An FPT Algorithm for Set Splitting

Frank Dehne[1], Michael R. Fellows[2], and Frances A. Rosamond[2]

[1] Carleton University, Ottawa, Canada,
frank@dehne.net
http://www.dehne.net
[2] Univ. of Newcastle, Australia
{mfellows,fran}@cs.newcastle.edu.au

Abstract. An FPT algorithm with a running time of $O(n^4 + 2^{O(k)}n^{2.5})$ is described for the SET SPLITTING problem, parameterized by the number k of sets to be split. It is also shown that there can be no FPT algorithm for this problem with a running time of the form $2^{o(k)}n^c$ unless the satisfiability of n-variable 3SAT instances can be decided in time $2^{o(n)}$.

1 Introduction

Consider a collection F of subsets of a finite set X. The task is to find a partition of X into two disjoint subsets X_0 and X_1 which maximizes the number of subsets of F that are split by the partition, i.e. not entirely contained in either X_0 or X_1. This problem, called the MAX SET SPLITTING problem, is NP-complete [9] and APX-complete [15].

Andersson and Engebretsen [3] as well as Zhang [16] presented approximation algorithms that provide solutions within a factor of 0.7240 and 0.7499, respectively. A 1/2 approximation algorithm for the special case of the MAX SET SPLITTING problem where the size of X_0 is given was presented in [1]. A variation of the SET SPLITTING problem, called MAX E_m SPLITTING, in which all sets in F contain the same number of elements m is NP-hard for any fixed $m \geq 2$ [13]. As pointed out in [16], MAX E_m SPLITTING is a special case of MAX E_m NAE-SAT. MAX E_m splitting is approximable within a factor of 0.8787 for $m \leq 3$ and approximable within a factor of $\frac{1}{1-2^{1-m}}$ for $m \geq 4$ [12,17,18], and it is NP-hard for factor $\frac{1}{\frac{8}{7}-1}$ [10]. In [4] it is shown that a polynomial time approximation scheme exists for $|F| = \Theta(|X|^k)$.

In this paper, we show that the SET SPLITTING problem is *fixed parameter tractable* [8] if we consider as parameter, k, the number of sets in F that are split by a given partition of X, for arbitrary size sets in F. We present an FPT algorithm with a running time $O(n^4 + 2^{O(k)}n^{2.5})$, parameterized by the number k of sets to be split. We also show that there can be no FPT algorithm for this problem with a running time of the form $2^{o(k)}n^c$ unless the satisfiability of n-variable 3SAT instances can be decided in time $2^{o(n)}$.

There are a variety of fundamental computational problems that concern families $\mathcal{F} \subseteq 2^X$ of subsets of a base set X, such as HITTING SET (finding a

H.L. Bodlaender (Ed.): WG 2003, LNCS 2880, pp. 180–191, 2003.

small subset X' of the base set such that every set in $\mathcal{F}$ contains an element of X'), Set Packing (finding a large pairwise disjoint subfamily $\mathcal{F}'$ of $\mathcal{F}$), and Set Splitting. Important applications of these problems on families of sets arise in the analysis of *micro-array* data. Micro-array data can be viewed as a matrix, where the columns represent "features" such as whether a gene is "on" or "off", or whether a sample is cancerous or not, while the rows represent the samples. Each row can be viewed as representing a subset of the column space. An example of how Set Splitting is relevant is given by the following scenario. The set X represents a set of genetic markers (such as SNPs) of an individual. A set $A \in \mathcal{F}$ represents a combination of markers associated with a phenotypic condition that is exhibited by an individual but not by either parent. The partition of X maximizing the number of sets that are split represents a parsimonious hypothesis explaining why the phenotypic signals are not present in the parents (since for a set of markers that is split, the entire set is not present in either parent).

Since the Set Splitting is NP-complete and APX-complete, it is important to find new approaches that can solve problem instances of practical importance. Fixed parameter tractability has provided such solutions for various other NP-complete problems [8], and FPT algorithms have considerably increased the size of solvable problem instances for some of these problems [6]. In this paper, we contribute the first exploration of the parameterized complexity of one of the classic problems about families of sets, Set Splitting, parameterized by the number of sets to be split.

2 A FPT Algorithm for Set k-Splitting

A set k-splitting, $\mathcal{SSP}(X, F, k)$, for a collection F of n subsets of a finite set X is formally defined as a partition $[X_0, X_1]$ of X into two disjoint subsets X_0 and X_1 such that at least k sets in F are split by the partition. See Figure 1(a) for an illustration. For the remainder we assume, w.l.o.g., that $X = \bigcup_{S \in F} S$. Let $|X| = n$. Define a predicate $\Pi_{\mathcal{SSP}}(X, F, k)$ by $\Pi_{\mathcal{SSP}}(X, F, k) =$ TRUE if $\mathcal{SSP}(X, F, k)$ exists and $\Pi_{\mathcal{SSP}}(X, F, k) =$ FALSE if $\mathcal{SSP}(X, F, k)$ does not exist.

