On the Hardness of Approximating Multicut and Sparsest-Cut*

Shuchi Chawla† Robert Krauthgamer‡ Ravi Kumar§ Yuval Rabani¶

D. Sivakumar‖

September 19, 2005

Abstract

We show that the Multicut, Sparsest-Cut, and Min-2CNF≡ Deletion problems are NP-hard to approximate within every constant factor, assuming the Unique Games Conjecture of Khot [STOC, 2002]. A quantitatively stronger version of the conjecture implies an inapproximability factor of $\Omega(\sqrt{\log \log n})$.

1 Introduction

In the Multicut problem the input is an undirected graph $G = (V, E)$ on $n = |V|$ vertices together with k pairs of vertices $\{s_i, t_i\}_{i=1}^k$, called *demand pairs*, and the goal is to find a minimum-size subset of the edges $M \subseteq E$ whose removal disconnects all the demand pairs, i.e., in the subgraph $(V, E \setminus M)$ every s_i is disconnected from its corresponding vertex t_i. In the weighted version of this problem, the input also specifies a positive cost $c(e)$ for each edge $e \in E$ and the goal is to find a multicut M whose total cost $c(M) = \sum_{e \in M} c(e)$ is minimal. This problem is known to be APX-hard [DJP+94].

We prove that if the Unique Games Conjecture of Khot [Kho02] is true, then for every constant $L > 0$ it is NP-hard to approximate Multicut within factor L. If a quantitatively stronger version of the conjecture is true, then Multicut is NP-hard to approximate within a factor of $\Omega(\sqrt{\log \log n})$.

Our methods also yield similar bounds for Sparsest-Cut and Min-2CNF≡ Deletion. The Sparsest-Cut problem has the same input as Multicut, but the goal is to find a subset of the edges $M \subseteq E$ that minimizes the ratio of $|M|$ (in the weighted version, the total cost of M) to the

*A preliminary version of this paper appeared in Proceedings of 20th Annual IEEE Conference on Computational Complexity (CCC 2005).

†Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA. This work is supported in part by an IBM PhD Fellowship, and by NSF grants CCR-0122581 and IIS-0121678, and part of it was done while the author was visiting IBM Almaden Research Center. Email: `shuchi@cs.cmu.edu`

‡IBM Almaden Research Center, 650 Harry Road, San Jose, CA 95120, USA. Email: `robi@almaden.ibm.com`

§Yahoo! Research, 701 First Avenue, Sunnyvale, CA 94089. This work was done while the author was at the IBM Almaden Research Center. Email: `ravikumar@yahoo-inc.com`

¶Computer Science Department, Technion—Israel Institute of Technology, Haifa 32000, Israel. Part of this work was done while the author was on sabbatical leave at Cornell University, while visiting IBM Almaden Research Center, and while visiting the Institute for Pure and Applied Mathematics at UCLA. Research at the Technion supported in part by ISF grant number 52/03 and BSF grant number 02-00282. Email: `rabani@cs.technion.ac.il`

‖Google, Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043. This work was done while the author was at the IBM Almaden Research Center. Email: `siva@google.com`

number of demand pairs that are disconnected in $(V, E \setminus M)$.[1] Since SPARSEST-CUT is not known to be APX-hard, our result gives the first indication that this problem might be hard to approximate. In the MIN-2CNF≡ DELETION problem the input is a weighted set of clauses on n variables, each clause of the form $(x \equiv y)$, where x and y are literals, and the goal is to find a Boolean assignment to the variables minimizing the total weight of unsatisfied clauses.[2] Our results immediately extend also to the CORRELATION CLUSTERING problem [BBC04, CGW03, DI03, EF03] of minimizing disagreements in a weighted graph, because the approximability of this problem is known to be equivalent to within constant factors to that of MULTICUT in weighted graphs [CGW03, EF03].

1.1 Known results on MULTICUT, SPARSEST-CUT, and MIN-2CNF≡ DELETION

MULTICUT and SPARSEST-CUT are fundamental combinatorial problems, with connections to multicommodity flow, edge expansion, and metric embeddings. Both problems can be approximated to within $O(\log k)$ factors through linear programming relaxations [LR88, GVY96, AR98, LLR95]. These bounds match the lower bounds on the integrality gaps of the corresponding relaxations up to constant factors [LR88, GVY96]. MIN-2CNF≡ DELETION can also be approximated to within an $O(\log n)$ factor, as implied by the results of Klein et al [KARR90] (see also [Vaz01, Section 20.4]), who give an approximation-preserving reduction from this problem to MULTICUT. Recently, improved approximation algorithms for the SPARSEST-CUT problem have been developed using a semidefinite programming relaxation [ARV04, CGR05, ALN05]. This started with the groundbreaking $O(\sqrt{\log n})$-approximation of Arora et al [ARV04] for the uniform demands case, and the best approximation factor currently known for general demands is $O(\sqrt{\log k}\log\log k)$ [ALN05]. The obvious modification of the semidefinite program used for SPARSEST-CUT to solve MULTICUT was recently shown to have an integrality ratio of $\Omega(\log k)$ [ACMM05, Section 6], which matches, up to constant factors, the approximation factor and integrality gap of previously analyzed linear programming relaxations for this problem.

On the hardness side, it is known that MULTICUT is APX-hard [DJP$^+$94], i.e., there exists a constant $c > 1$, such that it is NP-hard to approximate MULTICUT to within a factor smaller than c. It should be noted that this hardness of approximation holds even for $k = 3$, and that the value of c is not specified therein, but it is certainly smaller than 2. The MIN-2CNF≡ DELETION problem is also known to be APX-hard, as follows, e.g., from the hardness of approximating linear equations modulo 2 [Hås01].

Assuming the Unique Games Conjecture, Khot [Kho02, Theorem 3] essentially obtained an arbitrarily large constant-factor hardness for MIN-2CNF≡ DELETION, and this implies, using the aforementioned reduction of [KARR90], a similar hardness factor for MULTICUT. These results are not explicitly noted in [Kho02], and are weaker than our results in several respects. First, our quantitative bounds are better; thus if a stronger, yet almost as plausible, version of this conjecture is true, then our lower bound on the approximation factor improves to $L = \Omega(\sqrt{\log\log n})$, compared with the roughly $\Omega((\log\log n)^{1/4})$ hardness that can be inferred from [Kho02]; this can be viewed as progress towards proving tight inapproximability results for MULTICUT. Second, by qualitatively strengthening our MULTICUT result to a bicriteria version of the problem, we extend our hardness results to the SPARSEST-CUT problem. It is unclear whether Khot's reduction similarly leads to a hardness result for SPARSEST-CUT. Finally, our proof is simpler (both the reduction and its

[1] In general, the demand pairs may have positive weights (demands), but for our purpose of inapproximability results, it clearly suffices to consider the more restricted definition above. Our hardness results for SPARSEST-CUT do not apply to the special case of uniform-demand, in which every pair of vertices forms a demand pair.

