Min-Max Correlation Clustering via MultiCut*

Saba Ahmadi[1], Samir Khuller[2], and Barna Saha[3]

[1] University of Maryland, College Park
{saba}@cs.umd.edu
[2] Northwestern University
{samir.khuller}@northwestern.edu
[3] College of Information & Computer Science
University of Massachussetts Amherst
barna@cs.umass.edu

Abstract. Correlation clustering is a fundamental combinatorial optimization problem arising in many contexts and applications that has been the subject of dozens of papers in the literature. In this problem we are given a general weighted graph where each edge is labeled positive or negative. The goal is to obtain a partitioning (clustering) of the vertices that minimizes disagreements – weight of negative edges trapped inside a cluster plus positive edges between different clusters. Most of the papers on this topic mainly focus on minimizing total disagreement, a global objective for this problem.
In this paper we study a cluster-wise objective function that asks to minimize the maximum number of disagreements of each cluster, which we call min-max correlation clustering. The min-max objective is a natural objective that respects the quality of every cluster. In this paper, we provide the first nontrivial approximation algorithm for this problem achieving an $\mathcal{O}(\log n)$ approximation for general weighted graphs. To do so, we also obtain a corresponding result for multicut where we wish to find a multicut solution while trying to minimize the total weight of cut edges on every component. The results are then further improved to obtain an $\mathcal{O}(r^2)$-approximation for min-max correlation clustering and min-max multicut for graphs that exclude $K_{r,r}$ minors.

Keywords: Correlation Clustering · Multicut · Approximation Algorithms

1 Introduction

Correlation clustering is a fundamental optimization problem introduced by Bansal, Blum and Chawla [3]. In this problem, we are given a general weighted graph where each edge is labeled positive or negative. The goal is to obtain a

* A full version of this paper appears at http://cs.umd.edu/~samir/LCC.pdf
The first two authors are supported by NSF grant CNS 156019. Part of the research was done when the third author was visiting the Simons Institute of Theory of Computing and the author is supported by NSF CAREER 1652303, NSF CCF 1464310 and a Google faculty award.

partitioning of the vertices into an arbitrary number of clusters that agrees with the edge labels as much as possible. That is a clustering that minimizes disagreements, which is the weight of positive edges between the clusters plus the weight of negative edges inside the clusters. In addition, correlation clustering captures some fundamental graph cut problems including min s-t cut, multiway cut and multicut. Correlation clustering has been studied extensively for more than a decade [1,2,6,7,10]. Most of the papers have focused on a global min-sum objective function, i.e. minimizing total number of disagreements or maximizing the total number of agreements.

In recent work, Puleo and Milenkovic [14] introduced a local vertex-wise min-max objective for correlation clustering which bounds the maximum number of disagreements on each node. This problem arises in many community detection applications in machine learning, social sciences, recommender systems and bioinformatics [8,16,13]. This objective function makes sure each individual has a minimum quality within the clusters. They showed this problem is NP-hard even on un-weighted complete graphs, and developed an $O(1)$ approximation algorithm for unweighted complete graphs. Charikar et al. [5] improved upon the work by Puleo et al. [14] for complete graphs by giving a 7 approximation. For general weighted graphs, their approximation bound is $O(\sqrt{n})$ where n is the number of vertices. Both these algorithms rely on LP rounding, based on a standard linear program relaxation for the problem. In contrast, for the global minimization objective an $O(\log n)$-approximation can be obtained [10]. Therefore, the local objective for correlation clustering seems significantly harder to approximate than the global objective.

In this work, we propose a new local cluster-wise min-max objective for correlation clustering – minimizing the maximum number of disagreements of each cluster. This captures the case when we wish to create communities that are harmonious, as global min sum objectives could create an imbalanced community structure. This new local objective guarantees fairness to communities instead of individuals. To name a few applications for this new objective, consider a task of instance segmentation in an image which can be modeled using correlation clustering [12,11]. A cluster-wise min-max objective makes sure each detected instance has a minimum quality. Another example is in detecting communities in social networks, this objective makes sure there are no communities with lower quality compared to the other communities. No hardness results are known for the cluster-wise min-max objective.

A similar objective was proposed for the multiway cut problem by Svitkina and Tardos [15]. In the min-max multiway cut problem, given a graph G and k terminals the goal is to get a partitioning of G of size k that separates all terminals and the maximum weight of cut edges on each part is minimized. Svitkina and Tardos [15] showed an $\mathcal{O}(\log^3 n)$ approximation algorithm for min-max multiway cut on general graphs (this bound immediately improves to $\mathcal{O}(\log^2 n)$ by using better bisection algorithms). Bansal et al. [4] studied a graph partitioning problem called min-max k-partitioning from a similar perspective. In this problem, given a graph $G = (V, E)$ and $k \geq 2$ the goal is to partition the vertices into

k roughly equal parts $S_1, \cdots, S_k$ while minimizing $\max_i \delta(S_i)$. They showed an $\mathcal{O}(\sqrt{\log n \log k})$ approximation algorithm for this problem. They also improved the approximation ratio given by Svitkina et al. [15] for min-max multiway cut to $\mathcal{O}(\sqrt{\log n \log k})$. Bansal et al's seminal work [4] uses the concept of orthogonal separators introduced by Chlamtac et al. [9] to achieve their result.

2 Results & High Level Ideas

In this paper, we give an approximation algorithm for the problem of min-max correlation clustering.

Definition 1. *(Min-max Correlation Clustering) Let $G = (V, E)$ be an edge-weighted graph such that each edge is labeled positive or negative. The min-max correlation clustering problem asks for a partioning of the nodes (a clustering) where the maximum disagreement of a cluster is minimized. Disagreement of a cluster C is the weight of negative edges with both endpoints inside C plus the weight of positive edges with exactly one endpoint in C.*

We prove the following theorem for min-max correlation clustering.

