Approximate Hypergraph Vertex Cover and generalized Tuza's conjecture

Venkatesan Guruswami[*]   Sai Sandeep[†]

Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

A famous conjecture of Tuza states that the minimum number of edges needed to cover all the triangles in a graph is at most twice the maximum number of edge-disjoint triangles. This conjecture was couched in a broader setting by Aharoni and Zerbib who proposed a hypergraph version of this conjecture, and also studied its implied fractional versions. We establish the fractional version of the Aharoni-Zerbib conjecture up to lower order terms. Specifically, we give a factor $t/2 + O(\sqrt{t \log t})$ approximation based on LP rounding for an algorithmic version of the hypergraph Turán problem (AHTP). The objective in AHTP is to pick the smallest collection of $(t-1)$-sized subsets of vertices of an input t-uniform hypergraph such that every hyperedge contains one of these subsets.

Aharoni and Zerbib also posed whether Tuza's conjecture and its hypergraph versions could follow from non-trivial duality gaps between vertex covers and matchings on hypergraphs that exclude certain sub-hypergraphs, for instance, a "tent" structure that cannot occur in the incidence of triangles and edges. We give a strong negative answer to this question, by exhibiting tent-free hypergraphs, and indeed $\mathcal{F}$-free hypergraphs for any finite family $\mathcal{F}$ of excluded sub-hypergraphs, whose vertex covers must include almost all the vertices.

The algorithmic questions arising in the above study can be phrased as instances of vertex cover on *simple* hypergraphs, whose hyperedges can pairwise share at most one vertex. We prove that the trivial factor t approximation for vertex cover is hard to improve for simple t-uniform hypergraphs. However, for set cover on simple n-vertex hypergraphs, the greedy algorithm achieves a factor $(\ln n)/2$, better than the optimal $\ln n$ factor for general hypergraphs.

1 Introduction

The relationship between minimum vertex covers and maximum matchings of graphs and hypergraphs is a fundamental and well-studied topic in combinatorics and optimization. Even though the worst-case factor t gap between the two parameters cannot be improved on arbitrary t-uniform hypergraphs, there are some interesting special cases where the ratio between these quantities is smaller. A classic example of this phenomenon is the König's theorem on bipartite graphs, where the sizes of minimum vertex covers and maximum matchings are equal.

For the case of $t = 3$, a notorious open problem capturing this gap on special 3-uniform hypergraphs is Tuza's conjecture [Tuz81, Tuz90], which states that in any graph, the number of edges required to hit all triangles is at most twice the maximum number of edge-disjoint triangles. For a hypergraph H, let us denote by $\tau(H)$ and $\nu(H)$ the sizes of the minimum vertex cover and maximum matching respectively. Tuza's conjecture is then equivalent to the statement $\tau(H) \leq 2 \cdot \nu(H)$ for any 3-uniform hypergraph H obtained by taking the edges of a graph G as its vertices, and the triangles of G as its (hyper)-edges. (Taking $G = K_4$ shows that the factor 2 is best possible.) The conjecture has been verified for various classes of graphs such as graphs without $K_{3,3}$-subdivision [Kri95], graphs with maximum average degree less than 7 [Pul15], graphs with quadratic number of edge disjoint triangles [HR01, Yus12], graphs with treewidth at most 6 [BFG19], and random graphs in the $G_{n,p}$ model [BCD20, KP20]. On general graphs, the current best upper bound on the ratio is a factor of 2.87 due to Haxell [Hax99].

Aharoni and Zerbib [AZ20] introduced an extension of Tuza's conjecture to hypergraphs of larger uniformity. This generalized Tuza's conjecture states that for any t-uniform hypergraph H, the minimum vertex cover $\tau(H')$ of $H' = H^{(t-1)}$ is at most $\left\lceil \frac{t+1}{2} \right\rceil$ times that of the maximum matching $\nu(H')$. Here, for a t-uniform hypergraph $H = (V, E)$, the $(t-1)$-*blown-up hypergraph* $H' = H^{(t-1)}$ is a t-uniform hypergraph whose vertices are the set of all $(t-1)$ sized subsets that are contained in at least one edge of H, and corresponding to every edge e in H, all the $(t-1)$-sized subsets of e form an edge in H'. Tuza's conjecture is a special case of their conjecture when $t = 3$ and H has hyperedges corresponding to the triangles in a graph. As is the case with the original Tuza's conjecture, the conjectured value of $\left\lceil \frac{t+1}{2} \right\rceil$ is the best possible gap: when H is the complete t-uniform hypergraph on $(t+1)$ vertices, the $(t-1)$-blown-up hypergraph $H' = H^{(t-1)}$ has $\nu(H') = 1$ and $\tau(H') = \left\lceil \frac{t+1}{2} \right\rceil$.

1.1 Fractional Tuza's conjecture and the algorithmic hypergraph Turán problem

A first step towards non-trivially bounding $\tau(H)$ in terms of $\nu(H)$ for hypergraphs H from some structured family of hypergraphs is proving its *fractional version*, i.e., obtaining the same upper bound on the ratio between $\tau(H)$ and $\nu^*(H)$, the fractional maximum matching size. By LP duality, this is equivalent to bounding the ratio between $\tau(H)$ and $\tau^*(H)$, the fractional vertex cover value. As $\nu(H) \leq \nu^*(H) = \tau^*(H) \leq \tau(H')$ for any hypergraph H, establishing the fractional version is a necessary step toward bounding $\tau(H)/\nu(H)$. Note that understanding the extremal ratio between τ and τ^* on a given family of hypergraphs is equivalent to bounding the integrality gap of the natural linear programming relaxation of vertex cover on that class of hypergraphs.

Krivelevich [Kri95] proved the fractional version of Tuza's conjecture that $\tau(H^{(2)}) \leq 2\tau^*(H^{(2)})$ for any 3-uniform hypergraph H. A multi-transversal version of Krivelevich's result is proved in a recent work [CKSU20]. In this work, we prove the fractional version of the generalized Tuza's conjecture (upto $o(t)$ factors), establishing a non-trivial upper bound on the LP integrality gap for $(t-1)$-blown-up hypergraphs.

Theorem 1. *For any t-uniform hypergraph H, $\tau(H') \leq \left(\frac{t}{2} + 2\sqrt{t\ln t}\right)\tau^*(H')$, where $\tau(H')$ and $\tau^*(H')$ are respectively the size of the minimum vertex cover and minimum fractional vertex cover of the blown-up hypergraph $H' = H^{(t-1)}$. Furthermore, there is an efficient algorithm to approximate vertex cover on $(t-1)$-blown-up hypergraphs within a $\frac{t}{2} + 2\sqrt{t\ln t}$ factor.*

The vertex cover problem on $(t-1)$-blown-up hypergraphs is also intimately connected to the famous Hypergraph Turán Problem [Tur41,Tur61] in extremal combinatorics. In the Hypergraph Turán Problem, the goal is to find the minimum size of a family $\mathcal{F} \subseteq \binom{[n]}{t-1}$ of subsets of $[n]$ with cardinality $(t-1)$ such that for every subset S of $[n]$ of size t, there exists a set $T \in \mathcal{F}$ such that T is a subset of S. The best known upper bound is due to [Sid97]: there exists a family $\mathcal{F} \subseteq \binom{[n]}{t-1}$ of size $O(\frac{\log t}{t})\binom{n}{t-1}$ such that for every subset S of $[n]$ of size t, there exists a subset $T \in \mathcal{F}$ such that T is contained in S. On the other hand, the lower bound situation is rather dire, with only second-order improvements [CL99,LZ09] over the trivial $\frac{1}{t}\binom{n}{t-1}$ lower bound.

Note that the Hypergraph Turán Problem is precisely the minimum vertex cover problem on $H^{(t-1)}$ when H is the *complete* t-uniform hypergraph. Thus, for a general hypergraph H, finding vertex covers on the blown-up hypergraph $H^{(t-1)}$ can be viewed as an algorithmic version of the Hypergraph Turán problem.

Problem 2. *(Algorithmic Hypergraph Turán Problem (AHTP)) Given a t-uniform hypergraph $H = (V = [n], E)$, find the minimum size of a family $\mathcal{F} \subseteq \binom{[n]}{t-1}$ of subsets of V of size $(t-1)$ such that for every hyperedge $e \in E$, there exists $T \in \mathcal{F}$ such that T is a subset of e.*

The problem is a generalization of the minimum vertex cover on graphs, which corresponds to the case $t = 2$. As AHTP can be cast as a vertex cover problem on t-uniform hypergraphs, there is a trivial factor t approximation algorithm for this problem. We prove Theorem 1 by obtaining an improved algorithm for AHTP based on rounding the standard LP relaxation on $H' = H^{(t-1)}$.

We now briefly describe this rounding approach. First, using threshold rounding, we argue that one may focus on the case when the LP solution does not have any variables that are assigned values greater than $\frac{2}{t}$. Let S be the set of vertices of H' that are assigned non-zero LP value. The thresholding procedure ensures that every hyperedge $e \in E(H')$ intersects with S in at least $\frac{t}{2}$ vertices. We can bound the cardinality of S from above by $t \cdot \mathsf{OPT}$ using the dual matching LP, where OPT is the cost of the optimal LP solution. Our goal then becomes finding a vertex cover of size at most about $\frac{|S|}{2}$. We achieve this by a color-coding technique: we randomly assign a color from $\{0, 1\}$ to each vertex of H independently. Most edges of H are almost balanced under this coloring, in the sense that each color appears at least $t/2 - o(t)$ times. We then use this balance property to find a small vertex cover in H'.

1.2 Vertex cover vs. matching and excluded sub-hypergraphs

The generalized Tuza's conjecture concerns the relationship between τ and ν on the $(t-1)$-blown-up hypergraphs. There have been some works in the literature on the gap between τ and ν on other structured class of hypergraphs. An outstanding result of this type is Aharoni's proof [Aha01] that $\tau(H) \leq 2\nu(H)$ for all tripartite 3-uniform hypergraphs H. Aharoni and Zerbib [AZ20] asked if there is a structural explanation that unites the generalized Tuza's conjecture and the above result—for example, does the exclusion of a certain substructure in the hypergraph H imply better gaps between $\tau(H)$ and $\nu(H)$. A particular substructure they studied is the "tent" subhypergraph (Figure 1).
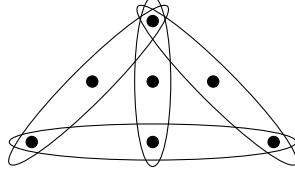
Figure 1: The 3-tent

They observed that both tripartite hypergraphs and 2-blown-up 3-uniform hypergraphs cannot contain the tent as a subhypergraph, and asked whether a generalization of Tuza's conjecture might hold for 3-uniform hypergraphs that exclude tents. If this is the case, it could give a common structural explanation of the existence of small vertex covers in 3-uniform hypergraphs.

