Local Guarantees in Graph Cuts and Clustering

Moses Charikar[1,†], Neha Gupta[1,†] and Roy Schwartz[2,‡]

[1] Stanford University, Stanford CA 94305, USA,
{moses,nehagupta}@cs.stanford.edu,
[2] Technion, Haifa, 3200003, Israel,
schwartz@cs.technion.ac.il

Abstract. Correlation Clustering is an elegant model that captures fundamental graph cut problems such as Min $s-t$ Cut, Multiway Cut, and Multicut, extensively studied in combinatorial optimization. Here, we are given a graph with edges labeled $+$ or $-$ and the goal is to produce a clustering that agrees with the labels as much as possible: $+$ edges within clusters and $-$ edges across clusters. The classical approach towards Correlation Clustering (and other graph cut problems) is to optimize a global objective. We depart from this and study local objectives: minimizing the maximum number of disagreements for edges incident on a single node, and the analogous max min agreements objective. This naturally gives rise to a family of basic min-max graph cut problems. A prototypical representative is Min Max $s-t$ Cut: find an $s-t$ cut minimizing the largest number of cut edges incident on any node. We present the following results: (1) an $O(\sqrt{n})$-approximation for the problem of minimizing the maximum total weight of disagreement edges incident on any node (thus providing the first known approximation for the above family of min-max graph cut problems), (2) a remarkably simple 7-approximation for minimizing local disagreements in complete graphs (improving upon the previous best known approximation of 48), and (3) a $1/(2+\varepsilon)$-approximation for maximizing the minimum total weight of agreement edges incident on any node, hence improving upon the $1/(4+\varepsilon)$-approximation that follows from the study of approximate pure Nash equilibria in cut and party affiliation games.

Keywords: Approximation Algorithms, Graph Cuts, Correlation Clustering, Linear Programming

1 Introduction

Graph cuts are extensively studied in combinatorial optimization, including fundamental problems such as Min $s-t$ Cut, Multiway Cut, and Multicut. Typically, given an undirected graph $G=(V,E)$ equipped with non-negative edge weights $c: E \rightarrow \mathcal{R}_+$ the goal is to find a *constrained* partition $\mathcal{S}=\{S_1,\ldots,S_\ell\}$ of V

† Supported by NSF grants CCF-1617577, CCF-1302518 and a Simons Investigator Award
‡ Supported by ISF grant 1336/16

minimizing the total weight of edges crossing between different clusters of $\mathcal{S}$. e.g., in Min $s-t$ Cut, $\mathcal{S}$ has two clusters, one containing s and the other containing t. Similarly, in Multiway Cut, $\mathcal{S}$ consists of k clusters each containing exactly one of k given special vertices $t_1, \ldots, t_k$. In Multicut, the clusters of $\mathcal{S}$ must separate k given pairs of special vertices $\{s_i, t_i\}_{i=1}^k$.

The elegant model of Correlation Clustering captures all of the above fundamental graph cut problems, and was first introduced by Bansal *et al.* [5] more than a decade ago. In Correlation Clustering, we are given an undirected graph $G = (V, E)$ equipped with non-negative edge weights $c : E \to \mathcal{R}_+$. Additionally, E is partitioned into E^+ and E^-, where edges in E^+ (E^-) are considered to be labeled as $+$ ($-$). The goal is to find a partition of V into an *arbitrary* number of clusters $\mathcal{S} = \{S_1, \ldots, S_\ell\}$ that agrees with the edges' labeling as much as possible: the endpoints of $+$ edges are supposed to be placed in the same cluster and endpoints of $-$ edges in different clusters. Typically, the objective is to find a clustering that minimizes the total weight of misclassified edges. This models, *e.g.*, Min $s-t$ Cut, since one can label all edges in G with $+$, and add (s,t) to E with a label of $-$ and set its weight to $c_{s,t} = \infty$ (Multiway Cut and Multicut are modeled in a similar manner).

Correlation Clustering has been studied extensively for more than a decade [1,2,9,10,13,26]. In addition to the simplicity and elegance of the model, its study is also motivated by a wide range of practical applications: image segmentation [26], clustering gene expression patterns [3,7], cross-lingual link detection [25], and the aggregation of inconsistent information [15], to name a few (refer to the survey [26] and the references therein for additional details).

Departing from the classical global objective approach towards Correlation Clustering, we consider a broader class of objectives that allow us to bound the number of misclassified edges incident on any node (or alternatively edges classified correctly). We refer to this class as Correlation Clustering with *local guarantees*. First introduced by Puleo and Milenkovic [20], Correlation Clustering with local guarantees naturally arises in settings such as community detection without antagonists, *i.e.*, objects that are inconsistent with large parts of their community, and has found applications in diverse areas, *e.g.*, recommender systems, bioinformatics, and social sciences [11,18,20,24].

