Electronic Colloquium on Computational Complexity, Report No. 155 (2015)

Nearly Optimal NP-Hardness of Unique Coverage

Venkatesan Guruswami* Euiwoong Lee†

Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213.

Abstract

The *Unique Coverage* problem, given a universe V of elements and a collection E of subsets of V, asks to find $S \subseteq V$ to maximize the number of $e \in E$ that intersects S in *exactly one* element. When each $e \in E$ has cardinality at most k, it is also known as 1*-in-k Hitting Set*, and admits a simple $\Omega(\frac{1}{\log k})$-approximation algorithm.

For constant k, we prove that 1-in-k Hitting Set is NP-hard to approximate within a factor $O(\frac{1}{\log k})$. This improves the result of Guruswami and Zhou [SODA'11, ToC'12], who proved the same result assuming the Unique Games Conjecture. For Unique Coverage, we prove that it is hard to approximate within a factor $O(\frac{1}{\log^{1-\varepsilon} n})$ for any $\varepsilon > 0$, unless NP admits quasipolynomial time algorithms. This improves the results of Demaine et al. [SODA'06, SICOMP'08], including their $\approx 1/\log^{1/3} n$ inapproximability factor which was proven under the Random 3SAT Hypothesis. Our simple proof combines ideas from two classical inapproximability results for Set Cover and Constraint Satisfaction Problem, made efficient by various derandomization methods based on bounded independence.

*Supported in part by NSF grant CCF-1115525. guruswami@cmu.edu
†Supported by a Samsung Fellowship and NSF CCF-1115525. euiwoonl@cs.cmu.edu

ISSN 1433-8092

1 Introduction

Given a universe V of n elements and a collection E of m subsets of V, the Unique Coverage problem asks to find $S \subseteq V$ to maximize the number of $e \in E$ that intersects S in exactly one element. When each $e \in E$ has size at most k, this problem is also known as 1-in-k Hitting Set (abbreviated as 1-in-k HS).

Unique Coverage models numerous practical situations where each element represents a *service* and each subset represents a *customer* interested in services it contains. We want to *activate* some services to satisfy customers, but customers want exactly one service from her list to be activated because more than one service may lead to confusion or high cost. These natural scenarios have been studied in many fields including wireless networks, radio broadcast, and envy-free pricing. We refer the reader to the work of Guruswami and Trevisan [6] and Demaine et al. [3] for a more detailed list of applications. Chalermsook et al. [2] showed an approximation-preserving reduction from Unique Coverage to a special case of envy-free pricing called the *Tollbooth Pricing* problem, so our result improves the hardness of Tollbooth Pricing as well.

There is a simple $\Omega(\frac{1}{\log k})$-approximation algorithm for 1-in-k HS. First, consider the case where each subset e has the same cardinality k (also known as 1-in-Ek HS). Independently adding each $v \in V$ to S with probability $\frac{1}{k}$ will ensure that each set $e \in E$ will intersect S in exactly one element with probability $(1-\frac{1}{k})^{k-1}$, which approaches $\frac{1}{e}$ as k grows. For the general case where each subset has cardinality at most k (assume k is a power of 2), randomly choosing a value $l \in \{2,4,8,\ldots,k\}$ first and independently adding each $v \in V$ to S with probability $\frac{1}{l}$ will give an $\Omega(\frac{1}{\log k})$-approximation algorithm. If there exists $S \subseteq V$ that intersects *every* subset in exactly one element, solving the standard LP relaxation and independently rounding with the resulting solution will guarantee a factor $1/e$-approximation even if the subsets have different sizes [6].

These approximation algorithms highlight interesting theoretical aspects of this problem. 1-in-k HS can be naturally interpreted as a Constraint Satisfaction Problem (CSP) where each element $v \in V$ becomes a variable taking a value from $\{0,1\}$ ($v \leftarrow 1$ corresponds to $v \in S$), and each subset becomes a constraint. Each constraint $e = (v_1,\ldots,v_l)$ is satisfied by an assignment $\sigma : V \rightarrow \{0,1\}$ if and only if $\sigma(v_1)+\cdots+\sigma(v_l) = 1$. An $\Omega(1)$-approximation for 1-in-Ek HS and an $\Omega(\frac{1}{\log k})$-approximation for 1-in-k HS exhibit an example where mixing predicates of different arities decreases the best approximation ratio significantly. The second $\Omega(1)$-approximation when every subset can be intersected exactly once shows a rare example where *perfect completeness* of a CSP allows a much better approximation. When k is a growing function of n, as pointed out in [3], Unique Coverage is one of few natural *maximization* problems for which the tight approximation threshold is (semi)-logarithmic.