In this paper, we answer this question in the negative. We prove that that there are hypergraphs H on n vertices that exclude tents with $\tau(H) \geq (1 - o(1))n$. Since $\nu(H) \leq n/3$ trivially, this shows that the ratio τ/ν can approach 3 on tent-free 3-uniform hypergraphs, and the extension of Tuza's conjecture as raised in [AZ20] does not hold. More generally, one might ask if there is some collection of hypergraphs which are excluded from blown-up hypergraphs whose absence implies a non-trivial gap between τ and ν. In fact, we prove a stronger statement showing that there is no 3-uniform hypergraph family $\mathcal{F}$ (that is absent from blown-up hypergraphs) whose exclusion alone could imply Tuza's conjecture. Our result applies for larger uniformity t and the fractional version of Tuza's conjecture.

Theorem 3. *For every $\varepsilon > 0$ and every finite family of t-uniform hypergraphs $\mathcal{F}$ such that no hypergraph from $\mathcal{F}$ appears in any $(t-1)$-blown-up hypergraph $H' = H^{(t-1)}$, there is a t-uniform hypergraph T such that T does not contain any hypergraph from $\mathcal{F}$ but $\tau(T) \geq (t - \varepsilon)\nu(T)$ (and in fact $\tau(T) \geq (t - \varepsilon)\tau^*(T)$).*

The above result rules out the possibility of a "local" proof of Tuza's conjecture. Our construction is a probabilistic one, first sampling each edge of the hypergraph independently with certain probability, and then removing all the copies of hypergraphs in $\mathcal{F}$. Using the fact that the family $\mathcal{F}$ satisfies certain sparsity requirements [FM08, BFM10], we can conclude that there is no large independent set in this construction.

We also provide an *explicit construction* that answers the tent-free question of [AZ20]: our counterexample is the hypergraph T with vertex set $[3]^n$ for large enough n and edges being the set of combinatorial lines. By the density Hales Jewett Theorem [FK91, Pol12], there is no large independent set in T, and using the structure of combinatorial lines, we can prove that T does not have any tent.

1.3 Vertex cover and set cover on simple hypergraphs

As mentioned earlier, AHTP is a special case of vertex cover on t-uniform hypergraphs. In fact, the blown-up hypergraph $H^{(t-1)}$ is a *simple*[1] hypergraph: any two edges intersect in at most one vertex. This is simply because any two distinct t-sized subsets of $[n]$ intersect in at most one $(t-1)$-sized subset. Simple hypergraphs have been well studied in Graph Theory, especially in the context of Erdős-Faber-Lovász conjecture [Erd81, Erd88], Ryser's conjecture [FHMW17] and chromatic number of bounded degree hypergraphs [DLR95, FM13].

---

[1]Simple hypergraphs are also referred to as linear hypergraphs.

A natural question is whether we can obtain an approximation ratio smaller than t for vertex cover on simple hypergraphs. However, Theorem 3 shows that the natural LP has an integrality gap approaching t on simple, and indeed a lot more structured, hypergraphs. But perhaps there are other algorithms that beat the trivial factor t approximation for this problem. We prove that this is not the case, and in fact, vertex cover on simple hypergraphs is as hard as vertex cover on general t-uniform hypergraphs.

Theorem 4. *For every $\varepsilon > 0$, unless NP $\subseteq$ BPP, no polynomial time algorithm can approximate vertex cover on simple t-uniform hypergraphs within a factor of $t - 1 - \varepsilon$. Under the Unique Games conjecture, the inapproximability factor improves to $t - \varepsilon$.*

We also study the set cover problem on simple set families where any two sets in the family intersect in at most one element. Equivalently, we want to pick the minimum number of edges to cover all vertices in a simple hypergraph. Interestingly, simplicity of the set family helps in getting an improved approximation factor for the set cover—in fact, the greedy algorithm itself delivers such an approximation.

Theorem 5. *For set cover on simple set systems over a universe of size n, the greedy algorithm achieves an approximation ratio $\frac{\ln n}{2} + 1$. Further, there are simple set systems where the greedy is off by a factor exceeding $\frac{\ln n}{2} - 1$.*

The same techniques used in proving Theorem 4 also gives an inapproximability factor of $(\ln n)^{\Omega(1)}$ for set cover on simple set families. Interestingly, the dual Maximum Coverage problem, where the goal is to cover as many elements as possible with a specified number of sets, does not become easier on simple set systems and is hard to approximate within a factor exceeding $(1 - 1/e)$ [CKL21], the factor achieved by the greedy algorithm on general set systems. In [CKL20], the authors conjecture the hardness of achieving an approximation factor beating $(1 - 1/e)$ even for the Maximum Coverage version of AHTP, and call this the *Johnson Coverage Hypothesis*. They show that this hypothesis implies strong inapproximability results for fundamental clustering problems like k-means and k-median on Euclidean metrics. For example, they showed that the hypothesis implies that k-median is hard to approximate within a factor of 1.73 on ℓ_1 metrics, matching the best hardness factor on general metrics due to Guha and Khuller [GK98].

1.4   Other improved hypergraph vertex cover algorithms

Algorithms beating the trivial factor t approximation have been obtained for the vertex cover problem on some other families of t-uniform hypergraphs. In his doctoral thesis, Lovász [Lov75] gave a LP rounding algorithm to obtain a factor $\frac{t}{2}$ approximation for vertex cover on t-uniform t-partite hypergraphs. This algorithm is shown to be optimal under the Unique Games Conjecture by Guruswami, Sachdeva, and Saket [GSS15], and an almost matching NP-hardness is also shown. Aharoni, Holzman, and Krivelevich [AHK96] generalized the above algorithmic result to other class of hypergraphs which have a partition of vertices obeying certain size restrictions. A factor $\frac{t}{2}$ approximation algorithm has also been obtained on subdense regular t-uniform hypergraphs [CKSV12].

For the problem of covering all paths of length t (t-Path Transversal), Lee [Lee19] gave a factor $O(\log t)$ approximation. For covering all copies of the star on t vertices, i.e., $K_{1,t-1}$, a factor $O(\log t)$ approximation is given in [GL17], and this is tight by a simple reduction from dominating set on degree t graphs. Covering 2-connected t-vertex pattern graphs (in particular t-cliques or t-cycles) is as hard as general t-uniform hypergraph vertex cover [GL17].

1.5 Open problems

A number of intriguing questions and directions come to light following our work, and we mention a few of them below.

The most obvious question is whether our algorithm for AHTP can be improved and yield approximation ratios smaller than $t/2$. Can stronger LP relaxations like Sherali-Adams help in this regard? On the hardness side, essentially nothing is known. There is a straightforward approximation preserving reduction from vertex cover on graphs to AHTP, but this only shows the hardness of beating a factor of 2. Can one show a better inapproximability factor? We do not know any good lower bound on the integrality gap of the LP either—for example, we do not know the existence of a hypergraph H for which $\tau(H^{(t-1)})/\tau^*(H^{(t-1)})$ grows with t. A natural candidate is the complete t-uniform hypergraph on n vertices for which de Cain conjectured [dC94] that in fact, $\tau(H^{(t-1)})/\tau^*(H^{(t-1)})$ grows with t. However, this is precisely the lower bound of hypergraph Turán problem, and is perhaps very hard to resolve. On the algorithmic side, obtaining $o(\log t)$ approximation algorithm for AHTP would lead to improvements on the hypergraph Turán problem: On the complete hypergraph instance, either the algorithm outputs a family $\mathcal{F} \subseteq \binom{[n]}{t-1}$ of size $o\left(\frac{\log t}{t-1}\right)\binom{n}{t}$ that covers every subset of size t, or gives a certificate that any such family should have size at least $\omega\left(\frac{1}{t}\right)\binom{n}{t-1}$. In the first case, we get an improvement on the upper bound of hypergraph Turán problem, and in the second case, we resolve de Cain's conjecture.

Similar to the $(t-1)$-blown-up hypergraphs, one can define the k-blown-up hypergraph of a t-uniform hypergraph—which will be a $\binom{t}{k}$-uniform hypergraph—and study the vertex cover problem on it. A special case of this problem when $k = 2$ is the analog of Tuza's problem for larger cliques, i.e., covering all copies of t-cliques in a graph by the fewest possible edges. Our algorithm for AHTP extends to this setting, and in particular gives an algorithm with ratio $t^2/4$ for the $k = 2$ case, beating the trivial $\binom{t}{2}$ factor (see Section 3.4 for details). Can one achieve a $o(t^2)$ factor algorithm? A simple reduction from vertex cover on t-uniform hypergraphs shows an inapproximability factor of $t - O(1)$, but can one show hardness or integrality gaps of $\omega(t)$?

In general, our work brings to the fore challenges about covering graph structures by *edges*, on both the algorithmic and hardness fronts. On the hardness side, we seem to have essentially no techniques to show strong inapproximability results, as the known PCP techniques where one naturally associates vertices with proof locations do not seem to apply. As mentioned earlier, covering all copies of a t-vertex pattern graph H with *vertices* is as hard to approximate as general t-uniform hypergraph vertex cover when H is 2-connected [GL17].

Our structural results show that the LP integrality gap (and therefore also the vertex cover to matching ratio) remains close to t on hypergraphs that exclude subgraphs absent in $(t-1)$-blown-up hypergraphs, and thus no "local" proof of Tuza-type conjectures is possible. Are there interesting families of t-uniform hypergraphs $\mathcal{F}$ such that vertex cover admits non-trivial approximation (with ratio less than t) on $\mathcal{F}$-free t-uniform hypergraphs?

What is the optimal approximation factor one can achieve for set cover on simple set systems? Is a $o(\ln n)$ approximation possible? For simple set systems with set sizes bounded by t, is there an algorithm with approximation ratio $c \ln t$ for some $c < 1$? We note that the greedy algorithm itself cannot provide such a guarantee, as our tight example for greedy in Theorem 5 uses sets of size at most $\sqrt{n}$.

For the maximization version of AHTP, where we seek to pick a specified number $(t-1)$-sized subsets to cover the largest number of edges in a t-uniform hypergraph, is there an algorithm that beats the $(1 - 1/e)$ factor (achieved by greedy for the general Max Coverage problem)? The

Johnson Coverage Hypothesis of [CKL20] asserts that for any $\varepsilon > 0$, a $(1 - 1/e + \varepsilon)$-approximation is hard to obtain for t large enough compared to ε.

