Smaller Parameters for Vertex Cover Kernelization*

Eva-Maria C. Hols[1] and Stefan Kratsch[2]

1 **Department of Computer Science, University of Bonn, Germany**
`hols@cs.uni-bonn.de`
2 **Department of Computer Science, University of Bonn, Germany**
`kratsch@cs.uni-bonn.de`

Abstract

We revisit the topic of polynomial kernels for VERTEX COVER relative to structural parameters. Our starting point is a recent paper due to Fomin and Strømme [WG 2016] who gave a kernel with $\mathcal{O}(|X|^{12})$ vertices when X is a vertex set such that each connected component of $G-X$ contains at most one cycle, i.e., X is a modulator to a pseudoforest. We strongly generalize this result by using modulators to d-quasi-forests, i.e., graphs where each connected component has a feedback vertex set of size at most d, and obtain kernels with $\mathcal{O}(|X|^{3d+9})$ vertices. Our result relies on proving that minimal blocking sets in a d-quasi-forest have size at most $d+2$. This bound is tight and there is a related lower bound of $\mathcal{O}(|X|^{d+2-\epsilon})$ on the bit size of kernels.

In fact, we also get bounds for minimal blocking sets of more general graph classes: For d-quasi-bipartite graphs, where each connected component can be made bipartite by deleting at most d vertices, we get the same tight bound of $d+2$ vertices. For graphs whose connected components each have a vertex cover of cost at most d more than the best fractional vertex cover, which we call d-quasi-integral, we show that minimal blocking sets have size at most $2d+2$, which is also tight. Combined with existing randomized polynomial kernelizations this leads to randomized polynomial kernelizations for modulators to d-quasi-bipartite and d-quasi-integral graphs. There are lower bounds of $\mathcal{O}(|X|^{d+2-\epsilon})$ and $\mathcal{O}(|X|^{2d+2-\epsilon})$ for the bit size of such kernels.

1998 ACM Subject Classification G.2.2 Graph Algorithms

Keywords and phrases Vertex Cover, Kernelization, Structural Parameterization

Digital Object Identifier 10.4230/LIPIcs.IPEC.2017.20

1 Introduction

The VERTEX COVER problem plays a central role in parameterized complexity. In particular, it has been very important for the development of new kernelization techniques and the study of structural parameters. As a result of this work, there is now a solid understanding of which parameterizations of VERTEX COVER lead to fixed-parameter tractability or existence of a polynomial kernelization. This is motivated by the fact that parameterization by solution size leads to large parameter values on many types of easy instances. Thus, while there is a well-known kernelization for instances of VERTEX COVER(k) to at most $2k$ vertices, it may be more suitable to apply a kernelization with a size guarantee that is a larger function but depends on a smaller parameter.

Jansen and Bodlaender [13] were the first to study kernelization for VERTEX COVER under different, smaller parameters. Their main result is a polynomial kernelization to instances

* A full version of the paper is available at `http://arxiv.org/abs/1711.04604`.

12th International Symposium on Parameterized and Exact Computation (IPEC 2017).
Editors: Daniel Lokshtanov and Naomi Nishimura; Article No. 20; pp. 20:1–20:12
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

with $\mathcal{O}(|X|^3)$ vertices when X is a feedback vertex set of the input graph, also called a *modulator* to the class of forests. Clearly, the size of X is a lower bound on the vertex cover size (as any vertex cover is a modulator to an independent set). Since then, their result has been generalized and complemented in several ways. The two main directions of follow-up work are to use modulators to other tractable cases instead of forests (see below) and parameterization above lower bounds (see related work).

For any graph class $\mathcal{C}$, we can define a parameterization of VERTEX COVER by distance to $\mathcal{C}$, i.e., by the minimum size of a modulator X such that $G - X$ belongs to $\mathcal{C}$. For fixed-parameter tractability and kernelization of the arising parameterized problem it is necessary that VERTEX COVER is tractable on inputs from $\mathcal{C}$. For hereditary classes $\mathcal{C}$, this condition is also sufficient for fixed-parameter tractability but not necessarily for the existence of a polynomial kernelization. Interesting choices for $\mathcal{C}$ are various well-studied hereditary graph classes, like forests, bipartite, or chordal graphs, and graphs of bounded treewidth, bounded treedepth, or bounded degree.

Majumdar et al. [16] studied VERTEX COVER parameterized by (the size of) a modulator X to a graph of maximum degree at most d. For $d \geq 3$ this problem is NP-hard but for $d = 2$ and $d = 1$ they obtained kernels with $\mathcal{O}(|X|^5)$ and $\mathcal{O}(|X|^2)$ vertices, respectively. Their result motivated Fomin and Strømme [9] to investigate a parameter that is smaller than both a modulator to degree at most two and the size of a feedback vertex set: They consider X being a modulator to a *pseudoforest*, i.e., with each connected component of $G - X$ having at most one cycle. For this they obtain a kernelization to $\mathcal{O}(|X|^{12})$ vertices, generalizing (except for the size) the results of Majumdar et al. [16] and Jansen and Bodlaender [13]. They also prove that the parameterization by a modulator to so-called *mock forests*, where no cycles share a vertex, admits no polynomial kernelization unless $\mathsf{NP} \subseteq \mathsf{coNP/poly}$.

For their kernelization, Fomin and Strømme [9] prove that minimal blocking sets in a pseudoforest have size at most three, which requires a lengthy proof. (A minimal blocking set is a set of vertices whose deletion decreases the independence number by exactly one.)[1] This allows to reduce the number of components of the pseudoforest such that one can extend the modulator X to a sufficiently small feedback vertex set by adding one (cycle) vertex per component to X. At this point, the kernelization of Jansen and Bodlaender [13] can be applied to get the result.

The results of Fomin and Strømme [9] suggest that the border for existence of polynomial kernels for feedback vertex set-like parameters may be much more interesting than expected previously. Arguably, there is still quite some room between allowing a single cycle per component and allowing an arbitrary number of cycles so long as they share no vertices. Do larger numbers of cycles per component still allow a polynomial kernelization? Similarly, cycles in the lower bound proof have odd length and it is known that absence of odd cycles is sufficient, i.e., a kernelization for modulators to bipartite graphs is known. Could this be extended to allowing bipartite graphs with one or more odd cycles per connected component?

