How to Play Unique Games against a Semi-Random Adversary

Study of Semi-Random Models of Unique Games

Alexandra Kolla
Microsoft Research

Konstantin Makarychev
IBM Research

Yury Makarychev
TTIC

Abstract

In this paper, we study the average case complexity of the Unique Games problem. We propose a natural semi-random model, in which a unique game instance is generated in several steps. First an adversary selects a completely satisfiable instance of Unique Games, then she chooses an ε–fraction of all edges, and finally replaces ("corrupts") the constraints corresponding to these edges with new constraints. If all steps are adversarial, the adversary can obtain any $(1 - \varepsilon)$ satisfiable instance, so then the problem is as hard as in the worst case. In our semi-random model, one of the steps is random, and all other steps are adversarial. We show that known algorithms for unique games (in particular, all algorithms that use the standard SDP relaxation) fail to solve semi-random instances of Unique Games.

We present an algorithm that with high probability finds a solution satisfying a $(1 - \delta)$ fraction of all constraints in semi-random instances (we require that the average degree of the graph is $\tilde{\Omega}(\log k)$). To this end, we consider a new non-standard SDP program for Unique Games, which is not a relaxation for the problem, and show how to analyze it. We present a new rounding scheme that simultaneously uses SDP and LP solutions, which we believe is of independent interest.

Our result holds only for ε less than some absolute constant. We prove that if $\varepsilon \geq 1/2$, then the problem is hard in one of the models, that is, no polynomial–time algorithm can distinguish between the following two cases: (i) the instance is a $(1 - \varepsilon)$ satisfiable semi–random instance and (ii) the instance is at most δ satisfiable (for every $\delta > 0$); the result assumes the 2–to–2 conjecture.

Finally, we study semi-random instances of Unique Games that are *at most* $(1 - \varepsilon)$ satisfiable. We present an algorithm that with high probability, distinguishes between the case when the instance is a semi-random instance and the case when the instance is an (arbitrary) $(1 - \delta)$ satisfiable instance if $\varepsilon > c\delta$ (for some absolute constant c).

1 Introduction

In this paper, we study the average case complexity of the Unique Games problem in a semi-random model. In the Unique Games problem, we are given a graph $G = (V, E)$ (denote $n = |V|$), a set of labels $[k] = \{0, \ldots, k-1\}$ and a set of permutations π_{uv} on $[k]$, one permutation for every edge (u, v). Our goal is to assign a label (or state) $x_u \in [k]$ to every vertex u so as to maximize the number of satisfied constraints of the form $x_v = \pi_{uv}(x_u)$. The value of the solution is the number of satisfied constraints.

The problem is conjectured to be very hard in the worst case. The Unique Games Conjecture (UGC) of Khot [17] states that for every positive ε, δ and sufficiently large k, it is NP-hard to distinguish between the case where at least a $1 - \varepsilon$ fraction of constraints is satisfiable, and the case where at most a δ fraction of all constraints is satisfiable.

One reason which makes UGC particularly intriguing is its numerous implications. The conjecture, if true, implies that the currently best known approximation algorithms for many important computational problems have optimal approximation ratios. Indeed, since its origin, UGC has been successfully used to prove often optimal hardness of approximation results for several important NP-hard problems such as MAX CUT [18], Vertex Cover [19], Maximum Acyclic Subgraph [14], Max k-CSP [23, 15, 25], which are not known to follow from standard complexity assumptions.

Arguably, a seemingly strong reason for belief in UGC is the failure of several attempts to design efficient algorithms for Unique Games using current state-of-the-art techniques, even though a large amount of research activity in recent years has focused on the design of such algorithms. One direction of research has concentrated on developing polynomial-time approximation algorithms for arbitrary instances of unique games. The first algorithm was presented by Khot in his original paper on the Unique Games Conjecture [17], and then several algorithms were developed in papers by Trevisan [26], Gupta and Talwar [10], Charikar, Makarychev and Makarychev [8], Chlamtac, Makarychev, and Makarychev [9]. Another direction of research has been to study subexponential approximation algorithms for Unique Games. The work was initiated by Kolla [21] and Arora, Impagliazzo, Matthews and Steurer [4] who proposed subexponential algorithms for certain families of graphs. Then, in a recent paper, Arora, Barak and Steurer [3] gave a subexponential algorithm for arbitrary instances of Unique Games.

These papers, however, do not disprove the Unique Games Conjecture. Moreover, Khot and Vishnoi [20] showed that it is impossible to disprove the Conjecture by using the standard semidefinite programming relaxation for Unique Games, the technique used in the best currently known polynomial-time approximation algorithms for general instances of Unique Games. Additionally, Khot, Kindler, Mossel, and O'Donnell [18] proved that the approximation guarantees obtained in [8] cannot be improved if UGC is true (except possibly for lower order terms).

All that suggests that Unique Games is a very hard problem. Unlike many other problems, however, we do not know any specific families of hard instances of Unique Games. In contrast, we do know many specific hard instances of other problems. Many such instances come from cryptography; for example, it is hard to invert a one-way function f on a random input, it is hard to factor the product $z = xy$ of two large prime numbers x and y. Consequently, it is hard to satisfy SAT formulas that encode statements "$f(x) = y$" and "$xy = z$". There are even more natural families of hard instances of optimization problems; e.g.

- **3-SAT:** Feige's 3-SAT Conjecture [11] states that no randomized polynomial time algorithm can distinguish random instances of 3-SAT (with a certain clause to variable ratio) from $1 - \varepsilon$ satisfiable instances of 3-SAT (with non-negligible probability).

- **Linear Equations in $\mathbb{Z}/2\mathbb{Z}$:** Alekhnovich's Conjecture [1] implies that given a random $(1 - \varepsilon)$ satisfiable instance of a system of linear equations in $\mathbb{Z}/2\mathbb{Z}$, no randomized polynomial time algorithm

can find a solution that satisfies a $1/2 + \delta$ fraction of equations (for certain values of parameters ε and δ).

- **Maximum Clique Problem:** It is widely believed [16] that no randomized polynomial time algorithm can find a clique of size $(1 + \varepsilon) \log_2 n$ in a $G(n, 1/2)$ graph with a planted clique of size $m = n^{1/2-\delta}$ (for every constant $\varepsilon, \delta > 0$).

No such results are known or conjectured for Unique Games. In order to better understand Unique Games, we need to identify, which instances of the problem are easy and which are potentially hard. That motivated the study of specific families of Unique Games. Barak, Hardt, Haviv, Rao, Regev and Steurer [6] showed that unique game instances obtained by parallel repetition are "easy" (we say that a family of $1 - \varepsilon$ satisfiable instances is *easy* if there is a randomized polynomial-time algorithm that satisfies a constant fraction of constraints) . Arora, Khot, Kolla, Steurer, Tulsiani, and Vishnoi [5] showed that unique games on spectral expanders are easy (see also Makarychev and Makarychev [22], and Arora, Impagliazzo, Matthews and Steurer [4]).

In this paper, we investigate the hardness of semi-random (semi-adversarial) instances of Unique Games. In a semi-random model, an instance is generated in several steps; at each step, choices are either made adversarially or randomly. Semi-random models were introduced by Blum and Spencer [7] (who considered semi-random instances of the k-coloring problem) and then studied by Feige and Kilian [12], and Feige and Krauthgamer [13].

In this paper, we propose and study a model, in which a $1 - \varepsilon$ satisfiable unique game instance is generated as follows:

1. **Graph Selection Step.** Choose the constraint graph $G = (V, E)$ with n vertices and m edges.

2. **Initial Instance Selection Step.** Choose a set of constraints $\{\pi_{uv}\}_{(u,v)\in E}$ so that the obtained instance is completely satisfiable.

3. **Edge Selection Step.** Choose a set of edges E_ε of size $\varepsilon m = \varepsilon |E|$.

4. **Edge Corruption Step.** Replace the constraint for every edge in E_ε with a new constraint.

Note that if an adversary performs all four steps, she can obtain an arbitrary $1 - \varepsilon$ satisfiable instance, so, in this fully–adversarial case, the problem is as hard as in the worst case. The four most challenging semi-random cases are when choices at one out of the four steps are made randomly, and all other choices are made adversarially. The first case — when the graph G is random and, in particular, is an expander — was studied by Arora, Khot, Kolla, Steurer, Tulsiani, and Vishnoi [5], who showed that this case is easy.

We present algorithms for the other three cases that with high probability (w.h.p.) satisfy a $1 - \delta$ fraction of constraints (if the average degree of G is at least $\tilde{\Omega}(\log k)$ and ε is less than some absolute constant).

Theorem 1.1. *For every $k \geq k_0$, $\varepsilon_0 > 0$ and $\delta > C \max(\varepsilon_0, \log k/\sqrt{k})$ (where C and k_0 are absolute constants), there exists a randomized polynomial time algorithm that given a semi-random instance of Unique Games with $\varepsilon = \varepsilon_0$ (generated in one of the three models; see Section 2.3 for details) on a graph G with average degree at least $\tilde{\Omega}(\log k)\delta^{-3}$, finds a solution of value at least $1 - \delta$ with probability[1] $1 - o(1)$.*

---

[1]The probability is over both the random choices that we make when we generate the semi-random instance, and the random choices that the algorithm does. That is, the probability that the model generates a semi-random instance $\mathcal{I}$, such that the algorithm finds a solution of $\mathcal{I}$ of value at least $1 - \delta$ with probability $1 - o(1)$, is $1 - o(1)$.

The theorem follows from Theorems 3.1, 4.1, and 6.1, in which we analyze each model separately, and establish more precise bounds on the parameters for each model.

In our opinion, this is a very surprising result since the adversary has a lot of control over the semi-random instance. Moreover, our results suggest that the Unique Games problem is different in nature than several NP-hard problems like SAT, which are thought to be hard on average.

We want to point out that previously known approximation algorithms for Unique Games cannot find good solutions of semi-random instances. Also techniques developed for analyzing semi-random instances of other problems such as local analysis, statistical analysis, spectral gap methods, standard semidefinite programming techniques seem to be inadequate to deal with semi-random instances of Unique Games. To illustrate this point, consider the following example. Suppose that the set of corrupted edges is chosen at random, and all other steps are adversarial ("the random edges, adversarial constraints case"). The adversary generates a semi-random instance as follows. It first prepares two instances $\mathcal{I}_1$ and $\mathcal{I}_2$ of Unique Games. The first instance $\mathcal{I}_1$ is the Khot–Vishnoi instance [20] on a graph G with the label set $[k] = \{0, \ldots, k-1\}$ and permutations $\{\pi^1_{uv}\}$ whose SDP value is $\varepsilon' < \varepsilon/2$ but which is only $k^{-\Omega(\varepsilon')}$ satisfiable. The second instance $\mathcal{I}_2$ is a completely satisfiable instance on the same graph G with the label set $\{k, \ldots, 2k-1\}$ and permutations $\pi^2_{uv} = \mathrm{id}$. She combines these instances together: the combined instance is an instance on the graph G with the label set $[2k] = \{0, \ldots, 2k-1\}$, and permutations $\{\pi_{uv} : \pi_{uv}(i) = \pi^1_{uv}(i)$ if $i \in [k]$, and $\pi_{uv}(i) = \pi^2_{uv}(i)$, otherwise$\}$. Once the adversary is given a random set of edges E_ε, she randomly changes ("corrupts") permutations $\{\pi^2_{uv}\}_{(u,v) \in E_\varepsilon}$ but does not change π^1_{uv}, and then updates permutations $\{\pi_{uv}\}_{(u,v) \in E_\varepsilon}$ accordingly. It turns out that the SDP value of $\mathcal{I}_2$ with corrupted edges is very close to ε, and therefore, it is larger than ε', the SDP value of $\mathcal{I}_1$ (if we choose parameters properly). So in this case the SDP solution assigns positive weight only to the labels in $[k]$ from the first instance. That means that the SDP solution does not reveal any information about the optimal solution (the only integral solution we can obtain from the SDP solution has value $k^{-\Omega(\varepsilon)}$). Similarly, algorithms that analyze the spectral gap of the label extended graph cannot deal with this instance. Of course, in this example, we let our first instance, $\mathcal{I}_1$, to be the Khot–Vishnoi instance because it "cheats" SDP based algorithms. Similarly, we can take as $\mathcal{I}_1$ another instance that cheats another type of algorithms. For instance, if UGC is true, we can let $\mathcal{I}_1$ to be a $1 - \varepsilon'$ satisfiable unique game that is indistinguishable in polynomial-time from a δ–satisfiable unique game.

Our algorithms work only for values of ε less than some absolute constants. We show that this restriction is essential. For every $\varepsilon \geq 1/2$ and $\delta > 0$, we prove that no polynomial time algorithm satisfies a δ fraction of constraints in the "adversarial constraints, random edges" model (only the third step is random), assuming the 2–to–2 conjecture.

One particularly interesting family of semi-random unique games (captured by our model) are *mixed instances*. In this model, the adversary prepares a satisfiable instance, and then chooses a δ fraction of edges and replaces them with adversarial constraints (corrupted constraints); i.e. she performs all four steps in our model and can obtain an arbitrary $1 - \delta$ satisfiable instance. Then the "nature" replaces every corrupted constraint with the original constraint with probability $1 - \varepsilon$. In our model, this case corresponds to an adversary who at first prepares a list of corrupted constraints π'_{uv}, and then at the fourth step replaces constraints for edges in E_ε with constraints π'_{uv} (if an edge from E_ε is not in the list, the adversary does not modify the corresponding constraint).

Distinguishing Semi-Random At Most $(1 - \varepsilon)$ Satisfiable Instances From Almost Satisfiable Instances. We also study whether semi-random instances of Unique Games that are at most $(1 - \varepsilon)$ satisfiable can be distinguished from almost satisfiable instances of Unique Games. This question was studied for other problems. In particular, Feige's "Random 3-SAT Conjecture" states that it is impossible to distinguish be-

tween random instances of 3-SAT (with high enough clause density) and $1-\delta$ satisfiable instances of 3-SAT. In contrast, we show that in the "adversarial edges, random constraints" case (the fourth step is random), semi-random $(1-\varepsilon)$-satisfiable instances can be efficiently distinguished from (arbitrary) $(1-\delta)$-satisfiable instances when $\varepsilon > c\delta$ (for some absolute constant c). (This problem, however, is meaningless in the other two cases — when the adversary corrupts the constraints — since then she can make the instance almost satisfiable even if ε is large.)

