Randomized Approximation for the Set Multicover Problem in Hypergraphs

**Mourad El Ouali · Peter Munstermann ·
Anand Srivastav**

Abstract Let $b \in \mathbb{N}_{\geq 1}$ and let $\mathcal{H} = (V, \mathcal{E})$ be a hypergraph with maximum vertex degree Δ and maximum edge size l. A set b-multicover in $\mathcal{H}$ is a set of edges $C \subseteq \mathcal{E}$ such that every vertex in V belongs to at least b edges in C. SET b-MULTICOVER is the problem of finding a set b-multicover of minimum cardinality, and for $b = 1$ it is the fundamental set cover problem. Peleg et al. (Algorithmica 18(1):44–66, 1997) gave a randomized algorithm achieving an approximation ratio of $\delta \cdot \left(1 - \left(\frac{c}{n}\right)^{\frac{1}{\delta}}\right)$, where $\delta := \Delta - b + 1$ and $c > 0$ is a constant. As this ratio depends on the instance size n and tends to δ as n tends to ∞, it remained an open problem whether an approximation ratio of $\delta\alpha$ with a *constant* $\alpha < 1$ can be proved. In fact, the authors conjectured that for any fixed Δ and b, the problem is not approximable within a ratio smaller than δ, unless $\mathcal{P} = \mathcal{NP}$. We present a randomized algorithm of hybrid type for SET b-MULTICOVER, $b \geq 2$, combining LP-based randomized rounding with greedy repairing, and achieve an approximation ratio of $\delta \cdot \left(1 - \frac{11(\Delta - b)}{72l}\right)$ for hypergraphs with maximum edge size $l \in \mathcal{O}\left(\max\left\{(nb)^{\frac{1}{5}}, n^{\frac{1}{4}}\right\}\right)$. In particular, for all hypergraphs where l is *constant*, we get an $\alpha\delta$-ratio with constant $\alpha < 1$. Hence the above stated conjecture does not hold for hypergraphs with constant l and we have identified the boundedness of the maximum hyperedge size as a relevant parameter responsible for approximations below δ.

M. El Ouali · P. Munstermann (✉) · A. Srivastav
Department of Computer Science, Christian-Albrechts-Universität, 24118 Kiel, Germany
e-mail: pmu@informatik.uni-kiel.de

M. El Ouali
e-mail: meo@informatik.uni-kiel.de

A. Srivastav
e-mail: asr@informatik.uni-kiel.de

1 Introduction

SET b-MULTICOVER is a basic covering problem in combinatorial optimization. It is convenient to formulate the problem in terms of hypergraphs. A hypergraph $\mathcal{H} = (V, \mathcal{E})$ consists of a finite set V and a set $\mathcal{E}$ of subsets of V. We call the elements of V vertices and the elements of $\mathcal{E}$ (hyper-)edges. Further let $n := |V|$, $m := |\mathcal{E}|$. Let l be the maximum edge size and let Δ be the maximum vertex degree, where the degree of a vertex is the number of edges containing that vertex.

For $b \in \mathbb{N}$ a set b-multicover in $\mathcal{H}$ is a set of edges $C \subseteq \mathcal{E}$ such that every vertex in V belongs to at least b edges in C.[1] SET b-MULTICOVER is the problem of finding a set b-multicover of minimum cardinality. For $b = 1$ we have the well-known set cover problem. We denote by Opt the cardinality of a minimum set b-multicover implicitly assuming the dependence of Opt on b.

1.1 Previous Work

While the set cover problem ($b = 1$) has been intensively explored over decades [2,4,6,12,13,16,17], relatively less is known for $b \geq 2$. In the following we give a brief overview of the known approximability results. We say that a polynomial-time algorithm A achieves an approximation ratio $\alpha \geq 1$, or is an α-approximation, if for all instances I it holds that the set b-multicover $A(I)$ returned by A satisfies $|A(I)| / \text{Opt} \leq \alpha$. With a greedy algorithm an approximation ratio of $1 + \ln(l)$ can be obtained for constant l in $\mathcal{O}(nmb)$ time [15,20]. This result was improved by Berman et al. [3], who presented a randomized algorithm that yields an expected approximation ratio of $(1 + o(1)) \ln \left(\frac{l}{b-1} \right)$, when $\frac{l}{b}$ is large enough, and an expected approximation ratio of $1 + 2\sqrt{\frac{l}{b}}$, if $\frac{l}{b}$ is small. Let $\delta := \Delta - b + 1$. An approximation ratio of δ has been achieved in $\mathcal{O}(m \cdot \max\{n, m\})$ time by Hall and Hochbaum [11]. The first approximation below δ has been achieved with randomized rounding by Peleg et al. [18], who showed an approximation ratio of $\delta \cdot \left(1 - \left(\frac{c}{n} \right)^{\frac{1}{\delta}} \right)$, where $c > 0$ is a constant bounded from below roughly by $2^{-2^{50}}$. The essential point of this $\delta \cdot \left(1 - \left(\frac{c}{n} \right)^{\frac{1}{\delta}} \right)$-approximation is that it depends on the instance size n and tends to δ as n grows. The authors conjectured that unless $\mathcal{P} = \mathcal{NP}$, there does not exist a polynomial time approximation algorithm with a *constant* approximation ratio smaller than δ. For $b = 1$ it follows from the work of Khot and Regev [19] that for any arbitrary constant $\epsilon > 0$ an approximation ratio of $\delta - \epsilon = \Delta - \epsilon$ cannot be achieved in polynomial time under the unique games conjecture.

---

[1] We may assume that the minimum vertex degree is at least b, because otherwise the problem has no solution.

So, the challenge is to give an algorithm for SET b-MULTICOVER, $b \geq 2$, with a worst-case ratio of the form $\delta\alpha$, for a *constant* $\alpha < 1$, or to prove that no such algorithm exists, unless some complexity classes collapse. To our best knowledge no such results have been proved since 1997.

