Polynomial kernels for 3-leaf power graph modification problems*

S. Bessy   C. Paul   A. Perez

CNRS, LIRMM, Université Montpellier 2, France †

September 17, 2008

Abstract

A graph $G = (V, E)$ is a 3-leaf power iff there exists a tree T whose leaves are V and such that $(u, v) \in E$ iff u and v are at distance at most 3 in T. The 3-leaf power graph edge modification problems, *i.e.* edition (also known as the CLOSEST 3-LEAF POWER), completion and edge-deletion, are FTP when parameterized by the size of the edge set modification. However polynomial kernel was known for none of these three problems. For each of them, we provide cubic kernels that can be computed in linear time for each of these problems. We thereby answer an open problem first mentioned by Dom, Guo, Hüffner and Niedermeier [6].

*Research supported by the ANR ANR-BLAN-0148-06project "*Graph Decomposition on Algorithm* (GRAAL)"

†e-mail: bessy@lirmm.fr, paul@lirmm.fr, perez@lirmm.fr

Introduction

An edge modification problem aims at changing the edge set of an input graph $G = (V, E)$ in order to get a certain property Π satisfied (see [17] for a recent study). Edge modification problems cover a broad range of graph optimization problems among which completion problems (*e.g.* MINIMUM CHORDAL COMPLETION), edition problems (*e.g.* CLUSTER EDITION) and edge deletion problems (*e.g.* MAXIMUM PLANAR SUBGRAPH). For completion problems, the set F of modified edges is constraint to be disjoint from E, whereas for edge deletion problems F has to be a subset of E. If no restriction applies to F, then we obtain an edition problem. These problems are fundamental in graph theory and they play an important role in computational theory (*e.g.* they represent a large number of the earliest NP-Complete problems [9]). Edge modification problems have recently been extensively studied in the context of fixed parameterized complexity [7, 18]. The natural parameterization is the number $k = |F|$ of modified edges. The generic question is thereby whether for fixed k, the considered edge modification problem is tractable. More formally:

PARAMETERIZED Π-MODIFICATION PROBLEM
Input: An undirected graph $G = (V, E)$.
Parameter: An integer $k \geqslant 0$.
Question: Is there a subset $F \subseteq V \times V$ with $|F| \leqslant k$ such that the graph $G + F = (V, E \triangle F)$ satisfies Π.

This paper studies the parameterized version of edge modification problems and more precisely the existence of a polynomial *kernel*. A problem is *fixed parameterized tractable* (FPT for short) with respect to parameter k whenever it can be solved in time $f(k).n^{O(1)}$, where $f(k)$ is an arbitrary function. The membership to the FPT complexity class is equivalent to the property of having a kernel (see [18] for example). A problem is kernalizable if any instance (G, k) can be reduced in polynomial time into an instance (G', k') such that the size of G' is bounded by a function of k. Having a kernel of small size is clearly highly desirable [12]. Indeed preprocessing the input in order to reduce its size while preserving the existence of a solution is an important issue in the context of various applications ([12]). However the equivalence mentioned above only provides an exponential bound on the kernel size. For a parameterized problem the challenge is to know whether it admits or not a polynomial, even linear (in k) kernel (see *e.g.* [18]). The k-VERTEX COVER problem is the classical example of a problem with a linear kernel. Only recently, parameterized problems, among which the TREEWIDTH-k problem, have been shown to not have polynomial kernel [1] (unless some collapse occurs in the computational complexity hierarchy).

In this paper we focus on graph modification problems with respect to the property Π being 3-*leaf power*. The p-power of a graph $G = (V, E)$ is the graph $G^p = (V, E')$ with $(u, v) \in E'$ iff there is a path of length at most p between u and v in G. We say that G^p is the p-power of G and G the p-root of G^p. Deciding whether a graph is a power of some other graphs is a well-studied problem which is NP-Complete in general [16], but cubic for p-power of trees (fixed p) [13]. The notion of *leaf power* has been introduced in [19]: $G = (V, E)$ is a p-leaf power if there exists a tree T whose leaves are V and $(u, v) \in E$ iff u and v belong to V and are at distance at most p in T. The p-leaf power recognition problem has application in the context of phylogenetic tree reconstruction [15]. For $p \leqslant 5$, the p-leaf power recognition is polynomial [14]. Parameterized p-leaf power edge modification problems have been studied so far for $p \leqslant 4$. The edition problem for $p = 2$ is known as the Cluster Editing problem for which the bound of a polynomial kernel has been successively improved in a series of recent papers [8, 10, 22], culminating in [11] with a $4k$ kernel size. For larger values of p, the edition problem is known as the closest p-leaf power problem. For $p = 3$ and 4, the closest p-leaf power problem is known to be FPT [3, 6], while its fixed parameterized tractability is still open for larger values of p. But the existence of a polynomial kernel for $p \neq 2$ remained an open question [4, 5]. For the completion and edge-deletion, the problems are also know to be FTP for $p \leqslant 4$ [4, 6] and again polynomial kernel are only known for $p = 2$ [11].

Our results. We prove that the closest 3-leaf power, the 3-leaf power completion and the 3-leaf power edge-deletion admit a cubic kernel. We thereby answer positively to the open question of Dom, Guo, Hüffner and Niedermeier [4, 6].

Outlines. First section is dedicated to some known and new structural results of 3-leaf powers and their related critical clique graphs. Section 2 describes the data-reduction rules for the closest 3-leaf power problem. The cubic kernels for the other two variants, the 3-leaf power completion and the 3-leaf power edge-deletion problems, are presented in Section 3.

1 Preliminaries

The graphs we consider in this paper are undirected and loopless graphs. The vertex set of a graph G is denoted by $V(G)$, with $|V(G)| = n$, and its edge set by $E(G)$, with $|E(G)| = m$ (or V and E when the context is clear). The neighborhood of a vertex x is denoted by $N_G(x)$ (or $N(x)$). We write $d_G(u, v)$ the distance between two vertices u and v in G. Two vertices x and y of G are *true twins* if they are adjacent and $N(x) \cup \{x\} = N(y) \cup \{y\}$. A subset S of vertices is a *module* if for any distinct vertices x and y of S, $N(x) \setminus S = N(y) \setminus S$. Given a subset S of vertices, $G[S]$ denotes the subgraph of G induced by S. If H is a subgraph of G, $G \setminus H$ stands for $G[V(G) \setminus V(H)]$. A graph family $\mathcal{F}$ is *hereditary* if for any graph $G \in \mathcal{F}$, any induced subgraph H of G also belongs to $\mathcal{F}$. For a set $\mathcal{S}$ of graphs, we say that G is $\mathcal{S}$*-free* if none of the graphs of $\mathcal{S}$ is an induced subgraph of G.

As all the paper deals with undirected graphs, we abusively denote by $X \times Y$ the set of pairs (and not couples) containing one element of X and one of Y. Let $G = (V, E)$ be a graph and F be a subset of $V \times V$, we denote by $G + F$ the graph on vertex set V, the edge set of which is $E \vartriangle F$ (the symmetric difference between E and F). Such a set F is called an *edition* of G (we may also abusively say that $G + F$ is an edition). A vertex $v \in V$ is *affected* by an edition F whenever F contains an edge incident to v. Given a graph family $\mathcal{F}$ and given a graph $G = (V, E)$, a subset $F \subseteq V \times V$ is an *optimal* $\mathcal{F}$*-edition* of G if F is a set of minimum cardinality such that $G + F \in \mathcal{F}$. Whenever we constrain F to be disjoint from E, we say that F is a *completion*, whereas if F is asked to be a subset of E, then F is an *edge deletion*.

1.1 Critical cliques

The notions of critical clique and critical clique graph have been introduced in [15] in the context of leaf power of graphs. More recently, it has been successfully used in various editing problems such as Cluster Editing [11], Bicluster Editing [22].

Definition 1.1 *A* critical clique *of a graph G is a clique K which is a module and is maximal under this property.*

It follows from definition that the set $\mathcal{K}(G)$ of critical cliques of a graph G defines a partition of its vertex set V.

