# New Lower Bound on Max Cut of Hypergraphs with an Application to $r$-Set Splitting

Archontia C. Giannopoulou[1,⋆], Sudeshna Kolay[2], and Saket Saurabh[2]

[1] National and Kapodistrian University of Athens, Athens, Greece
arcgian@math.uoa.gr
[2] The Institute of Mathematical Sciences, Chennai, India
{skolay,saket}@imsc.res.in

**Abstract.** A classical result by Edwards states that every connected graph $G$ on $n$ vertices and $m$ edges has a cut of size at least $\frac{m}{2}+\frac{n-1}{4}$. We generalize this result to $r$-hypergraphs, with a suitable notion of connectivity that coincides with the notion of connectivity on graphs for $r=2$. More precisely, we show that for every "partition connected" $r$-hypergraph (every hyperedge is of size at most $r$) $H$ over a vertex set $V(H)$, and edge set $E(H)=\{e_1,e_2,\ldots e_m\}$, there always exists a 2-coloring of $V(H)$ with $\{1,-1\}$ such that the number of hyperedges that have a vertex assigned 1 as well as a vertex assigned $-1$ (or get "split") is at least $\mu_H+\frac{n-1}{r2^{r-1}}$. Here $\mu_H=\sum_{i=1}^{m}(1-2/2^{|e_i|})=\sum_{i=1}^{m}(1-2^{1-|e_i|})$. We use our result to show that a version of $r$-Set Splitting, namely, Above Average $r$-Set Splitting (AA-$r$-SS), is fixed parameter tractable (FPT). Observe that a random 2-coloring that sets each vertex of the hypergraph $H$ to 1 or $-1$ with equal probability always splits at least $\mu_H$ hyperedges. In AA-$r$-SS, we are given an $r$-hypergraph $H$ and a positive integer $k$ and the question is whether there exists a 2-coloring of $V(H)$ that splits at least $\mu_H+k$ hyperedges. We give an algorithm for AA-$r$-SS that runs in time $f(k)n^{O(1)}$, showing that it is FPT, even when $r=c_1\log n$, for every fixed constant $c_1<1$. Prior to our work AA-$r$-SS was known to be FPT only for constant $r$. We also complement our algorithmic result by showing that unless NP $\subseteq$ DTIME$(n^{\log\log n})$, AA-$\lceil\log n\rceil$-SS is not in $X$P.

## 1 Introduction

Max Cut is a well known classical problem. Here, the input is a graph $G$ and a positive integer $k$ and the objective is to check whether there is a cut of size at least $k$. A cut of a graph is a bipartition of the vertices of a graph into two disjoint subsets. The size of the cut is the number of edges whose end points are in different subsets of the bipartition. Max Cut is NP-hard and has been the focus of extensive study, from the algorithmic perspective in computer science as well as the extremal perspective in combinatorics. In this paper we focus on

⋆ Supported by a grant of the Special Account for Research Grants of the National and Kapodistrian University of Athens (project code: 70/4/10311).

D. Fernández-Baca (Ed.): LATIN 2012, LNCS 7256, pp. 408–419, 2012.

a generalization of Max Cut to hypergraphs and study this generalization with respect to extremal combinatorics and parameterized complexity.

A *hypergraph* $H$ consists of a vertex set $V(H)$ and a set $E(H)$ of hyperedges. A *hyperedge* $e \in E(H)$ is a subset of the vertex set $V(H)$. By $V(e)$ we denote the subset of vertices corresponding to the edge $e$. A hypergraph is called an $r$*-hypergraph* if the size of each hyperedge is upper bounded by $r$. Given a hypergraph 2-coloring, $\phi: V(H) \rightarrow \{-1, 1\}$, we say that it splits a hyperedge $e$ if $V(e)$ has a vertex assigned 1 as well as a vertex assigned $-1$ under $\phi$. In Max $r$-Set Splitting, a generalization of Max Cut, we are given a hypergraph $H$ and a positive integer $k$ and the objective is to check whether there exists a coloring function $\phi: V(H) \rightarrow \{-1, 1\}$ such that at least $k$ hyperedges are split. This problem is the main topic of this article.

For a graph $G$, let $\zeta(G)$ be the size of a maximum cut. Erdős [9] observed that $\zeta(G) \geq m/2$ for graphs with $m$ edges. To see this notice that a random bipartition of the vertices of a graph $G$ with $m$ edges gives a cut with size at least $m/2$. A natural question was whether the bound on $\zeta$ could be improved. Answering a question of Erdős [9], Edwards [8] proved that for any graph $G$ on $m$ edges $\zeta(G) \geq \left\lceil \frac{m}{2} + \sqrt{\frac{m}{8} + \frac{1}{64}} - \frac{1}{16} \right\rceil$. In the same paper Edwards also showed that for every connected graph $G$ on $n$ vertices and $m$ edges, $\zeta(G) \geq \frac{m}{2} + \frac{n-1}{4}$. These bounds are known to be tight (see [2] for a survey on this area). Our first result generalizes this classical result. For an $r$-hypergraph $H$, let $\zeta(H)$ be the maximum number of edges that can be split by a hypergraph 2-coloring. Let $H$ be a hypergraph with vertex set $V(H)$, and edge set $E(H) = \{e_1, e_2, \ldots e_m\}$. Observe that a random 2-coloring that sets each vertex of hypergraph $H$ to 1 or $-1$ with equal probability always splits at least $\mu_H = \sum_{i=1}^{m}(1 - 2/2^{|e_i|}) = \sum_{i=1}^{m}(1 - 2^{1-|e_i|})$ number of hyperedges. We show that if an $r$-hypergraph $H$ is "partition connected" then $\zeta(H) \geq \mu_H + \frac{n-1}{r2^{r-1}}$.