There are even more theoretically interesting developments from the hardness side. Demaine, Feige, Hajiaghayi, and Salavatipour [3] showed it is hard to approximate Unique Coverage within a factor of $\Omega(\frac{1}{\log^{\varepsilon} n})$ for some constant $\varepsilon > 0$ depending on δ, assuming that $\mathsf{NP} \not\subseteq \mathsf{BPTIME}(2^{n^{\delta}})$ for some constant $\delta > 0$. Their second result proved that the inapproximability can be strengthened to $\Omega(\frac{1}{\log^{1/3-\varepsilon} n})$ for any $\varepsilon > 0$ assuming Feige's Random 3SAT Hypothesis [5]. For 1-in-k HS for constant k, Guruswami and Zhou [7] recently proved that the $\Omega(\frac{1}{\log k})$-approximation is optimal, assuming Khot's Unique Games Conjecture [8]. As many other problems whose strong inapproximabilities are known only under one of Feige's or Khot's conjecture, it was open whether we were able to *bypass* these conjectures to show almost optimal inapproximability only assuming $\mathsf{P} \neq \mathsf{NP}$ or $\mathsf{NP} \not\subseteq \mathsf{QP}$.

Our main contribution in this work is a positive answer to this question. For 1-in-k HS for constant k, we prove the following theorem.

Theorem 1.1. *Assuming* $\mathsf{P} \neq \mathsf{NP}$*, for large enough constant k, there is no polynomial time algorithm that approximates* 1*-in-k HS within a factor better than* $O(\frac{1}{\log k})$.

This result bypasses the Unique Games Conjecture to show that the simple $\Omega(\frac{1}{\log k})$-approximation algorithm is the best polynomial time algorithm up to a constant factor. For Unique Coverage, we prove that following theorem. Recall that $\mathsf{QP} = \cup_{c \in \mathbb{N}} \mathsf{DTIME}(2^{\log^c n})$.

Theorem 1.2. *Assuming* $\mathsf{NP} \not\subseteq \mathsf{QP}$*, for any* $\varepsilon > 0$*, there is no polynomial time algorithm that approximates Unique Coverage within a factor better than* $\frac{1}{\log^{1-\varepsilon} n}$.

Compared to the first result of Demaine et al., we replace their assumption $\mathsf{NP} \not\subseteq \mathsf{BPTIME}(2^{n^\delta})$ for *some* δ by a much weaker assumption $\mathsf{NP} \not\subseteq \mathsf{QP}$ and at the same time show an improved (and near-optimal) inapproximability factor, which is near-optimal and also improves their second result conditioned on the Random 3SAT Hypothesis.

Besides these improvements, our proof is also significantly simpler than previous works. The result of Guruswami and Zhou for constant k is obtained by constructing a gap instance for a semidefinite programming (SDP) relaxation for the problem, and using the sophisticated result of Raghavendra [12] that converts a SDP gap to a Unique Games hardness. Demaine et al. first showed a reduction from Unique Coverage to an intermediate problem called Balanced Bipartite Independent Set (BBIS), and uses the Random 3SAT Hypothesis or Khot's Quasirandom PCP [9] to prove hardness of BBIS. Our two theorems are corollaries of one simple reduction from the basic Label Cover, whose hardness relies only on the PCP theorem and the Parallel Repetition Theorem.

1.1 Techniques

While Unique Coverage can be interpreted as a CSP, it also seems similar to the Max k-Coverage problem, where given a set system (V, E), we want to find a subset $S \subseteq V$ with $|S| = k$ that intersects as many $e \in E$ as possible.[1] Max k-Coverage is tightly related to the more famous Set Cover problem, and admits an $\frac{e-1}{e}$-approximation algorithm which is proved to be tight [11, 4]. It can be also interpreted as a variant of CSPs where each element becomes a variable taking a value from $\{0, 1\}$, and each subset becomes a constraint that is satisfied if at least one of its variables is assigned 1, and we additionally require that at most k variables have to be assigned 1.

A weaker but simple inapproximability of Max k-Coverage can be proved via the Label Cover problem. An instance of Label Cover consists of a biregular bipartite graph $G = (U_G \cup V_G, E_G)$ where each edge $e = (u, v)$ is associated with a projection $\pi_e : [R] \mapsto [L]$ for some positive integers R and L, and we look for a labeling $l : U_G \cup V_G \mapsto [R]$ that satisfies as many $e \in E_G$ as possible

[1] Max k-Coverage is usually stated in terms of the dual set system, where we want to find a subcollection $E' \subseteq E$ of subsets with $|E'| = k$ that maximizes the number of elements covered.

($e = (u,v)$ is satisfied when $\pi_e(l(v)) = l(u)$). Given an instance of Label Cover, the reduction to Max k-Coverage makes every (vertex, label)-pair of Label Cover as an element of the set system, and for each projection $e = (u,v) \in E_G$ and $b \in \{0,1\}^L$, there is a subset corresponding to (e,b) containing $\{(u,j) : b_j = 0\} \cup \{(v,j) : b_{\pi_e(j)} = 1\}$. It is a simple but useful exercise to check that if a labeling l satisfies every projection, its *canonical set* $\{(v,l(v)) : v \in U_G \cup V_G\}$ will intersect every subset exactly once. However, it is also easy to see that for any labeling l, its canonical set will intersect at least half of subsets exactly once.