Definition 1.2 *Given a graph $G = (V, E)$, its* critical clique graph $\mathcal{C}(G)$ *has vertex set $\mathcal{K}(G)$ and edge set $E(\mathcal{C}(G))$ with*

$$(K, K') \in E(\mathcal{C}(G)) \Leftrightarrow \forall v \in K, v' \in K', (v, v') \in E(G)$$

Let us note that given a graph G, its critical clique graph $\mathcal{C}(G)$ can be computed in linear time with modular decomposition algorithm (see [23] for example).

The following lemma was used in the construction of quadratic kernels for Cluster Editing and Bicluster Editing problems in [22].

Lemma 1.3 *[22] Let $G = (V, E)$ be a graph. If H is the graph $G + \{(u, v)\}$ with $(u, v) \in V \times V$, then $|\mathcal{K}(H)| \leqslant |\mathcal{K}(G)| + 4$.*

The following lemma shows that for special graph families, critical cliques are robust under optimal edition.

Lemma 1.4 *Let $\mathcal{F}$ be an hereditary graph family closed under true twin addition. For any graph $G = (V, E)$, there exists an optimal $\mathcal{F}$-edition (resp. $\mathcal{F}$-deletion, $\mathcal{F}$-completion) F such that any critical clique of $G + F$ is the disjoint union of a subset of critical cliques of G.*

Proof. We prove the statement for the edition problem. Same arguments applies for edge deletion and edge completion problem.

Let F be an optimal $\mathcal{F}$-edition of G such that the number i of critical cliques of G which are clique modules in $H = G + F$ is maximum. Assume by contradiction that $i < c$ and denote $\mathcal{K}(G) = \{K_1, \ldots, K_c\}$, where $K_1, \ldots, K_i$ are clique modules in H and $K_{i+1}, \ldots, K_c$ are no longer clique modules in H. So, let x be a vertex of K_{i+1} such that the number of edges of F incident to x is minimum among the vertices of K_{i+1}. Roughly speaking, we will modify F by editing all vertices of K_{i+1} like x. Let V_x be the subset $V \setminus (K_{i+1} \setminus \{x\})$. As $\mathcal{F}$ is hereditary, $H_x = H[V_x]$ belongs to $\mathcal{F}$ and, as $\mathcal{F}$ is closed under true twin addition, reinserting $|K_{i+1}|-1$ true twins of x in H_x results in a graph H' belonging to $\mathcal{F}$. It follows that $F' = E(G) \vartriangle E(H')$ is an $\mathcal{F}$-edition of G. By the choice of x, we have $|F'| \leqslant |F|$. Finally let us remark that by construction $K_1, \ldots, K_i$ and K_{i+1} are clique modules of H': contradicting the choice of F. □

In other words, for an hereditary graph family $\mathcal{F}$ which is closed under true twin addition and for any graph G, there always exists an optimal $\mathcal{F}$-edition F (resp. $\mathcal{F}$-deletion, $\mathcal{F}$-completion) such that 1) any edge between two vertices of a same critical clique of G is an edge of $G + F$, and 2) between two distinct critical cliques $K, K' \in \mathcal{K}(G)$, either $K \times K' \in E(G + F)$ or $(K \times K') \cap E(G + F) = \emptyset$. From now on, every considered optimal edition (resp. deletion, completion) is supposed to verify this property.

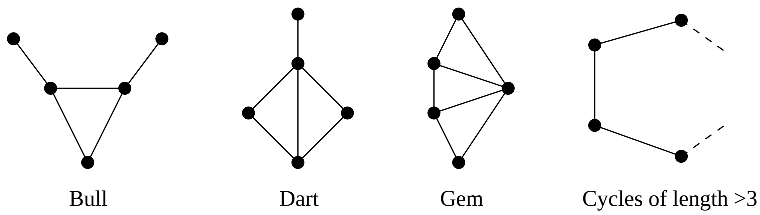

Figure 1: Forbidden induced subgraphs of a 3-leaf power.

1.2 Leaf powers

Definition 1.5 *Let T be an unrooted tree whose leaves are one-to-one mapped to the elements of a set V. The k-*leaf power *of T is the graph T^k, with $T^k = (V, E)$ where $E = \{(u, v) \mid u, v \in V$ and $d_T(u, v) \leqslant k\}$. We call T a k-*leaf root *of T^k.*

It is easy to see that for any k, the k-leaf power family of graphs satisfies Lemma 1.4. In this paper we focus on the class of 3-leaf powers for which several characterizations is known.

Theorem 1.6 *[3] For a graph G, the following conditions are equivalent:*

1. *G is a 3-leaf power.*
2. *G is $\{$bull, dart, gem, $C_{>3}\}$-free, $C_{>3}$ being the cycles of length at least 4. (see Figure 1).*
3. *The critical clique graph $\mathcal{K}(G)$ is a forest.*

For the fixed parameterized tractability of the 3-LEAF POWER EDITION, with respect to parameter k being the size of the editing set, the complexity bound of $O((2k + 8)^k.nm)$ is proposed in [5]. The proofs of our kernel for the 3-LEAF POWER EDITION problem is based on the critical clique graph characterization and on the following new one which uses the join composition of two graphs.

Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be two disjoint graphs and let $S_1 \subseteq V_1$ and $S_2 \subseteq V_2$ be two non empty subsets of vertices. Then the *join composition* of G_1 and G_2 on S_1 and S_2, denoted $(G_1, S_1) \otimes (G_2, S_2)$, results in the graph $H = (V_1 \cup V_2, E_1 \cup E_2 \cup (S_1 \times S_2))$ (see Figure 2).

Theorem 1.7 *Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be two connected 3-leaf powers. Then the graph $H = (G_1, S_1) \otimes (G_2, S_2)$, with $S_1 \subseteq V_1$ and $S_2 \subseteq V_2$, is a 3-leaf power if and only if one of the following conditions holds:*

1. *S_1 and S_2 are two cliques respectively of G_1 and G_2 and if S_1 (resp. S_2) is not critical, then G_1 (resp. G_2) is a clique.*
2. *there exists a vertex $v \in V_1$ such that $S_1 = N(v) \cup \{v\}$ and $S_2 = V_2$ is a clique.*

Proof.

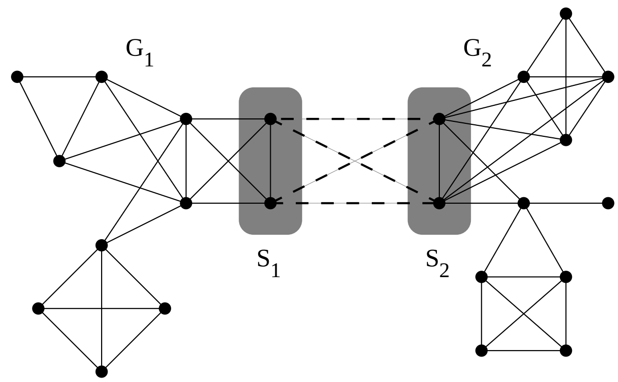

Figure 2: The join composition $H = (G_1, S_1) \otimes (G_2, S_2)$ creates the doted edges. As G_1 and G_2 are two 3-leaf powers and as the subsets S_1 and S_2 of vertices are critical cliques of respectively G_1 and G_2, by Theorem 1.7, H is also a 3-leaf power.

$\Leftarrow$ If condition (2) holds, then we simply add true twins to v and H is a 3-leaf power. Assume S_1 and S_2 are two cliques. If S_1 and S_2 are both critical cliques of respectively G_1 and G_2, then the critical clique graph $\mathcal{C}(H)$ is clearly the tree obtained from $\mathcal{C}(G_1)$ and $\mathcal{C}(G_2)$ by adding the edges between S_1 and S_2. By Theorem 1.6, H is a 3-leaf power. For $i = 1$ or 2, if G_i is a clique and $S_i \subset V(G_i)$, then S_i and $V(G_i) \setminus S_i$ are critical cliques in H. Again, it is easy to see that $\mathcal{C}(H)$ is a tree.