Local Minimization of Disagreements and Graph Cuts: A prototypical example when considering minimization of disagreements with local guarantees is the Min Max Disagreements problem, whose goal is to find a clustering that minimizes the maximum total weight of misclassified edges incident on any node.

Formally, given a partition $\mathcal{S} = \{S_1, \ldots, S_\ell\}$ of V, for $u \in S_i$, define:

$$\text{disagree}_{\mathcal{S}}(u) \triangleq \sum_{v \notin S_i : (u,v) \in E^+} c_{u,v} + \sum_{v \in S_i : (u,v) \in E^-} c_{u,v} \ .$$

The objective of Min Max Disagreements is: $\min_{\mathcal{S}} \max_{u \in V} \{\text{disagree}_{\mathcal{S}}(u)\}$. This is NP-hard even on complete unweighted graphs and approximations are known for only a few special cases [20]. No approximation is known for general graphs.

Just as minimization of total disagreements in Correlation Clustering models fundamental graph cut problems, Min Max Disagreements gives rise to a variety of basic min-max graph cut problems. A natural problem here is Min Max $s-t$ Cut: Its input is identical to that of Min $s-t$ Cut, however its objective is to find an $s-t$ cut $(S, \overline{S})$ minimizing the total weight of cut edges incident on any node: $\min_{S \subseteq V: s \in S, t \notin S} \max_{u \in V} \{\sum_{v:(u,v) \in \delta(S)} c_{u,v}\}$.[3] Despite the fact that Min Max $s-t$ Cut is a natural graph cut problem, no approximation is known for it. Min Max Disagreements also gives rise to Min Max Multiway Cut and Min Max Multicut, defined similarly; no approximation is known for these. One of our goals is to highlight this family of min-max graph cut problems which we believe deserve further study. Other graph cut problems were studied from the min-max perspective, *e.g.*, [6,22]. However, the goal there is to find a constrained partition that minimizes the total weight of cut edges incident on any *cluster* (as opposed to incident on any *node*).

Min Max Disagreements is a special case of the more general Min Local Disagreements problem. Given a clustering $\mathcal{S}$, consider the vector of all disagreement values $\text{disagree}_{\mathcal{S}}(V) \in \mathcal{R}_+^V$, where $(\text{disagree}_{\mathcal{S}}(V))_u = \text{disagree}_{\mathcal{S}}(u)\ \forall u \in V$. The objective of Min Local Disagreements is to find a partition $\mathcal{S}$ that minimizes $f(\text{disagree}_{\mathcal{S}}(V))$ for a given function f. For example, if f is the max function Min Local Disagreements reduces to Min Max Disagreements, and if f is the summation function Min Local Disagreements reduces to the classic objective of minimizing total disagreements.

Local Maximization of Agreements: Another natural objective of Correlation Clustering is that of maximizing the total weight of edges correctly classified [5,23]. A prototypical example for local guarantees is Max Min Agreements, *i.e.* finding a clustering that maximizes the minimum total weight of correctly classified edges incident on any node. Formally, given a partition $\mathcal{S} = \{S_1, \ldots, S_\ell\}$ of V, for $u \in S_i$, define:

$$\text{agree}_{\mathcal{S}}(u) \triangleq \sum_{v \in S_i:(u,v) \in E^+} c_{u,v} + \sum_{v \notin S_i:(u,v) \in E^-} c_{u,v} \ .$$

The objective of Max Min Agreements is: $\max_{\mathcal{S}} \min_{u \in V} \{\text{agree}_{\mathcal{S}}(u)\}$.

This is a special case of the more general Max Local Agreements problem. Given a clustering $\mathcal{S}$, consider the vector of all agreement values $\text{agree}_{\mathcal{S}}(V) \in \mathcal{R}_+^V$, where $(\text{agree}_{\mathcal{S}}(V))_u = \text{agree}_{\mathcal{S}}(u)\ \forall u \in V$. The objective of Max Local Agreements is to find a partition $\mathcal{S}$ that maximizes $g(\text{agree}_{\mathcal{S}}(V))$ for a given function g ,where, g is required to satisfy the following two conditions: (1) for any $\mathbf{x}, \mathbf{y} \in \mathcal{R}_+^V$ if $\mathbf{x} \leq \mathbf{y}$ then $g(\mathbf{x}) \leq g(\mathbf{y})$ (monotonicity), and (2) $g(\alpha \mathbf{x}) \geq \alpha g(\mathbf{x})$ for any $\alpha \geq 0$ and $\mathbf{x} \in \mathcal{R}_+^V$ (reverse scaling). Note that g is not required to be concave. For example, if g is the min function Max Local Agreements reduces to Max Min Agreements, and if g is the summation function Max Local Agreements reduces to the classic objective of maximizing total agreements.

