# A Randomized Polynomial Kernel for Subset Feedback Vertex Set

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

1 **Institut für Informatik, Universität Bonn, Germany**
hols@cs.uni-bonn.de

1 **Institut für Informatik, Universität Bonn, Germany**
kratsch@cs.uni-bonn.de

## Abstract

The SUBSET FEEDBACK VERTEX SET problem generalizes the classical FEEDBACK VERTEX SET problem and asks, for a given undirected graph $G = (V, E)$, a set $S \subseteq V$, and an integer $k$, whether there exists a set $X$ of at most $k$ vertices such that no cycle in $G - X$ contains a vertex of $S$. It was independently shown by Cygan et al. (ICALP '11, SIDMA '13) and Kawarabayashi and Kobayashi (JCTB '12) that SUBSET FEEDBACK VERTEX SET is fixed-parameter tractable for parameter $k$. Cygan et al. asked whether the problem also admits a polynomial kernelization.

We answer the question of Cygan et al. positively by giving a randomized polynomial kernelization for the equivalent version where $S$ is a set of edges. In a first step we show that EDGE SUBSET FEEDBACK VERTEX SET has a randomized polynomial kernel parameterized by $|S| + k$ with $\mathcal{O}(|S|^2 k)$ vertices. For this we use the matroid-based tools of Kratsch and Wahlström (FOCS '12). Next we present a preprocessing that reduces the given instance $(G, S, k)$ to an equivalent instance $(G', S', k')$ where the size of $S'$ is bounded by $\mathcal{O}(k^4)$. These two results lead to a polynomial kernel for SUBSET FEEDBACK VERTEX SET with $\mathcal{O}(k^9)$ vertices.

**1998 ACM Subject Classification** F.2.2 Nonnumerical Algorithms and Problems

**Keywords and phrases** parameterized complexity, kernelization, subset feedback vertex set

**Digital Object Identifier** 10.4230/LIPIcs.STACS.2016.43

## 1 Introduction

In the SUBSET FEEDBACK VERTEX SET (SUBSET FVS) problem we are given an undirected graph $G = (V, E)$, a set of vertices $S \subseteq V$, and an integer $k$, and have to determine whether there is a set $X$ of at most $k$ vertices that intersects all cycles that contain at least one vertex of $S$. Clearly, because we can choose $S = V$, this is a generalization of the well-studied FEEDBACK VERTEX SET (FVS) problem where, given $G$ and $k$, we have to determine whether some set $X$ of at most $k$ vertices intersects *all cycles* in $G$. FEEDBACK VERTEX SET has been extensively studied in parameterized complexity: It is known to be fixed-parameter tractable (FPT) with parameter $k$, i.e., solvable in time $f(k) \cdot |V|^c$, and after a series of improvements the fastest known algorithms take deterministic time $\mathcal{O}^*(3.619^k)$ [11] and randomized time $\mathcal{O}^*(3^k)$ [2]. It is also known to admit a *polynomial kernelization* [1], i.e., there is an efficient algorithm that reduces any instance $(G, k)$ of FVS to an equivalent instance of size polynomial in $k$; the best known kernelization creates an equivalent instance with $\mathcal{O}(k^2)$ vertices [19].

In 2011, Cygan et al. [3] and Kawarabayashi and Kobayashi [10] independently showed that SUBSET FVS is FPT. The algorithm of Cygan et al. runs in time $2^{\mathcal{O}(k \log k)} n^{\mathcal{O}(1)}$, while the one of Kawarabayashi and Kobayashi runs in time $\mathcal{O}(f(k) \cdot n^2 m)$. Wahlström [21] then gave the first single-exponential algorithm with running time $4^k \cdot n^{\mathcal{O}(1)}$; an algorithm with

33rd Symposium on Theoretical Aspects of Computer Science (STACS 2016).
Editors: Nicolas Ollinger and Heribert Vollmer; Article No. 43; pp. 43:1–43:14
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

subexponential dependence on $k$ is ruled out under the Exponential-Time Hypothesis (e.g., because SUBSET FVS generalizes VERTEX COVER). More recently, Lokshtanov et al. [13] gave algorithms with deterministic time $2^{\mathcal{O}(k \log k)} \cdot (n+m)$ and randomized time $\mathcal{O}(25.6^k \cdot (n+m))$.

Cygan et al. [3] ask whether the SUBSET FVS problem also admits a polynomial kernelization and suggest that the matroid-based tools of Kratsch and Wahlström [12] could be applicable. The latter work uses representative sets of independent sets in matroids to obtain, amongst others, polynomial kernels for $s$-MULTIWAY CUT and DELETABLE TERMINAL MULTIWAY CUT (DTMWC) with $\mathcal{O}(k^{s+1})$ and $\mathcal{O}(k^3)$ vertices, respectively. In MULTIWAY CUT we are given a graph $G = (V, E)$, a set $T \subseteq V$ of terminals, and an integer $k$ and have to determine whether deletion of at most $k$ non-terminal vertices separates all terminals. In $s$-MULTIWAY CUT the terminal set has size at most $s$, and in DTMWC we are also allowed to delete terminals (which is essentially the same as restricting terminals to be degree one).

Interestingly, Cygan et al. [3] also provide a polynomial-time reduction from MULTIWAY CUT to SUBSET FVS that does not change the parameter value and, hence, is known to imply that SUBSET FVS is at least as hard as MULTIWAY CUT regarding existence of polynomial kernels. Accordingly, MULTIWAY CUT would be the natural next target problem for attempting to find a polynomial kernelization (after $s$-MULTIWAY CUT and DELETABLE TERMINAL MULTIWAY CUT). It appears, however, that the reduction of Cygan et al. is from DELETABLE TERMINAL MULTIWAY CUT rather than from the more general MULTIWAY CUT, and it is not obvious whether similar ideas could yield a reduction from MULTIWAY CUT to SUBSET FVS.

We apply the matroid-based tools of Kratsch and Wahlström [12] and develop a randomized polynomial kernelization that reduces instances $(G, S, k)$ of SUBSET FVS to equivalent instances with at most $\mathcal{O}(k^9)$ vertices; this is our main result. Like Cygan et al. [3] we also work on EDGE SUBSET FVS where $S$ is a set of edges of $G$ and $X$ needs to intersect all cycles that contain at least one edge of $S$; EDGE SUBSET FVS and SUBSET FVS are equivalent [3]. The result is obtained in two parts. In the first part (Section 3) we establish a randomized polynomial kernelization for EDGE SUBSET FVS parameterized by $|S| + k$ that reduces to equivalent instances with at most $\mathcal{O}(|S|^2 k)$ vertices. Note that nontrivial instances have $k < |S|$ since one could otherwise remove $S$ by deleting one endpoint of each edge in $S$. Thus, parameter $|S|$ suffices, but $\mathcal{O}(|S|^2 k)$ gives a tighter overall bound than $\mathcal{O}(|S|^3)$.

At high level, this part is similar to the polynomial kernelization for DELETABLE TERMINAL MULTIWAY CUT. We show that certain solutions $X$, later called *dominant* solutions, allow particular path packings in the underlying graph $G$. For DTMWC this is achieved by a fairly simple replacement argument for solutions $X$ that are not sufficiently well connected to connected components of $G - X$. For EDGE SUBSET FVS the endpoints $T = V(S)$ of edges in $S$ can be regarded as terminals, but this gives a different separation property: Solutions $X$ need not generate many connected components in $G - X$ since only $S$-cycles need to be prevented, and components may contain many vertices of $T$. Rather, in $G - X$ there must be a tree-like (or forest-like) structure with components without $S$-edges playing the role of nodes and with edges given by $S$. Nevertheless, using the tree-like structure, a replacement argument can be found, implying that dominant solutions must create many components in $(G - X) - S$ containing vertices of $T$ and be well connected to them. This allows to set up a gammoid on $G - S$ with sources $T$ and apply, as in [12], a result of Lovász [14] on representative sets in (linear) matroids that is then guaranteed to generate a superset of $X$. Randomization is only needed to generate a matrix representation for the gammoid.

