Electronic Colloquium on Computational Complexity, Report No. 111 (2010)

The complexity of finding independent sets in bounded degree (hyper)graphs of low chromatic number

Venkatesan Guruswami* Ali Kemal Sinop*

Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213.

Abstract

We prove almost tight hardness results for finding independent sets in bounded degree graphs and hypergraphs that admit a good coloring. Our specific results include the following (where Δ, assumed to be a constant, is a bound on the degree, and n is the number of vertices):

- *NP-hardness* of finding an independent set of size larger than $O\left(n\left(\frac{\log\Delta}{\Delta}\right)^{\frac{1}{r-1}}\right)$ in a 2-colorable r-uniform hypergraph for $r \geqslant 4$. A simple algorithm is known to find independent sets of size $\Omega\left(\frac{n}{\Delta^{1/(r-1)}}\right)$ in *any* r-uniform hypergraph of maximum degree Δ. Under a combinatorial conjecture on hypergraphs, the $(\log\Delta)^{1/(r-1)}$ factor in our result is necessary.
- Conditional hardness of finding an independent set with more than $O\left(\frac{n}{\Delta^{1-c/(k-1)}}\right)$ vertices in a k-colorable graph for some absolute constant $c \leqslant 4$, under Khot's 2-to-1 Conjecture. This suggests the near-optimality of Karger, Motwani and Sudan's graph coloring algorithm which finds an independent set of size $\Omega\left(\frac{n}{\Delta^{1-2/k}\sqrt{\log\Delta}}\right)$ in k-colorable graphs.
- Conditional hardness of finding independent sets of size $n\Delta^{-1/8+o_\Delta(1)}$ in almost 2-colorable 3-uniform hypergraphs, under the Unique Games Conjecture. This suggests the optimality of the known algorithms to find an independent set of size $\tilde{\Omega}(n\Delta^{-1/8})$ in 2-colorable 3-uniform hypergraphs.
- Conditional hardness of finding an independent set of size more than $O\left(n\left(\frac{\log\Delta}{\Delta}\right)^{\frac{1}{r-1}}\right)$ in r-uniform hypergraphs that contain an independent set of size $n\Big(1-O\big(\frac{\log r}{r}\big)\Big)$ assuming the Unique Games Conjecture.

*Supported in part by a Packard Fellowship and US-Israel BSF grant 2008293. Email: `guruswami@cmu.edu`, `asinop@cs.cmu.edu`

ISSN 1433-8092

1 Introduction

The Independent Set problem is one of the most well-known NP-complete problems. On general graphs (and hypergraphs), even approximating the solution to a factor of $n^{1-\varepsilon}$ (as a function of the number n of vertices) is known to be NP-hard. For graphs with maximum degree at most Δ, a simple greedy algorithm finds an independent set of size $\Omega(n/\Delta)$. For graphs which are promised to have small chromatic number, a better guarantee is known: given a k-colorable graph, an algorithm due to Karger *et al.* [12] uses semidefinite programming (SDP) to find an independent set of size about $\Omega(n/\Delta^{1-2/k})$. Similar SDP based algorithms for finding independent sets of non-trivial size are also known for 2-colorable 3-uniform hypergraphs [1, 4].

In this paper, we investigate the inapproximability of the independent set problem on bounded degree graphs and hypergraphs which are promised to be colorable with few colors (and therefore in particular have a large independent set). We obtain several strong hardness results even for such hypergraphs, and our inapproximability factors as a function of the degree bound Δ almost match known algorithms.

Before stating our results and comparing them to the known algorithmic bounds, we recall the basic terminology. An *r-uniform hypergraph* $G = (V, E)$ consists of a collection of vertices $v \in V$ and a collection of (hyper)edges $e \in E$ where e is a size r subset of V ($e \in \binom{V}{r}$). The *degree* of a vertex $v \in V$, Δ_v is the number of times it appears in edges. A hypergraph G is *degree-Δ bounded* if all vertices have degree at most Δ. Notice that for $r = 2$, this definition corresponds to a graph. An *independent set* of G is a subset I of vertices such that no edge is completely contained in I. Similarly, a subset of vertices S is a *vertex cover* of G if S has non-empty intersection with all edges. A hypergraph G is called *k-colorable* if its vertices can be partitioned into at most k many disjoint independent sets. A related weaker notion is *(k, ε)-semicolorability*, where there is a k coloring so that at most ε-fraction of edges are monochromatic.

1.1 Our results

We obtain several hardness results for finding independent sets of the following general form: given a degree-Δ bounded (hyper)graph G with some strong structural property that guarantees the existence of a large independent set (such as k-colorability for some small k), it is nevertheless hard to find an independent set of size $\frac{n}{f(\Delta)}$ for an appropriate function $f(\cdot)$ of the degree Δ. Furthermore the bounds $f(\Delta)$ in our hardness results almost match the bounds achieved by known efficient algorithms. Note that a result of this form also implies a factor $\Omega(f(\Delta))$ inapproximability result for finding independent sets in degree-Δ bounded (hyper)graphs. The formal result statements follow. In the following, Δ (assumed to be a constant) denotes the bound on the degree of the hypergraphs and n the number of vertices.

For r-uniform hypergraphs, a simple randomized algorithm finds an independent set of size $\Omega(\frac{n}{\Delta^{1/(r-1)}})$ [1]. We prove the NP-hardness of finding independent sets larger than $O\left(n\left(\frac{\log \Delta}{\Delta}\right)^{1/(r-1)}\right)$ even for 2-*colorable* r-uniform hypergraphs when $r \geqslant 4$. To the best of our knowledge, such a $\Omega\left(\left(\frac{\Delta}{\log \Delta}\right)^{1/(r-1)}\right)$ hardness result for independent sets on degree-Δ bounded r-uniform hypergraphs was not known. Also, the factor $(\log \Delta)^{1/(r-1)}$ slack in our hardness result (compared to the algorithmic bound of $\frac{n}{\Delta^{1/(r-1)}}$) is inherent, assuming P $\neq$ NP and a conjecture of Frieze and Mubayi [7] on the existence of independent sets in hypergraphs with forbidden subgraphs. Our work highlights a natural open question on finding $o(\Delta^{1/(r-1)})$-factor approximation for independent set on such hypergraphs (see Remark 1).

Turning to graphs, the famous algorithm of Karger, Motwani and Sudan (KMS) finds an independent set of size $\Omega\left(\frac{n}{\Delta^{1-2/k}\sqrt{\log \Delta}}\right)$ in a k-colorable degree-Δ bounded graph. Assuming Khot's 2-to-1 Conjecture [13], and using a construction (albeit with a different analysis) from the authors' hardness result for Maximum k-Colorable Subgraph [9], we show that the performance of the KMS algorithm is nearly best possible. Formally, we prove that for some absolute constant $c \leqslant 4$, it is hard to find an independent set larger than $O\left(\frac{n}{\Delta^{1-c/(k-1)}}\right)$ in a k-colorable graph for $k \geqslant 7$.

For 3-uniform hypergraphs, our results for 2-colorable hypergraphs mentioned above do not apply. This is for a good reason, since there are algorithms to find an independent set of size about $n/\Delta^{1/8}$ in the 2-colorable case. For 2-colorable 3-uniform hypergraphs, we obtain a conditional hardness results based on the Unique Games Conjecture (UGC) [13]. Due to the nature of this conjecture, our result only shows hardness for $(2, \varepsilon)$ semicolorable hypergraphs. Using a SDP based algorithm, it was shown by several authors [1, 4, 15] how to find an independent set of size $\tilde{\Omega}(n\Delta^{-1/8})$ in 2-colorable 3-uniform hypergraphs. By analyzing an appropriate dictatorship test using tools from Gaussian Noise Stability [16], we show that this performance is essentially best possible (up to $\log^{O(1)} \Delta$ factors).

