



Simplified inapproximability of hypergraph coloring via t -agreeing families

Per Austrin* Amey Bhangale † Aditya Potukuchi ‡

April 2, 2019

Abstract

We reprove the results on the hardness of approximating hypergraph coloring using a different technique based on bounds on the size of extremal t -agreeing families of $[q]^n$. Specifically, using theorems of Frankl-Tokushige [FT99], Ahlswede-Khachatrian [AK98] and Frankl [Fra76] on the size of such families, we give simple and unified proofs of quasi NP-hardness of the following problems:

- coloring a 3 colorable 4-uniform hypergraph with $(\log n)^\delta$ many colors
- coloring a 3 colorable 3-uniform hypergraph with $\tilde{O}(\sqrt{\log \log n})$ many colors
- coloring a 2 colorable 6-uniform hypergraph with $(\log n)^\delta$ many colors
- coloring a 2 colorable 4-uniform hypergraph with $\tilde{O}(\sqrt{\log \log n})$ many colors

where n is the number of vertices of the hypergraph and $\delta > 0$ is a universal constant.

1 Introduction

We study the fundamental problem of coloring hypergraphs with minimum number of colors. A k -uniform hypergraph $H(V, E)$ consists of a collection of vertices V and a set of hyperedges $E \subseteq \binom{V}{k}$.

A coloring $\chi : V \rightarrow [c]$ is called a *proper coloring* of H if no hyperedge $e \in E$ is monochromatic with respect to the coloring χ . The chromatic number of a hypergraph is the minimum number of colors needed to properly color it.

In this paper we study the problem of *approximating* the chromatic number of a given hypergraph. More specifically, we study the following problem: Given a c -colorable k -uniform hypergraph, find a proper D -coloring of it in polynomial time for a given $D \geq c$. There has been rich history of studying the computational complexity of finding or approximating the chromatic number of graphs as well as hypergraphs. It is known that unless $P = NP$, approximating the chromatic number of an n vertex graph to within a factor of $n^{1-\varepsilon}$ is hard

*KTH Royal Institute of Technology. austrin@kth.se. Research supported by the Approximability and Proof Complexity project funded by the Knut and Alice Wallenberg Foundation.

†Weizmann Institute of Science, Rehovot, Israel. amey.bhangale@weizmann.ac.il. Research supported by Irit Dinur's ERC-CoG grant 772839.

‡Department of Computer Science, Rutgers University. aditya.potukuchi@cs.rutgers.edu. Research supported by Mario Szegedy's NSF grant CCF-1514164

Uniformity k	Colors c	Hardness D
3	3	$\Omega((\log \log n)^{1/9})$ [Kho02b]
3	3	$2^{\frac{\log \log n}{\log \log \log n}}$ [GHH ⁺ 14]
3	2	$\Omega((\log \log n)^{1/9})$ [DRS02] *no independent set guarantee
3	3	$\Omega((\log \log n)^{1/2})$ [this paper]
4	2	$\Omega(\frac{\log \log n}{\log \log \log n})$ [Hol02, GHS02]
4	$q \geq 7$	$(\log n)^{\Omega(q)}$ [Kho02a]
4	2	$(\log n)^{\Omega(1)}$ [Sak14]
4	4	$2^{2^{\sqrt{\log \log n}}}$ [GHH ⁺ 14]
4	4	$2^{(\log n)^{\Omega(1)}}$ [Var15]
4	2	$(\log n)^{\Omega(1)}$ [Bha18] *no independent set guarantee
4	2	$\Omega((\log n \log n)^{1/2})$ [this paper]
4	3	$(\log n)^{\Omega(1)}$ [this paper]
6	2	$(\log n)^{\Omega(1)}$ [DG15]
6	2	$(\log n)^{\Omega(1)}$ [this paper]
8	2	$2^{2^{\sqrt{\log \log n}}}$ [GHH ⁺ 14]
8	2	$2^{(\log n)^{\Omega(1)}}$ [Var15]
8	2	$2^{(\log n)^{1/10-o(1)}}$ [Hua15]
12	2	$2^{(\log n)^{\Omega(1)}}$ [KS14]

Figure 1: Comparison of the known results and our results on the hardness of hypergraph coloring.

for all $\varepsilon > 0$ [FK98, Zuc06]. The best approximation algorithms currently known for chromatic number give an approximation factor guarantee of $O\left(\frac{n(\log \log n)^2}{(\log n)^3}\right)$ [Hal93].

Given these results, a lot of attention has been devoted to understanding the complexity of finding a proper coloring given the guarantee that the hypergraph has very small, even constant, chromatic number. In fact, it is NP-hard to decide if a given hypergraph is 2-colorable or not, unlike the graph case where it is easy to decide if the graph is bipartite or not in polynomial time. The best polynomial time algorithms currently known require $n^{\Omega(1)}$ colors to color a 2-colorable hypergraph [KNS01, CF96, AKMH96, KT17]. Since finding a proper D -coloring is at least as hard as finding the independent set of size n/D , a lot of attention went into studying the following (computationally easier) problem:

Definition 1.1 ($\text{HYPCOL}_n(k, c, D)$). *Given a k -uniform hypergraph $H(V, E)$ on n vertices which is c -colorable, find an independent set of size $\frac{n}{D}$.*

The study of the complexity of approximate hypergraph coloring was initiated by Guruswami *et al.* [GHS02]. Holmerin [Hol02] and Guruswami *et al.* [GHS02] showed that $\text{HYPCOL}_n(4, 2, \frac{\log \log n}{\log \log \log n})$ is quasi NP-hard¹. Khot [Kho02b, Kho02a] showed quasi NP-hardness of $\text{HYPCOL}_n(4, q, (\log n)^{cq})$ for some $c > 0$ and all $q \geq 7$ and $\text{HYPCOL}_n(3, 3, O(\log \log n)^{1/9})$. Dinur-Guruswami [DG15] showed $\text{HYPCOL}_n(6, 2, (\log n)^c)$ is quasi NP-hard. Saket [Sak14] improved the state for 2-colorable 4-uniform hypergraph by showing $\text{HYPCOL}_n(4, 2, (\log n)^c)$ is quasi NP-hard for some $c > 0$.

¹Recall that if a problem is quasi NP-hard then it cannot be solved in quasipolynomial time $2^{\text{poly} \log n}$ unless all problems in NP can be solved in quasipolynomial time.

Guruswami *et al.* [GHH⁺14] broke the logarithmic barrier for the first time and showed quasi NP-hardness of $\text{HYPCOL}_n(8, 2, 2^{2^{\sqrt{\log \log n}}})$, $\text{HYPCOL}_n(4, 4, 2^{2^{\sqrt{\log \log n}}})$ and $\text{HYPCOL}_n(3, 3, 2^{\frac{\log \log n}{\log \log \log n}})$ using the *short code*. Finally, Khot and Saket [KS14] improved the factor to almost polynomial by showing quasi NP-hardness of $\text{HYPCOL}_n(12, 2, 2^{(\log n)^{\Omega(1)}})$. Building on the work of Khot and Saket [KS14], Varma [Var15] showed quasi NP-hardness of $\text{HYPCOL}_n(8, 2, 2^{(\log n)^{\Omega(1)}})$ as well as $\text{HYPCOL}_n(4, 4, 2^{(\log n)^{\Omega(1)}})$. In the 2-colorable case, Huang [Hua15] independently showed quasi NP-hardness of $\text{HYPCOL}_n(8, 2, 2^{(\log n)^{1/20-o(1)}})$.

