On a Cut-Matching Game for the Sparsest Cut Problem

Rohit Khandekar
Subhash A. Khot
Lorenzo Orecchia
Nisheeth K. Vishnoi

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2007-177
http://www.eecs.berkeley.edu/Pubs/TechRpts/2007/EECS-2007-177.html

December 22, 2007

ON A CUT-MATCHING GAME FOR THE SPARSEST CUT PROBLEM

ROHIT KHANDEKAR [1], SUBHASH A. KHOT [2], LORENZO ORECCHIA [3], AND NISHEETH K. VISHNOI [4]

[1] IBM T.J. Watson Research Center
E-mail address: rohitk@us.ibm.com

[2] New York University
E-mail address: khot@cs.nyu.edu

[3] U.C. Berkeley
E-mail address: orecchia@cs.berkeley.edu

[4] U.C. Berkeley and IBM India Research Lab
E-mail address: nisheeth.vishnoi@gmail.com

Abstract. We study the following game between a "cut" player $\mathcal{C}$ and a "matching" player $\mathcal{M}$. The game starts with an empty graph G on n vertices. In each round, the cut player chooses a bisection $(S, \overline{S})$ of vertices and the matching player then adds a perfect matching M (not necessarily belonging to G) between S and $\overline{S}$ to the (multi-)graph G. The choices of the players in each round may depend on those in the previous rounds. The game ends when G becomes an edge-expander. The value of this game, denoted by $\text{VAL}(n, \mathcal{C}, \mathcal{M})$, is the total number of rounds in the game before it ends. We study this game for its connection with the Sparsest Cut problem in undirected graphs: if there is a polynomial-time cut player $\mathcal{C}_f$ such that $\text{VAL}(n, \mathcal{C}_f, \mathcal{M}) \leq f(n)$ for all $\mathcal{M}$, then there is a polynomial-time $O(f(n))$-approximation algorithm for the Sparsest Cut problem.

We show that there is no cut player $\mathcal{C}$, even unbounded-time, that can ensure $\text{VAL}(n, \mathcal{C}, \mathcal{M}) = o(\sqrt{\text{GAP}(n)})$ for all matching players $\mathcal{M}$, where $\text{GAP}(n)$ is the integrality gap of the well-studied SDP with triangle inequality constraints for the Sparsest Cut problem. Recall that $\text{GAP}(n) = \Omega(\log\log n)$ [5]. Thus, we prove that this approach cannot yield a $o(\sqrt{\text{GAP}(n)})$-approximation (and in particular, $o(\sqrt{\log\log n})$-approximation) algorithm for this problem. Furthermore, we show that there is a (super-polynomial time) cut player $\mathcal{C}^*$ such that, for all $\mathcal{M}$, we have $\text{VAL}(n, \mathcal{C}^*, \mathcal{M}) = O(\log n)$.

1. Introduction

1.1. The game of expansion

In this paper, we study the following game between two players called the "cut" player and the "matching" player. The game starts with an empty graph G on n vertices (where n is an even integer) and goes in several rounds. In each round,

- The cut player chooses a bisection $(S, \overline{S})$ of the vertex set $[n]$.
- The matching player then chooses a perfect matching M between S and $\overline{S}$.
- The matching M is then added to the (multi-)graph G.

In any round, the choices made by the players could depend on those made by the players in the previous rounds and can be assumed to be deterministic. We identify the cut player with the strategy $\mathcal{C}$ it picks and the matching player with the strategy $\mathcal{M}$ it picks from the set of all possible strategies available for them respectively. The game stops when G becomes an edge-expander, i.e., the ratio of the number of edges across U and the number of vertices in U is $\Omega(1)$ for all subsets U with $0 < |U| \leq n/2$. For a cut player $\mathcal{C}$ and a matching player $\mathcal{M}$, the value of the game, denoted by

$\textsc{val}(n, \mathcal{C}, \mathcal{M})$, is the number of rounds in the game. The goal of the cut player is to use a strategy $\mathcal{C}_{\text{opt}}$ which minimizes $\max_{\mathcal{M}} \textsc{val}(n, \mathcal{C}_{\text{opt}}, \mathcal{M})$ or ensure that independent of the strategy employed by the matching player, the graph G becomes an expander as quickly as possible. On the contrary, the goal of the matching player is to use a strategy $\mathcal{M}_{\text{opt}}$ that maximizes $\min_{\mathcal{C}} \textsc{val}(n, \mathcal{C}, \mathcal{M}_{\text{opt}})$.

1.2. Significance and History

The significance of the game of expansion lies in the fact that a polynomial-time cut player $\mathcal{C}_f$ that ensures that $\max_{\mathcal{M}} \textsc{val}(n, \mathcal{C}_f, \mathcal{M}) \leq f(n)$ gives a polynomial-time $O(f(n))$-approximation algorithm for the Sparsest Cut problem[1] in undirected graphs. Furthermore, if the cut player's strategy takes time $T(n)$ in each round, this approximation algorithm runs in time $O(f(n) \cdot (T(n) + T_{flow}))$ where T_{flow} is the time needed to compute a single commodity maximum flow in the input graph.

The Sparsest Cut problem is very useful as a sub-routine in designing graph theoretic algorithms via the divide-and-conquer paradigm. In practice, typical inputs to this problem consist of very large graphs, and hence, it is imperative to find algorithms that run fast and have a guarantee about the quality of the cut they produce. A comprehensive survey of the applications of the Sparsest Cut problem can be found in [10]. The seminal work of Leighton and Rao [8] gave an $O(\log n)$-approximation algorithm for Sparsest Cut via an LP relaxation based on multicommodity flows. A breakthrough result of Arora, Rao and Vazirani [2] showed how to achieve an $O(\sqrt{\log n})$-approximation to the Sparsest Cut problem using semidefinite programming (SDP). Though their algorithm improves the quality of the solution produced, it does worse as far as the running time is concerned. Subsequent attempts by Arora, Hazan and Kale [1], and more recently by Arora and Kale [3], have reduced the running time of the Arora, Rao and Vazirani algorithm to that of computing poly-logarithmic number of multi-commodity flows.