Since a 2-colorable hypergraph in particular has an independent set of size $n/2$, our result shows the hardness of finding non-trivial independent sets in such hypergraphs. Once a hypergraph has an independent set of size at least $(1 - 1/r + \gamma)n$ for some constant $\gamma > 0$, the factor-r approximation algorithm for vertex cover on r uniform hypergraphs gives an efficient algorithm to find an independent set of size $r\gamma n = \Omega(n)$. In the quest to pin down the largest α for which strong hardness results apply for finding independent sets in r-uniform hypergraphs which are promised to have an independent set of αn, we show that even if there is an independent set of size $n\left(1 - \frac{\log r}{r}\right)$, it is hard (under UGC) to find independent sets larger than $O\Big(n\big(\frac{\log \Delta}{\Delta}\big)^{1/(r-1)}\Big)$.

1.2 Proof methods

Our main technique involves combining certain "robust" hardness results for independent sets on graphs or hypergraphs of small chromatic number (with no restrictions on the degree) with a sparsification procedure to reduce the degree of the graph. The "robustness" refers to the fact that in the soundness case, not only is there no independent set of size γn (for any desired $\gamma > 0$), but every set of size γn contains a constant fraction $\alpha(\gamma)$ of the total number of edges in the graph. This approach was used for the results shown in [2] to establish hardness of approximating independent set size on degree-Δ bounded graphs within a $\Omega\left(\frac{\Delta}{\log^2 \Delta}\right)$ factor assuming UGC. In the soundness analysis of these density arguments, we use a soft decoding type approach to argue about the existence of a good labeling, unlike the usual thresholding-based approach.

This paper extends the results shown in [2] in several directions. The results from [2] are conditioned on UGC and only apply to graphs with big independent sets. Moreover they do not shed light on the approximation guarantee possible for k-colorable graphs. In this work, we borrow their sparsification procedure and combine it with different robust hardness results for finding independent sets in graphs and hypergraphs with small chromatic number. Some of these robust hardness results for independent set (e.g. the one for 2-colorable 4-uniform hypergraphs) are borrowed directly from the literature, and the others (e.g. for k-colorable graphs and 2-semicolorable 3-uniform hypergraphs) are shown in this paper. Our final quantitative results show that it is not possible to beat the performance of the known coloring algorithms (up to some small factors) as a function of the degree. Some of our hypergraph results do not rely on any conjecture, and the result for k-colorable graphs relies on the 2-to-1 Conjecture [13]. Also conditioned on UGC, we show hardness results for hypergraphs which have extremely large independent sets.

2 Independent sets in 2-colorable hypergraphs

Our first result is an unconditional NP-hardness result for 2-colorable $r \geqslant 4$-uniform hypergraphs. The next result uses the 2-to-1 Conjecture for proving a similar result on k-colorable graphs. Both of these results are based on the idea of proving a certain density property and then using this to show that a simple sparsification procedure based on random sampling of (hyper)edges will not produce a big independent set. We first introduce the sparsification procedure. We then use it to first give the results on hypergraphs as it is easier and unconditional.

2.1 Sparsification Procedure

Both of our hardness results use a random sampling based sparsification procedure which relies on the underlying hypergraph having a certain density property, as outlined in [2].

For a given set S, we will denote $x \sim S$ as choosing an element of S uniformly at random. We now give the density definition:

Definition 2.1. *Let H be an r-uniform hypergraph $H = (V(H), E(H))$, with $r \geqslant 2$. Density of a set $U \subseteq V$ on H is defined as*

$$\mathsf{Density}_H[U] \triangleq \mathsf{Pr}_{e\sim E(H)}[e \subseteq U].$$

In other words, density is the probability that a randomly selected hyperedge will lie completely inside this set.

Moreover H is said to be (α,γ)-dense if for every $S \in \binom{V}{\alpha|V|}$, $\mathsf{Density}_H[S] \geqslant \gamma$.

We construct $H' = (V, E')$ by randomly sampling β fraction of edges with replacement, so that $|E'| = \beta|E|$. The new graph H' inherits any (semi)coloring properties of the original hypergraph H. First, we introduce the definition of semicolorability:

Definition 2.2. *An r-uniform hypergraph G is called (k,ε) semicolorable if there is a k-coloring of the vertices in G such that all but at most ε-fraction of edges are monochromatic.*

Observation 2.3. *If H is (k,ε)-semicolorable, H' is also (k,ε)-semicolorable.*

Lemma 2.4. *If H is $(\mu, \mu^{\lambda}/3)$-dense, then H' has no independent set of size μn with probability $\frac{1}{2}$ for the choice $\beta|E| = \Theta(n\mathrm{h}(\mu)\mu^{-\lambda})$, where $\mathrm{h}(\mu) = -\mu\log\mu - (1-\mu)\log(1-\mu)$.*

Proof. Consider a fixed subset U. There are at least $|E|\mu^\lambda$ many hyperedges completely inside U. The probability that U is an independent set in H' is given by $(1-\mu^\lambda/3)^{\beta|E|} \approx \exp\left(-\mu^\lambda\beta|E|/3\right)$. Taking union probability over all such subsets, the probability of having a size μn-independent set is:

$$\binom{n}{\mu n}\exp(-\beta\mu^\lambda|E|/3) = \exp\left(O\left(n\mathrm{h}(\mu) - \mu^\lambda\beta|E|/3\right)\right).$$

For suitable $\beta|E| = \Theta(n\mathrm{h}(\mu)\mu^{-\lambda})$, this probability is $\leqslant \frac{1}{2}$. $\square$

The average degree of a vertex in H' is $\Delta_{\mathrm{avg}} = O\left(\beta|E|/n\right) = O(\mathrm{h}(\mu)\,\mu^{-\lambda})$. By a simple Chernoff bound, it is easy to see that $\mathsf{Pr}_{E'}\left[\Delta_u \geqslant 2\Delta_{\mathrm{avg}}\right] \leqslant \exp(-\Delta_{\mathrm{avg}}) = o(\mu)$ where Δ_u is the degree of vertex u in H'. We can remove all such vertices, as there is no independent set of size $\mu n - o(\mu n)$ in H'. Hence we can make sure that in the final hypergraph, all vertices have degree $\Delta_u \leqslant 2H(\mu)\mu^{-\lambda}$.

As a summary, we have the following lemma:

Lemma 2.5. *Given positive reals, μ, λ, for any r-uniform hypergraph $H = (V, E)$, in polynomial time, we can construct an r-uniform hypergraph $H' = (V, E')$ with maximum degree $\Delta = O(\mathrm{h}(\mu)\;\mu^{-\lambda})$ such that:*

- *Completeness: If H is k-colorable, then H' is k-colorable also.*
- *Soundness: If H is $(\mu, \frac{\mu^\lambda}{3})$-dense, then H' has no independent set of size $\mu n/2$.*

In order to avoid repeating calculations, we summarize how to relate the degree bound with independent set inapproximability in the following corollary, whose proof appears in Appendix A.

Corollary 2.6. *Given an r-uniform hypergraph $H = (V, E)$, any positive constant $\varepsilon \geqslant 0$, $\lambda > 1$ and Δ some large integer, there exists a constant $\mu = \mu(\Delta, \lambda)$ and an r-uniform hypergraph $H' = (V, E')$ with maximum degree Δ such that:*

- *Completeness: If H is (k, ε)-semicolorable, then H' is (k, ε)-semicolorable also.*
- *Soundness: If H is $(\mu, \frac{\mu^\lambda}{3})$-dense, H' has no independent set of size $O\left(n\left(\frac{\log\Delta}{\Delta}\right)^{1/(\lambda-1)}\right)$.*

2.2 Results for 2-Colorable r-Uniform Hypergraphs

For this result, we will use the PCP verifier given by Holmerin [11] for Vertex-Cover hardness on 4-uniform hypergraphs, which we will then extend to higher uniformities by a simple direct product. Notice that this is the same PCP verifier originally used by Guruswami *et al.* [8] for proving hardness of approximate 4-uniform hypergraph coloring. The following is a simple corollary of the results from these papers [8, 11]:

Theorem 2.7. *For any $\mu > 0$ and small enough $0 < \delta = o(\mu^4)$, the following holds: Given a 4-uniform hypergraph $H = (V, E)$ on n vertices, it is NP-hard to distinguish if*

- *H is 2-colorable or,*

- *H is $(\mu, \mu^4 - \delta)$-dense.*

In other words, it is NP-hard to find a subset with size μn of vertices such that at most $(\mu^4 - \delta)$-fraction of the edges lie completely inside that set even *when the hypergraph is 2-colorable. In particular, it is hard to find independent set of size μn.*

Proof. Consider Test FHS-δ given in [11] (Proposition 3.14), where there is a vertex corresponding the entry in each table, $A_u(f)$. For every query of the form $(A_{w_1}(g_1), A_{w_1}(h_1), A_{w_2}(g_2), A_{w_2}(g_2))$, there is a hyperedge on the corresponding set of vertices. Notice that in [8], it is proven that for this hypergraph will be 2-colorable if the underlying instance ϕ is satisfiable. Also it can easily be shown that the hypergraph produced will have all vertex degrees equal.