In terms of approximating the chromatic number for 2 colorable hypergraphs, Dinur *et al.* [DRS02] showed that it is quasi NP-hard to color 2-colorable 3-uniform hypergraph by $(\log \log n)^{1/9}$. This result is weaker than showing hardness for $\text{HYPCOL}_n(3, 2, (\log \log n)^{1/9})$. In fact, it is still open to determine the complexity of $\text{HYPCOL}_n(3, 2, c)$ for any $c = \omega(1)$. Very recently, the second author [Bha18] showed NP-hardness of coloring 2-colorable 4-uniform hypergraph with $(\log n)^c$ colors for some $c > 0$.

1.1 Our results

We give unified proofs of many of the known results on the hardness of approximate hypergraph coloring. Our analysis makes a novel use of the maximum size of t -agreeing families.

We now state the theorems that we (re)prove. See Section 1.3 for the comparison with the previous works on hypergraph coloring.

Our first result gives an alternate proof of the result of Khot [Kho02a] for q -colorable 4-uniform hypergraph for smaller values of q .

Theorem 1.2. *There exists a constant $\delta > 0$ such that $\text{HYPCOL}_n(4, 3, \log^\delta n)$ is quasi NP-hard.*

For the 3-uniform hypergraph, Khot [Kho02b] and Dinur *et al.* [DRS02] start with a multi-layered Label Cover instance, which was one of the highlights of their proofs. In our proof, we also start with this multi-layered Label Cover instance, but simplify² the inner verification step. Our proof is almost along the lines of proof of Dinur *et al.* [DRS02], but we get a stronger independent set guarantee recovering (slightly improving) the result of Khot [Kho02b].

Theorem 1.3. $\text{HYPCOL}_n(3, 3, \tilde{O}(\sqrt{\log \log n}))$ is quasi NP-hard.

We note that the above theorem gives a weaker bound than the result of [GHH⁺14], $\text{HYPCOL}_n(3, 3, 2^{\frac{\log \log n}{\log \log \log n}})$, which is the best known result for 3-colorable 3-uniform hypergraph.

Both the previous results have in completeness case a hypergraph which is 3-colorable. Our next theorem shows hardness of finding independent sets for 2-colorable hypergraphs, but with slightly larger uniformity. This gives another proof of [DG15].

Theorem 1.4. *There exists a constant $\delta > 0$ such that $\text{HYPCOL}_n(6, 2, \log^\delta n)$ is quasi NP-hard.*

We also extend our techniques to prove the following hardness for 2-colorable 4-uniform hypergraph which, up to a quadratic factor, recovers the result of [GHS02] and [Hol02].

Theorem 1.5. $\text{HYPCOL}_n(4, 2, \tilde{O}(\sqrt{\log \log n}))$ is quasi NP-hard.

²modulo the theorem on t -agreeing family, which we use as a black box.

In this case too, Saket [Sak14] gets a better guarantee.

In addition to giving alternate proofs of the known results, our proofs give rise to interesting questions about *t-agreeing families*. If solved in a positive way could lead to improved inapproximability of hypergraph coloring for lower uniformity hypergraphs. See Section 6 for more details.

1.2 Proof overview

For the proof overview, we are going to think of a Label Cover instance as a regular graph $G(V, E, [L])$ on V with the alphabet $[L]$. The edges of the graph are labeled with d -to- d constraints ϕ_e for every $e \in E$, for some $1 \leq d \leq L$. The instance is always *regular* which means that there exists $\delta > 0$ such that any induced graph on $\alpha|V|$ vertices has at least $\Omega(\alpha)$ fraction of the constraints, for $\alpha \geq 1/L^\delta$.

We now give a proof overview of quasi-NP-hardness of $\text{HYPCOL}_n(4, 3, \log^\delta n)$. The starting point is the gap Label Cover problem where distinguishing between the cases when the Label Cover instance is satisfiable vs. no assignment can satisfy more than $L^{-\varepsilon}$ fraction of the edges is NP-hard, for some constant $\varepsilon > 0$. Given an instance $G(V, E, L)$ of gap Label Cover, we reduce it to a 4-uniform hypergraph as follows. We replace every vertex v with a cloud $C[v]$ of size 3^L , where the vertex in a cloud is referred by a pair (v, a) for $a \in [3]^L$. Thus, the number of vertices in the hypergraph is $|V| \cdot 3^L$. Now for every edge $e(u, v)$ in G we put a hyperedge between $(u, a_u), (u, b_u)$ and $(v, a_v), (v, b_v)$ iff for every label i and j such that (i, j) satisfies the constraint ϕ_e , $\{a_u(i), b_u(i), a_v(j), b_v(j)\}$ are not all equal.

The completeness case is easy: Given a labeling $\ell : V \rightarrow [L]$ to G satisfying all the edges, we color (v, a) with the color $a_{\ell(v)}$. It is easy to see that any hyperedge $\{(u, a_u), (u, b_u), (v, a_v), (v, b_v)\}$ will get at least two distinct colors *by construction*.

Let us consider the more interesting soundness case. Suppose the Label Cover instance is only $1/L^\varepsilon$ satisfiable. Suppose the hypergraph has an independent set \mathcal{I} of fractional size $\alpha \in [0, 1]$. Then by simple averaging argument, there exist at least $\alpha/2$ fraction of the vertices $v \in V$ such that $|\mathcal{I} \cap C[v]| \geq \frac{\alpha}{2} \cdot 3^L$. Let $X \subseteq V$ be the set of such vertices. For every $v \in X$, using a theorem on t -agreeing families (Theorem 3.5), there must exist $a_v, b_v \in \mathcal{I} \cap C[v]$ such that they agree on at most $t := O(\log(1/\alpha))$ coordinates. Denote L_v be the set of such coordinates. Now, for an edge (u, v) such that both $u, v \in X$, it must be the case that there exist $i \in L_u$ and $j \in L_v$ such that (i, j) satisfies $\pi_{(u,v)}$. In fact, if this was not the case then the t -agreeing pairs from the clouds $C[v]$ and $C[u]$ (which was used to define the sets L_u and L_v) form a valid hyperedge! Thus, for every vertex $v \in X$, we have a small list of at most t labels. If we assign a random label from L_v to v for all $v \in X$ then any edge (u, v) , such that both $u, v \in X$, is satisfied with probability at least $1/t^2$. By the *regularity* of the Label Cover instance, there are $\Omega(\alpha)$ fraction of edges in X and hence the labeling satisfies at least $\Omega(\alpha/t^2)$ fraction of the total edges in G in expectation. Setting $\alpha \approx L^{-\varepsilon/3}$ gives a contradiction. Thus, if we set L such that $3^L \gg |V|$, there is an inverse logarithmic dependency between the size of the hypergraph ($\approx 3^L$) and the lower bound on the fractional size of the independent set $\Omega(1/L^{\varepsilon/3})$ which proves Theorem 1.2.

The reduction and the analysis of 6-uniform hypergraph construction (Theorem 1.4) is exactly the same as before. Instead of 2-wise t -agreeing family of $[3]^L$, we use 3-wise t -agreeing family of $\{0, 1\}^L$ (Theorem 3.6).

The reduction and analysis in the proofs of Theorems 1.3 and 1.5 are similar to each other where the starting point is the layered Label Cover instance.

1.3 Comparison with previous works

Previous reductions [Kho02b, Kho02a, GHH⁺14, KS14] showing independent set guarantee require *smoothness* property of the Label Cover instance (See Section 3.1 for the definition of Label Cover). The smoothness property roughly says that for any *small* list of labels to $u \in U$, these labels are projected to different labels, with high probability, if we choose a random constraint attached to u . The reason they need this property is rather technical and there is no intuitive reason for it. Saket’s work [Sak14] uses following structural property of the Label Cover instance: There is a universal constant $c_0 > 0$, such that for any $u \in U$ and $S \subseteq [L]$

$$\Pr_{v \sim N(u)} [|\phi_{u \rightarrow v}(S)| < |S|^{c_0}] \leq |S|^{-c_0}.$$

Therefore, all the previous proofs of Theorem 1.2 and 1.3 exploited the special structure of the Label Cover instance constructed from the PCP Theorem [AS98, FGL⁺96, ALM⁺98] and the parallel repetition theorem of Raz [Raz98].