**Theorem 1.** *Let $H$ be a partition connected $r$-hypergraph with an $n$ sized vertex set $V(H)$, and edge set $E(H) = \{e_1, e_2, \ldots, e_m\}$. Then $\zeta(H) \geq \mu_H + \frac{n-1}{r2^{r-1}}$. Here, $\mu_H = \sum_{i=1}^{m}(1 - 2^{1-|e_i|})$.*

Since the definition of partition connectivity coincides with the definition of connectivity on graphs, for partition connected uniform 2-hypergraphs (every hyperedge has size exactly 2), $\zeta(H) \geq \frac{m}{2} + \frac{n-1}{4}$. The notion of uniform 2-hypergraphs is same as that of ordinary graphs, thus, for $r = 2$, we get the old result of Edwards. Proof of Theorem 1 could also be thought of as a generalization of a similar proof obtained in [3] for ordinary graphs.

We use our combinatorial result to study an above guarantee version of Max $r$-Set Splitting in the realm of parameterized complexity. The goal of parameterized complexity is to find ways of solving NP-hard problems more efficiently than brute force: here the aim is to restrict the combinatorial explosion to a parameter that is hopefully much smaller than the input size. Formally, a *parameterization* of a problem is assigning an integer $k$ to each input instance and we say that a parameterized problem is *fixed-parameter tractable (FPT)* if there is an algorithm that solves the problem in time $f(k) \cdot |I|^{O(1)}$, where $|I|$ is the size of the input and $f$ is an arbitrary computable function depending on the parameter $k$ only. Another notion from parameterized complexity that will be useful to our article is kernelization. A parameterized problem $\Pi$ is said to admit a $g(k)$ *kernel* if there is a polynomial time algorithm that transforms any instance $(x, k)$ to an equivalent instance $(x', k')$ such that $|x'| \leq g(k)$ and $k' \leq g(k)$. If $g(k) = k^{O(1)}$ or $g(k) = O(k)$ we say that $\Pi$ admits a polynomial kernel and linear kernel respectively.

Just as NP-hardness is used as evidence that a problem probably is not polynomial time solvable, there exists a hierarchy of complexity classes above FPT, and showing that a parameterized problem is hard for one of these classes gives evidence that the problem is unlikely to be fixed-parameter tractable. The main classes in this hierarchy are:

$$FPT \subseteq W[1] \subseteq W[2] \subseteq \cdots \subseteq W[P] \subseteq XP$$

The principal analogue of the classical intractability class NP is $W[1]$, which is a strong analogue, because a fundamental problem complete for $W[1]$ is the $k$-Step Halting Problem for Nondeterministic Turing Machines (with unlimited nondeterminism and alphabet size) — this completeness result provides an analogue of Cook's Theorem in classical complexity. In particular this means that an $FPT$ algorithm for any $W[1]$ hard problem would yield a $O(f(k)n^c)$ time algorithm for $k$-Step Halting Problem for Nondeterministic Turing Machines. A convenient source of $W[1]$-hardness reductions is provided by the result that $k$-Clique is complete for $W[1]$. Other highlights of the theory include the fact that $k$-Dominating Set, by contrast, is complete for $W[2]$. $X$P is the class of all problems that are solvable in time $O(n^{g(k)})$. The book by Downey and Fellows [7] provides a good introduction to the topic of parameterized complexity. For recent developments see the books by Flum and Grohe [10] and Niedermeier [20].

Studies on problems parameterized above guaranteed combinatorial bounds are in vogue. A simple example of such a problem is the decision problem that takes as input a planar graph on $n$ vertices and asks if there is an independent set of size at least $\frac{n}{4} + k$. An independent set of size at least $n/4$ is guaranteed by the Four Color Theorem. Could this problem be solved in time $O(n^{g(k)})$, for some function $g$? Is there an FPT algorithm? No one knows. This is a nice and simple example of this research theme, which is quite well-motivated and that has developed strongly since it was introduced by Mahajan and Raman [17]. They showed that several above guarantee versions of Max Cut and Max Sat are FPT. Later, Mahajan et al. [18] published a paper with several new results and open problems around parameterizations beyond guaranteed lower and upper bounds. In a breakthrough paper Gutin et al. [12] developed a probabilistic approach to problems parameterized above or below tight bounds. Alon et al. [1] combined this approach with methods from algebraic combinatorics and Fourier analysis to obtain an FPT algorithm for parameterized Max $r$-SAT beyond the guaranteed lower bound. Other significant results in this direction include quadratic kernels for ternary permutation constraint satisfaction problems parameterized above average and results around systems of linear equations over field with two elements [3,4,13,15].

