Kernelization Dichotomies for Hitting Subgraphs Under Structural Parameterizations

Marin Bougeret
LIRMM, Université de Montpellier, CNRS, France

Bart M. P. Jansen
Eindhoven University of Technology, The Netherlands

Ignasi Sau
LIRMM, Université de Montpellier, CNRS, France

Abstract

For a fixed graph H, the H-Subgraph Hitting problem consists in deleting the minimum number of vertices from an input graph to obtain a graph without any occurrence of H as a subgraph. This problem can be seen as a generalization of Vertex Cover, which corresponds to the case $H = K_2$. We initiate a study of H-Subgraph Hitting from the point of view of characterizing structural parameterizations that allow for polynomial kernels, within the recently active framework of taking as the parameter the number of vertex deletions to obtain a graph in a "simple" class $\mathcal{C}$. Our main contribution is to identify graph parameters that, when H-Subgraph Hitting is parameterized by the vertex-deletion distance to a class $\mathcal{C}$ where any of these parameters is bounded, and assuming standard complexity assumptions and that H is biconnected, allow us to prove the following sharp dichotomy: the problem admits a polynomial kernel if and only if H is a clique. These new graph parameters are inspired by the notion of *$\mathcal{C}$-elimination distance* introduced by Bulian and Dawar [Algorithmica 2016], and generalize it in two directions. Our results also apply to the version of the problem where one wants to hit H as an induced subgraph, and imply in particular, that the problems of hitting minors and hitting (induced) subgraphs have a substantially different behavior with respect to the existence of polynomial kernels under structural parameterizations.

2012 ACM Subject Classification Theory of computation → Graph algorithms analysis; Theory of computation → Parameterized complexity and exact algorithms

Keywords and phrases hitting subgraphs, hitting induced subgraphs, parameterized complexity, polynomial kernel, complexity dichotomy, elimination distance

Digital Object Identifier 10.4230/LIPIcs.ICALP.2024.33

Category Track A: Algorithms, Complexity and Games

Related Version *Full Version*: `https://arxiv.org/abs/2404.16695` [11]

Funding *Bart M. P. Jansen*: Supported by the Dutch Research Council (NWO) through Gravitation-grant NETWORKS-024.002.003.
Ignasi Sau: Supported by the French project ELIT (ANR-20-CE48-0008-01).

1 Introduction

The theory of parameterized complexity deals with *parameterized problems*, which are decision problems in which a positive integer k, called the *parameter*, is associated with every instance x. One of the pivotal notions in the domain is that of *kernelization* [5, 15, 20, 24, 39], which is a polynomial-time algorithm that reduces any instance (x, k) of a parameterized problem to an equivalent instance (x', k') of the same problem whose size is bounded by $f(k)$ for some function f, which is the *size* of the kernelization. A kernelization algorithm, or just *kernel*, can be seen as a preprocessing procedure with provable guarantees, and it is fundamental to find kernels of the smallest possible size, ideally polynomial. Identifying which parameterized problems admit polynomial kernels is one of the most active areas within Parameterized Complexity (cf. for instance [24]).

When dealing with a problem where the goal is to find a (say, small) subset of vertices S of an input graph G satisfying some property, such as Vertex Cover, it is natural to consider as the parameter the size of the desired set S. Assuming that the problem admits a polynomial kernel parameterized by $|S|$, as it is the case for Vertex Cover and many other problems [15, 24], we can ask whether the problem still admits polynomial kernels when the parameter is (asymptotically) smaller than the solution size. The goal of this approach is to provide better preprocessing guarantees, as well as to understand what is the limit of the polynomial-time "compressibility" of the considered problem. For problems defined on graphs, apart from using the solution size as a parameter, it is common to consider so-called *structural parameters*, which quantify some structural property of the input graph that can be seen as a measure of its "complexity". Among structural parameters, the most successful is probably *treewidth* [15, 35], but unfortunately taking the treewidth of the input graph as the parameter does not allow for polynomial kernels for essentially all natural optimization problems, unless $\mathsf{NP} \subseteq \mathsf{coNP/poly}$ [6, 8]. The same applies to another relevant graph parameter called *treedepth*, denoted by td and defined as the minimum number of rounds needed to obtain the empty graph, where each round consists of removing one vertex from each connected component.

In fact, the lower bound proofs go through for parameterizations for which the value on a disconnected graph is the *maximum*, rather than the *sum*, of the values of its components, and whose value is polynomially bounded in the size of the graph. Hence, to be able to obtain positive kernelization results, we need to turn to parameterizations other than *width measures*. This motivates to consider structural parameters that quantify the "distance from triviality", a concept first coined by Guo, Hüffner, and Niedermeier [28]. The idea is to take as the parameter the vertex-deletion distance of a graph to a "trivial" graph class where the considered problem can be solved efficiently. This paradigm has proved very successful for a number of problems, in particular for Vertex Cover. In an influential work, Jansen and Bodlaender [30] showed that Vertex Cover admits a polynomial kernel when parameterized by the feedback vertex number of the input graph, which is the vertex-deletion distance to the "trivial" class of forests. This result triggered a number of results in the area, aiming to characterize the "trivial" families $\mathcal{F}$ for which Vertex Cover admits a polynomial kernel under this parameterization [10, 12, 25, 29, 40].

Let us mention some of these results that are relevant to our work. Bougeret and Sau [12] proved that Vertex Cover admits a polynomial kernel parameterized by the vertex-deletion distance to a graph of bounded treedepth. This result was further generalized into two orthogonal directions, namely by considering a more general problem or a more general target graph class $\mathcal{F}$. For the former generalization, Jansen and Pieterse [33] proved that the following problem also admits a polynomial kernel parameterized by the vertex-deletion distance to a graph of bounded treedepth: for a fixed finite set of connected graphs $\mathcal{M}$, the $\mathcal{M}$-Minor Deletion problem consists in deleting the minimum number of vertices from an input graph to obtain a graph that does not contain any of the graphs in $\mathcal{M}$ as a minor. Note that this problem (vastly) generalizes Vertex Cover, which corresponds to the case $\mathcal{M} = \{K_2\}$. For the latter generalization, Bougeret, Jansen, and Sau [10] proved that Vertex Cover admits a polynomial kernel parameterized by the vertex-deletion distance to a graph of bounded *bridge-depth*, which is a parameter that generalizes treedepth and the feedback vertex number. It turns out that, under the assumption that the target graph class $\mathcal{F}$ is minor-closed, the property of $\mathcal{F}$ having bounded bridge-depth is also a *necessary* condition for Vertex Cover admitting a polynomial kernel. Another complexity dichotomy of this flavor has been achieved by Dekker and Jansen [18] for the Feedback Vertex Set problem (with another characterization of the target graph class $\mathcal{F}$). These complexity dichotomies, while precious, are unfortunately quite hard to obtain, and the current knowledge seems still far from obtaining dichotomies of this type for general families of problems, such as for $\mathcal{M}$-Minor Deletion for any finite family of graphs $\mathcal{M}$. Indeed, it is wide open whether $\mathcal{M}$-Minor Deletion admits a polynomial kernel parameterized by the solution size for any set $\mathcal{M}$ containing only non-planar graphs [23, 34], so considering parameters smaller than the solution size is still out of reach.

Our contribution. We consider an alternative generalization of Vertex Cover by considering (induced) *subgraphs* instead of *minors*. Namely, for a fixed graph H, the H-Subgraph Hitting problem is defined as deleting the minimum number of vertices from an input graph to obtain a graph without any occurrence of H as a subgraph. The H-Induced Subgraph Hitting problem is defined analogously by forbidding occurrences of H as an *induced* subgraph. (As we shall see later, both problems behave in the same way with respect to our results.) Note that both problems correspond to Vertex Cover for the case $H = K_2$ and therefore are indeed generalization of it. As opposed to the case for hitting minors, it is well-known that both the H-Subgraph Hitting and H-Induced Subgraph Hitting problems admit polynomial kernels parameterized by the solution size for any graph H [1, 21].

Therefore, it does make sense to parameterize these problems by structural parameters in the "distance from triviality" spirit, and this is the main focus of this article. To the best of our knowledge, this is an unexplored topic, besides all the literature for Vertex Cover discussed above. Our main result is to identify structural parameters that allow to provide sharp dichotomies for these problems *depending on the forbidden (induced) subgraph H*.

Before presenting our results, we proceed to motivate and define these structural parameters. They are inspired by the following parameter, first introduced by Bulian and Dawar [13, 14] and further studied, for instance, in [2, 22, 29, 31, 41, 42]. For a fixed class of graphs $\mathcal{H}$, the *$\mathcal{H}$-elimination distance* of a graph G, denoted by $\mathsf{ed}_{\mathcal{H}}(G)$, is defined by mimicking the above definition of treedepth and replacing "empty graph" with "a graph in $\mathcal{H}$". The recursive definitions of treedepth and $\mathcal{H}$-elimination distance suggest the notion of *elimination forest*, which is the forest-like process of vertex removals from the considered graph to obtain either an empty graph for treedepth, or a graph in $\mathcal{H}$ for $\mathcal{H}$-elimination distance. Suppose now that $\mathcal{H}$ is defined by the exclusion of a fixed graph H as a subgraph or as an induced subgraph. Formally, let $\mathcal{F}_{\bar{H}}$ (resp. $\mathcal{F}^{\mathsf{ind}}_{\bar{H}}$) be the class of graphs that exclude a fixed graph H as a subgraph (resp. induced subgraph). For this particular case, the notion of $\mathcal{F}_{\bar{H}}$-elimination distance (or $\mathcal{F}^{\mathsf{ind}}_{\bar{H}}$-elimination distance) can be interpreted as a generalization of treedepth where, in the last round of the elimination process, the vertices that do not belong to any occurrence of $\mathcal{H}$ as a subgraph (or induced subgraph) can be deleted "for free". We generalize the notion of $\mathcal{H}$-elimination distance by allowing "free removal" of vertices not contained in a copy of H in *every round* of the elimination process, rather than just the last; we denote the corresponding parameter by $\mathsf{ved}^{+}_{\mathcal{F}_{\bar{H}}}$ (or $\mathsf{ved}^{+}_{\mathcal{F}^{\mathsf{ind}}_{\bar{H}}}$), where "v" stands for the removal of *vertices*, in order to distinguish this parameter from the one defined below (see Section 2 for the formal definitions of these parameters). Our first main result is the following somehow surprising dichotomy, which states that, under the assumption that H is biconnected, whenever H has a non-edge, the problem is unlikely to admit a polynomial kernel. Our result applies to both the induced and non-induced versions of the problem.

