A Randomised Approximation Algorithm for the Hitting Set Problem

Mourad El Ouali[1], Helena Fohlin[2], and Anand Srivastav[1]

[1] Departement of Computer Science. University of Kiel. Germany
{meo,asr}@informatik.uni-kiel.de
[2] Department of Clinical and Experimental Medicine. Linköping University, Sweden
Helena.Fohlin@lio.se

Abstract. Let $\mathcal{H} = (V, \mathcal{E})$ be a hypergraph with vertex set V and edge set $\mathcal{E}$, where $n := |V|$ and $m := |\mathcal{E}|$. Let l be the maximum size of an edge and Δ be the maximum vertex degree. A hitting set (or vertex cover) in $\mathcal{H}$ is a set of vertices from V in which all edges are incident. The hitting set problem is to find a hitting set of minimum cardinality. It is known that an approximation ratio of l can be achieved easily. On the other side, for constant l, an approximation ratio better than l cannot be achieved in polynomial time under the unique games conjecture (Khot and Ragev 2008). Thus breaking the l-barrier for significant classes of hypergraphs is a complexity-theoretic and algorithmically interesting problem, which has been studied by several authors (Krivelevich (1997), Halperin (2000), Okun (2005)). We propose a randomised algorithm of hybrid type for the hitting set problem, which combines LP-based randomised rounding, graphs sparsening and greedy repairing and analyse it in different environments. For hypergraphs with $\Delta = O(n^{\frac{1}{4}})$ and $l = O(\sqrt{n})$ we achieve an approximation ratio of $l\left(1 - \frac{c}{\Delta}\right)$, for some constant $c > 0$, with constant probability. In the case of l-uniform hypergraphs, l and Δ being constants, we prove by analysing the expected size of the hitting set and using concentration inequalities, a ratio of $l\left(1 - \frac{l-1}{4\Delta}\right)$. Moreover, for quasi-regularisable hypergraphs, we achieve an approximation ratio of $l\left(1 - \frac{n}{8m}\right)$. We show how and when our results improve over the results of Krivelevich, Halperin and Okun.

Keywords: Approximation algorithms, probabilistic methods, randomised rounding, hitting set, vertex cover, greedy.

1 Introduction

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 $n := |V|$ and $m := |\mathcal{E}|$. A hitting set (or vertex cover) of a hypergraph $\mathcal{H}$ is a set C of vertices such that for every $E \in \mathcal{E}$ there exists a vertex $v \in E \cap C$. The hitting set problem in hypergraphs is the task of finding a hitting set of minimum cardinality. A set $\mathcal{S} \subseteq \mathcal{E}$ is called a set cover, if all vertices of $\mathcal{H}$ are contained in edges of $\mathcal{S}$, and the set cover problem is to find a set cover of minimum

S.K. Ghosh and T. Tokuyama (Eds.): WALCOM 2013, LNCS 7748, pp. 101–113, 2013.

cardinality. Note that the hitting set problem in hypergraphs is equivalent to the set cover problem by changing the role of vertices and edges.

A number of inapproximability results are known. Lund and Yannakakis [20] proved for the set cover problem that for any $\alpha < \frac{1}{4}$, the existence of a polynomial-time $(\alpha \ln n)$-ratio approximation algorithm would imply that $\mathcal{NP}$ has a quasipolynomial, i.e., $n^{\mathcal{O}(\text{poly}(\ln n))}$ deterministic algorithm. This result was improved to $(1 - o(1)) \ln n$ by Feige [7]. A $c \cdot \ln n$-approximation under the assumption that $\mathcal{P} \neq \mathcal{NP}$ was established by Safra and Raz [24], where c is a constant. A similar result for larger values of c was proved by Alon, Moshkovitz and Safra [1].

The hitting set problem remains hard for many hypergraph classes. Most interesting are l-uniform hypergraphs with a constant l, because for them under the unique games conjecture (UGC), it is $\mathcal{NP}$-hard to approximate the problem within a factor of $l - \epsilon$, for any *fixed* $\epsilon > 0$, see [17], while an approximation ratio of l can be easily achieved by finding a maximal matching. Therefore, the problem of breaking the l-barrier for significant and interesting classes of hypergraphs received much attention.

Let us briefly give an overview of the known approximability results for the problem. The earliest published approximation algorithms for the hitting set problem achieve an approximation ratio of the order $\ln m + 1$ [6,16,19] by using a greedy heuristic. For l-uniform hypergraphs, several authors achieved the ratio of l using different techniques (see e.g. [3,11,13,14]). The first and important result breaking the barrier of l for l-uniform hypergraphs, is due to Krivelevich [18]. He proved an approximation ratio of $l(1 - cn^{\frac{1-l}{l}})$, for some constant $c > 0$, using a combination of the LP-based algorithm and the local ratio approach described by Bar-Yehuda and Even [4]. Later, for l-uniform hypergraphs with $l^3 = o(\frac{\ln \ln n}{\ln \ln \ln n})$ and $\Delta = O(n^{l-1})$, Halperin [12] presented a semidefinite programming based algorithm with an approximation ratio of $l - (1 - o(1)) \frac{l \ln \ln n}{\ln n}$. Note that this condition enforces the doubly exponential bound, $n \geq 2^{2^{l^2}}$, and already for $l = 3$ the hypergraph is very large and is hardly suitable for practical purposes.