To prove a stronger inapproximability result, we have a subset for each tuple $(e_1,\ldots,e_q,b)$ for various values of q where $e_1,\ldots,e_q$ share an endpoint in U_G. If l satisfies *all* $e_1,\ldots,e_q$, its canonical set will intersect $(e_1,\ldots,e_q,b)$ in exactly one element for many (but not all) b, but if l does not even *approximately* satisfy $e_1,\ldots,e_q$, there is no way to intersect many subsets in exactly one element. Even though our technique is different from traditional hardness results for Max-CSPs (e.g., no long code consisting of variables), the idea of *probabilistic checking* (i.e., subsets have weights summing up to 1, and the instance is interpreted as a probabilistic procedure where we sample a subset e according to weights and check $|S \cap e| = 1$) conceptually simplifies the proof, and technically makes the reduction efficient by appealing to various derandomization methods based on bounded independence.

1.2 Preliminaries

An instance of Unique Coverage is simply a set system. We view the set system as a hypergraph $H = (V_H, E_H)$ where V_H is the universe of elements and E_H is a collection of hyperedges. Unless stated otherwise, every log in this work indicates a logarithm base 2. We use $a \sim \mathscr{D}$ to indicate that a random variable a is sampled from a distribution $\mathscr{D}$. When a random variable a is sampled uniformly from a set A, we write $a \in A$. For a positive integer m, we denote $[m] := \{1,2,\ldots,m\}$.

2 Reduction from Label Cover

Our main reduction is from Label Cover. An instance of Label Cover consists of a biregular bipartite graph $G = (U_G \cup V_G, E_G)$ where each edge $e = (u,v)$ is associated with a projection $\pi_e : [R] \mapsto [L]$ for some positive integers R and L. For $u \in U_G$, let $N(u)$ denote its neighbors and $D := |N(u)|$ be the left degree. We additionally require that every projection π_e is d-regular, i.e., $R = dL$ and for every $l \in [L]$, $|\pi_e^{-1}(l)| = d$. A labeling $l : U_G \cup V_G \mapsto [R]$ satisfies $e = (u,v)$ when $\pi_e(l(v)) = l(u)$. The standard application of PCP Theorem, Parallel Repetition Theorem, and the trick of Wenner [13] to make each projection d-regular[2], implies the following theorem.

Theorem 2.1. *There exists an absolute constant $\tau < 1$ such that the following is true. For any positive integer $r > 0$, there is a reduction that given an instance ϕ of 3SAT with n variables, outputs an instance of Label Cover $(G, \{\pi_e\}_e)$ with $|U_G|, |V_G| = n^{O(r)}$, $R = 10^r, L = 2^r, d = D = 5^r$ in time $n^{O(r)}$, and satisfies the following.*

- *Completeness: If ϕ is satisfiable, there exists a labeling that satisfies every projection.*

[2]The basic 2-prover game based on 3SAT does not make the projections d-regular, but a simple trick allows us to assume this without loss of generality. See Theorem 1.17 of Wenner [13] for the formal proof.

- *Soundness: If ϕ is not satisfiable, every labeling satisfies at most τ^r fraction of projections.*

Given an instance of Label Cover $G = (U_G \cup V_G, E_G)$ with projections $\{\pi_e\}_{e \in E_G}$ with parameters R, L, D, d, we produce an instance $H = (V_H, E_H)$ of Unique Coverage. The set of vertices V_H is defined to be $V_G \times [R]$. In the following, we describe a probabilistic procedure to sample a hyperedge e. E_H is defined to be the set of hyperedges with nonzero probability, with these probabilities as weights. We abuse notation and let E_H also denote the distribution. There are three distributions used to describe the entire procedure.

1. Let Q be a positive integer to be determined later. We will take Q to be a power of 2 and $Q < D$. Let $\mathscr{D}$ be a uniform distribution on $\{2, 4, 8, \ldots, Q\}$.

2. For each $u \in U_G$, let $\mathscr{D}_{u,Q}$ be a *uniform pairwise independent* distribution on $(v_1, \ldots, v_Q) \in N(u)^Q$ such that

$$\Pr_{(v_1,\ldots,v_Q)\sim\mathscr{D}_{u,Q}}[v_i = v, v_j = v'] = \frac{1}{D^2} \qquad \text{for all } i \neq j \in [Q] \text{ and } v, v' \in N(u),$$

   and its support has size D^2. Note that it implies that $\Pr[v_i = v] = \frac{1}{D}$ for all $i \in [Q]$ and $v \in N(u)$.