Linear Unique Games We separately consider the case of Linear Unique Games (MAX Γ-LIN). In the semi-random model for Linear Unique Games, we require that constraints chosen at the second and fourth steps are of the form $x_u - x_v = s_{uv}(\text{mod } k)$. Note that in the "random edges, adversarial constraints" model, the condition that constraints are of the form $x_u - x_v = s_{uv}(\text{mod } k)$ only restricts the adversary (and does not change how the random edges are selected). Therefore, our algorithms for semi-random general instances still applies to this case. However, in the "adversarial edges, random constraints" case, we need to sample constraints from a different distribution of permutations at the fourth step: for every edge (u, v) we now choose a random shift permutation $x_v = x_u - s_{uv}$, where $s_{uv} \in_U \mathbb{Z}/k\mathbb{Z}$. We show that our algorithm still works in this case; the analysis however is different. We believe that it is of independent interest. We do not consider the case where only the initial satisfying assignment is chosen at random, since for Linear Unique Games, the initial assignment uniquely determines the constraints between edges (specifically, $s_{uv} = x_u - x_v(\text{mod } k)$). Thus the case when only the second step is random is completely adversarial.

It is interesting that systems of linear equations affected by noise with more than two variables per equations are believed to be much harder. Suppose we have a consistent system of linear equations $Ax = b$ over $\mathbb{Z}/2\mathbb{Z}$. Then we randomly change an ε fraction of entries of b. Alekhnovich [1] conjectured that no polynomial-time algorithm can distinguish the obtained instance from a completely random instance even if $\varepsilon \approx n^{-c}$, for some constant c (Alekhnovich stated his conjecture both for systems with 3 variables per equation and for systems with an arbitrary number of variables per equation).

Our results can be easily generalized to Unique Games in arbitrary Abelian groups. We omit the details in the conference version of this paper.

1.1 Brief Overview of Techniques

In this paper, we develop new powerful algorithmic techniques for solving semi-random instances of unique games. We use different algorithms for different models. First, we outline how we solve semi-random unique games in the "adversarial constraints, random edges" model (see Section 3 for details). As we explained above, we cannot use the standard SDP relaxation (or other standard techniques) to solve semi-random instances in this model. Instead, we consider a very unusual SDP program for the problem, which we call "Crude SDP" (C-SDP). This SDP is not even a relaxation for Unique Games and its value can be large when the instance is satisfiable. The C-SDP assigns a unit vector u_i to every vertex (u, i) of the label–extended graph (for a description of the label–extended graph we refer the reader to Section 2). We use vectors u_i to define the length of edges of the label–extended graph: the length of $((u, i), (v, j))$ equals $\|u_i - v_j\|^2$. Then we find *super short edges* w.r.t. the C-SDP solution, those edges that have length $O(1/\log k)$. One may expect that there are very few short edges since for a given C-SDP most edges will be long if we choose the unique games instance at random. We prove, however, that for every C-SDP solution $\{u_i\}$, with high probability (over the semi-random instance) either

1. there are many super short edges w.r.t. $\{u_i\}$ in the satisfiable layer of the semi-random game,

2. or there is another C-SDP solution of value less than the value of the solution $\{u_i\}$.

Here, as we describe later on in section 2, the "satisfiable layer" corresponds to the representation of the satisfying assignment in the label–extended graph.

Note that if our instance is completely satisfiable, then in the optimal (integral) C-SDP solution all the edges that correspond to the satisfiable layer have length zero and, therefore are super short.

Our proof shows how to combine the C-SDP solution with an integral solution so that the C-SDP value goes down unless almost all edges in the satisfiable layer are super short. We then show that this claim holds with high probability not only for one C-SDP solution but also for all C-SDP solutions simultaneously. The idea behind this step is to find a family $\mathcal{F}$ of representative C-SDP solutions and then use the union bound. One of the challenges is to choose a very small family $\mathcal{F}$, so that we can prove our result under the assumption that the average degree is only $\tilde{\Omega}(\log k)$. The result implies that w.h.p. there are many super short edges w.r.t. the optimal C-SDP solution.

Now given the set of super short edges, we need to find which of them lie in the satisfiable layer. We write and solve an LP relaxation for Unique Games, whose objective function depends on the set of super short edges. Then we run a rounding algorithm that rounds the C-SDP and LP solutions to a combinatorial solution using a variant of the "orthogonal separators" technique developed in [9].

Our algorithm for the "adversarial edges, random constraints" model is quite different. First, we solve the standard SDP relaxation for Unique Games. Now, however, we cannot claim that many edges of the label–extended graph are short. We instead find the "long" edges of the graph G w.r.t. the SDP solution. We prove that most corrupted edges are long, and there are at most $O(\varepsilon)$ long edges in total (Theorem 4.4). We remove all long edges and obtain a highly–satisfiable instance. Then we write and solve an SDP for this instance, and round the SDP solution using the algorithm of [8].

We also present algorithms for two more cases: the case of "adversarial edges, random constraints" where the constraints are of the special form MAX-Γ-LIN and the case of "random initial constraints".

In this paper, we develop several new techniques. In particular, we propose a novel C-SDP program, and then show how to exploit the optimality of a C-SDP solution in the analysis. We develop a rounding algorithm that simultaneously uses SDP and LP solutions. We demonstrate how to bound the number of different SDP solutions using dimension reduction methods and other tricks. We believe that our techniques are of independent interest and that they will prove useful for solving semi-random instances of other problems.

2 Notation and Preliminaries

2.1 The Label-Extended Graph

For a given instance of Unique Games on a constraint graph $G = (V, E)$, with alphabet size k and constraints $\{\pi_{uv}\}_{(u,v) \in E}$ we define the *Label-Extended* graph $M(V' = V \times [k], E')$ associated with that instance as follows: M has k vertices $B_v = \{v_0, \cdots, v_{k-1}\}$ for every vertex $v \in V$. We refer to this set of vertices as the block corresponding to v. M has a total of $|V|$ blocks, one for each vertex of G. Two vertices $u_i, v_j \in V'$ are connected by an edge if $(u, v) \in E$ and $\pi_{uv}(i) = j$. We refer to a set of nodes $L = \{u^{(z)}{}_{i(z)}\}_{z=1}^{|V|}$ as a "layer" if L contains exactly one node from each block $B_{u^{(z)}}$. We note that a layer L can be seen as an assignment of labels to each vertex of G. If a layer L consists of vertices with the same index i, i.e. $L = \{u_i^{(z)}\}_{z=1}^{|V|}$, we will call L the i-th layer.

2.2 Standard SDP for Unique Games

Our algorithms use the following standard SDP relaxation for Unique Games (see also [17, 20, 8, 9]).

$$\min \qquad \frac{1}{2|E|} \sum_{(u,v)\in E} \sum_{i\in[k]} \|u_i - v_{\pi_{uv}(i)}\|^2$$

$$\text{subject to} \qquad \sum_{i=1}^{k} \|u_i\|^2 = 1 \qquad\qquad\qquad \text{for all } u \in V$$

$$\langle u_i, u_j \rangle = 0 \qquad\qquad\qquad \text{for all } u \in V,\ i,j \in [k](i \neq j)$$

$$\langle u_i, v_j \rangle \geq 0 \qquad\qquad\qquad \text{for all } u,v \in V,\ i,j \in [k](i \neq j)$$

$$\|u_i - v_j\|^2 \leq \|u_i - w_l\|^2 + \|w_l - v_j\|^2 \qquad \text{for all } u,v,w \in V,\ i,j,l \in [k].$$

In this relaxation, we have a vector variable u_i for every vertex u and label i. In the intended solution, u_i is an indicator variable for the event "$x_u = i$". That is, if $x_u = i$ then $u_i = e$; otherwise, and $u_i = 0$; where e is a fixed unit vector. The objective function measures the fraction of *unsatisfied* constraints: if the unique game has value $1 - \varepsilon$, then the value of the intended SDP solution equals ε (and, therefore, the value of the optimal SDP solution is at most ε).

Given an SDP solution of value ε, the approximation algorithm of Charikar, Makarychev, and Makarychev [8] finds a solution of value $1 - O(\sqrt{\varepsilon \log k})$. We will use this approximation algorithm as a subroutine (we will refer to it as CMMa). We will also use the following fact.

Lemma 2.1 (see e.g. Lemmas A.1 and A.2 in [9])**.** *Suppose, we are given two random Gaussian variables γ_1 and γ_2 with mean 0 and variance 1 (not necessarily independent), and a parameter $k \geq 2$. Let $\alpha = 1/(2k^2)$. Consider a threshold t s.t. $\Pr(\gamma_1 \geq t) = \Pr(\gamma_2 \geq t) = \alpha$. Then $\Pr(\gamma_1 \geq t$ and $\gamma_2 \geq t) \geq \alpha\left(1 - \sqrt{\frac{1}{c^*} \operatorname{Var}(\gamma_1 - \gamma_2) \log k}\right)$ for some absolute constant c^*.*

2.3 Models

In what follows, we will use several models for generating semi-random $(1 - \varepsilon)$ satisfiable instances of Unique Games[2]:

1. **"Random Edges, Adversarial Constraints" Model.** The adversary selects a graph $G(V, E)$ on n vertices and m edges and an initial set of constraints $\{\pi_{uv}\}_{(u,v)\in E}$ so that the instance is completely satisfiable. Then she adds every edge of E to a set E_ε with probability ε (the choices for different edges are independent). Finally, the adversary replaces the constraint for every edge in E_ε with a new constraint of her choice. Note that this model also captures the case where at the last step the constraints for every edge in E_ε are replaced with a new random constraint (random adversary).

2. **"Adversarial Edges, Random Constraints" Model.** The adversary selects a graph $G(V, E)$ on n vertices and m edges and an initial set of constraints $\{\pi_{uv}\}_{(u,v)\in E}$ so that the instance is completely satisfiable. Then she chooses a set E_ε of $\varepsilon|E|$ edges. Finally, the constraint for every edge in E_ε is randomly replaced with a new constraint. We will also consider some variations of this model, where at all steps the constraints are MAX Γ-LIN. In particular, at the last step, choosing a random constraint of the form MAX Γ-LIN, corresponds to choosing a random value $s \in [|\Gamma|]$.

---

[2]The parameters of the models are the number of vertices n, the number of edges m, and the probability ε that an edge is corrupted.

3. **"Random Initial Constraints" Model.** The adversary chooses the constraint graph $G = (V, E)$ and a "planted solution" $\{x_u\}$. Then for every edge $(u, v) \in E$, she randomly chooses a permutation (constraint) π_{uv} such that $\pi_{uv}(x_u) = x_v$ (among $(k-1)!$ possible permutations). Then the adversary chooses an arbitrary set E_ε of edges of size at most $\varepsilon|E|$ and replaces constraint π_{uv} with a constraint π'_{uv} of her choice for $(u, v) \in E_\varepsilon$.

Remark 2.2. *Without loss of generality, we will assume, when we analyze the algorithms, that the initial completely satisfying assignment corresponds to the "zero" layer. I.e. for every edge (u, v), $\pi_{uv}(0) = 0$. Note that in reality, the real satisfying assignment is hidden from us.*

3 Random Edges, Adversarial Constraints

In this section, we study the "random edges, adversarial constraints" model and prove the following result.

Theorem 3.1. *Let $k \in \mathbb{N}$ ($k \geq 2$), $\varepsilon \in (0, 1/3)$, and $\eta \in (0, 1)$. There exists a polynomial-time approximation algorithm, that given an instance of Unique Games from the "random edges, adversarial constraints" model on graph G with $C\eta^{-3}(1/3 - \varepsilon)^{-4}n \log k(\log(\eta^{-1} \log k))^2$ edges (C is a sufficiently large absolute constant), finds a solution of value $(1 - \varepsilon - \eta)/(1 + \varepsilon + \eta) + O(1/k)$, with probability $1 - o(1)$.*

Corollary 3.2. *There exists a polynomial-time approximation algorithm, that given an instance of unique games from the "random edges, adversarial constraints" model on graph G with $Cn \log k(\log \log k)^2$ edges (C is a sufficiently large absolute constant) and $\varepsilon \leq 1/4$, finds a solution of value $1/2$, with probability $1 - o(1)$.*

Remark 3.3. *We can make the constant in the $O(1/k)$ term in $(1 - \varepsilon - \eta)/(1 + \varepsilon + \eta) + O(1/k)$ arbitrarily small by increasing the value of C (and decreasing the value of α in the proof). We omit the details to simplify the exposition.*

The main challenge in solving "random edges, adversarial constraints" unique games is that the standard SDP relaxation may assign zero vectors to layers corresponding to the optimal solution (as well as to some other layers) and assign non-zero vectors to layers, where every integral solution satisfies very few constraints. To address this issue, we introduce a new slightly modified SDP. As usual the SDP has a vector u_i for every vertex–label pair $(u, i) \in V \times [k]$. We require that vectors u_i, $u_{i'}$ corresponding to the same vertex u are orthogonal: $\langle u_i, u_{i'} \rangle = 0$ for all $u \in V$ and $i, i' \in [k]$, $i \neq i'$. We also impose triangle inequality constraints:

$$\frac{1}{2}\|u_i - v_j\|^2 + \frac{1}{2}\|u_{i'} - v_j\|^2 \geq 1,$$

for all $(u, v) \in E$ and $i, i', j \in [k]$, $i \neq i'$; and require that all vectors have unit length: $\|u_i\| = 1$ for all $u \in V$ and $i \in [k]$. Observe, that our SDP is not a relaxation[3], since the integer solution does not satisfy the last constraint. The objective function is

$$\min \sum_{\substack{(u,v) \in E}} \sum_{\substack{i \in [k] \\ j = \pi_{uv}(i)}} \frac{\|u_i - v_j\|^2}{2}.$$

Usually, this objective function measures the number of unsatisfied unique games constraints. However, in our case it does not. In fact, it does not measure any meaningful quantity. Note, that the value of the SDP

---

[3]Unless, the unique game is from a special family like Linear Unique Games (see Section 5).

can be arbitrary large even if the unique games instance is satisfiable. We call this SDP—the Crude SDP or C-SDP. Given a C-SDP solution, we define the set of *super short* edges, which play the central role in our algorithm.