A further important class consists of hypergraphs with Δ and l being constants. In this case Krivelevich [18] gave an LP-based algorithm that provides an approximation ratio of $l(1 - c\Delta^{\frac{1}{1-l}})$ for some constant $c > 0$. An improved approximation ratio of $l - (1 - o(1)) \frac{l(l-1) \ln \ln \Delta}{\ln \Delta}$ was presented by Halperin [12], provided that $l^3 = o(\frac{\ln \ln \Delta}{\ln \ln \ln \Delta})$. For hypergraphs which are not necessarily uniform, but with size of edges bounded from above by a constant l, an improvement of the result of Krivelevich was given by Okun [23]. He proved an approximation ratio of $l(1 - c(\beta, l)\Delta^{-\frac{1}{\beta l}})$ for $\beta \in (0, 1)$ and a constant $c(\beta, l) \in (0, 1)$ depending on β and l, by a modification of the algorithm presented in [18].

Our Results. We consider hypergraphs with maximum edge size l and maximum vertex degree Δ, at the moment not necessarily assumed to be constants. In Section 3 we present a randomised algorithm, combining LP-based randomised rounding, sparsening of the hypergraph and greedy repairing. Such a hybrid approach is frequently used in practice and it has been analysed for many problems, e.g., maximum graph bisection [9], maximum graph partitioning problems [8,15] and the vertex cover and partial vertex cover problem in graphs [11,12]. In Section 4.1 we show that our algorithm achieves for $l = O(\sqrt{n})$ and $\Delta = O(n^{\frac{1}{4}})$ an approximation ratio of $l\left(1 - \frac{c}{\Delta}\right)$, for some constant $c > 0$, with constant probability. In this case our result improves the result of Krivelevich, for any function $f(n)$ satisfying $f(n) = O(n^{\frac{1}{4}})$, since $n^{\frac{1}{4}} < n^{1-\frac{1}{l}}$ for $l \geq 2$, and the approximation is the better the smaller $f(n)$ becomes. For $\Delta \leq \frac{\ln n}{\ln \ln n}$ we obtain a better approximation than Halperin. In Section 5.1 we analyse the algorithm for the class of uniform, quasi-regularisable hypergraphs, which are known and useful in the combinatorics of hypergraphs (see Berge [5]). We prove an approximation ratio of $l\left(1 - \frac{n}{8m}\right)$ provided that $\Delta = O(n^{\frac{1}{3}})$. This result improves the approximation ratio given by Krivelevich and Halperin for sparse hypergraphs (roughly speaking sparseness means, $m \leq n^{\alpha}$, $\alpha \leq 2$, see section 5.1, page 9 for details). In Section 5.2 we consider l-uniform hypergraphs, where l and Δ are constants, and achieve a ratio of $l\left(1 - \frac{l-1}{4\Delta}\right)$. This improves over the result of Krivelevich for Δ smaller than $(l-1)^{1+\frac{1}{l-2}}$ and of Okun for Δ smaller than $(l-1)^{1+\frac{1}{\beta l-1}}$, respectively.

The paper is organised as follows: In Section 2 we give definitions and probabilistic tools. In Section 3 we present our randomised algorithm for the hitting set problem. In Section 4 we analyse the approximation ratio for hypergraphs with non-constant size of edges and non-constant vertex degree. In Section 5 we analyse the algorithm in a different way and prove an approximation ratio for the subclass of uniform quasi-regularisable hypergraphs (Section 5.1) and uniform hypergraphs with bounded vertex degree (Section 5.2). In Section 6 we comment on some future works.

2 Preliminaries and Definitions

Graph-theoretical Notions. Let $\mathcal{H} = (V, \mathcal{E})$ be a hypergraph, with V and $\mathcal{E}$ its set of vertices and edges. For $v \in V$ we define $d(v) = |\{E \in \mathcal{E}; v \in E\}|$ and $\Delta = \max_{v \in V}\{d(v)\}$. Here $d(v)$ is the vertex-degree of v and Δ is the maximum vertex degree of $\mathcal{H}$. Further for a set $X \subseteq V$ we denote by $\Gamma(X) := \{E \in \mathcal{E}\,;\, X \cap E \neq \emptyset\}$ the set of edges incident to the set X. Let $l, \Delta \in \mathbb{N}$ be two given constants. We call $\mathcal{H}$ l-uniform, if $|E| = l$ for all $E \in \mathcal{E}$, and with bounded degree Δ, if for every $v \in V$ it holds $d(v) \leq \Delta$. 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.

For an integer $k \geq 0$, multiplying the edge E_i by k means replacing the edge E_i in $\mathcal{H}$ by k identical copies of E_i. If $k = 0$, this operation is the deletion of the edge E_i. A hypergraph $\mathcal{H}$ is called *regularisable* if a regular hypergraph can be obtained from $\mathcal{H}$ by multiplying each edge E_i by an integer $k_i \geq 1$. Finally, a hypergraph $\mathcal{H}$ is called *quasi-regularisable* if a regular hypergraph is obtained by multiplying each edge E_i by an integer $k_i \geq 0$ where $\sum_i^m k_i > 0$. Regular implies regularisable and this implies quasi-regularisable (see [5]). Note that quasi-regularisable hypergraphs play an important role in the study of matching and covering in hypergraphs. e.g. [10].

Concentration Inequalities. For the one-sided deviation the following Chebychev-Cantelli inequality will be frequently used:

Theorem 1 ([2]). *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 independent bounded differences inequality theorem:

Theorem 2 (see [21]). *Let $X = (X_1, X_2, ..., X_n)$ 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 $\Pi_{k=1}^{n} A_k$ satisfies $|f(x) - f(x')| \leq c_k$ if the vector x and x' differ only in the k-th coordinate. Let $\mathbb{E}(X)$ be the expected value of the random variable $f(X)$. Then for any $t > 0$ it holds*

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

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 [2]). *Let X be the sum of finitely many $0/1$ random variables, i.e. $X = X_1 + \ldots + X_n$, and let $p_i = \mathbb{E}(X_i)$ for all $i = 1, \ldots, n$. For a pair $i, j \in \{1, \ldots, n\}$ we write $i \sim j$, if X_i and X_j are dependent. Let Γ be the set of all unordered dependent pairs i, j, i.e. 2-element sets $\{i, j\}$, and let $\gamma = \sum_{\{i,j\} \in \Gamma} \mathbb{E}(X_i X_j)$, then it holds:* $\mathrm{Var(X)} \leq \mathbb{E}(\mathrm{X}) + 2\gamma$.