In our proofs of Theorem 1.2, and 1.4, we do not need any structure on the projection constraints of the Label Cover. In fact, for Theorem 1.2 and 1.4, we can start with a gap instance of 2-CSP over alphabet $[L]$ and *arbitrary* constraints with completeness 1 and soundness $\frac{1}{L^\varepsilon}$ for some $\varepsilon > 0$.³ In the current exposition though, we start with the usual gap Label Cover instance to keep things simple. For our proofs of Theorem 1.3 and Theorem 1.5 though, we need the smoothness property of the layered Label Cover instance.

Previous works including [DRS02] and [Bha18] which use agreement based decoding (like ours), only show hardness for approximating chromatic number. In fact, their hypergraphs always contain an independent set of size around half in both the completeness and soundness cases. Our reductions and analyses are very similar to [DRS02] and [Bha18], but we get independent set guarantee by using theorems on the extremal t -agreeing families.

We also like to point out that although our proofs are modular and require weaker conditions on the Label Cover instance (in two cases), the earlier reductions which use Fourier analysis often prove a stronger statement. More specifically, in the soundness case, we show that the maximum sized independent set is upper bounded by αn , whereas the previous reductions usually prove a robust statement of the form - every subset of vertices of size αn contains at least $f(\alpha)$ fraction of edges. We make no attempt to confirm the stronger soundness guarantees. Nonetheless, it would be interesting to know if this holds for our reduction.

2 Organization

We start with preliminaries first, where we define Label Cover and variants of it in Section 3.1. We then state and prove results on the size of k -wise t -agreeing of $[q]^n$ in Section 3.2. In Section 4, we prove a general Theorem 4.1, where the starting point is the Label Cover instance. Theorem 1.2 and 1.4 follow as corollaries of Theorem 4.1.

³to get the quasi NP-hardness, we would still need the reduction from 3-SAT of size n to the gap instance of 2-CSP to run in time $n^{(\log L)^C}$ for some constant $C \geq 0$.

In Section 5, we prove a general Theorem 5.1, where the starting point is a multi-layered Label Cover instance. Theorem 1.3 and 1.5 follow as corollaries of Theorem 5.1.

3 Preliminaries

We use $[q]$ to denote the set $\{1, 2, \dots, q\}$ and for a string $x \in [q]^n$, we use $x(i)$ to denote the element at its i^{th} location. We use $A \sqcup B$ to denote a disjoint union of two sets A and B .

3.1 Label Cover

A *Label Cover instance* is given by a tuple $\mathcal{L} = (U, V, E, [L], [R], \Phi)$. The variables of the instance are $U \sqcup V$, with the variables in U taking values in $[L]$ and the variables in V taking values in $[R]$. We have a set of edges $E \subseteq U \times V$ and for every $(x, y) \in E$, there is a constraint $\phi_{x \rightarrow y} \in \Phi$. Moreover, these constraints are *projection constraints*: this means (slightly abusing notation), that there is a map $\phi_{x \rightarrow y} : [L] \rightarrow [R]$ for every $(x, y) \in E$ such that, for every $i \in [L]$, $\phi_{x \rightarrow y}(i)$ is the unique assignment to y that satisfies $\phi_{x \rightarrow y}$.

We will use $\phi_{x \rightarrow y}$ to denote both the projection map as well as the constraint itself, when there is no ambiguity. For $A \subseteq U$ and $B \subseteq V$, we define

$$\Phi(A, B) := \{\phi_{x \rightarrow y} \mid (x, y) \in (A \times B) \cap E\}.$$

For an $x \in U$ and $A \subseteq V$, we define

$$N_A(x) = \{y \mid (x, y) \in E \text{ and } y \in A\}.$$

In the case that $A = V$, we drop the subscript and denote $N(x) = N_V(x)$. These notations are symmetric and by a slight abuse of notation, will also be used when $x \in V$ and $A \subseteq U$.

We have the following theorem as a result of the PCP Theorem [AS98, FGL⁺96, ALM⁺98] and the parallel repetition theorem of Raz [Raz98].

Theorem 3.1. *There is a $C > 0$ such that the following holds: For any parameter $b \in \mathbb{N}$, there exists a reduction from a 3-SAT instance of n variables to a Label Cover instance with at most $n^{O(b)}$ variables over an alphabet of size at most $2^{O(b)}$. The Label Cover instance has the following completeness and soundness conditions:*

1. *If the 3-SAT instance is satisfiable, then there exists an assignment to the Label Cover instance that satisfies all the constraints.*
2. *If the 3-SAT instance is not satisfiable, then every assignment to the Label Cover instance satisfies at most a 2^{-Cb} fraction of the constraints.*

Moreover, the Label Cover instance produced is bi-regular and the reduction runs in time $n^{O(b)}$.

3.1.1 Multi-layered Label Cover

We also have the multilayered Label Cover problem, which will be used as a starting point for proving Theorems 1.3 and 1.5. An ℓ -layered Label Cover instance is a tuple $\mathcal{L} = (\mathcal{U}, \mathcal{R}, E, \Phi)_\ell$.

Here, $\mathcal{U} = \{U_i \mid i \in [\ell]\}$ where each U_i is a set of variables. The variables in U_i take values in $[R_i]$ and $\mathcal{R} = \{[R_i] \mid i \in [\ell]\}$ are the sets of labels. We also have a set of edges $E = \{E_{i,j} \subseteq U_i \times U_j \mid 1 \leq i < j \leq \ell\}$ and a set of constraints Φ where for $1 \leq i < j \leq \ell$, the constraints in $\Phi(U_i, U_j)$ project from the variables in U_i to the variables in U_j .

In using multi-layered Label Covers, we will require a bit more of structure on them.

Definition 3.2 (Weakly Dense). *A multi-layered Label Cover instance $\mathcal{L} = (\mathcal{U}, \mathcal{R}, E, \Phi)_\ell$ is weakly dense if for any $1 \leq m \leq \ell$, any sequence of distinct integers $1 \leq i_1 < \dots < i_m \leq \ell$, and any sequence of sets $S_j \subseteq U_{i_j}$ such that $|S_j| \geq \frac{2}{m}|U_{i_j}|$ for all $j \in [m]$, there are two sets S_k and $S_{k'}$ such that $\Phi(S_k, S_{k'}) \geq \frac{1}{m^2}|\Phi(U_{i_k}, U_{i_{k'}})|$.*

Definition 3.3 (Smoothness). *A multi-layered Label Cover instance $\mathcal{L} = (\mathcal{U}, \mathcal{R}, E, \Phi)_\ell$ is T -smooth if for every $1 \leq i < j \leq \ell$, $x \in U_i$ and $S \subseteq [R_i]$, it holds that*

$$\Pr_{y \in N_{U_j}(x)} [|\phi_{x \rightarrow y}(S)| < |S|] \leq \frac{|S|^{2\ell}}{T}.$$

We have the following theorem which gives us hardness for the smooth weakly dense multi-layered Label Cover problem.