[2] Note that the constraints in MIN-2CNF≡ DELETION are restricted to equality (and effectively non-equality) constraints.

analysis), and makes direct connections to cuts (in a hypercube), and thus may prove useful in further investigation of such questions.

For Sparsest-Cut, no hardness of approximation result was previously known. Independently of our work, Khot and Vishnoi [KV05] have recently used a different construction to show an arbitrarily large constant factor hardness for Sparsest-Cut assuming the Unique Games Conjecture; their hardness factor could, in principle, be pushed to $(\log \log n)^{1/4-\delta}$, for every fixed $\delta > 0$, assuming a stronger quantitative version of the conjecture. Additionally, they prove an integrality ratio lower bound of $(\log \log n)^{1/4-\delta}$, for the semidefinite program relaxations used in the recent approximation algorithms for Sparsest-Cut.

1.2 The Unique Games Conjecture

Unique 2-prover game is the following problem. The input is a bipartite graph $G_Q = (Q, E_Q)$, where each side $p = 1, 2$ contains $n = |Q|/2$ vertices denoted $q_1^p, \cdots, q_n^p$, and represents n possible questions to prover p. In addition, the input contains for each edge $(q_i^1, q_j^2) \in E_Q$ a non-negative weight $w(q_i^1, q_j^2)$. These edges will be called *question edges*, to distinguish them from edges in the Multicut instance. Each question to a prover is associated with a set of d distinct answers, denoted by $[d] = \{1, \dots, d\}$. The input also contains, for every edge $(q_i^1, q_j^2) \in E_Q$, a bijection $b_{ij} : [d] \to [d]$, which maps every answer of question q_i^1 to a distinct answer for q_j^2.

A *solution* A to the 2-prover game consists of an answer $A_i^p \in [d]$ for each question q_i^p (i.e., a sequence $\{A_i^p\}$ over all $p \in [2]$ and $i \in [n]$). The solution is said to satisfy an edge $(q_i^1, q_j^2) \in E_Q$ if the answers A_i^1 and A_j^2 agree, i.e., $A_j^2 = b_{ij}(A_i^1)$. We assume that the total weight of all the edges in E_Q is 1 (by normalization). The *value of a solution* is the total weight of all the edges satisfied by the solution. The *value of the game* is the maximum value achievable by any solution to the game.

Conjecture 1.1 (Unique Games [Kho02]). *For every fixed $\eta, \delta > 0$ there exists $d = d(\eta, \delta)$ such that it is NP-hard to determine whether a unique 2-prover game with answer set size d has value at least $1 - \eta$ or at most δ.*

We will also consider stronger versions of the Unique Games Conjecture in which η, δ, and d are functions of n. Specifically, we will consider versions with $\eta \le O(1/\sqrt{\log \log n})$, $\delta \le 1/(\log n)^{\Omega(\eta)}$, and $d = d(\eta, \delta) \le O(\log n)$. The reason for the latter upper bound is that our construction size is exponential in d.

Plausibility of the conjecture and stronger versions of it. The Unique Games Conjecture has been used to show optimal inapproximability results for Vertex Cover [KR03] and Max-Cut [KKMO04, MOO05]. Proving the conjecture using current techniques appears quite hard. In particular, the asserted NP-hardness is much stronger than what we can obtain via standard constructions using the PCP theorem [AS98, ALM+98] and the parallel repetition theorem [Raz98], two deep results in computational complexity.

Although the conjecture seems difficult to prove in general, some special cases are well-understood. In particular, if at all the Unique Games Conjecture is true (and assuming P $\neq$ NP), then necessarily $d \ge \max\{1/\eta^{1/10}, 1/\delta\}$. This follows from a rounding procedure for a semidefinite programming relaxation presented in [Kho02]. Recently, Trevisan [Tre05] used a different semidefinite programming relaxation to show that a stronger version of the Unique Games Conjecture can only be true if $\eta(n) \ge \Omega(1/\log n)$ (assuming P $\neq$ NP). Very recently, we were informed [CMM05] of a different lower bound on η expressed in terms of δ and d. Feige and Reichman [FR04] showed that for every constant $L > 0$ there exists a constant $\delta > 0$, such that it is NP-hard to distinguish whether a unique 2-prover game (with $d = d(L, \delta)$) has value at least $L\delta$ or at most δ; this result falls short of the Unique Games Conjecture in that $L\delta$ is bounded away from 1. Our $\Omega(\sqrt{\log \log n})$ hardness result (Corollary 1.4 below) requires $\eta \le O(1/\sqrt{\log \log n})$, $\delta \le 1/(\log n)^{\Omega(\eta)}$, and $d = d(\eta, \delta) \le O(\log n)$, which is not excluded by known results. In fact, the recent integrality ratio of [KV05] for Unique Games Conjecture may indicate that the unique 2-prover game is hard for $d = \Theta(\log n)$, $\delta = d^{-\Omega(\eta)}$, and every $\Omega(\frac{\log d}{d}) \le \eta < 1/2$.

1.3 Our results

We prove the following hardness of approximation for Multicut, Sparsest-Cut, and Min-2CNF≡ Deletion based on the Unique Games Conjecture.

Theorem 1.2. *Suppose that for $\eta = \eta(n)$, $\delta = \delta(n)$, and $d = d(\eta, \delta) \le O(\log n)$, it is NP-hard to determine whether a unique 2-prover game with $|Q| = 2n$ vertices and answer set size d has value at least $1 - \eta(n)$ or at most $\delta(n)$. Then there exists $L(n) = \Omega(\min\{\frac{1}{\eta(n^{\Omega(1)})}, \log \frac{1}{\delta(n^{\Omega(1)})}\})$ such that it is NP-hard to approximate* Multicut, Sparsest-Cut, *and* Min-2CNF≡ Deletion *to within factor $L(n)$.*

This theorem immediately implies the following two specific hardness results.

Corollary 1.3. *The Unique Games Conjecture implies that, for every constant $L > 0$, it is NP-hard to approximate* Multicut, Sparsest-Cut, *and* Min-2CNF≡ Deletion *to within factor L.*

Corollary 1.4. *A stronger version of the Unique Games Conjecture in which $\eta \le O(1/\sqrt{\log \log n})$, $\delta \le 1/(\log n)^{\Omega(\eta)}$, and $d = d(\eta, \delta) \le O(\log n)$, implies that for some fixed $c > 0$, it is NP-hard to approximate* Multicut, Sparsest-Cut, *and* Min-2CNF≡ Deletion *to within factor $c\sqrt{\log \log n}$.*

For Sparsest-Cut our hardness results hold only for the search version (in which the algorithm needs to produce a cutset and not only its value), since our proof employs a Cook reduction. As noted before, a similar (but slightly weaker) bound is proved in [KV05].

