, v_3, v_4 be its vertices, assuming without loss of generality that $\{v_1, v_2\} \subset A$ and that $\{v_3, v_4\} \subset B$. Since $v_1v_3, v_2v_4 \notin E(H)$ and H is a supergraph of G , in particular $v_1v_3, v_2v_4 \notin E(G)$. Moreover, $\{v_2v_3, v_4v_1\} \subset E(G)$ since H is a minimum co-bipartite-completion. It follows that the subgraph $G[\{v_1, v_2, v_3, v_4\}]$ is isomorphic to a $2K_2$, P_4 or C_4 , thus contradicting the fact that G is threshold. Hence, C must be a cycle of length 3, and therefore there are no induced claws, asteroidal triples, or induced cycles of length more than 3 in H , which implies that H is a PIG-completion of G . ■

Let us denote by I , the set of isolated vertices in $V(G)$. Given any bipartition $\{A, B\}$ of $V(G)$, we call the (A, B) -co-bipartite-completion of G the completion obtained by adding every possible edge between the vertices in A and every possible edge between the vertices in B . For any set $S \subseteq V(G)$ and $i \in \{1, 2, \dots, n\}$, define $S_{<i} = S \cap \{v_1, v_2, \dots, v_{i-1}\}$ and $S_{>i} = S \cap \{v_{i+1}, v_{i+2}, \dots, v_n\}$. The sets $S_{\leq i}$ and $S_{\geq i}$ are similarly defined. Observe that, for any optimal co-bipartite completion of G , the edges between the resulting cliques are in $E(G)$.

Lemma 4.2. *Let H be the (A, A') -co-bipartite-completion of G . Let $A^{(j)} = \begin{cases} A & \text{if } v_j \in A \\ A' & \text{otherwise.} \end{cases}$*

Then,

$$|E(H)| = |E(G)| + \sum_{v_j \in I} |A_{<j}^{(j)}| = |E(G)| + \sum_{v_i \in V(G)} |A_{>i}^{(i)} \cap I|$$

Proof. We obtain the (A, A') -co-bipartite completion H of G by adding, for each isolated vertex $v_j \in A^{(j)} \in \{A, A'\}$, all edges from v_j to the vertices of smaller index that are also in $A^{(j)}$, which are exactly the vertices of $A_{<j}^{(j)}$. In other words, all the edges between a vertex v_i and the isolated vertices of $A_{>i}^{(i)}$ are added. ■

Lemma 4.3. *Let $\{B, B'\}$ be a bipartition of $V(G)$ such that the (B, B') -co-bipartite-completion of G is optimal. Let v_t be a dominating vertex for some $t \in \{1, 2, \dots, n\}$. Suppose that $|B_{<t}| \geq |B'_{<t}|$. Then there exists a bipartition $\{A, A'\}$ of $V(G)$ such that $A_{<t} = B_{<t}$, $A'_{<t} = B'_{<t}$, $v_t \in A$, and the (A, A') -co-bipartite-completion of G is optimal.*

Proof. Denote by H' the (B, B') -co-bipartite-completion of G . If $v_t \in B$, then we can simply let $A = B$ and $A' = B'$. Suppose instead that $v_t \in B'$. Consider the sets $D = B_{<t} \cup B'_{\geq t}$ and $D' = B'_{<t} \cup B_{\geq t}$, and let H be the (D, D') -co-bipartite-completion of G . It follows from Lemma 4.2 that $|E(H)| = |E(H')| + (|B_{\geq t} \cap I| - |B'_{\geq t} \cap I|)(|B'_{<t}| - |B_{<t}|)$.

Suppose that $|B'_{<t}| < |B_{<t}|$, otherwise $|B'_{<t}| = |B_{<t}|$ and thus H is an optimal co-bipartite-completion of G and therefore $\{A, A'\}$, with $A = D$ and $A' = D'$, is a bipartition of $V(G)$ that fulfils the desired properties. Since H' is an optimal co-bipartite-completion of G by hypothesis, it follows from Lemma 4.2 that $|B_{\geq t} \cap I| \leq |B'_{\geq t} \cap I|$. Moreover, v_t is a dominating vertex, thus $|B_{>t} \cap I| \leq |B'_{>t} \cap I|$. Consider H'' to be the $(B \cup \{v_t\}, B' \setminus \{v_t\})$ -co-bipartite-completion of G . By using Lemma 4.2 again, we have that $|E(H'')| = |E(H')| - |B'_{>t} \cap I| + |B_{>t} \cap I|$. The latter combined with the fact that $|B_{>t} \cap I| \leq |B'_{>t} \cap I|$ implies that $|E(H'')| \leq |E(H')|$, hence H'' is also an optimal co-bipartite-completion of G . So in this case we can define $A = B \cup \{v_t\}$ and $A' = B' \setminus \{v_t\}$ that fulfil the desired properties. ■

Lemma 4.4. *There exists a bipartition $\{A, A'\}$ of G such that $|A| \geq |A'|$, every dominating vertex is in A and the (A, A') -co-bipartite-completion of G is optimal.*

Proof. Let $\{B, B'\}$ be a bipartition of $V(G)$ such that the (B, B') -co-bipartite-completion of G is optimal. Let $X^0 = B$ and $Y^0 = B'$. We will build X^i and Y^i for each $i \in \{0, 1, \dots, n\}$ maintaining the following invariants in each step: for each i , $|X_{\leq i}^i| \geq |Y_{\leq i}^i|$, and the (X^i, Y^i) -co-bipartite-completion of G is optimal. We inductively define step i as follows. If v_i is a dominating vertex and $v_i \in Y^{i-1}$, since $|X_{< i}^{i-1}| = |X_{\leq i-1}^{i-1}| \geq |Y_{\leq i-1}^{i-1}| = |Y_{< i}^{i-1}|$ and the (X^{i-1}, Y^{i-1}) -co-bipartite-completion of G was optimal, it follows from Lemma 4.3 that there exists a bipartition $\{A, A'\}$ of $V(G)$ such that $A_{< i} = X_{< i}^{i-1}$, $A'_{< i} = Y_{< i}^{i-1}$ where $X^i = A$ and $Y^i = A'$, $v_i \in A$, and the (A, A') -co-bipartite-completion of G is optimal. It is easy to see that we have $|X_{\leq i}^i| > |Y_{\leq i}^i|$.