3 The Randomised Algorithm

An integer, linear programming formulation of the hitting set problem in a hypergraph $\mathcal{H}$ is the following.

$$\text{(ILP-VC)} \qquad \min \sum_{j=1}^{n} x_j$$
$$\sum_{j \in E} x_j \geq 1 \text{ for all } E \in \mathcal{E},$$
$$x_j \in \{0, 1\} \text{ for all } j \in [n] := \{1, \ldots, n\}.$$

Its linear programming relaxation, denoted by LP-VC, is given by relaxing the integrality constraints to $x_j \in [0, 1]\ \forall j \in [n]$. Let Opt and Opt* be the value of an optimal solution to ILP-VC and LP-VC, respectively. Clearly, Opt$^* \leq$ Opt. Let x^* be an optimal solution of LP-VC. Let $\epsilon \in [0, 1]$ be a parameter that will be chosen based on the application, we set $\lambda = l(1 - \epsilon)$.

Algorithm 1. VC-$\mathcal{H}$

Input : A hypergraph $\mathcal{H} = (V, \mathcal{E})$
Output: A hitting set C

1. Initialise $C := \emptyset$.
2. Solve the LP relaxation of ILP-VC
3. Set $S_0 := \{j \in [n] \mid x_j^* = 0\}$, $S_1 := \{j \in [n] \mid x_j^* = 1\}$,
   $S_{\geq} := \{j \in [n] \mid 1 \neq x_j^* \geq \frac{1}{\lambda}\}$ and $S_{\leq} := \{j \in [n] \mid 0 \neq x_j^* < \frac{1}{\lambda}\}$.
4. Delete the vertices in S_0 from $\mathcal{H}$, and set $V := V \setminus S_0$ and $\mathcal{E} := \{E \cap V \mid E \in \mathcal{E}\}$.
5. Take all vertices of S_1 and $S_{\geq}$ into the hitting set C.
   Set $V := V \setminus S_1$ and $\mathcal{E} := \mathcal{E} \setminus \Gamma(S_1)$.
6. (Randomised Rounding) For all vertices $j \in S_{\leq}$ include the vertex j in the hitting set C, independently for all such j, with probability $x_j^* \lambda$.
7. (Repairing) Repair the Hitting Set C (if necessary) as follows:
   a) If $|\{E \in \mathcal{E} \mid E \cap C \neq \emptyset\}| = |\mathcal{E}|$, then return C.
   b) If $|\{E \in \mathcal{E} \mid E \cap C \neq \emptyset\}| < |\mathcal{E}|$, then pick arbitrary at most $|\mathcal{E}| - |C|$ additional vertices from not covered edges in the hitting set.
8. Return the hitting set C of $\mathcal{H}$

Let us briefly explain the ingredients of the algorithm. Usually, as in [8,9,11], the LP or semidefinite program is solved and randomised rounding or random hyperplane techniques are used followed by a repairing step. In our algorithm we thin out the hypergraph by removing vertices and edges corresponding to LP-variables with zero value, which will not be taken into the hitting set by randomised rounding (Step 4), *before* entering randomised rounding and repairing. This is an intuitively meaningful sparsening, and in fact will be necessary in Section 5 where we estimate the expected size of the repaired hitting set (one step analysis), while in Section 4 it is sufficient to analyse randomised rounding and repairing separately.

4 Two-Step Analysis of the Algorithm VC-$\mathcal{H}$

Let $X_1, ..., X_n$ be 0/1-random variables defined as follows: X_j is 1 if the vertex v_j was picked into the hitting set after the rounding step and 0 otherwise.
For all $i \in [m]$ we define the 0/1- random variables Z_i as follows: Z_i is 1 if the edge E_i is covered after the rounding step and 0 otherwise.
Then $Y := \sum_{j=1}^{n} X_j$ is the cardinality of the hitting set after the randomised rounding step in the algorithm and $W = \sum_{j=1}^{m} Z_j$ is the number of covered edges after this step.

For the expected size of the hitting set we have the following upper bound:

$$\mathbb{E}(|C|) \leq \mathbb{E}(Y) + \mathbb{E}(m - W). \tag{1}$$

For the computation of the expectation of W we need the following lemma (See Lemma 2.2 [22]).

Lemma 2. *For all $n \in \mathbb{N}$, $\lambda > 0$ and $x_1, \cdots, x_n, z \in [0,1]$ with $\sum_{i=1}^n x_i \geq z$ and $\lambda x_i < 1$ for all $i \in \mathbb{N}$, we have $\prod_{i=1}^n (1-\lambda x_i) \leq (1-\lambda\frac{z}{n})^n$, and this bound is the tight maximum.*

Lemma 3. *Let l and Δ be integers, not necessarily constant and let $\epsilon > 0$.*

(i) $\mathbb{E}(W) \geq (1-\epsilon^2)m$.
(ii) $\text{Opt}^* \geq \frac{m}{\Delta}$.
(iii) *Let hypergraph $\mathcal{H} = (V, \mathcal{E})$ with $x_j^* > 0$ for all $j \in [n]$ it holds* $\text{Opt}^* \geq \frac{n}{l}$, *where l is the maximum size of a edge.*
(iv) $\text{Opt}^* \leq \mathbb{E}(Y) \leq \lambda \text{Opt}^*$.