1.2 Our Results

We present a hybrid randomized algorithm, combining LP-based randomized rounding and a greedy repairing, if the randomized solution is infeasible. Such a hybrid approach is frequently used in practice. The analysis of such an algorithm is a theoretical challenge and has been successfully carried out for, e.g. maximum graph bisection [8], maximum graph partitioning problems [7,14], the vertex cover and partial vertex cover problem in graphs and hypergraphs [5,6,9,12]. In Theorem 4 we show that our algorithm achieves for hypergraphs with $l = \mathcal{O}\left((nb)^{\frac{1}{5}}\right)$ an approximation ratio of $\delta \cdot \left(1 - \frac{11(\Delta - b)}{72l}\right)$ with constant probability. Using a different technique for the analysis we prove the same approximation ratio again with constant probability for hypergraphs with $l = \mathcal{O}\left(n^{\frac{1}{4}}\right)$ (Theorem 5). Depending on the value of b both results outperform each other, for different ranges of l. While Theorem 4 covers a wider range for l if $b = \Omega(n^{\frac{1}{4}})$, Theorem 5 is the better result for smaller values of b. Together we achieve the approximation ratio stated above for hypergraphs with $l = \mathcal{O}\left(\max\left\{(nb)^{\frac{1}{5}}, n^{\frac{1}{4}}\right\}\right)$.

Note that in both cases our results improve over the ratio of Peleg et al. [18]: for any l satisfying $l = \mathcal{O}\left((\Delta - b) \cdot n^{\frac{1}{\delta}}\right)$ our ratio is at most the ratio of Peleg et al., and it is the better the smaller l is. In the important case of l being a constant, our ratio is $\delta \cdot (1 - c)$, $c \in (0, 1)$ a constant *independent* of n.

This shows that the above stated non-approximability conjecture does not hold for hypergraphs with constant l, including the class of l-uniform hypergraphs (with constant l). The conjecture might be true in general, but according to our result reductions must use edge sizes depending on n.

In conclusion, the crucial algorithmic point of this paper is the identification of the maximum edge size l as a complexity-theoretic relevant hypergraph parameter governing the approximability of SET b-MULTICOVER and leading to approximations below δ in worst case.

The methods developed in this paper go beyond the well-know analysis of *randomized rounding* with Chernoff bounds for sums of independent random variables. The challenge is to combine the different subroutines of the hybrid algorithm in the analysis as the repairing step depends on the randomized cover generation, modeled by sums of *dependent* random variables. For example, we invoke the powerful bounded difference inequality based on the Azuma-Hoeffding martingale bound. Its particular effect is the identification of local hypergraph parameters important for the analysis, in our setting it is the maximum edge size l.

The paper is organized as follows: in Sect. 2 we give definitions and probabilistic tools. In Sect. 3 we present our hybrid randomized algorithm for SET b-MULTICOVER. In Sects. 4 and 5 we analyze the algorithm in two different ways which together provide

the proof for our stated approximation ratio result. Finally in Sect. 6 we sketch some future work.

2 Definitions and Tools

Let $\mathcal{H} = (V, \mathcal{E})$ be a hypergraph, with V and $\mathcal{E}$ being its sets of vertices and hyperedges. For every vertex $v \in V$ we define the vertex degree of v as $d(v) := |\{E \in \mathcal{E} \mid v \in E\}|$. The maximum vertex degree is $\Delta := \max_{v \in V} d(v)$. Let l denote the maximum cardinality of a hyperedge from $\mathcal{E}$. For a set $X \subseteq V$ we denote by $\Gamma(X) := \{E \in \mathcal{E} \mid X \cap E \neq \emptyset\}$ the set of edges incident in X. It is convenient to order the vertices and edges, i.e., $V = \{v_1, \ldots, v_n\}$ and $\mathcal{E} = \{E_1, \ldots, E_m\}$, and to identify the vertices and edges with their indices.

SET b-MULTICOVER:

Let $\mathcal{H} = (V, \mathcal{E})$ be a hypergraph and $b \in \mathbb{N}$. We call $C \subseteq \mathcal{E}$ a set b-multicover if no vertex $i \in V$ is contained in fewer than b hyperedges of C. SET b-MULTICOVER is the problem of finding a set b-multicover with minimum cardinality.

For the one-sided deviation we will use the Chebychev–Cantelli inequality:

Theorem 1 (see [1]) *Let X be a non-negative random variable with finite mean $\mathbb{E}(X)$ and variance $\mathrm{Var}(X)$. Then for any $a > 0$ we have*

$$\Pr(X \geq \mathbb{E}(X) + a) \leq \frac{\mathrm{Var}(X)}{\mathrm{Var}(X) + a^2}.$$

A further useful concentration result is the bounded differences inequality:

Theorem 2 (See [10]) *Let $X = (X_1, X_2, \ldots, X_n)$ be a family of independent random variables with X_k taking values in a set A_k for each k. Suppose that the real-valued function f defined on $A_1 \times \cdots \times A_n$ satisfies $|f(x) - f(x')| \leq c_k$ for every pair of vectors x and x' that differ only in the kth coordinate. Then for any $t > 0$*

$$\Pr\left[f(X) \leq \mathbb{E}(f(X)) - t\right] \leq \exp\left(\frac{-2t^2}{\sum_{k=1}^{n} c_k^2}\right).$$

The following estimate on the variance of a sum of dependent random variables can be proved as in the book of Alon and Spencer:

Lemma 1 (see [1]) *Let X be the sum of n 0/1 random variables, i.e. $X = X_1 + \ldots + X_n$. Let Λ be the set of all unordered pairs $i, j \in \{1, \ldots, n\}$, $i \neq j$ such that X_i and X_j are not independent. Let $\gamma = \sum_{\{i,j\} \in \Lambda} \mathbb{E}(X_i X_j)$, then it holds that $\mathrm{Var}(X) \leq \mathbb{E}(X) + 2\gamma$.*