In the second part (Section 4) we give a (deterministic) polynomial-time preprocessing that, given an instance $(G, S, k)$ of EDGE SUBSET FVS, returns an equivalent instance $(G', S', k')$ with $k' \leq k$ and $|S'| \in \mathcal{O}(k^4)$. Together with the randomized kernelization from the first part this implies the claimed randomized kernelization to $\mathcal{O}(k^9)$ vertices.

A reduction of the number of $S$-edges is also a crucial ingredient in the FPT algorithm for EDGE SUBSET FVS by Cygan et al. [3]. They achieve $|S| \in \mathcal{O}(k^3)$, but it is in a slightly more favorable setting: Using iterative compression, it suffices to solve the task of finding a solution $X'$ of size $k$ when given a solution $X$ of size $k+1$. (This is well known in parameterized complexity, and we prefer not to repeat it here.) Considering some unknown solution $X'$ of size $k$, one can guess the intersection $D$ of $X'$ with $X$, by trying all $\mathcal{O}(2^{k+1})$ possibilities. For the correct guess $D = X' \cap X$, the remaining problem is to find for $(G - D, S \setminus D, k - |D|)$ a solution $Z'$ of size at most $k - |D|$ that is disjoint from $Z = X \setminus D$, since $Z' = X' \setminus D$ would be such a solution; here $S \setminus D$ denotes the set of edges in $S$ with no endpoint in $D$. Cygan et al. make the nice observation that the guessing also allows to assume that there is no other solution $X'$ with an even larger intersection with $X$.

In contrast, we cannot afford to run iterative compression for a kernelization to get a starting solution of size $k+1$ and, as is common, we have to start with an approximate solution $Z$, which can be assumed to be of size at most $8k$ using an 8-approximation algorithm of Even et al. [6]. The idea of guessing the intersection of an optimal solution with $Z$ is infeasible regarding both time and the number of created instances. Thus, while several structures like $z$-flowers or disjoint $x$,$y$-paths containing $S$-edges appear in both approaches, many things have to be handled differently. For example, having $k+2$ disjoint $x$,$y$-paths containing $S$-edges for $x, y \in Z$ implies that one of $x$ and $y$ must be in every solution of size $k$; Cygan et al. can stop here because the solution would not be disjoint from $Z$; we need to instead store the information about $x$ and $y$ to later detect $S$-edges that can be safely removed. Like Cygan et al., we also use Gallai's $A$-path Theorem but we avoid the 2-expansion lemma by using the properties of a blocking set of size at most $2k$ differently. Moreover, we observe that $z$-flowers can be found by solving a matroid parity instance on an appropriate gammoid; this can be done in deterministic polynomial time using a specialized matroid parity algorithm by Tong et al. [20].

Proofs omitted in this extended abstract can be found in Hols and Kratsch [9].

## 2 Preliminaries

We use standard graph notation, mostly following Diestel [5]. All graphs are undirected and may contain multi-edges and loops; accordingly, they may contain cycles of length one and two (formed by loops and multi-edges, respectively.) An edge $e \in E$ is called a *bridge* if $(V, E \setminus \{e\})$ has more connected components than $G$. For a set $X \subseteq V$, let $G[X]$ denote the subgraph of $G$ induced by $X$ and let $N_G(X)$ denote the neighborhood of $X$ in $G$, i.e., $N_G(X) = \{v \in V \setminus X \mid \exists u \in X \colon \{u, v\} \in E\}$. Given two sets $X, Y \subseteq V$, by $E(X, Y)$ we denote the set of edges that have one endpoint in $X$ and one endpoint in $Y$. For a set $E' \subseteq E$ of edges let $V(E')$ be the set of vertices that are incident with at least one edge in $E'$. For $X \subseteq V$ and $F \subseteq E$ we shorthand $G - X$ for $G[V \setminus X]$ and $G - F$ for $(V(G), E(G) \setminus F)$; if $X = \{x\}$ then we may also write $G - x$ instead of $G - \{x\}$. Note that the graph $(G - X) - F$ is the same graph as the graph $(G - F) - X$ and we will drop the parentheses.

For $A \subseteq V$ a path with endpoints in $A$ and internal vertices not in $A$ is called an *$A$-path*. The following theorem about $A$-paths was already used by Cygan et al. [3] for SUBSET FVS and in the quadratic kernelization for FEEDBACK VERTEX SET by Thomassé [19].

▶ **Theorem 1** (Gallai [8]). *Let $A \subseteq V$ and $k \in \mathbb{N}$. If the maximum number of vertex-disjoint $A$-paths is strictly less than $k+1$, then there exists a set $B \subseteq V$ of at most $2k$ vertices that intersect every $A$-path.*

In particular it is possible to find either $(k+1)$-disjoint $A$-paths or a set $B$ that intersects all $A$-paths in polynomial time. This follows from Schrijver's proof of Gallai's theorem [18].

Let $(G, S, k)$ be an instance of the EDGE SUBSET FVS problem. We call a cycle $C$ an *$S$-cycle*, if at least one edge of $S$ is contained in $C$. Let $x$ be a vertex of $V$. A set $\{C_1, C_2, \ldots, C_t\}$ of $S$-cycles that contain $x$ is called an *$x$-flower* of order $t$, if the sets of vertices $C_i \setminus \{x\}$ are pairwise disjoint. Note that if there exists a $x$-flower of order at least $k+1$, then the vertex $x$ must be in every solution for $(G, S, k)$, if one exists. A set $B \subseteq V \setminus \{x\}$ of size $t$ is called an *$x$-blocker* of size $t$, if each $S$-cycle through $x$ also contains at least one vertex of $B$.

**Matroids, gammoids, and representative sets.** A matroid $M = (U, \mathcal{I})$ consists of a finite set $U$ and a family $\mathcal{I}$ of subsets of $U$, called *independent sets*, fulfilling the following properties: (i) $\emptyset \in \mathcal{I}$; (ii) if $X \subseteq Y$ and $Y \in \mathcal{I}$ then also $X \in \mathcal{I}$; and (iii) if $X, Y \in \mathcal{I}$ with $|X| < |Y|$ then there exists $y \in Y \setminus X$ such that $X \cup \{y\} \in \mathcal{I}$.

The *rank* of of a matroid $M$, denoted by $r(M)$, is the size of the largest independent set of the matroid $M$.

Let $A$ be a matrix over an arbitrary field $F$. Let $U$ be the set of columns of $A$ and let $\mathcal{I}$ be the family of all sets $X \subseteq U$ of columns that are linearly independent over $F$. Then $M = (U, \mathcal{I})$ is a matroid, called the *linear matroid* or *vector matroid* of $A$, and we say that $A$ *represents* $M$. If $M = (U, \mathcal{I})$ is representable over some field, then it is also representable by an $r(M) \times |U|$ matrix; by Gaussian elimination we can always reduce a representing matrix for $M$ to one with $r(M)$ many rows (cf. [15]). Let $M_1 = (U_1, \mathcal{I}_1)$ and $M_2 = (U_2, \mathcal{I}_2)$ be two matroids with $U_1 \cap U_2 = \emptyset$. The *direct sum* $M_1 \oplus M_2$ is a matroid over $U = U_1 \cup U_2$ with independent sets $\mathcal{I} = \{X \subseteq U \mid X \cap U_1 \in \mathcal{I}_1, X \cap U_2 \in \mathcal{I}_2\}$. If $A_1$ and $A_2$ represent the two matroids over the same field $F$, then matrix $A = \operatorname{diag}(A_1, A_2)$ represents $M_1 \oplus M_2$.