Now suppose that v_i is an isolated vertex, $v_i \in Y^{i-1}$, and $|X_{\leq i-1}^{i-1}| = |Y_{\leq i-1}^{i-1}|$: in this case, let $X^i = X_{\leq i-1}^{i-1} \cup Y_{\geq i}^{i-1}$ and $Y^i = Y_{\leq i-1}^{i-1} \cup X_{\geq i}^{i-1}$. Observe that the edges induced respectively by $X_{\geq i}^{i-1}, X_{\leq i-1}^{i-1}, Y_{\geq i}^{i-1}, Y_{\leq i-1}^{i-1}$ are preserved in the (X^i, Y^i) -co-bipartite-completion. Thus, the remaining fill edges in both completions are the ones between those sets, and there are $|X_{\geq i}^{i-1} \cap I| \cdot |X_{\leq i-1}^{i-1}| + |Y_{\geq i}^{i-1} \cap I| \cdot |Y_{\leq i-1}^{i-1}|$ of them in the (X^{i-1}, Y^{i-1}) -co-bipartite-completion, and $|Y_{\geq i}^{i-1} \cap I| \cdot |X_{\leq i-1}^{i-1}| + |X_{\geq i}^{i-1} \cap I| \cdot |Y_{\leq i-1}^{i-1}|$ of them in the (X^i, Y^i) -co-bipartite-completion. Since $|X_{\leq i-1}^{i-1}| = |Y_{\leq i-1}^{i-1}|$, the (X^i, Y^i) -co-bipartite-completion of G has exactly the same number of edges as the (X^{i-1}, Y^{i-1}) -co-bipartite-completion of G , which means that it is also optimal. Moreover it is clear that we have $|X_{\leq i}^i| > |Y_{\leq i}^i|$.

For every other case, we simply consider $X^i = X^{i-1}$ and $Y^i = Y^{i-1}$. Clearly, the desired invariants hold in this case as well.

Finally, notice that X^n contains all the dominating vertices, $|X^n| \geq |Y^n|$, and that the (X^n, Y^n) -co-bipartite-completion of G is optimal. ■

Lemma 4.5. *There exists a bipartition $\{B, B'\}$ of G such that $|B| \geq |B'|$, every dominating vertex is in B , the (B, B') -co-bipartite-completion of G is optimal, and if $t = \min\{i \in \{1, 2, \dots, n\} : v_i \in B'\}$, then $|I_{>t}| < |B_{<t}| = t - 1$ and $I \cap B_{>t} = \emptyset$.*

Proof. Consider a bipartition $\{B, B'\}$ of $V(G)$ such that $|B| \geq |B'|$, every dominating vertex is in B , and the (B, B') -co-bipartite-completion of G is optimal. The existence of such a partition is a consequence of Lemma 4.4. Moreover, consider one such that $|I_{>t}| - |B_{<t}|$ is minimum, where $t = \min\{i \in \{1, 2, \dots, n\} : v_i \in B'\}$. Let H' be the (B, B') -co-bipartite-completion of G . It follows from the choice of the bipartition $\{B, B'\}$ that H' is an optimal co-bipartite-completion of G .

Towards a contradiction, suppose that $I \cap B_{>t} \neq \emptyset$. This implies in particular that there is a vertex $v_{t'} \in I \cap B_{>t}$, where $t' = \min\{i \in \{t+1, t+2, \dots, n\} : v_{t'} \in I \cap B\}$. We will now define another partition that will yield a co-bipartite-completion that uses less edges, thus reaching a contradiction. Let $A = (B \cup \{v_{t'}\}) \setminus \{v_{t'}\}$ and $A' = (B' \setminus \{v_{t'}\}) \cup \{v_{t'}\}$. Clearly, this is a partition of $V(G)$. If H be the (A, A') -co-bipartite-completion of G , then it follows that

$$|E(H)| = |E(H')| + |B_{<t}| + |I \cap B_{>t}| - 1 + |B'_{<t'}| - 1 + |I \cap B'_{>t'}| - |I \cap B_{>t'}| - |B_{<t'}| - |I \cap B'_{>t}|.$$

Furthermore, notice that every vertex in B' is isolated, thus we know that $|B'_{<t'}| + |I \cap B'_{>t'}| = |I \cap B'_{>t}| + 1$. It follows that,

$$|E(H)| = |E(H')| + |B_{<t}| + |I \cap B_{>t}| - 1 - |I \cap B_{>t'}| - |B_{<t'}|.$$

Since $v_{t'}$ is in $I \cap B_{>t}$, then $|I \cap B_{>t'}| = |I \cap B_{>t}| - 1$. This means that $|E(H)| = |E(H')| + |B_{<t}| - |B_{<t'}|$, which implies that $|B_{<t}| = |B_{<t'}|$ since H' is an optimal co-bipartite-completion of G . Thus, for each $i \in \{t+1, \dots, t'-1\}$, $v_i \in B'$ and by construction, $v_i \in A'$. But then $t+1 = \min\{i \in \{1, 2, \dots, n\} : v_i \in A'\}$ and $|I_{>t+1}| - |A_{<t+1}| = |I_{>t+1}| - |B_{<t}|$, which is strictly smaller than $|I_{>t}| - |B_{<t}|$, and thus contradicts the minimality of H' . Hence, necessarily $I \cap B_{>t} = \emptyset$.

Finally, suppose to the contrary that $|I_{>t}| - |B_{<t}| \geq 0$. Let $A = B \cup \{v_t\}$ and $A' = B' \setminus \{v_t\}$, and denote by H the (A, A') -co-bipartite completion of G . It is easy to see that H is an optimal co-bipartite-completion of G : we know that $|E(H)| = |E(H')| - |I \cap B'_{>t}| + |B_{<t}| + |I \cap B_{>t}|$, and since $|B_{<t}| \leq |I_{>t}| = |I \cap B_{>t}| + |I \cap B'_{>t}|$ and $|I \cap B_{>t}| = 0$, it follows that $|E(H)| \leq |E(H')|$. Therefore, we found a bipartition $\{A, A'\}$ of $V(G)$ such that $|A| \geq |A'|$, every dominating vertex is in A , the (A, A') -co-bipartite-completion H of G is optimal, and $|I_{>t'}| - |A_{<t'}| < |I_{>t}| - |B_{<t}|$, where $t' = \min\{i \in \{1, 2, \dots, n\} : v_i \in A'\} > t$. This results in a contradiction since we assumed $|I_{>t}| - |B_{<t}|$ to be minimum. This finishes the proof. ■

Theorem 4.6. *Given a threshold graph G on n vertices together with a threshold ordering of its vertices in which each vertex is either marked as “dominating” or “isolated”, there is an $\mathcal{O}(n)$ -time algorithm that computes a bipartition $\{X, Y\}$ of $V(G)$ such that the (X, Y) -co-bipartite-completion of G is optimal.*

Proof. Let $v_1, v_2, \dots, v_n$ be the given threshold ordering of the vertices of the graph G . We denote by I the vertices marked as isolated in this ordering. We construct a bipartition $\{X, Y\}$ of $V(G)$ as follows. Let $X^1 = \{v_1\}$ and $Y^1 = \emptyset$. For each $i \in \{2, 3, \dots, n\}$, we define X^i, Y^i inductively as follows:

1. If v_i is a dominating vertex (that is, v_i is not in I), or if $|I_{>i}| \geq |X^{i-1}|$, we add v_i to X^i . More precisely, $X^i = X^{i-1} \cup \{v_i\}$, and $Y^i = Y^{i-1}$.
2. Else, we place v_i in Y^i . That is, $X^i = X^{i-1}$, and $Y^i = Y^{i-1} \cup \{v_i\}$.

