

Polynomial kernels for PROPER INTERVAL COMPLETION and related problems *

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

September 18, 2021

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 [14, 22]. First announced by Kaplan, Tarjan and Shamir in FOCS '94, this problem is known to be FPT [14], but no polynomial kernel was known to exist. We settle this question by proving that PROPER INTERVAL COMPLETION admits a kernel with at most $O(k^5)$ 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 [12].

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 [18, 21, 22]. Well-known examples include the VERTEX COVER [8], FEEDBACK VERTEX SET [24], or CLUSTER EDITING [5] problems. These problems find applications in various domains, such as computational biology [14, 22], image processing [21] or relational databases [23].

Due to these applications, one may be interested in computing an *exact* solution for such problems. *Parameterized complexity* provides a useful theoretical framework to that aim [9, 19]. 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 any constant $c > 0$. 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 been extensively studied in the last decade (see [2] for a survey). A *kernelization algorithm* is a

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

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 parameterized problem is FPT if and only if it has a kernelization algorithm [19]. 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 [20]. 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 [24] 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 [11, 16]. 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 [17, 28]. In particular, there exists an *infinite* set of forbidden induced subgraphs that characterizes proper interval graphs [28] (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 [13, 14], 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 [13]). 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 [10], but fixed-parameter tractable due to a result of Kaplan, Tarjan and Shamir in FOCS '94 [14, 15].¹ The fixed-parameter tractability of the PROPER INTERVAL COMPLETION can also be seen as a corollary of a characterization of Wegner [28] combined with Cai's result [4]. Nevertheless, it was not known whether this problem admit a polynomial kernel or not.

Our results We prove that the PROPER INTERVAL COMPLETION problem admits a kernel with at most $O(k^5)$ 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

¹Notice also that the *vertex deletion* of the problem is fixed-parameter tractable [26].

problem where one is given a graph $G = (V, E)$ and seeks a set of at most k edges whose deletion from E result 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 [12] 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 $O(k^5)$ vertices. Finally, we prove that our techniques can be applied to BIPARTITE CHAIN DELETION to obtain a quadratic-vertex kernel, completing a previous result of Guo [12].

1 Preliminaries

1.1 Proper interval graphs

We consider simple, loopless, undirected graphs $G = (V, E)$ where $V(G)$ denotes the vertex set of G and $E(G)$ its edge set². 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 $\{N_G(s) \setminus S : s \in S\}$. 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 $x_1, \dots, x_{|X|}$ on X such that for all $i = 1, \dots, |X| - 1$, $N_Y(x_i) \subseteq N_Y(x_{i+1})$. 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 [28]). *A graph is a proper interval graph if and only if it does not contain any $\{\text{hole}, \text{claw}, \text{net}, \text{3-sun}\}$ as an induced subgraph (see Figure 1).*

The claw graph is the bipartite graph $K_{1,3}$. Denoting the 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 [17]). *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 $v_i v_j \in E$ with $i < j$ then $v_i v_l, v_l v_j \in E$ 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

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

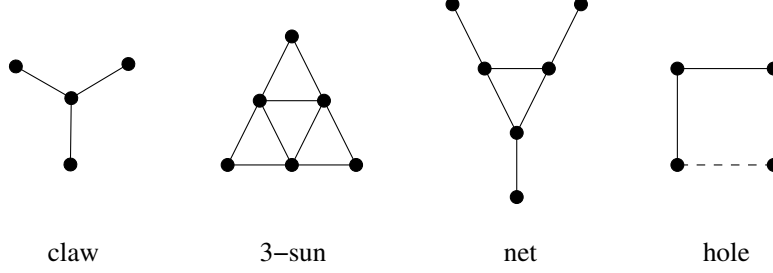


Figure 1: The forbidden induced subgraphs of proper interval graphs. A *hole* is an induced cycle of length at least 4.

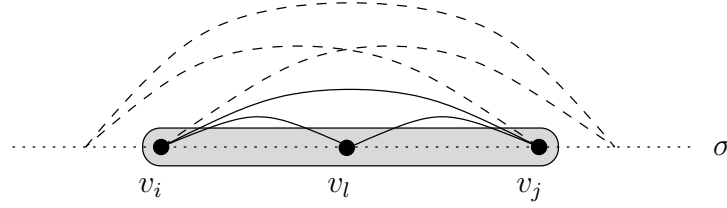


Figure 2: Illustration of the umbrella property. The edge $v_i v_j$ is extremal.³

maximal set of true twins of G is consecutive, and that σ_G is unique up to permutation of true twins of 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 $1 \leq i < l$, $(\{v_1, \dots, v_i\}, \{v_{i+1}, \dots, v_n\})$ 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 of 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 . \square

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.

³In 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.

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):

- (i) The graph $G[B]$ is a connected proper interval graph admitting an umbrella ordering $\sigma_B = b_1, \dots, b_{|B|}$ 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 $\{b_1, \dots, b_{l-1}\}$ 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 $b_{|B|}$). We call B_1 the *attachment clique* of B , and use B^R to denote $B \setminus B_1$.

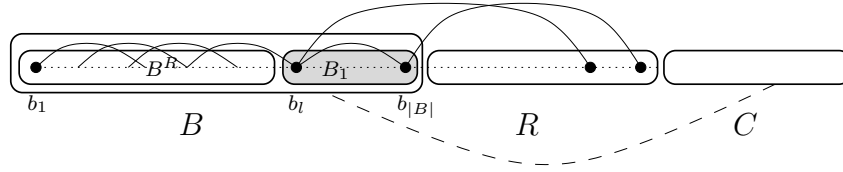


Figure 3: A 1-branch of a graph $G = (V, E)$. The vertices of B are ordered according to the umbrella ordering σ_B .

