Online Correlation Clustering: Simultaneously Optimizing All ℓ_p-Norms

Sami Davies[*] Benjamin Moseley[†] Heather Newman[‡]

Abstract

The ℓ_p-norm objectives for correlation clustering present a fundamental trade-off between minimizing total disagreements (the ℓ_1-norm) and ensuring fairness to individual nodes (the ℓ_∞-norm). Surprisingly, in the offline setting it is possible to simultaneously approximate all ℓ_p-norms with a single clustering. Can this powerful guarantee be achieved in an online setting? This paper provides the first affirmative answer. We present a single algorithm for the online-with-a-sample (AOS) model that, given a small constant fraction of the input as a sample, produces one clustering that is *simultaneously* $O(\log^4 n)$-competitive for all ℓ_p-norms with high probability, $O(\log n)$-competitive for the ℓ_∞-norm with high probability, and $O(1)$-competitive for the ℓ_1-norm in expectation. This work successfully translates the offline "all-norms" guarantee to the online world.

Our setting is motivated by a new hardness result that demonstrates a fundamental separation between these objectives in the standard random-order (RO) online model. Namely, while the ℓ_1-norm is trivially $O(1)$-approximable in the RO model, we prove that any algorithm in the RO model for the fairness-promoting ℓ_∞-norm must have a competitive ratio of at least $\Omega(n^{1/3})$. This highlights the necessity of a different beyond-worst-case model. We complement our algorithm with lower bounds, showing our competitive ratios for the ℓ_1- and ℓ_∞- norms are nearly tight in the AOS model.

1 Introduction

Clustering is a fundamental task in unsupervised learning. In this paper, we study *correlation clustering*, where the goal is to partition a set of n items based on pairwise similarity $(+)$ and dissimilarity $(-)$ labels. Given a *complete* graph where each edge has a label of positive or negative, a clustering is a partition of the vertices. A positive edge is a *disagreement* if its endpoints are in different clusters, and a negative edge is a disagreement if its endpoints are in the same cluster. The goal is to find a partition that minimizes some function of these disagreements.

The choice of objective function is critical. The most well-studied objective, the ℓ_1-*norm*, minimizes the total number of disagreements. However, ℓ_1-optimized solutions can be "unfair," leaving some nodes with a very large number of incident disagreements. To address this, the ℓ_∞-*norm* objective seeks to minimize the maximum number of disagreements incident to any single node, ensuring a worst-case fairness guarantee. While constant-factor approximation algorithms have been known for both norms in the offline setting since their introductions [4, 25], it was until recently unknown (even existentially) whether it is possible to always produce a clustering that simultaneously is constant factor approximate for the ℓ_1- and ℓ_∞- norms. A recent work [13] shows (perhaps surprisingly) that in the offline setting, there is an algorithm achieving this goal, and moreover its output is simultaneously constant factor approximate for *all* ℓ_p-norms for $1 \leqslant p \leqslant \infty$. In the literature, this is sometimes referred to as the *all-norms* objective.

[*]Department of EECS at UC Berkeley and RelationalAI.

[†]Tepper School of Business, Carnegie Mellon University. Benjamin Moseley is supported in part by a Google Research Award, NSF grants CCF-2121744 and CCF-1845146, and ONR Grant N000142212702.

[‡]Department of Computer Science, Vassar College.

This raises a natural and challenging question: Can a simultaneous, all-norms guarantee be achieved in an online setting, where nodes arrive sequentially and must be irrevocably assigned to a cluster?

The lower bounds for correlation clustering are pessimistic in the standard online model where nodes arrive one at a time in adversarial order—a simple construction due to Mathieu, Sankur, and Schudy shows it is impossible to achieve any sublinear competitive ratio for any ℓ_p-norm objective while maintaining cluster consistency [24]. To circumvent this, prior works have explored relaxed models, but they have focused exclusively on the ℓ_1-norm. These models include allowing limited recourse (changing a node's cluster assignment) [9] or, as we study here, providing the algorithm with a small, random sample of the input upfront [22], where the remaining part of the instance then arrives online. This latter model, which we call the *online-with-a-sample* model (AOS[1] for short), is motivated by applications where historical data can inform decisions on new arrivals. The model has been used to study, for example, the Secretary problem [20], online bipartite matching [21], and online set cover [16]. In the AOS model, an $O(1/\varepsilon)$-competitive algorithm was shown by Lattanzi et al. for ℓ_1-norm correlation clustering, given an ε-fraction of the nodes as a sample [22]. However, the online landscape for other crucial objectives, like the fairness-promoting ℓ_∞-norm, has remained entirely unexplored. In fact, up until recently, all approximation algorithms for the ℓ_∞-norm (and in fact for any ℓ_p-norms with $p > 1$) were based on rounding the solution to a convex program, making them less amenable to the online setting than, say, combinatorial algorithms.

We are interested in simultaneously approximating all ℓ_p-norms online. To motivate our model, we show that the ℓ_∞-norm exhibits fundamentally different characteristics online from the ℓ_1-norm. This contrast is clearest in the *random-order (RO) model*, another popular online relaxation. For the ℓ_1-norm, achieving an $O(1)$-approximation in the RO model is straightforward by simulating the classic Pivot algorithm [1]. Thus, the RO model and the AOS model (as shown in [22]) both admit $O(1)$-competitive algorithms for ℓ_1-norm correlation clustering. This is in line with recent work by Gupta, Kehne, and Levin [16], which shows that for a certain class of minimization problems, the AOS model can be reduced to the RO model with loss $1/\varepsilon$ in the competitive ratio, for ε the fraction of input given as sample in the AOS model. In contrast, we establish that for the ℓ_∞-norm, no algorithm in the RO model can be better than $\Omega(n^{1/3})$-competitive. This hardness separation highlights a crucial limitation of the RO model. Further, this provides strong motivation for using the AOS model given the interest in the ℓ_∞-norm, as well as in simultaneously optimizing all ℓ_p-norms.

In all, we are motivated by the following questions: Can a small sample provide enough structural information to break the online hardness barrier for the ℓ_∞-norm? Further, in the AOS model, is it possible to approximate the all-norms objective online?

Our Contributions

This paper provides the first non-trivial results for ℓ_p-norm correlation clustering in a semi-online setting and, in doing so, translates the offline all-norms guarantee into the online world. Our main contribution is a single algorithm in the AOS model that simultaneously achieves strong approximation guarantees for all ℓ_p-norms. Let OPT_p denote the cost of an optimal clustering for the ℓ_p-norm objective. We assume in the AOS model that an adversary fixes the online input. However, we see upfront a random sample of size εn, where n is the number of nodes in the instance and each node is sampled uniformly and independently with probability ε. The parameter $0 < \varepsilon < 1$ can be anything in our upper bound results (Theorem 1), and as small as $n^{-1/4}$ in our lower bound results (Theorem 2).

A single algorithm for all ℓ_p-norms. We present a deterministic online algorithm that, given the sample, produces one clustering that is simultaneously a good approximation for all ℓ_p-norms.

Theorem 1. *Given $0 < \varepsilon < 1$, there is a **single** algorithm in the AOS model that produces a clustering with cost:*

1. $O\left(\frac{1}{\varepsilon^8} \cdot \log^4 n\right) \cdot \mathsf{OPT}_p$ *for all $1 \leqslant p < \infty$ with probability at least $1 - 1/n$.*
2. $O\left(\frac{1}{\varepsilon^6} \cdot \log n\right) \cdot \mathsf{OPT}_\infty$ *for the ℓ_∞-norm with probability at least $1 - 1/n$.*

[1] We note the abbreviation AOS stands for adversarial-order model with a sample.

3. $O\left(\frac{1}{\varepsilon^6}\right) \cdot \mathsf{OPT}_1$ *in expectation for the* ℓ_1*-norm.*

This result demonstrates that balancing the global quality of the ℓ_1-norm and the local fairness of the ℓ_∞-norm is achievable online, given a small sample of the offline underlying input.

Near-optimal lower bounds. We show that the competitive ratios of our algorithm are nearly optimal for the ℓ_1- and ℓ_∞- norms. In particular, the logarithmic factor for the ℓ_∞-norm and a dependence on $1/\varepsilon$ for the ℓ_1- and ℓ_∞- norms are necessary. While the ℓ_1-norm lower bound is known [22], our contribution is the new, and more involved, lower bound for the ℓ_∞-norm.

Theorem 2. *For any* $\frac{1}{n^{1/4}} \leqslant \varepsilon \leqslant 3/4$, *any* randomized *algorithm in the AOS model has a worst-case expected cost of at least* $\Omega\left(\frac{1}{\varepsilon} \cdot \log n\right) \cdot \mathsf{OPT}_\infty$ *for the* ℓ_∞*-norm and* $\Omega\left(\frac{1}{\varepsilon}\right) \cdot \mathsf{OPT}_1$ *for the* ℓ_1*-norm.*

Hardness of the random-order (RO) model. Finally, we prove the hardness of the ℓ_∞-norm in the RO model, justifying our focus on the AOS model.

Theorem 3. *Any* randomized *algorithm for* ℓ_∞*-norm correlation clustering in the random-order model has a competitive ratio of at least* $\Omega(n^{1/3})$.

This result establishes a fundamental separation between the ℓ_1- and ℓ_∞- norm objectives in the RO model, where an $O(1)$-competitive ratio for ℓ_1-norm trivially follows from the Pivot algorithm. It underscores that even with random arrival order, achieving fairness online is hard without giving the online algorithm any additional power. Recent work [16] shows that for a general class of minimization problems (called *augmentable integer programs*), there is a reduction from the RO model to the AOS model with loss of a factor at most $1/\varepsilon$ in the competitive ratio, i.e., if there is an algorithm with competitive ratio Δ in the random-order model, then there is an algorithm with competitive ratio Δ/ε in the AOS model. In contrast, ℓ_p-norm correlation clustering is an example of a minimization problem where the AOS model is actually *much stronger* at breaking through worst-case instances than the RO model.

1.1 Related work

Prior work offline. Bansal, Blum, and Chawla [4] proposed correlation clustering for the goal of minimizing the ℓ_1-norm of the disagreement vector. The problem is NP-hard, and numerous approximation algorithms have been developed [1, 8, 10, 6]. A 1.437-approximation is known for the ℓ_1-norm [6], which improves upon the work that beat the threshold of 2 [10]. Puleo and Milenkovic [25] proposed the ℓ_p-norm objective for $p > 1$ and for each fixed p they gave a 48-approximation. This factor has since been improved in a series of works, first to 7 [7], and then to 5 [19]. Notably, this entire line of work on ℓ_p-norm objectives relies on rounding solutions to convex programs.

Davies, Moseley, and Newman [11] introduced the first *combinatorial* $O(1)$-approximation algorithm for the ℓ_∞-norm. Heidrich, Irmai, and Andres [17] built off of the techniques in [11] to prove a combinatorial 4-approximation for the ℓ_∞-norm. Then, Davies, Moseley, and Newman [13] offered a new combinatorial algorithm proving there exists a single clustering that is an $O(1)$-approximation for all ℓ_p-norms simultaneously (also known as the all-norms objective). Cao, Li, and Ye [5] modified their algorithm in order to improve the factor for the all-norms objective, as well as to run in near-linear time in the MPC model in polylogarithmic rounds.

Prior work online. In the online setting, nodes arrive over time and reveal the signs of all of their edges to nodes that have previously arrived. Upon arrival of a node, an algorithm must irrevocably assign the node to a cluster. In the popular *competitive analysis* framework, the algorithm's clustering cost is compared to the best optimal offline algorithm that is aware of the entire instance in advance. Recall that no constant-competitive algorithm exists in the purely online setting for any ℓ_p-norm objective of correlation clustering [24]. There has been much work in clustering in the online and streaming settings. For the popular k-clustering problem (which includes k-median, k-means, and k-center), which likewise face pessimistic lower bounds in the online setting, a popular remedy of choice is recourse [23, 18, 14, 15]. Likewise, to the best of our knowledge, the only previous beyond-worst-case results on correlation clustering in the online setting, besides the AOS model [22], allow recourse [9, 3], and these only consider the ℓ_1-norm objective.

Prior work in the online-with-a-sample model. The AOS model was initiated by Kaplan, Naori, and Raz [20] in the context of the secretary problem.[2] Lattanzi et al. [22] then used the AOS model for ℓ_1-norm correlation clustering. They showed that the classic Pivot algorithm [1] can be modified to be seeded with an offline sample of size εn, thus leading to an $O(1/\varepsilon)$-competitive algorithm, and this guarantee matches the lower bound up to constant factors. Since the works of Kaplan, Naori, and Raz [20] and Lattanzi et. al. [22] in the AOS model, the model has been applied to Steiner tree, load balancing, and facility location [2]; bipartite matching [21]; and set cover [16]. A similar "semi-online" setting also appears in [26] for bipartite matching; like the AOS model, the semi-online model described there also contains predicted and adversarial parts of the input, but the predicted part is not necessarily a random sample, and further, the parts may be interleaved in an arbitrary manner.

1.2 Technical overview

Our primary challenge is to adapt an offline algorithm that requires complete, global knowledge of the graph into an online setting with limited information. The offline algorithm for all-norms correlation clustering due to Davies, Moseley and Newman [13] relies on two offline-only steps. Step (1) is to compute a semi-metric d^* over all vertex pairs; d^* is an (almost) feasible solution to the canonical convex relaxation for the problem, but can be computed using explicit combinatorial properties of the graph. Step (2) is to, in place of an optimal solution to the relaxation, feed d^* into the convex program rounding algorithm by Kalhan, Makarychev, and Zhou [19] (from now on, the *KMZ algorithm*). This is a ball-cutting procedure that iteratively cuts out clusters based on "suggestions" from d^* (i.e., if d^*_{uv} is small, then nodes u and v "want" to be clustered together). As in Step (1), the whole graph is required to determine the order in which clusters are cut out. So, both steps require significant changes in order to be adapted to the online setting.

To replace step (1), we compute a semi-metric $\tilde{d}$, only using the sample S, as a proxy for d^*. This will imply that when a node v arrives online, its distances $\tilde{d}_{uv}$ can immediately be computed for all u that have already arrived. To replace step (2), we note that the offline all-norms result holds (up to constants) when d^* is fed into *any* constant-approximate convex program rounding algorithm. We adapt the offline rounding algorithm of Charikar, Gupta, and Schwartz [7] (from now on, the *CGS algorithm*) instead of the KMZ algorithm, as we find the former easier to adapt to our online setting. The CGS algorithm is as follows: choose the unclustered node v that has the most unclustered vertices in the ball of radius r around it, then let v and all unclustered nodes in its ball of radius $3r$ form a new cluster (where the balls are w.r.t the semi-metric that is the solution to the convex program). Notably, the algorithm is dynamic in that the sizes of the balls changes at each iteration, because nodes are removed when they are clustered.

Estimating a semi-metric via sampling. Prior to the works of Davies, Moseley, and Newman [11, 13], *all* algorithms for ℓ_p-norm correlation clustering with $p > 1$ relied on solving the canonical convex program relaxation, and feeding the optimal semi-metric solution x^* (for the chosen ℓ_p-norm objective) to a rounding algorithm (e.g., the KMZ or CGS algorithms). Arguably, these techniques are less amenable to the online setting, because solving the convex program for an optimal solution is a black box. Moreover, for our dual goal of finding a solution that is good for *all* ℓ_p-norms, these techniques do not work, as they require specifying the ℓ_p-norm objective over which to optimize.

Our key insight is that the semi-metric d^* in [13] – which, crucially, is defined using explicit combinatorial properties of the graph – can be approximated by computing it only on $G[S]$, the graph we see upfront on the random sample S. We first show this estimate, $\tilde{d}$, is of high quality (Section 4). Analyzing $\tilde{d}$ is non-trivial, as $\tilde{d}$ is defined based on non-linear calculations as well as on thresholds, both of which are highly sensitive to error. More specifically, in the offline world, the term $|N_u^+ \cap N_v^+|/|N_u^+ \cup N_v^+|$ (where N_u^+ is the set of positive neighbors of u) is crucial for computing d^*_{uv}. So, in computing $\tilde{d}$, we must study an estimate of $|N_u^+ \cap N_v^+|/|N_u^+ \cup N_v^+|$, which is a quotient of correlated random variables, and is prone to high variance and bias, especially for nodes with small positive neighborhoods.

[2]The analysis, however, differs from that presented here, in that we define competitiveness on the whole instance, whereas they restrict to the online portion of the input.

Further, in the offline setting, the definition of d^* involves rounding based on whether the values $|N_u^+ \cap N_v^+|/|N_u^+ \cup N_v^+|$ lie above or below certain thresholds. When applied now to an estimate of this quotient for the online setting, small estimation errors can trigger large changes in the semi-metric. This means we *cannot* compute pointwise bounds, e.g., of the form $\mathbb{E}[\tilde{d}_{uv}] \approx d^*_{uv}$. Instead, our analysis requires a deep dive into the probabilistic events that cause large errors and a novel charging scheme to bound the impact of these unavoidable estimation failures.

We pause to note that our technique, of using the sample S to estimate combinatorial quantities of interest, is quite different from that for ℓ_1-norm correlation clustering, which recall was previously studied in the AOS model [22]. That work adapts the Pivot algorithm, a 3-competitive algorithm in the RO model. In some sense, the Pivot algorithm is more readily adaptable to the AOS setting, since it is already an online algorithm, and the distribution of the (first ε-fraction of) input matches that of the random sample. Our present work provides an example of how the AOS sample can be useful in other situations. In particular, we believe this general strategy of estimating quantities for an algorithm, and then charging the cost of an algorithm's objective to these estimates, is of broader interest for other problems in this model.

Preprocessing the sample. As discussed above, the sample S is used to estimate d^* with a proxy metric $\tilde{d}$. It is also used to adapt the CGS algorithm. Specifically, we use S to select *centers* – such centers v are used to cut out clusters in the ball-cutting procedure. We also use S to determine the order in which clusters are cut out around these centers. In the CGS algorithm, these are cut in decreasing order of $|\text{Ball}_{x*}(v, r)|$. We instead can only estimate the analogous quantity, $|\text{Ball}_{\tilde{d}}(v, r)|$, so we do so using S. Importantly, to avoid correlational issues, we show how to simulate four independent subsamples on S. Then, we estimate different random variables of interest using different subsamples, rather than on the whole common sample S. The subsamples are seemingly essential for making the analysis tractable.

Feeding the proxy metric into a static, online version of the CGS algorithm. The algorithm has two phases. The first is an online version of the offline ball-cutting CGS algorithm. Vertices clustered in this phase are said to be *pre-clustered*. The second phase handles the remaining vertices.

Pre-clustering phase: When a vertex v arrives, we first determine if the sample S is trustworthy for v. We check whether the distances $\tilde{d}$ are sufficiently accurate for edges incident to v. If so, we then check if v is close to one of the pre-selected centers.

- If the checks pass, v is assigned to a center that it is close to, particularly the earliest in the ordering of centers (see above). We note that to take the CGS algorithm online, this ordering is static, meaning it is pre-computed before the vertices arrive, unlike in the offline CGS algorithm, which dynamically orders the centers using the unclustered vertices at each iteration.
- If the test fails, the algorithm falls back to one of two versions of the Pivot algorithm.

(Modified) Pivot phases: We perform the classic Pivot algorithm of [1] on the vertices that fail the first check. We perform a modified version of the Pivot algorithm on the vertices that only fail the second check. While classic Pivot only takes into account the signs of edges, our modified version needs to take into account the distances $\tilde{d}$. We note that, even for the classic Pivot subroutine, the analysis looks different from that of classic Pivot. This is because classic Pivot has previously only been used for the ℓ_1-norm objective, and further assumes random order, rather than the adversarial order we have here.

1.3 Organization

In Section 2, we discuss the online-with-a-sample model and define the correlation metric and adjusted correlation metric. In Section 3, we define Algorithm 1, whose output satisfies Theorem 1. We prove item 1 of Theorem 1 (the statement for all finite p) in Sections 4 and 5, though a few proofs of lemmas and constructions stated in these sections are deferred to Appendix A. Then we prove item 3 of Theorem 1 (the statement for $p = 1$) in Appendix B, and item 2 of Theorem 1 (the statement for $p = \infty$) in Appendix C. Lastly, the lower bound result, Theorem 2, is in Appendix D.

2 Preliminaries

Let $G = (V, E)$ be a complete graph, where E is partitioned into positive edges (E^+) and negative edges (E^-). Let N_u^+ and N_u^- denote the positive and negative neighborhoods, respectively, of vertex u. That is, $N_u^+ = \{v \in V : uv \in E^+\}$ and $N_u^- = \{v \in V : uv \in E^-\}$. For convenience, assume that each vertex has a positive self-loop, i.e., for all $u \in V$, $u \in N_u^+$.

Recall that OPT_p is the optimal objective value of an integral solution for the ℓ_p-norm objective, where here $p \in [1, \infty]$.

We set some parameters. Let $\delta = 10/7$, and define $c = c(\delta) := 2\delta^2 + \delta = 270/49$ and $r = r(\delta) := \frac{1}{2c\delta^2} = \frac{2401}{54000}$.

2.1 The online-with-a-sample model

In the *online-with-a-sample* model, an adversary fixes the online input, then we are given a sample S, where each element of the universe V is in S independently with probability $\varepsilon > 0$. We assume ε is known[3]. So for our problem, we see the induced subgraph on S a priori.

We will estimate various quantities using the sample S, and, for the analysis to be tractable, these estimations should not be correlated with each other. To this end, we show how to "split" the sample S into four independent subsamples S_p, S_d, S_b, S_r (where vertices in S may be in more than one subsample). We define this notion of independence more precisely in Appendix A.1, but it will suffice to think of these subsamples as having the same joint distribution as four random subsets of vertices, each obtained by taking a uniformly random sample of (expected) size $\Theta(\varepsilon^2 n)$, and repeating this procedure *independently* four times.[4]

We defer the construction of the four independent subsamples S_p, S_d, S_b, S_r, and the precise notions of independence they satisfy, to Lemma 13 and Corollary 2 in Appendix A.1.

Definition 1. *We call S_d the* distance sample*, S_p the* pre-clustering sample*, S_b the* counting sample*, and S_r the* rounding sample*. We call the vertices in S_p* centers*.*

Since the four samples are constructed from S, edges in the (complete) subgraph induced by $S_d \cup S_p \cup S_b \cup S_r$ are known to the online algorithm a priori. Then, the remaining vertices arrive one by one in adversarial order. When a vertex arrives, it reveals the signs of its incident edges to all vertices that have previously arrived (including all of S), and it must be irrevocably assigned a cluster.

We define $q(\varepsilon) := \mathbb{P}[v \in S_i] = \varepsilon^2/2$ for any subsample $S_i \in \{S_d, S_p, S_b, S_r\}$, per the construction in Lemma 13.

2.2 Correlation metric, adjusted correlation metric, and their estimates

The *correlation metric* is a near-optimal feasible solution to the canonical convex program[5] for ℓ_p-norm correlation clustering. The correlation metric d is feasible for this convex program, meaning here that it satisfies the triangle inequality. So we may think of d_{uv} as specifying a distance between u and v, where the smaller the distance, the more that u and v would like to be clustered together. Relevant theorems on the (adjusted) correlation metric from [11, 13] are in Section 4.

Below, we generalize the definition of the correlation metric by defining it on a subgraph of G. Taking $U = V$ below recovers the correlation metric in [11].

Definition 2 (Correlation metric)**.** *Let $G = (V, E)$ be a complete, signed graph, and let $U \subseteq V$. For every (unordered) pair $u, v \in V$, define the* correlation metric on U*, denoted $d^U : V \times V \to [0, 1]$, by*

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

[3]One can immediately remove this assumption if we consider an "online-with-samples" model, suggested in the next footnote.

[4]Alternatively, one could consider a type of "online-with-samples" model, where an algorithm is given access to a constant number k of independent random samples of size $\varepsilon_i \cdot n$ of V, where $\sum_{i \in [k]} \varepsilon_i = \varepsilon$. We find this to be a perfectly reasonable model, and the reader may find it simpler to assume this model.

[5]See [13] for more details on this convex program, which, while motivating, is not necessary for understanding the work herein.

To make this well-defined, we take $d^U_{uv} = 1$ if $(N^+_u \cup N^+_v) \cap U = \varnothing$ for $u \neq v$, and $d^U_{uu} = 0$ always. When $U = V$, we simply write d for d^V and refer to d as the correlation metric.

The correlation metric distances are near-optimal for the convex program with the ℓ_∞-norm objective (see Theorem 4), but the metric must be adjusted in order to be simultaneously near-optimal for all ℓ_p-norms. We likewise generalize the definition of the so-called *adjusted correlation metric* in [13]; taking $U = W = V$ recovers their definition.

Definition 3 (Adjusted correlation metric)**.** *Let $G = (V, E)$ be a complete, signed graph, and let $U, W \subseteq V$. Let d^U be the correlation metric on U as in Definition 2. Compute the* adjusted correlation metric on U and W*, denoted $d^{U,W} : E \to [0, 1]$, as follows:*

- *If $uv \in E^-$ and $d^U_{uv} > 7/10$, set $d^{U,W}_{uv} = 1$. (We say uv is* rounded up*.)*
- *For $u \in V$ such that $|\{v \in N^-_u : d^U_{uv} \leqslant 7/10\} \cap W| \geqslant 10/3 \cdot |N^+_u \cap U|$, set $d^{U,W}_{uv} = 1$ for all $v \in V \backslash \{u\}$. (We say u is* isolated *by $d^{U,W}$.)*

When $U = W = V$, we write d^* for $d^{V,V}$ and refer to d^* as the adjusted correlation metric.

A key takeaway is that the correlation metric and the adjusted correlation metric, while intended as surrogates for *non-combinatorial* optimal solutions to a convex program, are based solely on *combinatorial* properties of the graph – thus making them more amenable to the online setting. In the online-with-a-sample model, we obviously cannot exactly compute the correlation metric or the adjusted correlation metric as we go, as the full positive neighborhood of a vertex is not necessarily known upon its arrival. But, for any $u, v \in V$, we *can* compute, e.g., d^U_{uv} in the case that U is a subset of S, *as soon as u, v have arrived*, since S is known to the algorithm upfront! The hope is that d^U and $d^{U,W}$ should be good approximations of $d = d^V$ and $d^* = d^{V,V}$, respectively, since S is a random sample of V – while also being usable online.

To estimate the adjusted correlation metric, we proceed in two steps. First, we estimate the correlation metric using the distance sample S_d; call this $\bar{d}$. Then, using the rounding sample S_r, we round $\bar{d}$ to estimate the adjusted correlation metric; call this $\tilde{d}$.

Definition 4 (Estimated correlation metric)**.** *For every (unordered) pair $u, v \in V$, define the* estimated correlation metric $\bar{d} : V \times V \to [0, 1]$ *by $\bar{d}_{uv} := d^{S_d}_{uv}$.*

Definition 5 (Estimated adjusted correlation metric)**.** *Let $\bar{d}$ be the estimated correlation metric as in Definition 4. Define the* estimated adjusted correlation metric $\tilde{d} : E \to [0, 1]$ *by $\tilde{d}_{uv} := d^{S_d,S_r}_{uv}$. Moreover, we let R_1 be the (random) subset of V for which bullet 2 of Definition 3 applies (i.e., the set of vertices that are isolated by $\tilde{d}$), $R_2 = V \backslash R_1$, and $R_1(u) := \{v \in N^-_u : \bar{d}_{uv} \leqslant 7/10\}$.*

The next observation ensures our algorithm for the AOS model is an online algorithm.

Observation 1. *As soon as both u and v have arrived (including if one or both is in S), $\bar{d}_{uv}$ and $\tilde{d}_{uv}$ can be computed.*

We see that if u has no positive neighbors in the sample S_d, then $\tilde{d}$ isolates u from all other vertices.

Fact 1. *Fix $u \in V$ such that $N^+_u \cap S_d = \varnothing$. Then $\tilde{d}_{uv} = 1$ for all $v \neq u$.*

Both $\bar{d}$ and $\tilde{d}$ enjoy similar properties to d and d^*, respectively, in that they are a semi-metric and near semi-metric, respectively. Thus, we can still view $\bar{d}$ and $\tilde{d}$ as specifying distances between vertices.

Definition 6. *We say a symmetric function $f : V \times V \to \mathbb{R}_+$ is a δ-*semi-metric *if $f_{uv} \leqslant \delta \cdot (f_{uw} + f_{wv})$ for all $u, v, w \in V$ (along with the usual requirement that $f_{uu} = 0$). We say in this case that f satisfies an* approximate triangle inequality*, or f is a* near semi-metric.

Lemma 1. *Let $\bar{d}$ and $\tilde{d}$ be as in Definitions 4 and 5, respectively. Then*

- *$\bar{d} : V \times V \to [0, 1]$ is a 1-semi-metric, that is, $\bar{d}$ satisfies the triangle inequality.*
- *$\tilde{d} : V \times V \to [0, 1]$ is a $\frac{10}{7}$-semi-metric.*

Algorithm 1 Main Algorithm

Input: $G = (V, E)$, S_d, $\tilde{d}$, *ordered* $S_p = \{u_1, \ldots, u_{|S_p|}\}$, radius r, constant c
Initialize: empty clusters $\mathcal{C}_{\mathsf{ALG}} = \{C_1, \ldots, C_{|S_p|}\}$, sets $V_0 = V$ and $V' = \varnothing$
for each arriving $v \in V$ **do**
  // Pre-clustering phase
  if $|N_v^+ \cap S_d| \neq \varnothing$ and there exists $u_i \in S_p$ such that $\tilde{d}_{u_i v} \leqslant c \cdot r$ **then**
    Let i^* be the earliest in the ordering among all such u_i
    Add v to cluster C_{i*}
  // Pivot phase
  else
    Add v to V'
    if v has $|N_v^+ \cap S_d| = \varnothing$ **then**
      Remove v from V_0, order $\overline{V}_0 = V \backslash V_0$ by arrival order
      Set $E_c := E^+(G[\overline{V}_0])$
      Add v to $\mathcal{C}_{\mathsf{ALG}}$ by running $\mathsf{ModifiedPivot}(G[\overline{V}_0], \mathcal{C}_{\mathsf{ALG}}, \overline{V}_0 \backslash \{v\}, E_c)$
    else
      Order $V' \backslash \overline{V}_0$ by arrival order
      Set $E_c := E^+(G[V' \backslash \overline{V}_0]) \cap \{uv \in E : \tilde{d}_{uv} < c \cdot r\}$
      Add v to $\mathcal{C}_{\mathsf{ALG}}$ by running $\mathsf{ModifiedPivot}(G[V' \backslash \overline{V}_0], \mathcal{C}_{\mathsf{ALG}}, V' \backslash (\overline{V}_0 \cup \{v\}), E_c)$
Output: Clustering $\mathcal{C}_{\mathsf{ALG}}$

2.3 Ordering the centers

Given a map $f : V \times V \to [0, 1]$ on the vertices of G, for $c \in V$, $U \subseteq V$, and $\rho \geqslant 0$ we define: $\mathrm{Ball}_f(c, \rho) := \{v \in V : f_{cv} \leqslant \rho\}$ and $\mathrm{Ball}_f^U(c, \rho) := \{v \in U : f_{cv} \leqslant \rho\}$.