Let $G = (V, E)$ be a graph that may have both directed and undirected edges and let $S \subseteq V$. A set $T \subseteq V$ is *linked* to $S$ if there exist $|T|$ vertex-disjoint paths from $S$ to $T$. Thus every vertex in $T$ is endpoint of a different path from $S$. It holds that $M = (U, \mathcal{I})$, where $U \subseteq V$ and $\mathcal{I}$ contains all sets $T \subseteq U$ that are linked to $S$ in $G$, is a matroid [17]. The matroid $M$ is also called the *gammoid* on $G$ with sources $S$ and ground set $U$; if $U = V$ then $M$ is also called a *strict gammoid.* Marx [15] gave a randomized polynomial-time procedure for finding a matrix representation of a strict gammoid. The error probability can be made exponentially small in the size of the graph. (This is the only source of randomness and error in our kernelization.) A matrix representation for a gammoid for graph $G = (V, E)$ with ground set $U \subsetneq V$ and sources $S$ can be obtained from one for the strict gammoid for $G$ and $S$ by simply deleting columns corresponding to elements of $V \setminus U$.

Let $A, B$ be independent sets in a matroid. We say that $A$ *extends* $B$ if $A \cap B = \emptyset$ and $A \cup B$ is again an independent set. Note that from the independence of $A \cup B$ follows the independence of $A$ and $B$ due to the second matroid property.

▶ **Definition 2.** Let $M = (U, \mathcal{I})$ be a matroid, let $\mathcal{A} \subseteq \mathcal{I}$, and let $q \in \mathbb{N}$. A set $\mathcal{A}' \subseteq \mathcal{A}$ is *$q$-representative* for $\mathcal{A}$ if for every independent set $B$ of size at most $q$ there is a set $A \in \mathcal{A}$ that extends $B$ if and only if there is also a set $A' \in \mathcal{A}'$ that extends $B$.

Observe that if $\mathcal{A}'$ is $q$-representative for $\mathcal{A}$ and there exists a set $A \in \mathcal{A}$ that *uniquely extends* some given independent set $I$ of size at most $q$, then this implies that $A \in \mathcal{A}'$.

The following theorem of Lovász [14] proves that for any linear matroid there exist small representative sets. It was made algorithmic by Marx [15] and, thus, permits to find representative sets in polynomial time when given a matrix representation of the matroid. A faster algorithm for this task was developed recently by Fomin et al. [7].

▶ **Lemma 3** (Lovász [14], Marx [15]). *Let $M$ be a linear matroid of rank $q+p$, and let $\mathcal{T} = \{I_1, I_2, \ldots, I_t\}$ be a collection of independent sets, each of size $p$. If $|\mathcal{T}| > \binom{q+p}{p}$, then there is a set $I \in \mathcal{T}$ such that $\mathcal{T} \setminus \{I\}$ is $q$-representative for $\mathcal{T}$. Furthermore, given a representation $A$ of $M$, we can find such a set $I$ in $f(q,p) \cdot (\|A\|t)^{\mathcal{O}(1)}$ time.*

## 3 Randomized polynomial kernelization for parameter $|S| + k$

In this section we present a randomized polynomial kernelization for EDGE SUBSET FVS parameterized by $|S| + k$. Because deletion of one endpoint of each edge in $S$ always constitutes a feasible solution, nontrivial instances have $|S| > k$. Thus, our kernelization also works for parameter $|S|$ alone. However, to achieve a better bound for EDGE SUBSET FVS parameterized by $k$ it is beneficial to give the size in terms of $|S|$ and $k$ rather than $|S|$ alone.

We use representative sets of independent sets of matroids to obtain a kernel of size $\mathcal{O}(|S|^2k)$. Our approach is similar to the kernelization of DELETABLE TERMINAL MULTIWAY CUT($k$) [12]. As in that paper we construct path packings such that certain vertices can be shown to be in a representative set. Note that, unlike for multiway cut-type problems, a solution $X \subseteq V$ will not necessarily create many connected components. Rather, as used also in the FPT algorithm of Cygan et al. [3], it creates a particular tree-like structure in $G - X$. Nevertheless, endpoints of edges in S, denoted $T := V(S)$, will play the role of terminals that need to be separated in a certain way; hence a vertex $x$ in $T$ is called a *terminal*. We will focus on the graph $G - S$, i.e., with edges of $S$ deleted, in which a solution $X$ creates a grouping of (not deleted) terminals into connected components. The structure of these components will be crucial for a replacement argument (Lemma 5) that leads to the required path packing; this constitutes one of the key arguments for our result.

The kernelization consists of four steps. In the first step we show that if an instance is YES then there exists a solution $X$ with a certain path packing from $T$ to $X$. Then we define an appropriate gammoid to find in a next step a representative set of size $\mathcal{O}(|S|^2k)$ which is (essentially) a superset of $X$ using Lemma 3. Finally we explain how to reduce the graph $G$, using the superset of the last step, to obtain an equivalent instance of EDGE SUBSET FVS.

**Analyzing solutions.** Let $(G, S, k)$ be a yes-instance of EDGE SUBSET FVS $(k+|S|)$. We say that a solution $X$ for $(G, S, k)$ is *dominant*, if it has minimum size and contains a maximal number of vertices from $T$ among solutions of minimum size. The vertices in $X \cap T$ correspond to endpoints of edges in $S$ that we delete and the vertices in $X_0 = X \setminus T$ block all $x$-$y$ paths with $\{x, y\} \in S_0 = \{e \in S \mid e \cap X = \emptyset\}$, except the one that consists of the edge $\{x, y\}$. We show that $X$ is linked to $T$ in a strong sense.

▶ **Lemma 4.** *Let $X$ be a dominant solution for $(G, S, k)$ and $x$ any vertex in the set $X_0 = X \setminus T$. There exist $|X| + 2$ paths from $T$ to $X$ in $G - S$ that are vertex-disjoint except for three paths ending in vertex $x$. Moreover, the paths can be chosen in such a way that each connected component of $G - X - S$ is intersected by at most one path.*

We use Hall's Theorem and the lemma below to prove this. For this purpose we consider two graphs $G - X$ and $G - X - S$. We call a connected component $K$ of $G - X - S$ *interesting* if it contains a terminal, i.e., if $T \cap V(K) = (T \setminus X) \cap V(K) \neq \emptyset$, and we say that $x \in X_0$ *sees* a connected component $K$ if $x$ is adjacent to a vertex of $K$ in $G$. We extend this definition by saying that $Y \subseteq X_0$ sees an interesting component $K$ if at least one vertex $y \in Y$ sees $K$.

▶ **Lemma 5.** *If $X$ is a dominant solution then every nonempty set $Y \subseteq X_0$ sees at least $|Y| + 2$ interesting components of $G - X - S$.*

**Proof.** Assume for contradiction that there exists a nonempty set $Y \subseteq X_0$ that sees at most $|Y|+1$ interesting components of $G-X-S$. Let $\mathcal{C}$ denote the set of connected components of $G-X$, let $\mathcal{C}_i \subseteq \mathcal{C}$ denote the set of interesting components seen by $Y$, and let $\mathcal{C}_o \subseteq \mathcal{C}$ denote the other components seen by $Y$. We will show that there is an alternative solution $X' = (X \setminus Y) \cup Y'$ that is smaller than $X$ or that contains more vertices of $T$, contradicting the choice of $X$ as a dominant solution. Let us consider the graphs $G-X$ and $G-(X \setminus Y)$.

We study the structure of $G-(X \setminus Y)$ to find a set $Y'$ that intersects all $S$-cycles in $G-(X \setminus Y)$. Accordingly we are interested in the structure that is induced by the $S$-edges in $G-(X \setminus Y)$. To study them we define, for any subgraph $G-Z$ of $G$, the $S$-component graph $H_Z$ which has a vertex for each connected component of $G-Z-S$ and for every $S$-edge $e$ an edge between two (not necessary different) vertices which correspond to the connected components that contain the endpoints of $e$; note that $H_Z$ can have parallel edges and loops. We say that $G-Z$ is an *$S$-forest* if the $S$-component graph $H_Z$ is a forest. Observe that a set $Z$ is a solution if and only if $G-Z$ is an $S$-forest. Note that vertices that correspond to components without terminals in $G-Z$ are isolated in $H_Z$ because they are not incident with $S$-edges; e.g., this is true in $H_X$ for non-interesting components of $G-X-S$.