We have considered covering problems in this work, and there are interesting questions concerning the dual packing problems as well. For instance, what is the approximability of packing *edge-disjoint* copies, of say t-cliques, in a graph? This is a special case of the matching problem on 2-blown-up hypergraphs. For the maximum matching problem on general k-uniform hypergraphs, also known as k-set packing, Cygan [Cyg13] gave a local search algorithm that achieves an approximation factor of $\frac{k+1}{3} + \varepsilon$ for any constant $\varepsilon > 0$. Can we get better algorithms for the maximum matching problem on blown-up hypergraphs?

On the hardness front, k-set packing is inapproximable to a $\Omega(k/\log k)$ factor [HSS06]. Known inapproximability results for the independent set problem on graphs with maximum degree k [AKS11, Cha16] imply that the maximum matching problem on k-uniform *simple* hypergraphs is hard to approximate within a $\Omega\left(\frac{k}{\log^2 k}\right)$ factor. Could maximum matching on simple hypergraphs be easier to approximate than general hypergraphs?

Organization of the paper. In Section 2, we introduce some notation and definitions. In Section 3, we describe and analyze our algorithm for AHTP and prove Theorem 1. Then, in Section 4, we prove that the analog of (generalized) Tuza's conjecture does not hold based only on local forbidden sub-hypergraph characterizations, proving Theorem 3, and also giving an explicit construction for the tent-free case posed in [AZ20]. Finally, in Section 5, we consider simple hypergraphs and prove Theorems 4 and 5.

2 Preliminaries

Notation. We use $[n]$ to denote the set $\{1, 2, \ldots, n\}$. We use $\mathbb{Z}_n$ to denote the set $\{0, 1, \ldots, n-1\}$. For a set S and an integer $1 \le k \le |S|$, we use $\binom{S}{k}$ to denote the family of all the k-sized subsets of S. A hypergraph $H' = (V', E')$ is called a subhypergraph of $H = (V, E)$ if $V' \subseteq V$ and $E' \subseteq E$. For a hypergraph $H = (V, E)$, we use $\tau(H), \nu(H)$ to denote the size of the minimum vertex cover and the maximum matching respectively. Similarly, we use $\tau^*(H)$ to denote the minimum fractional vertex cover of H:

$$\tau^*(H) = \min \left\{ \sum_{v \in V} x_v : x_v \in \mathbb{R}_{\ge 0} \ \forall v \in V, \sum_{v \in e} x_v \ge 1 \ \forall e \in E \right\}$$

We define the k-blown up hypergraph formally:

Definition 6. *For a t-uniform hypergraph $G = (V, E)$ and for an integer $1 \le k < t$, we define the k-blown up hypergraph $H = G^{(k)} = (V', E')$ as follows:*

1. *The vertex set $V' \subseteq \binom{V}{k}$ is the set of all k-sized subsets of V that are contained in an edge of G:*

$$V' = \{U : U \subseteq V, |U| = k, \exists e \in E : U \subseteq e\}$$

2. *For every edge $e \in E$, we include in E' all the k-sized subsets of e, so that*

$$E' = \left\{ e' : e' = \binom{e}{k}, e \in E \right\}$$

We will need the following Chernoff bound:

Lemma 7. *(Multiplicative Chernoff bound) Suppose $X_1, X_2, \ldots, X_n$ are independent random variables taking values in $\{0, 1\}$. Let $X = X_1 + X_2 + \ldots + X_n$, and let $\mu = \mathbb{E}[X]$. Then, for any $0 \leq \delta \leq 1$,*

$$Pr(X \leq (1 - \delta)\mu) \leq e^{-\frac{\delta^2 \mu}{2}}$$

3 LP rounding algorithm for AHTP

In this section, we present our algorithm for the AHTP and prove Theorem 1. Given a t-uniform hypergraph G as an input to the AHTP, let $H = G^{(t-1)}$ be the $(t-1)$-blown-up hypergraph of G.

3.1 Color-coding based small vertex cover

We first prove a lemma that in any $(t-1)$-blown-up hypergraph $H = ([n], E)$, there is a vertex cover of size at most $O(\frac{\log t}{t})n$ using a color-coding argument. This lemma illustrates the color-coding idea well, and is also useful later in the context of structural characterization of the blown-up hypergraphs. This lemma is not used in the main algorithm, and the reader can skip to Section 3.2 for the algorithm.

Lemma 8. *Suppose $G = ([n], E(G))$ is a t-uniform hypergraph and $H = G^{(t-1)} = (V(H), E(H))$. Then, there exists a randomized polynomial time algorithm that outputs a vertex cover of H with expected size at most $|V|\left(\frac{2\ln t}{t} + O\left(\frac{1}{t}\right)\right)$.*

Proof. Our algorithm is based on the color-coding technique used to get upper bounds for the hypergraph Turán problem [KR83, Sid95]. Let $P = \left\lceil \frac{t-1}{2\ln t} \right\rceil$. Color each vertex of G with $c : [n] \to [P]$ uniformly independently at random. For $v \in V(H)$ and $i \in [P]$, let $C_i(v)$ denote the number of nodes of v that are colored with i, i.e., $C_i(v) := |\{j \in v : c(j) = i\}|$.

We define a function $f : V(H) \to \mathbb{Z}_P$ as

$$f(v) = C_1(v) + 2C_2(v) + \ldots + (P-1)C_{(P-1)}(v) \mod P$$

For an element $i \in \mathbb{Z}_P$, let $f^{-1}(i)$ denote the set $\{v \in V(H) : f(v) = i\}$. Let $p \in \mathbb{Z}_P$ be such that $|f^{-1}(p)| \leq |f^{-1}(i)|$ for all $i \in \mathbb{Z}_P$. Note that by definition, $|f^{-1}(p)| \leq \frac{|V|}{P}$. Let $U \subseteq V(H)$ be defined as follows:

$$U = \{v : v \in V(H), \exists i \in [P] \text{ such that } C_i(v) = 0\}$$

We claim that $S = f^{-1}(p) \cup U$ is a vertex cover of H. Consider an arbitrary edge $e = \{v_1, v_2, \ldots, v_t\} \in E(H)$. Let the corresponding edge in G be equal to $e(G) = \bigcup_{j \in [t]} v_j = (u_1, u_2, \ldots, u_t) \in E(G)$ where $u_1, u_2, \ldots, u_t$ are elements of $[n]$. Without loss of generality, let $v_j = e(G) \setminus \{u_j\}$. For a color $i \in [P]$, let $C_i(e) = |\{j \in [t] : c(u_j) = i\}|$. We consider two cases separately:

1. First, if there exists a color $i \in [P]$ such that $C_i(e) = 0$, then for every $j \in [t]$, $C_i(v_j) = 0$, and thus, for every $j \in [t]$, $v_j \subseteq U$, and thus, $e \cap S \neq \phi$.

2. Suppose that for every color $i \in [P]$, $C_i(e) > 0$. We define $f(e) \in \mathbb{Z}_P$ as

$$f(e) = C_1(e) + 2C_2(e) + \ldots + (P-1)C_{(P-1)}(e) \mod P$$

Note that for every $j \in [t]$, we have

$$f(v_j) = f(e) - c(u_j) \mod P$$

As the size of $\{c(u_1), c(u_2), \ldots, c(u_t)\}$ is equal to P, the size of the set $\{f(v_1), f(v_2), \ldots, f(v_t)\}$ is equal to P as well. Thus, there exists a $j \in [t]$ such that $f(v_j) = p$ which implies that $v_j \in S$.

Thus, our goal is to upper bound the expected value of $|S|$. Note that $P \leq \frac{t-1}{\ln t}$. By taking union bound over all the colors, we get

$$\mathbb{E}[U] \leq P\left(1 - \frac{1}{P}\right)^{t-1}|V| \leq \frac{t-1}{\ln t}e^{-2\ln t}|V| \leq \left(\frac{1}{t \ln t}\right)|V| \leq O\left(\frac{1}{t}\right)|V|$$

Thus, the expected value of S is at most $|f^{-1}(p)| + \mathbb{E}[|U|]$ which is at most $\left(\frac{2\ln t}{t-1} + O\left(\frac{1}{t}\right)\right)|V|$. $\quad\square$

3.2   LP rounding based algorithm for AHTP

Consider the standard LP relaxation for vertex cover in H:

$$\text{Minimize} \sum_{v \in V(H)} x_v$$

$$\text{such that} \quad \sum_{v \in e} x_v \geq 1 \; \forall e \in E(H)$$

$$x_v \geq 0 \; \forall v \in V(H)$$

Let $\bar{x}$ be an optimal solution to the above Linear Program, and let $\mathsf{OPT} = \sum_{v \in V(H)} \bar{x}_v$. Let $S \subseteq V(H)$ be the set of vertices that are assigned positive LP value i.e.

$$S = \{v \in V(H) : \bar{x}_v > 0\}$$

We need a lemma relating $|S|$ and OPT:

Lemma 9. *The cardinality of S is at most $t \cdot \mathsf{OPT}$.*

Proof. Consider the dual of the vertex cover LP:

$$\text{Maximize} \sum_{e \in E(H)} y(e)$$

$$\text{such that} \sum_{e \ni v} y(e) \leq 1 \; \forall v \in V(H)$$

$$y(e) \geq 0 \; \forall e \in E(H)$$

Let $\bar{y}$ be an optimal solution to the above matching LP. By LP-duality, we get $\sum_{e \in E(H)} \bar{y}_e = \mathsf{OPT}$. Recall that for all $v \in S$, $\bar{x}_v \neq 0$. By the complementary slackness conditions, we get that for all $v \in S$, $\sum_{e \ni v} \bar{y}_e = 1$. Summing over all $v \in S$, we obtain

$$|S| = \sum_{v \in S} \sum_{e \ni v} \bar{y}_e \leq t \sum_{e \in E(H)} \bar{y}_e = t \cdot \mathsf{OPT}. \qquad\square$$

8

In general, OPT could be much smaller than $|V(H)|$, and thus we cannot use Lemma 8 directly to obtain algorithm for AHTP. However, we can obtain a simple $(t-1)$-factor approximation algorithm for AHTP using Lemma 8, extending the proof of fractional Tuza's conjecture of Krivelevich [Kri95]. We consider two different cases:

1. Suppose that there is a vertex $v \in V(H)$ such that $\overline{x}_v = 0$. Consider an arbitrary edge $e \in E(H)$ with $v \in e$. As $\sum_{u \in e} \overline{x}_u \geq 1$, we can infer that there is a vertex $v' \in e$ such that $\overline{x}_{v'} \geq \frac{1}{t-1}$. We round v' to 1 i.e. add v' to our vertex cover solution, delete all the edges containing v' and recursively proceed.