Max Local Agreements is closely related to the computation of local optima for Max Cut, and the computation of pure Nash equilibria in cut and party

[3] $\delta(S)$ denotes the collection of edges crossing the cut $(S, \overline{S})$.

affiliation games [4,8,12,14,21] (a well studied special class of potential games [19]). In the setting of party affiliation games, each node of G is a player that can choose one of two sides of a cut. The player's payoff is the total weight of edges incident on it that are classified correctly. It is well known that such games admit a pure Nash equilibria via the *best response dynamics* (also known as *Nash dynamics*), and that each such pure Nash equilibrium is a ($1/2$)-approximation for Max Local Agreements. Unfortunately, in general the computation of a pure Nash equilibria in cut and party affiliation games is PLS-complete [17], and thus it is widely believed no polynomial time algorithm exists for solving this task. Nonetheless, one can apply the algorithm of Bhalgat *et al.* [8] for finding an approximate pure Nash equilibrium and obtain a $1/(4+\varepsilon)$-approximation for Max Local Agreements (for any constant $\varepsilon > 0$). This approximation is also the best known for the special case of Max Min Agreements.

Our Results: Focusing first on Min Max Disagreements on general graphs we prove that both the natural LP and SDP relaxations admit a large integrality gap of $n/2$. Nonetheless, we present an $O(\sqrt{n})$-approximation for Min Max Disagreements, bypassing the above integrality gaps.

Theorem 1. *The natural LP and SDP relaxations for Min Max Disagreements have an integrality gap of $n/2$.*

Theorem 2. *Min Max Disagreements admits an $O(\sqrt{n})$-approximation for general weighted graphs.*

Since Min Max $s-t$ Cut, along with Min Max Multiway Cut and Min Max Multicut, are a special case of Min Max Disagreements, Theorem 2 applies to them as well, thus providing the first known approximation for this family of cut problems.[4]

When considering the more general Min Local Disagreements problem, we present a remarkably simple approach that achieves an improved approximation of 7 for both complete graphs and complete bipartite graphs (where disagreements are measured w.r.t one side only). This improves upon and simplifies [20] who presented an approximation of 48 for the former and 10 for the latter.

Theorem 3. *Min Local Disagreements admits a 7-approximation for complete graphs.*
where f is required to satisfy the following three conditions: (1) *for any* $\mathbf{x}, \mathbf{y} \in \mathcal{R}_+^V$ *if* $\mathbf{x} \leq \mathbf{y}$ *then* $f(\mathbf{x}) \leq f(\mathbf{y})$ *(monotonicity),* (2) $f(\alpha\mathbf{x}) \leq \alpha f(\mathbf{x})$ *for any* $\alpha \geq 0$ *and* $\mathbf{x} \in \mathcal{R}_+^V$ *(scaling), and* (3) f *is convex.*

Theorem 4. *Min Local Disagreements admits a 7-approximation for complete bipartite graphs where disagreements are measured w.r.t. one side of the graph.*
where f is required to satisfy the following three conditions: (1) *for any* $\mathbf{x}, \mathbf{y} \in \mathcal{R}_+^V$ *if* $\mathbf{x} \leq \mathbf{y}$ *then* $f(\mathbf{x}) \leq f(\mathbf{y})$ *(monotonicity),* (2) $f(\alpha\mathbf{x}) \leq \alpha f(\mathbf{x})$ *for any* $\alpha \geq 0$ *and* $\mathbf{x} \in \mathcal{R}_+^V$ *(scaling), and* (3) f *is convex.*

[4] Theorem 1 can be easily adapted to apply also for Min Max $s-t$ Cut, Min Max Multiway Cut, and Min Max Multicut, resulting in a gap of $(n-1)/2$.

Focusing on local maximization of agreements, we present a $1/(2+\varepsilon)$ approximation for Max Min Agreements without any assumption on the edge weights. This improves upon the previous known $1/(4+\varepsilon)$-approximation that follows from the computation of approximate pure Nash equilibria in party affiliation games [8]. As before, we show that both the natural LP and SDP relaxations for Max Min Agreements have a large integrality gap of $\frac{n}{2(n-1)}$.

Theorem 5. *For any $\varepsilon > 0$, Max Min Agreements admits a $1/(2+\varepsilon)$-approximation for general weighted graphs, where the running time of the algorithm is poly$(n, 1/\varepsilon)$.*

Theorem 6. *The natural LP and SDP relaxations for Max Min Agreements have an integrality gap of $\frac{n}{2(n-1)}$.*

Table 1: Results for Correlation Clustering with local guarantees.