On the other hand, a seemingly different and more efficient approach was proposed by Khandekar, Rao and Vazirani [7]. They reduced the problem of finding the sparsest cut in a graph to a poly-logarithmic number of single commodity max-flow computations. This, for the first time, brought the Sparsest Cut problem into the domain where an algorithm, with theoretical guarantees on the quality of the cut produced, could possibly compete with the heuristics such as METIS [6] in terms of the running time. More precisely, Khandekar, Rao and Vazirani [7] presented an $O(\log^2 n)$-approximation algorithm for the Sparsest Cut problem that runs in time $\tilde{O}(m + n^{1.5})$. The cut-matching game was implicit in their work and they achieve this approximation by presenting an $\tilde{O}(n)$-time cut player $\mathcal{C}_{KRV}$ that ensures $\max_{\mathcal{M}} \textsc{val}(n, \mathcal{C}_{KRV}, \mathcal{M}) = O(\log^2 n)$. Their approximation guarantee has been improved to $O(\log n)$ independently by Arora and Kale [3] and by Schulman, Orecchia, Vazirani and Vishnoi [9]. Both these algorithms follow the Khandekar-Rao-Vazirani paradigm and reduce the problem of Sparsest Cut to single commodity max-flow computations.

Given this connection, the following questions are interesting:

(1) Is there is a polynomial-time cut player $\mathcal{C}^*$ such that $\max_{\mathcal{M}} \textsc{val}(n, \mathcal{C}^*, \mathcal{M}) = o(\sqrt{\log n})$? If true, it would improve the best known approximation [2] for the Sparsest Cut problem. A priori, this seems possible, since the approach of [7] seems different from that [2].

(2) From a point of view of practical algorithms, is there a *fast* strategy $\mathcal{C}^*$ for the cut player for which $\max_{\mathcal{M}} \textsc{val}(n, \mathcal{C}^*, \mathcal{M})$ is *small*? For example, a near-linear time strategy $\mathcal{C}^*$ with $\max_{\mathcal{M}} \textsc{val}(n, \mathcal{C}^*, \mathcal{M}) = o(\log n)$ would be very interesting.

(3) Are there inherent bottlenecks in this approach in obtaining better approximations and better running times for the Sparsest Cut problem?

[1] Recall that the Sparsest Cut problem is to find a cut $(U, \overline{U})$ with $0 < |U| \leq n/2$ such that $|E(U, \overline{U})|/|U|$ is minimum.

1.3. Our results

In this paper, we study the quantities $\min_{\mathcal{C}} \max_{\mathcal{M}} \textsc{val}(n, \mathcal{C}, \mathcal{M})$ and $\max_{\mathcal{M}} \min_{\mathcal{C}} \textsc{val}(n, \mathcal{C}, \mathcal{M})$. The goal of the paper is to understand this min-max (or max-min) quantity as a function of n. We prove the following two results regarding the abstraction of the cut-matching game presented in this paper:

- **Lower Bound.** There is a matching player $\mathcal{M}^*$ such that

$$\min_{\mathcal{C}} \textsc{val}(n, \mathcal{C}, \mathcal{M}^*) = \Omega(\sqrt{\textsc{gap}(n)}) \quad (= \Omega(\sqrt{\log \log n}) \text{ [5]}),$$

  where $\textsc{gap}(n)$ is the integrality gap of the well-studied SDP with triangle inequality constraints for the Sparsest Cut problem. Thus, in particular we rule out this approach for obtaining a $o(\sqrt{\textsc{gap}(n)})$-approximation (and hence $o(\sqrt{\log \log n})$-approximation) algorithm for this problem.
- **Upper Bound.** There is a cut player $\mathcal{C}^*$ such that

$$\max_{\mathcal{M}} \textsc{val}(n, \mathcal{C}^*, \mathcal{M}) = O(\log n).$$

  The strategy $\mathcal{C}^*$ is not known to have a polynomial time implementation.

1.4. Our techniques

Our lower bound argument, in Section 3, is based on the dual of the standard SDP relaxtion for the Sparsest Cut problem and the integrality gap instances for this SDP. The strategy $\mathcal{M}^*$ of the matching player is similar to [7] when the input graph is the integrality gap example. We argue that a feasible dual solution can be extracted from the set of matchings computed by the matching player. Using Cheeger's relationship [4] between the sparsity and the eigenvalue gap for the graphs, we compare the value of the dual solution with the sparsity of the integrality gap example. This enables us to obtain a lower bound on the number of rounds in terms of the integrality gap of the SDP.

Our upper bound argument, in Section 4, is based on an alternate formulation of edge expanders [7]. We associate, with the sequence of matchings added by the matching player, a natural random walk; and prove that G becomes an expander when the random walks mix. To measure the progress of mixing, we consider the *entropy* potential. This is in deviation from [7] who use the ℓ_2^2 potential and show that their potential becomes small enough to imply the expansion of G in $O(\log^2 n)$ rounds. We, on the other hand, argue that $O(\log n)$ rounds suffice for the entropy potential to become large enough to imply the expansion of G.

2. The Cut-Matching Game

Preliminaries. All graphs considered in this paper are undirected, they may have parallel edges but have no loops. Let $G(V, E)$ be a graph on n vertices, where n is assumed to be even throughout the paper. For $S \subseteq V$, the cut $(S, \overline{S} := V \setminus S)$ is called c-balanced if $\min\{|S|, |\overline{S}|\} \geq cn$. A "bisection" is a 1/2-balanced cut. The *expansion* of a cut $(S, \overline{S})$, where $0 < |S| \leq n/2$, is defined as $|E(S, \overline{S})|/|S|$. Here $E(S, \overline{S})$ denotes the set of edges that have one end point in S and another in $\overline{S}$, counted with multiplicities. For $0 \leq c \leq 1/2$, let

$$\phi_c(G) := \min_{\substack{\emptyset \neq S \subset V \\ cn \leq |S| \leq n/2}} \frac{|E(S, \overline{S})|}{|S|}$$

Cut-Matching Game $\mathcal{G}(n, c)$:

- $G := (V, \emptyset)$ be an empty graph where $|V| = n$ is even.
- Fix a cut player $\mathcal{C}$ and a matching player $\mathcal{M}$.
- While $\phi_c(G) < 1/4$ do:
  (1) $\mathcal{C}$ chooses a bisection $(S, \overline{S})$ in G.
  (2) $\mathcal{M}$ chooses a perfect matching M between S and $\overline{S}$. The edges of M may or may not belong to G.
  (3) $G \leftarrow G + M$.
- The value of the game, $\text{VAL}(n, \mathcal{C}, \mathcal{M})$, is the number of iterations of the while loop.