To construct an alternative solution $X'$ we compare the $S$-component graphs of $G-X$ and $G-(X \setminus Y)$; let $\mathcal{C}'$ denote the set of connected components of $G-(X \setminus Y)-S$. A component in $\mathcal{C}'$ either fully contains some components in $\mathcal{C}_i \cup \mathcal{C}_o$ and additionally it may contain vertices of $Y$ or it is equal to a connected component in $\mathcal{C}$: This follows from the fact that we only delete the subset $X \setminus Y$ of $X$ from $G-S$ instead of $X$. However, in $G-(X \setminus Y)$ the set of $S$-edges incident with components in $\mathcal{C}'$ is the same as the set of $S$-edges incident with $\mathcal{C}$ in $G-X$, because $Y \subseteq X_0 = X \setminus T$ and hence there are no additional vertices of $T$, i.e., $T \setminus X = T \setminus (X \setminus Y)$. Altogether, $H_{X \setminus Y}$ is obtained from $H_X$ by merging vertices in $H_X$ whose corresponding connected components are connected in $G-(X \setminus Y)-S$. In general, $G-(X \setminus Y)$ will not be an $S$-forest: The merging of vertices may lead to loops (from $S$-edges with both ends in the same component) and longer cycles in $H_{X \setminus Y}$.

We will see that deleting at most $|Y|$ edges of $S$, i.e., deleting a set $Y'$ of at most $|Y|$ endpoints of $S$-edges, will suffice for $G-((X \setminus Y) \cup Y')$ to be an $S$-forest, making $(X \setminus Y) \cup Y'$ a valid solution. To see this consider an arbitrary connected component $C^+$ in $G-(X \setminus Y)$ whose corresponding connected component in $H_{X \setminus Y}$ is not cycle-free. Note that $C^+$ is a union of connected components in $\mathcal{C}'$ that are connected by $S$-edges. Therefore $C^+$ must contain connected components in $\mathcal{C}$ that are seen by $Y$. Let $C_i^1, \dots C_i^a \in \mathcal{C}_i$ and $C_o^1, \dots C_o^b \in \mathcal{C}_o$ be the connected components in $\mathcal{C}$ that are contained in $C^+$ and that are seen by $Y$.

In $G-X$ the connected component $C^+$ may decompose into several separate connected components because we additionally delete the vertices of $Y$. Since $Y$ sees only components in $\mathcal{C}_i \cup \mathcal{C}_o$ the set $C^+$ decomposes into at most $a+b$ separate components by deleting $Y$. Recall that components in $\mathcal{C}_o$ are isolated in $G-X$ and contain no vertices of $T$ and, thus, they do not contribute any $S$-edges to $C^+$. It remains to consider the components $C_i^1, \dots, C_i^a$ that are contained in $C^+$.

The connected components $C_i^1, \dots, C_i^a$ are part of at least one connected component in $G-X$. Thus, they correspond to a subforest $F$ of $H_X$ and not deleting $Y$ corresponds to merging $a$ vertices in this forest into $d \geq 1$ new vertices; let $F'$ be the connected subgraph in $H_{X \setminus Y}$ that results from $F$ by this operation. If the subforest $F$ consists of $c$ vertices and, thus, at most $c-1$ $S$-edges then we obtain $c-a+d$ vertices that are connected by at most $c-1$ edges for $F'$. It therefore suffices to delete at most $(c-1)-((c-a+d)-1) = a-d \leq a-1$ $S$-edges, i.e., to delete one endpoint of each of at most $a-1$ $S$-edges, to obtain a forest-structure in $F'$. (We cannot delete just *any* $a-1$ edges but we can keep any $c-a+d-1$ $S$-edges spanning the $c - a + d$ components and delete the at most $a - 1$ remaining $S$-edges.)

Overall, we get that a connected component $C^+$ in $G - (X \setminus Y)$ that fully contains $a$ interesting components from $\mathcal{C}_i$ requires at most $a-1$ vertex deletions of endpoints of $S$-edges to obtain an $S$-forest. Since $Y$ sees at most $|Y| + 1$ such components, the worst case is achieved by a single component $C^+$ containing all $|Y| + 1$ interesting components in $\mathcal{C}_i$; this still costs at most $(|Y| + 1) - 1 = |Y|$ vertex deletions, as claimed.

Let $Y'$ contain all the endpoints of $S$-edges that we delete to get an $S$-forest. We know that $|Y'| \leq |Y|$ and thus $|(X \setminus Y) \cup Y'| \leq |X|$. Moreover, by the initial considerations, we know that $X' = (X \setminus Y) \cup Y'$ is a feasible solution as $G - X'$ has the required $S$-forest. If $|Y'| < |Y|$, including the case that $Y' = \emptyset$, then $|X'| < |X|$ as $Y \neq \emptyset$; this contradicts optimality of $X$ (required for being a dominant solution). If $|Y'| = |Y|$ then $Y' \neq \emptyset$ and $X'$ is an optimal solution that contains more vertices of $T \supseteq Y'$, contradicting the choice of $X$ as a dominant solution. Thus, every nonempty set $Y$ must see at least $|Y| + 2$ connected components, as claimed. ◀

Lemma 4 can now be obtained via Hall's Theorem; the proof is similar to the one for DELETABLE TERMINAL MULTIWAY CUT [12].

**Setting up the gammoid.** The gammoid $M$ that we use is the direct sum of two gammoids $M_1$ and $M_2$. To construct gammoid $M_1$ we define a graph $G_1 = (V_1, E_1)$ that is obtained from $G - S$ by adding two so called *sink-only copies* $v'$ and $v''$ for every vertex $v \in V$. A sink-only copy of a vertex $v$ is a vertex $v'$ (or $v''$) that has a directed edge $(u, v')$ for each edge $\{u, v\}$; these were already used in previous work [12]. Note that adding sink-only copies of vertices does not affect the possible path packings to other vertices since they can only be endpoints of paths; however, they are convenient to capture multiple vertex-disjoint paths that, intuitively, end in the same vertex. Matroid $M_1$ is defined as the gammoid on $G_1$ with sources $T = V(S)$ and ground set $V_1 = \{v, v', v'' \mid v \in V\}$; note that the sink-only copies of vertices in $T$ are not sources of $M_1$. The rank of matroid $M_1$ is $|T|$, because the set of all trivial paths is independent and at most $|T|$ vertices can be linked to $T$.

Matroid $M_2$ is the gammoid on the directed graph $G_2 = K_{k,n} = (S_2 \dot{\cup} \hat{V}, E_2)$ with sources $S_2$ and ground set $\hat{V} = \{\hat{v} \mid v \in V\}$; the edges in $E_2$ are directed from $S_2$ to $\hat{V}$. In other words, gammoid $M_2$ is a uniform matroid and a (deterministic) matrix representation could also be obtained by using a Vandermonde matrix. The rank of $M_2$ is $k = |S_2|$ as no more than $|S_2|$ vertices can be linked to $S_2$ and every set of at most $k$ vertices of $\hat{V}$ is linked to $S_2$.

For the application of Lemma 3 we will use the matroid $M = M_1 \oplus M_2$, which has rank $|T| + k$. Representations $A_1$ and $A_2$ for both $M_1$ and $M_2$ can be computed by a randomized polynomial-time algorithm with exponentially small error chance [15]; hence we get a representation for $M$ by $\operatorname{diag}(A_1, A_2)$, i.e., the block-diagonal matrix with blocks $A_1$ and $A_2$. We may assume that $A_1$ has $|T|$ rows and $A_2$ has $k$ rows (cf. [15]).