For a sum of independent random variables we will use the large deviation inequality due to Angluin and Valiant:

Theorem 3 (see [10]) *Let* $X_1, \ldots, X_n$ *be independent* $\{0, 1\}$-*random variables. Let* $X = \sum_{i=1}^n X_i$. *For every* $\beta > 0$ *it holds that*

$$\Pr[X \geq (1 + \beta) \cdot \mathbb{E}(X)] \leq \exp\left(-\frac{\beta^2 \mathbb{E}(X)}{3}\right).$$

3 The Randomized Algorithm

Let $\mathcal{H} = (V, \mathcal{E})$ be a hypergraph and let $(a_{ij})_{(i,j) \in [n] \times [m]}$ be its incidence matrix, where $a_{ij} = 1$ if vertex i is contained in edge j, and $a_{ij} = 0$ otherwise. An integer, linear programming formulation of SET b-MULTICOVER is the following:

$$(b - \text{ILP}) \quad \min \sum_{j=1}^m x_j$$

$$\sum_{j=1}^m a_{ij} x_j \geq b \quad \text{for all } i \in [n]$$

$$x_j \in \{0, 1\} \quad \text{for all } j \in [m].$$

Its linear programming relaxation, denoted by b−LP, is given by relaxing the integrality constraints to $x_j \in [0, 1]$ for all $j \in [m]$. We already defined Opt as the value of an optimal solution for b−ILP. Let Opt* be the value of an optimal solution to b−LP. Clearly, Opt* $\leq$ Opt. Let $x^* = (x_1^*, \ldots, x_m^*)$ be an optimal solution of b−LP. So Opt* $= \sum_{j=1}^m x_j^*$.

Algorithm 1: SET b-MULTICOVER

Input : $b \in \mathbb{N}, \epsilon \in (0, 1)$, a hypergraph $\mathcal{H} = (V, \mathcal{E})$ with maximum degree Δ
Output: A set b-multicover C

1. Initialize $C := \emptyset$ and set $\lambda := (1 - \epsilon)\delta$.
2. Obtain an optimal solution $x^* \in [0, 1]^m$ by solving the LP relaxation.
3. Set $S_0 := \{E_j \in \mathcal{E} \mid x_j^* = 0\}$, $S_\geq := \{E_j \in \mathcal{E} \mid x_j^* \geq \frac{1}{\lambda}\}$
   and $S_< := \{E_j \in \mathcal{E} \mid 0 \neq x_j^* < \frac{1}{\lambda}\}$.
4. Take all edges of $S_\geq$ into the cover C and set $\mathcal{E} = \mathcal{E} \setminus S_0$.
5. (Randomized Rounding) For all edges $E_j \in S_<$ include the edge E_j in
   the cover C, independently for all such E_j, with probability λx_j^*.
6. (Repairing) Repair the cover C (if necessary) as follows: Include arbitrary
   edges from $S_<$, incident to vertices not covered by b edges, to C until all
   vertices are covered by at least b edges.
7. Return the cover C.

Let us briefly explain the ingredients of the algorithm SET b-MULTICOVER. We start with an empty set C, which we extend to a feasible cover. First we solve the LP-relaxation b-LP in polynomial time, say with some known polynomial-time procedure, e.g. the interior point method. Next, we remove the edges with a corresponding LP-variable of value zero, because they will not be taken into the set b-multicover by

randomized rounding. Then randomized rounding is performed followed by a greedy repairing step, if the solution is infeasible, so the resulting cover is always feasible.

4 The First Analysis

Let $X_1, \ldots, X_m$ be $\{0, 1\}$-random variables defined as follows:

$$X_j = \begin{cases} 1 & \text{if the edge } E_j \text{ was picked into the cover before repairing} \\ 0 & \text{otherwise.} \end{cases}$$

Note that the $X_1, \ldots, X_m$ are independent for a given $x^* \in [0, 1]^m$. For all $i \in [n]$ we define the $\{0, 1\}$- random variables Z_i as follows:

$$Z_i = \begin{cases} 1 & \text{if the vertex } v_i \text{ is covered by at least } b \text{ edges before repairing} \\ 0 & \text{otherwise.} \end{cases}$$

Then $Y := \sum_{j=1}^m X_j$ is the cardinality of the cover after randomized rounding and $W := \sum_{i=1}^n Z_i$ is the number of at least b time covered vertices after this step.

The following lemma shows that all vertices are almost covered after step 4 of Algorithm 1.

Lemma 2 Let $0 < \epsilon \leq \frac{1}{4}$ and $b, d, \Delta \in \mathbb{N}$ with $2 \leq b \leq d - 1 \leq \Delta - 1$. Let $\lambda = (1 - \epsilon)\delta$ and let $x_j \in [0, 1]$, $j \in [d]$, such that $\sum_{j=1}^d x_j \geq b$. Then at least $b - 1$ of the x_j fulfill the inequality $x_j \geq \frac{1}{\lambda}$.

Proof Suppose that there exist at most $b-2$ values x_j that fulfill the inequality. W.l.o.g. let $x_{b-1}, \ldots, x_d$ be the variables which do not fulfill it. Then we have

$$\sum_{j=1}^d x_j \leq b - 2 + \sum_{j=b-1}^d x_j < b - 2 + \frac{d - b + 2}{\lambda}$$

$$= b - 2 + \underbrace{\frac{d - b + 2}{\Delta - b + 1}}_{\leq \frac{3}{2}} \cdot \frac{1}{1 - \epsilon} \leq b - 2 + \frac{3}{2} \cdot \frac{4}{3} = b.$$

This contradicts $\sum_{j=1}^d x_j \geq b$. $\qquad\square$

Let us consider a fixed vertex $v \in V$ with $d = deg(v) > b$. By Lemma 2 there do exist at least $b - 1$ edges in $S_\geq$ that contain v. Thus at most one additional edge per vertex is needed to obtain a feasible cover. We are now able to bound the number of additional edges taken into the cover in the repairing step by $n - W$ and receive in total

$$|C| \leq Y + n - W. \tag{1}$$

For the computation of the expectation of W we need the following lemma (see Lemma 2.2, [18]).

