Meta-kernelization using Well-structured Modulators*

Eduard Eiben, Robert Ganian, and Stefan Szeider

Algorithms and Complexity Group, TU Wien
Vienna, Austria

Abstract

Kernelization investigates exact preprocessing algorithms with performance guarantees. The most prevalent type of parameters used in kernelization is the solution size for optimization problems; however, also structural parameters have been successfully used to obtain polynomial kernels for a wide range of problems. Many of these parameters can be defined as the size of a smallest *modulator* of the given graph into a fixed graph class (i.e., a set of vertices whose deletion puts the graph into the graph class). Such parameters admit the construction of polynomial kernels even when the solution size is large or not applicable. This work follows up on the research on meta-kernelization frameworks in terms of structural parameters.

We develop a class of parameters which are based on a more general view on modulators: instead of size, the parameters employ a combination of rank-width and split decompositions to measure structure inside the modulator. This allows us to lift kernelization results from modulator-size to more general parameters, hence providing smaller kernels. We show (i) how such large but well-structured modulators can be efficiently approximated, (ii) how they can be used to obtain polynomial kernels for any graph problem expressible in Monadic Second Order logic, and (iii) how they allow the extension of previous results in the area of structural meta-kernelization.

1998 ACM Subject Classification F.1.3 Complexity Measures and Classes, G.2.1 Combinatorics

Keywords and phrases Kernelization, Parameterized complexity, Structural parameters, Rank-width, Split decompositions

Digital Object Identifier 10.4230/LIPIcs.IPEC.2015.114

1 Introduction

Kernelization investigates exact preprocessing algorithms with performance guarantees. Similarly as in parameterized complexity analysis, in kernelization we study *parameterized problems*: decision problems where each instance I comes with a parameter k. A parameterized problem is said to admit a kernel of size $f : \mathbb{N} \to \mathbb{N}$ if every instance (I, k) can be reduced in polynomial time to an equivalent instance (called the *kernel*) whose size and parameter are bounded by $f(k)$. For practical as well as theoretical reasons, we are mainly interested in the existence of *polynomial kernels*, i.e., kernels whose size is polynomial in k. The study of kernelization has recently been one of the main areas of research in parameterized complexity, yielding many important new contributions to the theory.

The by far most prevalent type of parameter used in kernelization is the *solution size*. Indeed, the existence of polynomial kernels and the exact bounds on their sizes have been studied for a plethora of distinct problems under this parameter, and the rate of advancement

* The authors acknowledge support by the Austrian Science Fund (FWF, projects P26696 and W1255-N23).

10th International Symposium on Parameterized and Exact Computation (IPEC 2015).
Editors: Thore Husfeldt and Iyad Kanj; pp. 114–126
LIPICS Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

achieved in this direction over the past 10 years has been staggering. Important findings were also obtained in the area of *meta-kernelization* [4, 12, 18], which is the study of general kernelization techniques and frameworks used to establish polynomial kernels for a wide range of distinct problems.

In parameterized complexity analysis, an alternative to parameterization by solution size has traditionally been the use of *structural parameters*. But while parameters such as *treewidth* and the more general *rank-width* allow the design of FPT algorithms for a range of important problems, it is known that they cannot be used to obtain polynomial kernels for problems of interest. Instead, the structural parameters used for kernelization often take the form of the size of minimum *modulators* (a modulator of a graph is a set of vertices whose deletion puts the graph into a fixed graph class). Examples of such parameters include the size of a minimum vertex cover [11, 5] (modulators into the class of edgeless graphs) or of a minimum feedback vertex set [6, 17] (modulators into the class of forests). While such structural parameters are not as universal as the structural parameters used in the context of fixed-parameter tractability, these results nonetheless allow efficient preprocessing of instances where the solution size is large and for problems where solution size simply cannot be used (such as 3-coloring).

This paper follows up on the recent line of research which studies meta-kernelization in terms of structural parameters. Gajarský et al. [13] developed a meta-kernelization framework parameterized by the size of a modulator to the class of graphs of bounded treedepth on sparse graphs. Ganian et al. [15] independently developed a meta-kernelization framework using a different parameter based on rank-width and modular decompositions (see Subsection 2.4 for details). Our results build upon both of the aforementioned papers by fully subsuming the meta-kernelization framework of [15] and lifting the meta-kernelization framework of [13] to more general graph classes. The class of problems investigated in this paper are problems which can be expressed using *Monadic Second Order* (MSO) logic (see Subsection 2.5).

The parameters for our kernelization results are also based on modulators. However, instead of parameterizing by the *size* of the modulator, we instead measure the *structure* of the modulator through a combination of rank-width and split decompositions. Due to its technical nature, we postpone the definition of our parameter, the *well-structure number*, to Section 3; for now, let us roughly describe it as the number of sets one can partition a modulator into so that each set induces a graph with bounded rank-width and a simple neighborhood. We call modulators which satisfy our conditions *well-structured*. A less restricted variant of the well-structure number has recently been used to obtain meta-theorems for FPT algorithms on graphs of unbounded rank-width [10].

After formally introducing the parameter, in Section 4 we showcase its applications on the special case of generalizing the *vertex cover number* by considering well-structured modulators to edgeless graphs. While it is known that there exist MSO-definable problems which do not admit a polynomial kernel parameterized by the vertex cover number on general graphs, on graphs of bounded expansion this is no longer the case (as follows for instance from [13]). On the class of graphs of bounded expansion, we prove that every MSO-definable problem admits a linear kernel parameterized by the well-structure number for edgeless graphs. As a corollary of our approach, we also show that every MSO-definable problem admits a linear kernel parameterized by the well-structure number for the empty graph (without any restriction on the expansion). We remark that the latter result represents a direct generalization of the results in [15]. The proof is based on a combination of a refined version of the replacement techniques developed in [10] together with the annotation framework used in [15].