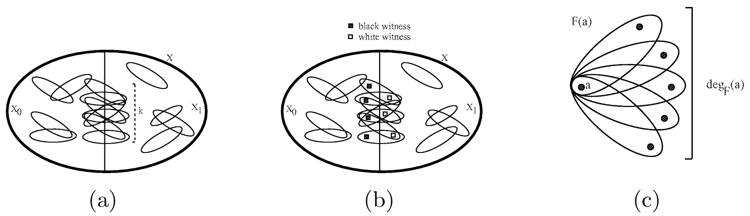

Fig. 1. (a)Illustration Of A Set k-Splitting $\mathcal{SSP}(X, F, k) = [X_0, X_1]$. (b) A k-Witness Structure. (c) Illustration Of F(a) And $\deg_F(a)$.

Definition 1. *Consider a problem instance* (X, F, k)*. A sequence* $W = (b_1, \ldots, b_k, w_1, \ldots, w_k) \in X^{2k}$ *is called a* **k-witness structure** *for* F *if and only if* $(b_1, \ldots, b_k) \cap (w_1, \ldots, w_k) = \emptyset$ *and there exist* k *subsets* $S_i \in F$ *such that* $\{b_i, w_i\} \subseteq S_i$ *for all* $i = 1, \ldots, k$*.*

For a k-witness structure $W = (b_1, \ldots, b_k, w_1, \ldots, w_k)$, we will call b_i the *black* witness element for S_i, and w_i the *white* witness element for S_i. See Figure 1(b) for an illustration. Note that, some black [white] witness elements b_i, b_j $[w_i, w_j]$ may be identical. A witness element in a k-witness structure $W = (b_1, \ldots, b_k, w_1, \ldots, w_k)$ is called *private* if it is unique in W; otherwise it is called *shared.*

For any $W = (b_1, \ldots, b_k, w_1, \ldots, w_k)$, we refer to the process of replacing all occurrences of an element b_i or w_j by another element $b_{i'}$ or $w_{j'}$, respectively, as *deleting* b_i or w_j. A k-witness structure $W = (b_1, \ldots, b_k, w_1, \ldots, w_k)$ is *non-redundant* if any W' obtained from W by deleting any single element b_i or w_j is *not* a k-witness structure.

Observation 1 *Consider a problem instance* (X, F, k)*. If there exists a* k*-witness structure then there also exists a non-redundant* k*-witness structure.*

Lemma 1. *(a) Every* k*-splitting* $\mathcal{SSP}(X, F, k) = [X_0, X_1]$ *implies at least one* k*-witness structure* $(b_1, \ldots, b_k, w_1, \ldots, w_k)$*. (b) Every* k*-witness structure* $(b_1, \ldots, b_k, w_1, \ldots, w_k)$ *implies at least one* k*-splitting* $\mathcal{SSP}(X, F, k) = [X_0, X_1]$

Proof. (a) Consider a k-splitting $\mathcal{SSP}(X, F, k) = [X_0, X_1]$ which splits k sets $S_1, \ldots, S_k$. This implies a k-witness structure $(b_1, \ldots, b_k, w_1, \ldots, w_k)$ where each b_i is an arbitrary element in $S_i \cap X_0$ and each w_i is an arbitrary element in $S_i \cap X_1$, $i = 1, \ldots, k$. (b) A k-witness structure $(b_1, \ldots, b_k, w_1, \ldots, w_k)$ implies a k-splitting $[X_0, X_1]$ where $X_0 = \{b_1, \ldots, b_k\}$ and $X_1 = X - X_0$.

For a k-witness structure $W = (b_1, \ldots, b_k, w_1, \ldots, w_k)$ let $F(W) = \{S \in F | \{b_i, w_i\} \subseteq S \text{ for some } 1 \le i \le k\}$ be the collection of (at least k) sets $S \in F$ which are split by W. For each $a \in X$ let $\mathrm{F}(a) \subset F$ be the collection of sets $S \in F$, $|S| \ge 2$, which contain a, and let $\deg_F(a) = |\mathrm{F}(a)|$. See Figure 1(c) for an illustration.

Algorithm 1 *Kernelization*: Convert the given problem instance (X, F, k) into an equivalent *reduced* problem instance.

(1) Apply the following rules as often as possible.

Rule 1: IF there exists an element $a \in X$ with $\deg_F(a) > k$ THEN report $\mathcal{SSP}(X, F, k) = [\{a\}, X - \{a\}]$ and STOP.

Rule 2: IF there exists a set $S \in F$ with $|S| \le 1$ THEN set $F \leftarrow F - \{S\}$.