Once we have gone through all the vertices, we define $X = X^n$ and $Y = Y^n$.

Clearly the sets X and Y can be computed in time $\mathcal{O}(n)$, thus it suffices to show that the (X, Y) -co-bipartite-completion of G is optimal.

Let $\{B, B'\}$ be a bipartition of G such that $|B| \geq |B'|$, every dominating vertex is in B , the (B, B') -co-bipartite-completion of G is optimal, $|I_{>t}| < |B_{<t}| = t - 1$, and $I \cap B_{>t} = \emptyset$, where $t = \min\{i \in \{1, 2, \dots, n\} : v_i \in B'\}$. The existence of such partition follows from Lemma 4.5.

Now, it suffices to show that $X = B$ and $Y = B'$. Let v_i in X be any vertex, where $i \in \{1, 2, \dots, n\}$.

If v_i is a dominating vertex, then v_i is in B by definition of B . If instead v_i is an isolated vertex, then $|I_{>i}| \geq |X^{i-1}|$ which follows by construction of X . In this case, we know that $|I_{>j}| \geq |I_{>i}|$ and $|X^{j-1}| \leq |X^{i-1}|$, for every $j \in \{1, 2, \dots, i\}$. It follows that $|I_{>j}| \geq |X^{j-1}|$ for every $j \leq i$, thus by construction of X , v_j is also in X for every $j \in \{1, 2, \dots, i\}$. Hence, $|X^{i-1}| = i - 1$. Towards a contradiction, suppose that $v_i \in B'$. It follows from the definition of t that $t \leq i$. This implies that $|I_{>t}| \geq |I_{>i}| \geq |X^{i-1}| = i - 1 \geq t - 1$, which contradicts the assumption that stipulated that $|I_{>t}| < |B_{<t}| = t - 1$ for the chosen bipartition $\{B, B'\}$. Hence, v_i lies in B , thus $X \subseteq B$.

Suppose now to the contrary that $Y \cap B \neq \emptyset$, and let $i = \max\{j \in \{1, 2, \dots, n\} : v_j \in Y \cap B\}$. Observe that, if $v_k \in Y$, then for every $j \geq k$ such that $v_j \in I$, we have $v_j \in Y$ since v_j is an isolated vertex and $|I_{>j}| \leq |I_{>k}| < |X^{k-1}| \leq |X^{j-1}|$. In other words, $I_{>k} \subseteq Y$. It follows from the previous remark that $I_{>i} \cap B = \emptyset$. Moreover, since v_i is in Y , in particular v_i is an isolated vertex such that $|I_{>i}| < |X^{i-1}|$. The latter implies that $|I_{>i}| < |B_{<i}|$, since $X \subseteq B$. Furthermore, v_i is in B and $I \cap B_{>t} = \emptyset$, hence $i < t$. Since $I_{>i} \cap B = \emptyset$, then $|I_{>i} \cap B'| = |I_{>i}|$. Consider H to be the (B, B') -co-bipartite-completion of G , and let H' be the $(B \setminus \{v_i\}, B' \cup \{v_i\})$ -co-bipartite-completion of G . It follows from the previous remarks that

$$|E(H')| = |E(H)| + |I_{>i} \cap B'| - |I_{>i} \cap B| - |B_{<i}| = |E(H)| + |I_{>i}| - |B_{<i}| < |E(H)|,$$

which contradicts the fact that H' is an optimal co-bipartite-completion of G . The contradiction came from assuming that $Y \cap B \neq \emptyset$, therefore we have that $X = B$ and $Y = B'$. ■

To conclude this section, we mention the connection to the max-cut problem in the case of threshold graphs. Given a partition $\{A, B\}$ of the vertices of a graph G , consider these two sets:

1. Let C be the set of pairs of nonadjacent vertices, where one element is in A and the other in B .
2. Let F be the set of pairs of nonadjacent vertices where either both are in A or both in B (notice that F is a set of fill edges corresponding to a co-bipartite completion of G)

Notice that $E(G)$, C , and F gives a partition of the set of all pairs of distinct vertices from $V(G)$, hence $|E(G)| + |C| + |F| = \binom{n}{2}$. Two of these four values depend only on the input graph, not on the partition. Hence, choosing a partition minimizing $|F|$ is the same as maximizing $|C|$. Now take the complement graph and consider the same partition. Clearly, C is a cut between A and B . This implies that finding a minimum co-bipartite completion of a graph is equivalent to finding a maximum cut of its complement. By Lemma 4.1, this implies that for a threshold graph, finding a minimum PIG-completion is equivalent to finding a maximum cut of its complement.

We remark that an $\mathcal{O}(n^2)$ -time algorithm for calculating max-cut for cographs (and thus for threshold graphs) has been presented in [2]. This result, together with the fact that threshold graphs are closed under complementation, gives an alternative algorithm for PIG-completion for this class. Furthermore, even though our algorithm strongly relies on having a threshold ordering, in the worst case scenario where we do not have such an ordering, this can be computed in $\mathcal{O}(n + m)$ -time, thus yielding a $\mathcal{O}(n + m)$ -time algorithm for PIG-completion on threshold graphs using our proposed approach.

4.1. Completion from quasi-threshold graphs

A natural candidate for generalizing the previous result is the class of quasi-threshold graphs. Recall that these are precisely the $\{P_4, C_4\}$ -free graphs. However, for this particular class we encounter the problem that a crucial result from the previous section does not hold: a minimum co-bipartite-completion is not the same as a minimum PIG-completion. We present a counterexample for this in Fig. 4, in which we can see that a minimum co-bipartite-completion of the graph has an induced C_4 .

We will not give a full proof of the fact that the graph on the right of Fig. 4 is a minimal completion of the graph on the left. This was checked by a computer program. Let us call H_2, H_3, H_4 the cliques of size 4, 6 and 8 represented by circles in the figure, and H_1 be the clique composed of only the one vertex in the top left corner. The intuition is that since the values $|H_1| + |H_4|$ and $|H_2| + |H_3|$ are much closer than any other bipartition of the H_i 's would allow, and since the number of edges that we could possibly spare outside of the subgraph induced by $\cup_{i \in \{1, 2, 3, 4\}} H_i$ is comparatively small, the

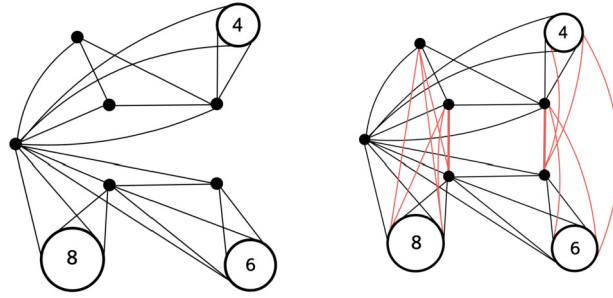


Fig. 4. A quasi-threshold graph G and a minimum co-bipartite-completion of G that is not PIG. Numbered circles correspond to cliques of size equal to the number inside it.

completion on the right of Fig. 4 is the best possible. It most notably beats the partition that has H_1 and H_2 on one side and H_3 and H_4 on the other, as well as the one that has H_1 , H_2 and H_3 on one side and H_4 on the other.