   Claim 2.2. *Such a distribution $\mathscr{D}_{u,Q}$ exists.*

   Proof. $\mathscr{D}_{u,Q}$ can be described by the following standard procedure. Fix a bijection f from $N(u)$ to the finite field $\mathbb{F}_D$ (recall D is a power of 5), another injective mapping g from $[Q]$ to a subset of $\mathbb{F}_D$, sample $a, b \in \mathbb{F}_D$ independnetly, and output $v_i \leftarrow f^{-1}(a \cdot g(i) + b)$. It is a standard fact that this distribution is uniform pairwise independent. ☐

3. Let $\mathscr{D}_L$ be a uniform 4-wise independent distribution on $(c_1, \ldots, c_L) \in \{0,1\}^L$ such that

$$\Pr_{(c_1,\ldots,c_L)\sim\mathscr{D}_L}[c_{j_i} = b_i \text{ for } 1 \leqslant i \leqslant 4] = \frac{1}{2^4} \ \forall\, (j_1, j_2, j_3, j_4) \in \binom{[L]}{4} \text{ and } (b_1, b_2, b_3, b_4) \in \{0,1\}^4,$$

   and its support has size $2L^2$.

   Claim 2.3. *Such a distribution $\mathscr{D}_L$ exists.*

   Proof. $\mathscr{D}_L$ can be described by the following standard procedure. Fix a bijection f from $[L]$ to the finite field $\mathbb{F}_L$ (recall L is a power of 2). Sample $a, b \in \mathbb{F}_L$ and $d \in \mathbb{F}_2$ independently, and output $c_i \leftarrow \mathrm{Tr}(a \cdot f(i)^3 + b \cdot f(i)) + d$ where Tr is the Trace map from $\mathbb{F}_L$ to $\mathbb{F}_2$. This distribution is uniform over codewords of a dual BCH code (the dual of the extended form of the BCH code of designed distance 5 to be precise). It is well known that this distribution is uniform 4-wise independent (eg. see the monograph of Luby and Wigderson [10]). ☐

Given these distributions, a random hyperedge e is sampled by the following procedure.

- Sample $u \in U_G$.

- Sample $q \sim \mathscr{D}$.
- Sample $(v_1, \dots, v_Q) \sim \mathscr{D}_{u,Q}$. Note that only $v_1, \dots, v_q$ are used in the reduction. This slightly redundant sampling reduces the number of distributions involved and simplifies our analysis.
- Sample $(c_1, \dots, c_L) \sim \mathscr{D}_L$.
- For $j \in [L]$, consider a *block* of vertices $\cup_{i=1}^{q}(\{v_i\} \times \pi^{-1}_{(u,v_i)}(j))$. Every block has cardinality at most qd. It has exactly qd vertices when $v_1, \dots, v_Q$ are pairwise distinct.
  - If $c_j = 0$, add d vertices in $\{v_1\} \times \pi^{-1}_{(u,v_1)}(j)$ to e.
  - If $c_j = 1$, add the entire block to e.

Note that the maximum cardinality of any hyperedge is RQ. Also note that $|V_H| = |V_G| \cdot R$ and the total number of hyperedges with nonzero probability is bounded by $|U_G| \cdot \log Q \cdot D^2 \cdot 2L^2$.

2.1 Completeness

Lemma 2.4. *If the instance $(G, \{\pi_e\}_e)$ of Label Cover admits a labeling l that satisfies every projection, there exists $S \subseteq V_H$ such that* $\Pr_{e \sim E_H}[|e \cap S| = 1] \geqslant \frac{1}{2}$.

Proof. Given a labeling $l : U_G \cup V_G \mapsto [R]$, let $S := \cup_{v \in V_G}\{(v, l(v))\}$. From the above probabilistic procedure to sample a hyperedge e, every choice of u, q, $v_1, \dots, v_q$ satisfies $\pi_{(u,v_1)}(l(v_1)) = \cdots = \pi_{(u,v_q)}(l(v_q))$. In particular, S and $\cup_{i=1}^{q}(\{v_i\} \times [R])$ intersect in exactly one block corresponding to $l(u) \in [L]$. Therefore, if $c_{l(u)} = 0$, which happens with probability $\frac{1}{2}$, $(v_1, l(v_1))$ is the only element in $S \cap e$. $\square$

2.2 Soundness

For soundness, we prove that if the Label Cover instance $(G, \{\pi_e\})$ does not admit a good labeling, the Unique Coverage instance H does not have a good solution either.

Lemma 2.5. *If every labeling $l : U_G \cup V_G \mapsto [R]$ satisfies at most ε fraction of projections, for every $S \subseteq V_H$,* $\Pr_{e \sim E_H}[|e \cap S| = 1] \leqslant 2Q\sqrt{\varepsilon} + \frac{Q^2}{D} + O\left(\frac{1}{\log Q}\right)$.

Proof. Fix $S \subseteq V_H$. We construct a partial labeling $l : V_G \mapsto [R]$ as follows. For each $v \in V_G$ and $j \in [R]$, we say that v *picks* j if $(v, j) \in S$. If v picked at least one label, we choose an arbitrary picked label j and set $l(v) = j$. Otherwise, we set $l(v) = \emptyset$, which means that every projection including v will not be satisfied. Note that we have not defined labels for U_G yet.