Before we can proceed to wider applications of our parameter in kernelization, it is first necessary to deal with the subproblem of finding a suitable well-structured modulator in polynomial time. We resolve this question for well-structured modulators to a vast range of graph classes. In particular, in Subsection 5.1 we obtain a 3-approximation algorithm for finding well-structured modulators to acyclic graphs, and in the subsequent Subsection 5.2 we show how to approximate well-structured modulators to any graph class characterized by a finite set of forbidden induced subgraphs within a constant factor.

Section 6 then contains our most general result, Theorem 11, which is the key for lifting kernelization results from modulators to well-structured modulators. The theorem states that whenever a modulator to a graph class $\mathcal{H}$ can be used to poly-kernelize some MSO-definable problem, this problem also admits a polynomial kernel when parameterized by the well-structure number for $\mathcal{H}$ as long as well-structured modulators to $\mathcal{H}$ can be approximated in polynomial time. The remainder of Section 6 then deals with the applications of this theorem. Since the class of graphs of treedepth bounded by some fixed integer can be characterized by a finite set of forbidden induced subgraphs, we can use well-structured modulators to lift the results of [13] from modulators to well-structured modulators for all MSO-definable decision problems. Furthermore, by applying the *protrusion* machinery of [4, 18] we show that, in the case of bounded degree graphs, parameterization by a modulator to acyclic graphs (i.e., a feedback vertex set) allows the computation of a linear kernel for any MSO-definable decision problem. By our framework it then follows that such modulators can also be lifted to well-structured modulators.

2 Preliminaries

The set of natural numbers (that is, positive integers) will be denoted by $\mathbb{N}$. For $i \in \mathbb{N}$ we write $[i]$ to denote the set $\{1, \ldots, i\}$. If $\sim$ is an equivalence relation over a set A, then for $a \in A$ we use $[a]_\sim$ to denote the equivalence class containing a.

2.1 Graphs

We will use standard graph theoretic terminology and notation (cf. [8]). All graphs in this document are simple and undirected.

Given a graph $G = (V(G), E(G))$ and $A \subseteq V(G)$, we denote by $N(A)$ the set of neighbors of A in $V(G) \setminus A$; if A contains a single vertex v, we use $N(v)$ instead of $N(\{v\})$. We use V and E as shorthand for $V(G)$ and $E(G)$, respectively, when the graph is clear from context. $G - A$ denotes the subgraph of G obtained by deleting A. For $A \subseteq V(G)$ we use $G[A]$ to denote the subgraph of G induced by the set A.

2.2 Splits and Split-Modules

A *split* of a connected graph $G = (V, E)$ is a vertex bipartition $\{A, B\}$ of V such that every vertex of $A' = N(B)$ has the same neighborhood in $B' = N(A)$. The sets A' and B' are called *frontiers* of the split.

Let $G = (V, E)$ be a graph. To simplify our exposition, we will use the notion of *split-modules* instead of splits where suitable. A set $A \subseteq V$ is called a *split-module* of G if there exists a connected component $G' = (V', E')$ of G such that $\{A, V' \setminus A\}$ forms a split of G'. Notice that if A is a split-module then A can be partitioned into A_1 and A_2 such that $N(A_2) \subseteq A$ and for each $v_1, v_2 \in A_1$ it holds that $N(v_1) \cap (V' \setminus A) = N(v_2) \cap (V' \setminus A)$; A_1 is then called the frontier of A. For technical reasons, V and $\emptyset$ are also considered split-modules. We say that two disjoint split-modules $X, Y \subseteq V$ are *adjacent* if there exist $x \in X$ and $y \in Y$ such that x and y are adjacent. We use $\lambda(A)$ to denote the frontier of split-module A.

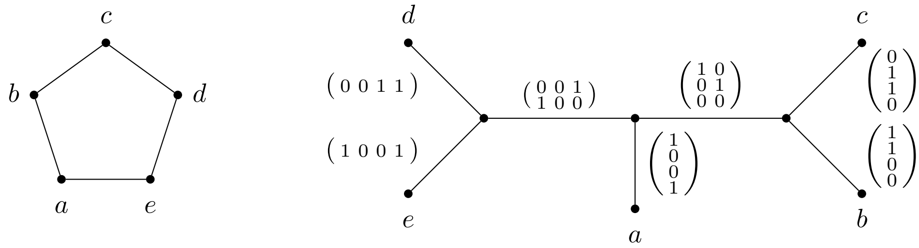

Figure 1 A rank-decomposition of the cycle C_5.

2.3 Rank-Width

For a graph G and $U, W \subseteq V(G)$, let $\boldsymbol{A}_G[U,W]$ denote the $U \times W$-submatrix of the adjacency matrix over the two-element field GF(2), i.e., the entry $a_{u,w}$, $u \in U$ and $w \in W$, of $\boldsymbol{A}_G[U,W]$ is 1 if and only if $\{u,w\}$ is an edge of G. The *cut-rank* function ρ_G of a graph G is defined as follows: For a bipartition (U,W) of the vertex set $V(G)$, $\rho_G(U) = \rho_G(W)$ equals the rank of $\boldsymbol{A}_G[U,W]$ over GF(2).

