Simultaneously Approximating All ℓ_p-Norms in Correlation Clustering

Sami Davies
Department of EECS and Simons Institute, University of California at Berkeley, CA, USA

Benjamin Moseley
Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA, USA

Heather Newman
Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA, USA

Abstract

This paper considers correlation clustering on unweighted complete graphs. We give a combinatorial algorithm that returns a *single* clustering solution that is *simultaneously* $O(1)$-approximate for all ℓ_p-norms of the disagreement vector; in other words, a combinatorial $O(1)$-approximation of the *all-norms* objective for correlation clustering. This is the first proof that minimal sacrifice is needed in order to optimize different norms of the disagreement vector. In addition, our algorithm is the first combinatorial approximation algorithm for the ℓ_2-norm objective, and more generally the first combinatorial algorithm for the ℓ_p-norm objective when $1 < p < \infty$. It is also faster than all previous algorithms that minimize the ℓ_p-norm of the disagreement vector, with run-time $O(n^\omega)$, where $O(n^\omega)$ is the time for matrix multiplication on $n \times n$ matrices. When the maximum positive degree in the graph is at most Δ, this can be improved to a run-time of $O(n\Delta^2 \log n)$.

2012 ACM Subject Classification Theory of computation → Approximation algorithms analysis

Keywords and phrases Approximation algorithms, correlation clustering, all-norms, lp-norms

Digital Object Identifier 10.4230/LIPIcs.ICALP.2024.52

Category Track A: Algorithms, Complexity and Games

Related Version *Full Version*: https://arxiv.org/abs/2308.01534 [20]

Funding *Sami Davies*: Supported by an NSF Computing Innovation Fellowship while at Northwestern University.
Benjamin Moseley: Supported in part by a Google Research Award, Inform Research Award, Carnegie Bosch Junior Faculty Chair, NSF grants CCF-2121744 and CCF-1845146, and ONR Award N000142212702.
Heather Newman: Supported in part by a Google Research Award, Inform Research Award, Carnegie Bosch Junior Faculty Chair, NSF grants CCF-2121744 and CCF-1845146, and ONR Award N000142212702.

1 Introduction

Correlation clustering is one of the most prominent problems in clustering, as it cleanly models community detection problems [38, 36] and provides a way to decompose complex network structures [39, 32]. The input to the unweighted correlation clustering problem is a complete graph $G = (V, E)$, where $|V| = n$ and each edge $e \in E$ is labeled positive (+) or negative (−). If the edge (u, v) is positive, this indicates that u and v are similar, and analogously if the edge (u, v) is negative, this indicates that u and v are dissimilar. The output of the problem is a partition of the vertex set into parts $C_1, C_2, \ldots$, where each part represents a cluster.

The output should cluster similar vertices together and separate dissimilar vertices. Specifically, for a fixed clustering (i.e., partition of the vertices), a positive edge (u, v) is a *disagreement* with respect to the clustering if u and v are in different clusters and an *agreement* if u and v are in the same cluster. Similarly, a negative edge (u, v) is a disagreement with respect to the clustering if u and v are in the same cluster and an agreement if u and v are in different clusters. The goal is to find a clustering that minimizes some objective that is a function of the disagreements.[1] For example, the most commonly studied objective minimizes the total number of disagreements.

As an easy example to illustrate the problem, consider a social network. Every pair of people has an edge between them, and the edge is positive if the two people have ever met before, and negative otherwise. The goal of correlation clustering translates to partitioning all the people into clusters so that people are in the same cluster as their friends/acquaintances and in different clusters than strangers. The difficulty in constructing a clustering is that the labels may not be consistent, making disagreements unavoidable. Consider in the social network what happens when there is one person with two friends who have never met each other (u, v, w with (u, v) and (u, w) positive but (v, w) negative). The choice of objective matters in determining the best clustering.

For a given clustering $\mathcal{C}$, let $y_{\mathcal{C}}(u)$ denote the number of edges incident to u that are disagreements with respect to $\mathcal{C}$ (we drop $\mathcal{C}$ and write y when it is clear from context). The most commonly considered objectives are $\|y_{\mathcal{C}}\|_p = \sqrt[p]{\sum_{u \in V} y_{\mathcal{C}}(u)^p}$ for $p \in \mathbb{R}_{\geq 1} \cup \{\infty\}$, the ℓ_p-norms of the *disagreement vector* y. Note that the optimal objective values may drastically vary for different norms too. (For instance, in the example in Appendix A of [35], $V = A \sqcup B$[2], where $|A| = |B| = n/2$, and all edges are positive except for a negative matching between A and B. The optimal ℓ_∞-norm objective value is 1 whereas the optimal for ℓ_1 is $\Theta(n)$.) When $p = 1$, this objective minimizes the total number of disagreements. Setting $p = \infty$ minimizes the maximum number of disagreements incident to any node, ensuring a type of worst-case fairness.[3] Balancing these two extremes – average welfare on one hand and fairness on the other – is the ℓ_2-norm, which minimizes the variance of the disagreements at each node.

Correlation clustering was proposed by Bansal, Blum, and Chawla [7] with the objective of minimizing the ℓ_1-norm of the disagreement vector. The problem is NP-hard and several approximation algorithms have been proposed [7, 4, 16, 18]. Puleo and Milenkovic [35] proposed studying ℓ_p-norms of the disagreement vector for $p > 1$, and they give a 48-approximation for any fixed p. Charikar, Gupta, and Schwartz [15] introduced an improved 7-approximation, which Kalhan, Makarychev, and Zhou [29] further improved to a 5-approximation. When $p > 1$, up until recently, the only strategies were LP or SDP rounding, and it has been of interest to develop fast combinatorial algorithms [37]. Davies, Moseley, and Newman [19] introduced a combinatorial $O(1)$-approximation algorithm for $p = \infty$ (see also [25] for a different combinatorial algorithm), and leave open the question of discovering a combinatorial $O(1)$-approximation algorithm for $1 < p < \infty$.

In all prior work, solutions obtained for ℓ_p-norms are tailored to each norm (i.e., p is part of the input to the algorithm), and it was not well-understood what the trade-offs were between solutions that optimize different norms. Solutions naively optimizing one norm can be arbitrarily bad for other norms (see Figure 1). A natural question is whether this loss from using a solution to one objective for another is avoidable. More specifically:

[1] Note that the sizes and number of clusters are unspecified.

[2] $\sqcup$ denotes disjoint union.

[3] In the social network example, minimizing the ℓ_1-norm corresponds to finding a clustering that minimizes the total number of friends who are separated plus the total number of strangers who are in the same cluster. The ℓ_∞-norm corresponds to finding a clustering minimizing the number of friends any person is separated from plus the number of strangers in that person's same cluster.

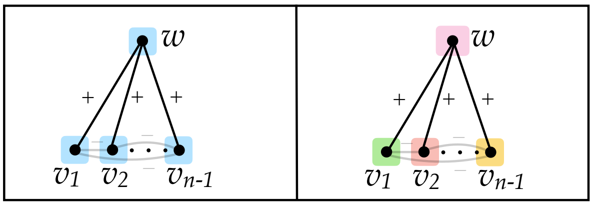

Figure 1 Two clusterings of the star graph, which has one node (w) with positive edges to all nodes, and the rest of the edges negative. **Left:** Clustering assigns all nodes to one (blue) cluster, and is (almost) optimal for the ℓ_∞-norm with cost $\Theta(n)$. **Right:** Clustering assigns all nodes to different clusters and is (almost) optimal for the ℓ_1-norm with cost $\Theta(n)$. The left solution is terrible for the ℓ_1-norm, as the negative clique has $\Theta(n^2)$ edges that are disagreements.

For any graph input to unweighted, complete correlation clustering, does there exist a partition (clustering) that is ***simultaneously*** *$O(1)$-approximate for all ℓ_p-norm objectives?*

Phrased another way, does there exist a *universal algorithm* for ℓ_p-norm correlation clustering – one which is guaranteed to produce a solution that well-approximates many objectives at once? When the goal is to simultaneously optimize every ℓ_p-norm, this is known as the *all-norms* objective.[4] Universal algorithms and the all-norms objective are well-studied in combinatorial optimization problems, such as load balancing and set cover (see Section 1.2 for more discussion). In the context of correlation clustering, such an algorithm outputs a partition that has good global performance (i.e. ℓ_1-norm) and also has no individual node with too many adjacent disagreements (i.e. ℓ_∞-norm). Universal algorithms exist for some problems and are provably impossible for others. The question looms, what can be said about universal algorithms for correlation clustering?

As far as we are aware, there are no known results for the all-norms objective in other clustering problems. In fact, for the popular k-median and k-center problems, it is actually *impossible* to $O(1)$-approximate (or even $o(\sqrt{n})$-approximate) these two objectives simultaneously [5].

1.1 Results

This paper is focused on optimizing all ℓ_p-norms ($p \geq 1$) for correlation clustering at the same time. The main result of the paper answers the previous question positively: perhaps surprisingly, there is a single clustering that simultaneously $O(1)$-approximates the optimal for all ℓ_p-norms. Further, it can be found through an *efficient combinatorial algorithm.* This is also the first known combinatorial approximation algorithm for the ℓ_2-norm objective and more generally ℓ_p-norm objective for fixed $2 \leq p < \infty$.

In what follows, let $O(n^\omega)$ denote the run-time of $n \times n$ matrix multiplication.

▶ **Theorem 1.** *Let $G = (V, E)$ be an instance of unweighted, complete correlation clustering on $|V| = n$ nodes. There exists a combinatorial algorithm returning a single clustering that is simultaneously an $O(1)$-approximation[5] for all ℓ_p-norm objectives, for all $p \in \mathbb{R}_{\geq 1} \cup \{\infty\}$, and its run-time is $O(n^\omega)$.*

[4] In some of the literature, for instance that of Golovin et al. [24], it is called the *all-ℓ_p-norms* objective.
[5] Note this is independent of p.

