Theoretical Computer Science 555 (2014) 23–34

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

A randomised approximation algorithm for the hitting set problem

Mourad El Ouali [a,*], Helena Fohlin [b], Anand Srivastav [a,**]

[a] Department of Computer Science, Kiel University, 24118 Kiel, Germany
[b] Department of Clinical and Experimental Medicine, Linköping University, S-58185 Linköping, Sweden

ARTICLE INFO

Article history:
Received 19 April 2013
Received in revised form 12 November 2013
Accepted 25 March 2014
Available online 28 March 2014

Keywords:
Approximation algorithms
Probabilistic methods
Randomised rounding
Hitting set
Vertex cover
Greedy algorithms

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 subset of 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 hand, for constant l, an approximation ratio better than l cannot be achieved in polynomial time under the unique games conjecture (Khot and Regev, 2008 [17]). Thus breaking the l-barrier for significant classes of hypergraphs is a complexity-theoretically and algorithmically interesting problem, which has been studied by several authors (Krivelevich, 1997 [18], Halperin, 2000 [12], Okun, 2005 [23]). 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 for different classes of hypergraphs. For hypergraphs with $\Delta = O(n^{\frac{1}{4}})$ and $l = O(\sqrt{n})$ we achieve an approximation ratio of $l(1 - \frac{c}{\Delta})$, for some constant $c > 0$, with constant probability. For the case of hypergraphs where l and Δ are constants, we prove a ratio of $l(1 - \frac{l-1}{8\Delta})$. The latter is done by analysing the expected size of the hitting set and using concentration inequalities. Moreover, for quasi-regularisable hypergraphs, we achieve an approximation ratio of $l(1 - \frac{n}{8m})$. We show how and when our results improve over the results of Krivelevich, Halperin and Okun.

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 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 $\mathcal{NP}$ problems

* Principal corresponding author.
** Corresponding author.
E-mail addresses: meo@informatik.uni-kiel.de (M. El Ouali), Helena.Fohlin@lio.se (H. Fohlin), asr@informatik.uni-kiel.de (A. Srivastav).

http://dx.doi.org/10.1016/j.tcs.2014.03.029
0304-3975/

have a quasi-polynomial, i.e., $n^{\mathcal{O}(\mathrm{poly}(\ln n))}$ 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 $c > 0$ 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. For l-uniform hypergraphs with a constant l 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]. On the other hand an approximation ratio of l can be 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 $l(1-c/n^{\frac{\ell-1}{\ell}})$, for some constant $c > 0$, using a combination of an 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, $n \geqslant 2^{2^{l^2}}$, and already for $l = 3$ the hypergraph is very large, 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 $l(1-c/\Delta^{\frac{1}{l-1}})$ 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 $l^3 = o(\frac{\ln\ln\Delta}{\ln\ln\ln\Delta})$. Recently Saket and Srividenko [25] showed that the cost of the approximate cover is bounded from above by $(l(1-1/\Delta^{\frac{1}{l-1}}) + 1/\Delta^{\frac{1}{l-1}})\,\mathrm{Opt}$ in expectation.

For hypergraphs which are not necessarily uniform, but with edge size 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 $\lceil \beta l \rceil$ where $\beta \in (0,1)$ satisfies $1-\beta = (\frac{\beta l}{\beta l+1})\Delta^{-\frac{1}{\beta l}}$, by a modification of the algorithm presented in [18].

In this work we investigate the approximability of the hitting set problem in hypergraphs with maximum edge size l and maximum vertex degree Δ. 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 practise 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]. We show that our algorithm achieves for $l = O(\sqrt{n})$ and $\Delta = O(n^{\frac{1}{4}})$ an approximation ratio of $l(1-\frac{c}{\Delta})$, 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 $l \geqslant 2$, and the approximation is the better the smaller $f(n)$ becomes. For $\Delta \leqslant \frac{\ln n}{\ln\ln n}$ we obtain a better approximation than Halperin. Furthermore 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 $l(1-\frac{n}{8m})$ 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, $m \leqslant n^{\alpha}$, $\alpha \leqslant 2$, see Remark 1 at the end of Section 4 for more details). Finally, we consider l-uniform hypergraphs, where l and Δ are constants, and achieve a ratio of $l(1-\frac{l-1}{8\Delta})$ where $l \leqslant \frac{16}{3}\Delta$. 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 sketch some future work.

2. Preliminaries and definitions

Graph-theoretical notions. For an integer $n \in \mathbb{N}$, we define $[n] := \{1,\dots,n\}$. Let $\mathcal{H} = (V,\mathcal{E})$ be a hypergraph, where V is a finite set V and $\mathcal{E}$ is a set of subsets of V. We call the elements of V vertices and the elements of $\mathcal{E}$ hyperedges or simply 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 $E \in \mathcal{E}$. $\mathcal{H}$ has bounded degree Δ, if for every $v \in V$ it holds $d(v) \leqslant \Delta$. Some times it is convenient to order the vertices and edges, i.e., $V = \{v_1,\dots,v_n\}$ and $\mathcal{E} = \{E_1,\dots,E_m\}$, and to identify the vertices and edges with their indices.

For an integer $k \geqslant 0$, 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 $k_i \geqslant 1$. Finally, a hypergraph $\mathcal{H}$ is called *quasi-regularisable* if a regular hypergraph is obtained by multiplying each edge E_i by an integer $k_i \geqslant 0$ where $\sum_i^m k_i > 0$. Clearly, a regular