Theorem 3.4. *[DRS02, DGKR05, Kho02b] For any parameters $T \gg \ell \geq 2$, $r \in \mathbb{N}$, there exists a reduction from 3-SAT instances of size n to ℓ -layered weakly dense T -smooth Label Cover instances with layers U_1, \dots, U_ℓ , and $n^{O((T+\ell)r)}$ variables over a range of size $2^{O(\ell r)}$. The Label Cover instance has the following completeness and soundness conditions:*

1. *If the 3-SAT instance is satisfiable, then there exists an assignment to the Label Cover instance that satisfies all the constraints.*
2. *If the 3-SAT instance is not satisfiable, then for every $1 \leq i < j \leq \ell$, every assignment to the Label Cover instance satisfies at most $2^{-\Omega(r)}$ fraction of the constraints between layers U_i and U_j .*

Moreover, the reduction runs in $n^{O((T+\ell)r)}$ time.

3.2 Bounds on t -agreeing Families

For an alphabet Σ and strings $a_1, \dots, a_k \in \Sigma^n$ we define the *agreement* $\text{ag}(a_1, \dots, a_k)$ to be the set of coordinates where all k strings agree, i.e.,

$$\text{ag}(a_1, \dots, a_k) = \{i \in [n] \mid a_1(i) = a_2(i) = \dots = a_k(i)\}.$$

If $|\text{ag}(a_1, \dots, a_k)| \geq t$, we say that a_1, \dots, a_k are *t -agreeing*. We say that a family $\mathcal{F} \subseteq \Sigma^n$ is *k -wise t -agreeing* if all subsets of k strings $a_1, \dots, a_k \in \mathcal{F}$ are t -agreeing. For the special case $k = 2$ we drop the “ k -wise” and simply call the family t -agreeing for brevity.

We have the following bound on the maximal t -agreeing subfamily of $[3]^n$:

Theorem 3.5 ([FT99, AK98]). *Let n and t be integers such that $n \geq 3t - 1$. Then for every t -agreeing family $\mathcal{F} \subset [3]^n$ it holds that*

$$|\mathcal{F}| \leq 3^{n-3t+1} \sum_{i=0}^{t-1} \binom{3t-1}{i} 2^i.$$

It is easy to check that such a family \mathcal{F} as described above has size at most $3^{n-t/10}$ for large enough t .

For the proofs of Theorems 1.4 and 1.5, we need a similar theorem, but for a family of subsets of $\{0, 1\}^n$. Note that here, there are t -agreeing families of $\{0, 1\}^n$ of size at least $2^{n-1}(1 - o(1))$ for $t = o(\sqrt{n})$ (by taking all strings of Hamming weight $\leq (n-t)/2$). However, for the maximum size of a 3-wise t -agreeing family we have a similar upper bound as Theorem 3.5.

Theorem 3.6. *Let n and t be integers such that $n \geq t$. Then for every 3-wise t -agreeing $\mathcal{F} \subset \{0, 1\}^n$ it holds that*

$$\frac{|\mathcal{F}|}{2^n} \leq \left(\frac{\sqrt{5} - 1}{2} \right)^t.$$

We say that a family $\mathcal{F} \subset \{0, 1\}^n$ is 3-wise t -intersecting if for every $a, b, c \in \mathcal{F}$ it holds that $|\{i \in [n] \mid a(i) = b(i) = c(i) = 1\}| \geq t$.

Theorem 3.7 ([Fra76]). *Let n and t be integers such that $n \geq t$. Then for every 3-wise t -intersecting $\mathcal{F} \subset \{0, 1\}^n$ it holds that*

$$\frac{|\mathcal{F}|}{2^n} \leq \left(\frac{\sqrt{5} - 1}{2} \right)^t.$$

Theorem 3.7 essentially follows⁴ from a beautiful proof of Frankl [Fra76] (see also [Fra18] where this statement is made explicit), on the size of 3-wise t -intersecting families. The proof of Theorem 3.6 now follows using a standard *shifting* argument.

Proof of Theorem 3.6. Let $\mathcal{F} \subseteq \{0, 1\}^n$ be any 3-wise t -agreeing family. Iteratively for every $i = 1, 2, \dots, n$ we do the following shifting of $\mathcal{F}_0 \stackrel{\text{def}}{=} \mathcal{F}$ to get families $\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_n$. If $x \in \mathcal{F}_{i-1}$ is such that $x(i) = 0$ and $x \oplus e_i \notin \mathcal{F}_{i-1}$ ⁵, then in \mathcal{F}_i we replace x with $x \oplus e_i$ (and otherwise we keep x in \mathcal{F}_i).

We have following two invariants after every iteration: (1) the size of the family remains unchanged. (2) the modified family is still a 3-wise t -agreeing family. (1) is obvious. To see that (2) holds, we assume for contradiction that \mathcal{F}_i is no longer a 3-wise t -agreeing family, whereas \mathcal{F}_{i-1} is a 3-wise t -agreeing family. This means that there exists $a', b', c' \in \mathcal{F}_i$ such that $\text{ag}(a', b', c') < t$. Since in the i^{th} iteration, only the i^{th} location gets affected, it must be the case that $a'(i), b'(i), c'(i)$ are not all the same. Without loss of generality, assume $a'(i) = 0$ and $a' \in \mathcal{F}_{i-1}$. This means that $a' \oplus e_i \in \mathcal{F}_{i-1}$. Up to symmetry between b' and c' we have two cases to handle:

Case 1: $b' \in \mathcal{F}_{i-1}$ and $c \in \mathcal{F}_{i-1}$ such that $c(i) = 0$ and $c' = c \oplus e_i$.

In this case, $\text{ag}(a', b', c') = \text{ag}(a' \oplus e_i, b', c) < t$, regardless of $b'(i)$, contradicting the fact that \mathcal{F}_{i-1} is a 3-wise t -agreeing family, as $\{a' \oplus e_i, b', c\} \subseteq \mathcal{F}_{i-1}$.

Case 2: $b, c \in \mathcal{F}_{i-1}$ such that $b(i) = c(i) = 0$ and $b' = b \oplus e_i$ and $c' = c \oplus e_i$.

In this case, $\text{ag}(a' \oplus e_i, b, c) = \text{ag}(a', b', c') < t$, contradicting the fact that \mathcal{F}_{i-1} is a 3-wise t -agreeing family, as $\{a' \oplus e_i, b, c\} \subseteq \mathcal{F}_{i-1}$.

⁴Frankl proves this for $t = 3$ but the proof can easily be seen to work by a minor modification of Proposition 3 in [Fra76].

⁵the operation $x \oplus e_i$ flips the i^{th} bit of x .

We keep reiterating the above n -step process until $\mathcal{F}_0 = \mathcal{F}_n$ (by letting $\mathcal{F}_0 \leftarrow \mathcal{F}_n$ at the start). The process must halt, since at each n -step iteration if $\mathcal{F}_0 \neq \mathcal{F}_n$, then we have increased the total hamming weights of strings in \mathcal{F}_0 by at least 1. If at the end of the n -step process, we have $\mathcal{F}_0 = \mathcal{F}_n$, then this condition means that \mathcal{F}_n is a *monotone* 3-wise t -agreeing family.⁶

We now claim that \mathcal{F}_n is in fact a 3-wise t -intersecting family. Suppose not, this means that there are $a, b, c \in \mathcal{F}_n$ such that $S_1 = \{i \in [n] \mid a(i) = b(i) = c(i) = 1\}$ and $|S_1| < t$. Let $S_0 = \{i \in [n] \mid a(i) = b(i) = c(i) = 0\}$. Now, the monotonicity of \mathcal{F}_n gives that there exists $a' \in \mathcal{F}_n$ such that $a'(i) = a(i)$ for all $i \in [n] \setminus S_0$ and $a'(i) = 1$ for all $i \in S_0$. We can conclude that $\text{ag}(a', b, c) = |S_1| < t$, contradicting the fact that \mathcal{F}_n is a 3-wise t -agreeing family.