Theorem 1 *Given an edge weighted graph $G = (V, E)$ on n vertices such that each edge is labeled positive or negative, there exists a polynomial time algorithm which outputs a clustering $\mathcal{C} = \{C_1, \cdots, C_{\mathcal{C}}\}$ of G such that the disagreement on each $C_i \in \mathcal{C}$ is at most $\mathcal{O}(\log(n)) \cdot OPT$; where OPT is the maximum disagreement on each cluster in an optimal solution of min-max correlation clustering.*

In order to prove Theorem 1, we give a reduction from the problem of min-max correlation clustering to a problem which we call min-max multicut.

Definition 2. *(Min-max Multicut) Given an edge weighted graph $G = (V, E)$ and a set of source-sink pairs $\{(s_1, t_1), \cdots, (s_T, t_T\}$, the goal is to give a partitioning $\mathcal{P} = \{P_1, P_2, \cdots, P_{|\mathcal{P}|}\}$ of G such that all the source sink pairs are separated, and $\max_{1 \le i \le |\mathcal{P}|} \delta(P_i)$ is minimized.*

In min-max multicut, we do not force each part of the partitioning to have a terminal and there could be some parts without any terminals in the final solution. However, in the min-max multiway cut problem introduced by Svitkina and Tardos [15], each part needs to have exactly one terminal. We prove the following theorem for min-max multicut:

Theorem 2 *Given an edge weighted graph $G = (V, E)$ on n vertices, and a set of source sink pairs $S_G = \{(s_1, t_1), \cdots, (s_T, t_T)\}$, there exists a polynomial time algorithm which outputs a partitioning $\mathcal{P} = \{P_1, \cdots, P_{|\mathcal{P}|}\}$ of G such that all source sink pairs are separated, and $\max_{1 \le i \le |\mathcal{P}|} \delta(P_i) \le \mathcal{O}(\log(n)) \cdot OPT$; where OPT is the value of the optimum solution of min-max multicut.*

We also consider the following variation of min-max multicut called min-max constrained multicut. In this variation, the goal is to partition a graph into a minimum number of parts to separate all source-sink pairs.

Definition 3. *(Min-max Constrained Multicut) An edge weighted graph $G = (V, E)$ and a set of source-sink pairs $\{(s_1, t_1), \cdots, (s_T, t_T)\}$ is given. Given k the* *minimum number of parts needed to separate all source sink pairs, the goal is to partition* G *into* k *parts* $\{P_1, \cdots, P_k\}$ *which separate all source-sink pairs, and* $\max_{1 \leq i \leq k} \delta(P_i)$ *is minimized.*

We defer our results for this problem to the full version of this paper. Finally, we get improved approximation ratios for min-max correlation clustering, min-max multicut on graphs excluding a fixed minor.

Theorem 3 *Given an edge weighted graph* G *excluding* $K_{r,r}$ *minors, there exist polynomial time* $\mathcal{O}(r^2)$*-approximation algorithms for min-max correlation clustering and min-max multicut.*

2.1 High Level Ideas

Most algorithms for correlation clustering with the global minimizing disagreement objective use a linear programming relaxation [7,10,6]. The recent work of Charikar, Gupta and Scharwtz also uses a similar linear programming relaxation for the vertex-wise min-max objective [5]. Surprisingly, these relaxations do not work for the min-max correlation clustering problem considered in this paper. Indeed, simply obtaining a linear programming relaxation for the cluster-wise min-max objective looks challenging!

Bansal et al. [4] considered a semidefinite programming (SDP) based approximation algorithm for min-max k balanced partitioning and min-max multiway cut with k terminals. In their approach, instead of finding the entire solution in one shot, they obtain a single part at a time. It is possible to encode the same problem with a linear program albeit with a worse approximation guarantee. They use SDP rounding to obtain a part with low cut capacity, and repeat the process until the parts produce a covering of all the vertices. By properly adjusting the weight of each part, the covering can be obtained efficiently. Finally, they convert the covering to partitioning.

The problem of extracting a single cluster of min-max correlation clustering can be captured by a semidefinite programming formulation. Here it is not over a cut capacity objective, instead we need to simultaneously consider the intra-cluster negative edges as well as inter-cluster positive edges. Indeed, even for the global minimization objective, we are not aware of any good rounding algorithm based on SDP relaxation of correlation clustering. Therefore, rounding the SDP formulation directly looks difficult. To overcome this, we instead consider a new problem of *min-max multicut*. Demaine et al. [10] have shown an approximation preserving reduction between multicut and correlation clustering (for the global objective function). By solving the min-max multicut problem and then using the aforementioned reduction, we solve the min-max correlation clustering problem.

First, the reduction of Demaine et al. [10] is for the global objective, and an equivalence in global objective does not necessarily correspond to equivalency in local min-max objective. Fortunately, we could show indeed such an equivalency can be proven (the details are deferred to the full version). Thus, the “multicut” route seems promising as it optimizes over a cut objective. We consider obtaining each component of the min-max multicut problem, repeat this process to obtain a covering [4], and finally convert the covering to a partitioning.

The major technical challenge comes in rounding the SDP relaxation for the multicut instance where we seek to find a single component with good cut property. In order for the relaxation to be valid, we have to add new constraints so that no source-sink pair (s_i, t_i) appears together. We also need to ensure that the component obtained satisfies a weight lower bound by assigning weights to each vertex. This is important in the next step when we wish to get a covering of all the vertices: we will decrease the weight of the vertices in the component recovered and again recompute the SDP relaxation with the same weight lower bound. This ensures the same component is not repeatedly recovered and a final covering can be obtained. To solve min-max multiway cut, Bansal et al. [4] need to separate k terminals. To do so, they can just guess which of the k terminals if any should appear in the current component with only $k+1$ guesses. For us, the number of such guesses would be 3^T where T is the number of source sink pairs since for every pair (s_i, t_i), either s_i or t_i or none would be part of the returned component. Since T could be $O(n^2)$ such a guessing is prohibitive. We need to come up with a new approach to address this issue.