Lemma 3 *For all $n \in \mathbb{N}$, $\alpha > 0$ and $x_1, \ldots, x_n, z \in [0, 1]$ with $\sum_{i=1}^{n} x_i \geq z$ and $\alpha x_i < 1$ for all $i \in \mathbb{N}$, we have $\prod_{i=1}^{n}(1 - \alpha x_i) \leq (1 - \alpha \frac{z}{n})^n$, and this bound is the tight maximum.*

The next lemma is an adaption of Lemma 2 in [6] to SET b-MULTICOVER.

Lemma 4 *Let l and Δ be the maximum size of an edge resp. the maximum vertex degree, not necessarily constants. Let $\epsilon > 0$ and $\lambda = (1 - \epsilon)\delta$ as in Algorithm 1.*

  (i) $\mathbb{E}(W) \geq (1 - \epsilon^2)n$.
  (ii) $\text{Opt}^* \geq \frac{nb}{l}$.
  (iii) *If $x_j^* > 0$ for all $j \in [m]$, then $\text{Opt}^* \geq \frac{mb}{\Delta}$.*
  (iv) *If $\lambda \geq 1$, then $\text{Opt}^* \leq \mathbb{E}(Y) \leq \lambda \text{Opt}^*$.*

Proof (i) Let $i \in [n]$, $r = d(i) - b + 1$ and $z_i^* := \sum_{E_j \in (\Gamma(v_i) \setminus S_{\geq})} x_j^*$.

If $|S_{\geq} \cap \Gamma(v_i)| \geq b$, then $\Pr(Z_i = 0) = 0$. Otherwise we get by Lemma 2 $|S_{\geq} \cap \Gamma(v_i)| = b - 1$ and $z_i^* \geq 1$. Therefore

$$
\begin{aligned}
\Pr(Z_i = 0) &= \prod_{E_j \in \Gamma(v_i) \setminus S_{\geq}} \left(1 - \lambda x_j^*\right) \\
&\leq \prod_{E_j \in \Gamma(v_i) \setminus S_{\geq}} \left(1 - \lambda \frac{x_j^*}{z_i^*}\right) \\
&\overset{\text{Lem 3}}{\leq} \left(1 - \frac{\lambda}{r}\right)^r \leq \left(1 - \frac{\lambda}{\delta}\right)^r \\
&= (1 - (1 - \epsilon))^r \leq \epsilon^2.
\end{aligned}
$$

So we get for the expectation of W

$$
\begin{aligned}
\mathbb{E}(W) &= \sum_{i=1}^{n} \Pr(Z_i = 1) = \sum_{i=1}^{n}(1 - \Pr(Z_i = 0)) \\
&\geq \sum_{i=1}^{n} \left(1 - \epsilon^2\right) \\
&= \left(1 - \epsilon^2\right)n.
\end{aligned}
$$

  (ii) Using the ILP constraints we have

$$
nb \leq \sum_{i=1}^{n} \sum_{E_j \in \Gamma(i)} x_j^* = \sum_{j=1}^{m} |E_j| x_j^* \leq l \cdot \sum_{j=1}^{m} x_j^* = l \cdot \text{Opt}^*.
$$

(iii) Let us consider the dual LP of b-LP:

$$(D) \qquad \max \sum_{i=1}^{n} y_i b$$

$$\sum_{i \in E} y_i \leq 1 \ \text{ for all } E \in \mathcal{E},$$

$$y_i \in [0,1] \ \text{ for all } i \in [n].$$

Let $(y_i^*)_{i \in [n]}$ be an optimal solution for (D) with value $\text{Opt}^*(D) = \text{Opt}^*$ (duality theorem). By complementary slackness if $x_j^* > 0$, then $\sum_{i \in E_j} y_i^* = 1$. Therefore

$$mb = \sum_{E \in \mathcal{E}} b \overset{\sum y_i^* = 1}{=} \sum_{E \in \mathcal{E}} \sum_{i \in E} b y_i^* = \sum_{i \in V} d(i) b y_i^* \leq \Delta \sum_{i \in V} b y_i^* = \Delta \cdot \text{Opt}^*.$$

(iv) For $S \subseteq \mathcal{E}$ we set $\text{Opt}^*(S) := \sum_{E_j \in S} x_j^*$. By using the LP relaxation and the definition of the sets $S_{\geq}$ and $S_{<}$, and since $\lambda \geq 1$, we get

$$\text{Opt}^* \leq \underbrace{|S_{\geq}| + \lambda \text{Opt}^*(S_{<})}_{= \mathbb{E}(|C|)} \leq \underbrace{\lambda \text{Opt}^*(S_{\geq})}_{\geq |S_{\geq}|} + \lambda \text{Opt}^*(S_{<}) = \lambda \text{Opt}^*.$$

$\square$

We wish to get an approximation ratio strictly better than δ. Before we continue let us rule out the case where already a trivial solution leads to a good approximation.