Rule 3: IF there exists a set $S \in F$ with $|S| \ge 2k$ THEN set $F \leftarrow F - \{S\}$ and $k \leftarrow k - 1$.

Rule 4: IF there exists a set $S \in F$, $|S| \ge 2$, which contains an element $a \in S$ with $\deg_F(a) = 1$ THEN set $F \leftarrow F - \{S\}$ and $k \leftarrow k - 1$.

Rule 5: IF there exist three different elements $a_1, a_2, a_3 \in X$ with $\emptyset \neq \mathrm{F}(a_1) \subset \mathrm{F}(a_2) \subset \mathrm{F}(a_3)$ THEN set $S \leftarrow S - \{a_1\}$ for all $S \in F$. (Note: May need to re-apply Rule 2.)

(2) Set $X \leftarrow \bigcup_{S \in F} S$.

— End of Algorithm —

In the following, we prove the correctness of the above rules. The parts marked "(⇒)" show that $\Pi_{\mathcal{SSP}}(X, F, k) = \text{TRUE}$ before the application of a rule implies $\Pi_{\mathcal{SSP}}(X, F', k') = \text{TRUE}$ after the application of that rule. The sections marked "(⇐)" show the opposite direction, i.e. $\Pi_{\mathcal{SSP}}(X, F', k') = \text{TRUE}$ after the application of a rule implies $\Pi_{\mathcal{SSP}}(X, F, k) = \text{TRUE}$ before the application of that rule.

Proof Of Correctness For Rule 1. If $\deg_F(a) > k$ then $[\{a\}, X - \{a\}]$ splits k or more sets. □

Proof Of Correctness For Rule 2. Consider a set $S \in F$ with $|S| \leq 1$. (⇒) Since S can not be split by any $\mathcal{SSP}(X, F, k) = [X_0, X_1]$, any such $[X_0, X_1]$ also splits k sets in $F - \{S\}$. (⇐) If there exists a $\mathcal{SSP}(X, F - \{S\}, k) = [X_0, X_1]$ then the same $[X_0, X_1]$ is also a set k-splitting $\mathcal{SSP}(X, F, k)$. □

Proof Of Correctness For Rule 3. Consider a set $S \in F$ with $|S| \geq 2k$. (⇒) If $[X_0, X_1]$ is a k-splitting for F then $[X_0, X_1]$ is either a k-splitting for $F - \{S\}$ (if S is not split by $[X_0, X_1]$) or $[X_0, X_1]$ is a $k-1$-splitting for $F - \{S\}$ (if S is split by $[X_0, X_1]$). (⇐) Consider a $k - 1$-splitting $[X_0, X_1]$ of $F - \{S\}$ with witness structure $W = (b_1, \ldots, b_{k-1}, w_1, \ldots, w_{k-1})$ by Lemma 1(a). Since $|S| \geq 2k$, there exist two elements $b_k, w_k \in S - W$. Hence $(b_1, \ldots, b_k, w_1, \ldots, w_k\}$ is a witness structure for a k-splitting of F; see Lemma 1(b). □

Proof Of Correctness For Rule 4. Consider an $S \in F$ with $|S| \geq 2$ which contains an element $a \in S$ with $\deg_F(a) = 1$. (⇒) If $[X_0, X_1]$ is a k-splitting for F then $[X_0, X_1]$ is a $k - 1$-splitting for $F - \{S\}$. (⇐) Consider a $k - 1$-splitting $[X_0, X_1]$ for $F - \{S\}$ with witness structure $W = (b_1, \ldots, b_{k-1}, w_1, \ldots, w_{k-1})$ by Lemma 1(a). Since $\deg_F(a) = 1$, a is not an element of W. Since $|S| \geq 2$, S contains another element $a' \neq a$. Assume $a' \in W$, w.l.o.g. $a' = b_j$ for some $1 \leq j \leq k - 1$, then $W' = (b_1, \ldots, b_{k-1}, w_1, \ldots, w_{k-1}, w_k)$ with $w_k = a$ is a k-witness structure for a k-splitting of F (Lemma 1(b)). Assume $a' \notin W$, then $W'' = (b_1, \ldots, b_{k-1}, b_k, w_1, \ldots, w_{k-1}, w_k)$ with $b_k = a'$ and $w_k = a$ is a k-witness structure and implies a k-splitting of F (Lemma 1(b)). □