Nevertheless, we can still find an algorithm that solves the minimum co-bipartite-completion problem when the input is a quasi-threshold graph. Notice that, even though this does not necessarily yields a minimum PIG-completion for a quasi-threshold graph, indeed provides a lower bound for it. Recall that every connected quasi-threshold graph has a dominating vertex, and hence any PIG-completion is also a co-bipartite-completion. The fact that a minimum co-bipartite-completion gives a lower bound for the minimum PIG-completion follows from this.

Let $G = (V, E)$ be a connected quasi-threshold graph and $T = (V, E_T)$ be the tree rooted in r that stems from its definition.

For $v \in V$, let $n_v = |V(T_v)|$ be the number of vertices in T_v , the maximal subtree rooted at v , and let c_v be the number of children of v .

For $v \in V$, let $v_1, \dots, v_{c_v}$ be the children of v in T . For $0 \leq i \leq c_v$, let $X_{i,v} := \bigcup_{w \in \{v_1, \dots, v_i\}} V(T_w)$ and denote by $x_{i,v} = |X_{i,v}| = \sum_{1 \leq k \leq i} n_{v_k}$. For $0 \leq j \leq x_{i,v}$, let us define $C(v, i, j)$ as the minimum number of edges in a co-bipartite-completion in the subgraph induced by $X_{i,v}$, such that one clique has cardinality j and the other has cardinality $x_{i,v} - j$. Notice that $v \notin X_{i,v}$.

Finally, we define $D(v, j)$ as the minimum number of edges in a co-bipartite-completion of the subgraph induced by $V(T_v)$ such that one of the cliques has cardinality j .

Proposition 4.7. $D(v, 0) = C(v, c_v, n_v - 1)$ and for $1 \leq j \leq n_v$, it holds that $D(v, j) = C(v, c_v, j - 1)$.

Proof. Note that $V(T_v) = X_{c_v,v} \cup \{v\}$, and that v is adjacent to every vertex in $X_{c_v,v}$, hence we can always add it to any clique of the co-bipartite-completion of $X_{c_v,v}$ without increasing the cost of the completion. ■

It follows from those definitions that the value we are looking for is given by

$$\min_{0 \leq j \leq |V(G)|} D(r, j). \quad (1)$$

We now show how to compute $C(v, i, j)$, for every v and all possible values of $0 \leq i \leq c_v$ and $0 \leq j \leq x_{i,v}$. We assume that the children of v , if any, are $v_1, \dots, v_{c_v}$.

$$C(v, i, j) = \begin{cases} 0, & \text{if } v \text{ is a leaf or } i = 0. \\ \min_{\max(j - n_{v_i}, 0) \leq k \leq \min(j, x_{i-1,v})} \left(C(v, i-1, k) + \right. \\ \left. + D(v_i, j - k) + k(j - k) + (x_{i-1,v} - k)(n_{v_i} - j + k) \right), & \text{otherwise.} \end{cases} \quad (2)$$

Proposition 4.8. Equation (2) computes $C(v, i, j)$ correctly.

Proof. For the first case, the graph is either trivial or empty. For the second case, let C_1 and C_2 be the cliques of a co-bipartite-completion of $X_{i,v}$ such that $|C_1| = j$, and let $k = |X_{i-1,v} \cap C_1|$. Notice that $k \leq x_{i-1,v}$. It follows that $|V(T_{v_i}) \cap C_1| = j - k$, and thus that $n_{v_i} \geq j - k$. Notice that we need $C(v, i-1, k)$ edges to turn each of $X_{i-1,v} \cap C_1$ and $X_{i-1,v} \cap C_2$ into cliques. Similarly, we need $D(v_i, j - k)$ edges to turn $V(T_{v_i}) \cap C_1$ and $V(T_{v_i}) \cap C_2$ into cliques. Finally, we need $k(j - k)$ edges to connect $X_{i-1,v} \cap C_1$ and $V(T_{v_i}) \cap C_1$, and $(x_{i-1,v} - k)(n_{v_i} - j + k)$ edges to connect $X_{i-1,v} \cap C_2$ and $V(T_{v_i}) \cap C_2$. Since we try all possible valid values of k , we eventually find the smallest possible completion. ■

We arrive thus at the main result of this section, which we state below:

Theorem 4.9. *Let G be a quasi-threshold graph. There exists an $\mathcal{O}(n^4)$ -time dynamic programming algorithm that computes the minimum co-bipartite-completion number for G .*

Proof. The algorithm is given by expressions (1) and (2), and its correctness is immediate by Propositions 4.7 and 4.8. For the complexity, notice that the algorithm proceeds in a bottom-up manner, from the leaves of T up to the root, and from left to right for vertices located at the same level of the tree. For every vertex $v \in V$, every $0 \leq i \leq c_v$ and every $0 \leq j \leq x_{i,v}$, we compute the recurrence relation (2). This means that the values of $C(v, i-1, k)$ and $D(v_i, j-k)$ were already calculated for every $\max(j - n_{v_i}, 0) \leq k \leq \min(j, x_{i-1,v})$ by the time we compute $C(v, i, j)$, so they are obtainable in $\mathcal{O}(1)$ time. Hence, given v, i , and j , computing $C(v, i, j)$ is $\mathcal{O}(n)$. Since this is required for every i and j as defined above, we perform $\mathcal{O}(n^3)$ operations for every vertex of T . The complete algorithm is thus $\mathcal{O}(n^4)$. ■

5. An algorithm for PIG-completion on caterpillars

In the previous sections, we studied PIG-completion within subclasses of interval graphs that contain very dense graphs, such as complete graphs. Another possible approach is to restrict the input to more sparse graph classes. Namely, caterpillars are those interval graphs that are also trees, and thus the sparsest subclass of interval, when we consider connected graphs. We show that caterpillars admit minimum PIG-completions having a very particular structure, and we provide a quadratic-time algorithm to compute one of them. Recall that, as defined in Section 2, a *caterpillar* is a tree such that the deletion of its leaves results in a path, called its *central path*, or *spine*. From now on, for a caterpillar G on n vertices, we denote by P its central path, and the *children* of a vertex of P are the leaves to which said vertex is adjacent to.

Let us consider a PIG-completion H of G , and a unit interval model $\mathbf{I}_H$ of H . First, we transform $\mathbf{I}_H$ into the model of a PIG-completion H' of G with at most as many edges as H , such that the vertices of the central path are represented by unit intervals of the form $[i, i+1]$ for some integer i . Once we have that, we will describe an $\mathcal{O}(n^2)$ -time algorithm to obtain an optimal completion.