The algorithm gives the *fastest run-time* of any $O(1)$-approximation algorithm for the ℓ_p-norm objective when $p \in \mathbb{R}_{>1}$. Further, the run-time can be improved when the positive degree of the graph is bounded, as shown in the following corollary.

▶ **Corollary 2.** *Let Δ denote the maximum positive degree in an instance $G = (V, E)$ of unweighted, complete correlation clustering on $|V| = n$ nodes. Suppose G is given as an adjacency list representation of its positive edges. There exists a combinatorial algorithm returning a single clustering that is simultaneously an $O(1)$-approximation for all ℓ_p-norm objectives, for all $p \in \mathbb{R}_{\geq 1} \cup \{\infty\}$, and its run-time is $O(n\Delta^2 \log n)$.*

The run-time of the algorithm matches the fastest known algorithm for the ℓ_∞-norm objective [19], in both the general case and when the maximum positive degree is bounded. The best-known algorithm before our work relied on solving a convex relaxation on $|V|^2$ variables and $|V|^3$ constraints. We improve the run-time by avoiding this bottleneck.

In the setting when the positive edges form a regular graph, the interested reader may also find a clean proof (which is much simpler than that of Theorem 1) in Section 3 showing there is a solution that is simultaneously $O(1)$-approximate for the ℓ_1-norm and ℓ_∞-norm objectives.

1.2 Related work

Correlation clustering was introduced by Bansal, Blum, and Chawla [7]. The version they introduced also studies the problem on unweighted, complete graphs, but is concerned with minimizing the ℓ_1-norm of the disagreement vector. For this problem, Ailon, Charikar, and Newman [4] designed the Pivot algorithm, which is a randomized algorithm that in expectation obtains a 3-approximation. While we know algorithms with better approximations for ℓ_1 correlation clustering than Pivot [16, 18], the algorithm remains a baseline in correlation clustering due to its simplicity. (However, Pivot can perform arbitrarily badly – i.e., give $\Omega(n)$ approximation ratios – for other ℓ_p-norms; see again the example in Appendix A of [35].) It is an active area of research to develop algorithms for the ℓ_1-norm that focus on practical scalability [11, 17, 33, 36, 14]. Correlation clustering has also been studied on non-complete, weighted graphs [15, 29], with conditions on the cluster sizes [34], and with asymmetric errors [26]. In fact, in recent work Veldt [37] highlighted the need for deterministic techniques in correlation clustering that do not use linear programming. Much interest in correlation clustering stems from its connections to applications, including community detection, natural language processing, location area planning, and gene expression [38, 36, 39, 32, 9, 21].

Puleo and Milenkovic [35] introduced correlation clustering with the goal of minimizing the ℓ_p-norm of the disagreement vector. They show that even for minimizing the ℓ_∞-norm on complete, unweighted graphs, the problem is NP-hard (Appendix C in [35]). Several groups found $O(1)$-approximation algorithms for minimizing the ℓ_p-norm on complete, unweighted graphs [35, 15, 29], the best of which is currently the 5-approximation of Kalhan, Makarychev, and Zhou [29]. Many other interesting objectives for correlation clustering focus on finding solutions that are (in some sense) fair or locally desirable [3, 8, 1, 22, 27, 2]. All of these previous works that study general ℓ_p-norms or other notions of fairness or locality rely on solving a convex relaxation. This has two downsides: (1) the run-time of the algorithms are bottle-necked by the time it takes to solve the relaxation with at least $\Omega(n^2)$ many variables and $\Omega(n^3)$ constraints; in fact, it is time-consuming to even enumerate the $\Omega(n^2)$ variables and $\Omega(n^3)$ constraints; and (2) the solution is only guaranteed to be good for one particular value of p.

Several problems have been studied with the goal of finding a solution that is a good approximation for several objectives simultaneously. The all-norms objective was introduced by Azar et al. [6], where the goal is to design a ρ-approximation algorithm for all ℓ_p-norm objectives of a problem. They originally introduced the objective for the restricted assignment load balancing problem and showed an all-norms 2-approximation. Further follow-up on the all-norms objective has been done for load balancing [30, 10, 31], and for set cover [24]. The term "universal" algorithm has also been used for Steiner tree [12, 13], TSP [28], and clustering [23], though in these settings the goal is different, namely, to find a solution that is good for any potential *input*; e.g., in Universal Steiner Tree, the goal is to find a spanning tree where for any set of terminals, the sub-tree connecting the root to the terminals is a good approximation of the optimal.

2 Preliminaries

We will introduce notation, and then we will discuss two relevant works – the papers by Kalhan, Makarychev, and Zhou [29] and Davies, Moseley, and Newman [19].

2.1 Notation

Recall our input to the correlation clustering problem is $G = (V, E)$, an unweighted, complete graph on n vertices, and every edge is assigned a label of either positive (+) or negative (−). Let the set of positive edges be denoted E^+ and the set of negative edges E^-. Then, we can define the *positive neighborhood* and *negative neighborhood* of a vertex u as $N_u^+ = \{v \in V \mid (u,v) \in E^+\}$ and $N_u^- = \{v \in V \mid (u,v) \in E^-\}$, respectively. We further assume without loss of generality that every vertex has a positive self-loop to itself.

A *clustering* $\mathcal{C}$ is a partition of V into *clusters* $C_1, \ldots, C_k$ (but recall that k is *not* pre-specified). Let $C(u)$ denote the cluster that vertex u is in, i.e., if $\mathcal{C}$ has k clusters, there exists exactly one $i \in [k]$ such that $C(u) = C_i$. It is also helpful to consider the vertices in a different cluster than u, and so we let $\overline{C(u)} = V \setminus C(u)$ denote this. We say that a positive edge $e = (u,v) \in E^+$ is a *disagreement* with respect to $\mathcal{C}$ if $v \in \overline{C(u)}$. On the other hand, we say that a negative edge $e = (u,v) \in E^-$ is a disagreement with respect to $\mathcal{C}$ if $v \in C(u)$. For a fixed clustering $\mathcal{C}$, we denote the *disagreement vector* of $\mathcal{C}$ as $y_{\mathcal{C}} \in \mathbb{Z}^n_{\geq 0}$, where for $u \in V$, $y_{\mathcal{C}}(u)$ is the number of edges incident to u that are disagreements with respect to $\mathcal{C}$. We omit the subscript throughout the proofs when a clustering is clear.

Throughout, we let OPT be the optimal objective value, and the ℓ_p-norm to which it corresponds will be clear from context. The next fact follows from the definitions seen so far (recalling also the positive self-loops).

▶ **Fact 3.** *For any $u, v \in V$, $n = |N_u^+ \cap N_v^+| + |N_u^- \cap N_v^-| + |N_u^+ \cap N_v^-| + |N_u^- \cap N_v^+|$.*

2.2 Summary of work by Kalhan, Makarychev, and Zhou

The standard linear program relaxation for correlation clustering is given in (P) below.[6] In the *integer* LP, the variable x_{uv} indicates whether vertices u and v will be in the same cluster (0 for yes, 1 for no), and the disagreement vector is y; the optimal solution to the integer LP has value OPT, while the optimal solution to the relaxation gives a lower bound on OPT. Note the triangle inequality is enforced on all triples of vertices, inducing a semi-metric space

[6] Technically this is a convex program as the objective is convex; we say LP as the constraints are linear.

on V. Throughout this paper, as in [19], we refer to the algorithm by Kalhan, Makarychev, and Zhou as the KMZ algorithm. The *KMZ algorithm* has two phases: it solves (P), and then uses the *KMZ rounding algorithm* to obtain an integral assignment of vertices to clusters. At a high-level, the KMZ rounding algorithm is an iterative, ball-growing algorithm that uses the semi-metric to guide its choices. Their algorithm is a 5-approximation, and produces different clusterings for different p, since the optimal solution x^* to (P) depends on p.

$$\begin{aligned} & \min \|y\|_p && \\ \text{s.t. } & y_u = \sum_{v \in N_u^+} x_{uv} + \sum_{v \in N_u^-} (1 - x_{uv}) && \forall u \in V \\ & x_{uv} \le x_{uw} + x_{vw} && \forall u, v, w \in V \\ & 0 \le x_{uv} \le 1 && \forall u, v \in V. \end{aligned} \tag{P}$$

▶ **Definition 4.** *Let f be a semi-metric on V, i.e., taking $x = f$ gives a feasible solution to (P). The* fractional cost of f in the ℓ_p-norm objective *is the value of (P) that results from setting $x = f$. When p is clear from context, we will simply call this the fractional cost of f.*

2.3 Summary of work by Davies, Moseley, and Newman

The main take-away from the work of Kalhan, Makarychev, and Zhou [29] is that one only requires a semi-metric on the set of vertices, whose cost is comparable to the cost of an optimal solution, as input to the KMZ rounding algorithm. Thus, the insight of Davies, Moseley, and Newman [19] for the ℓ_∞-norm objective is that one can combinatorially construct such a semi-metric without solving an LP, and at small loss in the quality of the fractional solution. They do this by introducing the *correlation metric.*

▶ **Definition 5** ([19]). *For all $u, v \in V$, the correlation metric defines the distance between u and v as*

$$d_{uv} = 1 - \frac{|N_u^+ \cap N_v^+|}{|N_u^+ \cup N_v^+|} = \frac{|N_u^+ \cap N_v^-| + |N_u^- \cap N_v^+|}{|N_u^+ \cap N_v^+| + |N_u^+ \cap N_v^-| + |N_u^- \cap N_v^+|}.$$

Note that the rewrite in the second equality is apparent from Fact 3.

