Polynomial Kernels for Proper Interval Completion and Related Problems

Stéphane Bessy, Anthony Perez

Polynomial kernels for Proper Interval Completion and related problems *

Stéphane Bessy Anthony Perez
bessy@lirmm.fr perez@lirmm.fr
LIRMM – Université Montpellier II - FRANCE

November 17, 2011

Abstract

Given a graph $G = (V, E)$ and a positive integer k, the Proper Interval Completion problem asks whether there exists a set F of at most k pairs of $(V \times V) \setminus E$ such that the graph $H = (V, E \cup F)$ is a proper interval graph. The Proper Interval Completion problem finds applications in molecular biology and genomic research [16, 24]. First announced by Kaplan, Tarjan and Shamir in FOCS '94, this problem is known to be FPT [16], but no polynomial kernel was known to exist. We settle this question by proving that Proper Interval Completion admits a kernel with at most $O(k^3)$ vertices. Moreover, we prove that a related problem, the so-called Bipartite Chain Deletion problem, admits a kernel with at most $O(k^2)$ vertices, completing a previous result of Guo [13].

Introduction

The aim of a graph modification problem is to transform a given graph in order to get a certain property Π satisfied. Several types of transformations can be considered: for instance, in *vertex deletion* problems, we are only allowed to delete vertices from the input graph, while in *edge modification problems* the only allowed operation is to modify the edge set of the input graph. The optimization version of such problems consists in finding a *minimum* set of edges (or vertices) whose modification makes the graph satisfy the given property Π. Graph modification problems cover a broad range of NP-Complete problems and have been extensively studied in the literature [20, 23, 24]. Well-known examples include the Vertex Cover [8], Feedback Vertex Set [26], or Cluster Editing [5] problems. These problems find applications in various domains, such as computational biology [16, 24], image processing [23] or relational databases [25].

A natural approach to deal with such problems is to measure their difficulty with respect to some parameter such as ,for instance, the number of allowed modifications. *Parameterized complexity* provides a useful theoretical framework to that aim [10, 21]. A problem *parameterized* by some integer k is said to be *fixed-parameter tractable* (FPT for short) whenever it can be solved in time $f(k) \cdot n^c$ for some constant $c > 0$, where n is the size of the instance (for problems on graphs, usually, n is the number of vertices of the input graph). A natural parameterization for graph modification problems thereby consists in the number of allowed transformations. As one of the most powerful technique to design fixed-parameter algorithms, *kernelization algorithms* have

*Research supported by the AGAPE project (ANR-09-BLAN-0159).

been extensively studied in the last decade (see [2] for a survey). A *kernelization algorithm* is a polynomial-time algorithm (called *reduction rules*) that given an instance (I,k) of a parameterized problem P computes an instance (I',k') of P such that (i) (I,k) is a YES-instance if and only if (I',k') is a YES-instance and (ii) $|I'| \leq h(k)$ for some computable function $h()$ and $k' \leq k$. The instance (I',k') is called the *kernel* of P. We say that (I',k') is a *polynomial kernel* if the function $h()$ is a polynomial. It is well-known that a decidable parameterized problem is FPT if and only if it has a kernelization algorithm [21]. But this equivalence only yields kernels of super-polynomial size. To design efficient fixed-parameter algorithms, a kernel of small size - polynomial (or even linear) in k - is highly desirable [22]. However, recent results give evidence that not every parameterized problem admits a polynomial kernel, unless $NP \subseteq coNP/poly$ [3]. On the positive side, notable kernelization results include a less-than-$2k$ kernel for VERTEX COVER [8], a $4k^2$ kernel for FEEDBACK VERTEX SET [26] and a $2k$ kernel for CLUSTER EDITING [5].

We follow this line of research with respect to graph modification problems. It has been shown that a graph modification problem is FPT whenever Π is hereditary and can be characterized by a finite set of forbidden induced subgraphs [4]. However, recent results proved that several graph modification problems do not admit a polynomial kernel even for such properties Π [12, 18]. In this paper, we are in particular interested in *completion* problems, where the only allowed operation is to add edges to the input graph. We consider the property Π as being the class of *proper interval graphs*. This class is a well-studied class of graphs, and several characterizations are known to exist [19, 30]. In particular, there exists an *infinite* set of forbidden induced subgraphs that characterizes proper interval graphs [30] (see Figure 1). More formally, we consider the following problem:

PROPER INTERVAL COMPLETION:
Input: A graph $G = (V,E)$ and a positive integer k.
Parameter: k.
Output: A set F of at most k pairs of $(V \times V) \setminus E$ such that the graph $H = (V, E \cup F)$ is a proper interval graph.

Interval completion problems find applications in molecular biology and genomic research [15, 16], and in particular in *physical mapping* of DNA. In this case, one is given a set of long contiguous intervals (called *clones*) together with experimental information on their pairwise overlaps, and the goal is to reconstruct the relative position of the clones along the target DNA molecule. We focus here on the particular case where all intervals have equal length, which is a biologically important case (e.g. for cosmid clones [15]). In the presence of (a small number of) unidentified overlaps, the problem becomes equivalent to the PROPER INTERVAL COMPLETION problem. It is known to be NP-Complete for a long time [11], but fixed-parameter tractable due to a result of Kaplan, Tarjan and Shamir in FOCS '94 [16, 17][1]. The fixed-parameter tractability of the PROPER INTERVAL COMPLETION can also be seen as a corollary of a characterization of Wegner [30] combined with Cai's result [4]. Nevertheless, it was not known whether this problem admits a polynomial kernel or not.

Our results We prove that the PROPER INTERVAL COMPLETION problem admits a kernel with at most $O(k^3)$ vertices. To that aim, we identify *nice* parts of the graph that induce proper interval graphs and can hence be safely reduced. Moreover, we apply our techniques to the so-called BIPARTITE CHAIN DELETION problem, closely related to the PROPER INTERVAL COMPLETION problem where one is given a graph $G = (V,E)$ and seeks a set of at most k edges whose deletion

[1]Notice also that the *vertex deletion* of the problem is fixed-parameter tractable [28].

from E results in a bipartite chain graph (a graph that can be partitioned into two independent sets connected by a join). We obtain a kernel with $O(k^2)$ vertices for this problem. This result completes a previous result of Guo [13] who proved that the Bipartite Chain Deletion With Fixed Bipartition problem admits a kernel with $O(k^2)$ vertices.

Outline We begin with some definitions and notations regarding proper interval graphs. Next, we give the reduction rules the application of which leads to a kernelization algorithm for the Proper Interval Completion problem. These reduction rules allow us to obtain a kernel with at most $O(k^3)$ vertices. Finally, we prove that our techniques can be applied to Bipartite Chain Deletion to obtain a quadratic-vertex kernel.

1 Preliminaries

1.1 Proper interval graphs

We consider simple, loopless, undirected graphs $G = (V(G), E(G))$ where $V(G)$ denotes the vertex set of G and $E(G)$ its edge set[2]. Given a vertex $v \in V$, we use $N_G(v)$ to denote the *open neighborhood* of v and $N_G[v] = N_G(v) \cup \{v\}$ for its *closed neighborhood*. Two vertices u and v are *true twins* if $N[u] = N[v]$. If u and v are not true twins but $uv \in E$, we say that a vertex of $N[u] \triangle N[v]$ *distinguishes* u and v. Given a subset of vertices $S \subseteq V$, $N_S(v)$ denotes the set $N_G(v) \cap S$ and $N_G(S)$ denotes the set $(\cup_{s\in S} N_G(s)) \setminus S$. Moreover, $G[S]$ denotes the subgraph *induced* by S, i.e. $G[S] = (S, E_S)$ where $E_S = \{uv \in E \ : \ u, v \in S\}$. A *join* in a graph $G = (V, E)$ is a bipartition (X, Y) of G and an order $x_1, \ldots, x_{|X|}$ on X such that for all $i = 1, \ldots, |X|-1$, $N_Y(x_i) \subseteq N_Y(x_{i+1})$. The edges between X and Y are called the *edges of the join*, and a subset $F \subseteq E$ is said to *form a join* if F corresponds to the edges of a join of G. Finally, a graph is an *interval graph* if it admits a representation on the real line such that: (i) the vertices of G are in bijection with intervals of the real line and (ii) $uv \in E$ if and only if $I_u \cap I_v \neq \emptyset$, where I_u and I_v denote the intervals associated to u and v, respectively. Such a graph is said to admit an *interval representation*. A graph is a *proper interval graph* if it admits an interval representation such that $I_u \not\subset I_v$ for every $u, v \in V$. In other words, no interval strictly contains another interval.
We will make use of the two following characterizations of proper interval graphs to design our kernelization algorithm.

Theorem 1.1 (Forbidden subgraphs [30])**.** *A graph is a proper interval graph if and only if it does not contain any* $\{hole, claw, net, 3\text{-}sun\}$ *as an induced subgraph (see Figure 1).*

The claw graph is the bipartite graph $K_{1,3}$. Denoting its bipartition by $(\{c\}, \{l_1, l_2, l_3\})$, we call c the *center* and $\{l_1, l_2, l_3\}$ the *leaves* of the claw.

Theorem 1.2 (Umbrella property [19])**.** *A graph is a proper interval graph if and only if its vertices admit an ordering* σ *(called* umbrella ordering*) satisfying the following property: given* $v_iv_j \in E$ *with* $i < j$ *then* $v_iv_l, v_lv_j \in E$ *for every* $i < l < j$ *(see Figure 2).*

In the following, we associate an umbrella ordering σ_G to any proper interval graph $G = (V, E)$. There are several things to remark. First, note that in an umbrella ordering σ_G of a graph G, every maximal set of true twins of G is consecutive. Moreover, it is known [9] that σ_G is unique up to permutation of true twins of G or by reversal of the ordering induced on a connected component of

[2] In all our notations, we forget the mention to the graph G whenever the context is clear.

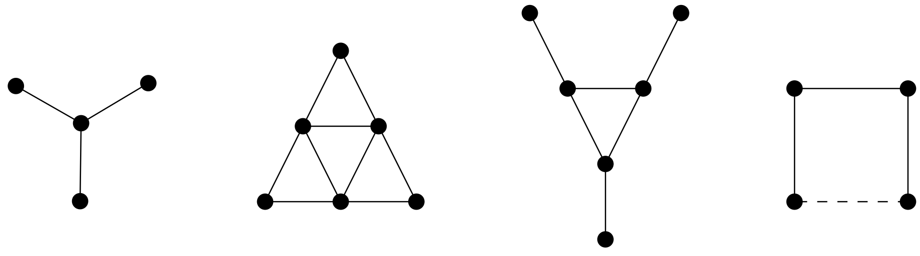


Figure 1: The forbidden induced subgraphs of proper interval graphs. A *hole* is an induced cycle of length at least 4.

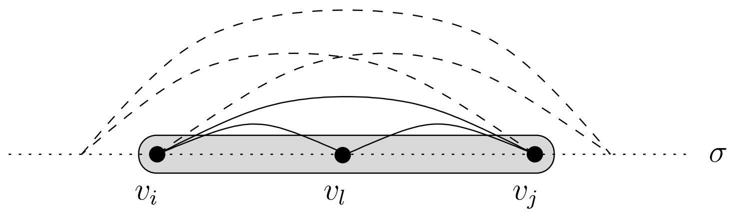


Figure 2: Illustration of the umbrella property. The edge v_iv_j is extremal. [3]

G. Remark also that for any edge uv with $u <_{\sigma_G} v$, the set $\{w \in V \;:\; u \leq_{\sigma_G} w \leq_{\sigma_G} v\}$ is a clique of G, and for every i with $1 \leq i < l$, $(\{v_1, \ldots, v_i\}, \{v_{i+1}, \ldots, v_n\})$ is a join of G.
According to this ordering, we say that an edge uv is *extremal* if there does not exist any edge $u'v'$ different from uv such that $u' \leq_{\sigma_G} u$ and $v \leq_{\sigma_G} v'$ (see Figure 2).

Let $G = (V, E)$ be an instance of Proper Interval Completion. A *completion* of G is a set $F \subseteq (V \times V) \setminus E$ such that the graph $H = (V, E \cup F)$ is a proper interval graph. In a slight abuse of notation, we use $G + F$ to denote the graph H. A k-*completion* of G is a completion such that $|F| \leq k$, and an *optimal completion* F is such that $|F|$ is minimum. We say that $G = (V, E)$ is a *positive* instance of Proper Interval Completion whenever it admits a k-completion. We state a simple observation that will be very useful for our kernelization algorithm.

Observation 1.3. *Let $G = (V, E)$ be a graph and F be an optimal completion of G. Given an umbrella ordering σ of $G + F$, any extremal edge of σ is an edge of G.*

Proof. Assume that there exists an extremal edge e in σ that belongs to F. By definition, σ is still an umbrella ordering if we remove the edge e from F, contradicting the optimality of F. □

1.2 Branches

We now give the main definitions of this Section. The branches that we will define correspond to some parts of the graph that already behave like proper interval graphs. They are the parts of the graph that we will reduce in order to obtain a kernelization algorithm.

Definition 1.4 (1-branch)**.** *Let $B \subseteq V$. We say that B is a* 1-*branch if the following properties hold (see Figure 3):*

[3]In all the figures, (non-)edges between blocks stand for all the possible (non-)edges between the vertices that lie in these blocks, and the vertices within a gray box form a clique of the graph.