A standard parameterized version of Max $r$-Set Splitting is defined by asking whether there exists a hypergraph 2-coloring that splits at least $k$ hyperedges. This version of Max $r$-Set Splitting, called $p$-Set Splitting, has been extensively studied in parameterized algorithms. In $p$-Set Splitting we *do not* restrict the size of hyperedges to at most $r$ as in the case of Max $r$-Set Splitting. Dehne, Fellows and Rosamond [6] initiated the study of $p$-Set Splitting and gave an algorithm running in time $O^*(72^k)$ (the $O^*()$ notation suppresses the polynomial factor). After several rounds of improvement the current fastest algorithm is given by Nederlof and van Rooij [19] and runs in time $O^*(1.8213^k)$.

From now onwards we only consider $r$-hypergraphs. If we have a hyperedge of size one then it can never be split and hence we can remove it from consideration. So we assume that every hyperedge is of size at least 2 and at most $r$. Let $H$ be a hypergraph with vertex set $V(H)$, and edge set $E(H) = \{e_1, e_2, \dots e_m\}$. Since every hyperedge is of size at least 2, we have that $\mu_H \geq m/2$. Thus, the standard parameterization of Max $r$-Set Splitting is trivially FPT because of the following argument. If $k \leq m/2$ then the answer is yes else we have that $m \leq 2k$ and hence $n \leq 2kr$. In this case we can enumerate all the $\{1, -1\}$ colorings to $V(H)$ and check whether anyone of them splits at least $k$ hyperedges and answer accordingly. Thus given an $r$-hypergraph $H$, the more meaningful question is whether there exists a $\{1, -1\}$ coloring of $V(H)$ that splits at least $\mu_H + k$ clauses. In other words, we are interested in the following above average version of Max $r$-Set Splitting.

Above Average $r$-Set Splitting (AA-$r$-SS)
*Instance:* An $r$-hypergraph $H$ and a non-negative integer $k$.
*Parameter:* $k$.
*Question:* Does there exist 2-coloring of $V(H)$ that splits at least $\mu_H + k$ hyperedges?

It is known by the results in [15] that AA-$r$-SS is FPT for a constant $r$ ($r = O(1)$). From an algorithmic point of view, a natural question is whether AA-$r$-SS is FPT if the sizes of hyperedges is at most $r(n)$ for some function of $n$. If yes, how far can we push the function $r(n)$? On the algorithmic side, using Theorem 1 we get the following result.

**Theorem 2.** *For every fixed constant* $\alpha < 1$, AA-$\alpha \log n$-SS *is FPT.*

We complement the algorithmic result by a matching lower bound result which states the following.

**Theorem 3.** *Unless* NP $\subseteq$ DTIME$(n^{\log \log n})$, AA-$\lceil \log n \rceil$-SS *is not in X*P.

Theorems 2 and 3 are in sharp contrast to a similar question about AA-Max-$r$-SAT. Let $F$ be a CNF formula on $n$ variables and $m$ clauses and let $r_1, \dots, r_m$ be the number of literals in the clauses of $F$. Then $\text{asat}(F) = \sum_{i=1}^{m}(1 - 2^{-r_i})$

is the expected number of clauses satisfied by a random truth assignment (the truth values to the variables are distributed uniformly and independently). In AA-Max-$r$-SAT we are given a $r$-CNF formula $F$ (all clauses are of size at most $r$) and a positive integer $k$ and the question is whether there is an assignment that satisfies at least $\text{asat}(F) + k$ clauses. Here $k$ is the parameter. In [5], it is shown that AA-Max-$r(n)$-SAT is not FPT unless Exponential Time Hypothesis fails [14], where $r(n) \geq \log\log n + \phi(n)$ and $\phi(n)$ is any unbounded strictly increasing function. However, they also show that Max-$r(n)$-Sat-AA is FPT for any $r(n) \leq \log\log n - \log\log\log n - \phi(n)$, where $\phi(n)$ is any unbounded strictly increasing function.

The proof of Theorem 2 also shows that AA-$r$-SS admits a kernel with $O(k)$ vertices for fixed $r$. Earlier, as per our understanding, only a linear "bikernel" was known [15]. The proofs of Theorem 1 and 2 combine the properties of Fourier coefficients of pseudo-Boolean functions, observed by Crowston et al. [3], with results on a certain kind of connectivity of hypergraphs. The proof of Theorem 3 is inspired by a similar proof given in [5].

## 2 New Lower Bound on $\zeta(H)$ and Proof of Theorem 1

In this section we obtain the new lower bound on $\zeta(H)$, the maximum number of hyperedges that can be split in an $r$-hypergraph $H$ by a hypergraph 2-coloring. Towards this we first define the notion of hypergraph connectivity and hypergraph spanning tree.

**Hypergraph Connectivity and Hypergraph Spanning Tree.** Firstly, for every positive integer $n$, let $[n] = \{1, 2, \ldots, n\}$ and for every set $S$ we denote its powerset by $2^S$. With every hypergraph $H$ we can associate the following graph: The *primal graph*, also called the *Gaifman graph*, $P(H)$ has the same vertices $V(H)$ as $H$ and, two vertices $u, v \in V(H)$ are connected by an edge in $P(H)$ if there is a hyperedge $e \in E(H)$, such that $\{u, v\} \subseteq V(e)$. We say that $H$ is *connected* or has $r$ components if the corresponding primal graph $P(H)$ is connected or has $r$ components. Now we define the notions of *strong cut-sets* and forests in hypergraphs.