Definition 7 (Density)**.** *Given a semi-metric $f : V \times V \to [0, 1]$ and vertex $c \in V$, define $|Ball_f(c, r)|$ to be the* density *of c w.r.t f and r.*

A subroutine of our algorithm will be an adaptation of the CGS algorithm (see Section 1.2). This algorithm takes as input a metric f on the vertices and orders the vertices in decreasing order of their densities with respect to f and some radius r. Due to our online setting, we will only be able to estimate these densities, which we do using the counting sample S_b.

We note that $|\mathrm{Ball}_{\tilde{d}}^{S_b}(c, r)|$ depends on the randomness of S_b, S_d, and S_r. The following observation will ensure our algorithm is well-defined.

Observation 2. *For any center $c \in S_p$, the density $|Ball_{\tilde{d}}^{S_b}(c, r)|$ can be computed using only the information given* a priori *in the AOS model, i.e., the sample S.*

Lastly, we order the centers S_p based on their estimated densities.

Definition 8 (Ordered center sample)**.** *Let $S_p = \{u_1, \ldots, u_{|S_p|}\}$. We assume the u_i are labeled so that $|Ball_{\tilde{d}}^{S_b}(u_1, r)| \geqslant \cdots \geqslant |Ball_{\tilde{d}}^{S_b}(u_{|S_p|}, r)|$ (with ties broken arbitrarily).*

3 Algorithm Description

Now we are ready to describe our main algorithm (Algorithm 1) for the AOS model. Note that the for loop of Algorithm 1 considers the vertices in S in arbitrary order. So, we assume the algorithm considers the vertices in S first, and then the vertices in $V \backslash S$ as they arrive online.

Our main algorithm is an online version of the offline CGS algorithm [7] (see the first two paragraphs in Section 1.2). The CGS algorithm must be adapted in several important ways in order to be taken online. First, the CGS algorithm takes as input an optimal solution to the convex program for the ℓ_p-norm objective of interest. Since we are not given the graph upfront, this is impossible to compute on the fly. Moreover, the convex program requires specifying an ℓ_p-norm objective, whereas we would like to optimize for all ℓ_p-norms simultaneously. In place of this optimal solution, we use $\tilde{d}$, which *can* be computed on the fly (Observation 1), and does *not* depend on p.

Algorithm 2 ModifiedPivot($H, \mathcal{C}_{\mathsf{ALG}}, P, E_c$)

Input: $H = (V_H, E_H)$ where V_H is *ordered*, $\mathcal{C}_{\mathsf{ALG}}$ is a clustering on $P \subseteq V_H$, and $E_c \subseteq E_H$ is a set of *clusterable* edges
for each $v_i \in V_H \backslash P$ (in order) **do**
  if there exists v_j that is before v_i in the ordering with $v_j v_i \in E_c$ **then**
    Let j^* be the earliest in the ordering among all such v_j
    Add v_i to the same cluster as v_{j^*}
  else
    Open a new cluster and add v_i to it
Output: Clustering $\mathcal{C}_{\mathsf{ALG}}$

The CGS algorithm requires an ordering on *all* of V based on the ball densities around the vertices. We are only able to estimate these densities for vertices in the offline sample, so we use the ordered sample of centers S_p (Definition 8), and $V \backslash S_p$ remains unordered. In the offline CGS algorithm, each vertex in V is clustered by the earliest vertex in the ordering that is nearby (thus by itself if necessary). In our setting, we cannot guarantee every vertex will be clustered because only S_p is ordered. Further, $\tilde{d}_{uv}$ is meaningless as a distance when u or v does not have positive neighbors in S_d (see Definition 2). Thus, we will not cluster every vertex in this way, and instead throw the vertices that are *not* "pre-clustered" (for either of these two reasons) to a subroutine called ModifiedPivot (Algorithm 2).

ModifiedPivot is a generalization of the Pivot algorithm from [1]. Taking $E_c = E^+$ in Algorithm 2 (as we do in the first call to ModifiedPivot in Algorithm 1) recovers the original Pivot algorithm. In the second call to ModifiedPivot in Algorithm 1, we further restrict E_c to edges that are "short" according to $\tilde{d}$; this is not simply an optimization, but seemingly needed in the analysis.

Terminology

If the else statement in Algorithm 2 holds, we say v_i is a *pivot*, and that v_i's pivot is itself, v_i. If the if statement holds, we refer to v_{j^*} as v_i's *pivot*.

We say v is *pre-clustered* by u_i if v is added to C_i. Otherwise, $v \in V'$, and we say v is *unclustered* or *not pre-clustered*. Note that a vertex v may be unclustered for one of two reasons: either v has no positive neighbors in S_d ($v \notin V_0$ or as written in Algorithm 1, $v \in \overline{V}_0$, where $\overline{V}_0 = V \backslash V_0$), or $v \in V_0$ but v is not close to any vertex in S_p with respect to $\tilde{d}$. We call the vertices in V_0 *eligible for pre-clustering*, or simply *eligible*, and otherwise *ineligible*.

If v is pre-clustered by s, we denote s by $s^*(v)$. We may refer to $s^*(v)$ as v's *center*. We say u *is clustered after* v or v *is clustered before* u if either both u and v are pre-clustered, but $s^*(v)$ is before $s^*(u)$ (w.r.t the ordering of S_p), or if v is pre-clustered but u is not. Notationally, this will denoted as $v \succ u$.

The "good" event. For the cost analysis of Algorithm 1 when $p \neq 1$ (Sections 4, 5, and C), we condition on a certain *good event*, denoted B^c, that occurs with high probability (Lemma 15). For the formal definition of the good event, see Definition 12. Informally, the event B^c ensures via concentration that the size of sufficiently large sets of vertices can be well-estimated based on observing their intersection with the subsamples. It also ensures that for pairs of nodes u, v with large combined positive neighborhood, $\bar{d}_{uv}$ is a good estimate of d_{uv}.

4 Fractional cost of correlation metric

In the offline setting, the algorithm has access to the *true* correlation metric d and adjusted correlation metric d^*. We overview some notation and previous results that will be useful to us.

4.1 Preliminaries

Definition 9 (Fractional cost)**.** *For $u \in V$, the* fractional cost *of u w.r.t d (Definition 2) is $D(u) := \sum_{v \in N_u^+} d_{uv} + \sum_{v \in N_u^-}(1 - d_{uv})$, and $D = (D(u))_{u \in V}$ is the vector of fractional costs of vertices. $D^*(u)$ and D^* are defined analogously using d^* (Definition 3).*

Similarly, $\bar{D}(u)$ and $\bar{D}$ are defined analogously using $\bar{d}$ (Definition 4), and $\tilde{D}(u)$ and $\tilde{D}$ using $\tilde{d}$ (Definition 5), though these are estimated fractional cost *w.r.t the semi-metrics.*

The key result of Davies, Moseley, and Newman [11] is that the fractional cost of d in the ℓ_∞-norm is a constant approximation to the optimal solution for the ℓ_∞-norm. We will black-box this result in part of our analysis.

Theorem 4 (Lemma 4.2 in [11])**.** *Let D be as in Definition 9. Then $||D||_\infty \leqslant 8 \cdot \mathsf{OPT}_\infty$.*

While the correlation metric must be modified to give bounded fractional cost for other ℓ_p-norms, the following claim states that the correlation metric actually has bounded frational cost for the ℓ_1-norm objective *when restricted to positive edges*. This will be useful in our later analysis.

Lemma 2 (Lemma 8, Claim 1 in [12])**.** $\sum_{u \in V} \sum_{v \in N_u^+} d_{uv} \leqslant 3 \cdot \mathsf{OPT}_1$.

In the follow-up work [13], the authors show that taking $x = d^*$ gives bounded fractional cost for *all* ℓ_p-norms simultaneously (and thus a *simultaneous* approximation for all ℓ_p-norms). While we are not able to blackbox this result often, it does help in bounding the cost of disagreements in some settings.

Theorem 5 (Lemma 4 in [13])**.** *For $D^* = (D^*(u))_{u \in V}$ the fractional cost vector of the adjusted correlation metric d^*, there exists a universal constant M (independent of p) such that, for all $p \in [1, \infty]$,*

$$||D^*||_p \leqslant M \cdot \mathsf{OPT}_p.$$

4.2 Fractional cost of correlation metric for finite p

Recall that V_0 is the random set consisting of all vertices $u \in V$ with $N_u^+ \cap S_d = \varnothing$.

In this subsection, we show that the fractional cost of the estimated adjusted correlation metric $\tilde{d}$ in the ℓ_p-norm, *restricted to* V_0, is bounded in expectation against OPT_1 for $p = 1$, and bounded with high probability against OPT_p for finite p. This will allow us to charge the cost of some disagreements incurred by Algorithm 1 to $\tilde{d}$; other disagreements will require different surrogates for optimal.

For $p > 1$, we condition on the good event B^c (see Section 3).

Lemma 3. *Let $1 \leqslant p < \infty$. The estimated adjusted correlation metric $\tilde{d}$ satisfies the following:*

- $\mathbb{E}\left[\sum_{u \in V_0} \sum_{v \in N_u^+ \cap V_0} \tilde{d}_{uv}\right] \leqslant O(1/\varepsilon^4) \cdot \mathsf{OPT}_1$*, and*
- *Conditioned on the event B^c, $\sum_{u \in V_0} \left(\sum_{v \in N_u^+ \cap V_0} \tilde{d}_{uv}\right)^p \leqslant O\left(\left(1/\varepsilon^6 \cdot \log^3 n\right)^p\right) \cdot \mathsf{OPT}_p^p$.*

Lemma 4. *Let $1 \leqslant p < \infty$. The estimated adjusted correlation metric $\tilde{d}$ satisfies the following:*

- $\mathbb{E}\left[\sum_{u \in V_0} \sum_{v \in N_u^- \cap V_0} (1 - \tilde{d}_{uv})\right] \leqslant O(1/\varepsilon^2) \cdot \mathsf{OPT}_1$*, and*
- *Conditioned on the event B^c, $\sum_{u \in V_0} \left(\sum_{v \in N_u^- \cap V_0} (1 - \tilde{d}_{uv})\right)^p \leqslant O\left((1/\varepsilon^2 \cdot \log n)^p\right) \cdot \mathsf{OPT}_p^p$.*

For $u \in V_0$, let $\tilde{D}_0(u)$ be the fractional cost with respect to $\tilde{d}$ of the edges incident to u in the subgraph induced by V_0, i.e., $\tilde{D}_0(u) := \sum_{v \in N_u^+ \cap V_0} \tilde{d}_{uv} + \sum_{v \in N_u^- \cap V_0} (1 - \tilde{d}_{uv})$. By Lemmas 3 and 4, we obtain the following corollary.

Corollary 1. *Let $1 \leqslant p < \infty$. The estimated adjusted correlation metric $\tilde{d}$ satisfies the following:*

- $\mathbb{E}[\sum_{u \in V_0} \tilde{D}_0(u)] \leqslant O(1/\varepsilon^4) \cdot \mathsf{OPT}_1$*, and*
- *Conditioned on the event B^c, we have $\sum_{u \in V_0} \left(\tilde{D}_0(u)\right)^p \leqslant O\left(\left(1/\varepsilon^6 \cdot \log^3 n\right)^p\right) \cdot \mathsf{OPT}_p^p$.*

We prove Lemma 3 in Appendix A.4 and Lemma 4 in this section, as both proofs are similar in flavor, but the proof of Lemma 4 serves us again in Section 5.1.2. We give a brief overview here of both proofs. We will not be able to prove a pointwise bound comparing $\mathbb{E}[\tilde{d}_{uv}]$ to, e.g., d^*_{uv}, but instead, we will have to unbox the offline proof from [13] that the fractional cost of the adjusted correlation metric is bounded (see Theorem 5 above). While the casing will be similar to in the offline case, there are nontrivial technical details, e.g., independence issues, taking expectations of sums defined over random sets, and taking expectations over quotients, to name a few, that must be handled delicately.

Recall that $q(\varepsilon) = \mathbb{P}[v \in S_i] = \varepsilon^2/2$ for any subsample $S_i \in \{S_d, S_p, S_b, S_r\}$. In the following three lemma, we show that, with some constant factor loss, we can distribute an expectation over a quotient for certain random variables that will be of interest.

Lemma 5. *Fix a subsample $S_* \in \{S_d, S_b, S_r, S_p\}$ and fix $\mathcal{B} \subseteq \{S_d, S_b, S_r, S_p\} \backslash \{S_*\}$. Let $A \subseteq V$ be a subset of vertices that may depend on the randomness of the sets in $\mathcal{B}$, but it does not depend on the randomness of any other subsamples. Define the random variable $Z = \mathbf{1}_{\{|A \cap S_*| \geqslant 1\}} \cdot \frac{1}{|A \cap S_*|}$, where Z is defined to be 0 if $A \cap S_* = \varnothing$. Then,*

$$\mathbb{E}[Z \mid \mathcal{B}] \leqslant O(1/\varepsilon^2) \cdot \frac{1}{|A|} \quad \textit{and} \quad \mathbb{E}[Z^2 \mid \mathcal{B}] \leqslant O(1/\varepsilon^4) \cdot \frac{1}{|A|^2},$$

where the expectation is taken only over the randomness of S_*.

Note in the above that if $\mathcal{B} = \varnothing$, then $\mathbb{E}[Z \mid \mathcal{B}]$ is a deterministic value, and otherwise $\mathbb{E}[Z \mid \mathcal{B}]$ is a random variable.

Proof of Lemma 5. As in the lemma statement, $Z = \mathbf{1}_{\{|A \cap S_*| \geqslant 1\}} \cdot \frac{1}{|A \cap S_*|}$, where Z is defined to be 0 if $A \cap S_* = \varnothing$. Define $Y := |A \cap S_*| = \sum_{w \in A} X_w$, where X_w is 1 if $w \in S_*$ and 0 otherwise. Let $\mu := \mathbb{E}[Y] = q(\varepsilon) \cdot |A|$, where the expectation is taken only over the randomness of S_*. Again taking the expectation only over the randomness of S_*, we have that

$$\begin{aligned}
\mathbb{E}[Z \mid \mathcal{B}] &= \mathbb{E}\Big[\frac{1}{|A \cap S_*|} \cdot \mathbf{1}_{\{|A \cap S_*| \geqslant 1\}} \mid \mathcal{B}\Big] \\
&= \sum_{\ell=1}^{|A|} \mathbb{P}[Y = \ell \mid \mathcal{B}] \cdot \frac{1}{\ell} = \sum_{\ell=1}^{|A|} \mathbb{P}[Y = \ell] \cdot \frac{1}{\ell} \qquad (1) \\
&\leqslant \mathbb{P}[Y \leqslant \mu/2] + \sum_{\ell=\lceil \mu/2 \rceil}^{|A|} \mathbb{P}[Y = \ell] \cdot \frac{1}{\ell} \\
&\leqslant e^{-\mu/12} + \frac{2}{\mu} \cdot \sum_{\ell=\lceil \mu/2 \rceil}^{|A|} \mathbb{P}[Y = \ell] \leqslant \frac{5}{\mu} + \frac{2}{\mu} = \frac{7}{q(\varepsilon) \cdot |A|} = \frac{14}{\varepsilon^2 \cdot |A|}.
\end{aligned}$$

In the second line, if $|A| = 0$, we treat the sum as 0. Further, the first equality in (1) follows from the fact that A does not depend on the randomness of S_*, so we may move the sum outside of the probability and then we use Corollary 2 for the next equality (specifically saying that $\mathbb{P}[|A \cap S_*| = \ell \mid \mathcal{B}] = \mathbb{P}[|A \cap S_*| = \ell]$). In the last line, we applied a Chernoff bound (Theorem 6). The bound for $\mathbb{E}[Z^2 \mid \mathcal{B}]$ is derived similarly: starting in line (1), replace $1/\ell$ with $1/\ell^2$, and continuing as before we obtain a bound of $e^{-\mu/12} + 4/\mu^2 \leqslant 83/\mu^2 = 332/\varepsilon^4 \cdot |A|^2$. ☐

Proposition 1. *Define $\mathbf{1}_{\{u \in V_0\}} \cdot \frac{|N_u^+|}{|N_u^+ \cap S_d|}$ to be 0 if $u \notin V_0$, i.e., if $|N_u^+ \cap S_d| = 0$. Likewise, define $\mathbf{1}_{\{|R_1(u) \cap S_r| \geqslant 1\}} \cdot \frac{|R_1(u)|}{|R_1(u) \cap S_r|}$ to be 0 if $|R_1(u) \cap S_r| = 0$. Then*

- $\mathbb{E}\left[\mathbf{1}_{\{u \in V_0\}} \cdot \frac{|N_u^+|}{|N_u^+ \cap S_d|}\right] \leqslant O(1/\varepsilon^2)$ *, and* $\mathbb{E}\left[\mathbf{1}_{\{u \in V_0\}} \cdot \frac{|N_u^+|^2}{|N_u^+ \cap S_d|^2}\right] \leqslant O(1/\varepsilon^4)$
- $\mathbb{E}\left[\mathbf{1}_{\{|R_1(u) \cap S_r| \geqslant 1\}} \cdot \frac{|R_1(u)|}{|R_1(u) \cap S_r|}\right] \leqslant O(1/\varepsilon^2)$
- *Conditioned on the event B^c, we have* $\mathbf{1}_{\{u \in V_0\}} \cdot \frac{|N_u^+|}{|N_u^+ \cap S_d|} \leqslant O(1/\varepsilon^2) \cdot \log n$

- *Conditioned on the event B^c, we have $\mathbf{1}_{\{|R_1(u)\cap S_r|\geqslant 1\}}\cdot\frac{|R_1(u)|}{|R_1(u)\cap S_r|}\leqslant O(1/\varepsilon^2)\cdot\log n$.*

Proof of Proposition 1. The bounds in the first bullet follow directly from Lemma 5, taking $\mathcal{B}=\varnothing$, $S_*=S_d$, and $A=N_u^+$. The bound in the second bullet likewise follows from Lemma 5, taking $\mathcal{B}=\{S_d\}$, $S_*=S_r$, and $A=R_1(u)$, and then applying Law of Total Expectation.

For the third bullet, since B^c holds, we know that if $|N_u^+|\geqslant C\cdot\log n/\varepsilon^2$, then $|N_u^+\cap S_d|\geqslant\frac{\varepsilon^2}{4}\cdot|N_u^+|$, so $\frac{|N_u^+|}{|N_u^+\cap S_d|}\leqslant\frac{4}{\varepsilon^2}$. Otherwise, we use that $u\in V_0$ to bound $\frac{|N_u^+|}{|N_u^+\cap S_d|}\leqslant\frac{C\cdot\log n}{\varepsilon^2}$. Either way, we have the desired bound. The proof for the fourth bullet is nearly the same, again using the conditioning on B^c with $R_1(u)$ taking the place of $|N_u^+|$ and S_r taking the place of S_d. □

Fact 2. *Fix $u,v\in V$ and let $\mathcal{C}$ be as above. Suppose $\bar{d}_{uv}\leqslant 7/10$ and $|N_u^+\cap S_d\cap C(u)|\geqslant(17/20)\cdot|N_u^+\cap S_d|$. Then $|N_u^+\cap N_v^+\cap S_d\cap C(u)|/|N_u^+\cap S_d|\geqslant 3/20$.*

Proof of Fact 2. Since $\bar{d}_{uv}\leqslant 7/10$, we have that

$$|N_u^+\cap N_u^+\cap S_d|\geqslant\frac{3}{10}\cdot|(N_u^+\cup N_v^+)\cap S_d|\geqslant\frac{3}{10}\cdot|N_u^+\cap S_d|.$$

Further, since $|N_u^+\cap S_d\cap C(u)|\geqslant(17/20)\cdot|N_u^+\cap S_d|$, it follows that $|N_u^+\cap S_d\cap\overline{C(u)}|\leqslant(3/20)\cdot|N_u^+\cap S_d|$. Then we see that

$$\begin{aligned}|N_u^+\cap N_v^+\cap S_d\cap C(u)|&=|N_u^+\cap N_v^+\cap S_d|-|N_u^+\cap N_v^+\cap S_d\cap\overline{C(u)}|\\&\geqslant\frac{3}{10}\cdot|N_u^+\cap S_d|-\frac{3}{20}|N_u^+\cap S_d|=\frac{3}{20}|N_u^+\cap S_d|.\end{aligned}$$

□

4.2.1 Negative fractional cost of correlation metric for finite p

We next bound the ℓ_p-norm cost of the estimated adjusted correlation metric $\tilde{d}$ on negative edges.

Proof of Lemma 4. Let y be the vector of disagreements in a fixed clustering $\mathcal{C}$, let $C(u)$ denote the set of vertices in vertex v's cluster, and let $\overline{C(u)}:=V\backslash C(u)$.

By Definition 5, the only negative edges uv that contribute to the fractional cost are those with $\bar{d}_{uv}\leqslant 7/10$, i.e., with $1-\bar{d}_{uv}\geqslant 3/10$. Moreover, if $u\in R_1$, no negative edges incident to u contribute to the fractional cost. Thus, to prove the lemma, it suffices to bound $\sum_{u\in R_2\cap V_0}|R_1(u)|^p$, where recall $R_2=V\backslash R_1$, and $R_1(u)=\{v\in N_u^-:\bar{d}_{uv}\leqslant 7/10\}$. While $R_1(u)$ is a random set, we note its randomness only depends on subsample S_d.

We consider a few cases for $u\in R_2$. The sets below depend on the randomness of S_d.

$$V^1:=V_0\cap\{u\in V:|N_u^+\cap S_d\cap\overline{C(u)}|\geqslant 3/20\cdot|N_u^+\cap S_d|\}$$
$$V^2:=V_0\cap\{u\in V:|N_u^+\cap S_d\cap C(u)|\geqslant 17/20\cdot|N_u^+\cap S_d|\}$$
$$V^{2a}:=V^2\cap\{u\in V:|N_u^-\cap C(u)|\geqslant|N_u^+|\},\qquad V^{2b}:=V^2\cap\{u\in V:|C(u)|\leqslant 2\cdot|N_u^+|\}$$

Note that $V_0=V^1\cup V^{2a}\cup V^{2b}$, so

$$\sum_{u\in R_2\cap V_0}|R_1(u)|^p=\sum_{u\in R_2\cap V^1}|R_1(u)|^p+\sum_{u\in R_2\cap V^{2a}}|R_1(u)|^p+\sum_{u\in R_2\cap V^{2b}}|R_1(u)|^p.$$

Now we bound the three sums separately.

Claim 1. *Let $1\leqslant p<\infty$. The following bounds hold:*

- $\mathbb{E}\left[\sum_{u\in R_2\cap V^1}|R_1(u)|\right]\leqslant O\left(1/\varepsilon^2\right)\cdot\mathit{OPT}_1$*, and*
- *Conditioned on the event B^c, we have $\sum_{u\in R_2\cap V^1}|R_1(u)|^p\leqslant O\left((1/\varepsilon^2\cdot\log n)^p\right)\cdot\mathit{OPT}_p^p$.*

Proof of Claim 1. We have

$$\sum_{u\in R_2\cap V^1}|R_1(u)|^p=\sum_{u\in R_2\cap V^1}|R_1(u)|^p\cdot\mathbf{1}_{\{|R_1(u)\cap S_r|=0\}}+\sum_{u\in R_2\cap V^1}|R_1(u)|^p\cdot\mathbf{1}_{\{|R_1(u)\cap S_r|\geqslant 1\}}. \tag{2}$$

We bound the first term of the right-hand side as

$$\begin{aligned}&\sum_{u\in V}|R_1(u)|^p\cdot\mathbf{1}_{\{|R_1(u)\cap S_r|=0\}}\cdot\mathbf{1}_{\{u\in R_2\}}\cdot\mathbf{1}_{\{u\in V^1\}}\\ &\qquad\leqslant\sum_{u\in V}|R_1(u)|^p\cdot\mathbf{1}_{\{|R_1(u)\cap S_r|=0\}}\cdot\mathbf{1}_{\{u\in V^1\}}\cdot|N_u^+\cap S_d|\\ &\qquad\leqslant {}^{20}\!/\!_3\cdot\sum_{u\in V}|R_1(u)|^p\cdot\mathbf{1}_{\{|R_1(u)\cap S_r|=0\}}\cdot|N_u^+\cap S_d\cap\overline{C(u)}|^p\\ &\qquad\leqslant {}^{20}\!/\!_3\cdot\sum_{u\in V}(y(u))^p\cdot|R_1(u)|^p\cdot\mathbf{1}_{\{|R_1(u)\cap S_r|=0\}}\end{aligned}$$

which for $p=1$ is bounded in expectation by ${}^{20}\!/\!_3\cdot\sum_{u\in V}y(u)\cdot\mathbb{E}[|R_1(u)|\cdot(1-q(\varepsilon))^{|R_1(u)|}]$ (by Corollary 2), which in turn is bounded by ${}^{40}\!/\!_{3\varepsilon^2}\cdot\sum_{u\in V}y(u)$. For finite p, we obtain an upper bound of ${}^{20}\!/\!_3\cdot({}^{2C\cdot\log n}\!/\!_{\varepsilon^2})^p\cdot\sum_{u\in V}y(u)^p$, using the definition of B^c (see item 5 in Definition 12).

Next, we bound the second term of the right-hand side of (2):

$$\begin{aligned}&\sum_{u\in V}|R_1(u)|^p\cdot\mathbf{1}_{\{|R_1(u)\cap S_r|\geqslant 1\}}\cdot\mathbf{1}_{\{u\in R_2\}}\cdot\mathbf{1}_{\{u\in V^1\}}\\ &\leqslant({}^{10}\!/\!_3)^p\cdot\sum_{u\in V}|N_u^+\cap S_d|^p\cdot\mathbf{1}_{\{u\in V^1\}}\cdot\mathbf{1}_{\{|R_1(u)\cap S_r|\geqslant 1\}}\cdot\frac{|R_1(u)|^p}{|R_1(u)\cap S_r|^p}\\ &\leqslant({}^{200}\!/\!_9)^p\cdot\sum_{u\in V}|N_u^+\cap S_d\cap\overline{C(u)}|^p\cdot\mathbf{1}_{\{|R_1(u)\cap S_r|\geqslant 1\}}\cdot\frac{|R_1(u)|^p}{|R_1(u)\cap S_r|^p}\\ &\leqslant({}^{200}\!/\!_9)^p\cdot\sum_{u\in V}y(u)^p\cdot\mathbf{1}_{\{|R_1(u)\cap S_r|\geqslant 1\}}\cdot\frac{|R_1(u)|^p}{|R_1(u)\cap S_r|^p}\end{aligned}$$

Applying Proposition 1 we obtain a bound of ${}^{14}\!/\!_{\varepsilon^2}\cdot{}^{200}\!/\!_9\cdot\sum_{u\in V}y(u)$ in expectation for $p=1$, and of $\left({}^{(200\cdot C\cdot\log n)}\!/\!_{(9\cdot\varepsilon^2)}\right)^p\cdot\sum_{v\in V}y(v)^p$ conditioned on the event B^c for finite $p\geqslant 1$. ☐

Claim 2. *Let $1\leqslant p<\infty$. The following bounds hold:*

- $\mathbb{E}\left[\sum_{u\in R_2\cap V^{2a}}|R_1(u)|\right]\leqslant O({}^1\!/\!_{\varepsilon^2})\cdot \mathit{OPT}_1$
- *Conditioned on the event B^c, we have* $\left[\sum_{u\in R_2\cap V^{2a}}|R_1(u)|^p\right]\leqslant O\left(({}^1\!/\!_{\varepsilon^2}\cdot\log n)^p\right)\cdot \mathit{OPT}_p^p$.

Proof of Claim 2. The proof is identical to that of Claim 1 until the penultimate line, at which point we upper bound $|N_u^+\cap S_d|$ by $|N_u^+|$, and then use that $u\in V^{2a}$ implies $y(u)\geqslant|N_u^+|$. ☐

Claim 3. *Let $1\leqslant p<\infty$. The following bounds holds:*

- $\mathbb{E}\left[\sum_{u\in R_2\cap V^{2b}}|R_1(u)|\right]\leqslant O({}^1\!/\!_{\varepsilon^2})\cdot \mathit{OPT}_1$
- *Conditioned on the event B^c, we have* $\sum_{u\in R_2\cap V^{2b}}|R_1(u)|^p\leqslant O\left(({}^1\!/\!_{\varepsilon^2}\cdot\log n)^p\right)\cdot \mathit{OPT}_p^p$.

Proof of Claim 3. Fix $u\in R_2\cap V^{2b}$. For $v\in R_1(u)$, define

$$N_{u,v}:=N_u^+\cap N_v^+\cap C(u).$$

Since $\bar{d}_{uv}\leqslant 7/10$ and $|N_u^+\cap S_d\cap C(u)|\geqslant(17/20)\cdot|N_u^+\cap S_d|$, we have that $\frac{|N_{u,v}\cap S_d|}{|N_u^+\cap S_d|}\geqslant{}^3\!/\!_{20}$ by Fact 2. For $w\in N_u^+\cap C(u)$, define

$$\varphi(u,w):=|R_1(u)\cap N_w^+|.$$

Observe that $\varphi(u,w) \leqslant y(w) + y(u)$.