Definition 3.4. *We say that an edge $((u,i),(v,j))$ in the label–extended graph is η–super short, if $\|u_i - v_j\|^2 \leq c^*\eta^2/\log k$, here c^* is an absolute constant defined in Lemma 2.1. We denote the set of all η–super short edges by Γ_η.*

In Section 3.1, we prove the following surprising result (Theorem 3.5), which states that all but very few edges in the zero level of the label-extended graph are super short.

Theorem 3.5. *Let $k \in \mathbb{N}$ ($k \geq 2$), $c \in (0,1)$, $\varepsilon \in (0,1/3)$, $\eta \in (c,1)$ and $\gamma \in (\varepsilon + c, 1)$ and let $G = (V,E)$ be an arbitrary graph with at least $C\eta^{-2}(\gamma - \varepsilon)^{-1}(1/3 - \varepsilon)^{-4} \times n \log k (\log(c^{-1} \log k))^2$, edges. Consider a semi-random instance of Unique Games in the "random edges, adversarial constraints" model. Let $\{u_i\}$ be the optimal solution of the C-SDP. Then with probability $1 - o(1)$, the set*

$$\Gamma_\eta^0 = \Gamma_\eta \cap \{((u,0),(v,0)) : (u,v) \in E\}$$

contains at least $(1 - \gamma)|E|$ edges.

More concretely, we proceed as follows. First, we solve the C-SDP. Then, given the C-SDP solution, we write and solve an LP to obtain weights $x(u,i) \in [0,1]$ for every $(u,i) \in V \times [k]$. These weights are in some sense substitutes for lengths of vectors in the standard SDP relaxation. In the LP, for every vertex $u \in V$, we require that

$$\sum_{i \in [k]} x(u,i) = 1.$$

The objective function is

$$\max \sum_{((u,i),(v,j)) \in \Gamma_\eta} \min(x(u,i), x(v,j))$$

(note that the objective function depends on the C-SDP solution). Denote the value of the LP by LP. The intended solution of this LP is $x(u,0) = 1$ and $x(u,i) = 0$ for $i \neq 0$. Since the LP contribution of every edge in Γ_η^0 is 1, the value of the intended solution is at least $|\Gamma_\eta^0|$. Applying Theorem 3.5 with $\gamma = \varepsilon + \eta$, we get $|\Gamma_\eta^0| \geq (1 - \gamma)|E| = (1 - \varepsilon - \eta)|E|$, so $LP \geq (1 - \varepsilon - \eta)|E|$. In the next section, we present an approximation algorithm (which rounds C-SDP an LP solutions) and its analysis. We prove the approximation guarantee in Lemma 3.9, which implies Theorem 3.1.

3.1 Lower Bound on the Number of Super Short Edges: Proof of Theorem 3.5

We need the following lemma.

Lemma 3.6. *Let $G = (V,E)$ be an arbitrary graph on n vertices, and let $\varepsilon \in [0,1/3)$, $\rho = 1/3 - \varepsilon$, $\nu \in (0,\rho)$. Suppose, that $\{Z_{uv}\}_{(u,v) \in E}$ are i.i.d. Bernoulli random variables taking values 1 with probability ε and 0 with probability $(1 - \varepsilon)$. Define the payoff function $p : \{0,1\} \times \mathbb{R} \to \mathbb{R}$ as follows*

$$p(z,\alpha) = \begin{cases} -2\alpha, & \text{if } z = 1; \\ \alpha, & \text{if } z = 0. \end{cases}$$

Then, with probability at least $1 - o(1)$ for every set of vectors $\{u_0\}_{u \in V}$ satisfying (for some significantly large absolute constant C)

$$\frac{1}{2} \sum_{(u,v) \in E} \|u_0 - v_0\|^2 \geq C\nu|E| + C\rho^{-4} \log^2(1/\nu)n \tag{1}$$

the following inequality holds

$$\sum_{(u,v) \in E} p(Z_{uv}, \|u_0 - v_0\|^2) > 0. \tag{2}$$

Proof. We need the following dimension reduction lemma, which is based on the Johnson–Lindenstrauss Lemma and is fairly standard (for an example of using the Johnson–Lindenstrauss Lemma in SDP rounding see Raghavendra and Steurer [24]).

Lemma 3.7. *For every positive ζ, η and ν, there exists a set N of unit vectors of size at most*

$$\exp\left(O(\zeta^{-2} \log(1/\eta) \log(1/\nu))\right)$$

such that for every set of unit vectors $\mathcal{U}$ there exists a randomized mapping $\varphi : \mathcal{U} \to N$ satisfying the following property: for every $u, v \in \mathcal{U}$,

$$\Pr((1 + \zeta)^{-1}\|u - v\|^2 - \eta^2 \leq \|\varphi(u) - \varphi(v)\|^2 \leq (1 + \zeta)\|u - v\|^2 + \eta^2) \geq 1 - \nu. \tag{3}$$

Proof sketch. To prove the lemma, we consider an $O(\zeta^{-2} \log(1/\nu))$ dimensional space L and choose a $\eta^2/32$-net N in it. The size of N satisfies the bound in the lemma. To construct the mapping $\varphi : \mathcal{U} \to N$, we project all vectors from $\mathcal{U}$ to L using the Johnson–Lindenstrauss lemma and then define $\varphi(u)$ to be the closest vector $u^* \in N$ to the projection of u. $\qquad\square$

Proof of Lemma 3.6. Set $\zeta = \rho/5 \equiv (1/3 - \varepsilon)/5$. Define a new payoff function,

$$p_\zeta(z, \alpha) = \begin{cases} -2(1 + \zeta)\alpha, & \text{if } z = 1; \\ (1 + \zeta)^{-1}\alpha, & \text{if } z = 0. \end{cases}$$

Note, that $\zeta < 1/15$ and $p_\zeta(0, \alpha) - p_\zeta(1, \alpha) = ((1+\zeta)^{-1} - (-2(1+\zeta)))\alpha \leq 4\alpha$ (for $\alpha \geq 0$). Suppose, that for a given realization $\{Z_{uv}^*\}_{(u,v) \in E}$ of $\{Z_{uv}\}_{(u,v) \in E}$ there exists a set of unit vectors $\{u_0\}_{u \in V}$ satisfying condition (1) and violating (2). Embed vectors $\{u_0\}_{u \in V}$ into a net N of size $\exp\left(O(\rho^{-2} \log^2(1/\nu))\right)$ using Lemma 3.7, such that for $(1 - \nu^2/2)$ fraction of all edges $(u, v) \in E$, the following condition holds

$$(1 + \zeta)^{-1}\|u_0 - v_0\|^2 - \nu^2/8 \leq \|u^* - v^*\|^2 \leq (1 + \zeta)\|u_0 - v_0\|^2 + \nu^2/4,$$

here u^* is the image of u_0; v^* is the image of v_0. Then,

$$\sum_{(u,v) \in E} p_\zeta(Z_{uv}^*, \|u^* - v^*\|^2) < \nu^2|E| < \nu\rho|E| = 5\nu\zeta|E|. \tag{4}$$

and (since we choose C to be sufficiently large)

$$\sum_{(u,v) \in E} \|u^* - v^*\|^2 \geq 40\nu|E| + C\rho^{-4} \log^2(1/\nu)n. \tag{5}$$

9

Thus, the existence of vectors $\{u_0\}_{u \in V}$ satisfying condition (1) and violating (2) implies the existence of vectors $\{u^*\}_{u \in V}$ satisfying (4) and (5). We now show that for a random $\{Z_{uv}\}$ such vectors $\{u^*\}$ exist with exponentially small probability. Fix a set $\{u^*\}_{u \in V} \subset N$ and consider random $\{Z_{uv}\}$,

$$
\begin{aligned}
\mathbb{E} \sum_{(u,v) \in E} p_\zeta(Z_{uv}, \|u^* - v^*\|^2) &\geq \sum_{(u,v) \in E} ((1-\varepsilon)(1-\zeta) - 2(1+\zeta)\varepsilon) \|u^* - v^*\|^2 \\
&= (1 - 3\varepsilon - 3\varepsilon\zeta - \zeta) \sum_{(u,v) \in E} \|u^* - v^*\|^2 \\
&\geq (1 - 3\varepsilon - 4\zeta) \sum_{(u,v) \in E} \|u^* - v^*\|^2 \\
&\geq \zeta \sum_{(u,v) \in E} \|u^* - v^*\|^2.
\end{aligned}
$$

By Hoeffding's inequality (using that $p_\zeta(0, \|u^* - v^*\|^2) - p_\zeta(1, \|u^* - v^*\|^2) \leq 4\|u^* - v^*\|^2$, $\sum_{(u,v) \in E} \|u^* - v^*\|^2 - 10\nu|E| \geq \frac{3}{4} \sum_{(u,v) \in E} \|u^* - v^*\|^2$ and $\|u^* - v^*\|^2 \leq 4$),

$$
\begin{aligned}
\Pr\left(\sum_{(u,v) \in E} p_\zeta(Z_{uv}, \|u^* - v^*\|^2) < 2\nu\rho|E| \right) &\leq \exp\left(-\frac{2\zeta^2 \left(\sum_{(u,v) \in E} \|u^* - v^*\|^2 - 10\nu|E| \right)^2}{\sum_{(u,v) \in E} 16\|u^* - v^*\|^4} \right) \\
&\leq \exp\left(-\frac{\zeta^2 \left(\sum_{(u,v) \in E} \|u^* - v^*\|^2 \right)^2}{\sum_{(u,v) \in E} 64\|u^* - v^*\|^2} \right) \\
&\leq \exp\left(-\frac{\zeta^2}{64} \sum_{(u,v) \in E} \|u^* - v^*\|^2 \right) \\
&\leq \exp\left(-\frac{C\zeta^2}{64} \rho^{-4} \log^2(1/\nu) n \right) \\
&= \exp\left(-\frac{C}{25 \cdot 64} \rho^{-2} \log^2(1/\nu) n \right)
\end{aligned}
$$

The number of all possible subsets $\{u^*\}_{u \in V} \subset N$ is at most

$$
|N|^n \leq \exp\left(O(n\rho^{-2} \log^2(1/\nu)) \right).
$$

Hence, by the union bound with probability at least $1 - \exp(-n) = 1 - o(1)$ for random $\{Z_{uv}\}_{(u,v) \in E}$, there does not exist a set of vectors $\{u_0\}_{u \in V}$ satisfying condition (1) and violating (2). $\qquad\square$

$\hfill\square$

Proof of Theorem 3.5. Let $\{u_i^*\}$ be the optimal SDP solution. Pick a unit vector e orthogonal to all vectors u_i^*. Define a new SDP solution $u_0^{int} = e$ and $u_i^{int} = u_i^*$ for $i \neq 0$ (for all $u \in V$). Note that restricted to $\{u_0^{int}\}_{u \in V}$ this solution is integral. Since $\{u_i^*\}$ is the optimal solution,

$$
\sum_{\substack{(u,v) \in E}} \sum_{\substack{i \in [k] \\ j = \pi_{uv}(i)}} \|u_i^* - v_j^*\|^2 \leq \sum_{\substack{(u,v) \in E}} \sum_{\substack{i \in [k] \\ j = \pi_{uv}(i)}} \|u_i^{int} - v_j^{int}\|^2.
$$

10

Denote by E_ε the set of corrupted edges. Let $Z_{uv} = 1$, if $(u, v) \in E_\varepsilon$ and $Z_{uv} = 0$, otherwise. Let $\tilde{E}_\varepsilon = \{(u, v) \in E : \pi_{uv}(0) \neq 0\}$. Clearly, $\tilde{E}_\varepsilon \subset E_\varepsilon$. Write,

$$\sum_{\substack{(u,v)\in E}} \sum_{\substack{i\in[k]\\j=\pi_{uv}(i)}} \|u_i^* - v_j^*\|^2 \quad - \quad \|u_i^{int} - v_j^{int}\|^2 =$$

$$= \sum_{(u,v)\in E\setminus\tilde{E}_\varepsilon} \|u_0^* - v_0^*\|^2 + \sum_{(u,v)\in\tilde{E}_\varepsilon} \|u_0^* - v_{\pi_{uv}(0)}^*\|^2 + \|u_{\pi_{vu}(0)}^* - v_0^*\|^2$$

$$- \sum_{(u,v)\in\tilde{E}_\varepsilon} \|u_0^{int} - v_{\pi_{uv}(0)}^{int}\|^2 + \|u_{\pi_{vu}(0)}^{int} - v_0^{int}\|^2.$$

For $(u, v) \in \tilde{E}_\varepsilon$, we have $\|u_0^{int} - v_{\pi_{uv}(0)}^{int}\|^2 = \|u_{\pi_{vu}(0)}^{int} - v_0^{int}\|^2 = 2$ and $\|u_0^* - v_{\pi_{uv}(0)}^*\|^2 \geq 2 - \|u_0^* - v_0^*\|^2$. Thus,

$$\sum_{\substack{(u,v)\in E}} \sum_{\substack{i\in[k]\\j=\pi_{uv}(i)}} \|u_i^* - v_j^*\|^2 - \|u_i^{int} - v_j^{int}\|^2 \geq \sum_{(u,v)\in E\setminus\tilde{E}_\varepsilon} \|u_0^* - v_0^*\|^2 - 2\sum_{(u,v)\in\tilde{E}_\varepsilon} \|u_0^* - v_0^*\|^2$$

$$\geq \sum_{(u,v)\in E\setminus E_\varepsilon} \|u_0^* - v_0^*\|^2 - 2\sum_{(u,v)\in E_\varepsilon} \|u_0^* - v_0^*\|^2$$

$$= \sum_{(u,v)\in E} p(Z_{uv}, \|u_0^* - v_0^*\|^2),$$

where $p(\cdot, \cdot)$ is the function from Lemma 3.6. By Lemma 3.6, with probability $1 - o(1)$, for some absolute constant C', $\rho = (1/3 - \varepsilon)$, $\nu = \frac{c^* \eta^2 (\gamma - \varepsilon)}{2C' \log k}$,

$$\frac{1}{2|E|} \sum_{(u,v)\in E} \|u_0^* - v_0^*\|^2 < \frac{C'\nu|E| + C'\rho^{-4}\log^2(1/\nu)n}{|E|} \leq c^*\eta^2(\gamma - \varepsilon)\log^{-1}k.$$