Proof Of Correctness For Rule 5. Consider three different elements $a_1, a_2, a_3 \in X$ with $\emptyset \neq \mathrm{F}(a_1) \subset \mathrm{F}(a_2) \subset \mathrm{F}(a_3)$. Let $F' = \{S - \{a_1\} | S \in F\}$. (⇒) Assume that $[X_0, X_1]$ is a k-splitting for F with *non-redundant* k-witness structure $W = (b_1, \ldots, b_k, w_1, \ldots, w_k)$. If $a_1 \notin W$ then W is also a witness structure for a

k-splitting of F'. Assume $a_1 \in W$, w.l.o.g. $a_1 = b_i$ for some $1 \le i \le k-1$. If both, a_2 and a_3 were contained in W then they can not both be black [white] witness elements since, otherwise, W would not be minimal (a_2 could be replaced by a_3). However, if a_2 and a_3 were both contained in W and had different colors, then W would not be minimal as well since a_1 could be replaced by either a_2 or a_3. Hence, only either a_2 or a_3 can be in W and must have a color different from a_1 (otherwise a_1 could be replaced by either a_2 or a_3). Assume, w.l.o.g. $a_2 = w_j$ for some $1 \le j \le k-1$. Then, W' obtained from W by replacing a_1 by a_3 is also a k-witness structure and implies a k-splitting of F' (since W' does not contain a_1). ($\Leftarrow$) Assume that $[X_0, X_1]$ is a k-splitting for F' then it is also a k-splitting for F. □

Theorem 1. *Let (X, F, k) be any reduced problem instance. If $|F| \ge 2k$ then $\Pi_{\mathcal{SSP}}(X, F, k)$ = TRUE.*

Theorem 1 follows from Lemma 2, below.

Lemma 2. *"Boundary Lemma"*
If (X, F, k) is reduced and $\Pi_{\mathcal{SSP}}(X, F, k)$ = TRUE and $\Pi_{\mathcal{SSP}}(X, F, k+1)$ = FALSE then $|F| \le 2k$.

Proof. Assume there exists a counter example to Lemma 2, that is, a reduced problem instance (X, F, k) with $\Pi_{\mathcal{SSP}}(X, F, k)$ = TRUE and $\Pi_{\mathcal{SSP}}(X, F, k+1)$ = FALSE but $|F| > 2k$. The following Claims 2 to 2 show that this leads to a contradiction.

Since $\Pi_{\mathcal{SSP}}(X, F, k)$ = TRUE, there exists a k-splitting $\mathcal{SSP}(X, F, k) = [X_0, X_1]$ which splits k sets $S_1, \ldots, S_k \in F$ and, by Lemma 1(a), there exists a k-witness structure $W = (b_1, \ldots, b_k, w_1, \ldots, w_k)$ such that $\{b_i, w_i\} \subset S_i$ for all $i = 1, \ldots, k$.

A set $S \in F$ is called *chosen* if $S = S_i$ for some $i \in \{1, \ldots, k\}$, otherwise S is called *not chosen*. Recall that any element $b_1, \ldots, b_k$ is called *black* and any $w_1, \ldots, w_k$ is called *white*. All other elements $a \in X - W$ are called *grey*.

Claim. If a set $S \in F$ is not chosen then it cannot contain both a black and a white element.

Proof. Otherwise, $\Pi_{\mathcal{SSP}}(X, F, k+1)$ = TRUE.

Claim. If a set $S \in F$ is not chosen then it consists entirely of black elements or entirely of white elements.

Proof. Assume, w.l.o.g, that S contains a black element b_i and a grey element $a \in X - W$. If we color a white, i.e. add $b_{k+1} = b_i$ and $w_{k+1} = a$ to W, then we obtain a $k+1$-witness and, hence, $\Pi_{\mathcal{SSP}}(X, F, k+1)$ = TRUE. Assume that S contains only grey elements. Due to Rule 2, S contains at least two elements a and b. If we color a white and b black, then we obtain a $k+1$-witness and, hence, $\Pi_{\mathcal{SSP}}(X, F, k+1)$ = TRUE.

Two chosen sets S_i, S_j are called *connected* if $b_i = b_j$ or $w_i = w_j$. Let $C_1, \ldots, C_r \subset F$ be a partitioning of $\{S_1, \ldots, S_k\}$ into maximal collections of connected chosen sets (i.e. connected components with respect to the above *connected* relation). We will refer to $C_1, \ldots, C_r$ as *connected components.*

A set $S \in F$ *intersects* a component C_t if there exists a chosen set $S_i \in C_t$ with $b_i \in S$ or $w_i \in S$.

Claim. If $S \in F$ is not chosen then it does not intersect two different components C_t and $C_{t'}$. (See Figure 2 for an illustration.)