▶ **Theorem 1.1.** *Let H be a biconnected graph, let $\lambda \geq 1$ be an integer, and assume that* $\mathsf{NP} \not\subseteq \mathsf{coNP/poly}$. *$H$-Subgraph Hitting (resp. H-Induced Subgraph Hitting) admits a polynomial kernel parameterized by the size of a given vertex set X of the input graph G such that $\mathsf{ved}^{+}_{\mathcal{F}_{\bar{H}}}(G - X) \leq \lambda$ (resp. $\mathsf{ved}^{+}_{\mathcal{F}^{\mathsf{ind}}_{\bar{H}}}(G - X) \leq \lambda$) if and only if H is a clique.*

Note that a graph G satisfies $\mathsf{ved}^{+}_{\mathcal{F}_{\bar{H}}}(G) = 0$ (resp. $\mathsf{ved}^{+}_{\mathcal{F}^{\mathsf{ind}}_{\bar{H}}}(G) = 0$) if and only if G does not contain H as a subgraph (resp. induced subgraph), so the setting $\lambda = 0$ corresponds to the parameterization by solution size which always admits a polynomial kernel [1, 21]; this is why we assume that $\lambda \geq 1$ in the statement of Theorem 1.1.

Theorem 1.1 shows that the behavior of the considered problems in terms of the existence of polynomial kernels drastically changes as soon as one edge is missing from H (under the biconnectivity assumption, which is needed in the reduction). To the best of our knowledge, this is the first time that such a dichotomy, in terms of H, is found with respect to the existence of polynomial kernels. It is worth mentioning that, with respect to the existence of certain fixed-parameter tractable algorithms parameterized by treewidth, dichotomies of this flavor exist for hitting subgraphs [17], induced subgraphs [43], or minors [4].

The proof of Theorem 1.1 consists of two independent pieces. On the one hand, we need to prove that both problems admit a polynomial kernel when H is a clique (note that, in that case, both problems are equivalent, as any H-subgraph is induced). On the other hand, we need to provide a kernelization lower bound for all other graphs H (cf. Theorem 3.2), and here is where we need the hypothesis that $\mathsf{NP} \not\subseteq \mathsf{coNP/poly}$ and, for technical aspects of the reduction, that H is biconnected.

In fact, we provide a kernel that is more general than the one stated in Theorem 1.1. Also, on the negative side, we present another lower bound incomparable to that of Theorem 1.1. For the former, we provide a polynomial kernel, when H is a clique, for a parameter that is more powerful than $\mathsf{ved}^{+}_{\mathcal{F}_{\bar{H}}}$ (or $\mathsf{ved}^{+}_{\mathcal{F}^{\mathsf{ind}}_{\bar{H}}}$). This more powerful parameter is somehow inspired by the parameter bridge-depth mentioned before [10], which is a generalization of treedepth in which, in every round of the elimination process, we are allowed to remove subgraphs in each component that are more general than just single vertices. In our setting, it turns out that we can afford to remove vertex sets $T \subseteq V(G)$ that induce connected subgraphs that do not contain H as a subgraph (or induced subgraph) and that are "weakly attached" to the rest of the graph, meaning that each connected component of $G - T$ has at most one neighbor in T. If H is biconnected, it is easily seen that the "candidate" sets T to be removed can be assumed to be connected unions of blocks (biconnected components) of G, and this is why we call this parameter $\mathsf{bed}^{+}_{\mathcal{F}_{\bar{H}}}$ (or $\mathsf{bed}^{+}_{\mathcal{F}^{\mathsf{ind}}_{\bar{H}}}$), where "$\mathsf{b}$" stands for the removal of *blocks*. For any two graphs G and H, the following inequalities, as well as the corresponding ones for the induced version, follow easily from the definitions (cf. Section 2):

$$\mathsf{td}(G) \geq \mathsf{ed}_{\mathcal{F}_{\bar{H}}}(G) \geq \mathsf{ved}^{+}_{\mathcal{F}_{\bar{H}}}(G) \geq \mathsf{bed}^{+}_{\mathcal{F}_{\bar{H}}}(G). \tag{1}$$

We prove the following result, where K_t denotes the clique on t-vertices, and note that in this case the induced and non-induced versions of the problem coincide.

▶ **Theorem 1.2.** *Let $t \geq 3$ and $\lambda \geq 1$ be fixed integers. The K_t-Subgraph Hitting problem admits a polynomial kernel parameterized by the size of a given vertex set X of the input graph G such that $\mathsf{bed}^{+}_{\mathcal{F}_{\bar{K_t}}}(G - X) \leq \lambda$.*

Note that for $t = 2$, the parameter $\mathsf{bed}^{+}_{\mathcal{F}_{\bar{K_t}}}$ is exactly treedepth, and therefore Theorem 1.2 can be seen as a far-reaching generalization of the main result of [12], that is, a polynomial kernel for Vertex Cover parameterized by the vertex deletion distance to a graph of bounded treedepth. Also, note that Equation 1 and Theorem 1.2 imply that the same dichotomy stated in Theorem 1.1 holds if we substitute "$\mathsf{ved}^{+}_{\mathcal{F}_{\bar{H}}}(G - X) \leq \lambda$" for "$\mathsf{bed}^{+}_{\mathcal{F}_{\bar{H}}}(G - X) \leq \lambda$", and the same for the induced version.

As for strengthening our hardness results, we present kernelization lower bounds for H-Subgraph Hitting and H-Induced Subgraph Hitting, when H is not a clique, parameterized by the vertex-deletion distance to a graph of constant *treedepth*. By Equation 1, lower bounds for treedepth are stronger than the ones of Theorem 1.1. However, in our next main result, we need a condition on H that is stronger than biconnectivity, namely the non-existence of a *stable cutset*, that is, a vertex separator that induces an independent set.

▶ **Theorem 1.3.** *Let H be a graph on h vertices that is not a clique and that has no stable cutset. H-Subgraph Hitting and H-Induced Subgraph Hitting do not admit a polynomial kernel parameterized by the size of a given vertex set X of the input graph G such that $\mathsf{td}(G - X) = \mathcal{O}(h)$, unless $\mathsf{NP} \subseteq \mathsf{coNP/poly}$.*

Note that, for $t \geq 4$, the graph K_t minus one edge satisfies the conditions of Theorem 1.3. The mere existence of a graph H satisfying the conditions of Theorem 1.3 is remarkable, as it shows that (induced) subgraph hitting problems behave differently than minor hitting problems. Indeed, as mentioned before, it is known [33] that, for *every* finite family $\mathcal{M}$ of connected graphs, the $\mathcal{M}$-Minor Deletion problem admits a polynomial kernel parameterized by the vertex-deletion distance to a graph of constant treedepth.

Dekker and Jansen [18] asked if for every finite set of graphs $\mathcal{M}$, $\mathcal{M}$-Minor Deletion admits a polynomial kernel parameterized by the vertex-deletion distance to a graph with constant $\mathsf{exc}(\mathcal{M})$-elimination distance, where $\mathsf{exc}(\mathcal{M})$ is the class of graphs that excludes all the graphs in $\mathcal{M}$ as a minor. Theorem 1.3 shows that, by Equation 1, for the problems of excluding subgraphs or induced subgraphs, the answer to this question is negative.

Finally, let us mention another consequence of our results. Agrawal et al. [2] proved, among other results, that for every hereditary target graph class $\mathcal{C}$ satisfying some mild assumptions, parameterizing by the vertex-deletion distance to $\mathcal{C}$ and by the $\mathcal{C}$-elimination distance are equivalent from the point of view of the existence of fixed-parameter tractable algorithms. Our results imply, in particular, that the same equivalence does *not* hold with respect to the existence of polynomial kernels in this "distance from triviality" setting, namely for problems defined by the exclusion of (induced) subgraphs.

Organization of the paper. In Section 2 we provide an overview of the main ideas of the kernelization algorithm, which is our main technical contribution. The formal description of the kernel and its analysis, which are quite lengthy, can be found in the full version of the article [11]. In Section 3 we present our hardness results (with some proofs deferred to the full version as well), and in Section 4 we discuss some directions for further research.

2 Overview of the kernelization algorithm

In this section we sketch the main ideas of the kernelization algorithm stated in Theorem 1.2, along with intuitive explanations and the required definitions.

Preliminaries. Given a graph G and $C \subseteq V(G)$, we denote by $N(C) = \bigcup_{v \in C} N(v) \setminus C$ and $G - C = G[V(G) \setminus C]$. Given a graph H, and a subgraph (resp. induced subgraph) F of G, we say that F is a *copy* (resp. *induced copy*) of H if F is isomorphic to H. We say that G is H-free (resp. H-induced free) if there is no copy of H (resp. induced copy) in G. Given two disjoint subsets $A, B \subseteq V(G)$, we say that there is an edge between A and B if there exists $e \in E(G)$ such that $|e \cap A| = |e \cap B| = 1$. Otherwise, A and B are said to be *anticomplete*. When $\mathcal{P} = \{V_1, \ldots, V_m\}$ is a set of subsets of $V(G)$, we let $V(\mathcal{P}) = \bigcup_{V_i \in \mathcal{P}} V_i$. A *$t$-clique* is a set $K \subseteq V(G)$ such that $G[K]$ is a clique and $|K| = t$. For any integer $n \in \mathbb{N}$, we denote by $[n] = \{1, \ldots, n\}$. We study the following problem(s) for a fixed graph H.

> H-SH (for H-Subgraph Hitting)
> **Input**: A graph G.
> **Objective**: Find a set $S \subseteq V(G)$ of minimum size such that $G - S$ is H-free.

We denote by H-ISH (for H-Induced Subgraph Hitting) the variant of the above problem where we impose that $G - S$ is H-*induced* free. We denote by $\mathsf{opt}^H(G)$ the optimal value of the considered problem for G, or simply $\mathsf{opt}(G)$ when H is clear from the context.

Let us now introduce the main graph measures used in this paper.

▶ **Definition 2.1.** *Let H be a fixed graph. For a graph G, define $\mathsf{ved}^+_{\mathcal{F}_{\bar{H}}}(G)$ as*

$$\begin{cases} 0 & \text{if } V(G) = \emptyset, \\ \mathsf{ved}^+_{\mathcal{F}_{\bar{H}}}(G - v) & \text{if } v \text{ is a vertex that is not in any copy of } H, \\ \max_{C_i} \mathsf{ved}^+_{\mathcal{F}_{\bar{H}}}(C_i) & \text{if } G \text{ has connected components } C_1, \ldots, C_c \text{ with } c \geq 2, \\ 1 + \min_{v \in V(G)} \mathsf{ved}^+_{\mathcal{F}_{\bar{H}}}(G - v) & \text{otherwise.} \end{cases}$$

We define $\mathsf{ved}^+_{\mathcal{F}^{\mathsf{ind}}_{\bar{H}}}$ in the same way as $\mathsf{ved}^+_{\mathcal{F}_{\bar{H}}}$, except that we replace "in any copy of H" by "in any induced copy of H". Note that the notation $\mathsf{ved}^+_{\mathcal{F}_{\bar{H}}}$ is motivated by the fact that it corresponds to *vertex* elimination distance, with additional power of removing "free" vertices not in any copy of H. Note also that even though there could be multiple vertices v which satisfy the second criterion, the value is well-defined since it does not matter which one is picked; the second case will apply until all such vertices have been removed. As the case where $H = K_t$ plays an important role in this paper, for the sake of shorter notation we use the shortcut ved^+_t to denote the parameter $\mathsf{ved}^+_{\mathcal{F}_{\bar{K_t}}}$ (or $\mathsf{ved}^+_{\mathcal{F}^{\mathsf{ind}}_{\bar{K_t}}}$, which is the same).