Problem	Input Graph	Approximation	
		This Work	**Previous Work**
Min Local Disagreements	complete	7	48 [20]
	complete bipartite (one sided)	7	10 [20]
Min Max Disagreements	general weighted	$O(\sqrt{n})$	–
Min Max $s-t$ Cut Min Max Multiway Cut Min Max Multicut	general weighted	$O(\sqrt{n})$	–
Max Min Agreements	general weighted	$1/(2+\varepsilon)$	$1/(4+\varepsilon)$ [8]

Our main algorithmic results are summarized in Table 1.

Approach and Techniques: The non-linear nature of Correlation Clustering with local guarantees makes problems in this family much harder to approximate than Correlation Clustering with classic global objectives.

Firstly, LP and SDP relaxations are not always useful when considering local objectives. For example, the natural LP relaxation for the global objective of minimizing total disagreements on general graphs has a bounded integrality gap of $O(\log n)$ [9,13,16]. However, we prove that for its local objective counterpart, *i.e.*, Min Max Disagreements, both the natural LP and SDP relaxations have a huge integrality gap of $n/2$ (Theorem 1). To overcome this our algorithm for Min Max Disagreements on general weighted graphs uses a *combination* of the LP lower bound and a combinatorial bound. Even though each of these bounds on its own is bad, we prove that their combination suffices to obtain an approximation of $O(\sqrt{n})$, thus bypassing the huge integrality gaps of $n/2$.

Secondly, randomization is inherently difficult to use for local guarantees, while many of the algorithms for minimizing total disagreements, *e.g.*, [1,2,10], as well as maximizing total agreements, *e.g.*, [23], are all randomized in nature.

The reason is that a bound on the expected weight of misclassified edges incident on any node does not translate to a bound on the maximum of this quantity over all nodes (similarly the expected weight of correctly classified edges incident on any node does not translate to a bound on the minimum of this quantity over all nodes). To overcome this difficulty, all the algorithms we present are deterministic, *e.g.*, for Min Local Disagreements we propose a new remarkably simple method of clustering that greedily chooses a center node s^* and cuts a sphere of a fixed and predefined radius around s^*, and for Max Min Agreements we present a new *non-oblivious* local search algorithm that runs on a graph with modified edge weights and circumvents the need to compute approximate pure Nash equilibria in party affiliation games.

Paper Organization: Section 3 contains the improved approximations for Min Max Disagreements on general weighted graphs and for Min Local Disagreements on complete and complete bipartite graphs (Theorems 2, 3, and 4), along with the integrality gaps of the natural LP and SDP relaxations (Theorem 1). Section 4 contains the improved approximation for Max Min Agreements as well as the integrality gaps of the natural LP and SDP relaxations (Theorems 5 and 6).

2 Local Minimization of Disagreements and Graph Cuts

We consider the natural convex programming relaxation for Min Local Disagreements. The relaxation imposes a metric d on the vertices of the graph. For each node $u \in V$ we have a variable $D(u)$ denoting the total fractional *disagreement* of edges incident on u. Additionally, we denote by $\mathbf{D} \in \mathcal{R}_+^V$ the vector of all $D(u)$ variables. Note that the relaxation is solvable in polynomial time since f is convex.[5]

$$
\begin{aligned}
\min \quad & f(\mathbf{D}) && (1)\\
& \sum_{v:(u,v)\in E^+} c_{u,v} d(u,v) + \sum_{v:(u,v)\in E^-} c_{u,v}(1-d(u,v)) = D(u) && \forall u \in V\\
& d(u,v) + d(v,w) \geq d(u,w) && \forall u,v,w \in V\\
& D(u) \geq 0,\ 0 \leq d(u,v) \leq 1 && \forall u,v \in V
\end{aligned}
$$

For the special case of Min Max Disagreements, *i.e.*, f is the max function, (1) can be written as an LP. The proof of Theorem 1, which states that even for the special case of Min Max Disagreements the above natural LP and in addition the natural SDP both have a large integrality gap of $n/2$, appears in the full version of this paper. We note that Theorem 1 also applies to Min Max $s-t$ Cut, a further special case of Min Max Disagreements.

---

[5] The convexity of f is used only to show that relaxation (1) can be solved, and it is not required in the rounding process.

2.1 Min Max Disagreements on General Weighted Graphs

Our algorithm for Min Max Disagreements on general weighted graphs cannot rely solely on the the lower bound of the LP relaxation, since it admits an integrality gap of $n/2$ (Theorem 1). Thus, a different lower bound must be used. Let $c_{\max}$ be the maximum weight of an edge that is misclassified in some optimal solution $\mathcal{S}^*$. Clearly, $c_{\max}$ also serves as a lower bound on the value of an optimal solution. Hence, we can mix these two lower bounds and choose $\max\{\max_{u\in V}\{D(u)\}, c_{\max}\}$ to be the lower bound we use. Note that we can assume w.l.o.g. that $c_{\max}$ is known to the algorithm, as one can simply execute the algorithm for every possible value of $c_{\max}$ and return the best solution.