(i) The graph $G[B]$ is a connected proper interval graph admitting an umbrella ordering $\sigma_B = b_1, \ldots, b_{|B|}$ and,

(ii) The vertex set $V \setminus B$ can be partitioned into two sets R and C with: no edges between B and C, every vertex in R has a neighbor in B, no edges between $\{b_1, \ldots, b_{l-1}\}$ and R where b_l is the neighbor of $b_{|B|}$ with minimal index in σ_B, and for every $l \leq i < |B|$, we have $N_R(b_i) \subseteq N_R(b_{i+1})$.

We denote by B_1 the set of vertices $\{v \in V : \ b_l \leq_{\sigma_B} v \leq_{\sigma_B} b_{|B|}\}$, which is a clique (because b_l is a neighbor of $b_{|B|}$). This set is exactly the neighborhood of $b_{|B|}$ in B. We call B_1 the *attachment clique* of B, and use B^R to denote $B \setminus B_1$.

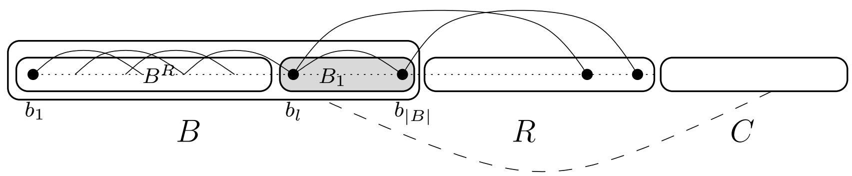


Figure 3: A 1-branch of a graph $G = (V, E)$. The vertices of B are ordered according to the umbrella ordering σ_B.

Definition 1.5 (2-branch). *Let $B \subseteq V$. We say that B is a 2-branch if the following properties hold (see Figure 4):*

(i) The graph $G[B]$ is a connected proper interval graph admitting an umbrella ordering $\sigma_B = b_1, \ldots, b_{|B|}$ and,

(ii) The vertex set $V \setminus B$ can be partitioned into sets L, R and C with:

- *no edges between B and C,*
- *every vertex in L (resp. R) has a neighbor in B,*
- *no edges between $\{b_1, \ldots, b_{l-1}\}$ and R where b_l is the neighbor of $b_{|B|}$ with minimal index in σ_B,*
- *no edges between $\{b_{l'+1}, \ldots, b_{|B|}\}$ and L where $b_{l'}$ is the neighbor of b_1 with maximal index in σ_B and,*
- *$N_R(b_i) \subseteq N_R(b_{i+1})$ for every $l \leq i < |B|$ and $N_L(b_{i+1}) \subseteq N_L(b_i)$ for every $1 \leq i < l'$.*

Again, we denote by B_1 (resp. B_2) the set of vertices $\{v \in V : \ b_1 \leq_{\sigma_B} v \leq_{\sigma_B} b_{l'}\}$ (resp. $\{v \in V : \ b_l \leq_{\sigma_B} v \leq_{\sigma_B} b_{|B|}\}$). We call B_1 and B_2 the *attachment cliques* of B, and use B^R to denote $B \setminus (B_1 \cup B_2)$. We assume that $L \neq \emptyset$ and $R \neq \emptyset$, otherwise B is a 1-branch. Finally, when $B^R = \emptyset$, it is possible that a vertex of L or R is adjacent to all the vertices of B. In this case, we will denote by N the set of vertices that are adjacent to every vertex of B, remove them from R and L and abusively still denote by L (resp. R) the set $L \setminus N$ (resp. $R \setminus N$). We will precise when we need to use the set N.

In both cases, in a 1- or 2-branch, whenever the proper interval graph $G[B]$ is a *clique*, we say that B is a *K-join*. Observe that, in a 1- or 2-branch B, for any extremal edge uv in σ_B, the set of vertices $\{w \in V : \ u \leq_{\sigma_B} w \leq_{\sigma_B} v\}$ defines a K-join. In particular, this means that a branch can

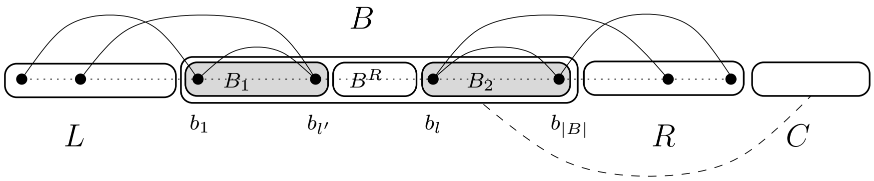


Figure 4: A 2-branch of a graph $G = (V, E)$. The vertices of B are ordered according to the umbrella ordering σ_B.

be decomposed into a sequence of K-joins. Observe however that the decomposition is not unique: for instance, the K-joins corresponding to all the extremal edges of σ_B are not disjoint. We will precise in Section 2.1.5, when we will reduce the size of 2-branches, how to fix a decomposition. Finally, we say that a K-join is *clean* whenever its vertices are not contained in any claw or 4-cycle. Remark that a subset of a K-join (resp. clean K-join) is also a K-join (resp. clean K-join).

2 Kernel for Proper Interval Completion

The basic idea of our kernelization algorithm is to detect the large enough branches and then to reduce them. This section details the rules we use for that.

2.1 Reduction rules

2.1.1 Basic rules

We say that a rule is *safe* if when it is applied to an instance (G, k) of the problem, (G, k) admits a k-completion if, and only if, the instance (G', k') reduced by the rule admits a k'-completion.

The first reduction rule gets rid of connected components that are already proper interval graphs. This rule is trivially safe and can be applied in $O(n + m)$ time using any recognition algorithm for proper interval graphs [6].

Rule 2.1 (Connected components)**.** *Remove any connected component of G that is a proper interval graph.*

The following reduction rule can be applied since proper interval graphs are closed under true twin addition and induced subgraphs. For a class of graphs satisfying these two properties, we know that this rule is safe [1] (roughly speaking, we edit all the large set of true twins in the same way). Furthermore, it is possible to compute every set of pairewise true twins using a modular decomposition algorithm or more easily, partition refinement (see [14] for example).

Rule 2.2 (True twins [1])**.** *Let T be a set of true twins in G such that $|T| > k$. Remove $|T| - (k+1)$ arbitrary vertices from T.*

We also use the classical *sunflower* rule, allowing to identify a set of edges that must be added in any optimal completion.

Rule 2.3 (Sunflower)**.** *Let $\mathcal{S} = \{C_1, \ldots, C_m\}$, $m > k$ be a set of claws having two leaves u, v in common but distinct third leaves. Add uv to F and decrease k by 1.*
Let $\mathcal{S} = \{C_1, \ldots, C_m\}$, $m > k$ be a set of distinct 4-cycles having a non-edge uv in common. Add uv to F and decrease k by 1.

Lemma 2.1. *Rule 2.3 is safe and can be carried out in polynomial time. More precisely, it is possible to detect all the 4-cycles and claws of G in time $O(n^2m)$.*

Proof. We only prove the first rule. The second rule can be proved similarly. Let F be a k-completion of G and assume that F does not contain (u, v). Since any two claws in $\mathcal{S}$ only share (u, v) as a common non-edge, F must contain one edge for every C_i, $1 \leq i \leq m$. Since $m > k$, we have $|F| > k$, which cannot be. Now, we brievely indicate how to compute all claws and the 4-cycles of G. For every edge xy of G, in time $O(n)$, we compute the sets $N_x = N_G(x) \setminus N_G[y]$ and $N_y = N_G(y) \setminus N_G[x]$. Each edge uv between N_x and N_y correspond to the 4-cycle $xyvu$. So, in time $O(m.(n + m))$ (less than $O(n^2m)$), we enumerate all the 4-cycles of G. On the other hand, for every vertex x of G, we compute all the three cycle in H_x, the complementary of $G[N_G(x)]$, what can be done in time $O(n(H_x)m(H_x))$ (for instance, by computing for every vertex y of H_x, a breadh search tree rooted on y). This gives all the claws with center x. And, in all, we enumerate all the claws of G in time $O(n^2m)$. Finaly, sparsing the claws and the 4-cycles, it is then easy to detect the sunflowers. □

2.1.2 Number of vertices in claws or 4-cycles

The general idea of our process is to reduce the size of the branches. However, we realized that is not always possible, even for K-join. We will see that this problem is due to the presence of claws or 4-cycles intersecting the branches. So, in this part, we give a bound of the number of vertices belonging to these obstructions in a positive instance of Proper Interval Completion.

Lemma 2.2. *Let $G = (V, E)$ be a positive instance of* Proper Interval Completion *on which Rule 2.3 has been applied. There are at most k^2 claws with distinct sets of leaves, and at most $k^2 + 2k$ vertices of G are leaves of claw. Furthermore, there are at most $2k^2 + 2k$ vertices of G that are vertices of a 4-cycle.*

Proof. As G is a positive instance of Proper Interval Completion, every claw or 4-cycle of G has a non-edge that will be completed and then is an edge of F. Let xy be an edge of F. As we have applied Rule 2.3 on G, there are at most k vertices in G that form the three leaves of a claw with x and y. So, at most $(k + 2)k$ vertices of G are leaves of claws. Similarly, there are at most k non-edges of G, implying at most $2k$ vertices, that form a 4-cycle with x and y. So, at most $(2k + 2)k$ vertices of G are in a 4-cycle. □

Lemma 2.3. *Let $G = (V, E)$ be a positive instance of* Proper Interval Completion *on which Rule 2.2 and Rule 2.3 have been applied. There are at most $4k^3 + 15k^2 + 16k$ vertices of G that belong to a claw or a 4-cycle.*

Proof. As G is a positive instance of Proper Interval Completion, there exists a set F of at most k edges such that $G + F$ is a proper interval graph and admits an umbrella ordering σ. We contract all the set of true twins of G and denote by G' the obtained graph. Remark that, as Rule 2.2 has been applied on G, every contracted set has size at most $k + 1$. As G' is also an induced subgraph of G, we denote by σ' the order induced by σ on G'.
Now, we define C to be the vertices of G' which are center of a claw in G', not incident to any edge of F, are not contained in a 4-cycle neither a leaf of a claw. We sort this set according to σ' and denote by $c_1, \dots, c_l$ its vertices in this order. As the vertices of C are not incident with edges of F, the edges incident with vertices of C respect the umbrella property.
We look for distinct vertices which distinguish the pairs of consecutive vertices of C. Remark that

it is possible that two consecutive vertices of C, c_i and c_{i+1} are twins, but not true twins. In this case, we can identify all the neighbors of c_i and c_{i+1}. Indeed, assume that c_i and c_{i+1} are not linked but that they have same neighborhood. Then, c_i has no neighbor x with $x <_{\sigma'} c_i$, otherwise x is also a neighbor of c_{i+1} and c_i and c_{i+1} would be neighbors, by the umbrella property. As c_i is not an isolated vertex, it has at least one neighbor. So, let x be the neighbor of c_i with maximal index in σ'. As c_i and c_{i+1} are not linked, then $x <_{\sigma'} c_{i+1}$. So, let Y denotes the set $\{y \in G' \ : \ x <_{\sigma'} y <_{\sigma'} c_{i+1}\}$. If $Y \neq \emptyset$, as x and c_{i+1} are linked by an edge, then Y is a set of neighbors of c_{i+1} and then a set of neighbors of c_i also, what contradicts the choice of x. So, $Y = \emptyset$, and c_{i+1} is the first non-neighbor of c_i after c_i according to σ'. Similarly, c_i is the last non-neighbor of c_{i+1} before c_{i+1} according to σ', and we conclude that $N_{G'}(c_i) = N_{G'}(c_{i+1}) = \{x \in G' \ : \ c_i <_{\sigma'} x <_{\sigma'} c_{i+1}\}$. So, c_i and c_{i-1} cannot be twins, so it is for c_{i+1} and c_{i+2}. It means that we can remove at most half of c_i and obtain $C' = \{c'_1, \ldots, c'_p\}$ (with $p \geq l/2$), a subset of C, sorted according to σ', in which every pair of consecutive vertices is not made of twins.

Now, let x be a vertex of G'. As no vertex of C' are incident to an edge of F, it means that the neighborhood of x in C' is consecutive according to the order $c'_1, \ldots, c'_p$. Then, x distinguishes at most two pairs $\{c'_i, c'_{i+1}\}$, for $1 \leq i \leq p-1$ So, for $1 \leq i \leq p-1$, we choose d_i a vertex of G' which distinguishes c'_i from c'_{i+1}. If, amongst all the vertices of G' which distinguishes c'_i from c'_{i+1}, one is the leaf of a claw, we preferably choose it for d_i. As seen previously, it is possible that a vertex has been chosen twice to be a vertex d_i, but no more than two times. So, the set $\{d_1, \ldots, d_{p-1}\}$ contains at least $(p-1)/2$ distinct vertices which we denote by $d'_1, \ldots, d'_q$ sorted according to σ', and with $q \geq (p-1)/2 \geq l/4 - 1$.

Now, for every $i = 1, \ldots, q$, we will find a claw containing d'_i as leaf. Assume that such a claw does not exist, we will derive a contradiction. Without loss of generality, we can assume that we have $d'_i c'_j \notin E(G')$ and $d'_i c'_{j+1} \in E(G')$, for some j with $1 \leq j \leq p-1$. By hypothesis, c'_{j+1} is the center of a claw in G'. We denote by x, y and z the leaves of this claw. As d'_i is not the leaf of a claw, it is disjoint from $\{x, y, z\}$, and by the choice of d'_i, no one of these vertices distinguishes c'_j from c'_{j+1}. It means that c'_j is linked to all vertices of $\{x, y, z\}$. If two elements of this set, say x and y, are adjacent to d'_i, then $\{x, d'_i, y, c'_j\}$ forms a 4-cycle that contains c'_j, which is not possible. So, at least two elements among $\{x, y, z\}$, say x and y, are not adjacent to d'_i and then, we find the claw $\{c'_{j+1}, x, y, d'_i\}$ of center c'_{j+1} that contains d'_i, which is also not possible, by assumption.