Proof. Consider a set $S \in F$ that is not chosen and assume that S intersects two different components C_t and $C_{t'}$. Hence, there exists a chosen set $S_i \in C_t$ with $b_i \in S$ or $w_i \in S$ and there exists a chosen set $S_j \in C_{t'}$ with $b_j \in S$ or $w_j \in S$. If $b_i \in S$ and $w_j \in S$ then S is split by $[X_0, X_1]$ and, hence, $\Pi_{\mathcal{SSP}}(X, F, k+1) =$ TRUE. Thus, assume w.l.o.g. that $b_i \in S$ and $b_j \in S$. However, if we now invert the colors of all witnesses for chosen sets in C_t then all chosen sets in C_t are still split and S is split as well. Thus, $\Pi_{\mathcal{SSP}}(X, F, k+1) =$ TRUE; a contradiction.

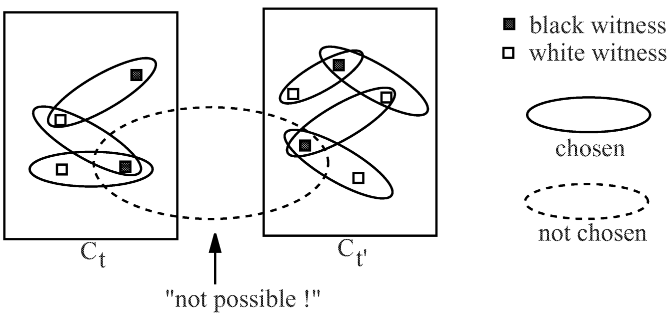

Fig. 2. Illustration of Claim 2

Claim. For any components C_t, the number of sets which intersect C_t and are chosen is larger than (or equal to) the number of sets which intersect C_t and are not chosen.

Proof. Assume that the number of sets which intersect C_t and are not chosen is larger than the number of sets which intersect C_t and are chosen. Consider the following undirected graph $G = (V, E)$ where V is the union of the witnesses $\{b_i, w_i\}$ of all sets S_i which intersect C_t and are chosen and $E = E_1 \cup E_2$ defined as follows. For each chosen set S_i which intersect C_t, E_1 contains an edge $e(S_i) = (b_i, w_i)$. For each set $S \in F$ which intersects C_t and is not chosen, E_2 contains an edge $e(S) = (a_1, a_2)$ where a_1, a_2 are two arbitrary (but different) elements of S. Note that, since (X, F, k) is reduced, $|S| \geq 2$. Furthermore, it follows from Claim 2 that S consists entirely of black elements or entirely of white elements.

Hence, a_1 and a_2 are either both black or white elements. See Figure 3 for an illustration. For any $v \in V$ let $deg_1(v)$ be the number of edges in E_1 that are incident to v and let $deg_2(v)$ be the number of edges in E_2 that are incident to v. From the assumption at the beginning of this proof it follows that $|E_2| > |E_1|$. Hence, there exists a vertex $v_0 \in V$ such that $deg_2(v_0) > deg_1(v_0)$. As a consequence, if we invert the color of the witness corresponding to v_0, we obtain a new witness structure which splits at least one more set, which implies $\Pi_{\mathcal{SSP}}(X, F, k + 1) =$ TRUE; a contradiction.

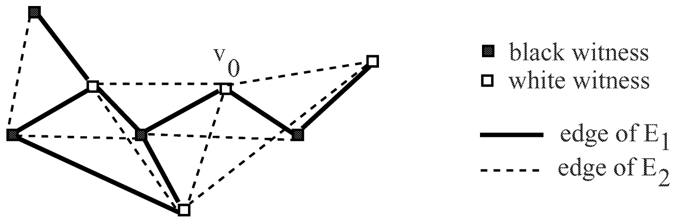

Fig. 3. Illustration of Claim 2

Claim. The number of sets which are chosen is larger than (or equal to) the number of sets which are not chosen.

Proof. Follows from Claims 2 and 2.

Since the number of chosen sets is k, it follows from Claim 2 that $F \leq 2k$; a contradiction. This concludes the proof of Lemma 2 and Theorem 1.

Let (X, F, k) be a *reduced* problem instance. If $|F| \geq 2k$ then $\Pi_{\mathcal{SSP}}(X, F, k) =$ TRUE by Theorem 1. Hence, for the remainder let us assume that $|F| < 2k$.

Since, by Rule 3 of Algorithm 1, every $S \in F$ is of size smaller than $2k$, it follows that X is of size at most $4k^2$. Thus, there are at most 2^{4k^2} possible k-witness structures which implies an $O(n^4 + n2^{4k^2})$ algorithm for set k-splitting. In the following we will show how to reduce the 2^{4k^2} term to $2^{O(k)}$.