We use a SDP relaxation to compute a metric on the graph vertices and add additional constraints to separate source sink pairs along with the spreading constraints from Bansal et al. [4] to recover a component of desired size. Next, we use the SDP separator technique introduced by Bansal et al. [4] to design a rounding algorithm that returns a set $S = \{S_1, S_2, \cdots, S_j\}$, such that for each $S_i \in S$, there are no source-sink pairs in S_i. Bansal et al. [4] need to glue the sets in S and report it as a single component, since they wish to get a solution with specified number of components at the end. However, in min-max multicut problem, the number of components does not matter. Therefore, we do not need to union the sets in S, and as a result no source-sink violations happen.

It is possible to use a linear programming formulation for the detour via multicut and use LP-separators of Bansal et al. [4] in place of orthogonal separators and follow our algorithm. This would achieve a similar bound for min-max multicut and min-max correlation clustering in general graphs, but a much better bound of $\mathcal{O}(r^2 \cdot OPT)$ for graphs that exclude $K_{r,r}$ minors. The details are deferred to the full version.

3 Min-Max Multicut

Given a subset $S \subseteq V$, let $\delta(S)$ denote the number of edges with exactly one end-point in S and let the number of source sink pairs (s_i, t_i) such that both s_i and t_i belong to S be $vio(S)$.

In order to prove Theorem 2, we first wish to find a set $S = \{S_1, \cdots, S_j\}$, such that $\forall S_i \in S, S_i \subseteq V$, and $\delta(S_i) \leq \mathcal{O}(\log(n)) \cdot OPT$, where OPT is the maximum number of cut edges on each part of the optimum partitioning for the min-max multicut problem on graph G. In addition, $\Pr[vio(S_i) \geq 1] \leq 1/n$, where n is the number of vertices in G.

Graph $G = (V, E)$ can have arbitrary edge weights, $w : E \to \mathbb{R}^+$. We assume graph $G = (V, E)$ is also a vertex-weighted graph, and there is a measure η on V such that $\eta(V) = 1$. This measure is used to get a covering of all the vertices. In Section 3.4, Theorem 4 is repeatedly applied to generate a family of sets that cover all the vertices. When a vertex is covered its weight is decreased so the uncovered vertices have a higher weight. Constraint $\eta(\mathcal{S}) \in \eta(\mathcal{S}) = \sum_{i=1}^{j} \eta(S_i) \in \left[H/4, 12H\right]$ makes sure the newly computed family of sets $\mathcal{S}$ has adequate coverage. Parameter $H \in (0,1)$ is equal to $1/k$ where k is the number of parts in the optimum partitioning which we guess. Since the maximum number of parts is at most n, $H \geq 1/n$.

After getting a covering of all the vertices, in Section 3.4, it is explained how to convert a covering into a partitioning with the properties desired in Theorem 2. In order to prove Theorem 1, in the full version of this paper we show how a $\mathcal{O}(\log n)$-approximation algorithm for min-max multicut implies a $\mathcal{O}(\log n)$-approximation algorithm for min-max correlation clustering.

First we prove the following theorem:

Theorem 4 *We are given an edge-weighted graph $G = (V, w)$, a set of source sink pairs S_G, a measure η on V such that $\eta(V) = 1$, and a parameter $H \in (0,1)$. Assume there exists a set $T \subseteq V$ such that $\eta(T) \in [H, 2H]$, and $vio(T) = 0$. We design an efficient randomized algorithm to find a set $\mathcal{S}$, where $\mathcal{S} = \{S_1, \cdots, S_j\}$ satisfying $\forall S_i \in \mathcal{S}, S_i \subseteq V$, $\eta(\mathcal{S}) = \sum_{i=1}^{j} \eta(S_i) \in \left[H/4, 12H\right]$, and $\forall S_i \in \mathcal{S}$, $\Pr[vio(S_i) \geq 1] \leq \frac{1}{n}$, and:*

$$\delta(S_i) \leq \mathcal{O}(\log(n)) \cdot \min\left\{\delta(T) : \eta(T) \in [H, 2H], \forall (s_i, t_i) \in S_G, |\{s_i, t_i\} \cap T| \leq 1\right\}$$

In order to prove this theorem, we use the notion of $m-$orthogonal separators, a distribution over subsets of vectors, introduced by Chlamtac et al. [9] which is explained in the following:

Definition 4. *Let X be an ℓ_2^2 space (i.e a finite collection of vectors satisfying ℓ_2^2 triangle inequalities with the zero vector) and $m > 0$. A distribution over subsets S of X is an $m-$orthogonal separator of X with probability scale $\alpha > 0$, separation threshold $0 < \beta < 1$, and distortion $D > 0$, if the following conditions hold:*

- $\forall u \in X, \Pr(u \in S) = \alpha \left\|u\right\|^2$
- $\forall u, v \in X$ *if* $\left\|u - v\right\|^2 \geq \beta \min\{\left\|u\right\|^2, \left\|v\right\|^2\}$ *then* $\Pr(u \in S$ *and* $v \in S) \leq \frac{\min\{\Pr(u \in S), \Pr(v \in S)\}}{m}$
- $\forall u, v \in X$, $\Pr(I_S(u) \neq I_S(v)) \leq \alpha D \cdot \left\|u - v\right\|^2$, *where I_S is the indicator function for the set S.*

Operator $\|.\|$ shows the ℓ^2 norm. Chlamtac et al. [9] proposed an algorithm for finding m-orthogonal separators.