Lemma 5 *If there exists a constant $c > 1$ with $\frac{\delta \cdot \text{Opt}^*}{c \cdot bn} > 1$, then it is trivial to compute a set b-multicover of size bn, which is an $\frac{\delta}{c}$-approximation.*

Proof It is always possible to choose for every vertex b arbitrary edges containing it, because the minimum vertex degree is at least b. Hence we need at most bn edges to cover all vertices. Since $bn < \frac{\delta}{c} \cdot \text{Opt}^*$, we have a $\frac{\delta}{c}$-approximation, which is strictly better than a δ-approximation by a constant factor. $\square$

By Lemma 5 we may restrict ourselves to the case where the above trivial situation does not hold. Therefore, we assume from now on

$$1 > \frac{\delta \cdot \text{Opt}^*}{c \cdot bn} \overset{\text{Lem.4}(ii)}{\geq} \frac{\delta}{c \cdot l}. \tag{2}$$

We will use (2) to guarantee non-trivial approximations in the following theorems. For example, in the next theorem we have by (2), $\frac{11(\Delta - b)}{72l} \in [0,1]$ taking $c = 72/11$, so $1 - \frac{11(\Delta - b)}{72l} \geq 0$.

Theorem 4 *Let $b \in \mathbb{N}_{\geq 2}$ and let $\mathcal{H}$ be a hypergraph with maximum vertex degree Δ and maximum edge size l, $3 \leq l \leq \underbrace{2^{-1/4} \cdot 3^{-1}}_{\approx 0.28} \cdot (nb)^{\frac{1}{5}}$.*

Algorithm 1 returns a set b-multicover of size

$$|C| \leq \left(1 - \frac{11(\Delta - b)}{72l}\right) \delta \cdot \text{Opt}^*$$

with probability at least $1 - (\exp(-2) + \exp(-6)) \approx 0.862$ *in polynomial time.*

Proof We choose

$$\epsilon := \frac{\delta \cdot \text{Opt}^*(1 + \beta)}{6bn} \quad \text{with} \quad \beta = \frac{\sqrt{2}l}{\sqrt{nb}}. \tag{3}$$

For $l < \frac{(nb)^{\frac{1}{2}}}{4\sqrt{2}}$ we have $\beta < \frac{1}{4}$. By the assumption on l in Theorem 4 this upper bound on l is satisfied. Further let us introduce the constants $c_1 := \frac{45}{8}$ and $c_2 := \frac{6}{c_1} > 1$. We can assume $\frac{\delta \cdot \text{Opt}^*}{c_2 \cdot bn} \leq 1$ (Lemma 5). Now

$$\epsilon = \frac{\delta \cdot \text{Opt}^*(1 + \beta)}{6bn} = \frac{\delta \cdot \text{Opt}^*(1 + \beta)}{c_1 c_2 bn} \leq \frac{1 + \beta}{c_1} < \frac{5}{4c_1} = \frac{4}{18} < \frac{1}{4}. \tag{4}$$

SET b-MULTICOVER is trivial or even unsolvable for $\Delta \leq b$. So we may assume $\Delta \geq b + 1$, hence $\lambda = (1 - \epsilon)\delta \geq (1 - \epsilon)2 > (1 - \frac{4}{18})2 > 1$, which is important for using Lemma 4 later on.

First we focus on two statements that will help us bounding $|C|$ with a constant probability.

Claim 1

$$\Pr\left(W \leq n(1 - \epsilon^2) - \sqrt{\delta \cdot \sum_{k=1}^{m} |E_k|^2}\right) \leq \exp(-6) \tag{5}$$

Proof of Claim 1 Consider the function $f(X_1, \ldots, X_m) := \sum_{i=1}^{n} Z_i$. Then we have for any two vectors $x = (x_1, \ldots, x_k, \ldots, x_n)$ and $x' = (x_1, \ldots, x_k', \ldots, x_n)$ that only differ in the k-th coordinate

$$\left|f(x_1, \ldots, x_k, \ldots, x_n) - f(x_1, \ldots, x_k', \ldots, x_n)\right| \leq |E_k|$$

for all $k \in [m]$. Therefore we can use Theorem 2 with $c_k = |E_k|$ for all $k \in [m]$ and $t = \sqrt{\delta \cdot \sum_{k=1}^{m} |E_k|}$ and receive

$$\Pr\left(W \leq n(1 - \epsilon^2) - \sqrt{\delta \cdot \sum_{k=1}^{m} |E_k|^2}\right) \leq \Pr\left(W \leq \mathbb{E}(W) - \sqrt{\delta \cdot \sum_{k=1}^{m} |E_k|^2}\right)$$

$$\leq \exp\left(\frac{-2\delta \cdot \sum_{k=1}^{m} |E_k|^2}{\sum_{k=1}^{m} |E_k|^2}\right)$$

$$= \exp(-2\delta)$$

$$\leq \exp(-6).$$

Claim 2 For $\beta = \frac{\sqrt{2l}}{\sqrt{nb}}$ it holds that

$$\Pr\left(Y \geq \lambda(1+\beta)\mathrm{Opt}^*\right) \leq \exp(-2) \tag{6}$$

Proof of Claim 2 To show the following inequalities we invoke Lemma 4 and Theorem 3:

$$
\begin{aligned}
\Pr\left(Y \geq \lambda(1+\beta)\mathrm{Opt}^*\right)
&\overset{\text{Lem 4}(iv)}{\leq} \Pr(Y \geq \mathbb{E}(Y)(1+\beta)) \\
&\overset{\text{Th 3}}{\leq} \exp\left(-\frac{\beta^2 \mathbb{E}(Y)}{3}\right) \\
&\overset{\text{Lem 4}(iv)}{\leq} \exp\left(-\frac{\beta^2 \mathrm{Opt}^*}{3}\right) \\
&\overset{\text{Lem 4}(ii)}{\leq} \exp\left(-\frac{nb\beta^2}{3l}\right) \\
&\overset{\text{Def.}\beta}{=} \exp\left(-\frac{2l}{3}\right) \\
&\overset{l\geq 3}{\leq} \exp(-2).
\end{aligned}
$$