Using Theorem 3.7, we then have

$$\frac{|\mathcal{F}_n|}{2^n} \leq \left(\frac{\sqrt{5} - 1}{2} \right)^t,$$

and since $|\mathcal{F}| = |\mathcal{F}_n|$, the theorem follows. \square

3.3 Other combinatorial lemmas

We will use another simple combinatorial lemma:

Lemma 3.8. *Let \mathcal{F} be a family of multisets of subsets of $[n]$ of size at most t . Suppose for every $S_1, \dots, S_d \in \mathcal{F}$, there are distinct $i, j \in [t]$ such that $S_i \cap S_j \neq \emptyset$, then there exists an $i \in [n]$ such that*

$$|\{S \in \mathcal{F} \mid i \in S\}| \geq \frac{1}{t(d-1)} |\mathcal{F}|.$$

Proof. Let $S_1, \dots, S_r \in \mathcal{F}$ be a maximal pairwise disjoint subfamily of \mathcal{F} , and note that $r \leq d - 1$. Consider $S \stackrel{\text{def}}{=} \bigsqcup_{i \in [r]} S_i$, and for $s \in S$ let $\mathcal{F}_s \stackrel{\text{def}}{=} \{T \in \mathcal{F} \mid s \in T\}$. Because S_1, \dots, S_r form a maximal pairwise disjoint family, we have that $\mathcal{F} = \cup_{s \in S} \mathcal{F}_s$. On the other hand $|S| = tr \leq t(d-1)$. Therefore, there is some $s \in S$ such that $|\mathcal{F}_s| \geq \frac{1}{t(d-1)} |\mathcal{F}|$. \square

In particular, setting $d = 2$ in the above lemma, we have the following simple corollary:

Corollary 3.9. *Let \mathcal{F} be a family of multisets of subsets of $[n]$ of size at most t that is also intersecting, then there exists an $i \in [n]$ such that*

$$|\{S \in \mathcal{F} \mid i \in S\}| \geq \frac{1}{t} |\mathcal{F}|.$$

4 Poly-logarithmic hardness with large uniformity

Theorem 4.1. *Let k, q , and c be such that for some $\delta = 1/\text{poly } m$ and all sufficiently large m , no family of $[q]^m$ of size at least $\delta \cdot q^m$ is k -wise δ^{-c} -agreeing. Then $\text{HYPOL}_n(2k, q, \text{poly } \log n)$ is quasi NP-hard.*

⁶In fact, after just the first n -step iteration, the family becomes monotone.

As immediate corollaries, it follows by plugging in Theorem 3.5 that $\text{HYPCOL}(4, 3, \text{poly log } n)$ is quasi NP-hard (Theorem 1.2), and plugging in Theorem 3.6 that $\text{HYPCOL}(6, 2, \text{poly log } n)$ is quasi NP-hard (Theorem 1.4). We note that those Theorems on t -agreeing families in fact give bounds of the form $O(\log 1/\delta)$ on the amount of agreement, much better than the $\text{poly } 1/\delta$ needed by Theorem 4.1.

We now proceed to prove the theorem and begin by describing the reduction. Consider a Label Cover instance $\mathcal{L} = (U, V, E, [R], [L], \Phi)$. We reduce it to a hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ whose vertices and edges are as follows:

Vertices \mathcal{V} : The vertex set is obtained by replacing each variable $x \in U$ by a cloud $[q]^L$ of vertices: for a variable $x \in U$, denote

$$\mathcal{V}_x := \{(x, a) \mid a \in [q]^L\}.$$

The vertex set of \mathcal{H} is given by

$$\mathcal{V} = \bigcup_{x \in U} \mathcal{V}_x.$$

Edges \mathcal{E} : For every $y \in V$ and $x_1, x_2 \in N(y)$, there is a hyperedge on a set of $2k$ vertices $\{(x_1, a_1), \dots, (x_1, a_k), (x_2, b_1), \dots, (x_2, b_k)\}$ if they have the property that for every $i_1, i_2 \in [L]$ such that $\phi_{x_1 \rightarrow y}(i_1) = \phi_{x_2 \rightarrow y}(i_2)$, it holds that

$$\left| \bigcup_{j=1}^k \{a_j(i_1), b_j(i_2)\} \right| \geq 2.$$

4.1 Completeness

Lemma 4.2. *If there is an assignment to the variables of $U \sqcup V$ that satisfies all the constraints in Φ , then $\chi(\mathcal{H}) \leq q$.*

Proof. Let $A : U \sqcup V \rightarrow [L] \cup [R]$ be the assignment that satisfies all the constraints of \mathcal{L} . Consider the coloring that colors vertex $(x, a) \in U \times [q]^L$ with the color $a(A(x))$. Suppose for contradiction that this yields a monochromatic hyperedge $\{(x_1, a_1), \dots, (x_1, a_k), (x_2, b_1), \dots, (x_2, b_k)\}$ for some $x_1, x_2 \in N(y)$ and some $y \in V$.

Monochromaticity implies that

$$\left| \bigcup_{j=1}^k \{a_j(A(x_1)), b_j(A(x_2))\} \right| = 1.$$

However, since A satisfies both $\phi_{x_1 \rightarrow y}$ and $\phi_{x_2 \rightarrow y}$, we also have that

$$\phi_{x_1 \rightarrow y}(A(x_1)) = \phi_{x_2 \rightarrow y}(A(x_2)) = A(y).$$

Taken together, these contradict the condition for being an edge. □

4.2 Soundness

Lemma 4.3. *If $\alpha(\mathcal{H}) > 2\delta$ then there exists an assignment to $U \sqcup V$ that satisfies at least a δ^{1+2c} fraction of the constraints Φ (where δ and c are as in Theorem 4.1).*

Proof. Let \mathcal{I} be an independent set in \mathcal{H} of size $|\mathcal{I}| \geq 2\delta|\mathcal{V}|$. For every variable $x \in U$, let $\mathcal{I}_x \stackrel{\text{def}}{=} \mathcal{I} \cap \mathcal{V}_x$. By an averaging argument there exists an $X \subseteq U$ such that:

1. $|X| > \delta|U|$.
2. $|\mathcal{I}_x| > \delta|\mathcal{V}_x|$ for every $x \in X$.

We will henceforth restrict our attention to variables in X .

By the Theorem assumption that all k -wise δ^{-c} -agreeing families of $[q]^n$ have size at most δq^n , it follows that for every $x \in X$, there are k vertices $(x, a_{x,1}), \dots, (x, a_{x,k})$ in \mathcal{I}_x such that $|\text{ag}(a_{x,1}, \dots, a_{x,k})| < \delta^{-c} \stackrel{\text{def}}{=} t$. Define $L_x = \text{ag}(a_{x,1}, \dots, a_{x,k}) \subseteq [L]$.

Another observation is that for $y \in V$,

$$\phi_{x_1 \rightarrow y}(L_{x_1}) \cap \phi_{x_2 \rightarrow y}(L_{x_2}) \neq \emptyset \text{ for every } x_1, x_2 \in N_X(y). \quad (1)$$

Indeed, if this were not the case, one can check that

$$\{(x_1, a_{x_1,1}), \dots, (x_1, a_{x_1,k}), (x_2, a_{x_2,1}), \dots, (x_2, a_{x_2,k})\}$$

is a hyperedge, contradicting our assumption that these vertices are from an independent set.

We now define a (randomized) labelling $A : U \sqcup V \rightarrow [L] \cup [R]$ of \mathcal{L} . For $x \in X$, pick a label $A(x)$ from L_x at random.

