Multicut and Sparesest-Cut problem

Lecturer: Dr. My T. Thai Scribe: Dung T. Nguyen

1 Unique Games Conjecture

1.1 The unique game

Recall the two-prover one round proof system which can be seen as a game between two provers and a verifier. There are two sets of all possible "questions" V and W that the verifier can ask the first and second prover respectively.

The strategy of the first prover is a map $L_V : V \to N$ where N is a set of all possible answers of the first prover. Given a question u, the first prover returns the answer $L_V(u)$ to the verifier. Similarly, the stratefy of the second prover is a map $L_W : W \to M$ where M is the set of its possible answers.

The decision of verifier is a map:

$$\Gamma : V \times N \times W \times M \to \{TRUE, FALSE\}$$

The game between the verifier and provers is described as follow. The verifier picks a pair of question (v, w), $v \in V$, $w \in W$ with a certain probability distribution on the set of all pairs. It asks question v to the first prover, question w to the second one. Two provers return answers $L_V(v), L_W(w)$ respectively. The verifier accepts iff :

$$\Gamma(v, L_V(v), w, L_W(w)) = TRUE$$

The value of game is defined as the maximum, over all possible prover strategies, of acceptance probability of the verifier.

Consider the game such that the answer of the second prover uniquely determines the answer of the first prover. It means that for every question pair (v, w) asked by the verifier and every answer $b \in M$ of the second prover, there is a unique answer $a \in N$ that makes the verifier accept. Thus, we can associate a function $\Pi_{vw} : M \to N$ to the question pair (v, w) such that the verifier accepts iff:

$$\Pi_{vw}(L_W(w)) = L_V(v)$$

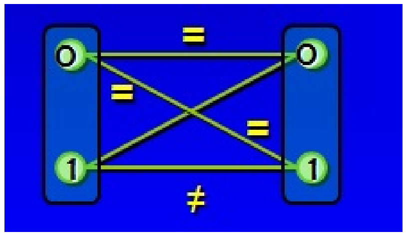

Figure 1: Map functions from the answer of the second prover to the answer of the first prover

The game is called unique if $M = N$ and every function Π_{vw} is a bijection. Following is an example of unique game. [1]:

- The verifier sends a random bit to each prover
- Each prover responds with a bit (so k=2 here)
- The verifier accepts iff the XOR of the answers is equal to the AND of the questions

The unique game can be represented on biparttitle graph $G(V, W, E)$ where $V = \{q_1^1, q_2^1, ..., q_n^1\}$ and $W = \{q_1^2, q_1^2, ..., q_n^2\}$ are sets of questions for the first and second prover respectively. Set of answer is $\{1, 2, ..., d\}$ where d is the size of answer sets. Each edge $(q_i^1, q_j^2) \in E$ has weight w_{ij} with the total weight is 1 and a associated bijection $\Pi_{ij} : [d] \to [d]$ which maps every answer for the question q_i^1 to a distinct answer for question q_j^2. Given an assignment $A = \{A_i^p | p \in [2], A_i^p \in [d]\}$ of answers to questions, the edge (q_i^1, q_j^2) is sastified if $A_j^2 = \Pi_{ij}(A_i^1)$. The goal is to find an assignment that maximizes the total weight of sastified edges.

Note 1.1. If the number of questions of two provers are different, add dummy questions that are never asked.

1.2 Unique Games Conjecture

Unique Games Conjecture [2]. 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.

Fact 1.1. For all $\delta > 0$, $\textit{Gap-Unique Game}_{1,\delta}$ is not NP-hard for any d.

Remark 1.1. Feige et al [3] proved that for every $\delta > 0$, there is some constant $0 < \eta < 1$ such that $\textit{Gap-Unique Game}_{\eta,\delta\eta}$ is NP-hard.

Remark 1.2. Charikar et al [4] gave an approximation algorithm for unique game. Given any unique game G, the algorithm outputs a value η s.t.:

$$1 - O((\eta \log d)^{\frac{1}{2}}) < OPT(G) < 1 - \eta$$

This establishes a lower bound value of d if the Unique Games Conjecture is true.

Definition 1.1. A unique 2-prover game is called regular if the total weight of question edges incident at any single vertex is the same.

Lemma 1.1. The Unique Games Conjecture implies that for every $\eta, \delta > 0$, there exists $d = d(\eta, \delta)$ such that it is NP-hard to decide if a regular unique 2-prover game with n vertices and d answers for each vertex has value at least $1 - \eta$ or at most δ.