Consider a *reduced* problem instance (X, F, k) with $|F| < 2k$. If $\Pi_{\mathcal{SSP}}(X, F, k) =$ TRUE then there exists a k-witness structure $W = (b_1, \ldots, b_k, w_1, \ldots, w_k)$. Assume that W has the largest number of private elements among all possible k-witness structures for (X, F, k). We shall call such a k-witness structure *maximal*. Let $X(W)$ denote the set of elements $a \in X$ that are contained in W. Consider the collection $F(W)$ of sets $S_i \in F$ which are split by W. Each set $S_i \in F(W)$ has a black witness b_i and a white witness w_i in W. We partition $F(W)$ into sets $A(W)$, $B(W)$, $C(W)$, $D(W)$ where $S_i \in A(W)$ if S_i has a private black witness and a private white witness, $S_i \in B(W)$ if S_i has a private black witness and a shared white witness, $S_i \in C(W)$ if S_i has a private white witness and a shared black witness, and $S_i \in D(W)$ if S_i has a shared black witness and a shared white witness. See Figure 4 for an illustration. Let X_A be the set of all elements $a \in X$ that are contained in at least one set $S_i \in A(W)$, and let X_{BCD} be the set of all elements $a \in X$ that are contained in at least one set $S_i \in B(W) \cup C(W) \cup D(W)$.

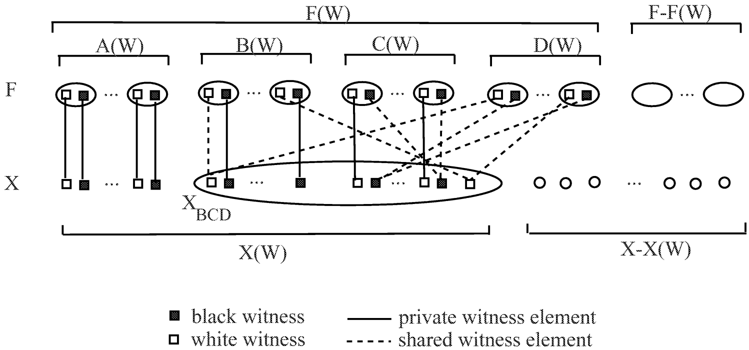

Fig. 4. Illustration of $A(W)$, $B(W)$, $C(W)$, $D(W)$ and X_{BCD}.

Lemma 3. *Consider a reduced problem instance* (X, F, k) *with* $|F| < 2k$ *and a maximal* k*-witness structure* W*, and let* X_{BCD} *and* $X(W)$ *be defined as above, then* $X_{BCD} \subset X(W)$*.*

Proof. Consider an element $a \in X_{BCD}$ which is not contained in $X(W)$. By assumption, a is contained in a set $S_i \in B(W) \cup C(W) \cup D(W)$. Note that, by definition, S_i has at most one private witness in W. However, since a is not in $X(W)$, we can add a to W as a witness for S_i, replacing a shared witness of S_i. This increases the number of private witnesses in W; a contradiction to the assumption that W is maximal.

For any $F' \subset F$ and $X' \subset X$ let $\hat{G}(F', X')$ be a bipartite graph with vertex set $V_{F'} \cup X'$, where $V_{F'}$ contains two elements b_S and w_S for each $S \in F'$. For every set $S \in F'$ and each element $a \in S \subset X'$, the graph $\hat{G}$ contains two edges: one edge between b_S and a, and one edge between w_S and a. A maximum matching [11] in the bipartite graph $\hat{G}(F', X')$ is called *complete* if every vertex in $V_{F'}$ is matched. For such a matching, we call those elements in X' matched to a $b_S \in V_{F'}$ the black elements and those elements in X' matched to a $w_S \in V_{F'}$ the white elements.

Algorithm 2 SET k-SPLITTING

(1) *Kernelization*

Using Algorithm 1, convert the given problem instance into an equivalent *reduced* problem instance (X, F, k). IF $|F| \geq 2k$ THEN report that $\Pi_{\mathcal{SSP}}(X, F, k) =$ TRUE and STOP.

(2) *Search*

FOR ALL collections $\{S_1, \ldots, S_k\}$ of k sets of F

FOR ALL bi-partitions of $\{S_1, \ldots, S_k\}$ into $A(W)$ and $\bar{A}(W) = B(W) \cup C(W) \cup D(W)$ such that $|X_{BCD}| \leq 2k$

FOR ALL tri-partitions of X_{BCD} into X^0_{BCD}, X^1_{BCD} and X^2_{BCD} for which $[X^0_{BCD}, X^1_{BCD}]$ splits all sets in $B(W) \cup C(W) \cup D(W)$

(Note: X^0_{BCD} represents those elements $a \in X_{BCD}$ that are black witnesses for sets in $B(W) \cup C(W) \cup D(W)$, X^1_{BCD} represents those elements $a \in X_{BCD}$ that are white witnesses for sets in $B(W) \cup C(W) \cup D(W)$, and X^2_{BCD} represents the remainder of X_{BCD}.)