$\Rightarrow$ First, let us notice that if S_1 and S_2 are not cliques, then H contains a chordless 4-cycle, which is forbidden. So let us assume that S_1 is not a clique but S_2 is. Then S_1 contains two non-adjacent vertices x and y. If $d_{G_1}(x, y) > 2$, then H contains a gem. To see this, consider π a shortest x, y-path in G_1. Together with any vertex $v \in S_2$, the vertices of π induce a cycle at length at least 5 in H. By construction the only possible chords are incident to v. So any 4 consecutive vertices on π plus the vertex v induce a gem. It follows that there exists in G_1 a vertex u which dominates x and y. Again, as H is chordal, u has to be adjacent to v and therebyF_2 $u \in S_1$. Now if there exists a vertex in $V(G_2) \setminus S_2$, as G_2 is connected, there exists two adjacent vertices, $v \in S_2$ and $w \in V(G_2) \setminus S_2$. But, $\{w, u, x, y, v\}$, induce a dart in H: contradicting that H is 3-leaf power. So, $S_2 = V(G_2)$ and G_2 is a clique. Finally, assume by contraction again that u has a neighbor $w \in V(G_1) \setminus S_1$. Considering a vertex v of S_2, the set of vertices $\{w, x, y, u, v\}$ induces an obstruction in H, whatever the adjacency between w and $\{x, y\}$ is. So, $N(u) \cup \{u\} \subset S_1$. Conversely, if S_1 contains a vertex $w \notin N(u)$, $\{w, x, y, u, v\}$ induces an obstruction in H. So, $S_1 = N(u) \cup \{u\}$, as expected in condition (2).

Assume now that both S_1 and S_2 are cliques. If S_1 and S_2 are not modules in respectively G_1 and G_2, then we can find a bull in H. Assume that only S_1 is not a module *i.e.* there exist $x, y \in S_1$ and $u \in V(G_1) \setminus S_1$ such that $(u, x) \in E(G_1)$ and $(u, y) \notin E(G_1)$. If $S_2 \neq V(G_2)$, then again H has a bull induced by $\{u, x, y, v, w\}$ with $v \in S_2$ and $w \in V(G_2) \setminus S_2$, w neighbor of v. Otherwise, either condition (2) holds or y has a neighbor w in $V(G_1) \setminus S_1$. The latter case is impossible since we find in H an obstruction induced by $\{u, x, y, v, w\}$ whatever the adjacency between w and $\{u, x\}$ is. Finally assume that S_1

and S_2 are modules. But consider the case that S_1 is not critical (the case S_2 is not critical is symmetric). Then there exists a critical clique $C_1 \in \mathcal{K}(G_1)$ containing S_1. Denote by x a vertex of S_1 and by y a vertex of $C_1 \setminus S_1$. If $V(G_1) \neq C_1$, then G_1 contains two non-adjacent vertices, say u and u'. If $u = x$ and $u' \notin C_1$, then as G_1 is connected, we can choose u' and $w \notin C_1$ such that $\{u', w, x, y, v\}$ with $v \in S_2$ is a bull in H. Otherwise we can choose u and u' both adjacent to the vertices of C_1, and then $\{u, u', x, y, v\}$ would induce a dart in H. It follows that if S_1 is not critical, then condition (1) holds.

□

In order to prove the correctness of the reduction rules, the following observation will be helpful to apply Theorem 1.7.

Observation 1.8 *Let C be a critical clique of a 3-leaf power $G = (V, E)$. For any $S \subseteq V$, if $C \setminus S$ is not a critical clique of the induced subgraph $G[V \setminus S]$, then the connected component of $G[V \setminus S]$ containing C is a clique.*

Proof. Assume that $C \setminus S$ is not a critical clique of $G[V \setminus S]$, *i.e.* though $C \setminus S$ is a clique module in $G[V \setminus S]$, it is not maximal. Let $x \notin S$ be a vertex such that $C \cup \{x\}$ is a clique module of $G[V \setminus S]$. Then x belongs to a critical clique C' of G adjacent to C in $\mathcal{C}(G)$. It follows that S has to contain the union of all the critical cliques of G adjacent to C in $\mathcal{C}(G)$ but C' (otherwise $C \cup \{x\}$ could not be a module of $G[V \setminus S]$), and all the critical cliques of G adjacent to C' in $\mathcal{C}(G)$ but C (for the same reason). This means that the connected component containing C in $G[V \setminus S]$ is a subset of $C \cup C'$ which is a clique. □

Finally, let us conclude this preliminary study of 3-leaf powers by a technical lemma required in the proof of the last reduction rule.

Lemma 1.9 *Let $G = (V, E)$ be a 3-leaf power. Any cycle C of length at least 5 in G contains four distinct vertices a, b, c, d (appearing in this order along C) with ab and cd edges of C such that $ad \in E$, $ac \in E$ and $bd \in E$.*

Proof. As the 3-leaf power graphs form an hereditary family, the subgraph H of G induced by the vertices of the cycle C is a 3-leaf power with at least 5 vertices. As H is not a tree, it contains a critical clique K of size at least 2. Let a and d be two distinct vertices of K. As $|C| \geq 5$, observe that there exist two distinct vertices b and c, distinct from a and d, such that a, b, c and d appear in this order along C and that ab and cd are edges of C. As K is a clique module, any vertex adjacent to some vertex in K neighbors all the vertices of K. It follows that $ad \in E$, $ac \in E$ and $bd \in E$. □

2 A cubic kernel for the 3-leaf power edition problem

In this section, we present five preprocessing rules the application of which leads to a cubic kernel for the 3-LEAF POWER EDITION problem. The first rule is the trivial one which gets rid of connected components of the input graph that are already 3-leaf powers. Rule 2.1 is trivially safe.

Rule 2.1 *If G has a connected component C such that $G[C]$ is 3-leaf power, then remove C from G.*

The next rule was already used to obtain a quadratic kernel for the parameterized cluster editing problem [22]. It bounds the size of any critical clique in a reduced instance by $k+1$.

Rule 2.2 *If G has a critical clique K of size $|K| > k+1$, then remove $|K| - k - 1$ vertices of K from $V(G)$.*

The safeness of Rule 2.2 follows from the fact that Lemma 1.4 applies to 3-leaf powers.

2.1 Branch reduction rules

We now assume that the input graph G is reduced under Rule 2.1 (*i.e.* none of the connected component is a 3-leaf power) and Rule 2.2 (*i.e.* critical cliques of G have size at most $k+1$). The next three reduction rules use the fact that the critical clique graph of a 3-leaf power is a tree. The idea is to identify induced subgraphs of G, called *branch*, which corresponds to subtrees of $\mathcal{C}(G)$. That is a branch of G is an induced subgraph which is already a 3-leaf power. More precisely:

Definition 2.1 *Let $G = (V, E)$ be a graph. An induced subgraph $G[S]$, with $S \subseteq V$, is a* branch *if S is the disjoint union of critical cliques $K_1, \dots, K_r \in \mathcal{K}(G)$ such that the subgraph of $\mathcal{C}(G)$ induced by $\{K_1, \dots, K_r\}$ is a tree.*

Let $B = G[S]$ be a branch of a graph G and let $K_1, \dots, K_r$ be the critical cliques of G contained in S. We say that K_i ($1 \leqslant i \leqslant r$) is an *attachment point* of the branch B if it contains a vertex x such that $N_G(x)$ intersects $V(G) \setminus S$. A branch B is a *l-branch* if it has a l attachment points. Our next three rules deal with 1-branches and 2-branches.

In the following, we denote by B^R the subbranch of B in which the vertices of the attachment points have been removed. If P is an attachment point of B, then the set of neighbors of vertices of P in B is denoted $N_B(P)$.

Lemma 2.2 *Let $G = (V, E)$ be a graph and B be a 1-branch of G with attachment point P. There exists an optimal 3-leaf power edition F of G such that*

1. *the set of affected vertices of B is a subset of $P \cup N_B(P)$ and*
2. *in $G + F$, the vertices of $N_B(P)$ are all adjacent to the same vertices of $V \setminus B^R$.*

Proof. Let F be an arbitrary edition of G into a 3-leaf power. We construct from F a (possibly) smaller edition which satisfies the two conditions above.