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 $\sigma_B = b_1, \dots, b_{|B|}$ 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 $\{b_1, \dots, b_{l-1}\}$ and R where b_l is the neighbor of $b_{|B|}$ with minimal index in σ_B ,
 - no edges between $\{b_{l'+1}, \dots, b_{|B|}\}$ 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 $B \setminus (B_1 \cup B_2)$. Observe that the cases where $L = \emptyset$ or $R = \emptyset$ are possible, and correspond to the definition of 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 .

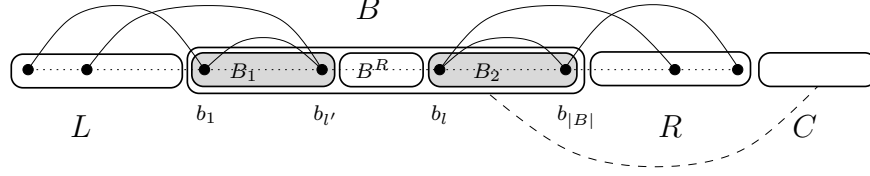


Figure 4: A 2-branch of a graph $G = (V, E)$. The vertices of B are ordered according to the umbrella ordering σ_B .

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 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 iff 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).

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 $\mathcal{S} = \{C_1, \dots, C_m\}$, $m > k$ 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 $\mathcal{S} = \{C_1, \dots, C_m\}$, $m > k$ 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.

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. Observe that a sunflower can be found in polynomial time once we have enumerated all the claws and 4-cycles of a graph, which can clearly be done in $O(n^4)$. \square

2.1.2 Extracting a clean K -join from a K -join

Now, we want to reduce the size of the 'simplest' branches, namely the K -joins. More precisely, in the next subsection we will bound the number of vertices in a clean K -join (whose vertices are not contain in any claw or 4-cycle), and so, we first indicate how to extract a clean K -join from a K -join.

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. \square

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 and B be a K -join of G . There are at most $k^3 + 4k^2 + 5k + 1$ vertices of B that belong to a claw or a 4-cycle.

Proof. By Lemma 2.2, there are at most $3k^2 + 4k$ vertices of B that are leaves of a claw or in 4-cycles. We remove these vertices from B and denote B' the set of remaining vertices, which forms a K -join. Now, we remove from B' all the vertices that do not belong to any claw and contract all the true twins in the remaining vertices. As Rule 2.2 has been applied on B , every contracted set has size at most $k + 1$. We denote by B'' the obtained set which can be seen as a subset of B and then, B'' is also a K -join of G . Remark that every vertex of B'' is the center of a claw. We consider an umbrella ordering b_1, \dots, b_l of B'' . We will find a set of $l - 1$ claws with distinct sets of leaves, which will bound l by $k^2 + 1$, by Lemma 2.2. As, for all $i = 1, \dots, l - 1$, b_i and b_{i+1} are not true twins, there exists c_i such that $b_i c_i \in E$ and $b_{i+1} c_i \notin E$ or $b_i c_i \notin E$ and $b_{i+1} c_i \in E$. As B'' is K -join, by definition, all the c_i 's are distinct. Now, for every $i = 1, \dots, l - 1$, we will find a claw containing c_i as leaf. Assume that $b_i c_i \notin E$ and $b_{i+1} c_i \in E$. As b_{i+1} is the center of a claw, there exists a set $\{x, y, z\}$ which is an independent set and is fully adjacent to b_{i+1} . If $c_i \in \{x, y, z\}$, we

are done. Assume this is not the case. This means that b_i is adjacent to any vertex of $\{x, y, z\}$ (otherwise one of this vertex would be adjacent to b_{i+1} and not to b_i , and we choose it to be c_i). Now, if two elements of this set, say x and y , are adjacent to c_i , then $\{x, c_i, y, b_i\}$ forms a 4-cycle that contains b_i , which is not possible. So, at least two elements among $\{x, y, z\}$, say x and y , are not adjacent to c_i and then, we find the claw $\{b_{i+1}, x, y, c_i\}$ of center b_{i+1} that contains c_i . In the case where $b_i c_i \in E$ and $b_{i+1} c_i \notin E$, we proceed similarly by exchanging the role of b_i and b_{i+1} and find also a claw containing c_i . Finally, all the considered claws have distinct sets of leaves and there are at most k^2 such claws by Lemma 2.2. What means that B'' has size at most $k^2 + 1$ and B' at most $(k + 1)(k^2 + 1)$. As we removed at most $3k^2 + 4k$ vertices of B that could be leaves of claws or contain in 4-cycles, we obtain $k^3 + 4k^2 + 5k + 1$ vertices of B that are possibly in claws or 4-cycles. \square

Since any subset of a K -join forms a K -join, Lemma 2.3 implies that it is possible to remove a set of at most $k^3 + 4k^2 + 5k + 1$ vertices from any K -join to obtain a clean K -join.

2.1.3 Bounding the size of the 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.

Rule 2.4 (K -join). *Let B be a clean K -join of size at least $2k + 2$. Let B_L be the $k + 1$ first vertices of B , B_R be its $k + 1$ last vertices 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 $G' = G \setminus M$. 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 $H = G' + F$ the resulting proper interval graph and by $\sigma_H = b_1, \dots, b_{|H|}$ 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).