Figure 1: The Cut-Matching game

denote the minimum expansion among all c-balanced cuts. Note that $\phi_0(G)$ denotes the minimum expansion among all cuts in G, i.e., the expansion of the sparsest cut in G. The SPARSEST CUT problem is, given a graph $G(V, E)$, find the cut that achieves $\phi_0(G)$.

The game. We consider the multi-round game $\mathcal{G}(n, c)$, given in Figure 1, between two players called a "cut" player and a "matching" player. We identify the cut player with the strategy $\mathcal{C}$ it uses and the matching player with the strategy $\mathcal{M}$ it uses. We start with an empty graph G on n vertices, where n is an even integer. We denote by $G_t(V, E_t)$ the graph after t rounds of the game. In each round $t \geq 1$, first the cut player $\mathcal{C}$ chooses a bisection $(S_t, \overline{S}_t)$ of V. This choice may depend on the actions of the players in the pervious rounds and, in particular, G_{t-1}. The matching player $\mathcal{M}$, then, chooses a perfect matching M_t that matches each vertex in S_t to a unique vertex in $\overline{S}_t$. The matching M_t is not required to belong to G_{t-1}. This choice may also depend on the actions of the players in the pervious rounds and also on $(S_t, \overline{S}_t)$. The matching is then added to the graph G_{t-1} to obtain a (multi-)graph G_t. Thus $G_t := G_{t-1} + M_t$ where the sum is interpreted as the multi-set union. The game $\mathcal{G}(n, c)$ ends when $\phi_c(G_t) \geq 1/4$, i.e., each c-balanced cut has expansion at least $1/4$. The bound of $1/4$ on the expansion is arbitrary and can be replaced by any constant at most 1.

The value $\text{VAL}(n, \mathcal{C}, \mathcal{M})$ of the game is the total number of rounds, i.e., iterations of the while loop before the game ends. The goal of the cut player is to minimize $\text{VAL}(n, \mathcal{C}, \mathcal{M})$ while that of the matching player is to maximize it.

Connection with the SPARSEST CUT problem. The following lemma states a connection between the game $\mathcal{G}(n, 0)$ and the SPARSEST CUT problem.

Lemma 2.1. *If there is a strategy $\mathcal{C}$ of the cut player in $\mathcal{G}(n, 0)$ that ensures $\max_{\mathcal{M}} \text{VAL}(n, \mathcal{C}, \mathcal{M}) \leq f(n)$ and that runs in time $T(n)$ per round, there is an $O(f(n))$-approximation algorithm for the* SPARSEST CUT *problem on any undirected graph H on n vertices and m edges that runs in time $O(f(n) \cdot (T(n) + T_{flow}))$ where $T_{flow} = \tilde{O}(m^{3/2})$ is the time to compute a single-commodity maximum flow in H.*

For the proof, we refer the reader to [7]. A similar connection exists between the game $\mathcal{G}(n, c)$ for $0 < c \leq 1/2$ and the balanced separator problem. We omit the details.

3. SDP Based Lower Bound

We first recall the well-studied SDP relaxation 3.1 for the SPARSEST CUT problem introduced in [2]:

$$\textsc{sdp}: \begin{array}{rrcl} \text{Minimize} & \frac{n}{4}\sum_{(i,j)\in E}\|\mathbf{v}_i-\mathbf{v}_j\|^2 & & \\ \text{s.t.} & & & \\ \forall\, i,j,k\in V & \|\mathbf{v}_i-\mathbf{v}_j\|^2+\|\mathbf{v}_j-\mathbf{v}_k\|^2 & \geq & \|\mathbf{v}_i-\mathbf{v}_k\|^2 \\ & \frac{1}{4}\sum_{i<j}\|\mathbf{v}_i-\mathbf{v}_j\|^2 & = & 1 \end{array} \tag{3.1}$$

With each vertex $i \in V$, we associate an n-dimensional vector $\mathbf{v}_i \in \mathbb{R}^n$ and try to minimize $\frac{n}{4}\sum_{(i,j)\in E}\|\mathbf{v}_i-\mathbf{v}_j\|^2$ subject to the given "triangle inequality w.r.t. ℓ_2^2" and normalization constraints. It is easy to see that half the optimum value of the SDP is a lower bound on the sparsity of the sparsest cut in G as follows. Given a cut $(S,\overline{S})$ with $|S| \leq \frac{n}{2}$ in G, let $\mathbf{v}_i := \frac{1}{\sqrt{|S||\overline{S}|}}(-1,0,\ldots,0)^\dagger$ if $i \in S$ and $\mathbf{v}_i := \frac{1}{\sqrt{|S||\overline{S}|}}(1,0,\ldots,0)^\dagger$ if $i \notin S$. It is easy to see that this forms a feasible solution to the SDP and has value $\frac{n}{|S||\overline{S}|}\cdot|E(S,\overline{S})| \leq 2\frac{|E(S,\overline{S})|}{|S|}$, where $E(S,\overline{S})$ is the set of edges crossing the cut $(S,\overline{S})$.

Let $\textsc{gap}(n)$ denote the integrality gap, i.e., the maximum ratio of the sparsity to the value of the above SDP, on graphs with n vertices. We know that $\textsc{gap}(n) = \Omega(\log\log n)$ from a result of Devanur, Khot, Saket and Vishnoi [5]. We prove the following theorem which implies that there is no cut player $\mathcal{C}$, even unbounded-time, such that $\textsc{val}(n,\mathcal{C},\mathcal{M}) = o(\sqrt{\textsc{gap}(n)})$ for all $\mathcal{M}$.