A *rank-decomposition* of a graph G is a pair (T,μ) where T is a tree of maximum degree 3 and $\mu : V(G) \to \{t : t \text{ is a leaf of } T\}$ is a bijective function. For an edge e of T, the connected components of $T - e$ induce a bipartition (X,Y) of the set of leaves of T. The *width* of an edge e of a rank-decomposition (T,μ) is $\rho_G(\mu^{-1}(X))$. The *width* of (T,μ) is the maximum width over all edges of T. The *rank-width* of G, $rw(G)$ in short, is the minimum width over all rank-decompositions of G. We denote by $\mathcal{R}_i$ the class of all graphs of rank-width at most i, and say that a graph class $\mathcal{H}$ is of *unbounded rank-width* if $\mathcal{H} \not\subseteq \mathcal{R}_i$ for any $i \in \mathbb{N}$.

▶ **Fact 1** ([16]). *Let $k \in \mathbb{N}$ be a constant and $n \geq 2$. For an n-vertex graph G, we can output a rank-decomposition of width at most k or confirm that the rank-width of G is larger than k in time $\mathcal{O}(n^3)$.*

More properties of rank-width can be found, for instance, in [21].

2.4 Fixed-Parameter Tractability and Kernels

We refer to the standard textbooks for basic notions in parameterized complexity such as *parameterized problem*, *kernelization* and *bikernelization* [9]. The following fact links the existence of bikernels to the existence of kernels.

▶ **Fact 2** ([1]). *Let $\mathcal{P}, \mathcal{Q}$ be a pair of decidable parameterized problems such that $\mathcal{Q}$ is in* NP *and $\mathcal{P}$ is* NP*-complete. If there is a bikernelization from $\mathcal{P}$ to $\mathcal{Q}$ producing a polynomial bikernel, then $\mathcal{P}$ has a polynomial kernel.*

Within this paper, we will also consider (and compare to) various structural parameters which have been used to obtain polynomial kernels. We provide a brief overview of these parameters below.

A *modulator* of a graph G to a graph class $\mathcal{H}$ is a vertex set $X \subseteq V(G)$ such that $G - X \in \mathcal{H}$. We denote the cardinality of a minimum modulator to $\mathcal{H}$ in G by $mod^{\mathcal{H}}(G)$. The *vertex cover number* of a graph G ($vcn(G)$) is a special case of $mod^{\mathcal{H}}(G)$, specifically for $\mathcal{H}$ being the set of edgeless graphs. The vertex cover number has been used to obtain polynomial kernels for problems such as Largest Induced Subgraph [11] or Long Cycle along with other path and cycle problems [5]. Similarly, a *feedback vertex set* is a modulator to the class of acyclic graphs, and the size of a minimum feedback vertex set has been used to kernelize, for instance, Treewidth [6] or Vertex Cover [17].

For the final considered parameter, we will need the notion of *module*, which can be defined as a split-module with the restriction that every vertex in the split-module lies in its frontier. Then the *rank-width$_c$ cover number* [15] of a graph G ($rwc_c(G)$) is the smallest number of *modules* the vertex set of G can be partitioned into such that each module induces a subgraph of rank-width at most c. A wide range of problems, and in particular all MSO-definable problems, have been shown admit linear kernels when parameterized by the rank-width$_c$ cover number [15].

2.5 Monadic Second Order Logic on Graphs

We assume that we have an infinite supply of individual variables, denoted by lowercase letters x, y, z, and an infinite supply of set variables, denoted by uppercase letters X, Y, Z. *Formulas* of *monadic second-order logic* (MSO) are constructed from atomic formulas $E(x, y)$, $X(x)$, and $x = y$ using the connectives $\neg$ (negation), $\wedge$ (conjunction) and existential quantification $\exists x$ over individual variables as well as existential quantification $\exists X$ over set variables. Individual variables range over vertices, and set variables range over sets of vertices. The atomic formula $E(x, y)$ expresses adjacency, $x = y$ expresses equality, and $X(x)$ expresses that vertex x in the set X. From this, we define the semantics of monadic second-order logic in the standard way (this logic is sometimes called MSO_1).

Free and bound variables of a formula are defined in the usual way. A *sentence* is a formula without free variables. We write $\varphi(X_1, \dots, X_n)$ to indicate that the set of free variables of formula φ is $\{X_1, \dots, X_n\}$. If $G = (V, E)$ is a graph and $S_1, \dots, S_n \subseteq V$ we write $G \models \varphi(S_1, \dots, S_n)$ to denote that φ holds in G if the variables X_i are interpreted by the sets S_i, for $i \in [n]$. The problem framework we are mainly interested in is formalized below.

MSO Model Checking (MSO-MC$_\varphi$)
Instance: A graph G.
Question: Does $G \models \varphi$ hold?

While MSO model checking problems already capture many important graph problems, there are some well-known problems on graphs that cannot be captured in this way, such as Vertex Cover, Dominating Set, and Clique. Many such problems can be formulated in the form of *MSO optimization problems.* Let $\varphi = \varphi(X)$ be an MSO formula with one free set variable X and $\diamondsuit \in \{\leq, \geq\}$.

MSO-Opt$_\varphi^\diamondsuit$
Instance: A graph G and an integer $r \in \mathbb{N}$.
Question: Is there a set $S \subseteq V(G)$ such that $G \models \varphi(S)$ and $|S| \diamondsuit r$?

It is known that MSO formulas can be checked efficiently as long as the graph has bounded rank-width.