Our algorithm consists of two main phases. In the first we compute the LP metric d but require additional constraints that ensure no *heavy* edge, *i.e.*, an edge e having $c_e > c_{\max}$, is (fractionally) misclassified by d. In the second phase, we perform a careful *layered clustering* of an auxiliary graph consisting of all $+$ edges whose length in the metric d is short. At the heart of the analysis lies a distinction between $+$ edges whose length in the metric d is short and all other edges. The contribution of the former is bounded using the combinatorial lower bound, *i.e.*, $c_{\max}$, whereas the contribution of the latter is bounded using the LP. Our algorithm also ensures that in the final clustering no heavy edge is misclassified. Let us now elaborate on the two phases, before providing an exact description of the algorithm (Algorithm 1).

Phase 1 (constrained metric computation): Denote by,

$$E^+_{\text{heavy}} \triangleq \{e \in E^+ : c_e > c_{\max}\} \quad \text{and} \quad E^-_{\text{heavy}} \triangleq \{e \in E^- : c_e > c_{\max}\}$$

the collection of all heavy $+$ and $-$ edges, respectively. We solve the LP relaxation (1) (recall that f is the max function) while adding the following additional constraints that ensure d does not (fractionally) misclassify heavy edges:

$$d(u,v) = 0 \qquad \forall e = (u,v) \in E^+_{\text{heavy}} \tag{2}$$

$$d(u,v) = 1 \qquad \forall e = (u,v) \in E^-_{\text{heavy}} \tag{3}$$

If no feasible solution exists then our current guess for $c_{\max}$ is incorrect.

Phase 2 (layered clustering): Denote the collections of $+$ and $-$ edges which are *almost* classified correctly by d as $E^+_{\text{bad}} \triangleq \{e = (u,v) \in E^+ : d(u,v) < 1/\sqrt{n}\}$ and $E^-_{\text{bad}} \triangleq \{e = (u,v) \in E^- : d(u,v) > 1 - 1/\sqrt{n}\}$, respectively. Intuitively, any edge $e \notin E^+_{\text{bad}} \cup E^-_{\text{bad}}$ can use its length d to pay for its contribution to the cost, regardless of what the output is. This is not the case with edges in E^+_{bad} and E^-_{bad}, therefore all such edges are considered *bad*. Additionally, denote by $E^+_0 \triangleq \{e = (u,v) \in E^+ : d(u,v) = 0\}$ the collection of $+$ edges for which d assigns a length of 0.[6]

We design the algorithm so it ensures that no mistakes are made for edges in E^+_0 and E^-_{bad}. However, the algorithm might make mistakes for edges in E^+_{bad}, thus a careful analysis is required. To this end we consider the auxiliary graph

[6] Note that $E^+_{\text{heavy}} \subseteq E^+_0 \subseteq E^+_{\text{bad}}$ and $E^-_{\text{heavy}} \subseteq E^-_{\text{bad}}$.

consisting of all edges in E^+_{bad}, *i.e.*, $G^+_{\text{bad}} \triangleq \left(V, E^+_{\text{bad}}\right)$, and equip it with the distance function dist_ℓ defined as the shortest path metric with respect to the length function $\ell : E^+_{\text{bad}} \to \{0, 1\}$:

$$\ell(e) \triangleq \begin{cases} 0 & e \in E^+_0 \\ 1 & e \in E^+_{\text{bad}} \setminus E^+_0 \end{cases}$$

Assume E^-_{bad} contains k edges and denote the endpoints of the i^{th} edge by s_i and t_i. The algorithm partitions every connected component X of G^+_{bad} into clusters as follows: as long as X contains s_i and t_i for some i, we examine the layers $\text{dist}_\ell(s_i, \cdot)$ defines and perform a carefully chosen level cut. This *layered clustering* suffices as we can prove that our choice of a level cut ensures (1) no mistakes are made for edges in E^+_0 and E^-_{bad}, and (2) the *number* of misclassified edges from $E^+_{\text{bad}} \setminus E^+_0$ incident on any node is at most $O(\sqrt{n})$. This ends the description of the second phase.