Theorem 3.1. *There is a matching player $\mathcal{M}^*$ for the game $\mathcal{G}(n,0)$ to ensure that*

$$\min_{\mathcal{C}} \textsc{val}(n,\mathcal{C},\mathcal{M}^*) = \Omega(\sqrt{\textsc{gap}(n)}).$$

The rest of the section is devoted to proving Theorem 3.1.

3.1. Preliminaries

For a symmetric matrix $M \in \mathbb{R}^{n\times n}$, let $\mathcal{L}(M) \in \mathbb{R}^{n\times n}$ be its Laplacian which is defined below.

$$\mathcal{L}(M)_{ij} = \begin{cases} \sum_{j\neq i} M_{ij} & \text{if } i=j \\ -M_{ij} & \text{if } i\neq j. \end{cases}$$

The following are simple facts about the Laplacian of a matrix.

Fact 3.2. *For any non-negative symmetric matrix M, the Laplacian $\mathcal{L}(M)$ is positive semidefinite.*

Proof. For all $\mathbf{x} = (x_1,\ldots,x_n) \in \mathbb{R}^n$, we have $\mathbf{x}^\dagger\mathcal{L}(M)\mathbf{x} = \sum_{i<j} M_{ij}(x_i-x_j)^2 \geq 0$, as $M_{ij} \geq 0$. ∎

Fact 3.3. *If M,N are two symmetric matrices such that $M_{ij} \geq N_{ij}$ for all i,j, then $\mathcal{L}(M) \succeq \mathcal{L}(N)$.*

Proof. Notice $\mathcal{L}(M)-\mathcal{L}(N) = \mathcal{L}(M-N)$. But $M-N$ is symmetric and non-negative as $M_{ij} \geq N_{ij}$. Hence, by the previous fact, $\mathcal{L}(M-N) \succeq 0$. ∎

For a graph G, let $\mathcal{L}(G)$ denote the Laplacian of the adjacency matrix of G. Let K_n denote the complete graph on n vertices and its adiacency matrix. The following is a simple characterization of the second smallest eigenvalue of M in terms of $\mathcal{L}(M)$ and $\mathcal{L}(K_n)$.

Lemma 3.4. *For a non-negative symmetric matrix $M \in \mathbb{R}^{n\times n}$, the maximum z such that $\mathcal{L}(M) - z\mathcal{L}(K_n) \succeq 0$ is $\frac{\lambda_2(M)}{n}$, where $\lambda_2(M)$ denotes the second smallest eigenvalue of $\mathcal{L}(M)$.*

Proof. First notice that the minimum eigenvalue of both $\mathcal{L}(M)$ and $\mathcal{L}(K_n)$ is 0, with eigenvector $\mathbf{1} := (1, 1, \ldots, 1)$. Let λ_i be the i-th smallest eigenvalue of $\mathcal{L}(M)$. Consider an orthonormal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ of eigenvectors for $\mathcal{L}(M)$, where $\mathbf{u}_i$ is associated with the eigenvalue λ_i. We have $\lambda_1 = 0$ with $\mathbf{u}_1 = \frac{\mathbf{1}}{\sqrt{n}}$. Finally, notice that for all $i > 1$, $\mathbf{u}_i^\dagger \mathcal{L}(K_n)\mathbf{u}_i = n$ as $\mathcal{L}(K_n) = nI - J$, where J is the matrix with all entries set to 1, and $\mathbf{u}_i \perp \mathbf{1}$. Further, for any $\mathbf{x} \in \mathbb{R}^n$, let $(x_1, \ldots, x_n)$ be such that $\mathbf{x} = \sum_i x_i \mathbf{u}_i$. Then

$$\mathbf{x}^\dagger(\mathcal{L}(M) - z\mathcal{L}(K_n))\mathbf{x} = \sum_{i=1}^{n} x_i^2 \mathbf{u}_i^\dagger(\mathcal{L}(M) - z\mathcal{L}(K_n))\mathbf{u}_i =$$

$$\sum_{i=2}^{n} x_i^2 \mathbf{u}_i^\dagger(\mathcal{L}(M) - z\mathcal{L}(K_n))\mathbf{u}_i = \sum_{i=2}^{n} x_i^2(\lambda_i - zn).$$

This will be positive for all $\mathbf{x}$ if and only if $\lambda_2 \geq zn$. Hence, the maximum z such that $\mathcal{L}(M) - z\mathcal{L}(K_n) \succeq 0$ is $z = \frac{\lambda_2}{n}$ as required. ∎

Another fundamental fact is the relation between $\lambda_2(G)$ and expansion for a t-regular graph [4]:

Theorem 3.5 (Cheeger's Bound). *For an undirected t-regular graph G, we have:*

$$\lambda_2(G) \geq \frac{(\phi_0(G))^2}{2t}$$

For the lower bound, it would be more convenient for us to work with the dual of the SDP (3.1). We first rewrite the primal replacing $\langle \mathbf{v}_i, \mathbf{v}_j \rangle$ with a variable z_{ij} for all i, j. Moreover, we let c_{ij} to be the (i, j) entry of the adjacency matrix of the instance graph G. We also let $\mathcal{P}_{ij}$ be the set of undirected paths between i and j in K_n. (We do not distinguish between the sets $\mathcal{P}_{ji}$ and $\mathcal{P}_{ij}$.) Then, the primal can be expressed as below.

$$\begin{array}{rccl}
\text{Minimize} & \frac{n}{4}\sum_{i<j} c_{ij}x_{ij} & & \\
\text{s.t.} & & & \\
\forall i,j\ \forall p \in \mathcal{P}_{ij} & \sum_{(u,v)\in p} x_{uv} \geq & x_{ij} & \\
& \frac{1}{4}\sum_{i<j} x_{ij} & = & 1 \\
\forall i,j & x_{ij} & = & z_{ii} + z_{jj} - 2z_{ij} \\
& Z \succeq 0 & &
\end{array} \tag{3.2}$$

The SDP-dual is:

$$\begin{array}{rlcl}
\text{Maximize} & 4z' & & \\
\text{s.t.} & & & \\
\forall i,j & z' + s_{ij} - \sum_{p\in\mathcal{P}_{ij}} f_{i,j,p} + \sum_{p\ni(i,j)} f_{i,j,p} & = & \frac{n}{4}c_{ij} \\
\forall p & f_p \geq 0 & & \\
& \mathcal{L}(S) \succeq 0 & &
\end{array} \tag{3.3}$$