To define the parameters $\mathsf{bed}^+_{\mathcal{F}_{\bar{H}}}$ and $\mathsf{bed}^+_{\mathcal{F}^{\mathsf{ind}}_{\bar{H}}}$, it is convenient to introduce the following definitions (see Figure 1).

▶ **Definition 2.2** (root and pending component). *Given a fixed graph H and a connected graph G, we say that a set $T \subseteq V(G)$ is a* root of G *if*

- *$T \neq \emptyset$, $G[T]$ is connected and H-free, and*
- *for any connected component C of $G - T$, $|N(C) \cap T| = 1$.*

We extend to notion of root to any graph G as follows. For any graph G with connected components $\mathcal{C}$, we say that a set $\mathcal{T} = \{T_C \mid C \in \mathcal{C}\}$ is a root of G *if for any $C \in \mathcal{C}$, T_C is a root of $G[C]$. We define $V(\mathcal{T}) = \bigcup_{C \in \mathcal{C}} T_C$, and $E(\mathcal{T})$ as the set of edges that have both their endpoints inside $V(\mathcal{T})$.*

Given a graph G, a root $\mathcal{T}$ of G, and a vertex $v \in V(\mathcal{T})$, we define the pending component of v relatively to $\mathcal{T}$*, denoted by $C^{\mathcal{T}}(v)$, as the connected component of v in the graph obtained from G by removing all edges $e \subseteq E(\mathcal{T})$. We extend the notation to any subset $Z \subseteq V(\mathcal{T})$ with $C^{\mathcal{T}}(Z) = \bigcup_{v \in Z} C^{\mathcal{T}}(v)$. When the root is clear from context, we use $C(v)$ instead of $C^{\mathcal{T}}(v)$.*

We define an *induced root* in the same way, except that we replace "$G[T]$ is connected and H-free" by "$G[T]$ is connected and H-induced free". Note that any graph admits a root, by taking for example a single vertex (to play the role of T_C) in each connected component.

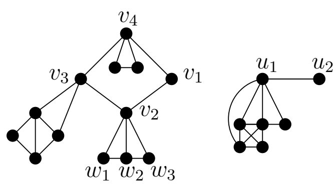

Figure 1 In this example we consider that $H = K_4$, and we denote by C_1 and C_2 the two connected components of G (where $v_1 \in C_1$). Observe that $\mathcal{T} = (T_1, T_2)$ is a root of the depicted graph with $T_1 = \{v_1, v_2, v_3, v_4\}$ and $T_2 = \{u_1, u_2\}$. We have $C(v_1) = \{v_1\}$ and $C(v_2) = \{v_2, w_1, w_2, w_3\}$. Finally, taking $T_1' = T_1 \cup \{w_1\}$ and $\mathcal{T}' = \{T_1', T_2\}$ would not be a root as T_1' is not a root of $G[C_1]$.

▶ **Observation 2.3.** *Let G be a graph and $\mathcal{T}$ be a root of G. For any $v \in V(\mathcal{T})$, there is no edge between $C(v) \setminus \{v\}$ and $V(G) \setminus C(v)$.*

▶ **Definition 2.4.** *Let H be a fixed graph. For a graph G, we define $\mathsf{bed}^+_{\mathcal{F}_{\bar{H}}}(G)$ as*

$$\begin{cases} 0 & \text{if } V(G) = \emptyset, \\ \mathsf{bed}^+_{\mathcal{F}_{\bar{H}}}(G - v) & \text{if } v \text{ is a vertex that is not in any copy of } H, \\ \max_{C_i} \mathsf{bed}^+_{\mathcal{F}_{\bar{H}}}(C_i) & \text{if } G \text{ has connected components } C_1, \ldots, C_c \text{ with } c \geq 2, \\ 1 + \min_{T \subset V(G)} \mathsf{bed}^+_{\mathcal{F}_{\bar{H}}}(G - T) & \text{otherwise, where } T \text{ ranges over all roots of } G. \end{cases}$$

We define $\mathsf{bed}^+_{\mathcal{F}^{\mathsf{ind}}_{\bar{H}}}$ in the same way as $\mathsf{bed}^+_{\mathcal{F}_{\bar{H}}}$, except that we replace "where T ranges over all roots of G" by "where T ranges over all induced roots of G". Again, as the case where $H = K_t$ plays an important role in this paper, for the sake of shorter notation we use the shortcut ved^+_t to denote the parameter $\mathsf{bed}^+_{\mathcal{F}_{\bar{K}_t}}$ (or $\mathsf{bed}^+_{\mathcal{F}^{\mathsf{ind}}_{\bar{K}_t}}$, which is the same). We point out that, to make the definition of $\mathsf{bed}^+_{\mathcal{F}_{\bar{H}}}$ as simple as possible, we allowed T to range over all roots of G. However, as shown in [11, Lemma 6.9], as soon as H is biconnected, there always exists a root that is the connected union of H-(induced-)free blocks of G, hence our choice of notation to differentiate $\mathsf{bed}^+_{\mathcal{F}_{\bar{H}}}$ from $\mathsf{ved}^+_{\mathcal{F}_{\bar{H}}}$.

Given $\lambda \in \mathbb{N}$, let us now define the following variant of the considered problem, where we suppose that we are given as an additional input a modulator (corresponding to set X) to a "simple" graph $G - X$, where the simplicity is captured by bed^+_t being at most λ.

> K_t-SHM$^\lambda$ (for K_t-Subgraph Hitting given a modulator to bed^+_t at most λ)
> **Input**: A graph G and a set $X \subseteq V(G)$ such that $\mathsf{bed}^+_t(G[R]) \leq \lambda$, where $R = V(G) \setminus X$.
> **Objective**: Find a set $S \subseteq V(G)$ of minimum size such that for any t-clique Z of G, $S \cap Z \neq \emptyset$.

We denote by K_t-SHM$^\lambda_{\mathsf{p}}$ the associated parameterized decision problem with an additional k in the input, where the goal to decide whether $\mathsf{opt}(G) \leq k$, and the parameter is $|X|$.

In [11, Section 5] we prove our main positive result that we restate here with less details, and which is a reformulation of Theorem 1.2 with the notation introduced in this section:

▶ **Theorem 2.5.** *There is a polynomial kernel for K_t-SHM$^\lambda_{\mathsf{p}}$ of size $\mathcal{O}_{\lambda,t}(|X|^{\delta(\lambda,t)})$ for some function $\delta(\lambda, t)$.*

Let us now present an overview of the techniques used to establish the above result.

Warming up with VERTEX COVER**.** As an extreme simplification of our set up, let us consider the case where $t = 2$, corresponding to VERTEX COVER, and assume that X is a modulator to the simplest graph class, namely an independent set. Our kernel uses a marking procedure (cf. [11, Definition 5.5]) that corresponds to the following marking algorithm for VERTEX COVER. For any $u \in X$, mark up to $|X| + 1$ vertices $v \in R$ such that $\{u, v\} \in E(G)$. Let $M \subseteq R$ be the set of marked vertices. Observe the following "packing property" of the marking algorithm: if there exists $v \in R \setminus M$ and $u \in X$ with $\{u, v\} \in E(G)$, then there exists a "packing" $\mathcal{P} \subseteq M$ of $|X| + 1$ vertices such that for any $v' \in \mathcal{P}$, $\{v', u\} \in E(G)$ (the term "packing" may seem inappropriate here, but becomes natural for $t > 2$ as the marking algorithm will mark disjoint sets of vertices instead of distinct vertices). Now, if $R = M$ then $|R| \leq \mathcal{O}(|X|^2)$ and the instance is kernelized. Otherwise, if there exists $v \in R \setminus M$, then define a *reduced instance* as $G' = G - v$ and $k' = k$. Let us sketch why this removing step is safe, as the arguments also correspond to a very simplified version of [11, Lemma 5.14]. The only non-trivial direction is that if (G', k') is a yes-instance, then (G, k) is also a yes-instance. Given a solution Z' of (G', k'), if there exists $u \in X \setminus Z'$ such that $\{u, v\} \in E(G)$, then the packing property implies the existence of the above set $\mathcal{P}$. Thus, we get that Z' *overpays for* u: it contains one extra vertex (in this case, one instead of zero) for each $v' \in \mathcal{P}$ as we must have $\mathcal{P} \subseteq Z'$. This implies that we can restructure Z' into $\tilde{Z} = X$, while ensuring that $|\tilde{Z}| \leq |Z'|$. Now, $\tilde{Z}$ can be easily completed to a solution of G of size k (in this case, by doing nothing). By repeating this reduction rule, we get the kernel of size $\mathcal{O}(|X|^2)$.

Parts, chunks, and conflicts. Let us now point out some important ideas used to lift up the previous kernel for VERTEX COVER to K_t-$\mathrm{SHM}_{\mathsf{p}}^{\lambda}$. In the previous setting, the key property when proving the safeness of the reduction rule, given a solution Z' of (G', k'), is the following: when "adding back" a non-marked vertex $v \in R \setminus M$ to G', either there exists $u \in X \setminus Z'$ such that Z' overpays for u, or there is no edge $\{u, v\}$ for any $u \in X \setminus Z'$.

Let us now rephrase this key property in the setting of hitting t-cliques using the adapted concepts of part, chunk, and conflict; we will formally define these terms later. When "adding back" a non-marked *part* $V' \subseteq R \setminus M$ to G', we know that either there exists *a chunk* $X' \subseteq X \setminus Z'$ such that Z' overpays for X', or there is no *conflict between* X' *and* V' for any chunk X'. Observe first that, as $G - X$ is now more general than an independent set, we have to consider a packing of "parts" (subsets of vertices of R), meaning that if there is a non-marked part V' that we remove, we now set $k' = k - \mathsf{opt}(G[V'])$. The second difference is the notion of "conflict between X' and V'" that plays the role of "edge $\{u, v\}$". We say that there is no *conflict* between X' and V' if $\mathsf{conf}^t_{X'}(V') = 0$, the condition $\mathsf{conf}^t_{X'}(V') = 0$ being equivalent to the fact that we can pick only $\mathsf{opt}(G[V'])$ vertices in V', while still hitting all t-cliques in $G[X' \cup V']$ (see [11, Definition 5.2] for the formal definition of conf^t). The third difference is the notion of chunk and blocking set. A good starting point when trying to complete a solution Z' of G' to a solution Z of G is that $\mathsf{conf}^t_{X \setminus Z'}(V') = 0$. Indeed, this condition implies that there exists a set $S^\star_{V'}$ of size $\mathsf{opt}(G[V'])$ such that $S^\star_{V'}$ hits all t-cliques in $G[V' \cup (X \setminus Z')]$. Thus, $S^\star_{V'}$ is a good candidate to build a solution $Z = Z' \cup S^\star_{V'}$ of (G, k). Note that this only remains a good starting point, as Z may not be a solution: it could miss cliques using V', $(X \setminus Z')$, and other vertices in $R \setminus V'$. This condition $\mathsf{conf}^t_{X \setminus Z'}(V') = 0$ could be achieved by a marking algorithm that, for any $U \subseteq X$, marks up to $|X| + 1$ parts V' such that $\mathsf{conf}^t_U(V') > 0$, which is a generalization of the previous marking algorithm for VERTEX COVER. However, the running time and the number of marked parts by such an algorithm would not be polynomial in $|X|$, as there are too many subsets U to consider.