Proof. (i) For this proof we consider an equivalent form of the LP relaxation of the problem given in section 2.

$$
\begin{aligned}
(\text{LP}-1)\ \ & \min \sum_{j=1}^n x_j \\
& \sum_{j=1}^n a_{ij} x_j \geq z_i \text{ for all } i \in [m] := \{1, \ldots, m\} \\
& \sum_{i=1}^m z_i \geq m \\
& x_j, z_i \in [0,1] \text{ for all } i \in [m], j \in [n].
\end{aligned}
$$

It is easy to show that an optimal solution of LP-1 is an optimal solution of LP and vice versa.

Let $i \in [m]$, $|E_i| = r$ and $z_i^* = \sum_{j \in E_i} x_j^*$. If there is a $j \in E_i$ with $\lambda x_j \geq 1$ then $\Pr(Z_i = 0) = 0$, else we have

$$
\begin{aligned}
\Pr(Z_i = 0) &= \prod_{j \in E_i} (1 - \lambda x_j^*) \underset{\text{Lem 2}}{\leq} \left(1 - \frac{\lambda z_i^*}{r}\right)^r \\
&\leq \left(1 - \frac{\lambda z_i^*}{l}\right)^r = (1-(1-\epsilon)z_i^*)^r \\
&\leq (1-(1-\epsilon)z_i^*)^2 \leq 1 - z_i^*(1-\epsilon^2)
\end{aligned}
$$

and we get

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

(ii) Let $d(v_j)$ the degree of the vertex v_j. With the ILP constraints we have

$$
m \leq \sum_{i=1}^m z_i^* \leq \sum_{i=1}^m \sum_{j \in E_i} x_j^* = \sum_{j=1}^n d(v_j) x_j^* \leq \Delta \sum_{j=1}^n x_j^* = \Delta \cdot \text{Opt}^*
$$

(iii) Let consider the LP problem dual to the hitting set LP problem

$$\text{(D-VC)} \qquad \max \sum_{j\in\mathcal{E}} y_j$$
$$\sum_{j\in\mathcal{E},\, i\in j} y_j \leq 1 \quad \text{for every } i \in V,$$
$$y_j \in [0,1] \quad \text{for all } j \in \mathcal{E}.$$

Let $(y_j^*)_{j\in[m]}$ resp. $\text{Opt}^*(D)$ be an optimal solution of D-VC resp. the value of the optimal solution, than the duality Theorem of Linear Programming applied to the (LP-VC) and (D-VC) implies:

(a) $\text{Opt}^* = \text{Opt}^*(D)$
(b) If $x_i^* > 0 \Rightarrow \sum_{j\in\mathcal{E},\, i\in j} y_j = 1$.

Therefore, we have
$n = \sum_{i\in V} 1 = \sum_{i\in V}\sum_{j\in\mathcal{E},\, i\in j} y_j^* = \sum_{j\in\mathcal{E}} y_j^*|j\cap V| \leq l\sum_{j\in\mathcal{E}} y_j^* \underset{\text{(a)}}{=} l\text{Opt}^*$.

(iv) By using the LP relaxation and the definition of the sets S_1, $S_{\geq}$ and $S_{\leq}$, and since $\lambda \geq 1$, we get

$$\text{Opt}^* \leq \overbrace{\underbrace{|S_1|}_{\text{Opt}^*(S_1)} + \underbrace{|S_{\geq}|}_{\leq\lambda\text{Opt}^*(S_{\geq})} + \lambda\text{Opt}^*(S_{\leq})}^{=\mathbb{E}(Y)} \leq \lambda\text{Opt}^*.$$

□

4.1 Hypergraphs with Non-constant l, Δ

In this section we will analyse the algorithm for hypergraphs with maximum degree and maximum edge size that are not constant but may be given as functions of n. The main result in this section is:

Theorem 3. *Let $\mathcal{H}$ be a hypergraph with maximum edge size $l = O(\sqrt{n})$ and maximum vertex degree $\Delta = O(n^{\frac{1}{4}})$. The algorithm VC-$\mathcal{H}$ returns a hitting set C such that,* $|C| \leq l\left(1 - \frac{\sqrt{2}-1}{4\sqrt{2}\Delta}\right)\text{Opt}$ *with probability at least $\frac{3}{5}$.*

Proof. *Case 1* : $S_0 = \emptyset$.
Let
$$\epsilon := \frac{l\text{Opt}^*(1+\beta)}{4m} \quad \text{for} \quad \beta = \frac{\sqrt{2}l}{\sqrt{n}}. \tag{2}$$

We can assume that

$$\epsilon \leq \frac{1+\beta}{4-\eta}, \quad \text{for all} \quad \eta \in (0,\, 1), \tag{3}$$

because otherwise it follows from the definition of ϵ in (2) that $l\mathrm{Opt}^* \geq \frac{4m}{4-\eta}$, hence $l(1-\frac{\eta}{4})\mathrm{Opt}^* \geq m$. Since a hitting set of size m can be trivially found by picking m arbitrary edges and taking one vertex from each of them pairwise distinct, we can get a $l(1-\frac{\eta}{4})$-approximation —i.e. a constant factor strictly better than l— in this case.
It is straightforward to check that (3) implies $\epsilon \leq \frac{2}{3}$, so $\lambda = l(1-\epsilon) \underset{l\geq 3}{>} 1$.

Claim 1. $\Pr\left(W \leq m(1-\epsilon^2) - \sqrt{\sum_{i=1}^n d^2(v_i)}\right) \leq \frac{1}{5}$.