Our work. We show that the answers to the above questions are largely positive and provide, essentially, a single elegant proof to cover them. To this end, it is convenient to take the perspective of feedback sets rather than the maximum size of a cycle packing. Say that a *d-quasi-forest* is a graph such that each connected component has a feedback vertex set of size at most d, whereas in a *d-quasi-bipartite graph* each connected component must have an odd cycle transversal (a feedback set for odd cycles) of size at most d.

[1] Like previous work [13, 9] we prefer to work with INDEPENDENT SET rather than VERTEX COVER, but this makes no important difference.

We show that VERTEX COVER admits a kernelization with $\mathcal{O}(|X|^{3d+9})$ vertices when X is a modulator to a d-quasi-forest (Section 3). The case for $d = 1$ strengthens the result of Fomin and Strømme [9] (as one cycle per component is stricter than feedback vertex set size one). For every fixed larger value of d we obtain a polynomial kernelization, though of increasing size. The result is obtained by proving that minimal blocking sets in a d-quasi-forest have size at most $d+2$ (and then applying [13]). Intuitively, having a large minimal blocking set implies getting a fairly small maximum independent set because there are optimal independent sets that avoid all but any chosen vertex of a minimal blocking set. In contrast, a d-quasi-forest always has a large independent set because each connected component is almost a tree.

The value $d + 2$ is tight already for cliques of size $d + 2$, which are permissible connected components in a d-quasi-forest. Such cliques also imply that our parameterization inherits a lower bound of $\mathcal{O}(|X|^{d+2-\epsilon})$ from the lower bound of $\mathcal{O}(|X'|^{r-\epsilon})$ (assuming $\mathsf{NP} \not\subseteq \mathsf{coNP/poly}$) for X' being a modulator to a cluster graph with component size at most r [16].

It turns out that our proof directly extends also to d-quasi-bipartite graphs, proving that their minimal blocking sets similarly have size at most $d + 2$ (Section 4). Thus, when given a modulator X such that $G - X$ is d-quasi-bipartite, we can extend it to an odd cycle transversal X' of size at most $d \cdot |X|^{d+3} + |X|$, which directly yields a randomized polynomial kernel by using a randomized polynomial kernelization for VERTEX COVER parameterized by an odd cycle transversal [15]. Motivated by this, we explore also modulators to graphs in which each connected component has vertex cover size at most d plus the size of a minimum fractional vertex cover, which we call d-quasi-integral (Section 4). This is stronger than the previous parameter because it allows connected components that have an odd cycle transversal of size at most d. We show that minimal blocking sets in any d-quasi-integral graph have size at most $2d + 2$. This bound is tight, as witnessed by the cliques with $2d + 2$ vertices, and the problem inherits a lower bound of $\mathcal{O}(|X|^{2d+2-\epsilon})$ from the lower bound for modulators to cluster graphs with clique size at most $r = 2d + 2$ [16]. Using the upper bound of $2d + 2$ one can remove redundant connected components until the obtained instance has vertex cover size at most $d \cdot |X|^{2d+3} + |X|$ more than the best fractional vertex cover. In other words, one can reduce to an instance of VERTEX COVER parameterized above LP with parameter value $d \cdot |X|^{2d+3} + |X|$ and apply the randomized polynomial kernelization of Kratsch and Wahlström [15] to get a randomized polynomial kernel.

Related work. Recent work of Bougeret and Sau [5] shows that VERTEX COVER admits a kernel of size $\mathcal{O}(|X|^{f(c)})$ when X is a modulator to a graph of treedepth at most c. Their result is incomparable to ours: Already the kernelization by feedback vertex set size [13], which we generalize, allows arbitrarily long paths in $G - X$; such paths are forbidden in a graph of bounded treedepth. Conversely, taking a star with d leaves and appending a 3-cycle at each leaf yields a graph with feedback vertex set and odd cycle transversal size equal to d but constant treedepth; d can be chosen arbitrarily large.

The fact that deciding whether a graph G has a vertex cover of size at most k is trivial when k is lower than the size $MM(G)$ of a largest matching in G has motivated the study of above lower bound parameters like $\ell = k - MM(G)$. The strongest lower bound employed so far is $2LP(G) - MM(G)$, where $LP(G)$ denotes the minimum cost of a fractional vertex cover, and Garg and Philip [10] gave an $\mathcal{O}^*(3^{k-(2LP(G)-MM(G))})$ time algorithm. Randomized polynomial kernels are known for parameters $k - MM(G)$ and $k - LP(G)$ [15] and for parameter $k - (2LP(G) - MM(G))$ [14]. Our present kernelizations are not covered even by the strongest parameter $k - (2LP(G) - MM(G))$ because already d-quasi-forests for any $d \geq 2$ can have a vertex cover size that is arbitrarily larger than $k - (2LP(G) - MM(G))$: Consider, for example, a disjoint union of cliques K_4 with four vertices each, where $2LP(K_4) - MM(K_4) = 2$ but vertex cover size is three per component.

Regarding lower bounds for kernelization (all assuming $\mathsf{NP} \nsubseteq \mathsf{coNP/poly}$), it is of course well known that there are no polynomial kernels for VERTEX COVER when parameterized by width parameters like treewidth, pathwidth, or treedepth (cf. [2]). Lower bounds similar to the one for modulators to mock forests by Fomin and Strømme [9] were already obtained by Cygan et al. [7] (modulators to treewidth at most two) and Jansen [12] (modulators to outerplanar graphs). Bodlaender et al. [3] showed that there is no polynomial kernelization in terms of the vertex deletion distance to a single clique, which is stronger than distance to cluster or perfect graphs for example. Majumdar et al. [16] ruled out kernels of size $\mathcal{O}(|X|^{r-\epsilon})$ when X is a modulator to a cluster graph with cliques of size bounded by r.

Due to space limitations several proofs are deferred to the full version.