By substituting the s variables out and rescaling, we obtain the following simpler form of the dual:

$$\text{DUAL-SDP}: \quad \begin{array}{cc}
\text{Maximize} & n \cdot z \\
\text{s.t.} & \\
& \mathcal{L}(G) - z\mathcal{L}(K_n) + \mathcal{L}(D) - \mathcal{L}(F) \succeq 0
\end{array} \tag{3.4}$$

Here, the variables are z, F and D. The matrices F and D arise as follows: F_{ij} is defined to be the the total flow on the edge (i, j), while D_{ij} denotes the total flow between i and j. Formally,

$$F_{ij} := \sum_{\substack{s,\ t,\ p \in \mathcal{P}_{st}: \\ p \ni (i,j)}} f_{s,t,p} \qquad \text{and} \qquad D_{ij} := \sum_{p \in \mathcal{P}_{ij}} f_{i,j,p}.$$

If F is "routable" in G, then $G_{ij} \geq F_{ij}$, and hence, $\mathcal{L}(G) \succeq \mathcal{L}(F)$ by Lemma 3.4. Therefore, the objective of the DUAL-SDP (3.4) is at least $\lambda_2(D)$. It follows from strong duality[2] that the optimum of the primal is the same as that of the dual. For a graph G, we denote this optimum to be $\text{SDP}(G)$.

Finally, let G_n be a graph on n vertices for which the integrality gap of the SDP relaxation in 3.1 is $\text{GAP}(n)$. Let $\phi_0(G_n) := \alpha_n$ and $\text{SDP}(G_n) = \sigma_n$, so that $\frac{\alpha_n}{\sigma_n} = \text{GAP}(n)$. Furthermore, we may assume that $\frac{1}{\alpha_n}$ is an integer. Recall that [5] has showed that $\text{GAP}(n) = \Omega(\log \log n)$.

3.2. Proof of Theorem 3.1

Strategy $\mathcal{M}^*$ of the Matching player in $\mathcal{G}(n, 0)$:

- Given a bisection $(S, \overline{S})$, find a perfect matching M across $(S, \overline{S})$ which is routable in $\frac{1}{\alpha_n} G_n$, where $\frac{1}{\alpha_n} G_n$ is the graph obtained from G_n by multiplying each edge-capacity by $\frac{1}{\alpha_n}$.
- Output the matching M.

Figure 2: A strategy $\mathcal{M}^*$ of the matching player that ensures $\min_{\mathcal{C}} \text{VAL}(n, \mathcal{C}, \mathcal{M}^*) = \Omega(\sqrt{\text{GAP}(n)})$

Now we show that there is a strategy for the matching player such that the cut player cannot end the game before $\Omega(\sqrt{\text{GAP}(n)})$ rounds. For a given n (recall that n is even in our setting), the matching player bases his strategy on the graph G_n. A bisection $(S, \overline{S})$ is interpreted as a cut in G_n, and the matching player produces a perfect matching across $(S, \overline{S})$ which is routable in $\frac{1}{\alpha_n} G_n$, where $\frac{1}{\alpha_n} G_n$ is the graph obtained from G_n by multiplying each edge-capacity by $\frac{1}{\alpha_n}$. The next lemma (from [7]) shows that there exists such a matching.

Lemma 3.6. *Given a graph G_n with expansion α_n such that $1/\alpha_n$ is integral, for every bisection $(S, \overline{S})$, it is possible to construct a pair of graphs (M_S, F_S) such that M_S is a perfect matching across $(S, \overline{S})$ and that F_S is a flow routing M_S in $\frac{1}{\alpha_n} G_n$, by doing a single commodity max-flow computation in G_n.*

Proof. Consider a max-flow problem set up on $\frac{1}{\alpha_n} G_n$ as follows: every vertex in S is connected to an auxiliary vertex s with an edge of capacity 1. Similarly, every vertex in $\overline{S}$ to an auxiliary vertex t with an edge of capacity 1. Since the capacity of the cut separating s from S is $n/2$, the min-cut in this new graph is at most $n/2$. Suppose now that the min-cut had value strictly less than $n/2$ and let the number of edges in the min-cut incident to the source s (resp. sink t) be n_s (resp. n_t). The remaining capacity of the cut is strictly less than $n/2 - n_s - n_t$, and thus uses at most $(n/2 - n_s - n_t)$ capacity in the graph $\frac{1}{\alpha_n} G_n$. Moreover, the cut consisting of edges in the graph separates at least $n/2 - n_s$ vertices in S from $n/2 - n_t$ vertices in $\overline{S}$. The expansion of this cut is strictly less than $(n/2 - n_s - n_t)/\min(n/2 - n_s, n/2 - n_t)$, i.e. strictly less than 1. But α_n is the expansion of G_n, so $\frac{1}{\alpha_n} G_n$ must have expansion equal to 1. Hence, the min-cut must have

[2] In our case we just need weak duality.

value $n/2$. By the max-flow-min-cut theorem for single commodity, there is a flow F_S of value $n/2$ between s and t. This flow F_S can be chosen to be integral as all the edge capacities are integral. Since F_S is integral and the capacity of each of the edges connecting s to S and t to $\overline{S}$ are 1, the flow paths connecting vertices in S to those in $\overline{S}$ must be edge disjoint. This gives rise to a matching M_S between the vertices in S and $\overline{S}$. ■

Let M_S be the matching output by the matching player when presented with partition $(S, \overline{S})$. We let F_S and M_S also denote the adjacency matrices corresponding to the flow F_S and the matching M_S respectively. The matching player gets the bisections $(S_1, \overline{S_1})$, $(S_2, \overline{S_2}), \ldots$ $(S_t, \overline{S_t})$, and it outputs the matchings $M_i := M_{S_i}$ (along with the corresponding flows $F_i := F_{S_i}$ routable in $\frac{1}{\alpha_n} G_n$) based on the procedure described above.