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 stand in N . So, if R contains two vertices u, v such that $uv \notin E$, we form the claw $\{b_{|B|}, b_1, u, v\}$ of 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 $M \subseteq N_B(l)$.*

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 b_f and b_l be respectively the first and last vertex of $B \setminus M$ appearing in σ_H . We let B_H denote the set $\{u \in V(H) : b_f \leq_{\sigma_H} u \leq_{\sigma_H} b_l\}$. Observe that B_H is a clique in

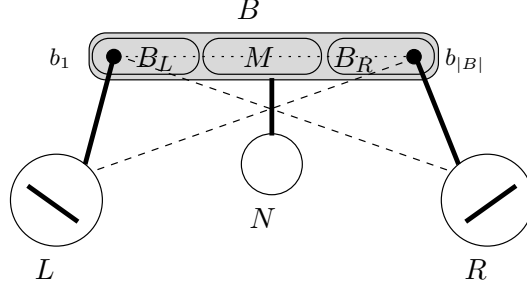


Figure 5: The structure of the K -join B .

H since $b_f b_l \in E(G)$ 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 $x <_{\sigma_H} y <_{\sigma_H} b_f$ and $N_M(y) \subset N_M(x)$, then we exchange x and y in σ_H . This process stops when the considered true 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 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 $vm \notin E(G)$ 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 $ub_l \in E(G)$. Since σ_H is an umbrella ordering for H , it follows that $vb_l \in E(H)$ 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 $u'w$ 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 $b_l <_{\sigma_H} w$ 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 $vm \in E(G)$ by Observation 2.6. Assume now that there exists a vertex w which distinguishes u and v with $v <_{\sigma_H} w <_{\sigma_H} b_f$. 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 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').

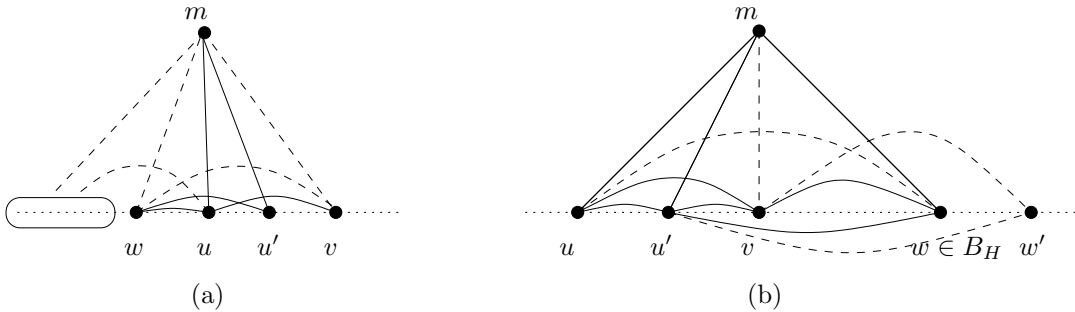


Figure 6: (a) u and v are distinguished by some vertex $w <_{\sigma_H} u$; (b) u and v are distinguished by a vertex $w \in B_H$.

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 (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 distinguish 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 $x <_{\sigma_H} y <_{\sigma_H} b_f$.

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

Claim 2.9. *Let $m \in M$. Then m can be added to the graph H while preserving an umbrella ordering.*

Proof. Let $m \in M$ and v_i (resp. v_j) be the vertex with minimal (resp. maximal) index in σ_H such that $v_im \in E(G)$ (resp. $v_jm \in E(G)$). By definition, we have $v_{i-1}m, v_{j+1}m \notin E(G)$ and through Claim 2.8, we know that $N_H(m) = \{w \in V : v_i \leq_{\sigma_H} w \leq_{\sigma_H} v_j\}$. Moreover, since $B_H \cup M$ is a clique by Claim 2.7, it follows that $v_{i-1} <_{\sigma_H} b_f$ and $b_l <_{\sigma_H} v_{j+1}$. Hence, by Claim 2.8, we know that $v_{i-1}v_{j+1} \notin E(G)$, 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 $\{w \leq_{\sigma_H} v_{i-1} : vw \in E(H)\}$ (resp. $\{w \geq_{\sigma_H} v_{j+1} : vw \in E(H)\}$). 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 $xv, yv \in E(G)$, then the set $\{v, x, y, m\}$ defines a claw containing m in G , which cannot be. We now consider $b_{v_{i-1}}$ the neighbor of v_{i-1} with maximal index in σ_H . Similarly we let $b_{v_{j+1}}$ be the neighbor of v_{j+1} with minimal index in σ_H . Since $v_{i-1}v_{j+1} \notin E(G)$, we have $b_{v_{i-1}}, b_{v_{j+1}} \in N_H(m)$. We study the behavior of $b_{v_{i-1}}$ and $b_{v_{j+1}}$ in order to conclude.

Assume first that $b_{v_{j+1}} <_{\sigma_H} b_{v_{i-1}}$. Let X be the set of vertices $\{w \in V : b_{v_{j+1}} \leq_{\sigma_H} w \leq_{\sigma_H} b_{v_{i-1}}\}$. Remark that we have $b_{v_{i-1}} \leq_{\sigma_H} b_l$ and $b_f \leq_{\sigma_H} b_{v_{j+1}}$, otherwise for instance, if we have $b_{v_{i-1}} >_{\sigma_H} b_l$, then $B_H \subseteq N_H(v_{i-1})$ implying, as usual, that $v_{i-1}m \in E(G)$ 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 $w'' \in N^+(x)$. We now 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 $H \setminus B_H$.