▶ **Fact 3** ([14]). *Let φ and $\psi = \psi(X)$ be fixed MSO formulas and let c be a constant. Then* MSO-MC$_\varphi$ *and* MSO-Opt$_\varphi^\diamondsuit$ *can be solved in $\mathcal{O}(n^3)$ time on the class of graphs of rank-width at most c, where n is the order of the input graph. Moreover, if G has rank-width at most c and and $S \subseteq V(G)$, it is possible to check whether $G \models \psi(S)$ in $\mathcal{O}(n^3)$ time.*

We review MSO *types* roughly following the presentation in [19]. The *quantifier rank* of an MSO formula φ is defined as the nesting depth of quantifiers in φ. For non-negative

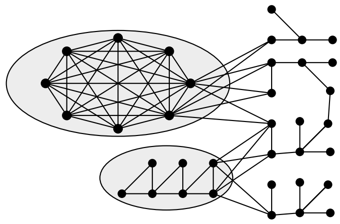

Figure 2 A graph with a $(2,1)$-well-structured modulator to forests (in the two shaded areas).

integers q and l, let $\mathrm{MSO}_{q,l}$ consist of all MSO formulas of quantifier rank at most q with free set variables in $\{X_1, \ldots, X_l\}$.

Let $\varphi = \varphi(X_1, \ldots, X_l)$ and $\psi = \psi(X_1, \ldots, X_l)$ be MSO formulas. We say φ and ψ are *equivalent*, written $\varphi \equiv \psi$, if for all graphs G and $U_1, \ldots, U_l \subseteq V(G)$, $G \models \varphi(U_1, \ldots, U_l)$ if and only if $G \models \psi(U_1, \ldots, U_l)$. Given a set F of formulas, let $F/\equiv$ denote the set of equivalence classes of F with respect to $\equiv$. A system of representatives of $F/\equiv$ is a set $R \subseteq F$ such that $R \cap C \neq \emptyset$ for each equivalence class $C \in F/\equiv$. The following statement has a straightforward proof using normal forms (see [19, Proposition 7.5] for details).

▶ **Fact 4** ([15]). *Let q and l be fixed non-negative integers. The set $MSO_{q,l}/\equiv$ is finite, and one can compute a system of representatives of $MSO_{q,l}/\equiv$.*

We will assume that for any pair of non-negative integers q and l the system of representatives of $\mathrm{MSO}_{q,l}/\equiv$ given by Fact 4 is fixed.

3 (k,c)-Well-Structured Modulators

▶ **Definition 1.** Let $\mathcal{H}$ be a graph class and let G be a graph. A set $\vec{X}$ of pairwise-disjoint split-modules of G is called a (k,c)-*well-structured modulator* to $\mathcal{H}$ if

1. $|\vec{X}| \leq k$, and
2. $\bigcup_{X_i \in \vec{X}} X_i$ is a modulator to $\mathcal{H}$, and
3. $rw(G[X_i]) \leq c$ for each $X_i \in \vec{X}$.

For the sake of brevity and when clear from context, we will sometimes identify $\vec{X}$ with $\bigcup_{X_i \in \vec{X}} X_i$ (for instance $G - \vec{X}$ is shorthand for $G - \bigcup_{X_i \in \vec{X}} X_i$). To allow a concise description of our parameters, for any hereditary graph class $\mathcal{H}$ we let the *well-structure number* ($wsn_c^{\mathcal{H}}$ in short) denote the minimum k such that G has a (k,c)-well-structured modulator to $\mathcal{H}$.

We conclude this section with a brief discussion on the choice of the parameter. The specific conditions restricting the contents of the modulator $\bigcup \vec{X}$ have been chosen as the most general means which allow both (1) the efficient finding of a suitable well-structured modulator, and (2) the efficient use of this well-structured modulator for kernelization. In this sense, we do not claim that there is anything inherently special about rank-width or split modules, other than being the most general notions which are currently known to allow the achievement of these two goals.

In some of the applications of our results, we will consider graphs which have bounded expansion or bounded degree. We remark that in these cases, our results could equivalently be stated in terms of treewidth (instead of rank-width) and MSO_2 logic (instead of MSO_1 logic).?

4 A Case Study: Vertex Cover

In this section we show how well-structured modulators to edgeless graphs can be used to obtain polynomial kernels for various problems. In particular, this special case can be viewed as a generalization of the vertex cover number. We begin by comparing the resulting parameter to known structural parameters. Let $c \in \mathbb{N}$ be fixed and $\mathcal{E}$ denote the class of edgeless graphs. The class $\mathcal{Z}$ containing only the empty graph will also be of importance later on in the section; we remark that while $mod^{\mathcal{Z}}$ represents a very weak parameter as it is equal to the order of the graph, this is not the case for $wsn_c^{\mathcal{Z}}$. We begin by comparing well-structured modulators to edgeless graphs with similar parameters used in kernelization.

▶ **Proposition 2.** *Let $\mathcal{E}$ be graph class of edgeless graphs. Then:*
1. *$rwc_c(G) \geq wsn_c^{\mathcal{E}}(G)$ for any graph G. Furthermore, for every $i \in \mathbb{N}$ there exists a graph G_i such that $rwc_c(G_i) \geq 2i$ and $wsn_c^{\mathcal{E}} = 2$.*
2. *$vcn(G) \geq wsn_1^{\mathcal{E}}(G)$ for any graph G. Furthermore, for every $i \in \mathbb{N}$ there exists a graph G_i such that $vcn(G) \geq i$ and $wsn_1^{\mathcal{E}} = 1$.*

It will be useful to observe that the above Proposition 2 also holds when restricted to the class of graphs of bounded expansion and bounded degree, and even when the graph class $\mathcal{E}$ is replaced by $\mathcal{Z}$.

As we have established that already $wsn_1^{\mathcal{E}} \leq vcn(G)$, it is important to mention that an additional structural restriction on the graph is necessary to allow the polynomial kernelization of MSO-Opt problems in general (as is made explicit in the following Fact 5).