Using the pairs (M_i, F_i), for $1 \le i \le t$, we produce a feasible solution to the SDP (3.4) for G_n as follows. We let $F := \frac{\alpha_n}{t}(F_1 + \cdots + F_t)$, and $D := \frac{\alpha_n}{t}(M_1 + \cdots + M_t)$. Since each M_i is routable in $\frac{1}{\alpha_n} G_n$ with routing F_i, the pair (F, D) is feasible for the DUAL-SDP (3.4). Since each F_i is routable in $\frac{1}{\alpha_n} G_n$, F is routable in G and hence we have $\text{SDP}(G_n) \ge \lambda_2(D)$. By Cheeger's Inequality, and the fact that the union of matchings is a t-regular graph, it follows that

$$\lambda_2(D) \ge \frac{\alpha_n}{t} \frac{(\phi_0(M_1 + \cdots + M_t))^2}{2t} \ge \frac{\alpha_n}{32t^2}$$

Hence,

$$\sigma_n \ge \text{SDP}(G_n) \ge \lambda_2(D) \ge \frac{\alpha_n}{32t^2}.$$

So, $t \ge \sqrt{\frac{\alpha_n}{32\sigma_n}} = \Omega(\sqrt{\text{GAP}(n)})$. This completes the proof of Theorem 3.1.

4. Upper Bound

The following theorem is the main result of this section.

Theorem 4.1. *There is a strategy $\mathcal{C}^*$ of the cut player in the game $\mathcal{G}(n, 0)$ to ensure that*

$$\max_{\mathcal{M}} \text{VAL}(n, \mathcal{C}^*, \mathcal{M}) = O(\log n).$$

To prove the above theorem, we first prove the following result.

Theorem 4.2. *In the game $\mathcal{G}(n, c)$, the strategy $\mathcal{C}^*$ of the cut player (see Figure 3) that chooses a bisection in the current graph G that does not "cross" the c-balanced cut with minimum expansion ensures that* $\max_{\mathcal{M}} \text{VAL}(n, \mathcal{C}^*, \mathcal{M}) = O(\frac{1}{c} \log n)$.

Strategy $\mathcal{C}^*$ of the cut player in $\mathcal{G}(n, c)$:

- Find a c-balanced cut $(T, \overline{T})$ with minimum expansion, with, say $|T| \le |\overline{T}|$. Let S be an *arbitrary* superset of T such that $|S| = n/2$.
- Output the bisection $(S, \overline{S})$.

Figure 3: A strategy $\mathcal{C}^*$ of the cut player in $\mathcal{G}(n, c)$ that ensures that $\max_{\mathcal{M}} \text{VAL}(n, \mathcal{C}^*, \mathcal{M}) = O(\frac{1}{c} \log n)$

We first introduce some technical tools to prove Theorem 4.2. These tools were also used by [7]. Recall that in the game, the graph G starts from being empty and after k rounds it is the (multi-set) union of k perfect matchings $M_1, M_2, \ldots, M_k$ each of size $n/2$. Given a sequence $\{M_1, \ldots, M_k\}$ of perfect matchings, we associate, with each starting vertex $u \in V$, a k-step random walk on the vertex set V. Imagine a particle located at vertex $u \in V$. Assume that in step t (where $t = 1, \ldots, k$), the particle stays at its current position with probability 1/2 and jumps across the (unique) edge incident to the current position in the tth matching M_t. Let $p_{u,v}(t)$ denote the probability that the particle starting at u reaches vertex v after t steps. It is easy to see that the probabilities $p_{u,v}(t)$ satisfy the following properties.

Lemma 4.3. *(1) $p_{u,u}(0) = 1$ for all u and $p_{u,v}(0) = 0$ for $u \neq v$.*
(2) $p_{u,v}(t) = \frac{p_{u,v}(t-1)+p_{u,w}(t-1)}{2}$ for $t \geq 1$ and $(v, w) \in M_t$.
(3) $\sum_{v \in V} p_{u,v}(t) = 1$ for all u and $t \geq 0$.
(4) $\sum_{v \in V} p_{v,u}(t) = 1$ for all u and $t \geq 0$.

Proof. Items 1 and 2 follow from the initialization and the averaging update explained above. Items 3 and 4 can be easily proved by induction on t. ■

We associate a potential $\Psi(k)$ with the sequence $\{M_1, \ldots, M_k\}$ of perfect matchings. Unlike [7] who use ℓ_2^2 potential, we use simply the sum of the *entropies* of all the n random walks:

$$\Psi(k) = \sum_{u \in V} \left(- \sum_{v \in V} p_{u,v}(k) \ln p_{u,v}(k) \right). \tag{4.1}$$

Here we use the usual convention that $-0 \ln 0 = \lim_{x \to 0^+} -x \ln x = 0$. Observe that the potential satisfies the following properties.

Lemma 4.4. *(1) $\Psi(0) = 0$.*
(2) $\Psi(t) \leq n \ln n$ for all t.
(3) $\Psi(t) \geq \Psi(t-1)$ for $t \geq 1$.

Proof. Item 1 follows from the definition. Item 2 follows from the fact that the entropy of a probability distribution on n points can be at most $\ln n$. Item 3 follows from the fact that the entropy does not decrease if we average the probabilities on two points: for $0 \leq p, q \leq 1$ such that $p + q \leq 1$ we have

$$-p \ln p - q \ln q \leq - \left(\frac{p+q}{2}\right) \ln \left(\frac{p+q}{2}\right) - \left(\frac{p+q}{2}\right) \ln \left(\frac{p+q}{2}\right).$$

■

4.1. Proof of Theorem 4.2

Proof Idea. Recall that the game $\mathcal{G}(n, c)$ ends when all the c-balanced cuts have expansion at least 1/4. We show that as long as there is a c-balanced cut with expansion less than 1/4, the potential Ψ increases by $\Omega(cn)$ in each round, i.e., $\Psi(t+1) - \Psi(t) = \Omega(cn)$. Since the potential does not increase beyond $n \ln n$, we get that the number of rounds is $O(\frac{1}{c} \log n)$.