Finally, for $1 \leq i \leq q$ every d'_i is the leaf of a claw. So, by Lemma 2.2, we have $q \leq k^2+2k$. Then, we conclude that $l \leq 4(k^2+2k+1)$ and that G contains at most $4(k^2+2k+1)(k+1)$ vertices which are center of a claw. Finally, using Lemma 2.2 G contains at most $4(k^2+2k+1)(k+1)+k^2+2k+2k^2+2k$ vertices belonging to a claw or a 4-cycle. □

Remark that, using Lemma 2.1, it possible to detect all the vertices of G which belongs to a claws or a 4-cycle in time $O(n^2m)$.

2.1.3 Bounding the size of the clean K-joins

Now, we set a rule that will bound the number of vertices in a clean K-join, once applied. Although quite technical to prove, this rule is the core tool of our process of kernelization. Remark that, if we remove the vertices contained in a claw or a 4-cycle from a (general) K-join, we obtain a clean K-join. So, by the result of the previous subsection, providing a bound on the size of the clean K-joins will give a bound on the size of K-joins.

Rule 2.4 (K-join)**.** *Let B be a clean K-join of size at least $2k+2$, provided with an umbrella ordering σ_B. Let B_L be the $k+1$* first vertices *of B (according to σ_B), B_R be its $k+1$* last vertices *(according to σ_B) and $M = B \setminus (B_R \cup B_L)$. Remove the set of vertices M from G.*

Lemma 2.4. *Rule 2.4 is safe.*

Proof. Let $G' = G \backslash M$. Observe that the restriction to G' of any k-completion of G is a k-completion of G', since proper interval graphs are closed under induced subgraphs. So, let F be a k-completion for G'. We denote by H the resulting proper interval graph $G' + F$ and by $\sigma_H = h_1, \ldots, h_{|H|}$ an umbrella ordering of H. We prove that we can insert the vertices of M into σ_H and modify it if necessary, to obtain an umbrella ordering for G without adding any edge (in fact, some edges of F might even be deleted during the process). This will imply that G admits a k-completion as well. To see this, we need the following structural description of G. As explained before, we denote by N the set $\cap_{b \in B} N_G(b) \setminus B$, and abusively still denote by L (resp. R) the set $L \setminus N$ (resp. $R \setminus N$) (see Figure 5). We also denote by $b_1, \ldots, b_{|B|}$ the umbrella ordering σ_B of B.

Claim 2.5. *The sets L and R are cliques of G.*

Proof. We prove that R is a clique in G. The proof for L uses similar arguments. No vertex of R is a neighbor of b_1, otherwise such a vertex must be adjacent to every vertex of B and then stands in N. So, if R contains two vertices u, v such that $uv \notin E$, we form the claw $\{b_{|B|}, b_1, u, v\}$ with center $b_{|B|}$, contradicting the fact that B is clean. $\diamond$

The following observation comes from the definition of a K-join.

Observation 2.6. *Given any vertex $r \in R$, if $N_B(r) \cap B_L \neq \emptyset$ holds then $M \subseteq N_B(r)$. Similarly, given any vertex $l \in L$, if $N_B(l) \cap B_R \neq \emptyset$ holds then $M \subseteq N_B(l)$.*

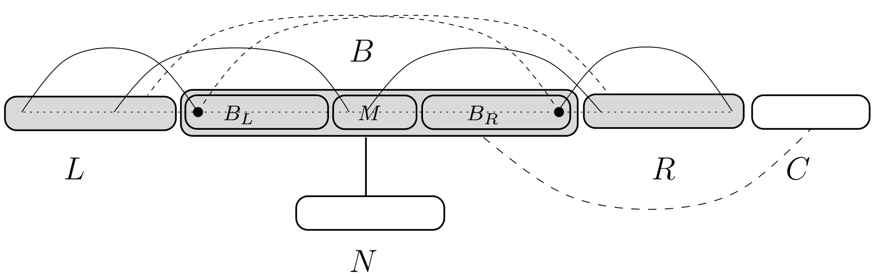


Figure 5: The structure of the K-join B.

We use these facts to prove that an umbrella ordering can be obtained for G by inserting the vertices of M into σ_H. Let h_f and h_l be respectively the first and last vertex of $B \setminus M$ appearing in σ_H. We let B_H denote the set $\{u \in V(H) \ : \ h_f \leq_{\sigma_H} u \leq_{\sigma_H} h_l\}$. Observe that B_H is a clique in H since $h_f h_l \in E(G)$ and that $B \setminus M \subseteq B_H$. Now, we modify σ_H by ordering the true twins in H according to their neighborhood in M: if x and y are true twins in H, are consecutive in σ_H, verify $x <_{\sigma_H} y <_{\sigma_H} h_f$ and $N_M(y) \subset N_M(x)$, then we exchange x and y in σ_H. This process stops when the considered true twins are ordered following the join between $\{u \in V(H) \ : \ u <_{\sigma_H} h_f\}$ and M. We proceed similarly on the right of B_H, i.e. for x and y consecutive twins with $h_l <_{\sigma_H} x <_{\sigma_H} y$ and $N_M(x) \subset N_M(y)$. The obtained order is clearly an umbrella ordering too (in fact, we just re-labeled some vertices in σ_H), and we abusively still denote it by σ_H.

Claim 2.7. *The set $B_H \cup \{m\}$ is a clique of G for any $m \in M$, and consequently $B_H \cup M$ is a clique of G.*

Proof. Let u be any vertex of B_H. We claim that $um \in E(G)$. Observe that if $u \in B$ then the claim trivially holds. So assume $u \notin B$. Recall that B_H is a clique in H. It follows that u is adjacent to every vertex of $B \setminus M$ in H. Since B_L and B_R both contain $k+1$ vertices, we have $N_G(u) \cap B_L \neq \emptyset$ and $N_G(u) \cap B_R \neq \emptyset$. Hence, u belongs to $L \cup N \cup R$ and $um \in E(G)$ by Observation 2.6. $\diamond$

Claim 2.8. *Let m be any vertex of M and σ'_H be the ordering obtained from σ_H by removing B_H and inserting m to the position of B_H. The ordering σ'_H respects the umbrella property.*

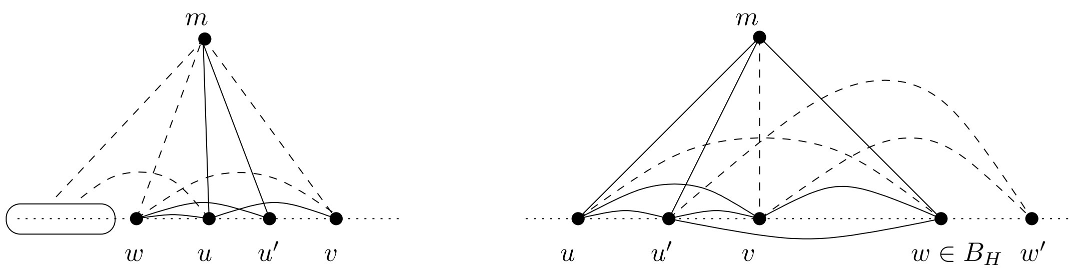
Proof. Assume that σ'_H does not respect the umbrella property, i.e. that there exist (w.l.o.g.) two vertices u and v of $H \setminus B_H$ such that either (1) $u <_{\sigma'_H} v <_{\sigma'_H} m$, $um \in E(H)$ and $uv \notin E(H)$ or (2) $u <_{\sigma'_H} m <_{\sigma'_H} v$, $um \notin E(H)$ and $uv \in E(H)$ or (3) $u <_{\sigma'_H} v <_{\sigma'_H} m$, $um \in E(H)$ and $vm \notin E(H)$. First, assume that (1) holds. Since $uv \notin E(H)$ and σ_H is an umbrella ordering, $uw \notin E(H)$ for any $w \in B_H$, and hence $uw \notin E(G)$. This means that $B_L \cap N_G(u) = \emptyset$ and $B_R \cap N_G(u) = \emptyset$, which is impossible since $um \in E(G)$. Then, assume that (2) holds. Since $uv \in E(H)$ and σ_H is an umbrella ordering, $B_H \subseteq N_H(u)$, and in particular B_L and B_R are included in $N_H(u)$. As $|B_L| = |B_R| = k+1$, we know that $N_G(u) \cap B_L \neq \emptyset$ and $N_G(u) \cap B_R \neq \emptyset$, but then, Observation 2.6 implies that $um \in E(G)$. So, (3) holds, and we choose the first u satisfying this property according to the order given by σ'_H. So we have $wm \notin E(G)$ for any $w <_{\sigma'_H} u$. Similarly, we choose v to be the first vertex after u satisfying $vm \notin E(G)$. Since $um \in E(G)$, we know that u belongs to $L \cup N \cup R$. Moreover, since $vm \notin E(G)$, $v \in C \cup L \cup R$. There are several cases to consider:

(i) $u \in N$: in this case we know that $B \subseteq N_G(u)$, and in particular that $uh_l \in E(G)$. Since σ_H is an umbrella ordering for H, it follows that $vh_l \in E(H)$ and $B_H \subseteq N_H(v)$. Since $|B_L| = |B_R| = k+1$, we know that $N_G(v) \cap B_L \neq \emptyset$ and $N_G(v) \cap B_R \neq \emptyset$. But, then Observation 2.6 implies that $vm \in E(G)$.

(ii) $u \in R$, $v \notin R$: since $um \in E(G)$, $B_R \subseteq N_G(u)$. Let $b \in B_R$ be the vertex such that $B_R \subseteq \{w \in V \;:\; u <_{\sigma_H} w \leq_{\sigma_H} b\}$. Since $ub \in E(G)$, this means that $B_R \subseteq N_H(v)$. Now, since $|B_R| = k+1$, it follows that $N_G(v) \cap B_R \neq \emptyset$. Observation 2.6 allows us to conclude that $vm \in E(G)$.

(iii) $u, v \in R$: in this case, $uv \in E(G)$ by Claim 2.7 but u and v are not true twins in H (otherwise v would be placed before u in σ_H due to the modification we have applied to σ_H). This means that there exists a vertex $w \in V(H)$ that distinguishes u from v in H.
Assume first that $w <_{\sigma_H} u$ and $uw \in E(H), vw \notin E(H)$. We choose the first w satisfying this according to the order given by σ_H. There are two cases to consider. First, if $uw \in E(G)$, then since $wm \notin E(G)$ for any $w <_{\sigma_H} u$ by the choice of u, $\{u, v, w, m\}$ is a claw in G containing a vertex of B (see Figure 6 (a) ignoring the vertex u'), which cannot be. So assume $uw \in F$. By Observation 1.3, uw is not an extremal edge of σ_H. By the choice of w and since $vw \notin E(H)$, there exists u' with $u <_{\sigma_H} u' <_{\sigma_H} v$ such that wu' is an extremal edge of σ_H (and hence belongs to $E(G)$, see Figure 6 (a)). Now, by the choice of v we have $u'm \in E(G)$ and hence $u' \in N \cup R \cup L$. Observe that $u'v \notin E(G)$: otherwise $\{u', v, w, m\}$ would form a claw in G. Since R is a clique of G, it follows that $u' \in L \cup N$. Moreover, since $u'm \in E(G)$, $B_L \subseteq N_G(u')$. We conclude like in configuration (ii) that v should be adjacent to a vertex of B_L and hence to m.
Hence we can assume that all the vertices that distinguish u and v are after u in σ_H and that $uw'' \in E(H)$ implies $vw'' \in E(H)$ for any $w'' <_{\sigma_H} u$. Now, suppose that there exists $w \in H$

such that $h_l <_{\sigma_H} w$ and $uw \notin E(H)$, $vw \in E(H)$. In particular, this means that $B_L \subseteq N_H(v)$. Since $|B_L| = k+1$ we have $N_G(v) \cap B_L \neq \emptyset$, implying $vm \in E(G)$ by Observation 2.6. Assume now that there exists a vertex w which distinguishes u and v with $v <_{\sigma_H} w <_{\sigma_H} h_f$. In this case, since $uw \notin E(H)$, $B \cap N_H(u) = \emptyset$ holds and hence $B \cap N_G(u) = \emptyset$, which cannot be since $u \in R$. Finally, assume that there is $w \in B_H$ with $wu \notin E(H)$ and $wv \in E(H)$. Recall that $wm \in E(G)$ as $B_H \cup \{m\}$ is a clique by Claim 2.7. We choose w in B_H distinguishing u and v to be the last according to the order given by σ_H (i.e. $vw' \notin E(H)$ for any $w <_{\sigma_H} w'$, see Figure 6 (b), ignoring the vertex u').



(a) (b)

Figure 6: (a) u and v are distinguished by some vertex $w <_{\sigma_H} u$; (b) u and v are distinguished by a vertex $w \in B_H$.

If $vw \in E(G)$ then $\{u, m, w, v\}$ is a 4-cycle in G containing a vertex of B, which cannot be. Hence $vw \in F$ and by the choice of w, there exists $u' \in V(H)$ such that $u <_{\sigma_H} u' <_{\sigma_H} v$ and $u'w$ is an extremal edge of σ_H (and then belongs to $E(G)$). By the choice of v we know that $u'm \in E(G)$. Moreover, by the choice of w, observe that u' and v are true twins in H (if a vertex s distinguishes u' and v in H, s cannot be before u, since otherwise s would distinguishes u and v, not between u and w because it would be adjacent to u' and v, and not after w, by choice of w). This leads to a contradiction since we assumed that $N_M(x) \subseteq N_M(y)$ for any true twins x and y with $x <_{\sigma_H} y <_{\sigma_H} h_f$.