With Claims 1 and 2 we have with probability at least $1 - (\exp(-2) + \exp(-6)) \approx 0.862$:

$$
|C| \overset{(1)}{\leq} Y + n - W
$$

$$
\leq \underbrace{\lambda(1+\beta)\mathrm{Opt}^* + n\epsilon^2}_{(*)} + \underbrace{\sqrt{\delta \cdot \sum_{k=1}^{m} |E_k|^2}}_{(**)}.
$$

Next we bound $(*)$ and $(**)$ separately and combine their bounds later. We start with $(*)$:

$$
\begin{aligned}
(*) &= \delta(1+\beta)(1-\epsilon)\mathrm{Opt}^* + n\epsilon^2 \\
&\overset{\text{Def.}\epsilon}{=} \left((1+\beta)(1-\epsilon) + \frac{\delta \cdot \mathrm{Opt}^*(1+\beta)^2}{36nb^2}\right) \delta \cdot \mathrm{Opt}^* \\
&\overset{\text{Def.}\epsilon}{=} \left((1+\beta)\left(1 - \frac{\delta \cdot \mathrm{Opt}^*(1+\beta)}{6nb}\right) + \frac{\delta \cdot \mathrm{Opt}^*(1+\beta)^2}{36nb^2}\right) \delta \cdot \mathrm{Opt}^* \\
&= \left((1+\beta) - \frac{\delta \cdot \mathrm{Opt}^*(1+\beta)^2}{6nb} + \frac{\delta \cdot \mathrm{Opt}^*(1+\beta)^2}{36nb^2}\right) \delta \cdot \mathrm{Opt}^* \\
&= \left(1 - \left(\frac{(6b-1)\delta(1+\beta)^2\mathrm{Opt}^*}{36nb^2} - \beta\right)\right) \delta \cdot \mathrm{Opt}^*
\end{aligned}
$$

$$\overset{\text{Lem.4}(ii)}{\leq} \left(1 - \frac{(6b-1)\delta(1+\beta)^2 - 36\beta lb}{36lb}\right)\delta \cdot \text{Opt}^*$$

$$= \left(1 - \frac{(6-\frac{1}{b})\delta(1+\beta)^2 - 36\beta l}{36l}\right)\delta \cdot \text{Opt}^*$$

$$\overset{\text{Def.}\beta}{=} \left(1 - \frac{\delta}{36l}\left((6-\frac{1}{b})(1+\beta)^2 - \frac{36\sqrt{2}l^2}{\delta\sqrt{nb}}\right)\right)\delta \cdot \text{Opt}^*$$

$$\overset{b\geq 2}{\leq} \left(1 - \frac{\delta}{36l}\left(\frac{11}{2}(1+\beta)^2 - \frac{36\sqrt{2}l^2}{\delta\sqrt{nb}}\right)\right)\delta \cdot \text{Opt}^*$$

$$\overset{\text{bound on } l}{\leq} \left(1 - \frac{\delta}{36l}\left(\frac{11}{2}(1+\beta)^2 - \frac{8}{2\delta}\right)\right)\delta \cdot \text{Opt}^*$$

$$\leq \left(1 - \frac{\delta}{36l}\left(\frac{11}{2} - \frac{8}{2\delta}\right)\right)\delta \cdot \text{Opt}^*$$

$$= \left(1 - \frac{11\delta - 8}{72l}\right)\delta \cdot \text{Opt}^*.$$

Now we continue with $(**)$. Note that

$$\frac{b}{\Delta}(\Delta - b + 1) = \left(\frac{\Delta - b}{\Delta}\right)b + \frac{b}{\Delta} \geq \frac{\Delta - b}{\Delta} + \frac{b}{\Delta} = 1.$$

From this we can immediately deduce

$$\frac{m}{\delta} \overset{\text{Def. }\delta}{=} \frac{m}{\Delta - b + 1} \leq \frac{mb}{\Delta}. \tag{7}$$

We make use of this in the following calculation:

$$(**) = \sqrt{\delta \cdot \sum_{k=1}^{m} |E_k|^2}$$

$$\leq \sqrt{\delta m l^2} = l\sqrt{m}\frac{\delta}{\sqrt{\delta}}$$

$$\overset{(7)}{\leq} l\sqrt{\frac{mb}{\Delta}}\delta$$

$$\overset{\text{Lem.4}(iii)}{\leq} l\sqrt{\text{Opt}^*} \cdot \delta$$

$$= \frac{l^{\frac{5}{2}}}{l\sqrt{l}}\sqrt{\text{Opt}^*} \cdot \delta$$

$$\overset{\text{bound on } l}{\leq} \frac{3\sqrt{nb}\sqrt{\text{Opt}^*}}{72 \cdot l\sqrt{l}}\delta$$

$$\overset{\text{Lem.4}(ii)}{\leq} \frac{3}{72l}\delta \cdot \text{Opt}^*.$$

Together we get

$$
\begin{aligned}
|C| &\leq (*) + (**) \\
&\leq \left(1 - \frac{11\delta - 8}{72l}\right)\delta \cdot \mathrm{Opt}^* + \frac{3}{72l}\delta \cdot \mathrm{Opt}^* \\
&= \left(1 - \frac{11\delta - 11}{72l}\right)\delta \cdot \mathrm{Opt}^* \\
&= \left(1 - \frac{11(\Delta - b)}{72l}\right)\delta \cdot \mathrm{Opt}^*.
\end{aligned}
$$

$\square$

5 An Alternative Analysis

In this section we estimate the expected size of the set b-multicover after repairing. This is a non-trivial new aspect, because we have to simultaneously couple the outcomes of randomized rounding and repairing in the expectation. After estimating the expectation, we will use concentration inequalities, leading to the same approximation ratio we accomplished in the last section, but for hypergraphs with $l = \mathcal{O}(n^{\frac{1}{4}})$. Together with Theorem 4 we achieve the claimed ratio for $l \in \mathcal{O}\left(\max\{(nb)^{\frac{1}{5}}, n^{\frac{1}{4}}\}\right)$ as desired.