We have, by way of double counting,

$$\sum_{w \in N_u^+ \cap S_d \cap C(u)} \varphi(u,w) = \sum_{w \in N_u^+ \cap S_d \cap C(u)} |R_1(u) \cap N_w^+| = \sum_{v \in R_1(u)} |N_{u,v} \cap S_d| \geqslant {}^3\!/\!_{20} \cdot |N_u^+ \cap S_d| \cdot |R_1(u)|,$$

which implies

$$\begin{aligned}
&\sum_{u \in V^{2b}} \frac{|N_u^+|^p}{|N_u^+ \cap S_d|^p} \cdot \frac{1}{|C(u)|} \cdot \sum_{w \in N_u^+ \cap C(u) \cap S_d} \varphi(u,w)^p \\
&\geqslant \sum_{u \in V^{2b}} \frac{|N_u^+|^p}{|N_u^+ \cap S_d|^p} \cdot \frac{1}{|C(u)|} \cdot \frac{1}{|N_u^+ \cap C(u) \cap S_d|^{p-1}} \cdot \left(\sum_{w \in N_u^+ \cap C(u) \cap S_d} \varphi(u,w) \right)^p \\
&\geqslant \sum_{u \in V^{2b}} \frac{|N_u^+|^p}{|N_u^+ \cap S_d|^p} \cdot \frac{1}{2|N_u^+|} \cdot \frac{1}{|N_u^+|^{p-1}} \cdot \left(\sum_{w \in N_u^+ \cap C(u) \cap S_d} \varphi(u,w) \right)^p \\
&\geqslant \sum_{u \in V^{2b}} \frac{|N_u^+|^p}{|N_u^+ \cap S_d|^p} \cdot \frac{1}{2|N_u^+|} \cdot \frac{1}{|N_u^+|^{p-1}} \cdot (3/20)^p \cdot |N_u^+ \cap S_d|^p \cdot |R_1(u)|^p \\
&= \sum_{u \in V^{2b}} {}^1\!/\!_2 \cdot ({}^3\!/\!_{20})^p \cdot |R_1(u)|^p.
\end{aligned}$$

In the second line we have used Jensen's inequality. So to bound $\sum_{u \in V^{2b}} |R_1(u)|^p$, it suffices to bound

$$\begin{aligned}
&\sum_{u \in V^{2b}} \frac{|N_u^+|^p}{|N_u^+ \cap S_d|^p} \cdot \frac{1}{|C(u)|} \cdot \sum_{w \in N_u^+ \cap C(u) \cap S_d} \varphi(u,w)^p \\
&\leqslant \sum_{u \in V^{2b}} \frac{|N_u^+|^p}{|N_u^+ \cap S_d|^p} \cdot \frac{1}{|C(u)|} \cdot \sum_{w \in N_u^+ \cap C(u) \cap S_d} (y(u) + y(w))^p \\
&\leqslant 2^{p-1} \cdot \sum_{u \in V} \frac{|N_u^+|^p}{|N_u^+ \cap S_d|^p} \cdot \mathbf{1}_{\{u \in V_0\}} \cdot \frac{1}{|C(u)|} \cdot \sum_{w \in N_u^+ \cap C(u) \cap S_d} y(u)^p \\
&\quad + 2^{p-1} \cdot \sum_{u \in V} \frac{|N_u^+|^p}{|N_u^+ \cap S_d|^p} \cdot \mathbf{1}_{\{u \in V_0\}} \cdot \frac{1}{|C(u)|} \cdot \sum_{w \in N_u^+ \cap C(u) \cap S_d} y(w)^p
\end{aligned}$$

by Jensen's inequality again. By Proposition 1, the first term is easily seen to be bounded by ${}^{14}\!/\!_{\varepsilon^2} \cdot \mathsf{OPT}_1$ in expectation for $p = 1$, and by $2^{p-1} \cdot ({}^{C \cdot \log n}\!/\!_{\varepsilon^2})^p \cdot \mathsf{OPT}_p^p$ conditioned on the event B^c for finite $p \geqslant 1$. We bound the second sum:

$$\begin{aligned}
&2^{p-1} \cdot \sum_{u \in V} \frac{|N_u^+|^p}{|N_u^+ \cap S_d|^p} \cdot \mathbf{1}_{\{u \in V_0\}} \cdot \frac{1}{|C(u)|} \cdot \sum_{w \in N_u^+ \cap C(u) \cap S_d} y(w)^p \\
&\leqslant 2^{p-1} \cdot \sum_{w \in V} y(w)^p \cdot \sum_{u \in N_w^+ \cap C(w)} \frac{|N_u^+|^p}{|N_u^+ \cap S_d|^p} \cdot \mathbf{1}_{\{u \in V_0\}} \cdot \frac{1}{|C(u)|} \\
&\leqslant 2^{p-1} \cdot \sum_{w \in V} y(w)^p \cdot \sum_{u \in N_w^+ \cap C(w)} \frac{|N_u^+|^p}{|N_u^+ \cap S_d|^p} \cdot \mathbf{1}_{\{u \in V_0\}} \cdot \frac{1}{|C(w)|}.
\end{aligned}$$

Again applying Proposition 1, we obtain the same upper bounds as above for the first term. □

Combining Claims 1, 2, and 3 completes the proof of Lemma 4. □

5 Cost of Algorithm 1 for Finite p

We bound the cost of Algorithm 1 for $p \in [1, \infty)$. We still use OPT_p to refer to the value of an optimal solution for the ℓ_p-norm objective, and additionally we often fix such an optimal clustering $\mathcal{C}_{\mathsf{OPT}}$.

At a high-level, we charge disagreements made by Algorithm 1 to the disagreements in $\mathcal{C}_{\mathsf{OPT}}$. Recall Algorithm 1 has several subroutines:

- a Pre-clustering phase that clusters nodes v that have (i) some of their positive neighborhood sampled into S_d, i.e. $v \in V_0$, and (ii) are close (with respect to $\tilde{d}$) to a cluster center S_p;
- a subroutine that runs the standard Pivot algorithm on nodes $v \in \overline{V_0}$, which are the nodes that did not have any positive neighbor sampled into S_d;
- and a subroutine that runs a modified version of the Pivot algorithm on nodes $v \in V_0'$, which are the nodes that have a positive neighbor sampled into S_d, but were far away from all cluster centers.

We note the last two subroutines are part of the Pivot phase of Algorithm 1. Further, the clusters output by each subroutine are totally disjoint. The disagreements incurred by Algorithm 1 can be partitioned into the disagreements made *within* each subroutine (e.g., u and v are both clustered in the Pre-clustering phase, but uv forms a disagreement in the solution output by Algorithm 1), and the disagreements *between* each subroutine (e.g., u is clustered in the Pre-clustering phase and v is clustered in the Pivot phase, but $uv \in E^+$). Therefore, we partition our analysis into the cost of the disagreements incurred during the Pre-clustering phase (Section 5.1), the cost of the disagreements incurred during the Pivot phase (see Section 5.2), and those cost incurred between these two phases (Section 5.3). Note the cost of the disagreements incurred during the Pivot phase includes the cost from running the standard Pivot algorithm on nodes $G[\overline{V_0}]$, the cost from running the modified Pivot algorithm on nodes $G[V_0']$, and the cost of disagreements between the two subroutines of the Pivot phase. See Figure 1 for references to the lemmas for each type of disagreement.

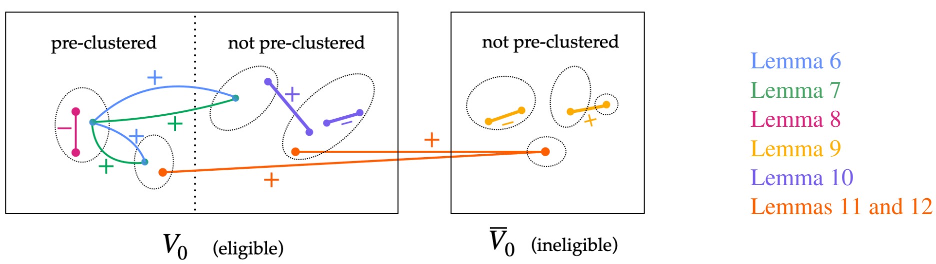

Figure 1: Overview of lemmas for the analysis of Algorithm 1 for finite p. Solid edges are disagreements, and dashed ovals are clusters. Vertices are partitioned into three sets based on whether or not they are eligible and pre-clustered, eligible and not pre-clustered, or ineligible. Charging the cost of a disagreement then depends on which set its endpoints belong to, its sign, and potentially which endpoint was higher with respect to the partial ordering $\succ$. Edges are partitioned by color, with edge types of the same color bounded by the correspondingly colored lemma. Note Lemmas 6 and 7 correspond to the same uv pair, but which lemma is relevant depends on (from the perspective of u's disagreements) whether $u \succ v$ or $v \succ u$.

Sometimes we are able to directly charge disagreements Algorithm 1 makes to $\mathcal{C}_{\mathsf{OPT}}$. More often, we use the estimated adjusted correlation metric, $\tilde{d}$, as an intermediary—specifically, we charge disagreements made by Algorithm 1 to $\tilde{d}$, then charge the cost of $\tilde{d}$ to OPT_p. The latter charging arguments can be found in Section 4.

As in Section 4, we condition on the good event B^c occurring with high probability (see the end of Section 3).

5.1 Cost of Pre-clustering phase

Throughout, we use the choices of δ, c, r in Algorithm 1, and let $t := r/(2\delta)$ be a threshold parameter, which will be used in our analysis.

Recall that for nodes u assigned to clusters during the Pre-clustering phase, there is some node in S_p that has close $\tilde{d}$ distance to u. The highest ordered, with respect to the ordering of S_p, is said to pre-cluster u and is denoted by $s^*(u)$. We may refer to $s^*(u)$ as u's *center*.

Recall we say v *is clustered before* u (or u *is clustered after* v) if either both u and v are pre-clustered, but $s^*(v)$ is before $s^*(u)$ (with respect to the ordering of S_p), or if v is pre-clustered but u is not. For shorthand, we write $v \succ u$ when v is clustered before u.

Fix a node u that is assigned a cluster during the Pre-clustering phase of Algorithm 1. Consider all nodes $v \in V_0$, so that uv is a disagreement in the output of Algorithm 1.

5.1.1 Cost of positive edges

We begin by bounding the cost of positive edges where at least one endpoint is pre-clustered, and both endpoints are eligible.

Fix a vertex $u \in V_0$. We partition the cost of positive disagreements incurred within the Pre-clustering phase based on whether $u \prec v$ or $v \prec u$. Lemma 6 handles the cost of disagreements uv incident to u when $v \succ u$, while Lemma 7 handles the cost of disagreements uv incident to u when $u \succ v$ and $v \in V_0$. In the proofs of both lemmas, we will see that it is easy to charge a disagreement uv to $\tilde{d}$ when $\tilde{d}_{uv}$ is sufficiently large, as we can then charge the ℓ_p-norm cost of $\tilde{d}$ to OPT_p using the lemmas in Section 4. On the other hand, the difficult settings for both lemmas are for edges uv where $\tilde{d}_{uv}$ is small and both $u, v \in V_0$, so this distance is actually a reliable indicator that u and v do have many positive neighbors in common. The key is that even though $\tilde{d}_{uv}$ is small, the fact that u and v are not clustered together indicates there must be some other vertices we can charge to that do have large distance from u.

Some of the future claims will use that for the choices of δ, c, r as in the algorithm,

$$\max\left\{\frac{1}{cr/\delta - r}, \frac{1}{1-(\delta\cdot r + \delta^2\cdot c\cdot r + \delta\cdot r/2)}\right\} \leqslant 8 \quad \text{and} \quad \max\left\{\frac{1}{r/2 + c\cdot\delta\cdot r}, \frac{1}{c\cdot r}\right\} \leqslant 5. \tag{3}$$

Lemma 6. *Condition on the good event B^c and fix $1 \leqslant p < \infty$. The ℓ_p-cost for $u \in V_0$ of the edges $uv \in E^+$ in disagreement with respect to $\mathcal{C}_{ALG}$, where $v \succ u$, is bounded by $O\left(\frac{1}{\varepsilon^8}\cdot\log^4 n\right)\cdot \mathsf{OPT}_p$.*

Proof of Lemma 6. Note by definition of $\succ$ that each v in the statement of the lemma is necessarily pre-clustered, thus also $v \in V_0$. We partition the set $\{v \in N_u^+ : v \succ u\}$ depending on whether $\tilde{d}_{uv} > t$ or $\tilde{d}_{uv} \leqslant t$. Define $E_1(u)$ to be the random set of u's close, positive neighbors that are clustered before u:

$$E_1(u) := \{v \in N_u^+ : v \succ u\} \cap \text{Ball}_{\tilde{d}}(u, t).$$

Define $E_2(u)$ to be the remaining positive neighbors of u that are clustered before u:

$$E_2(u) := \{v \in N_u^+ : v \succ u\} \backslash E_1(u).$$

We partition the sum we wish to bound using Jensen's inequality to see

$$\sum_{u\in V_0} \left|\{v \in N_u^+ : v \succ u\}\right|^p \leqslant 2^{p-1}\cdot \sum_{u\in V_0}|E_1(u)|^p + 2^{p-1}\cdot\sum_{u\in V_0}|E_2(u)|^p. \tag{4}$$

As we alluded to before the beginning of the proof, it is straightforward to bound the cost of disagreeing edges uv when $\tilde{d}_{uv}$ is large. In particular, we can bound the latter sum:

$$\sum_{u\in V_0}|E_2(u)|^p = \sum_{u\in V_0}|\{v\in N_u^+\cap V_0 : \tilde{d}_{uv} \geqslant t\}|^p \leqslant \frac{1}{t^p}\cdot\sum_{u\in V_0}\left(\sum_{v\in N_u^+\cap V_0}\tilde{d}_{uv}\right)^p$$
$$\leqslant \left({}^{(7215\cdot C^3\cdot\log^3 n)}\!/\!{}_{\varepsilon^6}\right)^p\cdot\mathsf{OPT}_p^p. \tag{5}$$

where the third inequality is from Lemma 3 and subbing in the values of δ and r.

Bounding $\sum_{u \in V_0} |E_1(u)|^p$ in line (4) is the more involved piece. We begin by partitioning $u \in V_0$ based on whether u has a close neighbor sampled by the center sample S_p; overall, we need to bound E_{1a} and E_{1b} where

$$\sum_{u \in V_0} |E_1(u)|^p = \underbrace{\sum_{u \in V_0 : \mathrm{Ball}_{\tilde{d}}^{S_p}(u,t) = \varnothing} |E_1(u)|^p}_{E_{1a}} + \underbrace{\sum_{u \in V_0 : \mathrm{Ball}_{\tilde{d}}^{S_p}(u,t) \neq \varnothing} |E_1(u)|^p}_{E_{1b}}$$

Bounding E_{1a}. Intuitively, the term E_{1a} will be easier to bound than E_{1b}, because, conditioned on B^c, the fact that $\mathrm{Ball}_{\tilde{d}}^{S_p}(u,t) = \varnothing$ implies that $|\mathrm{Ball}_{\tilde{d}}(u,t)|$ is small. So even though there are some nodes $v \in N_u^+$ that are close to u but assigned a different cluster than u, there cannot be that many of them. We use this insight together with the following claim, Claim 4, which proves that there is sufficient fractional cost incident to $u \in V_0$. In turn, the small number of disagreements incident to u can be charged to this fractional cost, via Lemma 3.

Recall that $\tilde{D}_0(u) := \sum_{v \in N_u^+ \cap V_0} \tilde{d}_{uv} + \sum_{v \in N_u^- \cap V_0} (1 - \tilde{d}_{uv})$.

Claim 4. *If $u \in V_0$ and $E_1(u) \neq \varnothing$, then $\tilde{D}_0(u) \geqslant 1/5$.*

Proof of Claim 4. Let v be any vertex in $E_1(u)$. Then $s^*(v)$ exists. We can upper and lower bound $\tilde{d}_{u\, s^*(v)}$ by a constant. Namely, by the approximate triangle inequality (Lemma 1), $\tilde{d}_{u\, s^*(v)} \leqslant \delta(\tilde{d}_{uv} + \tilde{d}_{v\, s^*(v)}) \leqslant \delta(t + c \cdot r)$. On the other hand, since u is eligible but $v \succ u$, we know that $\tilde{d}_{u\, s^*(v)} > c \cdot r$.

Since $\tilde{d}_{u\, s^*(v)} < 1$, we have that $s^*(v) \in V_0$ (Fact 1). This means that $D_0(u)$ is lower bounded by $1 - d_{u\, s^*(v)} \geqslant 1 - \delta(t + c \cdot r) \geqslant c \cdot r \geqslant 1/5$ in the case that $u\, s^*(v) \in E^-$ and by $d_{u\, s^*(v)} \geqslant c \cdot r \geqslant 1/5$ in the case that $u\, s^*(v) \in E^+$, using the lower bounds from line (3). $\square$

We now bound E_{1a} in the following claim.

Claim 5. *Condition on the good event B^c. For $t = \frac{r}{2\delta}$, it is the case that*

$$E_{1a} := \sum_{u \in V_0 : \mathit{Ball}_{\tilde{d}}^{S_p}(u,t) = \varnothing} |E_1(u)|^p \leqslant O\left(\left(1/\varepsilon^8 \cdot \log^4 n \right)^p \right) \cdot \mathit{OPT}_p^p.$$

Proof of Claim 5. It suffices here to upper bound $E_1(u)$ by $|\mathrm{Ball}_{\tilde{d}}(u,t)|$. Applying Claim 4, we obtain

$$\begin{aligned} E_{1a} &\leqslant \sum_{u \in V_0 : \mathrm{Ball}_{\tilde{d}}^{S_p}(u,t) = \varnothing} |\mathrm{Ball}_{\tilde{d}}(u,t)|^p \leqslant 5^p \cdot \sum_{u \in V_0 : \mathrm{Ball}_{\tilde{d}}^{S_p}(u,t) = \varnothing} |\mathrm{Ball}_{\tilde{d}}(u,t)|^p \cdot (\tilde{D}_0(u))^p \\ &\leqslant \left(5C' \cdot \log_{1/(1-q(\varepsilon))} n \right)^p \sum_{u \in V_0} (\tilde{D}_0(u))^p \\ &\leqslant \left((4440 \cdot C' \cdot C^3 \cdot \log^4 n)/\varepsilon^8 \right)^p \cdot \mathsf{OPT}_p^p. \end{aligned}$$

In the penultimate line, we use that conditioning on B^c, if $\mathrm{Ball}_{\tilde{d}}^{S_p}(u,t) = \varnothing$, it must be that $|\mathrm{Ball}_{\tilde{d}}(u,t)| \leqslant C' \cdot \log_{1/(1-q(\varepsilon))} n \leqslant 8C'/\varepsilon^2 \cdot \log n$. In the last line we use Corollary 1.

$\square$

Bounding E_{1b}. Recall

$$E_{1b} := \sum_{u \in V_0 : \mathrm{Ball}_{\tilde{d}}^{S_p}(u,t) \neq \varnothing} |E_1(u)|^p = \sum_{u \in V_0 : \mathrm{Ball}_{\tilde{d}}^{S_p}(u,t) \neq \varnothing} |\{v \in N_u^+ : v \succ u, \tilde{d}_{uv} \leqslant t\}|^p.$$

Intuitively, because u is both pre-clustered and has a close neighbor sampled in S_p, we are now closer to the offline setting. In particular, since $v \succ u$, we have $|\mathrm{Ball}_{\tilde{d}}^{S_b}(s^*(u), r)| \leqslant |\mathrm{Ball}_{\tilde{d}}^{S_b}(s^*(v), r)|$. The idea is that u lies in an annulus around $\mathrm{Ball}_{\tilde{d}}^{S_b}(s^*(v), r)$, and so the fractional cost of u can be lower

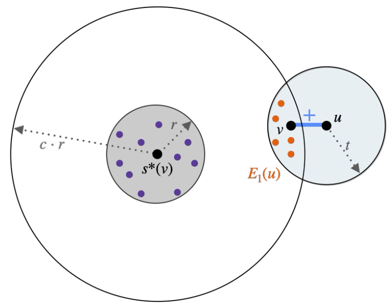

Figure 2: Bounding E_{1b} in the proof of Lemma 6, where $v \in N_u^+$ with $v > u$ and $\tilde{d}_{uv} \leqslant t$. We charge the disagreements between $v \in E_1(u)$ and u to the purple nodes in $\text{Ball}_{\tilde{d}}^{S_b}(s^*(v), r)$.

bounded by (a constant factor times) the $|\text{Ball}_{\tilde{d}}^{S_b}(s^*(v), r)|$. The subtlety is that the inequality above lower bounding $|\text{Ball}_{\tilde{d}}^{S_b}(s^*(v), r)|$ in turn only holds when the balls are restricted to S_b, unlike in the offline case, where $S_b = V$. So it is not a priori clear that there will be enough fractional cost to which to charge $|E_1(u)|^p$. See Figure 2 for an illustration.

For each $u \in V_0$ with $\text{Ball}_{\tilde{d}}^{S_p}(u, t) \neq \emptyset$, choose a fixed but arbitrary $z(u) \in \text{Ball}_{\tilde{d}}^{S_p}(u, t)$. Note that because $z(u) \in S_p$ and $\tilde{d}_{u\,z(u)} \leqslant t \leqslant c \cdot r$, we know that $z(u)$ is a *candidate* for clustering u, so in particular $s^*(u)$ exists. Note $z(u)$ is a random variable depending on S_p, S_d, and S_r.

Recall that $E_1(u) := \{v \in N_u^+ : v > u, \tilde{d}_{uv} \leqslant t\}$. Define

$$B(u) := \bigcup_{v \in E_1(u)} \text{Ball}_{\tilde{d}}^{S_b}(s^*(v), r),$$

that is, $B(u)$ is the union of balls in S_b, cut out around the vertices that cluster the vertices in $E_1(u)$. We will show that we can charge $|E_1(u)|^p$ to $|B(u)|^p$. Further, the set $B(u)$ is constructed so that every node $b \in B(u)$ lies in an annulus around u, so we can in turn charge $|B(u)|^p$ to the ℓ_p-cost of $\tilde{d}$.

The interesting case is when $|\text{Ball}_{\tilde{d}}(z(u), r)|$ is large. Here, we use Claims 6 and 7 to bound $|E_1(u)|$ in terms of $|B(u)|$. Then we use Claim 8 to relate $|B(u)|$ to the ℓ_p-cost of $\tilde{d}$.

Claim 6. *If $\text{Ball}_{\tilde{d}}^{S_p}(u, t) \neq \emptyset$, then $|E_1(u)| \leqslant |\text{Ball}_{\tilde{d}}(z(u), r)|$.*

Proof of Claim 6. The claim follows from an application of the approximate triangle inequality (Lemma 1). If $\text{Ball}_{\tilde{d}}^{S_p}(u, t) \neq \emptyset$, $z(u)$ exists and $\tilde{d}_{u\,z(u)} \leqslant t$. Further, by definition, if $v \in E_1(u)$, then $\tilde{d}_{uv} \leqslant t$. So for any $v \in E_1(u)$, we have $\tilde{d}_{v\,z(u)} \leqslant 2\delta t = r$, so $E_1(u) \subseteq \text{Ball}_{\tilde{d}}(z(u), r)$. $\square$

Claim 7. *If $u \in V_0$ and $\text{Ball}_{\tilde{d}}^{S_p}(u, t) \neq \emptyset$, then $|\text{Ball}_{\tilde{d}}^{S_b}(z(u), r)| \leqslant |B(u)|$.*

Proof of Claim 7. If $\text{Ball}_{\tilde{d}}^{S_p}(u, t) \neq \emptyset$, then $z(u)$ exists and is a center, and in particular is a candidate for pre-clustering u (since $\tilde{d}_{u,z(u)} \leqslant t < c \cdot r$ and $u \in V_0$). Thus $s^*(u)$ exists and $|\text{Ball}_{\tilde{d}}^{S_b}(z(u), r)| \leqslant |\text{Ball}_{\tilde{d}}^{S_b}(s^*(u), r)|$. Further, for any $v \in E_1(u)$, $|\text{Ball}_{\tilde{d}}^{S_b}(s^*(u), r)| \leqslant |\text{Ball}_{\tilde{d}}^{S_b}(s^*(v), r)|$, since v by definition is clustered before u. The claim then follows from the definition of $B(u)$. $\square$

Claim 8. *If $u \in V_0$, then $|B(u)| \leqslant 8 \cdot \tilde{D}_0(u)$.*

Proof of Claim 8. If $E_1(u) = \emptyset$, then $B(u) = \emptyset$, so the claim holds. Thus we may assume that $E_1(u) \neq \emptyset$. It then suffices to show that $\tilde{d}_{ub} \geqslant 1/8$ and $1 - \tilde{d}_{ub} \geqslant 1/8$ for every $b \in B(u)$. Then we will have (by Fact 1) that $b \in V_0$ (because $\tilde{d}_{ub} < 1$) for every $b \in B(u)$, and thus the claim follows.

To see that $\tilde{d}_{ub} \geqslant 1/8$ for any $b \in B(u)$, let $v \in E_1(u)$ be such that $b \in \text{Ball}_{\tilde{d}}(s^*(v), r)$ (such v exists by the definition of $B(u)$). Since $v \in E_1(u)$, u is clustered after v, so, using also that $u \in V_0$, we have that $\tilde{d}_{u\,s*(v)} > c \cdot r$. Also, by choice of v, $\tilde{d}_{b\,s*(v)} \leqslant r$. So by the approximate triangle inequality (Lemma 1), $\tilde{d}_{ub} \geqslant cr/\delta - r$, which is lower bounded by $1/8$ by line (3).

To see that $1 - \tilde{d}_{ub} \geqslant 1/8$ for any $b \in B(u)$, observe that $\tilde{d}_{b\,s*(v)} \leqslant r$ (by choice of v), $\tilde{d}_{v\,s*(v)} \leqslant c \cdot r$ (by definition of the algorithm and of $s^*(v)$), and $\tilde{d}_{vu} \leqslant t$ (since $v \in E_1(u)$). So by the approximate triangle inequality (Lemma 1), we have $\tilde{d}_{ub} \leqslant \delta \cdot [\tilde{d}_{b\,s*(v)} + \delta(\tilde{d}_{v\,s*(v)} + \tilde{d}_{vu})] \leqslant \delta \cdot r + \delta^2 \cdot c \cdot r + \delta^2 \cdot t$, so $1 - \tilde{d}_{ub} \geqslant 1 - (\delta \cdot r + \delta^2 \cdot c \cdot r + \delta^2 \cdot t)$, which is lower bounded by $1/8$ by line (3). $\square$

Claim 9. *Condition on the good event B^c. For $t = \frac{r}{2\delta}$, it is the case that*

$$E_{1b} := \sum_{u \in V_0 : \mathit{Ball}_{\tilde{d}}^{S_p}(u,t) \neq \varnothing} |E_1(u)|^p \leqslant O\left(\left(1/\varepsilon^8 \cdot \log^4 n\right)^p\right) \cdot \mathit{OPT}_p^p.$$

Proof of Claim 9. We use Claim 6, and then partition the sum

$$\begin{aligned} E_{1b} &\leqslant \sum_{u \in V_0 : \mathrm{Ball}_{\tilde{d}}^{S_p}(u,t) \neq \varnothing} \left|\mathrm{Ball}_{\tilde{d}}(z(u), r)\right|^p \\ &\leqslant \sum_{\substack{u \in V_0 : \mathrm{Ball}_{\tilde{d}}^{S_p}(u,t) \neq \varnothing, \\ |\mathrm{Ball}_{\tilde{d}}(z(u),r)| \geqslant 2C \log n/\varepsilon^2}} \left|\mathrm{Ball}_{\tilde{d}}(z(u), r)\right|^p + \sum_{\substack{u \in V_0 : \mathrm{Ball}_{\tilde{d}}^{S_p}(u,t) \neq \varnothing, \\ |\mathrm{Ball}_{\tilde{d}}(z(u),r)| < 2C \log n/\varepsilon^2}} \left|\mathrm{Ball}_{\tilde{d}}(z(u), r)\right|^p. \end{aligned} \tag{6}$$

First, we upper bound the first sum in line (6):

$$\begin{aligned} \sum_{\substack{u \in V_0 : \mathrm{Ball}_{\tilde{d}}^{S_p}(u,t) \neq \varnothing \\ |\mathrm{Ball}_{\tilde{d}}(z(u),r)| \geqslant 2C \log n/\varepsilon^2}} \left|\mathrm{Ball}_{\tilde{d}}(z(u), r)\right|^p &\leqslant \sum_{\substack{u \in V_0 : \\ |\mathrm{Ball}_{\tilde{d}}(z(u),r)| \geqslant 2C \log n/\varepsilon^2}} \left(|\mathrm{Ball}_{\tilde{d}}(z(u), r)| \cdot 8 \cdot \frac{\tilde{D}_0(u)}{|\mathrm{Ball}_{\tilde{d}}^{S_b}(z(u), r)|} \right)^p \\ &\leqslant (32/\varepsilon^2)^p \cdot \sum_{u \in V_0} \left(\tilde{D}_0(u)\right)^p \leqslant \left((3552 \cdot C^3 \cdot \log^3 n)/\varepsilon^8\right)^p \cdot \mathsf{OPT}_p^p, \end{aligned}$$

where in the first inequality, we have applied Claims 7 and 8, and in the second inequality, we use the conditioning on B^c, which implies that $|\mathrm{Ball}_{\tilde{d}}(z(u), r)|/|\mathrm{Ball}_{\tilde{d}}^{S_b}(z(u), r)| \leqslant 4/\varepsilon^2$, since $|\mathrm{Ball}_{\tilde{d}}(z(u), r)| \geqslant 2C \log n/\varepsilon^2$. In the last inequality we have used Corollary 1.

Finally, we bound the second sum in line (6):

$$\begin{aligned} \sum_{\substack{u \in V_0 : \mathrm{Ball}_{\tilde{d}}^{S_p}(u,t) \neq \varnothing, \\ |\mathrm{Ball}_{\tilde{d}}(z(u),r)| < 2C \log n/\varepsilon^2}} \left|\mathrm{Ball}_{\tilde{d}}(z(u), r)\right|^p &= \sum_{\substack{u \in V_0 : \\ |\mathrm{Ball}_{\tilde{d}}(z(u),r)| < 2C \log n/\varepsilon^2}} \left(5 \cdot |\mathrm{Ball}_{\tilde{d}}(z(u), r)|\right)^p \cdot \left(\tilde{D}_0(u)\right)^p \\ &\leqslant \left((1110 \cdot C^4 \cdot \log^4 n)/\varepsilon^8\right)^p \cdot \mathsf{OPT}_p^p. \end{aligned}$$

The first line follows from Claim 4. In the second line, we use the bound $|\mathrm{Ball}_{\tilde{d}}(z(u), r)| < 2C \log n/\varepsilon^2$ and Corollary 1. $\square$

Combining the bounds on the sums and using that C is sufficiently large (which is required by the good event, anyway), we conclude $\sum_{u \in V_0} \left|\{v \in N_u^+ : v \succ u\}\right|^p \leqslant \left((2900 \cdot C' \cdot C^4 \cdot \log^4 n)/\varepsilon^8\right)^p \cdot \mathsf{OPT}_p^p$. $\square$

The proof of Lemma 7 is similar in spirit to that of Lemma 6, but must nonetheless be handled separately (except in the case of $p = 1$, where we can sum over disagreements edge-wise rather than node-wise).