Proof. Roadmap:

- Eliminate all small weight edges i.e. the weight is less than a threshold.
- Round the weights of edges to multiple of a common divisor value.
- Add new vertices, new edges and devide the weight of each edge to new edges such that the total weight of every vertex is the same, the common divisor value.

Given a unique game G, denote $w_{max} = max_e\{w_e\}$. Let $r = max\{1/\eta, 1\}$. First we remove all edges of weight less than $(1/2n^2r)w_{max}$, then add an equal value to all left edges such that the total weight of all edges is 1. There is at most n^2 edges, thus the optimal value is reduced or increased at most $(1/2r)w_{max}$. In addition, w_{max} is less than optimal value. Therefore:

- If the $OPT(G) > 1 - \eta$ then $OPT(G') > 1 - \eta - 1/2r > 1 - 3\eta/2$
- If the $OPT(G) < \delta$ then $OPT(G') < \delta(1 + 1/2r) < 3\delta/2$

where G' is the obtained graph after removing light weight edges.

Next, let $t = (1/2r)w_{min}$. Round down the weight of every edge to the nearest multiple of t then increase the weight of some different edges to ensure that the total weight is still 1. This makes the weight of every edge increased or decreased at most $1/2r$ fraction of its value. This leads to the following result:

- If the $OPT(G') > 1 - 3\eta/2$ then $OPT(G'') > 1 - 3\eta/2 - 1/2r > 1 - 2\eta$
- If the $OPT(G') < 3\delta/2$ then $OPT(G'') < 3\delta/2(1 + 1/2r) < 9\delta/4$

where G'' is the obtained graph after rounding the weight of edges.

Now:

- For each vertice q_i^p, create $W(p,i)/t$ vertices $q_i^p(1), ..., q_i^p(W(p,i))$, where $W(p,i)$ is the total weight of all edges incident to q_i^p.
- For each edge (q_i^1, q_j^2), form edges $(q_i^1(x), q_j^2(y))$ with weight $w_{ij}(t/W(1,i))(t/W(2,j))$ for all possible x, y.
- The bijection on the edge $(q_i^1(x), q_j^2(y))$ is set to be the same as the original bijection b_{ij} on the edge (q_i^1, q_j^2).

In the obtained graph is G''', every vertex has total weight of incident edges is t.

Completeness. If $OPT(G'') > 1 - 2\eta$, we assign to $q_i^p(x)$ the answer of the question q_i^p to achieve a solution with the value is at least $OPT(G'')$. Thus $OPT(G''') \geq OPT(G'') > 1 - 2\eta$.

Soundness. Consider a optimal solution for G'''. Assign q_i^p the answer for $q_i^p(x_max)$ where $q_i^p(x_max)$ is the vertex in set of vertices $q_i^p(x)$ that has the maximum total weight of incident sastified edges. Because all $q_i^p(x)$ have the same constrain, thus the total weight of incident sastified edges of q_i^p is at least the total weight of all sastified edges that incident to any $q_i^p(x)$. Therefore, $OPT(G'') > OPT(G''')$. If $OPT(G'') < 9\delta/4$ then $OPT(G''') < 9\delta/4$

In conclusion, we reduce *Gap Uniqe Game*$_{1-\eta,\delta}$ to *Gap Regular Unique Game*$_{1-2\eta,9\delta/4}$. The proof is completed.

2 Hardness of bicriteria approximation for MULTICUT

Multicut problem. Given an undirected graph $G = (V, E)$ with k pairs $\{s_i, t_i\}_{i=1}^k$, called demand pairs.The goal is to find a minimum size subset of edges $M \subseteq E$ whose removal disconnects all the demand pairs.

Definition 2.1. A *d-dimensional hypercube*(for short, a $d-cube$) is the graph $G = (V_C, E_C)$ with the vertex set $V_C = \{0,1\}^d$ and an edge $(u, v) \in E_C$ iff $u, v \in \{0,1\}^d$ and differ in exactly one dimension.

Definition 2.2. A *dimension-a cut* is set of edges whose two ends differ in a^{th} dimension i.e. $\{((u_1, u_2, ..., u_d), (v_1, v_2, ..., v_d)) | u_i = v_i, \forall i \neq a \text{ and } u_a \neq v_a\}$.

Definition 2.3. Two vertices u, v are called *antipodal pair* if they differ in all cordinate i.e. $v_i \neq u_i, \forall i = 1, d$. Denote an *antipodal pair* as $(u, \bar{u})$.