To overcome this issue, the trick is the notion of *maximum minimal blocking sets*, denoted by mmbs_t (cf. [11, Definition 4.1]), which is a graph parameter for which we skip the definition for the moment. What is important to state about mmbs_t here is that in [11, Theorem 2.6] we prove that there exists a function $\beta : \mathbb{N}^2 \to \mathbb{N}$ such that for every graph G, $\mathsf{mmbs}_t(G) \leq \beta(\mathsf{bed}_t^+(G), t)$. As in the kernelization algorithm we apply this to $G[R]$ and $\mathsf{bed}_t^+(G[R]) \leq \lambda$, we obtain $\mathsf{mmbs}_t(G[R]) \leq \beta(\lambda, t)$. Moreover, [11, Lemma 5.4] implies that the previous marking condition "there exists $U \subseteq X$ such that $\mathsf{conf}_U^t(V') > 0$" is equivalent to "there exists $X' \in \mathcal{X}$ such that $\mathsf{conf}_{X'}^t(V') > 0$", where $\mathcal{X} = \{X' \subseteq X \mid |X'| \leq (t-1)\beta(\lambda, t)$ and X' does not contain a t-clique$\}$ is the set of *chunks*. Observe that, as the chunks have bounded size, the marking algorithm runs in time $\mathcal{O}^\star(|X|^{(t-1)\beta(\lambda,t)})$. The conclusion is that the "triviality" of $G[R]$ ($\mathsf{bed}_t^+(G[R]) \leq \lambda$) implies that $G[R]$ has bounded mmbs_t, which allows to certify the absence of conflict (for any $U \subseteq X$) in polynomial time.

These notions of conflict, chunk, and minimal blocking set were also critical in previous work on kernelization: the notion of conflict was introduced in [30], and bounds on mmbs_2 (for Vertex Cover) have been proved for different triviality measures of $G[R]$ [3, 10, 12, 29, 30]. A first difference here is the study of mmbs_t for K_t-Subgraph Hitting, whose behavior is more complex than Vertex Cover, as discussed below when mentioning the new challenges.

Decreasing $\mathsf{bed}_t^+(G[R])$ and using recursion. Let us now informally describe the main steps of the kernel (see Figure 2). Given a graph G, we denote by $N^t(G) = \{v \in V(G) \mid \nexists$ t-clique K with $v \in K\}$ the set of vertices of G that do not occur in any copy of K_t, called the *non-K_t-vertices*. Given an input (G, X, k) of K_t-$\text{SHM}_{\mathsf{p}}^{\lambda}$, we first compute $N^t(G[R])$ and, using the algorithm of [11, Lemma 6.9], a bed_t^+-root $\mathcal{T}$ of $G[R] - N^t(G[R])$, where a bed_t^+*-root* of G is a root $\mathcal{T}$ such that $\mathsf{bed}_t^+(G - V(\mathcal{T})) = \mathsf{bed}_t^+(G) - 1$. We point out that, unlike the case for treedepth or bridge-depth, computing such a root is not straightforward, as one cannot try the a priori exponentially many possible roots to find one that decreases bed_t^+. However, the algorithm of [11, Lemma 6.9] relies on the fact that it is possible to compute in polynomial time a set of size $\mathcal{O}(n)$ that contains a bed_t^+-root. Coming back to the kernel strategy, observe that there may be edges between some $C(v)$ and $N^t(G)$, but not between $C(v)$ and $C(u)$ for $u \neq v$, and that by definition of a bed_t^+-root, $\mathsf{bed}_t^+(G[R']) < \lambda$, where $R' = R - V(\mathcal{T})$. Then, we mark a small (polynomial in $|X|$) set $M(\mathcal{T}, N^t(G[R]), G, X)$ of vertices (cf. [11, Definition 5.8]) of $V(\mathcal{T})$ using the `mark` algorithm (cf. [11, Definition 5.5]). If there exists $v \in V(\mathcal{T}) \setminus M(\mathcal{T}, N^t(G[R]), G, X)$, then we can remove $C(v)$ and decrease k by $\mathsf{opt}(G[C(v)])$ (cf. [11, Lemma 5.14]). Otherwise, $|M(\mathcal{T}, N^t(G[R]), G, X)| = \mathcal{O}(|X|^{f(\lambda,t)})$ for some function f, and thus we can move $M(\mathcal{T}, N^t(G[R]), G, X)$ to the modulator and get a new modulator $X' = X \cup M(\mathcal{T}, N^t(G[R]), G, X)$ whose size is still polynomial in $|X|$. The key point is that $\mathsf{bed}_t^+(G - X') = \mathsf{bed}_t^+(G[R']) < \lambda$, and thus we use induction on λ and make a recursive call to (G, X'), which is an input of K_t-$\text{SHM}_{\mathsf{p}}^{\lambda-1}$, leading to a kernel polynomial in $|X'|$, and thus in $|X|$.

This idea of shrinking the "root" of a decomposition of $G - X$ to decrease the "triviality measure" (here, bed_t^+) and recurse originates in [27], and was used in [12] for treedepth. It was subsequently generalized in [10], where the triviality measure is a parameter called *bridge-depth* and the equivalent of a root is a so-called *tree of bridges* for each connected component of $G[R]$.

New challenges. With respect to the strategies followed in previous work on related topics [10, 12, 18, 29, 30, 32, 33], in our setting we encounter (at least) the following three orthogonal difficulties, for which we have to develop new ideas: dealing with the non-K_t-vertices, dealing with cliques K_t for arbitrary fixed t instead of $t = 2$, and proving that there exists a function β such that for every graph G, $\mathsf{mmbs}_t(G) \leq \beta(\mathsf{bed}_t^+(G))$.

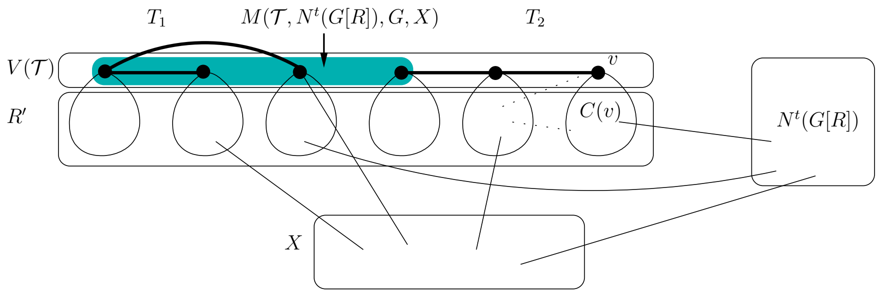

Figure 2 Main steps of the kernel. In this example $\mathcal{T} = \{T_1, T_2\}$ (edges inside T_i are in bold, and dotted edges cannot exist), and there exists a non-marked vertex v, implying that the pending component $C(v)$ will be removed.

The first difficulty is handling vertices of $N^t(G[R])$, which are vertices not belonging to a t-clique in $G[R]$. Indeed, observe that these non-K_t-vertices are "free" for bed_t^+, in the sense that $\mathsf{bed}_t^+(G[R]) = \mathsf{bed}_t^+(G[R] - N^t(G[R]))$. However, these vertices make the structure of $G[R]$ more complicated. Indeed, $\mathcal{T}$ being a root of $G[R] - N^t(G[R])$ implies that for any $T \in \mathcal{T}$ and $v \in T$, there are no edges between $C(v) \setminus \{v\}$ and other vertices in $C(u)$ for $u \in V(\mathcal{T}) \setminus \{v\}$, but there could be edges between $C(v)$ and $N^t(G[R])$. Thus, unlike in [10, 12], we cannot just bound the number of connected components of $G[R]$, and then assume that we have a single root T with simple properties (again, the root being a single vertex in [12], and a tree of bridges in [10]). Typically, here $G[R]$ could have only one connected component, but the nice structure given by $\mathcal{T}$ could be "polluted" by vertices of $N^t(G[R])$. We handle these vertices by considering a packing of "bidimensional" parts (V_i, N_i), where in particular $V_i \subseteq V(\mathcal{T})$ is a clique of size at most $t-1$ and $N_i \subseteq N^t(G[R])$, and we use a kind of generalized "sunflower-like" marking by first creating a maximal packing $\mathcal{P}$ of parts (V_i, N_i), of size at most $|X|+1$, and then recursively marking around each possible $g \in \bigcup N_i$ (see the last line of [11, Definition 5.5]).

The second difficulty is to handle t-cliques instead of edges. Indeed, assume that we just removed a pending component $C(v)$ for some $v \in V(\mathcal{T})$ and defined $k' = k - \mathsf{opt}(G[C(v)])$. Assume also that, given a solution Z' of (G', k'), we have the good starting point $\mathsf{conf}^t_{(X \cup N^t(G[R])) \setminus Z'}(C(v)) = 0$, implying, by the definition of conflict, that there exists a locally optimal solution $S_v^\star$ of $G[C(v)]$ that intersects all t-cliques of $G[C(v) \cup ((X \cup N^t(G[R])) \setminus Z')]$. However, there could also exist "spread" cliques K containing v and using vertices of $((X \cup N^t(G[R])) \setminus Z')$ and $M' \subseteq (V(\mathcal{T}) \setminus \{v\})$. These cliques may be spread across several vertices of $V(\mathcal{T})$, and by definition of a root they cannot use vertices in $C(u) \setminus \{u\}$ for any $u \in M' \cup \{v\}$ (according to Observation 2.3). To take into account the potential conflicts generated by these spread cliques, we perform $t-1$ marking steps (cf. [11, Definition 5.8]), where informally at each step we guess all possible subsets M', with $|M'| \leq t-1$, corresponding to a guessed intersection of a spread clique with previously marked vertices.

The last difficulty is to bound $\mathsf{mmbs}_t(G)$ as a function of $\mathsf{bed}_t^+(G)$ for any graph G. We first need to define the notion of blocking set adapted to our problem. Let EK_t-SH (for *Extended* K_t-Subgraph Hitting) be the problem where given $(G, \mathcal{F})$, where $\mathcal{F}$ is a set of subsets of $V(G)$ such that for any $Z \in \mathcal{F}, 1 \leq |Z| \leq t-1$ and $G[Z]$ is a clique, a solution must intersect all t-cliques of G and all $Z \in \mathcal{F}$. A *blocking set* $\mathcal{B}$ of G is a set of subsets of vertices of G such that $\mathsf{opt}(G, \mathcal{B}) > \mathsf{opt}(G)$, where $\mathsf{opt}(G, \mathcal{B})$ is the minimum size of a solution of the EK_t-SH problem with input $(G, \mathcal{B})$, meaning that any set hitting all t-cliques of G and all $Z \in \mathcal{B}$ cannot be an optimal solution of G for K_t-SH. Then, $\mathsf{mmbs}_t(G)$ is the maximum size of an inclusion-wise minimal blocking set of G. The "max-min" taste of this definition makes it difficult to handle, but fortunately we will use [11, Property 1] stating that for any β, $\mathsf{mmbs}_t(G) \leq \beta$ is equivalent to the fact that for any blocking set $\mathcal{B}$ of G, there exists $\overline{\mathcal{B}} \subseteq \mathcal{B}$ such that $\overline{\mathcal{B}}$ is still a blocking set of G and $|\overline{\mathcal{B}}| \leq \beta$. We obtain the following upper bound, which requires a considerable amount of technical work.