For the second part, assume that ϕ is not satisfiable and there is a set U with $|U| = \mu n$ such that $\mathsf{Pr}_{e\sim E}[e \subseteq U] < \mu^4 - \delta$. Let $f : V \to \{0,1\}$ be the indicator function for this set. Arithmetizing the density:

$$\mathsf{Pr}_{e\sim E}[e \subseteq U] = \mathbb{E}_{e\sim E}\left[\prod_{v\in e} f(v)\right]$$

which is the rejection probability of Test FHS-δ on proof given by $g = 2f - 1$. Since g is $(2\mu - 1)$-balanced, by Proposition 3.15 in [11], ϕ is satisfiable since $\mathsf{Pr}_{e\sim E}[e \subseteq U] \leqslant \mu^4 - \delta$, which is a contradiction. □

A simple extension of the above argument gives a similar result for hypergraphs with uniformities $\geqslant 4$.

Lemma 2.8. *For all $\delta_r > 0$ and $r \geqslant 4$, the following holds: Given an r-uniform regular hypergraph $H = (V, E)$ on n vertices, it is NP-hard to distinguish if*

- *H is 2-colorable or,*
- *H is $(\mu, \mu^r - \delta_r)$-dense .*

Proof. We will prove by induction on r. For $r = 4$, the result is given above. Assume for $r \geqslant 4$, the decision problem is hard for δ_r. Let $\delta_{r+1} = \delta_r \mu$. Given any r-uniform regular hypergraph $H^{(r)} = (V, E^{(r)})$, construct a new $(r+1)$-uniform hypergraph $H^{(r+1)} = (V, E^{(r+1)})$ on the same vertex set with the hyperedges defined as follows: For every $v \in V$, and every $e \in E^{(r)}$ such that $v \notin e$, add edge $e_v = e \cup \{v\}$ to $E^{(r+1)}$. Since all vertices had same degree in $H^{(r)}$, they have the same degree in $H^{(r+1)}$ also. Moreover if $H^{(r)}$ is 2-colorable, $H^{(r+1)}$ is 2-colorable also, as its edges strictly contain at least an edge from $E^{(r)}$.

Assume there is a set $U \in \binom{[n]}{\mu n}$ with $\mathsf{Density}_U\left[H^{(r+1)}\right] \leqslant \mu^{r+1} - \delta_{r+1}$. This implies

$$\begin{aligned}
\mu^{r+1} - \delta_{r+1} &\geqslant \mathsf{Pr}_{e\sim E(H^{(r+1)})}[e \subseteq U] = \mathsf{Pr}_{e'\sim E(H^{(r)}), v\sim V\setminus e'}[e \subseteq U \wedge v \in U] \\
&\geqslant \mathsf{Pr}_{e'\sim E(H^{(r)})}[e \subseteq U]\, \mathsf{Pr}_{v\sim V\setminus e'}[v \in U] \\
&> \mathsf{Density}_U\left[H^{(r)}\right]\mu \\
\mu^r - \delta_r &> \mathsf{Density}_U\left[H^{(r)}\right].
\end{aligned}$$

The first inequality follows because all vertices have same degree. This means the decision problem on $H^{(r+1)}$ implies the decision problem on $H^{(r)}$, which is hard by our induction hypothesis. $\square$

On this hypergraph, we can directly apply Corollary 2.6 and obtain the following:

Theorem 2.9. *For any integer $r \geqslant 4$ and integer constant Δ, given an r-uniform hypergraph, with maximum degree Δ, it is NP-hard to decide whether if this hypergraph is 2-colorable or largest independent set size is $O\Big(n\big(\frac{\log \Delta}{\Delta}\big)^{1/(r-1)}\Big)$.*

Proof. Given an r-uniform hypergraph instance H from Lemma 2.8 we can construct another hypergraph H' using Corollary 2.6 with $\lambda = r$. Completeness follows from how 2-colorability feature is preserved in all transformations. Soundness follows because if there is no independent set of size $O\left(\frac{\log^{1/(r-1)} \Delta}{\Delta^{1/(r-1)}} n\right)$, H is not $(\mu, \mu^r - \delta)$-dense for small enough δ. $\square$

Remark 1. The corollary is tightly related to a conjecture of Frieze and Mubayi [7], which (restated for suitability to our problem) says that r-uniform K_{2r-1}-free hypergraphs (K_{2r-1} is the hypergraph clique on $2r-1$ vertices) with maximum degree Δ have an independent set of size at least $c_r n \left(\frac{\log \Delta}{\Delta}\right)^{\frac{1}{r-1}}$. Since any 2-colorable hypergraph is K_{2r-1}-free, their conjecture implies that our hardness result is tight (assuming P $\neq$ NP).

3 Independent sets in k-colorable graphs

For this result, we will use the verifier given by Guruswami and Sinop [9] for the Maximum k-Colorable Subgraph problem. The difference is that we analyze density properties instead of estimating the fraction of miscolored edges. First we will give some definitions and introduce the d-to-1 Conjecture which is needed for this hardness result.

3.1 Preliminaries

We begin by reviewing some definitions and the d-to-1 Conjecture.

Definition 3.1. *An instance of a bipartite Label Cover problem represented as $\mathcal{L} = (U, V, E, W, R_U, R_V, \Pi)$ consists of a weighted bipartite graph over node sets U and V with edges $e = (u, v) \in E$ of non-negative real weight $w_e \in W$ and $\sum_e w_e = 1$. R_U and R_V are integers with $1 \leqslant R_U \leqslant R_V$. Π is a collection of projection functions for each edge: $\Pi = \{\pi_{vu} : \{1, \ldots, R_V\} \to \{1, \ldots, R_U\} \big| u \in U, v \in V\}$. A labeling ℓ is a mapping $\ell : U \to \{1, \ldots, R_U\}$, $\ell : V \to \{1, \ldots, R_V\}$. An edge $e = (u, v)$ is satisfied by labeling ℓ if $\pi_e(\ell(v)) = \ell(u)$. We define the value of a labeling as sum of weights of edges satisfied by this labeling normalized by the total weight. $\mathsf{Opt}(\mathcal{L})$ is the maximum value over any labeling.*

Definition 3.2. *A projection $\pi : \{1, \ldots, R_V\} \to \{1, \ldots, R_U\}$ is called d-to-1 if for each $i \in \{1, \ldots, R_U\}$, $|\pi^{-1}(i)| \leqslant d$. It is called* exactly *d-to-1 if $|\pi^{-1}(i)| = d$ for each $i \in \{1, 2, \ldots, R_U\}$.*

Definition 3.3. *A bipartite Label-Cover instance $\mathcal{L}$ is called d-to-1 Label-Cover if all projection functions, $\pi \in \Pi$ are d-to-1.*

Conjecture 3.4 (d-to-1 Conjecture [13]). *For any $\gamma > 0$, there exists a right regular d-to-1 Label-Cover instance $\mathcal{L}$ with $R_V = R(\gamma)$ and $R_V/d \leqslant R_U \leqslant R_V$ many labels such that it is NP-hard to decide between two cases, $\mathsf{Opt}(\mathcal{L}) = 1$ or $\mathsf{Opt}(\mathcal{L}) \leqslant \gamma$. Note that although the original conjecture involves d-to-1 projection functions, we will assume that it also holds for* exactly *d-to-1 functions (so $dR_V = R_U$) as in [5].*

The hardness results for these problems generally rely heavily on Gaussian stability bounds. We will include here the form we use in this paper. We use the same notation as Mossel [16]. Let $\phi(x) \triangleq (2\pi)^{-1/2}\exp(-x^2/2)$ and $N(x) \triangleq \int_{-\infty}^{x}\phi(x)dx$.