Definition 2.4. 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 demand pairs and has cost at most β times of the cost of optimum cut.

Theorem 2.1. Suppose that for $\eta = \eta(n)$, $\delta = \delta(n)$, and $d = d(\eta, \delta) \leq O(\log n)$, it is NP-hard to determine whether a unique 2-prover game with $|V| = 2n$ vertices and answer set size d has value at least $1 - \eta$ or at most δ. Then there exist constants $c_1, c_2 > 0$ such that it is NP-hard to approximate Multicut, Sparsest-Cut to within factor $L(n) = c_1 min\{\frac{1}{\eta(n^{c_2})}, \log \frac{1}{\delta(n^{c_2})}\}$.

2.1 Reduction from regular unique 2-prover game to Multicut

Given a regular unique 2-prover game instance $G_Q = (V_Q, W_Q, E_Q)$ with $n = |V_Q| = |W_Q|$ and the corresponding edge weights w_{ij} and bijections $\Pi_{ij} : [d] \to [d]$, we construct a Multicut instance $G = (V, E)$ as follows:

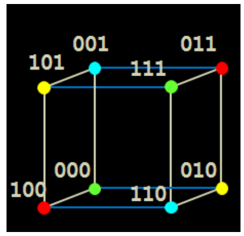

Figure 2: A 3-dimensional hypercube with antipodal pairs and dimension-2 cut

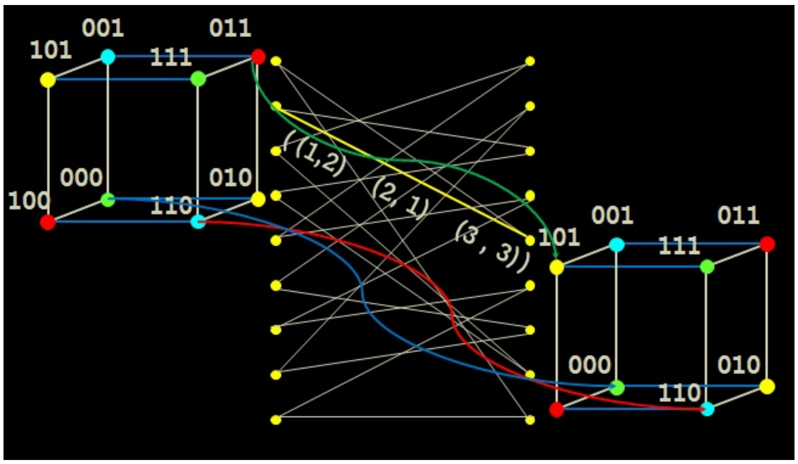

Figure 3: Cross edges

- For each vertex $q_i^p \in V_Q \cup W_Q$, construct a *d-dimensional hypercube* C_i^p. Edges inside the hypercube have weight of 1 and called hypercube edges.
- For each question edge (q_i^1, q_j^2), connect the vertex $u \in C_i^1$ to the vertex $v = \Pi'_{ij}(u) \in C_j^2$ where $\Pi'_{ij}(u)_k = u_{\Pi_{ij}^{-1}(k)}$ i.e. $\Pi'_{ij}(u)_{\Pi_{ij}^{-1}(k)} = u_k$ for all k from 1 to d. The weight of edge uv is $w_{ij}\Lambda$, where $\Lambda = n/\eta$. These edges are called cross edges.
- Set of demand pairs is all antipodal pairs in all hypercubes.

2.2 Completeness

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 1η, then the resulting Multicut instance G contains a multicut $M \subseteq E$ such that the cost $c(M) \leq 2^{d+1}n$.

Proof. Construct M as following:

- For each vertex q_i^p assigned answer $A_i^p \in [d]$, collect all edges in A_i^p*-dimension* cut to M.

- For each edge (q_i^1, q_j^2) which is not satisfied, collect all cross edges between C_i^1 and C_j^2 to M.

Let's consider an arbitrary antipodal pair $(u, \bar{u})$ in C_i^p, there are only two ways to travel from u to $\bar{u}$:

- In the same hypercube.
- Go to other hypercubes and come back.

A A_i^p-dimension cut seperates C_i^p into two sets based on the value of A_i^{pth} cordinate. Use the value of A_i^{pth} cordinate to label all vertices in the hypercube. If (q_i^1, q_j^2) is satisfied, two ends of all cross edges between C_i^1 and C_j^2 have the same label. Thus, removing M disconnects all demand pairs.