Remark. The conference version of our paper [CKK+05] presented a weaker bound than that of Theorem 1.2, with the dependence on η being $\Omega(\log 1/\eta)$. The reduction is exactly the same, but the analysis is different (and perhaps simpler), as the current version uses Friedgut's Junta Theorem [Fri98] rather than a theorem of Kahn, Kalai, and Linial [KKL88]. This improvement was motivated, in part, by the integrality ratio of [KV05] for unique games, which suggests a significant asymmetry between $\eta(n)$ and $\delta(n)$.

1.4 Preliminaries

Regular Unique Games. A unique 2-prover game is called *regular* if the total weight of question edges incident at any single vertex is the same, i.e., $1/n$, for every vertex in Q. We now show that we can assume without loss of generality that the graph in the Unique Games Conjecture is regular. For simplicity, we state this only for fixed η and δ. A similar result holds when they depend on n, because we increase the input size by no more than a polynomial factor, and increase η and δ by no more than a constant factor.

Lemma 1.5. *The Unique Games Conjecture implies that for every fixed $\eta, \delta > 0$, there exists $d = d(\eta, \delta)$ such that it is NP-hard to decide if a regular unique 2-prover game has value at least $1 - \eta$ or at most δ.*

The proof is given in Appendix A, and is based on an argument of Khot and Regev [KR03, Lemma 3.3]. The dependence of d on η and δ is important for our purposes. We thus point out that this argument does not change $d = d(\eta, \delta)$, and increases the size of the instance by at most a polynomial factor in n. This is acceptable in the setting of Theorem 1.2, since the requirement $d = d(\eta, \delta) \leq O(\log n)$ is maintained and only the unspecified constants therein are affected.

Bicriteria Multicut. Our proof for the hardness of approximating Sparsest-Cut relies on a generalization of Multicut, where the solution M is required to cut only a certain fraction of the demand pairs. For a given graph $G = (V, E)$, a subset of the edges $M \subseteq E$ will be called a *cutset* of the graph. A cutset whose removal disconnects all the demand pairs is a *multicut*.

An algorithm is called an (α, β)-bicriteria approximation for Multicut if, for every input instance, the algorithm outputs a cutset M that disconnects at least α fraction of the demands and has cost at most β times that of the optimum multicut. In other words, if M^* is a least cost cutset that disconnects all the k demand pairs, then M disconnects at least αk demand pairs and $c(M) \leq \beta \cdot c(M^*)$.

Hypercubes, dimension cuts, and antipodal vertices. As usual, the *d-dimensional hypercube* (in short a *d-cube*) is the graph $C = (V_C, E_C)$ with the vertex set $V_C = \{0,1\}^d$, and an edge $(u, v) \in E_C$ for every two vertices $u, v \in \{0,1\}^d$ that differ in exactly one dimension (coordinate). An edge (u, v) is called a *dimension-a edge*, for $a \in [d]$, if u and v differ in dimension a, i.e., $u \oplus v = 1_a$ where 1_a is a unit vector along dimension a. The set of all the dimension-a edges in the hypercube is called the *dimension-a cut* in the hypercube; a *dimension cut* is a dimension-a cut for some dimension a. The *antipode* of a vertex u is the (unique) vertex $\overline{u}$ all of whose coordinates are different from those of u, i.e., $\overline{u} = u \oplus \mathbf{1}$ where $\mathbf{1}$ is the vector with 1 in every coordinate. Notice that v is the antipode of u if and only if u is the antipode of v; thus, $\langle u, \overline{u} \rangle$ form an *antipodal pair*. The following simple fact will be key in our proof.

Fact 1.6. *In every hypercube, a single dimension cut disconnects every antipodal pair.*

Organization. In Section 2 we prove the part of Theorem 1.2 regarding the Multicut problem; our proof will actually hold for bicriteria approximation for Multicut. We will then show in Section 3 that this stronger result yields a similar hardness of approximation for Sparsest-Cut. Finally, in Section 4, we modify the reduction to obtain a hardness of approximation for Min-2CNF≡ Deletion.

2 Hardness of bicriteria approximation for Multicut

In this section we prove the part of Theorem 1.2 regarding the Multicut problem, namely, that the Unique Games Conjecture implies that it is NP-hard to approximate Multicut within a certain factor L. Our proof will actually show a stronger result—for every $\alpha \geq 7/8$ it is NP-hard to distinguish between whether there is a multicut of cost less than $n2^{d+1}$ (the YES instance) or whether every cutset that disconnects at least αk demand pairs has cost at least $n2^{d+1}L$ (the NO instance). This implies that it is NP-hard to obtain an (α, L)-bicriteria approximation for Multicut.

We start by describing a reduction from unique 2-prover game to Multicut (Section 2.1), and then proceed to analyze the YES instance (Section 2.2) and the NO instance (Sections 2.3 and 2.4). Finally, we discuss the gap that is created for a bicriteria approximation of Multicut (Section 2.5).

2.1 The reduction

Given a unique 2-prover game instance $G_Q = (Q, E_Q)$ with $n = |Q|/2$ and the corresponding edge weights $w(e)$ and bijections $b_{ij} : [d] \to [d]$, we construct a MULTICUT instance $G = (V, E)$ with demand pairs, as follows. For every vertex (i.e., question) $q_i^p \in Q$, construct a d-dimensional hypercube C_j^p; the dimensions in this cube correspond to answers for the question q_j^p.[3] For each of the $2n$ hypercubes, we let the edges inside the hypercube have cost 1, and call them *hypercube edges*.

For each question edge $(q_i^1, q_j^2) \in E_Q$, we extend b_{ij} (in the obvious way) to a bijection from the vertices of C_i^1 to the vertices of C_j^2, and denote the resulting bijection by $b'_{ij} : \{0,1\}^d \to \{0,1\}^d$. Formally, for every $u \in \{0,1\}^d$ (vertex in C_i^1) and every $a \in [d]$, the a-th coordinate of $b'_{ij}(u)$ is given by $(b'_{ij}(u))_a = u_{b_{ij}^{-1}(a)}$. Then, we connect every vertex $u \in C_i^1$ to the corresponding vertex $b'_{ij}(u) \in C_j^2$ using an edge of cost $w_{ij}\Lambda$, where $\Lambda = n/\eta$ is a scaling factor. These edges are called *cross edges*.

Denote the resulting graph by $G = (V, E)$. Notice that V is simply the union of the vertex sets of the hypercubes C_i^p, for all $p \in [2]$ and $i \in [n]$, and that the edge set E contains two types of edges, hypercube edges and cross edges.

To complete the reduction, it remains to define the demand pairs. For a vertex $u \in V$, the *antipode* of u in G, denoted $\overline{u}$, is defined to be the vertex antipodal to u in the hypercube C_i^p that contains u. The set D of demand pairs then contains every pair of antipodal vertices in G, and hence $k = |D| = n2^{d-1}$. Note that every vertex of G belongs to exactly one demand pair.