Algorithm 1: Layered Clustering $(G = (V, E), c_{\max})$

$\mathcal{C} \leftarrow \emptyset$.
let d be a solution to LP (1) with the additional constraints (2) and (3)
for every connected component X in G^+_{bad} **do**
    while X contains $\{s_i, t_i\}$ for some i **do**
        $r_i \leftarrow \text{dist}_\ell(s_i, t_i)$ and $L^i_j \leftarrow \{u : \text{dist}_\ell(s_i, u) = j\}$ for every $j = 0, 1, \ldots, r_i$.
        choose $j^* \le (\sqrt{n}-1)/2$ s.t. $|L^i_{j^*}|, |L^i_{j^*+1}|, |L^i_{j^*+2}| \le 16\sqrt{n}$.
        $S \leftarrow \cup_{j=0}^{j^*} L^i_j$.
        $X \leftarrow X \setminus S$ and $\mathcal{C} \leftarrow \mathcal{C} \cup \{S\}$.
    end while
    $\mathcal{C} \leftarrow \mathcal{C} \cup \{X\}$.
end for
Output $\mathcal{C}$.

Refer to Algorithm 1 for a precise description of the algorithm. The following Lemma states that the distance between any $\{s_i, t_i\}$ pair with respect to the metric dist_ℓ is large, its proof appears in the full version of this paper.

Lemma 1. *For every $i = 1, \ldots, k$, $\text{dist}_\ell(s_i, t_i) > \sqrt{n} - 1$.*

The following Lemma simply states that only a few layers could be too large, its proof appears in in the full version of this paper. It implies Corollary 1, whose proof appears in the full version of this paper.

Lemma 2. *For every $i = 1, \ldots, k$, the number of layers L^i_j for which $|L^i_j| > 16\sqrt{n}$ is at most $\sqrt{n}/16$.*

Corollary 1. *Algorithm 1 can always find j^* as required.*

Lemma 3 proves that no mistakes are made for edges in E_0^+ and E_{bad}^-, whereas Lemma 4 bounds the *number* of misclassified edges from $E_{\text{bad}}^+ \setminus E_0^+$ incident on any node. Their proofs appear in the full version of this paper.

Lemma 3. *Algorithm 1 never misclassifies edges in E_0^+ and E_{bad}^-.*

Lemma 4. *Let $u \in V$ and S be the cluster in $\mathcal{C}$ Algorithm 1 assigned u to. Then, $\left|\left\{e \in E_{bad}^+ \setminus E_0^+ : e = (u,v), v \notin S\right\}\right| \leq 48\sqrt{n}$.*

We are now ready to prove the main result, Theorem 2.

Proof (of Theorem 2). We prove that Algorithm 1 achieves an approximation of $49\sqrt{n}$. The proof considers edges according to their type: (1) E_0^+ and E_{bad}^- edges, (2) $E_{\text{bad}}^+ \setminus E_0^+$ edges, and (3) all other edges. It is worth noting that the contribution of edges of type (2) is bounded using the combinatorial lower bound, *i.e.*, $c_{\max}$, whereas the contribution of edges of type (3) is bounded using the LP, *i.e.*, $D(u)$ for every node $u \in V$ (as defined by the relaxation (1)).

First, consider edges of type (1). Lemma 3 implies Algorithm 1 does not make any mistakes with respect to these edges, thus their contribution to the value of the output $\mathcal{C}$ is always 0. Second, consider edges of type (2). Lemma 4 implies that every node u has at most $48\sqrt{n}$ edges of type (2) incident on it that are classified incorrectly. Additionally, the weight of every edge of type (2) is at most $c_{\max}$ since $E_{\text{heavy}}^+ \subseteq E_0^+$ and edges of type (2) do not contain any edge of E_0^+. Thus, we can conclude that for every node u the total weight of edges of type (2) that touch u and are misclassified is at most $48\sqrt{n} \cdot c_{\max}$.

Finally, consider edges of type (3). Fix an arbitrary node u and let $D(u)$ be the fractional disagreement value the LP assigned to u (see (1)). Edge e of type (3) is either an edge $e \in E^+$ whose d length is at least $1/\sqrt{n}$, or an edge $e \in E^-$ whose d length is at most $1 - 1/\sqrt{n}$. Hence, in any case the fractional contribution of such an edge e to $D(u)$ is at least $c_e/\sqrt{n}$. Therefore, regardless of what the output is, the total weight of misclassified edges of type (3) incident on u is at most $\sqrt{n} \cdot D(u)$.

Summing over all types of edges, we can conclude that the total weight of misclassified edges incident on u in $\mathcal{C}$ (the output of Algorithm 1) is at most $48\sqrt{n}c_{\max} + \sqrt{n} \cdot D(u)$. Since both $c_{\max}$ and $D(u)$ are lower bounds on the value of an optimal solution, the proof is concluded. □

2.2 Min Local Disagreements on Complete Graphs

We consider a simple deterministic greedy clustering algorithm for complete graphs that iteratively partitions the graph. In every step it does the following: (1) greedily chooses a center node s^* that has many nodes *close* to it, and (2) removes from the graph a sphere around s^* which constitutes a new cluster. The greedy choice of s^* is similar to that of [20]. However, our algorithm departs from the approach of [20], as it *always* cuts a large sphere around s^*. The algorithm of [20], on the other hand, outputs either a singleton cluster containing s^* or some other large sphere around s^* (the average distance within the large sphere determines which of the two options is chosen), thus mimicking the approach of [9]. Surprisingly, restricting the algorithm's choice enables us not only to obtain a simpler algorithm, but also to improve upon the approximation guarantee from 48 to 7.