Lemma 7. *Condition on the good event B^c and fix $1 \leqslant p < \infty$. The ℓ_p-cost for u that are pre-clustered (thus are necessarily in V_0) of the edges $uv \in E^+$ in disagreement with respect to $\mathcal{C}_{\mathsf{ALG}}$, where $u \succ v$, is bounded by*

$$\sum_{u \in V_0} \left|\{v \in V_0 \cap N_u^+ : u \succ v\}\right|^p \leqslant O\left(\left(1/\varepsilon^8 \cdot \log^4 n\right)^p\right) \cdot \mathit{OPT}_p^p.$$

Proof of Lemma 7. This proof will follow the same structure as that of Lemma 6, but the fact that v is clustered *after* u (which is the differentiating factor from Lemma 6) plays a key role in the analysis. Note that any u for which $\{v \in V_0 \cap N_u^+ : u \succ v\} \neq \varnothing$ is necessarily pre-clustered, by definition of $\succ$. We partition the set $\{v \in V_0 \cap N_u^+ \mid u \succ v\}$ depending on whether $\tilde{d}_{uv} > t$ or $\tilde{d}_{uv} \leqslant t$. The random set of u's close, positive neighbors that are clustered *after* u is

$$E_1(u) := \{v \in N_u^+ \cap V_0 : u \succ v\} \cap \mathrm{Ball}_{\tilde{d}}(u, t).$$

Define $E_2(u)$ to be the remaining positive neighbors of u that are clustered *after* u:

$$E_2(u) := \{v \in N_u^+ \cap V_0 : u \succ v\} \backslash E_1(u).$$

Then as in Lemma 6, we partition the sum and apply Jensen's inequality to see

$$\sum_{u \in V_0} \left|\{v \in V_0 \cap N_u^+ : u \succ v\}\right|^p \leqslant 2^{p-1} \cdot \sum_{u \in V_0} |E_1(u)|^p + 2^{p-1} \cdot \sum_{u \in V_0} |E_2(u)|^p. \tag{7}$$

Bounding $\sum_{u \in V_0} |E_2(u)|^p$ is the same as in the proof of Lemma 6, as that bound does not rely on whether $u \succ v$ or $v \succ u$. Specifically, line (5) is also an upper bound on $\sum_{u \in V_0} |E_2(u)|^p$, so

$$\sum_{u \in V_0} |E_2(u)|^p \leqslant \left({(7215 \cdot C^3 \cdot \log^3 n)}/{\varepsilon^6}\right)^p \cdot \mathsf{OPT}_p^p. \tag{8}$$

We now case on $|E_1(u)|$:

$$\sum_{u \in V_0} |E_1(u)|^p = \underbrace{\sum_{\substack{u \in V_0: \\ |E_1(u)| < 2C \log n/\varepsilon^2}} |E_1(u)|^p}_{E_{1a}} + \underbrace{\sum_{\substack{u \in V_0: \\ |E_1(u)| \geqslant 2C \log n/\varepsilon^2}} |E_1(u)|^p}_{E_{1b}}$$

Note that we may assume u is pre-clustered for all u in the summations above, since if $E_1(u) \neq \varnothing$, then u is necessarily pre-clustered, by definition of $E_1(u)$.

Bounding E_{1a}. As in the proof of Lemma 6, E_{1a} is easier to bound, because the number of disagreements incident to each u in the sum is small, but again we need to show there is sufficient fractional cost to which to charge these disagreements so that we can then invoke Lemma 3. We use the following claim, Claim 10, which is the analogue of Claim 4 in Lemma 6; the difference here is that the edge $(u, s^*(u))$ contributes to the fractional cost of $\tilde{d}$, rather than the edges $(u, s^*(v))$ for $v \in E_1(u)$.

Claim 10. *If $E_1(u) \neq \varnothing$, then $\tilde{D}_0(u) \geqslant 3/20$.*

Proof of Claim 10. We will lower bound both $\tilde{d}_{u\, s^*(u)}$ and $1 - \tilde{d}_{u\, s^*(u)}$ by $3/20$, and argue that $s^*(u) \in V_0$. Moreover, since $E_1(u) \neq \varnothing$, u is pre-clustered, so $u \in V_0$. Thus the claim follows.

First, since $\tilde{d}_{u\, s^*(u)} \leqslant c \cdot r$, we have that $1 - \tilde{d}_{u\, s^*(u)} \geqslant 1 - c \cdot r$. Then, fix some $v \in E_1(u)$, and apply the approximate triangle inequality (Lemma 1) to see that $\tilde{d}_{v\, s^*(u)} \leqslant \delta(\tilde{d}_{uv} + \tilde{d}_{u\, s^*(u)})$. Substituting in $\tilde{d}_{uv} \leqslant t$ and $\tilde{d}_{v\, s^*(u)} > c \cdot r$, we see $c \cdot r \leqslant \delta(t + \tilde{d}_{u\, s^*(u)})$, which rearranging shows $\frac{c \cdot r}{\delta} - t \leqslant \tilde{d}_{u\, s^*(u)}$.

It just remains to show that $s^*(u) \in V_0$, which follows from the contrapositive of Fact 1, because $\tilde{d}_{u\, s^*(u)} < 1$ and $u \neq s^*(u)$ since $d_{u\, s^*(u)} > 0$.

In total, since we take $t = \frac{r}{2\delta}$, we obtain $\tilde{D}_0(u) \geqslant \min\{1 - c \cdot r, \frac{c \cdot r}{\delta} - t\} \geqslant 3/20$. □

We can now bound E_{1a} in the next claim.

Claim 11. *Condition on the good event B^c. For $t = \frac{r}{2\delta}$, it is the case that*

$$E_{1a} := \sum_{\substack{u \in V_0: \\ |E_1(u)| < 2C \log n/\varepsilon^2}} |E_1(u)|^p \leqslant O\left(\left(1/\varepsilon^8 \cdot \log^4 n\right)^p\right) \cdot \mathit{OPT}_p^p.$$

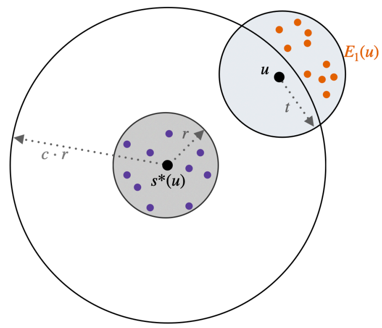

Figure 3: Bounding E_{1b} in the proof of Lemma 7. We show all the purple nodes are sufficiently far from u. Then, we will show that the number of orange nodes, $|E_1(u)|$, is roughly equal to the number of purple nodes, $\text{Ball}_{\tilde{d}}(s^*(u), r)$, so we can charge the disagreements uv for $v \in E_1(u)$ to the cost of $\tilde{d}$ on edges uw for $w \in \text{Ball}_{\tilde{d}}(s^*(u), r)$.

Proof of Claim 11. If $E_1(u) = \varnothing$, it contributes nothing to the sum. For $E_1(u) \neq \varnothing$, we apply Claim 10 to charge to $\tilde{D}_0(u)$ and use the upper bound on $|E_1(u)|$. Then we apply Corollary 1. Together, this leads to the following string of inequalities:

$$\sum_{\substack{u \in V_0: \\ |E_1(u)| \leqslant 2C \log n/\varepsilon^2}} |E_1(u)|^p \leqslant \left({}^{(40 \cdot C \log n)}/_{3\varepsilon^2} \right)^p \cdot \sum_{u \in V_0} \left(\tilde{D}_0(u)\right)^p \leqslant \left({}^{(1480 \cdot C^4 \cdot \log^4 n)}/_{\varepsilon^8} \right)^p \cdot \mathsf{OPT}_p^p.$$

□

Bounding E_{1b}. Recall

$$E_{1b} := \sum_{\substack{u \in V_0: \\ |E_1(u)| \geqslant 2C \log n/\varepsilon^2}} |E_1(u)|^p = \sum_{\substack{u \in V_0: \\ |E_1(u)| \geqslant 2C \log n/\varepsilon^2}} |\{v \in N_u^+ : u \succ v, \tilde{d}_{uv} \leqslant t\}|^p$$

We first show $|E_1(u)| = O(|\text{Ball}_{\tilde{d}}(s^*(u), r)|)$. This bound is useful because, although $s^*(u)$ was chosen as the cluster center for u, both $s^*(u)$ and the nodes that are close to it with respect to $\tilde{d}$ must actually be relatively far from u. Intuitively, this is because the nodes $v \in E_1(u)$ are all close to u, so u cannot be very close to $s^*(u)$—otherwise, those nearby nodes v would have been assigned to the same cluster as $s^*(u)$, which they were not. See Figure 3.

In the follow claim, we show $|E_1(u)| = O(|\text{Ball}_{\tilde{d}}(s^*(u), r)|)$. Then, we see all $a \in \text{Ball}_{\tilde{d}}(s^*(u), r)$ have sufficiently far distance from u, and so we can charge to those distances.

Claim 12. *Condition on the good event B^c. For u with $|E_1(u)| \geqslant 2C \log n/\varepsilon^2$, it is the case that $|E_1(u)| \leqslant {}^8/_{\varepsilon^2} \cdot |\mathit{Ball}_{\tilde{d}}(s^*(u), r)|$.*

Proof of Claim 12. By the good event, since $|\text{Ball}_{\tilde{d}}(u, t)| \geqslant |E_1(u)| \geqslant 2C \log n/\varepsilon^2 \geqslant C' \log_{1/(1-\varepsilon^2/2)} n$ is sufficiently large, there is some sampled center in $\text{Ball}_{\tilde{d}}(u, t)$, i.e. there exists some $z \in \text{Ball}_{\tilde{d}}^{S_p}(u, t)$. Since all $v \in E_1(u)$ have $\tilde{d}_{vz} \leqslant \delta(\tilde{d}_{vu} + \tilde{d}_{uz}) \leqslant 2t\delta = r$, we have $E_1(u) \subseteq \text{Ball}_{\tilde{d}}(z, r)$ and in particular, all $v \in E_1(u) \cap S_b$ are close enough to z to be counted in $\text{Ball}_{\tilde{d}}^{S_b}(z, r)$. But since $v \in E_1(u)$ are clustered after u it must be that

$$\left|\text{Ball}_{\tilde{d}}^{S_b}(z, r)\right| \leqslant \left|\text{Ball}_{\tilde{d}}^{S_b}(s^*(u), r)\right| \leqslant \left|\text{Ball}_{\tilde{d}}(s^*(u), r)\right|. \tag{9}$$

Since $|E_1(u)| > 2C \log n/\varepsilon^2$ and $\text{Ball}_{\tilde{d}}(z, r) \supseteq E_1(u)$, we have $|\text{Ball}_{\tilde{d}}(z, r)| > 2C \log n/\varepsilon^2$. Then we use the good event to see $\left|\text{Ball}_{\tilde{d}}^{S_b}(z, r)\right| \geqslant \frac{q(\varepsilon)}{4} \left|\text{Ball}_{\tilde{d}}(z, r)\right|$. Combining these inequalities, we see

$$|E_1(u)| \leqslant \left|\text{Ball}_{\tilde{d}}(z, r)\right| \leqslant {}^4/_{q(\varepsilon)} \cdot \left|\text{Ball}_{\tilde{d}}^{S_b}(z, r)\right|. \tag{10}$$

Then combining the inequalities in (9) and (10) gives the claim. □

We note the proof of the following claim is similar in structure to that of Claim 8, but we charge to different nodes.

Claim 13. *If $E_1(u) \neq \emptyset$, then $|Ball_{\tilde{d}}(s^*(u), r)| \leqslant 17 \cdot \tilde{D}_0(u)$.*

Proof of Claim 13. It suffices to show that $\tilde{d}_{ua} \geqslant 1/17$ and $1 - \tilde{d}_{ua} \geqslant 1/17$ for every $a \in \text{Ball}_{\tilde{d}}(s^*(u), r)$. Then we will have (by Fact 1) that $a \in V_0$ (because $\tilde{d}_{ua} < 1$) for every $a \in \text{Ball}_{\tilde{d}}(s^*(u), r)$, and thus the claim follows.

Since $\tilde{d}_{u\,s^*(u)} \geqslant \frac{c \cdot r}{\delta} - t$ (by the proof of Claim 10), by Lemma 1 we have that for $a \in \text{Ball}_{\tilde{d}}(s^*(u), r)$,

$$\frac{c \cdot r}{\delta} - t \leqslant \tilde{d}_{u\,s^*(u)} \leqslant \delta(\tilde{d}_{a\,s^*(u)} + \tilde{d}_{ua}) \leqslant \delta(r + \tilde{d}_{ua}),$$

which rearranging gives us that $\frac{c \cdot r}{\delta^2} - \frac{t}{\delta} - r \leqslant \tilde{d}_{ua}$. For our choice of constants, $\tilde{d}_{ua} \geqslant 1/17$.

To lower bound $1 - \tilde{d}_{ua}$, we again use Lemma 1 to see

$$\tilde{d}_{ua} \leqslant \delta(\tilde{d}_{u\,s^*(u)} + \tilde{d}_{a\,s^*(u)}) \leqslant \delta(c+1)r,$$

which rearranged is $1 - \tilde{d}_{ua} \geqslant 1 - \delta r(c+1) \geqslant 1/17$. □

We can now bound E_{1b} in the next claim.

Claim 14. *Condition on the good event B^c. For $t = \frac{r}{2\delta}$, it is the case that*

$$E_{1b} := \sum_{\substack{u \in V_0: \\ |E_1(u)| \geqslant 2C \log n / \varepsilon^2}} |E_1(u)|^p \leqslant O\left(\left(1/\varepsilon^8 \cdot \log^3 n\right)^p\right) \cdot \mathit{OPT}_p^p.$$

Proof of Claim 14. Now we combine the above claims with the bound on the cost of $\tilde{d}$ on V_0 to see

$$E_{1b} \leqslant (8/\varepsilon^2)^p \cdot \sum_{u \in V_0} \left|\text{Ball}_{\tilde{d}}(s^*(u), r)\right|^p \leqslant (136/\varepsilon^2)^p \cdot \sum_{u \in V_0} \left(\tilde{D}_0(u)\right)^p \leqslant \left((15096 \cdot C^3 \cdot \log^3 n)/\varepsilon^8\right)^p \cdot \mathsf{OPT}_p^p. \tag{11}$$

The first inequality is from Claim 12, the second is from Claim 13, and the last is from Corollary 1. □

Continuing the bound from line (7), we substitute in the inequalities from line (8) and Claims 11 and 14, and use the fact that C is sufficiently large by the good event, to see that

$$\sum_{u \in V_0} \left|\{v \in V_0 \cap N_u^+ \mid u \succ v\}\right|^p \leqslant \left((3414 \cdot C^4 \cdot \log^4 n)/\varepsilon^8\right)^p \cdot \mathsf{OPT}_p^p.$$

□

5.1.2 Cost of negative edges

The edges $uv \in E^-$, where at least one endpoint is pre-clustered, that are in disagreement with respect to $\mathcal{C}_{\mathsf{ALG}}$ are those where u is clustered with v. The proof actually follows easily.

Lemma 8. *Condition on the good event B^c and fix $1 \leqslant p < \infty$. The ℓ_p-cost for u in V_0 of the negative edges in the Pre-clustering phase is bounded by*

$$\sum_{u \in V_0} \left|\left\{v \in N_u^- : v \text{ clustered with } u\right\}\right|^p \leqslant O\left((1/\varepsilon^2 \cdot \log n)^p\right) \cdot \mathit{OPT}_p^p.$$

Proof of Lemma 8. If $u, v \in V_0$ are clustered together, there exists $s^* \in S_p$ such that $s^* = s^*(u) = s^*(v)$ (possibly with $s^* = u$ or $s^* = v$). By the approximate triangle inequality (Lemma 1),

$$\tilde{d}_{u,v} \leqslant \delta \cdot (\tilde{d}_{u\,s^*} + \tilde{d}_{v\,s^*}) \leqslant 2\delta c r \leqslant 7/10.$$

Since $\tilde{d}_{uv} < 1$, we know that $\tilde{d}_{uv} = \bar{d}_{uv}$ and also that $u \in R_2$, where we recall that $R_2 = V \backslash R_1$, for R_1 the set of vertices that are isolated by $\tilde{d}$. Thus, to prove the lemma, it suffices to bound $\sum_{u \in V_0 \cap R_2} |R_1(u)|^p$, where $R_1(u) := \{v \in N_u^- : \bar{d}_{uv} \leqslant 7/10\}$. Since we bound this exact quantity in Lemma 4, we find that

$$\sum_{u \in V_0 \cap R_2} |R_1(u)|^p \ \leqslant ((^{74 \cdot C \cdot \log n}/_{\varepsilon^2})^p \cdot \mathsf{OPT}_p^p.$$

□

5.2 Cost of Pivot phase

Let $G' = (V', E')$ be the subgraph induced by the vertices that are *not* pre-clustered in Algorithm 1. In this section, we bound the cost of disagreements in G'. Recall that V_0 is the set of eligible vertices, i.e., those vertices $v \in V$ such that $|N_v^+ \cap S_d| \neq \varnothing$. So V' contains the vertices that are *not* eligible (those in $\overline{V_0} = V \backslash V_0$), as well as vertices that are eligible but that are far from all vertices in S_p:

$$V' := \overline{V_0} \cup V_0', \text{ where } V_0' = V_0 \cap \{v \in V : \tilde{d}_{vu_i} > c \cdot r \text{ for all } u_i \in S_p\}.$$

Algorithm 1 runs the standard Pivot algorithm on $G[\overline{V_0}]$, and runs Modified Pivot on $G[V_0']$. In Lemma 9, we bound the disagreements incurred by Pivot on $G[\overline{V_0}]$, and in Lemma 10 we bound the disagreements incurred by Modified Pivot on $G[V_0']$.

We note the arguments in this section may look rather different than those in the Pre-clustering phase. This is a consequence of the fact that we are using totally different clustering subroutines in each phase. In this Pivot phase, we use two versions of the (modified) Pivot algorithm. Therefore, the analysis is more combinatorial; often the charging arguments use "bad triangles" as intermediaries:

Definition 10. *A* bad triangle *is a triple uvw such that $uv, uw \in E^+$ and $vw \in E^-$.*

Note that every clustering must incur a disagreement on at least one edge in a bad triangle.

Definition 11. *In Algorithm 2, for $v_i \in \overline{V_0}$ (analogously, $v_i \in V_0'$), we define v_i's* pivot *to be v_j^* if the* **if** *statement holds, and to be v_i otherwise.*

5.2.1 Disagreements in $G[\overline{V_0}]$

To bound disagreements incident to $u \in \overline{V_0}$, ideally we would relate the set of bad triangles that contain u to the disagreements incident to u in $\mathcal{C}_{\mathsf{ALG}}$. However, while the optimal must make a disagreement on each bad triangle it does not necessarily have any disagreements incident u. So in effect, we have to charge some of our disagreements incident to u to the optimal solution's disagreements on other vertices.

As before, fix an optimal clustering $\mathcal{C}_{\mathsf{OPT}}$ (for the entire graph G) for any fixed ℓ_p-norm. Let $\mathsf{OPT}(u)$ be the (positive or negative) neighbors inducing disagreements with u in $\mathcal{C}_{\mathsf{OPT}}$:

$$\mathsf{OPT}(u) := \{v \in V \mid uv \text{ a disagreement in } \mathcal{C}_{\mathsf{OPT}}\}.$$

Analogously, for $u \in \overline{V_0}$, define $\mathsf{Pivot}(u)$ to be the (positive or negative) neighbors inducing disagreements with u, *restricted to the clusters formed by the Pivot phase of Algorithm 1 on* $G[\overline{V_0}]$:

$$\mathsf{Pivot}(u) := \{v \in \overline{V_0} \mid uv \text{ a disagreement in } \mathcal{C}_{\mathsf{ALG}}\}.$$

Lemma 9. *Condition on the good event B^c and fix $1 \leqslant p < \infty$. The ℓ_p-cost for $u \in \overline{V_0}$ of the edges uv in disagreement with respect $\mathcal{C}_{\mathsf{ALG}}$ for $v \in \overline{V_0}$, is bounded by*

$$\sum_{u \in \overline{V_0}} |\mathsf{Pivot}(u)|^p \leqslant O\left((^1/_{\varepsilon^2} \cdot \log n)^p\right) \cdot \mathsf{OPT}_p^p.$$

Proof of Lemma 9. Let $\mathcal{T}$ denote the set of bad triangles in $G[\overline{V_0}]$, and let $\mathcal{T}(u)$ denote the set of bad triangles in $G[\overline{V_0}]$ that contain vertex u. As mentioned before the start of the proof, to bound disagreements uv where both u and v are in $\overline{V_0}$, we would like to relate $|\mathcal{T}(u)|$ to $|\mathsf{Pivot}(u)|$. However, this is not possible, so instead we charge some of the disagreements incident to u incurred by our Pivot phase to disagreements the optimal solution incurs on other vertices.

Let $\mathcal{T}_{\mathsf{OPT}}(u) \subseteq \mathcal{T}(u)$ be the bad triangles T for which $\mathcal{C}_{\mathsf{OPT}}$ has a disagreement incident to u in T, and let $\overline{\mathcal{T}_{\mathsf{OPT}}(u)} = \mathcal{T}(u) \backslash \mathcal{T}_{\mathsf{OPT}}(u)$ be the remaining bad triangles in $G[\overline{V_0}]$ containing u. We observe that for every $u \in \overline{V_0}$, every disagreement in $G[\overline{V_0}]$ incident to u can be mapped to some $T \in \mathcal{T}(u)$ – namely, the unique bad triangle containing the disagreement and u's pivot. Moreover, this mapping is injective because Pivot incurs *exactly* one disagreement on each bad triangle. Therefore we can bound the disagreements that our algorithm makes in $G[\overline{V_0}]$ that are incident to u by applying Jensen's inequality,

$$\sum_{u \in \overline{V_0}} |\mathsf{Pivot}(u)|^p \leqslant \sum_{u \in \overline{V_0}} |\mathcal{T}(u)|^p \leqslant 2^{p-1} \cdot \underbrace{\sum_{u \in \overline{V_0}} |\mathcal{T}_{\mathsf{OPT}}(u)|^p}_{S_1} + 2^{p-1} \cdot \underbrace{\sum_{u \in \overline{V_0}} \left|\overline{\mathcal{T}_{\mathsf{OPT}}(u)}\right|^p}_{S_2}. \tag{12}$$

Recall that $u \in \overline{V_0}$ because $N_u^+ \cap S_d = \varnothing$. So by the good event, it must be that $|N_u^+| < C \log n / \varepsilon^2$. This bound on $|N_u^+|$ will repeatedly be used.

First we bound S_1, which will be simpler to bound since these triangles directly correspond to a disagreement that $\mathcal{C}_{\mathsf{OPT}}$ has on u.

Claim 15. $\sum_{u \in \overline{V_0}} |\mathcal{T}_{\mathsf{OPT}}(u)|^p \leqslant O\left(({}^{1}\!/\!{}_{\varepsilon^2} \cdot \log n)^p\right) \cdot \mathit{OPT}_p^p$.

Proof of Claim 15. Fix $u \in \overline{V_0}$. We use the definition of $\mathcal{T}_{\mathsf{OPT}}(u)$ to see that

$$|\mathcal{T}_{\mathsf{OPT}}(u)| = \sum_{T \in \mathcal{T}_{\mathsf{OPT}}(u)} 1 \leqslant \sum_{v \in \mathsf{OPT}(u) \cap \overline{V_0}} \sum_{\substack{T \in \mathcal{T}_{\mathsf{OPT}}(u): \\ uv \in T}} 1 \leqslant \sum_{v \in \mathsf{OPT}(u) \cap \overline{V_0}} (|N_u^+ \cap \overline{V_0}| + |N_v^+ \cap \overline{V_0}|).$$

Then we use the bound $|N_u^+ \cap \overline{V_0}| \leqslant |N_u^+| \leqslant C \log n / \varepsilon^2$ for $u \in V_0$ and the fact that $|\mathsf{OPT}(u) \cap \overline{V_0}| \leqslant y(u)$ to upper bound S_1:

$$S_1 = \sum_{u \in \overline{V_0}} |\mathcal{T}_{\mathsf{OPT}}(u)|^p \leqslant \sum_{u \in \overline{V_0}} \left(|\mathsf{OPT}(u) \cap \overline{V_0}| \cdot {}^{(2C \cdot \log n)}\!/\!{}_{\varepsilon^2}\right)^p \leqslant \left({}^{(2C \cdot \log n)}\!/\!{}_{\varepsilon^2}\right)^p \cdot \mathsf{OPT}_p^p.$$

□

It remains to bound S_2, and this sum contains the bad triangles which we will charge to disagreements not incident to u.

Claim 16. $\sum_{u \in \overline{V_0}} \left|\overline{\mathcal{T}_{\mathsf{OPT}}(u)}\right|^p \leqslant O\left(({}^{1}\!/\!{}_{\varepsilon^2} \cdot \log n)^p\right) \cdot \mathit{OPT}_p^p$.

Proof of Claim 16. Fix $u \in \overline{V_0}$. By definition, for every $T \in \overline{\mathcal{T}_{\mathsf{OPT}}(u)}$, $\mathcal{C}_{\mathsf{OPT}}$ has an edge in disagreement on the *unique* edge of T not incident to u. So, no other triangle in $\overline{\mathcal{T}_{\mathsf{OPT}}(u)}$ contains this edge as a disagreement. Moreover, by the definition of a bad triangle, one of the endpoints of this disagreeing edge is a positive neighbor of u in $G[\overline{V_0}]$. So we have by the discussion above that

$$\begin{aligned}
S_2 = \sum_{u \in \overline{V_0}} \left|\overline{\mathcal{T}_{\mathsf{OPT}}(u)}\right|^p &\leqslant \sum_{u \in \overline{V_0}} \left(\sum_{\overline{V_0} \cap N_u^+} |\mathsf{OPT}(v) \cap \overline{V_0}| \right)^p \\
&\leqslant \sum_{u \in \overline{V_0}} |N_u^+|^{p-1} \sum_{v \in \overline{V_0} \cap N_u^+} |\mathsf{OPT}(v) \cap \overline{V_0}|^p \qquad (13) \\
&\leqslant \sum_{v \in \overline{V_0}} |\mathsf{OPT}(v) \cap \overline{V_0}|^p \sum_{u \in \overline{V_0} \cap N_v^+} |N_u^+|^{p-1} \\
&\leqslant ({}^{C}\!/\!{}_{\varepsilon^2} \cdot \log n)^p \cdot \mathsf{OPT}_p^p. \qquad (14)
\end{aligned}$$

Line (13) follows from Jensen's inequality. Line (14) follows from the fact that because we conditioned on the good event B^c, the maximum positive degree of any $u \in \overline{V_0}$ is $\frac{C}{\varepsilon^2} \cdot \log n$ and for $v \in \overline{V_0}$, there are at least $|\mathsf{OPT}(v) \cap \overline{V_0}|$ disagreements incident to v in $\mathcal{C}_{\mathsf{OPT}}$ by definition of $\mathsf{OPT}(v)$. □

The lemma statement follows from the claims, since continuing from line (12), we see that

$$\begin{aligned}\sum_{u \in \overline{V_0}} \left|\mathsf{Pivot}(u)\right|^p &\leqslant 2^{p-1} \cdot ({}^{2C}\!/\varepsilon^2 \cdot \log n)^p \cdot \mathsf{OPT}_p^p + 2^{p-1} \cdot ({}^{C}\!/\varepsilon^2 \cdot \log n)^p \cdot \mathsf{OPT}_p^p \\ &\leqslant ({}^{4C}\!/\varepsilon^2 \cdot \log n)^p \cdot \mathsf{OPT}_p^p.\end{aligned}$$

□

5.2.2 Disagreements in $G[V_0']$

Next we will bound the disagreements in $G[V_0']$. Recall that the vertices in V_0' are those that have sufficiently large positive neighborhood sampled in S_d, but were far from all cluster centers. As has been in the case for other disagreement types, the disagreements whose cost is most difficult to bound are on the positive edges uv, where $\tilde{d}_{uv}$ is quite small. Here, we are able to charge uv to some other edges uw, for uvw a bad triangle.

As before, fix an optimal clustering $\mathcal{C}_{\mathsf{OPT}}$ (for the entire graph G) for any fixed ℓ_p-norm. Let $\mathsf{OPT}(u)$ to be the (positive or negative) disagreements incident to u in $\mathcal{C}_{\mathsf{OPT}}$:

$$\mathsf{OPT}(u) := \{uv \mid uv \text{ a disagreement in } \mathcal{C}_{\mathsf{OPT}}\}.$$

For $u \in V_0'$, define $\mathsf{Pivot}(u)$ to be the (positive or negative) disagreements incident to u, *restricted to the clusters formed by the (Modified) Pivot phase of Algorithm 1 on* $G[V_0']$:

$$\mathsf{Pivot}(u) := \{uv \mid v \in V_0', uv \text{ a disagreement in } \mathcal{C}_{\mathsf{ALG}}\}.$$

Note that both sets are sets of *edges*, unlike in Lemma 9 where the analogous sets are sets of vertices.

We write Pivot for brevity, but recall that the algorithm on $G[V_0']$ is actually a modified version of the classic Pivot algorithm, as the pivots in our algorithm grab positive neighbors that are additionally required to be nearby with respect to $\tilde{d}$ (see the definition of E_c in the **else** statement for $V_0' = V' \backslash \overline{V_0}$ in Algorithm 1).