2.2 The YES instance

Lemma 2.1. *If there is a solution A for the unique 2-prover game G_Q such that the total weight of the satisfied questions is at least $1 - \eta$, then there exists a multicut $M \subseteq E$ for the* MULTICUT *instance G such that $c(M) \le 2^{d+1}n$.*

Proof. Let A be such a solution for G_Q. Construct M by taking the following edges. For every question $q_i^p \in Q$ and the corresponding answer A_i^p (of prover p), take the dimension-A_i^p cut in cube C_i^p. In addition, for every edge $(q_i^1, q_j^2) \in E_Q$ that the solution A does not satisfy, take all the cross edges between the corresponding cubes C_i^1 and C_j^2.

We first claim that removing M from G disconnects all the demand pairs. To see this, we define a Boolean function $f : V \to \{0,1\}$ on the graph vertices. For every cube C_i^p, consider the dimension-A_i^p cut; it disconnects the cube into two connected components, one containing the all zeros vector $\mathbf{0}$ and one containing the all ones vector $\mathbf{1}$. For every $v \in C_i^p$, let $f(v) = 0$ if v is in the same side as $\mathbf{0}$, and $f(v) = 1$ otherwise. This is exactly the A_i^p-th bit in v, i.e., $f(v) = v_{A_i^p}$. Now consider any demand pair $(v, \overline{v})$, and note that $f(v) = 1 - f(\overline{v})$. We will show below that every edge $(u, v) \notin M$ satisfies the property $f(u) = f(v)$. This will clearly complete the proof of the claim.

Consider first a hypercube edge (u, v) in C_i^p that is not a dimension-A_i^p edge. Then $f(u) = u_{A_i^p} = v_{A_i^p} = f(v)$, by the definition of f. Next consider a cross edge $(u, v) \notin M$. Then this edge lies between cubes C_i^1 and C_j^2, such that the question edge (q_i^1, q_j^2) satisfied by the unique 2-prover game solution A. Therefore, $b_{ij}(A_i^1) = A_j^2$. Then, $f(u) = u_{A_i^1} = v_{b_{ij}(A_i^1)} = v_{A_j^2} = f(v)$.

[3]This is a standard technique in PCP constructions for graph optimization problems. A hypercube can be interpreted as a "long code" [BGS98], and a dimension cut is the encoding of an answer in the 2-prover game.

Finally, we bound the cost of the solution. Let S be the set of question edges not satisfied by the solution A. The total cost of the multicut solution is thus

$$c(M) = 2n\,2^{d-1} + 2^d \Lambda \sum_{(q_i^1, q_j^2) \in S} w_{ij} \leq 2^d n + 2^d \Lambda \eta = 2^{d+1} n.$$

□

2.3 Hypercube cuts, Boolean functions, influences, and juntas

We will analyze the NO instance shortly, but first we set up some notation and present a few technical lemmas regarding cuts in hypercubes. In particular, we present Theorem 2.2, which will have a crucial role in the sequel.

Recall that the dimensions of the hypercubes in the multicut instance corresponds to answers to the 2-prover game. Therefore, we would like to determine which dimensions are the most significant participants in a cut on the cube, as follows. Let $C = (V_C, E_C)$ be a d-dimensional hypercube. It will be useful to represent cuts on the hypercube C as functions $f : V_C \to \mathbb{Z}$. Such a function f corresponds to a partition of V_C into sets $\{f^{-1}(r) \mid r \in f(V_C)\}$, which in turn corresponds to the cutset $\{(u,v) \in E_C \mid f(u) \neq f(v)\}$. Notice that f can be described as a function on d Boolean variables (corresponding to the dimensions of the hypercube), where the dimension $a \in [d]$ corresponds to the a-th variable. The *influence of a dimension (variable)* $a \in [d]$ on the function f, denoted I_a^f, is defined as the fraction of the dimension a-edges $(u,v) \in E_C$ for which $f(u) \neq f(v)$. In other words, $I_a^f = \Pr_{u \in V_C}[f(u) \neq f(u \oplus 1_a)]$ where 1_a is a unit vector along dimension a. The *total influence* (also called average sensitivity) of f is $\sum_{a \in [d]} I_a^f$. We say that the function f is a *k-junta* if there exists a subset $J \subseteq [d]$, $|J| \leq k$, such that for every variable (dimension) $a \notin J$ and for every $u \in V_C$, we have $f(u) = f(u \oplus 1_a)$. In other words, f depends on at most k variables, and the remaining variables have zero influence. Two functions f and f' are said to be *ε-close* if $\Pr_{u \in V_C}[f(u) \neq f'(u)] \leq \varepsilon$.

An important special case is that of Boolean functions, i.e., $g : V_C \to \{0,1\}$, which corresponds to a bipartition of V_C. The *balance* of a Boolean function g is defined as $\min\{\frac{|g^{-1}(0)|}{|V_C|}, \frac{|g^{-1}(1)|}{|V_C|}\}$, i.e., the minimum of $\Pr_{u \in V_C}[g(u) = 0]$ and $\Pr_{u \in V_C}[g(u) = 1]$.

The next theorem, due to Friedgut [Fri98], asserts that every function of low total influence is close to a junta. We will later use it to determine a set of dimensions that are the most significant participants in a low-cost cutset.

Theorem 2.2 (Friedgut's Junta Theorem). *Let g be a Boolean function defined on a hypercube and fix $\varepsilon > 0$. Then g is ε-close to a Boolean function h defined on the same cube and depending on only $2^{O(T/\varepsilon)}$ variables, where $T = \sum_{a \in [d]} I_a^g$ is the total influence of g.*

2.4 The NO instance

Lemma 2.3. *There exists $L = \Omega(\min\{\eta^{-1}, \log \delta^{-1}\})$ such that if the* Multicut *instance G has a cutset of cost at most $2^d nL$ whose removal disconnects $\alpha \geq 7/8$ fraction of the demand pairs, then there is a solution A for the unique 2-prover game G_Q whose value is larger than δ.*

Proof. Let $L = \min\{c/\eta, c \log(1/\delta)\}$ for a constant $c > 0$ to be determined later, and let $M \subseteq E$ be a cutset of cost $c(M) \leq 2^d nL$ whose removal disconnects $\alpha \geq 7/8$ fraction of the demand pairs. Using M, we will construct for the unique 2-prover game G_Q a randomized solution A whose expected value is larger than δ, thereby proving the existence of a solution of value larger than δ. The randomized solution A (i.e., a strategy for the two provers) is defined as follows. Label each connected component of $G \setminus M$ as either 0 or 1 independently at random with equal probabilities, and define a Boolean function $f : V \to \{0,1\}$ by letting $f(v)$ be the label of the connected component of $v \in V$. Next, for each vertex (question) $q_i^p \in Q$ consider the restriction of f to the hypercube $C_i^p \subset V$, denoted $f_{|C_i^p}$, and apply to it Theorem 2.2 (Friedgut's Junta Theorem) with $\varepsilon = 1/64$, to obtain a subset of the variables (dimensions) $J_i^p \subseteq [d]$; the idea is that for many hypercubes we obtain $|J_i^p| \leq 2^{O(L/\varepsilon)}$. Finally, choose the answer A_i^p uniformly at random from J_i^p, independently of all other events.