Theorem 5 *[9] There exists a polynomial-time randomized algorithm that given an ℓ_2^2 space X containing 0 and a parameter $m > 0$, and $0 < \beta < 1$, generates an $m-$orthogonal separator with distortion $D = \mathcal{O}_\beta(\sqrt{\log|X| \log m})$ and $\alpha \geq \frac{1}{poly(|X|)}$.*

3.1 SDP Relaxation

In order to prove Theorem 4, we use the following SDP relaxation which is inspired by Bansal et al. [4] except for Constraints 5 and 6. In this relaxation, we assign a vector $\bar{v}$ for each vertex $v \in V$. The objective is to minimize the total weight of cut edges. The set of Constraints 2 are ℓ_2^2 triangle inqualities, and the set of Constraints 3 and 4 are ℓ_2^2 triangle inequalities with the zero vector. The set of Constraints 5 and 6 make sure that for each source-sink pair (s_i, t_i), both s_i and t_i do not belong to S since both vectors $\bar{s_i}$ and $\bar{t_i}$ could not be $\mathbf{1}$ for some fixed unit vector simultaneously. Constraint 7 and the set of Constraints 8 make sure the returned subgraph has the desired size. Suppose now that we have approximately guessed the measure H of the optimal solution $H \le \eta(S) \le 2H$. We can ignore all vertices $v \in V$ with $\eta(v) > 2H$ since they do not participate in the optimal solution and thus write the set of Constraints 8. Constraints (9) are spreading constraints introduced by Bansal et al. [4] which ensure size of S is small.

$\min \sum_{(u,v)\in E} w(u,v) \left\|\bar{u}-\bar{v}\right\|^2$		(1)
$\left\|\bar{u}-\bar{w}\right\|^2 + \left\|\bar{w}-\bar{v}\right\|^2 \ge \left\|\bar{u}-\bar{v}\right\|^2$	$\forall u,v,w \in V$	(2)
$\left\|\bar{u}-\bar{w}\right\|^2 \ge \left\|\bar{u}\right\|^2 - \left\|\bar{w}\right\|^2$	$\forall u,w \in V$	(3)
$\left\|\bar{u}\right\|^2 + \left\|\bar{v}\right\|^2 \ge \left\|\bar{u}-\bar{v}\right\|^2$	$\forall u,v \in V$	(4)
$\left\|\bar{s_i}-\bar{t_i}\right\|^2 \ge \left\|\bar{s_i}\right\|^2$	$\forall (s_i,t_i) \in S_G$	(5)
$\left\|\bar{s_i}-\bar{t_i}\right\|^2 \ge \left\|\bar{t_i}\right\|^2$	$\forall (s_i,t_i) \in S_G$	(6)
$\sum_{v\in V} \left\|\bar{v}\right\|^2 \eta(v) \ge H$		(7)
$\left\|\bar{v}\right\|^2 = 0$	if $\eta(v) > 2H$	(8)
$\sum_{v\in V} \eta(v)\cdot \min\{\left\|\bar{u}-\bar{v}\right\|^2, \left\|\bar{u}\right\|^2\} \ge (1-2H)\left\|\bar{u}\right\|^2$	$\forall u \in V$	(9)

Lemma 1. *Given* $S^* = \arg\min \{\delta(T) : \eta(T) \in [H, 2H], \forall (s_i,t_i) \in S_G, |\{s_i,t_i\} \cap T| \le 1\}$*, the optimal value of SDP is at most* $\delta(S^*)$*.*

Proof. We defer proof to the full version of this paper.

3.2 Approximation Algorithm

In this section, we prove Theorem 4. We propose an approximation algorithm which is inspired by Bansal et al.'s [4] algorithm for small-set expansion (SSE). However, there is a significant difference between our algorithm and theirs. In the SSE problem, one does not need to worry about separating source sink pairs.

First, we solve the SDP relaxation, and then proceed iteratively. In each iteration, we sample an n^3-orthogonal separator S with $\beta = 1/2$ and return it (we repeatedly sample S, until a particular function[4] $f(S)$ has some positive value. Details are deferred to Section 3.3). Then, S is removed from graph G and the SDP solution, by zeroing the weight of edges incident on S (i.e discarding

[4] defined later

these edges), and zeroing the SDP variables corresponding to vertices in S. The algorithm maintains the subsets of vertices removed so far in a set $U \subseteq V$, by initializing $U = \emptyset$, and then at each iteration by updating $U = U \cup \{S\}$. We keep iterating until $\eta(U) = \sum_{S_i \in U} \eta(S_i) \geq H/4$. After the last iteration, if $\eta(U) > H$, we output $F = S$ where S is computed in the last iteration. Otherwise, we put $F = U$. Note that in this case, $U = \{S_1, \cdots, S_{|U|}\}$.

3.3 Analysis

First, let's see what is the effect of algorithm's changes to the SDP solution. By zeroing vectors in S and discarding the edges incident on S, the SDP value may only decrease. Triangle inequalities, and the source-sink constraints still hold. Constraint $\sum_{v \in V} \|\bar{v}\|^2 \eta(v) \geq H$ will be violated due to zeroing some variables. However, since before the last iteration $\eta(U) \leq \frac{H}{4}$, the following constraint still holds:

$$\sum_{v \in V} \|\bar{v}\|^2 \eta(v) \geq \frac{3H}{4} \tag{10}$$

Next, we show the set of spreading constraints (9) will remain satisfied after removing S. Consider the spreading constraint for a fixed vertex u, two cases might happen:

Case 1: If $\exists S \in U$ such that $u \in S$, then u will be removed and $\|\bar{u}\| = 0$, the spreading constraint will be satisfied since RHS is 0.