Proof of Claim 1. First we consider the function: $f(X_1, ..., X_n) = \sum_{j=1}^m Z_j$. f satisfies: $|f(X_1, .., X_k, .., X_n) - f(X_1, .., X_k^{'}, .., X_n)| \leq d(v_k)$, with $X_k^{'} \in \{0, 1\}$ and $X_k \neq X_k^{'}$.

Since the $X_1, ..., X_n$ are chosen independently at random, by Theorem 2 we get for any $t > 0$

$$\Pr(f(X) - \mathbb{E}(f(X)) \leq -t) \leq \exp\left(\frac{-2t^2}{\sum_{i=1}^n d^2(v_i)}\right). \quad (4)$$

Let us choose $t = \sqrt{\sum_{i=1}^n d^2(v_i)}$. By Lemma 3 (ii)

$$\Pr\left(W \leq m(1-\epsilon^2) - \sqrt{\sum_{i=1}^n d^2(v_i)}\right) \leq \Pr\left(W \leq \mathbb{E}(W) - \sqrt{\sum_{i=1}^n d^2(v_i)}\right)$$
$$\underset{\text{Ineq (4)}}{\leq} \exp\left(\frac{-2\sum_{i=1}^n d^2(v_i)}{\sum_{i=1}^n d^2(v_i)}\right) < \frac{1}{5}.$$

This concludes the proof of Claim 1.

Claim 2. For $\beta = \frac{\sqrt{2}l}{\sqrt{n}}$ it holds that $\Pr(Y \geq l \cdot \mathrm{Opt}^*(1-\epsilon)(1+\beta)) < \frac{1}{5}$.

Proof of Claim 2. The random variables $X_1, ..., X_n$ in the rounding step are independent. Moreover, since $l \leq \sqrt{2n}$ we have $\beta \in (0, 1)$. Thus the Angluin-Valliant form of Chernoff bound ([21], Theorem 2.3, p. 200) shows

$$\Pr(Y \geq l(1-\epsilon)(1+\beta)\mathrm{Opt}^*) \underset{\text{Lem3}(iv)}{\leq} \Pr(Y \geq \mathbb{E}(Y)(1+\beta)) \leq \exp\left(-\frac{\beta^2 \mathbb{E}(Y)}{3}\right).$$

On the other hand we have: $\frac{\mathbb{E}(Y)\beta^2}{3} \underset{\text{Lem3}(iv)}{\geq} \frac{\mathrm{Opt}^*\beta^2}{3} \underset{\text{Lem 3}(iii)}{\geq} \frac{n\beta^2}{3l} \geq \frac{2l^2 n}{3ln} \underset{l\geq 3}{\geq} 2$.
Finally we get: $\Pr(Y \geq l(1-\epsilon)(1+\beta)\mathrm{Opt}^*) \leq \exp(-2) < \frac{1}{5}$.
This concludes the proof of Claim 2.

By Claims 1 and 2 we get with probability at least $1 - (\frac{1}{5} + \frac{1}{5}) \geq \frac{3}{5}$ an upper bound for the final hitting set:

$$|C| \leq \underbrace{l(1-\epsilon)(1+\beta)\mathrm{Opt}^* + m\epsilon^2}_{(*)} + \underbrace{\sqrt{\sum_{i=1}^n d^2(v_i)}}_{(**)}.$$

By Lemma 3(iii) and the condition $\Delta \leq \frac{1}{32} n^{\frac{1}{4}}$ it holds:

$$(**) = \sqrt{\sum_{i=1}^{n} d^2(v_i)} \leq \Delta\sqrt{n} \leq \sqrt{\frac{n}{l}}\sqrt{l}\Delta \leq l\sqrt{\mathrm{Opt}^*}\sqrt{\mathrm{Opt}^*}\frac{1}{4\sqrt{2}\Delta} \leq l\mathrm{Opt}^*\frac{1}{4\sqrt{2}\Delta}.$$

Furthermore we have

$$(*) \underset{\mathrm{Eq}\,(2)}{=} l\left((1+\beta)(1-\epsilon) + \tfrac{l\mathrm{Opt}^*(1+\beta)}{16m}\right)\mathrm{Opt}^* = l(1+\beta)\left(1 - \tfrac{3l\mathrm{Opt}^*(1+\beta)}{16m}\right)\mathrm{Opt}^*$$
$$\underset{\mathrm{Lem}\,3(ii)}{\leq} l(1+\beta)\left(1 - \tfrac{3l(1+\beta)}{16\Delta}\right)\mathrm{Opt}^* = l\left(1+\beta - \tfrac{3l(1+\beta)^2}{16\Delta}\right)\mathrm{Opt}^*.$$

On the other hand we can easily check, that $\frac{3l(1+\beta)^2}{16\Delta} - \beta \geq \frac{1}{4\Delta}$, therefore

$$l(1-\epsilon)(1+\beta)\mathrm{Opt}^* + m\epsilon^2 \leq l\left(1 - \frac{1}{4\Delta}\right)\mathrm{Opt}^*.$$

Finally $(*) + (**) \leq l\left(1 - \frac{1}{4\Delta} + \frac{1}{4\sqrt{2}\Delta}\right)\mathrm{Opt}^* \leq l\left(1 - \frac{\sqrt{2}-1}{4\sqrt{2}\Delta}\right)\mathrm{Opt}^*$.

The randomised algorithm returns with probability at least $\frac{3}{5}$ a hitting set C with cardinality at most $l\left(1 - \frac{c}{\Delta}\right)\mathrm{Opt}^*$, where $c = \frac{1}{4}(1 - \frac{1}{\sqrt{2}})$.