2 Preliminaries and notation

Graphs. We use standard notation mostly following Diestel [8]. Let $G = (V, E)$ be a graph. For a set $X \subseteq V$, let $N_G(X)$ denote the neighborhood of X in G, i.e., $N_G(X) = \{v \in V \setminus X \mid \exists u \in X\colon \{u, v\} \in E\}$ and let $N_G[X]$ denote the neighborhood of X in G including X, i.e., $N_G[X] = N_G(X) \cup X$. We omit the subscript whenever the underlying graph is clear from the context. Furthermore, we use $G - X$ as shorthand for $G[V \setminus X]$. For a graph G we denote by $\mathrm{vc}(G)$ the vertex cover number of G and by $\alpha(G)$ the independence number of G. Let $Y \subseteq V$, we call Y a *blocking set* of G, if deleting the vertex set Y from the graph G decreases the size of a maximum independent set, hence if $\alpha(G) > \alpha(G - Y)$. A blocking set Y is *minimal*, if no proper subset $Y' \subsetneq Y$ of Y is a blocking set of G. We denote by K_n the clique of size n.

Linear Programming. We denote the linear program relaxation for VERTEX COVER resp. INDEPENDENT SET for a graph $G = (V, E)$ by $\mathrm{LP_{VC}}(G)$ resp. $\mathrm{LP_{IS}}(G)$. Recall that $\mathrm{LP_{VC}}(G) = \min\{\sum_{v \in V} x_v \mid \forall \{u, v\} \in E\colon x_u + x_v \geq 1 \wedge \forall v \in V\colon 0 \leq x_v \leq 1\}$ and $\mathrm{LP_{IS}}(G) = \max\{\sum_{v \in V} x_v \mid \forall \{u, v\} \in E\colon x_u + x_v \leq 1 \wedge \forall v \in V\colon 0 \leq x_v \leq 1\}$. A *feasible solution* to one of the above linear program relaxations is an assignment to the variables x_v for all vertices $v \in V$ which satisfies the conditions of the linear program. An optimum solution to $\mathrm{LP_{VC}}(G)$ resp. $\mathrm{LP_{IS}}(G)$ is a feasible solution x which minimizes resp. maximizes the objective function value $w(x) := \sum_{v \in V} x_v$. It follows directly from the definition that x is a feasible solution to $\mathrm{LP_{VC}}(G)$ if and only if $x' = 1 - x$ is a feasible solution to $\mathrm{LP_{IS}}(G)$; thus $w(x') = |V| - w(x)$. It is well known that there exists an optimum feasible solution x to $\mathrm{LP_{VC}}(G)$ with $x_v \in \{0, \frac{1}{2}, 1\}$; we call such a solution *half integral*. The same is, of course, true for $\mathrm{LP_{IS}}(G)$. Given a half integral solution x (to $\mathrm{LP_{VC}}(G)$ or $\mathrm{LP_{IS}}(G)$), we define $V_i^x = \{v \in V \mid x_v = i\}$ for each $i \in \{0, \frac{1}{2}, 1\}$. Note that if x is an optimum half integral solution to $\mathrm{LP_{VC}}(G)$, then it holds that $N(V_0^x) = V_1^x$, whereas, it holds that $N(V_1^x) = V_0^x$, when x is an optimum half integral solution to $\mathrm{LP_{IS}}(G)$. We omit the subscript x, when the solution x is clear from the context.

3 Vertex Cover parameterized by a modulator to a d-quasi-forest

In this section we present a polynomial kernel for VERTEX COVER parameterized by a modulator to a d-quasi-forest. More precisely, we develop a polynomial kernel for INDEPENDENT SET parameterized by a modulator to a d-quasi-forest which, by the relation between these two problems, directly yields a polynomial kernel for VERTEX COVER parameterized by a modulator to a d-quasi-forest.

Consider an instance (G, X, k) of the problem, which asks whether graph G, with $G - X$ is a d-quasi-forest, has an independent set of size k. Like Fomin and Strømme [9], we reduce the input instance (G, X, k) until the d-quasi-forest $G - X$ has at most polynomially many connected components in terms of $|X|$; see Rule 1. By adding for each component of the d-quasi-forest a feedback vertex set of size d to the modulator X, we polynomially increase the size of the modulator X. The resulting modulator is a feedback vertex set, hence we can apply the polynomial kernelization for INDEPENDENT SET parameterized by a modulator to a feedback vertex set from Jansen and Bodlaender [13].

Let (G, X, k) be an instance of INDEPENDENT SET parameterized by a modulator to a d-quasi-forest. Since d is a constant we can compute in polynomial time a maximum independent set in $G - X$. Choosing some vertices from the set X to be in an independent set will prevent some vertices in $G - X$ to be part of the same independent set; thus it may be that we can add less than $\alpha(G - X)$ vertices from $G - X$ to an independent set that contains some vertices of X. To measure this difference, we use the term of conflicts introduced by Jansen and Bodlaender [13]. Our definition is more general in order to use it also for modulators to d-quasi-bipartite resp. d-quasi-integral graphs.

▶ **Definition 1** (Conflicts). Let $G = (V, E)$ be a graph and $X \subseteq V$ be a subset of V, such that we can compute a maximum independent set in $G - X$ in polynomial time. Let F be a subgraph of $G - X$ and let $X' \subseteq X$. We define the number of conflicts on F which are induced by X' as $\textsc{Conf}_F(X') := \alpha(F) - \alpha(F - N(X'))$.

Now we can state our reduction rule, which deletes some components of the d-quasi-forest $G-X$. More precisely, we delete components H of which we know that there exists a maximum independent set in G that contains a maximum independent set of the component H.

Rule 1: If there exists a connected component H of $G - X$ such that for all independent sets $X_I \subseteq X$ of size at most $d+2$ with $\textsc{Conf}_H(X_I) > 0$ it holds that $\textsc{Conf}_{G-H-X}(X_I) \geq |X|$, then delete H from G and reduce k by $\alpha(H)$.

The proof of safeness will be given in the sequel. In particular, we delete connected components that have no conflicts. The goal of Rule 1 is to delete connected components of the d-quasi-forest $G - X$ such that we can bound the number of connected components by a polynomial in the size of X. Thus, if we cannot apply this reduction rule any more we should be able to find a good bound for the number of connected components in the d-quasi-forest $G - X$. The following lemma yields such a bound.