The last inequality holds, because

$$|E| \geq C\eta^{-2}(\gamma - \varepsilon)^{-1}\rho^{-4}n\log k(\log(c^{-1}\log k))^2 \geq \frac{C'}{2c^*}\eta^{-2}(\gamma - \varepsilon)^{-1}\rho^{-4}n\log k\log^2(1/\nu)^2.$$

By the Markov inequality, for all but $(\gamma - \varepsilon)/2$ fraction of edges $(u, v) \in E$, $\|u_0^* - v_0^*\|^2 \leq c^*\eta^2/\log k$. Recall that a pair $((u, 0), (v, 0))$ is an η–super short edge in the label–extended graph if $\pi_{uv}(0) = 0$ (i.e., (u_0, v_0) is an edge of the label–extended graph) and $\|u_0^* - v_0^*\|^2 \leq c^*\eta^2/\log k$. By the Chernoff Bound, $|E_\varepsilon| \leq (\varepsilon + (\gamma - \varepsilon)/2)|E|$ with probability $(1 - o(1))$; therefore, $\pi_{uv}(0) \neq 0$ for at most a $\varepsilon + (\gamma - \varepsilon)/2$ fraction of edges. Thus, with probability $(1 - o(1))$, there are at least $(1 - (\gamma - \varepsilon)/2 - (\varepsilon + (\gamma - \varepsilon)/2))m = (1 - \gamma)|E|$ η–super short edges. $\qquad\square$

3.2 SDP and LP Rounding

We now present an algorithm that given a C-SDP solution $\{u_i\}$ and an LP solution $\{x(u, i)\}$ finds an integer solution. We first present a procedure for sampling subsets of vertex–label pairs, which is an analog of the algorithm for finding orthogonal separators in CMMb [9].

LP Weighted Orthogonal Separators

Input: An SDP solution $\{u_i\}$, an LP solution $\{x(u, i)\}$.
Output: A set S of label vertex pairs (u, i).

1. Set a parameter $\alpha = 1/(2k^2)$, which we call the probability scale.

2. Generate a random Gaussian vector g with independent components distributed as $\mathcal{N}(0, 1)$.

3. Fix a threshold t s.t. $\Pr(\xi \geq t) = \alpha$, where $\xi \sim \mathcal{N}(0, 1)$.

4. Pick a random uniform value r in the interval $(0, 1)$.

5. Find set
$$S = \{(u, i) \in V \times [k] : \langle u_i, g \rangle \geq t \text{ and } x(u, i) \geq r\}.$$

6. Return S.

The rounding algorithm is given below.

LP and SDP Based Rounding Algorithm

Input: An instance of unique games.
Output: An assignment of labels to the vertices.

1. Solve the SDP.

2. Find the set of all super short edges Γ_η.

3. Solve the LP.

4. Mark all vertices unprocessed.

5. while (there are unprocessed vertices)

   - Sample a set S of vertex–label pairs using LP weighted orthogonal separators.

   - For all unprocessed vertices u:
     - Let $S_u = \{i : (u, i) \in S\}$
     - If S_u contains exactly one element i, assign label i to u and mark u as processed.

If after nk/α iterations, there are unprocessed vertices, the algorithm assigns arbitrary labels to them and terminates.

Lemma 3.8. *Let S be an LP weighted orthogonal separator. Then, for every $(u, i) \in V \times [k]$,*

- $\Pr((u, i) \in S) = \alpha x(u, i)$.

For every $((u, i), (v, j)) \in \Gamma_\eta$ and $(u, i') \in V \times [k]$ $(i' \neq i)$,

- $\Pr((u, i) \in S \text{ and } (v, j) \in S) \geq \alpha \min(x(u, i), x(v, j))(1 - \eta).$

- $\Pr((u, i) \in S; (v, j) \in S; (u, i') \in S) \leq \alpha \min(x(u, i), x(v, j))/(2k^2).$

Proof. We have

$$\Pr((u, i) \in S) = \Pr(\langle u_i, g \rangle \geq t \text{ and } x(u, i) \geq r) = \Pr(\langle u_i, g \rangle \geq t) \Pr(x(u, i) \geq r) = \alpha x(u, i).$$

Then, by Lemma 2.1,

$$\begin{aligned}
\Pr((u, i) \in S \text{ and } (v, j) \in S) &= \Pr(\langle u_i, g \rangle \geq t \text{ and } \langle v_j, g \rangle \geq t) \Pr(\min(x(u, i), x(v, j)) \geq r) \\
&\geq \alpha(1 - \eta) \min(x(u, i), x(v, j)).
\end{aligned}$$

Finally, we have (below we use that $\langle u_i, g \rangle$ and $\langle u_{i'}, g \rangle$ are independent random variables)

$$\begin{aligned}
\Pr((u, i) \in S; (v, j) \in S; (u, i') \in S) &\leq \Pr(\langle u_i, g \rangle \geq t) \Pr(\langle u_{i'}, g \rangle \geq t) \Pr(\min(x(u, i), x(v, j)) \geq r) \\
&= \alpha^2 \min(x(u, i), x(v, j)) \leq \alpha \min(x(u, i), x(v, j))/(2k^2).
\end{aligned}$$

$\square$

Lemma 3.9. *Given a C-SDP solution $\{u_i\}$ and an LP solution $\{x(u, i)\}$ of value at least $LP \geq x|E|$, the algorithm finds a solution to the unique games instance that satisfies $(1 - \eta)x/(2 - (1 - \eta)x) - O(1/k)$ fraction of all constraints in the expectation.*

Proof. Consider an arbitrary edge (u, v). We estimate the probability that the algorithm assigns labels that satisfy the constraint π_{uv}. For simplicity of presentation, assume that $\pi_{uv}(i) = i$ (we may always assume this by renaming the labels of v). Let $\delta_i(u, v) = \min(x(u, i), x(v, i))$ if $((u, i), (v, i)) \in \Gamma_\eta$; and $\delta_i(u, v) = 0$, otherwise. Let $\delta(u, v) = \sum_i \delta_i(u, v)$ and $\delta'(u, v) = \delta(u, v)(1 - \eta - 1/k)$.

Consider an arbitrary iteration at which both u and v have not yet been processed. By Lemma 3.8 (item 2), if $((u, i), (v, i)) \in \Gamma_\eta$, then

$$\Pr(i \in S_u \text{ and } i \in S_v) \geq \alpha \min(x(u, i), x(v, i))(1 - \eta).$$

Then, by Lemma 3.8 (3) and the union bound, the probability that S_u or S_v contains more than one element and $i \in S_u$, $i \in S_v$ is at most $\alpha \min(x(u, i), x(v, i))/k$. Hence, the algorithm assigns i to both u and v with probability at least

$$\alpha \min(x(u, i), x(v, i))(1 - \eta - 1/k) = \alpha \delta_i(u, v) \times (1 - \eta - 1/k).$$

The probability that the algorithm assigns the same label to u and v is at least

$$\sum_{i:((u,i),(v,i))\in\Gamma_\eta} \alpha \delta_i(u, v) \times (1 - \eta - 1/k) = \alpha \delta'(u, v).$$

The probability that the algorithm assigns a label to u is at most α and similarly the probability that the algorithm assigns a label to v is at most α. Thus the probability that it assigns a label to either u or v is at most $\alpha(2 - \delta'(u, v))$.

Hence, the probability that the algorithm assigns the same label to u and v at one of the iterations is at least (note that the probability that there are unlabeled vertices when the algorithm stops after nk/α

13

iterations is exponentially small, therefore, for simplicity we may assume that the number of iterations is infinite)

$$\sum_{t=0}^{\infty}(1 - \alpha(2 - \delta'(u, v)))^t \alpha \delta'(u, v) = \frac{\alpha \delta'(u, v)}{\alpha(2 - \delta'(u, v))} = \frac{\delta'(u, v)}{2 - \delta'(u, v)}.$$

The function $\delta \mapsto \delta/(2 - \delta)$ is convex on $(0, 2)$ and

$$\frac{1}{|E|}\sum_{(u,v)\in E}\delta'(u, v) = \frac{1}{|E|}\sum_{(u,v)\in E}\delta(u, v)(1 - \eta - 1/k) \geq x(1 - \eta - 1/k),$$

thus, by Jensen's inequality, the expected number of satisfied constraints is at least

$$\frac{x(1 - \eta - 1/k)}{2 - x(1 - \eta - 1/k)}|E|.$$

$\square$

3.3 Hardness: Semi-Random Instances for $\varepsilon \geq 1/2$

In this section, we show that the problem becomes hard when $\varepsilon \geq 1/2$ assuming the 2–to–2 conjecture. The 2–to–2 conjecture follows from Khot's 2–to–1 conjecture[4], which is a frequently used complexity assumption [17]. We prove the following theorem.

Theorem 3.10. *For every $\varepsilon \geq 1/2$ and $\delta > 0$, no polynomial-time algorithm can distinguish with probability greater than $o(1)$ between the following two cases:*

1. *Yes Case: the instance is a $1 - \varepsilon$ satisfiable semi-random instance (in the "random edges, adversarial constraints" model),*

2. *No Case: the instance is at most δ satisfiable.*

This result holds if the 2–to–2 conjecture holds.

Before we proceed with the proof we remind the reader the definition of 2–to–2 games and 2–to–2 conjecture.

Definition 3.11. *In a 2–to–2 game, we are given a graph $G = (V, E)$, a set of labels $[k] = \{0, \ldots, k-1\}$ (k is even) and set of constraints, one constraint for every edge (u, v). Each constraint is defined by a 2–to–2 predicate Π_{uv}: for every label i there are exactly two labels j such that $\Pi_{uv}(i, j) = 1$ (the predicate is satisfied); similarly, for every j there are exactly two labels i such that $\Pi_{uv}(i, j) = 1$. Our goal is to assign a label $x_u \in [k]$ to every vertex u so as to maximize the number of satisfied constraints $\Pi_{uv}(x_u, x_v) = 1$. The value of the solution is the number of satisfied constraints.*

Definition 3.12. *The 2–to–2 conjecture states that for every $\delta > 0$ and sufficiently large k, there is no polynomial time algorithm that distinguishes between the following two cases (i) the instance is completely satisfiable and (ii) the instance is at most δ satisfiable.*

---

[4] The 2–to–1 conjecture implies the 2–to–2 conjecture because every 2–to–1 game can be converted to a 2–to–2 game of the same value as follows. Let $\mathcal{I}$ be an instance of a 2–to–1 game on a graph (L, R, E), s.t. every constraint has degree 2 on the left side, and degree 1 on the right side. We create two copies a_1 and a_2 for every label a for vertices in L. We replace each constraint Π_{lr} with a constraint Π_{lr}^* defined by $\Pi_{lr}^*(a_1, b) = \Pi_{lr}^*(a_2, b) = \Pi_{lr}(a, b)$. We obtain a 2–to–2 game. It is clear that its value equals the value of the original 2–to–1 game.

Proof. We construct a randomized reduction that maps every satisfiable 2–to–2 game to a semi-random unique game, and every at most δ satisfiable 2–to–2 game to an at most δ satisfiable unique game. Suppose we are given an instance of a 2–to–2 game on a graph G with predicates Π_{uv}. For each predicate Π_{uv}, we find two permutations π^0_{uv} and π^1_{uv} such that $\Pi_{uv}(i, \pi^0_{uv}(i)) = 1$ and $\Pi_{uv}(i, \pi^1_{uv}(i)) = 1$, and $\pi^0_{uv}(i) \neq \pi^1_{uv}(i)$ for every label i as follows. We consider a bipartite graph on vertex set $\{(u, i) : i \in [k]\} \cup \{(v, j) : j \in [k]\}$, in which two vertices (u, i) and (v, j) are connected with an edge if $\Pi_{uv}(i, j) = 1$. Every vertex has degree 2 in this graph. Therefore, it is a union of two matchings; each of them defines a permutation. Now for every edge (u, v), we choose one of the two permutations, π^0_{uv} or π^1_{uv}, at random. We obtain a unique game instance.

First of all it is clear, that the value of the unique game is at most the value of the 2–to–2 game (if a labeling $\{x_u\}$ satisfies $x_v = \pi^0_{uv}(x_u)$ or $x_v = \pi^1_{uv}(x_u)$ then it also satisfies $\Pi_{uv}(x_u, x_v)$). Hence our reduction maps an at most δ satisfiable 2–to–2 game to an at most δ satisfiable unique game. Now suppose that the 2–to–2 game instance is completely satisfiable. We show how an adversary can generate semi-random instances that have the same distribution as instances generated by our reduction. The adversary finds a solution $\{x^*_u\}$ to the 2–to–2 game that satisfies all constraints Π_{uv}. For every edge (u, v), she chooses $r_{uv} \in \{0, 1\}$ such that $\pi^{r_{uv}}_{uv}(x^*_u) = x^*_v$. She obtains a completely satisfiable unique game on G with constraints $\pi^{r_{uv}}_{uv}(x_u) = x_v$. Now she performs the random step — she chooses a random set of edges E_ε; every edge belongs to E_ε with probability ε. She replaces every constraint $\pi^{r_{uv}}_{uv}(x_u) = x_v$ with the constraint $\pi^{1-r_{uv}}_{uv}(x_u) = x_v$ with probability $1/(2\varepsilon)$, and returns the obtained instance. Note that the constraint for the edge (u, v) is $\pi^0_{uv}(x_u) = x_v$ with probability $1/2$ and $\pi^1_{uv}(x_u) = x_v$ with probability $1/2$. Therefore, the distribution of instances coincides with the distribution generated by our reduction. $\qquad \square$

Our positive results (for $\varepsilon < 1/3$) apply only to graphs with average degree greater than $\varepsilon^{-1} \log k \log \log k$. We want to point out that our negative results (for $\varepsilon \geq 1/2$) also apply to graphs with high average degree. Indeed, note that our reduction does not change the constraint graph. As the following lemma shows, if the 2–to–2 conjecture is true then it is also true for instances on graphs with very high average degree (as large as $n^{1-\theta} \gg k \log k$).