Definition 3.5. *For $\mu, \nu \in [0,1]$ and $\rho \in [-1,1]$, Gaussian stability bound is defined as:*

$$\underline{\Gamma}_\rho(\mu,\nu) \triangleq \mathsf{Pr}_{x,y}\left[x < N^{-1}(\mu) \ \wedge \ \rho x + \sqrt{1-\rho^2}y > N^{-1}(1-\nu)\right].$$

In this expression, x and y are independent normal variables with mean 0, variance 1 and $N^{-1}(x)$ is the inverse cumulative normal distribution.

Given $(\rho_1,\ldots,\rho_{k-1}) \in [0,1]^{k-1}$, and $(\mu_1,\ldots,\mu_k) \in [0,1]^k$ for $k \geqslant 3$, we define, inductively,

$$\underline{\Gamma}_{\rho_1,\ldots,\rho_{k-1}}(\mu_1,\ldots,\mu_k) \triangleq \underline{\Gamma}_{\rho_1}(\mu_1, \underline{\Gamma}_{\rho_2,\ldots,\rho_{k-1}}(\mu_2,\ldots,\mu_k)).$$

In Dinur *et al.* [5], the noise stability of two functions with same domains are bounded by the spectral radius of underlying Markov operator mapping this domain to itself. In [16], the spectral radius is generalized to correlation constants so as to allow for functions with different domains:

Definition 3.6. *Given two finite probability spaces Ω_1 and Ω_2 with probability distribution $\mathbf{P}$ on $\Omega_1 \times \Omega_2$, the correlation between these two spaces is defined as:*

$$\rho(\Omega_1,\Omega_2;\mathbf{P}) \triangleq \sup\{\mathrm{Cov}[f,g] : f \in L^2(\Omega_1), g \in L^2(\Omega_2), \mathbf{Var}[f] = \mathbf{Var}[g] = 1\}.$$

Since this is a supremum over functions of bounded variance, we can further assume that $\mathbb{E}[f] = \mathbb{E}[g] = 0$.

We will restate the noise stability result of Mossel [16] in the form we use:

Proposition 3.7. *[16] Let $(\prod_{j=1}^{k}\Omega_i^{(j)}, \mathbf{P}_i)$, $1 \leqslant i \leqslant r$ be a sequence of finite probability spaces such that for all $1 \leqslant i \leqslant r$, the minimum probability of any atom in $\prod_{j=1}^{k}\Omega_i^{(j)}$ is at least α. Assume furthermore that there exists $\underline{\rho} \in [0,1]^{k-1}$ and $0 \leqslant \rho < 1$ such that $\rho(\Omega_i^{(1)},\ldots,\Omega_i^{(k)};\mathbf{P}_i) \leqslant \rho$, and*

$$\rho(\Omega_i^{(\{1,\ldots,j\})}, \Omega_i^{(\{j+1,\ldots,k\})}; \mathbf{P}_i) \leqslant \underline{\rho}(j) \tag{1}$$

for all i,j. Then for every $\varepsilon > 0$, there exists $\tau > 0$ and $\kappa = \kappa(\tau,\alpha) > 0$ such that if $f_j : \prod_{i=1}^{n}\Omega_i^{(j)} \to [0,1]$ for $1 \leqslant j \leqslant k$ satisfy $\max_{i,j}(\mathrm{Inf}_i^{\leqslant\kappa}(f_j)) \leqslant \tau$ then it holds that

$$\underline{\Gamma}_{\underline{\rho}}(\mathbb{E}[f_1],\ldots,\mathbb{E}[f_k]) - \varepsilon \leqslant \mathbb{E}\Big[\prod_{j=1}^{k} f_j\Big]. \tag{2}$$

If instead of Condition 1, all $\mathbf{P}_i$ are pairwise independent, then $\prod_{j=1}^{k}\mathbb{E}[f_j] - \varepsilon \leqslant \mathbb{E}\big[\prod_{j=1}^{k} f_j\big]$.

We will use the following estimate. Its proof appears in Appendix B:

Lemma 3.8. *For $\rho > 0$, $\lambda \geqslant 1$ and small enough $\mu > 0$, $\underline{\Gamma}_\rho(\mu,\mu^\lambda) = \Omega\left(\frac{\mu^{\frac{1+\lambda+2\sqrt{\lambda}\rho}{1-\rho^2}}}{\sqrt{\ln 1/\mu}}\right)$.*

3.2 Reduction from 2-to-1 Label Cover

The following lemma is based on a PCP verifier from Guruswami and Sinop [9]. Usually soundness analysis is performed by thresholding the influential coordinates for each vertex and then choosing a label from these. As mentioned in the introduction, we use a randomized decoding procedure based on influences for soundness analysis. This is almost equivalent to Håstad's decoding procedure in his classic paper [10]. We are interested in keeping the exponents as small as possible, which is not possible with usual thresholding based decoding methods. Also this decoding procedure makes the analysis bit more transparent.

Lemma 3.9. *Given any integer $k \geqslant 6$, any constant $\mu > 0$, there exists a constant $\gamma' = \gamma'(\mu, k)$ such that for any 2-to-1 Label-Cover instance $\mathcal{L}$, it is possible to construct a graph G on vertex set $V(\mathcal{L}) \times [k]^R$ with the following properties:*

- *Completeness: If $\mathcal{L}$ is satisfiable, then G is k-colorable.*
- *Soundness: If $\mathsf{Opt}(\mathcal{L}) < \gamma'$, then G is $\left(\mu, \Omega\left(\frac{\mu^{2/(1-\frac{c}{k-1})}}{\sqrt{\ln(1/\mu)}}\right)\right)$-dense for some $1 \leqslant c \leqslant 4$*

Proof. Consider the graph constructed from $\mathcal{L}$ by the verifier in Section 3.4 of [9]. Completeness of this verifier is already proven in [9]. For the soundness part, assume that there is an indicator function $f(v, x) : V \times [k^2]^R \to \{0, 1\}$ with $\mathbb{E}[f] \geqslant \mu$ and fraction of edges inside is $\gamma \leqslant \frac{1}{3}\underline{\Gamma}_{\rho(T)}(\mu, \mu)$. Here $\rho(T) = \frac{c}{k-1}$ corresponds to the spectral radius of Markov operator T from that construction with $c \leqslant 4$ is a positive real. We will show it is possible to construct a good labeling for $\mathcal{L}$ from this f. After arithmetization:

$$\mathbb{E}_{u\sim U}\left[\mathbb{E}_{v,v'\sim\delta(u)}\left[\mathbb{E}_{x\sim[k]^{2r}, y\sim T^{\otimes r}x}\left[f(v, x) \cdot f(v', y)\right]\right]\right] = \gamma \leqslant \underline{\Gamma}_{\rho(T)}(\mu, \mu)/3 \tag{3}$$

where $\delta(u)$ denotes the neighbors of u. As usual, if we let $g_u(x) = \mathbb{E}_{v\sim\Gamma(u)}[f(v, \overline{x})]$, we have $\mathbb{E}_{u,x}[g_u(x)] = \mathbb{E}_u[\mu_u] \geqslant \mu$. Here if $x \in [k]^{2r}$, then $\overline{x} \in [k^2]^r$ with $\overline{x}_i = x_{2i-1} + kx_{2i}$.