The cases where $u \in L$ are similar, what concludes the proof of Claim 2.8 $\diamond$

Now, we will insert vertices of M into the graph H while preserving an umbrella ordering. For simplicity, once one vertex of M is inserted into H, we still denote the obtained graph by H and consider the new vertex as a vertex of H, for the next add. We then prove the following.

Claim 2.9. *Let m be a vertex of M. Then m can be added to the graph H while preserving an umbrella ordering.*

Proof. Let m be a vertex of M and h_i (resp. h_j) be the vertex with minimal (resp. maximal) index in σ_H such that $h_i m \in E(G)$ (resp. $h_j m \in E(G)$). By definition, we have $h_{i-1}m \notin E(G)$, $h_{j+1}m \notin E(G)$ and through Claim 2.8, we know that $N_H(m) = \{w \in V(H) \ : \ h_i \leq_{\sigma_H} w \leq_{\sigma_H} h_j\}$. Moreover, since $B_H \cup M$ is a clique by Claim 2.7, it follows that $h_{i-1} <_{\sigma_H} h_f$ and $h_l <_{\sigma_H} h_{j+1}$. Hence, by Claim 2.8, we know that $h_{i-1}h_{j+1} \notin E(G)$, otherwise the ordering σ'_H defined in Claim 2.8 would not be an umbrella ordering. The situation is depicted in Figure 7 (a). For any vertex $v \in N_H(m)$, let $N^-(v)$ (resp. $N^+(v)$) denote the set of vertices $\{w \in V(H) \ : \ w \leq_{\sigma_H} h_{i-1}$ and $wv \in E(H)\}$ (resp. $\{w \in V(H) \ : \ w \geq_{\sigma_H} h_{j+1}$ and $wv \in E(H)\}$). Observe that for any vertex

$v \in N_H(m)$, if there exist two vertices $x \in N^-(v)$ and $y \in N^+(v)$ such that $xv \in E(G)$ and $yv \in E(G)$, then the set $\{v, x, y, m\}$ defines a claw containing m in G, which cannot be. We now consider $c_{h_{i-1}}$ the neighbor of h_{i-1} with maximal index in σ_H. Similarly we let $c_{h_{j+1}}$ be the neighbor of h_{j+1} with minimal index in σ_H. Since $h_{i-1}h_{j+1} \notin E(G)$, we have $c_{h_{i-1}}, c_{h_{j+1}} \in N_H(m)$. We study the behavior of $c_{h_{i-1}}$ and $c_{h_{j+1}}$ in order to conclude.

Assume first that $c_{h_{j+1}} \leq_{\sigma_H} c_{h_{i-1}}$. Let X be the set of vertices $\{w \in V(H) \ : \ c_{h_{j+1}} \leq_{\sigma_H} w \leq_{\sigma_H} c_{h_{i-1}}\}$. Remark that we have $c_{h_{i-1}} \leq_{\sigma_H} h_l$ and $h_f \leq_{\sigma_H} c_{h_{j+1}}$, otherwise for instance, if we have $c_{h_{i-1}} >_{\sigma_H} h_l$, then $B_H \subseteq N_H(h_{i-1})$ implying, as usual, that $h_{i-1}m \in E(G)$ which is not. So, we know that $X \subseteq B_H$. Then, let $X_1 \subseteq X$ be the set of vertices $x \in X$ such that there exists $w \in N^+(x)$ with $xw \in E(G)$ and $X_2 = X \setminus X_1$. Let $x \in X_1$: observe that by construction $xw' \in F$ for any $w' \in N^-(x)$. Similarly, given $x \in X_2$, $xw'' \in F$ for any $w'' \in N^+(x)$. Now, we reorder the vertices of X as follows: we first put the vertices from X_2 and then the vertices from X_1, preserving the order induced by σ_H for both sets. Moreover, we remove from $E(H)$ all edges between X_1 and $N^-(X_1)$ and between X_2 and $N^+(X_2)$. Recall that such edges have to belong to F. We claim that inserting m *between* X_2 and X_1 yields an umbrella ordering (see Figure 7 b). Indeed, by Claim 2.8, we know that the umbrella ordering is preserved between m and the vertices of $H \setminus B_H$.

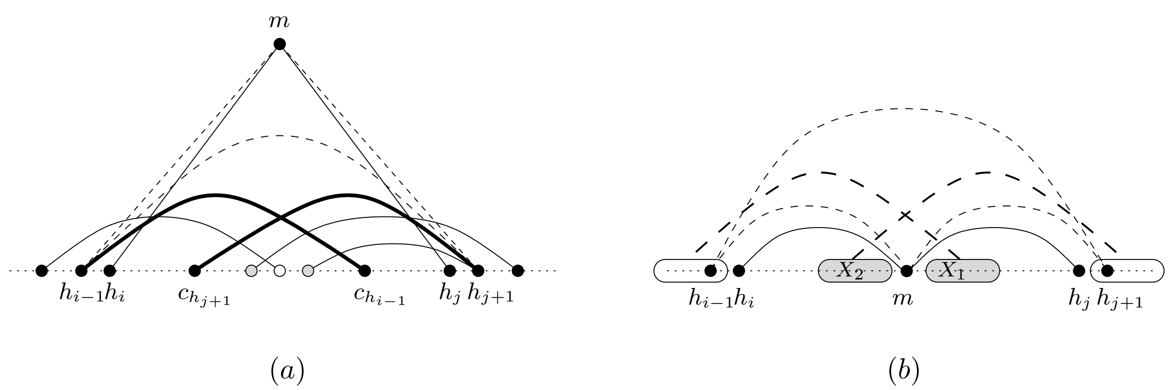


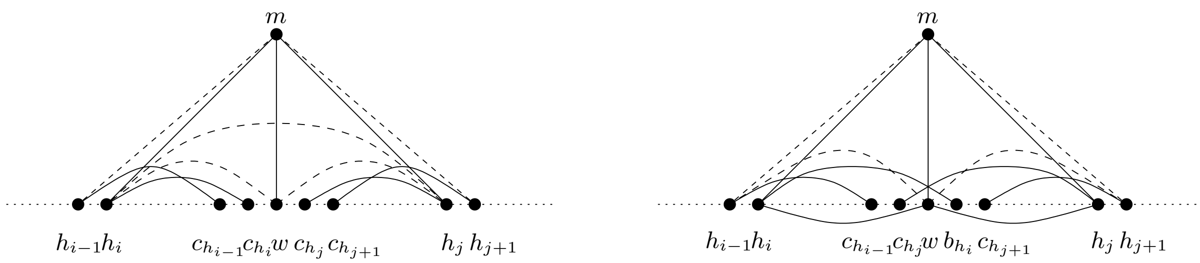
Figure 7: Illustration of the reordering applied to σ_H. The thin edges stand for edges of G. On the left, the gray vertices represent vertices of X_1 while the white vertex is a vertex of X_2.

Now, remark that there is no edge between X_1 and $\{w \in V(H) \ : \ w \leq_{\sigma_H} h_{i-1}\}$, that there is no edge between X_2 and $\{w \in V(H) \ : \ w \geq_{\sigma_H} h_{j+1}\}$), that there are still all the edges between $N_H(m)$ and $X_1 \cup X_2$ and that the edges between X_1 and $\{w \in V(H) \ : \ w \geq_{\sigma_H} h_{j+1}\}$ and the edges between X_2 and $\{w \in V(H) \ : \ w \leq_{\sigma_H} h_{i-1}\}$ are unchanged. So, it follows that the new ordering respects the umbrella property, and we are done.

Next, assume that $c_{h_{i-1}} <_{\sigma_H} c_{h_{j+1}}$. We let c_{h_i} (resp. c_{h_j}) be the neighbor of h_i (resp. h_j) with maximal (resp. minimal) index in $N_H(m)$. Notice that $c_{h_{i-1}} \leq_{\sigma_H} c_{h_i}$ and $c_{h_j} \leq_{\sigma_H} c_{h_{j+1}}$ (see Figure 8). Two cases may occur:

(i) First, assume that $c_{h_i} <_{\sigma_H} c_{h_j}$, case depicted in Figure 8 (a). In particular, this means that $h_ih_j \notin E(G)$. If c_{h_i} and c_{h_j} are consecutive in σ_H, then inserting m between c_{h_i} and c_{h_j} yields an umbrella ordering (since c_{h_j} (resp. c_{h_i}) does not have any neighbor before (resp. after) h_i (resp. h_j) in σ_H). Now, if there exists $w \in V(H)$ such that $c_{h_i} <_{\sigma_H} w <_{\sigma_H} c_{h_j}$, then one can see that the set $\{m, h_i, w, h_j\}$ forms a claw containing m in G, which is impossible.

(ii) The second case to consider is when $c_{h_j} \leq_{\sigma_H} c_{h_i}$. In such a case, one can see that m and the vertices of $\{w \in V(H) \ : \ c_{h_j} \leq_{\sigma_H} w \leq_{\sigma_H} c_{h_i}\}$ are true twins in $H \cup \{m\}$, because their common neighborhood is exactly $\{w \in V(H) \ : \ h_i \leq_{\sigma_H} w \leq_{\sigma_H} h_j\}$. Hence, inserting m just before c_{h_i} (or anywhere between c_{h_i} and c_{h_j} or just after c_{h_j}) yields an umbrella ordering.



(a) (b)

Figure 8: The possible cases for $c_{h_{i-1}} <_{\sigma_H} c_{h_{j+1}}$.

$\diamond$

As explained before, since the proof of Claim 2.9 does not use the fact that the vertices of H do not belong to M, it follows that we can iteratively insert the vertices of M into σ_H, preserving an umbrella ordering at each step. This concludes the proof of Lemma 2.4. $\square$

The complexity needed to compute Rule 2.4 will be discussed in the next section. The following observation results from the application of Rule 2.4 and from Section 2.1.2.

Observation 2.10. *Let $G = (V, E)$ be a positive instance of* Proper Interval Completion *reduced under Rules 2.2 to 2.4. Any K-join of G contains at most $2k + 2$ vertices which are not contained in any 4-cycle or claw of G.*

Proof. Let B be any K-join of G, and X be the set of vertices of B which are contained in a 4-cycle or a claw of G. As any subgraph of a K-join is a K-join, $B \setminus X$ is a clean K-join of G. Then, after having applied Rule 2.4, we have $|B \setminus X| \leq 2k + 2$. $\square$

2.1.4 Cutting the 1-branches

We now turn our attention to branches of a graph $G = (V, E)$, proving how they can be reduced.

Lemma 2.11. *Let $G = (V, E)$ be a connected graph which is a positive instance of* Proper Interval Completion*, and let B be a 1-branch of G associated with the umbrella ordering σ_B. Assume that $|B^R| \geq 2k + 1$ and let B_L be the $2k + 1$ last vertices of B^R according to σ_B. Then, there exists a k-completion F of G into a proper interval graph and a vertex $b \in B_L$ such that the umbrella ordering of $G + F$ preserves the order induced by σ_B on the set $B_b = \{w \in V(B) \ : \ b_1 \leq_{\sigma_B} w \leq_{\sigma_B} b_f\}$, where f is the maximal index in σ_B such that $bb_f \in E(G)$. Moreover, the vertices of B_b are the first in an umbrella ordering of $G + F$.*

Proof. Let F be any k-completion of G, $H = G + F$ and σ_H be the umbrella ordering of H. Since $|B_L| = 2k + 1$ and $|F| \leq k$, there exists a vertex $b \in B_L$ not incident to any added edge of F. We let N_b be the set of neighbors of b that are *after* b in σ_B, $B_b = \{w \in V(B) \ : \ b_1 \leq_{\sigma_B} w \leq_{\sigma_B} b_f\}$, where f is the maximal index in σ_B such that $bb_f \in E(G)$ (i.e. b_f is the last vertex of N_b), and $C = V \setminus B_b$ (see Figure 9, which depicts the case where $b_f \in B_1$, but $b_f \in B_L$ is possible too).

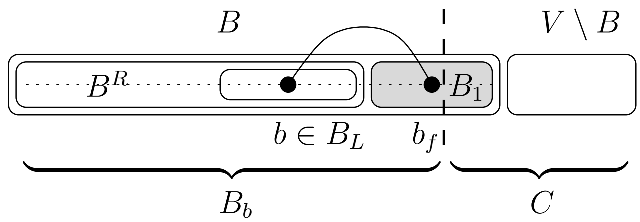


Figure 9: The different sets to cut a 1-branch (the set N_b, not shown in figure, is made by vertices lying under the edge bb_f, including b_f, not b).

Remark that, by the definition of the attachment clique B_1 (which is $B_1 = N_B(b_{|B|})$), we have $B_1 \nsubseteq B_b$ (because b is not a neighbor of $b_{|B|}$) and then $B \cap C \neq \emptyset$.