**Definition 1 (Strong Cut-Set and Partition Connected).** *A subset $X \subseteq E(H)$ is called a strong cut-set if the hypergraph $H' = (V, E(H) \setminus X)$ has at least $|X| + 2$ connected components. A hypergraph $H$ is partition connected if it does not have a strong cut-set.*

**Definition 2 (Hypergraph Forest).** *A forest $\mathcal{F}$ of a hypergraph $H$ is a pair $(F, g)$ where $F$ is a forest, in the normal graph theoretic sense, with vertex set $V(H)$ and edge set $E(F)$, and $g : E(F) \to E(H)$ is an injective map such that for every $uv \in E(F)$ we have $\{u, v\} \subseteq V(g(uv))$. The number of edges in $\mathcal{F}$ is $|E(F)|$.*

Observe that if a forest $\mathcal{F}$ has $|V(H)| - 1$ edges then $F$ is a spanning tree on $V(H)$. In this case we say that $\mathcal{F}$ is a *hypertree* of $H$. Frank, Király, and Kriesell proved the following duality result relating spanning trees and strong cut-set in hypergraphs [11, Corollary 2.6].

**Proposition 1 ([11]).** *A hypergraph $H$ contains a hypertree if and only if $H$ does not have a strong cut-set.*

A 2-coloring of a hypergraph $H$ is a function $c : V(H) \to \{-1, 1\}$. We say that a hyperedge $e$ of $H$ is split by $c$ if some vertex in $V(e)$ is assigned 1 and some vertex is assigned $-1$. We denote by *split*$(c, H)$ the number of hyperedges split by $c$. The maximum number of hyperedges split over all such 2-colorings is denoted by *split*$(H)$.

**Observation 1.** *Let $H$ be a hypergraph, $e$ be a hyperedge of $H$ and $v \in V(e)$ be a vertex of $H$. If $c$ is a 2-coloring of $H$ then $e$ is not split if and only if $c(v) \cdot c(u) = 1$ for every $u \in V(e) \setminus \{v\}$.*

For every $i \geq 2$, let $m_i$ be the number of hyperedges of $H$ that have size $i$ and for every $r$-hypergraph $H$, we rewrite $\mu_H$ as follows, $\mu_H = \sum_{i=2}^{r}(1 - 2^{-(i-1)})m_i$.

Let $H$ be a hypergraph that does not have a strong cut-set. Here, we will show that for such hypergraphs, there exists a 2-coloring that splits far more than the average. This will be crucial both for our kernelization (Theorem 6) and algorithmic (Theorem 2) results. For this we will also need a result on boolean functions.

**Results from Boolean Functions.** A function $f$ that maps $\{-1, 1\}^n$ to $\mathbb{R}$ is called a *pseudo-boolean* function. It is well known that every pseudo-boolean function $f$ can be uniquely written as

$$f(x_1, \ldots, x_n) = \hat{f}(\emptyset) + \sum_{I \in 2^{[n]} \setminus \emptyset} \hat{f}(I) \prod_{i \in I} x_i,$$

where each $\hat{f}(I)$ is a real. This formula is called the Fourier expansion of $f$ and the $\hat{f}(I)$ are the Fourier coefficients of $f$. See [21] for more details. By $\bar{x}$ we represent $(x_1, \ldots, x_n)$.

**Theorem 4 ([3]).** *Let $f(\bar{x}) = \hat{f}(\emptyset) + \sum_{I \in \mathcal{F}} \hat{f}(I) \prod_{i \in I} x_i$ be a pseudo-boolean function of degree $r > 0$, where $\mathcal{F}$ is a family of non-empty subsets of $[n]$ such that $I \in \mathcal{F}$ if and only if $\hat{f}(I) \neq 0$ and $\hat{f}(\emptyset)$ is the constant term of $f$. Then*

$$\max_{\bar{x} \in \{-1,1\}^n} f(\bar{x}) \geq \hat{f}(\emptyset) + \left\lfloor \frac{rankA - 1 + r}{r} \right\rfloor \cdot \min\{|\hat{f}(I)| \mid I \in \mathcal{F}\},$$

*where $A$ is a $(0, 1)$-matrix with entries $\alpha_{ij}$ such that $\alpha_{ij} = 1$ if and only if term $j$ of the sum contains $x_i$.*

Now we are ready to give the proof of Theorem 1.

*Proof (of Theorem 1).* Let $H$ be an $r$-hypergraph and $1,\ldots,n$ be an arbitrary ordering of vertices in $V(H)$. Let $x_1,\ldots,x_n$ be $n$ variables corresponding to $1,\ldots,n$ respectively. With every hyperedge $e\in E(H)$ we associate a polynomial $f_e(\bar{x})$. For a given $e\in E(H)$, let $j$ be the largest index inside $V(e)$, then

$$f_e(\bar{x}) = 1 - \frac{1}{2^{|e|-1}} \prod_{i\in V(e)\setminus\{j\}} (1+x_ix_j).$$

Notice that for every $\delta=(\delta_1,\delta_2,\ldots,\delta_n)\in\{-1,1\}^n$, we may define a 2-coloring $c_\delta$ of $V(H)$ such that $c_\delta(i)=\delta_i$ and, conversely, for every 2-coloring $c$ we may define a vector $\delta_c\in\{-1,1\}^n$. Observe then that, given a 2-coloring $c$ of $H$, $f_e(\delta_c)=1$ if and only if $e$ is split by $c$. Thus $f_e(\delta_c)=0$ if and only if $e$ is not split by $c$. Hence, it is enough to prove that $\max_{\bar{y}} f(\bar{y}) \geq \mu_H + \frac{n-2}{r\cdot 2^{r-1}}$, where $f(\bar{x}) = \sum_{e\in E(H)} f_e(\bar{x})$ is a pseudo-boolean function of degree $r>0$ and $\bar{y}\in\{-1,1\}^n$. Next we show that it indeed holds.