The cost of M is total cost of hypercube edges and cross edges:

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

2.3 Soundness

Given a *d-dimensional hypercube* $C = (V_C, E_C)$, a cut separates the vertex set to 2 partitions. Label every vertex 0 or 1 such that vertices in the same parition have the same value. This paritioning can be viewed as a boolean function $f : \{0,1\}^d \to \{0,1\}$ and the cut is $\{(u,v)|f(u) \neq f(v)\}$. The influence of a dimension (variable) a, $a \in [d]$, on the function f, denoted by I_a^f, is defined as the fraction of the *dimension-a* edges $(u,v) \in E_C$ for which $f(u) \neq f(v)$. The total influence of f is $\sum_{a \in [d]} I_a^f$.

Definition 2.5. We say that the function f is a *k-junta* if there exists a subset $J \subseteq [d]$ with $|J| \leq k$ such that the influence of any dimension $a \notin J$ is 0.

Definition 2.6. Two functions f and f' are said to be *ϵ-close* if fraction of vertices that the value of two functions are different is less than ϵ.

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

Lemma 2.2. There exists $L = \Omega(min\{1/\eta, \log 1/\delta\})$ such that if the Multicut instance G has a cutset of cost at most $2^{d+1} nL$ whose removal disconnects an $\alpha \geq 7/8$ 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 $L = c \ min\{1/\eta, \log 1/\delta\}$ and let $M \subseteq E$ be a cutset of cost $c(M) \leq 2^{d+1} nL$ whose removal disconnects an $\alpha \geq 7/8$ fraction of the demand pairs. Construct a randomized solution for unique 2-prover game G_Q

- Label each connected component of $G \backslash M$ 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 that include v.

- For each vertex (question) q_i^p consider the restriction of f to the hypercube C_i^p, denoted $f_{|C_i^p}$, and apply Friedgut's Junta Theorem) with $\epsilon = 1/60$, to obtain a subset of dimensions $J_i^p \subseteq [d]$. On hypercube such that we obtain $|J_i^p| \leq 2^{O(L/\epsilon)}$ choose the answer A_i^p uniformly at random from J_i^p

Randomly consider an edge (q_i^1, q_j^2) with probability w_{ij}, we will compute the probability such that the question edge is satisfied i.e. $A_j^2 = \Pi_{ij}(A_i^1)$. The probability such that a question q_i^p is selected in the question edge is $1/n$ because the G_Q is regular. Without loss of generality, we assume that removing M disconnects at least as many demand pairs inside the cubes $\{C_i^1\}_{i\in[d]}\}$ as inside the cubes $\{C_i^2\}_{i\in[d]}$. Consider 4 "bad" events:

- $E_1 =$ greater than a 7/8-fraction of the vertices $v \in C_i^1$ satisfy $f(v) = f(\bar{v})$.
- $E_2 = M$ contains more than $2^{d+5}L$ hypercube edges in C_i^1.
- $E_3 = M$ contains more than $2^{d+5}L$ hypercube edges in C_j^2.
- $E_4 = M$ contains more than $2^{d+5}\eta L$ cross edges between C_i^1 and C_j^2

We have:

- At least 7/8-fraction of demand pairs in $\{C_i^1\}_{i\in[d]}\}$ is disconnected, each cube is choosen randomly uniformly. Thus the expectated fraction of connected demand pairs in C_i^1 is at most 1/8. In addition, the probability $Pr[f(v) = f(\bar{v})|v \textit{ and } \bar{v} \textit{ disconnected }] = 1/2$. Then the expected fraction of vertices $v \in C_i^1$ for which $f(v) = f(\bar{v})$ is at most 5/8, and by Markov's inequality, $Pr[E_1] \leq 5/7$.
- M contains at most $2^{d+1}nL$ hypercube edges, thus the expected number of edges in $C_i^1 \cup C_j^2$ that are contained in M is at most $2^{d+1}L$ and by Markov's inequality $Pr[E_2 \cup E_3] \leq 1/16$
- $Pr[E_4] \leq 1/16$, as otherwise because the probability of the total cost of M along the cross edges corresponding to this event is more than $1/16(2^{d+5}\eta L)\Lambda = 2^{d+1}nL \geq c(M)$.