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

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

- (i) First, assume that $b_{v_i} <_{\sigma_H} b_{v_j}$, case depicted in Figure 8 (a). In particular, this means that $v_iv_j \notin E(G)$. If b_{v_i} and b_{v_j} are consecutive in σ_H , then inserting m between b_{v_i} and b_{v_j} yields

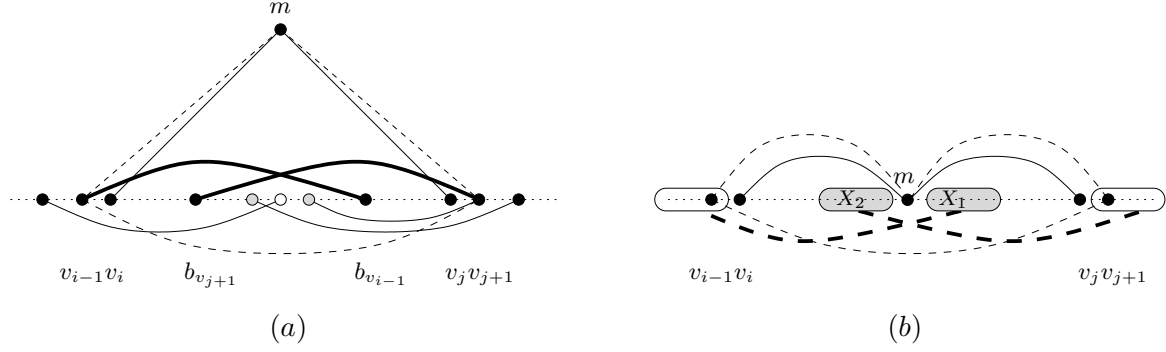


Figure 7: Illustration of the reordering applied to σ_H . The thin edges stand for edges of G . On the left, the gray vertices represent vertices of X_1 while the white vertex is a vertex of X_2 .

an umbrella ordering (since b_{v_j} (resp. b_{v_i}) does not have any neighbor before (resp. after) v_i (resp. v_j) in σ_H). Now, if there exists $w \in V$ such that $b_{v_i} <_{\sigma_H} w <_{\sigma_H} b_{v_j}$, then one can see that the set $\{m, v_i, w, v_j\}$ forms a claw containing m in G , which is impossible.

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

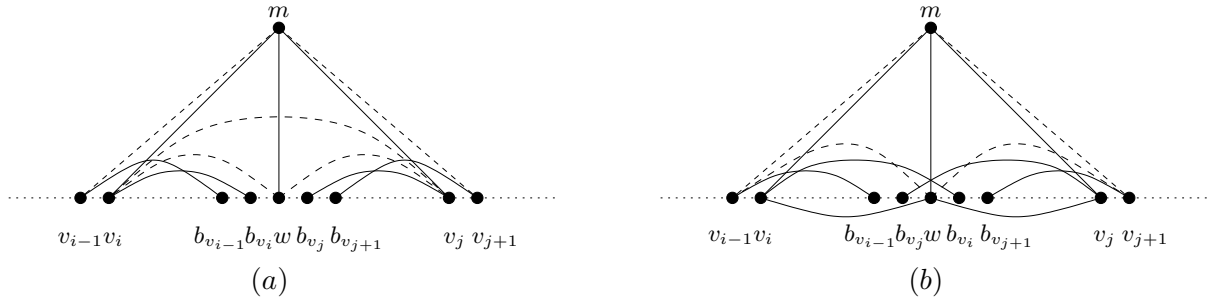


Figure 8: The possible cases for $b_{v_{i-1}} <_{\sigma_H} b_{v_{j+1}}$.

◇

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 has size at most $k^3 + 4k^2 + 7k + 3$.*

Proof. Let B be any K -join of G , and assume $|B| > k^3 + 4k^2 + 7k + 3$. By Lemma 2.2 we know that it is possible to extract a clean K -join from B of size at least $|B| - (k^3 + 4k^2 + 5k + 1) > 2(k + 1)$ what is impossible after having applied Rule 2.4. \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 and B be a 1-branch of G associated with the umbrella ordering σ_B . Assume that $|B^R| \geq 2k + 1$ and let B_f be the $2k + 1$ last vertices of B^R according to σ_B . For any k -completion F of G into a proper interval graph, there exists a k -completion F' of G with $F' \subseteq F$ and a vertex $b \in B_f$ such that the umbrella ordering of $G + F'$ preserves the order of the set $B_b = \{v \in V : b_1 \leq_{\sigma_B} v \leq_{\sigma_B} b_{l'}\}$, where l' is the maximal index such that $bb_{l'} \in E(G)$. 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 $|B_f| = 2k + 1$ and $|F| \leq k$, there exists a vertex $b \in B_f$ not incident to any added edge of F . We let N_D be the set of neighbors of b that are *after* b in σ_B , B' the set of vertices that are *before* $N_G[b]$ in σ_B , $B_b = B' \cup N_G[b]$ and $C = V \setminus B_b$ (see Figure 9).

Claim 2.12. (i) $G[C]$ is a connected graph and
(ii) Either $\forall u \in C \ b <_{\sigma_H} u$ holds or $\forall u \in C \ u <_{\sigma_H} b$ holds.

Proof. The first point follows from the fact that G is connected and that, by construction, $B_1 \subseteq C$ and B_1 is connected. 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 $u', v' \in C$ such that $u' <_{\sigma_H} b <_{\sigma_H} v'$ and $u'v' \in E(G)$. Then, we have $u'b, v'b \notin E(H)$, contradicting the fact that σ_H is an umbrella ordering for H . \diamond

In the following, we assume w.l.o.g. that $b <_{\sigma_H} u$ holds for any $u \in C$. 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 9). We construct a completion F' from F as follows: we remove from F the edges with both extremities in B_b , and remove all edges between $B_b \setminus N_D$ and C . In other words, we set:

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

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 $B_b \setminus N_D$ and C , it follows that S intersects N_D and C . We study the different cases:

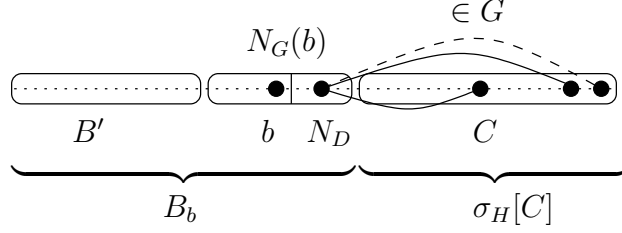


Figure 9: The construction of the ordering σ according to σ_H .