▶ **Lemma 2.** *Let (G, X, k) be an instance of* INDEPENDENT SET *parameterized by a modulator to a d-quasi-forest where Rule 1 is not applicable. Then the number of connected components in $G - X$ is at most $|X|^{d+3}$.*

Proof. Let H be a connected component of the d-quasi-forest $G - X$. Since Rule 1 is not applicable, there exists an independent set $X_I \subseteq X$ of size at most $d + 2$ such that $\textsc{Conf}_H(X_I) > 0$ and $\textsc{Conf}_{G-H-X}(X_I) < |X|$; otherwise Rule 1 would delete H (or another connected component with the same properties).

Observe, that there are at most $|X|$ connected components of the d-quasi-forest $G - X$ that have a conflict with an independent set $X_I \subseteq X$, when X_I is the reason that we cannot apply Rule 1 to one of these connected components: Assume for contradiction that there are $p > |X|$ connected components $H_1, H_2, \ldots, H_p$ of the d-quasi-forest $G - X$ that have a conflict with the same independent set $X_I \subseteq X$ of size at most $d + 2$; therefore it holds that

$\textsc{Conf}_{H_i}(X_I) > 0$ for all $i \in \{1, 2, \ldots, p\}$. But now, for all $i \in \{1, 2, \ldots, p\}$

$$\textsc{Conf}_{G-H_i-X}(X_I) \geq \sum_{\substack{j=1 \\ j \neq i}}^{p} \textsc{Conf}_{H_j}(X_I) \geq p - 1 \geq |X|,$$

where the first inequality corresponds to summing over some connected components of $G - H_i - X$. Thus, X_I could not be the reason why the connected components $H_1, H_2, \ldots, H_p$ are not reduced during Rule 1.

This leads to the claimed bound of at most $\binom{|X|}{\leq d+2} \cdot |X| \leq |X|^{d+3}$ connected components in $G - X$, because for every independent set $X_I \subseteq X$ of size at most $d + 2$ there are at most $|X|$ connected components for which X_I is the reason that we cannot apply Rule 1. ◀

It remains to show that Rule 1 is safe; i.e. that there exists a solution for (G, X, k) if and only if there exists a solution for (G', X, k'), where $G' = G - H$, $k' = k - \alpha(H)$ and H is the connected component of $G - X$ we delete during Rule 1. The main ingredient for this is to prove that any minimal blocking set has size at most $d + 2$ (Lemma 6). To bound the size of minimal blocking sets we need the existence of a half integral solution x to $\mathrm{LP}_{\mathrm{IS}}(G - Y)$ for which *every* maximum independent set I in $G - Y$ fulfills $V_1 \subseteq I \subseteq V_{\frac{1}{2}} \cup V_1$. This is similar to the result of Nemhauser and Trotter [17] and other results about the connection between maximum independent sets (resp. minimum vertex covers) and their fractional LP solutions [1, 4, 6, 11].

▶ **Lemma 3.** *Let $G = (V, E)$ be an undirected graph. There exists an optimum half integral solution $x \in \{0, \frac{1}{2}, 1\}^{|V|}$ to $\mathrm{LP}_{\mathrm{IS}}(G)$ such that for all maximum independent sets I in G it holds that $V_1^x \subseteq I \subseteq V \setminus V_0^x$.*

Proof. Let $x \in \{0, \frac{1}{2}, 1\}^{|V|}$ be an optimum half integral solution to $\mathrm{LP}_{\mathrm{IS}}(G)$, such that $V_{\frac{1}{2}}^x$ is maximal; this means, that there exists no optimum half integral solution $x' \neq x$ to $\mathrm{LP}_{\mathrm{IS}}(G)$ such that $V_{\frac{1}{2}}^x \subsetneq V_{\frac{1}{2}}^{x'}$. We will show that every independent set I in G with $V_1^x \nsubseteq I$ or $V_0^x \cap I \neq \emptyset$ is not a maximum independent set in G.

First, we observe that for all subsets $V_0' \subseteq V_0^x$ it must hold that the size of the neighborhood of V_0' in V_1^x is larger than the size of V_0', i.e. $|V_1^x \cap N(V_0')| > |V_0'|$; if this is not the case, then we can construct an optimum half integral solution x' to $\mathrm{LP}_{\mathrm{IS}}(G)$ with $V_{\frac{1}{2}}^x \subsetneq V_{\frac{1}{2}}^{x'}$ (which contradicts the fact that $V_{\frac{1}{2}}^x$ is maximal), by assigning a value of $\frac{1}{2}$ to all vertices in $(V_1^x \cap N(V_0')) \cup V_0'$. Obviously, it holds that $V_{\frac{1}{2}}^x \subsetneq V_{\frac{1}{2}}^{x'}$ and that

$$w(x') = w(x) - |V_1^x \cap N(V_0')| + \frac{1}{2}(|V_1^x \cap N(V_0')| + |V_0'|) \geq w(x).$$

In order to show that x' is indeed a feasible solution to $\mathrm{LP}_{\mathrm{IS}}(G)$, it suffices to consider edges $\{u, v\}$ of G that have at least one endpoint in V_0', say $v \in V_0'$, because these are the only vertices for which we increase the value of the half integral solution x to obtain x'. Since $x_v' = \frac{1}{2}$, the constraint $x_u' + x_v' \leq 1$ can only be violated if $x_u' = 1$. But then $x_u = 1$ must hold since the only changed values are $\frac{1}{2}$ in x'. This of course means that $u \in V_1^x \cap N(V_0')$ and $x_u' = \frac{1}{2}$; a contradiction.