Let C be the critical clique of $H = G + F$ that contains P and set $C' = C \setminus B^R$. By Lemma 1.4, the set of critical cliques of G whose vertices belong to $N_B(P)$ contains two kind of cliques: those, say $K_1, \dots, K_c$, whose vertices are in C or adjacent to the vertices of C in H', and those, say $K_{c+1}, \dots, K_h$ whose vertices are not adjacent to the vertices of C is H'. For $i \in \{1, \dots, h\}$, let C_i be the connected component of B^R containing K_i.

Let us consider the three following induced subgraphs: G_1 the subgraph of G whose connected components are $C_1, \dots, C_c$; G_2 the subgraph of G whose connected components are $C_{c+1}, \dots, C_h$;

and finally the subgraph G' of H induced by $V \setminus B^R$. Let us notice that these three graphs G_1, G_2 and G' are 3-leaf power. By Observation 1.8, if C' is not a critical clique of G', then the connected component of G' containing C' is a clique. Similarly, if K_i, for any $1 \leqslant i \leqslant c$, is not a critical clique of G_1, the connected component of G_1 in which it is contained is a clique. Thus, by Theorem 1.7, the disjoint union H' of G_2 and $(G', C') \otimes (G_1, \{K_1, \dots, K_c\})$ is a 3-leaf power. Now by construction, the edge edition set F' such that $H' = G + F'$ is a subset of F. Moreover the vertices of B affected by F' all belong to $P \cup N_B(P)$, which proves the first point.

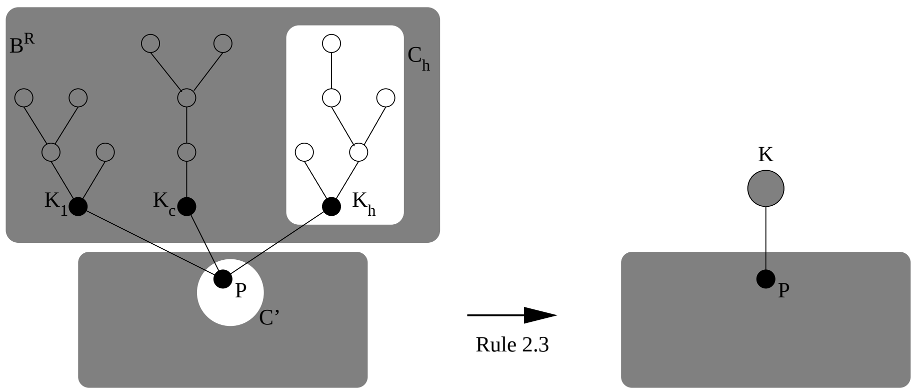

Figure 3: On the left, a 1-branch B, with attachment point P, in which the component C_h of B^R is distinguished in white. On the right, the effect of Rule 2.3 which replace B^R by a clique K of size $\min\{|N_B(P)|, k+1\}$.

We now consider an optimal edition F that satisfies the first point. To state the second point, we focus on the relationship between the critical cliques K_i and C' in $H = G + F$. If some K_i is linked to C' in H (i.e. $c > 1$), it means that the cost of adding the missing edges between K_i and C' (which, by Theorem 1.7, would also result in a 3-leaf power) is lower than the cost of removing the existing edge between K_i and C': $|K_i|.|C' \setminus P| \leqslant |K_i|.|P|$. On the other hand, if some K_j is not linked to C' in H (i.e. $c < h$), we conclude that $|P| \leqslant |C' \setminus P|$. Finally, if both cases occur, we have $|P| = |C' \setminus P|$, and we can choose to add all or none of the edges between K_i and C'. In both cases, we provide an optimal edition of G into a 3-leaf powerin which, the vertices of $N_B(P)$ are all adjacent to the same vertices of $V \setminus B^R$. □

We can now state the first 1-branch reduction rule whose safeness directly follows from Lemma 2.2.

Rule 2.3 *If G contains a 1-branch B with attachment point P, then remove from G the vertices of B^R and add a new critical clique of size $\min\{|N_B(P)|, k+1\}$ adjacent to P.*

Our second 1-branch reduction rule considers the case where several 1-branches are attached to the rest of the graph by a join. The following lemma shows that under certain cardinal condition, the vertices of such 1-branches are not affected by an optimal edition.

Lemma 2.3 *Let $G = (V, E)$ be a graph for which a 3-leaf power edition of size at most k exists. Let $B_1, \dots, B_l$ ($l \geqslant 2$) be 1-branches, the attachment points $P_1, \dots, P_l$ of which all have the same neighborhood N in $V \setminus \cup_{i=1}^{l} V(B_i)$. If $\sum_{i=1}^{l} |P_i| > 2k+1$, then, in any 3-leaf power optimal* *edition F of G, N has to be a critical clique of $G+F$ and none of the vertices of $\cup_{i=1}^{l} V(B_i)$ is affected.*

Proof. We just show that any optimal 3-leaf power edition F of G has to transform N into a critical clique, which directly implies the second part of the result. First, notice that since G is reduced under Rule 2.2, any attachment point P_i satisfies $|P_i| \leqslant k+1$.

Assume that F does not edit N into a clique: *i.e.* there are two vertices a and b of N such that $(a,b) \notin E(G+F)$. For any pair of vertices $u_i \in P_i$ and $u_j \in P_j$ with $i \neq j$, the set $\{a,b,u_i,u_j\}$ cannot induce a chordless cycle in $H = G+F$, which implies that the vertices of P_i or those of P_j are affected. It follows that among the attachment points, the vertices of at most one are not affected by F. As the P_i's have size at most $k+1$, the size of F has to be at least $k+1$: contradicting the assumptions. So N is a clique in $G+F$.

Now, assume that N is not a module of $G+F$: *i.e.* there exists $w \notin N$ such that for some $x,y \in N$ we have $(x,w) \in E(G+F)$ and $(y,w) \notin E(G+F)$. As $|F| \leqslant k$, there exist two vertices $u_i \in P_i$ and $u_j \in P_j$, such that $u_iu_j \notin E(G+F)$. But, together with x,y and w, u_i and u_j induce a dart in $G+F$, what contradicts Theorem 1.6. So, in $G+F$, the set of vertices N has to be a clique module.

Finally, let us notice that N has to be critical in $G+F$, otherwise it would imply that there exists a vertex $v \notin N$ that has been made adjacent to at least $k+1$ vertices of $\cup_{i=1}^{l} B_i$, implying that $|F| > k$: contradiction. □

By Lemma 2.3, if there exists a 3-leaf power edition F of G such that $|F| \leqslant k$, then the 1-branches $B_1, \ldots, B_l$ can be safely replaced by 2 critical cliques of size $k+1$. This gives us the second 1-branch reduction rule.

Rule 2.4 *If G has several 1-branches $B_1, \ldots, B_l$ ($l \geqslant 2$), the attachment points $P_1, \ldots, P_l$ of which all have the same neighborhood N in $V \setminus \cup_{i=1}^{l} V(B_i)$ and if $\sum_{i=1}^{l} |P_i| > 2k+1$, then remove from G the vertices of $\cup_{i=1}^{l} V(B_i)$ and add two new critical cliques of size $k+1$ neighboring exactly N.*

2.2 The 2-branch reduction rule

Let us consider a 2-branch B of a graph $G = (V,E)$ with attachment points P_1 and P_2. The subgraph of G induced by the critical cliques of the unique path from P_1 to P_2 in $\mathcal{C}(B)$ is called the *main path* of B and denoted *path*(B). We say that B is *clean* if P_1 and P_2 are leaves of $\mathcal{C}(B)$ and denote by Q_1 and Q_2 the critical cliques which respectively neighbor P_1 and P_2 in B.