- (i) (1) holds and $u \in N_D$, $v, w \in C$: since the edge set between N_D 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 $u, v \in N_D$, $w \in C$: this case cannot happen since N_D is a clique of H' .
- (iii) (2) holds and $u \in N_D$, $v, w \in C$: this case is similar to (i). Observe that we may assume $uv \in E(H)$ (otherwise (i) holds). By construction $vw \notin E(H)$ 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 $bu \in E(H)$. In the latter case, since σ_H is an umbrella ordering this means that $bv \in E(H)$. Since b is non affected vertex and $bv \notin E(G)$, this leads to a contradiction.
- (iv) (2) holds and $u, v \in N_D$, $w \in C$: 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) containing 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 10. 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, since $u'w \in E(G)$, we have a contradiction since $N_C(u') \subseteq N_C(v)$ in G (observe that $u' \in B$ by construction). Finally, in the third case, $uw', vw' \in E(G)$ since $N_C(u') \subseteq N_C(u) \subseteq N_C(v)$ in G , and we are in configuration (i) with vertex set $\{v, w, w'\}$.

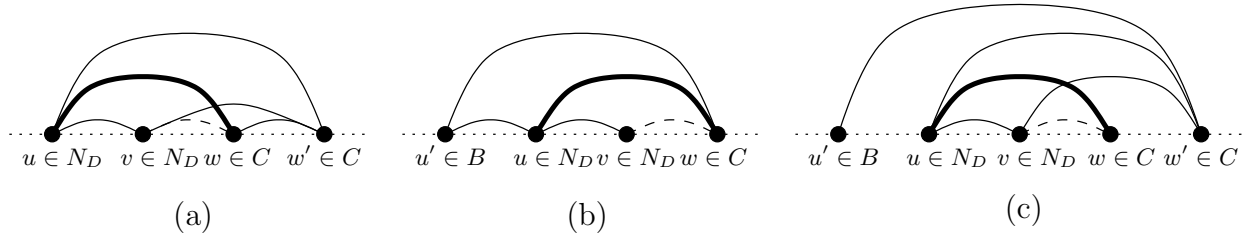


Figure 10: Illustration of the different cases of configuration (iv) (the bold edges belong to F').

◇

Altogether, we proved that for any k -completion F , there exists 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. □

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

Lemma 2.14. Rule 2.5 is safe.

Proof. Let $G' = G \setminus (B^R \setminus B_f)$ 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. By Lemma 2.11 we know that there exists a vertex $b \in B_f$ such that the order of $B_b = N_G[b] \cup \{v \in B_f : v <_{\sigma_B} N_G[b]\}$ is the same than in B and the vertices of B_b are the first of σ_H . Since $N_G(B^R \setminus B_f) \subseteq N_G[b]$, it follows that the vertices of $B^R \setminus B_f$ 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. The 1-branches of G contain at most $k^3 + 4k^2 + 9k + 4$ vertices.

Proof. Let B be a 1-branch of a graph $G = (V, E)$ reduced under Rules 2.2 to 2.5. Assume $|B| > k^3 + 4k^2 + 9k + 4$. Since G is reduced under Rule 2.4, we know by Observation 2.10 that the attachment clique B_1 of B , which is a K -join, contains at most $k^3 + 4k^2 + 7k + 3$ vertices. This implies that $|B^R| > 2k + 1$, which cannot be since G is reduced under Rule 2.5. \square

2.1.5 Cutting the 2-branches

To obtain a rule reducing the 2-branches, we need to introduce a particular decomposition of 2-branches into K -joins. Let B be a 2-branch with an umbrella ordering $\sigma_B = b_1, \dots, b_{|B|}$. As usual, we denote by $B_1 = b_1, \dots, b_{l'}$ its first attachment clique and by $B_2 = b_l, \dots, b_{|B|}$ its second. The reversal of the permutation σ_B gives a second possibility to fix B_1 and B_2 . We fix one of these possibilities and define \mathcal{B} , the K -join decomposition of B . The K -joins of \mathcal{B} are defined by $B'_i = b_{l_{i-1}+1}, \dots, b_{l_i}$ where b_{l_i} is the neighbor of $b_{l_{i-1}+1}$ with maximal index. The first K -join of \mathcal{B} is B_1 (so, $l_0 = 0$ and $l_1 = l'$), and once B'_{i-1} is defined, we set B'_i : if $b_{l_{i-1}+1} \in B_2$, then we set $B'_i = b_{l_{i-1}+1}, \dots, b_{|B|}$, otherwise, we choose $B'_i = b_{l_{i-1}+1}, \dots, b_{l_i}$ (see Figure 11).

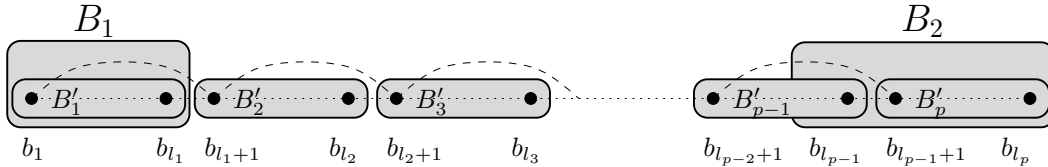


Figure 11: The K -join decomposition.