Let $\varepsilon \triangleq \underline{\Gamma}_{\rho(T)}(\mu, \mu)/3 = \Omega\left(\frac{\mu^{2/(1-\frac{c}{k-1})}}{\sqrt{\ln(1/\mu)}}\right)$ with corresponding $\kappa(\varepsilon, \mu)$ and $\tau(\varepsilon, \mu)$ from noise stability proposition. Consider the following labeling, ℓ: Label each $v \in V$ (resp. $u \in U$) using $i \in \{1, \ldots, 2R\}$ (resp. $j \in \{1, \ldots, R\}$) with probability $\mathrm{Inf}_i^{\leqslant 2\kappa}(f(v))/(2\kappa)$ (resp. $\mathrm{Inf}_j^{\leqslant\kappa}(g_u)/\kappa$). With the remaining probability, assign arbitrary label. Since for any function h, $\sum_j \mathrm{Inf}_j^{\leqslant d}(h) \leqslant d\mathbf{Var}[h]$ and that $\mathbf{Var}[f(v)], \mathbf{Var}[g] \leqslant 1$, this is a valid probability distribution. Given a vertex $u \in U$, without loss of generality, assume $\pi_{vu}(2i-1) = \pi_{vu}(2i) = i$. The fraction of satisfied edges incident to this vertex u is given by the following expression:

$$\begin{aligned}
\mathsf{Pr}_{v\in\delta(u)}\left[\ell(v) \in \{2\ell(u), 2\ell(u)+1\}\right] &\geqslant \frac{1}{2\kappa^2}\mathbb{E}_{v\in\delta(u)}\left[\sum_{i=1}^{r} \mathrm{Inf}_i^{\leqslant\kappa}(g_u)\left(\mathrm{Inf}_{2i-1}^{\leqslant 2\kappa}(f(v)) + \mathrm{Inf}_{2i}^{\leqslant 2\kappa}(f(v))\right)\right] \\
&\geqslant \frac{1}{2\kappa^2}\sum_{i=1}^{r} \mathrm{Inf}_i^{\leqslant\kappa}(g_u)\mathbb{E}_{v\in\delta(u)}\left[\mathrm{Inf}_{2i-1}^{\leqslant 2\kappa}(f(v)) + \mathrm{Inf}_{2i}^{\leqslant 2\kappa}(f(v))\right] \\
&\geqslant \frac{1}{2\kappa^2}\sum_{i=1}^{r} \mathrm{Inf}_i^{\leqslant\kappa}(g_u)\sum_{k=0}^{1} \mathrm{Inf}_{2i-k}^{\leqslant 2\kappa}\mathbb{E}_{v\in\delta(u)}\left[f(v)\right] \geqslant \frac{1}{2\kappa^2}\sum_{i=1}^{r} \mathrm{Inf}_i^{\leqslant\kappa}(g_u)^2,
\end{aligned}$$

where we used the convexity of Inf and the fact that for two functions $h_1(x) = h_2(\overline{x})$, under the uniform distribution, $\mathrm{Inf}_i^{\leqslant d}(h_1) \leqslant \mathrm{Inf}_{2i-1}^{\leqslant 2d}(h_2) + \mathrm{Inf}_{2i}^{\leqslant 2d}(h_2)$.

Using the noise stability result from Proposition 3.7 and the fact that the spectral radius of T, $\rho = \rho(T) \leqslant \frac{c}{k-1}$ is the correlation constant, if $\mathbb{E}_{x,y\sim T}[g_u(x)g_u(y)] \leqslant \underline{\Gamma}_{\rho(T)}(\mu_u, \mu_u) - \varepsilon$, then $\sum_i \mathrm{Inf}_i^{\leqslant \kappa}(g_u)^2 \geqslant \tau^2$. Otherwise $\sum_i \mathrm{Inf}_i^{\leqslant \kappa}(g_u)^2 \geqslant 0$. Here we used $\mu_u = \mathbb{E}_x[g_u(x)]$. Therefore

$$\begin{aligned}
\sum_i \mathrm{Inf}_i^{\leqslant \kappa}(g_u)^2 \geqslant & \tau^2 \left[\underline{\Gamma}_{\rho(T)}(\mu_u, \mu_u) - \varepsilon - \mathbb{E}_{x,y\sim T}[g_u(x)g_u(y)]\right] \\
\mathbb{E}_u\left[\sum_i \mathrm{Inf}_i^{\leqslant \kappa}(g_u)^2\right] \geqslant & \tau^2 \mathbb{E}_u \left[\underline{\Gamma}_{\rho(T)}(\mu_u, \mu_u) - \varepsilon - \mathbb{E}_{x,y\sim T}[g_u(x)g_u(y)]\right] \\
\geqslant & \tau^2 \left\{ \mathbb{E}_u \left[\Omega\left(\frac{\mu_u^{2/(1-\frac{c}{k-1})}}{\sqrt{\ln(1/\mu_u)}} \right) \right] - \varepsilon - \mathbb{E}_u\left[\mathbb{E}_{x,y\sim T}[g_u(x)g_u(y)]\right] \right\} \\
\geqslant & \Omega\left(\tau^2 \frac{\mu^{2/(1-\frac{c}{k-1})}}{\sqrt{\ln(1/\mu)}} \right).
\end{aligned}$$

In the last line, we appealed to the convexity of $\Omega\left(\frac{\mu^{2/(1-\frac{c}{k-1})}}{\sqrt{\ln(1/\mu)}}\right)$ and Equation 3. Hence this labeling will satisfy at least a constant (independent of n and R) fraction of edges. Taking γ' appropriately, we see that if the initial label cover instance $\mathcal{L}$ was not satisfiable beyond γ', then no such set can exist. $\square$

Theorem 3.10. *For any integer $k \geqslant 6$ and integer constant Δ, given a graph G with maximum degree Δ, it is NP-hard to decide whether if it is k-colorable or largest independent set size is $\tilde{O}_\Delta\left(n\Delta^{-1+c/(k-1)}\right)$ assuming the 2-to-1 Conjecture, for some constant $c \leqslant 4$.*

Proof. Given a regular graph instance G from Lemma 3.9, we can construct another graph G' using Corollary 2.6 with $\lambda = 2/(1-\rho)$ with $\rho \leqslant \frac{c}{k-1}$ being the correlation constant. In this new graph, if there is no such independent set, this implies H is not $(\mu, \mu^{2/(1-\rho)})$-dense, which is NP-hard to decide under the 2-to-1 Conjecture. $\square$

4 2-Semicolorable 3-Uniform Hypergraphs

In this section, we give a similar result to the previous result, but assuming only semicolorability. In this context, we say a hypergraph is 2-semicolorable if there exists a 2-coloring of the graph which leaves at most ε-fraction hyperedges monochromatic. Our reduction is conditioned on the Unique Games Conjecture, which is the primary reason for losing the perfect completeness. UGC is very similar to the d-to-1 Conjecture introduced above with $d = 1$ and the perfect completeness is replaced with $1 - \varepsilon$ completeness.

Conjecture 4.1 (Unique Games Conjecture [13])**.** *For any $\gamma > 0, \varepsilon > 0$, there exists a right regular 1-to-1 Label-Cover instance $\mathcal{L}$ with $R = R(\gamma, \varepsilon)$ many labels such that it is NP-hard to decide between two cases,* $\mathsf{Opt}(\mathcal{L}) \geqslant 1 - \varepsilon$ *or* $\mathsf{Opt}(\mathcal{L}) \leqslant \gamma$.

4.1 The Not-All-Equal test

Given a Unique Games instance $\mathcal{L}$, our verifier is:

1. Choose $u \sim U$, and three vertices v_1, v_2, v_3 independent at random from the neighbors of v.
2. Choose $x, y, z \in \{0,1\}^R$ such that for each $i \in [R]$, (x_i, y_i, z_i) are drawn uniformly at random from the NAE $= \{(0,0,1),(0,1,0),(1,0,0),(0,1,1),(1,0,1),(1,1,0)\}$
3. Accept iff Not-All-Equal$(f(u_1,x), f(u_2,y), f(u_3,z))$.

In soundness analysis, we will bound the density property of this verifier using a generalized noise stability theorem of Mossel [16]. We will first compute the correlation constant, ρ, for the probability space of accepting predicates of Not-All-Equal.