For $y \in N(X)$, define the (multi-)family $\mathcal{F}(y) \stackrel{\text{def}}{=} \{\phi_{x \rightarrow y}(L_x) \mid x \in N_X(y)\}$ of subsets of $[R]$. By (1), $\mathcal{F}(y)$ is an intersecting family where every set has size at most t . Therefore, Lemma 3.9 implies that there is a label $i \in [R]$ that is present in at least $\frac{1}{t} \cdot |\mathcal{F}(y)|$ sets in $\mathcal{F}(y)$. Define $A(y)$ to be that label. For the remaining variables $x \in U \setminus X$ and $y \in V \setminus N(X)$, assign a label arbitrarily.

By the choice of $A(y)$, for every $y \in N(X)$, a $\frac{1}{t}$ fraction of all $x \in N_X(y)$ have a label $i \in L_x$ such that $\phi_{x \rightarrow y}(i) = A(y)$, and each such constraint is satisfied by A with probability at least $1/|L_x| \geq 1/t$. It follows that the expected number of constraints satisfied by the labelling A is at least $\frac{1}{t^2} |\Phi(X, V)|$.

By the regularity of the Label Cover instance, $|\Phi(X, V)| = \frac{|X|}{|U|} |\Phi| = \delta |\Phi|$. Thus A satisfies (in expectation) at least a $\delta/t^2 = \delta^{1+2c}$ fraction of all constraints. This proves the existence of an assignment that achieves the above guarantee. \square

Proof of Theorem 1.2. Start with a 3-SAT instance Π on n variables. Setting $b = \log \log n$, Theorem 3.1 gives us a Label Cover instance $\mathcal{L} = (U, V, E, [L], [R], \Phi)$ with $n^{O(\log \log n)}$ variables taking values over an alphabet of size at most $2^{O(b)} = \text{poly } \log n$. Applying the reduction above gives us a hypergraph \mathcal{H} on $|\mathcal{V}| = q^{\text{poly } \log n} \cdot n^{O(b)} = 2^{\text{poly } \log n}$ vertices. In the completeness case, Lemma 4.2, gives us $\chi(\mathcal{H}) \leq q$. On the other hand, in the soundness case, then no assignment to \mathcal{L} will satisfy more than $2^{-Cb} = (\log n)^{-C}$ fraction of the constraints, where C is the constant from Theorem 3.1, and so by Lemma 4.3, $\alpha(\mathcal{H}) \leq (\log n)^{-C/(1+2c)} = 1/\text{poly } \log |\mathcal{V}|$. \square

Remark 4.4. *In the analysis, we never use the fact that the constraints are projection constraints. Thus, one can start with any gap 2-CSP over $[L]$ and carry out the reduction.*

5 poly log log-hardness with smaller uniformity

In this section, we prove a following general theorem.

Theorem 5.1. *Let k, q , and c be such that for some $\delta = 1/\text{poly } m$ and all sufficiently large m , no family of $[q]^m$ of size at least $\delta \cdot q^m$ is k -wise $(c \log(1/\delta))$ -agreeing. Then $\text{HYPCOL}_n(k+1, q, \tilde{\Omega}(\sqrt{\log \log n}))$ is quasi NP-hard.*

As immediate corollaries, it follows using Theorem 3.5 that $\text{HYPCOL}(3, 3, \tilde{\Omega}(\sqrt{\log \log n}))$ is quasi NP-hard (Theorem 1.3), and using Theorem 3.6 that $\text{HYPCOL}(4, 2, \tilde{\Omega}(\sqrt{\log \log n}))$ is quasi NP-hard (Theorem 1.5).

For this reduction, we start with an ℓ -layered Label Cover $\mathcal{L} = (\mathcal{U}, \mathcal{R}, E, \Phi)_\ell$, where $\mathcal{U} = \{U_i \mid i \in [\ell]\}$ is the set of layers, $\mathcal{R} = \{[R_i] \mid i \in [\ell]\}$ are the sets of labels, $\Phi = \{\Phi_{i \rightarrow j} \mid i, j \in [\ell]\}$ are the sets of constraints. We reduce it to a hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ whose vertices and edges as follows:

Vertices \mathcal{V} : The vertex set is obtained by replacing each variable x in each layer U_i by a cloud $[q]^{R_i}$ of vertices: for a variable $x \in U_i$, denote

$$\mathcal{V}_x := \{(x, a) \mid a \in [q]^{R_i}\}.$$

The vertex set is

$$\mathcal{V} = \bigcup_{U_i \in \mathcal{U}} \bigcup_{x \in U_i} \mathcal{V}_x.$$

Edges \mathcal{E} : For every $1 \leq i < j \leq \ell$, $y \in U_j$ and $x \in N_{U_i}(y)$, $k+1$ vertices

$$\{(x, a_1), (x, a_2), \dots, (x, a_k), (y, b)\}$$

forms a hyperedge if for every $r \in [R_i]$, it holds that

$$|\{a_1(r), a_2(r), \dots, a_k(r), b(\phi_{x \rightarrow y}(r))\}| \geq 2. \quad (2)$$

In what follows we write $U = \bigsqcup_{i=1}^{\ell} U_i$ to denote the set of all variables of \mathcal{L} .

5.1 Completeness

Lemma 5.2. *If there is an assignment to the variables of U that satisfy all the constraints of Φ , then the $\chi(\mathcal{H}) \leq q$.*

Proof. Let $A : U \rightarrow \bigcup_{i \in [\ell]} [R_i]$ be an assignment that satisfies all the constraints in Φ . A proper q -coloring is given by coloring a vertex $(x, a) \in U_i \times [q]^{R_i}$ with the color $a(A(x))$. Suppose for contradiction that this is not a proper coloring and that there is a monochromatic edge $\{(x, a_1), \dots, (x, a_k), (y, b)\}$ for some $y \in U_j$ and $x \in N_{U_i}(y)$.

Since A satisfies $\phi_{x \rightarrow y}$, we have $\phi_{x \rightarrow y}(A(x)) = A(y)$. However, monochromaticity implies that

$$|\{a_1(A(x)), \dots, a_2(A(x)), b(A(y))\}| = 1$$

This contradicts the condition for $\{(x, a_1), \dots, (x, a_k), (y, b)\}$ being a hyperedge. \square

5.2 Soundness

Lemma 5.3. *Let k, q, c and δ be as in Theorem 5.1, $t = c \log(1/\delta)$, $\ell = 2/\delta^2$, and suppose that \mathcal{L} is T -smooth for some $T \geq \frac{16(c \log 1/\delta)^2 \ell}{\delta^2}$. Then if $\alpha(\mathcal{H}) \geq 4\delta$ then there exists $1 \leq i < j \leq \ell$ and an assignment to \mathcal{L} which satisfies at least a $\frac{\delta^{4+2qc}}{\text{poly log } \delta}$ fraction of $\Phi_{i \rightarrow j}$.*

Proof. Let \mathcal{I} be an independent set of size $|\mathcal{I}| \geq 4\delta|\mathcal{V}|$. For $x \in U$, let us denote $\mathcal{I}_x := \mathcal{I} \cap \mathcal{V}_x$. An averaging argument gives us that there is a $U' \subseteq U$ such that:

1. $|U'| \geq 2\delta|U|$.
2. $|\mathcal{I}_x| \geq 2\delta|\mathcal{V}_x|$ for $x \in U'$.

A similar averaging argument gives us that there is a set $\mathcal{W} = \{U_{i_1}, \dots, U_{i_{|\mathcal{W}|}}\}$ of layers with the following properties:

1. $|\mathcal{W}| = \delta\ell$.
2. For every layer $W \in \mathcal{W}$, $|W \cap U'| \geq \delta|W|$.