Case 2: If $\nexists S \in U$ such that $u \in S$, the RHS will not change and we can show that $\min\{\|\bar{u} - \bar{v}\|^2, \|\bar{u}\|^2\}$ does not decrease. If $\nexists S' \in U$ such that $v \in S'$, then the term $\min\{\|\bar{u} - \bar{v}\|^2, \|\bar{u}\|^2\}$ does not change. If $\exists S' \in U$ such that $v \in S'$, then $\min\{\|\bar{u} - \bar{v}\|^2, \|\bar{u}\|^2\} = \|\bar{u}\|^2$ since $\|\bar{v}\| = 0$, and its value does not decrease.

Therefore, in both these cases, the spreading constraints will not be violated.

Lemma 2. *Let S be a sampled n-orthogonal separator. Fix a vertex u. We claim that* $\Pr[\eta(S) \leq 12H \mid u \in S] \geq \frac{7}{8}$.

Proof. We defer proof to Appendix A.1.

Next, we upper bound $\delta(S)$. By the third property of orthogonal separators:

$$\mathbb{E}[\delta(S)] \leq \alpha D \cdot \sum_{(u,v) \in E} \|\bar{u} - \bar{v}\|^2 \cdot w(u, v) \leq \alpha D \cdot SDP$$

Where $D = \mathcal{O}_\beta(\sqrt{\log n \log(n^3)}) = \mathcal{O}(\log n)$. Note that $\beta = 1/2$. Consider the function f:

$$f(S) = \begin{cases} \eta(S) - \delta(S) \cdot \frac{H}{4D \cdot SDP} & \text{if } S \neq \emptyset \text{ and } \eta(S) < 12H \\ 0 & \text{otherwise} \end{cases}$$

We wish to lower bound $\mathbb{E}[f(S)]$. First, we lower bound $\mathbb{E}[\eta(S)]$. As a result of Lemma 2 and the first property of orthogonal separators:

$$\mathbb{E}[\eta(S)] = \sum_{u \in V} \Pr[u \in S \wedge \eta(S) < 12H] \cdot \eta(u)$$

$$= \sum_{u \in V} \Pr[\eta(S) < 12H \mid u \in S] \cdot \Pr[u \in S] \cdot \eta(u) \geq \sum_{u \in V} \frac{7\alpha \|\bar{u}\|^2 \eta(u)}{8}$$

Since $\mathbb{E}[\delta(S)] \leq \alpha D \cdot SDP$ and using Constraint 10:

$$\mathbb{E}[f(S)] \geq \sum_{u \in V} \frac{7\alpha \|\bar{u}\|^2 \eta(u)}{8} - \alpha \cdot D \cdot SDP \cdot \frac{H}{4D \cdot SDP} \geq \frac{7\alpha \frac{3H}{4}}{8} - \frac{\alpha H}{4} = \frac{13}{32}\alpha H$$

We have $f(S) \leq 2nH$ since $\|\bar{u}\| = 0$ whenever $\eta(u) > 2H$. Therefore, $\Pr[f(S) > 0] \geq \frac{\frac{13}{32}\alpha H}{2nH} = \Omega(\frac{\alpha}{n})$. So after $\mathcal{O}(n^2/\alpha)$ samples, with probability exponentially close to 1, the algorithm finds a set S with $f(S) > 0$. $f(S) > 0$ implies $\eta(S) \geq \delta(S) \cdot \frac{H}{4D \cdot SDP}$, therefore $\delta(S) \leq \frac{4D \cdot SDP \cdot \eta(S)}{H}$. Consider the two possible cases for the output F:

Case 1: $F = U = \{S_1, S_2, \cdots, S_{|U|}\}$, and $\eta(F) = \sum_{i=1}^{|U|} \eta(S_i)$. In this case, $\frac{H}{4} \leq \eta(F) \leq H$. The set U is a set of orthogonal separators and each $S_i \in U$ forms a separate part.

Case 2: $F = S$. In this case, let's show the last iteration of step 1 as $U = U_{old} \cup \{S\}$. We know $\eta(U) > H$, and $\eta(U_{old}) < \frac{H}{4}$, therefore $\eta(S) > 3H/4$. Also $f(S) > 0$ implies $\eta(S) \leq 12H$. Therefore, $3H/4 < \eta(S) \leq 12H$.

In both cases, $\frac{H}{4} \leq \eta(F) \leq 12H$.

We showed when a set $S_i \in U$ is sampled, $\delta(S_i) \leq \frac{4D \cdot SDP \cdot \eta(S_i)}{H}$. However, in the LHS of this inequality, edges like (u, v) where $u \in S_j, v \in S_i$ and $j < i$ are not considered. We can show $\sum_{j=1}^{i-1} \delta(S_j, S_i) \leq \sum_{j=1}^{i-1} \frac{4D \cdot SDP \cdot \eta(S_j)}{H} \leq 4D \cdot SDP$ since $\sum_{j=1}^{i-1} \eta(S_j) \leq H$. Therefore, $\delta(S_i) \leq \frac{4D \cdot SDP \cdot \eta(S_i)}{H} + \sum_{j=1}^{i-1} \delta(S_j, S_i) \leq \mathcal{O}(D \cdot SDP)$ since $\eta(S_i) \leq 12H$.

In addition, by the second property of orthogonal separators, for each source-sink pair (s_j, t_j), the probability that both s_j and t_j belong to the orthogonal separator S_i is bounded by $\frac{1}{n^3}$. Therefore, $\Pr[vio(S_i) \geq 1] \leq \frac{T}{n^3} \leq \frac{n^2}{n^3} = \frac{1}{n}$. This completes the proof of Theorem 4.

The following corollary is implied from Theorem 4 and is used in the next section.