Lemma 2.4 *Let B be a clean 2-branch of a graph $G = (V,E)$ with attachment points P_1 and P_2 such that path(B) contains at least 5 critical cliques. Then there exists an optimal 3-leaf power edition F of G such that*

1. *if path(B) is a disconnected subgraph of $G+F$, then F may contain a min-cut of path(B);*
2. *and in any case, the other affected vertices of B belongs to $P_1 \cup Q_1 \cup P_2 \cup Q_2$.*

Proof. Let F be an arbitrary optimal 3-leaf power edition of G. We call C_1 and C_2 the critical cliques of $G+F$ that respectively contain P_1 and P_2 (possibly, C_1 and C_2 could be the same), and denote $C_1 \setminus B^R$ and $C_2 \setminus B^R$ respectively by C_1' and C_2' (see Figure 4). We will construct from F another optimal 3-leaf power edition F' of G satisfying the statement.

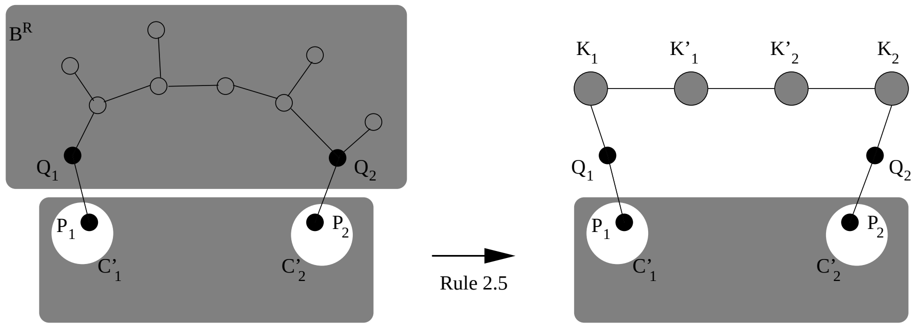

Figure 4: A 2-branch B on the left (only pendant critical cliques are hanging on $path(B)$ since we can assume that the graph is reduced by the previous rules). On the right, the way Rule 2.5 reduces B.

- *Assume that F disconnects $path(B)$.* First of all, it is clear that for any subset F_1 of F, if F_2 is an optimal edition of $H_1 = G + F_1$, then $F' = F_1 \cup F_2$ is an optimal 3-leaf power edition of G. We use this fact in the following different cases. Assume that F contains the edges $F_1 := P_1 \times Q_1$ and consider the graph $H_1 := G + F_1$. We call B_1 the 1-branch $B \setminus P_1$ of H_1 whose attachment point is P_2. Then, Lemma 2.2 applies to B_1 and provides from F an optimal 3-leaf power edition of H_1 F_2 where the edited vertices of B_1 are contained in $P_2 \cup Q_2$. By the previous observation, it follows that $F_1 \cup F_2$ is an optimal edition for G that respects conditions (1) and (2). We proceed similarly if F contains the edges $P_2 \times Q_2$. Now, we consider that F does not contain $P_1 \times Q_1$ or $P_2 \times Q_2$. In that case, there exists $F_1 \subset F$ which is a minimal cut of $path(B)$ disjoint from $P_1 \times Q_1$ and $P_2 \times Q_2$. Then, there are two connected component in $B + F_1$, the one containing P_1, say B_1, and the one containing P_2, say B_2. The subgraphs of $H_1 := G + F_1$, B_1 and B_2 are 1-branches with respectively P_1 and P_2 for attachment points. So, Lemma 2.2 applies to B_1 and B_2 and provides, from F an optimal 3-leaf power edition of H_1 F_2 where the edited vertices of B_1 and B_2 are contained in $P_1 \cup P_2 \cup Q_1 \cup Q_2$. To conclude, remark that, if F_1 is not a minimum (for cardinality) cut of $path(B)$, we could replace F_1 by such a minimum cut, and perform a similar 3-leaf power edition for G with size strictly lower than $|F|$, what contradicts the choice of F. It follows that $F_1 \cup F_2$ is an optimal edition for G that respects conditions (1) and (2).

- *Assume that F does not disconnect $path(B)$.* Let X_1 (resp. X_2) be the connected component of $(G + F) \setminus B^R$ containing P_1 (resp. P_2).
  We first consider the case where X_1 and X_2 are distinct connected components. By definition, B^R is a 3-leaf power and Q_1 and Q_2 are two of its critical cliques (since $path(B)$ contains at least 5 critical cliques). Moreover the subgraph X_1 (resp. X_2) is also a 3-leaf power which is a clique if C'_1 (resp. C'_2) is not a critical clique. It follows that by Theorem 1.7, the composition of these three subgraphs yields a 3-leaf power: $H' = (X_1, C'_1) \otimes (B^R, Q_1)$ is a 3-leaf power and $(H', Q_2) \otimes (X_2, C'_2)$ is a 3-leaf power. It follows that if F affects some vertices of $V(B^R) \setminus (Q_1 \cup Q_2)$, then a smaller edition could be found by removing from F the edges in $V(B^R) \times V(B^R)$. This would contradict the optimality of F.

  So assume that P_1 and P_2 belongs to the same connected component X of $(G + F) \setminus B^R$.

Let y_1 and y_2 be respectively vertices of P_1 and P_2 (in the case $C_1 = C_2$, choose $y_1 = y_2$). Let π_B and π_X be two distinct paths between y_1 and y_2 defined as follows: π_B is obtained by picking one vertex b_i in each critical clique H_i of $path(B)$ ($H_1 = P_1$ and $H_q = P_2$, with $q \geqslant 5$); π_X is a chordless path in X (thereby its vertices $x_1, \ldots, x_p$, with $x_1 = y_1$ and $x_p = y_2$ belong to distinct critical cliques, say $K_1, \ldots, K_p$ of $G + F$, with $K_1 = C'_1$ and $K_p = C'_2$). The union of these two paths results in a cycle C of length at least 5. So by Lemma 1.9, there are two disjoint edges $e = (a, b)$ and $f = (c, d)$ in C such that the edges (a, c) and (b, d) belong to $E \vartriangle F$. By construction of C, at most one of the edges e and f belongs to π_X.

- Either the edges e and f belong to π_B. W.l.o.g assume that $a = b_i$, $b = b_{i+1}$ and $c = b_j$, $d = b_{j+1}$ ($i + 1 < j$). By Lemma 1.4, F contains the set of edges $(H_i \times H_j) \cup (H_{i+1} \times H_{j+1})$. Notice that $\min\{|H_i|.|H_{i+1}|, |H_j|.|H_{j+1}|\} < |H_i|.|H_j| + |H_{i+1}|.|H_{j+1}|$. W.lo.g. assume that $\min\{|H_i|.|H_{i+1}|, |H_j|.|H_{j+1}|\} = |H_i|.|H_{i+1}|$. We will 'cut' the edges between H_i and H_{i+1}: consider the set

$$F' = (F \setminus (V \times V(B^R))) \cup (H_i \times H_{i+1})$$

  Moreover, if $H_i \neq P_1$, add to F' the edges $(C'_1 \setminus P_1) \times Q_1$ (which were previously in F) and, if $H_{i+1} \neq P_2$, add to F' the edges $(C'_2 \setminus P_2) \times Q_2$ (which were previously in F). In all cases, we have $|F'| < |F|$. As in the case where X_1 and X_2 were distinct, by Theorem 1.7, the graph $G + F'$ is a 3-leaf power: contradicting the optimality of F.
- Or the edge e belongs to π_B and f to π_X. W.l.o.g. assume that $a = b_i \in H_i$, $b = b_{i+1} \in H_{i+1}$ and $d = k_j \in K_j$, $c = k_{j+1} \in K_{j+1}$. As above, by Lemma 1.4, F contains $(H_i \times K_{j+1}) \cup (H_{i+1} \times K_j)$. Notice that $\min\{|H_i|.|H_{i+1}|, |K_j|.|K_{j+1}|\} < |H_i|.|K_{j+1}| + |H_{i+1}|.|K_j|$. If $\min\{|H_i|.|H_{i+1}|, |K_j|.|K_{j+1}|\} = |H_i|.|H_{i+1}|$, then we consider the set

$$F' = (F \setminus (V \times V(B^R))) \cup (H_i \times H_{i+1})$$

  Here again, if $H_i \neq P_1$, add to F' the edges $(C'_1 \setminus P_1) \times Q_1$ (which were previously in F) and, if $H_{i+1} \neq P_2$, add to F' the edges $(C'_2 \setminus P_2) \times Q_2$ (which were previously in F). As previously, $|F'|$ is smaller than $|F|$ and by Theorem 1.7, we can prove that $G + F'$ is a 3-leaf power. Finally, if $\min\{|H_i|.|H_{i+1}|, |K_j|.|K_{j+1}|\} = |K_j|.|K_{j+1}|$, then we consider the set

$$F' = (F \setminus (V \times V(B^R))) \cup (K_i \times K_{i+1}) \cup ((C'_1 \setminus P_1) \times Q_1) \cup ((C'_2 \setminus P_2) \times Q_2)$$

  Again $|F'|$ is smaller than $|F|$ and by Theorem 1.7, we can prove that $G + F'$ is a 3-leaf power. In any case, we found a better 3-leaf power edition F', contradicting the optimality of F.