2. Suppose that for every vertex $v \in V(H)$, we have $\overline{x}_v > 0$. In this case, using Lemma 8, we can find a vertex cover of size $O(\frac{\log t}{t})|V(H)|$, which can be bounded above by $O(\log t)$OPT using Lemma 9.

We now describe a randomized algorithm to round the LP to obtain an integral solution whose expected size is at most $\left(\frac{t}{2} + 2\sqrt{t \ln t}\right)$ OPT. As is evident from the second case in the above $(t-1)$-factor algorithm, the problem is easy when the set of vertices $S \subseteq V(H)$ with non-zero LP value is large. Instead of considering the two different cases based on whether $S = V(H)$ or not, we take a more direct approach by finding a vertex cover of size $\left(\frac{1}{2} + o(1)\right)|S|$. Combined with Lemma 9, we get our required approximation guarantee.

For ease of notation, let $t' = \frac{t}{2} + 2\sqrt{t \ln t}$. Our first step is to round all the variables above a certain threshold to 1 (Algorithm 1). However, we need to do it recursively to ensure that we can bound the optimal value of the remaining instance.

Algorithm 1 Recursive thresholding for AHTP

1: Let $\gamma = \frac{1}{t'}$.
2: Let $\overline{x}$ be an optimal solution of the LP and let $V' = \{v : \overline{x}_v \geq \gamma\}$.
3: Let $U = V'$.
4: **while** V' is non-empty **do**
5:     Delete V' from $V(H)$, and delete all the edges $e \in E(H)$ that contain at least one vertex $v \in V'$.
6:     Solve the LP with updated H. Update $\overline{x}$ to be the new LP solution.
7:     Update $V' = \{v \in V(H) : \overline{x}_v \geq \gamma\}$. Update $U \leftarrow U \cup V'$.
8: Output U and the updated H.

Let the final updated hypergraph H when Algorithm 1 terminates be denoted by H'. Let the optimal cost of the solution $\overline{x}$ for the vertex cover on H' be denoted by OPT$'$. We prove that the size of the vertex cover output by the algorithm is not too large:

Lemma 10. *When the above recursive thresholding algorithm (Algorithm 1) terminates, we have* $|U| \leq t' \cdot (OPT - OPT')$.

Proof. We will inductively prove the following: after line 6 in the while loop of the algorithm, $|U| \leq t' \cdot (\text{OPT} - \text{OPT}_{new})$ where OPT$_{new}$ is the cost of the current optimal solution $\overline{x}$. Let $\overline{x}'$ is the optimal solution before deleting V' from H. Let OPT$_{old}$ be the cost of the solution $\overline{x}'$. By inductive hypothesis, we have $|U| - |V'| \leq t' \cdot (\text{OPT} - \text{OPT}_{old})$.

We claim that $|V'| \leq t' \cdot (\text{OPT}_{old} - \text{OPT}_{new})$. As $\overline{x}$ is an optimal vertex cover of H, we have that $\overline{x}'$ restricted to H has cost at least OPT$_{new}$. This implies that $\sum_{v \in V'} \overline{x}'_v \geq \text{OPT}_{old} - \text{OPT}_{new}$. As each $\overline{x}'_v, v \in V'$ is at least $\frac{1}{t'}$, we obtain the required claim. $\qquad \square$

We are now ready to state our main algorithm for the AHTP. The input to the algorithm is a t-uniform hypergraph G, and the output is a vertex cover for the hypergraph $H = G^{(t-1)}$.

Algorithm 2 Main algorithm

1: Apply Algorithm 1 to obtain U and let $H' = (V(H'), E(H'))$ be the updated H. Let $\overline{x}$ be an optimal solution of the vertex cover LP on H' with $\overline{x}_v \leq \gamma$ for all $v \in V(H')$.
2: Let $S \subseteq V(H')$ be defined as $S = \{v : V(H') : \overline{x}_v > 0\}$.
3: Let $\delta = \sqrt{\frac{4\ln t}{t-1}}$.
4: Color the vertices $[n]$ of G using $c : [n] \to \{0, 1\}$ uniformly and independently at random.
5: For a vertex $v \in S$ and a color $i \in \{0, 1\}$, let $C_i(v)$ denote the number of nodes that are colored with the color i i.e.   ▷ Recall that $S \subseteq V(H') \subseteq \binom{[n]}{t-1}$.

$$C_i(v) = |\{j \in v : c(j) = i\}|$$

6: Let $S' \subseteq S$ be defined as the set of vertices in S where the discrepancy between two colors is high:

$$S' = \left\{v \in S : \exists i \in \{0, 1\} : C_i(v) \leq (1 - \delta)\frac{t-1}{2}\right\}$$

7: We now define a function $f : S \to \{0, 1\}$ as $f(v) = C_1(v) \mod 2$.
8: For $i \in \{0, 1\}$, let $f^{-1}(i)$ denote the set of all the vertices $v \in S$ such that $f(v) = i$.
9: Let $p \in \{0, 1\}$ be such that $|f^{-1}(p)| \leq |f^{-1}(1-p)|$.
10: Let $T \subseteq S$ be defined as $T = S' \cup f^{-1}(p)$.
11: Output $T \cup U$.

3.3 Analysis of the algorithm and proof of Theorem 1

We will first prove that Algorithm 2 indeed outputs a valid vertex cover of H.

Lemma 11. $T \cup U$ *is a vertex cover of* H.

Proof. It suffices to prove that T is a vertex cover of H'.

Consider an arbitrary edge $e = (v_1, v_2, \dots, v_t) \in E(H')$ corresponding to the edge $e(G) = \cup_{j \in [t]} v_j = \{u_1, u_2, \dots, u_t\} \in E(G)$. Since $\overline{x}_v \leq \gamma$ for all $v \in V(H')$, we can deduce that $|e \cap S| \geq \frac{1}{\gamma} = t'$.

Our goal is to show that there exists $j \in [t]$ such that $v_j \in T$. We consider two separate cases:

1. If there is a color $i \in \{0, 1\}$ such that there are at most $(1 - \delta)\frac{t-1}{2}$ nodes of color i in $e(G)$, then for all $j \in [t], C_i(v_j) \leq (1 - \delta)\frac{t-1}{2}$. Since $e \cap S$ is non-empty, there exists $j \in [t]$ such that $v_j \in S$. By definition of S', this implies that $v_j \in S'$ as well, and thus $e \cap T \neq \phi$.

2. Suppose that in the coloring c, both the colors $0, 1$ occur at least $(1 - \delta)\frac{t-1}{2}$ times in e. Let $e' = e \cap S$ and let $k = |e'| \geq t'$. Without loss of generality, let $e' = \{v_1, v_2, \dots, v_k\}$. For every

10

$j \in [k]$, let $v_j = e(G) \setminus \{u_j\}$ for $u_j \in [n]$. First, we claim that $t - k < (1-\delta)\frac{t-1}{2}$. We have

$$
\begin{aligned}
t - k - (1-\delta)\frac{t-1}{2} &\leq t - t' - (1-\delta)\frac{t-1}{2} \\
&= \frac{t}{2} - 2\sqrt{t\ln t} - \left(1 - \sqrt{\frac{4\ln t}{t-1}}\right)\frac{t-1}{2} \\
&= \frac{1}{2}\left(t - 4\sqrt{t\ln t} - (t-1) + 2\sqrt{(t-1)\ln t}\right) \\
&\leq \frac{1}{2}\left(1 - 2\sqrt{t\ln t}\right) < 0
\end{aligned}
$$

Since each color occurs at least $(1-\delta)\frac{t-1}{2}$ times in $e(G)$, using the above, we can infer that

$$
|\{c(u_1), c(u_2), \ldots, c(u_k)\}| \geq 2.
$$

We define the value $f(e)$ in the same fashion as we have defined $f(v)$ for $v \in S$: For $i \in \{0,1\}$, let $C_i(e)$ denote the number of nodes $j \in [t]$ such that $c(u_j) = i$, and let $f(e) = C_1(e) \mod 2$. Using this definition, we get

$$
f(v_j) = f(e) - c(u_j) \mod 2 \ \forall j \in [k].
$$

As $\{c(u_1), c(u_2), \ldots, c(u_k)\} = \{0,1\}$, we have $\{f(v_1), f(v_2), \ldots, f(v_k)\} = \{0,1\}$ as well. Thus, there exists $j \in [k]$ such that $f(v_j) = p$, which proves that $v_j \in f^{-1}(p) \subseteq T$. $\qquad\square$

Note that the expected number of nodes of each color $i \in \{0,1\}$ in a vertex $v = (u_1, u_2, \ldots, u_{t-1}) \in S$ is equal to $\frac{t-1}{2}$. The set S' is the set of vertices of S where there is a color that occurs much fewer than its expected value. We prove that this happens with low probability:

Lemma 12. *The expected cardinality of S' is at most $\frac{2}{t}|S|$.*

Proof. Let $v = (u_1, u_2, \ldots, u_{t-1}) \in S$ be an arbitrary vertex in S, where $u_1, u_2, \ldots, u_{t-1}$ are elements of $[n]$. For a color $i \in \{0,1\}$, let the random variable $X(i)$ denote to the number of nodes $j \in [t-1]$ such that $c(u_j) = i$. We can write $X(i) = \sum_{j \in [t-1]} X(i,j)$, where $X(i,j)$ is the indicator random variable of the event that $c(u_j) = i$. We have $\mu = \mathbb{E}[X(i)] = \frac{t-1}{2}$. Using multiplicative Chernoff bound (Lemma 7), we can upper bound the probability that $X(i) \leq (1-\delta)\frac{t-1}{2}$ by

$$
\Pr\left(X(i) \leq (1-\delta)\frac{t-1}{2}\right) \leq e^{-\frac{\delta^2(t-1)}{4}}
$$

For the choice $\delta = \sqrt{\frac{4\ln t}{t-1}}$, the above probability is at most $\frac{1}{t}$. By applying union bound over the two colors and adding the expectation over all the vertices in S, we obtain the lemma. $\qquad\square$

Finally, we bound the expected size of the output of the algorithm:

Lemma 13. *The expected cardinality of $T \cup U$ is at most $\left(\frac{t}{2} + 2\sqrt{t\ln t}\right) \cdot OPT$.*

Proof. Note that by definition, $|f^{-1}(p)| \leq \frac{|S|}{2}$. We bound the expected size of the output of the algorithm $T \cup U$ as

$$
\begin{aligned}
\mathbb{E}[|T \cup U|] \leq \mathbb{E}[|T|] + \mathbb{E}[|U|] &\leq \mathbb{E}[|S'|] + \frac{1}{2}|S| + \mathbb{E}[|U|] \\
&\leq \left(\frac{1}{2} + \frac{2}{t}\right)|S| + \mathbb{E}[|U|] \quad \text{(Using Lemma 12)} \\
&\leq \left(\frac{t}{2} + 2\right)\mathsf{OPT'} + \mathbb{E}[|U|] \quad \text{(Using Lemma 9)} \\
&\leq \left(\frac{t}{2} + 2\sqrt{t \ln t}\right)\mathsf{OPT} \quad \text{(Using Lemma 10)} . \qquad \square
\end{aligned}
$$

Lemma 11 and Lemma 13 together imply Theorem 1.