IF there exists a complete matching in $\hat{G}(A(W), X_A - (X^0_{BCD} \cup X^1_{BCD}))$ with black elements X^0_A and white elements X^1_A THEN report that $[X^0_{BCD} \cup X^0_A, X^1_{BCD} \cup X^1_A]$ splits k sets in F and $\Pi_{\mathcal{SSP}}(X, F, k) =$ TRUE and STOP.

Report that $\Pi_{\mathcal{SSP}}(X, F, k) =$ FALSE.

— End of Algorithm —

Theorem 2. *Algorithm 2 solves the* SET k-SPLITTING *problem in time* $O(n^4 + 2^{O(k)} n^{2.5})$.

Proof. The time for Step 1 is bounded by $O(n^4)$. For Step 2, observe that the number of collections $\{S_1, \ldots, S_k\}$ of k sets of F is at most $4^k = 2^{2k}$ and the number of bi-partitions of $\{S_1, \ldots, S_k\}$ is at most 2^k. Since $X_{BCD} \subset X(W)$ due to Lemma 3, and $|X(W)| \leq 2k$, the number of tri-partitions of X_{BCD} is at most $3^{2k} = 2^{\frac{2 \log 3}{\log 2} k}$. The computation of a maximum matching requires time $O(n^{5/2})$[11]. Thus, the time for Step 2 is bounded by $O(n^{5/2} 2^{(3 + \frac{2 \log 3}{\log 2}) k})$.

3 Optimality

In this section we employ newly developed methods of parameterized complexity analysis to prove "exponential optimality" for our main FPT result in Section 2.

To illustrate the issue, consider some recent results for the k-VERTEX COVER problem. An FPT algorithm of the form $2^{O(k)}$, based on search trees, was first described by Mehlhorn [14]. In [2] it was shown that the PLANAR k-VERTEX COVER problem can be solved in time $2^{O(\sqrt{k})} n$. This raises two natural questions: (1) Is there an FPT algorithm for the general k-VERTEX COVER problem with running time $2^{O(\sqrt{k})} n^c$? (2) Can the FPT algorithm for the PLANAR k-VERTEX COVER problem be further improved? For example, is an algorithm with time $2^{O(k^{1/3})} n^c$ possible? It has been shown that the answers to both questions is "no" [5]. There is no FPT algorithm with a running time of the form $2^{o(k)} n^c$ for the general k-VERTEX COVER PROBLEM, and there is no FPT algorithm with a running time $2^{o(\sqrt{k})} n^c$ for the PLANAR k-VERTEX COVER problem unless there is an unlikely collapse $FPT = MINI[1]$ in the hierarchy of parameterized complexity classes

$$FPT \subseteq MINI[1] \subseteq W[1].$$

Such a collapse is considered unlikely because because $FPT = MINI[1]$ if and only if n-variable 3SAT can be solved in time $2^{o(n)}$ [7,5].

In the remainder of this section we establish a similar result for the SET k-SPLITTING problem $\mathcal{SSP}(X, F, k)$.

Theorem 3. *There is no FPT algorithm for* SET k-SPLITTING *with running time* $2^{o(k)}n^c$ *unless* $FPT = MINI[1]$.

Proof. It is sufficient to show that the following problem is hard for $MINI[1]$:

> ***MINI*** SET SPLITTING
> Input: Integers k, n in unary format, a family $\mathcal{F} \subseteq 2^X$ where $|\mathcal{F}| \leq k \log n$, and an integer r. Parameter: k. Question: Is there a bipartition of X that splits at least r sets of F.

Proving that MINI SET SPLITTING is hard for $MINI[1]$ is sufficient to establish our theorem because (1) If there is a $2^{o(s)}n^c$ algorithm to determine if s sets can be split then this algorithm can be used to determine in time $2^{o(k \log n)}n^c$ if $r \leq k \log n$ sets can be split. (2) If $g(n, k) = o(k \log n)$ for any fixed k, i.e. $lim_{n\to\infty} \frac{f(k,n)}{k \log n} = 0$ for any fixed k, then $2^{f(k,n)}n^c$ is bounded by $g(k)n^{c'}$ for appropriately chosen constant c' and function $g(k)$.

To show that MINI SET SPLITTING is $MINI[1]$ hard, we reduce from the $MINI[1]$-complete problem MINI-3SAT [7]

> ***MINI 3SAT***
> Input: Integers k, n in unary format, a 3SAT expression $\mathcal{E}$ where $\mathcal{E}$ has at most $k \log n$ variables and $k \log n$ clauses. Parameter: k. Question: Is $\mathcal{E}$ satisfiable.