Theorem 5.1. *Let G be a caterpillar, and let k be the number of vertices of the central path P of G , and let $v_0, v_1, \dots, v_{k-1}$ be the vertices of P written in consecutive order. Then, there exists a minimum PIG-completion of G with a unit interval model such that, for each $i \in \{0, 1, \dots, k-1\}$:*

- v_i is represented by an interval of the form $[i, i+1]$, and
- each child of v_i is represented by either $[i \times (1 + \frac{1}{k}) - 1, i \times (1 + \frac{1}{k})]$ or $[(i+1) \times (1 + \frac{1}{k}) - 1, (i+1) \times (1 + \frac{1}{k})]$.

Proof. Let V be the vertex set of G . Let H be a PIG-completion of G and $\mathbf{I}_H = \{\mathbf{I}_u\}_{u \in V}$ be a unit interval model of H . Let us now transform H into a PIG-completion H' of G that requires at most as many edges as H (but H' is not necessarily a subgraph of H). To do this, we first obtain an auxiliary interval model $\mathbf{I}' = \{\mathbf{I}'_u\}_{u \in V}$ of H' that is not a unit interval model, in which some of the intervals are reduced to a single point, while the others remain as unit intervals. Once we obtain such an auxiliary model, we will transform it into a unit interval model that represents H' as well.

Consider the vertices of P , in the consecutive order $v_0, \dots, v_{k-1}$ in which they appear in the path. For $i \in \{0, \dots, k-1\}$, let V_i be the set of children of v_i .

We construct the model $\mathbf{I}'$ of H' iteratively using a sequence $H^{(0)}, H^{(1)}, \dots, H^{(k)}$ of supergraphs of G , where $H^{(0)} = H$ and $H^{(k)} = H'$, and for each j , $\mathbf{I}^{(j)} = \{\mathbf{I}^{(j)}_u\}_{u \in V}$ is an interval model of $H^{(j)}$.

For every $j \in \{0, \dots, k-1\}$ and every $i \in \{0, \dots, k-1\}$, let us denote by $V_{i,r}^{(j)}$ the set of children of v_i whose corresponding intervals contain the right endpoint of v_i in $\mathbf{I}^{(j)}$, including those that are represented by the same interval as v_i , and by $V_{i,\ell}^{(j)}$ the set of all the remaining children of v_i whose intervals contain the left endpoint of v_i in $\mathbf{I}^{(j)}$.

For each j , denote by S_j the set of vertices containing v_i and their children for every $i < j-1$, as well as every vertex in $V_{j-1,\ell}^{(j)}$ (if v_{j-1} exists). Analogously, denote by T_j the set of vertices containing every v_i for $i > j$ and all their children. Notice that neither v_j , v_{j-1} (if it exists), nor the children of v_j , nor the vertices in $V_{j-1,r}^{(j)}$ are in S_j or in T_j .

We define the set of *unimportant edges* in $H^{(j)}$ as those that fulfil one of the following conditions: (1) The edges having one endpoint in S_j and one in T_j , (2) The edge $v_{j-1}x$ (if v_{j-1} exists), for all $x \in T_j$ such that $v_jx \notin E(H^{(j)})$. Those edges that are not unimportant are referred to as *important edges*. We denote the set of important edges by F_j . In what follows, we will not count the unimportant edges in a model of the sequence. Nonetheless, we will show that the number of important edges never increases from $H^{(j)}$ to $H^{(j+1)}$, and that there are no unimportant edges in the end, that is, in $H^{(k)}$.

Given an integer $j \in \{0, \dots, k-1\}$, we say that the graph $H^{(j)}$ with interval model $\mathbf{I}^{(j)}$ satisfies *invariant A_j* if:

- $H^{(j)}$ is a supergraph of G with the same vertex set.
- For every integer $i \in \{0, \dots, j\}$, the vertex v_i in P is represented by the interval $[i, i+1]$.
- For every integer $i \in \{0, \dots, j-1\}$, the vertices in $V_{i,\ell}^{(j)}$ are represented by the point i , and the vertices in $V_{i,r}^{(j)}$ are represented by the point $i+1$.

- All the leaves of vertices of P from v_j onwards are represented by unit intervals, such that if two intervals intersect, then they intersect in more than one point.
- If $j > 0$, then $|F_j| \leq |F_{j-1}|$.

Note that we can assume that $\mathbf{I}^{(0)}$, the model of $H = H^{(0)}$, satisfies A_0 . Recall that we can always assume that two overlapping intervals intersect in more than one point. We can also suppose that v_0 is represented by $[0, 1]$, up to translation of the model.

Let j be an integer number smaller than $k - 1$. Suppose that $H^{(j)}$ is an interval graph with model $\mathbf{I}^{(j)}$ that satisfies A_j .

Let us first define the following vertex sets. Let p be the endpoint of $\mathbf{I}_{v_{j+1}}^{(j)}$ that lies between j and $j + 1$, (if $\mathbf{I}_{v_{j+1}}^{(j)} = [j, j + 1]$ then $p = j$). Let C be the set of all the vertices in T_j whose corresponding interval intersects p in $\mathbf{I}^{(j)}$. In turn, C is partitioned into two sets: C_1 , the vertices of C whose corresponding interval intersects j in $\mathbf{I}^{(j)}$, and C_2 , the remaining vertices (note that their intervals intersect $j + 1$ in $\mathbf{I}^{(j)}$). Let $V_L = V_{j,\ell}^{(j)} \cup V_{j-1,r}^{(j)} \cup \{v_{j-1}\}$ and V_R be an arbitrary subset of $V_{j,r}^{(j)}$ of size $\min(|V_{j,r}^{(j)}|, |C_1|)$.

We construct $H^{(j+1)}$ with model $\mathbf{I}^{(j+1)}$, starting with $\mathbf{I}^{(j)}$ and performing the following changes:

- (1) If p is the right endpoint of $\mathbf{I}_{v_{j+1}}^{(j)}$, then replace each interval corresponding to a vertex in T_j by the interval obtained after symmetry on point $j + \frac{1}{2}$. By symmetry on point x , we mean replacing an interval of the form $[x + a, x + b]$ by the interval $[x - b, x - a]$. Now the right endpoint of the interval corresponding to v_{j+1} lies to the right of the right endpoint of the interval corresponding to v_j .
Observe that after this step the intervals of the vertices of C all intersect the left endpoint of the interval of v_{j+1} . Moreover, this would also be true if instead of being the right endpoint, p was the left endpoint of $\mathbf{I}_{v_{j+1}}^{(j)}$.
- (2) Translate the intervals corresponding to vertices of T_j to the right, all by the same amount, so that the interval for v_{j+1} corresponds to $[j + 1, j + 2]$.
- (3) Let us set $V_{j,\ell}^{(j+1)} = V_{j,\ell}^{(j)} \cup V_R$ and $V_{j,r}^{(j+1)} = V_{j,r}^{(j)} \setminus V_R$, and replace the representation of every vertex in $V_{j,\ell}^{(j+1)}$ by point j and every vertex in $V_{j,r}^{(j+1)}$ by $j + 1$.

In what follows, we show that the resulting model $\mathbf{I}^{(j+1)}$ satisfies A_{j+1} . It is easy to see that with this construction, the vertex set of $H^{(j+1)}$ is V , and that for each i , the adjacencies between v_i and its children, as well as v_{i+1} and v_{i-1} if they exist, are preserved.