Claim 2.12. *We have:*
- *(i) $G[C]$ is a connected graph and*
- *(ii) Either for every vertex u of C we have $b <_{\sigma_H} u$ or for every vertex u of C we have $u <_{\sigma_H} b$.*

Proof. The first point follows from the fact that by definition of a 1-branch, every vertex of $V \setminus B$ which has a neighbor in B is a neighbor of $b_{|B|}$ (which belongs to C). So, as G is connected, every connected component of $G[V \setminus B]$ contains a neighbor of $b_{|B|}$. As, $C \cap B$ is a subset of the attachement clique B_1 and then linked to b_B, we conclude that $G[C]$ is a connected graph.
To see the second point, assume that there exist $u, v \in C$ such that w.l.o.g. $u <_{\sigma_H} b <_{\sigma_H} v$. Since $G[C]$ is a connected graph, there exists a path between u and v in G that avoids $N_G[b]$, which is equal to $N_H[b]$ since b is not incident to any edge of F. Hence there exist $u', v' \in C$, consecutive along this path, such that $u' <_{\sigma_H} b <_{\sigma_H} v'$ and $u'v' \in E(G)$. Then, as the neighborhood of b is the same in G than in H, we have $u'b, v'b \notin E(H)$, contradicting the fact that σ_H is an umbrella ordering for H. ◇

In the following, up to reversing the order σ_H, we assume that $b <_{\sigma_H} u$ holds for any $u \in C$. We will then find B_b at the beginning of σ_H. We now consider the following ordering σ of H: we first put the set B_b according to the order of B and then put the remaining vertices C according to σ_H (see Figure 10). We construct a coresponding completion F' of G from F as follows: we remove from F the edges with both extremities in B_b, and remove all edges between $B_b \setminus N_b$ and C. In other words, we set:

$$F' = F \setminus (F[B_b \times B_b] \cup F[(B_b \setminus N_b) \times C])$$

Finally, we inductively remove from F' any extremal edge of σ that belongs to F', and abusively still call F' the obtained edge set.

Claim 2.13. *The set F' is a k-completion of G.*

Proof. We prove that σ is an umbrella ordering of $H' = G+F'$. Since $|F'| \leq |F|$ by construction, the result will follow. Assume this is not the case. By definition of F', $H'[B_b]$ and $H'[C]$ induce proper interval graphs. This means that there exists a set of vertices $S = \{u, v, w\}$, $u <_\sigma v <_\sigma w$, intersecting both B_b and C and violating the umbrella property. We either have (1) $uw \in E, uv \notin E$ or (2) $uw \in E, vw \notin E$. Since neither F' nor G contain an edge between $B_b \setminus N_b$ and C, it follows that S intersects N_b and C. We study the different cases:

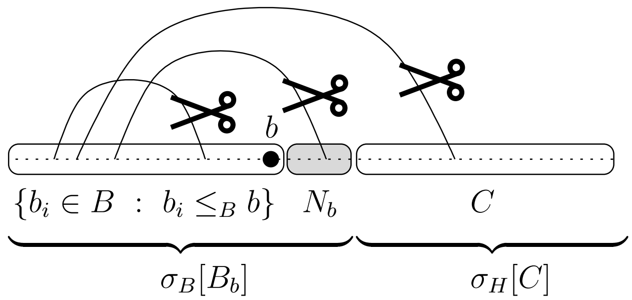


Figure 10: The construction of the ordering σ and the set F' (possible cut edges are from F).

(i) (1) holds and $u \in N_b,\ v, w \in C$: since the edge set between N_b and C is the same in H and H', it follows that $uv \notin E(H)$. Since σ_H is an umbrella ordering of H, we either have $v <_{\sigma_H} u <_{\sigma_H} w$ or $v <_{\sigma_H} w <_{\sigma_H} u$ (recall that C is in the same order in both σ and σ_H). Now, recall that $b <_{\sigma_H} \{v, w\}$ by assumption. In particular, since $bu \in E(G)$, this implies in both cases that σ_H is not an umbrella ordering, what leads to a contradiction.

(ii) (1) holds and $u, v \in N_b,\ w \in C$: this case cannot happen since N_b is a clique of H'.

(iii) (2) holds and $u \in N_b,\ v, w \in C$: this case is similar to (i). Observe that we may assume $uv \in E(H)$ (otherwise (i) holds). By construction of F', we have $vw \notin E(H)$ and hence $v <_{\sigma_H} w <_{\sigma_H} u$ or $v <_{\sigma_H} u <_{\sigma_H} w$. The former case contradicts the fact that σ_H is an umbrella ordering since $wu \in E(H)$. In the latter case, since σ_H is an umbrella ordering this means that $bv \in E(H)$ (as $bu \in E(H)$ and $v <_{\sigma_H} u <_{\sigma_H} w$). Since b is non affected vertex and $v \in C$, we have $bv \notin E(G)$, which leads to a contradiction.

(iv) (2) holds and $u, v \in N_b,\ w \in C$: first, if $uw \in E(G)$, then we have a contradiction since $N_C(u) \subseteq N_C(v)$. So, we have $uw \in F'$. By construction of F', we know that uw is not an extremal edge. Hence there exists an extremal edge (of G) above uw, which is either uw' with $w <_\sigma w'$, $u'w$ with $u' <_\sigma u$ or $u'w'$ with $u' <_\sigma u <_\sigma w <_\sigma w'$. The three situation are depicted in Figure 11. In the first case, $vw' \in E(G)$ (since $N_C(u) \subseteq N_C(v)$ in G) and hence we are in configuration (i) with vertex set $\{v, w, w'\}$. In the second case, $u'w \in E(G)$ and $vw \notin E(G)$ are in contradiction with $N_C(u') \subseteq N_C(v)$ in G (since $u' \in B_b$). Finally, in the third case, $vw' \in E(G)$ (since $N_C(u') \subseteq N_C(v)$ in G), and we are in configuration (i) with vertex set $\{v, w, w'\}$.

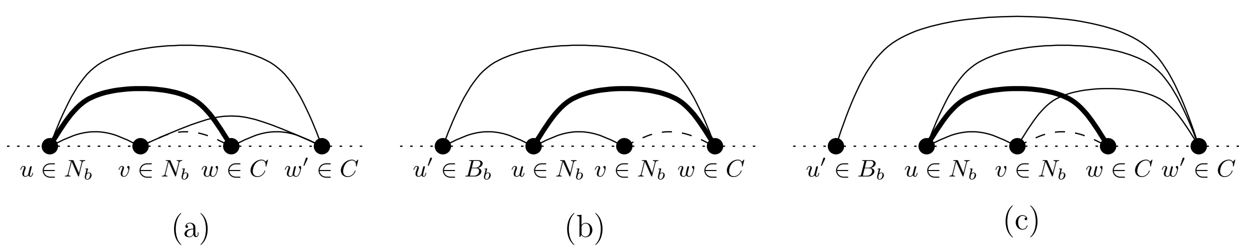


Figure 11: Illustration of the different cases of configuration (iv) (the bold edges belong to F').

$\diamond$

Altogether, we proved that there exists a k-completion of G associated with an umbrella ordering where the vertices of B_b are ordered in the same way than in the ordering of B and stand at the beginning of this ordering, what concludes the proof. $\square$

Rule 2.5 (1-branches)**.** *Let B be a 1-branch such that $|B^R| > 2k+1$. Remove $B^R \setminus B_L$ from G, where B_L denotes the $2k+1$ last vertices of B^R.*

Lemma 2.14. *Rule 2.5 is safe.*

Proof. Let $G' = G \setminus (B^R \setminus B_L)$ denote the reduced graph. Observe that any k-completion of G is a k-completion of G' since proper interval graphs are closed under induced subgraphs. So let F be a k-completion of G'. We denote by $H = G' + F$ the resulting proper interval graph and let σ_H be the corresponding umbrella ordering. Without loss of generality, we assume that the connected component of H containing B_L is the first according to σ_H. Remark that $B_1 \cup B_L$ forms a 1-branch of G', which we denote by B'. The umbrella ordering associated with B' is induced by σ_B. So, as previously, for a vertex b of B_L, we denote $\{wV(B') \ : \ b_1 \leq_{\sigma_{B'}} w \leq_{\sigma_{B'}} b_f\}$ by B_b. By Lemma 2.11 we know that there exists a vertex $b \in B_L$ such that the order of B_b in σ_H is the same than in $\sigma_{B'}$ and the vertices of B_b are the first of σ_H. Since $N_G(B^R \setminus B_L) \subseteq B_b$, it follows that the vertices of $B^R \setminus B_L$ can be inserted into σ_H while respecting the umbrella property. Hence, F is a k-completion for G, implying the result. $\square$

Here again, the time complexity needed to compute Rule 2.5 will be discussed in the next section. The following property of a reduced graph will be used to bound the size of our kernel.

Observation 2.15. *Let $G = (V, E)$ be a positive instance of* Proper Interval Completion *reduced under Rules 2.2 to 2.5. Every 1-branch of G contains at most $4k+3$ vertices which are not contained in any 4-cycle or claw of G.*

Proof. Let B be a 1-branch of a graph $G = (V, E)$ reduced under Rules 2.2 to 2.5. As B has been reduced under Rule 2.5, we know that $B \setminus B_1$ contains at most $2k+1$ vertices. Furthermore B_1 forms a K-join of G, and then, by Observation 2.10, contains at most $2k+2$ vertices which are not contained in any 4-cycle or claw of G. $\square$

2.1.5 Cutting the 2-branches

We now focus on 2-branches of the graph and explain how to reduce them. Let (G, k) be an instance of Proper Interval Completion and $B = \{b_1, \ldots, b_{|B|}\}$ be a 2-branch of G associated with the umbrella ordering σ_B. Recall that the attachment cliques of B are $B_1 = \{b \in V(B) \ : \ b_1 \leq_{\sigma_B} b \leq_{\sigma_B} b_{l'}\}$, where $b_{l'}$ is the neighbor of b_1 with maximal index in σ_B, and $B_2 = \{b \in V(B) \ : \ b_l \leq_{\sigma_B} b \leq_{\sigma_B} b_{|B|}\}$), where b_l is the neighbor of $b_{|B|}$ with minimal index in σ_B. Now, we define the next cliques in the 2-branch B (see Figure 12), namely $B_1' = \{b \in V(B) \ : \ b_{l'+1} \leq_{\sigma_B} b \leq_{\sigma_B} b_{\tilde{l}'}\}$, where $b_{\tilde{l}'}$ is the neighbor of $b_{l'+1}$ with maximal index in σ_B, and $B_2' = \{b \in V(B) \ : \ b_{\tilde{l}} \leq_{\sigma_B} b \leq_{\sigma_B} b_{l-1}\}$), where $b_{\tilde{l}}$ is the neighbor of b_{l-1} with minimal index in σ_B. Finaly, we denote by B_M the set $B \setminus (B_1 \cup B_1' \cup B_2' \cup B_2)$. Remark that by definition, we have $B^R = B_1' \cup B_M \cup B_2'$. Remark also that B_M could be empty if B is made with four K-join or less. However, we are interested in 2-branches B with B_M large enough, to reduce it.

Rule 2.6 (2-branches)**.** *Let G be a connected instance of* Proper Interval Completion *and B be a 2-branch such that $G[V \setminus B^R]$ is not connected. Assume that $|B_M| \geq 4k+2$ and let B_M^f be the set of the $2k+1$ vertices after B_1' according to σ_B and B_M^l be the set of the $2k+1$ vertices*

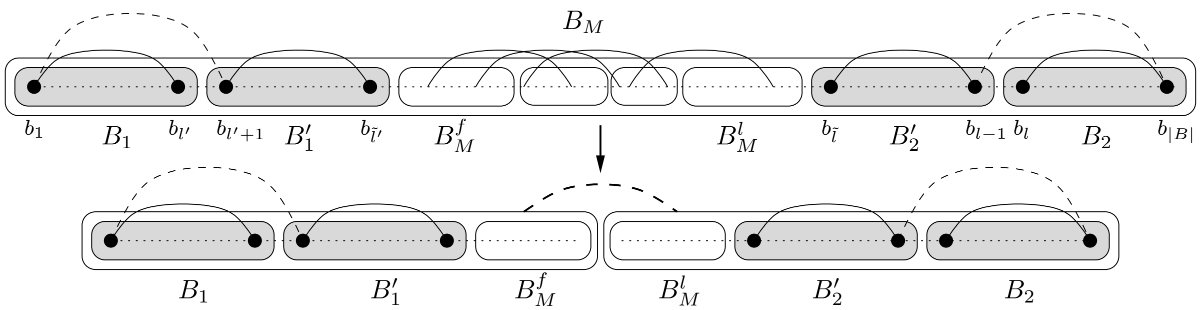


Figure 12: Applying Rule 2.6.

before B'_2 according to σ_B. Remove $B_M \setminus (B_M^f \cup B_M^l)$ from G (see Figure 12) and delete all the edges between $B'_1 \cup B_M^f$ and $B_M^l \cup B'_2$, if exist.