Lemma 10. *Condition on the good event B^c and fix $1 \leqslant p < \infty$. The ℓ_p-cost for $u \in V_0'$ of the edges uv in disagreement with respect to $\mathcal{C}_{\mathsf{ALG}}$ for $v \in V_0'$ is bounded by*

$$\sum_{u \in V_0'} \left|\mathsf{Pivot}(u)\right|^p \leqslant O\left(\left({}^{1}\!/\varepsilon^8 \cdot \log^4 n\right)^p\right) \cdot \mathsf{OPT}_p^p.$$

Proof of Lemma 10. All disagreements in $\mathsf{Pivot}(u) \cap \mathsf{OPT}(u)$ can be charged directly to $\mathsf{OPT}(u)$, so we focus on $\mathsf{Pivot}(u) \cap \overline{\mathsf{OPT}(u)}$. We partition the remaining disagreements into several sets:

$$\begin{aligned}\mathsf{Pivot}(u) \cap \overline{\mathsf{OPT}(u)} = &\underbrace{\left\{uv \in \mathsf{Pivot}(u) \cap \overline{\mathsf{OPT}(u)} \cap E^+ \mid \tilde{d}_{uv} \geqslant cr\right\}}_{S_1(u)} \\ &\dot\cup \underbrace{\left\{uv \in \mathsf{Pivot}(u) \cap \overline{\mathsf{OPT}(u)} \cap E^+ \mid \tilde{d}_{uv} < cr\right\}}_{S_2(u)} \dot\cup \underbrace{\left(\mathsf{Pivot}(u) \cap \overline{\mathsf{OPT}(u)} \cap E^-\right)}_{S_3(u)}.\end{aligned}$$

So all together, we can apply Jensen's inequality to see that

$$\begin{aligned}\sum_{u \in V_0'} \left|\mathsf{Pivot}(u)\right|^p \leqslant\ & 4^p \cdot \sum_{u \in V_0'} \left|\mathsf{Pivot}(u) \cap \mathsf{OPT}(u)\right|^p \\ &+ 4^p \cdot \underbrace{\sum_{u \in V_0'} \left|S_1(u)\right|^p}_{S_1} + 4^p \cdot \underbrace{\sum_{u \in V_0'} \left|S_2(u)\right|^p}_{S_2} + 4^p \cdot \underbrace{\sum_{u \in V_0'} \left|S_3(u)\right|^p}_{S_3}. \qquad (15)\end{aligned}$$

Bounding S_1. As has been the case for other types of disagreements, bounding the cost of edges that have large $\tilde{d}$ is relatively straightforward.

We charge the cost of all $uv \in \mathsf{Pivot}(u) \cap \overline{\mathsf{OPT}(u)} \cap E^+$ with $\tilde{d}_{uv} \geqslant c \cdot r$ to $\tilde{d}$. Specifically, we use that $V_0' \subseteq V_0$ and $\mathsf{Pivot}(u) \subseteq V_0 \cap N_u^+$, and then apply Lemma 3.

$$
\begin{aligned}
S_1 = \sum_{u \in V_0'} |S_1(u)|^p &\leqslant \frac{1}{(cr)^p} \cdot \sum_{u \in V_0'} \left(\sum_{uv \in \mathsf{Pivot}(u) \cap \overline{\mathsf{OPT}(u)} \cap E^+} \tilde{d}_{uv} \right)^p \\
&\leqslant \frac{1}{(cr)^p} \cdot \sum_{u \in V_0} \left(\sum_{v \in V_0 \cap N_u^+} \tilde{d}_{uv} \right)^p \leqslant \left({}^{(550 \cdot C^3 \cdot \log^3 n)}/_{\varepsilon^6} \right)^p \cdot \mathsf{OPT}_p^p. \qquad (16)
\end{aligned}
$$

Bounding S_2. For positive disagreements uv with small $\tilde{d}_{uv}$, we go through bad triangles to find a suitable edge to charge.

For $uv \in S_2(u)$, we have that $\tilde{d}_{uv} < c \cdot r$. Since $uv \in E^+$, there are only two ways for uv to be in $\mathsf{Pivot}(u)$.

- The first is if uv is involved in a bad triangle with some w which is u's pivot but not v's. In this case, $uw \in E^+$ and $vw \in E^-$. In the second case, uv is involved in a bad triangle with some w which is v's pivot but not u's. In this case, $vw \in E^+$ and $uw \in E^-$. Since uvw is a bad triangle, and $uv \notin \mathsf{OPT}(u)$, we have that either wu or wv is a disagreement in $\mathcal{C}_{\mathsf{OPT}}$. We will charge the cost of the disagreement uv made by $\mathcal{C}_{\mathsf{ALG}}$ to whichever one is a disagreement in $\mathcal{C}_{\mathsf{OPT}}$ (choosing one arbitrarily if both are disagreements).
- The second is if uv is on a triangle uvw of *all* positive edges with some w which is u's pivot but not v's in the case that $\tilde{d}_{wv} \geqslant c \cdot r$ but $\tilde{d}_{wu} < c \cdot r$; or, w which is v's pivot but not u's in the case that $\tilde{d}_{wu} \geqslant c \cdot r$ but $\tilde{d}_{wv} < c \cdot r$. In the former case, we can charge the cost of the disagreement uv made by $\mathcal{C}_{\mathsf{ALG}}$ to wv as $\tilde{d}_{wv} \geqslant c \cdot r$, and in the latter case, to wu as $\tilde{d}_{wu} \geqslant c \cdot r$.

Call the subset of disagreements in $S_2(u)$ satisfying the former case $S_{2a}(u)$ and those satisfying the latter case $S_{2b}(u)$.

We start by bounding the contribution of all $S_{2a}(u)$.

Claim 17. $\sum_{u \in V_0'} |S_{2a}(u)|^p \leqslant O\left(({}^1/_{\varepsilon^2} \cdot \log n)^p\right) \cdot \mathit{OPT}_p^p$.

Proof of Claim 17. For each disagreement in $\mathcal{C}_{\mathsf{OPT}}$, we have to bound how many times it can be charged by an edge uv in $S_{2a}(u)$. By the discussion above, a disagreement in $\mathcal{C}_{\mathsf{OPT}}$ can only be charged by uv if it is of the form uw or wv. If it is of the form uw, then it can only be charged by uv with $v \in N_u^+ \cap \mathrm{Ball}_{\tilde{d}}(u, c \cdot r)$. If it is of the form wv, then it can only be charged by once, namely, by uv where again $v \in N_u^+ \cap \mathrm{Ball}_{\tilde{d}}(u, c \cdot r)$. So

$$
|S_{2a}(u)| \leqslant |\mathsf{OPT}(u)| \cdot |N_u^+ \cap \mathrm{Ball}_{\tilde{d}}(u, c \cdot r)| + \sum_{v \in N_u^+ \cap \mathrm{Ball}_{\tilde{d}}(u, c \cdot r) \cap V_0'} |\mathsf{OPT}(v)|
$$

and so by Jensen's inequality,

$$
\begin{aligned}
\sum_{u \in V_0'} |S_{2a}(u)|^p \leqslant 2^{p-1} \cdot &\sum_{u \in V_0'} |\mathsf{OPT}(u)|^p \cdot |N_u^+ \cap \mathrm{Ball}_{\tilde{d}}(u, c \cdot r)|^p \\
&+ 2^{p-1} \cdot \sum_{u \in V_0'} \left(\sum_{v \in N_u^+ \cap \mathrm{Ball}_{\tilde{d}}(u, c \cdot r) \cap V_0'} |\mathsf{OPT}(v)| \right)^p.
\end{aligned}
$$

The first sum is bounded by $(C' \cdot \log_{1/(1-\varepsilon^2/2)} n)^p \cdot \mathsf{OPT}_p^p$. This is because $u \in V_0'$, which implies that $\mathrm{Ball}_{\tilde{d}}^{S_p}(u, c \cdot r) = \varnothing$, which in turn implies that $|\mathrm{Ball}_{\tilde{d}}(u, c \cdot r)| \leqslant C' \cdot \log_{1/(1-\varepsilon^2/2)} n$, since we have conditioned on the good event B^c. To bound the second sum, we apply Jensen's inequality and then flip it:

$$\sum_{u\in V_0'}\left(\sum_{v\in N_u^+\cap \mathrm{Ball}_{\tilde d}(u,c\cdot r)}|\mathsf{OPT}(v)|\right)^p \leqslant \sum_{u\in V_0'}|N_u^+\cap \mathrm{Ball}_{\tilde d}(u,c\cdot r)|^{p-1}\sum_{v\in N_u^+\cap \mathrm{Ball}_{\tilde d}(u,\cdot r)}|\mathsf{OPT}(v)|^p$$
$$\leqslant \sum_{v\in V_0'}|\mathsf{OPT}(v)|^p\sum_{u\in V_0'\cap \mathrm{Ball}_{\tilde d}(v,\cdot r)}|\mathrm{Ball}_{\tilde d}(u,c\cdot r)|^{p-1}$$
$$\leqslant (C'\cdot \log_{1/(1-\varepsilon^2/2)} n)^p\cdot \mathsf{OPT}_p^p$$

where we have used the same reasoning as above to bound the inner sum by $(C'\cdot\log_{1/(1-\varepsilon^2/2)} n)^p$, due to the good event B^c.

Combining the bounds on the first and second sums, and doing a change of base finishes the claim with $\sum_{u\in V_0'}|S_{2a}(u)|^p \leqslant ({(4\cdot C'\cdot\log n)}/{\varepsilon^2})^p\cdot \mathsf{OPT}_p^p$.

□

We now bound the contribution of all $S_{2b}(u)$, so that we may then finish off the bound of S_2.

Claim 18. $\sum_{u\in V_0'}|S_{2b}(u)|^p \leqslant O\left(\left({1}/{\varepsilon^8}\cdot\log^4 n\right)^p\right)\cdot \mathit{OPT}_p^p$.

Proof of Claim 18. For each disagreement in $\mathcal{C}_{\mathsf{OPT}}$, we have to bound how many times it can be charged by an edge uv in $S_{2b}(u)$. By the discussion above, a disagreement in $\mathcal{C}_{\mathsf{OPT}}$ can only be charged by uv if it is of the form uw or wv. If it is of the form uw, then it can only be charged by uv with $v\in N_u^+\cap \mathrm{Ball}_{\tilde d}(u,c\cdot r)$. If it is of the form wv, then it can only be charged by once, namely, by uv where again $v\in N_u^+\cap \mathrm{Ball}_{\tilde d}(u,c\cdot r)$. So

$$|S_{2b}(u)|\leqslant \sum_{\substack{w\in N_u^+\cap V_0':\\ \tilde d_{uw}\geqslant c\cdot r}}|N_u^+\cap \mathrm{Ball}_{\tilde d}(u,c\cdot r)| + \sum_{v\in N_u^+\cap \mathrm{Ball}_{\tilde d}(u,c\cdot r)\cap V_0'}|\{w\in N_v^+\cap V_0' : \tilde d_{wv}\geqslant c\cdot r\}|$$

and so by Jensen's inequality,

$$\sum_{u\in V_0'}|S_{2b}(u)|^p\leqslant 2^{p-1}\cdot\sum_{u\in V_0'}\left(\sum_{\substack{w\in N_u^+\cap V_0':\\ \tilde d_{uw}\geqslant c\cdot r}}|N_u^+\cap \mathrm{Ball}_{\tilde d}(u,c\cdot r)|\right)^p$$
$$+\,2^{p-1}\cdot\sum_{u\in V_0'}\left(\sum_{v\in N_u^+\cap \mathrm{Ball}_{\tilde d}(u,c\cdot r)\cap V_0'}|\{w\in N_v^+\cap V_0' : \tilde d_{wv}\geqslant c\cdot r\}|\right)^p$$

The first sum is bounded by

$$\sum_{u\in V_0'}|N_u^+\cap \mathrm{Ball}_{\tilde d}(u,c\cdot r)|^p\cdot\left(\sum_{w\in N_u^+\cap V_0'}{1}/{c\cdot r}\cdot\tilde d_{uw}\right)^p\leqslant \left({(2C'\cdot\log n)}/{(c\cdot r\cdot\varepsilon^2)}\right)^p\cdot\sum_{u\in V_0}\left(\sum_{w\in N_u^+\cap V_0}\tilde d_{uw}\right)^p$$

where again we have used the conditioning on B^c and the fact that $u\in V_0'$ to bound $|\mathrm{Ball}_{\tilde d}(u,c\cdot r)|$. The second sum is bounded by

$$\sum_{u\in V_0'}\left(\sum_{v\in N_u^+\cap \mathrm{Ball}_{\tilde d}(u,c\cdot r)\cap V_0'}|\{w\in N_v^+\cap V_0' : \tilde d_{wv}\geqslant c\cdot r\}|\right)^p$$
$$\leqslant \sum_{u\in V_0'}|\mathrm{Ball}_{\tilde d}(u,c\cdot r)|^{p-1}\sum_{v\in N_u^+\cap \mathrm{Ball}_{\tilde d}(u,c\cdot r)\cap V_0'}\left(\sum_{w\in N_v^+\cap V_0'}\frac{1}{c\cdot r}\cdot\tilde d_{wv}\right)^p$$

$$\leqslant \sum_{v\in V_0'}\left(\sum_{w\in N_v^+\cap V_0'}\frac{1}{c\cdot r}\cdot\tilde{d}_{wv}\right)^p \sum_{u\in V_0'\cap \mathrm{Ball}_{\tilde{d}}(v,c\cdot r)} |\mathrm{Ball}_{\tilde{d}}(u,c\cdot r)|^{p-1}$$

$$\leqslant \left({}^{(C'\cdot \log_{1/(1-\varepsilon^2/2)} n)}\!/\!_{c\cdot r}\right)^p\cdot \sum_{u\in V_0}\left(\sum_{w\in N_u^+\cap V_0}\tilde{d}_{uw}\right)^p.$$

Combining the bounds on the first and second sums, and applying Lemma 3, finishes the claim with $\sum_{u\in V_0'}|S_{2b}(u)|^p \leqslant \left({}^{(440\cdot C'\cdot C^3\cdot \log^4 n)}\!/\!_{\varepsilon^8}\right)^p\cdot \mathsf{OPT}_p^p$. □

Tying it all together using Claims 17 and 18, and using that C is sufficiently large to ensure the good event, we have

$$S_2 =\leqslant 2^{p-1}\cdot\sum_{u\in V_0'}|S_{2a}(u)|^p + 2^{p-1}\cdot\sum_{u\in V_0'}|S_{2b}(u)|^p \leqslant \left({}^{(882\cdot C'\cdot C^3\cdot \log^4 n)}\!/\!_{\varepsilon^8}\right)^p\cdot \mathsf{OPT}_p^p. \tag{17}$$

Bounding S_3. It remains to bound the cost of the negative edges. Fix $uv \in S_3(u)$. Then the only way u and v can be clustered together is if u and v are clustered by the same pivot w, where $w \in N_u^+ \cap N_v^+$ and $\tilde{d}_{uw}, \tilde{d}_{vw} \leqslant c\cdot r$. But then $1-\tilde{d}_{uv} \geqslant 1-2\delta cr$, so we have

$$S_3 = \sum_{u\in V_0'}|S_3(u)|^p \leqslant \left(\frac{1}{1-2\delta cr}\right)^p\cdot\sum_{u\in V_0}\left(\sum_{v\in N_u^-\cap V_0}(1-\tilde{d}_{uv})\right)^p \leqslant \left({}^{(296\cdot C\cdot\log n)}\!/\!_{\varepsilon^2}\right)^p\cdot \mathsf{OPT}_p^p, \tag{18}$$

where the last equality is by Lemma 4 and subbing the values for δ, c, r chosen in Section 2.

Continuing from line (15), and substituting in from lines (16), (17), and (18), we conclude the proof

$$\sum_{u\in V_0'}|\mathsf{Pivot}(u)|^p \leqslant \left({}^{(5728\cdot C'\cdot C^3\cdot\log^4 n)}\!/\!_{\varepsilon^8}\right)^p\cdot \mathsf{OPT}_p^p.$$

□

5.2.3 Disagreements between $G[\overline{V_0}]$ and $G[V_0']$

The only disagreements occurring on edges going between V_0' and $\overline{V_0}$ are from positive edges. We bound the cost of disagreements uv incident to $u \in \overline{V_0}$, for $v \in V_0 \cap N_u^+$, in Lemma 11, then we bound the cost of disagreements uv incident to $u \in V_0$, for $v \in \overline{V_0} \cap N_u^+$, in Lemma 12. Since $V_0' \subseteq V_0$, these lemmas immediately bound the cost of disagreements between $G[V_0']$ and $G[\overline{V_0}]$.

5.3 Cost between the Pre-clustering phase and the Pivot phase

It remains to bound the cost of edges that go between the Pre-clustering and Pivot phases. The disagreements uv incident to u can take several forms: u is pre-clustered and $v \in V_0'$, $u \in V_0'$ and v is pre-clustered, $u \in \overline{V_0}$ and v is pre-clustered, u is pre-clustered and $v \in \overline{V_0}$. We discuss each disagreement type in order of the above list.

For u pre-clustered and $v \in V_0'$, both u and v are in V_0, but $u \succ v$; recall we use the notation $u \succ v$ to mean u is clustered before v, or in other words u is either pre-clustered to a higher ordered center than v, or u is pre-clustered and v is not. These disagreements are already accounted for in Lemma 7.

Similarly, $u \in V_0'$ and v pre-clustered, both nodes are in V_0 again. Though this time, $v \succ u$. These disagreements are already accounted for in Lemma 6.

The cost of the next type of disagreement, when $u \in \overline{V_0}$ and v is pre-clustered, will be bounded in Lemma 11. Then the cost of disagreements where u is pre-clustered (thus $u \in V_0$) and $v \in \overline{V_0}$ will be bounded in Lemma 12.

Lemma 11. *Condition on the good event B^c and fix $1 \leqslant p < \infty$. The ℓ_p-cost for $u \in \overline{V_0}$ of edges uv in disagreement with respect to $\mathcal{C}_{ALG}$, for $v \in V_0$, is bounded by*

$$\sum_{u \in \overline{V_0}} |V_0 \cap N_u^+|^p \leqslant O\left(\left(1/\varepsilon^4 \cdot \log^2 n\right)^p\right) \cdot \mathit{OPT}_p^p.$$

As two consequences, we have that $\sum_{u \in \overline{V_0}} |\{v \in N_u^+ \mid v \text{ pre-clustered}\}|^p \leqslant O\left(\left(1/\varepsilon^4 \cdot \log^2 n\right)^p\right) \cdot \mathit{OPT}_p^p$ and $\sum_{u \in \overline{V_0}} |\{v \in N_u^+ \cap V_0'\}|^p \leqslant O\left(\left(1/\varepsilon^4 \cdot \log^2 n\right)^p\right) \cdot \mathit{OPT}_p^p$.

Proof of Lemma 11. It suffices to prove the first bound, since if v is pre-clustered then $v \in V_0$, and if $v \in V_0'$ then $v \in V_0$.

Define $D^+(u)$ to be the fractional cost of the positive edges incident to u with respect to the (actual) correlation metric d, that is, $D^+(u) := \sum_{v \in N_u^+} d_{uv}$. We partition the sum based on how large $D^+(u)$ is, and see that

$$\sum_{u \in \overline{V_0}} |V_0 \cap N_u^+|^p \leqslant \sum_{u \in \overline{V_0}:\ D^+(u) > \frac{\varepsilon^2}{2C \log n}} |N_u^+|^p + \sum_{u \in \overline{V_0}:\ D^+(u) < \frac{\varepsilon^2}{2C \log n}} \left(\sum_{v \in N_u^+ \cap V_0} 1 \right)^p. \tag{19}$$

Bounding the first term of Equation (19) is straightforward. Since we condition on the good event B^c, we know that $u \in \overline{V_0}$ implies $|N_u^+| \leqslant C \cdot \log n/\varepsilon^2$. Recalling that $D^* = (D^*(u))_{u \in V}$ is the fractional cost vector for the adjusted correlation metric, the following holds:

$$\begin{aligned}\sum_{u \in \overline{V_0}:\ D^+(u) > \varepsilon^2/2C \log n} |N_u^+|^p &\leqslant \left(2C \log n/\varepsilon^2\right)^p \cdot \sum_{u \in \overline{V_0}} |N_u^+|^p \cdot (D^+(u))^p \leqslant \left(2C \log n/\varepsilon^2\right)^{2p} \cdot \sum_{u \in \overline{V_0}} (D^+(u))^p \\ &\leqslant \left(2C \log n/\varepsilon^2\right)^{2p} \cdot \sum_{u \in \overline{V_0}} (D^*(u))^p \leqslant \left((4C^2 \cdot M \cdot \log^2 n)/\varepsilon^4\right)^p \cdot \mathsf{OPT}_p^p.\end{aligned}$$

In the first inequality on the last line, we have used the fact that $d_{uv} \leqslant d^*_{uv}$ for all $uv \in E^+$, and the last inequality uses the bound on $||D^*||_p$ from Lemma 5, for M the constant in Lemma 5.

Now consider the second term in Equation (19). We show this term is 0 since we conditioned on B^c.

Assume for sake of deriving a contradiction there exists $u \in \overline{V_0}$ with $D^+(u) \leqslant \varepsilon^2/(2C \log n)$ whose positive neighborhood does *not* form a perfect clique (i.e., u is incident to at least one bad triangle). We know u has one neighbor $v \in N_u^+$, where $N_u^+ \neq N_v^+$. Then either $N_v^+ \subset N_u^+$, in which case

$$d_{uv} = 1 - \frac{|N_u^+ \cap N_v^+|}{|N_u^+ \cup N_v^+|} \geqslant 1 - \frac{|N_u^+| - 1}{|N_u^+|} \geqslant 1 - \frac{\frac{C \log n}{\varepsilon^2} - 1}{\frac{C \log n}{\varepsilon^2}} \geqslant \frac{\varepsilon^2}{2C \log n},$$

or $|N_v^+ \backslash N_u^+| > 1$, in which case,

$$d_{uv} = 1 - \frac{|N_u^+ \cap N_v^+|}{|N_u^+ \cup N_v^+|} \geqslant 1 - \frac{|N_u^+|}{|N_u^+| + 1} \geqslant 1 - \frac{\frac{C \log n}{\varepsilon^2}}{\frac{C \log n}{\varepsilon^2} + 1} \geqslant \frac{\varepsilon^2}{2C \log n}.$$

Both equations use that $|N_u^+| < \frac{C \log n}{\varepsilon^2}$, which must be the case since $u \in \overline{V_0}$ and we conditioned on the good event. So no such v can exist (regardless of whether or not it is in $N_u^+ \cap V_0$), we as have contradicted the fact that $D^+(u) \leqslant \varepsilon^2/(2C \log n)$.

Therefore, we we may assume u is part of a perfect clique. So for $w \in N_v^+$ and $v \in N_u^+$, we have $w \in N_u^+$. If this is the case, then $v \notin V_0$ for each $v \in N_u^+$. This is because if $N_v^+ \cap S_d \neq \varnothing$, then $N_u^+ \cap S_d \neq \varnothing$, since $N_u^+ = N_v^+$. Thus, none of u's positive neighbors v will be in V_0, and u will not contribute to the second sum in Equation (19).

□

Lemma 12. *Condition on the good event B^c and fix $1 \leqslant p < \infty$. The ℓ_p-cost for $u \in V_0$ of edges uv in disagreement with respect to $\mathcal{C}_{\mathsf{ALG}}$, for $v \in \overline{V_0}$, is bounded by*

$$\sum_{u \in V_0} |\overline{V_0} \cap N_u^+|^p \leqslant O\left(\left(1/\varepsilon^4 \cdot \log^2 n\right)^p\right) \cdot \mathit{OPT}_p^p.$$

As two consequences, we have that $\sum_{u \text{ pre-clustered}} |\{v \in N_u^+ \cap \overline{V_0}\}|^p \leqslant O\left(\left(1/\varepsilon^4 \cdot \log^2 n\right)^p\right) \cdot \mathit{OPT}_p^p$ and $\sum_{u \in V_0'} |\{v \in N_u^+ \cap \overline{V_0}\}|^p \leqslant O\left(\left(1/\varepsilon^4 \cdot \log^2 n\right)^p\right) \cdot \mathit{OPT}_p^p$.

Proof of Lemma 12. It suffices to prove the first bound, since if u is pre-clustered then $u \in V_0$, and if $u \in V_0'$ then $u \in V_0$.

Recall $D^+(u)$ is the fractional cost of the positive edges incident to u with respect to the (actual) correlation metric d, that is, $D^+(u) := \sum_{v \in N_u^+} d_{uv}$. Since $u \in V_0$ and we conditioned on the good event, we know that $|N_u^+| \geqslant C \cdot \log n/\varepsilon^2$, but we further partition $u \in V_0$ based on whether $|N_u^+| \geqslant 2C \cdot \log n/\varepsilon^2$, and on how large $D^+(u)$ is

$$\begin{aligned}\sum_{u \in V_0} |\overline{V_0} \cap N_u^+|^p \leqslant & \underbrace{\sum_{u \in V_0 \colon |N_u^+| \geqslant \frac{2C \cdot \log n}{\varepsilon^2}} |N_u^+|^p}_{S_1} + \\ & \underbrace{\sum_{\substack{u \in V_0 \colon |N_u^+| < \frac{2C \cdot \log n}{\varepsilon^2}, \\ D^+(u) > \frac{\varepsilon^2}{4C \log n}}} \left|N_u^+ \cap \overline{V_0}\right|^p}_{S_2} + \underbrace{\sum_{\substack{u \in V_0 \colon |N_u^+| < \frac{2C \cdot \log n}{\varepsilon^2}, \\ D^+(u) \leqslant \frac{\varepsilon^2}{4C \log n}}} \left|N_u^+ \cap \overline{V_0}\right|^p}_{S_3}. \end{aligned} \tag{20}$$

Bounding S_1. Since we conditioned on the good event, for $v \in \overline{V_0}$ we have $|N_v^+| < C \log n/\varepsilon^2$. Therefore we find that $u \in V_0$ and $v \in \overline{V_0} \cap N_u^+$ are quite far apart, with

$$d_{uv} = 1 - \frac{|N_u^+ \cap N_v^+|}{|N_u^+ \cup N_v^+|} \geqslant 1 - \frac{\frac{C}{\epsilon^2} \log n}{\frac{2C}{\epsilon^2} \log n} \geqslant \frac{1}{2},$$

where we have used the fact that each u included in the sum for S_1 has $|N_u^+| \geqslant 2C \log n/\varepsilon^2$. We can then charge the disagreement directly to d_{uv}, since it is sufficiently large:

$$\begin{aligned}\sum_{u \in V_0 \colon |N_u^+| \geqslant \frac{2C}{\varepsilon^2 \log n}} |\overline{V_0} \cap N_u^+|^p &= \sum_{u \in V_0 \colon |N_u^+| \geqslant \frac{2C}{\varepsilon^2 \log n}} \left(\sum_{v \in \overline{V_0} \cap N_u^+} 1\right)^p \\ &\leqslant 2^p \cdot \sum_{u \in V_0} \left(D^+(u)\right)^p \leqslant 2^p \cdot \sum_{u \in V_0} \left(D^*(u)\right)^p \leqslant (2 \cdot M)^p \cdot \mathsf{OPT}_p^p\end{aligned}$$

In the last line, we used the fact that $d_{uv} \leqslant d^*_{uv}$ for all $uv \in E^+$, and then use the bound on $||D^*||_p$ from Lemma 5 (M is the constant from Lemma 5).

Bounding S_2. Bounding S_2 is almost identical to bounding the first sum in Equation (19). This is because u being ineligible implies $|N_u^+|$ is small by B^c, and here the u in the sum have small $|N_u^+|$ too.

$$\begin{aligned}\sum_{\substack{u \in V_0 \colon |N_u^+| < \frac{2C}{\varepsilon^2 \log n}, \\ D^+(u) > \varepsilon^2/4C \log n}} \left|N_u^+ \cap \overline{V_0}\right|^p &\leqslant \left(4C \log n/\varepsilon^2\right)^p \cdot \sum_{u \in V_0} |N_u^+|^p \cdot (D^+(u))^p \\ &\leqslant \left(4C \log n/\varepsilon^2\right)^{2p} \cdot \sum_{u \in V_0} (D^+(u))^p \leqslant \left((16 \cdot C^2 \cdot M \cdot \log^2 n)/\varepsilon^4\right)^p \cdot \mathsf{OPT}_p^p,\end{aligned}$$

where in the last inequality, we used that $d_{uv} \leqslant d^*_{uv}$ for all $uv \in E^+$, then apply Lemma 5.

Bounding S_3 We show $S_3 = 0$, since we conditioned on B^c.

The proof follows exactly as that for bounding the second sum in Equation (19), just replacing $C \log n/\varepsilon^2$ with $2C \log n/\varepsilon^2$.

Summing together S_1, S_2 and S_3, we find that

$$\sum_{u \in V_0} |\overline{V_0} \cap N_u^+|^p \leqslant \left({}^{(18 \cdot C^2 \cdot M \cdot \log^2 n)}/_{\varepsilon^4}\right)^p \cdot \mathsf{OPT}_p^p.$$

□

5.4 Proof of item 1 for Theorem 1

We are ready to combine the results proven so far in this section.

Proof of item 1 for Theorem 1. Let $\mathcal{C}_{\mathsf{ALG}}$ be the clustering output by Algorithm 1, and let $\text{cost}_p(\mathcal{C}_{\mathsf{ALG}})$ be the ℓ_p-norm of the disagreement vector of $\mathcal{C}_{\mathsf{ALG}}$. We further partition the edges in disagreement based on whether they are positive or negative, which phase in Algorithm 1 they are clustered in, and (if at least one endpoint of an edge is pre-clustered) whether or not $u \succ v$. These cases are exhaustive; see Figure 1. Combining the terms from Lemmas 6, 7, 8, 9,10, 11, and 12, and then applying Jensen's inequality and taking the p^{th} root, we see that with high probability (as we recall the good event B^c occurs with high probability)

$$||y_{\mathcal{C}_{\mathsf{ALG}}}||_p \leqslant O\left({}^1/_{\varepsilon^8} \cdot \log^4 n\right) \cdot \mathsf{OPT}_p.$$

□

6 Conclusion

We develop an algorithm for online correlation clustering which, given a sample of ε-fraction of the nodes from the underlying instance, returns a clustering that is simultaneously $O({}^1/_{\varepsilon^6})$-competitive for the ℓ_1-norm objective in expectation and $O({}^{\log n}/_{\varepsilon^6})$-competitive for the ℓ_∞-norm objective with high probability. This is the first positive result for the ℓ_∞-norm in the online setting. We also prove lower bounds that match our upper bounds up to constants and powers of ${}^1/_\varepsilon$ for either norm. Finally, we show that our algorithm is also $O({}^{\log^4 n}/_{\varepsilon^8})$-competitive for each finite ℓ_p-norm with high probability. Thus, we successfully translate the all-norms result of [13] to the online setting.