Lemma 4.2. *The correlation between probability spaces $XY := \{0,1\}^R \times \{0,1\}^R$ and $Z := \{0,1\}^R$, is $\rho(XY,Z) = \frac{\sqrt{3}}{3}$ and $\rho(X,Y) = \frac{1}{3}$.*

Proof. First observe that our distribution is invariant under permutations.

$$\begin{aligned}
\rho(XY,Z) &= \sup\{\mathrm{Cov}[f,g] : f \in L^2(\{0,1\}^2), g \in L^2(\{0,1\}), \mathbf{Var}[f] = \mathbf{Var}[g] = 1\} \\
&= \sup_{f,g,\mathbb{E}[f]=0,\mathbb{E}[g]=0,\mathsf{Var}[f]=1,\mathsf{Var}[g]=1} \frac{f(00)g(1) + [f(01)+f(10)]\overbrace{[g(0)+g(1)]}^{0} + f(11)g(0)}{6} \\
&= \sup_{f,g,\mathbb{E}[f]=0,\mathbb{E}[g]=0,\mathsf{Var}[f]=1,\mathsf{Var}[g]=1} \frac{f(00)g(1) + f(11)g(0)}{6}
\end{aligned}$$

This last expression is maximized when $f(00)g(1) \geqslant 0$, $f(11)g(0) \geqslant 0$ and $f(01) = f(10) = 0$. Thus we have $\mathsf{Var}\,[f] = 1 \implies f(11) = -\sqrt{6 - f(00)^2}$. Similarly $g(1) = -\sqrt{2 - g(0)^2}$. Hence

$$\rho(XY,Z) = \max_{f(00),g(0)} -\frac{f(00)\sqrt{2-g(0)^2} + \sqrt{6-f(00)^2}\,g(0)}{6}$$

This is maximized for $f(00) = -\sqrt{3}$ and $g(0) = -1$, so $\rho(XY,Z) = \sqrt{3}/3$.

Similarly, for $\rho(X,Y)$, we have

$$\begin{aligned}
\rho(X,Y) &= \sup\{\mathrm{Cov}[f,g] : f \in L^2(\{0,1\}), g \in L^2(\{0,1\}), \mathbf{Var}[f] = \mathbf{Var}[g] = 1\} \\
&= \sup_{f,g,\mathbb{E}[f]=0,\mathbb{E}[g]=0,\mathsf{Var}[f]=1,\mathsf{Var}[g]=1} \frac{f(0)g(0) + 2(f(0)g(1) + f(1)g(0)) + f(1)g(1)}{6} \\
&= \sup_{f,g,\mathbb{E}[f]=0,\mathbb{E}[g]=0,\mathsf{Var}[f]=1,\mathsf{Var}[g]=1} \frac{f(0)g(1) + f(1)g(0) + \overbrace{(f(0)+f(1))(g(0)+g(1))}^{0}}{6} \\
&= \frac{1}{3} \qquad (f(0) = -f(1) = -g(0) = g(1) = -1).
\end{aligned}$$

□

4.2 Reduction from Unique Games

The following lemma gives us the necessary density argument so as to make the sparsification procedure go through.

Lemma 4.3. *Given any positive reals μ, ε, there exists $\gamma'(\mu, \varepsilon) > 0$, such that for any Unique Games instance $\mathcal{L}$, it is possible to construct a 3-uniform hypergraph G on vertex set $V(\mathcal{L}) \times \{0,1\}^R$ with the following properties:*

- *Completeness If $\mathsf{Opt}(\mathcal{L}) \geqslant 1 - \varepsilon/3$, then G has a $(2, \varepsilon)$-semicoloring.*
- *Soundness If $\mathsf{Opt}(\mathcal{L}) \leqslant \gamma'$, then G is $\left(\mu, \Omega\left(\mu^9/\ln(1/\mu)\right)\right)$-dense.*

Proof. Consider the queries of verifier described above. Each query can be thought of as a 3-uniform edge. For the completeness part, assume we are given an labeling which satisfies $1 - \varepsilon/3$ fraction of the constraints in $\mathcal{L}$. The corresponding dictator functions will pass the Not-All-Equal test with probability $(1-\varepsilon/3)^3 \geqslant 1-\varepsilon$. In other words, the long codes will form a $(2, \varepsilon)$-semicoloring for the resulting 3-uniform hypergraph.

For the soundness part, consider an indicator function $f(v, x) : V \times \{0,1\}^R \to \{0,1\}$ with $\mathbb{E}[f] \geqslant \mu$ having density $\gamma \leqslant \frac{1}{3}\underline{\Gamma}_{\left(\sqrt{1/3}, 1/3\right)}(\mu, \mu, \mu)$. Let $\varepsilon = \frac{1}{3}\underline{\Gamma}_{\left(\sqrt{1/3}, 1/3\right)}(\mu, \mu, \mu)$ and κ,τ be the corresponding parameters from noise stability. After arithmetization:

$$\mathbb{E}_{u \sim U}\left[\mathbb{E}_{v_1, v_2, v_3 \sim \delta(u)}\left[\mathbb{E}_{x,y,z}\left[f(v_1, x) \cdot f(v_2, y) \cdot f(v_3, z)\right]\right]\right] = \gamma.$$

As usual, let $g_u(x) = \mathbb{E}_{v \sim \delta(u)}[f(v, x)]$. Pick label i for each vertex with probability $\mathrm{Inf}_i^{\leqslant \kappa}/\kappa$ (pick arbitrary label with remaining probability). For the same reasons with k-colorability case, this is a valid probability distribution. Using the same arguments, we can show that the fraction of satisfied edges for fixed $u \in U$ is (after ordering labels so that $\pi_{uv}(i) = i$):

$$\mathsf{Pr}_{v \in \delta(u)}\left[\ell(v) = \ell(u)\right] \geqslant \frac{1}{\kappa^2} \sum_{i=1}^{r} \mathrm{Inf}_i^{\leqslant \kappa}(g_u)^2.$$

Using the noise stability result from Theorem 3.7 and correlation constants of this test, if

$$\begin{aligned}\mathbb{E}_{x,y,z}\left[g_u(x) g_u(y) g_u(z)\right] \leqslant & \underline{\Gamma}_{\left(\sqrt{1/3}, 1/3\right)}(\mu_u, \mu_u, \mu_u) - \varepsilon = \underline{\Gamma}_{\sqrt{1/3}}(\mu_u, \Omega(\mu_u^3/\sqrt{\ln(1/\mu_u)})) - \varepsilon \\ \leqslant & \Omega(\mu_u^9/\ln(1/\mu_u)) - \varepsilon,\end{aligned}$$

then $\sum_i \mathrm{Inf}_i^{\leqslant \kappa}(g_u)^2 \geqslant \tau^2$. Otherwise $\sum_i \mathrm{Inf}_i^{\leqslant \kappa}(g_u)^2 \geqslant 0$ ($\mu_u = \mathbb{E}_x[g_u(x)]$). Therefore

$$\begin{aligned}\sum_i \mathrm{Inf}_i^{\leqslant \kappa}(g_u)^2 \geqslant & \tau^2\left\{\Omega(\mu_u^9/\ln(1/\mu_u)) - \varepsilon - \mathbb{E}_{x,y,z}\left[g_u(x) g_u(y) g_u(z)\right]\right\} \\ \mathbb{E}_u\left[\sum_i \mathrm{Inf}_i^{\leqslant \kappa}(g_u)^2\right] \geqslant & \tau^2\left\{\mathbb{E}_u\left[\Omega(\mu_u^9/\ln(1/\mu_u))\right]\right\} - \varepsilon - \gamma \geqslant \Omega(\tau^2\mu^9/\ln(1/\mu)).\end{aligned}$$

Taking $\gamma' = \Omega(\tau^2\kappa^{-2}\mu^9/\ln(1/\mu))$ shows that a good labeling exists. $\square$

Theorem 4.4. *Assuming UGC, the following holds: For any constant $\varepsilon > 0$ and integer constant Δ, given a 3-uniform hypergraph G with maximum degree Δ, it is NP-hard to decide whether if it is $(2, \varepsilon)$-semicolorable or largest independent set size is $\widetilde{O}_{\Delta}\left(n\Delta^{-1/8}\right)$.*

Proof. As usual, given a 3-uniform hypergraph instance H from Lemma 4.3, we can construct another hypergraph H' using Corollary 2.6 with $\lambda = 9$. In this new graph, if there is no such independent set, this implies H is not $(\mu, \Omega(\mu^9/\ln(1/\mu)))$-dense. $\square$

Remark 2. Up to poly $\log \Delta$ factors, this result is tight. See [1, 4, 15].