We proceed to analyze the expected value of this randomized solution A. Recall that the value of a solution is equal to the probability that, for a question edge (q_i^1, q_j^2) chosen at random with probability proportional to its weight, we have $a_j^2 = b_{ij}(a_i^1)$. Notice that although q_i^1 and q_j^2 are correlated, each one is uniformly distributed because Q is regular. Without loss of generality, we assume removing M disconnects at least as many demand pairs inside the cubes $\{C_l^1\}_{l\in[n]}$ as inside the cubes $\{C_l^2\}_{l\in[n]}$. Now we claim that with a constant probability over the choice of a question edge, the cut M has a low cost over edges incident on the corresponding hypercubes, and disconnects many demand pairs in the hypercubes. In other words, the quality of the cut locally is nearly as good as the quality of the cut globally. In particular, we upper bound the probability of the following four "bad" events (for a random question edge (q_i^1, q_j^2)):

$\mathcal{E}_1$ = fewer than 1/8-fraction of the vertices $v \in C_i^1$ satisfy $f(v) \neq f(\overline{v})$.

$\mathcal{E}_2$ = M contains more than $2^{d+4}L$ hypercube edges in C_i^1.

$\mathcal{E}_3$ = M contains more than $2^{d+4}L$ hypercube edges in C_j^2.

$\mathcal{E}_4$ = M contains more than $2^{d+4}\eta L$ cross edges between C_i^1 and C_j^2.

First, by our assumption above, removing M disconnects at least $\alpha \geq 7/8$ fraction of the demand pairs inside the cubes $\{C_l^1\}_{l\in[n]}$. Thus, the expected fraction of demand pairs $(v, \overline{v})$ in C_i^1 that are not disconnected in $G \setminus M$ (and thus clearly $f(v) = f(\overline{v})$) is at most 1/8. In addition, the expected fraction of demand pairs $(v, \overline{v})$ in C_i^1 that are disconnected in $G \setminus M$ and satisfy $f(v) = f(\overline{v})$, is at most 1/2, because different connected components of $G \setminus M$ are labeled independently. Thus, the expected fraction of vertices $v \in C_i^1$ for which $f(v) = f(\overline{v})$ is at most 5/8, and by Markov's inequality, $\Pr[\mathcal{E}_1] \leq 5/7$. Next, the cutset M contains at most $2^d nL$ hypercube edges, thus the expected number of edges in $C_i^1 \cup C_j^2$ that are contained in M is at most $2^d L$, and by Markov's inequality $\Pr[\mathcal{E}_2 \cup \mathcal{E}_3] \leq 1/16$. Finally, $\Pr[\mathcal{E}_4] \leq 1/16$, as otherwise the total cost of M along the cross edges corresponding to this event is more than $1/16 \cdot (2^{d+4}\eta L) \cdot \Lambda = 2^d nL \geq c(M)$. Taking a union bound, we upper bound the probability that any of the bad events occurs by

$$\Pr[\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3 \cup \mathcal{E}_4] \leq \frac{5}{7} + \frac{2}{16} < \frac{6}{7}.$$

In order to lower bound the expected value of the randomized solution A, we would like to show that if none of the above bad events happens, then the two sets of dimensions J_i^1 and J_j^2 obtained using Friedgut's Junta Theorem are relatively small, and further they are in "weak agreement", and these two properties will immediately imply that the randomized solution A satisfies $\Pr[b_{ij}(A_i^1) = A_j^2] > \delta$. Observe that if $(u,v) \in E$ and $f(u) \neq f(v)$, then $(u,v) \in M$. If the event $\mathcal{E}_2$ does not occur, then the total influence of $f_{|C_i^1}$ is at most $8L$, and thus $|J_i^1| \leq 2^{O(L/\varepsilon)}$. Similarly, if the event $\mathcal{E}_3$ does not occur, then the total influence of $f_{|C_j^2}$ is at most $8L$, and thus $|J_j^2| \leq 2^{O(L/\varepsilon)}$. In addition, if the event $\mathcal{E}_1$ does not occur, then the balance of $f_{|C_i^1}$ is at least 1/16.

We now claim that if none of the above bad events happens then there is $a \in J_i^1$ for which $b_{ij}(a) \in J_j^2$. Indeed, assume towards contradiction that $J_i^1 \cap b_{ij}^{-1}(J_j^2) = \emptyset$. Then by construction there is a Boolean function $g_i^1 : C_i^1 \to \{0,1\}$ that is ε-close to $f_{|C_i^1}$ and depends on only variables in J_i^1. Clearly, the balance of g_i^1 is close to that of $f_{|C_i^1}$, namely, at least $1/16 - \varepsilon$. Similarly, there is a Boolean function $g_j^2 : C_j^2 \to \{0,1\}$ that is ε-close to $f_{|C_j^2}$ and depends on only variables in J_j^2. We can relate these two functions via $b'_{ij} : C_1^i \to C_j^2$, namely by considering $h : C_i^1 \to \{0,1\}$ given by $h(v) = g_j^2(b'_{ij}(v))$.

Notice that h is ε-close to $f_{|C_j^2} \circ b'_{ij}$, and that it depends only on variables in $b_{ij}^{-1}(J_j^2)$. Therefore g_i^1 and h depend on disjoint sets of variables. It follows that $\Pr_{v \in C_i^1}[g_i^1(v) \neq h(v)] \geq 1/16 - \varepsilon$, because if we condition on the value of the variables in $b_{ij}^{-1}(J_j^2)$ we get that $h(v)$ is determined, but this does not affect the distribution of $g_i^1(v)$, which still attains each value (0 or 1) with probability at least $1/16 - \varepsilon$. Consequently, g_i^1 and $h = g_j^2 \circ b_{ij}$ are not $(1/16 - \varepsilon)$-close.

On the other hand, the event $\mathcal{E}_4$ not occurring implies that at most $16\eta L$ vertices $v \in C_i^1$ satisfy $f(v) \neq f(b_{ij}(v))$. In other words, $f_{|C_i^1}$ is $(16\eta L)$-close to $f_{|C_j^2} \circ b_{ij}$. The former is ε-close to g_i^1 while the latter is ε-close to $g_j^2 \circ b_{ij}$ (because b_{ij} is a bijection on the variables), and by the triangle inequality we get that g_i^1 and $h = g_j^2 \circ b_{ij}$ are $2\varepsilon + 16\eta L$ close. If $c > 0$ is sufficiently small, $2\varepsilon + 16\eta L < 1/16 - \varepsilon$, which yields a contradiction.