Note that, $S_{j+1} = S_j \cup V_L \cup V_R$ and $T_j = T_{j+1} \cup \{v_{j+1}\} \cup V_{j+1}$. Observe also that the intervals of the vertices in $S_j \cup V_{j-1,r}^{(j)} \cup \{v_{j-1}, v_j\}$ are identical in $\mathbf{I}^{(j+1)}$ and $\mathbf{I}^{(j)}$, and the subgraphs of $H^{(j)}$ and $H^{(j+1)}$ induced by this set plus $V_{j,\ell}^{(j)}$ (in other words induced by $S_{j+1} \cup \{v_j\} \setminus V_R$) are identical. Similarly the model induced by the intervals of the vertices in T_j is identical in $\mathbf{I}^{(j+1)}$ and $\mathbf{I}^{(j)}$, up to a potential symmetry and translation. In particular, for each $i \neq j$, either $V_{i,\ell}^{(j)} = V_{i,\ell}^{(j+1)}$ and $V_{i,r}^{(j)} = V_{i,r}^{(j+1)}$, or $V_{i,r}^{(j)} = V_{i,\ell}^{(j+1)}$ and $V_{i,\ell}^{(j)} = V_{i,r}^{(j+1)}$.

By construction, $\mathbf{I}_{v_{j+1}}^{(j+1)} = [j + 1, j + 2]$ and the vertices in $V_{j,\ell}^{(j+1)}$ and $V_{j,r}^{(j+1)}$ are represented respectively by the points j and $j + 1$.

Thus, from the previous observations and the fact that $H^{(j)}$ satisfies A_j , we deduce that $H^{(j+1)}$ satisfies the four first items of A_{j+1} . Let us now show that $H^{(j+1)}$ has at most as many important edges as $H^{(j)}$.

Recall that important edges are defined as the counterpart of unimportant edges in each $H^{(i)}$. To provide a more precise definition, observe that an edge in $H^{(i)}$ (here $i \in \{j, j + 1\}$) is important if one of the following assertions holds: (1) both its extremities are in S_i , or (2) both its extremities are in T_i , or (3) both extremities are in $\{v_{i-1}, v_i\} \cup V_{i-1,r}^{(i)} \cup V_i$, or (4) one extremity is in S_i and the other is in $\{v_{i-1}, v_i\} \cup V_{i-1,r}^{(i)} \cup V_i$, or (5) one extremity is some vertex x in T_i and, either the other lies in $\{v_i\} \cup V_{i-1,r}^{(i)} \cup V_i$, or the other extremity is precisely v_{i-1} and x, v_i are adjacent.

From the previous observation, all the important edges induced by $S_{j+1} \cup \{v_j\} \setminus V_R$ (resp. by T_j) are the same in $H^{(j)}$ and $H^{(j+1)}$. Similarly, V_R and $V_{j,r}^{(j+1)} \subseteq V_{j,r}^{(j)}$ each induce a clique in both $H^{(j)}$ and $H^{(j+1)}$, so the important edges they induce are also shared by the two graphs. Finally, all the edges between v_j and its children are common important edges. Let us call F the set of common important edges we just described (note that there are other common important edges, for example $v_j v_{j+1}$, that are not in F).

Claim 5.2. *There are at most $|F| + |V_L| \cdot |V_R| + |C| + |C| \cdot |V_{j,r}^{(j+1)}|$ edges in F_{j+1} .*

Proof. In the subgraph of $H^{(j+1)}$ induced by S_{j+1} , the only edges in $F_{j+1} \setminus F$ are edges with one endpoint in V_L and the other in V_R . Note that all such edges exist in $H^{(j+1)}$ but some of them can also exist in $H^{(j)}$. Thus, F_{j+1} contains $|V_L| \cdot |V_R|$ of such edges.

Now consider the edges of $H^{(j+1)}$ between S_{j+1} and T_j . Since the edges between S_{j+1} and T_{j+1} are not important, we only need to consider the subset $\{v_{j+1}\} \cup V_{j+1}$ of T_j . By A_j , the intervals of the children of v_{j+1} intersect $\mathbf{I}_{v_{j+1}}^{(j)} = [j+1, j+2]$ in more than one point, so in particular in some point strictly larger than $j+1$, and they have length one, so they do not intersect the point j . Note that the intervals of all the vertices of S_{j+1} , are contained in $[0, j]$. Therefore, there is no important edge between S_{j+1} and T_j .

Finally, the only vertices that are neither in T_j nor in S_{j+1} are the vertices of $V_{j,r}^{(j+1)} \cup \{v_j\}$. Every important edge between v_j and S_{j+1} is already counted in F , so for the important edges incident to v_j , we only need to focus on the one incident to the closed neighbourhood of v_{j+1} in T_{j+1} . The intervals of these vertices intersect $\mathbf{I}_{v_j}^{(j+1)} = [j, j+1]$ and $\mathbf{I}_{v_{j+1}}^{(j+1)} = [j+1, j+2]$, hence in particular they all contain $j+1$. The latter implies that they belong to C , thus there are $|C|$ such important edges incident to v_j . The intervals of the vertices in $V_{j,r}^{(j+1)}$ are reduced to the single point $j+1$, so they cannot be adjacent to any vertex in S_{j+1} , and in T_{j+1} , they are adjacent to only vertices in C . So there are at most $|C| \cdot |V_{j,r}^{(j+1)}|$ such edges in F_{j+1} and they are not counted in F .

Therefore, considering all the previous statements we have $|F_{j+1}| \leq |F| + |V_L| \cdot |V_R| + |C| + |C| \cdot |V_{j,r}^{(j+1)}|$. This finishes the proof of the claim. ■

Claim 5.3. *There are at least $|F| + |V_L| \cdot |C_1| + |C| + (|V_R| + |C_2|) \cdot |V_{j,r}^{(j+1)}| + |V_R| \cdot |C_2|$ edges in F_j .*

Proof. First, observe that in $\mathbf{I}^{(j)}$, all the intervals corresponding to vertices in V_L (that are not vertices in S_j by definition) intersect j , and the vertices of C_1 are vertices of T_j whose intervals intersect j . Moreover, the edges between v_{j-1} and C_1 are also important, since $\mathbf{I}^{(j)}(v_j)$ contains the point j . Thus, for every x in C_1 , we know that $v_j x \in E(H^{(j)})$, hence every possible edge between C_1 and V_L exists and is an important edge in $H^{(j)}$. This implies that there are at least $|V_L| \cdot |C_1|$ such edges in F_j .

The vertices in C_2 , in V_R , and in $V_{j,r}^{(j+1)}$ are all vertices whose intervals intersect $j+1$ in $\mathbf{I}^{(j)}$. Hence, these three sets are complete to each other in $H^{(j)}$. Notice that, whereas all the internal edges in each of the cliques C_2 , V_R , and $V_{j,r}^{(j+1)}$ are counted in F , the edges between these sets are not. This represents a total of $(|V_R| + |C_2|) \cdot |V_{j,r}^{(j+1)}| + |V_R| \cdot |C_2|$ edges in F_j .