□

Rule 2.5 *Let G be a graph having a clean 2-branch B such that $path(B)$ is composed by at least 8 critical cliques. Then remove from G all the vertices of $V(B)$ except those of $P_1 \cup Q_1 \cup P_2 \cup Q_2$ and add four new critical cliques:*

- *K_1 (resp. K_2) of size $k + 1$ adjacent to Q_1 (resp. Q_2);*

- *K'_1 (resp K'_2) adjacent to K_1 (resp. K_2) and such that K'_1 and K'_2 are adjacent and $|K'_1|.|K'_2|$ equals the min-cut of path(B).*

Proof. Let B' be the 2-branch replacing B after the application of the rule. It is easy to see that by construction the min-cut of B' equals the min-cut of *path*(B). Moreover the attachment points P_1 and P_2 and their respective neighbors Q_1 and Q_2 are unchanged. It follows by Lemma 2.4 that any optimal edition F of G corresponds to an optimal edition F' of G', the graph reduced by Rule 2.5, such that $|F| = |F'|$. □

2.3 Kernel size and time complexity

Let us discuss the time complexity of the reduction rules. The 3-leaf power recognition problem can be solved in $O(n + m)$ [2]. It follows that Rule 2.1 requires linear time. To implement the other reduction rules, we fist need to compute the critical clique graph $\mathcal{C}(G)$. As noticed in [22], $\mathcal{C}(G)$ can be built in $O(n+m)$. For instance, to do so, we can compute in linear time the modular decomposition tree of G, which is a classical and well-studied problem in algorithmic graph theory (see [23] for a recent paper). Given $\mathcal{K}(G)$, which is linear in the size of G, it is easy to detect the critical cliques of size at least $k+1$. So, Rule 2.2 requires linear time. A search on $\mathcal{C}(G)$ can identify the 1-branches. It follows that the two 1-branches reduction rules (Rule 2.3 and Rule 2.4) can also be applied in $O(n + m)$ time. Let us now notice that in a graph reduced by the first four reduction rules, a 2-branch is a path to which pendant vertices are possibly attached. It follows that to detect a 2-branch B, such that *path*(B) contains at least 5 critical cliques, we first prune the pendant vertices and then, identify in $\mathcal{C}(G)$ the paths containing at least 5 vertices (for instance, by proceeding a DFS starting on $\mathcal{C}(G)$ at a vertex of degree at least 3, if it exists, otherwise the problem is trivial). This shows that Rule 2.5 can be carried in linear time.

Theorem 2.5 *The parameterized* 3-LEAF POWER EDITION *problem admits a cubic kernel. Given a graph G, a reduced instance can be computed in linear time.*

Proof. The discussion above established the time complexity to compute a kernel. Let us determine the kernel size. Let $G = (V, E)$ be a reduced graph (*i.e.* none of the reduction rules applies to G) which can be edited into a 3-leaf power with a set $F \subseteq V \times V$ such that $|F| \leqslant k$. Let us denote $H = G + F$ the edited graph. We first show that $\mathcal{C}(H)$ has $O(k^2)$ vertices (*i.e.* $|\mathcal{K}(H)| \in O(k^2)$). Then Lemma 1.3 enables us to conclude.

We say that a critical clique is affected if it contains an affected vertex and denote by A the set of the affected critical cliques. As each edge of F affects two vertices, we have that $|A| \leqslant 2k$. Since H is a 3-leaf power, its critical clique graph $\mathcal{C}(H)$ is a tree. Let T be the minimal subtree of $\mathcal{C}(H)$ that spans the affected critical cliques. Let us observe that if B is a maximal subtree of $\mathcal{C}(H) - T$, then none of the critical cliques in B contains an affected vertices and thus B was the critical clique graph of a 1-branch of G, which has been reduced by Rule 2.3 or Rule 2.4. Let $A' \subset \mathcal{K}(H)$ be the critical cliques of degree at least 3 in T. As $|A| \leqslant 2k$, we also have $|A'| \leqslant 2k$. The connected components resulting from the removal of A and A' in T are paths. There are at most $4k$ such paths. Each of these paths is composed by non-affected critical cliques. It follows that each of them corresponds to *path*(B) for some 2-branch B of G, which has been reduced by Rule 2.5. From these observations, we can now estimate the size of the reduced graph. Attached to each of the critical cliques of $T \setminus A$, we can have 1 pendant critical clique resulting from the application of Rule 2.3. Remark that any 2-branch reduced by Rule 2.5 has no such pendant clique and that $path(B)$ contains at least 8 critical cliques. So, a considered 2-branch in $\mathcal{C}(H)$ is made of at most 14 critical cliques. Finally attached to each critical clique of A, we can have at most $(4k+2)$ extra critical cliques resulting from the application of Rule 2.4. See Figure 5 for an illustration of the shape of $\mathcal{C}(H)$. Summing up everything, we obtain that $\mathcal{K}(H)$ contains at most $4k.14 + 2k.2 + 2k.(4k+3) = 8k^2 + 66k$ critical cliques.

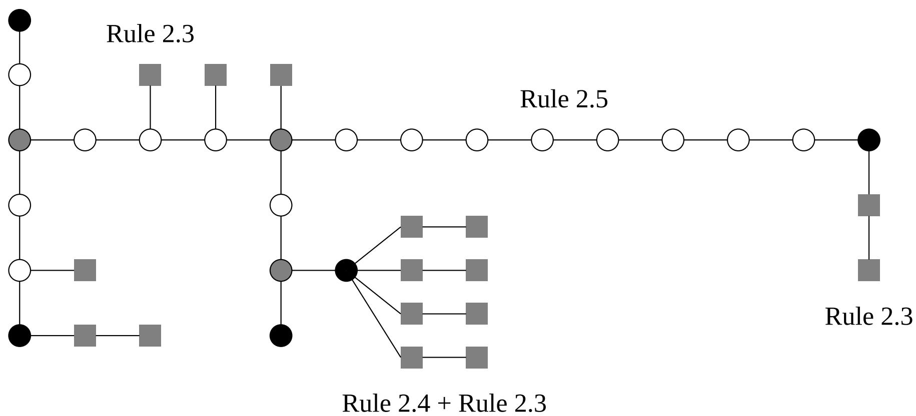

Figure 5: The black circles are the critical cliques of A, the grey ones belong to A', and the squares are the critical cliques not in T. On the figure, we can observe a 2-branch of size 8 reduced by Rule 2.5. There cannot be pendant critical cliques attached to its nodes. Application of Rule 2.3, may let a path of two critical cliques pendant to the elements of $A \cup A'$ and a single critical clique pendant to the elements of the small 2-branches. Finally, Rule 2.4 can only affect critical cliques of A.

By Lemma 1.3, we know that each edited edge in a graph, the number of critical cliques increase of at most 4. It follows that $\mathcal{K}(G)$ contains at most $8k^2 + 70k$ critical cliques. By Rule 2.2, each critical clique of the reduced graph has size at most $k+1$. This implies that the reduced graph contains at most $8k^3 + 78k^2 + 70k$ vertices, proving the $O(k^3)$ kernel size. □

We should notice that some small modifications of the branch reduction rules and a more precise analysis would improve the constants involved in the kernel size. However the cubic bound would not change.