Using the above claim we get that for a random question edge,

$$\begin{aligned}
\Pr[A_j^2 = b_{ij}(A_i^1)] &\geq \Pr[A_i^1 \in J_i^1 \cap b_{ij}^{-1}(J_j^2), A_j^2 = b_{ij}(A_i^1)] \\
&\geq \frac{1}{7} \cdot 2^{-O(L/\varepsilon)} \cdot 2^{-O(L/\varepsilon)} \\
&= 2^{-O(L)}.
\end{aligned}$$

We conclude that the expected value of the randomized solution A is $\Pr[A_j^2 = b_{ij}(A_i^1)] \geq 2^{-O(L)} > \delta$, where the last inequality holds if $c > 0$ is sufficiently small, and this completes the proof of Lemma 2.3. □

2.5 Putting it all together

The above reduction from unique 2-prover game to MULTICUT produces a gap of $L(n) = \Omega(\min\{\frac{1}{\eta(n)}, \log \frac{1}{\delta(n)}\})$. We assumed $d(\eta, \delta) \leq O(\log n)$, and thus the resulting MULTICUT instance G has size $N = (n2^d)^{O(1)} = n^{\Theta(1)}$. It follows that in terms of the instance size N, the gap is

$$L(N) = \Omega(\min\{\tfrac{1}{\eta(N^{\Theta(1)})}, \log \tfrac{1}{\delta(N^{\Theta(1)})}\}).$$

This completes the proof of the part of Theorem 1.2 regarding the MULTICUT problem, namely, that the Unique Games Conjecture implies that it is NP-hard to approximate MULTICUT within the above factor $L(N)$. In fact, the above proof shows that it is NP-hard to obtain even a $(7/8, L(N))$-bicriteria approximation.

Note that the number of demand pairs is $k = n2^{d-1} = n^{\Theta(1)}$, and thus the hardness of approximation factor is similar when expressed in terms of k as well. Note also that all edge weights in the MULTICUT instance constructed above are bounded by a polynomial in the size of the graph. Therefore, via a standard reduction, a similar hardness result holds for the unweighted MULTICUT problem as well.

3 Hardness of approximating Sparsest-Cut

In this section we prove the part of Theorem 1.2 regarding the Sparsest-Cut problem. The proof follows immediately from the next lemma in conjunction with the hardness of bicriteria approximation of Multicut (from the Section 2).

Lemma 3.1. *Let $0 < \alpha < 1$ be a constant. If there exists a polynomial-time algorithm for* Sparsest-Cut *that produces a cut whose value is within factor $\rho \geq 1$ of the minimum, then there is a polynomial time algorithm that computes an $(\alpha, \frac{2\rho}{1-\alpha})$-bicriteria approximation for* Multicut.

Proof. Fix $0 < \alpha < 1$, and suppose $\mathcal{A}$ is a polynomial-time algorithm for Sparsest-Cut that produces a cut whose value is within factor $\rho \geq 1$ of the minimum. Now suppose we are given an input graph $G = (V, E)$ and k demand pairs $\{s_i, t_i\}_{i=1}^k$. We may assume without loss of generality that every s_i is connected (in G) to its corresponding t_i. Let $c_{\min}$ and $c_{\max}$ be the smallest and largest edge costs in G, and let $n = |V|$.

We now describe the bicriteria approximation algorithm for Multicut. For every value $C \in [c_{\min},\, n^2 c_{\max}]$ that is a power of 2, execute a procedure that we will describe momentarily to compute a cutset $M_C \subseteq E$, and report, from all these cutsets M_C whose removal disconnects at least αk demand pairs, the one of least cost. For a given value $C > 0$, the procedure starts with $M_C = \emptyset$, and then iteratively "augments" M_C as follows: Take a connected component S of $G \setminus M_C$, apply algorithm $\mathcal{A}$ to $G[S]$ (the subgraph induced on S and all the demand pairs that lie inside S), and if the resulting cutset E_S has value (in $G[S]$) at most $\frac{\rho}{1-\alpha} \cdot \frac{C}{k}$, then add the edges E_S to M_C. Here, the value (ratio of cost to demands cut) of E_S is defined as $b_S = c(E_S)/|D_S|$, where D_S is the collection of demand pairs that lie in $G[S]$ and get disconnected (in $G[S]$) when E_S is removed. Proceed with the iterations until for every connected component S in $G \setminus M_C$ we have $b_S > \frac{\rho}{1-\alpha} \frac{C}{k}$, at which point the procedure returns the cutset M_C.

This algorithm clearly runs in polynomial time. To analyze its performance, we first claim that for every value C, the cutset M_C returned by the above procedure has sparsest-cut value (ratio of cost to demand disconnected, in G) at most $\frac{\rho}{1-\alpha} \frac{C}{k}$. Indeed, suppose the procedure performs t augmentation iterations. Denote by S_i the connected component S that is cut at iteration $i \in [t]$, by E_{S_i} the corresponding cutset output by $\mathcal{A}$, and by D_{S_i} the corresponding set of demand pairs that get disconnected. Clearly, M_C is the disjoint union $E_1 \cup \cdots \cup E_t$, and it is easy to verify that the collection D_C of demand pairs cut by the cutset M_C is the disjoint union $D_{S_1} \cup \cdots \cup D_{S_t}$. Thus,

$$\begin{aligned} c(M_C) = \sum_{i=1}^{t} c(E_{S_i}) &\leq \frac{\rho}{1-\alpha} \cdot \frac{C}{k} \sum_{i=1}^{t} |D_{S_i}| \\ &= \frac{\rho}{1-\alpha} \cdot \frac{C}{k} |D_C|, \end{aligned}$$

which proves the claim.