For $j \in [|\mathcal{W}|]$, let us denote $Z_j \stackrel{\text{def}}{=} U_{i_j} \cap U'$. From the two properties of \mathcal{W} we have $|Z_j| \geq \delta|U_{i_j}| = \frac{\delta^2 \ell}{|\mathcal{W}|} |U_{i_j}|$. By assumption $\ell = 2/\delta^2$ and thus $|Z_j| \geq \frac{2}{|\mathcal{W}|} |U_{i_j}|$, so from the Weakly Dense property of \mathcal{L} it follows that there are $Z_k \subseteq U_{i_k}$ and $Z_{k'} \subseteq U_{i_{k'}}$ such that

$$|\Phi(Z_k, Z_{k'})| \geq \frac{1}{|\mathcal{W}|^2} \cdot |\Phi(U_{i_k}, U_{i_{k'}})| = \frac{\delta^2}{4} |\Phi(U_{i_k}, U_{i_{k'}})|. \quad (3)$$

By yet another averaging argument, at least a $\frac{\delta^2}{8}$ fraction of $x \in Z_k$ has at least a $\frac{\delta^2}{8}$ fraction of their constraints in $Z_{k'}$. Let us denote those x by $X \subseteq Z_k$ and let $Y \stackrel{\text{def}}{=} Z_{k'}$. We will henceforth restrict our attention to the constraints between X and Y . We have that for every $x \in X$, \mathcal{I}_x there are k vertices, $(x, a_{x,1}), \dots, (x, a_{x,k})$ such that $|\text{ag}(a_{x,1}, \dots, a_{x,k})| \leq c \log 1/\delta \stackrel{\text{def}}{=} t$. Let us denote $L_x = \text{ag}(a_{x,1}, \dots, a_{x,k})$.

We now trim bad constraints from $\Phi(X, Y)$ as follows. For every $x \in X$, we have from the smoothness property that at most $\frac{t^2 \ell}{T} \leq \frac{\delta^2}{16}$ fraction of constraints $\phi_{x \rightarrow y}$ where $y \in U_{i_{k'}}$ are such that $|\phi_{x \rightarrow y}(L_x)| < |L_x|$; Call such constraints *bad*. We remove all such bad constraints from $\Phi(X, Y)$. This gives that at least $\frac{\delta^2}{16}$ fraction of the $y \in N_Y(x)$ are such that $|\phi_{x \rightarrow y}(L_x)| = |L_x|$, i.e., no two labels from L_x project on the same label and we retain all such constraints. We use $N'_X(y)$ to denote only the neighbors with no bad constraints. Thus we are left with $\frac{\delta^2}{4} \cdot \frac{\delta^2}{16} \cdot |\Phi(U_{i_k}, U_{i_{k'}})|$ constraints in $\Phi(X, Y)$. Let us call these Φ_{good} .

Claim 5.4. *For any $y \in Y$, consider any set $\{x_1, \dots, x_s\} \subseteq N'_X(y)$ such that for every $i, j \in [s]$, we have $\phi_{x_i \rightarrow y}(L_{x_i}) \cap \phi_{x_j \rightarrow y}(L_{x_j}) = \emptyset$. Then*

$$s \leq \frac{\log(1/\delta)}{\delta^{2qc}}$$

Proof. For every $i \in [s]$ and $(y, b) \in \mathcal{I}_y$, condition (2) implies that $b(\cdot)$ at locations $\cup_{r \in L_{x_i}} \phi_{x_i \rightarrow y}(r)$ can take at most $q^{|L_{x_i}|} - (q-1)^{|L_{x_i}|}$ different values. This is because if all values at $b(\phi_{x_i \rightarrow y}(r))$ are from $[q] \setminus a_{x_i,1}(r)$ (which by definition equals $[q] \setminus a_{x_i,j}(r)$ for all $j \in [k]$), then $(x_i, a_{x_i,1}), \dots, (x_i, a_{x_i,k}), (y, b)$ forms a hyperedge in \mathcal{H} , which is a contradiction as they all belong to \mathcal{I} . Since $\phi_{x_i \rightarrow y}(L_{x_i}) \cap \phi_{x_j \rightarrow y}(L_{x_j}) = \emptyset$, these restrictions for each $i \in [s]$ are disjoint from each other. Given this, we can upper bound the size of \mathcal{I}_y by

$$|\mathcal{I}_y| \leq \left(1 - \left(\frac{q-1}{q}\right)^t\right)^s |\mathcal{V}_y| \leq \exp(-s \exp(-2qt)) |\mathcal{V}_y| = \exp(-s \delta^{2qc}) |\mathcal{V}_y|.$$

On the other hand, we have that $|\mathcal{I}_y| \geq \delta |\mathcal{V}_y|$. Combining these two facts gives us our desired bound. \square

The rest of the proof proceeds exactly as the proof of Lemma 4.3. We now define a (randomized) labelling $A : X \sqcup Y \rightarrow [R_{i_k}] \cup [R_{i_{k'}}]$. For $x \in X$, pick a label $A(x)$ from L_x at random.

For $y \in Y$, define the (multi-)family $\mathcal{F}(y) \stackrel{\text{def}}{=} \{\phi_{x \rightarrow y}(L_x) \mid x \in N'_X(y)\}$ of subsets of $[R_{k'}]$. By Claim 5.4 and Lemma 3.8, it follows that there is a label $\ell_y \in [R_{i_{k'}}]$ that is present in at least a $\frac{1}{t(s-1)} |\mathcal{F}(y)|$ sets in $\mathcal{F}(y)$. Define $A(y)$ to be that label.

By the choice of $A(y)$, for every $y \in N'(X)$, a $\frac{1}{t(s-1)}$ fraction of all $x \in N_X(y)$ have a label $\ell_x \in L_x$ such that $\phi_{x \rightarrow y}(i) = A(y)$, and each such constraint is satisfied by A with probability at least $1/|L_x| \geq 1/t$. It follows that the expected number of constraints satisfied by the labelling A is at least $\frac{1}{t^2(s-1)} |\Phi_{\text{good}}|$.

As noted earlier, $|\Phi_{\text{good}}| \geq \frac{\delta^4}{64} |\Phi(U_{i_k}, U_{i_{k'}})|$ and thus A satisfies at least a $\frac{\delta^{4+2qc}}{\text{poly log } \delta}$ fraction of all constraints between layers U_{i_k} and $U_{i_{k'}}$. \square

Proof of Theorem 5.1. Start with a 3-SAT instance Π on n variables. Set $\ell = c_1 \frac{\log \log n}{\log \log \log n}$, $T = (\log \log n)^3$ and $r = c_2 \log \log \log n$. Theorem 3.4 gives us an ℓ -layered Label Cover instance $\mathcal{L} = (\mathcal{U}, \mathcal{R}, E, \Phi)$ on $n^{O((T+\ell)r)} = n^{\text{poly log log } n}$ vertices over alphabets of size $2^{O(\ell r)} = 2^{O(c_1 c_2 \log \log n)}$ with soundness $2^{-\Omega(r)}$, and we choose the constants c_1 and c_2 such that the alphabet size is $\log n$ and the soundness is a sufficiently large power of $\log \log n$ to apply Lemma 5.3 below.