▶ **Fact 5** ([7])**.** Clique *parameterized by the vertex cover number does not admit a polynomial kernel, unless* NP $\subseteq$ coNP/poly*.*

However, it turns out that restricting the inputs to graphs of bounded expansion completely changes the situation: under this condition, it is not only the case that all all MSO-MC and MSO-Opt problems admit a linear kernel when parameterized by the vertex cover number, but also when parameterized by the more general parameter $wsn_c^{\mathcal{E}}$. To prove these claims, we begin by stating the following result.

▶ **Fact 6** ([13])**.** *Let $\mathcal{K}$ be a graph class with bounded expansion. Suppose that for $G \in \mathcal{K}$ and $S \in V(G)$, $\mathcal{C}_1, \dots, \mathcal{C}_s$ are sets of connected components of $G - S$ such that for all pairs $C, C' \in \cup_i \mathcal{C}_i$ it holds that $C, C' \in \mathcal{C}_j$ for some j if and only if $N_S(C) = N_S(C')$. Let $\delta \geq 0$ be a constant bound on the diameter of these components, i.e., for all $C \in \cup_i \mathcal{C}_i$, $diam(G[V(C)]) \leq \delta$. Then there can be only at most $\mathcal{O}(|S|)$ such sets $\mathcal{C}_i$.*

This allows us to establish a key link between $wsn_c^{\mathcal{E}}$ and $wsn_c^{\mathcal{Z}}$ on graphs of bounded expansion.

▶ **Lemma 3.** *Let $\mathcal{K}$ be a graph class with bounded expansion. Then there exists a constant d such that for every $G \in \mathcal{K}$ it holds that $wsn_c^{\mathcal{Z}}(G) \leq d \cdot (wsn_c^{\mathcal{E}}(G))$.*

Proof. Let $k = wsn_c^{\mathcal{E}}(G)$ and let $\vec{H}$ be a (k, c)-well-structured modulator to $\mathcal{E}$. Let S be a set of vertices containing exactly one vertex from the frontier of every split-module in $\vec{H}$. The graph $G' = G - (\vec{H} - S)$ is a graph with bounded expansion and S is its vertex cover. Clearly, the diameter of every connected component of $G' \setminus S$ is at most 1 (every connected component is a singleton). Therefore, by Fact 6 there exists a constant d' such that there are at most $d' \cdot |S| = d' \cdot wsn_c^{\mathcal{E}}(G)$ sets of vertices $\mathcal{C}_1, \dots, \mathcal{C}_s$ in $G' - S$ such that for all pairs $v, v' \in \cup_i \mathcal{C}_i$ it holds that $v, v' \in \mathcal{C}_j$ for some j if and only if $N_S(v) = N_S(v')$. Clearly each such $\mathcal{C}_i$ is a split-module in G', and hence also in G. Furthermore, each such $\mathcal{C}_i$ has rank-width at most 1. Hence $wsn_c^{\mathcal{Z}}(G) \leq wsn_c^{\mathcal{E}}(G) + d' \cdot wsn_c^{\mathcal{E}}(G)$. ◀

The above lemma allows us to shift our attention from modulators to $\mathcal{E}$ to a partition of the vertex set into split-modules of bounded rank-width. The rest of this section is then dedicated to proving our results for well-structured modulators to $\mathcal{Z}$. Our proof strategy for this special case of well-structured modulators closely follows the replacement techniques used to obtain the kernelization results for the rank-width cover number [15], with the distinction that many of the tools and techniques had to be generalized to cover splits instead of modules.

▶ **Theorem 4.** *Let $\mathcal{K}$ be a graph class of bounded expansion, $\mathcal{E}$ be the class of edgeless graphs and $\mathcal{Z}$ be the class of empty graphs. For every MSO sentence φ the problem* MSO-MC$_\varphi$ *admits a linear kernel parameterized by $wsn_c^{\mathcal{Z}}$. Furthermore, the problem* MSO-MC$_\varphi$ *admits a linear kernel parameterized by $wsn_c^{\mathcal{E}}$ on $\mathcal{K}$.*

Sketch of Proof. By Lemma 3 it is sufficient to show that MSO-MC$_\phi$ admits a linear kernel parameterized by $wsn_c^{\mathcal{Z}}$. Let G be a graph, $k = wsn_c^{\mathcal{Z}}(G)$, and q be the nesting depth of quantifiers in ϕ. By Fact 7 (given in the following section) we can find the set $\vec{X}$ of equivalence classes of $\sim_c^G$ in polynomial time. Clearly, the set $\vec{X}$ is a (k, c)-well-structured modulator to the empty graph. We proceed by using replacement techniques to construct an equivalent graph $(G', \vec{X}')$ such that each $X'_i \in \vec{X}'$ has size bounded by a constant. Since $|\vec{X}'| \le k$ and $\bigcup \vec{X}' = V(G')$, it follows that G' is an instance of MSO-MC$_\varphi$ of size $\mathcal{O}(k)$. ◀

Next, we combine the approaches used in [15] and [10] to handle MSO-OPT$_\varphi^\diamond$ problems by using our more general parameters. Similarly as in [15], we use a more involved replacement procedure which explicitly keeps track of the original cardinalities of sets and results in an *annotated version* of MSO-OPT$_\varphi^\diamond$. However, some parts of the framework (in particular the replacement procedure) had to be reworked using the techniques developed in [10], since we now use split-modules instead of simple modules.