For $q \in \{2, 4, 8, \dots, Q\}$, $u \in U_G$, and $v_1, \dots, v_q \in N(u)$, we say that $(u, v_1, \dots, v_q)$ is *weakly satisfied* by a partial labeling to V_G if there exist $1 \leqslant i < j \leqslant q$ such that $l(v_i) \neq \emptyset, l(v_j) \neq \emptyset$, and $\pi_{(u,v_i)}(l(v_i)) = \pi_{(u,v_j)}(l(v_j))$. Note that if the Label Cover instance admitted a labeling l^* that satisfied every projection, every tuple $(u, v_1, \dots, v_q)$ with $v_1, \dots, v_q \in N(u)$ would satisfy $\pi_{(u,v_1)}(l^*(v_1)) = \cdots = \pi_{(u,v_q)}(l^*(v_q))$. The following claim shows that since the Label Cover instance does not admit a good labeling, if we sample $u, v_1, \dots, v_Q$ as in the reduction, $(u, v_1, \dots, v_Q)$ is unlikely to be even weakly satisfied.

Claim 2.6. *Suppose we sample $u \in U_G$ and $(v_1, \dots, v_Q) \sim \mathscr{D}_{u,Q}$. The probability that $(u, v_1, \dots, v_Q)$ is weakly satisfied is at most $Q^2\varepsilon$.*

Proof. Fix $1 \leqslant i < j \leqslant Q$. By uniform pairwise independence of $\mathscr{D}_{u,Q}$, the probability that $\pi_{(u,v_i)}(l(v_i)) = \pi_{(u,v_j)}(l(v_j))$ is equal to the expected fraction of satisfied projections of the randomized extension to the current labeling l where each $u \in U_G$ picks a random neighbor $v \in N(u)$ and set $l(u) \leftarrow \pi_{(u,v)}(l(v))$. Since every labeling satisfies at most ε fraction of constraints, the probability that $\pi_{(u,v_i)}(l(v_i)) = \pi_{(u,v_j)}(l(v_j))$ is at most ε. The claim follows by taking union bound over $\binom{Q}{2}$ pairs. $\square$

By an averaging argument, the fraction of $u \in U_G$ such that

$$\Pr_{(v_1,\dots,v_Q)\sim\mathscr{D}_{u,Q}}[(u, v_1, \dots, v_Q) \text{ is weakly satisfied}] > Q\sqrt{\varepsilon}$$

is at most $Q\sqrt{\varepsilon}$. Call these u *bad*, and fix a *good* u. Since the probability that $v_i = v_j$ is exactly $\frac{1}{D}$ for fixed $i \neq j$,

$$\Pr_{(v_1,\dots,v_Q)\sim\mathscr{D}_{u,Q}}[\exists i \neq j \text{ s.t. } v_i = v_j] \leqslant \binom{Q}{2}\frac{1}{D} \leqslant \frac{Q^2}{D}. \tag{1}$$

Fix $q \in \{2, 4, 8, \dots, Q\}$. For fixed u, by the definition of weak satisfaction, the probability of weak satisfaction decreases as the number of considered neighbors q decreases, i.e.,

$$\begin{aligned}
&\Pr_{(v_1,\dots,v_Q)\sim\mathscr{D}_{u,Q}}[(u, v_1, \dots, v_q) \text{ is weakly satisfied}] \\
\leqslant &\Pr_{(v_1,\dots,v_Q)\sim\mathscr{D}_{u,Q}}[(u, v_1, \dots, v_Q) \text{ is weakly satisfied}] \\
\leqslant & Q\sqrt{\varepsilon}.
\end{aligned} \tag{2}$$

Fix $v_1, \dots, v_q \in N(u)$ such that $(u, v_1, \dots, v_q)$ is not weakly satisfied and $v_1, \dots, v_q$ are pairwise distinct. Let $p := |\{i \in [q] : l(v_i) \neq \emptyset\}|$. The fact that $(u, v_1, \dots, v_q)$ is not weakly satisfied implies that for at least p values of $j \in [L]$, the corresponding block $\cup_{i=1}^{q}(\{v_i\} \times \pi^{-1}_{(u,v_i)}(j))$ intersects S. Consider the probabilistic procedure to sample a hyperedge e as in the reduction.

Claim 2.7. *Given u, q, $v_1, \dots, v_q$, and p satisfying the conditions above, the probability that $|e \cap S| = 1$ is at most*

$$\begin{cases} 0 & p = 0 \\ 1 & 1 \leqslant p \leqslant 4 \\ O\left(\frac{1}{p^2}\right) & p > 4. \end{cases}$$

Proof. The above upper bounds are clear when $p = 0$ or $1 \leqslant p \leqslant 4$. For $p \geqslant 5$, note that if at least 2 of p different blocks see $c_j = 1$ and decide to add the whole block to e, $|e \cap S| \geqslant 2$. Therefore, if we let $\{i \in [q] : l(v_i) \neq \emptyset\} = \{i_1, \dots, i_p\}$, $\Pr[|e \cap S| = 1]$ is at most the probability that $c_{i_1} + \cdots + c_{i_p} \leqslant 1$ when $(c_1, \dots, c_L) \sim \mathscr{D}_L$. We use the following concentration inequality for the sum of k-wise independent random variables by Bellare and Rompel [1].