The above reduction gives us a $(k+1)$ -uniform hypergraph \mathcal{H} on $|\mathcal{V}| = 3^{\log n} \cdot n^{\text{poly log log } n} = n^{\text{poly log log } n}$ vertices and in particular $\log \log |\mathcal{V}| = O(\log \log n)$. In the completeness case, we have $\chi(\mathcal{H}) \leq q$. Setting $\delta = \sqrt{2/\ell} = (\log \log n)^{-1/2+o(1)}$ in the soundness case, we get that from Lemma 5.3 that $\alpha(\mathcal{H}) \leq (\log \log n)^{-1/2+o(1)} = (\log \log |\mathcal{V}|)^{-1/2+o(1)}$. \square

6 Conclusion

We hope that although we do not improve upon previous results, our proof technique will be useful in improving hypergraph coloring hardness for lower uniformity. The reduction has an outer verifier (Label Cover instance) and an inner verifier (gadget) as two components. It might be the case that our inner verifier is stronger than the previous inner verifiers and hence we do not require extra structural property on the Label Cover instance (for Theorem 1.2, and 1.4). It will be interesting to understand this trade-off.

It is interesting that all of our results follow from a general framework and uses t-agreeing families. This can be thought as a unified proofs for results which otherwise had somewhat different proofs. We hope that this sheds light on the hardness of hypergraph coloring results regardless of its uniformity and the completeness guarantee.

We now give a concrete open problem which will improve the proven hardness results in this paper. Consider the following property of a family $\mathcal{F} \subset [q]^n$, parameterized by t and $T = T(t)$: For any subset $S \subseteq \mathcal{F}$ such that $|S| \geq \frac{1}{T}|\mathcal{F}|$ there exists $a, b \in S$ such that $|\text{ag}(a, b)| \leq t$. We say that such a family \mathcal{F} has property $\mathfrak{P}(t, T)$. Theorem 3.5 shows that $[3]^n$ satisfies $\mathfrak{P}(t, 2^{O(t)})$ for any t . We pose the following problem, which, to the best of our knowledge has not been investigated yet:

Open Problem: For $q \geq 3$, is there a family $\mathcal{F} \subset [q]^n$ such that $|\mathcal{F}| \leq q^{\text{poly}(\log n)}$ that has property $\mathfrak{P}(t, t^{\omega(1)})$?

References

- [AK98] Rudolf Ahlswede and Levon H Khachatrian. The diametric theorem in hamming spaces—optimal anticodes. *Advances in Applied Mathematics*, 20(4):429–449, 1998.
- [AKMH96] Noga Alon, Pierre Kelsen, Sanjeev Mahajan, and Ramesh Hariharan. Approximate hypergraph coloring. *Nord. J. Comput.*, 3(4):425–439, 1996.
- [ALM⁺98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof Verification and the Hardness of Approximation Problems. *Journal of the ACM*, 45(3):501–555, 1998. (Preliminary version in *33rd FOCS*, 1992).
- [AS98] Sanjeev Arora and Shmuel Safra. Probabilistic Checking of Proofs: A new characterization of NP. *JACM*, 45(1):70–122, January 1998. (Preliminary version in *33rd FOCS*, 1992).
- [Bha18] Amey Bhangale. NP-Hardness of Coloring 2-Colorable Hypergraph with Poly-Logarithmically Many Colors. In *45th International Colloquium on Automata, Languages, and Programming, ICALP 2018, July 9-13, 2018, Prague, Czech Republic*, pages 15:1–15:11, 2018.
- [CF96] Hui Chen and Alan Frieze. Coloring bipartite hypergraphs. In *International Conference on Integer Programming and Combinatorial Optimization*, pages 345–358. Springer, 1996.
- [DG15] Irit Dinur and Venkatesan Guruswami. Pcps via the low-degree long code and hardness for constrained hypergraph coloring. *Israel Journal of Mathematics*, 209(2):611–649, 2015.
- [DGKR05] Irit Dinur, Venkatesan Guruswami, Subhash Khot, and Oded Regev. A new multilayered pcp and the hardness of hypergraph vertex cover. *SIAM Journal on Computing*, 34(5):1129–1146, 2005.
- [DRS02] I Dinur, O Regev, and C Smyth. The hardness of 3-uniform hypergraph coloring. In *Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on*, pages 33–40. IEEE, 2002.

- [FGL⁺96] Uriel Feige, Shafi Goldwasser, Laszlo Lovász, Shmuel Safra, and Mario Szegedy. Interactive proofs and the hardness of approximating cliques. *Journal of the ACM (JACM)*, 43(2):268–292, 1996.
- [FK98] Uriel Feige and Joe Kilian. Zero knowledge and the chromatic number. *Journal of Computer and System Sciences*, 57(2):187–199, 1998.
- [Fra76] Péter Frankl. Families of finite sets satisfying an intersection condition. *Bulletin of the Australian Mathematical Society*, 15(1):73–79, 1976.
- [Fra18] Peter Frankl. Some exact results for multiply intersecting families. *Journal of Combinatorial Theory, Series B*, 2018.
- [FT99] Peter Frankl and Norihide Tokushige. The erdős–ko–rado theorem for integer sequences. *Combinatorica*, 19(1):55–63, 1999.
- [GHH⁺14] Venkatesan Guruswami, Prahladh Harsha, Johan Håstad, Srikanth Srinivasan, and Girish Varma. Super-polylogarithmic Hypergraph Coloring Hardness via Low-degree Long Codes. In *Proceedings of the Forty-sixth Annual ACM Symposium on Theory of Computing, STOC '14*, pages 614–623, New York, NY, USA, 2014. ACM.
- [GHS02] Venkatesan Guruswami, Johan Håstad, and Madhu Sudan. Hardness of approximate hypergraph coloring. *SIAM Journal on Computing*, 31(6):1663–1686, 2002.
- [Hal93] Magnús M Halldórsson. A still better performance guarantee for approximate graph coloring. *Information Processing Letters*, 45(1):19–23, 1993.
- [Hol02] Jonas Holmerin. Vertex cover on 4-regular hyper-graphs is hard to approximate within $2-\epsilon$. In *Proceedings of the thirty-fourth annual ACM symposium on Theory of computing*, pages 544–552. ACM, 2002.
- [Hua15] Sangxia Huang. $2^{(\log N)^{1/10-o(1)}}$ hardness for hypergraph coloring. *CoRR*, abs/1504.03923, 2015.
- [Kho02a] Subhash Khot. Hardness results for approximate hypergraph coloring. In *Proceedings of the thirty-fourth annual ACM symposium on Theory of computing*, pages 351–359. ACM, 2002.
- [Kho02b] Subhash Khot. Hardness results for coloring 3-colorable 3-uniform hypergraphs. In *Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on*, pages 23–32. IEEE, 2002.
- [KNS01] Michael Krivelevich, Ram Nathaniel, and Benny Sudakov. Approximating coloring and maximum independent sets in 3-uniform hypergraphs. *Journal of Algorithms*, 41(1):99–113, 2001.
- [KS14] Subhash Khot and Rishi Saket. Hardness of Coloring 2-Colorable 12-Uniform Hypergraphs with $\exp(\log \hat{\Omega}(1))n$ Colors. In *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*, pages 206–215. IEEE, 2014.

- [KT17] Ken-ichi Kawarabayashi and Mikkel Thorup. Coloring 3-colorable graphs with less than $n^{1/5}$ colors. *Journal of the ACM (JACM)*, 64(1):4, 2017.
- [Raz98] Ran Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, 1998.
- [Sak14] Rishi Saket. Hardness of finding independent sets in 2-colorable hypergraphs and of satisfiable CSPs. In *IEEE 29th Conference on Computational Complexity (CCC)*, pages 78–89, 2014.
- [Var15] Girish Varma. Reducing uniformity in Khot-Saket hypergraph coloring hardness reductions. *Chicago Journal OF Theoretical Computer Science*, 3:1–8, 2015.
- [Zuc06] David Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 681–690. ACM, 2006.