3.4 $(t, 2)$-**version of** AHTP

An interesting generalization of AHTP is the (t, k)-version, the problem of vertex cover on the k-blown-up hypergraph $H = G^{(k)}$ for a t-uniform hypergraph G, for an arbitrary $1 \leq k < t$. The case of $k = 1$ is the standard vertex cover on t-uniform hypergraphs, and $k = t - 1$ is the AHTP. Note that there is a trivial $\binom{t}{k}$-factor approximation algorithm for this problem as it can be cast as an instance of vertex cover on a $\binom{t}{k}$-uniform hypergraph. The above algorithm can be shown to achieve a $\binom{t}{k}c(k)$ approximation guarantee for the general problem where $c(k) \to \frac{1}{2} + o(1)$ as $k \to t - 1$.

We now turn our attention to the interesting case of $k = 2$. When the hypergraph G consists of t-cliques in a graph, the vertex cover problem on $G^{(2)}$ is the generalization of Tuza's problem where we try to hit all t-cliques with the fewest possible edges. Note that in this case, the trivial hypergraph vertex cover algorithm achieves a $\binom{t}{2}$-factor approximation. We describe how a simplified version of our algorithm can be used to get a $\frac{t^2}{4}$-factor guarantee: Let $H = G^{(2)} = (V(H), E(H))$, and we iteratively solve the Vertex Cover LP on H to round all the vertices with value at least $\frac{4}{t^2}$. In the remaining instance, we let $S \subseteq V(H)$ to be the vertices of H that are assigned non-zero LP value i.e. $S = \{v \in V(H) : x_v > 0\}$. We use a color coding function $c : [n] \to \{0, 1\}$ picked uniformly and independently at random, and we output all the vertices $T = \{\{i, j\} \in S : c(i) = c(j)\}$. The expected size of T is at most $\frac{1}{2}\binom{t}{2}\mathsf{OPT} \leq \frac{t^2}{4}\mathsf{OPT}$ as the cardinality of S is at most $\binom{t}{2}\mathsf{OPT}$.

We now argue that T is indeed a vertex cover of H. Consider an edge $e = \{v_{i,j} : i \neq j \in [t]\} \in E(H)$ corresponding to the edge $e' = (u_1, u_2, \ldots, u_t) \in E(G)$. Recall that every element of e corresponds to a subset of size 2 of e', and thus, without loss of generality, let $v_{i,j} = \{u_i, u_j\}$ for all $i, j \in [t]$. As $x_{v_{i,j}} < \frac{4}{t^2}$ for all $i, j \in [t]$, there are greater than $\frac{t^2}{4}$ pairs of indices i, j such that $x_{v_{i,j}} > 0$, or equivalently, $v_{i,j} \in S$. Thus, $|e \cap S| > \frac{t^2}{4}$. For every function $c : [t] \to \{0, 1\}$, the number of pairs of indices $i \neq j \in [t]$ such that $c(i) \neq c(j)$ is at most $\frac{t^2}{4}$. Thus, there are at least $\binom{t}{2} - \frac{t^2}{4}$ pairs of indices $i \neq j \in [t]$ such that $c(u_i) = c(u_j)$. As $|e \cap S| > \frac{t^2}{4}$, there exists a pair of indices $i \neq j \in [t]$ such that $v_{i,j} \in S, c(u_i) = c(u_j)$, which implies that $v_{i,j} \in T$. Thus, for every edge $e \in E(H)$ of H, there exists an element $v \in e$ such that $v \in T$, which completes the proof that T is a vertex cover of H.

As a corollary of the algorithm, we deduce that for all hypergraphs $H = G^{(2)}$ of a t-uniform

hypergraph G,

$$\tau(H) \leq \frac{t^2}{4}\tau^*(H)$$

This proves the fractional version of a conjecture due to Aharoni and Zerbib [AZ20] (Conjecture 1.4) for the case when $k = 2$.

4   Forbidden sub-hypergraphs and Tuza's conjecture

Since AHTP is the problem of vertex cover on $H = G^{(t-1)}$ for a given t-uniform hypergraph G, an interesting question is to characterize the t-uniform hypergraphs H that can arise as the blown-up hypergraph $G^{(t-1)}$ of some t-uniform hypergraph G. A very simple necessary condition is that the hypergraph H should be simple. However, this is not sufficient—there are simple t-uniform hypergraphs H that cannot be written as $H = G^{(t-1)}$ for any G. For example, the t-tent hypergraph (Definition 14) is a simple hypergraph, but cannot be written as a $(t-1)$-blown-up hypergraph. A natural question in this context is the following:

> Is there a finite set of hypergraphs $\mathcal{F}$ such that every hypergraph that does not have any member of $\mathcal{F}$ as a sub-hypergraph can be represented as $G^{(t-1)}$ for some t-uniform hypergraph G?

In addition to its inherent structural interest, the above question can shed light on Tuza's conjecture. Recall that Aharoni and Zerbib [AZ20] proposed a generalization of Tuza's conjecture stating that $\tau\left(G^{(2)}\right) \leq 2 \cdot \nu\left(G^{(2)}\right)$ for all 3-uniform hypergraphs G. They suggested that understanding the structure of blown-up hypergraphs, and specifically, the sub-hypergraphs that it excludes might be a promising approach to establish this conjecture. In particular, they observed that the blown-up hypergraphs do not contain "tents" as a sub-hypergraph.

Definition 14. *A t-tent (Figure 1) is a set of four t-uniform edges e_1, e_2, e_3, e_4 such that*

1. $\cap_{i=1}^3 e_i \neq \phi$.

2. $|e_4 \cap e_i| = 1$ *for all $i \in [3]$.*

3. $e_4 \cap e_i \neq e_4 \cap e_j$ *for all $i \neq j \in [3]$.*

In [AZ20], the authors pose the following question. Note that an answer in the affirmative would resolve Tuza's conjecture, and in fact its above generalization that $\tau\left(G^{(2)}\right) \leq 2\nu\left(G^{(2)}\right)$ for all 3-uniform hypergraphs G.

Problem 15. *Is it true that for every 3-uniform hypergraph H without a 3-tent, $\tau(H) \leq 2 \cdot \nu(H)$?*

We answer this question in the negative. In fact, we prove a stronger statement that there can be no forbidden substructure-based Tuza's theorem.

Theorem 16. *Let $\mathcal{F} = \{F_1, F_2, \ldots, F_\ell\}$ be an arbitrary set of t-uniform hypergraphs such that for every t-uniform hypergraph G, the blown-up hypergraph $G^{(t-1)}$ does not contain any $F_i \in \mathcal{F}$ as a sub-hypergraph.*

Then, for every $\varepsilon > 0$, there exists a hypergraph H' that does not contain any member of $\mathcal{F}$ as a sub-hypergraph and which satisfies $\tau(H') \geq (t - \varepsilon)\tau^*(H') \geq (t - \varepsilon)\nu(H')$.

By setting $\mathcal{F}$ to be the single 3-tent hypergraph, we obtain a counterexample to Problem 15. Furthermore, when $t = 3$, the construction we give to prove Theorem 16 will belong to the class of 3-uniform hypergraphs H obtained from a given graph G with the vertex set of H being the edge set of G, and every triangle in G forming an edge in H. Thus, there is no "local" proof of Tuza's conjecture that uses only substructure properties of the underlying hypergraph.

We call a hypergraph non-trivial if it has at least two edges. Before we prove the above theorem, we use a definition from [FM08].

Definition 17. *Let F be a non-trivial t-uniform hypergraph. Then,*

$$\rho(F) = \max_{F' \subseteq F} \frac{e' - 1}{v' - t}$$

where F' is a non-trivial subhypergraph of F with $e' > 1$ edges and v' vertices.

We now return to the proof of Theorem 16.

Proof. (of Theorem 16) We will first prove that $\rho(F_i) > \frac{1}{t-1}$ for all $i \in [\ell]$. Suppose for contradiction that there exists a t-uniform hypergraph $F_i \in \mathcal{F}$ such that $\rho(F_i) \leq \frac{1}{t-1}$. Without loss of generality, we can assume that F_i is connected. Order the edges of F_i as $\{e_1, e_2, \ldots, e_m\}$ such that for every $j > 1$, $e_j \cap (e_1 \cup e_2 \cup \ldots \cup e_{j-1}) \neq \phi$. For every $j \geq 1$, let F'_j be the subhypergraph induced by $\{e_1, e_2, \ldots, e_j\}$. As $\rho(F_i) \leq \frac{1}{t-1}$, we can infer that for every $j > 1$,

$$\frac{|E(F'_j)| - 1}{|V(F'_j)| - t} \leq \frac{1}{t - 1}$$

which implies that $|V(F'_j)| \geq (t-1)|E(F'_j)| + 1 = (t-1)j + 1$. As $|V(F'_j)| = |V(F'_{j-1})| + t - |e_j \cap (e_1 \cup e_2 \cup \ldots \cup e_{j-1})|$, we get that $|V(F'_j)| \leq |V(F'_{j-1})| + t - 1$, which combined with the above shows that the inequality is in fact tight for every $j > 1$. Thus, for every $j > 1$, $|V(F'_j)| = |V(F'_{j-1})| + t - 1$, which implies that for every $j > 1$,

$$|e_j \cap (e_1 \cup e_2 \cup \ldots \cup e_{j-1})| = 1.$$

We now construct a t-uniform hypergraph H such that F_i is isomorphic to $H^{(t-1)}$. We construct the hypergraph H inductively via $H_1, H_2, \ldots, H_m = H$ such that $H_j^{(t-1)}$ is isomorphic to F'_j for all $j \in [m]$. First, we set the hypergraph H_1 to be equal to the t-uniform hypergraph on t vertices with a single edge. H_1 is trivially isomorphic to F'_1. Assume by inductive hypothesis that there is a hypergraph H_k such that F'_k is isomorphic to $H_k^{(t-1)}$ for some $k \in [m-1]$. Let $\phi : F'_k \to H_k^{(t-1)}$ be the isomorphism between the two hypergraphs. The hypergraph F'_{k+1} is obtained from F'_k by adding an edge e_{k+1} such that e_{k+1} intersects with F'_k in exactly one vertex $v \in V(F'_k)$. Recall that the vertex set of $H_k^{(t-1)}$ is the set of subsets of vertices of H_k of size $t - 1$. Thus, $\phi(v) = \{(p_1, p_2, \ldots, p_{t-1})\}$ for a set of vertices $p_1, p_2, \ldots, p_{t-1} \in V(H_k)$. We construct H_{k+1} by introducing a new vertex v' and adding the edge $\{v', p_1, p_2, \ldots, p_{t-1}\}$ to the hypergraph H_k. Thus, $H_{k+1}^{(t-1)}$ is obtained from $H_k^{(t-1)}$ by adding single edge that intersects with $H_k^{(t-1)}$ at exactly one vertex, that is $\{p_1, p_2, \ldots, p_{t-1}\}$. Hence, $H_{k+1}^{(t-1)}$ is isomorphic to F'_{k+1}, completing the proof. This proves that there exists a t-uniform hypergraph $H = H_m$ such that $F_i = H^{(t-1)}$, contradicting the fact that no $(t-1)$-blown-up hypergraph contains F_i as a subhypergraph.