Let,

$$\begin{aligned} f(\bar{x}) &= \sum_{e\in E(H)} \left(1 - \frac{1}{2^{|e|-1}} \prod_{i\in V(e)\setminus\{j\}} (1+x_ix_j)\right) \\ &= \sum_{i=2}^{r} m_i - \sum_{e\in E(H)} \frac{1}{2^{|e|-1}} \prod_{i\in V(e)\setminus\{j\}} (1+x_ix_j). \end{aligned}$$

Notice, for every $e\in E(H)$, $\frac{1}{2^{|e|-1}}x_px_j$ and $\frac{1}{2^{|e|-1}}x_j^2x_px_q$ appear in the terms of $\prod_{i\in V(e)\setminus\{j\}}(1+x_ix_j)$ for every $\{p,q\}\subseteq V(e)\setminus\{j\}$. We use this fact later. We rewrite $f(\bar{x})$ as,

$$\begin{aligned} f(\bar{x}) &= \sum_{i=2}^{r} m_i - \sum_{i=2}^{r} \frac{1}{2^{i-1}} m_i + \sum_{I\in\mathcal{F}} c_I \prod_{i\in I} x_i^{\lambda_{(I,i)}} \\ &= \sum_{i=2}^{r} \left(1-\frac{1}{2^{i-1}}\right) m_i + \sum_{I\in\mathcal{F}} c_I \prod_{i\in I} x_i^{\lambda_{(I,i)}}, \end{aligned}$$

where $\mathcal{F}$ is a family of subsets of $[n]$ such that for each set $I\in\mathcal{F}$,

1. $2\leq |I| \leq r$,
2. $|c_I| \geq \frac{1}{2^{r-1}}$, and
3. for every $i\in I$, $\lambda_{(I,i)}$ is a positive integer.

Then, as above, for every $e\in E(H)$, $\frac{1}{2^{|e|-1}}x_px_j$ and $\frac{1}{2^{|e|-1}}x_j^2x_px_q$ appear in $f(\bar{x})$ for every $\{p,q\}\subseteq V(e)\setminus\{j\}$.

Let,

$$f_p(\bar{x}) = \sum_{i=2}^{r} \left(1 - \frac{1}{2^{i-1}}\right) m_i + \sum_{I \in \mathcal{F}} c_I \prod_{i \in I} x_i^{\lambda(I,i) \mod 2}.$$

Clearly $f_p : \{-1,1\}^n \to \mathbb{R}$ is a pseudo-boolean function. Then, for every $\bar{x} \in \{-1,1\}^n$, $f(\bar{x}) = f_p(\bar{x})$. Therefore, $\max_{\bar{x}} f(\bar{x}) = \max_{\bar{x}} f_p(\bar{x})$. Notice that $f_p(\bar{x})$ can also be written as

$$f_p(\bar{x}) = \sum_{i=2}^{r} \left(1 - \frac{1}{2^{i-1}}\right) m_i + \sum_{I \in \mathcal{F}'} c'_I \prod_{i \in I} x_i,$$

where $\mathcal{F}'$ is a family of subsets of $[n]$ such that

1. $2 \le |I| \le r$ and
2. $|c'_I| \ge \frac{1}{2^{r-1}}$ for every $I \in \mathcal{F}'$.

Then, for every hyperedge $e \in E(H)$, the term $x_p x_q$, $p, q \in V(e)$ with $p \ne q$ appears in $\sum_{I \in \mathcal{F}'} c'_I \prod_{i \in I} x_i$. Before we proceed we rewrite $f_p(\bar{x})$ as

$$f_p(\bar{x}) = \hat{f}(\emptyset) + \sum_{I \in \mathcal{F}'} \hat{f}(I) \prod_{i \in I} x_i,$$

where $\hat{f}(\emptyset) = \mu_H$ is the constant term of $f_p$ and $\hat{f}(I) = c'_I$, for every $I \in \mathcal{F}'$. Note that $f_p(\bar{x})$ has degree $r_p$ with $2 \le r_p \le r$. From Theorem 4, it follows that

$$\max_{\bar{x}} f_p(\bar{x}) \ge \hat{f}(\emptyset) + \left\lfloor \frac{\mathrm{rank} A - 1 + r_p}{r_p} \right\rfloor \cdot \min\{|\hat{f}(I)| : I \in \mathcal{F}'\},$$