The correlation metric captures useful information succinctly. Intuitively, if u and v have relatively large positive intersection, i.e., $N_u^+ \cap N_v^+$ is large compared to their other relevant joint neighborhoods $(N_u^+ \cap N_v^-) \cup (N_u^- \cap N_v^+)$, then from the perspective of u and v, fewer disagreements are incurred by putting u and v in the same cluster than by putting them in different clusters. This is because if u and v are in the same cluster, then they have disagreements on edges (u, w) and (v, w) for $w \in (N_u^+ \cap N_v^-) \cup (N_u^- \cap N_v^+)$, but if they are in different clusters, then u and v have disagreements on edges (u, w) and (v, w) for $w \in N_u^+ \cap N_v^+$. For more on intuition behind the correlation metric, see Section 2 in [19].

Davies, Moseley, and Newman [19] prove that the correlation metric d can be used as input to the KMZ rounding algorithm by showing that (1) d satisfies the triangle inequality and (2) the fractional cost of d in the ℓ_∞-norm (recall Definition 4) is no more than 8 times the value of the optimal integral solution (OPT). Since the KMZ rounding algorithm loses a factor of at most 5, inputting d to that algorithm returns a 40-approximation algorithm. A benefit of the correlation metric is that it can be computed in time $O(n^\omega)$, and even faster when the subgraph on positive edges is sparse.

2.4 Technical overview

It is not hard to see that the correlation metric cannot be used as input to the KMZ algorithm for ℓ_p-norms other than $p = \infty$, as one cannot bound the fractional cost of the correlation metric against the optimal with only an $O(1)$-factor loss. To see why, consider the star again, as in Figure 1. Here, for all $u, v \in \{v_1, \ldots, v_{n-1}\}$, $d_{uv} = 1 - 1/(n - (n-3)) = 2/3$, but for the ℓ_1-norm, we need the semi-metric to have the value $1 - d_{uv}$ be close to 0, i.e. $O(1/n)$, for such u, v, in order for the fractional cost to be comparable to the value of OPT for $p = 1$.

There are several possible fixes one could try to make to the correlation metric. One idea is that since one can interpret the correlation metric as a coarse approximation of the probability the Pivot algorithm[7] separates u and v, one could try to adapt the correlation metric to more accurately approximate this probability.[8] Another idea, inspired by an observation below, is that one could define a semi-metric for edges in E^+ and another semi-metric for edges in E^-, but then there is the difficulty of showing the triangle inequality holds when positive and negative edges are mixed. Both of these ideas were, for us, unsuccessful.

Instead, the following two observations of how the correlation metric works with respect to the ℓ_1-norm led us to an effective adaptation:

1. One can bound the fractional cost, *restricted to positive edges*, of the correlation metric in the ℓ_1-norm by an $O(1)$-factor times the optimal solution's cost (see Claim 1 in Appendix C of the full version [20]). Negative edges still pose a challenge.
2. If the subgraph of positive edges is regular, then we can actually bound the fractional cost of the correlation metric in the ℓ_1-norm on negative edges as well.[9] See Section 3.

These observations led us to ask whether some adjustments to the correlation metric might yield a semi-metric with bounded fractional cost in the ℓ_1-norm or even the ℓ_p-norm more generally (while still remaining bounded in the ℓ_∞-norm). Moreover, since the KMZ rounding algorithm does not depend on p (whereas in the KMZ algorithm, the solution to the LP *does* depend on p), inputting the same semi-metric to the rounding algorithm produces the same clustering for all ℓ_p-norms!

We are ready to define the *adjusted correlation metric.* Let Δ_u denote the positive degree of u (the degree of u in the subgraph of positive edges).

▶ **Definition 6.** *Define the* adjusted correlation metric $f : E \to [0,1]$ *as follows:*

1. *For d the correlation metric, i.e.,* $d_{uv} = 1 - \frac{|N_u^+ \cap N_v^+|}{|N_u^+ \cup N_v^+|}$, *initially set* $f = d$.
2. *If $e \in E^-$ and $d_e > 0.7$, set $f_e = 1$ (round up).*
3. *For $u \in V$ such that $|N_u^- \cap \{v : d_{uv} \le 0.7\}| \ge \frac{10}{3}\Delta_u$, set $f_{uv} = 1$ for all $v \in V \setminus \{u\}$.*

The idea in Step 3 is that if the fractional cost of negative edges incident to u is sufficiently large, we instead trade this for the cost of positive disagreements, as the rounding algorithm will now put u in its own cluster. For the ℓ_∞-norm, this trade-off is innocuous. For ℓ_p-norms in general, a refined charging argument is needed to show that post-processing d in this way sufficiently curbs the (too large) fractional cost of d.

In Section 3, we start with a warm-up exercise and show that if the graph on positive edges is regular, then the (original) correlation metric d has $O(1)$-approximate fractional cost. Note this section is not necessary to understanding the rest of the paper, but we

---

[7] Pivot operates as follows: Choose a random unclustered $u \in V$. Take u and all its unclustered positive neighbors and let this be the newest cluster. Continue until all vertices in V are clustered.

[8] If in (P), x_{uv} is set exactly to the probability that u and v are separated by Pivot, then x will be a feasible solution with cost at most $3\mathsf{OPT}$. However, this probability seems difficult to express in closed form, even approximately.

[9] In contrast, one cannot bound the fractional cost of the correlation metric on the star (Figure 1).

include it in the main body because we find the proof here is clean and lends insight into the challenges for the irregular case. The main technical result of the paper is Section 4, where we prove Theorem 1 by showing that the adjusted correlation metric can be input to the KMZ rounding algorithm. Namely, we will first show (quite easily) that the adjusted correlation metric satisfies an approximate triangle inequality. Then, it remains to upper bound the fractional cost of the adjusted correlation metric against OPT. We tackle this with a combinatorial charging argument. This argument leverages a somewhat different approach from that used in [19] and is simpler than their proof for only the ℓ_∞-norm. The *constant* approximation factor obtained from inputting the adjusted correlation metric to the KMZ rounding algorithm is bounded above (and below) by universal constants for all p (this is the worst case and one can get better constants for each p).

3 A Special Case: Regular Graphs

In general, the original correlation metric d (Definition 5) does not necessarily have bounded fractional cost for the ℓ_1-norm objective (or more generally for ℓ_p-norm objectives). So, we use the adjusted correlation metric f (Definition 6) as input to the KMZ rounding algorithm. In this section, we show that *if the subgraph of positive edges is regular*, then the correlation metric d can be used as is (i.e., without the adjustments in Steps 2 and 3 of Definition 6) to yield a clustering that is constant approximate for the ℓ_1-norm and ℓ_∞-norm simultaneously:

▶ **Theorem 7.** *Let $G = (V, E)$ be an instance of unweighted, complete correlation clustering, and let E^+ denote the set of positive edges. Suppose that the subgraph induced by E^+ is regular. The fractional cost of d in the ℓ_1-norm objective is within a constant factor of OPT:*

$$\sum_{u\in V}\sum_{v\in N_u^+} d_{uv} + \sum_{u\in V}\sum_{v\in N_u^-}(1-d_{uv}) = O(\mathit{OPT}).$$

Therefore, the clustering produced by inputting d to the KMZ rounding algorithm is a constant-factor approximation simultaneously for the ℓ_1-norm and ℓ_∞-norm objectives.

Proof. Let Δ be the (common) degree of the positive subgraph. To show that the fractional cost of d in the ℓ_1-norm objective is $O(\mathsf{OPT})$ for regular graphs, we will use a dual fitting argument. The LP relaxation we consider is from [4], which uses a dual fitting argument to show constant approximation guarantees for Pivot (although the proof here does not otherwise resemble the proof for Pivot). The primal is given by

$$\min\Big\{\sum_{e\in E} x_e \mid x_{ij}+x_{jk}+x_{ki}\ge 1, \forall ijk\in\mathcal{T}, x\ge 0\Big\} \tag{P'}$$

where $\mathcal{T}$ is the set of bad triangles (i.e. triangles with exactly two positive edges and one negative edge). For $x \in \{0,1\}^{|E|}$, x corresponds to disagreements in a clustering: we set $x_e = 1$ if e is a disagreement and $x_e = 0$ otherwise. The constraints state that every clustering must make a disagreement on every bad triangle. Thus, (P') is a relaxation for the ℓ_1-norm objective. In fact, we will prove the stronger statement that the fractional cost is $O(\mathsf{OPT}_{P'})$, where $\mathsf{OPT}_{P'}$ is the optimal objective value of (P').

The dual is given by

$$\max\Big\{\sum_{T\in\mathcal{T}} y_T \mid \sum_{T\in\mathcal{T}:T\ni e} y_T\le 1, \forall e\in E, y\ge 0\Big\}. \tag{D'}$$

We show that by setting $y_T = \frac{1}{2\Delta}$ for all $T \in \mathcal{T}$, y satisfies the following properties:

1. y is feasible in (D').
2. The fractional cost of d is at most $6 \cdot \sum_{T \in \mathcal{T}} y_T$.

Letting $\mathsf{OPT}_{D'}$ be the optimal objective value of (D'), we have $6 \cdot \sum_{T\in\mathcal{T}} y_T \leq 6 \cdot \mathsf{OPT}_{D'} = 6 \cdot \mathsf{OPT}_{P'} \leq 6 \cdot \mathsf{OPT}$, which will conclude the proof.

To prove feasibility, we case on whether e is positive or negative.

If $e \in E^-$, then $|\{T \in \mathcal{T} : T \ni e\}| = |N_u^+ \cap N_v^+| \leq \Delta$, where equality is by the definition of a bad triangle. So $\sum_{T\in\mathcal{T}:T\ni e} y_T \leq \frac{\Delta}{2\Delta} \leq 1$.