▶ **Theorem 2.6.** *For any graph G and any integer $t \geq 3$, it holds that $\mathsf{mmbs}_t(G) \leq \beta(\mathsf{bed}_t^+(G), t)$, where $\beta(x,t) = \underbrace{2^{t2^{t2^{\cdots t}}}}_{x \text{ times}}$ (i.e., $\beta(1) = 2^t, \beta(2) = 2^{t2^t}$, etc.)*

To explain the difficulty of proving the above theorem, let us sketch how such a bound is obtained for Vertex Cover as a function of treedepth, for example in the proof in [12] that, for every graph G, $\mathsf{mmbs}_2(G) \leq 2^{\mathsf{td}(G)}$.

Observe first that for the Vertex Cover problem, a blocking set $\mathcal{B}$ is a set of singletons, which we consider as a subset of vertices, such that any vertex cover S containing $\mathcal{B}$ is not optimal. Let us now use [11, Property 1] and consider a blocking set $\mathcal{B}$ of G, and let us show that there exists $\overline{\mathcal{B}} \subseteq \mathcal{B}$ such that $\overline{\mathcal{B}}$ is still a blocking set of G and $|\overline{\mathcal{B}}| \leq 2^{\mathsf{td}(G)}$. Consider a graph G and a "root" v of a treedepth decomposition of G, meaning that $\mathsf{td}(G - v) < \mathsf{td}(G)$. Let us consider the most complex case where there exists an optimal solution using v, another avoiding v, and $v \notin \mathcal{B}$. It is not difficult to prove that, as $\mathcal{B}$ is a blocking set of G, $\mathcal{B}_1 = \mathcal{B}$ is a blocking set of $G_1 := G - v$, and $\mathcal{B}_2 = \mathcal{B} \setminus N(v)$ is a blocking set of $G_2 := G - (\{v\} \cup N(v))$. Thus, as for any $i \in [2]$, $\mathsf{td}(G_i) < \mathsf{td}(G)$, by induction we get that there exists $\overline{\mathcal{B}}_i \subseteq \mathcal{B}_i$ such that $\overline{\mathcal{B}}_i$ is a blocking set of G_i and $|\overline{\mathcal{B}}_i| \leq 2^{\mathsf{td}(G)-1}$. As $\overline{\mathcal{B}}_1 \cup \overline{\mathcal{B}}_2$ is a blocking set of G, we get the desired bound. The problem when lifting this idea to K_t-Subgraph Hitting instead of Vertex Cover is that, when considering an optimal solution S that avoids a root v, we do not know which vertex the solution S will pick in $N(v)$. This is the reason for which we consider the more general version of the problem, namely EK_t-SH, to encode the fact that a solution of $(G, \emptyset)$ avoiding a root v must be a solution of $(G - v, \mathsf{pr}_v^t(V(G) \setminus \{v\}))$, where given two disjoint sets $A, B \subseteq V(G)$, $\mathsf{pr}_A^t(B) = \{K \cap B \mid K \text{ is a } t\text{-clique in } G[A \cup B] \text{ and } K \cap A \neq \emptyset \text{ and } K \cap B \neq \emptyset\}$. We also need to define the corresponding generalized notion of blocking set of an instance $(G, \mathcal{F})$ of EK_t-SH (cf. [11, Definition 4.1]), and not only of a graph G. Moreover, we have to keep track of the structure of $\mathcal{F}$, as there is no hope to bound $\mathsf{mmbs}_t(G, \mathcal{F})$ as a function of $\mathsf{bed}_t^+(G)$ for an arbitrary set $\mathcal{F}$. Indeed, for example, let G_ℓ be a chain of triangles of length ℓ, as depicted in Figure 4. We have $\mathsf{mmbs}_2(G_\ell) \geq \ell$, as if we let $\mathcal{B}$ be the set of top vertices of the ℓ triangles, then it can be easily seen that $\mathcal{B}$ is a minimal blocking set of G_ℓ with $|\mathcal{B}| = \ell$. Now, take $\mathcal{F} = E(G_\ell)$, $t = 4$, and observe that any solution of instance $(G_\ell, \mathcal{F})$ of EK_t-SH is a vertex cover of G_ℓ, and thus $\mathcal{B}$ is also a minimal blocking set of size ℓ for the input $(G_\ell, \mathcal{F})$ with $t = 4$, while $\mathsf{bed}_t^+(G_\ell) = 0$ as G_ℓ is K_4-free.

We resolve this problem by proving bounds on $\mathsf{mmbs}_t(G, \mathcal{F})$ only for a special type of instances that we call *clean*, which are pairs $(G, \mathcal{F})$ such that $\mathsf{opt}(G, \mathcal{F}) = \mathsf{opt}(G)$. The first main difficulty is that, when starting with a blocking set $\mathcal{B}$ of $(G, \mathcal{F})$, reducing to a graph G' with $\mathsf{bed}_t^+(G') < \mathsf{bed}_t^+(G)$ requires to remove the entire root $\mathcal{T}$ of $G - N^t(G)$, instead of just one vertex as in the treedepth case. As $|V(\mathcal{T})|$ may be arbitrarily large, we need to prove (see [11, Lemma 4.8]) that it is enough to "zoom in" on a small number of subgraphs (pending components here), allowing us to extract (by induction) a small blocking set only in each of these subgraphs. The second main difficulty is to ensure that we can reduce via recursion to smaller *clean* instances. Indeed, even if we initially consider a clean instance $(G, \emptyset)$, and even in the favorable case where $\mathcal{T}$ is just one vertex v (as in treedepth), we have to consider the case where there exists an optimal solution of G using v, another avoiding v, and $v \notin \mathcal{B}$, where $\mathcal{B}$ is fixed blocking set from which we try to extract a small one. However, observe that optimal solutions avoiding v are optimal solutions of $(G - v, \mathsf{pr}_v^t(V(G) \setminus \{v\}))$, and that $\mathsf{opt}(G - v, \mathsf{pr}_v^t(V(G) \setminus \{v\})) = \mathsf{opt}(G) = 1 + \mathsf{opt}(G - v)$ (the last equality holds since there are optimal solutions taking v). Thus, we observe that this situation leads to a *non*-clean instance $(G - v, \mathsf{pr}_v^t(V(G) \setminus \{v\}))$, but "almost clean" as $\mathsf{opt}(G - v, \mathsf{pr}_v^t(V(G) \setminus \{v\})) = \mathsf{opt}(G - v) + 1$. We treat these almost clean instances in [11, Lemma 4.5], which is the main cause of the huge growth of function β in the final bound $\mathsf{mmbs}_t(G) \leq \beta(\mathsf{bed}_t^+(G), t)$ given in Theorem 2.6. As this bound directly reverberates both in the running time and the size of the kernel (see [11, Theorem 5.16], where $\delta(\lambda, t)$ is dominated by $\beta(\lambda, t)$ for an instance (G, X) of K_t-SHM where $\mathsf{bed}_t^+(G - X) \leq \lambda$), improving this bound is crucial in order to improve the kernel size. In this direction, we provide in [11, Lemma 4.12] a significantly better upper bound for minimal blocking sets of the K_t-SH problem as a function of td instead of bed_t^+. As the proof technique is also different, we believe that this result might be of independent interest.

Finally, let us mention that in several earlier papers on kernelization using structural parameterizations, it was also crucial to understand the maximum size of an inclusion-minimal set with additional requirements on the solution to a vertex-deletion problem, for which no optimal solution can satisfy all additional requirements; these correspond to variations on the notion of blocking sets. They were explored for the problems of hitting forbidden connected minors in graphs of bounded treedepth [33], and for hitting cycles in graphs of bounded elimination distance to a forest [18], both of which lead to super-exponential bounds in terms of the graph parameter.

3 Hardness results

In this section we present two reductions from CNF-SAT, and to transfer the non-existence of polynomial kernels (under reasonable complexity assumptions), we use the notion of *polynomial parameter transformation*, introduced by Bodlaender, Thomassé, and Yeo [9]. A polynomial parameter transformation from a parameterized problem P to a parameterized problem Q is an algorithm that, given an instance (x, k) of P, computes in polynomial time an equivalent instance (x', k') of Q such that k' is bounded by a polynomial depending only on k. It follows easily from the definition that if P does not admit a polynomial generalized[1] kernel under some complexity assumption, then the same holds for Q. The complexity hypothesis in the following proposition builds on the results by Fortnow and Santhanam [26].

▶ **Proposition 3.1** (Dell and van Melkebeek [19]). CNF-SAT *does not admit a polynomial generalized kernel parameterized by the number of variables of the input formula, unless* $\mathsf{NP} \subseteq \mathsf{coNP}/\mathsf{poly}$.

We are ready to present our main hardness result, which is inspired by other reductions for related problems [10, 16, 18, 25, 32]. The crucial issue of this reduction, and its main conceptual novelty, is the following fact: when H is not a clique, the intersection of an occurrence of H with (a subgraph of) $G - X$ may be *disconnected*. We exploit this fact by

[1] A *generalized kernel* for a parameterized problem P, also called sometimes *compression* in the literature [15], is a polynomial-time algorithm reducing any instance (x, k) of P to an equivalent instance (x', k') with size bounded by a function $f(k)$ depending only on k of a fixed but potentially *different* parameterized problem Q. A generalized kernel is *polynomial* if $f(k)$ is a polynomial function.

creating clause gadgets with large minimal blocking sets whose elements are disconnected (in Figure 3, each pair of consecutive non-adjacent vertices u, v is an element of a blocking set), and this results in clause gadgets behaving as "chains" where the propagation of information (that is, the vertices picked by the solution) is done without needing edges connecting the elements of the chain (thus, in a "wireless" fashion), easily implying that the corresponding parameter in $G - X$ is bounded by a constant.

▶ **Theorem 3.2.** *Let H be a biconnected graph that is not a clique. The H-*Subgraph Hitting *(resp. H-*Induced Subgraph Hitting*) problem does not admit a polynomial kernel parameterized by the size of a given vertex set X of the input graph G such that $\mathsf{ved}^+_{\mathcal{F}_{\bar{H}}}(G - X) \leq 1$ (resp. $\mathsf{ved}^+_{\mathcal{F}^{\mathsf{ind}}_{\bar{H}}}(G - X) \leq 1$), unless $\mathsf{NP} \subseteq \mathsf{coNP/poly}$.*

Proof. We present a polynomial parameter transformation from the CNF-SAT problem parameterized by the number of variables, which does not admit a polynomial generalized kernel by Proposition 3.1, unless $\mathsf{NP} \subseteq \mathsf{coNP/poly}$. We present our reduction for the H-Subgraph Hitting problem, and at the end of the proof we observe that the same reduction applies to H-Induced Subgraph Hitting as well.