5 r-Uniform Hypergraphs with Small Vertex Covers

In the previous sections, we gave strong inapproximability results for independent sets on bounded degree 2-colorable (or 3-colorable) hypergraphs. A natural question is how far can we push in the other direction: What is the best inapproximability we can get when all but a small fraction of vertices form an independent set? We know that when there is an independent set of size greater than $n(1 - 1/r + \delta)$, we can find one of size at least $r\delta n$ (by finding an r-approximate vertex cover). Although we do not obtain results for this case, we will present results for the case of hypergraphs which have an independent set of size $n(1 - O(\frac{\log r}{r}) - \varepsilon)$. Our formal result is the following.

Theorem 5.1. *Assuming UGC, the following holds: For all large enough integer constants r, Δ, given an r-uniform hypergraph G with maximum degree Δ, it is NP-hard to decide whether G has an independent set of size $n\left(1 - O\left(\frac{\log r}{r}\right)\right)$ or no independent set of size $O\left(n\left(\frac{\log \Delta}{\Delta}\right)^{\frac{1}{r-1}}\right)$.*

Our proof uses the following "independent set" version of the Unique Games Conjecture, which is shown to be equivalent to the original one by Khot and Regev [14]:

Conjecture 5.2. *For any $\gamma > 0, \varepsilon > 0$, there exists a right regular 1-to-1 Label-Cover instance $\mathcal{L}$ with $R = R(\gamma, \varepsilon)$ many labels such that it is NP-hard to decide between two cases:*

1. *Completeness There is a set $V' \subseteq V$ with $|V'| \geqslant (1-\varepsilon)|V|$ and a labeling function $\ell : U \cup V \to [R]$ which has $\ell(u) = \pi_{v \to u}(\ell(v))$ for every edge (u, v) with $u \in U$ and $v \in V'$.*
2. *Soundness $\mathsf{Opt}(\mathcal{L}) \leqslant \gamma$.*

First, we state the properties of a certain distribution whose existence is proved by Benjamini *et al.* (Theorem 27 from [3]):

Theorem 5.3. *There exists a pairwise independent distribution $\mathcal{D}$ on $\{0, 1\}^r$ with the following properties:*

1. $\mathsf{Pr}_{x \sim \mathcal{D}}[x = 1] = 0$.
2. *For all $i \in \{1, \ldots, r\}$, $\mathsf{Pr}_{x \sim \mathcal{D}}[x_i = 1] = 1 - O\left(\frac{\log r}{r}\right)$.*

Our verifier will, given a Unique Games instance $\mathcal{L}$, make queries in the usual way:

1. Choose vertex $u \sim U$,
2. Choose $v_1, \ldots, v_r \sim \Gamma(u)$,
3. Choose $x_1, \ldots, x_r \sim \mathcal{D}$,
4. Reject iff $f(v_1, \pi_{v_1 \to u}(x_1)) = \ldots = f(v_r, \pi_{v_r \to u}(x_r)) = 1$.

Due to the biased nature of this test, all vertices (v, x) have associated weights $\mathrm{wt}(v, x) = (1 - \rho)^{\mathrm{wt}(x)} \rho^{r - \mathrm{wt}(x)}$. Here $\mathrm{wt}(x)$ is the number of 1's in string x and $\rho = O(\log r / r)$. Note that it is easy to go back to the unweighted instances by simply duplicating each vertex proportional to its weight as observed by Dinur and Safra [6].

Lemma 5.4. *For any $r \geqslant C$, $\varepsilon > 0$ and $\mu > 0$, there exists $\gamma'(\mu, \varepsilon) > 0$, such that for any Unique Games instance $\mathcal{L}$, it is possible to construct an r-uniform hypergraph G on vertex set $V(\mathcal{L}) \times \{0, 1\}^R$ with the following properties:*

- *Completeness If there is a set V', $|V'| \geqslant |V|(1 - \varepsilon)$ as described in the completeness part of Conjecture 5.2, then G has an independent set of size $n\left(1 - O\left(\frac{\log r}{r}\right)\right)$.*
- *Soundness If $\mathsf{Opt}(\mathcal{L}) \leqslant \gamma'$, then G is $(\mu, \mu^r/3)$-dense.*

Proof. Completeness part is easy to see. Given a good labeling function ℓ with set V', let

$$f(v, x) = \begin{cases} x_{\ell(v)} & \text{if } v \in V', \\ 0 & \text{else.} \end{cases}$$

Assume that $f^{-1}(1)$ is not an independent set. Then $\exists u \in U$ with $v_1, \ldots, v_r \in V'$ such that $x_{\ell(v_i)} = 1$ and (u, v_i) are all constraint edges. But we have $\pi_{v_i \to u}(\ell(v_i)) = \pi_{v_j \to u}(\ell(v_j))$, which is impossible due to the first property of distribution $\mathcal{D}$.

For the soundness part, the procedure is exactly same with 3-uniform 2-semicolorability case. We will only point out the differences. Assume we are given a set of size μ with density $\leqslant \frac{1}{3}\mu^r$. We also pick $\varepsilon = \frac{1}{3}\mu^r$. Using the stability bounds for pairwise independent functions, at the last step, we obtain that the value of this labeling is at least $\tau^2 \kappa^{-2} \left[\mathbb{E}_u \left[\mu_u^r\right] - \frac{2}{3}\mu^r\right]$. For $r \geqslant 2$, μ_u^r is convex, hence $\mathbb{E}_u \left[\mu_u^r\right] - \frac{2}{3}\mu^r \geqslant \mu^r/3$. Thus we take $\gamma' = \tau^2 \kappa^{-2} \mu^r / 3$ which completes the proof. $\square$

Theorem 5.1 now follows by combining Lemma 5.4 and Corollary 2.6.

6 Acknowledgments

We thank Rishi Saket for pointing us to the reference [3] on the pairwise independent distribution used in the proof of Theorem 5.1. We thank Eric Blais, Anupam Gupta, and Ryan O'Donnell for useful comments on an earlier version of this write-up.

References

[1] N. Alon, P. Kelsen, S. Mahajan, and R. Hariharan. Approximate hypergraph coloring. *Nord. J. Comput.*, 3(4):425–439, 1996. 1, 2, 12

[2] P. Austrin, S. Khot, and M. Safra. Inapproximability of vertex cover and independent set in bounded degree graphs. *IEEE Conference on Computational Complexity*, pages 74–80, 2009. 2, 3

[3] I. Benjamini, O. Gurel-Gurevich, and R. Peled. On k-wise independent distributions and boolean functions. *in preparation*, 2008. 12, 13

[4] H. Chen and A. M. Frieze. Coloring bipartite hypergraphs. In *IPCO*, pages 345–358, 1996. 1, 2, 12

[5] I. Dinur, E. Mossel, and O. Regev. Conditional hardness for approximate coloring. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, pages 344–353, 2006. 7

[6] I. Dinur and S. Safra. On the hardness of approximating minimum vertex cover. *Annals of Mathematics*, 162(1):439–485, 2005. 13

[7] A. Frieze and D. Mubayi. Coloring simple hypergraphs. *arXiv:0809.2979v2 [math.CO], under submission*, 2008. 2, 6

[8] V. Guruswami, J. Håstad, and M. Sudan. Hardness of approximate hypergraph coloring. *SIAM J. Comput.*, 6(31):1663–1686, 2002. 4, 5

[9] V. Guruswami and A. K. Sinop. Improved inapproximability results for maximum k-colorable subgraph. *APPROX-RANDOM*, pages 163–176, 2009. 2, 6, 8