Thus, $\rho(F_i) > \frac{1}{t-1}$ for all $i \in [\ell]$. Let $\rho = \min_{i \in [\ell]} \rho(F_i) > \frac{1}{t-1}$. Consider a random t-uniform hypergraph H' on n vertices sampled by picking each edge independently with probability $p =$

$n^{-\frac{1}{\rho}}$. We now delete the edges in a maximal collection of edge disjoint copies of members of $\mathcal{F}$ from H'. It has been proved [BFM10, FM08] that the maximum independent set $\alpha(H')$ of this construction satisfies

$$\alpha(H') \leq \tilde{O}\left(n^{\frac{1}{(t-1)\rho}}\right)$$

with high probability. Thus, there exists a t-uniform hypergraph H' with n vertices without any substructure from $\mathcal{F}$ such that $\tau(H') \geq (1 - o(1))n$. Since for any t-uniform hypergraph H' on n vertices, $\nu(H') \leq \tau^*(H') \leq \frac{1}{t}n$, this proves the claimed factor $(t - \varepsilon)$ gap between $\tau(H')$ and $\tau^*(H')$ for every positive constant $\varepsilon > 0$. $\qquad\square$

4.1 Explicit construction of tent-free hypergraphs

We now describe an explicit hypergraph giving negative answer to Problem 15. Our counterexample is a hypergraph with vertex set $[3]^n$ for large enough n and the edge set is the set of all combinatorial lines that we formally define below:

Definition 18. *(Combinatorial lines in $[3]^n$) A set of three distinct vectors $u = (u_1, u_2, \ldots, u_n), v = (v_1, v_2, \ldots, v_n), w = (w_1, w_2, \ldots, w_n) \in [3]^n$ forms a combinatorial line if there exists a subset $S \subseteq [n]$ such that*

1. *For all $i \in [n] \setminus S, u_i = v_i = w_i$.*

2. *There exist three distinct integers $u', v', w' \in [3]$ such that for all $i \in S, u_i = u', v_i = v', w_i = w'$.*

We will use the following seminal result about combinatorial lines:

Theorem 19. *(Density Hales Jewett Theorem [FK91], [Pol12]) For every positive integer k and every real number $\delta > 0$ there exists a positive integer $DHJ(k, \delta)$ such that if $n \geq DHJ(k, \delta)$ and A is any subset of $[k]^n$ of density at least δ, then A contains a combinatorial line.*

We now prove Theorem 3.

Theorem 3. *For every $\varepsilon > 0$, there exists a 3-uniform hypergraph H without a 3-tent such that $\tau(H) > (3 - \varepsilon)\nu(H)$.*

Proof. The hypergraph that we use $H = (V, E)$ has $V = [3]^n$ for n large enough to be set later, and the edges are all the combinatorial lines in $[3]^n$. First, we claim that the above defined hypergraph does not have a 3-tent. Suppose for contradiction that there are edges e_1, e_2, e_3, e_4 satisfying the properties of Definition 14. Let $u = (u_1, u_2, \ldots, u_n) \in e_4 \cap e_1, v = (v_1, v_2, \ldots, v_n) \in e_4 \in e_2, w = (w_1, w_2, \ldots, w_n) \in e_4 \cap e_3$. Note that $e_4 = \{u, v, w\}$. Thus, there exists a subset $S \subseteq [n]$ such that for all $i \in [n] \setminus S, u_i = v_i = w_i$. Without loss of generality, we can also assume that for all $i \in S, u_i = 1, v_i = 2, w_i = 3$.

Let $x = (x_1, x_2, \ldots, x_n) \in e_1 \cap e_2 \cap e_3$. Note that $\{x, u\} \subseteq e_1, \{x, v\} \subseteq e_2, \{x, w\} \subseteq e_3$. Consider an arbitrary element $p \in S$, and without loss of generality, let $x_p = 1$. Thus, we have that $x_p = 1, v_p = 2$ and both x, v share the combinatorial line e_2. This implies that there exist a subset $S_2 \subseteq [n]$ such that for all $i \in [n] \setminus S_2, x_i = v_i$ and for all $i \in S_2, x_i = 1, v_i = 2$. Similarly, there exists a subset $S_3 \subseteq [n]$ such that for all $i \in [n] \setminus S_3, x_i = w_i$ and for all $i \in S_3, x_i = 1, w_i = 3$.

Note that $S_2 \subseteq S$. Suppose for contradiction that there exists $j \in S_2 \setminus S$. Then, we have $v_j = 2, x_j = 1$. However, since $v_i = w_i$ for all $i \in [n] \setminus S$, we get that $w_j = 2$, and thus, $j \notin S_3$,

which implies that $x_j = w_j = 2$, a contradiction. Thus, $S_2 \subseteq S$, and similarly $S_3 \subseteq S$. We can also observe that $S_2 \neq S$ since in that case, $x = u$ which cannot happen since $|e_4 \cap e_2| = 1$. By the same argument on e_3, we can deduce that $S_3 \neq S$. As S_2 is a strict subset of S, there exists $j \in S \setminus S_2$. As $v_i = x_i$ for all $i \in [n] \setminus S_2$, $x_j = v_j = 2$. As $j \in S$, we have $w_j = 3$. However, as $w_j \neq x_j$, this implies that $j \in S_3$, which then implies that $x_j = 1$, a contradiction.

Now, we will prove that for large enough n, $\tau(H) > (3 - \varepsilon)\nu(H)$. Let $N = 3^n$. Since the cardinality of V is equal to N, we have $\nu(H) \leq \frac{N}{3}$. We apply Theorem 19 with $k = 3, \delta = \frac{\varepsilon}{3}$, and set $n \geq \mathsf{DHJ}(k, \delta)$. Thus, we can infer that in any subset $T \subseteq V$ of size $\frac{\varepsilon}{3}N$, there exists an edge of H fully contained in T. Thus, we get that $\tau(H) > (1 - \frac{\varepsilon}{3})N$, which gives $\tau(H) > (3 - \varepsilon)\nu(H)$. $\square$

5 Vertex cover and set cover on simple hypergraphs

As mentioned earlier, the edges in a $(t-1)$-blown-up hypergraph of a t-uniform hypergraph can intersect on at most one element, so such hypergraphs are simple. In this section, we will take a step back and address to what extent improved approximation algorithms are possible for vertex cover on simple hypergraphs. We will also consider the dual problem, of covering the vertices by the fewest possible hyperedges, namely the set cover problem, on simple hypergraphs but without any restriction on the size of the hyperedges. Note that a hypergraph is simple if and only if the edge-vertex incidence bipartite graph does not contain a copy of $K_{2,2}$. Thus, a hypergraph is simple if and only if its dual is simple.

5.1 Vertex cover on simple t-uniform hypergraphs

We now prove Theorem 4 which shows that simple hypergraphs are still rich enough to preclude a non-trivial approximation to vertex cover. Our hardness is established using a reduction from the general problem of vertex cover on t-uniform hypergraphs. In particular, we use the following result:

Theorem 20. *([DGKR05]) For every constant $\varepsilon > 0$ and $t \geq 3$, the following holds: Given a t-uniform hypergraph $G = (V, E)$, it is NP-hard to distinguish between the following cases:*

1. *Completeness: G has a vertex cover of measure $\frac{1+\varepsilon}{t-1}$.*

2. *Soundness: Any subset of V of measure ε contains an edge from E.*

We give a randomized reduction from Theorem 20 to Theorem 4. The approach is similar to the one used in [GL17] for showing the inapproximability of H-Transversal in graphs. It was also used in the recent tight hardness for Max Coverage on simple set systems [CKL21].

Let us instantiate Theorem 20 with ε replaced by $\varepsilon' = \frac{\varepsilon}{4}$, and let the resulting hypergraph be denoted by G. Now, given this t-uniform hypergraph $G = (V, E)$, we output a t-uniform hypergraph $H = (V', E')$ as follows: Let $n = |V|, m = |E|$. We have integer parameters B, P depending on ε, t, n, m to be set later. The vertex set of H is $V' = V \times [B]$–we have a cloud of B vertices $v^1, v^2, \ldots, v^B$ in V' corresponding to every vertex $v \in V$. For every edge $e = (v_1, v_2, \ldots, v_t) \in E$, we pick P edges $e^1, e^2, \ldots, e^P$ with $e^i = ((v_1)_i, (v_2)_i, \ldots, (v_t)_i)$ and add them to E', where for each $j \in [t]$ and $i \in [P]$, $(v_j)_i$ is chosen uniformly and independently at random from $(v_j)^1, (v_j)^2, \ldots, (v_j)^B$. Thus, so far, we have added mP edges to E'.

We first upper bound the expected value of the number of pairs of edges in E' that intersect in more than one vertex. Order the edges in E' as $e_1, e_2, \ldots, e_{mP}$. Let X denote the random

variable that counts the number of pairs of edges in E' that intersect in more than one vertex. For every pair of indices $i, j \in [mP]$, let the random variable X_{ij} be the indicator variable of the event that the edges e_i and e_j of E' intersect in greater than one vertex. Note that the edges in E corresponding to e_i and e_j have at most t vertices in common. Thus, the probability that e_i and e_j intersect in at least two vertices is upper bounded by $\binom{t}{2} \frac{1}{B^2}$. Summing over all the pairs i, j, we get

$$\mathbb{E}[X] \le \binom{mP}{2} \binom{t}{2} \frac{1}{B^2} \le \frac{m^2 t^2 P^2}{B^2}.$$

By Markov's inequality, with probability at least $\frac{9}{10}$, X is at most $\frac{10 m^2 t^2 P^2}{B^2}$.