Given a CNF-SAT formula ϕ with n variables $x_1, \ldots, x_n$ and m clauses $C_1, \ldots, C_m$, we proceed to construct in polynomial time an instance G of H-Subgraph Hitting, together with a set $X \subseteq V(G)$ with $\mathsf{ved}^+_H(G - X) \leq 1$ and $|X| = |V(H)| \cdot n$, such that ϕ is satisfiable if and only if G contains a solution of H-Subgraph Hitting of size at most $n - m + \sum_{j=1}^{m} c_j$, where c_j denotes the number of literals in clause C_j. Since $|V(H)|$ is a constant, this would indeed define a polynomial parameter transformation from CNF-SAT parameterized by the number of variables to H-Subgraph Hitting parameterized by the size of a given vertex set X of the input graph G such that $\mathsf{ved}^+_H(G - X) \leq 1$.

For each variable x_i, we add a disjoint copy of H to G. We call such a copy of H the *i-variable-copy* of H. For each clause C_j, we add $c_j - 1$ disjoint copies of H to G, and we order them arbitrarily from 1 to $c_j - 1$. Moreover, we add two new vertices u_j^0 and $v_j^{c_j}$ to G. We call each of these $c_j - 1$ copies of H a *j-clause-copy* of H. Note that, so far, we have introduced $n - m + \sum_{j=1}^{m} c_j$ disjoint copies of H in G.

We now proceed to interconnect these copies of H according to ϕ. Since H is a biconnected graph that is not a clique (hence, it is 2-connected), it follows that $|V(H)| \geq 4$. Thus, in particular there exist two *non-adjacent* vertices $u, v \in V(H)$ and another vertex $w \in V(H)$ distinct from u and v. Let $H' = H - \{u, v, w\}$. (Even if it is not critical for the proof, note that $|V(H')| \geq 1$.) Let also z^+ and z^- be two distinct vertices of H (not necessarily different from u, v, w). We will use the copies of these vertices in the variable-copies and clause-copies of H to interconnect them in G. To this end, for three distinct vertices $a, b, c \in V(G)$ and a subgraph F of G isomorphic to H' not containing any of a, b, c, by *adding an $(a, b, c, V(F))$-copy of H to G* we mean the operation of, starting from $G[\{a, b, c\} \cup V(F)]$, adding the missing edges to complete a copy of H, where vertex a (resp. b, c) of G plays the role of vertex w (resp. u, v) of H, and F plays the role of H', with a fixed isomorphism that we suppose to have at hand.

For each clause C_j of ϕ, consider an arbitrary ordering of its literals as $\ell_1, \ldots, \ell_{c_j}$, and recall that G contains $c_j - 1$ ordered disjoint j-clause-copies of H together with two extra vertices u_j^0 and $v_j^{c_j}$. For $i \in [c_j]$, we add a new copy of H' to G, which we denote by F_j^i. For $i \in [c_j - 1]$, let u_j^i and v_j^i be the copies of vertices u and v of H, respectively, in the i-th j-clause-copy of H. For $i \in [c_j]$, if literal ℓ_i of clause C_j corresponds to a positive (resp. negative) occurrence of a variable x_p, let z be the copy of vertex $z^+ \in V(H)$ (resp. $z^- \in V(H)$) in the p-variable-copy of H. Then we add a $(z, u_j^{i-1}, v_j^i, V(F_j^i))$-copy of H to G,

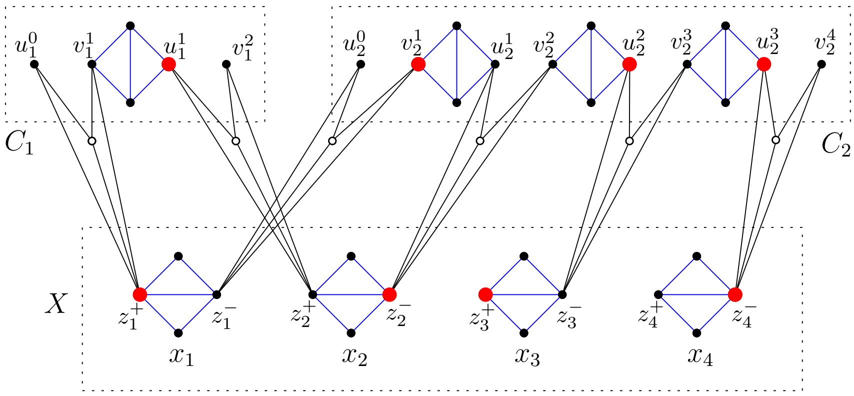

Figure 3 Example of the construction of graph G in the proof of Theorem 3.2 for H-Subgraph Hitting. In this example, H is the diamond (that is, K_4 minus one edge), u and v are the only pair of non-adjacent vertices in H, and w is any other vertex. The construction corresponds to a CNF-SAT formula ϕ consisting of two clauses $C_1 = (x_1 \vee x_2)$ and $C_2 = (\bar{x}_1 \vee \bar{x}_2 \vee \bar{x}_3 \vee \bar{x}_4)$, and the satisfying assignment $\alpha(x_1) = 1$, $\alpha(x_2) = 0$, $\alpha(x_3) = 1$, and $\alpha(x_4) = 0$. The variable-copies and clause-copies of H are depicted in blue, the vertices in the copies of $H' = H - \{u, v, w\}$ (which is a single vertex) are the white ones, and the vertices in the solution S are the large red ones. Note that clause C_2 is satisfied by both $\bar{x}_2$ and $\bar{x}_4$; in the example we have taken $\bar{x}_2$ as the satisfying literal.

and we call such a copy of H a *transversal-copy* of H, denoted by H_j^i. We define $X \subseteq V(G)$ to be the union of the vertex sets of all the variable-copies of H, and note that $|X| = |V(H)| \cdot n$. This completes the construction of G and X, which is illustrated in Figure 3.

In the next claim we prove one of the properties claimed in the statement of the theorem.

▷ Claim 3.3. $\mathsf{ved}^+_{\mathcal{F}_{\bar{H}}}(G - X) = 1$ and $\mathsf{ved}^+_{\mathcal{F}^{\mathsf{ind}}_{\bar{H}}}(G - X) = 1$.

Proof. Note that each connected component of $G - X$ corresponds to the clause-copies of H associated with a clause C_j and isolated vertices, together with the copies of H' between those j-clause-copies of H and isolated vertices. By construction of G, each such a copy of H', say F_j^i, has at most two neighbors in $G - X$, namely vertices u_j^{i-1} and v_j^i. If an occurrence of H as a subgraph in $G - X$, say F, contained a vertex of F_j^i, since $|V(H')| = |V(H)| - 3$, necessarily F contains at least one of u_j^{i-1} and v_j^i, and at least one more vertex in the $(i-1)$-th or i-th j-clause-copies of H. Thus, u_j^{i-1} or v_j^i is a separator of size one of F, contradicting the hypothesis that H is biconnected.

That is, we have proved that no vertex in a copy F_j^i of H' in $G - X$ is contained in an occurrence of H as a subgraph, hence neither as an induced subgraph. Therefore, those vertices can be removed while preserving the value of $\mathsf{ved}^+_{\mathcal{F}_{\bar{H}}}(G - X)$. Formally,

$$\mathsf{ved}^+_{\mathcal{F}_{\bar{H}}}(G - X) = \mathsf{ved}^+_{\mathcal{F}_{\bar{H}}}(G - X - \bigcup_{j=1}^{m} \bigcup_{i=1}^{c_j} V(F_j^i)),$$

and the same holds for $\mathsf{ved}^+_{\mathcal{F}^{\mathsf{ind}}_{\bar{H}}}(G - X)$. To conclude the proof of the claim, it suffices to note that $G - X - \bigcup_{j=1}^{m} \bigcup_{i=1}^{c_j} V(F_j^i)$ consists of a disjoint union of clause-copies of H and isolated vertices, and using the fact that $\mathsf{ved}^+_{\mathcal{F}_{\bar{H}}}$ (resp. $\mathsf{ved}^+_{\mathcal{F}^{\mathsf{ind}}_{\bar{H}}}$) of a disconnected graph is the maximum of $\mathsf{ved}^+_{\mathcal{F}_{\bar{H}}}$ (resp. $\mathsf{ved}^+_{\mathcal{F}^{\mathsf{ind}}_{\bar{H}}}$) over its connected components, by removing one arbitrary vertex from each such a copy of H we get that $\mathsf{ved}^+_{\mathcal{F}_{\bar{H}}}(G - X) = 1$ and $\mathsf{ved}^+_{\mathcal{F}^{\mathsf{ind}}_{\bar{H}}}(G - X) = 1$. ◁

We now claim that ϕ is satisfiable if and only if G contains a solution $S \subseteq V(G)$ of H-Subgraph Hitting of size at most $n - m + \sum_{j=1}^{m} c_j$.

Suppose first that ϕ is satisfiable and let $\alpha\colon \{x_1, \ldots, x_n\} \to \{0,1\}$ be a satisfying assignment of the variables. We define a set $S \subseteq V(G)$ of size $n - m + \sum_{j=1}^{m} c_j$ as follows (cf. the red vertices in Figure 3). For each variable x_i, if $\alpha(x_i) = 1$ (resp. $\alpha(x_i) = 0$), we add to S the copy of z^+ (resp. z^-) in the i-variable-copy of H, which we denote by z_i^+ (resp. z_i^-). For each clause C_j, let ℓ_{s_j} be a literal in C_j that is satisfied by the assignment α. Recall that the $c_j - 1$ j-clause-copies of H are ordered (arbitrarily) from 1 to $c_j - 1$. We add to S the vertex set

$$\{v_j^i \mid 1 \leq i \leq s_j - 1\} \cup \{u_j^i \mid s_j \leq i \leq c_j - 1\}.$$

In words, we add to S the copy of vertex v in all the j-clause-copies of H from 1 to $s_j - 1$, and the copy of vertex u in all the j-clause-copies of H from s_j to $c_j - 1$. Note that $|S| = n - m + \sum_{j=1}^{m} c_j$, and it remains to prove that $G - S$ does not contain H as a subgraph. Note that each variable-copy and clause-copy of H contains exactly two vertices that have neighbors outside of that copy – let us call these vertices *boundary vertices* of that copy – , and that S contains exactly one of these two boundary vertices for each of these copies of H. Hence, since H is biconnected, no occurrence of H in $G - S$ can contain a non-boundary vertex in a variable-copy or clause-copy of H.

Moreover, there do not exist two pairs of integers (i_1, j_1) and (i_2, j_2), with $i_1 \neq i_2$ or $j_1 \neq j_2$, such that there exists an occurrence F of H in $G - S$ with $F \cap (V(H_{j_1}^{i_1}) \setminus X) \neq \emptyset$ and $F \cap (V(H_{j_2}^{i_2} \setminus X) \neq \emptyset$. Indeed, if such (i_1, j_1) and (i_2, j_2) existed, then, as $|N(V(H_j^i) \setminus X) \cap X| = 1$ for any two indices i, j, and as F cannot contain a non-boundary vertex of a variable-copy of H, there would exist $z \in X$ such that $F \cap X = \{z\}$, implying that z is a separator of F, contradicting the 2-connectivity of H.