Now, we assume that there exists a maximum independent set I that contains a vertex of the set V_0^x. Let $V_0' = V_0^x \cap I \neq \emptyset$. We will show that deleting the set V_0' from the independent set I and adding the set $N(V_0') \cap V_1^x$ to the independent set I leads to a larger independent I' of G, i.e. $I' = I \setminus V_0' \cup (N(V_0') \cap V_1^x)$. First we show that I' has larger cardinality than I. Since I is an independent set, we know that $(N(V_0') \cap V_1^x) \cap I = \emptyset$ and hence that the cardinality of I' is $|I| - |V_0'| + |N(V_0') \cap V_1^x|$. From the above observation, we know that $|N(V_0') \cap V_1^x| > |V_0'|$ and it follows that I' has larger cardinality than I. To prove that I' is an independent set in G, it is enough to show that any vertex $v \in N(V_0') \cap V_1^x$ has no neighbor in I'; this holds because V_1^x is an independent set, $N(V_1^x) \subseteq V_0^x$ and $V_0^x \cap I' = \emptyset$. Thus, I' is an independent set which has larger cardinality than I; this contradicts the assumption that I is a maximum independent set.

It remains to show that there exists no maximum independent set I in G with $V_1^x \nsubseteq I \subseteq V_1^x \cup V_{\frac{1}{2}}^x$. Let $v \in V_1^x \setminus I$. Since I is a maximum independent set, there exists a vertex $w \in N(V_1^x) \cap I$ (otherwise $I \cup \{v\}$ would be a larger independent set in G). But $N(V_1^x) \subseteq V_0^x$ and hence $w \in V_0^x \cap I$, which contradicts the assumption that $I \subseteq V_1^x \cup V_{\frac{1}{2}}^x$. ◀

Using the above lemma, we can show that every minimal blocking set in a d-quasi-forest has size at most $d+2$. This generalizes the result of Fomin and Stromme [9], who showed that a minimal blocking set in a pseudoforest has size at most three. Furthermore, we can show that this bound is tight.

▶ **Theorem 4.** *Minimal blocking sets have a tight upper bound of $d+2$ in d-quasi-forests.*

The crucial part of Theorem 4 is to prove the upper bound.

▶ **Lemma 5.** *Let $H = (V, E)$ be a d-quasi-forest and let Z be a feedback vertex set in H of size at most d. Then it holds that a minimal blocking set Y in the d-quasi-forest H has size at most $|Z| + 2 \leq d + 2$.*

Proof. We consider an optimum half integral solution x to $\mathrm{LP}_{\mathrm{IS}}(H - Y)$ which fulfills the properties of Lemma 3; let $V_i = \{v \in V(H - Y) \mid x_v = i\}$ for $i \in \{0, \frac{1}{2}, 1\}$. We know that every maximum independent set I of $H - Y$ contains the set V_1 and no vertex of the set V_0 (because x fulfills the properties of Lemma 3).

Observe that for all vertices $y \in Y$ it holds that $\alpha(H - (Y \setminus \{y\})) = \alpha(H)$; otherwise, the set Y would not be a minimal blocking set. Furthermore, from the above observation it follows that $\alpha(H) = \alpha(H - Y) + 1$, because

$$\alpha(H - Y) < \alpha(H) = \alpha(H - (Y \setminus \{y\})) \leq \alpha(H - Y) + 1 \text{ for all } y \in Y.$$

The key observation of our proof is that $N_H(Y) \subseteq V_0 \cup V_{\frac{1}{2}}$; this follows from the fact that Y is minimal: As observed above, we know that $\alpha(H - (Y \setminus \{y\})) = \alpha(H)$. Thus, for all vertices $y \in Y$ there exists a maximum independent set I_y in H that contains the vertex y and no other vertex from the set Y. Consider the sets $I'_y = I_y \setminus \{y\}$ for all vertices $y \in Y$. Obviously, the sets I'_y are independent sets in $H - Y$ for all vertices $y \in Y$, because $y \in Y$ is the only vertex of the set Y that is contained in I_y. Furthermore, we know that the sets I'_y are maximum independent sets in $H - Y$ because

$$|I'_y| + 1 = |I_y| = \alpha(H) = \alpha(H - Y) + 1.$$

The fact that I'_y is a maximum independent set for all vertices $y \in Y$ implies that $V_1 \subseteq I'_y = I_y \setminus \{y\} \subseteq I_y$ (by the choice of the solution x to $\mathrm{LP}_{\mathrm{IS}}(H - Y)$). Thus, for all vertices $y \in Y$ it holds that $V_1 \subseteq I_y$ and therefore that $N_H(I_y) \cap V_1 = \emptyset$ which implies that $N_H(\{y\}) \cap V_1 = \emptyset$ (because $V_1 \cup \{y\} \subseteq I_y$). Since this holds for all vertices $y \in Y$ it follows that $N_H(Y) \cap V_1 = \emptyset$, hence $N_H(Y) \subseteq V_0 \cup V_{\frac{1}{2}}$.

To bound the size of Y we try to find an upper bound for the size of a maximum independent set in $H - Y$ and a lower bound for the size of a maximum independent set in H. An obvious upper bound for the size of a maximum independent set in $H - Y$ is the optimum value of $\mathrm{LP_{IS}}(H-Y)$ which is equal to $|V_1| + \frac{1}{2}|V_{\frac{1}{2}}|$. This leads to an upper bound for $\alpha(H-Y)$:

$$\alpha(H-Y) \leq w(x) = |V_1| + \frac{1}{2}|V_{\frac{1}{2}}| = |V_1| + \frac{1}{2}|H - V_0 - V_1 - Y|$$
$$= |V_1| + \frac{|H - V_0 - V_1|}{2} - \frac{|Y|}{2}, \qquad (1)$$

because $V_0 \cup V_1 \subseteq H - Y$.