where $A$ is a $(0,1)$-matrix with entries $\alpha_{ij}$ such that $\alpha_{ij} = 1$ if and only if term $j \in I$ contains $x_i$. As $H$ does not contain a strong cut-set, $H$ has a hypertree $T$ (Hypothesis and Proposition 1). Moreover, recall that for every hyperedge $e \in E(H)$, the term $x_p x_q$, $p, q \in V(e)$ with $p \ne q$ appears in $f_p(\bar{x})$. Thus, the edge-vertex incidence matrix of $T$ is a submatrix of $A$. It is known that the edge incidence matrix of a connected graph on $n$ vertices has rank at least $n-1$, thus we have that the rank$T$ is $n-1$. We also know that the rank of a matrix is at least as much as any of its submatrices. This implies that rank$A \ge n-1$ and,

$$\max_{\bar{x}} f_p(\bar{x}) \ge \hat{f}(\emptyset) + \left\lfloor \frac{n-1-1+r}{r} \right\rfloor \cdot \min\{|\hat{f}(I)| \mid I \in \mathcal{F}'\} \ge \mu_H + \frac{n-1}{r \cdot 2^{r-1}}.$$

To see the last inequality let us assume that $n = pr + q$ where $0 \le q \le r-1$. Then if $q \ge 2$ we have that $\lfloor p + \frac{q+r-2}{r} \rfloor \ge p+1$ and this gives the desired result. In other cases we have $q \le 1$ and that gives us that $\lfloor p + \frac{q+r-2}{r} \rfloor \ge p \ge \frac{n-1}{r}$. As $\max_{\bar{x}} f(\bar{x}) = \max_{\bar{x}} f_p(\bar{x})$, this completes the proof by applying Theorem 4. $\square$

## 3 Linear Kernel for Fixed $r$ and Proof of Theorem 2

In this section we combine our results from the previous section with known reduction rules obtained in [16] for $p$-SET SPLITTING to obtain the desired kernel for AA-$r$-SS when $r = O(1)$. Finally, we give the proof of Theorem 2. Towards this we need the notion of *reduction rule*. A reduction rule is a polynomial time algorithm that takes an input instance $(I, k)$ of a problem $\Pi$ and outputs an equivalent instance $(I', k')$ of $\Pi$.

When the hypergraph $H$ is disconnected we can give a simple reduction rule.

REDUCTION RULE 1 ([16]) : *Let $(H, k)$ be an instance of* AA-$r$-SS *such that $P(H)$ has connected components $P(H)[C_1], \dots, P(H)[C_t]$. Let $v_1, \dots, v_t$ be vertices such that $v_i \in C_i$. Construct a hypergraph $H'$ from $H$ by unifying the vertices $v_1, \dots, v_t$. In particular $V(H') = V(H) \setminus \{v_i \mid 2 \le i \le t\}$ and for every hyperedge $e \in E(H)$ make the edge $e' \in E(H')$ where $e' = e$ if $v_i \notin e$ for every $i \in [t]$ and $e' = (V(e) \setminus \{v_i \mid 2 \le i \le t\}) \cup \{v_1\}$ otherwise. We obtain $(H', k)$.*

For a hypergraph $H$ and a coloring $\chi$, let $E(\chi, H)$ denote the set of hyperedges that are split by $\chi$. Our next reduction rule takes care of the case when the hypergraph has a strong cut-set. It is based on the following lemma.

**Theorem 5 ([16]).** *There is a polynomial time algorithm that given a strong cut-set $X$ of a connected hypergraph $H$ finds a cut-set $X' \subseteq X$ such that $X' \neq \emptyset$ and there exists a coloring $\chi$ such that split($\chi$,$H$)=split($H$) and $\chi$ splits all the hyperedges in $X'$. In fact, it shows that given any coloring $c$, there exists a coloring $\chi$ such that $E(\chi, H) = E(c, H) \cup X'$.*

This results in the following reduction rule.

REDUCTION RULE 2 : *Let $(H, k)$ be an instance of* AA-$r$-SS *and $X'$ be a set as defined in Theorem 5. Remove $X'$ from the set of hyperedges and reduce $k$ to $k - \sum_{e \in X'} \frac{1}{2^{|e|-1}}$, that is, obtain an instance $(H', k - \sum_{e \in X'} \frac{1}{2^{|e|-1}})$. Here $E(H') = E(H) \setminus X'$.*

Now we argue the correctness of Reduction Rule 2. Let $(H, k)$ be an instance of AA-$r$-SS and $X'$ be as in the Theorem 5. By Theorem 5 we know that there exists a coloring $\chi$ such that split($\chi$,$H$)=split($H$) and $\chi$ splits all the hyperedges in $X'$. This implies that in $H'$ at least

$$\mu_H + k - |X'| \ge \mu_{H'} + \sum_{e \in X'} \left(1 - \frac{1}{2^{|e|-1}}\right) + k - |X'| \ge \mu_{H'} + k - \sum_{e \in X'} \frac{1}{2^{|e|-1}}$$

hyperedges are split. For the other direction observe that if in $H'$ we have $\mu_{H'} + k - \sum_{e \in X'} \frac{1}{2^{|e|-1}}$ hyperedges split then in $H$ we have $\mu_{H'} + k - \sum_{e \in X'} \frac{1}{2^{|e|-1}} + |X'|$ hyperedges split. The last inequality implies that in $H$, we have $\mu_H + k$ split hyperedges. This proves the correctness of the Reduction Rule 2.