Lemma 3.13. *Suppose that the 2–to–2 conjecture holds. Then it also holds for graphs with average degree at least $n^{1-\theta}$ (for every $\theta > 0$).*

Proof. Suppose we are given an instance $\mathcal{I}$ of a 2–to–2 game on a graph $G = (V, E)$ with constraints $\Pi_{uv}(x_u, x_v)$. We construct a new instance $\mathcal{I}'$ of a 2–to–2 game as follows. For every vertex u, we create a "cloud" of $N = |V|^{2/\theta}$ new vertices V_u. We connect all vertices in V_u with all vertices in V_v for every edge (u, v). For every edge between vertices $a \in V_u$ and $b \in V_v$, we add the constraint $\Pi_{uv}(x_a, x_b) = 1$ (where Π_{uv} is the predicate for the edge (u, v) in $\mathcal{I}$). We obtain a 2–to–2 game on a graph with average degree $2N^2|E|/(N|V|) \geq 2N/|V| \geq (N|V|)^{1-\theta}$.

Clearly, this is a polynomial–time reduction. We show that the values of $\mathcal{I}$ and $\mathcal{I}'$ are equal. If a solution $\{x_u\}$ for $\mathcal{I}$ has value t then the solution defined by $x_a = x_u$ if $a \in V_u$ for $\mathcal{I}'$ has also value t. On the other hand, given a solution $\{x_a\}$ for $\mathcal{I}'$ of value t, we construct a solution for $\mathcal{I}$ as follows: for every u, we choose a random vertex a from V_u and let $x_u = x_a$. Then the expected value of this solution is t. Therefore, the value of the optimal solution for $\mathcal{I}$ is at least t. $\qquad \square$

4 Adversarial Edges, Random Constraints

Theorem 4.1. *There exists a polynomial-time approximation algorithm that given $k \in \mathbb{N}$ ($k \geq k_0$), $\varepsilon \in (0, 1)$, $\eta \in (c\sqrt{\varepsilon \log k/k}, 1)$ and a semi-random instance of unique games from the "adversarial edges,*

random constraints" model on graph $G = (V, E)$ with at least $C\eta^{-2}n \log k$ edges (C is a sufficiently large absolute constant) finds a solution of value $1 - O(\varepsilon + \eta)$, with probability $1 - o(1)$.

Our algorithm proceeds in several steps. First, it solves the standard SDP relaxation for Unique Games. Then it removes "η–long edges" (see below) with respect to the SDP solution, and finally it runs the CMMa [8] algorithm to solve the unique game on the remaining graph (the CMMa algorithm will again solve the SDP relaxation for Unique Games — it cannot reuse our SDP solution).

Definition 4.2. *We say that an edge (u, v) is η–long with respect to an SDP solution $\{u_i\}$, if*

$$\frac{1}{2} \sum_{i \in [k]} \|u_i - v_{\pi_{uv}(i)}\|^2 > \eta.$$

Otherwise, we say that the edge (u, v) is η–short.

Now we formally present the algorithm.

Input: An instance of unique games.
Output: An assignment of labels to the vertices.

1. Solve the SDP and obtain an SDP solution $\{u_i^*\}$.

2. Remove all $1/16$–long (with respect to $\{u_i^*\}$) edges $(u, v) \in E$ from the graph G. Denote the new graph G^*.

3. Solve the SDP on the graph G^* and run the CMMa algorithm.

In Theorem 4.4, we show that after removing all $1/16$–long edges from the graph G, the unique games instance contains at most $O(\gamma)|E|$ corrupted constraint w.h.p, where $\gamma = \eta^2 / \log k$. Since the value of the optimal SDP is at least ε, the algorithm removes at most 16ε edges at step 2. In the remaining graph, G', the CMMa algorithm finds an assignment satisfying $1 - O(\sqrt{\gamma \log k}) = 1 - O(\eta)$ fraction of all constraints. This assignment satisfies at least $1 - O(\varepsilon + \eta)$ fraction of all constraints in G.

Remark 4.3. *In the previous section we proved that a typical instance of unique games in the "random edges, adversarial constraints" model contains many "super short" edges of the label-extended graph. Then we showed how we can find an assignment satisfying many super short edges. Note, that edges in the set E_ε are not necessarily short or long. In this section, we show something very different: in the typical instance of unique games in the "adversarial edges, random constraints" model, most edges in the set E_ε are long. However, note that the label-extended graph does not have to have any super short edges at all.*

4.1 Almost All Corrupted Edges are Long

Theorem 4.4. *Let $k \in \mathbb{N}$ ($k \geq k_0$), $\varepsilon \in (0, 1]$, $\gamma \in (c\varepsilon/\sqrt{k}, 1/\log k)$. Consider a graph $G = (V, E)$ with at least $C\gamma^{-1}n$ edges and a unique game instance on G (c, C, and k_0 are absolute constants). Suppose that all constraints for edges in E_ε are chosen at random; where E_ε is a set of edges of size $\varepsilon|E|$. Then, the set E_ε contains less than $\gamma|E|$ $1/16$–short edges w.r.t. every SDP solution $\{u_i\}$ with probability $1 - o(1)$.*

Proof. Consider a semi-random instance of unique games. Let $\{u_i\}$ be an SDP solution. Suppose, that at least $\gamma|E|$ edges in E_ε is $1/16$–short. Define,

$$f_{\alpha,\{u_i\}}(u,i,v) = \begin{cases} 1, & \text{if } \|u_i - v_{\pi_{uv}(i)}\|^2 < \alpha(\|u_i\|^2 + \|v_{\pi_{uv}(i)}\|^2); \\ 0, & \text{otherwise.} \end{cases}$$

We shall omit the second index of f, when it is clear that we are talking about the set $\{u_i\}$. For every short $1/16$–short edge $(u,v) \in E_\varepsilon$, we have

$$\frac{1}{2} \sum_{\substack{i \in [k] \\ j = \pi_{uv}(i)}} \frac{\|u_i\|^2 + \|v_j\|^2}{8} \times (1 - f_{1/8}(u,i,v)) \leq \frac{1}{2} \sum_{\substack{i \in [k] \\ j = \pi_{uv}(i)}} \|u_i - v_j\|^2 \leq \frac{1}{16}, \tag{6}$$

and, hence,

$$\frac{1}{2} \sum_{\substack{i \in [k] \\ j = \pi_{uv}(i)}} (\|u_i\|^2 + \|v_j\|^2) f_{1/8}(u,i,v) \geq \frac{1}{2} \sum_{\substack{i \in [k] \\ j = \pi_{uv}(i)}} (\|u_i\|^2 + \|v_j\|^2) - \frac{1}{2} = \frac{1}{2}. \tag{7}$$

We get (from (7) and the assumption that there are at least $\gamma|E|$ $1/16$–short edges in E_ε)

$$\sum_{(u,i) \in V \times [k]} \sum_{v:(u,v) \in E_\varepsilon} \|u_i\|^2 f_{1/8}(u,i,v) = \sum_{(u,v) \in E_\varepsilon} \sum_{\substack{i \in [k] \\ j = \pi_{uv}(i)}} (\|u_i\|^2 + \|v_j\|^2) f_{1/8}(u,i,v) \geq \gamma|E|.$$

$$= \frac{\gamma}{2}|E| + \frac{\gamma}{8} \sum_{(u,i) \in V \times [k]} \|u_i\|^2 (\varepsilon^{-1} \deg_{E_\varepsilon}(u) + \Delta).$$

Here $\deg_{E_\varepsilon}(u)$ denotes the number of edges in E_ε incident to u; and Δ denotes the average degree in the graph G. We used that $\sum_{i \in [k]} \|u_i\|^2 = 1$. We now group all vertices of the label–extended graph $(u,i) \in E \times [k]$ depending on the value of $\log_2(\|u_i\|^2)$. Specifically, we pick a random $r \in (1,2)$ distributed with density $(x \ln 2)^{-1}$ and for every $t \in \mathbb{N}$, define

$$V_{t,r} = \{(u,i) \in V \times [k] : \|u_i\|^2 \in [r2^{-(t+1)}, r2^{-t}]\}.$$

We define $V_{\infty,r} = \{(u,i) \in V \times [k] : \|u_i\|^2 = 0\}$. Let $g_r(u,i,v) = 1$, if (u,i) and $(v, \pi_{uv}(i))$ lie in the same group $V_{t,r}$; and $g_r(u,i,v) = 0$, otherwise. We claim, that if $f_{1/8}(u,i,v) = 1$, then $\Pr_r(g_r(u,i,v) = 1) \geq 1/2$.

Claim 4.5. *If $f_{1/8}(u,i,v) = 1$, then $\Pr_r(g_r(u,i,v) = 1) \geq 1/2$.*

Proof. Denote $j = \pi_{uv}(i)$. If $f_{1/8}(u,i,v) = 1$, then $u_i \neq 0$ and $v_j \neq 0$. Assume w.l.o.g. that $\|u_i\|^2 \geq \|v_j\|^2$. Note, that $g_r(u,i,v) = 0$, if and only if $r2^{-t} \in [\|v_j\|^2, \|u_i\|^2]$ for some $t \in \mathbb{N}$. The probability of this event is bounded by $\ln(\|u_i\|^2/\|v_j\|^2)/\ln 2$, since the probability that an interval $[x, x + \Delta x]$ contains a point from $\{r2^{-t} : t \in \mathbb{N}\}$ is $\Delta x/(x \ln(2)) + o(1)$. Using the inequality

$$\|u_i\|^2 - \|v_j\|^2 \leq \|u_i - v_j\|^2 \leq \frac{\|u_i\|^2 + \|v_j\|^2}{8},$$

we get $\|u_i\|^2/\|v_j\|^2 \leq 9/7$. Thus,

$$\Pr_r(g_r(u,i,v) = 0) \leq \ln(9/7)/\ln 2 < 1/2.$$

$\square$

Therefore, for some $r \in [1, 2]$,

$$\sum_{(u,i) \in V \times [k]} \sum_{v:(u,v) \in E_\varepsilon} \|u_i\|^2 f_{1/8}(u,i,v) g_r(u,i,v) \geq \frac{\gamma}{4}|E| + \frac{\gamma}{16} \sum_{(u,i) \in V \times [k]} \|u_i\|^2 (\varepsilon^{-1} \deg_{E_\varepsilon}(u) + \Delta).$$

Rewrite this inequality as follows

$$\sum_{t \in \mathbb{N}} \sum_{(u,i) \in V_{t,r} \times [k]} \sum_{v:(u,v) \in E_\varepsilon} \|u_i\|^2 f_{1/8}(u,i,v) g_r(u,i,v) \geq$$

$$\sum_{t \in \mathbb{N}} 2^{-(t+1)} \frac{\gamma}{4}|E| + \sum_{t \in \mathbb{N}} \sum_{(u,i) \in V_{t,r}} \|u_i\|^2 (\varepsilon^{-1} \deg_{E_\varepsilon}(u) + \Delta).$$

For some $t \in \mathbb{N}$,

$$\sum_{(u,i) \in V_{t,r}} \sum_{v:(u,v) \in E_\varepsilon} \|u_i\|^2 f_{1/8}(u,i,v) g_r(u,i,v) \geq \frac{2^{-(t+1)} \gamma |E|}{4} + \frac{\gamma}{16} \sum_{(u,i) \in V_{t,r}} \|u_i\|^2 (\varepsilon^{-1} \deg_{E_\varepsilon}(u) + \Delta).$$

We now replace each term $\|u_i\|^2$ in the left hand side with the upper bound $r2^{-t}$ and each term $\|u_i\|^2$ in the right hand side with the lower bound $r2^{-(t+1)}$. Then, we divide both sides by $r2^{-t} \leq 2^{-t}$.

$$\sum_{(u,i) \in V_{t,r}} \sum_{v:(u,v) \in E_\varepsilon} f_{1/8}(u,i,v) g_r(u,i,v) \geq \frac{\gamma}{16}|E| + \frac{\gamma}{32} \sum_{(u,i) \in V_{t,r}} \varepsilon^{-1} \deg_{E_\varepsilon}(u) + \frac{\gamma}{32}|V_{t,r}|\Delta.$$

Define a new set of vectors $u_i^* = u_i/\|u_i\|$, if $(u,i) \in V_{t,r}$; $u_i^* = 0$, otherwise. Observe, that if $f_\alpha(u,i,v) g_r(u,i,v) = 1$, then $f_{\alpha,\{u_i^*\}}(u,i,v) = 1$ since for $j = \pi_{uv}(i)$,

$$\begin{aligned}
\|u_i^* - v_j^*\|^2 &= 2 - 2\langle u_i^*, v_j^* \rangle = 2 - 2\frac{\langle u_i, v_j \rangle}{\|u_i\| \|v_j\|} = 2 - \frac{\|u_i\|^2 + \|v_j\|^2 - \|u_i - v_j\|^2}{\|u_i\| \|v_j\|} \\
&< 2 - (1-\alpha)\frac{\|u_i\|^2 + \|v_j\|^2}{\|u_i\| \|v_j\|} \leq 2 - (1-\alpha) \cdot 2 = 2\alpha = \alpha(\|u_i^*\|^2 + \|v_j^*\|^2).
\end{aligned}$$

Therefore,

$$\sum_{(u,i) \in V_{t,r}} \sum_{v:(u,v) \in E_\varepsilon} f_{1/8,\{u_i^*\}}(u,i,v) \geq \frac{\gamma}{16}|E| + \frac{\gamma}{32} \sum_{(u,i) \in V_{t,r}} \varepsilon^{-1} \deg_{E_\varepsilon}(u) + \frac{\gamma}{32}|V_{t,r}|\Delta. \tag{8}$$