Our work highlights two key insights. First, it demonstrates the robustness of the adjusted correlation metric: even an estimated version suffices to guide near-optimal decisions in the AOS model. Second, it identifies structural properties that make problems amenable to this online model. Specifically, the ability to estimate key quantities from a small but uniformly sampled subset of the input is crucial for solving problems in the AOS model.

Overall, our results suggest that the AOS model is a promising framework for problems where limited but well-distributed information allows for effective decision-making. In particular, ℓ_∞-norm clustering is an example of a problem when the AOS model is much stronger than the popular RO model. We remark that the model is still relatively new, and we believe the techniques from this paper can be of use to understand a wider range of problems in this setting. For instance, our idea of leveraging different subsamples independently to help mitigate correlation effects across estimating different quantities may be useful. It is of further interest to determine which problems with strong lower bounds on the competitive ratio in the strictly online setting and/or the random-order (RO) model admit small competitive ratios in the AOS model.

References

[1] N. Ailon, M. Charikar, and A. Newman. "Aggregating inconsistent information: ranking and clustering". In: *Journal of the ACM (JACM)* 55.5 (2008), pp. 1–27.

[2] C. Argue, A. Frieze, A. Gupta, and C. Seiler. “Learning from a sample in online algorithms”. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 13852–13863.

[3] E. Balkanski, J. Chatzitheodorou, and A. Maggiori. “Cost-Free Fairness in Online Correlation Clustering”. In: *36th International Conference on Algorithmic Learning Theory*. 2025.

[4] N. Bansal, A. Blum, and S. Chawla. “Correlation clustering”. In: *Machine Learning* 56.1 (2004), pp. 89–113.

[5] N. Cao, S. Li, and J. Ye. “Simultaneously Approximating All Norms for Massively Parallel Correlation Clustering”. In: *arXiv preprint arXiv:2410.09321* (2024).

[6] N. Cao et al. “Understanding the Cluster Linear Program for Correlation Clustering”. In: *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*. 2024, pp. 1605–1616.

[7] M. Charikar, N. Gupta, and R. Schwartz. “Local guarantees in graph cuts and clustering”. In: *International Conference on Integer Programming and Combinatorial Optimization*. Springer. 2017, pp. 136–147.

[8] S. Chawla, K. Makarychev, T. Schramm, and G. Yaroslavtsev. “Near Optimal LP Rounding Algorithm for Correlation Clustering on Complete and Complete k-partite Graphs”. In: *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*. 2015, 219–228.

[9] V. Cohen-Addad, S. Lattanzi, A. Maggiori, and N. Parotsidis. “Online and consistent correlation clustering”. In: *International Conference on Machine Learning*. PMLR. 2022, pp. 4157–4179.

[10] V. Cohen-Addad, E. Lee, and A. Newman. “Correlation Clustering with Sherali-Adams”. In: *Symposium on Foundations of Computer Science (FOCS).* (2022).

[11] S. Davies, B. Moseley, and H. Newman. “Fast Combinatorial Algorithms for Min Max Correlation Clustering”. In: *International Conference on Machine Learning*. PMLR. 2023.

[12] S. Davies, B. Moseley, and H. Newman. “Simultaneously Approximating All ℓ_p-norms in Correlation Clustering”. In: *arXiv preprint arXiv:2308.01534* (2023).

[13] S. Davies, B. Moseley, and H. Newman. “Simultaneously Approximating All lp-Norms in Correlation Clustering”. In: *51st International Colloquium on Automata, Languages, and Programming (ICALP 2024)*. Schloss Dagstuhl–Leibniz-Zentrum für Informatik. 2024.

[14] H. Fichtenberger, S. Lattanzi, A. Norouzi-Fard, and O. Svensson. “Consistent k-clustering for general metrics”. In: *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM. 2021, pp. 2660–2678.

[15] X. Guo, J. Kulkarni, S. Li, and J. Xian. “Consistent k-median: Simpler, better and robust”. In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2021, pp. 1135–1143.

[16] A. Gupta, G. Kehne, and R. Levin. “Set covering with our eyes wide shut”. In: *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM. 2024, pp. 4530–4553.

[17] H. S. Heidrich, J. Irmai, and B. Andres. "A 4-Approximation Algorithm for Min Max Correlation Clustering." In: *AISTATS*. 2024, pp. 1945–1953.

[18] M. R. K. Jaghargh, A. Krause, S. Lattanzi, and S. Vassilvtiskii. "Consistent online optimization: Convex and submodular". In: *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR. 2019, pp. 2241–2250.

[19] S. Kalhan, K. Makarychev, and T. Zhou. "Correlation clustering with local objectives". In: *Advances in Neural Information Processing Systems* 32 (2019).

[20] H. Kaplan, D. Naori, and D. Raz. "Competitive analysis with a sample and the secretary problem". In: *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM. 2020, pp. 2082–2095.

[21] H. Kaplan, D. Naori, and D. Raz. "Online weighted matching with a sample". In: *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM. 2022, pp. 1247–1272.

[22] S. Lattanzi, B. Moseley, S. Vassilvitskii, Y. Wang, and R. Zhou. "Robust online correlation clustering". In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 4688–4698.

[23] S. Lattanzi and S. Vassilvitskii. "Consistent k-clustering". In: *International Conference on Machine Learning*. PMLR. 2017, pp. 1975–1984.

[24] C. Mathieu, O. Sankur, and W. Schudy. "Online Correlation Clustering". In: *27th International Symposium on Theoretical Aspects of Computer Science, STACS 2010, March 4-6, 2010, Nancy, France*. Ed. by J. Marion and T. Schwentick. Vol. 5. LIPIcs. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2010, pp. 573–584. URL: `https://doi.org/10.4230/LIPIcs.STACS.2010.2486`.

[25] G. J. Puleo and O. Milenkovic. "Correlation clustering with constrained cluster sizes and extended weights bounds". In: *SIAM Journal on Optimization* 25.3 (2015), pp. 1857–1872.

[26] A. Schild, E. Vee, M. Purohit, R. K. Ravikumar, and Z. Svitkina. "Semi-online bipartite matching". In: *Proc. 10th Innovations in Theoretical Computer Science Conference*. 2019.

A Omitted Proofs and Constructions

This section contains proofs and constructions that were omitted from the main body.

A.1 Subsampling to simulate several independent samples

In Lemma 13, we show how to construct four independent subsamples S_p, S_d, S_b, S_r from the given sample S. The idea is to independently add each $v \in S$ to some subset of $\{S_p, S_d, S_b, S_r\}$, where the probability for each possible subset of subsamples is recursively defined so as to ensure the required independence properties.

Lemma 13. *For a fixed $0 < \varepsilon < 1$, let $S \subseteq V$ independently contain each element of V with probability ε. From S, we construct 4 samples S_p, S_d, S_b, and S_r such that:*

1. *The probability of an element being in any one of the 4 samples is $\varepsilon^2/2$, i.e., for any $S_i \in \{S_p, S_d, S_b, S_r\}$ and any $v \in V$, we have that $\mathbb{P}(v \in S_i) = \varepsilon^2/2$.*

2. *(Independence between elements) For any $\mathcal{S} \subseteq \{S_p, S_d, S_b, S_r\}$, the events $\{v \in \bigcap_{S_i \in \mathcal{S}} S_i\}_{v \in V}$ are mutually independent. That is, $\mathbb{P}(\bigcap_{v \in V'} \{v \in \bigcap_{S_i \in \mathcal{S}} S_i\}) = \prod_{v \in V'} \mathbb{P}(v \in \bigcap_{S_i \in \mathcal{S}} S_i)$, for any $V' \subseteq V$.*
3. *(Independence between samples) For any $v \in S$, and for any $\mathcal{S} \subseteq \{S_p, S_d, S_b, S_r\}$, the events $\{v \in S_i\}_{S_i \in \mathcal{S}}$ are mutually independent. That is, $\mathbb{P}\left(\bigcap_{S_i \in \mathcal{S}} \{v \in S_i\}\right) = \prod_{S_i \in \mathcal{S}} \mathbb{P}(v \in S_i)$.*

Proof of Lemma 13. We form the samples with the following procedure. Fix $x > 0$ to be specified as a function of ε later; it may be helpful to know that we will eventually show that $x = \mathbb{P}[v \in S_d \mid v \in S]$, where one could replace S_d here with S_b, S_p, or S_r. Consider each element $v \in S$ independently:

- Add v to only S_p (or S_d, S_b, or S_r) and to no other sample with probability $x - 3\varepsilon x^2 + 3\varepsilon^2 x^3 - \varepsilon^3 x^4$. Thus the probability that v is in exactly one sample is $4(x - 3\varepsilon x^2 + 3\varepsilon^2 x^3 - \varepsilon^3 x^4)$.
- Add v to exactly S_p and S_d (or S_p and S_b, S_p and S_r, S_d and S_b, S_d and S_r, or S_b and S_r) with probability $\varepsilon x^2 - 2\varepsilon^2 x^3 + \varepsilon^3 x^4$. Thus the probability that v is in exactly two samples is $6(\varepsilon x^2 - 2\varepsilon^2 x^3 + \varepsilon^3 x^4)$.
- Add v to exactly S_p and S_d and S_r (or S_p and S_d and S_b, S_p and S_r and S_b, or S_d and S_r and S_b) with probability $\varepsilon^2 x^3 - \varepsilon^3 x^4$. Thus the probability that v is in exactly three samples is $4(\varepsilon^2 x^3 - \varepsilon^3 x^4)$.
- Add v to all 4 samples with probability $\varepsilon^3 x^4$.
- Add v to none of the samples with probability

$$\begin{aligned}
&1 - \varepsilon^3 x^4 - 4(\varepsilon^2 x^3 - \varepsilon^3 x^4) - 6(\varepsilon x^2 - 2\varepsilon^2 x^3 + \varepsilon^3 x^4) - 4(x - 3\varepsilon x^2 + 3\varepsilon^2 x^3 - \varepsilon^3 x^4) \\
&= 1 - 4x + 6\varepsilon x^2 - 4\varepsilon^2 x^3 + \varepsilon^3 x^4.
\end{aligned}$$

The above is a probability distribution so long as x is chosen small enough relative to ε so that all probabilities above are between 0 and 1. For instance, the choice of $x = \varepsilon/2$ works. Further, since each v is sampled into S independently and is then considered independently in deciding where to allocate it among the subsamples in $\{S_p, S_d, S_b, S_r\}$, property (2) in the lemma statement holds by construction.

For a fixed $v \in S$, the probability that $v \in S_p \cap S_d \cap S_b$ is the probability that v is in S_p, S_d, S_b and not S_r plus the probability that v is in all 4 subsamples:

$$\mathbb{P}[v \in S_p \cap S_d \cap S_b \mid v \in S] = \varepsilon^2 x^3 - \varepsilon^3 x^4 + \varepsilon^3 x^4 = \varepsilon^2 x^3. \tag{21}$$

This is symmetric for all triples of subsamples.

Then, for a fixed $v \in S$, the probability that $v \in S_p \cap S_d$ is the probability that v is in S_p and S_d but neither S_b nor S_r, plus the probability that v is in S_p, S_d and exactly one of S_b or S_r, plus the probability that v is in all 4 subsamples:

$$\mathbb{P}[v \in S_p \cap S_d \mid v \in S] = \varepsilon x^2 - 2\varepsilon^2 x^3 + \varepsilon^3 x^4 + 2(\varepsilon^2 x^3 - \varepsilon^3 x^4) + \varepsilon^3 x^4 = \varepsilon x^2. \tag{22}$$

This is symmetric for all pairs of subsamples.

Then, for a fixed $v \in S$, the probability that $v \in S_p$ is the probability that v is in S_p but none of the other subsamples, plus the probability that v is in S_p and exactly 1 of the other subsamples, plus the probability that v is in S_p and exactly 2 of the other subsamples, plus the probability that v is in all 4 subsamples:

$$\mathbb{P}[v \in S_p \mid v \in S] = x - 3\varepsilon x^2 + 3\varepsilon^2 x^3 - \varepsilon^3 x^4 + 3(\varepsilon x^2 - 2\varepsilon^2 x^3 + \varepsilon^3 x^4) + 3(\varepsilon^2 x^3 - \varepsilon^3 x^4) + \varepsilon^3 x^4 = x.$$

In particular,

$$\mathbb{P}[v \in S_p] = \varepsilon x, \tag{23}$$

and this is symmetric for S_d, S_b, and S_r. We have proved that property (1) in the lemma statement is true for our choice of $x = \varepsilon/2$.

It remains to prove independence between subsamples (property (3) in the lemma statement). We begin with pairs of subsamples. Using Equations (22) and (23),

$$\mathbb{P}[v \in S_p \cap S_d] = \varepsilon^2 x^2 = \mathbb{P}[v \in S_p] \cdot \mathbb{P}[v \in S_d].$$

For triples, we use Equations (21) and (23),

$$\mathbb{P}[v \in S_p \cap S_d \cap S_b] = \varepsilon^3 x^3 = \mathbb{P}[v \in S_p] \cdot \mathbb{P}[v \in S_d] \cdot \mathbb{P}[v \in S_b].$$

Lastly, we use the fact that the probability $v \in S$ is added to all 4 subsamples is $\varepsilon^3 x^4$ together with Equation (23),

$$\mathbb{P}[v \in S_p \cap S_d \cap S_b \cap S_r] = \varepsilon^4 x^4 = \mathbb{P}[v \in S_p] \cdot \mathbb{P}[v \in S_d] \cdot \mathbb{P}[v \in S_b] \cdot \mathbb{P}[v \in S_r].$$

□

The independence properties satisfied by our construction yield the following corollary, which we apply heavily in the analysis.

Corollary 2. *Let $\mathcal{S} \subseteq \{S_p, S_d, S_b, S_r\}$. Let f be any function that depends on the subsamples in $\mathcal{S}$ and let $\mathcal{S}' \subseteq \mathcal{S}$. Let $\{s'_i\}_{i:S_i \in \mathcal{S}}$ be any collection of subsets of V, that is, a collection of possible realizations of $\{S_i\}_{i \in \mathcal{S}'}$. Define $g(\{s'_i\}_{i:S_i \in \mathcal{S}'}) := \mathbb{E}\left[f(\{s'_i\}_{i:S_i \in \mathcal{S}'}, \{S_i\}_{i:S_i \in \mathcal{S} \setminus \mathcal{S}'}\})\right]$, that is, the expectation of f when the subsamples in $\mathcal{S}'$ are fixed. Then*

$$\mathbb{E}\left[f(\mathcal{S}) \mid \mathcal{S}'\right] = g(\mathcal{S}').$$

In other words, to compute the conditional expectation, we can compute an unconditional expectation where we only integrate over the subsamples that are not part of the conditioning.

Proof of Corollary 2. We use the following claim.

Claim 19. *Let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \{S_p, S_d, S_b, S_r\}$ with $\mathcal{S}_1 \cap \mathcal{S}_2 = \varnothing$. Let $\{s_i\}_{i:S_i \in \mathcal{S}_1 \cup \mathcal{S}_2}$ be a collection of subsets of V. Then*

$$\mathbb{P}(\{S_i = s_i\}_{i:S_i \in \mathcal{S}_1 \cup \mathcal{S}_2}) = \mathbb{P}(\{S_i = s_i\}_{i:S_i \in \mathcal{S}_1}) \cdot \mathbb{P}(\{S_i = s_i\}_{i:S_i \in \mathcal{S}_2})$$

The claim is a straightforward consequence of properties (2) and (3) in Lemma 13. Having established the claim, the corollary then follows from the following well-known fact from probability theory.

Fact 3. *Let X, Y be* independent, discrete *random vectors, and let f be a function of X and Y. Define $g(x) := \mathbb{E}[f(x, Y)]$ for every x in the support of X. Then $\mathbb{E}[f(X, Y) \mid X] = g(X)$.*

□

A.2 Proof that $\bar{d}$ and $\tilde{d}$ are semi-metrics

Throughout this section, for $z \in V$ and S_* a chosen subsample, define the random variable X_z to be 1 if $z \in S_*$ and 0 otherwise. Most often, the analysis will choose S_* to be S_d.

Recall that Lemma 1 states that $\bar{d}$ and $\tilde{d}$ are semi-metric. We prove this lemma next.

Lemma 1. *Let $\bar{d}$ and $\tilde{d}$ be as in Definitions 4 and 5, respectively. Then*

- *$\bar{d} : V \times V \to [0, 1]$ is a 1-semi-metric, that is, $\bar{d}$ satisfies the triangle inequality.*
- *$\tilde{d} : V \times V \to [0, 1]$ is a $\frac{10}{7}$-semi-metric.*

Proof of Lemma 1. First, we prove that the triangle inequality holds for $\bar{d}$, i.e., for vertices $u, v, w \in V$, $\bar{d}_{uv} \leqslant \bar{d}_{uw} + \bar{d}_{wv}$. We already know that the correlation metric satisfies the triangle inequality by Lemma 2 in [11]. When the sample S_d intersects all of N_u^+, N_v^+, N_w^+, then $\bar{d}$ is equal to the correlation metric on the subgraph induced by the sample. To see this formally, recall S_d is the distance sample, and define the random variable X_z to be 1 if $z \in S_d$ and 0 otherwise. For $G[S_d]$ the subgraph induced on the vertices in S_d, we observe that

$$\bar{d}_{uv} = 1 - \frac{\sum_{z \in N_u^+ \cap N_v^+} X_z}{\sum_{z \in N_u^+ \cup N_v^+} X_z} = 1 - \frac{|(N_u^+ \cap N_v^+) \cap S_d|}{|(N_u^+ \cup N_v^+) \cap S_d|}.$$

This is almost exactly the correlation metric in S_d, except that u and v may no longer be in the subgraph $G[S_d]$. However, this subtle difference does not matter, as the vertices in S_d can still be partitioned based on their membership to $N_u^+ \cap S_d$ and $N_u^- \cap S_d$, or further partitioned based on their membership to $(N_u^+ \cap N_v^+) \cap S_d$, $(N_u^+ \cap N_v^-) \cap S_d$, $(N_u^- \cap N_v^+) \cap S_d$, or $(N_u^- \cap N_v^-) \cap S_d$, allowing the exact same proof as of Lemma 2 in [11] to go through.

Next, we observe that when at least one of N_u^+, N_v^+, N_w^+ do not intersect S_d, the triangle inequality still holds. The only case to check is when $N_w^+ \cap S_d = \varnothing$, but $N_u^+ \cap S_d \neq \varnothing$ and $N_v^+ \cap S_d \neq \varnothing$. Here, we see that $\bar{d}_{uw} + \bar{d}_{wv} \geqslant 1$ because $|(N_u^+ \cap N_w^+) \cap S_d|$ and $|(N_v^+ \cap N_w^+) \cap S_d|$ are 0.

The proof that $\tilde{d}$ is a 10/7 semi-metric is the same as the proof of Lemma 1 in [13]. There are three steps to computing $\tilde{d}$. The first step is to take $\tilde{d} = \bar{d}$ (for which we know the triangle inequality holds). Then in step 2, we round up negative edges with $\bar{d} \geqslant 7/10$, thus only gaining a factor of at most $10/7$ in an approximate triangle inequality. In the third and final step, some vertices are put in their own cluster. Here, if step 3 results in the left hand side of the approximate triangle inequality being changed to 1, then so is at least one term on the right hand side; so we do not gain any additional factor in the approximate triangle inequality from step 3. ☐

A.3 The good event B^c

For the ℓ_p-norm analysis with $p \in (1, \infty]$, we will condition on a "good" event occurring with high probability. The good event will be the complement of the following bad event, B.

Definition 12. *Let the* bad event B *be the union of the following events, which depend on the randomness of the subsamples:*

1. $|N_u^+ \cap S_d| < \varepsilon^2/4 \cdot |N_u^+|$ *for some* u *with* $|N_u^+| \geqslant C \cdot \log n/\varepsilon^2$.
2. *For* $t = \frac{r}{2\delta}$, $Ball_{\tilde{d}}^{S_p}(u,t) = \varnothing$ *for some* u *with* $|Ball_{\tilde{d}}(u,t)| \geqslant C' \cdot \log_{1/(1-\varepsilon^2/2)} n$.
3. $Ball_{\tilde{d}}^{S_p}(u, c \cdot r) = \varnothing$ *for some* u *with* $|Ball_{\tilde{d}}(u, c \cdot r)| \geqslant C' \cdot \log_{1/(1-\varepsilon^2/2)} n$.
4. $|Ball_{\tilde{d}}^{S_b}(w,r)| \leqslant {\varepsilon^2}/{4} \cdot |Ball_{\tilde{d}}(w,r)|$ *for some* w *with* $|Ball_{\tilde{d}}(w,r)| \geqslant 2C \cdot \log n/\varepsilon^2$
5. $|R_1(u) \cap S_r| < {\varepsilon^2}/{4} \cdot |R_1(u)|$ *for some* u *with* $|R_1(u)| \geqslant C \cdot \log n/\varepsilon^2$.
6. $|N_u^+ \cap S_p| < {\varepsilon^2}/{4} \cdot |N_u^+|$ *for some* u *with* $|N_u^+| \geqslant C \cdot \log n/\varepsilon^2$.
7. $\bar{d}_{uv} > \frac{(1+C)}{3\varepsilon^2} \cdot \log n \cdot d_{uv}$ *or* $1 - \bar{d}_{uv} > \frac{(1+C)}{3\varepsilon^2} \cdot \log n \cdot (1 - d_{uv})$ *for some* u, v *with* $|N_u^+ \cup N_v^+| \geqslant C \cdot \frac{\log n}{\varepsilon^2}$.

Define the complement B^c *to be the* good event.

To analyze the probability of the good event, we will use the following well-known Chernoff-Hoeffding bound.

Theorem 6 (Chernoff-Hoeffding)**.** *Let* $X = X_1 + \cdots + X_m$ *where* $\{X_1, \ldots, X_m\}$ *is a set of i.i.d. indicator random variables (*$X_i \in \{0,1\}$ *for* $i \in [m]$*). Define* $\mu = \mathbb{E}[X]$*. Then the following tail bounds hold:*

$$\mathbb{P}(X \geqslant (1+\delta)\mu) \leqslant e^{-\delta\mu/2} \quad \text{for } \delta \geqslant 2$$
$$\mathbb{P}(X \leqslant (1-\delta)\mu) \leqslant e^{-\delta^2\mu/3} \text{for } 0 < \delta < 1.$$

We will show that the good event B^c occurs with high probability in Lemma 15. For almost every event composing B in Definition 12, the fact that it happens with high probability is a consequence of previously proven statements. However, the last item (item 6) in Definition 12 requires a bit more work.

Specifically, in Lemma 14 we show that with high probability the estimated (unadjusted) correlation metric $\bar{d}$ is a good estimate of the (unadjusted) correlation metric[6] d under a technical condition, and thus the last item follows.

Lemma 14. *Let $\bar{d}$ be the estimated correlation metric. Fix u, v such that $|N_u^+ \cup N_v^+| \geqslant C \cdot \frac{\log n}{\varepsilon^2}$. Then, each of the following happens with probability at least $1 - \frac{3}{n^{C/24}}$:*

$$\bar{d}_{uv} \leqslant \frac{C+1}{3\varepsilon^2} \cdot \log n \cdot d_{uv} \quad \textit{and} \quad 1 - \bar{d}_{uv} \leqslant \frac{C+1}{3\varepsilon^2} \cdot \log n \cdot (1 - d_{uv}).$$

In particular, for fixed $u \in V$, if $|N_u^+| \geqslant C \cdot \frac{\log n}{\varepsilon^2}$, then with probability at least $1 - \frac{3}{n^{C/24-1}}$,

$$\bar{D}(u) \leqslant \frac{C+1}{3\varepsilon^2} \cdot \log n \cdot D(u) \leqslant \frac{8(C+1)}{3\varepsilon^2} \cdot \log n \cdot \mathit{OPT}_\infty.$$

We will use the following proposition, stating that the size of a sufficiently large random subset is well-concentrated, repeatedly. In all of our applications, we will take U to be one of S_p, S_d, S_b, or S_r.

Proposition 2. *Let $T \subseteq V$, and U be a subset such that each element $v \in V$ is in U independently with probability p. Suppose $|T| \geqslant (C \cdot \log n)/p$. Then, with probability at least $1 - \frac{1}{n^{C/12}}$,*

$$3 \cdot |T| \geqslant \frac{1}{p} \cdot |T \cap U| \geqslant \frac{1}{2} \cdot |T|.$$

Proof of Proposition 2. The proof follows from Theorem 6. Note that in what follows, $\mu = |T|p$. For the upper bound,

$$\mathbb{P}(|T \cap U| \geqslant 3|T|p) \leqslant e^{-2|T| \cdot p/2} \leqslant e^{-C \log n} = 1/n^C.$$

For the lower bound,

$$\mathbb{P}(|T \cap U| \leqslant |T|p/2) \leqslant e^{-|T|p/12} \leqslant e^{-C \log n/12} = \frac{1}{n^{C/12}}.$$

□

Now, we use Proposition 2 to prove Lemma 14. Recall that X_w is defined to be 1 if $w \in S_d$ and 0 otherwise. Recall also that $q(\varepsilon) = \mathbb{P}(v \in S_d) = \varepsilon^2/2$ for any vertex v.

Proof of Lemma 14. We may rewrite Definition 4 as

$$\bar{d}_{uv} = 1 - \frac{Y_{u,v}^{+,+}}{Y_{u,v}} = \frac{Y_{u,v}^{+,-} + Y_{u,v}^{-,+}}{Y_{u,v}} \tag{24}$$

for $Y_{u,v} > 0$, where

$$Y_{u,v}^{+,-} = \frac{1}{q(\varepsilon)} \cdot \sum_{w \in N_u^+ \cap N_v^-} X_w, \qquad Y_{u,v}^{-,+} = \frac{1}{q(\varepsilon)} \cdot \sum_{w \in N_u^- \cap N_v^+} X_w,$$

and

$$Y_{u,v}^{+,+} = \frac{1}{q(\varepsilon)} \cdot \sum_{w \in N_u^+ \cap N_v^+} X_w, \qquad Y_{u,v} = \frac{1}{q(\varepsilon)} \cdot \sum_{w \in N_u^+ \cup N_v^+} X_w.$$

Recall further that $\bar{d}_{uv}$ is defined to be 1 when $Y_{u,v} = 0$. We will show that, under the hypotheses of the proposition, the numerator and the denominator are each well-concentrated.

First we handle the denominator of $\bar{d}_{uv}$ in the right-hand side of (24). Applying Proposition 2 and the hypothesis that $|N_u^+ \cup N_v^+| \geqslant C \cdot \frac{\log n}{\varepsilon^2}$, we have

$$\mathbb{P}\left(Y_{u,v} \leqslant \frac{1}{2} \cdot |N_u^+ \cup N_v^+|\right) \leqslant \frac{1}{n^{C/24}}. \tag{25}$$

[6] We note that we do not, however, obtain an analogous statement for the estimate $\tilde{d}$ of the adjusted correlation metric d^*. We discuss this more in Appendix C.1.

In particular, $Y_{u,v} > 0$ with high probability, meaning that $\bar{d}_{uv}$ is defined as in line (24) with high probability.

Next we handle the numerator of $\bar{d}_{uv}$ in the right-hand side of (24). Observe that

$$\mathbb{E}[Y_{u,v}^{+,-}] = |N_u^+ \cap N_v^-| \quad \text{and} \quad \mathbb{E}[Y_{u,v}^{-,+}] = |N_u^- \cap N_v^+|.$$

Now we apply Theorem 6 to $Y_{u,v}^{+,-}$ and $Y_{u,v}^{-,+}$. By symmetry, it suffices to bound the former random variable:

$$\mathbb{P}\left(Y_{u,v}^{+,-} \geqslant (1+\delta) \cdot \mathbb{E}[Y_{u,v}^{+,-}]\right) = \mathbb{P}\Big(\sum_{w \in N_u^+ \cap N_v^-} X_w \geqslant (1+\delta) \cdot \mathbb{E}\big[\sum_{w \in N_u^+ \cap N_v^-} X_w\big]\Big). \tag{26}$$

We may assume without loss of generality that $|N_u^+ \cap N_v^-| \geqslant 1$ (as otherwise $\bar{d}_{uv} = d_{uv} = 0$) so that $\mu := \mathbb{E}\left[\sum_{w \in N_u^+ \cap N_v^-} X_w\right] \geqslant q(\varepsilon) = \varepsilon^2/2$. Thus, taking $\delta \geqslant \frac{C}{6} \cdot \frac{\log n}{\varepsilon^2}$ and applying Theorem 6 to line (26), we have that

$$\mathbb{P}\left(Y_{u,v}^{+,-} \geqslant (1+\delta) \cdot |N_u^+ \cap N_v^-|\right) \leqslant e^{-\delta\mu/2} \leqslant e^{-(C/24)\log n} = \frac{1}{n^{C/24}}. \tag{27}$$

and by an identical argument,

$$\mathbb{P}\left(Y_{u,v}^{-,+} \geqslant (1+\delta) \cdot |N_u^- \cap N_v^+|\right) \leqslant \frac{1}{n^{C/24}}. \tag{28}$$

Finally, combining (24), (25), (27), and (28), we have that, with probability at least $1 - \frac{3}{n^{C/24}}$,

$$\bar{d}_{uv} \leqslant \frac{C+1}{3\varepsilon^2} \cdot \log n \cdot \frac{|N_u^+ \cap N_v^-| + |N_u^- \cap N_v^+|}{|N_u^+ \cup N_v^+|},$$

as desired. The proof that, with high probability, $1 - \bar{d}_{uv} = O(\log n/\varepsilon^2) \cdot (1 - d_{uv})$ is similar, because $1 - \bar{d}_{uv}$ simplifies to $|N_u^+ \cap N_v^+|/|N_u^+ \cup N_v^+|$, and the concentration bound for $Y_{u,v}^{+,+}$ is identical to that of, say, $Y_{u,v}^{+,-}$.

Theorem 4 implies that with high probability $\bar{D}(u) \leqslant \frac{8(C+1)}{3\varepsilon^2} \cdot \log n \cdot \mathsf{OPT}_\infty$. $\square$

Now, we are ready to prove the good event occurs with high probability.