If $e \in E^+$, then $|\{T : T \ni e\}| = |(N_u^+ \cap N_v^-) \cup (N_u^- \cap N_v^+)| \leq 2\Delta$. We conclude y is feasible, since $\sum_{T\in\mathcal{T}:T\ni e} y_T \leq \frac{2\Delta}{2\Delta} = 1$.

Now we need to show that the fractional cost of d is bounded in terms of the objective value of (D'). First we bound the fractional cost of the negative edges:

$$\sum_{(u,v)\in E^-} (1 - d_{uv}) \leq \sum_{(u,v)\in E^-} |N_u^+ \cap N_v^+|/\Delta = \sum_{e\in E^-} \sum_{T\in\mathcal{T}:T\ni e} 1/\Delta = \sum_{e\in E^-} \sum_{T\in\mathcal{T}:T\ni e} 2y_T,$$

where in the first inequality we have used that $|N_u^+ \cup N_v^+| \geq \Delta$. Next we bound the fractional cost of the positive edges:

$$\sum_{(u,v)\in E^+} d_{uv} \leq \sum_{(u,v)\in E^+} (|N_u^+ \cap N_v^-| + |N_u^- \cap N_v^+|)/\Delta = \sum_{e\in E^+} \sum_{T\in\mathcal{T}:T\ni e} 1/\Delta = \sum_{e\in E^+} \sum_{T\in\mathcal{T}:T\ni e} 2y_T.$$

So the total fractional cost is bounded by $\sum_{e\in E} \sum_{T\in\mathcal{T}:T\ni e} 2y_T = 6 \cdot \sum_{T\in\mathcal{T}} y_T$, since each triangle contains three edges. This is what we sought to show. Since the fractional cost of d is bounded for the ℓ_1-norm objective (and the ℓ_∞-norm objective by [19]), using d as input to KMZ rounding algorithm produces a clustering that is simultaneously $O(1)$-approximate for the ℓ_1- and ℓ_∞-norm objectives. ◀

4 Proof of Theorem 1

The goal of this section is to prove Theorem 1 and the subsequent Corollary 2. We begin by outlining that the adjusted correlation metric satisfies an approximate triangle inequality in Subsection 4.1. Then in Subsection 4.2, we prove the fractional cost of the adjusted correlation metric in any ℓ_p-norm objective is an $O(1)$ factor away from the optimal solution's value. We tie it all together to prove Theorem 1 and Corollary 2 in Subsection 4.3.

We start with an easy but key proposition. Loosely, it states that if two vertices are close to each other according to d, then they have a large shared positive neighborhood.

▶ **Proposition 8.** *Fix vertices $u, v \in V$ and a clustering $\mathcal{C}$ on V such that $d_{uv} \leq 0.7$ and $|N_u^+ \cap C(u)| \,/\, |N_u^+| \geq 0.85$. Then $|N_u^+ \cap N_v^+ \cap C(u)| \geq 0.15 \cdot |N_u^+|$.*

4.1 Triangle inequality

Recall that the correlation metric d satisfies the triangle inequality (see Section 4.2 in [19]). We will show that the adjusted correlation metric f satisfies an approximate triangle inequality, which is sufficient for the KMZ rounding algorithm. Formally, we say that a function g is a *δ-semi-metric* on some set S if it is a semi-metric on S, except instead of satisfying the triangle inequality, g satisfies $g(u,v) \leq \delta \cdot (g(u,w) + g(v,w))$ for all $u, v, w \in S$.

▶ **Lemma 9** (Triangle Inequality). *The adjusted correlation metric f is a $\frac{10}{7}$-semi-metric.*

The proof of Lemma 9 is straightforward given that d satisfies the triangle inequality.

Lemma 3 in [19] proves that one can input a semi-metric that satisfies an approximate triangle inequality (instead of the exact triangle inequality) to the KMZ rounding algorithm (with some loss in the approximation factor). We summarize the main take-away below.

▶ **Lemma 10** ([19]). *If g is a δ-semi-metric on the set V, instead of a true semi-metric (i.e., 1-semi-metric), then the KMZ algorithm loses a factor of* $1 + \delta + \delta^2 + \delta^3 + \delta^4$.[10]

Since we show in Lemma 9 that f is a $\frac{10}{7}$-semi-metric, we lose a factor of 12 in inputting f to the KMZ algorithm (along with the factor loss from the fractional cost).

4.2 Bounding the fractional cost of ℓ_p-norms

This section bounds the fractional cost of the adjusted correlation metric for the ℓ_p-norms. The following lemma considers the case where $p = \infty$. The general case is handled after.

▶ **Lemma 11.** *The fractional cost of the adjusted correlation metric f in the ℓ_∞-norm objective is at most $56 \cdot \mathit{OPT}$, where OPT is the cost of the optimal integral solution.*

The lemma follows from the fact that the fractional cost of the correlation metric d in the ℓ_∞-norm is known to be bounded by [19], and that it only decreases when d is replaced by f due to Definition 6. See Appendix B in the full version [20] for a proof.

We use two primary lemmas – one for the positive edge fractional cost and one for the negative edge fractional cost – to show that the adjusted correlation metric well approximates the optimal for general ℓ_p-norms.

▶ **Lemma 12.** *The fractional cost of the adjusted correlation metric f in the ℓ_p-norm objective is a constant factor (independent of p) away from the cost of the optimal integral ℓ_p solution.*

Proof. Let y be the disagreement vector for an optimal clustering $\mathcal{C}$ in the ℓ_p-norm, for any $p \in \mathbb{R}_{\geq 1} \cup \{\infty\}$. When $p = \infty$, see Lemma 11. For $p \in \mathbb{R}_{\geq 1}$, by definition $\mathsf{OPT}^p = \sum_{w \in V} (y(w))^p$, and the pth power of the fractional cost of f is given by

$$\mathit{cost}(f)^p = \sum_{u \in V} \left[\sum_{v \in N_u^+} f_{uv} + \sum_{v \in N_u^-} (1 - f_{uv}) \right]^p .$$

Observe that $$\mathit{cost}(f)^p \leq 2^p \underbrace{\sum_{u \in V} \Big(\sum_{v \in N_u^+} f_{uv} \Big)^p}_{(S^+)^p} + 2^p \underbrace{\sum_{u \in V} \Big(\sum_{v \in N_u^-} (1 - f_{uv}) \Big)^p}_{(S^-)^p} .$$

We refer to bounding $(S^+)^p$ as bounding the fractional cost of the positive edges, and likewise $(S^-)^p$ for the negative edges. The first sum, $(S^+)^p$, is bounded in Lemma 13 and the second sum, $(S^-)^p$, is bounded in Lemma 16. Using those two bounds, together we have $\mathit{cost}(f) \leq [2^p((S^+)^p + (S^-)^p)]^{1/p} \leq 529$, for $p \in [1, \infty)$. Specifically, the middle term is maximized at $p = 1$, giving the bound of 529, and tends to below 214 as $p \to \infty$. (A more tailored analysis gives a constant of 74 for $p = 1$; see Appendix C in the full version [20].) ◀

We note that, as our main interest is determining whether a simultaneous constant approximation is even possible (and a combinatorial one, at that), we did not pay particular attention to optimizing constants, but suspect these could be greatly reduced.

[10] When $\delta = 1$, this factor equals 5, which is the loss in the KMZ algorithm.

4.2.1 Fractional cost of positive edges in ℓ_p-norms

We first bound the fractional cost of the positive edges.

▶ **Lemma 13.** *For $p \in \mathbb{R}_{\geq 1}$, the fractional cost of the adjusted correlation metric f in the ℓ_p-norm objective for the set of positive edges is a constant factor approximation to the optimal, i.e.,*

$$(S^+)^p = \sum_{u \in V} \Big(\sum_{v \in N_u^+} f_{uv} \Big)^p \leq 2^p \cdot [(8^p/2 + 1)((20/3)^p + 2 + 2 \cdot 4^p) + 8^p + 1] \cdot OPT^p.$$

One of the challenges in bounding the cost of f is that disagreements in the ℓ_p-norm objective for $p \neq 1$ are asymmetric, in that a disagreeing edge charges $y(u)$ and $y(v)$ (whereas for $p = 1$ we can just sum the number of disagreeing edges). Step 3 rounds up the edges incident to u when the tradeoff is good *from u's perspective*. However, an edge (u, v) may be rounded up to 1 when this tradeoff is good from v's perspective, but not from u's perspective. The high-level idea for why this is fine is that if u and v are close under d, their positive neighborhoods overlap significantly and, in some average sense, u can charge to v. Proving this requires a double counting argument using a bipartite auxiliary graph. If u and v are far under d, on the other hand, we can charge to the cost of the correlation metric, which will be bounded on an appropriate subgraph. The second challenge is showing that the ℓ_p-norm of the disagreement vector, restricted to vertices u that are made singletons in Step 3, is bounded. This again requires a double counting argument.

Proof. Fix an optimal clustering $\mathcal{C}$. We partition vertices based on membership in $C(u)$ or $\overline{C(u)}$ (as defined in Subsection 2.1). Let y denote the disagreement vector of $\mathcal{C}$. We have

$$(S^+)^p = \sum_{u \in V} \Big(\sum_{v \in N_u^+} f_{uv} \Big)^p \leq 2^p \underbrace{\sum_{u \in V} \Big(\sum_{v \in N_u^+ \cap C(u)} f_{uv} \Big)^p}_{S_1^+} + 2^p \underbrace{\sum_{u \in V} \Big(\sum_{v \in N_u^+ \cap \overline{C(u)}} f_{uv} \Big)^p}_{S_2^+}.$$

It is easy to bound S_2^+ by using the trivial upper bound $f_{uv} \leq 1$:

$$\boxed{S_2^+ = \sum_{u \in V} \Big(\sum_{v \in N_u^+ \cap \overline{C(u)}} f_{uv} \Big)^p \leq \sum_{u \in V} \Big(\sum_{v \in N_u^+ \cap \overline{C(u)}} 1 \Big)^p \leq \sum_{u \in V} (y(u))^p = \mathsf{OPT}^p,}$$

where we used that every edge $(u, v) \in E^+$ with $v \notin C(u)$ is a disagreement incident to u.