Theorem 2.8 ([1]). *Let k be an even integer, and let X be the sum of n k-wise independent random variables taking values in $[0,1]$. Let $\mu = \mathbb{E}[X]$ and $a > 0$. Then we have*

$$\Pr[|X-\mu| > a] < 1.1\left(\frac{nk}{a^2}\right)^{k/2}.$$

Applying the above theorem with $n \leftarrow p, \mu = p/2, k \leftarrow 4, a \leftarrow p/4$ gives

$$\Pr[c_{i_1} + \cdots + c_{i_p} \leqslant 1] < 1.1\left(\frac{64}{p}\right)^2 = O\left(\frac{1}{p^2}\right). \qquad \square$$

Let $\alpha := \alpha(u)$ be the fraction of $v \in N(u)$ such that $l(v) \neq \emptyset$. For $1 \leqslant i \leqslant q$, let $b_i \in \{0,1\}$ be the random variable such that $b_i = 1$ if and only if $l(v_i) \neq \emptyset$. By pairwise independence of $\{v_1, \dots, v_q\}$, $\{b_1, \dots, b_q\}$ are also pairwise independent, and $\Pr[b_i = 1] = \alpha$ for each i. Let $\mathscr{B}_{u,q,\alpha}$ be the distribution on $b_1 + \cdots + b_q$. By (1), (2), and Claim 2.7, for fixed good u and q, $\Pr_e[|e \cap S| = 1 \mid u, q]$ is at most

$$Q\sqrt{\varepsilon} + \frac{Q^2}{D} + \sum_{p=1}^{q} \Pr_{X \sim \mathscr{B}_{u,q,\alpha}}[X = p] \cdot O\left(\frac{1}{p^2}\right). \tag{3}$$

We now consider the expected value of (3) over $q \sim \mathscr{D}$, with u still fixed.

Claim 2.9.

$$\mathop{\mathbb{E}}_{q\sim\mathscr{D}}\left[\sum_{p=1}^{q} \Pr_{X \sim \mathscr{B}_{u,q,\alpha}}[X = p] \cdot O\left(\frac{1}{p^2}\right)\right] = O\left(\frac{1}{\log Q}\right).$$

Proof. Since $\Pr_{X \sim \mathscr{B}_{u,q,\alpha}}[X = p] = 0$ when $p > q$,

$$\begin{aligned}
\mathop{\mathbb{E}}_{q\sim\mathscr{D}}\left[\sum_{p=1}^{q} \Pr_{X \sim \mathscr{B}_{u,q,\alpha}}[X = p] \cdot O\left(\frac{1}{p^2}\right)\right] &= \mathop{\mathbb{E}}_{q\sim\mathscr{D}}\left[\sum_{p=1}^{Q} \Pr_{X \sim \mathscr{B}_{u,q,\alpha}}[X = p] \cdot O\left(\frac{1}{p^2}\right)\right] \\
&= \sum_{p=1}^{Q} O\left(\frac{1}{p^2}\right) \mathop{\mathbb{E}}_{q\sim\mathscr{D}}\left[\Pr_{X \sim \mathscr{B}_{u,q,\alpha}}[X = p]\right].
\end{aligned}$$

Since $\sum_{p=1}^{Q} O\left(\frac{1}{p^2}\right) = O(1)$, it suffices to prove that for any $1 \leqslant p \leqslant Q$,

$$\mathop{\mathbb{E}}_{q\sim\mathscr{D}}\left[\Pr_{X \sim \mathscr{B}_{u,q,\alpha}}[X = p]\right] = O\left(\frac{1}{\log Q}\right). \tag{4}$$

We analyze it by considering how $\Pr[X = p]$ changes as q gets smaller or larger. For the lower tail where $q\alpha \leqslant \frac{p}{2}$, let y be the biggest integer such that $2^y \in [2, Q]$ and $\alpha 2^y \leqslant \frac{p}{2}$. For every $x = y, y-1, \dots, 1$, by Markov's inequality,

$$\Pr_{X \sim \mathscr{B}_{u,2^x,\alpha}}[X = p] \leqslant \Pr_{X \sim \mathscr{B}_{u,2^x,\alpha}}[X \geqslant p] \leqslant \frac{\alpha 2^x}{p}.$$

By our choice of y, when $x = y$, $\frac{\alpha 2^x}{p} \leqslant \frac{1}{2}$, and it decreases by a factor of 2 as we decrease x by 1. Therefore,

$$\mathop{\mathbb{E}}_{q}\left[\Pr_{X\sim\mathscr{B}_{u,q,\alpha}}[X = p]\,\middle|\, q\alpha \leqslant \frac{p}{2}\right]\Pr_{q}\left[q\alpha \leqslant \frac{p}{2}\right] \leqslant \frac{1}{\log Q}\sum_{x=y}^{1}\left(\frac{\alpha 2^x}{p}\right) \leqslant O\left(\frac{1}{\log Q}\right). \tag{5}$$