Corollary 1. *Given an edge-weighted graph $G = (V, w)$, a set of source sink pairs S_G, a measure η on V such that $\eta(V) = 1$, and a parameter τ, a set $S = \{S_1, \cdots, S_j\}$ could be found satisfying $\forall S_i \in S, S_i \subseteq V, \Pr[vio(S_i) \geq 1] \leq 1/n$, and $\delta(S_i) \leq \mathcal{O}(\log(n)) \cdot OPT$, where $OPT = \arg\min\{\delta(T) : \frac{\eta(T)}{\eta(V)} \geq \tau, vio(T) = 0\}$. In addition, $\eta(S) = \sum_{i=1}^{j} \eta(S_i) \geq \Omega(\tau \cdot \eta(V))$.*

Proof. The algorithm guesses $H \geq \tau$ such that $H \leq \eta(OPT) \leq 2H$. Guessing is feasible since $0 \leq \eta(OPT) \leq n \cdot \eta(u)$, where u is the weight of the heaviest element in OPT, and H can be chosen from the set $\{2^t \eta(u) : u \in V, t = 0, \cdots, \log(n)\}$ of size $\mathcal{O}(n \log(n))$. Theorem 4 is invoked with parameter H. The obtained solution S satisfies the properties of this corollary.

3.4 Covering & Aggregation

Once we find F, we follow the multiplicative update algorithm of [4] with some minor modifications, to get a covering of all the vertices. Then, we use the aggregation step to convert the covering to a partitioning. This step is simpler than [4] since we are not required to maintain any size bound on the subgraphs returned after aggregation.

Theorem 6 *Given graph $G = (V, E)$ and T pairs of source and sink, running Algorithm 1 on this instance outputs a multiset $\mathcal{S}$ that satisfies the following conditions:*

- *for all $S \in \mathcal{S}$: $\delta(S) \leq D \cdot OPT$ where $D = \mathcal{O}(\log(n)), \Pr[vio(S) \geq 1] \leq 1/n$*
- *for all $v \in V$, $\frac{|\{S \in \mathcal{S}: v \in S\}|}{|\mathcal{S}|} \geq \frac{1}{5\gamma kn}$, where $\gamma = \mathcal{O}(1)$ and k is the number of parts in the optimal solution which we guess.*

Proof. We defer proof to Appendix A.2.

Algorithm 1: Covering Procedure for Min-Max Multicut

```
Set t = 1, S = ∅ and y^1(v) = 1 ∀v ∈ V;
Guess k, which is the number of parts in the optimal solution;
while Σ_{v∈V} y^t(v) > 1/n do
    Find set S^t = {S_1, ··· , S_j} using Corollary 1, where τ = 1/k and
      ∀v ∈ V, η(v) = y^t(v)/ Σ_{v∈V} y^t(v);
    𝒮 = S^t ∪ 𝒮;
    // Update the weights of the covered vertices;
    for v ∈ V do
        Set y^{t+1}(v) = 1/2 · y^t(v) if ∃S_i ∈ S^t such that v ∈ S_i, and
          y^{t+1}(v) = y^t(v) otherwise.;
    Set t = t + 1;
return 𝒮;
```

Now the covering of G is converted into a partitioning of G without violating min-max objective by much.

Theorem 7 *Given a weighted graph $G = (V, E)$, a set of source-sink pairs $(s_1, t_1), \cdots, (s_T, t_T)$, and a cover $\mathcal{S}$ consisting of subsets of V such that:*

- *$\forall v \in V$, v is covered by at least a fraction $\frac{c}{nk}$ of sets $S \in \mathcal{S}$, where k is the number of partitions of the optimum solution which we guessed in the covering section, and $c \in (0, 1]$.*
- *$\forall S \in \mathcal{S}$, $\delta(S) \leq B$, $\Pr[vio(S) \geq 1] \leq 1/n$.*

We propose a randomized polynomial time algorithm which outputs a partition $\mathcal{P}$ of V such that $\forall P_i \in \mathcal{P}$, $\delta(P_i) \leq 2B$, and $\Pr[vio(P_i) \geq 1] \leq 1/n$.

Algorithm 2: Aggregation Procedure For Min-Max Multicut

```
Step 1: Sort sets in 𝒮 in a random order: S_1, S_2, ··· , S_{|𝒮|}. Let
  P_i = S_i \ (∪_{j<i} S_j).
Step 2: while There is a set P_i such that δ(P_i) > 2B do
    Set P_i = S_i and for all j ≠ i, set P_j = P_j \ S_i;
```

Proof. We defer proof to Appendix A.3.

Acknowledgements: We are grateful to Nikhil Bansal for useful discussions during a Dagstuhl workshop on scheduling (18101) .