**Applying the representative set lemma.** Let $\mathcal{T} := \{\{v', v'', \hat{v}\} \mid v \in V\}$. For clarity, by the above notation, this means that $v', v'' \in V_1$ and $\hat{v} \in \hat{V}$ for each $v \in V$. Let $\mathcal{T}'$ be obtained by applying Lemma 3 to $\mathcal{T}$ using the above matrix representation for $M$; we have $|\mathcal{T}'| \in \mathcal{O}((|T|+k)^3) = \mathcal{O}(|S|^3)$. We will see later that we can find a $(|T|+k-3)$-representative set of size $\mathcal{O}(|S|^2 k)$ by a careful look at the proof of Lemma 3, using the fact that $M$ is the direct sum of two gammoids and that all sets $\{v', v'', \hat{v}\}$ in $\mathcal{T}$ have two elements from the first and one element from the second gammoid; a similar argument for getting a smaller representative set was already used by Kratsch and Wahlström [12].

We will prove that for each dominant solution $X$ we have $\{x', x'', \hat{x}\} \in \mathcal{T}'$ for each $x \in X_0 = X \setminus T$. To this end, we show that for each such set $\{x', x'', \hat{x}\}$ there exists an independent set $I$ of size at most $|T| + k - 3$ such that $\{x', x'', \hat{x}\}$ uniquely extends $I$ among triplets in $\mathcal{T}$. Thus, $\{x', x'', \hat{x}\}$ must be in every $(|T| + k - 3)$-representative set $\mathcal{T}'$ of $\mathcal{T}$.

▶ **Lemma 6.** *Let $X$ be a dominant solution for $(G, S, k)$ and let $T = V(S)$. For all $x \in X_0 = X \setminus T$ there exists an independent set $I$ of size at most $|T| + k - 3$ in $M$ such that $\{x', x'', \hat{x}\}$ uniquely extends $I$.*

We know now that for every vertex $x \in V \setminus T$ that is a vertex in a dominant solution the set $\{x', x'', \hat{x}\}$ is in every $(|T| + k - 3)$-representative set $\mathcal{T}'$. If we define $V(\mathcal{T}') = \{v \mid \{v', v'', \hat{v}\} \in \mathcal{T}'\}$ then this implies that $X_0 \subseteq V(\mathcal{T}')$ for each dominant solution $X$. Thus, every dominant solution $X$ is contained in $V(\mathcal{T}') \cup T$.

*Shrinking the input graph to $\boldsymbol{\mathcal{O}(|V(\mathcal{T}') \cup T|)}$ vertices.* In the previous parts we have shown that if there exists a solution for $(G, S, k)$, then there exists a solution that is completely contained in $W := V(\mathcal{T}') \cup T$. Using this we can make all vertices in $V \setminus W$ undeletable. We achieve this by applying the so-called torso operation to vertex set $W$ in $G$; let $G' = \text{torso}(G, W)$. By definition of $\text{torso}(G, W)$, the resulting graph $G'$ has vertex set $W$ and is derived from $G[W]$ by making each pair $\{u, v\} \subseteq W$ adjacent if there is a $u$,$v$-path in $G$ with internal vertices from $V \setminus W$. Note that we do not create double edges or loops in $G'$ and that all edges of $S$ are preserved in $G'$ because $T \subseteq W$. (The same can be achieved by iteratively selecting a vertex $v \in V \setminus W$, making its neighbors a clique, and deleting $v$.)

▶ **Lemma 7.** *$(G', S, k)$ has a solution if and only if $(G, S, k)$ has a solution.*

It follows from Lemma 7 that $(G', S, k)$ is an equivalent instance and the graph of this instance contains at most $|W|$ vertices. The correctness of Lemma 7 follows from the fact that the torso operation preserves the separators that are contained in $W$ (cf. [16]).

So far we have a kernelization that creates an equivalent instance $(G', S, k)$ such that $G'$ has $|W|$ vertices. As mentioned above, Lemma 3 guarantees that $|W| \in \mathcal{O}(|S|^3)$ and this implies a polynomial kernel for EDGE SUBSET FVS parameterized by $|S|$. If we use the fact that the gammoid $M$ is the direct sum of two gammoids $M_1$ and $M_2$, and that all sets $\{v', v'', \hat{v}\} \in \mathcal{T}$ contain exactly two elements of $M_1$ and one element of $M_2$, then we can prove that $|W| \in \mathcal{O}(|S|^2 k)$, this is an improvement for all nontrivial instances with $k < |S|$.

▶ **Lemma 8.** *Let $M = M_1 \oplus M_2$ be the gammoid of rank $|T| + k$ as defined above and $\mathcal{T} = \{I_1, I_2, \ldots, I_t\}$ be the set of independent sets of $M$ that we use for the kernelization. Let $A$ be represented by $\text{diag}(A_1, A_2)$ as above. If $|\mathcal{T}| > \binom{|T|}{2} \cdot \binom{k}{1}$, then there exists a set $I \in \mathcal{T}$ such that $\mathcal{T} \setminus \{I\}$ is $(|T| + k - 3)$-representative for $\mathcal{T}$.*

The proof of Lemma 8 is similar to Marx [15, Lemma 4.2]. We additionally use the fact that $M$ is the direct sum of two gammoids to get that the vectors in the exterior algebra which represent the sets in $\mathcal{T}$ span a space of smaller dimension. As mentioned above, Marx [15] showed that one can find in randomized polynomial-time a matrix with $r(M)$ rows that represents a given gammoid $M$. We can make this proof algorithmic in the same way. Combined with Lemma 8 it follows that we can find a $(|T| + k - 3)$-representative subset $\mathcal{T}'$ of $|\mathcal{T}|$ of size at most $\binom{|T|}{2} \cdot \binom{k}{1} \in \mathcal{O}(|S|^2 k)$. This implies a randomized polynomial kernel with $\mathcal{O}(|S|^2 k)$ vertices for EDGE SUBSET FVS parameterized by $|S|$ and $k$.

## 4 Reducing the size of $S$

We have seen that EDGE SUBSET FVS parameterized by $|S|$ and $k$ has a polynomial kernel. Now the goal is to reduce the size of the set $S$ until $|S|$ is polynomially bounded in $k$. This will lead to a polynomial kernel of EDGE SUBSET FVS parameterized by $k$.

To begin, we do some initial modifications to ensure that we can always find a solution of size at most $k$ that contains no vertex of the set $V(S)$, if one exists. For this we first delete all vertices $v \in V$ with the property that $e = \{v, v\} \in S$ is a loop in $G$ and we decrease the value $k$ by one. Next we delete all remaining loops. We also reduce the number of edges between two vertices $v, w \in V(G)$. If no edge that is incident to $v$ and $w$ is contained in the set $S$, then we delete all except one edge. On the other hand, if at least one edge between $v$ and $w$ is contained in $S$, then we delete all except two edges. One of these edges is contained in $S$ and the other not. In the next step we add for every edge $e = \{v, w\} \in S$ two new vertices $v_e, u_e$ to the graph, subdivide the edge $e$ into three edges $\{v, v_e\}, \{v_e, w_e\}, \{w_e, w\}$, and edit $S$ by replacing edge $e$ by the edge $\{v_e, w_e\}$ in $S$. If a solution $X$ of EDGE SUBSET FVS contains a vertex $x_e \in V(S)$, then we can instead add the vertex $x$ to $X$ and delete $x_e$ from $X$, because every cycle that contains vertex $x_e$ also contains vertex $x$; hence we can always find an optimal solution that is disjoint from $V(S)$.

Let $(G, S, k)$ be an instance of EDGE SUBSET FVS, such that $G$ is a graph with the above properties. Analogous to the paper of Cygan et al. [3] we consider a solution $Z$ of the EDGE SUBSET FVS, with the difference that our solution is an 8-approximation of the problem, to reduce the size of $S$. Even et al. [6] show that there exists an 8-approximation algorithm for SUBSET FVS. Since SUBSET FVS and EDGE SUBSET FVS are equivalent (cf. [3]), we can compute in polynomial time an 8-approximation for EDGE SUBSET FVS and we can assume that $Z \cap V(S) = \emptyset$. If $|Z| > 8k$, then we can stop immediately because no solution of size at most $k$ can exist. On the other hand, if $|Z| \leq k$, then $Z$ is a solution and we are done.