Now, we can prove the next lemma, which bounds the number of K -joins in the K -join decomposition of a 2-branch providing that some connectivity assumption holds.

Lemma 2.16. Let $G = (V, E)$ be an instance of PROPER INTERVAL COMPLETION and B be a 2-branch containing $p \geq (k + 4)$ K -joins in its K -join decomposition. Assume the attachment

cliques B_1 and B_2 of B belong to the same connected component of $G[V \setminus B^R]$. Then there is no k -completion for G .

Proof. Let B be a 2-branch of an instance $G = (V, E)$ of PROPER INTERVAL COMPLETION respecting the conditions of Lemma 2.16. Since B_1 and B_2 belong to the same connected component in $G[V \setminus B^R]$, let π be a shortest path between B_1 and B_2 in $G[V \setminus B^R]$. As B has $p \geq k + 4 \geq 3$ K -joins in its decomposition, no vertex of B_1 is adjacent to a vertex of B_2 and π has length at least two. We denote by $u \in B_1$ and $v \in B_2$ the extremities of such a path. We now construct an induced path P_{uv} of length at least $p - 1$ between u and v within B . To do so, considering the K -join decomposition $\mathcal{B} = \{B'_1, \dots, B'_p\}$ of B , we know that $u \in B'_1$ and that $v \in B'_{p-1} \cup B'_p$. We define $u_1 = u$ and while $v \notin N[u_i]$, we choose u_{i+1} the neighbor of u_i with maximum index in the umbrella ordering of B . In this case, we have $u_i \in \cup_{j=1}^i B'_j$ for every $1 \leq i \leq p$. Indeed, the neighbor of a vertex of $\cup_{j=1}^{i-1} B'_j$ with maximum index in the umbrella ordering of B is in $\cup_{j=1}^i B'_j$. Finally, when $v \in N[u_i]$, we just choose $u_{i+1} = v$. So, the path $P_{uv} = u_1, \dots, u_l$ is an induced path of length at least $p - 1$, with $u_1 = u$, $u_l = v$ and the only vertices that could have neighbors in $G \setminus B$ are u_1, u_{l-1} and u_l ($u_1 \in B_1$, $u_l \in B_2$ and u_{l-1} is possibly in B_2). Using π , we can form an induced cycle of length at least $p \geq k + 4$ in G . Since at least $q - 3$ completions are needed to triangulate any induced cycle of length q [14], it follows that there is no k -completion for G . \square

The following observation is a straightforward implication of Lemma 2.16.

Observation 2.17. *Let $G = (V, E)$ be a connected positive instance of PROPER INTERVAL COMPLETION, reduced by Rule 2.4 and B be a 2-branch such that $G[V \setminus B^R]$ is connected. Then B contains at most $k + 3$ K -joins in its K -join decomposition and hence at most $(k + 3)(k^3 + 4k^2 + 5k + 1)$ vertices.*

Rule 2.6 (2-branches). *Let G be a connected graph and B be a 2-branch such that $G[V \setminus B^R]$ is not connected. Assume that $|B^R| \geq 4(k + 1)$ and let B'_1 be the $2k + 1$ vertices after B_1 and B'_2 the $2k + 1$ vertices before B_2 . Remove $B \setminus (B_1 \cup B'_1 \cup B'_2 \cup B_2)$ from G .*

Lemma 2.18. *Rule 2.6 is safe.*

Proof. As usual, we denote by $\sigma_b = b_1, \dots, b_{|B|}$ the umbrella ordering defined on B , with $B_1 = \{b_1, \dots, b_{l'}\}$ and $B_2 = \{b_l, \dots, b_{|B|}\}$. We partition B^R into two sets $B' = \{b_{l'+1}, \dots, b_i\}$ and $B'' = \{b_{i+1}, \dots, b_{l-1}\}$ such that $|B'| \geq |B''| \geq 2k + 1$. We now remove the edges $E(B', B'')$ between B' and B'' , obtaining two connected components of G , G_1 and G_2 . Observe that B' defines a 1-branch of G_1 with attachment clique B_1 such that $B' \setminus B_1$ contains at least $2k + 1$ vertices. Similarly B'' defines a 1-branch of G_2 with attachment clique B_2 such that $B'' \setminus B_2$ contains at least $2k + 1$ vertices. Hence Lemma 2.11 can be applied to both G_1 and G_2 and we continue as if Rule 2.5 has been applied to G_1 and G_2 , preserving exactly $2k + 1$ vertices B'_f and B''_f , respectively. We denote by G' the reduced graph. Let F be a k -completion of G . Let F_1 and F_2 be the completions of G_1 and G_2 such that $|F_1| + |F_2| \leq k$. Moreover, let $H_1 = G_1 + F_1$ and $H_2 = G_2 + F_2$. By Lemma 2.11, we know that the vertices of $B' \setminus B'_f$ (resp. $B'' \setminus B''_f$) can be inserted into the umbrella ordering σ_{H_1} of H_1 (resp. σ_{H_2}) in the same order than in B' (resp. B''). We thus obtain two proper interval graphs H'_1 and H'_2 whose respective umbrella ordering preserve the order of B' and B'' . We now connect H'_1 and H'_2 by putting back the edges contained in $E(B', B'')$, obtaining a graph H with ordering σ_H . Since $G[B]$ is a proper interval graph and B' and B'' are ordered according to B in H'_1 and H'_2 , it follows that H is a proper interval graph, and hence $F = F_1 \cup F_2$ is a k -completion of G . \square

Observation 2.19. *Let $G = (V, E)$ be a positive instance of PROPER INTERVAL COMPLETION reduced under Rules 2.2 to 2.6. The 2-branches of G contain at most $(k+3)(k^3 + 4k^2 + 5k + 1)$ vertices.*