**Theorem 6.** *For a fixed* $r$, AA-$r$-SS *admits a kernel with* $O(k)$ *vertices.*

*Proof.* Let $(H, k)$ be a reduced instance of AA-$r$-SS, that is we cannot apply Reduction Rules 1 and 2. It is important to note that we can find a strong cut-set $X$ of a hypergraph $H$, if it exits, in polynomial time [16]. Thus, we can apply the Reduction Rule 2 in polynomial time. As Reduction Rule 1 does not apply, $H$ is connected. Moreover, as Reduction Rule 2 does not apply $H$ does not have a strong cut-set. From Theorem 1, it follows that if $k \le \frac{n-1}{r \cdot 2^{r-1}}$ then it is a YES-instance. Otherwise, $\frac{n-1}{r \cdot 2^{r-1}} \le k$, thus $n \le r \cdot 2^{r-1}k + 1 = O(k)$. □

*Proof (Proof of Theorem 2).* As in the proof of Theorem 6 we assume that $(H, k)$ is a reduced instance and hence $H$ is partition connected. For the simplicity of an argument choose $\alpha = 1/2$ and thus $r = \log \sqrt{n}$. From Theorem 1, it follows that if $k \le \frac{n-1}{r \cdot 2^{r-1}}$ then it is a YES-instance. Otherwise, $\frac{n-1}{r \cdot 2^{r-1}} \le k$, thus $n \le r \cdot 2^{r-1}k+1$. Substituting $r = \log \sqrt{n}$, we get that $2n \le (\log \sqrt{n})\sqrt{n}k + 1$. This implies that $k \ge n^{\frac{1}{2}-\epsilon}$ for every fixed $\epsilon > 0$. Since we can always solve AA-$r$-SS for any $r$ in time $2^n$, we get that AA-$\alpha \log n$-SS can be solved in time $O^*(2^{k^{\frac{2}{1-\epsilon}}})$. We remark that we could have chosen $\alpha = 1 - \delta$ for any fixed constant $\delta$. □

## 4 Lower Bound Result and Proof of Theorem 3

In this Section we will show that AA-$\lceil \log n \rceil$-SS is not in $X$P unless NP$\subseteq$ DTIME$[n^{\log \log n}]$. Towards this we will give a suitable reduction from $r$-NAE-SAT. A $r$-CNF formula $\phi = C_1 \wedge \cdots \wedge C_m$ is a boolean formula where each clause has size at least 2 and at most $r$ and each clause is a disjunction of literals. $r$-NAE-SAT is a variation of $r$-SAT, where given a $r$-CNF formula $\phi = c_1 \wedge \cdots \wedge c_m$ on $n$ variables, say $V(\phi) = \{x_1, \ldots, x_n\}$, the objective is to find a $\{0, 1\}$ assignment to $V(\phi)$ such that all the clauses get split. An assignment splits a clause if at least one of its literals gets the value 1 and at least one of its literals gets the value 0. We call an assignment that splits every clause a *splitting assignment.*

*Proof (of Theorem 3 (Sketch)).* Set $r = \lceil \log n \rceil + 1$ for the proof. We prove the theorem in three steps. First, we prove that $r$-Nae-Sat is NP-complete for $r = \lceil \log n \rceil + 1$. It is known that $\lceil \log n \rceil$-Sat is NP-complete even when the input has at most $cn$ clauses [5]. We combine this fact to give a reduction from $\lceil \log n \rceil$-Sat to $r$-Nae-Sat that shows NP-completeness of the latter when the input formula to it contains at most $cn$ clauses.

Our second step is to show a many one reduction from $r$-Nae-Sat to $r$-Set-Splitting running in time $O(n^{\log \log n})$.

For the third step we show that if we have an algorithm for AA-$(\lceil \log n \rceil + 1)$-SS running in time $n^{g(k)}$ then we can solve $(\lceil \log n \rceil + 1)$-Nae-Sat with $n$ variables and $\gamma n$ clauses (the number of clauses is linear in number of variables) in $n^{O(\log \log n)}$ time. Notice that if we consider an instance of AA-$(\lceil \log n \rceil + 1)$-SS with $\gamma n$ hyperedges, where $n$ is the number of vertices, then for any random colouring the number of split hyperedges is at least $\left(1 - \frac{1}{2^{\log n}}\right) \cdot \gamma n = \gamma n - \gamma$. Therefore, if AA-($\lceil \log n \rceil + 1$)-SS was in $X$P then by setting $k = \gamma$ we would be able to solve ($\lceil \log n \rceil + 1$)-Set-Splitting in polynomial time. Hence we would obtain that $\lceil \log n \rceil + 1$-Nae-Sat can be solved in $n^{O(\log \log n)}$ time. This is contradictory unless NP $\subseteq$ DTIME($n^{\log \log n}$). $\square$

## 5 Conclusion

In this paper we generalized an old result by Edwards on the size of max-cut on connected graphs to partition connected $r$-hypergraphs. We then used this result to show an above guarantee version of Max $r$-Set Splitting FPT. Our algorithmic results fit well with the current trend of studying problems above guaranteed lower bounds. There are several interesting problems that are still open in parameterized study of problems above guaranteed lower bounds, as well as in the specific directions pursued in this paper. Most notable ones are:

– Does the lower bound of $\mu_H + \frac{n-1}{r2^{r-1}}$ on $\zeta(H)$ for partition connected $r$-hypergraphs tight? That is, is there an infinite family of partition connected $r$-hypergraphs where $\zeta(H) = \mu_H + \frac{n-1}{r2^{r-1}}$.
– Is $\lceil \log n \rceil$-Set-Splitting with linear number of clauses NP-complete?
– Does the question of finding an independent set of size $\frac{n}{4} + k$ on planar graphs FPT? Even obtaining an algorithm in $X$P remains elusive.