Case 2: If S_0 is not empty, we can consider the sub-hypergraph $\mathcal{H}$ constructed in step 4 of algorithm VC-$\mathcal{H}$. Let $\tilde{\Delta}$ resp. $\tilde{l}$ be the maximum vertex degree resp. the maximum edge size of this sub-hypergraph. Now for this hypergraph we have $S_0 = \emptyset$ and with Case 1 we get a hitting set of cardinality at most $\tilde{l}(1 - \frac{c}{\tilde{\Delta}})$Opt. Since $\tilde{l} \leq l$ and $\tilde{\Delta} \leq \Delta$, the assertion of Theorem 3 holds. □

Remark 1. For hypergraphs addressed in Theorem 3 we have an improvement over the result of Krivelevich [18], for any function $f(n)$ satisfying $f(n) = O(n^{\frac{1}{4}})$, since $n^{\frac{1}{4}} < n^{1-\frac{1}{l}}$ for $l \geq 2$, and our approximation is the better the smaller $f(n)$ becomes. For $\Delta \leq \frac{\ln(n)}{\ln\ln(n)}$ we obtain a better approximation than Halperin [12].

5 One-Step Analysis of the Algorithm VC-$\mathcal{H}$

Instead bounding the error probability of the randomised rounding step and the repairing step separately as above, in this section we analyse the expected size of the hitting set including repairing, and then use concentration inequalities.

For a set $S \subset \{1, ..., n\}$ let $\mathrm{Opt}^*(S) := \sum_{j \in S} x_j^*$. By (1) it holds

$$\mathbb{E}(|C|) \leq \mathrm{Opt}^*(S_1) + l(1-\epsilon)(\mathrm{Opt}^*(S_{\geq}) + \mathrm{Opt}^*(S_{\leq})) + m\epsilon^2 \qquad (5)$$

Let us choose:

$$\epsilon = \frac{l(\mathrm{Opt}^*(S_{\geq}) + \mathrm{Opt}^*(S_{\leq}))}{2m}. \qquad (6)$$

We can assume that $\frac{l(\mathrm{Opt}^*(S_\geq)+\mathrm{Opt}^*(S_\leq))}{2m} \in [0,1]$. Otherwise, if $\frac{l(\mathrm{Opt}^*(S_\geq)+\mathrm{Opt}^*(S_\leq))}{2m} > 1$ then $\frac{l}{2}\mathrm{Opt}^* \geq \frac{l}{2}\left((\mathrm{Opt}^*(S_\geq) + \mathrm{Opt}^*(S_\leq)\right) > m$. Since any hitting set of cardinality m can be found trivially, this approximates the optimum within a factor of $\frac{l}{2} < l$.
Let $S_f := S_\geq \cup S_\leq \backslash \{j \in [n] | x_j^* = 0\}$. Plugging in ϵ from (6) into (5), we get

$$\mathbb{E}(|C|) \leq \mathrm{Opt}^*(S_1) + l\left(1 - \frac{l\mathrm{Opt}^*(S_f)}{4m}\right)\mathrm{Opt}^*(S_f). \quad (7)$$

We observe here that the LP-based sparsening of the instance becomes relevant. At next we compute the variance of the size of the hitting set. We get,

Lemma 4. *Let $X_1, \ldots, X_n$ be the 0/1-random variables returned by algorithm VC-$\mathcal{H}$. Then we have* $\mathrm{Var}(|C|) \leq l\Delta\mathbb{E}(|C|)$.

Proof. Let Γ and γ like in Lemma 1. Furthermore for every $v_i, v_j \in V, X_i, X_j$ are dependent iff they belong to the same edge. Thus, for a fixed v_i, there are at the most $(l-1)d(v_i)$ random variables X_j depending on X_i. Furthermore it holds for every $v_i, v_j \in V$:

$$\begin{aligned}\mathbb{E}(X_iX_j) = \Pr(X_i = 1 \wedge X_j = 1) &\leq \min\{\Pr(X_i = 1), \Pr(X_j = 1)\} \\ &\leq \frac{\Pr(X_i = 1) + \Pr(X_j = 1)}{2}.\end{aligned}$$

Moreover
$$\begin{aligned}\gamma &= \sum_{\{v_i,v_j\}\in\Gamma} \mathbb{E}(X_iX_j) \leq \sum_{\{v_i,v_j\}\in\Gamma} \frac{\Pr(X_i = 1) + \Pr(X_j = 1)}{2} \\ &\leq \sum_{i=1}^{n} \frac{(l-1)d(v_i)}{2}\Pr(X_i = 1) = \frac{(l-1)d(v_i)}{2}\mathbb{E}(|C|)\end{aligned}$$

by Lemma 1 we have: $\mathrm{Var}(|C|) \leq \mathbb{E}(|C|) + (l-1)d(v_i)\mathbb{E}(|C|) \leq l\Delta\mathbb{E}(|C|)$. □

5.1 Quasi-Regularisable *l*-Uniform Hypergraphs

Recall that S_1 is the set $S_1 = \{j \in [n] \mid x_j^* = 1\}$, containing those vertices for which the LP-optimal solution is tight (see algorithm VC-$\mathcal{H}$, step 3).
The next theorem is the main result of this section and it is proved using the above stated estimation (7) of $\mathbb{E}(|C|)$ and the Chebychev-Cantelli inequality.

Theorem 4. *Let $\mathcal{H}$ be a l-uniform, quasi-regularisable hypergraph with arbitrary l and maximum vertex degree $\Delta = O(n^{\frac{1}{3}})$, then the algorithm VC-$\mathcal{H}$ returns a hitting set C such that,* $|C| \leq l\left(1 - \frac{n}{8m}\right)\mathrm{Opt}^*$ *with probability at least $\frac{3}{4}$.*

We need the following theorem of Berge [5].