3 Cubic kernels for edge completion only and edge deletion only

We now prove and/or adapt the previous rules to the cases where only insertions or only deletions of edges are allowed. First, observe that Rules 2.1 and 2.2 are also safe in 3-LEAF POWER COMPLETION and 3-LEAF POWER DELETION (Rule 2.2 directly follows from Lemma 1.4). We have a similar result for the 1-branches reduction rules.

Lemma 3.1 *Rule 2.3 is safe for both* 3-LEAF POWER COMPLETION *and* 3-LEAF POWER DELETION*.*

Proof. In the following, we consider an optimal solution F such that $H := G + F$ is a 3-leaf power, denote by C the critical clique containing P in H and set $C' = C \setminus B^R$.

- 3-LEAF POWER COMPLETION. To show the safeness of Rule 2.3 in this case, we will build from F an optimal completion that respects conditions of Lemma 2.2. By Lemma 1.4, we know that the set of critical cliques $\{K_1, \dots, K_h\}$ of G whose vertices belong to $N_B(P)$ are in C or adjacent to the vertices of C in H (in this case, there is no critical cliques K_i disconnected from C in H because we cannot remove edges from G). In both cases, K_i is adjacent to C' in H. For $i \in \{1, \dots, h\}$, let C_i be the connected component of B^R containing K_i. As previously, we consider G_1 the subgraph of G whose connected components are $C_1, \dots, C_h$. By Observation 1.8, if C' is not a critical clique of G', then G' is a clique. Similarly, if K_i, for any $1 \leq i \leq h$, is not a critical clique of G_1, the connected component of G_1 in which it is contained is a clique. By Theorem 1.7, it follows that the graph $H' := (H \setminus B^R, C') \otimes (G_1, \{K_1, \dots, K_h\})$ is a 3-leaf power. By construction, the edge completion set F' such that $H' = G + F'$ is a subset of F and the vertices of B affected by F all belong to $P \cup N_B(P)$. Finally, as every K_i is connected to C' in H', the vertices of $N_B(P)$ are all adjacent to the same vertices of $V \setminus B^R$.

- 3-LEAF POWER DELETION. In the case where only edges deletion are allowed, we will build from F an optimal deletion respecting the conditions of Lemma 2.2 by studying the behavior of P in H. First of all, note that if P forms a bigger critical clique in H with some vertex $x \in V \setminus B^R$, this means that F contains $P \times N_B(P)$. Thus, there is no need to do extra deletions in B^R and then we are done.

  Now consider the cases where P is critical in H or form a bigger critical clique with some K_i (*i.e.* F contains $P \times (\{K_1, \dots, K_c\} \setminus K_i$ for some i). In both cases, we have $C' = P$. By Theorem 1.7, the graph $H' := (H \setminus B^R, C') \otimes (G_1, \{K_1, \dots, K_c\})$ is a 3-leaf power, and the edge set F' used to transform G into H' is a subset of F (all the edges between C' and $\{K_1, \dots, K_c\}$) are present in H), and then we are done.

□

Lemma 3.2 *Rule 2.4 is safe for both* 3-LEAF POWER COMPLETION *and* 3-LEAF POWER DELETION.

Proof. As in Lemma 2.3, we consider $B_1, \dots, B_l$ 1-branches of G, the attachment points $P_1, \dots, P_l$ of which all have the same neighborhood N and satisfy $\sum_{i=1}^{l} |P_i| > 2k + 1$.

- 3-LEAF POWER COMPLETION. In this case, same arguments as the ones used in the proof of Lemma 2.3 hold. We briefly detail them. First, assume that N was not transformed into a clique by an optimal completion F. To get rid of all the C_4's involving 2 non-adjacent vertices of N and P_i, P_j, $i \neq j$, the only possibility is to transform $\cup_{i=1}^{l} P_i$ into a clique, which requires more than $k + 1$ completions. Moreover, N must also become a clique module, otherwise we would find darts that once again would imply to transform $\cup_{i=1}^{l} P_i$ into a clique which is impossible. Finally, N must be critical (otherwise, at least one insertion for each vertex of $\cup_{i=1}^{l} P_i$ must be done), thus implying that no vertex in $\cup_{i=1}^{l} P_i$ is affected by an optimal edition.

- 3-LEAF POWER DELETION. Firstly, observe that if N is not a clique, then any optimal deletion in that case would have to destroy at least $k+1$ C_4 with edges deletion only, which is impossible. The same arguments used previously hold again in this case to conclude that N must become a critical clique in the modified graph.

□

Now, observe that the 2-branch reduction rule can apply directly to 3-LEAF POWER DELETION, but will not be safe for 3-leaf power completion. Indeed, in the proof of Lemma 2.4, if we look at the cycle C of G containing vertices of B, it might be needed to delete edges between two consecutive critical cliques along C. to transform $\mathcal{C}(C)$ into a tree. Nevertheless, it is possible to bound the number of vertices of $path(B)$ in the case of 3-LEAF POWER COMPLETION by looking at the edges modifications needed to make a cycle chordal (see Lemma 3.4).

Lemma 3.3 *Rule 2.5 is safe for* 3-LEAF POWER DELETION.

Proof. Let F be an arbitrary optimal 3-leaf power deletion of G. We call C_1 and C_2 the critical cliques of $H := G + F$ that respectively contain P_1 and P_2, $C'_1 := C_1 \setminus B^R$ and $C'_2 := C_2 \setminus B^R$. We will construct from F another optimal 3-leaf power edition F' of G satisfying the conditions of Lemma 2.4.

We have two cases to consider : 1) either $path(B)$ is disconnected in H or 2) $path(B)$ is still connected in H. Case 1) works exactly as the first case studied in the proof of Lemma 2.4, and thus there exists an optimal deletion on which conditions of Lemma 2.4 holds.

If case (2) holds, *i.e.* if $path(B)$ is still connected in H, then P_1 and P_2 must belong to distinct connected components of $H \setminus B^R$, say X_1 and X_2 (otherwise H would admit a chordless cycle as induced subgraph). Furthermore, notice that we must have $P_1 = C_1$ and $P_2 = C_2$ in H. Indeed, if P_1 forms a critical clique with some vertex $x \in V \setminus B^R$, this means F must contain $P_1 \times Q_1$ which is not, by hypothesis. Similarly, if P_1 forms a critical clique with some vertex $x \in Q_1$, then F must contain edges between Q_1 and $N_{B^R}(Q_1)$ which is not (the cases for P_2 are symmetric). By definition, B^R is a 3-leaf power, and so are X_1 and X_2. By Theorem 1.7, it follows that the composition of these three subgraphs yields a 3-leaf power : $H' = (X_1, P_1) \otimes (B^R, Q_1)$ and $(H', Q_2) \otimes (X_2, P_2)$ are 3-leaf powers. It follows that if F affects some vertices of $B^R \setminus (Q_1 \cup Q_2)$, then a smaller deletion could be found, what contradicts the optimality of F. □

We now prove a result usefull to conclude on the size of the kernel in the 3-LEAF POWER COMPLETION problem.

Lemma 3.4 *Let G be a graph admitting a clean 2-branch B such that $path(B)$ is composed by at least $k+4$ critical cliques. If P_1 and P_2 belong to the same connected component in G, then there is no 3-leaf power completion of size at most k.*

Proof. Let G be a graph with a clean 2-branch B on which conditions of the Lemma 3.4 applies, and let F be an optimal 3-leaf power completion of G. As P_1 and P_2 belong to the same connected component in G, we have a cycle C of size at least $k+4$ in $\mathcal{K}(G)$. Consider the subgraph of $\mathcal{K}(G)$ induced by the critical cliques of C. By Lemma 1.4 we know that $\mathcal{C}(C)$ must be a tree ; let us call F' the set of edges transforming C into a tree. It is known that F' is a triangulation of this cycle [5]. Moreover, every triangulation of a n-cycle needs at least $n-3$ chords, what implies that $|F'| > k$, which is impossible. □

Rule 3.1 *Let G be a graph having a clean 2-branch B with attachment points P_1 and P_2 such that path(B) is composed by at least $k+3$ critical cliques.*

- *if P_1 and P_2 belong to the same connected component in $G \backslash B^R$, then there is no completion of size at most k.*
- *otherwise, remove from G all the vertices of $V(B)$ except those of $P_1 \cup Q_1 \cup P_2 \cup Q_2$ and add all possible edges between Q_1 and Q_2.*

Proof. The first point follows directly from Lemma 3.4. To see the second point, notice that we are in the case where P_1 and P_2 belong to different connected components (which corresponds to the second case of the proof of Lemma 2.4). As edges insertion are allowed, the safeness of this rule is due to this particular case. □

Theorem 3.5 *The parameterized* 3-LEAF POWER COMPLETION *and* 3-LEAF POWER DELETION *problem admit cubic kernels. Given a graph G a reduced instance can be computed in linear time.*

Proof. We detail separately completion and deletion.