Proof. Let B be a 2-branch of a graph $G = (V, E)$ and C be the connected component containing B . If $G[C \setminus B^R]$ is connected, then Observation 2.17 implies the result. Otherwise, as G has been reduced under Rules 2.2 to 2.6, we know that $|B^R| \leq 4k + 4$ and then that $|B| \leq 2(k^3 + 4k^2 + 5k + 1) + (4k + 4)$ which is less than $(k+3)(k^3 + 4k^2 + 5k + 1)$, provided that $k \geq 1$. \square

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 2.20. *Let $G = (V, E)$ be a graph and x a vertex of G . In time $O(n^2)$, 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 $1 \leq j < |C_i|$, 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 $\cup_{l=1}^{i-1} B_l^R$). 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 x to y if: xy is an edge of G , $N_{B^R}(y) \subseteq N_{B^R}(x)$ and $N_{G \setminus B^R}[x] \subseteq N_{G \setminus B^R}[y]$. This graph can be computed in time $O(n^2)$. 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. \square

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 2.21. *Let $G = (V, E)$ be a graph and x and y two adjacent vertices of G . It is possible to compute in cubic 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. Now, we construct on N' an oriented graph, 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 $O(n^3)$. 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 $O(n^3)$ time. \square

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 $O(n^3m)$ 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 2.22. *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 $B' \subseteq \{u \in B : b_f \leq_{\sigma_B} u \leq_{\sigma_B} b_l\}$. Furthermore, as $\{u \in B : b_f \leq_{\sigma_B} u \leq_{\sigma_B} b_l\}$ is a K -join and B' is maximal, we have $B' = \{u \in B : b_f \leq_{\sigma_B} u \leq_{\sigma_B} b_l\}$. Now, if b_fb_l was not an extremal edge of σ_B , it would be possible to extend B' , contradicting the maximality of B' . \square

Now, we can detect the 2-branches B with a set B^R non empty. Observe that this is enough for our purpose since we want to detect 2-branches of size at least $(k+3)(k^3+4k^2+5k+1)$ and the attachment cliques contain at most $2(k^3+4k^2+7k+3)$ vertices.

Lemma 2.23. *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 quadratic 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 2.22, 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_fb_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 2.20, 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 . \square

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 a $O(n^4)$ brute force detection to localize all the 4-cycles and the claws, we obtain the following result.

Lemma 2.24. *Given a graph $G = (V, E)$, the reduction rules 2.4 to 2.6 can be carried out in polynomial time, namely in time $O(n^3m)$.*

2.3 Kernelization algorithm

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

Theorem 2.25. *The PROPER INTERVAL COMPLETION problem admits a kernel with $O(k^5)$ vertices.*

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 $A = \{a_1 <_{\sigma_H} \dots <_{\sigma_H} a_i <_{\sigma_H} \dots <_{\sigma_H} a_p\}$ be the set of such vertices, with $p \leq 2k$. The size of the kernel is due to the following observations (see Figure 12):

- Let $L_0 = \{l \in V : l <_{\sigma_H} a_1\}$ and $R_{p+1} = \{r \in V : a_p <_{\sigma_H} r\}$. Since the vertices of L_0 and R_{p+1} are not affected, it follows that $G[L_0]$ and $G[R_{p+1}]$ induce a proper interval graph. As Rule 2.1 has been applied, $G[L_0]$ and $G[R_{p+1}]$ both contain one connected component, and L_0 and R_{p+1} are 1-branches of G . So, by Observation 2.15, L_0 and R_{p+1} both contain at most $k^3 + 4k^2 + 9k + 4$ vertices.
- Let $S_i = \{s \in V : a_i <_{\sigma_H} s <_{\sigma_H} a_{i+1}\}$ for every $1 \leq i < p$. Again, since the vertices of S_i are not affected, it follows that $G[S_i]$ is a proper interval graph. As Rule 2.1 has been applied, there are at most two connected components in $G[S_i]$. If $G[S_i]$ is connected, then, S_i is a 2-branch of G and, by Observation 2.19, S_i contains at most $(k+3)(k^3 + 4k^2 + 5k + 1)$ vertices. Otherwise, if $G[S_i]$ contains two connected components, they correspond to two 1-branches of G , and by Observation 2.15, S_i contains at most $2(k^3 + 4k^2 + 9k + 4)$ vertices. In both cases, we bound the number of vertices of S_i by $(k+3)(k^3 + 4k^2 + 5k + 1)$, provided that $k \geq 1$.

Altogether, the proper interval graph H (and hence G) contains at most:

$$2(k^3 + 4k^2 + 9k + 4) + (2k - 1)((k + 3)(k^3 + 4k^2 + 5k + 1))$$

vertices, which implies the claimed $O(k^5)$ bound. The complexity directly follows from Lemma 2.24. \square

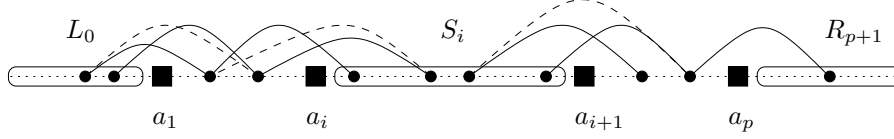


Figure 12: Illustration of the size of the kernel. The figure represents the graph $H = G + F$, the square vertices stand for the *affected* vertices, L_0 and R_{p+1} are 1-branches of G , and, on the figure, S_i defines a 2-branch.

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 [29] (see Figure 13). In [12], 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 13). Observe in particular that bi-clique chain graphs are proper interval graphs, and hence admit an umbrella ordering.