Lemma 15. *The good event B^c in Definition 12 happens with probability at least* $1 - \left(\frac{3}{n^{C/24-2}} + \frac{4}{n^{C/12-1}} + \frac{2}{n^{C'-1}}\right)$.

Proof of Lemma 15. We will go item by item from Definition 12 to upper bound the probability of the bad event B.

Item 1. Fix some u with $|N_u^+| \geqslant \frac{C}{\varepsilon^2} \cdot \log n$. Applying Proposition 2 by choosing T to be N_u^+ and U to be S_d, with probability at least $1 - \frac{1}{n^{C/12}}$

$$3 \cdot |N_u^+| \geqslant \frac{2}{\varepsilon^2} \cdot |N_u^+ \cap S_d| \geqslant \frac{1}{2} \cdot |N_u^+|.$$

Thus $|N_u^+ \cap S_d| < \varepsilon^2/4 \cdot |N_u^+|$ with probability less than $\frac{1}{n^{C/12}}$, and union bounding over all such potential vertices gives an upper bound of $\frac{1}{n^{C/12-1}}$.

Item 2. Fix some u with $|\mathrm{Ball}_{\tilde{d}}(u,t)| > C' \cdot \log_{1/(1-q(\varepsilon))} n$. Note we consider $\tilde{d}$ fixed here, meaning that some choice of S_d, S_r is fixed here. Then we can use Corollary 2 (and in particular the independence of S_p from S_d, S_r) to see that

$$\mathbb{P}[S_p \cap \mathrm{Ball}_{\tilde{d}}(u,t) = \varnothing \mid S_d, S_r] = (1-q(\varepsilon))^{|\mathrm{Ball}_{\tilde{d}}(u,t)|} \leqslant (1-q(\varepsilon))^{C' \cdot \log_{1/(1-q(\varepsilon))} n} = \frac{1}{n^{C'}},$$

where the randomness is over S_p. Taking a union bound over all such possible u, we upper bound the probability this happens for any u by $\frac{1}{n^{C'-1}}$.

Item 3. Same as proof of the previous item.
Item 4. Fix some w with $|\text{Ball}_{\tilde{d}}(w,r)| \geqslant C \cdot \log n/\varepsilon^2$. Note we consider $\tilde{d}$ fixed here, meaning that some choice of S_d and S_r are fixed here. Since S_b is independent of S_d and S_r, as in Corollary 2, we can apply Proposition 2 by letting T be $|\text{Ball}_{\tilde{d}}(w,r)|$ and U to be S_b, so that with probability at least $1 - \frac{1}{n^{C/12}}$

$$3 \cdot |\text{Ball}_{\tilde{d}}(w,r)| \geqslant \frac{2}{\varepsilon^2} \cdot |\text{Ball}_{\tilde{d}}(w,r) \cap S_b| \geqslant \frac{1}{2} \cdot |\text{Ball}_{\tilde{d}}(w,r)|.$$

Thus $|\text{Ball}_{\tilde{d}}(w,r) \cap S_b| \leqslant {}^{\varepsilon^2}\!/4 \cdot |\text{Ball}_{\tilde{d}}(w,r)|$ with probability less than $\frac{1}{n^{C/12}}$ for any one such w, and union bounding over any such possible w gives an upper bound of $\frac{1}{n^{C/12-1}}$.
Item 5. Fix some u with $|N_u^- \cap \{v \in V : \bar{d}_{uv} \leqslant 7/10\}| \geqslant C \cdot \log n/q(\varepsilon)$, thus fixing some choice of S_d here. Subsamples are independent by Corollary 2, and in particular S_r and S_d are independent. Thus we can apply Proposition 2 by letting T be $|N_u^- \cap \{v \in V : \bar{d}_{uv} \leqslant 7/10\}|$ and U to be S_r, so that with probability at least $1 - \frac{1}{n^{C/12}}$

$$\begin{aligned} 3 \cdot |N_u^- \cap \{v \in V : \bar{d}_{uv} \leqslant 7/10\}| &\geqslant \frac{2}{\varepsilon^2} \cdot |N_u^- \cap \{v \in V : \bar{d}_{uv} \leqslant 7/10\} \cap S_r| \\ &\geqslant \frac{1}{2} \cdot |N_u^- \cap \{v \in V : \bar{d}_{uv} \leqslant 7/10\}|. \end{aligned}$$

It follows that $|N_u^- \cap \{v \in V : \bar{d}_{uv} \leqslant 7/10\}| > \frac{4}{\varepsilon^2} \cdot |N_u^- \cap S_r \cap \{v \in V : \bar{d}_{uv} \leqslant 7/10\}|$ with probability less than $\frac{1}{n^{C/12}}$ and union bounding over all possible w gives an upper bound of with probability less than $\frac{1}{n^{C/12-1}}$.
Item 6. Same proof as item 1, just replacing S_d with S_p.
Item 7. The probability that some fixed u, v with $|N_u^+ \cup N_v^+| \geqslant C \cdot \frac{\log n}{\varepsilon^2}$ have $\bar{d}_{uv} > \frac{2(1+C)}{6\varepsilon^2} \log n \cdot d_{uv}$ is less than $\frac{3}{n^{C/24}}$ by Lemma 14. Union bounding over any such u, v pair, we see that any such event occurs with probability at most $\frac{3}{n^{C/24-2}}$.

Combining the probabilities, we see that the probability that event B occurs is upper bounded by

$$\frac{3}{n^{C/24-2}} + \frac{4}{n^{C/12-1}} + \frac{3}{n^{C'-1}}.$$

For choices of constants where $C' = 5$ and $C = 100$, the probability that the good event B^c occurs is at least $1 - 1/n$ for sufficiently large n. □

A.4 Positive fractional cost of correlation metric for finite p

We first bound the ℓ_p-norm cost of the estimated adjusted correlation metric $\tilde{d}$ on positive edges. We restate the lemma for convenience.

Lemma 3. *Let $1 \leqslant p < \infty$. The estimated adjusted correlation metric $\tilde{d}$ satisfies the following:*

- $\mathbb{E}\left[\sum_{u \in V_0} \sum_{v \in N_u^+ \cap V_0} \tilde{d}_{uv}\right] \leqslant O(1/\varepsilon^4) \cdot \mathit{OPT}_1$*, and*
- *Conditioned on the event B^c,* $\sum_{u \in V_0} \left(\sum_{v \in N_u^+ \cap V_0} \tilde{d}_{uv}\right)^p \leqslant O\left(\left(1/\varepsilon^6 \cdot \log^3 n\right)^p\right) \cdot \mathit{OPT}_p^p$.

Proof of Lemma 3. Let y be the vector of disagreements in a fixed clustering $\mathcal{C}$, let $C(u)$ denote the set of vertices in vertex v's cluster, and let $\overline{C(u)} := V \backslash C(u)$.

We first bound the ℓ_p-norm cost of the estimated *unadjusted* correlation metric, $\bar{d}$, on positive edges:

Claim 20. *Let $1 \leqslant p < \infty$. The following bounds hold:*

- $\mathbb{E}\left[\sum_{u \in V_0} \sum_{v \in N_u^+ \cap V_0} \bar{d}_{uv}\right] \leqslant O(1/\varepsilon^2) \cdot \mathit{OPT}_1$*, and*
- *Conditioned on the event B^c,* $\sum_{u \in V_0} \left(\sum_{v \in N_u^+ \cap V_0} \bar{d}_{uv}\right)^p \leqslant O\left((1/\varepsilon^2 \cdot \log n)^p\right) \cdot \mathit{OPT}_p^p$.

Proof of Claim 20. Define

$$E_1 = \sum_{u \in V_0} \left(\sum_{v \in N_u^+ \cap C(u) \cap V_0} \frac{|N_u^+ \cap N_v^- \cap S_d| + |N_u^- \cap N_v^+ \cap S_d|}{|(N_u^+ \cup N_v^+) \cap S_d|} \right)^p$$

$$E_2 = \sum_{u \in V_0} \left(\sum_{v \in N_u^+ \cap \overline{C(u)} \cap V_0} \frac{|N_u^+ \cap N_v^- \cap S_d| + |N_u^- \cap N_v^+ \cap S_d|}{|(N_u^+ \cup N_v^+) \cap S_d|} \right)^p.$$

Notice that the numerator in E_1 is bounded above by $y(u) + y(v)$. This is because $v \in C(u)$ implies that (u, w) or (v, w) is a disagreement for every $w \in (N_u^+ \cap N_v^-) \cup (N_u^- \cap N_v^+)$. So we define

$$E_{1a} = \sum_{u \in V_0} \left(\sum_{v \in N_u^+ \cap C(u) \cap V_0} \frac{y(u)}{|(N_u^+ \cup N_v^+) \cap S_d|} \right)^p$$

$$E_{1b} = \sum_{u \in V_0} \left(\sum_{v \in N_u^+ \cap C(u) \cap V_0} \frac{y(v)}{|(N_u^+ \cup N_v^+) \cap S_d|} \right)^p.$$

By Jensen's inequality, we can, at a loss a factor of 3^{p-1}, bound the quantities E_{1a}, E_{1b}, and E_2 individually.

Using that $\bar{d}_{uv} \leqslant 1$, we have $E_2 = \sum_{u \in V} \left(\sum_{v \in N_u^+ \cap \overline{C(u)} \cap V_0} \bar{d}_{uv} \right)^p \leqslant \sum_{u \in V} y(u)^p$, always. Next,

$$E_{1a} \leqslant \sum_{u \in V_0} \frac{1}{|N_u^+ \cap S_d|^p} \cdot \left(\sum_{v \in N_u^+} y(u) \right)^p \leqslant \sum_{u \in V} \frac{|N_u^+|^p}{|N_u^+ \cap S_d|^p} \cdot \mathbf{1}_{\{u \in V_0\}} \cdot y(u)^p.$$

We can now apply Proposition 1 to obtain a bound of $14/\varepsilon^2 \cdot \sum_{u \in V} y(u)$ in expectation for $p = 1$, and of $(C \cdot \log n/\varepsilon^2)^p \cdot \sum_{u \in V} y(u)^p$ conditioned on the event B^c for finite $p \geqslant 1$. Finally we bound E_{12}:

$$\begin{aligned}
E_{1b} &\leqslant \sum_{u \in V_0} \sum_{v \in N_u^+ \cap V_0 \cap C(u)} \frac{|N_u^+ \cap V_0 \cap C(u)|^{p-1}}{|(N_u^+ \cup N_v^+) \cap S_d|^p} \cdot y(v)^p \\
&\leqslant \sum_{u \in V_0} \sum_{v \in N_u^+ \cap V_0 \cap C(u)} \frac{|N_u^+|^{p-1}}{|N_u^+ \cap S_d|^{p-1} \cdot |N_v^+ \cap S_d|} \cdot y(v)^p \\
&\leqslant \left(\frac{C \cdot \log n}{\varepsilon^2} \right)^{p-1} \cdot \sum_{u \in V_0} \sum_{v \in N_u^+ \cap V_0 \cap C(u)} \frac{1}{|N_v^+ \cap S_d|} \cdot y(v)^p \\
&\leqslant \left(\frac{C \cdot \log n}{\varepsilon^2} \right)^{p-1} \cdot \sum_{v \in V_0} y(v)^p \sum_{u \in V_0 \cap N_v^+} \frac{1}{|N_v^+ \cap S_d|} \\
&\leqslant \left(\frac{C \cdot \log n}{\varepsilon^2} \right)^{p-1} \cdot \sum_{v \in V} y(v)^p \cdot \frac{|N_v^+|}{|N_v^+ \cap S_d|} \cdot \mathbf{1}_{\{u \in V_0\}}
\end{aligned}$$

where in the first line we have applied Jensen's inequality. Applying Proposition 1 we obtain a bound of $14/\varepsilon^2 \cdot \sum_{u \in V} y(u)$ in expectation for $p = 1$, and of $(C \cdot \log n/\varepsilon^2)^p \cdot \sum_{v \in V} y(v)^p$ conditioned on the event B^c for finite $p \geqslant 1$. $\square$

Recall R_1 is the set of vertices isolated by $\tilde{d}$ (see Definition 5), and $R_2 = V \backslash R_1$. Next, we need to bound the contribution of vertices $u \in R_1$ to the fractional cost. So, we need to bound $\sum_{u \in V_0 \cap R_1} |N_u^+|^p$. Define

$$V^1 := V_0 \cap \left\{ u \in V : |N_u^+ \cap S_d \cap \overline{C(u)}| \geqslant 3/20 \cdot |N_u^+ \cap S_d| \right\}$$

$$V^2 := V_0 \cap \left\{u \in V : |N_u^+ \cap S_d \cap C(u)| \geqslant {}^{17}\!/_{20} \cdot |N_u^+ \cap S_d|\right\}$$
$$V^{2a} := V^2 \cap \left\{u \in V : |N_u^- \cap C(u)| \geqslant |N_u^+|\right\}, \qquad V^{2b} := V^2 \cap \left\{u \in V : |C(u)| \leqslant 2 \cdot |N_u^+|\right\}.$$

Claim 21. *Let $1 \leqslant p < \infty$. The following bounds hold:*

- $\mathbb{E}\left[\sum_{u \in R_1 \cap V^1} |N_u^+|^p\right] \leqslant O({}^1\!/_{\varepsilon^2}) \cdot \mathit{OPT}_1$*, and*
- *Conditioned on the event B^c, we have $\sum_{u \in R_1 \cap V^1} |N_u^+|^p \leqslant O\left(({}^1\!/_{\varepsilon^2} \cdot \log n)^p\right) \cdot \mathit{OPT}_p^p$.*

Proof of Claim 21. Using the definition of V^1, we have

$$\begin{aligned}\sum_{u \in R_1 \cap V^1} |N_u^+|^p &\leqslant (20/3)^p \cdot \sum_{u \in V_0} |N_u^+|^p \cdot \frac{|N_u^+ \cap S_d \cap \overline{C(u)}|^p}{|N_u^+ \cap S_d|^p} \\ &\leqslant (20/3)^p \cdot \sum_{u \in V} \frac{|N_u^+|^p}{|N_u^+ \cap S_d|^p} \cdot \mathbf{1}_{\{u \in V_0\}} \cdot y(u)^p.\end{aligned}$$

We can now apply Proposition 1 to obtain a bound of ${}^{280}\!/_{3\varepsilon^2} \cdot \sum_{u \in V} y(u)$ in expectation for $p = 1$, and of $\left(\frac{20}{3} \cdot \frac{C \cdot \log n}{\varepsilon^2}\right)^p \cdot \sum_{u \in V} y(u)^p$ conditioned on the event B^c for finite $p \geqslant 1$. ☐

Claim 22. *Let $1 \leqslant p < \infty$. It is always the case that $\sum_{u \in R_1 \cap V^{2a}} |N_u^+|^p \leqslant \mathit{OPT}_p^p$.*

Proof of Claim 22. By definition of V^{2a}, $\sum_{u \in R_1 \cap V^{2a}} |N_u^+|^p \leqslant \sum_{u \in R_1 \cap V^{2a}} |N_u^- \cap C(u)|^p \leqslant \sum_{u \in V} y(u)^p$. ☐

Claim 23. *Let $1 \leqslant p < \infty$. The following bounds hold:*

- $\mathbb{E}\left[\sum_{u \in R_1 \cap V^{2b}} |N_u^+|\right] \leqslant O\left({}^1\!/_{\varepsilon^4}\right) \cdot \mathit{OPT}_1$*, and*
- *Conditioned on the event B^c, we have $\sum_{u \in R_1 \cap V^{2b}} |N_u^+|^p \leqslant O\left(\left({}^1\!/_{\varepsilon^4} \cdot \log^2 n\right)^p\right) \cdot \mathit{OPT}_p^p$.*

Proof of Claim 23. Recall $R_1(u) := N_u^- \cap \{v : \bar{d}_{uv} \leqslant 7/10\}$ (Definition 5), and that for $u \in R_1$, $|R_1(u)| \geqslant \frac{10}{3} \cdot |N_u^+ \cap S_d|$. For convenience, define $N_{u,v} := N_u^+ \cap N_v^+ \cap C(u)$.

Define $\varphi(u, w) := |R_1(u) \cap N_w^+|$. Note that for $w \in N_u^+$, we have $\varphi(u, w) \leqslant y(u) + y(w)$, since for any $v \in R_1(u) \cap N_w^+$, either the edge vw is a disagreement or the edge vu is a disagreement. We will charge the cost of $u \in R_1$ to these disagreements, using a double-counting argument. First we lower bound the number of such disagreements that arise in this way for fixed u:

$$\sum_{w \in N_u^+ \cap C(u) \cap S_d} \varphi(u, w) = \sum_{v \in R_1(u)} |N_{u,v} \cap S_d| \geqslant \sum_{v \in R_1(u)} \frac{3}{20} \cdot |N_u^+ \cap S_d| \geqslant \frac{1}{2} \cdot |N_u^+ \cap S_d|^2$$

where in the first inequality we have used Fact 2, and in the last inequality we have used the lower bound on $R_1(u)$ from above.

Next, applying Jensen's inequality, we have

$$|N_u^+ \cap C(u) \cap S_d|^{p-1} \cdot \sum_{w \in N_u^+ \cap C(u) \cap S_d} \varphi(u, w)^p \geqslant \left(\sum_{w \in N_u^+ \cap C(u) \cap S_d} \varphi(u, w)\right)^p \geqslant \frac{1}{2^p} \cdot |N_u^+ \cap S_d|^{2p}.$$

Now we may bound the contribution of $u \in R_1 \cap V_{2b}$:

$$\sum_{u \in R_1 \cap V^{2b}} |N_u^+|^p \leqslant 2^p \cdot \sum_{u \in R_1 \cap V^{2b}} \frac{|N_u^+|^p}{|N_u^+ \cap S_d|^{2p}} \cdot |N_u^+ \cap C(u) \cap S_d|^{p-1} \cdot \sum_{w \in N_u^+ \cap C(u) \cap S_d} \varphi(u, w)^p.$$

As $\varphi(u,w) \leqslant y(u) + y(w)$, it suffices by Jensen's inequality to individually bound the following terms, at a loss of an additional factor of 2^{p-1}:

$$\sum_{u \in R_1 \cap V^{2b}} \frac{|N_u^+|^p}{|N_u^+ \cap S_d|^{2p}} \cdot |N_u^+ \cap C(u) \cap S_d|^{p-1} \cdot \sum_{w \in N_u^+ \cap C(u) \cap S_d} y(u)^p$$

and

$$\sum_{u \in R_1 \cap V^{2b}} \frac{|N_u^+|^p}{|N_u^+ \cap S_d|^{2p}} \cdot |N_u^+ \cap C(u) \cap S_d|^{p-1} \cdot \sum_{w \in N_u^+ \cap C(u) \cap S_d} y(w)^p.$$

The first term is upper bounded by

$$\sum_{u \in V_0 : |C(u)| \leqslant 2 \cdot |N_u^+|} \frac{|N_u^+|^p}{|N_u^+ \cap S_d|^{2p}} \cdot |C(u)|^p \cdot y(u)^p \leqslant 2^p \cdot \sum_{u \in V} \frac{|N_u^+|^{2p}}{|N_u^+ \cap S_d|^{2p}} \cdot \mathbf{1}_{\{u \in V_0\}} \cdot y(u)^p.$$

Applying Proposition 1, we obtain a bound of ${}^{664}/_{\varepsilon^4} \cdot \sum_{u \in V} y(u)$ in expectation for $p = 1$, and of $2^p \cdot \left(\frac{C^2 \cdot \log^2 n}{\varepsilon^4}\right)^p \cdot \sum_{u \in V} y(u)^p$ conditioned on the event B^c for finite $p \geqslant 1$. The second term is upper bounded by

$$\begin{aligned}
&\sum_{w \in S_d} y(w)^p \sum_{u \in V_0 \cap C(w) : |C(u)| \leqslant 2 \cdot |N_u^+|} \frac{|N_u^+|^p}{|N_u^+ \cap S_d|^{2p}} \cdot |N_u^+ \cap C(u) \cap S_d|^{p-1} \\
\leqslant &\sum_{w \in S_d} y(w)^p \sum_{u \in V_0 \cap C(w) : |C(u)| \leqslant 2 \cdot |N_u^+|} \frac{|N_u^+|^p}{|N_u^+ \cap S_d|^{2p}} \cdot |C(w)|^{p-1} \\
\leqslant &\sum_{w \in S_d} y(w)^p \sum_{u \in V_0 \cap C(w) : |C(w)| \leqslant 2 \cdot |N_u^+|} \frac{|N_u^+|^{2p}}{|N_u^+ \cap S_d|^{2p}} \cdot \frac{|C(w)|^{p-1}}{|N_u^+|^p} \\
\leqslant &\, 2^p \cdot \sum_{w \in V} y(w)^p \sum_{u \in C(w)} \frac{|N_u^+|^{2p}}{|N_u^+ \cap S_d|^{2p}} \cdot \mathbf{1}_{\{u \in V_0\}} \cdot \frac{|C(w)|^{p-1}}{|C(w)|^p}.
\end{aligned}$$

Again applying Proposition 1, we obtain the same upper bounds as above for the first term. □

In total, we combine the results from Claims 21, 22, and 23 to see

$$\mathbb{E}\left[\sum_{u \in V_0 \cap R_1} |N_u^+|\right] \leqslant ({}^1/_{\varepsilon^4}) \cdot \mathsf{OPT}_1 \tag{29}$$

and conditioning on B^c, for finite p

$$\sum_{u \in V_0 \cap R_1} |N_u^+|^p \leqslant O\left(\left({}^1/_{\varepsilon^4} \cdot \log^2 n\right)^p\right) \cdot \mathsf{OPT}_p^p. \tag{30}$$

Finally, for $u \in R_2$, we need to bound the contribution of $uv \in E^+$ such that $v \in R_1$, i.e., of those v for which $\tilde{d}_{uv} = 1 \neq \bar{d}_{uv}$. We note that this case is not needed for $p = 1$, since Claims 20, 21, 22, and 23 suffice to compute the ℓ_1-norm of the fractional cost.

Claim 24. *Let $1 \leqslant p < \infty$. Conditioned on the event B_c, we have*

$$\sum_{u \in R_2 \cap V_0} |N_u^+ \cap R_1 \cap V_0|^p \leqslant O\left(\left({}^1/_{\varepsilon^6} \cdot \log^3 n\right)^p\right) \cdot \mathit{OPT}_p^p.$$

Proof of Claim 24. Observe that

$$\sum_{u \in R_2 \cap V_0} |\{v : v \in N_u^+ \cap R_1 \cap V_0 \text{ and } \bar{d}_{uv} \geqslant 1/4\}|^p \leqslant 4^p \cdot \sum_{u \in V_0} \left(\sum_{v \in V_0 \cap N_u^+} \bar{d}_{uv} \right)^p .$$

By Claim 20, this is bounded for $p = 1$ in expectation by $\frac{116}{\varepsilon^2} \cdot \mathsf{OPT}_1$ and for finite p by $\left(12C \cdot \log n/\varepsilon^2\right)^p \cdot \mathsf{OPT}_p^p$, since we conditioned on B^c .

Given the above observation, with loss of a factor of 2^{p-1} due to Jensen's inequality, it suffices to bound the analogous quantity with $\bar{d}_{uv} \leqslant 1/4$. We create a bipartite auxiliary graph $H = (R_2 \cap V_0, R_1 \cap V_0, F)$. An edge $uv \in F$ for $u \in R_2 \cap V_0$ and $v \in R_1 \cap V_0$ if $uv \in E^+$ and $\bar{d}_{uv} \leqslant 1/4$. It then suffices to show

$$\sum_{u \in R_2 \cap V_0} \deg_H(u)^p = O\left(\log n/\varepsilon^2\right)^p \cdot \sum_{v \in R_1 \cap V_0} |N_v^+|^p,$$

since then we may bound the right-hand side via Claims 21, 22, and 23 (specifically, we combine them using Equation (29)). We will bound via double counting the quantity $\sum_{f=uv \in F} \left(\deg_H(u) + \deg_H(v)\right)^{p-1}$. Let $N_H(\cdot)$ denote the neighborhoods in H of the vertices.

$$\begin{aligned}
\sum_{f=uv \in F} (\deg_H(u) + \deg_H(v))^{p-1} &\leqslant \sum_{v \in R_1 \cap V_0} \sum_{u \in N_H(v)} (\deg_H(v) + \deg_H(u))^{p-1} \\
&\leqslant \sum_{v \in R_1 \cap V_0} \sum_{u \in N_H(v)} \left(|N_v^+| + C \cdot \log n/\varepsilon^2 \cdot |N_u^+ \cap S_d|\right)^{p-1} && (31) \\
&\leqslant \sum_{v \in R_1 \cap V_0} \sum_{u \in N_H(v)} \left(|N_v^+| + C \cdot \log n/\varepsilon^2 \cdot 4/3 \cdot |N_v^+ \cap S_d|\right)^{p-1} && (32) \\
&\leqslant \sum_{v \in R_1 \cap V_0} \sum_{u \in N_H(v)} (1 + 4/3 \cdot C \cdot \log n/\varepsilon^2)^{p-1} \cdot |N_v^+|^{p-1} \\
&\leqslant (1 + 4/3 \cdot C \cdot \log n/\varepsilon^2)^{p-1} \cdot \sum_{v \in R_1 \cap V_0} |N_v^+| \cdot |N_v^+|^{p-1} \\
&\leqslant (3 \cdot C \cdot \log n/\varepsilon^2)^{p-1} \cdot \sum_{v \in R_1 \cap V_0} |N_v^+|^p
\end{aligned}$$

where (31) follows from Proposition 1 and (32) follows from Fact 4. It now just remains to lower bound the sum:

$$\begin{aligned}
\sum_{f=uv \in F} (\deg_H(u) + \deg_H(v))^{p-1} &= \sum_{u \in R_2 \cap V_0} \sum_{v \in N_H(u)} (\deg_H(u) + \deg_H(v))^{p-1} \\
&\geqslant \sum_{u \in R_2 \cap V_0} \sum_{v \in N_H(u)} \deg_H(u)^{p-1} \\
&= \sum_{u \in R_2 \cap V_0} \deg_H(u)^p,
\end{aligned}$$

which is what we sought to show. □

Combining Claims 20, 21, 22, 23, and 24 completes the proof of Lemma 3 (using only the first four claims for $p = 1$) with $\mathbb{E}\left[\sum_{u \in V_0} \sum_{v \in N_u^+ \cap V_0} \tilde{d}_{uv}\right] \leqslant 1451/\varepsilon^4 \cdot \mathsf{OPT}_1$, and conditioned on the event B^c, $\sum_{u \in V_0} \left(\sum_{v \in N_u^+ \cap V_0} \tilde{d}_{uv}\right)^p \leqslant \left(110 \cdot C^3 \cdot \log^3 n/\varepsilon^6\right)^p \cdot \mathsf{OPT}_p^p$. □

B Tight analysis of Algorithm 1 for $p = 1$

While the proof of item 1 of Theorem 1 is valid when $p = 1$, the analysis is rather lossy for this case. We know there exist $O(1)$ algorithms for ℓ_1-norm correlation clustering in the AOS model [22], and we will show Algorithm 1 achieves this too.

As in the finite ℓ_p-norm analysis (Section 5), we partition our analysis into the cost of the disagreements incurred during the Pre-clustering phase (Section B.1), the cost of the disagreements incurred during the two individual Pivot phases (Section B.2), and those cost incurred between the Pre-clustering phase and the Pivot phases (Section B.3). See Figure 4 for references to the lemmas for each type of disagreement. Note there are fewer cases than in Figure 1 for the finite ℓ_p-norm analysis, because we need not consider, e.g., whether $u \succ v$ or $v \succ u$ for the ℓ_1-norm objective.

The analysis in Section B.1 is for disagreements where both endpoints are eligible and at least one is pre-clustered (see the blue and pink edges in Figure 4). The analysis here is similar to the corresponding analysis (Section 5.1) for ℓ_p-norms, $p \in (1, \infty)$, but involves more probabilistic subtleties in order to obtain an upper bound that both holds in expectation and nearly matches the lower bound.

The analysis in Sections B.2 and B.3 looks most dissimilar to the corresponding ℓ_p-norm analyses (Sections 5.2 and 5.3). For edges that have at least one endpoint in $\overline{V_0}$ (see the orange edges and yellow edges in Figure 4), the key idea will be that the expected positive degree of a vertex in $G[\overline{V_0}]$ is small. This will help us charge the cost of disagreements to bad triangles (Definition 10) and to the (true) correlation metric d (Definition 2).

Finally, for the disagreements incurred from running Modified Pivot on $G[V_0']$, we will again charge to both bad triangles and to $\tilde{d}$, as we did for the ℓ_p-norm analysis (Lemma 10), but the analysis will be simpler as for the ℓ_1-norm we may charge on a per-edge basis rather than a per-vertex basis.

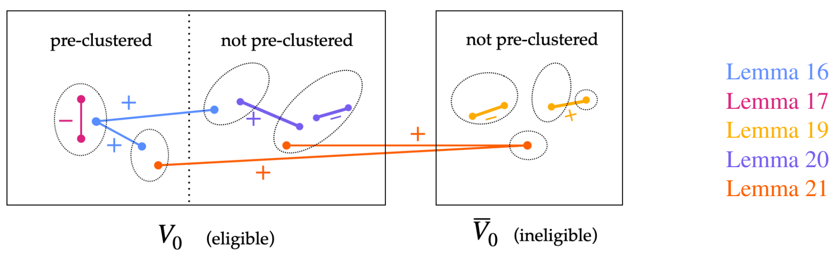

Figure 4: An overview of the cost analysis of Algorithm 1 for $p = 1$. As in Figure 1, solid edges are disagreements and dashed ovals are clusters, and different types of disagreements are color coded with the corresponding lemmas where we handle their charging arguments.

B.1 Cost of Pre-clustering phase

In this section, we bound the cost of the pre-clustering phase in Algorithm 1. By this we mean that we bound the cost of disagreeing edges that have at least one endpoint that is pre-clustered. We refer the reader to Section 5.1 for a reminder on notations and definitions related to the pre-clustering phase.

Define $t := r/(2\delta)$ to be a threshold parameter which will be used in our analysis.

B.1.1 Cost of positive edges

We begin in Lemma 16 by bounding the cost of positive edges that are disagreements in $\mathcal{C}_{\mathsf{ALG}}$ where both endpoints are eligible and at least one is pre-clustered. This lemma is an analog of Lemmas 6 / 7 in the cost analysis for ℓ_p-norms, $p \in (1, \infty)$. Here, we need not consider separate lemmas because the ℓ_1-norm objective is equivalent to minimizing the total number of edges that are disagreements.