[10] J. Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48:798–859, 2001. 8

[11] J. Holmerin. Vertex cover on 4-regular hyper-graphs is hard to approximate within 2-epsilon. In *STOC*, pages 544–552, 2002. 4, 5

[12] D. R. Karger, R. Motwani, and M. Sudan. Approximate graph coloring by semidefinite programming. *J. ACM*, 45(2):246–265, 1998. 1

[13] S. Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 767–775, 2002. 2, 3, 7, 9

[14] S. Khot and O. Regev. Vertex cover might be hard to approximate to within 2- epsilon. *J. Comput. Syst. Sci.*, 74(3):335–349, 2008. 12

[15] M. Krivelevich, R. Nathaniel, and B. Sudakov. Approximating coloring and maximum independent sets in 3-uniform hypergraphs. *J. Algorithms*, 41(1):99–113, 2001. 2, 12

[16] E. Mossel. Gaussian bounds for noise correlation of functions. To Appear in GAFA, 2010. 2, 7, 10

A Proof of Corollary 2.6

Corollary 2.6 (restated) *Given an r-uniform hypergraph $H = (V, E)$, any positive constant $\varepsilon \geqslant 0$, $\lambda > 1$ and Δ some large integer, there exists a constant $\mu = \mu(\Delta, \lambda)$ and an r-uniform hypergraph $H' = (V, E')$ with maximum degree Δ such that:*

- *Completeness: If H is (k, ε)-semicolorable, then H' is (k, ε)-semicolorable also.*
- *Soundness: If H is $(\mu, \frac{\mu^\lambda}{3})$-dense, H' has no independent set of size $O\left(n\left(\frac{\log \Delta}{\Delta}\right)^{1/(\lambda-1)}\right)$.*

Proof. Choose $\mu = \left(\frac{\Delta}{\log \Delta}\right)^{1/(1-\lambda)}$. For large enough Δ, $(1-\mu)\log(1-\mu) \geqslant -\mu/2$.

$$-\mu \log \mu = \frac{1}{\lambda - 1}\left(\frac{\log^\lambda \Delta}{\Delta}\right)^{1/(\lambda-1)} - \frac{\log\log \Delta}{\lambda - 1}\left(\frac{\log \Delta}{\Delta}\right)^{1/(\lambda-1)}.$$

$$\begin{aligned}
\mathrm{h}(\mu) &= -\mu \log \mu - (1-\mu)\log(1-\mu) \\
&\leqslant \frac{1}{\lambda-1}\left(\frac{\log^\lambda \Delta}{\Delta}\right)^{1/(\lambda-1)} - \frac{\log\log\Delta}{\lambda-1}\left(\frac{\log\Delta}{\Delta}\right)^{1/(\lambda-1)} + \frac{1}{2}\left(\frac{\log\Delta}{\Delta}\right)^{1/(\lambda-1)} \\
&\leqslant \frac{1}{\lambda-1}\left(\frac{\Delta}{\log^\lambda \Delta}\right)^{1/(1-\lambda)}.
\end{aligned}$$

Applying the sparsification procedure from Lemma 2.5, we obtain another hypergraph H' with degree of any node, Δ_u satisfying

$$\Delta_u \leqslant 2\mathrm{h}(\mu)\mu^{-\lambda} = O\left(\Delta^{1/(1-\lambda)}\left(\log^{-\lambda/(1-\lambda)}\Delta\right)\Delta^{-\lambda/(1-\lambda)}\left(\log^{\lambda/(1-\lambda)}\Delta\right)\right) = O(\Delta).$$

Consequently, using the previous lemma, if H' has an independent set of size $\Omega\left(n\left(\frac{\Delta}{\log\Delta}\right)^{1/(1-\lambda)}\right)$, then H is not $(\mu, \mu^\lambda/3)$-dense. $\square$

B Gaussian Stability Estimates

Lemma B.1. *For $\rho > 0$, $\lambda \geqslant 1$ fixed constants, and small enough $\mu > 0$, we have*

$$\underline{\Gamma}_\rho(\mu, \mu^\lambda) = \Omega\left(\frac{\mu^{\frac{1+\lambda+2\sqrt{\lambda}\rho}{1-\rho^2}}}{\sqrt{\ln 1/\mu}}\right)$$

Proof. Let t, t' be such that $N(-t) = \mu$ and $N(-t') = \mu^\lambda$ respectively. We know that $t = \sqrt{2\ln(1/\mu) - O(\ln\ln(1/\mu))}$ For $\mu < \frac{1}{2}, \lambda \geqslant 1, t, t' > 0$. We have the following inequality $N(-\sqrt{\lambda}t) =$

$O\big(N(-t)^\lambda\sqrt{\ln 1/N(t)}^{\lambda-1}\big) = O\left(\mu\left(\mu\sqrt{\ln 1/\mu}\right)^{\lambda-1}\right) = O(N(-t')\sqrt{\ln(1/N(-t))})$. Hence $\frac{N(-\sqrt{\lambda}t)}{\sqrt{\ln(1/N(-t))}} = O(N(-t'))$.

Thus;

$$\begin{aligned}
\underline{\Gamma}_\rho(\mu,\mu^\lambda) =&\mathsf{Pr}_{X,Y}\left[X \leqslant -t \wedge -\rho X + \sqrt{1-\rho^2}Y \leqslant -t'\right]\\
=&N(-t)\mathsf{Pr}_{X,Y}\left[-\rho X + \sqrt{1-\rho^2}Y \leqslant -t' \middle| X \leqslant -t\right]\\
=&\mu\mathsf{Pr}_{X,Y}\left[\sqrt{1-\rho^2}Y \leqslant -t' + \rho X \middle| X \leqslant -t\right]\\
\geqslant&\mu\mathsf{Pr}_{X,Y}\left[\sqrt{1-\rho^2}Y \leqslant -t' + \rho X \wedge X \geqslant -t(1+\varepsilon) \middle| X \leqslant -t\right]\\
\geqslant&\mu\mathsf{Pr}_{X}\left[X \geqslant -t(1+\varepsilon) \middle| X \leqslant -t\right]\\
&\ \mathsf{Pr}_{X,Y}\left[\sqrt{1-\rho^2}Y \leqslant -t' + \rho X \middle| X = -t(1+\varepsilon)\right]\\
\geqslant&\mu\mathsf{Pr}_{X}\left[X \geqslant -t(1+\varepsilon) \middle| X \leqslant -t\right]\mathsf{Pr}_{Y}\left[\sqrt{1-\rho^2}Y \leqslant -t' - \rho t(1+\varepsilon)\right]\\
\geqslant&\mu(1-\exp(-t^2(\varepsilon+\varepsilon^2/2)))N(-(t'+\rho t(1+\varepsilon))/\sqrt{1-\rho^2})\\
\geqslant&\mu(1-\exp(-t^2(\varepsilon+\varepsilon^2/2)))\Omega\left(\frac{1}{t'+\rho t(1+\varepsilon)}\exp(-t^2(\sqrt{\lambda}+\rho(1+\varepsilon))^2/[2(1-\rho^2)])\right).
\end{aligned}$$

Let $\varepsilon = 1/t^2$. Then:

$$\begin{aligned}
\underline{\Gamma}_\rho(\mu,\mu^\lambda) \geqslant&\mu(1-\exp(-(1+1/(2t^2))))\\
&\Omega\left(\frac{1}{t'+\rho t(1+\varepsilon)}\exp(-\frac{t^2}{2}(\sqrt{\lambda}+\rho(1+\varepsilon))^2/[(1-\rho^2)])\right)\\
\geqslant&\Omega\left(\frac{\mu}{\sqrt{\ln(1/\mu)}}\mu^{\frac{(\sqrt{\lambda}+\rho)^2}{1-\rho^2}}\right) = \Omega\left(\frac{\mu^{\frac{1+\lambda+2\sqrt{\lambda}\rho}{1-\rho^2}}}{\sqrt{\ln 1/\mu}}\right).
\end{aligned}$$

□

ECCC ISSN 1433-8092

http://eccc.hpi-web.de