Next, we bound S_1^+. Let R_1 be the set of u for which Step 3 of Definition 6 applies. For these u, we have $f_{uv} = 1$ for all $v \in V \setminus \{u\}$. Let $R_2 = V \setminus R_1$. For $u \in R_2$ and $v \in N_u^+$, we have that either $v \in R_2$, in which case $f_{uv} = d_{uv}$; or $v \in R_1$, in which case $f_{uv} = 1$. (Note that V is the disjoint union of R_1 and R_2.) So

$$S_1^+ = \underbrace{\sum_{u \in R_1} \Big(\sum_{v \in N_u^+ \cap C(u), v \neq u} 1 \Big)^p}_{S_{11}^+} + \underbrace{\sum_{u \in R_2} \Big(\sum_{v \in N_u^+ \cap C(u)} f_{uv} \Big)^p}_{S_{12}^+},$$

and in particular

$$
\begin{aligned}
S_{12}^{+} &= \sum_{u\in R_2}\Big(\sum_{v\in N_u^+\cap C(u)\cap R_1} 1 + \sum_{v\in N_u^+\cap C(u)\cap R_2} d_{uv}\Big)^p \\
&= \sum_{u\in R_2}\Big(\sum_{\substack{v\in N_u^+\cap C(u)\cap R_1\\ d_{uv}\le 1/4}} 1 + \sum_{\substack{v\in N_u^+\cap C(u)\cap R_1\\ d_{uv}\ge 1/4}} 1 + \sum_{v\in N_u^+\cap C(u)\cap R_2} d_{uv}\Big)^p \\
&\le \sum_{u\in R_2}\Big(\sum_{\substack{v\in N_u^+\cap C(u)\cap R_1\\ d_{uv}\le 1/4}} 1 + \sum_{\substack{v\in N_u^+\cap C(u)\cap R_1\\ d_{uv}\ge 1/4}} 4\cdot d_{uv} + \sum_{v\in N_u^+\cap C(u)\cap R_2} d_{uv}\Big)^p \\
&\le \sum_{u\in R_2}\Big(\sum_{\substack{v\in N_u^+\cap C(u)\cap R_1\\ d_{uv}\le 1/4}} 1 + \sum_{v\in N_u^+\cap C(u)} 4\cdot d_{uv}\Big)^p \\
&\le 2^p \underbrace{\sum_{u\in R_2}\Big(\sum_{\substack{v\in N_u^+\cap R_1\\ d_{uv}\le 1/4}} 1\Big)^p}_{S_{13}^+} + 8^p\cdot \underbrace{\sum_{u\in R_2}\Big(\sum_{v\in N_u^+\cap C(u)} d_{uv}\Big)^p}_{S_{14}^+}.
\end{aligned}
$$

First we bound S_{13}^+. We will strongly use that $d_{uv} \le 1/4$ in the inner sum. Observe:

▶ **Proposition 14.** *Let d be the correlation metric, and $d_{uv} \le 1/4$. Then $|N_u^+| \le \frac{7}{3}\cdot|N_v^+|$.*

Next, we will need to create a bipartite auxiliary graph $H = (R_2, R_1, F)$ with R_2 and R_1 being the sides of the partition, and F being the edge set. We will then use a double counting argument. Place an edge between $u \in R_2$ and $v \in R_1$ if $uv \in E^+$ and $d_{uv} \le 1/4$. Then we have precisely that $S_{13}^+ = \sum_{u\in R_2} \deg_H(u)^p$. We will show that

$$\boxed{S_{13}^+ = \sum_{u\in R_2}\deg_H(u)^p \le 4^{p-1}\cdot\sum_{v\in R_1}|N_v^+|^p \le 4^{p-1}\cdot((20/3)^p+2+2\cdot 4^p)\cdot \mathsf{OPT}^p} \tag{1}$$

where the last bound follows from Proposition 15, which we establish separately below. We will bound via double counting the quantity L, defined below. Let $N_H(\cdot)$ denote the neighborhoods in H of the vertices.

$$
\begin{aligned}
L &:= \sum_{f=uv\in F}(\deg_H(u)+\deg_H(v))^{p-1} \le \sum_{v\in R_1}\sum_{u\in N_H(v)}(\deg_H(v)+\deg_H(u))^{p-1} \\
&\le \sum_{v\in R_1}\sum_{u\in N_H(v)}\big(|N_v^+|+|N_u^+|\big)^{p-1} \le \sum_{v\in R_1}\sum_{u\in N_H(v)} 4^{p-1}\cdot|N_v^+|^{p-1} \qquad (2)\\
&\le 4^{p-1}\cdot\sum_{v\in R_1}|N_v^+|\cdot|N_v^+|^{p-1} = 4^{p-1}\cdot\sum_{v\in R_1}|N_v^+|^p
\end{aligned}
$$

where in (2) we've used Proposition 14. Note that L is upper bounded by the right-hand side in (1). Now it just remains to show that L is lower bounded by the left-hand side in (1).

$$
\begin{aligned}
L &= \sum_{f=uv\in F}(\deg_H(u)+\deg_H(v))^{p-1} = \sum_{u\in R_2}\sum_{v\in N_H(u)}(\deg_H(u)+\deg_H(v))^{p-1} \\
&\ge \sum_{u\in R_2}\sum_{v\in N_H(u)}\deg_H(u)^{p-1} = \sum_{u\in R_2}\deg_H(u)\cdot\deg_H(u)^{p-1} = \sum_{u\in R_2}\deg_H(u)^p,
\end{aligned}
$$

which is what we sought to show. Now we bound S_{14}^+.

$$\begin{aligned} S_{14}^+ &\leq \sum_{u \in V} \left(\sum_{v \in N_u^+ \cap C(u)} \frac{|N_u^+ \cap N_v^-| + |N_u^- \cap N_v^+|}{|N_u^+ \cup N_v^+|} \right)^p \\ &\leq \sum_{u \in V} \left(\sum_{v \in N_u^+} \frac{y(u) + y(v)}{|N_u^+ \cup N_v^+|} \right)^p \leq \sum_{u \in V} |N_u^+|^{p-1} \sum_{v \in N_u^+} \frac{(y(u) + y(v))^p}{|N_u^+ \cup N_v^+|^p} \\ &\leq 2^p \sum_{u \in V} \sum_{v \in N_u^+} |N_u^+|^{p-1} \cdot \frac{y(u)^p}{|N_u^+ \cup N_v^+|^p} + 2^p \sum_{u \in V} \sum_{v \in N_u^+} |N_u^+|^{p-1} \cdot \frac{y(v)^p}{|N_u^+ \cup N_v^+|^p}. \end{aligned}$$

In the second line, the first inequality uses the fact that for $w \in (N_u^+ \cap N_v^-) \cup (N_u^- \cap N_v^+)$, then at least one of $(u, w), (v, w)$ is a disagreement, since $v \in C(u)$ in the inner summation of the first line. The second inequality in the second line uses Jensen's inequality.

To bound the first double sum above, we use an averaging argument:

$$\sum_{u \in V} \sum_{v \in N_u^+} |N_u^+|^{p-1} \cdot \frac{y(u)^p}{|N_u^+ \cup N_v^+|^p} \leq \sum_{u \in V} \sum_{v \in N_u^+} \frac{y(u)^p}{|N_u^+|} = \sum_{u \in V} y(u)^p = \mathsf{OPT}^p.$$

To bound the second double sum, we first have to flip it:

$$\begin{aligned} \sum_{u \in V} \sum_{v \in N_u^+} |N_u^+|^{p-1} \cdot \frac{y(v)^p}{|N_u^+ \cup N_v^+|^p} &= \sum_{v \in V} \sum_{u \in N_v^+} |N_u^+|^{p-1} \cdot \frac{y(v)^p}{|N_u^+ \cup N_v^+|^p} \\ &\leq \sum_{v \in V} \sum_{u \in N_v^+} |N_u^+|^{p-1} \frac{y(v)^p}{|N_u^+|^{p-1} \cdot |N_v^+|} = \sum_{v \in V} y(v)^p = \mathsf{OPT}^p. \end{aligned}$$

In total, we have

$$\boxed{S_{14}^+ \leq 2 \cdot 2^p \cdot \mathsf{OPT}^p = 2^{p+1} \cdot \mathsf{OPT}^p}$$

and $\boxed{S_{12}^+ \leq 2^p \cdot S_{13}^+ + 8^p \cdot S_{14}^+ \leq 2^p \cdot 4^{p-1} \cdot ((20/3)^p + 2 + 2 \cdot 4^p) \cdot \mathsf{OPT}^p + 8^p \cdot \mathsf{OPT}^p.}$

Next we turn to bounding S_{11}^+. Recall that $R_1 = \{u : |N_u^- \cap \{v : d_{uv} \leq 0.7\}| \geq \frac{10}{3} \cdot \Delta_u\}$ and

$$S_{11}^+ \leq \sum_{u \in R_1} |N_u^+ \cap C(u)|^p \leq \sum_{u \in R_1} |N_u^+|^p.$$

So it suffices to bound the right-hand side, which we do in the following proposition.