First we compute the expectation and the variance of the cardinality of the set b-multicover returned by Algorithm 1. We keep in mind that we do not consider instances where a trivial cover with nb edges yields a good approximation (Lemma 5).

Lemma 6 *Let $\mathcal{H}$ be a hypergraph with maximum vertex degree Δ and maximum edge size $l \geq 3$. Let $\epsilon = \frac{\delta \cdot \mathrm{Opt}^*}{6bn}$. Then Algorithm 1 returns a set b-multicover C with*

$$
\mathbb{E}(|C|) \leq \delta \cdot \left(1 - \frac{(6 - 1/b)\delta}{36l}\right)\mathrm{Opt}^*.
$$

Proof By Lemma 5 we can assume that

$$
\frac{\delta \mathrm{Opt}^*}{6bn} \in [0, \frac{1}{4}].
$$

Therefore it holds that $\lambda = \delta\left(1 - \frac{\delta \mathrm{Opt}^*}{6bn}\right) \geq \frac{3}{4}\delta > 1$.
Now we can bound the expectation by

$$
\begin{aligned}
\mathbb{E}(|C|) &\overset{(1)}{\leq} \mathbb{E}(Y) + n - \mathbb{E}(W) \\
&\overset{\mathrm{Lem.4}}{\leq} \lambda \mathrm{Opt}^* + \epsilon^2 n \\
&= \delta(1 - \epsilon)\mathrm{Opt}^* + \epsilon^2 n
\end{aligned}
$$

$$
\overset{\text{Def.}\epsilon}{=} \quad \delta \cdot \left(1 - \frac{\delta \text{Opt}^*}{6bn} + \frac{\delta \text{Opt}^*}{36b^2n} \right) \text{Opt}^*
$$

$$
= \quad \delta \cdot \left(1 - \frac{(6b-1)\delta \text{Opt}^*}{36b^2n} \right) \text{Opt}^*
$$

$$
\overset{\text{Lem.4}(ii)}{\leq} \quad \delta \cdot \left(1 - \frac{(6 - 1/b)\delta}{36l} \right) \text{Opt}^*.
$$

$\square$

The variance is estimated as follows.

Lemma 7 $\text{Var}(|C|) \leq \delta l \mathbb{E}(|C|)$.

Proof For $j \in [m]$ let $\hat{X}_j = 1$ if and only if the edge E_j is contained in the set b-multicover after repairing. Thus, it holds that $|C| = \sum_{j=1}^{m} \hat{X}_j$. Further let Λ, γ be defined as in Lemma 1. As $\hat{X}_j = 1$ for all $j \in S_\geq$, $\hat{X}_j$ is a non-trivial variable only for $j \in S_<$, and we may restrict us to such j. Let $i, j \in S_<$. For any pair of edges from $S_<$ we have

$$
\mathbb{E}(\hat{X}_i \hat{X}_j) \leq \min\{\Pr(\hat{X}_i = 1), \Pr(\hat{X}_j = 1)\} \tag{8}
$$
$$
\leq \frac{\Pr(\hat{X}_i = 1) + \Pr(\hat{X}_j = 1)}{2}.
$$

According to Lemma 2, there exist for every vertex at least $b - 1$ edges from $S_\geq$ containing it. Consequently, the maximum vertex degree in the subhypergraph induced by $S_<$, which is considered for randomized rounding and repairing, is at most $\Delta - (b - 1) = \delta$. So for a fixed edge $E_i \in S_<$

$$
\sum_{j \in E_i} \left| \Gamma(\{v_j\}) \cap S_< \right| - 1 \leq \sum_{j \in E_i} (\delta - 1) \leq l(\delta - 1)
$$

random variables $\hat{X}_j$ depend on $\hat{X}_i$. Now

$$
\gamma = \sum_{\{E_i, E_j\} \in \Lambda} \mathbb{E}(\hat{X}_i \hat{X}_j) \overset{(8)}{\leq} \sum_{\{E_i, E_j\} \in \Lambda} \frac{\Pr(\hat{X}_i = 1) + \Pr(\hat{X}_j = 1)}{2}
$$
$$
< \sum_{i=1}^{m} \frac{\sum_{j \in E_i} \left| \Gamma(\{v_j\}) \cap S_< \right| - 1}{2} \Pr(\hat{X}_i = 1) \leq \frac{l(\delta - 1)}{2} \mathbb{E}(|C|).
$$

By Lemma 1 we have:

$$
\text{Var}(|C|) = \text{Var}(|C \setminus S_\geq|) \leq \mathbb{E}(|C \setminus S_\geq|) + l(\delta - 1)\mathbb{E}(|C|) \leq l\delta \mathbb{E}(|C|).
$$

$\square$

Finally we use the statements about expectation and variance to obtain our second main result. Note that due to Lemma 5 and (2), $\frac{11(\Delta-b)}{72l} \in [0, 1]$, so the approximation factor in the following theorem is non-negative.

Theorem 5 *Let $b \in \mathbb{N}$, $b \geq 2$, let $\mathcal{H}$ be a hypergraph with maximum edge size $l \leq \sqrt{\frac{11}{72}}n^{\frac{1}{4}}$ and maximum vertex degree Δ and let $\epsilon = \frac{\delta \cdot Opt^*}{6bn}$. Algorithm 1 returns a set b-multicover C of size*

$$|C| \leq \delta \left(1 - \frac{11(\Delta - b)}{72l}\right) Opt^*$$

with probability at least $1 - \frac{1}{1+b} \geq 2/3$ in polynomial time.