▶ **Theorem 5.** *Let $\mathcal{E}$ be a class of edgeless graphs and $\mathcal{Z}$ be the class containing the empty graph. For every MSO formula φ the problem* MSO-OPT$_\varphi^\le$ *admits a linear bikernel parameterized by $wsn_c^{\mathcal{E}}$ on any class of graphs of bounded expansion, and a linear bikernel parameterized by $wsn_c^{\mathcal{Z}}$.*

5 Finding (k, c)-Well-Structured Modulators

For the following considerations, we fix c and assume that the graph G has rank-width at least $c + 2$ (this is important for Fact 7). This assumption is sound, since the considered problems can be solved in polynomial time on graphs of bounded rank-width. Recall that given a split-module A in G, we use $\lambda(A)$ to denote the frontier of A. This section will show how to efficiently approximate well-structured modulators to various graph classes; in particular, we give algorithms for the class of forests and then for any graph class which can be characterized by a finite set of forbidden induced subgraphs.

The following Fact 7 linking rank-width and split-modules will be crucial for approximating our well-structured modulators.

▶ **Definition 6.** Let G be a graph and $c \in \mathbb{N}$. We define a relation $\sim_c^G$ on $V(G)$ by letting $v \sim_c^G w$ if and only if there is a split-module M of G with $v, w \in M$ and $rw(G[M]) \le c$. We drop the superscript from $\sim_c^G$ if the graph G is clear from context.

▶ **Fact 7** ([10]). *Let $c \in \mathbb{N}$ be fixed and G be a graph of rank-width at least $c + 2$. The relation $\sim_c^G$ is an equivalence, and any graph G has its vertex set uniquely partitioned by* *the equivalence classes of* $\sim_c$ *into inclusion-maximal split-modules of rank-width at most* c*. Furthermore, for* $a, b \in V(G)$ *it is possible to test* $a \sim_c b$ *in* $\mathcal{O}(n^3)$ *time.*

5.1 Finding (k, c)-Well-Structured Modulators to Forests

Our starting point is the following lemma, which shows that long cycles which hit a non-singleton frontier imply the existence of short cycles.

▶ **Lemma 7.** *Let* C *be a cycle in* G *such that* C *intersects at least three distinct equivalence classes of* $\sim_c$*, one of which has a frontier of cardinality at least* 2*. Let* Z *be the set of equivalence classes of* $\sim_c$ *which intersect* C*. Then there exists a cycle* C' *such that the set* Z' *of equivalence classes it intersects is a subset of* Z *and has cardinality at most* 3*.*

We will use the following observation to proceed when Lemma 7 cannot be applied.

▶ **Observation 1.** *Assume that for each equivalence class* B *of* $\sim_c$ *it holds that* $G[B]$ *is acyclic, and that no cycle intersects* B *if* $|\lambda(B)| \geq 2$*. Then for every cycle* C *in* G *and every vertex* $a \in C$*, it holds that* a *is in the frontier of some equivalence class of* $\sim_c$*.*

Fact 8 below is the last ingredient needed for the algorithm.

▶ **Fact 8** ([3])**.** FEEDBACK VERTEX SET *can be* 2*-approximated in polynomial time.*

▶ **Theorem 8.** *Let* $c \in \mathbb{N}$ *and* $\mathcal{F}$ *be the class of forests. There exists a polynomial algorithm which takes as input a graph* G *of rank-width at least* $c+2$ *and computes a set* $\vec{X}$ *of split-modules such that* $\vec{X}$ *is a* (k, c)*-well-structured modulator to* $\mathcal{F}$ *and* $k \leq 3 \cdot wsn_c^{\mathcal{F}}$*.*

Sketch of Proof. The algorithm proceeds in three steps.

1. By deciding $a \sim_c b$ for each pair of vertices in G as per Fact 7, we compute the equivalence classes of $\sim_c$.
2. For each set of up to three equivalence classes $\{A_1, A_2, A_3\}$ of $\sim_c$, we check if $G[A_1 \cup A_2 \cup A_3]$ is acyclic; if it's not, then we add A_1, A_2 and A_3 to $\vec{X}$ and set $G := G - (A_1 \cup A_2 \cup A_3)$.
3. We use Fact 8 to 2-approximate a feedback vertex set S of G in polynomial time; let S' contain every equivalence class of $\sim_c$ which intersects S. We then set $\vec{X} := \vec{X} \cup S'$, and output $\vec{X}$. ◀

5.2 Finding (k, c)-Well-Structured Modulators via Obstructions

Here we will show how to efficiently find a sufficiently small (k, c)-well-structured modulator to any graph class which can be characterized by a finite set of forbidden induced subgraphs. Let us fix a graph class $\mathcal{H}$ characterized by a set $\mathcal{R}$ of forbidden induced subgraphs, and let r be the maximum order of a graph in $\mathcal{R}$. Our first step is to reduce our problem to the classical HITTING SET problem, the definition of which is recalled below.

> d-HITTING SET
> *Instance*: A ground set S and a collection $\mathcal{C}$ of subsets of S, each of cardinality at most d.
> *Notation*: A hitting set is a subset of S which intersects each set in $\mathcal{C}$.
> *Task*: Find a minimum-cardinality hitting set.

Given a graph G (of rank-width at least $c+2$), we construct an instance W_G of r-HITTING SET as follows. The ground set of W contains each equivalence class $A \subseteq V(G)$ of $\sim_c$. For each induced subgraph $R \subseteq G$ isomorphic to an element of $\mathcal{R}$, we add the set C_R of equivalence classes of $\sim_c$ which intersect R into $\mathcal{C}$. This completes the construction of W_G; we let $hit(W_G)$ denote the cardinality of a solution of W_G.

▶ **Lemma 9.** *For any graph G of rank-width at least $c+2$, the instance W_G is unique and can be constructed in polynomial time. Every hitting set Y in W_G is a $(|Y|, c)$-well-structured modulator to $\mathcal{H}$ in G. Moreover, $wsn_c^{\mathcal{H}} = hit(W_G)$.*

The final ingredient we need for our approximation algorithm is the following result.