To this end, we fix a round t and assume that the c-balanced cut $(T, \overline{T})$ after t rounds has expansion less than 1/4. Say $\min\{|T|, |\overline{T}|\} = |T| = bn$ where $c \leq b \leq 1/2$. This implies that $|E(T, \overline{T})| < bn/4$. We next argue that the total probability that has "crossed" from T-side to $\overline{T}$-side is at most $bn/8$, i.e., $\sum_{u \in T, v \in \overline{T}} p_{u,v}(t) \leq bn/8$. Using this, we argue that there are at least $bn/2$ vertices $u \in T$ such that the total probability starting at u that has crossed the cut $(T, \overline{T})$ is at most 1/4. We then prove that after adding the matching M_{t+1} in the sequence, the entropy of the probability distribution associated with the random walk starting at each such u increases by $\Omega(1)$. This, in turn, implies that the overall potential increases in this round by $\Omega(bn)$.

For the purpose of this section, we assume that the cut player uses the strategy given in Figure 3. The following lemma directly implies Theorem 4.2, since the potential is always at most $n \ln n$.

Lemma 4.5. *During any round of the game, the potential increases significantly:* $\Psi(t+1) - \Psi(t) = \Omega(cn)$.

To prove the above lemma, fix a round t and assume that the c-balanced cut $(T, \overline{T})$ in G_t with minimum expansion has expansion less than 1/4. Let, without loss of generality, $\min\{|T|, |\overline{T}|\} = |T| = bn$ where $c \leq b \leq 1/2$. Thus $|E(T, \overline{T})| \leq bn/4$. Recall that $p_{u,v}(t)$ denotes the probability that the particle starting at vertex u reaches v after t steps of the random walk defined above. For $u \in T$, let $q_u(t) = \sum_{v \in \overline{T}} p_{u,v}(t)$ be the total probability of the random walk starting at u that has crossed the cut $(T, \overline{T})$.

Lemma 4.6. $\sum_{u \in T} q_u(t) < bn/8$.

Proof. Note that when we add an edge $(v, w) \in M_r$ to G in round r, a total of 1/2 units of probability currently present at v (from all the starting points summed up) travels from v to w and 1/2 units of probability present at w travels from w to v. Moreover each edge is used exactly once to mix the random walks. Thus each edge in $E(T, \overline{T})$ contributes at most 1/2 unit of probability cross over from T to $\overline{T}$. This completes the proof since $E(T, \overline{T})$ has less than $bn/4$ edges. ∎

Using averaging argument, it is clear from Lemma 4.6 that there are at least $bn/2$ vertices $u \in T$ such that $q_u(t) \leq 1/4$. Fix a vertex $u \in T$ such that $q_u(t) \leq 1/4$. We prove that the contribution of u to the potential Ψ increases by $\Omega(1)$ in this round $t+1$, i.e., the entropy of the distribution $\{p_{u,v}(t+1)\}$ is more than that of the distribution $\{p_{u,v}(t)\}$ by an amount at least $\Omega(1)$.

In round $t+1$, we add a matching M_{t+1} to G. Since $T \subseteq S_{t+1}$, each $v \in T$ is matched to a unique $\pi(v) \in \overline{T}$ in M_{t+1}. Call a vertex $v \in T$ "good" if $p_{u,v}(t) \geq 2p_{u,\pi(v)}(t)$, or "bad" otherwise. The total probability present in $\overline{T}$ is$\sum_{w \in \overline{T}} p_{u,w}(t) \leq 1/4$. Therefore the total probability present on the bad vertices is at most $2 \cdot 1/4 = 1/2$. Since the total probability on T is 3/4, at least 1/4 probability is present on good vertices: $\sum_{v:\text{good}} p_{u,v}(t) \geq 1/4$.

In round $t+1$, each good vertex $u \in T$ averages its probability $p_{u,v}(t)$ with $p_{u,\pi(v)}(t)$. Note that if $p \geq 2q$, we have

$$(-p \ln p - q \ln q) - \left(-2\left(\frac{p+q}{2}\right) \ln \left(\frac{p+q}{2}\right)\right) = \Omega(p).$$

Therefore the good vertices together contribute $\Omega(\sum_{v:\text{good}} p_{u,v}(t)) = \Omega(1)$ increase in the entropy. This, in turn, implies Lemma 4.5, and therefore Theorem 4.2.

4.2. Proof of Theorem 4.1

The cut player first uses the strategy in Figure 3 for $c = 1/4$. As proved in Theorem 4.2, after $O(\log n)$ rounds, the expansion of the every c-balanced cut becomes at least 1/4. The cut player repeats this strategy 4 times to ensure that every c-balanced cut has expansion, in fact, at least 1. After this, the cut player iteratively computes the sparsest cuts in the graph as described in Figure 4. Let T be the union of all the vertices removed in the above procedure. The following lemma proves that $|T| < n/4$. In the next and the *last* round, the cut player outputs a bisection $(S, \overline{S})$ where S is an arbitrary superset of T such that $|S| = n/2$.

Lemma 4.7. $|T| < n/4$.

Strategy $\mathcal{C}^*$ of the cut player in $\mathcal{G}(n,0)$:

- Use the strategy $\mathcal{C}^*$ in Figure 3, four times, to ensure that every c-balanced cut in G has expansion at least 1.
- In the next and the *last* round do:
  (1) Let $G' \leftarrow G$.
  (2) While there is a cut in G' with expansion less than 1/2 do:
    (a) Let $(U,\overline{U})$ be the sparsest cut in G', with $|U| \leq |\overline{U}|$.
    (b) Remove vertices U (and their incident edges) from G'.
  (3) Let T be the union of all the vertices removed in the while loop above.
  (4) Let S be an arbitrary superset of T such that $|S| = n/2$.
  (5) Output the bisection $(S,\overline{S})$.

Figure 4: A strategy $\mathcal{C}^*$ of the cut player in $\mathcal{G}(n,0)$ that ensures that $\max_{\mathcal{M}} \text{VAL}(n,\mathcal{C}^*,\mathcal{M}) = O(\log n)$

Proof. Assume on the contrary that $|T| \geq n/4$. We handle two cases separately: (1) $|T| \leq n/2$, (2) $|T| > n/2$. In case 1, it is easy to argue that the expansion of $(T,\overline{T})$ in the graph G is at most 1/2. This follows from the fact that the edges in $E(T,\overline{T})$ can be charged to the vertices in the sets U whose union forms T. This, in turn, implies that $|T| < 1/4$, since every 1/4-balanced in G has expansion at least 1. This yields a contradiction.