The set $Z$ is a feasible solution to EDGE SUBSET FVS on $(G, S, |Z|)$. This implies that every edge $e \in S$ is a bridge in $G - Z$. In a next step we also remove all edges in $S$ from $G - Z$. Every connected component in $G - Z - S$ contains no edge from $S$ and, following Cygan et al. [3], we call such a component a *bubble*. We denote the set of bubbles by $\mathcal{D}_Z$ and define a graph $H_Z = (\mathcal{D}_Z, E_{\mathcal{D}_Z})$ whose vertices are bubbles and with bubbles $I$ and $J$ being adjacent, i.e., $\{I, J\} \in E_{\mathcal{D}_Z}$, if and only if the components $I$ and $J$ are connected by an edge from $S$. The graph $H_Z$ is a forest, because $Z$ is a solution for $(G, S, |Z|)$ and a cycle in $H_Z$ would give rise to an $S$-cycle in $G - Z$. Similarly, no two bubbles can be connected by more than one edge of $S$. By $V_I$ we denote the vertices that are contained in bubble $I$. Since $|E(V_I, V_J) \cap S| \leq 1$ for all $I, J \in \mathcal{D}_Z$ and equality holds if and only if $\{I, J\} \in E_{\mathcal{D}_Z}$, we can associate an edge $e = \{I, J\} \in E_{\mathcal{D}_Z}$ with the one edge $e_S = \{v_I, v_J\}$ in $E(V_I, V_J) \cap S$. If we add the vertex set $Z$ and all edges $\{z, I\}$ with the property that $z \in Z, I \in \mathcal{D}_Z$ and $E(z, V_I) \neq \emptyset$ to the graph $H_Z$ we obtain a graph $H_Z^+$ that contains $S$-cycles. Note that every $S$-cycle must contain a vertex of the set $Z$. We partition the set of bubbles according to the number of bubbles they are connected with.

▶ **Definition 9.** A bubble $I \in \mathcal{D}_Z$ is called (i) solitary, if $\deg_{H_Z}(I) = 0$; (ii) leaf, if $\deg_{H_Z}(I) = 1$; and (iii) inner, if $\deg_{H_Z}(I) \geq 2$. By $\mathcal{D}_Z^s, \mathcal{D}_Z^l, \mathcal{D}_Z^i$ we denote the corresponding sets of of bubbles.

Let $X \subseteq V \setminus V(S)$ be a superset of $Z$. We define $H_X$, $H_X^+, \mathcal{D}_X$ and $E_{\mathcal{D}_X}$ analogously to $H_Z, H_Z^+, \mathcal{D}_Z$ and $E_{\mathcal{D}_Z}$. Observe that the number of edges in $S$ is at most $|\mathcal{D}_Z \setminus \mathcal{D}_Z^s|$, because $H_Z$ is a forest, any two bubbles are connected by at most one $S$-edge, and $V(S) \cap Z = \emptyset$.

So far our setup is essentially the same as the one used by Cygan et al. [3]. However, instead of an 8-approximate solution they use the framework of iterative compression, which provides a solution $Z$ of size $k + 1$ and leaves them with the task of reducing the number of $S$-edges for the problem of finding a solution $Z^*$ that is *disjoint* from $Z$. Moreover, it suffices for them to consider the case that every feasible solution (if one exists) is disjoint from $Z$. In this setting they are able to reduce to an equivalent instance (or find that some assumption was violated) with only $\mathcal{O}(k^3)$ edges in $S$. Thus, while many relevant structures like $z$-flowers or parallel $x$-$y$ paths containing $S$-edges are the same, many things have to be handled differently. In particular, if we find that at least one out of two vertices $x, y \in Z$ must be in the solution then we cannot stop (using the maximality condition) but need to continue and use this information in a more direct way.

During the reduction we detect certain pairs $\{x, y\}$ of different vertices with the property that each solution of size at most $k$ must contain at least one of the vertices (if one exists). We store this fact as a *pair-constraint*. We keep and enforce this information in the final instance, unless we decide earlier to delete $x$ or $y$. By $\mathcal{P}$ we denote the set of pair-constraints that we have found so far. We can interpret this set as a set of edges and by $V(\mathcal{P})$ we denote all vertices that are contained in a pair-constraint. Note that vertices from the set $V(S)$ are never contained in a pair-constraint from $\mathcal{P}$, because there always exists a solution that is disjoint from $V(S)$. We need the set $\mathcal{P}$ to detect edges in $S$ that may be safely deleted. To this end, we generalize the EDGE SUBSET FVS problem by adding a set of pair-constraints $\mathcal{P}$ to the input; we call this problem PAIR-CONSTRAINED EDGE SUBSET FVS.

PAIR-CONSTRANED EDGE SUBSET FEEDBACK VERTEX SET **Parameter:** $k$
**Input:** An undirected graph $G$, a set $S \subseteq E$ of edges, a set $\mathcal{P}$ of pair-constraints and an integer $k$.
**Question:** Does there exist a set $X \subseteq V$ of size at most $k$ such that $G - X$ contains no $S$-cycle and such that for each pair-constraint $\{x, y\} \in \mathcal{P}$ we have $x \in X$ or $y \in X$?

Clearly, instances $(G, S, k)$ of EDGE SUBSET FVS and $(G, S, \emptyset, k)$ of PAIR-CONSTRAINED EDGE SUBSET FVS are equivalent. Our goal is to reduce the size of $S$ by detecting $S$-edges that we can delete from $S$ without changing the outcome. This leads to the following definition:

▶ **Definition 10.** Let $(G, S, \mathcal{P}, k)$ be an instance of PAIR-CONSTRANED EDGE SUBSET FVS. We call an edge $e \in S$ *irrelevant*, if $X \subseteq V(G)$ is a solution for $(G, S, \mathcal{P}, k)$ if and only if $X$ is a solution for $(G, S \setminus \{e\}, \mathcal{P}, k)$.

Note that if two different $S$-edges $e$ and $e'$ are irrelevant in $(G, S, \mathcal{P}, k)$, then $e'$ is not necessarily irrelevant in $(G, S \setminus \{e\}, \mathcal{P}, k)$. Also, we do not expect to find all irrelevant edges.

*The reduction rules.* We now present our reduction rules; we assume that always the lowest numbered applicable rule is applied first. Correctness and efficiency of the overall reduction process are deferred to the full version. Let $(G, S, \mathcal{P} = \emptyset, k)$ be an instance for PAIR-CONSTRAINED EDGE SUBSET FVS and let $Z$ be an 8-approximation of this problem with $k < |Z| \leq 8k$ that is disjoint from $V(S)$. In the following the graphs $G - Z$, $G - Z - S$, $H_Z$, and $H_Z^+$ are always defined with respect to the current instance $(G, S, \mathcal{P}, k)$ of PAIR-CONSTRAINED EDGE SUBSET FVS. Note that $Z \subseteq V$ and we delete a vertex from $Z$ if we delete the corresponding vertex in $V$.

**Rule 1:** If $k < 0$, or if $k = 0$ and there exists an $S$-cycle, then reduce $(G, S, \mathcal{P}, k)$ to some trivial false instance, i.e. $G' := (\{x\}, \{e = \{x, x\}\})$, $S' := \{e\}$, $\mathcal{P}' = \emptyset$ and $k' := 0$.

**Rule 2:** Delete all bridges and all connected components not containing any edge from $S$.

**Rule 3:** If edge $e \in S$ is a bridge in $(V, E \setminus (S \setminus \{e\}))$, then reduce to $S' = S \setminus \{e\}$.

**Rule 4:** If vertex $v \in V(\mathcal{P})$ is contained in at least $k + 1$ pair-constraints of $\mathcal{P}$, then we reduce to $G' = G - v$ and $k' = k - 1$.