▶ **Proposition 15.** *Let R_1 be the set of u for which Step 3 of Definition 6 applies. Then*

$$\sum_{u \in R_1} |N_u^+|^p \leq ((20/3)^p + 2 + 2 \cdot 4^p) \cdot OPT^p.$$

Proof of Proposition 15. For $u \in R_1$, define $R_1(u) = N_u^- \cap \{v : d_{uv} \leq 0.7\}$, so in particular $|R_1(u)| \geq \frac{10}{3} \cdot \Delta_u$. Fix a vertex $u \in R_1$. We consider a few cases. The crux is Case 2a(ii).

Case 1. *At least a 0.15 fraction of N_u^+ is in clusters other than $C(u)$.*
Let $u \in V^1$ be the vertices in this case. This means that $0.15 \cdot |N_u^+| \leq y(u)$, so

$$\boxed{\sum_{u \in V^1} |N_u^+|^p \leq \sum_{u \in V^1} \frac{1}{0.15^p} y(u)^p \leq (20/3)^p \cdot \mathsf{OPT}^p.}$$

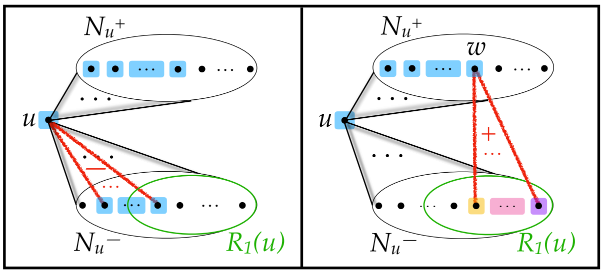

Figure 2 Left: Case 2a(i). For $v \in |N_u^- \cap C(u)|$, (u, v) is a disagreement. **Right:** Case 2a(ii). For $w \in |N_u^+ \cap C(u)|$ and $v \in N_w^+ \cap R_1'(u)$, (w, v) is a disagreement.

Case 2. *At least a 0.85 fraction of N_u^+ is in $C(u)$.*
We further partition the cases based on how much $R_1(u)$ intersects $C(u)$.

Case 2a: *At least half of $R_1(u)$ is in clusters other than $C(u)$.*
We partition into cases (just one more time!) based on the size of $N_u^- \cap C(u)$. See Figure 2.

Case 2a(i): *At least half of $R_1(u)$ is in clusters other than $C(u)$ and $|N_u^- \cap C(u)| \geq \Delta_u$.*
Let $u \in V^{2a(i)}$ be the vertices in this case. Note that $y(u) \geq |N_u^- \cap C(u)|$. Then

$$\boxed{\sum_{u \in V^{2ai}} |N_u^+|^p = \sum_{u \in V^{2ai}} \Delta_u^p \leq \sum_{u \in V^{2ai}} |N_u^- \cap C(u)|^p \leq \sum_{u \in V^{2ai}} y(u)^p \leq \mathsf{OPT}^p.}$$

Case 2a(ii): *At least half of $R_1(u)$ is in clusters other than $C(u)$ and $|N_u^- \cap C(u)| \leq \Delta_u$.*

Let $u \in V^{2a(ii)}$ be the vertices in this case. Denote the vertices in $R_1(u)$ that are in clusters other than $C(u)$ by $R_1'(u)$. By definition of Case 2a(ii), $|R_1'(u)| \geq \frac{5}{3} \cdot \Delta_u$. A key fact we will use is that $|C(u)| \leq 2 \cdot \Delta_u$:

$$|C(u)| = |N_u^- \cap C(u)| + |N_u^+ \cap C(u)| \leq \Delta_u + \Delta_u = 2 \cdot \Delta_u.$$

For $u \in V^{2a(ii)}$ and $w \in N_u^+ \cap C(u)$, define $\varphi(u, w) = |R_1'(u) \cap N_w^+|$.
Each $w \in N_u^+ \cap C(u)$ dispenses $\varphi(u, w)^p / |C(u)|$ charge to u. Also, observe that for $v \in R_1'(u)$, we have that $d_{uv} \leq 0.7$, so we know by Proposition 8 that $|N_u^+ \cap N_v^+ \cap C(u)| \geq 0.15 \cdot |N_u^+|$. This implies that

$$\begin{aligned}
\sum_{w \in N_u^+ \cap C(u)} |R_1'(u) \cap N_w^+| &= \sum_{w \in N_u^+ \cap C(u)} \sum_{v \in R_1'(u) \cap N_w^+} 1 = \sum_{v \in R_1'(u)} \sum_{\substack{w \in \cap N_v^+ \\ \cap C(u) \cap N_u^+}} 1 \\
&= \sum_{v \in R_1'(u)} |C(u) \cap N_u^+ \cap N_v^+| \geq \sum_{v \in R_1'(u)} 0.15 \cdot |N_u^+| \\
&= 0.15 \cdot |N_u^+| \cdot |R_1'(u)| \geq 0.15 \cdot \Delta_u \cdot \frac{5}{3} \Delta_u = 0.25 \cdot \Delta_u^2.
\end{aligned}$$

By Jensen's inequality, the amount of charge each u satisfying Case 2a(ii) receives is at least

$$\begin{aligned}
\frac{1}{|C(u)|} \sum_{w \in N_u^+ \cap C(u)} \varphi(u, w)^p &\geq \frac{1}{|C(u)|} \cdot \frac{1}{|N_u^+ \cap C(u)|^{p-1}} \cdot \Big(\sum_{w \in N_u^+ \cap C(u)} \varphi(u, w) \Big)^p \\
&\geq \frac{1}{2\Delta_u} \cdot \frac{1}{\Delta_u^{p-1}} \cdot \Big(0.25 \cdot \Delta_u^2 \Big)^p \geq \frac{1}{2} \cdot 0.25^p \cdot |N_u^+|^p,
\end{aligned}$$

Next we need to upper bound the amount of charge dispensed in total to *all* u satisfying Case 2a(ii). Note by definition that $\varphi(u,w) \le y(w)$. Each vertex $w \in V$ dispenses at most $y(w)^p/|C(u)| = y(w)^p/|C(w)|$ charge to each $u \in C(w) \cap N_w^+$. So in total each w dispenses at most $|C(w)| \cdot y(w)^p/|C(w)| = y(w)^p$ charge to all u satisfying Case 2a(ii). Now we put together the lower and upper bounds on the total charge dispensed:

$$\sum_{w\in V} y(w)^p \ge \text{charge dispensed} \ge \sum_{u\in V^{2a(ii)}} \frac{1}{|C(u)|} \sum_{w\in N_u^+\cap C(u)} \varphi(u,w)^p$$
$$\ge \sum_{u\in V^{2a(ii)}} \frac{1}{2} \cdot 0.25^p \cdot |N_u^+|^p.$$

In all, $\boxed{\sum_{u\in V^{2a(ii)}} |N_u^+|^p \le 2\cdot 4^p \cdot \sum_{w\in V} y(w)^p \le 2\cdot 4^p \cdot \mathsf{OPT}^p.}$

Case 2b: *At least half of $R_1(u)$ is in $C(u)$.*
Let $u \in V^{2b}$ be the vertices in this case. Denote the vertices in $R_1(u)$ that are in $C(u)$ by $R_1''(u)$. By definition of Case 2b, $|R_1''(u)| \ge \frac{5}{3}\cdot\Delta_u$. Since every vertex in $R''(u)$ is in N_u^-, there are at least $|R''(u)|$ disagreements incident to u. So $y(u) \ge |R''(u)| \ge \frac{5}{3}\cdot\Delta_u$, giving

$$\boxed{\sum_{u\in V^{2b}} |N_u^+|^p = \sum_{u\in V^{2b}} \Delta_u^p \le \sum_{u\in V^{2b}} y(u)^p \le \mathsf{OPT}^p.}$$

Adding the terms in the boxed expressions across all cases, the proposition follows. ◀

So we have $\boxed{S_{11}^+ \le \sum_{u\in R_1} |N_u^+|^p \le ((20/3)^p + 2 + 2\cdot 4^p)\cdot \mathsf{OPT}^p.}$

Adding together all the cases, we conclude that

$$(S^+)^p \le 2^p\cdot(S_1^+ + S_2^+) \le 2^p\cdot(S_{11}^+ S_{12^+} + S_2^+) \le 2^p\cdot[(8^p/2+1)((20/3)^p+2+2\cdot4^p)+8^p+1]\cdot\mathsf{OPT}^p.$$

◀

4.2.2 Fractional cost of negative edges in ℓ_p-norms

This section bounds the cost of negative edges. The meanings of $\mathcal{C}$, $C(\cdot)$, and y are as before.