Next, we try to find a lower bound for the size of a maximum independent set in H. We will construct an independent set I_H in H and the size of this independent set is a lower bound for the size of a maximum independent set in H. First of all, we add all vertices from the independent set V_1 to I_H; this will prevent every vertex from $N_H(V_1)$ to be part of the independent set I_H. Now, we can extend the independent V_1 by an independent set in $H - N_H[V_1]$. First, observe that $N_H[V_1] \cap Y = \emptyset$, because $V_1 \subseteq (H-Y)$ and $N_H(Y) \cap V_1 = \emptyset$. From this follows that $H - N_H[V_1] = H - V_0 - V_1$, because $N(V_1) = V_0$. Instead of adding an independent set of $H - V_0 - V_1$ to I_H, we add a maximum independent set I_F of the forest $H - V_0 - V_1 - Z$ to I_H; such an independent set I_F has size at least $\frac{1}{2}|H - V_0 - V_1 - Z|$. This leads to the following lower bound for $\alpha(H)$:

$$\alpha(H) \geq |I_H| = |V_1| + |I_F| \geq |V_1| + \frac{|H - V_0 - V_1 - Z|}{2}$$
$$= |V_1| + \frac{|H - V_0 - V_1|}{2} - \frac{|Z \setminus (V_0 \cup V_1)|}{2} \geq |V_1| + \frac{|H - V_0 - V_1|}{2} - \frac{|Z|}{2} \qquad (2)$$

Using the equation $\alpha(H) = \alpha(H-Y) + 1$ together with the upper bound for $\alpha(H-Y)$ and the lower bound for $\alpha(H)$ leads to the requested upper bound for the size of Y:

$$|V_1| + \frac{|H - V_0 - V_1|}{2} - \frac{|Z|}{2} \overset{(2)}{\leq} \alpha(H) = \alpha(H-Y) + 1 \overset{(1)}{\leq} |V_1| + \frac{|H - V_0 - V_1|}{2} - \frac{|Y|}{2} + 1$$
$$\implies |Y| \leq |Z| + 2. \qquad ◀$$

We showed that every minimal blocking set in a d-quasi-forest has size at most $d+2$. To proof Theorem 4 it remains to show that the bound is tight:

Proof of Theorem 4. We show the remaining part of Theorem 4, namely that the bound is tight. Consider the connected graph $H = K_{d+2}$. It holds that H is a d-quasi-forest, because any d vertices from H are a feedback vertex set. It holds that the size of a maximum independent set in a clique is 1, hence $\alpha(H - Y') = 1$ for all subsets $Y' \subsetneq V(H)$. Therefore, $Y = V(H)$ is the only, and hence a minimal, blocking set in H. ◀

Recall that Rule 1 considers the conflicts that a connected component H of the d-quasi-forest $G - X$ has with subsets of X. So far, we only talked about the size of minimal blocking sets instead of the size of minimal subset of X that leads to a conflict. Since every independent set $X_I \subseteq X$ that has a conflict with H, has some neighbors in this component, we know that these vertices are a blocking set of H. Using Lemma 5 we can argue that only a subset of at most $d+2$ vertices (of the neighborhood of X_I in H) is important. Like Jansen and Bodlaender [13] resp. Fomin and Strømme [9] we show how a smaller subset of $V(H) \cap N(X_I)$ leads to a smaller subset of X_I that has a conflict with the connected component H.

▶ **Lemma 6.** *Let (G, X, k) be an instance of* INDEPENDENT SET *parameterized by a modulator to a d-quasi-forest. Let H be a connected component of $G - X$ and let $X_I \subseteq X$ be an independent set in G. If $\mathrm{CONF}_H(X_I) > 0$, then there exists a set $X' \subseteq X_I$ of size at most $d+2$ such that $\mathrm{CONF}_H(X') > 0$.*

We showed that if a component H of $G - X$ has a conflict with a subset $X' \subseteq X$ of the modulator, then there always exists a set $X'' \subseteq X'$ of size at most $d + 2$ that has a conflict with the component H. Knowing this, we can show that Rule 1 is safe using Lemma 6 as well as some observations that where already used in earlier work [9, 13].

▶ **Lemma 7.** *Rule 1 is safe; let (G, X, k) be the instance before applying Rule 1 and let (G', X, k') be the reduced instance. Then there exists a solution for (G, X, k) if and only if there exists a solution for (G', X, k').*

Recall that if we have an instance (G, X, k) of INDEPENDENT SET parameterized by a modulator to a d-quasi-forest where Rule 1 is not applicable then $G - X$ has at most $|X|^{d+3}$ connected components. To apply the kernelization for INDEPENDENT SET parameterized by a modulator to a forest from Jansen and Bodlaender [13], we have to add vertices from each connected component of the d-quasi-forest $G - X$ to the modulator X, getting a set $X' \supseteq X$, such that the connected components of $G - X'$ are trees.

We know that every connected component of the d-quasi-forest $G - X$ has a feedback vertex set of size at most d, which we can find in polynomial time, since d is a constant. Let $Z \subseteq V(G - X)$ be the union of these feedback vertex sets; it holds that $|Z| \leq d \cdot |X|^{d+3}$. Now, the instance (G', X', k') with $G' = G$, $X' = X \cup Z$ and $k' = k$ is an instance of INDEPENDENT SET parameterized by a modulator to feedback vertex set. Obviously, it holds that (G, X, k) has a solution if and only if (G', X', k') has a solution. Applying the following result of Jansen and Bodlaender [13] will finish our kernelization.

▶ **Proposition 8** ([13, Theorem 2])**.** INDEPENDENT SET *parameterized by a modulator to a* FEEDBACK VERTEX SET *has a kernel with a cubic number of vertices: there is a polynomial-time algorithm that transforms an instance (G, X, k) into an equivalent instance (G', X', k') such that $|X'| \leq 2|X|$ and $|V(G')| \leq 2|X| + 28|X|^2 + 56|X|^3$.*

▶ **Theorem 9.** INDEPENDENT SET *parameterized by a modulator to a d-quasi-forest admits a kernel with $\mathcal{O}(d^3|X|^{3d+9})$ vertices.*

▶ **Corollary 10.** VERTEX COVER *parameterized by a modulator to a d-quasi-forest admits a kernel with $\mathcal{O}(d^3|X|^{3d+9})$ vertices.*

4 Two other graph classes with small blocking sets

In this section we consider VERTEX COVER parameterized by a modulator to a d-quasi-bipartite graph and by a modulator to a d-quasi-integral graph. As in the case of VERTEX COVER parameterized by a modulator to a d-quasi-forest, we prove that the size of a minimal blocking set is bounded linearly in d to reduce the number of connected components in the d-quasi-bipartite graph resp. the d-quasi-integral graph. Having only polynomial in the modulator many connected components we show that we can apply the randomized polynomial kernelizations for VERTEX COVER parameterized by a modulator to a bipartite graph, resp. VERTEX COVER above $\mathrm{LP}_{\mathrm{VC}}$.