## References

1. Alon, N., Gutin, G., Kim, E.J., Szeider, S., Yeo, A.: Solving max-r-sat above a tight lower bound. In: SODA, pp. 511–517 (2010)
2. Bollobás, B., Scott, A.D.: Better bounds for Max Cut. In: Contemporary Combinatorics. Bolyai Soc. Math. Stud., vol. 10, pp. 185–246. János Bolyai Math. Soc., Budapest (2002)
3. Crowston, R., Fellows, M.R., Gutin, G., Jones, M., Rosamond, F.A., Thomassé, S., Yeo, A.: Simultaneously satisfying linear equations over $F_2$: Maxlin2 and max-r-lin2 parameterized above average. In: FSTTCS. LIPIcs, vol. 13, pp. 229–240 (2011)
4. Crowston, R., Gutin, G., Jones, M., Kim, E.J., Ruzsa, I.Z.: Systems of Linear Equations over $\mathbb{F}_2$ and Problems Parameterized above Average. In: Kaplan, H. (ed.) SWAT 2010. LNCS, vol. 6139, pp. 164–175. Springer, Heidelberg (2010)
5. Crowston, R., Gutin, G., Jones, M., Raman, V., Saurabh, S.: Parameterized complexity of maxSat above average. In: Fernández-Baca, D. (ed.) LATIN 2012. LNCS, vol. 7256, pp. 184–194. Springer, Heidelberg (2012)
6. Dehne, F., Fellows, M.R., Rosamond, F.A.: An FPT Algorithm for Set Splitting. In: Bodlaender, H.L. (ed.) WG 2003. LNCS, vol. 2880, pp. 180–191. Springer, Heidelberg (2003)
7. Downey, R.G., Fellows, M.R.: Parameterized Complexity. Springer, New York (1999)
8. Edwards, C.S.: Some extremal properties of bipartite subgraphs. Canad. J. Math. 25, 475–485 (1973)

9. Erdős, P.: On some extremal problems in graph theory. Israel J. Math. 3, 113–116 (1965)
10. Flum, J., Grohe, M.: Parameterized Complexity Theory. Texts in Theoretical Computer Science. An EATCS Series. Springer, Berlin (2006)
11. Frank, A., Király, T., Kriesell, M.: On decomposing a hypergraph into k connected sub-hypergraphs. Discrete Applied Mathematics 131(2), 373–383 (2003)
12. Gutin, G., Kim, E.J., Szeider, S., Yeo, A.: A probabilistic approach to problems parameterized above or below tight bounds. J. Comput. Syst. Sci. 77(2), 422–429 (2011)
13. Gutin, G., van Iersel, L., Mnich, M., Yeo, A.: All Ternary Permutation Constraint Satisfaction Problems Parameterized above Average Have Kernels with Quadratic Numbers of Variables. In: de Berg, M., Meyer, U. (eds.) ESA 2010, Part I. LNCS, vol. 6346, pp. 326–337. Springer, Heidelberg (2010)
14. Impagliazzo, R., Paturi, R., Zane, F.: Which problems have strongly exponential complexity? J. Comput. Syst. Sci. 63(4), 512–530 (2001)
15. Kim, E.J., Williams, R.: Improved Parameterized Algorithms for above Average Constraint Satisfaction. In: Rossmanith, P. (ed.) IPEC 2011. LNCS, vol. 7112, pp. 118–131. Springer, Heidelberg (2012)
16. Lokshtanov, D., Saurabh, S.: Even Faster Algorithm for Set Splitting! In: Chen, J., Fomin, F.V. (eds.) IWPEC 2009. LNCS, vol. 5917, pp. 288–299. Springer, Heidelberg (2009)
17. Mahajan, M., Raman, V.: Parameterizing above guaranteed values: Maxsat and maxcut. J. Algorithms 31(2), 335–354 (1999)
18. Mahajan, M., Raman, V., Sikdar, S.: Parameterizing above or below guaranteed values. J. Comput. Syst. Sci. 75(2), 137–153 (2009)
19. Nederlof, J., van Rooij, J.M.M.: Inclusion/Exclusion Branching for Partial Dominating Set and Set Splitting. In: Raman, V., Saurabh, S. (eds.) IPEC 2010. LNCS, vol. 6478, pp. 204–215. Springer, Heidelberg (2010)
20. Niedermeier, R.: Invitation to Fixed-Parameter Algorithms. Oxford Lecture Series in Mathematics and its Applications, vol. 31. Oxford University Press, Oxford (2006)
21. O'Donnell, R.: Some topics in analysis of boolean functions. In: STOC, pp. 569–578 (2008)