Lemma 2.16. *Rule 2.6 is safe.*

Proof. We denote by G' the reduced graph, and first remark that G' is no more a connected graph. Indeed, by assumption $G \setminus B^R$ is not connected and we denote by G_1 and G_2 its two connected components containing respectively B_1 and B_2. As B is a 2-branch, all the neighbors of G_1 in B stand in B'_1 (that is why we need B'_1). Similarly, all the neighbors of G_2 in B stand in B'_2. As, in G' we have removed all the edges between $B'_1 \cup B_M^f$ and $B_M^l \cup B'_2$, $G'_1 = G[G_1 \cup B'_1 \cup B_M^f]$ and $G'_2 = [B_M^l \cup B'_2 \cup G_2]$ form two connected components of G'.
Now, observe that any k-completion of G induces a k-completion of G'. Indeed, since proper interval graphs are closed under induced subgraphs, any k-completion of G induces a k_1-completion of G'_1 and a k_2-completion of G'_2 with $k_1 + k_2 \leq k$ and then a $k_1 + k_2$-completion of G'. Conversely, let F' be a k-completion of G'. We denote by F'_1 (resp. F'_2) the edges of F' the extremities of which lie in G'_1 (resp. G'_2). Then, remark that $B_1 \cup B'_1 \cup B_M^f$ forms a 1-branch of G'_1 (and then of G') with attachment clique B_1 and with $|B'_1 \cup B_M^f| \geq 2k+1$ (that is why we need B_M^f). So, by Lemma 2.11, there exist a k_1-completion F'_1 of G'_1 with $k_1 \leq |F_1|$ and a vertex $b_1 \in B'_1 \cup B_M^f$ such that B_{b_1}, which is the set of vertices of $B'_1 \cup B_M^f$ which are neighbors of b_1 or lie after b_1 according to σ_B, is in the same order in σ_B than in an umbrella ordering of $G'_1 + F'_1$, and say, at the end of this ordering. Similarly, there exist a k_2-completion F'_2 of G'_2 with $k_2 \leq |F_2|$ and a vertex $b_2 \in B'_2 \cup B_M^l$ such that B_{b_2}, which is the set of vertices of $B'_2 \cup B_M^2$ which are neighbors of b_2 or lie before b_2 according to σ_B, is in the same order in σ_B than in an umbrella ordering of $G'_2 + F'_2$, and say, at the beginning of this ordering. Now, we can insert back the vertices and edges removed from G to obtain G'. Indeed, as B is 2-branch, the neighbors of $B_M \setminus (B_M^f \cup B_M^l)$ in $G'_1 \cup G'_2$ are in $B_{b_1} \cup B_{b_2}$, and similarly the removed edge between $B'_1 \cup B_M^f$ and $B_M^l \cup B'_2$ have their extremities in $B_{b_1} \cup B_{b_2}$. So, as B_{b_1} and B_{b_2} lie as in σ_B, we can put back the removed edges and vertices in order to obtain a $k'_1 + k'_2$-completion of G, with $k'_1 + k'_2 \leq k$. ☐

The following observation bounds the number of vertices in a 2-branch of a positive instance of Proper Interval Completion.

Observation 2.17. *Let $G = (V, E)$ be a connected positive instance of* Proper Interval Completion*, reduced under Rules 2.2 to 2.6, and B be a 2-branch of G such that $G[C \setminus B^R]$ is not connected, where C is the connected component of G containing B. Then B contains at most $12k + 10$ vertices which are not contained in any 4-cycle or claw of G.*

Proof. Let B be a 2-branch of G,reduced under Rules 2.2 to 2.6, and C be the connected component containing B. The sets B_1, B_1', B_2' and B_2 form four K-joins of G, and then by Observation 2.10, they contain in all at most $4.(2k+2) = 8k+8$ vertices which are not contained in any 4-cycle or claw of G. Furthermore, if $G[C \setminus B^R]$ is not connected, then, as G is reduced under Rule 2.6, $B \setminus (B_1 \cup B_1' \cup B_2' \cup B_2)$ contains at most $4k+2$ vertices, what provides the announced bound. □

2.2 Detecting the branches

We now turn our attention to the complexity needed to compute reduction rules 2.4 to 2.6. Mainly, we indicate how to obtain the maximum branches in order to reduce them. The detection of a branch is straightforward except for the attachment cliques, where several choices are possible.
So, first, we detect the maximum 1-branches of G. Remark that for every vertex x of G, the set $\{x\}$ is a 1-branch of G. The next lemma indicates how to compute a maximum 1-branch that contains a fixed vertex x as first vertex.

Lemma 2.18. *Let $G = (V, E)$ be a graph and x a vertex of G. In time $O(nm)$, it is possible to detect a maximum 1-branch of G containing x as first vertex.*

Proof. To detect such a 1-branch, we design an algorithm which has two parts. Roughly speaking, we first try to detect the set B^R of a 1-branch B containing x. We set $B_0^R = \{x\}$ and $\sigma_0 = x$. Once B_{i-1}^R has been defined, we construct the set C_i of vertices of $G\setminus(\cup_{l=1}^{i-1} B_l^R)$ that are adjacent to at least one vertex of B_{i-1}^R. Two cases can appear. First, assume that C_i is a clique and that it is possible to order the vertices of C_i such that for every $1 \leqslant j < |C_i|$, we have $N_{B_{i-1}^R}(c_{j+1}) \subseteq N_{B_{i-1}^R}(c_j)$ and $(N_G(c_j) \setminus B_{i-1}^R) \subseteq (N_G(c_{j+1}) \setminus B_{i-1}^R)$. In this case, the vertices of C_i correspond to a new K-join of the searched 1-branch (remark that, along this inductive construction, there is no edge between C_i and $\cup_{l=1}^{i-2} B_l^R$). So, we let $B_i^R = C_i$ and σ_i be the concatenation of σ_{i-1} and the ordering defined on C_i. In the other case, such an ordering of C_i can not be found, meaning that while detecting a 1-branch B, we have already detected the vertices of B^R and at least one (possibly more) vertex of the attachment clique B_1 with neighbors in B^R. Assume that the process stops at step p and let C be the set of vertices of $G \setminus \cup_{l=1}^{p} B_l^R$ which have neighbors in $\cup_{l=1}^{p} B_l^R$ and $B_1' \subseteq B_p^R$ be the set of vertices that are adjacent to all the vertices of C. Remark that $B_1' \neq \emptyset$, as B_1' contains at least the last vertex of σ_p. We denote by B^R the set $(\cup_{l=1}^{p} B_l^R) \setminus B_1'$ and we will construct the largest K-join containing B_1' in $G \setminus B^R$ which is compatible with σ_p, in order to define the attachment clique B_1 of the desired 1-branch. The vertices of C are the candidates to complete the attachment clique. On C, we define the following oriented graph: there is an arc from u to v if: uv is an edge of G, $N_{B^R}(v) \subseteq N_{B^R}(u)$ and $N_{G\setminus B^R}[u] \subseteq N_{G\setminus B^R}[v]$. This graph can be computed in time $O(nm)$. Now, it is easy to check that the obtained oriented graph is a transitive graph, in which the equivalent classes are made of true twins in G. A path in this oriented graph corresponds, by definition, to a K-join containing B_1' and compatible with σ_p. As it is possible to compute a longest path in linear time in this oriented graph, we obtain a maximum 1-branch of G that contains x as first vertex. □

So, we detect all the maximum 1-branches of G in time $O(n^2m)$.
Now, to detect the 2-branches, we first detect for all pairs of vertices a maximum K-join with these vertices as ends. More precisely, if $\{x, y\}$ are two vertices of G linked by an edge, then $\{x, y\}$ is a K-join of G, with $N = N_G(x) \cap N_G(y)$, $L = N_G(x) \setminus N_G[y]$ and $R = N_G(y) \setminus N_G[x]$. So, there exist K-joins with x and y as ends, and we will compute such a K-join with maximum cardinality.

Lemma 2.19. *Let $G = (V, E)$ be a graph and x and y two adjacent vertices of G. It is possible to compute in $O(nm)$ time a maximum (in cardinality) K-join that admits x and y as ends.*

Proof. We denote $N_G[x] \cap N_G[y]$ by N, $N_G(x) \setminus N_G[y]$ by L and $N_G(y) \setminus N_G[x]$ by R. Let us denote by N' the set of vertices of N that contains N in their closed neighborhood. The vertices of N' are the candidates to belong to the desired K-join, and we can identify them in time $O(n^2)$. Now, we construct on N' an oriented graph D, putting, for every vertices u and v of N', an arc from u to v if: $N_G(v) \cap L \subseteq N_G(u) \cap L$ and $N_G(u) \cap R \subseteq N_G(v) \cap R$. Basically, it could take a $O(n)$ time to decide if there is an arc from u to v or not, and so the whole oriented graph could be computed in time $O(n.|N'|^2)$. As N' is a clique of G, we have $|N'|^2 = O(m)$. Now, it is easy to check that the obtained oriented graph is a transitive graph in which the equivalent classes are made of true twins in G. In this oriented graph, it is possible to compute a longest path from x to y in linear time. Such a path corresponds to a maximal K-join that admits x and y as ends. It follows that the desired K-join can be identified in $O(nm)$ time. □

Now, for every edge xy of G, we compute a maximum K-join that contains x and y as ends and a reference to all the vertices that this K-join contains. This computation takes a $O(nm^2)$ time and gives, for every vertex, some maximum K-joins that contain this vertex. These K-joins will be useful to compute the 2-branches of G, in particular through the next lemma.

Lemma 2.20. *Let B be a 2-branch of G with $B^R \neq \emptyset$, and x a vertex of B^R. Then, for every maximal (by inclusion) K-join B' that contains x there exists an extremal edge uv of σ_B such that $B' = \{w \in B \ : \ u \leq_{\sigma_B} w \leq_{\sigma_B} v\}$.*

Proof. As usually, we denote by L, R and C the partition of $G \setminus B$ associated with B and by σ_B the umbrella ordering associated with B. Let B' be a maximal K-join that contains x and define by b_f (resp. b_l) the first (resp. last) vertex of B' according to σ_B. As there is no edge between $\{u \in B \ : \ u <_{\sigma_B} b_f\} \cup L \cup C$ and b_l and no edge between $\{u \in B \ : \ b_l <_{\sigma_B} u\} \cup R \cup C$ and b_f, we have $B' \subseteq \{u \in B \ : \ b_f \leq_{\sigma_B} u \leq b_l\}$. Furthermore, as $\{u \in B \ : \ b_f \leq_{\sigma_B} u \leq b_l\}$ is a K-join and B' is maximal, we have $B' = \{u \in B \ : \ b_f \leq_{\sigma_B} u \leq b_l\}$. Now, if $b_f b_l$ was not an extremal edge of σ_B, it would be possible to extend B', contradicting the maximality of B'. □

Now, we can detect the 2-branches B with a set B^R non empty.

Lemma 2.21. *Let $G = (V, E)$ be a graph, x a vertex of G and B' a given maximal K-join that contains x. There is a $O(nm)$ time algorithm to decide if there exists a 2-branch B of G which contains x as a vertex of B^R, and if it exists, to find a maximum 2-branch with this property.*

Proof. By Lemma 2.20, if there exists a 2-branch B of G which contains x as a vertex of B^R, then B' corresponds to a set $\{u \in B \ : \ b_f \leq_{\sigma_B} u \leq_{\sigma_B} b_l\}$ where $b_f b_l$ is an extremal edge of B. We denote by L', R' and C' the usual partition of $G \setminus B'$ associated with B', and by $\sigma_{B'}$ the umbrella ordering of B'. In G, we remove the set of vertices $\{u \in B' \ : \ u <_{\sigma_{B'}} x\}$ and the edges between L' and $\{u \in B' \ : \ x \leq_{\sigma_{B'}} u\}$ and denote by H_1 the resulting graph. From the definition of the 2-branch B, $\{u \in B \ : \ x \leq_{\sigma_B} u\}$ is a 1-branch of H_1 that contains x as first vertex. So, using Lemma 2.18, we find a maximal 1-branch B_1 that contains x as first vertex. Remark that B_1 has to contain $\{u \in B \ : \ x \leq_{\sigma_B} u\} \cap B^R$ at its beginning. Similarly, we define H_2 from G by removing the vertex set $\{u \in B' \ : \ x <_{\sigma_{B'}} u\}$ and the edges between R' and $\{u \in B' \ : \ u \leq_{\sigma_{B'}} x\}$. We detect in H_2 a maximum 1-branch B_2 that contains x as last vertex, and as previously, B_2 has to contain $\{u \in B \ : \ u \leq_{\sigma_B} x\} \cap B^R$ at its end. So, $B_1 \cup B_2$ forms a maximum 2-branch of G containing x. □

We would like to mention that it could be possible to improve the execution time of our detecting branches algorithm, using possibly more involved techniques (as for instance, inspired from [7]).

However, this is not our main objective here.
Anyway, using the $O(n^2m)$ time algorithm explained in Lemma 2.1 to localize all the 4-cycles and the claws, we obtain the following result.

Lemma 2.22. *Given a graph $G = (V, E)$, the reduction rules 2.4 to 2.6 can be carried out in polynomial time, namely in time $O(nm(n + m))$.*

2.3 Kernelization algorithm

We are now ready to the state the main result of this Section. The kernelization algorithm consists of an exhaustive application of Rules 2.1 to 2.6.

Theorem 2.23. *The* Proper Interval Completion *problem admits a kernel with $O(k^3)$ vertices, computable in time $O(nm(n + m))$.*

Proof. Let $G = (V, E)$ be a positive instance of Proper Interval Completion reduced under Rules 2.1 to 2.6. Let F be a k-completion of G, $H = G + F$ and σ_H be the umbrella ordering of H. Since $|F| \leq k$, G contains at most $2k$ *affected* vertices (i.e. incident to an added edge). Let $A = \{a_1 <_{\sigma_H} \ldots <_{\sigma_H} a_i <_{\sigma_H} \ldots <_{\sigma_H} a_p\}$ be the set of such vertices, with $p \leq 2k$. The size of the kernel is due to the following observations, which we admit without proof (see Figure 13).

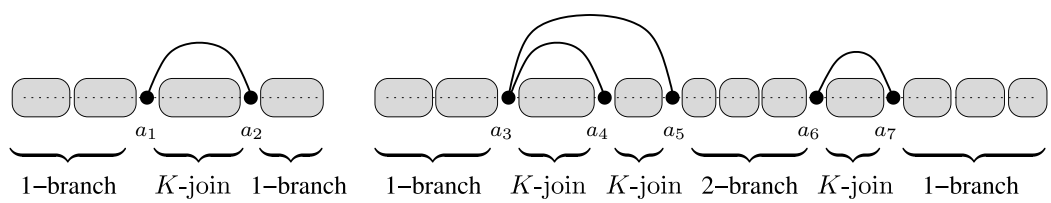


Figure 13: Illustration of the size of the kernel. The figure represents the graph $H = G + F$, the a_i's are the *affected* vertices, and the bold edges are edges of F.