The proof that there exists a kernelization for VERTEX COVER parameterized by a modulator to a d-quasi-bipartite graph works just the same as the kernelization for VERTEX COVER parameterized by a modulator to a d-quasi-forest, except for the last step. Here we apply the kernelization of VERTEX COVER parameterized by a modulator to a bipartite graph.

▶ **Corollary 11.** *In a d-quasi-bipartite graph the size of a minimal blocking set has a tight upper bound of $d + 2$.*

▶ **Corollary 12.** VERTEX COVER *parameterized by a modulator to a d-quasi-bipartite graph admits a randomized polynomial kernel.*

In contrast to d-quasi-forests and d-quasi-bipartite graphs, where every minimal blocking set is of size at most $d+2$, d-quasi-integral graphs have minimal blocking sets of size up to $2d+2$. Nevertheless, all proofs, to show that there exists a polynomial kernel, still work, because we only need the existence of a small blocking set.

▶ **Lemma 13.** *Let $H = (V, E)$ be a d-quasi-integral graph. Then it holds that a minimal blocking set Y in the d-quasi-integral graph H has size at most $2d + 2$.*

Proof. Like in the proof of Lemma 5 we consider an optimum half integral solution x to $\mathrm{LP_{IS}}(H - Y)$ which fulfills the properties of Lemma 3. Let $V_i = \{v \in V(H - Y) \mid x_v = i\}$ for $i \in \{0, \frac{1}{2}, 1\}$.

Note that the upper bound $\alpha(H - Y) \overset{(1)}{\leq} |V_1| + \frac{1}{2}|H - V_0 - V_1| - \frac{1}{2}|Y|$ also holds in this case, because the value of an optimum half integral solution is always a valid upper bound.

In this case the lower bound for $\alpha(H)$ works slightly differently. Instead of constructing an independent set in H we construct a feasible solution to $\mathrm{LP_{IS}}(H)$. We first use the fact that H is a d-quasi-integral graph, hence $\mathrm{vc}(H) \leq \mathrm{LP_{VC}}(H) + d$, which is equivalent to $\alpha(H) \geq \mathrm{LP_{IS}}(H) - d$, because $\alpha(H) = |H| - \mathrm{vc}(H)$ and $|H| - \mathrm{LP_{VC}}(H) = \mathrm{LP_{IS}}(H)$. Now, we construct a feasible solution x' to $\mathrm{LP_{IS}}(H)$. First, we assign every vertex v in the independent set V_1 the value 1 and every vertex w in $N_H(V_0)$ the value 0. Like in the proof of Lemma 5, it holds that $N_H[V_1] = H - V_0 - V_1$, because $N_H[V_1] \cap Y = \emptyset$. Finally, we assign the value $\frac{1}{2}$ to every vertex in $H - V_0 - V_1$. Obviously, x' is a feasible solution to $\mathrm{LP_{IS}}(H)$. This leads to the following lower bound for $\alpha(H)$:

$$\alpha(H) \geq \mathrm{LP_{IS}}(H) - d \geq |V_1| + \mathrm{LP_{IS}}(H - V_0 - V_1) - d \geq |V_1| + \frac{|H - V_0 - V_1|}{2} - d \tag{3}$$

Again, using the equation $\alpha(H) = \alpha(H - Y) + 1$ together with the upper bound for $\alpha(H - Y)$ and the lower bound for $\alpha(H)$ leads to the requested bound for the size of Y:

$$|V_1| + \frac{|H - V_0 - V_1|}{2} - d \overset{(3)}{\leq} \alpha(H) = \alpha(H - Y) + 1 \overset{(1)}{\leq} |V_1| + \frac{|H - V_0 - V_1|}{2} - \frac{|Y|}{2} + 1$$
$$\implies |Y| \leq 2d + 2.$$ ◀

▶ **Theorem 14.** *In a d-quasi-integral graph the size of a minimal blocking set has a tight upper bound of $2d + 2$.*

▶ **Theorem 15.** VERTEX COVER *parameterized by a modulator to a d-quasi-integral graph admits a randomized polynomial kernel.*

Proof sketch. Let (G, X, k) be an instance of VERTEX COVER parameterized by a modulator to a d-quasi-integral graph. We can obtain in polynomial time an equivalent instance $(\widetilde{G}, X, \widetilde{k})$ of VERTEX COVER parameterized by a modulator to a d-quasi-integral graphs by applying Rule 1 exhaustively, but instead of decreasing k by $\alpha(H)$ we decrease k by $\mathrm{vc}(H)$. Now, $\widetilde{G} - X$ has at most $|X|^{2d+3}$ connected components, because Lemma 2 uses only the fact that a minimal blocking set in $G - X$ has size at most $d + 2$ (we only have to replace $d + 2$ by $2d+2$). These instances are equivalent, since we only delete connected components H (during Rule 1) of which we know that there exists a minimum vertex cover in G that contains a minimum vertex cover of H. We can assume that $\mathrm{vc}(\widetilde{G} - X) + |X| > \widetilde{k}$. If this is not the case, then we can compute in polynomial time a vertex cover in $\widetilde{G} - X$ of size $\mathrm{vc}(\widetilde{G} - X)$ which together with the set X is a vertex cover in $\widetilde{G}$ of size at most $\widetilde{k}$.