**Rule 5:** If $|\mathcal{P}| > k^2$, then reduce $(G, S, \mathcal{P}, k)$ to some trivial false instance.

**Rule 6:** If there exists a $z$-flower of order $k + 1$ in $G$ for a vertex $z \in Z$, then we reduce to $G' := G - z$ and $k' := k - 1$.

Rules 2 and 3 ensure that each bubble $I \in \mathcal{D}_Z$ is adjacent to a vertex in $Z$ in graph $H_Z^+$ and Rules 4 and 5 make sure that $\mathcal{P}$ remains small. For the next rules we need a maximal matching $M$ in $H_Z$ that covers all inner bubbles $\mathcal{D}_Z^i$ in $H_Z$. Note that two adjacent leaf bubbles $I_1, I_2$ form a $K_2$ in $H_Z$, hence the edge $\{I_1, I_2\} \in E_{\mathcal{D}_Z}$ is contained in every maximal matching in $H_Z$. We use this matching to detect pair-constraints in $Z$. To this end we introduce the following definition: Let $e = \{I, J\}$ be an edge in the matching $M$. We say $e$ *sees* the pair $\{x, y\}$ of different vertices $x, y \in Z$ respectively the vertex $x \in Z$, if $\{I, x\}, \{J, y\} \in E(H_Z^+)$ or $\{I, y\}, \{J, x\} \in E(H_Z^+)$ respectively $\{I, x\}, \{J, x\} \in E(H_Z^+)$.

**Rule 7:** If at least $(k + 2)$ edges in $M$ see a pair $\{x, y\}$ of different vertices in $Z$, then we add $\{x, y\}$ to the set of pair-constraints $\mathcal{P}$.

**Rule 8:** If there exists an edge $e \in M$ such that $e$ sees no single vertex $z \in Z$ and for every pair $\{x, y\}$ seen by $e$ the pair $\{x, y\}$ is a pair-constraint in $\mathcal{P}$, then remove $e_S$ from $S$ and $e$ from $M$. (Recall: If $e = \{I, J\} \in E(H_Z)$, then $e_S$ is the unique edge in $E(V_I, V_J) \cap S$.)

The matching $M$ is always recomputed if, through application of rules, it does no longer cover every inner bubble or is not maximal when testing whether Rules 7 or 8 apply.

Let $L = \mathcal{D}_Z^l \setminus V(M)$ be the set of leaf bubbles that are not covered by $M$. Because the matching covers at least all inner bubbles, we know that $|S| \leq 2|M| + |L|$. Therefore we have to find a reduction rule that reduces the number of leaf bubbles in $L$. Every leaf bubble in $L$ is adjacent to an inner bubble in $H_Z$, because $M$ covers all leaf bubbles that are not adjacent to an inner bubble. To bound the number of leaf bubbles in $L$ we define for each $z \in Z$ a graph $G_z$ with the help of the following two sets. The first one, $L_z = N_{H_Z^+}(z) \cap L$, is the set of all leaf bubbles $I$ that are adjacent to $z$ in $H_Z^+$. The other $V_z^i = \{v \in V \mid \exists J \in N_{H_Z^+}(L_z)\colon v \in V_J\}$ consists of all vertices that are contained in an inner bubble that is adjacent to a leaf bubble in $L_z$. Let $V(G_z) = \{z\} \cup L_z \cup V_z^i$ and

$$E(G_z) = E_{H_Z^+}(z, L_z) \cup \{\{I, w\} \mid \exists I \in L_z, v \in V_I, w \in V_z^i\colon \{v, w\} \in S\} \cup (E(G[V_z^i]) \setminus S).$$

In the graph $G_z$ each leaf bubble $I \in L_z$ is a single vertex. We are not interested in the internal structure of leaf bubbles in $L_z$, whereas we are interested in the structure of the inner bubbles that are adjacent to the leaf bubbles in $L_z$. Thus we add the connected component that corresponds to an inner bubble which is adjacent to a bubble in $L_z$ to $G_z$. In order to apply the concept of flowers and blocking sets in $G_z$, an edge $e \in E(G_z)$ is an $S$-edge in $G_z$ if $e = \{I, w\}$ with $I \in L_z$ and $w \in V_z^i$. Note that $e$ is an edge in $G_z$, because there exists an $S$-edge $e' = \{v, w\}$ in $G$ with $v \in V_I$.

Since no previous rule is applicable and a $z$-flower in $G_z$ gives rise to a $z$-flower in $G$ of same order, one can show, using Gallai's $A$-path Theorem, that there exists a $z$-blocker $B_z \subseteq V_z^i \setminus V(S)$ of size at most $2k$ for every vertex $z \in Z$ in $G_z$. Let $B = \bigcup_{z \in Z} B_z$ be the union of all $z$-blockers $B_z$ of size at most $2k$. Note that the set $L$ is the union of all sets $L_z$ with $z \in Z$, because every leaf bubble is adjacent to a vertex in $Z$, i.e., $L = \bigcup_{z \in Z} L_z$.

Because $B \subseteq V_Z^i \setminus V(S)$ we know that $L \subseteq \mathcal{D}_{Z \cup B}^l$; thus we can use $H_{Z \cup B}$ to bound the number of leaf bubbles in $L$. Let $\boldsymbol{I} = \{J \in \mathcal{D}_{Z \cup B}^i \mid E(L, J) \neq \emptyset\}$ be the set of inner bubbles in $H_{Z \cup B}$ that are adjacent to a leaf bubble in $L$. Clearly the number of edges between $\boldsymbol{I}$ and $L$ in $H_{Z \cup B}$ equals the number $|L|$. Instead of again using a matching to reduce this number we consider more carefully the properties of these edges (more details in full version). For this we define the property of seeing a pair in a slightly different way. Let $e = \{I, J\}$ be an edge with $I \in \boldsymbol{I}$ and $J \in L$. We say that $e = \{I, J\}$ with $I \in \boldsymbol{I}$ and $J \in L$ sees the pair $\{x, y\}$ of different vertices $x \in Z \cup B$ and $y \in Z$, if $\{I, x\}, \{J, y\} \in E(H_{Z \cup B}^+)$. Observe that a bubble in $L$ is never adjacent to a vertex in $B$ in the graph $H_{Z \cup B}$, because $B \subseteq \bigcup_{z \in Z} V_z^i \setminus V(S)$.

**Rule 9:** If at least $(k+2)$ edges $\{I_1, J_1\}, \dots, \{I_l, J_l\}$ with $l \geq k+2$, $I_i \in \boldsymbol{I}$ and $J_i \in L$ for $1 \leq i \leq l$ see a pair $\{x, y\}$ of different vertices, such that $x \in Z \cup B$ is adjacent to $I_i$, $y \in Z$ is adjacent to $J_i$ for all $i \in \{1, \dots, l\}$, then we add $\{x, y\}$ to the set of pair-constraints $\mathcal{P}$.

**Rule 10:** If there exists an edge $e = \{I, J\}$ with $I \in \boldsymbol{I}$ and $J \in L$ such that $e$ sees no single vertex $z \in Z$ and for every pair $\{x, y\}$ seen by $e$ the pair $\{x, y\}$ is a pair-constraint in $\mathcal{P}$, then remove $e_S$ from $S$, delete $J$ from $L$ and replace $I$ by $I \cup J$ in $\boldsymbol{I}$.

Note that if we delete an edge $e = \{I, J\}$ from $S$ by applying Rule 10, then the consequence is that bubbles $I$ and $J$ are now merged into a single bubble.

If no reduction rule is applicable, then $|M| \in \mathcal{O}(k^3)$ and $|L| \in \mathcal{O}(k^4)$. (A proof of these results is deferred to the full version.) As mentioned above, $|S| = |\mathcal{D}^i_Z| + |\mathcal{D}^l_Z| \leq 2|M| + |L| \in \mathcal{O}(k^4)$, because $H_Z$ is a forest, because there is at most one edge of $S$ between any two bubbles, and because $V(S) \cap Z = \emptyset$.