Theorem 5. *For an l-uniform hypergraph $\mathcal{H}$, the following properties are equivalent:*

1. *$\mathcal{H}$ is quasi-regularisable;*
2. $\mathrm{Opt}^* = \frac{n}{l}$ *(i.e. the vector* $x^* = (\frac{1}{l}, ..., \frac{1}{l})$ *is an optimal solution for the LP relaxation and* l *is the size of the edges).*

By this theorem, the condition $S_1 = \emptyset$ becomes a graph-theoretical meaning.

Proof of Theorem 4. By (7) and Theorem 5 we get for quasi-regularisable l-uniform hypergraphs with arbitrary l and bounded degree Δ the approximation

$$\mathbb{E}(|C|) \leq l\left(1 - \frac{n}{4m}\right)\mathrm{Opt}^*. \tag{8}$$

Hence

$$\Pr\left(|C| \geq l\left(1 - \frac{n}{8m}\right)\mathrm{Opt}^*\right) \underset{\mathrm{Ineq}\,(8)}{\leq} \Pr\left(|C| \geq \mathbb{E}(|C|) + \frac{nl\mathrm{Opt}^*}{8m}\right) \underset{\mathrm{Th}\,1}{\leq} \frac{1}{4}.$$

Namely for $n \geq 8^2\Delta^3$ we get with a straightforward calculation that $\frac{\left(\frac{ln\mathrm{Opt}^*}{8m}\right)^2}{\mathrm{Var}(|C|)} \geq l \geq 3$. So we obtain a hitting set C of size at most $l\left(1 - \frac{n}{8m}\right)\mathrm{Opt}^*$ with probability at least $\frac{3}{4}$. $\square$

Remark 2. In Theorem 4, we can assume that $n < 8m$, because otherwise we have $\mathrm{Opt}^* = \frac{n}{l} \geq \frac{8m}{l}$ thus $m \leq \frac{l}{8}\mathrm{Opt}^*$. By taking one vertex for each edge we obtain a hitting set of cardinality $\frac{l}{8}\mathrm{Opt}^*$, which gives already an approximation ratio of $l/8$. For hypergraphs addressed in Theorem 4 we have an improvement over the ratio of Krivelevich if $m \leq cn^{\frac{2l-1}{l}}$ and the ratio of Halperin if $m \leq \frac{(1-o(1))^{-1}\ln(n)n}{\ln\ln(n)}$.

5.2 l-Uniform Hypergraphs with Bounded Vertex Degree

In this section l and Δ are constants and $\mathcal{H}$ is an l-uniform hypergraph. Let $\tilde{\mathcal{H}} = (\tilde{V}, \tilde{\mathcal{E}})$ be the sub-hypergraph of $\mathcal{H}$ constructed in step 5 of the algorithm VC-$\mathcal{H}$ with $|\tilde{V}| = \tilde{n}$ and $|\tilde{\mathcal{E}}| = \tilde{m}$. We denote by $\tilde{l}$ and $\tilde{\Delta}$ the maximum size of all edges and the maximum vertex degree in $\tilde{\mathcal{H}}$. We consider the LP relaxation of the ILP formulation of the hitting set problem in $\tilde{\mathcal{H}}$ which we denote by $\mathrm{LP}(\tilde{\mathcal{H}})$. By $\mathrm{Opt}^*(\tilde{\mathcal{H}})$ we denote the value of the optimal solution of $\mathrm{LP}(\tilde{\mathcal{H}})$. The optimal LP solution for $\mathcal{H}$ is Opt^*. Then the following holds.

Lemma 5. $\mathrm{Opt}^*(\tilde{\mathcal{H}}) = \mathrm{Opt}^* - |S_1|$ *and* $\mathbb{E}(|C|) \leq |S_1| + \mathbb{E}(|\tilde{C}|)$.

Lemma 6. *Let l and Δ be constants, and let $\mathcal{H}$ be a l-uniform hypergraph with maximum vertex degree Δ. Then:* $\mathbb{E}(|C|) \leq l\left(1 - \frac{l}{4\Delta}\right)\mathrm{Opt}^*$.

Proof . Since there is no tight $\mathrm{LP}(\tilde{\mathcal{H}})$-variable, because there are no 1's in the solution $(\tilde{x}_1, \ldots, \tilde{x}_{\tilde{n}})$, we get using (5)

$$\mathbb{E}(|\tilde{C}|) \leq \tilde{l}\left(1 - \frac{\tilde{l}\mathrm{Opt}^*(\tilde{\mathcal{H}})}{4\tilde{m}}\right)\mathrm{Opt}^*(\tilde{\mathcal{H}}) \underset{\mathrm{Lem}\,3(ii)}{\leq} \tilde{l}\left(1 - \frac{\tilde{l}}{4\tilde{\Delta}}\right)\mathrm{Opt}^*(\tilde{\mathcal{H}}).$$

Furthermore,

$$\begin{aligned}
\mathbb{E}(|C|) &\leq |S_1| + \mathbb{E}(|\tilde{C}|) \leq |S_1| + \tilde{l}\left(1 - \frac{\tilde{l}}{4\tilde{\Delta}}\right)\mathrm{Opt}^*(\tilde{\mathcal{H}}) \\
&\underset{\lambda \geq 1}{\leq} \tilde{l}\left(1 - \frac{\tilde{l}\mathrm{Opt}^*(\tilde{\mathcal{H}})}{4\tilde{m}}\right)|S_1| + \tilde{l}\left(1 - \frac{\tilde{l}}{4\tilde{\Delta}}\right)\mathrm{Opt}^*(\tilde{\mathcal{H}}) \\
&\underset{\mathrm{Lem}\,3(ii)}{\leq} \tilde{l}\left(1 - \frac{\tilde{l}}{4\tilde{\Delta}}\right)(\mathrm{Opt}^*(\tilde{\mathcal{H}}) + |S_1|) \underset{\mathrm{Lem}\,5}{=} \tilde{l}\left(1 - \frac{\tilde{l}}{4\tilde{\Delta}}\right)\mathrm{Opt}^*.
\end{aligned}$$

and because $\mathcal{H}$ is uniform and $\Delta \geq \tilde{\Delta}$ we have: $\mathbb{E}(|C|) \leq l\left(1 - \frac{l}{4\Delta}\right)\mathrm{Opt}^*$ □

Lemma 6 and Lemma 4 imply the following theorem using the Chebyshev-Cantelli inequality and standard calculations.