▶ **Lemma 16.** *For $p \in \mathbb{R}_{\ge 1}$, the fractional cost of the adjusted correlation metric f in the ℓ_p-norm objective for the set of negative edges is a constant factor away from optimal:*

$$(S^-)^p = \sum_{u\in V}\Big(\sum_{v\in N_u^-} (1-f_{uv})\Big)^p \le 2^p((200/9)^p + 1 + (10/3)^p + 2\cdot(20/3)^p)\cdot OPT^p.$$

Proof. We have

$$(S^-)^p = \sum_{u\in V}\Big(\sum_{v\in N_u^-} (1-f_{uv})\Big)^p \le 2^p \underbrace{\sum_{u\in V}\Big(\sum_{v\in N_u^-\cap C(u)} (1-f_{uv})\Big)^p}_{S_1^-}$$
$$+ 2^p \underbrace{\sum_{u\in V}\Big(\sum_{v\in N_u^-\cap \overline{C(u)}} (1-f_{uv})\Big)^p}_{S_2^-}.$$

It is easy to bound S_1^- by using the trivial upper bound $1 - f_{uv} \le 1$:

$$\boxed{S_1^- = \sum_{u \in V} \Big(\sum_{v \in N_u^- \cap C(u)} (1 - f_{uv}) \Big)^p \le \sum_{u \in V} \Big(\sum_{v \in N_u^- \cap C(u)} 1 \Big)^p \le \sum_{u \in V} y(u)^p = \mathsf{OPT}^p,}$$

where we have used that every edge $(u,v) \in E^-$ with $v \in C(u)$ is a disagreement incident to u. Next, we bound S_2^-. Let R_1 and R_2 be as in the previous subsection: $R_1 = \{u : |N_u^- \cap \{v : d_{uv} \le 0.7\}| \ge \frac{10}{3} \cdot \Delta_u\}$ and $R_2 = V \setminus R_1$. For $u \in R_2$, define

$$V_u = \{v : v \in N_u^- \cap \overline{C(u)}, d_{uv} \le 0.7\}.$$

Note that the definition of V_u is the same as $R_1'(u)$ in the previous subsection, but here V_u is only defined for $u \in R_2$, while $R_1'(u)$ was defined for $u \in R_1$. For $u \in R_1$, we have $1 - f_{uv} = 0$ for every $v \in V \setminus \{u\}$. So the outer sum in S_2^- only need be taken over $u \in R_2$:

$$S_2^- = \sum_{u \in R_2} \Big(\sum_{v \in N_u^- \cap \overline{C(u)}} (1 - f_{uv}) \Big)^p \le \sum_{u \in R_2} \Big(\sum_{\substack{v : v \in N_u^- \cap \overline{C(u)}, \\ d_{uv} \le 0.7}} (1 - d_{uv}) \Big)^p \le \sum_{u \in R_2} |V_u|^p$$

In the second equality, we have used that if $u \in R_2$ and $v \in N_u^-$, then $f_{uv} = d_{uv}$, unless f_{uv} was rounded up to 1 in Step 2 of Definition 6 (which happens when $d_{uv} > 0.7$), or f_{uv} was rounded up to 1 in Step 3 (in which case $1 - f_{uv} = 0 \le 1 - d_{uv}$).

A key observation is that since $u \in R_2$, it is the case that $|V_u| \le \frac{10}{3} \cdot \Delta_u$.

Fix a vertex $u \in R_2$. We consider a few cases.

Case 1. *At least a 0.15 fraction of N_u^+ is in clusters other than $C(u)$.*
Define V^1 to be the set of $u \in R_2$ satisfying Case 1. Then for $u \in V^1$, $0.15 \cdot |N_u^+| \le y(u)$, and

$$|V_u| \le \frac{10}{3} \Delta_u \le \frac{1}{0.15} \cdot \frac{10}{3} y(u) = \frac{200}{9} y(u).$$

so $\boxed{\sum_{u \in V^1} |V_u|^p \le (200/9)^p \cdot \sum_{u \in V^1} y(u)^p \le (200/9)^p \cdot \mathsf{OPT}^p.}$

Case 2. *At least a 0.85 fraction of N_u^+ is in $C(u)$.*
Define V^2 to be the set of $u \in R_2$ that satisfy Case 2. Fix $u \in V^2$ and $v \in V_u$. Define $N_{u,v} = N_u^+ \cap N_v^+ \cap C(u)$. Since $d_{uv} \le 0.7$ and by the assumption of this case, using Proposition 8 we have

$$|N_{u,v}| = |N_u^+ \cap N_v^+ \cap C(u)| \ge 0.15 \cdot \Delta_u.$$

Observe that since $v \notin C(u)$ for $v \in V_u$, (v, w) is a (positive) disagreement for all $w \in N_{u,v}$.
Case 2a: $|N_u^- \cap C(u)| \ge \Delta_u$.

Define V^{2a} to be the set of $u \in V^2$ that satisfy Case 2a. Since all edges (u,v) with $v \in N_u^- \cap C(u)$ are disagreements, we have $y(u) \ge \Delta_u$. Recalling that $|V_u| \le \frac{10}{3} \cdot \Delta_u$ for $u \in R_2$, we have

$$\boxed{\sum_{u \in V^{2a}} |V_u|^p \le \sum_{u \in V^{2a}} (10/3 \cdot \Delta_u)^p \le (10/3)^p \cdot \sum_{u \in V^{2a}} y(u)^p \le (10/3)^p \cdot \mathsf{OPT}^p.}$$

Case 2b: $|C(u)| \leq 2\Delta_u$.

Define V^{2b} to be the $u \in V^2$ satisfying Case 2b. Fix $w \in N_u^+ \cap C(u)$ and $u \in V^{2b}$. Define

$$\varphi(u,w) = |V_u \cap N_w^+|,$$

i.e. $\varphi(u,w)$ is the number of $v \in V_u$ with $w \in N_{u,v}$. Each $w \in N_u^+ \cap C(u)$ dispenses $\frac{\varphi(u,w)^p}{|C(u)|}$ charge to u. Also,

$$\begin{aligned}\sum_{w\in N_u^+\cap C(u)} \varphi(u,w) &= \sum_{w\in N_u^+\cap C(u)} |V_u \cap N_w^+| = \sum_{w\in N_u^+\cap C(u)} \sum_{v\in V_u\cap N_w^+} 1 \\ &= \sum_{v\in V_u}\sum_{w\in N_{u,v}} 1 = \sum_{v\in V_u} |N_{u,v}| \geq |V_u|\cdot 0.15\cdot\Delta_u.\end{aligned}$$

By Jensen's inequality, the amount of charge each u satisfying Case 2b receives is at least

$$\begin{aligned}\frac{1}{|C(u)|}\sum_{w\in N_u^+\cap C(u)} \varphi(u,w)^p &\geq \frac{1}{|C(u)|}\cdot\frac{1}{|N_u^+\cap C(u)|^{p-1}}\Big(\sum_{w\in N_u^+\cap C(u)}\varphi(u,w)\Big)^p \\ &\geq \frac{1}{2\Delta_u}\cdot\frac{1}{\Delta_u^{p-1}}\,(|V_u|\cdot 0.15\cdot\Delta_u)^p = \frac{1}{2}\cdot 0.15^p\cdot|V_u|^p,\end{aligned}$$

To upper bound the amount of charge dispensed in total to *all* u satisfying Case 2b, first note that $\varphi(u,w) \leq y(w)$. Also, each vertex $w \in V$ only distributes charge to $u \in C(w) \cap N_w^+$, and the amount of charge distributed to each such u is

$$\frac{\varphi(u,w)^p}{|C(u)|} = \frac{\varphi(u,w)^p}{|C(w)|} \leq \frac{y(w)^p}{|C(w)|},$$

so that in total each w dispenses at most $\frac{y(w)^p}{|C(w)|}\cdot|C(w)| \leq y(w)^p$ charge. Putting together the lower and upper bounds on the amount of charge dispensed:

$$\begin{aligned}\sum_{w\in V} y(w)^p \geq \text{total charge dispensed} &\geq \sum_{u\in V^{2b}}\frac{1}{|C(u)|}\sum_{w\in N_u^+\cap C(u)}\varphi(u,w)^p \\ &\geq \sum_{u\in V^{2b}}\frac{1}{2}\cdot 0.15^p\cdot|V_u|^p.\end{aligned}$$

In all, $\boxed{\sum_{u\in V^{2b}} |V_u|^p \leq 2\cdot(20/3)^p\cdot\sum_{w\in V} y(w)^p = 2\cdot(20/3)^p\cdot\mathsf{OPT}^p.}$

Combining the cases, we see that $(S^-)^p \leq 2^p((200/9)^p + 1 + (10/3)^p + 2\cdot(20/3)^p)\cdot\mathsf{OPT}^p$. ◀

4.3 Proofs of Theorem 1 and Corollary 2

Here we show that Theorem 1 follows directly from the preceding lemmas. We defer the proof of Corollary 2 to the full version [20].

Proof of Theorem 1. First we show that Lemma 12 implies that the clustering resulting from inputting f into the KMZ rounding algorithm is $O(1)$-approximate in any ℓ_p-norm. Since the rounding algorithm does not depend on p, the clustering will be the same for all p. Let $\mathcal{C}^*$ be the clustering produced by running the KMZ rounding algorithm with the adjusted correlation metric f as input. Let $\mathsf{ALG}(u)$ be the number of edges incident to u that are disagreements with respect to $\mathcal{C}^*$. From [29] and Lemmas 9 and 10, we have that for every $u \in V$, $\mathsf{ALG}(u) \leq 12 \cdot y_u$ where y_u is as in LP P when taking $x = f$. So $||y||_p$ is the fractional cost of f in the ℓ_p-norm. In what follows, $||\mathsf{ALG}||_p$ is the objective value of $\mathcal{C}^*$ in the ℓ_p-norm and $\mathsf{OPT}(p)$ is the optimal objective value in the ℓ_p-norm. Thus using Lemma 12 in the last inequality, we have

$$||\mathsf{ALG}||_p \leq 12 \cdot ||y||_p \leq 12 \cdot 529 \cdot \mathsf{OPT}(p) = 6348 \cdot \mathsf{OPT}(p).$$

The overall run-time is $O(n^\omega)$. From the analysis in [19], computing the correlation metric takes time $O(n^\omega)$, and the KMZ rounding algorithm takes time $O(n^2)$. We just have to show the post-processing of d in Steps 2 and 3 of Definition 6 that were done in order to obtain the adjusted correlation metric f can be done quickly. Indeed, Step 2 takes $O(n^2)$ time as it simply iterates through the edges. Step 3 also takes $O(n^2)$ time, since it visits each vertex and iterates through the neighbors. Thus, the run-time remains $O(n^\omega)$. ◀

5 Conclusion

This paper considered correlation clustering on unweighted, complete graphs, a problem that arises in many settings including community detection and the study of large networks. All previous works that study minimizing the ℓ_p-norm (for $p \in \mathbb{R}_{>1}$) of the disagreement vector rely on solving a large, convex relaxation (which is costly to the algorithm's run-time) and produce a solution that is only $O(1)$-approximate for one specific value of p. We innovate upon this rich line of work by (1) giving the first combinatorial algorithm for the ℓ_p-norms for $p \in \mathbb{R}_{>1}$, (2) designing scalable algorithms for this practical problem, and (3) obtaining solutions that are $O(1)$-approximate for all ℓ_p-norms (for $p \in \mathbb{R}_{\geq 1} \cup \{\infty\}$) simultaneously. We emphasize this last point, as such solutions are good in both global and local senses, and thus may be more desirable than typical optimal or approximate solutions. The existence of these solutions reveals a surprising structural property of correlation clustering.