Finally, we apply the kernelization algorithm for VERTEX COVER above $\mathrm{LP_{VC}}$ to the instance $(\widetilde{G}, \widetilde{k})$ and obtain an instance (G', k') in polynomial time. Note that we can bound the parameter $\widetilde{k} - \mathrm{LP_{VC}}(\widetilde{G})$ by a polynomial in the size of X as follows:

$$\begin{aligned}\widetilde{k} - \mathrm{LP_{VC}}(\widetilde{G}) &\leq \widetilde{k} - \mathrm{LP_{VC}}(\widetilde{G} - X) \\ &= \widetilde{k} - \sum_{H \text{ c.c of } \widetilde{G}-X} \mathrm{LP_{VC}}(H) \\ &\leq \widetilde{k} - \sum_{H \text{ c.c of } \widetilde{G}-X} (\mathrm{vc}(H) - d) \\ &= \widetilde{k} + |X|^{2d+3}d - \mathrm{vc}(\widetilde{G} - X) \\ &\leq |X| + |X|^{2d+3}d\end{aligned}$$

Since (G', k') is polynomially bounded in the size of $\widetilde{k} - \mathrm{LP_{VC}}(\widetilde{G})$, which is bounded by a polynomial in the size of $|X|$, we know that the instance $(G', X' = V(G'), k')$ is an equivalent instance of VERTEX COVER parameterized by a modulator to a d-quasi-integral graph. ◀

5 Conclusion

Starting from the work of Fomin and Strømme [9] we have presented new results for polynomial kernels for VERTEX COVER subject to structural parameters. Our results for modulators to d-quasi-forests show that bounds on the feedback vertex set size are more meaningful for kernelization than the treewidth of $G - X$ (recalling that there is a lower bound for treewidth of $G - X$ being at most two). By extending our kernelization to work for modulators to (d-quasi-bipartite and) d-quasi-integral graphs, we have encompassed existing kernelizations for parameterization by distance to forests [13], distance to max degree two [16] (both previously subsumed by), distance to pseudoforests [9], and parameterization above fractional optimum [15]. It would be interesting whether there is a single positive result that encompasses all parameterizations with polynomial kernels.

To obtain our results we have established tight bounds for the size of minimal blocking sets in d-quasi-forests, d-quasi-bipartite graphs, and d-quasi-integral graphs. Tightness comes from the fact that cliques of size $d + 2$ respectively $2d + 2$ are contained in these classes. The presence of these cliques also implies lower bounds ruling out kernels of size $\mathcal{O}(|X|^{r-\epsilon})$, assuming $\mathsf{NP} \nsubseteq \mathsf{coNP/poly}$, when $r = r(d)$ is the maximum size of minimal blocking sets as a consequence of a lower bound by Majumdar et al. [16]. It would be interesting whether there are matching upper bounds for kernelization, e.g., whether the kernelization of Jansen and Bodlaender [13] for modulators to forests can be improved to size $\mathcal{O}(|X|^2)$.

References

1. Faisal N. Abu-Khzam, Michael R. Fellows, Michael A. Langston, and W. Henry Suters. Crown structures for vertex cover kernelization. *Theory Comput. Syst.*, 41(3):411–430, 2007. `doi:10.1007/s00224-007-1328-0`.
2. Hans L. Bodlaender, Rodney G. Downey, Michael R. Fellows, and Danny Hermelin. On problems without polynomial kernels. *J. Comput. Syst. Sci.*, 75(8):423–434, 2009. `doi:10.1016/j.jcss.2009.04.001`.
3. Hans L. Bodlaender, Bart M. P. Jansen, and Stefan Kratsch. Kernelization lower bounds by cross-composition. *SIAM J. Discrete Math.*, 28(1):277–305, 2014. `doi:10.1137/120880240`.

4 Endre Boros, Martin Charles Golumbic, and Vadim E. Levit. On the number of vertices belonging to all maximum stable sets of a graph. *Discrete Applied Mathematics*, 124(1-3):17–25, 2002. doi:10.1016/S0166-218X(01)00327-4.

5 Marin Bougeret and Ignasi Sau. How much does a treedepth modulator help to obtain polynomial kernels beyond sparse graphs? *CoRR*, abs/1609.08095, 2016. arXiv:1609.08095.

6 Miroslav Chlebík and Janka Chlebíková. Crown reductions for the minimum weighted vertex cover problem. *Discrete Applied Mathematics*, 156(3):292–312, 2008. doi:10.1016/j.dam.2007.03.026.

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

8 Reinhard Diestel. *Graph Theory, 4th Edition*, volume 173 of *Graduate texts in mathematics*. Springer, 2012.

9 Fedor V. Fomin and Torstein J. F. Strømme. Vertex cover structural parameterization revisited. In Pinar Heggernes, editor, *Graph-Theoretic Concepts in Computer Science - 42nd International Workshop, WG 2016, Istanbul, Turkey, June 22-24, 2016, Revised Selected Papers*, volume 9941 of *Lecture Notes in Computer Science*, pages 171–182, 2016. doi:10.1007/978-3-662-53536-3_15.

10 Shivam Garg and Geevarghese Philip. Raising the bar for vertex cover: Fixed-parameter tractability above A higher guarantee. In Robert Krauthgamer, editor, *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 1152–1166. SIAM, 2016. doi:10.1137/1.9781611974331.ch80.

11 Peter L. Hammer, Pierre Hansen, and Bruno Simeone. Vertices belonging to all or to no maximum stable sets of a graph. *SIAM Journal on Algebraic Discrete Methods*, 3(4):511–522, 1982.

12 Bart M. P. Jansen. *The power of data reduction: Kernels for fundamental graph problems*. PhD thesis, Utrecht University, 2013.

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

14 Stefan Kratsch. A randomized polynomial kernelization for vertex cover with a smaller parameter. In Piotr Sankowski and Christos D. Zaroliagis, editors, *24th Annual European Symposium on Algorithms, ESA 2016, August 22-24, 2016, Aarhus, Denmark*, volume 57 of *LIPIcs*, pages 59:1–59:17. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2016. doi:10.4230/LIPIcs.ESA.2016.59.

15 Stefan Kratsch and Magnus Wahlström. Representative sets and irrelevant vertices: New tools for kernelization. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 450–459. IEEE Computer Society, 2012. doi:10.1109/FOCS.2012.46.

16 Diptapriyo Majumdar, Venkatesh Raman, and Saket Saurabh. Kernels for structural parameterizations of vertex cover - case of small degree modulators. In Thore Husfeldt and Iyad A. Kanj, editors, *10th International Symposium on Parameterized and Exact Computation, IPEC 2015, September 16-18, 2015, Patras, Greece*, volume 43 of *LIPIcs*, pages 331–342. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2015. doi:10.4230/LIPIcs.IPEC.2015.331.

17 George L. Nemhauser and Leslie E. Trotter Jr. Vertex packings: Structural properties and algorithms. *Math. Program.*, 8(1):232–248, 1975. doi:10.1007/BF01580444.