Between two consecutive affected vertices a_i and a_{i+1}, the interval of vertices of G, denoted by I, forms:

- Either a K-join, if I lies under an edge of F. For instance, on Figure 13, it corresponds to intervals of vertices between a_1 and a_2, or between a_3 and a_4, or between a_4 and a_5 or between a_6 and a_7. So, by Observation 2.10, we know that such a I contains at most $2k + 2$ vertices which are not contained in any claw or 4-cycle of G.

- Either a 1-branch or two disjoint 1-branch. If I lies at the beginning or at the end of σ_H, then I forms a 1-branch (for instance, on Figure 13, it corresponds to intervals of vertices before a_1 or after a_7). If I lies between two vertices a_i and a_{i+1} which are respectively the last (according to σ_H) of a connected component of G and the first (according to σ_H) of another connected component of G, then I forms two disjoint 1-branches (for instance, on Figure 13, it corresponds to the interval of vertices between a_2 and a_3). So, by Observation 2.15, we know that such a I contains at most $2.(4k + 3) = 8k + 6$ vertices which are not contained in any claw or 4-cycle of G.

- Or a 2-branch, if I lies between two vertices a_i and a_{i+1} which belongs to the same connected component C of G and such that there is no edge of F standing above I. In this case the

2-branch B forms by the vertices of I is such that $G[C \setminus B^R]$ is not connected, and then by Observation 2.17, we know that I contains at most $12k+10$ vertices which are not contained in any claw or 4-cycle of G.

Finaly, as there is at most $2k+1$ such intervals I, the graph H (and hence G) contains at most $(2k+1).(12k+10)$ vertices different from the a_i's and which are not contained in any claw or 4-cycle of G. Moreover, by Lemma 2.2, there is at most $4k^3+15k^2+16k$ vertices of G contained in any claw or 4-cycle. Alltogether, G contains at most $4k^3+15k^2+16k+(2k+1).(12k+10)+2k+1 = 4k^3+39k^2+50k+11$ vertices, which implies the claimed $O(k^3)$ bound. The complexity directly follows from Lemma 2.22. □

3 A special case: Bi-clique Chain Completion

Bipartite chain graphs are defined as bipartite graphs whose parts are connected by a join. Equivalently, they are known to be the graphs that do not admit any $\{2K_2, C_5, K_3\}$ as an induced subgraph [31] (see Figure 14). In [13], Guo proved that the so-called Bipartite Chain Deletion With Fixed Bipartition problem, where one is given a *bipartite* graph $G=(V,E)$ and seeks a subset of E of size at most k whose deletion from E leads to a bipartite chain graph, admits a kernel with $O(k^2)$ vertices. We define *bi-clique chain graph* to be the graphs formed by two disjoint cliques linked by a join. They correspond to interval graphs that can be covered by two cliques. Since the complement of a bipartite chain graph is a bi-clique chain graph, this result also holds for the Bi-clique Chain Completion With Fixed Bi-clique Partition problem. Using similar techniques than in Section 2, we prove that when the bipartition is not fixed, both problems admit a quadratic-vertex kernel. For the sake of simplicity, we consider the completion version of the problem, defined as follows.

Bi-clique Chain Completion:
Input: A graph $G=(V,E)$ and a positive integer k.
Parameter: k.
Output: A set $F \subseteq (V \times V) \setminus E$ of size at most k such that the graph $H=(V, E \cup F)$ is a bi-clique chain graph.

It follows from definition that bi-clique chain graphs do not admit any $\{C_4, C_5, 3K_1\}$ as an induced subgraph, where a $3K_1$ is an independent set of size 3 (see Figure 14). Observe in particular that bi-clique chain graphs are proper interval graphs, and hence admit an umbrella ordering.

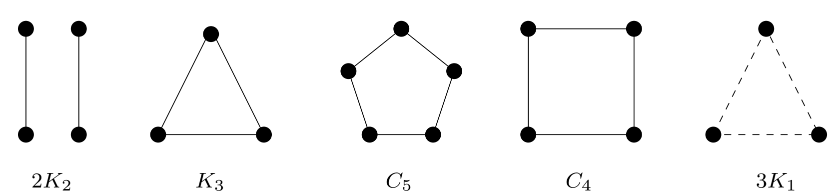


Figure 14: The forbidden induced subgraphs for bipartite and bi-clique chain graphs.

We provide a kernelization algorithm for the Bi-clique Chain Completion problem which follows the same lines that the one in Section 2.

Rule 3.1 (Sunflower)**.** *Let $\mathcal{S} = \{C_1, \ldots, C_m\}$, $m > k$ be a set of $3K_1$ having two vertices u, v in common but distinct third vertex. Add uv to F and decrease k by 1.*
Let $\mathcal{S} = \{C_1, \ldots, C_m\}$, $m > k$ be a set of distinct 4-cycles having a non-edge uv in common. Add uv to F and decrease k by 1.

The following result is similar to Lemma 2.2.

Lemma 3.1. *Let $G = (V, E)$ be a positive instance of* Bi-clique Chain Completion *on which Rule 3.1 has been applied. There are at most k^2+2k vertices of G contained in $3K_1$'s. Furthermore, there at most $2k^2 + 2k$ vertices of G that are vertices of a 4-cycle.*

We say that a K-join is *simple* whenever $L = \emptyset$ or $R = \emptyset$. In other words, a simple K-join consists in a clique connected to the rest of the graph by a join. We will see it as a 1-branch which is a clique and use for it the classical notation devoted to the 1-branch. Moreover, we (re)define a *clean K-join* as a K-join whose vertices do not belong to any $3K_1$ or 4-cycle. The following reduction rule is similar to Rule 2.4, the main ideas are identical, only some technical arguments change. Anyway, to be clear, we give the proof in all details.

Rule 3.2 (K-join)**.** *Let B be a simple clean K-join of size at least $2(k + 1)$ associated with an umbrella ordering σ_B. Let B_L (resp. B_R) be the $k + 1$ first (resp. last) vertices of B according to σ_B, and $M = B \setminus (B_L \cup B_R)$. Remove the set of vertices M from G.*

Lemma 3.2. *Rule 3.2 is safe and can be computed in polynomial time.*

Proof. Let $G' = G \setminus M$. Observe that any k-completion of G is a k-completion of G' since bi-clique chain graphs are closed under induced subgraphs. So, let F be a k-completion for G'. We denote by $H = G' + F$ the resulting bi-clique chain graph and by σ_H an umbrella ordering of H. We prove that we can always insert the vertices of M into σ_H and modify it if necessary, to obtain an umbrella ordering of a bi-clique chain graph for G without adding any edge. This will imply that F is a k-completion for G. To see this, we need the following structural property of G. As usual, we denote by R the neighbors in $G \setminus B$ of the vertices of B, and by C the vertices of $G \setminus (R \cup B)$. For the sake of simplicity, we let $N = \cap_{b \in B} N_G(b) \setminus B$, and remove the vertices of N from R. We abusively still denote by R the set $R \setminus N$, see Figure 15.

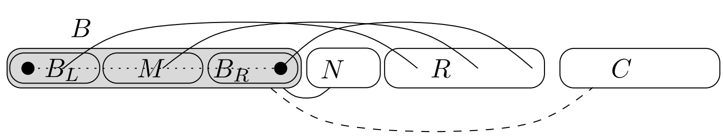


Figure 15: The K-join decomposition for the Bi-clique Chain Completion problem.

Claim 3.3. *The set $R \cup C$ is a clique of G.*

Proof. Observe that no vertex of R is a neighbor of b_1, since otherwise such a vertex must be adjacent to all the vertices of B and then must stand in N. So, if $R \cup C$ contains two vertices u, v such that $uv \notin E$, we form the $3K_1$ $\{b_1, u, v\}$, contradicting the fact that B is clean. $\diamond$

The following observation comes from the definition of a simple K-join.

Observation 3.4. *Given any vertex $r \in R$, if $N_B(r) \cap B_L \neq \emptyset$ holds then $M \subseteq N_B(r)$.*

We use these facts to prove that an umbrella ordering of a bi-clique chain graph can be obtained for G by inserting the vertices of M into σ_H. Let b_f, b_l be the first and last vertex of $B \setminus M$ appearing in σ_H, respectively. We let B_H denote the set $\{u \in V(H) \ : \ b_f <_{\sigma_H} u <_{\sigma_H} b_l\}$. Now, we modify σ_H by ordering the twins in H according to their neighborhood in M: if x and y are twins in H, are consecutive in σ_H, verify $x <_{\sigma_H} y <_{\sigma_H} b_f$ and $N_M(y) \subset N_M(x)$, then we exchange x and y in σ_H. This process stops when the considered twins are ordered following the join between $\{u \in V(H) \ : \ u <_{\sigma_H} b_f\}$ and M. We proceed similarly on the right of B_H, i.e. for x and y consecutive twins with $b_l <_{\sigma_H} x <_{\sigma_H} y$ and $N_M(x) \subset N_M(y)$. The obtained order is clearly an umbrella ordering of a bi-clique chain graph too (in fact, we just re-labeled some vertices in σ_H, and we abusively still denote it by σ_H).

Claim 3.5. *The set $B_H \cup \{m\}$ is a clique of G for any $m \in M$, and consequently $B_H \cup M$ is a clique of G.*

Proof. Let u be any vertex of B_H.We claim that $um \in E(G)$. Observe that if $u \in B$ then the claim trivially holds. So, assume that $u \notin B$. By definition of σ_H, B_H is a clique in H since $b_f b_l \in E(G)$. It follows that u is incident to every vertex of $B \setminus H$ in H. Since B_L contains $k+1$ vertices, it follows that $N_G(u) \cap B_L \neq \emptyset$. Hence, u belongs to $N \cup R$ and $um \in E$ by Observation 2.6. ◇

Claim 3.6. *Let m be any vertex of M and σ'_H be the ordering obtained from σ_H by removing B_H and inserting m to the position of B_H. The ordering σ'_H respects the umbrella property.*

Proof. Assume that σ'_H does not respect the umbrella property, i.e. that there exist (w.l.o.g.) two vertices $u, v \in H \setminus B_H$ such that either (1) $u <_{\sigma'_H} v <_{\sigma'_H} m$, $um \in E(H)$ and $uv \notin E(H)$ or (2) $u <_{\sigma'_H} m <_{\sigma'_H} v$, $um \notin E(H)$ and $uv \in E(H)$ or (3) $u <_{\sigma'_H} v <_{\sigma'_H} m$, $um \in E(H)$ and $vm \notin E(H)$. First, assume that (1) holds. Since $uv \notin E$ and σ_H is an umbrella ordering, $uw \notin E(H)$ for any $w \in B_H$, and hence $uw \notin E(G)$. This means that $B_R \cap N_G(u) = \emptyset$, which is impossible since $um \in E(G)$. If (2) holds, since $uv \in E(H)$ and σ_H is an umbrella ordering of H, we have $B_H \subseteq N_H(u)$. In particular, $B_L \subseteq N_H(u)$ holds, and as $|B_L| = k+1$, we have $B_L \cap N_G(u) \neq \emptyset$ and um should be an edge of G, what contradicts the assumption $um \notin E(H)$. So, (3) holds, and we choose the first u satisfying this property according to the order given by σ'_H. So we have $wm \notin E(G)$ for any $w <_{\sigma'_H} u$. Similarly, we choose v to be the first vertex satisfying $vm \notin E(G)$. Since $um \in E(G)$, we know that u belongs to $N \cup R$. Moreover, since $vm \notin E(G)$, $v \in R \cup C$. There are several cases to consider:

(i) $u \in N$: in this case we know that $B \subseteq N_G(u)$, and in particular that $ub_l \in E(G)$. Since σ_H is an umbrella ordering for H, it follows that $vb_l \in E(H)$ and that $B_L \subseteq N_H(v)$. Since $|B_L| = k+1$ we know that $N_G(v) \cap B_L \neq \emptyset$ and hence $v \in R$. It follows from Observation 2.6 that $vm \in E(G)$.

(ii) $u \in R, v \in R \cup C$: in this case $uv \in E(G)$, by Claim 3.3, but u and v are not true twins in H (otherwise v would be placed before u in σ_H due to the modification we have applied to σ_H). This means that there exists a vertex $w \in V(H)$ that *distinguishes* u from v in H.

Assume first that $w <_{\sigma_H} u$ and that $uw \in E(H)$ and $vw \notin E(H)$. We choose the first w satisfying this according to the order given by σ'_H. Since $vm, wm, vw \notin E(H)$, it follows that $\{v, w, m\}$ defines a $3K_1$ of G, which cannot be since B is clean. Hence we can assume that for any $w'' <_{\sigma_H} u$, $uw'' \in E(H)$ implies that $vw'' \in E(H)$. Now, suppose that $b_l <_{\sigma_H} w$ and $uw \notin E(H)$, $vw \in E(H)$. In particular, this means that $B_L \subseteq N_H(v)$. Since $|B_L| = k+1$

we have $N_G(v) \cap B_L \neq \emptyset$, implying $vm \in E(G)$ (Observation 2.6). Assume now that $v <_{\sigma_H} w <_{\sigma_H} b_f$. In this case, since $uw \notin E(H)$, $B \cap N_H(u) = \emptyset$ holds and hence $B \cap N_G(u) = \emptyset$, which cannot be since $u \in R$. Finally, assume that $w \in B_H$ and choose the last vertex w satisfying this according to the order given by σ'_H (i.e. $vw' \notin E(H)$ for any $w <_{\sigma_H} w'$ and $w' \in B_H$). If $vw \in E(G)$ then $\{u, m, w, v\}$ is a 4-cycle in G containing a vertex of B, which cannot be (recall that $B_H \cup \{m\}$ is a clique of G by Claim 2.7). Hence $vw \in F$ and there exists an extremal edge above vw. The only possibility is that this edge is some edge $u'w$ for some u' with $u' \in V(H)$, $u <_{\sigma_H} u' <_{\sigma_H} v$ and $u'w \in E(G)$. By the choice of v we know that $u'm \in E(G)$. Moreover, by the choice of w, observe that u' and v are true twins in H (if a vertex s distinguishes u' and v in H, s cannot be before u, since otherwise s would distinguish u and v, and not before w, by choice of w). This leads to a contradiction because v should have been placed before u through the modification we have applied to σ_H. $\diamond$

Claim 3.7. *Every vertex $m \in M$ can be added to the graph H while preserving an umbrella ordering.*

Proof. Let m be any vertex of M. The graph H is a bi-clique chain graph. So, we know that in its associated umbrella ordering $\sigma_H = b_1, \ldots, b_{|H|}$, there exists a vertex b_i such that $H_1 = \{b_1, \ldots, b_i\}$ and $H_2 = \{b_{i+1}, \ldots, b_{|H|}\}$ are two cliques of H linked by a join. We study the behavior of B_H according to the partition (H_1, H_2).