We now embed vectors $\{u_i^*\}$ in a net N of size $\exp(O(\log k))$ using a randomized mapping φ (see Lemma 3.7, a variant of the Johnson–Lindenstrauss lemma), so that for some small absolute constant β and every $(u,i), (v,j) \in V_{t,r}$,

$$\Pr((1+\beta)^{-1}\|u-v\|^2 - \beta \leq \|\varphi(u) - \varphi(v)\|^2 \leq (1+\beta)\|u-v\|^2 + \beta) \geq 1 - \frac{\beta}{k^2}.$$

We say that a vertex $u \in V$ is good if for all $i', i'' \in \{i : (u,i) \in V_{t,r}\}$, the following inequality holds: $\|\varphi(u_{i'}^*) - \varphi(u_{i''}^*)\|^2 \geq 2 - 3\beta$ for $i' \neq i''$. I.e., if u is good then all vectors $\varphi(u_i^*)$ are almost orthogonal. By Lemma 3.7, vertex u is good with probability at least $1 - \beta$. We let $u_i^{**} = \varphi(u_i^*)$, if u is good; and $u_i^{**} = 0$, otherwise. To slightly simplify the proof, for every vertex u, we also zero out a random subset of $\lceil \beta k/4 \rceil$ vectors u_i^{**}. Thus, if $f_{1/8,\{u_i^*\}}(u,i,v) = 1$, then $\Pr(f_{1/4,\{u_i^{**}\}}(u,i,v) = 1) \geq 1 - 4\beta$ (unless u or

18

v are not good vertices; u_i or $v_{\pi_{uv}(i)}$ has been zeroed; or the distance between u_i^* and $v_{\pi_{uv}(i)}^*$ is significantly distorted by ϕ). Hence, for some sample $\{u_i^{**}\}$,

$$R \equiv \sum_{(u,i):\|u_i^{**}\|=1} \sum_{v:(u,v)\in E_\varepsilon} f_{1/4,\{u_i^{**}\}}(u,i,v) \geq \frac{\gamma}{64}\Big(|E| + \sum_{(u,i)\in V_{t,r}} \varepsilon^{-1}\deg_{E_\varepsilon}(u) + |V_{t,r}|\Delta\Big) \equiv S. \quad (9)$$

We denote the left hand side by R. We denote the right hand side by S.

We now fix the set of vectors $\{u_i^{**}\}$ and estimate the probability that for a given set $\{u_i^{**}\}$ and a random set of constraints $\{\pi_{uv}\}$ on edges $(u,v) \in E_\varepsilon$ inequality (9) holds. For each $(u,v) \in E_\varepsilon$ and i, $\Pr_{\{\pi_{uv}\}}(f_{1/4,\{u_i^{**}\}}(u,i,v) = 1) \leq 1/k$ (since for every i, there is at most one $j \in [k]$ such that $\|u_i^{**} - v_j^{**}\|^2 \leq 1/4$; here we use that all non-zero vectors v_j are almost orthogonal). Thus,

$$\mathbb{E}R = \mathbb{E}\sum_{(u,i):\|u_i^{**}\|^2=1} \sum_{v:(u,v)\in E_\varepsilon} f_{1/4,\{u_i^{**}\}}(u,i,v) \leq \frac{1}{k}\sum_{(u,i):\|u_i^{**}\|^2=1} \deg_{E_\varepsilon}(u) < 64S/(\sqrt{k}c).$$

In the last inequality, we used the bound $\gamma \geq c\varepsilon/\sqrt{k}$ and (9). We assume that $64/c \ll 1$. We would like to apply the Bernstein inequality, which would give us an upper bound of

$$\exp(-\Omega(S)) \equiv \exp\Big(-\Omega\Big((\gamma|E| + \gamma\sum_{(u,i)\in V_{t,r}} \varepsilon^{-1}\deg_{E_\varepsilon}(u) + \gamma|V_{t,r}|\Delta\Big)\log(k)\Big) \quad (10)$$

on the probability that inequality (9) holds (note: $\mathbb{E}[R] \ll S$; $\log(S/\mathbb{E}[R]) \geq \Theta(\log k)$). Formally, we cannot use this inequality, since random variables $f_{1/4,\{u_i^{**}\}}(u,i,v)$ are not independent. So, instead, we consider a random process, where we consequently pick edges $(u,v) \in E_\varepsilon$ and then for every i such that $u_i^{**} \neq 0$ pick a random yet unchosen $j \in [k]$ and set $\pi_{uv}(i) = j$. For every u and i, there are at least $\lceil \beta k/4 \rceil$ different candidates j (since at least $\lceil \beta k/4 \rceil$ u_i's equal 0) and for at most one of them $\|u_i^{**} - v_j^{**}\|^2 \leq 1/4$. Hence, at every step $\Pr(f_{1/4,\{u_i^{**}\}}(u,i,v) = 1) \leq 4/(\beta k)$. We now apply a martingale concentration inequality and get the same bound (10).

Finally, we bound the number of possible solutions $\{u_i^{**}\}$ and then apply the union bound to show that w.h.p. the set E_ε contains less than $\gamma|E|$ 1/16-short edges with respect to every SDP solution. Let $W = |V_{t,r}|$ be the number of pairs (u,i) such that $u_i^{**} \neq 0$. We choose W variables among $\{u_i^{**}\}_{u,i}$ and assign them values from N as follows. Let $a_u = |\{u_i^{**} \neq 0 : i\}|$. First, we choose the values of a_u for all vertices (note that $\sum_{u\in V} a_u = W$). The number of ways to choose $\{a_u\}$ is at most the number of ways to write W as the sum of n numbers, which is $\binom{n+W-1}{n-1} < 2^{n+W}$. Then for each vertex u, we choose a_u labels $i_1, \ldots, i_{a_u}$ for which $u_{i_1}^{**} \neq 0, \ldots, u_{i_{a_u}}^{**} \neq 0$. The number of ways to choose labels is at most $\prod_{u\in V} k^{a_u} = k^W$. Finally, we assign a vector from N to each chosen variable u_i^{**}. The number of ways to do so is $|N|^W$. Thus there are at most

$$2^{n+W} \times |N|^W \times k^W = \exp(O(W \log k + n))$$

ways to choose vectors $\{u_i^{**}\}$ for given W. Since $\gamma|V_{t,r}|\Delta \log k \geq \gamma W \times (C\gamma^{-1})\log k \geq CW \log k$ and $\gamma|E|\log k \geq Cn$ inequality (9) holds for some collection $\{u^{**}\}$ with exponentially small probability $\exp(-\Omega(n))$ for every fixed value of W. Since there are only n possible values on W, inequality (9) holds for some value of W with exponentially small probability. $\qquad\square$

5 Random Corrupted Constraints in Linear Unique Games

In this section we study a special family of unique games, so-called Linear Unique Games (or MAX Γ-Lin). In Linear Unique Games all labels $i \in [k]$ are elements of the cyclic group $\mathbb{Z}/k\mathbb{Z}$; all constraints have the form $x_u = x_v - s_{uv}$, where $s_{uv} \in \mathbb{Z}/k\mathbb{Z}$. It is known that if the Unique Games Conjecture holds for general unique games, then it also holds for this family. We note that the results from the previous subsection are not directly applicable for Linear Unique Games, since random permutations are restricted to be shifts $\pi_{uv} : x_u \mapsto x_v + s_{uv}$ and, thus, a random unique game from the "adversarial edges, random constraints" model is most likely not a linear unique game. This is not the case in the "random edges, adversarial constraints" model, since there the adversary may choose corrupted constraints in any fashion she wants, particularly to be shifts. We note that the results of this section can be easily generalized to any finite Abelian group using the decomposition of the group into a sum of cyclic subgroups. We omit the details from the conference version of the paper.

For Linear Unique Games, we use a shift invariant variant of the SDP relaxation (see e.g., a paper by Andersson, Engebretsen, and Håstad [2]). In this relaxation, all vectors u_i are unit vectors, satisfying ℓ_2^2 triangle inequalities, and an extra constraint $\|u_i - v_j\| = \|u_{i+s} - v_{j+s}\|$. The objective is to minimize

$$\frac{1}{k} \times \frac{1}{2|E|} \sum_{(u,v) \in E} \sum_{i \in [k]} \|u_i - v_{i+s_{uv}}\|^2.$$

Given an arbitrary SDP solution $\{u_i\}$ one can obtain a uniform solution using symmetrization (this trick works only for unique games on groups, not arbitrary unique games):

$$u_i' = \frac{1}{\sqrt{k}} \bigoplus_{j \in \mathbb{Z}/k\mathbb{Z}} u_{i+j}.$$

It is easy to verify (and it is well-known, see e.g., [8]) that vectors $\{u_i'\}$ satisfy all SDP constraints and the objective value of the SDP does not change. We adapt the definition of η–short edges for this variant of SDP as follows: we say that an edge (u, v) is η–short with respect to the SDP solution $\{u_i\}$ if $1/2 \|u_i - v_{i+s}\|^2 \leq \eta$ for every $i \in [k]$ (or, equivalently, for some $i \in [k]$, since $\|u_i - v_{i+s}\| = \|u_j - v_{j+s}\|$).

We prove the following analog of Theorem 4.4, which immediately implies that our general algorithm for the adversarial edges, random constraints model works for linear unique games.

Theorem 5.1. *Let $k \in \mathbb{N}$ ($k \geq k_0$), $\varepsilon \in (0, 1)$, $\eta \in (\varepsilon / \log k, 1/ \log k)$. Consider a graph $G = (V, E)$ with at least $Cn\gamma^{-1}$ edges (k_0, c, C are absolute constants) and a random linear unique game from the "adversarial edges, random constraints" model. Let E_ε be the set of corrupted edges; and let $\{u_i\}$ be the optimal SDP solution. Then, with probability $1 - o(1)$, at most $\gamma|E|$ edges in E_ε are $1/32$–short.*

Proof. As mentioned in the beginning of this section the result we would like to prove holds not only for cyclic groups but for arbitrary Abelian groups. In fact, the proof is simpler if the group can be decomposed into a direct sum of many cyclic groups of small size. So, in some sense, our plan is to represent the cyclic group as a "pseudo-direct sum" of such groups. Before proceeding to the proof, we note that all transformations described below are only required for the proof and are not performed by the approximation algorithm.

Let S be a circle (or a one dimensional torus). We identify S with the segment $[0, 1]$ with "glued" ends 0 and 1. Define an embedding $\theta : [k] \to S$ as follows $\theta(i) = i/k$. We denote the distance between two adjacent labels by $\tau = 1/k$. Below, all "$+$" and "$-$" operations on elements of S are "modulo 1" and all operations on $[k]$ (and later $[T]$) are modulo k (and T respectively).

Recall, that in the "adversarial edges, random constraints" model of Linear Unique Games, the adversary first chooses a set of edges E_ε, then the "nature" randomly picks a shift $s_{uv} \in [k]$ for every $(u, v) \in E_\varepsilon$. We assume that the nature picks s_{uv} in the following way: it chooses a random $\tilde{s}_{uv} \in S$, and then defines $s_{uv} = \lfloor \tilde{s}_{uv} k \rfloor$. Then, clearly, s_{uv} is distributed uniformly in $[k]$.

If M is a subset of S, then $M + s$ denotes the shift of M by s. We denote by $L(M)$ the set of labels covered by M:

$$L(M) = \{i \in [k] : \theta(i) \in M\}.$$

We let $\Psi_u(M) = \sum_{i \in L(M)} u_i$, and $\Psi'_u(M) = \Psi_u(M) / \|\Psi_u(M)\|$.

Fix $T = R^2$ (where $R \in \mathbb{N}$) to be a sufficiently large absolute constant (not depending on k or any other parameters). We represent each $\tilde{s} \in S$ as

$$\tilde{s} = \sum_{p=1}^{\infty} \tilde{s}_p / T^p,$$

where every $\tilde{s}_p$ belongs to $[T]$, in other words, $0.\tilde{s}_1\tilde{s}_2\tilde{s}_3 \ldots$ is a representation of $\tilde{s}$ in the numeral system with base T. (This representation is not unique for a set of measure 0. If it is not unique, we pick the representation that ends in zeros.) We define $D_p(\tilde{s}) = \tilde{s}_p$. Note, that if $\tilde{s}$ is uniformly distributed random variable in S, then $D_p(\tilde{s})$ is uniformly distributed in $[T]$ and all $D_p(\tilde{s})$ are independent for different p. We are interested in the first $P = \lceil 2 \ln \ln k \rceil$ digits of $\tilde{s}_{uv}$. We have chosen P so that $T^{-P} = \log^{-\Theta(1)} k > \tau R \equiv R/k$, but $e^P \geq \ln^2 k$.

For every integer $p \in \{1, \ldots, P\}$ and $d \in [T]$ define

$$M_{p,d} = \{\tilde{s} \in S : D_p(\tilde{s}) \in [d, d + R - 1]\},$$

that is, $M_{p,d}$ is the set of numbers in $[0, 1]$, whose p-th digit is in the range $[d, d + R - 1]$. Note, that the set $\{\tilde{s} \in S : D_p(\tilde{s}) = d\}$ is a union of segments of length $T^{-P} > \tau R \equiv R/k$ (here and below, all segments are left-closed and right-open).

Lemma 5.2. *Let $M \subset S$ be a union of segments of length at least $R\tau$ and let $\tilde{s} \in [0, 1]$. Then,*

$$(1 - 1/R)\, \mu(M)/\tau \leq |L(M)| \leq (1 + 1/R)\, \mu(M)/\tau.$$

and

$$|L(M) \triangle L(M + \tilde{s})| \leq \min(|L(M)|, |L(M + \tilde{s})|)/R \times (2\tilde{s}/\tau + 2),$$

here $\triangle$ denotes the symmetric difference of two sets; and μ is the uniform measure on $[0, 1]$.