For the sake of analysis, fix an optimal multicut $M^* \subseteq E$, i.e., a cutset of G whose removal disconnects all the demand pairs and has the least cost. The sparsest-cut value of M^* is $b^* = c(M^*)/k$. We will show that if $C \in [c(M^*), 2c(M^*)]$, then the above procedure produces a cutset M_C whose removal disconnects a collection D_C containing $|D_C| \geq \alpha k$ demand pairs; this will complete the proof of the lemma, because it immediately follows that

$$c(M_C) \leq \frac{\rho}{1-\alpha} \cdot \frac{C}{k} |D_C| \leq \frac{\rho}{1-\alpha} \cdot 2c(M^*),$$

and clearly $c(M^*) \in [c_{\min}, \binom{n}{2} \cdot c_{\max}]$. So suppose now $C \in [c(M^*), 2c(M^*)]$ and assume for contradiction that $|D_C| < \alpha k$. Denote by $V_1, \ldots, V_p \subseteq V$ the connected components of $G \setminus M_C$, and let D_j contain the demand pairs that lie inside V_j. It is easy to see that $\sum_{j=1}^{p} |D_j| = k - |D_C| > (1-\alpha)k$. Similarly, let M_j^* be the collection of edges in M^* that lie inside V_j. Then $c(M^*) \geq \sum_{j=1}^{p} c(M_j^*)$. Notice that, in every induced graph $G[V_j]$, the edges of M_j^* form a cutset (of $G[V_j]$) that cuts all the demand pairs in D_j. Using the stopping condition of the procedure, and since $\mathcal{A}$ provides an approximation within factor ρ, we have $c(M_j^*) \geq \frac{1}{1-\alpha}\frac{C}{k}|D_j|$ (the inequality is not strict because D_j might be empty). We thus derive the contradiction

$$c(M^*) \geq \sum_{j=1}^{p} c(M_j^*) \geq \frac{1}{1-\alpha} \cdot \frac{C}{k} \sum_{j=1}^{p} |D_j| > c(M^*).$$

This shows that when $C \in [c(M^*), 2c(M^*)]$, the procedure stops with a cutset M_C whose removal disconnects $|D_C| \geq \alpha k$ demand pairs, and concludes the proof of the lemma. □

4 Hardness of approximating Min-2CNF≡ Deletion

In this section, we modify the reduction in Section 2.1 to obtain a hardness of approximation for Min-2CNF≡ Deletion. In particular, we reduce the Multicut instance obtained in Section 2.1 to Min-2CNF≡ Deletion, such that a solution to the latter gives a Multicut of the same cost in the former.

The Min-2CNF≡ Deletion instance contains $2^{d-1}n$ variables, one for each demand pair $(u, \overline{u})$. In particular, for every demand pair $(u, \overline{u}) \in D$, we associate the literal x_u with u and the literal $x_{\overline{u}} = \neg x_u$ with $\overline{u}$. For every edge $e = (u, v)$ in the graph G there is a clause $(x_u \equiv x_v)$ whose weight is equal to the edge-weight w_e.

The following lemma is immediate from the construction and implies an analog of Lemma 2.3 for Min-2CNF≡ Deletion.

Lemma 4.1. *Given an assignment S of cost W to the above instance of* Min-2CNF≡ Deletion, *we can construct a solution of cost W to the* Multicut *instance G.*

Proof. Let M be the set of edges (u, v) for which $S(x_u) \neq S(x_v)$. Then M corresponds to the clauses that are not satisfied by S and has weight W. The lemma follows from observing that M is indeed a multicut—S is constant over connected components in $G \setminus M$, and for every demand pair $(u, \overline{u})$, we have $S(x_u) \neq S(x_{\overline{u}})$. □

We now give an analog of Lemma 2.1.

Lemma 4.2. *If there is a solution A for the unique 2-prover game G_Q such that the total weight of the satisfied questions is at least $1-\eta$, then there exists an assignment S for the above* Min-2CNF≡ Deletion *instance such that $c(S) \leq 2^{d+1}n$.*

Proof. Given the solution A for G_Q, we construct an assignment S as follows. For every question q_i^p and for every vertex u in the corresponding hypercube C_i^p, define $S(x_u)$ to be the A_i^p-th bit of u, i.e., $S(x_u) = u_{A_i^p}$. Note that this is a valid assignment, i.e., $S(x_u) = 1 - S(x_{\overline{u}})$ for all vertices u, as $u_{A_i^p} = 1 - \overline{u}_{A_i^p}$.

We bound the cost of the solution by first analyzing the clauses corresponding to hypercube edges in the corresponding Multicut instance. Consider unsatisfied clauses containing both variables in the same hypercube C_i^p, and note that the hypercube edges corresponding to these clauses form a dimension-A_i^p cut in the cube C_i^p. Therefore, the total weight of these clauses is at most $(2^{d-1})(2n) = 2^d n$.

Finally, consider an unsatisfied clause $(x_u \equiv x_v)$ corresponding to vertices in different hypercubes C^1_i and C^2_j. Then $S(x_u) \neq S(x_v)$ implies that $u_{A^1_i} = v_{b_{ij}(A^1_i)} \neq v_{A^2_j}$, or, $b_{ij}(A^1_i) \neq A^2_j$. There are at most 2^d such clauses for each question pair not satisfied by the solution A. Therefore, the total weight of such clauses is at most $2^d \eta \Lambda = 2^d n$.

The lemma follows from adding the two costs. □

Lemmas 4.1 and 4.2 along with Lemma 2.3 imply the part of Theorem 1.2 regarding Min-2CNF≡ Deletion.

5 Concluding remarks

Several important questions are left open. First, one would like to eliminate the dependence on the Unique Games Conjecture, and obtain a "standard" hardness of approximation result. Yet another challenge is to improve the hardness factor. For Multicut, the $\Omega(\log n)$ integrality ratio lower bound of [ACMM05] suggests that the inapproximability bound may be improved. In particular, $(\log n)^c$ hardness for a constant $c > 1/2$ will separate the approximability of Multicut from that of Sparsest-Cut (in light of the recent approximation due to [ALN05]).

The main bottleneck to improving the hardness factor lies in Friedgut's Junta Theorem (and similarly in the result of [KKL88] that we used in the conference version). These bounds are tight in general, as shown by the tribes function [BL90] (see also [Fri98, Section 2]).

A third challenge is to obtain hardness of approximation results for the uniform-demand case of the Sparsest-Cut problem or for the Balanced-Cut problem. Our results do not apply to this special but important case; in particular, if a 2-prover system has a low-cost balanced cut, then the corresponding graph on hypercubes would have a low-cost balanced cut regardless of the value of the 2-prover game. Alternatively, of course, one might improve the approximation algorithms for any of these problems.

References

[ACMM05] A. Agarwal, M. Charikar, K. Makarychev, and Y. Makarychev. $O(\sqrt{\log n})$ approximation algorithms for Min UnCut, Min 2CNF Deletion, and directed cut problems. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 573–581, 2005. 1.1, 5

[ALM+98] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998. 1.2

[ALN05] S. Arora, J. R. Lee, and A. Naor. Euclidean distortion and the sparsest cut. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 553–562, 2005. 1.1, 5

[AR98] Y. Aumann and Y. Rabani. An $O(\log k)$ approximate min-cut max-flow theorem and approximation algorithm. *SIAM Journal on Computing*, 27(1):291–301, 1998. 1.1

[ARV04] S. Arora, S. Rao, and U. Vazirani. Expander flows, geometric embeddings, and graph partitionings. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, pages 222–231, 2004. 1.1

[AS98] S. Arora and S. Safra. Probabilistic checking of proofs: A new characterization of NP. *Journal of the ACM*, 45(1):70–122, 1998. 1.2