Thus:

$$Pr[E_1 \cup E_2 \cup E_3 \cup E_4] \leq \tfrac{5}{7} + \tfrac{2}{16} < \tfrac{6}{7}$$

Suppose that all four above events don't occur. We have:

- E_2 does not occur, then the total influence of $f_{|C_i^1}$ is at most $2^{d+5}L/2^{d-1} = 64L$. There exists ϵ-*close* function g^1 of $f_{|C_i^1}$ with junta $|J_i^1| \leq 2^{O(L/\epsilon)}$.
- E_3 does not occur, then the total influence of $f_{|C_j^2}$ is at most $2^{d+5}L/2^{d-1} = 64L$. There exists ϵ-*close* function g^2 of $f_{|C_j^2}$ with junta $|J_j^2| \leq 2^{O(L/\epsilon)}$.

- E_1 does not occur, then the fraction of the smaller partition of C_i^1 based on $f_{|C_i^1}$ is at least $1/16$. Call the fraction of smaller partition is the balance of function.

We will prove that there is a contradiction if $|J_i^1 \cap \Pi_{ij}^{-1}(J_j^2)| = 0$. Let the function $h : C_i^1 \to 0, 1$ with $h(v) = g_j^2(\Pi'_{ij}(v))$ then h depends only on variables in $\Pi_{ij}^{-1}(J_j^2)$. Thus h and g^1 depends on disjoint sets of variables. The balance of h and g^1 is at least $1/16 - \epsilon$, then probability $Pr_{v \in C_i^1}[g^1(v) \neq h(v)] \geq 1/16 - \epsilon$. This means g^1 and h are not $(1/16 - \epsilon)$-*close*.

Because E_4 doesnt't occur, at most $32\eta L 2^d$ vertices $v \in C_i^1$ satisfies $f(v) \neq f(\Pi'_{ij}(v))$ i.e. $f_{|C_i^1}$ is $32\eta L 2^d$-*close* to $f_{|C_j^2} o \Pi'_{ij}$. While g^1 and h are ϵ-*close* to $f_{|C_i^1}$ and $f_{|C_j^2} o \Pi'_{ij}$ respectively, therefore g^1 and h are $2\epsilon + 32\eta L 2^d$-*close*. With small enough c, $2\epsilon + 32\eta L 2^d < 1/16$, that leads to the contradiction.

Because $|J_i^1 \cap \Pi_i^{-1} j(J_j^2)| \geq 1$, the probability such that chosen edge (q_i^1, q_j^2) sastified is at least:

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

From above result, the expected value of randomized solution A is at least $2^{-O(L)} > \delta$ with small enough $c > 0$.

In conclusion, the achieved gap is $L(n) = \Omega(min\{1/\eta(n), \log(1/\delta(n))\})$. Assume $d(\eta, \delta) \leq O(\log n)$, and thus the resulting Multicut instance G has size $N = 2n2^d = n^{\theta(1)}$. It follows that in terms of the Multicut instance size N, the gap is

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

3 Hardness of approximating SPARSEST-CUT

SPARSEST-CUT problem Given an undirected graph $G = (V, E)$, each edge has a associated weight and k demand pair $\{s_i, t_i\}_{i=1}^k$. The goal is to find a cut M that disconnected $1 \leq \alpha \leq k$ pairs such that the ratio between total weight of edges in M over the number of disconnected pairs i.e. α is minimum.

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 (α, $\frac{\rho}{1-\alpha}$-*bicriteria*) approximation for Multicut.

Proof.

- Initialize $D = \emptyset$, $D = \emptyset$

- Each iteration, if $|D| < \alpha k$, where k is the nubmer of demand pairs, find a sparsest cut M' which disconnects a set D' of left demand pairs. Assign $|D| = |D| \cup D'$, $|M| = |M| \cup M'$. Eliminate D' from set of demand pairs.

At each step, at least $(1 - \alpha)k$ demand pairs left which are disconnected by optimal multicut M^*. Thus $c(M') \leq \rho \frac{c(M^*)}{(1-\alpha)k} |D'|$. Sum up all we have the proof of lemma.

References

[1] J. Kempe, O. Regev, and B. Toner. Unique games with entangled provers are easy. *Proc. 49th FOCS*.

[2] S. Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 767–775. ACM New York, NY, USA, 2002.

[3] U. Feige and D. Reichman. On systems of linear equations with two variables per equation. *Lecture notes in computer science*, pages 117–127, 2004.

[4] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Near-optimal algorithms for unique games. In *STOC '06: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 205–214, New York, NY, USA, 2006. ACM.