For the upper tail where $q\alpha \geqslant 2p$, let y be the smallest integer such that $2^y \in [2, Q]$ and $\alpha 2^y \geqslant 2p$. For every $x = y, y+1, \ldots, \log Q$, let X be a random variable sampled from $X \sim \mathscr{B}_{u,2^x,\alpha}$. By pairwise independence of $(b_1, \ldots, b_q)$, $\mathrm{Var}[X] \leqslant \mathbb{E}[X] = \alpha 2^x$. By Chebyshev's inequality,

$$\Pr_{X\sim\mathscr{B}_{u,2^x,\alpha}}[X = p] \leqslant \Pr_{X\sim\mathscr{B}_{u,2^x,\alpha}}[X \leqslant p] \leqslant \Pr_{X\sim\mathscr{B}_{u,2^x,\alpha}}[|X - \mathbb{E}[X]| \geqslant |\mathbb{E}[X] - p|] \leqslant \frac{\alpha 2^x}{(\alpha 2^x - p)^2}.$$

By our choice of y, for any $x \geqslant y$, $\frac{\alpha 2^x}{(\alpha 2^x - p)^2} \leqslant \frac{\alpha 2^x}{(\frac{\alpha 2^x}{2})^2} = \frac{4}{\alpha 2^x}$, and it is decreased by a factor of 2 as we increase x by 1. Therefore,

$$\mathop{\mathbb{E}}_{q}\left[\Pr_{X\sim\mathscr{B}_{u,q,\alpha}}[X = p]\,\middle|\, q\alpha \geqslant 2p\right]\Pr_{q}\left[q\alpha \geqslant 2p\right] \leqslant \frac{1}{\log Q}\sum_{x=y}^{\log Q}\frac{4}{\alpha 2^x} \leqslant O\left(\frac{1}{\log Q}\right). \tag{6}$$

Finally,

$$\Pr\left[\frac{p}{2} \leqslant q \leqslant 2p\right] \leqslant O\left(\frac{1}{\log Q}\right). \tag{7}$$

Equations (5), (6), (7) imply (4), which completes the proof of the claim. □

Therefore, for a good u, $\Pr_e[|e \cap S| = 1 \,|\, u]$ is at most

$$Q\sqrt{\varepsilon} + \frac{Q^2}{D} + O\left(\frac{1}{\log Q}\right),$$

and the overall probability $\Pr_e[|e \cap S| = 1]$ is at most

$$2Q\sqrt{\varepsilon} + \frac{Q^2}{D} + O\left(\frac{1}{\log Q}\right), \tag{8}$$

as desired in the Lemma. □

3 Main Results

We compose our reduction from Label Cover to Unique Coverage with the standard reduction from 3SAT to Label Cover. We restate Theorem 2.1 that shows the properties of the reduction from 3SAT to Label Cover.

Theorem 3.1 (Restatement of Theorem 2.1)**.** *There exists an absolute constant $\tau < 1$ such that the following is true. For any positive integer $r > 0$, there is a reduction that given an instance ϕ of 3SAT with n variables, outputs an instance of Label Cover $(G, \{\pi_e\}_e)$ with $|U_G|, |V_G| = n^{O(r)}$, $R = 10^r, L = 2^r, d = D = 5^r$ in time $n^{O(r)}$, and satisfies the following.*

- *Completeness: If ϕ is satisfiable, there exists a labeling that satisfies every projection.*
- *Soundness: If ϕ is not satisfiable, every labeling satisfies at most τ^r fraction of projections.*

Let $\gamma > 1$ be an absolute constant such that $\gamma\tau^{1/2} < \frac{1}{\gamma}$ and $\frac{\gamma^2}{5} < \frac{1}{\gamma}$, and for each r, let $Q = Q(r)$ be the largest power of 2 at most γ^r. We run our reduction given the Label Cover instance $(G, \{\pi_e\}_e)$ to produce an instance of Unique Coverage $H = (V_H, E_H)$. Recall that $|V_H| = |V_G| \cdot R = n^{O(r)}$, and $|E_H| = |U_G| \cdot \log Q \cdot D^2 \cdot 2L^2 = n^{O(r)}$, and the cardinality of each hyperedge is at most RQ.

If ϕ is satisfiable, $(G, \{\pi_e\}_e)$ admits a labeling that satisfies every constraint, so by Lemma 2.4, there exists $S \subseteq V_H$ such that the total weight of the hyperedges intersecting S in exactly one element is at least $\frac{1}{2}$. If ϕ is not satisfiable, every labeling of $(G, \{\pi_e\}_e)$ satisfies at most $\varepsilon = \tau^r$ fraction of projections, and by Lemma 2.5, for any $S \subseteq V_H$, the total weight of hyperedges intersecting S in exactly one element is at most

$$2Q\sqrt{\varepsilon} + \frac{Q^2}{D} + O\big(\frac{1}{\log Q}\big).$$

As r increases, (8) becomes

$$2Q\sqrt{\varepsilon} + \frac{Q^2}{D} + O\big(\frac{1}{\log Q}\big) \leqslant 2(\gamma\tau^{1/2})^r + \big(\frac{\gamma^2}{5}\big)^r + O\big(\frac{1}{\log Q}\big) \leqslant \frac{3}{Q} + O\big(\frac{1}{\log Q}\big) = O\big(\frac{1}{\log Q}\big) = O\big(\frac{1}{\log(RQ)}\big),$$

using the fact that $\log(RQ) = \Theta(\log Q)$.