Theorem 6. *Let $\mathcal{H}$ be an l-uniform hypergraph with bounded vertex degree, then the algorithm VC-$\mathcal{H}$ returns a hitting set C such that, $|C| \leq l\left(1 - \frac{l-1}{4\Delta}\right)\mathrm{Opt}^*$ with probability at least $\frac{3}{4}$.*

Proof. Assuming that $m \geq 16\Delta^5$ the proof is similar to the proof of Theorem 4. □

This improves over the result of Krivelevich [18] for Δ smaller then $(l-1)^{1+\frac{1}{l-2}}$ and of Okun [23] for Δ smaller then $(l-1)^{1+\frac{1}{\beta l-1}}$. The approximation ratio in this result is little weaker than the ratio of Halperin [12]. But the advantage here is that l and Δ are not coupled anymore, so a significantly larger class of hypergraphs than in [12] is covered.

6 Further Work

We believe that the analysis presented in this paper can incorporate other hypergraph parameters in a natural way, like bounded VC-dimension, uncrowdnedness, or exclusion of subgraphs. We hope that this may lead to new and better approximation results for the hitting set problem in such hypergraphs.

References

1. Alon, N., Moshkovitz, D., Safra, S.: Algorithmic construction of sets for k-restrictions. ACM Trans. Algorithms (ACM) 2, 153–177 (2006)
2. Alon, N., Spencer, J.: The probabilistic method, 2nd edn. Wiley Interscience (2000)
3. Bar-Yehuda, R., Even, S.: A linear-time approximation algorithm for the weighted vertex cover problem. Journal of Algorithms 2, 198–203 (1981)
4. Bar-Yehuda, R., Even, S.: A local ratio theorem for approximating weighted vertex cover problem. In: Ausiello, G., Lucertini, M. (eds.) Analysis and Design of Algorithms for Combinatorial Problems. Annals of Discrete Math., vol. 25, pp. 27–46. Elsevier, Amsterdam (1985)

5. Berge, C.: Hypergraphs-combinatorics of finite sets. North Holland Mathematical Library (1989)
6. Chvatal, V.: A greedy heuristic for the set covering problem. Math. Oper. Res. 4(3), 233–235 (1979)
7. Feige, U.: A treshold of $\ln n$ for approximating set cover. Journal of the ACM 45(4), 634–652 (1998)
8. Feige, U., Langberg, M.: Approximation algorithms for maximization problems arising in graph partitioning. Journal of Algorithms 41(2), 174–201 (2001)
9. Frieze, A., Jerrum, M.: Improved approximation algorithms for max k-cut and max bisection. Algorithmica 18, 67–81 (1997)
10. Füredi, Z.: Matchings and covers in hypergraphs. Graphs and Combinatorics 4(1), 115–206 (1988)
11. Gandhi, R., Khuller, S., Srinivasan, A.: Approximation Algorithms for Partial Covering Problems. J. Algorithms 53(1), 55–84 (2004)
12. Halperin, E.: Improved approximation algorithms for the vertex cover problem in graphs and hypergraphs. In: ACM-SIAM Symposium on Discrete Algorithms, vol. 11, pp. 329–337 (2000)
13. Hochbaum, D.S.: Approximation algorithms for the set covering and vertex cover problems. SIAM J. Computation 11(3), 555–556 (1982)
14. Hall, N.G., Hochbaum, D.S.: A fast approximation for the multicovering problem. Discrete Appl. Math. 15, 35–40 (1986)
15. Jäger, G., Srivastav, A.: Improved approximation algorithms for maximum graph partitioning problems. Journal of Combinatorial Optimization 10(2), 133–167 (2005)
16. Johnson, D.S.: Approximation algorithms for combinatorial problems. J. Comput. System Sci. 9, 256–278 (1974)
17. Khot, S., Regev, O.: Vertex cover might be hard to approximate to within 2-epsilon. J. Comput. Syst. Sci. 74(3), 335–349 (2008)
18. Krivelevich, J.: Approximate set covering in uniform hypergraphs. J. Algorithms 25(1), 118–143 (1997)
19. Lovász, L.: On the ratio of optimal integral and fractional covers. Discrete Math. 13, 383–390 (1975)
20. Lund, C., Yannakakis, M.: On the hardness of approximating minimization problems. J. Assoc. Comput. Mach. 41, 960–981 (1994)
21. McDiarmid, C.: Concentration. In: Habib, M., McDiarmid, C., Ramirez-Alfonsin, J., Reed, B. (eds.) Probabilistic Methods for Algorithmic Discrete Mathematics, pp. 195–248. Springer, Berlin (1998)
22. Peleg, D., Schechtman, G., Wool, A.: Randomized approximation of bounded multicovering problems. Algorithmica 18(1), 44–66 (1997)
23. Okun, M.: On approximation of the vertex cover problem in hypergraphs. Discrete Optimization (DISOPT) 2(1), 101–111 (2005)
24. Raz, R., Safra, S.: A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In: Proc. 29th ACM Symp. on Theory of Computing, pp. 475–484 (1997)