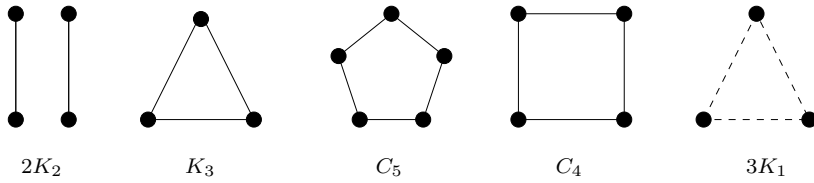


Figure 13: The forbidden induced subgraphs for bipartite and bi-clique chain graphs.

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 $\mathcal{S} = \{C_1, \dots, C_m\}$, $m > k$ 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 $\mathcal{S} = \{C_1, \dots, C_m\}$, $m > k$ 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 are 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 14.

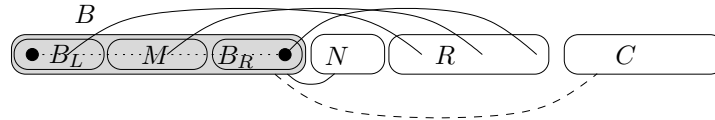


Figure 14: The K -join decomposition for the BI-CLIQUE CHAIN COMPLETION problem.

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. \diamond

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 $um \notin E(G)$ 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 $\sigma_H = b_1, \dots, b_{|H|}$, there exists a vertex b_i such that $H_1 = \{b_1, \dots, b_i\}$ and $H_2 = \{b_{i+1}, \dots, b_{|H|}\}$ 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 2.21. \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 15). 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.

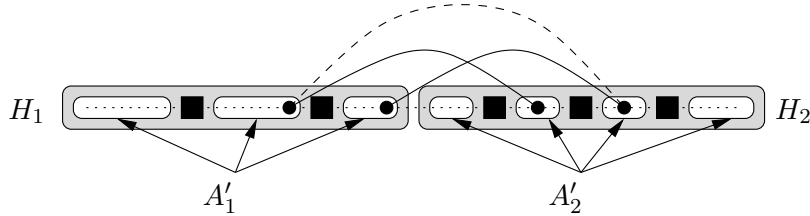


Figure 15: Illustration of the bi-clique chain graph H . The square vertices stand for affected vertices, and the sets $A'_1 = H_1 \setminus A_1$ and $A'_2 = H_2 \setminus A_2$ are simple K -joins of G , respectively. \square

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 $O(k^5)$ 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 [27]. 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 [25], 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 [12]. 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] Derek G. Corneil. A simple 3-sweep lbfs algorithm for the recognition of unit interval graphs. *Discrete Appl. Math.*, 138:371–379, April 2004.
- [7] Derek G. Corneil, Hiryoung Kim, Sridhar Natarajan, Stephan Olariu, and Alan P. Sprague. Simple linear time recognition of unit interval graphs. *Information Processing Letters*, 55(2):99 – 104, 1995.
- [8] F. K. H. A. 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] R.G. Downey and M.R. Fellows. *Parameterized complexity*. Springer, 1999.
- [10] M. C. Golumbic, H. Kaplan, and R. Shamir. On the complexity of DNA physical mapping. *ADVAM: Advances in Applied Mathematics*, 15, 1994.
- [11] 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.
- [12] J. Guo. Problem kernels for NP-complete edge deletion problems: Split and related graphs. In *ISAAC*, volume 4835 of *LNCS*, pages 915–926, 2007.
- [13] 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.

- [14] 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.
- [15] 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.
- [16] S. Kratsch and M. Wahlström. Two edge modification problems without polynomial kernels. In *IWPEC*, volume 5917 of *LNCS*, pages 264–275. Springer, 2009.
- [17] P. J. Looges and S. Olariu. Optimal greedy algorithms for indifference graphs. *Computers & Mathematics with Applications*, 25(7):15 – 25, 1993.
- [18] F. Mancini. *Graph modification problems related to graph classes*. PhD thesis, University of Bergen, Norway, 2008.
- [19] R. Niedermeier. *Invitation to fixed parameter algorithms*, volume 31 of *Oxford Lectures Series in Mathematics and its Applications*. Oxford University Press, 2006.
- [20] R. Niedermeier and P. Rossmanith. A general method to speed up fixed-parameter-tractable algorithms. *Inf. Process. Lett.*, 73(3-4):125–129, 2000.
- [21] R. Shamir, R. Sharan, and D. Tsur. Cluster graph modification problems. *Discrete Applied Mathematics*, 144(1-2):173–182, 2004.
- [22] R. Sharan. *Graph modification problems and their applications to genomic research*. PhD thesis, Tel-Aviv University, 2002.
- [23] 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.
- [24] S. Thomassé. A $4k^2$ kernel for feedback vertex set. *ACM Transactions on Algorithms*, 6(2), 2010.
- [25] Y. Villanger. www.lirmm.fr/~paul/ANR/CIRM-TALKS-2010/Villanger-cirm-2010.pdf, 2010.
- [26] Y. Villanger. Proper interval vertex deletion. In *IPEC*, volume 6478 of *LNCS*, pages 228–238, 2010.
- [27] Y. Villanger, P. Heggernes, C. Paul, and J. A. Telle. Interval completion is fixed parameter tractable. *SIAM J. Comput.*, 38(5):2007–2020, 2009.
- [28] G. Wegner. *Eigenschaften der nerven homologische-einfactor familien in R^n* . PhD thesis, Universität Göttingen, Göttingen, Germany, 1967.
- [29] M. Yannakakis. Computing the minimum fill-in is NP-Complete. *SIAM J. Alg. and Discr. Meth.*, 2(1):77–79, 1981.