References

1. N. Ailon, N. Avigdor-Elgrabli, E. Liberty, and A. Van Zuylen. Improved approximation algorithms for bipartite correlation clustering. *SIAM Journal on Computing*, 41(5):1110–1121, 2012.
2. N. Ailon, M. Charikar, and A. Newman. Aggregating inconsistent information: ranking and clustering. *Journal of the ACM (JACM)*, 55(5):23, 2008.
3. N. Bansal, A. Blum, and S. Chawla. Correlation clustering. *Machine Learning*, 56(1-3):89–113, 2004.
4. N. Bansal, U. Feige, R. Krauthgamer, K. Makarychev, V. Nagarajan, S. Naor, and R. Schwartz. Min-max graph partitioning and small set expansion. *SIAM Journal on Computing*, 43(2):872–904, 2014.
5. M. Charikar, N. Gupta, and R. Schwartz. Local guarantees in graph cuts and clustering. In *International Conference on Integer Programming and Combinatorial Optimization*, pages 136–147. Springer, 2017.
6. M. Charikar, V. Guruswami, and A. Wirth. Clustering with qualitative information. In *Foundations of Computer Science, 2003. Proceedings. 44th Annual IEEE Symposium on*, pages 524–533. IEEE, 2003.
7. S. Chawla, K. Makarychev, T. Schramm, and G. Yaroslavtsev. Near optimal lp rounding algorithm for correlationclustering on complete and complete k-partite graphs. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 219–228. ACM, 2015.
8. Y. Cheng and G. M. Church. Biclustering of expression data. In *Ismb*, volume 8, pages 93–103, 2000.
9. E. Chlamtac, K. Makarychev, and Y. Makarychev. How to play unique games using embeddings. In *Foundations of Computer Science, 2006. FOCS'06. 47th Annual IEEE Symposium on*, pages 687–696. IEEE, 2006.
10. E. D. Demaine, D. Emanuel, A. Fiat, and N. Immorlica. Correlation clustering in general weighted graphs. *Theoretical Computer Science*, 361(2-3):172–187, 2006.
11. S. Kim, S. Nowozin, P. Kohli, and C. D. Yoo. Higher-order correlation clustering for image segmentation. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 24*, pages 1530–1538. Curran Associates, Inc., 2011.
12. A. Kirillov, E. Levinkov, B. Andres, B. Savchynskyy, and C. Rother. Instancecut: From edges to instances with multicut. In *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*, pages 7322–7331. IEEE, 2017.
13. H.-P. Kriegel, P. Kröger, and A. Zimek. Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 3(1):1, 2009.
14. G. Puleo and O. Milenkovic. Correlation clustering and biclustering with locally bounded errors. In *International Conference on Machine Learning*, pages 869–877, 2016.
15. Z. Svitkina and É. Tardos. Min-max multiway cut. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 207–218. Springer, 2004.
16. P. Symeonidis, A. Nanopoulos, A. Papadopoulos, and Y. Manolopoulos. Nearest-biclusters collaborative filtering with constant values. In *International Workshop on Knowledge Discovery on the Web*, pages 36–55. Springer, 2006.

A Missing Proofs

A.1 Proof of Lemma 2

Proof. Consider a vertex u and let $A_u = \{v : \|\bar{u} - \bar{v}\|^2 \geq \beta \|\bar{u}\|^2\}$ and $B_u = \{v : \|\bar{u} - \bar{v}\|^2 < \beta \|\bar{u}\|^2\}$. Assume for now that $u \in S$. We show with high probability $\eta(A_u \cap S)$ is small, and $\eta(B_u)$ is also small. Vertex u satisfies the spreading constraint. It is easy to see that:

$$(1 - 2H) \|u\|^2 \leq \sum_{v \in V} \eta(v) \cdot \min\{\|\bar{u} - \bar{v}\|^2, \|\bar{u}\|^2\} \leq \beta \|\bar{u}\|^2 \eta(B_u) + \|\bar{u}\|^2 \eta(A_u)$$

Since $\eta(V) = 1$ and $A_u \cup B_u = V$, $\eta(A_u) + \eta(B_u) = 1$, and $\beta = 1/2$ therefore:

$$(1 - 2H) \leq \beta\eta(B_u) + (1 - \eta(B_u)) \tag{11}$$

$$\therefore \eta(B_u) \leq \frac{2H}{1 - \beta} = 4H \tag{12}$$

Consider an arbitrary vertex $v \in A_u$ where $\|\bar{v}\| \neq 0$. By definition of A_u, $\|\bar{u} - \bar{v}\|^2 \geq \beta \|\bar{u}\|^2 \geq \beta \min\{\|\bar{u}\|^2, \|\bar{v}\|^2\}$. Therefore, by the second property of orthogonal separators and since we assumed $u \in S$, then $\Pr[v \in S \mid u \in S] \leq \frac{1}{n^3} \leq H$. The second inequality holds since $H \geq 1/n$.

Now we show a bound for $\mathbb{E}[\eta(A_u \cap S) \mid u \in S]$:

$$\mathbb{E}[\eta(A_u \cap S) \mid u \in S] = \sum_{v \in A_u} \eta(v) \Pr[v \in S \mid u \in S] \leq H$$

Now, we want to bound $\Pr[\eta(S) \geq 12H \mid u \in S]$. The event $\{\eta(S) \geq 12H \mid u \in S\}$ implies the event $\{\eta(A_u \cap S) \geq 8H \mid u \in S\}$ since $\eta(B_u \cap S) \leq \eta(B_u) \leq 4H$. (The second inequality holds by (12)). Now we are ready to complete the proof.

$$\Pr[\eta(S) \geq 12H | u \in S] \leq \Pr[\eta(A_u \cap S) \geq 8H | u \in S] \leq \frac{\mathbb{E}[\eta(A_u \cap S) \mid u \in S]}{8H} \leq \frac{H}{8H} = 1/8$$

We showed $\Pr[\eta(S) \geq 12H | u \in S] \leq 1/8$, therefore $\Pr[\eta(S) \leq 12H | u \in S] \geq 7/8$ and the proof is complete.

A.2 Proof of Theorem 6

Proof. For an iteration t, let $Y^t = \sum_{v \in V} y^t(v)$. Consider the optimal solution $\{S_i^*\}_{i=1}^k$ to the min-max multicut problem. There exists at least a $S_j^* \in \{S_i^*\}_{i=1}^k$ with weight greater than or equal to the average ($y_t(S_j^*) \geq \frac{Y^t}{k}$), $vio(S_j^*) = 0$, and $\delta(S_j^*) \leq OPT$. Therefore by Corollary 1 where $H = \frac{1}{k}$, a set $S_t = \{S_1, S_2, \cdots, S_j\}$ could be found where $\forall S_i \in S_t$, $\delta(S_i) \leq \mathcal{O}(\log n) \cdot OPT$, $\Pr[vio(S_i) \geq 1] \leq 1/n$.