We consider all the pairs of edges that intersect in more than one vertex in E', and arbitrarily delete one of those edges. Let the resulting set of edges be denoted by E''. The final hypergraph resulting in this reduction is $H = (V', E'')$. Note that H is indeed a simple hypergraph. We will prove the following:

1. (Completeness) If G has a vertex cover of measure μ, then there is a vertex cover of measure μ in H.

2. (Soundness) If every subset of V of measure ε' contains an edge from E, then with probability at least $\frac{4}{5}$, every subset of V' of measure ε contains an edge from E''.

Completeness. If G has a vertex cover of size μn, then picking all the vertices in V' in the cloud corresponding to these vertices ensures that H has a vertex cover of size $\mu n B$. Thus, in the completeness case, there is a vertex cover of measure μ in H.

Soundness. Suppose that every ε' measure subset of V contains an edge from E. Our goal is to show that with probability at least $\frac{4}{5}$, every ε measure subset of V' contains an edge from E''. We first prove the following lemma:

Lemma 21. *With probability at least $\frac{9}{10}$ over the choice of E', the following holds: For every edge $e = (v_1, v_2, \ldots, v_t) \in E$, and every subset $S \subseteq V'$ such that for each $i \in [t]$, S contains at least $\frac{\varepsilon}{4} B$ vertices from $\{v_i^1, v_i^2, \ldots, v_i^B\}$, there exists an edge $e' \in E'$ all of whose vertices are in S.*

Proof. The probability that there exists an edge $e = (v_1, v_2, \ldots, v_t)$ and a subset S which contains at least $\frac{\varepsilon}{4} B$ vertices from each cloud and does not contain any edge from E' is at most

$$m 2^{tB} \left(1 - \left(\frac{\varepsilon}{4}\right)^t\right)^P \le m 2^{tB - \log e \frac{\varepsilon^t P}{4^t}} \le \frac{1}{10}$$

when $P = maB$ where $a := a(t, \varepsilon) = \frac{4^{t+2} t}{\varepsilon^t}$. $\qquad\square$

Using the above lemma, we can conclude that with probability at least $\frac{4}{5}$, $X \le \frac{10 m^2 t^2 P^2}{B^2} = 10 m^4 t^2 a^2$ and for every edge $e \in E$ and every subset $S \subseteq V'$ such that for each $i \in [t]$, S contains at least $\frac{\varepsilon}{4} B$ vertices from $\{v_i^1, v_i^2, \ldots, v_i^B\}$, there exists an edge $e' \in E'$ all of whose vertices are in S. We claim that this implies that with probability at least $\frac{4}{5}$, every ε measure subset of V' contains an edge of E''. Consider an arbitrary subset $U \subseteq V'$ such that $|U| \ge \varepsilon n B$. We choose B large enough such that $t(10 m^4 t^2 a^2) \le \frac{\varepsilon}{2} n B$. Thus, the set of the vertices W in the edges deleted from E' to obtain E'' has cardinality at most $\frac{\varepsilon}{2} n B$.

17

Let $U' = U \setminus W$. Note that all the edges in U' that are in E' are present in E'' as well. As U' has a measure of at least $\frac{\varepsilon}{2}$ in V', for at least $\frac{\varepsilon}{4}n$ vertices v in V, U' should contain at least $\frac{\varepsilon}{4}$ fraction of the vertices in the cloud $\{v^1, v^2, \ldots, v^B\}$. Since otherwise, the cardinality of U' is at most $\left(n - \frac{\varepsilon n}{4}\right) \cdot \frac{\varepsilon B}{4} + \frac{\varepsilon n}{4} \cdot B < \frac{\varepsilon n B}{2}$, a contradiction. By Lemma 21, we can deduce that there exists an edge $e \in E'$ all of whose vertices are in U', which implies that the edge e is in E'' as well. This proves that in the soundness case, with probability at least $\frac{4}{5}$, there exists an edge in every ε measure subset of V'.

This completes the proof of Theorem 4. Under the Unique Games Conjecture [Kho02], the hardness of vertex cover in t-uniform hypergraphs can be improved to $t - \varepsilon$. We remark that we can get the same hardness for simple hypergraphs by our reduction.

5.2 Set Cover on Simple Set Systems

In the set cover problem, there is a set family $\mathcal{S} \subseteq 2^X$ on a universe $X = [n]$, and the goal is to cover the universe $[n]$ with as few sets from the family as possible. The greedy algorithm where we repeatedly pick the set that covers the maximum number of new elements achieves a $\ln n$-factor approximation algorithm for the problem, and this is known to be optimal. We consider the same problem under the restriction that the family $\mathcal{S}$ is a simple set system i.e. for every $i \neq j, |S_i \cap S_j| \leq 1$. Surprisingly, in contrast with the hardness result for vertex cover, simplicity of the set family helps in achieving better approximation factor for the set cover problem.

Theorem 22. *(Theorem 5 restated) The greedy algorithm achieves a $\left(\frac{\ln n}{2} + 1\right)$-approximation guarantee for the set cover problem on simple set systems over a universe of size n. Furthermore, the bound is essentially tight for the greedy algorithm—there is a simple set system on which the approximation factor of greedy exceeds $\left(\frac{\ln n}{2} - 1\right)$.*

Proof. First, we prove the upper bound. Let the optimal solution size be equal to k i.e. there is $\mathcal{T} = \{S_1, S_2, \ldots, S_k\} \subseteq \mathcal{S}$ such that the union of sets in $\mathcal{T}$ is equal to $[n]$. For every set $S \in \mathcal{S} \setminus \mathcal{T}$, $|S \cap S_i| \leq 1$ by the simplicity of the set system, and thus, we get that

$$\forall S \in \mathcal{S} \setminus \mathcal{T}, |S| \leq k. \tag{1}$$

We now consider two different cases:

1. Suppose that $k \geq \sqrt{n}$. We recall that the greedy algorithm in fact achieves a $\log |S_{max}|$-factor approximation algorithm for set cover on general instances where $|S_{max}|$ is the size of the largest set in the family. Thus, after the greedy algorithm picks t sets each of which cover at least $\sqrt{n}$ new elements, in the remaining instance, we have $|S_{max}| \leq \sqrt{n}$. As there are k sets that cover the remaining instance, the greedy algorithm picks at most $k\frac{\ln n}{2}$ sets after picking the t sets. As each of the t sets cover at least $\sqrt{n}$ new elements, $t \leq \sqrt{n}$. Overall, the total number of sets used by the greedy algorithm is equal to

$$t + k\frac{\ln n}{2} \leq \sqrt{n} + k\frac{\ln n}{2} \leq \left(1 + \frac{\ln n}{2}\right)k$$

2. Suppose that $k < \sqrt{n}$. In this case, using (1), we can infer that there are at most k sets with size at least k in the family. Thus, after the greedy algorithm picks k sets, in the remaining instance, each set has size at most k, and thus, greedy algorithm picks at most $k \ln k$ sets. Overall, the total number of sets picked by the greedy algorithm is equal to

$$k + k\ln k \leq k + k\frac{\ln n}{2} = \left(1 + \frac{\ln n}{2}\right)k$$

18

Thus, in both the cases, the greedy algorithm picks at most $k\left(1 + \frac{\ln n}{2}\right)$ sets.

A hard instance for the greedy algorithm. We now prove that the above bound is tight for the greedy algorithm. Fix a large integer k, and let $n = k^2$, $X = [n]$. We first add k sets to the family $\mathcal{S}$ $S_1, S_2, \ldots, S_k$ where $S_j = \{(j-1)k+1, (j-1)k+2, \ldots, jk\}$. Note that these k sets together cover the whole universe X. We view the universe X as k blocks, with the j'th block comprising of the set S_j.

We now add $m = (k-1)\ln k$ additional pairwise disjoint sets $T_1, T_2, \ldots, T_m$ to $\mathcal{S}$ such that the greedy algorithm picks the set T_i in the i'th iteration. We choose the sets $T_i, i \in [m]$, as follows:

1. For $j \in [k]$, let the set X_j be the uncovered elements of the block S_j. Let $a_j = |X_j|$ for all $j \in [k]$. We initially set $X_j = S_j$ for all $j \in [k]$.

2. At every iteration $i \in [m]$:

   (a) Sort the elements $\{a_1, a_2, \ldots, a_k\}$ such that $a_{\alpha_1} \geq a_{\alpha_2} \geq \ldots \geq a_{\alpha_k}$ where $(\alpha_1, \alpha_2, \ldots, \alpha_k)$ is a permutation of $[k]$. Let $p = a_{\alpha_1} \leq k$ be the largest number of uncovered elements in a block.

   (b) Let $P = \{\alpha_1, \alpha_2, \ldots, \alpha_p\}$. For $l \in [p]$, let $u_l \in [n]$ be equal to the largest element in X_{α_l}.

   $$u_l = \max\{b \mid b \in X_{\alpha_l}\}$$

   We set

   $$T_i = \{u_l \mid l \in [p]\}$$

   Furthermore, we set $X_{\alpha_l} = X_{\alpha_l} \setminus u_l$ for all $l \in [p]$. We also update $a_j, j \in [k]$ as $a_j = |X_j|$ for all $j \in [k]$.

In the above procedure to output the sets $T_i, i \in [m]$, the cardinality of $|T_i| \geq |T_{i+1}|$ for all i. Furthermore, in the ith iteration of the above procedure, the cardinality of T_i is at least the number of elements in any block that are not covered yet. This ensures that the greedy algorithm in the ith iteration picks the set T_i. Furthermore, as the sets T_is are all mutually disjoint, and intersect each block at most once, the resulting set system is indeed a simple set system.

Our goal is to prove that after all the m sets are picked, there are still uncovered elements in $[n]$. For an integer $i \in [m]$, we let $s_i \in [n]$ denote the number of elements not covered by the greedy algorithm before the set T_i is picked. For $i \in [m]$, the size of the set T_i picked by the greedy algorithm in the ith iteration is equal to the largest number of uncovered elements in a block i.e. the value of a_{α_1} in the ith iteration. Based on the updating procedure followed above, we can infer that this value is equal to $|T_i| = \left\lceil \frac{s_i}{k} \right\rceil$. This follows from the fact that at any iteration of the above procedure, the sorted values a_{α_1} and a_{α_k} satisfy $a_{\alpha_1} \leq a_{\alpha_k} + 1$.