Proof. Let M be the disjoint union of segments $I \in \mathcal{I}$ of length at least $R\tau$ each. Consider one of the segments $I \subset M$. This segment covers at least $\lfloor \mu(I)/\tau \rfloor \geq (\mu(I)/\tau - 1)$ and at most $\lceil \mu(I)/\tau \rceil \leq (\mu(I)/\tau + 1)$ points $\theta(i)$. In other words, $\mu(I)/\tau - 1 \leq |L(I)| \leq \mu(I)/\tau + 1$. Observe that $1/R \cdot \mu(I)/\tau \geq 1$, because $\mu(I) \geq R\tau$, thus

$$(1 - 1/R)\mu(I)/\tau \leq |L(I)| \leq (1 + 1/R)\mu(I)/\tau.$$

Using equalities $|L(M)| = \sum_{I \in \mathcal{I}} |L(I)|$ and $\mu(M) = \sum_{I \in \mathcal{I}} \mu(I)$, we get $(1 - 1/R)\mu(M)/\tau \leq |L(M)| \leq (1 + 1/R)\mu(M)/\tau$.

To prove the second inequality, observe that

$$L(M) \triangle L(M + \tilde{s}) \subset \bigcup_{I \in \mathcal{I}} (L(I) \triangle L(I + \tilde{s})).$$

21

For each $I \in \mathcal{I}$, $L(I) \triangle L(I+\tilde{s}) = L(I \triangle (I+\tilde{s}))$. The set $I \triangle (I+\tilde{s})$ is the union of two segments of length at most $\tilde{s}$, thus $I \triangle (I+\tilde{s})$ covers at most $2(\tilde{s}/\tau+1)$ points $\theta(i)$ $(i \in [k])$. Every interval covers at least R points. Thus, the size of the family $\mathcal{I}$ is at most $L(I)/R$. Therefore, $|L(M) \triangle L(M+\tilde{s})| \le |L(M)|/R \times (2\tilde{s}/\tau+2)$. The same argument shows $|L(M) \triangle L(M+\tilde{s})| \le |L(M+\tilde{s})|/R \times (2\tilde{s}/\tau+2)$. $\qquad\square$

Lemma 5.3. *Let* $s_{uv} = \lfloor k\tilde{s}_{uv} \rfloor$ $(s_{uv} \in [k]$, $\tilde{s}_{uv} \in S)$, $p \in \{1, \ldots, P\}$ *and let* $d = D_p(\tilde{s}_{uv})$ *be the p-th digit of* $\tilde{s}_{uv}$. *Then* $|L(M_{p,0})|, |L(M_{p,d})| \in [(1-1/R)/(R\tau), (1+1/R)/(R\tau)]$ *and*

$$|L(M_{p,0}) \triangle (L(M_{p,d}) - s_{uv})| \le 8 \min(|L(M_{p,0})|, |L(M_{p,d})|)/R.$$

Proof. We have $\mu(M_{p,d}) = R/T = 1/R$, thus by Lemma 5.2, $|L(M_{p,0})| \in [(1 - 1/R)/(R\tau), (1 + 1/R)/(R\tau)]$. Write,

$$
\begin{aligned}
|L(M_{p,0}) \triangle (L(M_{p,d}) - s_{uv})| &= |(L(M_{p,0}) + s_{uv}) \triangle L(M_{p,d})| = |L(M_{p,0} + \theta(s_{uv})) \triangle L(M_{p,d})| \\
&\le |L(M_{p,0} + \theta(s_{uv})) \triangle L(M_{p,0} + \tilde{s}_{uv})| + |L(M_{p,0} + \tilde{s}_{uv}) \triangle L(M_{p,d})|.
\end{aligned}
$$

Since $\tilde{s}_{uv} - \theta(s_{uv}) \in [0, \tau]$, by Lemma 5.2,

$$|L(M_{p,0} + \theta(s_{uv})) \triangle L(M_{p,0} + \tilde{s}_{uv})| \le |L(M_{p,0} + \theta(s_{uv}))|/R \cdot (2(\tilde{s}_{uv} - \theta(s_{uv}))/\tau + 2) \le 4|L(M_{p,0})|/R.$$

The p-th digit of $\tilde{s}_{uv}$ is $D_p(\tilde{s}_{uv}) = d$. Hence, the p-th digit of every number $\tilde{s}$ in $M_{p,0} + \tilde{s}_{uv}$ is in the range $[d, (R-1) + d + 1] = [d, R + d]$. Moreover, all numbers with p-th digit in the interval $[d+1, (R-1) + d]$ are covered by $M_{p,0} + \tilde{s}_{uv}$. Thus, $(M_{p,0} + \tilde{s}_{uv}) \triangle M_{p,d} \subset \{\tilde{s} : D_p(\tilde{s}) \in \{d, d+R\}\}$. The measure of the set $\{\tilde{s} : D_p(\tilde{s}) \in \{d, d+R\}\}$ is $2/T$. It is a union of segments of lengths $T^{-p} \ge \tau R$. By Lemma 5.2,

$$L((M_{p,0} + \tilde{s}_{uv}) \triangle M_{p,d}) \subset L(\{\tilde{s} : D_p(\tilde{s}) \in \{d, d+R\}\}) \le (1 + 1/R) \cdot 2/(T\tau) \le 4L(M_{p,0})/R.$$

$\qquad\square$

Lemma 5.4. *Let* $s_{uv} = \lfloor k\tilde{s}_{uv} \rfloor$ $(s_{uv} \in [k]$, $\tilde{s}_{uv} \in S)$, $p \in \{1, \ldots, P\}$, *and let* $d = D_p(\tilde{s}_{uv})$ *be the p-th digit of* $\tilde{s}_{uv}$. *Suppose that an edge* $(u, v) \in E_\varepsilon$ *is $1/32$–short, then*

$$\|\Psi_u(M_{p,0}) - \Psi_v(M_{p,d})\|^2 \le 1/8 \ \min(\|\Psi_u(M_{p,0})\|^2, \|\Psi_v(M_{p,D_p(\tilde{s})})\|^2); \tag{11}$$

and

$$\|\Psi_u'(M_{p,0}) - \Psi_v'(M_{p,d})\|^2 \le 1/8. \tag{12}$$

Proof. Write,

$$
\begin{aligned}
\|\Psi_u(M_{p,0}) - \Psi_v(M_{p,d})\|^2 &= \Big\| \sum_{i \in L(M_{p,0})} u_i - \sum_{i \in L(M_{p,d})} v_i \Big\|^2 = \Big\| \sum_{i \in L(M_{p,0}) \cap (L(M_{p,d}) - s_{uv})} (u_i - v_{i+s_{uv}}) + \\
&\quad \sum_{i \in L(M_{p,0}) \setminus (L(M_{p,d}) - s_{uv})} u_i - \sum_{i \in (L(M_{p,d}) - s_{uv}) \setminus L(M_{p,0})} v_{i+s_{uv}} \Big\|^2.
\end{aligned}
$$

Using ℓ_2^2 triangle inequalities $\langle u_i - v_{i+s_{uv}}, u_j - v_{j+s_{uv}} \rangle = -\langle u_i, v_{j+s_{uv}} \rangle - \langle v_{i+s_{uv}}, u_j \rangle \le 0$, $\langle u_i, -v_{j+s_{uv}} \rangle \le 0$ (for $i \ne j$) and then Lemma 5.3, we get (for sufficiently large R)

$$
\begin{aligned}
\|\Psi_u(M_{p,0}) - \Psi_v(M_{p,d})\|^2 &\le \sum_{i \in L(M_{p,0}) \cap (L(M_{p,d}) - s_{uv})} \|u_i - v_{i+s_{uv}}\|^2 + |L(M_{p,0}) \triangle (L(M_{p,d}) - s_{uv})| \\
&\le 1/16 \ |L(M_{p,0})| + 8|L(M_{p,0})|/R \le 1/8 \ |L(M_{p,0})| = \|\Psi_u(M_{p,0})\|^2/8.
\end{aligned}
$$

Similarly, $\|\Psi_u(M_{p,0}) - \Psi_v(M_{p,d})\|^2 \le \|\Psi_u(M_{p,d})\|^2/8$.

Inequality (12) immediately follows from inequality (11): let $\psi_u = \Psi_u(M_{p,0})$, $\psi_v = \Psi_v(M_{p,d})$ and assume $\|\psi_u\| \le \|\psi_v\|$, then $\| \psi_u/\|\psi_u\| - \psi_v/\|\psi_u\| \|^2 \le 1/8$. Vector $\psi_u/\|\psi_u\|$ has length 1, and vector $\psi_v/\|\psi_u\|$ has length at least 1, hence $\| \psi_u/\|\psi_u\| - \psi_v/\|\psi_v\| \|^2 \le \| \psi_u/\|\psi_u\| - \psi_v/\|\psi_u\| \|^2 \le 1/8$. $\qquad\square$

Observe, that vectors $\Psi_v'(M_{p,d'})$, $\Psi_v'(M_{p,d''})$ are orthogonal if $|d'' - d'| > R$, and thus

$$\|\Psi_u'(M_{p,d'}) - \Psi_u'(M_{p,d''})\|^2 = 2.$$

We now proceed the same way as in the proof of Theorem 4.4. We embed $O(n \log \log k)$ vectors $\Psi_u'(M_{p,d})$ ($u \in V$, $p \in P$, $d \in T$) in a net N of size $O(1)$ using a randomized mapping φ (see Lemma 3.7), so that for some small absolute constant β and every $u, v \in V$; $d', d'' \in T$; and $p \in \{1, \dots, P\}$,

$$\Pr((1 + \beta)^{-1}\|\Psi_u'(M_{p,d'}) - \Psi_v'(M_{p,d''})\|^2 - \beta \le \|\Phi(u,p,d') - \Phi(v,p,d'')\|^2$$
$$\le (1 + \beta)\|\Psi_u'(M_{p,d'}) - \Psi_v'(M_{p,d''})\|^2 + \beta) \ge 1 - \beta/T^2,$$

where $\Phi(u,p,d) = \varphi(\Psi_u'(M_{p,d}))$. We say that a pair $(u,p) \in V \times \{1, \dots, P\}$ is good if the following inequality holds: $\|\Phi(u,p,d') - \Phi(u,p,d'')\|^2 \ge 2 - 3\beta$ for all $d', d'' \in [T]$ such that $|d'' - d'| > R$. By Lemma 3.7, a pair (u,p) is good with probability at least $1 - \beta$. Then, for a fixed $1/32$–short edge $(u,v) \in E_\varepsilon$, the expected fraction of p's for which both pairs (u,p) and (v,p) are good and

$$\|\Phi(u,p,0) - \Phi(v,p,D_p(\tilde{s}_{uv}))\|^2 \le 1/4 \tag{13}$$

is at least $1 - 3\beta$.

Assume that $\gamma|E| = \gamma\varepsilon^{-1}|E_\varepsilon|$ edges in E_ε are $1/32$–short. We say that an edge $(u,v) \in E_\varepsilon$ is good with respect to the set $\{\Phi(u',p,d)_{u,p,d}\}$ if the following statement holds: "for at least $(1 - 6\beta)$ fraction of p's in $\{1, \dots, P\}$, the pairs (u,p) and (v,p) are good, and inequality (13) holds". By the Markov inequality, there exists a realization of random variables $\Phi(u,p,d)$ (random with respect to a random embedding in the net N) such that for at least $\gamma/2$ fraction of edges $(u,v) \in E_\varepsilon$ the previous statement holds.

Thus we have shown that for every linear unique game with $\gamma|E|$ $1/32$–short edges there always exists a witness—a collection of vectors $\{\Phi(u,p,d)\} \subset N$, such that at least $\gamma|E|/2$ edges in E_ε are good with respect to this collection (Remark: so far we have not used that the instance is semi-random). Now, we will prove that for a fixed witness, the probability that $\gamma|E|/2$ edges in E_ε is good in a semi-random unique game is exponentially small.

Fix an edge $(u,v) \in E_\varepsilon$ and compute the probability that it is good with respect to a fixed witness $\{\Phi(u,p,d)\} \subset N$. The probability that

$$\|\Phi(u,p,0) - \Phi(v,p,D_p(\tilde{s}_{uv}))\|^2 \le 1/4 \tag{14}$$

for a random $(\tilde{s}_{uv})$ is at most $1/R$ if pairs (u,p) and (v,p) are good, since among every R values $d \in \{d_0, d_0 + R, d_0 + 2R, \cdots\}$ there is at most one d satisfying

$$\|\Phi(u,p,0) - \Phi(v,p,d)\|^2 \le 1/4.$$

(Recall, that $\|\Phi(v,p,d') - \Phi(v,p,d'')\|^2 \ge 2 - 3\beta$ if $|d' - d''| > R$). By the Chernoff bound the probability that for $(1 - 6\beta)$ fraction of p's the inequality (14) is satisfied and (u,p), (v,p) are good is at most

$$e^{-P \ln((1 - 6\beta)R)} \le e^{-P} \le \ln^{-2} k.$$

Hence, the expected number of good edges in E_ε is at most $|E_\varepsilon|/\ln^2 k = \varepsilon|E|/\ln^2 k$, and the probability that $\gamma|E|/2$ edges in E_ε are good is at most

$$\exp\left(-\gamma|E|/2 \cdot \ln\left(\gamma\varepsilon^{-1}\ln^2 k/2\right)\right) \le \exp(-Cn/2 \cdot \ln(1/2\ln k)).$$

The total number of possible witnesses $\{\Phi(u, p, d)\} \subset N$ is $\exp(O(n \log \log k))$. So by the union bound (for sufficiently large absolute constant C) with probability $1 - \exp(-n) = 1 - o(1)$, less than $\gamma|E|$ edges in E_ε are $1/32$–short. $\qquad\square$

6 Random Initial Constraints

In this section, we consider the model, in which the initial set of constraints is chosen at random and other steps are controlled by the adversary. Specifically, in this model the adversary chooses the constraint graph $G = (V, E)$ and a "planted solution" $\{x_u\}$. Then for every edge $(u, v) \in E$, she randomly chooses a permutation (constraint) π_{uv} such that $\pi_{uv}(x_u) = x_v$ (each of $(k-1)!$ possible permutations is chosen with the same probability $1/(k-1)!$; choices for different edges are independent). Then the adversary chooses an arbitrary set E_ε of edges of size at most $\varepsilon|E|$ and (adversarially) changes the corresponding constraints: replaces constraint π_{uv} with a constraint π'_{uv} for $(u, v) \in E_\varepsilon$. Note that the obtained semi-random instance is $1 - \varepsilon$ satisfiable since the "planted solution" x_u satisfies constraints for edges in $E \setminus E_\varepsilon$. The analysis of this model is much simpler than the analysis of the other two models that we study.