Thus, if an occurrence of H in $G - S$ existed, say F, then the above discussion and the construction of G imply that F should be one of the transversal-copies of H. But such an F cannot exist in $G - S$ by the choice of S: either S contains one of the boundary vertices in the two j-clause-copies intersected by F for some $j \in [m]$ or, if it is not the case, then S contains the vertex in a variable-copy of H corresponding to the literal that satisfies clause C_j.

Conversely, let $S \subseteq V(G)$ be a solution of H-Subgraph Hitting of size $n - m + \sum_{j=1}^{m} c_j$. Since G contains $|S|$ disjoint variable-copies and clause-copies of H, necessarily S consists of exactly one vertex in each of these copies. Since the boundary vertices in each of the variable-copies and clause-copies of H are the only vertices with neighbors outside of the corresponding copy, we may assume that all the vertices in S are boundary vertices. We define from S a satisfying assignment α of ϕ as follows. If S contains z_i^+ (resp. z_i^-) we set $\alpha(x_i) = 1$ (resp. $\alpha(x_i) = 0$). Let us verify that α indeed satisfies all the clauses of ϕ. Consider an arbitrary clause C_j with c_j literals, and note that S contains $c_j - 1$ vertices in the j-clause-copies of H. Therefore, since by construction no two transversal-copies intersect a clause-copy in a common vertex, there exists $s_j \in [c_j]$ such that the s_j-th transversal-copy of H associated with C_j, say F, is *not* hit by a vertex in a clause-copy of H. Thus, since $S \cap V(F) \neq \emptyset$, necessarily there exists an index $i \in [n]$ such that $S \cap V(F)$ is equal to either z_i^+ or z_i^-, and thus the defined assignment of variable x_i satisfies clause C_j.

To conclude the proof, we claim that the same reduction presented above proves the hardness result for the H-Induced Subgraph Hitting problem. Indeed, in the proof of the equivalence between the satisfiability of ϕ and the existence of a solution S of H-Subgraph Hitting with the appropriate size, all that is relevant to the proof are the variable-copies, clause-copies, and transversal-copies of H. As all these occurrences of H in G occur as induced subgraphs, the same reduction implies the non-existence of polynomial kernels for H-Induced Subgraph Hitting. ◀

In Theorem 1.3 we replace the condition "$\mathsf{ved}^+_H(G-X) \leq 1$" of Theorem 3.2 with the condition that $\mathsf{td}(G-X)$ is bounded by a constant. However, in the proof of Theorem 1.3 we need an extra condition on H stronger than biconnectivity, namely the non-existence of a stable cutset. The reduction in the proof of Theorem 1.3 follows essentially the same lines as the one described in Theorem 3.2, but in order to guarantee that $\mathsf{td}(G-X)$ is bounded by a constant (depending on H), we need to be more careful. Namely, in the interconnection among the variable and clause gadgets, now we cannot afford to add a distinct gadget for each literal in a clause, as it was the case for the copies of H' in the proof of Theorem 3.2 (cf. the white vertices in Figure 3). Indeed, these copies of H' can be removed "for free" when dealing with ved^+_H, but it is not the case anymore when dealing with treedepth, as they may blow up the value of $\mathsf{td}(G-X)$. In a nutshell, we overcome this issue by "reusing" these copies of a (now, carefully chosen) subgraph $H' \subseteq H$ for all the literals of the same clause. However, having a single common H' for each clause may create undesired occurrences of H (other than the variable-copies, clause-copies, and transversal-copies, as we wish), and preventing the existence of these undesired copies is the reason why we need an assumption on H stronger than biconnectivity. The proof of Theorem 1.3 can be found in the full version [11].

4 Further research

In this paper we studied the existence of polynomial kernels for the H-Subgraph Hitting and H-Induced Subgraph Hitting problems under structural parameterizations, namely parameterized by the size of a modulator to a graph class $\mathcal{C}$ that has a "simple structure". Our main achievement is the identification of two arguably natural graph parameters $\mathsf{ved}^+_{\mathcal{F}_{\bar{H}}}$ and $\mathsf{bed}^+_{\mathcal{F}_{\bar{H}}}$ (or $\mathsf{ved}^+_{\mathcal{F}^{\mathsf{ind}}_{\bar{H}}}$ and $\mathsf{bed}^+_{\mathcal{F}^{\mathsf{ind}}_{\bar{H}}}$ for the induced version) that allowed us to prove complexity dichotomies in terms of the forbidden graph H. Our results pave the way to a systematic investigation of this topic, where we identify the following avenues for further research.

Getting rid of the hypothesis on H. In our hardness results we need additional assumptions on H, mainly that H is biconnected in Theorem 3.2. Observe that the requirement that H is connected is necessary to obtain polynomial kernels. Indeed, when H is the union of a K_5 and a $K_{1,3}$, it is known [31] that H-Subgraph Hitting is para-NP-hard, even for $\mathsf{ed}_{\mathcal{H}} = 0$. Moreover, when H is a non-edge, H-Induced Subgraph Hitting parameterized by vertex cover (which is a larger parameter than $\mathsf{ed}_{\mathcal{H}}$) is equivalent to maximum clique parameterized by vertex cover, which does not admit a polynomial kernel under standard complexity assumptions [7]. Thus, it is natural to wonder whether the biconnectivity hypothesis could be replaced by just connectivity.

Improving the degree of the kernel. The degree of our polynomial kernel depends on the size t of the excluded clique and on the value λ of the promised upper bound $\mathsf{bed}^+_t(G-X) \leq \lambda$. Namely, as stated in Theorem 2.5, the kernel has size $\mathcal{O}_{\lambda,t}(|X|^{\delta(\lambda,t)})$, where function δ mainly depends on the upper bound on mmbs_t given in Theorem 2.6. This function behaves as a tower of exponents in t of height λ. Hence, improving the bound on mmbs_t directly translates to an improvement of the kernel size. We did obtain such an improvement if instead of assuming that $\mathsf{bed}^+_t(G-X) \leq \lambda$, one assumes that $\mathsf{td}(G-X) \leq \lambda$, namely with a function $\lambda^\lambda \cdot 2^{\lambda^2}$; see [11, Lemma 4.12]. We leave as an open problem to obtain improved upper bounds for mmbs_t in terms of ved^+_t and bed^+_t.

Computing the modulator. In our kernelization algorithm we assume that we are *given* a modulator, namely a set $X \subseteq V(G)$ such that $\mathsf{bed}_t^+(G-X) \le \lambda$. Note that this hypothesis appears also in the related work dealing with Feedback Vertex Set [18]. Obtaining a constant factor or even $\mathsf{poly}(\mathsf{opt})$-approximation of the modulator in polynomial time for fixed t and λ, which will be enough for our kernelization algorithm (note that minimizing its size is NP-hard [38]), remains an interesting direction. One may start with the probably simpler cases of a modulator to bounded $\mathcal{F}_{\bar{H}}$-elimination distance or bounded ved_t^+.

Finding the right measure. The focus of this article is on obtaining kernelization dichotomies as a function of the forbidden (induced) subgraph H. Of course, it is also relevant to characterize, for a fixed graph H, which is the most general (monotone or hereditary) target family $\mathcal{C}_H$ such that H-(Induced) Subgraph Hitting admits a polynomial kernel parameterized by the size of a modulator to a graph in $\mathcal{C}_H$. Needless to say, solving this general problem seems quite challenging. Indeed, even the case of Vertex Cover, that is, $H = K_2$, is far from being well understood for monotone or hereditary target graph classes, as for instance the only known polynomial kernel for Vertex Cover parameterized by a modulator to a bipartite graph (i.e., an odd cycle transversal) is randomized and relies on quite powerful tools [36, 37]. One may hope that larger cliques allow for simpler characterizations, the natural first candidate being the case where H is a triangle. Let $\mathcal{C}_\Delta$ be the, say, hereditary target graph class that we want to characterize. Following the approach of [10] that characterized the target *minor-closed* graph classes for Vertex Cover, one may hope that K_3-Subgraph Hitting admits a polynomial kernel parameterized by a modulator to $\mathcal{C}_\Delta$ if and only if the graphs in $\mathcal{C}_\Delta$ have bounded mmbs_3. With no extra assumption on $\mathcal{C}_\Delta$, this property is probably false due to the results of Hols, Kratsch, and Pieterse [29], but we conjecture that it is true if we ask $\mathcal{C}_\Delta$ to be hereditary and closed under disjoint union, even for hitting K_t for every $t \ge 3$, replacing mmbs_3 by mmbs_t. Toward an eventual proof of this conjecture, having unbounded minimal blocking sets seems to permit a generic reduction to obtain the lower bound, in the spirit of the one of Theorem 3.2 or any similar one in previous work [10, 16, 18, 25, 32]. Indeed, Hols, Kratsch, and Pieterse [29, Thm 1.1] show that for Vertex Cover, lower bounds on kernel sizes directly follow from lower bounds on mmbs_2. However, the opposite direction seems way more challenging. In [10], this fact was established for Vertex Cover and minor-closed target classes via the notion of bridge-depth by proving, in particular, that there is a *single* minor-obstruction for having large maximum minimal blocking sets, namely the chains of triangles (cf. left part of Figure 4). Unfortunately, we cannot hope the same nice behavior for K_3-Subgraph Hitting and monotone or hereditary graph classes, as chains of triangles are still an obstruction in this setting, but there exist other incomparable ones, as depicted in Figure 4.

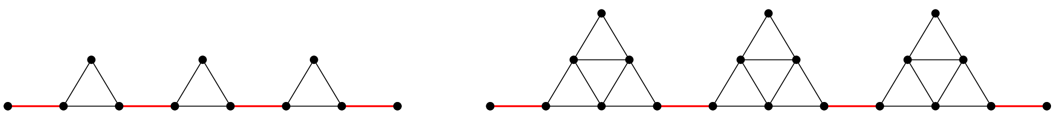

Figure 4 Two chains of length three incomparable with respect to the (induced) subgraph relation. In both graphs, it can be verified that the set of four red thicker edges is a minimal blocking set.

Finally, in the ambitious quest for finding the appropriate measures that characterize the hereditary or monotone classes $\mathcal{C}_t$ for which K_t-Subgraph Hitting admits a polynomial kernel parameterized by the size of a modulator X to $\mathcal{C}_t$, we hope that the techniques we developed to provide a polynomial kernel for the case $\mathsf{bed}_t^+(G-X) \le \lambda$ will play an important role. A natural attempt to generalize bed_t^+ to a more powerful measure is to relax, or even drop, the "weak attachment" condition on the sets to be removed in every round of the elimination process. This raises new challenges for obtaining a polynomial kernel that do not seem easy to overcome, at least with the existing tools in this area.