*Finding an equivalent instance for Edge Subset Feedback Vertex Set.* Up to now we can only bound the number of edges in $S$ for the PAIR-CONSTRAINED EDGE SUBSET FVS problem. As mentioned above the instance $(G, S, \mathcal{P} = \emptyset, k)$ for PAIR-CONSTRAINED EDGE SUBSET FVS is equivalent to the instance $(G, S, k)$ of EDGE SUBSET FVS. Therefore we only have to show that we can find in polynomial time an instance of EDGE SUBSET FVS that is equivalent to the instance $(G, S, \mathcal{P}, k)$ of PAIR-CONSTRAINED EDGE SUBSET FVS and has at most $\mathcal{O}(k^4)$ $S$-edges. Let $\{x, y\} \in \mathcal{P}$ be a pair-constraint. If there are two edges between $x$ and $y$ of which at least one is contained in $S$, then $x$ or $y$ must be in any solution, because $xy$ is an $S$-cycle. For this reason, the instance $(G', S' = S \cup \mathcal{P}, k)$ of EDGE SUBSET FVS is equivalent to the instance $(G, S, \mathcal{P}, k)$ of PAIR-CONSTRAINED EDGE SUBSET FVS, where $G'$ is created from $G$ by adding one edge $\{x, y\}$ between every two vertices $x$ and $y$ with $\{x, y\} \in \mathcal{P}$ when $\{x, y\} \notin E$ and by adding an edge $\{x, y\}$ between $x$ and $y$ that is also contained in $S'$; hence there are two edges between $x$ and $y$ with $\{x, y\} \in \mathcal{P}$ in graph $G'$ and we add exactly one edge between $x$ and $y$ to $S'$. Because we cannot apply Rule 4 or 5 to $(G, S, \mathcal{P}, k)$, we know that $|\mathcal{P}| \leq k^2$. This leads to a bound of $|S| + |\mathcal{P}| \in \mathcal{O}(k^4)$ edges in $S'$ for the EDGE SUBSET FVS problem after the reduction. Together with the kernel with $\mathcal{O}(|S|^2 k)$ vertices for EDGE SUBSET FVS parameterized by $|S|$ and $k$, we obtain a kernelized instance with $\mathcal{O}(k^9)$ vertices for EDGE SUBSET FVS parameterized by $k$.

Note that it is no problem that we use in Section 4 the existence of a solution disjoint from $V(S)$ and that we only preserve dominant solutions in Section 3, because the reduction rules in Section 4 as well as the kernelization in Section 3 lead to equivalent instances and because every instance has a dominant solution (if a solution exists).

## 5 Conclusions

We have shown that the SUBSET FVS problem has a randomized polynomial kernelization using the matroid-based tools of Kratsch and Wahlström [12], positively answering the question of Cygan et al. [3]. As in previous work [12] the error-probability can be made exponentially small without increasing the kernel size. Nevertheless, it would of course be very interesting whether the use of randomization and/or matroids can be avoided. Furthermore, there is quite a gap between $\mathcal{O}(k^9)$ vertices and a lower bound of *size* $\mathcal{O}(k^{2-\varepsilon})$ that is inherited from VERTEX COVER [4], conditioned on non-collapse of the polynomial hierarchy.

Other open problems regarding existence of polynomial kernels, possibly amenable to the matroid tools, are MULTIWAY CUT and DIRECTED FEEDBACK VERTEX SET (DFVS). There is also a directed version of SUBSET FVS, called DIRECTED SUBSET FEEDBACK VERTEX SET, but it generalizes DFVS, whose kernel status has remained open for quite some time now.

## References

1 Kevin Burrage, Vladimir Estivill-Castro, Michael R. Fellows, Michael A. Langston, Shev Mac, and Frances A. Rosamond. The undirected feedback vertex set problem has a poly($k$) kernel. In *IWPEC 2006*, volume 4169 of *LNCS*, pages 192–202. Springer, 2006. doi:10.1007/11847250_18.

2 Marek Cygan, Jesper Nederlof, Marcin Pilipczuk, Michal Pilipczuk, Johan M. M. van Rooij, and Jakub Onufry Wojtaszczyk. Solving connectivity problems parameterized by treewidth in single exponential time. In *FOCS 2011*, pages 150–159. IEEE Computer Society, 2011. doi:10.1109/FOCS.2011.23.

3 Marek Cygan, Marcin Pilipczuk, Michal Pilipczuk, and Jakub Onufry Wojtaszczyk. Subset feedback vertex set is fixed-parameter tractable. *SIAM J. Discrete Math.*, 27(1):290–309, 2013. doi:10.1137/110843071.

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

5 Reinhard Diestel. Graph theory. *Graduate texts in mathematics*, 2005.

6 Guy Even, Joseph Naor, and Leonid Zosin. An 8-approximation algorithm for the subset feedback vertex set problem. *SIAM J. Comput.*, 30(4):1231–1252, 2000. doi:10.1137/S0097539798340047.

7 Fedor V. Fomin, Daniel Lokshtanov, and Saket Saurabh. Efficient computation of representative sets with applications in parameterized and exact algorithms. In *SODA 2014*, pages 142–151. SIAM, 2014. doi:10.1137/1.9781611973402.10.

8 Tibor Gallai. Maximum-minimum sätze und verallgemeinerte faktoren von graphen. *Acta Mathematica Hungarica*, 12(1-2):131–173, 1961.

9 Eva-Maria C. Hols and Stefan Kratsch. A randomized polynomial kernel for subset feedback vertex set. *CoRR*, abs/1512.02510, 2015. URL: http://arxiv.org/abs/1512.02510.

10 Ken-ichi Kawarabayashi and Yusuke Kobayashi. Fixed-parameter tractability for the subset feedback set problem and the s-cycle packing problem. *J. Comb. Theory, Ser. B*, 102(4):1020–1034, 2012. doi:10.1016/j.jctb.2011.12.001.

11 Tomasz Kociumaka and Marcin Pilipczuk. Faster deterministic feedback vertex set. *Inf. Process. Lett.*, 114(10):556–560, 2014. doi:10.1016/j.ipl.2014.05.001.

12 Stefan Kratsch and Magnus Wahlström. Representative sets and irrelevant vertices: New tools for kernelization. In *FOCS 2012*, pages 450–459. IEEE Computer Society, 2012. doi:10.1109/FOCS.2012.46.

13 Daniel Lokshtanov, M. S. Ramanujan, and Saket Saurabh. Linear time parameterized algorithms for subset feedback vertex set. In *ICALP 2015*, volume 9134 of *LNCS*, pages 935–946. Springer, 2015. doi:10.1007/978-3-662-47672-7_76.

14 László Lovász. Flats in matroids and geometric graphs. *Combinatorial surveys*, pages 45–86, 1977.

15 Dániel Marx. A parameterized view on matroid optimization problems. *Theor. Comput. Sci.*, 410(44):4471–4479, 2009. doi:10.1016/j.tcs.2009.07.027.

16 Dániel Marx, Barry O'Sullivan, and Igor Razgon. Finding small separators in linear time via treewidth reduction. *ACM Transactions on Algorithms*, 9(4):30, 2013. doi:10.1145/2500119.

17 Hazel Perfect. Applications of Menger's graph theorem. *J. Math. Anal. Appl.*, 22:96–111, 1968.

18 Alexander Schrijver. A short proof of Mader's sigma-paths theorem. *J. Comb. Theory, Ser. B*, 82(2):319–321, 2001. doi:10.1006/jctb.2000.2029.

19 Stéphan Thomassé. A $4k^2$ kernel for feedback vertex set. *ACM Transactions on Algorithms*, 6(2), 2010. doi:10.1145/1721837.1721848.

20 Po Tong, Eugene L. Lawler, and Vijay V. Vazirani. *Solving the weighted parity problem for gammoids by reduction to graphic matching*. Computer Science Division, University of California, 1982.

21 Magnus Wahlström. Half-integrality, lp-branching and FPT algorithms. In *SODA 2014*, pages 1762–1781. SIAM, 2014. doi:10.1137/1.9781611973402.128.