The standard reduction from 3SAT to the variant NOT-ALL-EQUAL 3SAT FOR ALL POSITIVE LITERALS is, in fact, a linear-size reduction. That is, the expression $\mathcal{E}'$ to which $\mathcal{E}$ is transformed satisfies $|\mathcal{E}'| \leq c|\mathcal{E}|$ for some constant c. This yields immediately an FPT reduction from MINI 3SAT to MINI NOT-ALL-EQUAL 3SAT FOR ALL POSITIVE LITERALS. The latter is a special case of MINI SET SPLITTING (where all sets in the family $\mathcal{F}$ have size 3).

4 Conclusion

In this paper, we have presented the first exploration of the parameterized complexity of one of the classic problems about families of sets, SET SPLITTING, parameterized by the number of sets to be split. We have presented an FPT algorithm with a running time of $2^{O(k)}n^{2.5}$ and shown that there can be no FPT algorithm for this problem with a running time of the form $2^{o(k)}n^c$ unless the satisfiability of n-variable 3SAT instances can be decided in time $2^{o(n)}$.

The "final" goal of FPT methods is to increase the size of solvable problem instances for NP-complete problems like SET SPLITTING. In this context, an important open question is whether the $2^{(3+\frac{2\log 3}{\log 2})k}$ term in the running time of Algorithm 2 can be further reduced to some 2^{ck} with $c < 3 + \frac{2\log 3}{\log 2}$. We emphasize that this is a first study of the parameterized complexity of SET SPLITTING. For a practical implementation, our algorithm will probably also require more reduction rules in order to "shrink" problem instances as much as possible during kernelization and thereby further increase the size of solvable problem instances.

References

1. A.A. Ageev and M.I. Sviridenko. An approximation algorithm for hypergraph max k-cut with given sizes of parts. In *Proc. ESA 2000*, volume 1879 of *Lecture Notes in Computer Science*, pages 32–41, 2000.
2. J. Alber, H.L. Bodlaender, H. Fernau, and R. Niedermeier. Fixed parameter algorithms for planar dominating set and related problems. In *Proc. Scandinavian Workshop on Algorithm Theory (SWAT 2000)*, Lecture Notes in Computer Science, pages 97–110, 2000.
3. G. Andersson and L. Engebretsen. Better approximation algorithms and tighter analysis for set splitting and not-all-equal sat. In *ECCCTR: Electronic Colloquium on Computational Complexity*, 1997.
4. S. Arora, D. Karger, and M. Karpinski. Polynomial time approximation schemes for dense instances of NP-hard problems. In *Proc. ACM STOC*, pages 284–293, 1995.
5. L. Cai and D. Juedes. Subexponential parameterized algorithms collapse the w-hierarchy. In *Proc. ICALP 2001*, volume 2076 of *Lecture Notes in Computer Science*, 2001.
6. J. Cheetham, F. Dehne, A. Rau-Chaplin, U. Stege, and P. J. Taillon. Solving large FPT problems on coarse grained parallel machines. *Journal of Computer and System Sciences.* to appear.
7. R.D. Downey, V. Estivill-Castro, M.R. Fellows, E. Prieto-Rodriguez, and F.A. Rosamond. Cutting up is hard to do: The parameterized complexity of k-cut and related problems. In *Proc. Computing: The Australasian Theory Symposium (CATS 2003), to appear*, Elsevier Electronic Notes in Computer Science, Adelaide, 2003.
8. R.G. Downey and M.R. Fellows. *Parameterized Complexity.* Springer-Verlag, 1998.
9. M. Garey and D. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness.* W.H. Freeman and Company, 1979.
10. J. Hastad. Some optimal inapproximability results. In *Proc. ACM STOC*, pages 1–10, 1997.
11. J.E. Hopkroft and R.M. Karp. An $n^{5/2}$ algorithm for maximum matching in bipartite graphs. *SIAM J.Comput.*, 2:225–231, 1973.
12. V. Kann, J. Lagergren, and A. Panconesi. Approximability of maximum splitting of k-sets and some other apx-complete problems. *Information Processing Letters*, 58(3):105–110, 1996.
13. L. Lovasz. Coverings and colorings of hypergraphs. In *Proc. 4th Southeastern Conf. on Combinatorics, Graph Theory, and Computing*, pages 3–12, 1973.

14. K. Mehlhorn. *Data Structures And Algorithms.* Springer Verlag, 1990.
15. E. Petrank. The hardness of approximation: Gap location. *Computational Complexity*, 4:133–157, 1994.
16. H. Zhang and C. X. Ling. An improved learning algorithm for augmented naive Bayes. In *Proc. Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, volume 2035 of *Lecture Notes in Computer Science*, pages 581–586, 2001.
17. U. Zwick. Approximation algorithms for constraint satisfaction problems involving at most three variables per constraint. In *Proc. SODA*, pages 201–220, 1998.
18. U. Zwick. Outward rotations: A tool for rounding solutions of semidefinite programming relaxations, with applications to max cut and other problems. In *Proc. ACM STOC*, pages 679–687, 1999.