Finally, v_j is adjacent to all the vertices in C , which give $|C|$ edges in F_j .

All the sets of edges described here are pairwise disjoint, so in total, there are at least $|F| + |V_L| \cdot |C_1| + |C| + (|V_R| + |C_2|) \cdot |V_{j,r}^{(j+1)}| + |V_R| \cdot |C_2|$ edges in F_j . This ends the proof of this claim. ■

Now, when we compare the number of important edges in $H^{(j)}$ and $H^{(j+1)}$, by the previous claims and the fact that $|C| = |C_1| + |C_2|$ we obtain that:

$$|F_j| - |F_{j+1}| \geq |V_L| \cdot (|C_1| - |V_R|) + |V_{j,r}^{(j+1)}| \cdot (|V_R| - |C_1|) + |V_R| \cdot |C_2|. \quad (3)$$

Recall that $|V_R| = \min(|V_{j,r}^{(j)}|, |C_1|)$. We have two possibilities to analyze:

- If $|V_{j,r}^{(j)}| \geq |C_1|$, then $|V_R| = |C_1|$. By replacing this in the inequality (3), we obtain $|F_j| - |F_{j+1}| \geq |V_R| \cdot |C_2| \geq 0$.
- If $|C_1| > |V_{j,r}^{(j)}|$, then $|V_{j,r}^{(j+1)}| = 0$ and $|C_1| > |V_R|$. Thus the inequality (3) becomes $|F_j| - |F_{j+1}| \geq |V_L| \cdot (|C_1| - |V_R|) + |V_R| \cdot |C_2| \geq |V_R| \cdot |C_2| \geq 0$.

Therefore, in both cases, $|F_j| \geq |F_{j+1}|$, which concludes the proof that $\mathbf{I}^{(j+1)}$ satisfies A_{j+1} .

We can thus build the sequence $H^{(0)}, H^{(1)}, \dots, H^{(k-1)}$ such that for each j , $\mathbf{I}^{(j)}$, the interval model of $H^{(j)}$ satisfies A_j .

Construct $\mathbf{I}^{(k)}$, the model of $H^{(k)} = H'$, from of $\mathbf{I}^{(k-1)}$ by replacing the interval corresponding to every vertex in $V_{k-1,\ell}^{(k-1)}$ by the point $k-1$, and the interval corresponding to every vertex in $V_{k-1,r}^{(k-1)}$ by the point k . All the other intervals remain the same.

Notice that since $\mathbf{I}^{(k-1)}$ satisfies A_{k-1} , $H^{(k-1)}$ is a supergraph of G and $\mathbf{I}_{v_{k-1}}^{(k)} = \mathbf{I}_{v_{k-1}}^{(k-1)} = [k-1, k]$. Thus, the children of v_{k-1} are also connected to v_{k-1} in H' . Moreover, observe that H' has at most as many edges as $H^{(k-1)}$. Notice that $H^{(k-1)}$ has no unimportant edges and, by A_{k-1} , at most as many important edges as $H^{(0)} = H$. Hence, H' is a supergraph of G with at most as many edges as H and its model $\mathbf{I}^{(k)}$ fulfils the following:

- For every $i \in \{0, \dots, k-1\}$, the vertex v_i is represented by the interval $[i, i+1]$.
- For every $i \in \{0, \dots, k-1\}$, the vertices in $V_{i,\ell}^{(k)}$ are represented by the point i , and the vertices in $V_{i,r}^{(k)}$ are represented by the point $i+1$.

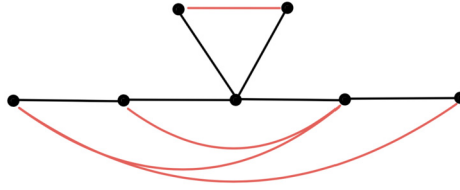


Fig. 5. A caterpillar graph (black edges) and an inclusion-wise minimal PIG-completion (red edges) for which the characterization given in Theorem 5.1 does not hold.

To conclude, let us show that H' is a PIG-completion of G by exhibiting a unit interval model I' of H' , that we construct as follows: For each interval in $I^{(k)}$, if it is a unit interval (i.e., a vertex of the central path), then we add the same interval to I' . Otherwise, the interval contains a single point i , thus we add the interval $[i \times (1 + \frac{1}{k}) - 1, i \times (1 + \frac{1}{k})]$ to I' . Note that such an interval will intersect exactly the same vertices of the central path and will not intersect intervals built from other single-point intervals unless they already coincided. Therefore H' is a unit interval graph, and thus a PIG-completion of G , with at most as many edges as H . ■

An interesting question that arises is whether this suffices to characterize all inclusion-wise minimal completions or not. In Fig. 5 we can see depicted a caterpillar graph G , represented by the black edges, and a supergraph H obtained by adding the red edges that is indeed an inclusion-wise minimal PIG-completion of G . The central path of G is not an induced path in H , therefore this cannot be a PIG-completion of the form specified in Theorem 5.1.

Now that we know that there exists a minimum PIG-completion as the one stated in Theorem 5.1, we design an algorithm that will always output such a solution.

Theorem 5.4. *There exists an algorithm that outputs a minimum PIG-completion of a caterpillar in $\mathcal{O}(n^2)$ -time.*

Proof. As before, let G be a caterpillar and let $v_0, \dots, v_{k-1}$ be the vertices of the central path, considered in the order in which they appear on the central path. Let V_i be the children of v_i . In order to build a minimum completion as the one presented in the statement of Theorem 5.1, we just need to determine how many elements of V_i are represented by $[i \times (1 + \frac{1}{k}) - 1, i \times (1 + \frac{1}{k})]$ and by $[(i + 1) \times (1 + \frac{1}{k}) - 1, (i + 1) \times (1 + \frac{1}{k})]$, which we will denote by V_i^ℓ and V_i^r respectively. We will determine the size of V_i^ℓ (and of V_i^r) via dynamic programming. More precisely, for every $i \in \{0, \dots, k-1\}$, and for every $j \in \{0, \dots, |V_i|\}$, we compute the optimal number $N_{i,j}$ of edges added to complete the caterpillar induced by $\{v_i, v_{i+1}, \dots, v_{k-1}\}$ and their children, except for j children of v_i , into a proper interval graph, assuming that all its remaining children are in V_i^r . We use the following formula to compute it:

- For $j \in \{0, \dots, |V_{k-1}|\}$, $N_{k-1,j} = \binom{|V_{k-1}|-j}{2}$;
- For $i \in \{0, \dots, k-2\}$ and $j \in \{0, \dots, |V_i|\}$, $N_{i,j} = \min_{j' \in \{0, \dots, |V_{i+1}|\}} \left(\binom{|V_i|-j+j'}{2} + j' + N_{i+1,j'} \right)$.