▶ **Fact 9** (Folklore). *There exists a polynomial-time algorithm which takes as input an instance W of r-*Hitting Set* and outputs a hitting set Y of cardinality at most $r \cdot hit(W)$.*

▶ **Theorem 10.** *Let $c \in \mathbb{N}$ and $\mathcal{H}$ be a class of graphs characterized by a finite set of forbidden induced subgraphs of order at most r. There exists a polynomial algorithm which takes as input a graph G of rank-width at least $c+2$ and computes a (k, c)-well-structured modulator to $\mathcal{H}$ such that $k \leq r \cdot wsn_c^{\mathcal{H}}$.*

Proof. We proceed in two steps: first, we compute the r-Hitting Set instance W_G, and then we use Fact 9 to compute an r-approximate solution Y of W_G in polynomial time. We then set $\vec{X} := Y$ and output. Correctness follows from Lemma 9. ◀

6 Applications of (k, c)-Well-Structured Modulators

We now proceed by outlining the general applications of our results. Our algorithmic framework is captured by the following Theorem 11.

▶ **Theorem 11.** *Let p, q be polynomial functions. For every MSO sentence ϕ and every graph class $\mathcal{H}$ such that*
1. MSO-MC$_\phi$ *admits a (bi)kernel of size $p(mod^{\mathcal{H}}(G))$, and*
2. *there exists a polynomial algorithm which finds a $(q(wsn_c^{\mathcal{H}}), c)$-well-structured modulator to $\mathcal{H}$.*

Then MSO-MC$_\phi$ *admits a (bi)kernel of size $p(q(wsn_c^{\mathcal{H}}(G)))$.*

Let us briefly discuss the limitations of the above theorem. The condition that MSO-MC$_\phi$ admits a polynomial (bi)kernel parameterized by $mod^{\mathcal{H}}(G)$ is clearly necessary for the rest of the theorem to hold, since $wsn^{\mathcal{H}}(G) \leq mod^{\mathcal{H}}(G)$. One might wonder whether a weaker necessary condition could be used instead; specifically, would it be sufficient to require that MSO-MC$_\phi$ is polynomial-time tractable in $\mathcal{H}$? This turns out not to be the case, as follows from the following fact.

▶ **Fact 10** ([10]). *There exists an MSO sentence ϕ and a graph class $\mathcal{H}$ characterized by a finite set of forbidden induced subgraphs such that* MSO-MC$_\phi$ *is polynomial-time tractable on $\mathcal{H}$ but NP-hard on the class of graphs with $mod^{\mathcal{H}}(G) \leq 2$.*

Condition 2 is also necessary for our approach to work, as we need some (approximate) well-structured modulator; luckily, Section 5 shows that a wide variety of studied graph classes satisfy this condition. Finally, one can also rule out an extension of Theorem 11 to MSO-Opt problems (which was possible in the special case considered in Section 4), as we show below.

▶ **Lemma 12.** *There exists an MSO formula φ and a graph class $\mathcal{H}$ characterized by a finite obstruction set such that* MSO-Opt$_\varphi^{\leq}$ *admits a bikernel parameterized by $mod^{\mathcal{H}}$ but is* paraNP*-hard parameterized by $wsn_1^{\mathcal{H}}$.*

Sketch of Proof. Consider the formula $\varphi(S) = \text{fvs}(S) \vee \deg(S)$, where $\text{fvs}(S)$ expresses that S is a feedback vertex set in G and $\deg(S)$ expresses that S is a modulator to graphs with maximum degree 4. Let $\mathcal{H}$ be the class of graphs of maximum degree 4. ◀

6.1 Applications of Theorem 11

As our first general application, we consider the results of Gajarský et al. in [13]. Their main result is summarized below.

▶ **Fact 11** ([13]). *Let* Π *be a problem with finite integer index,* $\mathcal{K}$ *a class of graphs of bounded expansion,* $d \in \mathbb{N}$*, and* $\mathcal{H}$ *be the class of graphs of treedepth at most* d*. Then there exist an algorithm that takes as input* $(G, \xi) \in \mathcal{K} \times \mathbb{N}$ *and in time* $\mathcal{O}(|G| + \log \xi)$ *outputs* (G', ξ') *such that*

1. $(G, \xi) \in \Pi$ *if and only if* $(G', \xi') \in \Pi$*;*
2. G' *is an induced subgraph of* G*; and*
3. $|G'| = \mathcal{O}(mod^{\mathcal{H}}(G))$*.*

The following fact provides a link between the notion of *finite integer index* used in the above result and the MSO-MC$_\varphi$ problems considered in this paper.

▶ **Fact 12** ([2], see also [4]). *For every MSO sentence* φ*, it holds that* MSO-MC$_\varphi$ *is finite-state and hence has finite integer index.*

Finally, the following well-known fact is the last ingredient we need to apply our machinery.

▶ **Fact 13** ([20], page 138). *Let* $d \in \mathbb{N}$ *and* $\mathcal{H}$ *be the class of graph of treedepth at most* d*. Then* $\mathcal{H}$ *can be characterized by finite set of forbidden induced subgraphs.*

▶ **Theorem 13.** *Let* $c, d \in \mathbb{N}$ *and* $\mathcal{H}$ *be the class of graphs of treedepth at most* d*. For every MSO sentence* φ*, it holds that* MSO-MC$_\varphi$ *admits a linear kernel parameterized by* $wsn_c^{\mathcal{H}}$ *on any class of graphs of bounded expansion.*

As our second general application, we consider well-structured modulators to the class of forests. Lemma 14 shows that feedback vertex set may be used to kernelize any MSO-definable decision problem on graphs of bounded degree.