[BBC04] N. Bansal, A. Blum, and S. Chawla. Correlation Clustering. *Machine Learning, Special Issue on Clustering*, 56(1-3):89–113, 2004. 1

[BGS98] M. Bellare, O. Goldreich, and M. Sudan. Free bits, PCP's and non-approximability – towards tight results. *SIAM Journal on Computing*, 27(3):804–915, 1998. 3

[BL90] M. Ben-Or and N. Linial. Collective coin flipping. In *Randomness and Computation*, pages 91–115, 1990. 5

[CGR05] S. Chawla, A. Gupta, and H. Räcke. Improved approximations to sparsest cut. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 102–111, 2005. 1.1

[CGW03] M. Charikar, V. Guruswami, and A. Wirth. Clustering with qualitative information. In *Proceedings of the 44th IEEE Symposium on Foundations of Computer Science*, pages 524–533, 2003. 1

[CKK+05] S. Chawla, R. Krauthgamer, R. Kumar, Y. Rabani, and D. Sivakumar. On the hardness of approximating multicut and sparsest-cut. In *Proceedings of the 20th Annual IEEE Conference on Computational Complexity*, pages 144–153, June 2005. 1.3

[CMM05] M. Charikar, K. Makarychev, and Y. Makarychev. Private Communication, September 2005. 1.2

[DI03] E. Demaine and N. Immorlica. Correlation clustering with partial information. In *Proceedings of the 6th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX)*, pages 1–13, 2003. 1

[DJP+94] E. Dahlhaus, D. S. Johnson, C. H. Papadimitriou, P. D. Seymour, and M. Yannakakis. The complexity of multiterminal cuts. *SIAM Journal on Computing*, 23(4):864–894, 1994. 1, 1.1

[EF03] D. Emanuel and A. Fiat. Correlation clustering—minimizing disagreements on arbitrary weighted graphs. In *Proceedings of 11th Annual European Symposium on Algorithms*, pages 208–220, 2003. 1

[FR04] U. Feige and D. Reichman. On systems of linear equations with two variables per equation. In *Proceedings of the 7th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX)*, pages 117–127, 2004. 1.2

[Fri98] E. Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–35, 1998. 1.3, 2.3, 5

[GVY96] N. Garg, V. V. Vazirani, and M. Yannakakis. Approximate max-flow min-(multi)cut theorems and their applications. *SIAM Journal on Computing*, 25(2):235–251, 1996. 1.1

[Hås01] J. Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, 2001. 1.1

[KARR90] P. Klein, A. Agrawal, R. Ravi, and S. Rao. Approximation through multicommodity flow. In *Proceedings of the 31st IEEE Symposium on Foundations of Computer Science*, pages 726–737, 1990. 1.1

[Kho02] S. Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 767–775, 2002. 1, 1.1, 1.1, 1.2

[KKL88] J. Kahn, G. Kalai, and N. Linial. The influence of variables on boolean functions. In *Proceedings of the 29th IEEE Symposium on Foundations of Computer Science*, pages 68–80, 1988. 1.3, 5

[KKMO04] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. Optimal inapproximability results for MAX-CUT and other 2-variable CSPs. In *Proceedings of the 45th IEEE Symposium on Foundations of Computer Science*, pages 146–154, 2004. 1.2

[KR03] S. Khot and O. Regev. Vertex cover might be hard to approximate to within 2-ϵ. In *Proceedings of the 18th Annual IEEE Conference on Computational Complexity*, pages 379–386, 2003. 1.2, 1.4

[KV05] S. Khot and N. K. Vishnoi. The unique games conjecture, integrality gap for cut problems and the embeddability of negative type metrics into ℓ_1. In *Proceedings of the 46th IEEE Symposium on Foundations of Computer Science*, 2005. To appear. 1.1, 1.2, 1.3, 1.3

[LLR95] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995. 1.1

[LR88] F. T. Leighton and S. Rao. An approximate max-flow min-cut theorem for uniform multicommodity flow problems with applications to approximation algorithms. In *Proceedings of the 29th IEEE Symposium on Foundations of Computer Science*, pages 422–431, 1988. 1.1

[MOO05] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: Invariance and optimality. In *Proceedings of the 46th IEEE Symposium on Foundations of Computer Science*, 2005. To appear. 1.2

[Raz98] R. Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, 1998. 1.2

[Tre05] L. Trevisan. Approximation algorithms for unique games. In *Proceedings of the 46th IEEE Symposium on Foundations of Computer Science*, 2005. To appear; see also comment for ECCC TR05-34. 1.2

[Vaz01] V. V. Vazirani. *Approximation algorithms.* Springer-Verlag, Berlin, 2001. 1.1

A Regularity of the Unique Games instance

Proof of Lemma 1.5. Given a unique 2-prover game Q, we describe how to convert it to a regular game while preserving its completeness and soundness. First we claim that we can assume that the ratio between the maximum weight $\max_e w_e$ and the minimum weight $\min_e w_e$ is bounded by n^3. This is because we can remove all edges with weight less than $\frac{1}{n^3} \max_e w_e$ from the graph, changing the soundness and completeness parameters by at most $\frac{1}{n}$. By a similar argument, we can assume that all weights in the graph are integral multiples of $t = \frac{1}{n^2} \min_e w_e$.

Now we convert Q to a regular graph Q' as follows. For each prover $p \in \{1, 2\}$ and question q_i^p, form $W(p,i)/t$ vertices $q_i^p(1), \cdots, q_i^p(W(p,i)/t)$, where $W(p,i)$ is the total weight of all the edges incident on q_i^p. For every pair of vertices (q_i^1, q_j^2), connected by an edge e in Q, we form an edge between $q_i^1(x)$ and $q_j^2(y)$, for all possible values of x and y, with weight $w_e \frac{t}{W(1,i)} \frac{t}{W(2,j)}$.

Note that the total weight of all the edges remains the same as before. Each new vertex $q_i^1(x)$ has total weight $\sum_e w_e \frac{t}{W(1,i)} \frac{t}{W(2,j)} \frac{W(2,j)}{t} = t$, where the sum is over all edges e incident on q_i^1. Therefore, the graph is regular. Furthermore, the number of vertices increases by a factor of at most n^5.

It only remains to show that the soundness and completeness parameters are preserved. To see this, note that any solution on the original graph Q can be transformed to a solution of the same value on Q', by picking the same answer for every vertex $q_i^p(x)$ in Q' as the answer picked for q_i^p in Q. Likewise, consider a solution in Q'. Note that the answers for the questions $q_i^p(x)$ with different values of x must all be the same, because all these questions are connected to identical sets of vertices, with the same weights. Therefore, the solution in Q that picks the same answer for q_i^p as the answer for $q_i^p(x)$ in Q' has the same weight as the given solution in Q'.

Thus for every solution in Q, there is a solution of the same weight in Q' and vice versa. This proves that the two games have exactly the same soundness and completeness parameters. $\square$