Now, the best PIG-completion of the entire caterpillar graph uses exactly

$$\min_{j \in \{0, \dots, |V_0|\}} \left(\binom{j}{2} + N_{0,j} \right)$$

fill edges. One can get such a completion from the choices of j obtained to get the minimal values.

Since $\sum_{i \in \{0, \dots, k-1\}} (|V_i| + 1) = n$, we know that there are only n different values of $N_{i,j}$ to be computed. Each one is computed using a minimum over at most n different expressions. Therefore the complexity of this algorithm is in $\mathcal{O}(n^2)$. ■

6. Conclusions and future work

In this work, we study the Π -completion problem when Π is the class of proper interval graphs. Given that the problem is NP-complete in general graphs, we focus our analysis on the case in which the input graph lies in some particular subclasses of chordal graphs. We prove that the problem remains hard in split graphs, whereof we conclude the same for chordal graphs. We present efficient algorithms for PIG-completion for both threshold graphs and caterpillar graphs, and an efficient algorithm for co-bipartite-completion for quasi-threshold graphs.

A future line of work is to continue studying the PIG-completion problem in other subclasses of chordal graphs that have bounded cliquewidth to obtain practical polynomial-time algorithms. This may lead to finding common properties that could be useful when it comes to designing efficient algorithms and heuristics to solve the problem within other chordal subclasses.

On the other hand, given that all the graph classes for which we give a polynomial-time algorithm are also subclasses of interval graphs, and that interval and proper interval graphs are very closely related, it raises as a natural question whether the PIG-completion problem can be solved in polynomial time when the input graph already belongs to this particular class.

In addition, since we studied the PIG-completion problem within caterpillars motivated by the fact that these graphs are precisely those interval graphs that are also trees, this gives way to an analogous question regarding the complexity of the PIG-completion problem when the input is a tree. Given that our current algorithm heavily relies on an interval model of the graph, we do not expect that our approach could be generalized to trees. This implies that the problem on trees should involve different techniques.

An interesting question arises also in the relation with the max-cut problem: for which other classes besides threshold graphs does it hold that PIG-completion is equivalent to co-bipartite-completion? For these possible classes, an algorithm for max-cut in the complement would also solve the PIG-completion problem, by the same argument given in Section 4.

Another possible continuation for this work may be to study the *PIG-deletion problem*, i.e., the removal of a set of edges F from an input graph $G = (V, E)$, so that the resulting subgraph $H = (V, E \setminus F)$ is a proper interval graph. It is known that this problem is hard for general graphs [8], but it would be interesting to investigate whether efficient algorithms could also be devised for this problem restricted to the aforementioned subclasses of chordal graphs.

Declaration of competing interest

No conflicts of interest.

Data availability

No data was used for the research described in the article.

Acknowledgements

We thank the reviewers for their time and helpful comments, which certainly improved this work.

Partially supported by Programa Regional MATHAMSUD MATH190013, by CNPq grants 311679/2018-8, 312069/2021-9 and 406036/2021-7, by FAPEMIG grant APQ-01707-21, by Argentina PIP 2021-2023 20020190200124BA, by Argentina UBACyT 20020170100495BA, by DFG grant VI 1045/1-1, by The Slovenian Research and Innovation Agency (research project J1-4008).

References

- [1] I. Bliznets, F.V. Fomin, M. Pilipczuk, M. Pilipczuk, A subexponential parameterized algorithm for proper interval completion, *SIAM J. Discrete Math.* 29 (4) (2015) 1961–1987.
- [2] H. Bodlaender, K. Jansen, On the complexity of the maximum cut problem, *Nord. J. Comput.* 7 (1) (2000) 14–31.
- [3] J.A. Bondy, U.S.R. Murty, *Graph Theory*, 1st edition, Springer Publishing Company, Incorporated, 2008.
- [4] B. Courcelle, J.A. Makowsky, U. Rotics, Linear time solvable optimization problems on graphs of bounded clique-width, 2000.
- [5] C. Crespelle, I. Todinca, An $O(n^2)$ -time algorithm for the minimal interval completion problem, *Theor. Comput. Sci.* 494 (2013) 75–85.
- [6] J. Diaz, A.M. Gibbons, M.S. Paterson, J. Toran, The MINSUMCUT problem, in: *Algorithms and Data Structures*, in: *Lecture Notes in Computer Science*, vol. 519, Springer, Berlin, 1991, pp. 65–79.
- [7] M.R. Garey, D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman and Company, N.Y., 1979.
- [8] P. Goldberg, M.C. Golumbic, H. Kaplan, R. Shamir, Four strikes against physical mapping of DNA, *J. Comput. Biol.* 2 (1) (1995) 139–152.
- [9] Phil Hanlon, Counting interval graphs, *Trans. Am. Math. Soc.* 272 (2) (1982) 383–426.
- [10] P. Heggernes, K. Suchan, I. Todinca, Y. Villanger, Characterizing minimal interval completions, in: Wolfgang Thomas, Pascal Weil (Eds.), *STACS 2007, 24th Annual Symposium on Theoretical Aspects of Computer Science, Proceedings, Aachen, Germany, February 22–24, 2007*, in: *Lecture Notes in Computer Science*, vol. 4393, Springer, 2007, pp. 236–247.
- [11] H. Kaplan, R. Shamir, R. Tarjan, Tractability of parameterized completion problems on chordal, strongly chordal and proper interval graphs, *SIAM J. Comput.* 28 (5) (1999) 1906–1922.
- [12] T. Kashiwabara, T. Fujisawa, An NP-complete problem on interval graphs, in: *IEEE Symp. Of Circuits and Systems*, 1979, pp. 82–83.
- [13] C. Lekkerkerker, J. Boland, Representation of a finite graph by a set of intervals on the real line, *Fundam. Math.* 51 (1962) 45–64.
- [14] A. Natanzon, R. Shamir, R. Sharan, Complexity classification of some edge modification problems, *Discrete Appl. Math.* 113 (2001) 109–128.
- [15] I. Rapaport, K. Suchan, I. Todinca, Minimal proper interval completions, *Inf. Process. Lett.* 106 (5) (2008) 195–202.
- [16] F.S. Roberts, *Representations of indifference relations*, PhD thesis, 1968.
- [17] F.S. Roberts, Indifference graphs. Proof techniques in graph theory, in: *Proceedings of the Second Ann Arbor Graph Conference*, Academic Press, New York, 1969.
- [18] B. Seymour, On the topology of the genetic fine structure, *Proc. Natl. Acad. Sci. USA* 45 (1959) 1607–1620.
- [19] P. Sheng-Lung, C. Chi-Kang, On the interval completion of chordal graphs, *Discrete Appl. Math.* 154 (6) (2006) 1003–1010.
- [20] M. Yannakakis, Computing the minimum fill-in is NP-complete, *SIAM J. Algebraic Discrete Methods* 2 (1) (1981) 77–79.
- [21] J. Yuan, Y. Lin, Y. Liu, S. Wang, NP-completeness of the profile problem and the fill-in problem on cobipartite graphs, *J. Math. Study* 31 (1998) 239–243.