References

1. Faisal N. Abu-Khzam. A kernelization algorithm for d-Hitting Set. *Journal of Computer and System Sciences*, 76(7):524–531, 2010. doi:10.1016/j.jcss.2009.09.002.
2. Akanksha Agrawal, Lawqueen Kanesh, Daniel Lokshtanov, Fahad Panolan, M. S. Ramanujan, Saket Saurabh, and Meirav Zehavi. Deleting, Eliminating and Decomposing to Hereditary Classes Are All FPT-Equivalent. In *Proc. of the 32st Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1976–2004, 2022. doi:10.1137/1.9781611977073.79.
3. Júlio Araújo, Marin Bougeret, Victor A. Campos, and Ignasi Sau. Parameterized complexity of computing maximum minimal blocking and hitting sets. *Algorithmica*, 85(2):444–491, 2023. doi:10.1007/s00453-022-01036-5.
4. Julien Baste, Ignasi Sau, and Dimitrios M. Thilikos. Hitting minors on bounded treewidth graphs. IV. an optimal algorithm. *SIAM Journal on Computing*, 52(4):865–912, 2023. doi:10.1137/21m140482x.
5. Hans L. Bodlaender. Kernelization, exponential lower bounds. In *Encyclopedia of Algorithms*, pages 1013–1017. Springer, 2016. doi:10.1007/978-1-4939-2864-4_521.
6. Hans L. Bodlaender, Rodney G. Downey, Michael R. Fellows, and Danny Hermelin. On problems without polynomial kernels. *Journal of Computer and System Sciences*, 75(8):423–434, 2009. doi:10.1016/j.jcss.2009.04.001.
7. Hans L. Bodlaender, Bart M. P. Jansen, and Stefan Kratsch. Cross-composition: A new technique for kernelization lower bounds. In *Proc. of the 28th 28th International Symposium on Theoretical Aspects of Computer Science (STACS)*, volume 9 of *LIPIcs*, pages 165–176, 2011. doi:10.4230/LIPIcs.STACS.2011.165.
8. Hans L. Bodlaender, Bart M. P. Jansen, and Stefan Kratsch. Kernelization lower bounds by cross-composition. *SIAM Journal on Discrete Mathematics*, 28(1):277–305, 2014. doi:10.1137/120880240.
9. Hans L. Bodlaender, Stéphan Thomassé, and Anders Yeo. Kernel bounds for disjoint cycles and disjoint paths. *Theoretical Computer Science*, 412(35):4570–4578, 2011. doi:10.1016/j.tcs.2011.04.039.
10. Marin Bougeret, Bart M. P. Jansen, and Ignasi Sau. Bridge-depth characterizes which minor-closed structural parameterizations of vertex cover admit a polynomial kernel. *SIAM Journal on Discrete Mathematics*, 36(4):2737–2773, 2022. doi:10.1137/21m1400766.
11. Marin Bougeret, Bart M. P. Jansen, and Ignasi Sau. Kernelization dichotomies for hitting subgraphs under structural parameterizations. *CoRR*, abs/2404.16695, 2024. arXiv:2404.16695.
12. Marin Bougeret and Ignasi Sau. How much does a treedepth modulator help to obtain polynomial kernels beyond sparse graphs? *Algorithmica*, 81(10):4043–4068, 2019. doi:10.1007/s00453-018-0468-8.
13. Jannis Bulian and Anuj Dawar. Graph isomorphism parameterized by elimination distance to bounded degree. *Algorithmica*, 75(2):363–382, 2016. doi:10.1007/s00453-015-0045-3.
14. Jannis Bulian and Anuj Dawar. Fixed-parameter tractable distances to sparse graph classes. *Algorithmica*, 79(1):139–158, 2017. doi:10.1007/s00453-016-0235-7.
15. Marek Cygan, Fedor V. Fomin, Lukasz Kowalik, Daniel Lokshtanov, Dániel Marx, Marcin Pilipczuk, Michal Pilipczuk, and Saket Saurabh. *Parameterized Algorithms*. Springer, 2015. doi:10.1007/978-3-319-21275-3.

16 Marek Cygan, Daniel Lokshtanov, Marcin Pilipczuk, Michal Pilipczuk, and Saket Saurabh. On the hardness of losing width. *Theory of Computing Systems*, 54(1):73–82, 2014. doi:10.1007/s00224-013-9480-1.

17 Marek Cygan, Dániel Marx, Marcin Pilipczuk, and Michal Pilipczuk. Hitting forbidden subgraphs in graphs of bounded treewidth. *Information and Computation*, 256:62–82, 2017. doi:10.1016/j.ic.2017.04.009.

18 David Dekker and Bart M. P. Jansen. Kernelization for Feedback Vertex Set via Elimination Distance to a Forest. In *Proc. of the 48th International Workshop Graph-Theoretic Concepts in Computer Science (WG)*, volume 13453 of *LNCS*, pages 158–172, 2022. doi:10.1007/978-3-031-15914-5_12.

19 Holger Dell and Dieter van Melkebeek. Satisfiability allows no nontrivial sparsification unless the polynomial-time hierarchy collapses. *Journal of the ACM*, 61(4):23:1–23:27, 2014. doi:10.1145/2629620.

20 Michael R. Fellows, Lars Jaffke, Aliz Izabella Király, Frances A. Rosamond, and Mathias Weller. What Is Known About Vertex Cover Kernelization? In *Adventures Between Lower Bounds and Higher Altitudes - Essays Dedicated to Juraj Hromkovič on the Occasion of His 60th Birthday*, volume 11011 of *LNCS*, pages 330–356, 2018. doi:10.1007/978-3-319-98355-4_19.

21 Jörg Flum and Martin Grohe. *Parameterized Complexity Theory*. Springer-Verlag, 2006. doi:10.1007/3-540-29953-X.

22 Fedor V. Fomin, Petr A. Golovach, and Dimitrios M. Thilikos. Parameterized complexity of elimination distance to first-order logic properties. *ACM Transactions on Computational Logic*, 23(3):17:1–17:35, 2022. doi:10.1145/3517129.

23 Fedor V. Fomin, Daniel Lokshtanov, Neeldhara Misra, and Saket Saurabh. Planar $\mathcal{F}$-Deletion: Approximation, Kernelization and Optimal FPT Algorithms. In *Proc. of the 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2012. doi:10.1109/FOCS.2012.62.

24 Fedor V. Fomin, Daniel Lokshtanov, Saket Saurabh, and Meirav Zehavi. *Kernelization: Theory of Parameterized Preprocessing*. Cambridge University Press, 2019. doi:10.1017/9781107415157.

25 Fedor V. Fomin and Torstein J. F. Strømme. Vertex cover structural parameterization revisited. In *Proc. of the 42nd International Workshop on Graph-Theoretic Concepts in Computer Science (WG)*, volume 9941 of *LNCS*, pages 171–182, 2016. doi:10.1007/978-3-662-53536-3_15.

26 Lance Fortnow and Rahul Santhanam. Infeasibility of instance compression and succinct PCPs for NP. *Journal of Computer and System Sciences*, 77(1):91–106, 2011. doi:10.1016/j.jcss.2010.06.007.

27 Jakub Gajarský, Petr Hlinený, Jan Obdrzálek, Sebastian Ordyniak, Felix Reidl, Peter Rossmanith, Fernando Sánchez Villaamil, and Somnath Sikdar. Kernelization using structural parameters on sparse graph classes. In *Proc. of the 21st Annual European Symposium on Algorithms (ESA)*, volume 8125 of *LNCS*, pages 529–540, 2013. doi:10.1007/978-3-642-40450-4_45.

28 Jiong Guo, Falk Hüffner, and Rolf Niedermeier. A structural view on parameterizing problems: Distance from triviality. In *Proc. of the 1st International Workshop on Parameterized and Exact Computation (IWPEC)*, volume 3162 of *LNCS*, pages 162–173, 2004. doi:10.1007/978-3-540-28639-4_15.

29 Eva-Maria C. Hols, Stefan Disc.tsch, and Astrid Pieterse. Elimination distances, blocking sets, and kernels for vertex cover. *SIAM Journal on Discrete Mathematics*, 36(3):1955–1990, 2022. doi:10.1137/20m1335285.

30 Bart M. P. Jansen and Hans L. Bodlaender. Vertex cover kernelization revisited - upper and lower bounds for a refined parameter. *Theory of Computing Systems*, 53(2):263–299, 2013. doi:10.1007/s00224-012-9393-4.

31 Bart M. P. Jansen, Jari J. H. de Kroon, and Michal Włodarczyk. Vertex deletion parameterized by elimination distance and even less. In *Proc. of the 53rd Annual ACM Symposium on Theory of Computing (STOC)*, pages 1757–1769, 2021. doi:10.1145/3406325.3451068.

32 Bart M. P. Jansen and Stefan Kratsch. On Polynomial Kernels for Structural Parameterizations of Odd Cycle Transversal. In *Proc. of the 6th International Symposium on Parameterized and Exact Computation (IPEC)*, volume 7112 of *LNCS*, pages 132–144, 2011. doi:10.1007/978-3-642-28050-4_11.

33 Bart M. P. Jansen and Astrid Pieterse. Polynomial kernels for hitting forbidden minors under structural parameterizations. *Theoretical Computer Science*, 841:124–166, 2020. doi:10.1016/j.tcs.2020.07.009.

34 Bart M. P. Jansen and Michal Wlodarczyk. Lossy planarization: a constant-factor approximate kernelization for planar vertex deletion. In *Proc of the 54th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 900–913, 2022. doi:10.1145/3519935.3520021.

35 Ton Kloks. *Treewidth, Computations and Approximations*, volume 842 of *LNCS*. Springer, 1994. doi:10.1007/BFb0045375.

36 Stefan Kratsch. A randomized polynomial kernelization for vertex cover with a smaller parameter. *SIAM Journal on Discrete Mathematics*, 32(3):1806–1839, 2018. doi:10.1137/16M1104585.

37 Stefan Kratsch and Magnus Wahlström. Representative sets and irrelevant vertices: New tools for kernelization. *Journal of the ACM*, 67(3):16:1–16:50, 2020. doi:10.1145/3390887.

38 John M. Lewis and Mihalis Yannakakis. The node-deletion problem for hereditary properties is NP-complete. *Journal of Computer and System Sciences*, 20(2):219–230, 1980.

39 Daniel Lokshtanov, Neeldhara Misra, and Saket Saurabh. Kernelization - Preprocessing with a guarantee. In *The Multivariate Algorithmic Revolution and Beyond*, pages 129–161, 2012. doi:10.1007/978-3-642-30891-8_10.

40 Diptapriyo Majumdar, Venkatesh Raman, and Saket Saurabh. Polynomial kernels for vertex cover parameterized by small degree modulators. *Theory of Computing Systems*, 62(8):1910–1951, 2018. doi:10.1007/s00224-018-9858-1.

41 Laure Morelle, Ignasi Sau, Giannos Stamoulis, and Dimitrios M. Thilikos. Faster parameterized algorithms for modification problems to minor-closed classes. In *Proc. of the 50th International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 261 of *LIPIcs*, pages 93:1–93:19, 2023. doi:10.4230/LIPIcs.ICALP.2023.93.

42 Michal Pilipczuk, Nicole Schirrmacher, Sebastian Siebertz, Szymon Torunczyk, and Alexandre Vigny. Algorithms and data structures for first-order logic with connectivity under vertex failures. In *Proc. of the 49th International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 229 of *LIPIcs*, pages 102:1–102:18, 2022. doi:10.4230/LIPICS.ICALP.2022.102.

43 Ignasi Sau and Uéverton dos Santos Souza. Hitting forbidden induced subgraphs on bounded treewidth graphs. *Information and Computation*, 281:104812, 2021. doi:10.1016/j.ic.2021.104812.