Algorithm 2 receives as input the metric d as computed by the relaxation (1), whereas the variables $D(u)$ are required only for the analysis. Additionally, we denote by $\text{Ball}_S(u,r) \triangleq \{v \in S : d(u,v) < r\}$ the sphere of radius r around u in subgraph S.

Algorithm 2: Greedy Clustering $(\{d(u,v)\}_{u,v\in V})$

$S \leftarrow V$ and $\mathcal{C} \leftarrow \emptyset$.
while $S \neq \emptyset$ **do**
  $s^* \leftarrow \text{argmax}\,\{|\text{Ball}_S(s, 1/7)| : s \in S\}$.
  $\mathcal{C} \;\leftarrow \mathcal{C} \cup \{\text{Ball}_S(s^*, 3/7)\}$.
  $S \;\leftarrow S \setminus \text{Ball}_S(s^*, 3/7)$.
end while
Output $\mathcal{C}$.

The following lemma summarizes the guarantee achieved by Algorithm 2 (its proof appears in the full version of this paper, which also contains an overview of our charging scheme).

Lemma 5. *Assuming the input is a complete graph, Algorithm 2 guarantees that* $\text{disagree}_{\mathcal{C}}(u) \leq 7D(u)$ *for every* $u \in V$*.*

Proof (of Theorem 3). Apply Algorithm 2 to the solution of the relaxation (1). Lemma 5 guarantees that for every node $u \in V$ we have that $\text{disagree}_{\mathcal{C}}(u) \leq 7D(u)$, *i.e.*, $\text{disagree}_{\mathcal{C}}(V) \leq 7\mathbf{D}$. The value of the output of the algorithm is $f\left(\text{disagree}_{\mathcal{C}}(V)\right)$ and one can bound it as follows:

$$f\left(\text{disagree}_{\mathcal{C}}(V)\right) \overset{(1)}{\leq} f\left(7\mathbf{D}\right) \overset{(2)}{\leq} 7f\left(\mathbf{D}\right) \ .$$

Inequality (1) follows from the monotonicity of f, whereas inequality (2) follows from the scaling property of f. This concludes the proof since $f\left(\mathbf{D}\right)$ is a lower bound on the value of any optimal solution. □

2.3 Min Local Disagreements on Complete Bipartite Graphs

Our algorithm for Min Local Disagreements on complete bipartite graphs (with one sided disagreements) is a natural extension of Algorithm 2. Similarly to the complete graph case, we are able to present a remarkably simple algorithm achieving an improved approximation of 7. The description of the algorithm and the proof of Theorem 4 appear in the full version of this paper.

3 Local Maximization of Agreements

As previously mentioned, Max Local Agreements is closely related to the computation of local optima for Max Cut and pure Nash equilibria in cut and party affiliation games, both of which are PLS-complete problems. We focus on the special case of Max Min Agreements.

The natural local search algorithm for Max Min Agreements can be defined similarly to that of Max Cut: it maintains a single cut $S \subseteq V$; a node u moves to the other side of the cut if the move increases the total weight of correctly classified edges incident on u. This algorithm terminates in a local optimum that is a ($1/2$)-approximation for Max Min Agreements. Unfortunately, it is known that such a local search algorithm can take exponential time, even for Max Cut.

When considering Max Cut, this can be remedied by altering the local search step as follows: a node u moves to the other side of the cut S if the move increases the total weight of edges crossing S by a multiplicative factor of at least $(1+\varepsilon)$ (for some $\varepsilon > 0$). This approach *fails* for the computation of (approximate) pure Nash equilibria in party affiliation games, as well as for Max Min Agreements. The reason is that both of these problems have *local* requirements from nodes, as opposed to the *global* objective of Max Cut. Thus, not surprisingly, the current best known $1/(4+\varepsilon)$-approximation for Max Min Agreements follows from [8] who present the state of the art algorithm for finding approximate pure Nash equilibria in party affiliation games.

We propose a direct approach for approximating Max Min Agreements that circumvents the need to compute approximate pure Nash equilibria in party affiliation games. We improve upon the $1/(4+\varepsilon)$-approximation by considering a *non-oblivious* local search that is executed with altered edge weights. We are able to change the edges' weights in such a way that: (1) any local optimum is a $1/(2+\varepsilon)$-approximation, and (2) the local search performs at most $O(n/\varepsilon)$ iterations. The proof of Theorem 5 appears in the full version of this paper, along with some intuition for our non-oblivious local search algorithm. Additionally, we prove that the natural LP and SDP relaxations for Max Min Agreements on general graphs admit an integrality gap of $\frac{n}{2(n-1)}$ (Theorem 6). This appears in the full version of this paper.