We have $s_1 = n = k^2$, and

$$s_{i+1} = s_i - |T_i| = s_i - \left\lceil \frac{s_i}{k} \right\rceil \geq s_i \left(1 - \frac{1}{k}\right) - 1$$

By setting $t_i = s_i + k$ for $i \in [m]$, we get

$$t_{i+1} \geq t_i \left(1 - \frac{1}{k}\right)$$

19

Thus, we get

$$
\begin{aligned}
t_{m+1} &\geq t_1 \left(1 - \frac{1}{k}\right)^m \\
&= (k^2 + k)\left(1 - \frac{1}{k}\right)^{(k-1)\ln k} \\
&\geq (k^2 + k)\exp\left(-\frac{\frac{1}{k}}{1 - \frac{1}{k}}(k-1)\ln k\right) \quad \text{(Using } 1 - x \geq e^{\frac{-x}{1-x}} \, \forall \, 0 \leq x < 1) \\
&= k + 1
\end{aligned}
$$

Thus, $s_{m+1} \geq 1$, which proves that there are elements that are not covered after the greedy algorithm uses $m = (k-1)\ln k$ sets. This completes the proof that there are simple set systems on n elements with $k = \sqrt{n}$ sets covering all the elements where as the greedy algorithm picks at least $(k-1)\ln k \geq k\left(\frac{\ln n}{2} - 1\right)$ sets. $\square$

Turning to the hardness of set cover on simple set systems, we note that the inapproximability result of Theorem 20 holds for $t = (\ln n)^c$ for an absolute constant $c > 0$, with a weaker completeness guarantee of a vertex cover of measure $\approx 2/t$. Combining it with our reduction in the proof of Theorem 4, and recalling that the dual of a simple hypergraph is also simple, we can deduce that set cover on simple set systems is NP-hard to approximate to a factor better than $(\ln n)^{\Omega(1)}$. It remains an interesting question to determine the exact approximability of set cover on simple set systems.

Acknowledgment

References

[Aha01]   Ron Aharoni. Ryser's conjecture for tripartite 3-graphs. *Comb.*, 21(1):1–4, 2001. 2

[AHK96]   Ron Aharoni, Ron Holzman, and Michael Krivelevich. On a theorem of Lovász on covers in tau-partite hypergraphs. *Combinatorica*, 16(2):149–174, 1996. 4

[AKS11]   Per Austrin, Subhash Khot, and Muli Safra. Inapproximability of vertex cover and independent set in bounded degree graphs. *Theory Comput.*, 7(1):27–43, 2011. 6

[AZ20]   Ron Aharoni and Shira Zerbib. A generalization of Tuza's conjecture. *Journal of Graph Theory*, 94(3):445–462, 2020. 1, 2, 3, 6, 13

[BCD20]   Patrick Bennett, Ryan Cushman, and Andrzej Dudek. Closing the random graph gap in tuza's conjecture through the online triangle packing process, 2020. 1

[BFG19]   Fábio Botler, Cristina G. Fernandes, and Juan Gutiérrez. On tuza's conjecture for triangulations and graphs with small treewidth. In *Proceedings of the tenth Latin and*

American Algorithms, Graphs and Optimization Symposium, LAGOS 2019, Belo Horizonte, Brazil, June 2-7, 2019, volume 346 of *Electronic Notes in Theoretical Computer Science*, pages 171–183. Elsevier, 2019. 1

[BFM10]   Tom Bohman, Alan M. Frieze, and Dhruv Mubayi. Coloring Hlo-free hypergraphs. *Random Struct. Algorithms*, 36(1):11–25, 2010. 3, 15

[Cha16]   Siu On Chan. Approximation resistance from pairwise-independent subgroups. *J. ACM*, 63(3):27:1–27:32, 2016. 6

[CKL20]   Vincent Cohen-Addad, Karthik C.S., and Euiwoong Lee. On approximability of k-means, k-median, and k-minsum clustering. *Manuscript*, 2020. 4, 6, 20

[CKL21]   Vincent Cohen-Addad, Karthik C. S., and Euiwoong Lee. On approximability of clustering problems without candidate centers. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, 2021. To appear. 4, 16

[CKSU20]   Parinya Chalermsook, Samir Khuller, Pattara Sukprasert, and Sumedha Uniyal. Multi-transversals for triangles and the tuza's conjecture. In Shuchi Chawla, editor, *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA 2020, Salt Lake City, UT, USA, January 5-8, 2020*, pages 1955–1974. SIAM, 2020. 1

[CKSV12]   Jean Cardinal, Marek Karpinski, Richard Schmied, and Claus Viehmann. Approximating vertex cover in dense hypergraphs. *Journal of discrete algorithms*, 13:67–77, 2012. 4

[CL99]   Fan Chung and Linyuan Lu. An upper bound for the Turán number t3(n, 4). *Journal of Combinatorial Theory, Series A*, 87(2):381–389, 1999. 2

[Cyg13]   Marek Cygan. Improved approximation for 3-dimensional matching via bounded pathwidth local search. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, Berkeley, CA, USA*, 2013. 6

[dC94]   D. de Caen. The current status of Turán's problem on hypergraphs. *Extremal Problems for Finite Sets*, 1991, Bolyai Soc. Math. Stud., Vol. 3, pp. 187–197, János Bolyai Math. Soc., Budapest, 1994. 5

[DGKR05]   Irit Dinur, Venkatesan Guruswami, Subhash Khot, and Oded Regev. A new multilayered PCP and the hardness of hypergraph vertex cover. *SIAM J. Comput.*, 34(5):1129–1146, 2005. 16

[DLR95]   Richard A. Duke, Hanno Lefmann, and Vojtech Rödl. On uncrowded hypergraphs. *Random Struct. Algorithms*, 6(2/3):209–212, 1995. 3

[Erd81]   Paul Erdős. On the combinatorial problems which I would most like to see solved. *Combinatorica*, 1(1):25–42, 1981. 3

[Erd88]   Paul Erdös. Problems and results in combinatorial analysis and graph theory. *Graph Theory and Applications, Proceedings of the First Japan Conference on Graph Theory and Applications*, pages 81–92, 1988. 3

[FHMW17]   Nevena Francetic, Sarada Herke, Brendan D. McKay, and Ian M. Wanless. On Ryser's conjecture for linear intersecting multipartite hypergraphs. *Eur. J. Comb.*, 61:91–105, 2017. 3

[FK91]   H. Furstenberg and Y. Katznelson. A density version of the Hales-Jewett theorem. *J. Anal. Math.*, 57:64–119, 1991. 3, 15

[FM08]   Alan M. Frieze and Dhruv Mubayi. On the chromatic number of simple triangle-free triple systems. *Electron. J. Comb.*, 15(1), 2008. 3, 14, 15

[FM13]   Alan M. Frieze and Dhruv Mubayi. Coloring simple hypergraphs. *J. Comb. Theory, Ser. B*, 103(6):767–794, 2013. 3

[GK98]   Sudipto Guha and Samir Khuller. Greedy strikes back: Improved facility location algorithms. In Howard J. Karloff, editor, *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, 25-27 January 1998, San Francisco, California, USA,* pages 649–657. ACM/SIAM, 1998. 4

[GL17]   Venkatesan Guruswami and Euiwoong Lee. Inapproximability of H-transversal/packing. *SIAM J. Discret. Math.*, 31(3):1552–1571, 2017. 4, 5, 16

[GSS15]   Venkatesan Guruswami, Sushant Sachdeva, and Rishi Saket. Inapproximability of minimum vertex cover on k-uniform k-partite hypergraphs. *SIAM J. Discret. Math.*, 29(1):36–58, 2015. 4

[Hax99]   Penny E. Haxell. Packing and covering triangles in graphs. *Discret. Math.*, 195(1-3):251–254, 1999. 1

[HR01]   Penny E. Haxell and Vojtech Rödl. Integer and fractional packings in dense graphs. *Comb.*, 21(1):13–38, 2001. 1

[HSS06]   Elad Hazan, Shmuel Safra, and Oded Schwartz. On the complexity of approximating k-set packing. *Computational Complexity*, 15(1):20–39, 2006. 6

[Kho02]   Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings on 34th Annual ACM Symposium on Theory of Computing,* pages 767–775, 2002. 18

[KP20]   Jeff Kahn and Jinyoung Park. Tuza's conjecture for random graphs, 2020. 1

[KR83]   K. H. Kim and F. W. Roush. *On a problem of Turán,* pages 423–425. Birkhäuser Basel, Basel, 1983. 7

[Kri95]   Michael Krivelevich. On a conjecture of Tuza about packing and covering of triangles. *Discret. Math.*, 142(1-3):281–286, 1995. 1, 9

[Lee19]   Euiwoong Lee. Partitioning a graph into small pieces with applications to path transversal. *Math. Program.*, 177(1-2):1–19, 2019. 4

[Lov75]   Láslzó Lovász. On minmax theorems of combinatorics, Doctoral thesis. *Mathematiki Lapok*, 26, 1975. 4

[LZ09]   Linyuan Lu and Yi Zhao. An exact result for hypergraphs and upper bounds for the Turán density of K_{r+1}^r. *SIAM J. Discret. Math.*, 23(3):1324–1334, 2009. 2

[Pol12]   D.H.J. Polymath. A new proof of the density Hales-Jewett theorem. *Annals of Mathematics*, 175(3):1283–1327, May 2012. 3, 15

[Pul15]   Gregory J. Puleo. Tuza's conjecture for graphs with maximum average degree less than 7. *Eur. J. Comb.*, 49:134–152, 2015. 1

[Sid95]   Alexander Sidorenko. What we know and what we do not know about Turán numbers. *Graphs Comb.*, 11(2):179–199, 1995. 7

[Sid97]   Alexander Sidorenko. Upper bounds for Turán numbers. *J. Comb. Theory, Ser. A*, 77(1):134–147, 1997. 2

[Tur41]   Paul Turán. On an extremal problem in graph theory. *Mat. Fiz. Lapok*, 48:436–452, 1941. 2

[Tur61]   Paul Turán. Research problem. *Közl MTA Mat. Kutató Int.*, 6:417–423, 1961. 2

[Tuz81]   Zsolt Tuza. Conjecture. *Finite and Infinite Sets*, Proc. Colloq. Math. Soc. Janos Bolyai:888, 1981. 1

[Tuz90]   Zsolt Tuza. A conjecture on triangles of graphs. *Graphs Comb.*, 6(4):373–380, 1990. 1

[Yus12]   Raphael Yuster. Dense graphs with a large triangle cover have a large triangle packing. *Comb. Probab. Comput.*, 21(6):952–962, 2012. 1