▶ **Lemma 14.** *Let* $\mathcal{F}$ *be the class of forests and* $d \in \mathbb{N}$*. For every MSO sentence* φ*, it holds that* MSO-MC$_\varphi$ *admits a linear kernel parameterized by* $mod^{\mathcal{F}}$ *on any class of graphs of degree at most* d*.*

With Lemma 14, the proof of the theorem below is analogous to the proof of Theorem 13.

▶ **Theorem 15.** *Let* $c \in \mathbb{N}$ *and* $\mathcal{F}$ *be the class of forests. For every MSO sentence* φ*, it holds that* MSO-MC$_\varphi$ *admits a linear kernel parameterized by* $wsn_c^{\mathcal{F}}$ *on any class of graphs of bounded degree.*

7 Conclusion

Our results show that measuring the structure of modulators can lead to an interesting and, as of yet, relatively unexplored spectrum of structural parameters. Such parameters have the potential of combining the best of decomposition-based techniques and modulator-based techniques, and can be applied both in the context of kernelization (as demonstrated in this work) and FPT algorithms [10]. We believe that further work in the direction of modulators will allow us to push the frontiers of tractability towards new, uncharted classes of inputs.

One possible direction for future research is the question of whether the class of MSO-definable problems considered in Theorem 11 can be extended to other finite-state problems. It would of course also be interesting to see more applications of Theorem 11 and new methods for approximating well-structured modulators. Last but not least, we mention that the split-modules used in the definition of our parameters could in principle be refined to less restrictive notions (for instance cuts of constant *cut-rank* [21]); such a relaxed parameter could still be used to obtain polynomial kernels, as long as there is a way of efficiently approximating or computing such modulators.

References

1 Noga Alon, Gregory Gutin, Eun Jung Kim, Stefan Szeider, and Anders Yeo. Solving max-r-sat above a tight lower bound. *Algorithmica*, 61(3):638–655, 2011.

2 Stefan Arnborg, Bruno Courcelle, Andrzej Proskurowski, and Detlef Seese. An algebraic theory of graph reduction. *J. of the ACM*, 40(5):1134–1164, 1993.

3 Ann Becker and Dan Geiger. Optimization of pearl's method of conditioning and greedy-like approximation algorithms for the vertex feedback set problem. *Artif. Intell.*, 83(1):167–188, 1996.

4 Hans L. Bodlaender, Fedor V. Fomin, Daniel Lokshtanov, Eelko Penninkx, Saket Saurabh, and Dimitrios M. Thilikos. (meta) kernelization. In *FOCS 2009*, pages 629–638. IEEE Computer Society, 2009.

5 Hans L. Bodlaender, Bart M. P. Jansen, and Stefan Kratsch. Kernel bounds for path and cycle problems. *Theor. Comput. Sci.*, 511:117–136, 2013.

6 Hans L. Bodlaender, Bart M. P. Jansen, and Stefan Kratsch. Preprocessing for treewidth: A combinatorial analysis through kernelization. *SIAM J. Discrete Math.*, 27(4):2108–2142, 2013.

7 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.

8 Reinhard Diestel. *Graph Theory*, volume 173 of *Graduate Texts in Mathematics*. Springer Verlag, New York, 2nd edition, 2000.

9 Rodney G. Downey and Michael R. Fellows. *Fundamentals of Parameterized Complexity*. Texts in Computer Science. Springer Verlag, 2013.

10 Eduard Eiben, Robert Ganian, and Stefan Szeider. Solving problems on graphs of high rank-width. In *Algorithms and Data Structures – 14th International Symposium*, volume 9214 of *Lecture Notes in Computer Science*, pages 314–326. Springer, 2015.

11 Fedor V. Fomin, Bart M. P. Jansen, and Michal Pilipczuk. Preprocessing subgraph and minor problems: When does a small vertex cover help? *J. Comput. Syst. Sci.*, 80(2):468–495, 2014.

12 Fedor V. Fomin, Daniel Lokshtanov, Saket Saurabh, and Dimitrios M. Thilikos. Bidimensionality and kernels. In *SODA*, pages 503–510, 2010.

13 Jakub Gajarský, Petr Hliněný, Jan Obdržálek, Sebastian Ordyniak, Felix Reidl, Peter Rossmanith, Fernando Sanchez Villaamil, and Somnath Sikdar. Kernelization using structural parameters on sparse graph classes. In *ESA 2013*, volume 8125 of *Lecture Notes in Computer Science*, pages 529–540. Springer, 2013.

14 Robert Ganian and Petr Hliněný. On parse trees and Myhill-Nerode-type tools for handling graphs of bounded rank-width. *Discr. Appl. Math.*, 158(7):851–867, 2010.

15 Robert Ganian, Friedrich Slivovsky, and Stefan Szeider. Meta-kernelization with structural parameters. In *MFCS*, pages 457–468, 2013.

16 Petr Hliněný and Sang il Oum. Finding branch-decompositions and rank-decompositions. *SIAM J. Comput.*, 38(3):1012–1032, 2008.

17 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.

18 Eun Jung Kim, Alexander Langer, Christophe Paul, Felix Reidl, Peter Rossmanith, Ignasi Sau, and Somnath Sikdar. Linear kernels and single-exponential algorithms via protrusion decompositions. In *ICALP (1)*, pages 613–624, 2013.

19 Leonid Libkin. *Elements of Finite Model Theory.* Springer, 2004.

20 Jaroslav Nesetril and Patrice Ossona de Mendez. *Sparsity – Graphs, Structures, and Algorithms*, volume 28 of *Algorithms and Combinatorics.* Springer, 2012.

21 Sang-il Oum and P. Seymour. Approximating clique-width and branch-width. *J. Combin. Theory Ser. B*, 96(4):514–528, 2006.