(i) Assume first that $B_H \subseteq H_1$ (the case $B_H \subseteq H_2$ is similar). We claim that the set $H_1 \cup \{m\}$ is a clique. Indeed, let $v \in H_1 \setminus B_H$: since H_1 is a clique, $B_H \subseteq N_H(v)$ and hence $N_G(v) \cap B_L \neq \emptyset$. In particular, this means that $vm \in E(G)$ by Observation 3.4. Since $B_H \cup \{m\}$ is a clique by Claim 3.5, the result follows. Now, let u be the neighbor of m with maximal index in σ_H, and b_u the neighbor of u with minimal index in σ_H. Observe that we may assume $u \in H_2$ since otherwise $N_H(m) \cap H_2 = \emptyset$ and hence we insert m at the beginning of σ_H. First, if $b_u \in H_1$, we prove that the order σ_m obtained by inserting m directly before b_u in σ_H yields an umbrella ordering of a bi-clique chain graph. Since $H_1 \cup \{m\}$ is a clique, we only need to show that $N_{H_2}(v) \subseteq N_{H_2}(m)$ for any $v \leq_{\sigma_m} b_u$ and $N_{H_2}(m) \subseteq N_{H_2}(w)$ for any $w \in H_2$ with $w \geq_{\sigma_m} b_u$. Observe that by Claim 3.6 the set $\{w \in V \ : \ m \leq_{\sigma_m} w \leq_{\sigma_m} u\}$ is a clique. Hence the former case holds since $vu' \notin E(G)$ for any $v \leq_{\sigma_m} b_u$ and $u' \geq_{\sigma_m} u$. The latter case also holds since $N_H(m) \subseteq N_H(b_u)$ by construction. Finally, if $b_u \in H_2$, then $b_u = b_{|H_1|+1}$ since H_2 is a clique. Hence, using similar arguments one can see that inserting m directly after $b_{|H_1|}$ in σ_H yields an umbrella ordering of a bi-clique chain graph.

(ii) Assume now that $B_H \cap H_1 \neq \emptyset$ and $B_H \cap H_2 \neq \emptyset$. In this case, we claim that $H_1 \cup \{m\}$ or $H_2 \cup \{m\}$ is a clique in H. Let u and u' be the neighbors of m with minimal and maximal index in σ_H, respectively. If $u = b_1$ or $u' = b_{|H|}$ then Claims 3.5 and 3.6 imply that $H_1 \cup \{m\}$ or $H_2 \cup \{m\}$ is a clique and we are done. So, none of these two conditions hold and $mb_1 \notin E(H)$ and $mb_{|H|} \notin E(H)$ Then, by Claim 3.6, we know that $b_1 b_{|H|}$ and the set $\{b_1, b_{|H|}, m\}$ defines a $3K_1$ containing m in G, which cannot be. This means that we can assume w.l.o.g. that $H_1 \cup \{m\}$ is a clique, and we can conclude using similar arguments than in (i).

$\diamond$

Since the proof of Claim 3.7 does not use the fact that the vertices of H do not belong to M, it follows that we can iteratively insert the vertices of M into σ_H, preserving an umbrella ordering at each step. To conclude, observe that the reduction rule can be computed in polynomial time using Lemma 2.19. $\square$

Observation 3.8. *Let $G = (V, E)$ be a positive instance of* Bi-clique Chain Completion *reduced under Rule 3.2. Any simple K-join B of G has size at most $3k^2 + 6k + 2$.*

Proof. Let B be any simple K-join of G, and assume $|B| > 3k^2 + 6k + 2$. By Lemma 3.1 we know that at most $3k^2 + 2k$ vertices of B are contained in a $3K_1$ or a 4-cycle. Hence B contains a set B' of at least $2k + 3$ vertices not contained in any $3K_1$ or a 4-cycle. Now, since any subset of a K-join is a K-join, it follows that B' is a *clean* simple K-join. Since G is reduced under rule 3.2, we know that $|B'| \leq 2(k + 1)$ what gives a contradiction. $\square$

Finally, we can prove that Rules 3.1 and 3.2 form a kernelization algorithm.

Theorem 3.9. *The* Bi-clique Chain Completion *problem admits a kernel with $O(k^2)$ vertices.*

Proof. Let $G = (V, E)$ be a positive instance of Bi-clique Chain Completion reduced under Rules 3.1 and 3.2, and F be a k-completion for G. We let $H = G + F$ and H_1, H_2 be the two cliques of H. Observe in particular that H_1 and H_2 both define simple K-joins. Let A be the set of affected vertices of G. Since $|F| \leq k$, observe that $|A| \leq 2k$. Let $A_1 = A \cap H_1$, $A_2 = A \cap H_2$, $A'_1 = H_1 \setminus A_1$ and $A'_2 = H_2 \setminus A_2$ (see Figure 16). Observe that since H_1 is a simple K-join in H, $A'_1 \subseteq H_1$ is a simple K-join of G (recall that the vertices of A'_1 are not affected). By Observation 3.8, it follows that $|A'_1| \leq 3k^2 + 6k + 2$. The same holds for A'_2 and H contains at most $2(3k^2 + 6k + 2) + 2k$ vertices.

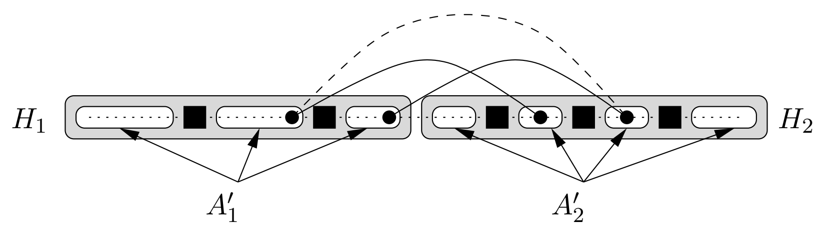


Figure 16: Illustration of the bi-clique chain graph H. The square vertices stand for affected vertices, and the sets $A'_1 = H_1 \setminus A_1$ and $A'_2 = H_2 \setminus A_2$ are simple K-joins of G, respectively.

$\square$

Corollary 3.10. *The* Bipartite Chain Deletion *problem admits a kernel with $O(k^2)$ vertices.*

4 Conclusion

In this paper we prove that the Proper Interval Completion problem admits a kernel with $O(k^3)$ vertices. Two natural questions arise from our results: firstly, does the Interval Completion problem admit a polynomial kernel? Observe that this problem is known to be FPT not for long [29]. The techniques we developed here intensively use the fact that there are few claws in the graph, what help us to reconstruct parts of the umbrella ordering. Of course, these considerations no more hold in general interval graphs. The second question is: does the Proper Interval Edge-Deletion problem admit a polynomial kernel? Again, this problem admits a fixed-parameter algorithm [27], and we believe that our techniques could be applied to this problem as well. Finally, we proved that the Bi-clique Chain Completion problem admits a kernel with $O(k^2)$ vertices, which completes a result of Guo [13]. In all cases, a natural question is thus whether these bounds can be improved?

References

[1] S. Bessy, C. Paul, and A. Perez. Polynomial kernels for 3-leaf power graph modification problems. *Discrete Applied Mathematics*, 158(16):1732–1744, 2010.

[2] H. L. Bodlaender. Kernelization: New upper and lower bound techniques. In *IWPEC*, pages 17–37, 2009.

[3] H. L. Bodlaender, R. G. Downey, M. R. Fellows, and D. Hermelin. On problems without polynomial kernels. *J. Comput. Syst. Sci*, 75(8):423–434, 2009.

[4] L. Cai. Fixed-parameter tractability of graph modification problems for hereditary properties. *Inf. Process. Lett*, 58(4):171–176, 1996.

[5] J. Chen and J. Meng. A $2k$ kernel for the cluster editing problem. In *COCOON*, volume 6196 of *LNCS*, pages 459–468, 2010.

[6] D. G. Corneil. A simple 3-sweep LBFS algorithm for the recognition of unit interval graphs. *Discrete Appl. Math.*, 138:371–379, April 2004.

[7] D. G. Corneil, H. Kim, S. Natarajan, S. Olariu, and A. P. Sprague. Simple linear time recognition of unit interval graphs. *Information Processing Letters*, 55(2):99 – 104, 1995.

[8] F. K. Dehne, M. R. Fellows, F. A. Rosamond, and P. Shaw. Greedy localization, iterative compression, modeled crown reductions: New FPT techniques, an improved algorithm for set splitting, and a novel 2k kernelization for vertex cover. In *IWPEC*, volume 3162 of *LNCS*, pages 271–280, 2004.

[9] X. Deng, P. Hell, and J. Huang. Linear-time representation algorithms for proper circular-arc graphs and proper interval graphs. *SIAM J. Comput.*, 25(2):390–403, 1996.

[10] R. G. Downey and M. R. Fellows. *Parameterized complexity*. Springer, 1999.

[11] M. C. Golumbic, H. Kaplan, and R. Shamir. On the complexity of DNA physical mapping. *ADVAM: Advances in Applied Mathematics*, 15, 1994.

[12] S. Guillemot, C. Paul, and A. Perez. On the (non-)existence of polynomial kernels for P_l-free edge modification problems. In *IPEC*, volume 6478 of *LNCS*, pages 147–157, 2010.

[13] J. Guo. Problem kernels for NP-complete edge deletion problems: Split and related graphs. In *ISAAC*, volume 4835 of *LNCS*, pages 915–926, 2007.

[14] M. Habib and C. Paul. A survey on algorithmic aspects of modular decomposition. *Computer Science Review*, 4(1):41–59, 2010.

[15] P. Hell, R. Shamir, and R. Sharan. A fully dynamic algorithm for recognizing and representing proper interval graphs. *SIAM Journal on Computing*, 31(1):289–305, 2001.

[16] H. Kaplan, R. Shamir, and R. E. Tarjan. Tractability of parameterized completion problems on chordal and interval graphs: Minimum fill-in and physical mapping. In *FOCS*, pages 780–791, 1994.

[17] H. Kaplan, R. Shamir, and R. E. Tarjan. Tractability of parameterized completion problems on chordal, strongly chordal, and proper interval graphs. *SIAM J. Comput*, 28(5):1906–1922, 1999.

[18] S. Kratsch and M. Wahlström. Two edge modification problems without polynomial kernels. In *IWPEC*, volume 5917 of *LNCS*, pages 264–275. Springer, 2009.

[19] P. J. Looges and S. Olariu. Optimal greedy algorithms for indifference graphs. *Computers & Mathematics with Applications*, 25(7):15 – 25, 1993.

[20] F. Mancini. *Graph modification problems related to graph classes.* PhD thesis, University of Bergen, Norway, 2008.

[21] R. Niedermeier. *Invitation to fixed parameter algorithms*, volume 31 of *Oxford Lectures Series in Mathematics and its Applications.* Oxford University Press, 2006.

[22] R. Niedermeier and P. Rossmanith. A general method to speed up fixed-parameter-tractable algorithms. *Inf. Process. Lett*, 73(3-4):125–129, 2000.

[23] R. Shamir, R. Sharan, and D. Tsur. Cluster graph modification problems. *Discrete Applied Mathematics*, 144(1-2):173–182, 2004.

[24] R. Sharan. *Graph modification problems and their applications to genomic research.* PhD thesis, Tel-Aviv University, 2002.

[25] R. E. Tarjan and M. Yannakakis. Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs. *SIAM J. Comput*, 13(3):566–579, 1984.

[26] S. Thomassé. A $4k^2$ kernel for feedback vertex set. *ACM Transactions on Algorithms*, 6(2), 2010.

[27] Y. Villanger. `www.lirmm.fr/~paul/ANR/CIRM-TALKS-2010/Villanger-cirm-2010.pdf`, 2010.

[28] Y. Villanger. Proper interval vertex deletion. In *IPEC*, volume 6478 of *LNCS*, pages 228–238, 2010.

[29] Y. Villanger, P. Heggernes, C. Paul, and J. A. Telle. Interval completion is fixed parameter tractable. *SIAM J. Comput*, 38(5):2007–2020, 2009.

[30] G. Wegner. *Eigenschaften der nerven homologische-einfactor familien in* $\mathbb{R}^n$. PhD thesis, Universität Gottigen, Gottingen, Germany, 1967.

[31] M. Yannakakis. Computing the minimum fill-in is NP-Complete. *SIAM J. Alg. and Discr. Meth.*, 2(1):77–79, 1981.