One question is whether there is a simpler existential (not necessarily algorithmic) proof that there exists an $O(1)$-approximation for the all-norm objective for correlation clustering.

It is also of interest to implement the KMZ algorithm with the adjusted correlation metric as input, and empirically gain an understanding of how good the adjusted correlation metric is for different ℓ_p-norms. We suspect that our analysis is lossy (we did not focus on optimizing constants), and that the approximation obtained would be of much better quality.

References

1. Saba Ahmadi, Sainyam Galhotra, Barna Saha, and Roy Schwartz. Fair correlation clustering. *arXiv preprint*, 2020. `arXiv:2002.03508`.
2. Saba Ahmadi, Samir Khuller, and Barna Saha. Min-max correlation clustering via multicut. In *International Conference on Integer Programming and Combinatorial Optimization*, pages 13–26. Springer, 2019.
3. Sara Ahmadian, Alessandro Epasto, Ravi Kumar, and Mohammad Mahdian. Fair correlation clustering. In *International Conference on Artificial Intelligence and Statistics*, pages 4195–4205. PMLR, 2020.
4. Nir Ailon, Moses Charikar, and Alantha Newman. Aggregating inconsistent information: ranking and clustering. *Journal of the ACM (JACM)*, 55(5):1–27, 2008.
5. Soroush Alamdari and David Shmoys. A bicriteria approximation algorithm for the k-center and k-median problems. In *Approximation and Online Algorithms: 15th International Workshop, WAOA 2017, Vienna, Austria, September 7–8, 2017, Revised Selected Papers 15*, pages 66–75. Springer, 2018.

6 Yossi Azar, Leah Epstein, Yossi Richter, and Gerhard J Woeginger. All-norm approximation algorithms. *Journal of Algorithms*, 52(2):120–133, 2004.
7 Nikhil Bansal, Avrim Blum, and Shuchi Chawla. Correlation clustering. *Machine learning*, 56(1):89–113, 2004.
8 Mohammadhossein Bateni, Vincent Cohen-Addad, Alessandro Epasto, and Silvio Lattanzi. Scalable and improved algorithms for individually fair clustering. In *Workshop on Trustworthy and Socially Responsible Machine Learning, NeurIPS 2022*, 2022.
9 Amir Ben-Dor and Zohar Yakhini. Clustering gene expression patterns. In *Proceedings of the third annual international conference on computational molecular biology*, pages 33–42, 1999.
10 Aaron Bernstein, Tsvi Kopelowitz, Seth Pettie, Ely Porat, and Clifford Stein. Simultaneously load balancing for every p-norm, with reassignments. In *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2017.
11 Francesco Bonchi, David Garcia-Soriano, and Edo Liberty. Correlation clustering: from theory to practice. In *KDD*, page 1972, 2014.
12 Costas Busch, Chinmoy Dutta, Jaikumar Radhakrishnan, Rajmohan Rajaraman, and Srivathsan Srinivasagopalan. Split and join: Strong partitions and universal steiner trees for graphs. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 81–90. IEEE, 2012.
13 Ostas Busch, Arnold Filtser, Daniel Hathcock, D Ellis Hershkowitz, Rajmohan Rajaraman, et al. One tree to rule them all: Poly-logarithmic universal steiner tree. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 60–76. IEEE, 2023.
14 Sayak Chakrabarty and Konstantin Makarychev. Single-pass pivot algorithm for correlation clustering. keep it simple! *arXiv preprint*, 2023. `arXiv:2305.13560`.
15 Moses Charikar, Neha Gupta, and Roy Schwartz. Local guarantees in graph cuts and clustering. In Friedrich Eisenbrand and Jochen Könemann, editors, *Integer Programming and Combinatorial Optimization - 19th International Conference, IPCO 2017, Waterloo, ON, Canada, June 26-28, 2017, Proceedings*, volume 10328 of *Lecture Notes in Computer Science*, pages 136–147. Springer, 2017. `doi:10.1007/978-3-319-59250-3_12`.
16 Shuchi Chawla, Konstantin Makarychev, Tselil Schramm, and Grigory 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, 2015.
17 Flavio Chierichetti, Nilesh Dalvi, and Ravi Kumar. Correlation clustering in mapreduce. In *Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 641–650, 2014.
18 Vincent Cohen-Addad, Euiwoong Lee, and Alantha Newman. Correlation clustering with sherali-adams. *Symposium on Foundations of Computer Science (FOCS).*, 2022.
19 Sami Davies, Benjamin Moseley, and Heather Newman. Fast combinatorial algorithms for min max correlation clustering. In *International Conference on Machine Learning*. PMLR, 2023.
20 Sami Davies, Benjamin Moseley, and Heather Newman. Simultaneously approximating all ℓ_p-norms in correlation clustering, 2024. `arXiv:2308.01534`.
21 Erik D Demaine and Nicole Immorlica. Correlation clustering with partial information. In *Approximation, Randomization, and Combinatorial Optimization.. Algorithms and Techniques*, pages 1–13. Springer, 2003.
22 Zachary Friggstad and Ramin Mousavi. Fair correlation clustering with global and local guarantees. In *Workshop on Algorithms and Data Structures*, pages 414–427. Springer, 2021.
23 Arun Ganesh, Bruce M Maggs, and Debmalya Panigrahi. Universal algorithms for clustering problems. *ACM Transactions on Algorithms*, 19(2):1–46, 2023.
24 Daniel Golovin, Anupam Gupta, Amit Kumar, and Kanat Tangwongsan. All-norms and all-l_p-norms approximation algorithms. In *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science*. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2008.

25 Holger Heidrich, Jannik Irmai, and Bjoern Andres. A 4-approximation algorithm for min max correlation clustering. *arXiv preprint*, 2023. arXiv:2310.09196.

26 Jafar Jafarov, Sanchit Kalhan, Konstantin Makarychev, and Yury Makarychev. Local correlation clustering with asymmetric classification errors. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pages 4677–4686. PMLR, 2021. URL: http://proceedings.mlr.press/v139/jafarov21a.html.

27 Jafar Jafarov, Sanchit Kalhan, Konstantin Makarychev, and Yury Makarychev. Local correlation clustering with asymmetric classification errors. In *International Conference on Machine Learning*, pages 4677–4686. PMLR, 2021.

28 Lujun Jia, Guolong Lin, Guevara Noubir, Rajmohan Rajaraman, and Ravi Sundaram. Universal approximations for tsp, steiner tree, and set cover. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 386–395, 2005.

29 Sanchit Kalhan, Konstantin Makarychev, and Timothy Zhou. Correlation clustering with local objectives. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 9341–9350, 2019. URL: https://proceedings.neurips.cc/paper/2019/hash/785ca71d2c85e3f3774baaf438c5c6eb-Abstract.html.

30 Jon Kleinberg, Yuval Rabani, and Éva Tardos. Fairness in routing and load balancing. In *40th Annual Symposium on Foundations of Computer Science (Cat. No. 99CB37039)*, pages 568–578. IEEE, 1999.

31 Zach Langley, Aaron Bernstein, and Sepehr Assadi. Improved bounds for distributed load balancing. In *34th International Symposium on Distributed Computing (DISC 2020). Schloss Dagstuhl – Leibniz-Zentrum für Informatik*, 2020.

32 Andrew McCallum and Ben Wellner. Conditional models of identity uncertainty with application to noun coreference. *Advances in neural information processing systems*, 17, 2004.

33 Xinghao Pan, Dimitris Papailiopoulos, Benjamin Recht, Kannan Ramchandran, and Michael I Jordan. Scaling up correlation clustering through parallelism and concurrency control. In *DISCML Workshop at International Conference on Neural Information Processing Systems*, 2014.

34 Gregory J Puleo and Olgica Milenkovic. Correlation clustering with constrained cluster sizes and extended weights bounds. *SIAM Journal on Optimization*, 25(3):1857–1872, 2015.

35 Gregory J. Puleo and Olgica Milenkovic. Correlation clustering and biclustering with locally bounded errors. In Maria-Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, volume 48 of *JMLR Workshop and Conference Proceedings*, pages 869–877. JMLR.org, 2016. URL: http://proceedings.mlr.press/v48/puleo16.html.

36 Jessica Shi, Laxman Dhulipala, David Eisenstat, Jakub Lacki, and Vahab S. Mirrokni. Scalable community detection via parallel correlation clustering. *Proc. VLDB Endow.*, 14:2305–2313, 2021.

37 Nate Veldt. Correlation clustering via strong triadic closure labeling: Fast approximation algorithms and practical lower bounds. In *International Conference on Machine Learning*, pages 22060–22083. PMLR, 2022.

38 Nate Veldt, David F Gleich, and Anthony Wirth. A correlation clustering framework for community detection. In *Proceedings of the 2018 World Wide Web Conference*, pages 439–448, 2018.

39 Anthony Wirth. Correlation clustering. In Claude Sammut and Geoffrey I. Webb, editors, *Encyclopedia of Machine Learning and Data Mining*, pages 280–284. Springer, 2017. doi:10.1007/978-1-4899-7687-1_176.