Now we show the second property of the theorem holds. Let ℓ denote the number of iterations in the while loop. Let $|\{S \in \mathcal{S} : v \in S\}| = N_v$. By the updating rules $y^{\ell+1}(v) = 1/2^{N_v}$. Therefore $\frac{1}{2^{N_v}} = y^{\ell+1}(v) \leq 1/n$, which implies $N_v \geq \log(n)$. By Corollary 1, $y^t(S^t) \geq \frac{1}{\gamma k} Y^t$ where $\gamma = \mathcal{O}(1)$. Therefore:

$$Y^{t+1} = Y^t - \frac{1}{2} y^t(S^t) \leq (1 - \frac{1}{2\gamma k}) Y^t$$

Which implies $Y^\ell \leq (1 - \frac{1}{2\gamma k})^{\ell-1} Y^1 = (1 - \frac{1}{2\gamma k})^{\ell-1} n$. Also $Y^\ell > 1/n$ therefore, $\ell \leq 1 + 4\gamma k \ln(n) \leq 5\gamma k \log(n)$. In each iteration t, the number of sets in S_t is at most n (since all the sets in S_t are disjoint), therefore $|\mathcal{S}| \leq 5\gamma k n \log(n)$, and the second property is proved.

A.3 Proof of Theorem 7

Proof. A similar proof to the one given by Bansal et al. [4] shows after step 2, for each $P_i \in \mathcal{P}$, $\delta(P_i) \leq 2B$. We start by analyzing Step 1. Observe that after Step 1, the collection of sets $\{P_i\}$ is a partition of V and $P_i \subseteq S_i$ for every i. Particularly, $vio(P_i) \leq vio(S_i)$. Note, however, that the bound $\delta(P_i) \leq B$ may be violated for some i since P_i might be a strict subset of S_i.
We finish the analysis of Step 1 by proving that $\mathbb{E}[\sum_i \delta(P_i)] \leq 2knB/c$. Fix an $i \leq |\mathcal{S}|$ and estimate the expected weight of edges $E(P_i, \cup_{j>i} P_j)$, given that the i^{th} set in the random ordering is S. If an edge (u,v) belongs to $E(P_i, \cup_{j>i} P_j)$, then $(u,v) \in E(S_i, V \setminus S_i) = E(S, V \setminus S)$ and both $u, v \notin \cup_{j<i} S_j$. For any edge $(u,v) \in \delta(S)$ (with $u \in S, v \notin S$), $\Pr((u,v) \in E(P_i, \cup_{j>i} P_j) \mid S_i = S) \leq \Pr(v \notin \cup_{j<i} S_j \mid S_i = S) \leq (1 - \frac{c}{nk})^{i-1}$, since v is covered by at least $\frac{c}{nk}$ fraction of sets in $\mathcal{S}$ and is not covered by $S_i = S$. Hence,

$$\mathbb{E}[w(E(P_i, \cup_{j>i} P_j)) \mid S_i = S] \leq (1 - \frac{c}{nk})^{i-1} \delta(S) \leq (1 - \frac{c}{nk})^{i-1} B$$

and $\mathbb{E}[w(E(P_i, \cup_{j>i} P_j)) \leq (1 - \frac{c}{nk})^{i-1} B$. Therefore:

$$\mathbb{E}\left[\sum_i \delta(P_i)\right] = 2 \cdot \mathbb{E}\left[\sum_i w(E(P_i, \cup_{j>i} P_j))\right] \leq 2 \sum_{i=0}^{\infty} (1 - \tfrac{c}{nk})^i B = 2knB/c$$

Now we want to analyze step 2. Consider potential function $\sum_i \delta(P_i)$, we showed after step 1, $\mathbb{E}\left[\sum_i \delta(P_i)\right] \leq 2knB/c$. We prove that this potential function reduces quickly over the iterations of Step 2, thus, Step 2 terminates after a small number of steps. After each iteration of Step 2, the following invariant holds: the collection of sets $\{P_i\}$ is a partition of V and $P_i \subseteq S_i$ for all i. Particularly, $vio(P_i) \leq vio(S_i)$. Using an uncrossing argument, we show at every iteration of the while loop in step 2, $\sum_i \delta(P_i)$ decreases by at least $2B$.

$$\delta(S_i) + \sum_{j \neq i} \delta(P_j \setminus S_i) \leq \delta(S_i) + \sum_{j \neq i} \Big(\delta(P_j) + w(E(P_j \setminus S_i, S_i)) - w(E(S_i \setminus P_j, P_j))\Big)$$
$$\leq \delta(S_i) + \sum_{j \neq i} \Big(\delta(P_j)\Big) + w(E(V \setminus S_i, S_i)) - w(E(P_i, V \setminus P_i))$$
$$= \sum_j \Big(\delta(P_j)\Big) + 2\delta(S_i) - 2\delta(P_i) \leq \sum_j \Big(\delta(P_j)\Big) - 2B$$

The above inequalities use the facts that $P_i \subseteq S_i$ for all i and that all the P_j's are disjoint. The second inequality uses the facts that $\sum_{j \neq i} w(E(P_j \setminus S_i, S_i)) = w(E(V \setminus S_i, S_i))$, and $\sum_{j \neq i} w(E(S_i \setminus P_j, P_j)) \geq w(E(P_i, V \setminus P_i))$, which hold since the collection of sets $\{P_i\}$ is a partition of V, and $P_i \subseteq S_i$. In particular, $\sum_{j \neq i} w(E(S_i \setminus P_j, P_j)) \geq w(E(P_i, V \setminus P_i))$ holds since for each edge e if $e \in E(P_i, P_j)$ then $e \in E(S_i \setminus P_j, P_j)$. The last inequality holds since $\delta(S_i) \leq B$ and $\delta(P_i) > 2B$.
This proves that the number of iterations of the while loop is polynomially bounded and after step 2, $\delta(P_i) \leq 2B$ for each P_i.
In addition, since each P_i is a subset of S_i, $vio(P_i) \leq vio(S_i)$. Therefore $\Pr[vio(P_i) \geq 1] \leq 1/n$.