- 3-LEAF POWER COMPLETION. As in the proof of Theorem 2.5, we consider $H := G + F$ with G being reduced and F being an optimal completion and we denote by T the minimal subtree of $\mathcal{C}(H)$ spanning the set of affected critical cliques A. As noticed before, we have $|A| \leqslant 2k$.

  First, remark that the only difference between this case and 3-LEAF POWER EDITION concerns the 2-branch reduction rule. This means that the only difference will occur in the number of vertices of the paths resulting from the removal of A and A' in T (A' being critical cliques of degree at least 3 in T). Due to both Lemma 3.4 and Rule 3.1 we know that a 2-branch in $\mathcal{C}(H)$ is made of at most $2k+6$ critical cliques: corresponding to a path of at most $k+3$ critical cliques (otherwise there is no optimal completion), each one having a pendant critical clique (by Rule 2.3). This means that $\mathcal{C}(H)$ contains at most $4k.(2k+6)+2k.2+2k.(4k+3) = 16k^2+34k$ critical cliques. By Lemma 1.3, we know that each edited edge creates at most 4 new critical cliques. If follows that $\mathcal{K}(G)$ contains at most $16k^2+38k$ critical cliques. By Rule 2.2, each critical clique of the reduced graph has size at most $k+1$, thus implying that the reduced graph contains at most $16k^3+54k^2+38k$ vertices, proving the $O(k^3)$ kernel size.
- 3-LEAF POWER DELETION. The rules used for the 3-leaf power deletion problem are exactly the same than the one used to obtain a cubic kernel for 3-LEAF POWER EDITION. Thus, the size of a reduced instance of 3-leaf power deletion will be exactly the same as one of a reduced instance of 3-LEAF POWER EDITION.

□

4 Conclusion

By proving the existence of a cubic kernel of the 3-LEAF POWER EDITION problem, we positively answered an open problem [4, 6]. The natural question is now whether the cubic bound could be improved. A strategy could be, as for the quadratic kernel of 3-HITTING SET [20] which is based of the linear kernel of VERTEX COVER [18], to consider the following subproblem:

PARAMETERIZED FAT STAR EDITION PROBLEM
Input: An undirected graph $G = (V, E)$.
Parameter: An integer $k \geqslant 0$.
Question: Is there a subset $F \subseteq V \times V$ with $|F| \leqslant k$ such that the graph $G + F = (V, E \vartriangle F)$ is a 3-leaf power and its critical clique graph $\mathcal{C}(G+F)$ is a star (we say that $G+F$ is a *fat star*).

It can be shown that small modifications of the Rule 2.1, 2.2 and 2.4 yield a quadratic kernel for the FAT STAR EDITION problem [21]. A linear bound may be helpful to improve the kernel of the CLOSEST 3-LEAF POWER since it can be shown that the neighborhood of any big enough critical clique of the input graph as to be edited into a fat star.

References

[1] H. Bodlaender, R.G. Downey, M.R. Fellows, and D Hermelin. On problems without polynomial kernels (extended abstract). In *Automata, Languages and Programming, 35th International Colloquium (ICALP)*, volume 5125 of *Lecture Notes in Computer Science*, pages 563–574, 2008.

[2] A. Brandstädt and V.B. Le. Structure and linear time recognition of 3-leaf powers. *Information Processing Letters*, 98(4):133–138, 2006.

[3] M. Dom, J. Guo, F. Hüffner, and R. Neidermeier. Error compensation in leaf root problems. *Algorithmica*, 44:363–381, 2006.

[4] M. Dom, J. Guo, F. Hüffner, and R. Neidermeier. Closest 4-leaf power is fixed-parameter tractable. *Discrete Applied Mathematics*, 2008. To appear.

[5] M. Dom, J. Guo, F. Hüffner, and R. Niedermeier. Error compensation in leaf root problems. In *International Symposium on Algorithms and Computation (ISAAC)*, number 3341 in Lecture Notes in Computer Science, pages 389–401, 2004.

[6] M. Dom, J. Guo, F. Hüffner, and R. Niedermeier. Extending the tractability border for closest leaf powers. In *International Workshop on Graph Theoretical Concepts in Computer Science (WG)*, number 3787 in Lecture Notes in Computer Science, pages 397–408, 2005.

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

[8] M.R. Fellows, M. Langston, F. Rosamond, and P. Shaw. Efficient parameterized preprocessing for cluster editing. In *International Symposium on Fundamentals of Computation Theory (FCT)*, number 4639 in Lecture Notes in Computer Science, pages 312–321, 2007.

[9] M.R. Garey and D.S. Jonhson. *Computers and intractability: a guide to the theory on NP-Completeness*. W.H. Freeman and Co., 1979.

[10] J. Gramm, J. Guo, F. Hüffner, and R. Niedermeier. Graph-modeled data clustering: Exact algorithms for clique generation. *Theory Comput. Syst.*, 38(4):373–392, 2005.

[11] J. Guo. A more effective linear kernelization for cluster editing. In *International Symposium on Combinatorics, Algorithms, Probabilistic and Experimental Methodologies (ESCAPE)*, volume 4614 of *Lecture Notes in Computer Science*, pages 36–47, 2007.

[12] J. Guo and R. Niedermeier. Invitation to data reduction and problem kernelization. *SIGACT News*, 38(1):31–45, 2007.

[13] P. Kearney and D. Corneil. Tree powers. *Journal of Algorithms*, 29(1):111–131, 1998.

[14] W. Kennedy and G. Lin. 5-th phylogenetic root construction for strictly chordal graphs. In *International Symposium on Algorithms and Computation (ISAAC)*, number 3827 in Lecture Notes in Computer Science, pages 738–747, 2005.

[15] G.H. Lin, P.E. Kearney, and T. Jiang. Phylogenetic k-root and steiner k-root. In *International Symposium on Algorithms and Computation (ISAAC)*, number 1969 in Lecture Notes in Computer Science, pages 539–551, 2000.

[16] R. Motwani and M. Sudan. Computing roots of graphs is hard. *Discrete Applied Mathematics*, 54(1):81–88, 1994.

[17] A. Natazon, R. Shamir, and R. Sharan. Complexity classification of some edge modification problems. *Discrete Applied Mathematics*, 113:109–128, 2001.

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

[19] N. Nishimura, P. Ragde, and D. Thilikos. On graph powers for leaf-labeled trees. *Journal of Algorithms*, 42(1):69–108, 2002.

[20] N. Nishimura, P. Ragde, and D. Thilikos. Smaller kernels for hitting set problems of constant arity. In *International Workshop on Parameterized and Exact Computation (IWPEC)*, number 3162 in Lecture Notes in Computer Science, pages 121–126, 2004.

[21] A. Perez. Complexité paramétrique et recherche de noyaux. Master's thesis, Université de Montpellier 2, France, 2008.

[22] F. Protti, M. Dantas da Silva, and J.L. Szwarcfiter. Applying modular decomposition to parameterized cluster editing problems. *Theory of Computing Systems*, 2007.

[23] M. Tedder, D. Corneil, M. Habib, and C. Paul. Simpler linear-time modular decomposition via recursive factorizing permutations. In *Automata, Languages and Programming (ICALP)*, volume 5115 of *Lecture Notes in Computer Science*, pages 634–645, 2008.