Proof Let $\mathcal{A}$ be the event $|C| > \delta \left(1 - \frac{11(\Delta-b)}{72l}\right) Opt^*$. To prove the claimed upper bound for $|C|$, we estimate the concentration of $|C|$ around its expectation. This can be done as follows:

$$
\begin{aligned}
\Pr\left(\mathcal{A}\right) &= \Pr\left(|C| > \delta\left(1 - \frac{11\delta}{72l}\right)Opt^* + \frac{11\delta Opt^*}{72l}\right) \\
&\overset{b\geq 2}{\leq} \Pr\left(|C| > \delta\left(1 - \frac{(6-1/b)\delta}{36l}\right)Opt^* + \frac{11\delta Opt^*}{72l}\right) \\
&\overset{Lem\ 6}{\leq} \Pr\left(|C| > \mathbb{E}(|C|) + \frac{11\delta Opt^*}{72l}\right) \\
&\overset{Th\ 1}{\leq} \frac{Var(|C|)}{Var(|C|) + \left(\frac{11\delta Opt^*}{72l}\right)^2} = \frac{1}{1 + \frac{\left(\frac{11\delta Opt^*}{72l}\right)^2}{Var(|C|)}}.
\end{aligned}
$$

We may continue,

$$\frac{\left(\frac{11\delta Opt^*}{72l}\right)^2}{Var(|C|)} \overset{Lem\ 7}{\geq} \left(\delta\frac{121(Opt^*)^2}{(72)^2l^3\mathbb{E}(|C|)}\right) \overset{\mathbb{E}(|C|)\leq \delta Opt^*}{\geq} \frac{121 \cdot Opt^*}{(72)^2l^3} \overset{Lem.4(iii)}{\geq} \frac{121bn}{(72)^2l^4},$$

since $l \leq \sqrt{\frac{11}{72}}n^{\frac{1}{4}}$ we have $\frac{121bn}{(72)^2l^4} \geq b$.

Therefore we get $\Pr\left(|C| \geq l\left(1 - \frac{11(\Delta-b)}{72l}\right)Opt^*\right) \leq \frac{1}{1+b} \leq \frac{1}{3}$. $\qquad\square$

6 Future Work

An interesting problem is the derandomization of our algorithms, where the challenge comes from the fact that we have a hybrid algorithm. Furthermore, attempts should be made to prove (or disprove) the non-approximability conjecture of Peleg et al. [18]. Our work indicates that hypergraphs where l is a function of n would be helpful for this task. Another interesting problem is the PARTIAL SET MULTICOVER, where only a subset of vertices needs to be covered.

References

1. Alon, N., Spencer, J.: The Probabilistic Method, 2nd edn. Wiley, New York (2000)
2. Bar-Yehuda, R.: Using homogeneous weights for approximating the partial cover problem. J. Algorithms **39**(2), 137–144 (2001)
3. Berman, P., DasGupta, B., Sontag, E.: Randomized approximation algorithms for set multicover problems with applications to reverse engineering of protein and gene networks. Discrete Appl. Math. **155**(6–7), 733–749 (2007)
4. Chvátal, V.: A greedy heuristic for the set covering problem. Math. Oper. Res. **4**(3), 233–235 (1979)
5. El Ouali, M., Fohlin, H., Srivastav, A.: An Approximation Algorithm for the Partial Vertex Cover Problem in Hypergraphs. J. Comb. Optim. (2014). doi:10.1007/s10878-014-9793-2
6. El Ouali, M., Fohlin, H., Srivastav, A.: A randomised approximation algorithm for the hitting set problem. Theor. Comput. Sci. **555**, 23–34 (2014)
7. Feige, U., Langberg, M.: Approximation algorithms for maximization problems arising in graph partitioning. J. Algorithms **41**(2), 174–201 (2001)
8. Frieze, A., Jerrum, M.: Improved Approximation Algorithms for MAX k-CUT and MAX BISECTION. Algorithmica **18**, 67–81 (1997)
9. Gandhi, R., Khuller, S., Srinivasan, A.: Approximation algorithms for partial covering problems. J. Algorithms **53**(1), 55–84 (2004)
10. Habib, M., McDiarmid, C., Ramirez-Alfonsin, J., Reed, B.: Probabilistic Methods for Algorithmic Discrete Mathematics, pp. 195–248. Springer Berlin Heidelberg (1998)
11. Hall, N.G., Hochbaum, D.S.: A fast approximation algorithm for the multicovering problem. Discrete Appl. Math. **15**, 35–40 (1986)
12. Halperin, E.: Improved approximation algorithms for the vertex cover problem in graphs and hypergraphs. SIAM J. Comput. **31**(5), 1608–1623 (2002)
13. Hochbaum, D.S.: Approximation algorithms for the set covering and vertex cover problems. SIAM J. Comput. **11**(3), 555–556 (1982)
14. Jäger, G., Srivastav, A.: Improved approximation algorithms for maximum graph partitioning problems. J. Comb. Optim. **10**(2), 133–167 (2005)
15. Johnson, D.S.: Approximation algorithms for combinatorial problems. J. Comput. Syst. Sci. **9**, 256–278 (1974)
16. Krivelevich, J.: Approximate set covering in uniform hypergraphs. J. Algorithms **25**(1), 118–143 (1997)
17. Lovász, L.: On the ratio of optimal integral and fractional covers. Discrete Math. **13**(4), 383–390 (1975)
18. Peleg, D., Schechtman, G., Wool, A.: Randomized approximation of bounded multicovering problems. Algorithmica **18**(1), 44–66 (1997)
19. Khot, S., Regev, O.: Vertex cover might be hard to approximate to within 2-epsilon. J. Comput. Syst. Sci. **74**(3), 335–349 (2008)
20. Vazirani, V.V.: Approximation Algorithms. Springer, Berlin (2003)