3.1 1-in-k Hitting Set for Constant k

We set parameters to show inapproximability of 1-in-k for constant k, proving Theorem 1.1. Given a large constant k, take the largest r such that $\frac{k}{20} \leqslant RQ \leqslant k$ (R is always a power of 10 and Q is a power of 2). Since r is a constant, the combined reduction from 3SAT to 1-in-k HS runs in time polynomial in n. Therefore, if we approximate 1-in-k HS within a factor better than $O(\frac{1}{\log(RQ)}) = O(\frac{1}{\log k})$ in polynomial time in $|V_H|$, we can decide whether a given formula ϕ is satisfiable or not in time polynomial in n.

3.2 Unique Coverage

We set parameters to show inapproximability of Unique Coverage, proving Theorem 1.2. Given $\varepsilon > 0$, let $r = \log^{1/\varepsilon} n$. For some absolute constant $\alpha > 1$, the combined reduction from 3SAT in Unique Coverage runs in time $n^{\alpha \log^{1/\varepsilon} n} = 2^{\alpha \log^{1/\varepsilon+1} n}$, which is quasipolynomial in n. Note that $|V_H| \leqslant 2^{\alpha \log^{1/\varepsilon+1} n}$ and $RQ \geqslant 2^{\beta r} = 2^{\beta \log^{1/\varepsilon} n}$ for another absolute constant $\beta > 0$. Therefore,

$$\log RQ \geqslant \beta \log^{1/\varepsilon} n = \beta \cdot \alpha^{-\frac{1}{1+\varepsilon}} \cdot (\alpha \log^{1/\varepsilon+1} n)^{\frac{1}{1+\varepsilon}} = \Omega(\log^{\frac{1}{1+\varepsilon}} |V_H|),$$

so if we approximate Unique Coverage within a factor better than $O(\frac{1}{\log^{1-\varepsilon} |V_H|})$ in time polynomial in $|V_H|$, we can decide whether a given formula ϕ is satisfiable or not in quasipolynomial time in n.

References

[1] M. Bellare and J. Rompel. Randomness-efficient oblivious sampling. In *Proceedings of the 35th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '94, pages 276–287, Nov 1994. 6, 7

[2] P. Chalermsook, J. Chuzhoy, S. Kannan, and S. Khanna. Improved hardness results for profit maximization pricing problems with unlimited supply. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, volume 7408 of *Lecture Notes in Computer Science*, pages 73–84. 2012. 1

[3] E. D. Demaine, U. Feige, M. Hajiaghayi, and M. R. Salavatipour. Combination can be hard: Approximability of the unique coverage problem. *SIAM Journal on Computing*, 38(4):1464–1483, 2008. Preliminary version in SODA '06. 1

[4] U. Feige. A threshold of ln n for approximating set cover. *Journal of the ACM*, 45(4):634–652, July 1998. 2

[5] U. Feige. Relations between average case complexity and approximation complexity. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, STOC '02, pages 534–543, 2002. 1

[6] V. Guruswami and L. Trevisan. The complexity of making unique choices: Approximating 1-in-k Ssat. In *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*, pages 99–110. Springer, 2005. 1

[7] V. Guruswami and Y. Zhou. Tight bounds on the approximability of almost-satisfiable horn SAT and exact hitting set. *Theory of Computing*, 8(1):239–267, 2012. Preliminary version in SODA '11. 1

[8] S. Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the 34th annual ACM Symposium on Theory of Computing*, STOC '02, pages 767–775, 2002. 1

[9] S. Khot. Ruling out PTAS for graph min-bisection, dense k-subgraph, and bipartite clique. *SIAM Journal on Computing*, 36(4):1025–1071, 2006. 2

[10] M. Luby and A. Wigderson. Pairwise independence and derandomization. *Foundation and Trends in Theoretical Computer Science*, 1(4):237–301, Aug. 2006. 4

[11] C. Lund and M. Yannakakis. On the hardness of approximating minimization problems. *Journal of the ACM*, 41(5):960–981, 1994. 2

[12] P. Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *Proceedings of the 40th annual ACM symposium on Theory of computing*, STOC '08, pages 245–254, New York, NY, USA, 2008. ACM. 2

[13] C. Wenner. Circumventing d-to-1 for approximation resistance of satisfiable predicates strictly containing parity of width four. *Theory of Computing*, 9(23):703–757, 2013. 3

ECCC ISSN 1433-8092

http://eccc.hpi-web.de