References

1. Ailon, N., Avigdor-Elgrabli, N., Liberty, E., van Zuylen, A.: Improved approximation algorithms for bipartite correlation clustering. SIAM Journal on Computing 41(5), 1110–1121 (2012)
2. Ailon, N., Charikar, M., Newman, A.: Aggregating inconsistent information: ranking and clustering. Journal of the ACM (JACM) 55(5), 23 (2008)
3. Amit, N.: The bicluster graph editing problem. Ph.D. thesis, Tel Aviv University (2004)
4. Balcan, M.F., Blum, A., Mansour, Y.: Improved equilibria via public service advertising. In: SODA' 09. pp. 728–737 (2009)

5. Bansal, N., Blum, A., Chawla, S.: Correlation clustering. Machine Learning 56(1-3), 89–113 (2004)
6. Bansal, N., Feige, U., Krauthgamer, R., Makarychev, K., Nagarajan, V., Naor, J., Schwartz, R.: Min-max graph partitioning and small set expansion. SIAM Journal on Computing 43(2), 872–904 (2014)
7. Ben-Dor, A., Shamir, R., Yakhini, Z.: Clustering gene expression patterns. Journal of computational biology 6(3-4), 281–297 (1999)
8. Bhalgat, A., Chakraborty, T., Khanna, S.: Approximating pure nash equilibrium in cut, party affiliation, and satisfiability games. In: EC' 10. pp. 73–82 (2010)
9. Charikar, M., Guruswami, V., Wirth, A.: Clustering with qualitative information. In: FOCS' 03. pp. 524–533 (2003)
10. Chawla, S., Makarychev, K., Schramm, T., Yaroslavtsev, G.: Near optimal LP rounding algorithm for correlationclustering on complete and complete k-partite graphs. In: STOC' 15. pp. 219–228 (2015)
11. Cheng, Y., Church, G.M.: Biclustering of expression data. In: Ismb. vol. 8, pp. 93–103 (2000)
12. Christodoulou, G., Mirrokni, V.S., Sidiropoulos, A.: Convergence and approximation in potential games. In: STACS' 06. pp. 349–360 (2006)
13. Demaine, E.D., Emanuel, D., Fiat, A., Immorlica, N.: Correlation clustering in general weighted graphs. Theoretical Computer Science 361(2), 172–187 (2006)
14. Fabrikant, A., Papadimitriou, C., Talwar, K.: The complexity of pure nash equilibria. In: Proceedings of the thirty-sixth annual ACM symposium on Theory of computing. pp. 604–612. ACM (2004)
15. Filkov, V., Skiena, S.: Integrating microarray data by consensus clustering. International Journal on Artificial Intelligence Tools 13(04), 863–880 (2004)
16. Garg, N., Vazirani, V.V., Yannakakis, M.: Approximate max-flow min-(multi) cut theorems and their applications. In: STOC' 93. pp. 698–707 (1993)
17. Johnson, D.S., Papadimitriou, C.H., Yannakakis, M.: How easy is local search? Journal of computer and system sciences 37(1), 79–100 (1988)
18. Kriegel, H.P., Kröger, P., Zimek, A.: 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)
19. Monderer, D., Shapley, L.S.: Potential games. Games and economic behavior 14(1), 124–143 (1996)
20. Puleo, G., Milenkovic, O.: Correlation clustering and biclustering with locally bounded errors. In: Proceedings of The 33rd International Conference on Machine Learning. pp. 869–877 (2016)
21. Schäffer, A.A., Yannakakis, M.: Simple local search problems that are hard to solve. SIAM journal on Computing 20(1), 56–87 (1991)
22. Svitkina, Z., Tardos, É.: Min-max multiway cut. In: Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, pp. 207–218. Springer (2004)
23. Swamy, C.: Correlation clustering: maximizing agreements via semidefinite programming. In: SODA' 04. pp. 526–527 (2004)
24. Symeonidis, P., Nanopoulos, A., Papadopoulos, A., Manolopoulos, Y.: Nearest-biclusters collaborative filtering with constant values. In: International Workshop on Knowledge Discovery on the Web. pp. 36–55. Springer (2006)
25. Van Gael, J., Zhu, X.: Correlation clustering for crosslingual link detection. In: IJCAI. pp. 1744–1749 (2007)
26. Wirth, A.: Correlation clustering. C. Sammut and G. Webb, editors, Encyclopedia of Machine Learning pp. 227–231 (2010)