Theorem 6.1. *There exists a polynomial-time algorithm that given $k \in \mathbb{N}$ $(k \ge k_0)$, $\varepsilon \in (0, 1)$, $\eta \in (c \log k/\sqrt{k}, 1)$ and a semi-random instance of unique games from the "random initial instance" model on graph $G = (V, E)$ with at least $C\eta^{-1}n \log k$ edges finds a solution of value $1 - O(\varepsilon + \eta/\log k)$ with probability $1 - o(1)$ (where c, C and k_0 are some absolute constants).*

Proof. We solve the standard SDP relaxation for the problem. Then we use a very simple rounding procedure. For every vertex u, if $\|u_i\|^2 > 1/2$ for some label i, we label u with i; otherwise, we label u with an arbitrary label (since for every two labels $i_1 \ne i_2$, $\|u_{i_1}\|^2 + \|u_{i_2}\|^2 \le 1$, we do not label any vertex with two labels).

We now show that our labeling satisfies a $1 - O(\varepsilon + \eta/\log k)$ fraction of constraints w.h.p. Without loss of generality, we assume that the planted solution is $x_u = 0$ for every $u \in V$. For every $t \in (1/2; 3/4)$, let $S_t = \{u : \|u_0\|^2 \ge t\}$. Let t_0 be the value of t that minimizes the size of the cut between S_t and $V \setminus S_t$; let $S = S_{t_0}$. Note that if $u \in S$ then we label vertex u with 0. Therefore, our labeling satisfies all constraints π_{uv} for edges (u, v) within S (but not necessarily constraints π'_{uv}). We conservatively assume that constraints for all edges from S to $V \setminus S$ and edges within $V \setminus S$ are not satisfied. We now estimate their number. First, we bound the number of edges leaving S. Note that since the unique game instance is $1 - \varepsilon$ satisfiable the cost of the SDP solution is at most ε. In particular,

$$\frac{1}{2} \sum_{(u,v)\in E} \left|\|u_0\|^2 - \|v_0\|^2\right| \le \frac{1}{2} \sum_{(u,v)\in E} \|u_0 - v_0\|^2 \le \varepsilon.$$

On the other hand, if we choose t uniformly at random from $(1/2; 3/4)$, then the probability that $u \in S_t$ and $v \notin S_t$ or $u \notin S_t$ and $v \in S_t$ is at most $4\left|\|u_0\|^2 - \|v_0\|^2\right|$ for every $(u, v) \in E$. Therefore, the expected size of the cut between S_t and $V \setminus S_t$ is at most $8\varepsilon|E|$. Hence the size of the cut between S and $V \setminus S$ is at most $8\varepsilon|E|$.

24

Now we estimate the number of edges within $V \setminus S$. We consider a new instance of Unique Games on G with the label set $\{1, \ldots, k-1\}$ and constraints $\pi^*_{uv} = \pi_{uv}|_{\{1,\ldots,k-1\}}$ (the restriction of π_{uv} to the set $\{1, \ldots, k-1\}$). Note that each π^*_{uv} is a permutation of $\{1, \ldots, k-1\}$ since $\pi_{uv}(0) = 0$. Moreover, each π^*_{uv} is a random permutation uniformly chosen among all permutations on $\{1, \ldots, k-1\}$. For each vertex u and label $i \in \{1, \ldots, k-1\}$, we define a vector u^*_i as follows. If $u \notin S$, we let $u^*_i = u_i$. Otherwise, we let $u^*_1 = e$ and $u^*_i = 0$ for $i > 1$, where e is a fixed unit vector orthogonal to all vectors v_j in the SDP solution.

We say that $\{u^*_i\}$ is a relaxed SDP solution if it satisfies all SDP conditions except possibly for the condition that $\sum_i \|u^*_i\|^2 = 1$ for every vertex u. We require instead that $1/4 \leq \sum_i \|u^*_i\|^2 \leq 1$. Note that the set of vectors $\{u^*_i\}$ is a relaxed SDP solution since for every $u \notin S$, $\sum_{i=1}^{k-1} \|u^*_i\|^2 = 1 - \|u_0\|^2 \geq 1 - t_0 \geq 1/4$; for every $u \in S$, $\sum_{i=1}^{k-1} \|u^*_i\|^2 = \|e\|^2 = 1$.

We now use a slightly modified version of Theorem 4.4.

Lemma 6.2. *Consider a unique game on a graph $G = (V, E)$ with at least $C\eta^{-1}n \log k$ edges with random set of constraints π^*_{uv}, where $\eta \in (c \log k/\sqrt{k}, 1)$ (where c and C are some absolute constants). Let $\{u^*_i\}$ be a relaxed SDP solution. Then, there are at most $O(\eta/\log k)|E|$ 1/64–short edges with probability $1 - o(1)$.*

The proof of the lemma almost exactly repeats the proof of Theorem 4.4 for $\varepsilon = 1$ (we only need to change inequalities (6) and (7) slightly).

We apply this lemma to the solution $\{u^*_i\}$. We get that there are at most $O(\eta/\log k)|E|$ 1/64–short edges. In particular, there are at most $O(\eta/\log k)|E|$ 1/64–short edges in $E(V \setminus S)$. Thus

$$\frac{1}{2} \sum_{(u,v) \in E} \sum_{i=1}^{k-1} \|u_i - v_{\pi_{uv}(i)}\|^2 \geq \frac{1}{2} \sum_{(u,v) \in E(V \setminus S)} \sum_{i=1}^{k-1} \|u_i - v_{\pi_{uv}(i)}\|^2 = \frac{1}{2} \sum_{(u,v) \in E(V \setminus S)} \sum_{i=1}^{k-1} \|u^*_i - v^*_{\pi^*_{uv}(i)}\|^2$$

$$\geq \frac{|E(V \setminus S)|}{64} - O(\eta/\log k)|E|.$$

However, the left hand side is at most $\varepsilon|E|$. Therefore, $E(V \setminus S) = O(\varepsilon + \eta/\log k)|E|$.

We conclude that the solution that our algorithm finds satisfies a $1 - O(\varepsilon + \eta/\log k)$ fraction of constraints π_{uv}. Since there are at most $\varepsilon|E|$ corrupted constraints, the solution also satisfies a $1 - O(\varepsilon + \eta/\log k) - \varepsilon = 1 - O(\varepsilon + \eta/\log k)$ fraction of corrupted constraints. $\square$

7 Distinguishing Between Semi-Random Unsatisfiable Games and Almost Satisfiable Games

In this section, we study the following question, Is it possible to distinguish between $(1 - \varepsilon)$ satisfiable semi-random games and $(1 - \delta)$ satisfiable (non-random) games if $\delta \ll \varepsilon$? This question is interesting only in the model where the corrupted constraints are chosen at random (i.e. step 4 is random), since in the other two semi-random models (when the initial constraints are random, and when the set of corrupted edges E_ε is random), the semi-random instance can be $1 - \delta$ satisfiable, therefore, the answer is trivially negative.

Specifically, we consider the following model. The adversary chooses a constraint graph G and a set of constraints π_{uv}. We do not require that this instance is completely satisfiable. Then she chooses a set of edges E_ε of size $\varepsilon|E|$. She replaces constraints for edges in E_ε with *random* constraints (each constraint is chosen uniformly at random among all $k!$ possible constraints).

We show that such semi-random instance can be distinguished w.h.p. from a $(1 - \delta)$ satisfiable instance if $\delta < c\varepsilon$ (where c is an absolute constant) if $|E| \geq Cn \max(\varepsilon^{-1}, \log k)$. To this end, we consider the

standard SDP relaxation for Unique Games. We prove that the SDP value of a semi-random instance is at least $c\varepsilon$; whereas, of course, the SDP value of a $(1 - \delta)$ satisfiable instance is at most δ.

Theorem 7.1. *Let $k \in \mathbb{N}$ ($k \geq k_0$) and $\varepsilon \in (0, 1]$. Consider a graph G with at least $Cn \max(\varepsilon^{-1}, \log k)$ edges, and a semi-random unique games instance $\mathcal{I}$ on G with $\varepsilon|E|$ randomly corrupted constraints (k_0 and C are absolute constants). Then the SDP value of $\mathcal{I}$ is at least $\varepsilon/32$ with probability $1 - o(1)$.*

Proof. We apply Theorem 4.4 to our instance of Unique Games, with $\gamma = \min(\varepsilon/2, 1/(2\log k))$. We get that at least half of all edges in E_ε are $1/16$–long w.h.p. The contribution of these edges to the sum $\frac{1}{2}\sum_{(u,v)}\sum_i \|u_i - v_{\pi_{uv}(i)}\|^2$ in the SDP objective function is at least $1/16 \times (|E_\varepsilon|/2) = \varepsilon|E|/32$. Therefore, the SDP value is at least $\varepsilon/32$. $\square$

References

[1] M. Alekhnovich. *More on Average Case vs Approximation Complexity.* In Proceedings of the 44st IEEE Symposium on Foundations of Computer Science: pp, 298–307, 2003.

[2] G. Andersson, L. Engebretsen, and J. Håstad. *A new way to use semidefinite programming with applications to linear equations mod p,* Journal of Algorithms, Vol. 39, 2001, pp. 162–204.

[3] S. Arora, B. Barak, and D. Steurer. *Subexponential Algorithms for Unique Games and Related problems.* In Proceedings of the 51st IEEE Symposium on Foundations of Computer Science, 2010.

[4] S. Arora, R. Impagliazzo, W. Matthews, and D. Stuerer. *Improved algorithms for unique games via divide and conquer,* In Electronic Colloquium on Computational Complexity, TR10-041, 2010.

[5] S. Arora, S. Khot, A. Kolla, D. Steurer, M. Tulsiani, and N. Vishnoi. *Unique games on expanding constraint graphs are easy.* In Proceedings of the 40th ACM Symposium on Theory of Computing, pp. 21–28, 2008.

[6] B. Barak, M. Hardt, I. Haviv, A. Rao, O. Regev and D. Steurer. *Rounding Parallel Repetitions of Unique Games,* In Proceedings of the 49th IEEE Symposium on Foundations of Computer Science, pp. 374–383, 2008.

[7] A. Blum and J. Spencer. *Coloring Random and Semi-Random k-Colorable Graphs,* J. Algorithms, vol. 19, no. 2, pp. 204–234, 1995.

[8] M. Charikar, K. Makarychev, and Y. Makarychev. *Near-Optimal Algorithms for Unique Games.* In Proceedings of the 38th ACM Symposium on Theory of Computing, pp. 205–214, 2006.

[9] E. Chlamtac, K. Makarychev, and Y. Makarychev. *How to Play Unique Games Using Embeddings.* In Proceedings of the 47th IEEE Symposium on Foundations of Computer Science, pp. 687–696, 2006.

[10] A. Gupta and K. Talwar. *Approximating Unique Games.* In Proceedings of the 17th ACM-SIAM Symposium on Discrete Algorithms, pp. 99–106, 2006.

[11] U. Feige. *Relations Between Average Case Complexity and Approximation Complexity* In Proceedings of the 34th Annual ACM Symposium on Theory of Computing, pp. 534–543, 2002.

[12] U. Feige and J. Kilian, *Heuristics for Semirandom Graph Problems*, Journal of Computing and System Sciences, vol. 63, pp. 639–671, 2001.

[13] U. Feige and R. Krauthgamer. *Finding and Certifying a Large Hidden Clique in a Semi-Random Graph* Random Structures and Algorithms, vol. 16(2), pp. 195–208, 2000.

[14] V. Guruswami, R. Manokaran, and P. Raghavendra. *Beating the Random Ordering is Hard: Inapproximability of Maximum Acyclic Subgraph.* In Proceedings of the 49th IEEE Symposium on Foundations of Computer Science, pp. 573–582, 2008.

[15] V. Guruswami and P. Raghavendra. *Constraint satisfaction over a non-boolean domain: Approximation algorithms and unique-games hardness.* In Proceedings of APPROX-RANDOM, 77–90, 2008.

[16] M. Jerrum. *Large Cliques Elude the Metropolis Process,* Random Structures and Algorithm, vol. 3 (4), pp. 347–359, 1992.

[17] S. Khot. *On the power of unique 2-prover 1-round games.* In Proceedings of the 34th ACM Symposium on Theory of Computing, pp. 767–775, 2002.

[18] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. *Optimal inapproximability results for MAX-CUT and other 2-variable CSPs?* ECCC Report TR05-101, 2005.

[19] S. Khot and O. Regev. *Vertex cover might be hard to approximate to within $2 - \varepsilon$.* In Proceedings of the 18th IEEE Annual Conference on Computational Complexity, 2003.

[20] S. Khot and N. Vishnoi. *The Unique Games Conjecture, Integrality Gap for Cut Problems and Embeddability of Negative Type Metrics into ℓ_1,* Foundations of Computer Science, pp. 53–62, 2005.

[21] A. Kolla. *Spectral Algorithms for Unique Games* In Proceedings of the Conference on Computational Complexity, pp. 122–130, 2010.

[22] K. Makarychev and Y. Makarychev. *How to Play Unique Games on Expanders,* In Proceedings of the eighth Workshop on Approximation and Online Algorithms, 2010.

[23] P. Raghavendra, *Optimal algorithms and inapproximability results for every CSP?,* Proceedings of the 40th Annual ACM Symposium on Theory of Computing, pp. 245–254, 2008.

[24] P. Raghavendra, D. Steurer, *How to Round Any CSP,* In Proceedings of the 50th IEEE Symposium on Foundations of Computer Science, pp. 586–594, 2009.

[25] A. Samorodnitsky and L. Trevisan. *Gowers uniformity, influence of variables, and PCPs.* In Proceedings of the 38th annual ACM symposium on Theory of computing, pp. 11–20, 2006.

[26] L. Trevisan. *Approximation Algorithms for Unique Games.* In Proceedings of the 46th IEEE Symposium on Foundations of Computer Science, pp. 197–205, 2005.