In the second case, let $U_1, U_2, \ldots$ be the sets of vertices removed in the above procedure. Let k be the smallest index such that $|\cup_{i=1}^{k} U_i| > n/2$. Clearly we have $k \geq 2$. Let $T' = \cup_{i=1}^{k-1} U_i$. Since $|T'| \leq n/2$, we have, by an argument similar to the one above, that the expansion of $(T',\overline{T}')$ in the graph G is at most 1/2. This implies that $|T'| < n/4$, again since every 1/4-balanced cut in G has expansion at least 1. Thus we have that $|U_k| \geq n/4 > |\cup_{i=1}^{k-1} U_i|$. Since each U_i has expansion less than 1/2 in the graph $G \setminus \cup_{j=1}^{i-1} U_j$, we get that the expansion of $(U_k,\overline{U}_k)$ in G is less than 1. This follows from the fact that $|U_k| \leq n/2$ and that the edges in $E(U_k,\overline{U}_k)$ can be charged to the vertices in $\cup_{i=1}^{k} U_i$ and from the fact that $|U_k| > |\cup_{i=1}^{k-1} U_i|$. This leads to a contradiction, again, since every 1/4-balanced cut in G has expansion at least 1. ∎

We now show that the graph G becomes an expander after the last round described in Figure 4. Let l be the index of the last round and let G_l and M_l be the graph after l rounds and the perfect matching added in the lth round. We show that the expansion of any cut in G_l is at least 1/3. Let $(W,\overline{W})$ be any cut in G_l with $|W| \leq |\overline{W}|$.

Lemma 4.8. *The expansion on* $(W,\overline{W})$ *in* G_l *is at least* 1/3*:* $|E(W,\overline{W})|/|W| \geq 1/3$.

Proof. We consider two cases: (1) $|W \cap \overline{T}| \geq |W|/3$, (2) $|W \cap \overline{T}| < |W|/3$. Consider the first case. From the procedure to construct T in the last round, it is clear that every cut in $G \setminus T$ has expansion at least 1. Since at least $|W|/3$ vertices in W lie in $G \setminus T$, we have that $|E(W,\overline{W})|$ in G_l is at least $|W|/3$, establishing the expansion of 1/3.

In the second case, it is easy to see that at least $|W| - 2|W \cap \overline{T}| \geq |W|/3$ edges in M_l belong to $E(W,\overline{W})$ in G_l, thereby again establishing the expansion of 1/3. ∎

This completes the proof of Theorem 4.1.

5. Concluding Remarks and Open Problems

- The main open problem is to establish tight bounds for $\max_{\mathcal{M}} \min_{\mathcal{C}} \textsc{val}(n, \mathcal{C}, \mathcal{M})$ and $\min_{\mathcal{C}} \max_{\mathcal{M}} \textsc{val}(n, \mathcal{C}, \mathcal{M})$ for the cut-matching game. From the lower bound perspective, it would be extremely interesting to improve the lower bound beyond $\Omega(\sqrt{\log \log n})$. The current lower bound relies on the non-trivial integrality gap construction for SDP (3.1). Ideally, one would like to improve the lower bound to $\Omega(\sqrt{\log n})$ without relying on the integrality gap constructions (which are harder to come by), while at the same time establishing that the Khandekar, Rao and Vazirani paradigm is no more powerful than that of Arora, Rao and Vazirani.
- Another variant, which we refer to as the game of *flow* may be of independent interest. The game of flow proceeds like the expansion game, but stops when G (with unit capacity on each of its edges) can support a concurrent uniform multi-commodity flow in which each vertex sends $1/n$ flow to every other vertex. This is equivalent to saying that the complete graph K_n is routable in G with congestion at most n. This game has a connection with the max-flow-min-cut gap for the uniform multi-commodity flows in undirected graphs. In particular, a strategy $\mathcal{C}^*$ of the cut player that ensures $\max_{\mathcal{M}} \textsc{val}(n, \mathcal{C}^*, \mathcal{M}) \leq f(n)$ for the flow game implies a $f(n)$ max-flow-min-cut gap for the (uniform) multi-commodity flows. Thus a cut player strategy with $f(n) = O(\log n)$ would yield an alternate proof of $O(\log n)$ max-flow-min-cut gap for undirected graphs [8].
- All our results can be appropriately generalized to the case when the cut player is supposed to output only a balanced cut and not necessarily a bisection.

References

[1] S. Arora, E. Hazan, and S. Kale. $O(\sqrt{\log n})$ approximation to sparsest cut in $\tilde{O}(n^2)$ time. In *Proceedings, IEEE Symposium on Foundations of Computer Science*, pages 238–247, 2004.

[2] S. Arora, S. Rao, and U. Vazirani. Expander flows, geometric embeddings, and graph partitioning. In *Proceedings, ACM Symposium on Theory of Computing*, pages 222–231, 2004.

[3] Sanjeev Arora and Satyen Kale. A combinatorial, primal-dual approach to semidefinite programs. In *STOC*, pages 227–236, 2007.

[4] Fan R.K. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.

[5] N. R. Devanur, S. Khot, R. Saket, and N. K. Vishnoi. Integrality gaps for sparsest cut and minimum linear arrangement problems. In *Proceedings, ACM Symposium on Theory of Computing*, pages 537–546, 2006.

[6] George Karypis and Vipin Kumar. A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM Journal on Scientific Computing*, 20:359 – 392, 1999.

[7] R. Khandekar, S. Rao, and U. Vazirani. Graph partitioning using single commodity flows. In *Proceedings, ACM Symposium on Theory of Computing*, pages 385–390, 2006.

[8] Tom Leighton and Satish Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *Journal of the ACM (JACM)*, 46(6):787–832, 1999.

[9] Lorenzo Orecchia Leonard Schulman, Umesh V. Vazirani and Nisheeth K. Vishnoi. On partitioning graphs using single commodity flows. Manuscript.

[10] David Shmoys. *Approximation algorithms for NP-hard problems*, chapter Cut problems and their application to divide and conquer, pages 192–235. PWS Publishing Company, 1997.