Lemma 16. *The expected cost of disagreements $uv \in E^+$ where u and v are eligible and at least one is pre-clustered is*

$$\mathbb{E}\Big[\sum_{u \in V_0} |\{v \in N_u^+ : v \succ u\}|\Big] \leqslant O\left({}^{1}\!/\!_{\varepsilon^6}\right) \cdot \mathit{OPT}_1.$$

Proof of Lemma 16. Observe first that each v in the statement of the lemma is in V_0, since $v \succ u$ implies v is pre-clustered. As in Lemma 6, for fixed $u \in V_0$ partition the disagreements we wish to bound into two sets:

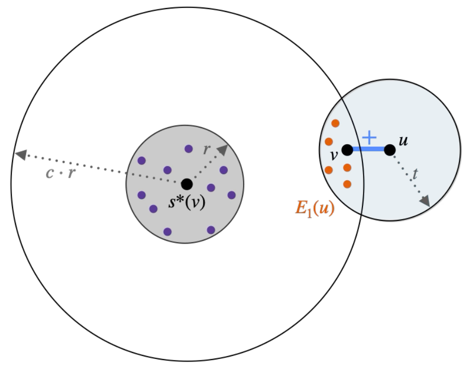

Figure 5: Bounding E_{1b} in the proof of Lemma 16. The orange points $v \in E_1(u)$ are clustered before u, so edges uv are disagreements. To bound the cost of the orange points, it suffices to show that there are more purple points in expectation, and that the cost of the purple points can be charged to the fractional cost $\tilde{D}_0(u)$ of u.

$$E_1(u) := \{v \in N_u^+ : v \succ u\} \cap \mathrm{Ball}_{\tilde{d}}(u, t)$$

and

$$E_2(u) := \{v \in N_u^+ : v \succ u\} \backslash E_1(u).$$

So we wish to bound

$$\mathbb{E}\Big[\sum_{u \in V_0} |\{v \in N_u^+ : v \succ u\}|\Big] = \mathbb{E}\Big[\sum_{u \in V_0} |E_1(u)|\Big] + \mathbb{E}\Big[\sum_{u \in V_0} |E_2(u)|\Big]$$

By Lemma 3, the second term is at most

$$\mathbb{E}\Big[\sum_{u \in V_0} |E_2(u)|\Big] \leqslant \frac{1}{t} \cdot \mathbb{E}\Big[\sum_{u \in V_0} \sum_{v \in N_u^+ \cap V_0} \tilde{d}_{uv}\Big] \leqslant O(1/\varepsilon^4) \cdot \mathsf{OPT}_1. \tag{33}$$

To bound the first term, $\mathbb{E}[\sum_{u \in V_0} |E_1(u)|]$, we case on whether u has a close neighbor sampled by S_p. We note that the sums below are taken over *random* sets.

$$\mathbb{E}\Big[\sum_{u \in V_0} |E_1(u)|\Big] = \underbrace{\mathbb{E}\Big[\sum_{u \in V_0 : \mathrm{Ball}_{\tilde{d}}^{S_p}(u,t) = \varnothing} |E_1(u)|\Big]}_{E_{1a}} + \underbrace{\mathbb{E}\Big[\sum_{u \in V_0 : \mathrm{Ball}_{\tilde{d}}^{S_p}(u,t) \neq \varnothing} |E_1(u)|\Big]}_{E_{1b}}$$

Bounding E_{1a}. As in the proof of Lemma 6, bounding E_{1a} is simpler because the expected size of $E_1(u)$ must be small. We repeat the following claim, introduced in the proof of Lemma 6, that shows there is sufficient fractional cost incident to u to which to charge $E_1(u)$.

Recall $\tilde{D}_0(u) := \sum_{v \in N_u^+ \cap V_0} \tilde{d}_{uv} + \sum_{v \in N_u^- \cap V_0} (1 - \tilde{d}_{uv})$.

Claim 4. *If $u \in V_0$ and $E_1(u) \neq \varnothing$, then $\tilde{D}_0(u) \geqslant 1/5$.*

Claim 25. *For $t = r/2\delta$, it is the case that*

$$E_{1a} := \mathbb{E}\Big[\sum_{u \in V_0 : Ball_{\tilde{d}}^{S_p}(u,t) = \varnothing} |E_1(u)|\Big] \leqslant O(1/\varepsilon^6) \cdot \mathit{OPT}_1.$$

Proof of Claim 25. In this case we may upper bound $|E_1(u)|$ by $|\mathrm{Ball}_{\tilde{d}}(u, t)|$. Applying Claim 4, we obtain:

$$
\begin{aligned}
E_{1a} &\leqslant \mathbb{E}\Big[\sum_{u\in V_0:\mathrm{Ball}_{\tilde d}^{S_p}(u,t)=\varnothing} |\mathrm{Ball}_{\tilde d}(u,t)|\Big] \leqslant 5\cdot\mathbb{E}\Big[\sum_{u\in V_0:\mathrm{Ball}_{\tilde d}^{S_p}(u,t)=\varnothing} |\mathrm{Ball}_{\tilde d}(u,t)|\cdot \tilde D_0(u)\Big] \\
&= 5\cdot\mathbb{E}\Big[\mathbb{E}\Big[\sum_{u\in V_0} |\mathrm{Ball}_{\tilde d}(u,t)|\cdot \tilde D_0(u)\cdot \mathbf{1}_{\{\mathrm{Ball}_{\tilde d}^{S_p}(u,t)=\varnothing\}}\Big| S_d, S_r\Big]\Big] \\
&\leqslant 5\cdot\mathbb{E}\Big[\sum_{u\in V_0} |\mathrm{Ball}_{\tilde d}(u,t)|\cdot \tilde D_0(u)\cdot \mathbb{E}[\mathbf{1}_{\{\mathrm{Ball}_{\tilde d}^{S_p}(u,t)=\varnothing\}} \mid S_d, S_r]\Big] \\
&= 5\cdot\mathbb{E}\Big[\sum_{u\in V_0} |\mathrm{Ball}_{\tilde d}(u,t)|\cdot \tilde D_0(u)\cdot (1-q(\varepsilon))^{|\mathrm{Ball}_{\tilde d}(u,t)|}\Big] \\
&= {}^{10}\!/\!_{\varepsilon^2}\cdot\mathbb{E}\Big[\sum_{u\in V_0} \tilde D_0(u)\Big] = {}^{10}\!/\!_{\varepsilon^2}\cdot {}^{2128}\!/\!_{\varepsilon^4}\cdot \mathsf{OPT}_1 = {}^{21280}\!/\!_{\varepsilon^6}\cdot \mathsf{OPT}_1.
\end{aligned}
$$

In the second line, the outer expectation is over S_d and S_r while the inner is over S_p. Then in the second to last line, we use Corollary 2, and in the last line we use Corollary 1. □

Bounding E_{1b}. For each $u \in V_0$ with $\mathrm{Ball}_{\tilde d}^{S_p}(u,t) \neq \varnothing$, choose a fixed but arbitrary $z(u) \in \mathrm{Ball}_{\tilde d}^{S_p}(u,t)$. Note that because $z(u) \in S_p$ and $\tilde d_{u,z(u)} \leqslant t \leqslant c\cdot r$, we know that $z(u)$ is a *candidate* for clustering u, so in particular $s^*(u)$ exists. Note $z(u)$ is a random variable depending on S_p, S_d, and S_r, but *not* on S_b.

Define

$$B(u) := \bigcup_{v\in E_1(u)} \mathrm{Ball}_{\tilde d}^{S_b}(s^*(v), r),$$

that is, $B(u)$ is the union of balls in S_b, cut out around the vertices that cluster the vertices in $E_1(u)$.

We repeat the following three claims from the proof of Lemma 6.

Claim 6. *If $\mathit{Ball}_{\tilde d}^{S_p}(u,t) \neq \varnothing$, then $|E_1(u)| \leqslant |\mathit{Ball}_{\tilde d}(z(u), r)|$.*

Claim 7. *If $u \in V_0$ and $\mathit{Ball}_{\tilde d}^{S_p}(u,t) \neq \varnothing$, then $|\mathit{Ball}_{\tilde d}^{S_b}(z(u), r)| \leqslant |B(u)|$.*

Claim 8. *If $u \in V_0$, then $|B(u)| \leqslant 8\cdot \tilde D_0(u)$.*

Claim 26. *For $t = {}^{r}\!/\!_{2\delta}$, it is the case that*

$$E_{1b} := \mathbb{E}\Big[\sum_{u\in V_0:\mathit{Ball}_{\tilde d}^{S_p}(u,t)\neq\varnothing} |E_1(u)|\Big] \leqslant O({}^{1}\!/\!_{\varepsilon^6})\cdot \mathit{OPT}_1.$$

The following proof will look a bit different from the analogous claim in the proof of Lemma 6, Claim 9. There, we split E_{1b} based on whether $|\mathrm{Ball}_{\tilde d}(z(u), r)|$ was above or below a threshold of $\Theta({}^{\log n}\!/\!_{\varepsilon^2})$. Here we will need to be less lossy.

Proof of Claim 26. We split the sum based on whether or not the ball around u's candidate cluster center $z(u)$, restricted to points in S_b, is non-empty. (Note this ball may be empty because $z(u)$ is not necessarily in S_b.) We then apply Claim 6.

$$
\begin{aligned}
E_{1b} &\leqslant \mathbb{E}\Big[\sum_{u\in V_0:\mathrm{Ball}_{\tilde d}^{S_p}(u,t)\neq\varnothing} |\mathrm{Ball}_{\tilde d}(z(u), r)|\Big] \\
&\leqslant \mathbb{E}\Big[\sum_{u\in V_0:\mathrm{Ball}_{\tilde d}^{S_p}(u,t)\neq\varnothing} |\mathrm{Ball}_{\tilde d}(z(u), r)|\cdot \mathbf{1}_{\{|\mathrm{Ball}_{\tilde d}^{S_b}(z(u),r)|\geqslant 1\}}\Big] \qquad (34)
\end{aligned}
$$

$$+ \mathbb{E}\Big[\sum_{u \in V_0 : \mathrm{Ball}_{\tilde{d}}^{S_p}(u,t) \neq \varnothing} |E_1(u)| \cdot \mathbf{1}_{\{|\mathrm{Ball}_{\tilde{d}}^{S_b}(z(u),r)|=0\}} \Big] \tag{35}$$

Define $\frac{1}{|\mathrm{Ball}_{\tilde{d}}^{S_b}(z(u),r)|} \cdot \mathbf{1}_{\{|\mathrm{Ball}_{\tilde{d}}^{S_b}(z(u),r)| \geqslant 1\}}$ to be 0 when $|\mathrm{Ball}_{\tilde{d}}^{S_b}(z(u),r)| = 0$. Then we upper bound the quantity in line (34):

$$\mathbb{E}\Big[\sum_{u \in V_0 : \mathrm{Ball}_{\tilde{d}}^{S_p}(u,t)} |\mathrm{Ball}_{\tilde{d}}(z(u),r)| \cdot \mathbf{1}_{\{|\mathrm{Ball}_{\tilde{d}}^{S_b}(z(u),r)| \geqslant 1\}} \Big]$$

$$\leqslant \mathbb{E}\Big[\sum_{u \in V_0 : \mathrm{Ball}_{\tilde{d}}^{S_p}(u,t)} 8 \cdot \tilde{D}_0(u) \cdot |\mathrm{Ball}_{\tilde{d}}(z(u),r)| \cdot \frac{1}{|\mathrm{Ball}_{\tilde{d}}^{S_b}(z(u),r)|} \cdot \mathbf{1}_{\{|\mathrm{Ball}_{\tilde{d}}^{S_b}(z(u),r)| \geqslant 1\}} \Big]$$

$$= \mathbb{E}\Bigg[\sum_{u \in V_0 : \mathrm{Ball}_{\tilde{d}}^{S_p}(u,t)} 8 \cdot \tilde{D}_0(u) \cdot |\mathrm{Ball}_{\tilde{d}}(z(u),r)| \cdot \mathbb{E}\Big[\frac{1}{|\mathrm{Ball}_{\tilde{d}}^{S_b}(z(u),r)|} \cdot \mathbf{1}_{\{|\mathrm{Ball}_{\tilde{d}}^{S_b}(z(u),r)| \geqslant 1\}} \Big| S_d, S_r, S_p \Big] \Bigg].$$

In the second line we have applied Claims 7 and 8. In the last line we have pulled out known factors. Finally, applying Lemma 5 with $A = |\mathrm{Ball}_{\tilde{d}}(z(u),r)|$, $S_* = S_b$, and $\mathcal{B} = \{S_d, S_r, S_p\}$, we continue upper bounding (34).

$$\mathbb{E}\Big[\sum_{u \in V_0 : \mathrm{Ball}_{\tilde{d}}^{S_p}(u,t)} \tilde{D}_0(u) \cdot |\mathrm{Ball}_{\tilde{d}}(z(u),r)| \cdot \frac{112}{\varepsilon^2 \cdot |\mathrm{Ball}_{\tilde{d}}(z(u),r)|} \Big]$$

$$= {}^{112}\!/\!{}_{\varepsilon^2} \cdot \mathbb{E}\Big[\sum_{u \in V_0} \tilde{D}_0(u) \Big] = {}^{112}\!/\!{}_{\varepsilon^2} \cdot {}^{2128}\!/\!{}_{\varepsilon^4} \cdot \mathsf{OPT}_1 = {}^{238336}\!/\!{}_{\varepsilon^6} \cdot \mathsf{OPT}_1,$$

where in the last line we have used Corollary 1.

Finally, we upper bound the quantity in line (35):

$$\mathbb{E}\Big[\sum_{u \in V_0 : \mathrm{Ball}_{\tilde{d}}^{S_p}(u,t) \neq \varnothing} |E_1(u)| \cdot \mathbf{1}_{\{|\mathrm{Ball}_{\tilde{d}}^{S_b}(z(u),r)|=0\}} \Big]$$

$$\leqslant 5 \cdot \mathbb{E}\Big[\sum_{u \in V_0 : \mathrm{Ball}_{\tilde{d}}^{S_p}(u,t) \neq \varnothing} \tilde{D}_0(u) \cdot |\mathrm{Ball}_{\tilde{d}}(z(u),r)| \cdot \mathbf{1}_{\{|\mathrm{Ball}_{\tilde{d}}^{S_b}(z(u),r)|=0\}} \Big]$$

$$= 5 \cdot \mathbb{E}\Bigg[\sum_{u \in V_0 : \mathrm{Ball}_{\tilde{d}}^{S_p}(u,t) \neq \varnothing} \tilde{D}_0(u) \cdot |\mathrm{Ball}_{\tilde{d}}(z(u),r)| \cdot \mathbb{E}\Big[\mathbf{1}_{\{|\mathrm{Ball}_{\tilde{d}}^{S_b}(z(u),r)|=0\}} \Big| S_d, S_r, S_p \Big] \Bigg]$$

$$= 5 \cdot \mathbb{E}\Big[\sum_{u \in V_0 : \mathrm{Ball}_{\tilde{d}}^{S_p}(u,t)} \tilde{D}_0(u) \cdot |\mathrm{Ball}_{\tilde{d}}(z(u),r)| \cdot (1 - q(\varepsilon))^{|\mathrm{Ball}_{\tilde{d}}(z(u),r)|} \Big]$$

$$= {}^{10}\!/\!{}_{\varepsilon^2} \cdot \mathbb{E}\Big[\sum_{u \in V_0} \tilde{D}_0(u) \Big] = {}^{10}\!/\!{}_{\varepsilon^2} \cdot {}^{2128}\!/\!{}_{\varepsilon^4} \cdot \mathsf{OPT}_1 = {}^{21280}\!/\!{}_{\varepsilon^6} \cdot \mathsf{OPT}_1.$$

In the first line we have used Claims 4 and 6. In the third line, we have taken out known quantities, as the outer expectation is over S_d, S_r, S_p and the inner expectation is over S_b. Then in the penultimate line we have used Corollary 2, and in the last line we have used Corollary 1.

Combining the upper bounds on (34) and (35) finishes the claim.

□

Combining the upper bounds in line (33) and Claims 25 and 26 finishes the lemma. □

B.1.2 Cost of negative edges

Lemma 17. *The expected cost of negative disagreements incurred during the Pre-clustering phase is*

$$\mathbb{E}\Big[\sum_{u \in V_0} |\{ v \in N_u^- : v \text{ clustered with } u \}| \Big] \leqslant O({}^{1}\!/\!{}_{\varepsilon^2}) \cdot \mathit{OPT}_1.$$

Proof of Lemma 17. If $u, v \in V_0$ are clustered together, there exists $s^* \in S_p$ such that $s^* = s^*(u) = s^*(v)$ (possibly with $s^* = u$ or $s^* = v$). By the approximate triangle inequality (Lemma 1),

$$\tilde{d}_{uv} \leqslant \delta \cdot (\tilde{d}_{us*} + \tilde{d}_{vs*}) \leqslant 2\delta cr \leqslant 7/10.$$

Since $\tilde{d}_{uv} < 1$, we know that $\tilde{d}_{uv} = \bar{d}_{uv}$ and also that $u \in R_2$.

Thus, to prove the lemma, it suffices to bound $\mathbb{E}\left[\sum_{u \in V_0 \cap R_2} |R_1(u)|\right]$. Since we bound this exact quantity in the proof of Lemma 4, the proof is complete. □

B.2 Cost of Pivot phase

Let $G' = (V', E')$ be the subgraph induced by the unclustered vertices, where V' is as defined in Algorithm 1. In this section, we bound the expected cost of disagreements in G'. Recall that V_0 is the set of eligible vertices, i.e., those vertices $v \in V$ such that $|N_v^+ \cap S_d| \neq \emptyset$. So V' contains the vertices that are *not* eligible, as well as vertices that are eligible but that are far from all vertices in S_p:

$$V' = \overline{V_0} \cup V_0', \text{ where } V_0' = V_0 \cap \{v \in V : \tilde{d}_{vu_i} > c \cdot r \text{ for all } u_i \in S_p\}.$$

Algorithm 1 runs the standard Pivot algorithm on $G[\overline{V_0}]$, and runs Modified Pivot on $G[V_0']$. In Lemma 19, we bound the disagreements incurred by Pivot on $G[\overline{V_0}]$, and in Lemma 20, we bound the disagreements incurred by Modified Pivot on $G[V_0']$.

B.2.1 Disagreements in $G[\overline{V_0}]$

The key idea is that if the expected positive degree of every vertex in a graph is small, then the cost of Pivot on that graph is small as well.

Fix an optimal clustering $\mathcal{C}_{\mathsf{OPT}}$ (on the entire graph G) for the ℓ_1-norm objective, and let $E^* \subseteq E$ be the disagreements in G with respect to $\mathcal{C}_{\mathsf{OPT}}$. For any $V_H \subseteq V$, let $H = (V_H, E_H)$ be the subgraph induced by V_H. Then $E^* \cap E_H$ is the set of disagreements in H with respect to $\mathcal{C}_{\mathsf{OPT}}$. Note that E^* is deterministic, but in context E_H may not be. Let $\mathsf{cost}_H(\mathsf{Pivot})$ be the number of disagreements incurred by running Pivot on H. For $i \in V_H$, let $\mathsf{deg}_H^+(i)$ be the positive degree of vertex i in H. The following lemma is proved in [22]:

Lemma 18 (Lemmas 2 and 3 in [22])**.** *Let $\mathcal{C}_{\mathsf{OPT}}, E^*, H, V_H, E_H$ be as above. Then the expected number of disagreements that (unmodified) Pivot[7] makes on H, $\mathbb{E}[\mathsf{cost}_H(\mathsf{Pivot})]$, is at most*

$$\mathbb{E}\Big[\sum_{uv \in E_H \cap E^*} (\mathsf{deg}_H^+(u) + \mathsf{deg}_H^+(v))\Big].$$

We bound the expected cost of the disagreements within $G[\overline{V_0}]$ in the following claim.

Lemma 19. *The expected cost of disagreements in $G[\overline{V_0}]$ is $O(1/\varepsilon^2) \cdot \mathsf{OPT}_1$.*

Proof of Lemma 19. To prove the claim, we apply Lemma 18 with $H = G[\overline{V_0}]$. We have

$$\begin{aligned}
\mathbb{E}[\mathsf{cost}_{G[\overline{V_0}]}(\mathsf{Pivot})] &\leqslant \sum_{uv \in E^*} \mathbb{E}\left[\mathbf{1}_{\{u \in \overline{V_0}\}} \cdot |N_u^+| + \mathbf{1}_{\{v \in \overline{V_0}\}} \cdot |N_v^+|\right] \\
&= \sum_{uv \in E^*} \left(|N_u^+| \cdot \mathbb{P}(u \in \overline{V_0}) + |N_v^+| \cdot \mathbb{P}(v \in \overline{V_0})\right) \\
&= \sum_{uv \in E^*} \left(|N_u^+| \cdot (1 - q(\varepsilon))^{|N_u^+|} + |N_v^+| \cdot (1 - q(\varepsilon))^{|N_v^+|}\right) \\
&= \sum_{uv \in E^*} 2/q(\varepsilon) = 4/\varepsilon^2 \cdot 1/2 \cdot \mathsf{OPT}_1 = 2/\varepsilon^2 \cdot \mathsf{OPT}_1.
\end{aligned}$$

□

[7]We do *not* assume that the vertices arrive in random order here, unlike, e.g., in the 3-approximation of [1].

B.2.2 Disagreements in $G[V_0']$

Next, we bound the expected cost of the disagreements in $G[V_0']$. Recall that on this graph, we run a *modified* version of the Pivot algorithm, where clusters are formed by vertices grabbing their *close*, positive neighbors (rather than just their positive neighbors as is done in classic Pivot). (See the definition of E_c in the second else statement in Algorithm 1).

Lemma 20. *The expected cost of disagreements in $G[V_0']$ is at most $O(1/\varepsilon^6) \cdot \mathit{OPT}_1$.*

Proof of Lemma 20. Let E_0' denote the edge set of $G[V_0']$. Let $E_{\mathsf{ALG}} \subseteq E_0'$ denote the disagreements made by $\mathcal{C}_{\mathsf{ALG}}$ on $G[V_0']$, that is, these are the disagreements incurred by running the modified version of Pivot on $G[V_0']$. Let E^* is the set of disagreements in an optimal solution $\mathcal{C}_{\mathsf{OPT}}$ for the ℓ_1-norm.

All disagreements in $E_{ALG} \cap E^*$ can be charged directly to OPT_1. We partition the remaining disagreements in E_{ALG} into three sets:

$$S_1 := \{uv \in E^+ \cap E_{\mathsf{ALG}} \cap \overline{E^*} \mid \tilde{d}_{uv} > c \cdot r\}, \qquad S_2 := \{uv \in E^+ \cap E_{\mathsf{ALG}} \cap \overline{E^*} \mid \tilde{d}_{uv} \leqslant c \cdot r\},$$
$$S_3 := E^- \cap E_{\mathsf{ALG}} \cap \overline{E^*}.$$

So

$$|E_{\mathsf{ALG}}| = |E_{\mathsf{ALG}} \cap E^*| + |S_1| + |S_2| + |S_3| \leqslant 1/2 \cdot \mathsf{OPT}_1 + |S_1| + |S_2| + |S_3|. \tag{36}$$

Bounding S_1. We charge the cost of all $uv \in S_1$ to $\tilde{d}$. Specifically,

$$\mathbb{E}[|S_1|] \leqslant 1/c \cdot r 1/2 \cdot \mathbb{E}\Big[\sum_{u \in V_0} \sum_{v \in N_u^+ \cap V_0} \tilde{d}_{uv}\Big] = 2962/\varepsilon^4 \cdot \mathsf{OPT}_1, \tag{37}$$

where we have applied Lemma 3 in the last equality.

Bounding S_2. For $uv \in S_2$, we have $\tilde{d}_{uv} \leqslant c \cdot r$. So there are only two ways for uv to be in E_{ALG}:

- The first is if uv is on a bad triangle (recall Definition 10) uvw where, WLOG, $wu \in E^+ \cap E_0'$, $wv \in E^- \cap E_0'$, and Pivot clusters w and u together, but not v. Since uvw is a bad triangle, and $uv \notin E^*$ by assumption, we have that either $wu \in E^*$ or $wv \in E^*$. We charge the cost of the disagreement uv made by $\mathcal{C}_{\mathsf{ALG}}$ to whichever one is in E^* (choosing one arbitarily if both wu and wv are in E^*).
- The second is if uv is on a triangle uvw of *all* positive edges such that, WLOG, $\tilde{d}_{wv} < c \cdot r$ but $\tilde{d}_{wu} \geqslant c \cdot r$, and Pivot clusters w and v together, but not u. Since $\tilde{d}_{wu} \geqslant c \cdot r$, we can charge the cost of the disagreement uv made by $\mathcal{C}_{\mathsf{ALG}}$ to $1/c \cdot r \cdot \tilde{d}_{wu}$.

We need to show that the edges in E^* are not charged too many times in the former case, and that $\tilde{d}$ is not charged too many times in the latter case. Call the subset of disagreements in S_2 satisfying the former case S_{2a}, and those satisfying the latter case S_{2b}.

Claim 27. $\mathbb{E}[|S_{2a}|] \leqslant O(1/\varepsilon^2) \cdot \mathit{OPT}_1$.

Proof of Claim 27. Fix $xy \in E^* \cap E_0'$. We need to compute how many times xy can be charged by edges in S_{2a}. For an edge in S_{2a} to charge xy, it must be that either one of its endpoints is x and the other is in $\mathrm{Ball}_{\tilde{d}}(x, c \cdot r)$; or, one of its endpoints is y and the other endpoint is in $\mathrm{Ball}_{\tilde{d}}(y, c \cdot r)$. Therefore,

$$\begin{aligned}
\mathbb{E}[|S_{2a}|] &\leqslant \mathbb{E}\Big[\sum_{xy \in E^* \cap E_0'} \big(|\mathrm{Ball}_{\tilde{d}}(x, c \cdot r)| + |\mathrm{Ball}_{\tilde{d}}(y, c \cdot r)|\big)\Big] \\
&\leqslant \sum_{xy \in E^*} \mathbb{E}\left[\mathbf{1}_{\{x \in V_0'\}} \cdot |\mathrm{Ball}_{\tilde{d}}(x, c \cdot r)| + \mathbf{1}_{\{y \in V_0'\}} \cdot |\mathrm{Ball}_{\tilde{d}}(y, c \cdot r)|\right]
\end{aligned}$$

$$
\begin{aligned}
&\leqslant \sum_{xy\in E*} \mathbb{E}\Big[\mathbf{1}_{\{\mathrm{Ball}_{\tilde d}^{S_p}(x,c\cdot r)=\varnothing\}}\cdot|\mathrm{Ball}_{\tilde d}(x,c\cdot r)|\Big] + \mathbb{E}\Big[\mathbf{1}_{\{\mathrm{Ball}_{\tilde d}^{S_p}(y,c\cdot r)\neq\varnothing\}}\cdot|\mathrm{Ball}_{\tilde d}(y,c\cdot r)|\Big]\\
&\leqslant \sum_{xy\in E*}\Bigg(\mathbb{E}\Big[|\mathrm{Ball}_{\tilde d}(x,c\cdot r)|\cdot\mathbb{E}\Big[\mathbf{1}_{\{\mathrm{Ball}_{\tilde d}^{S_p}(x,c\cdot r)=\varnothing\}}|S_d,S_r\Big]\Big]\\
&\qquad\qquad + \mathbb{E}\Big[|\mathrm{Ball}_{\tilde d}(y,c\cdot r)|\cdot\mathbb{E}\Big[\mathbf{1}_{\{\mathrm{Ball}_{\tilde d}^{S_p}(y,c\cdot r)=\varnothing\}}|S_d,S_r\Big]\Big]\Bigg)\\
&= \sum_{xy\in E*}\mathbb{E}[|\mathrm{Ball}_{\tilde d}(x,c\cdot r)|\cdot(1-q(\varepsilon))^{|\mathrm{Ball}_{\tilde d}(x,c\cdot r)|}] + \mathbb{E}[|\mathrm{Ball}_{\tilde d}(y,c\cdot r)|\cdot(1-q(\varepsilon))^{|\mathrm{Ball}_{\tilde d}(y,c\cdot r)|}]\\
&\leqslant {}^4\!/\!\varepsilon^2\cdot{}^1\!/\!2\cdot\mathsf{OPT}_1 = {}^2\!/\!\varepsilon^2\cdot\mathsf{OPT}_1,
\end{aligned}
$$

where in the penultimate line we used Corollary 2. ☐

Claim 28. $\mathbb{E}[|S_{2b}|]\leqslant O(1/\varepsilon^6)\cdot \mathit{OPT}_1$.

Proof of Claim 28. Fix $xy\in\{e\in E_0' \mid \tilde d_e\geqslant c\cdot r\}$. We need to upper bound how many times xy is charged by edges in S_{2b}. The situation is the same as in Claim 27 for S_{2a}: For an edge in S_{2b} to charge xy, it must be that either one of its endpoints is x and the other is in $\mathrm{Ball}_{\tilde d}(x,c\cdot r)$; or, one of its endpoints is y and the other endpoint is in $\mathrm{Ball}_{\tilde d}(y,c\cdot r)$. Therefore,

$$
\begin{aligned}
\mathbb{E}[|S_{2b}|] &\leqslant \mathbb{E}\Big[\sum_{\substack{xy\in E^+\cap E_0':\\ \tilde d_{xy}\geqslant c\cdot r}}\big(|\mathrm{Ball}_{\tilde d}(x,c\cdot r)|+|\mathrm{Ball}_{\tilde d}(y,c\cdot r)|\big)\Big]\\
&\leqslant \sum_{xy\in E^+} {}^1\!/\!c\cdot r\cdot\mathbb{E}\Big[\mathbf{1}_{\{x\in V_0'\}}\cdot\tilde d_{xy}\cdot|\mathrm{Ball}_{\tilde d}(x,c\cdot r)| + \mathbf{1}_{\{y\in V_0'\}}\cdot\tilde d_{xy}\cdot|\mathrm{Ball}_{\tilde d}(y,c\cdot r)|\Big]\\
&\leqslant {}^1\!/\!c\cdot r\cdot\sum_{xy\in E^+}\Bigg(\mathbb{E}\Big[\mathbf{1}_{\{\mathrm{Ball}_{\tilde d}^{S_p}(x,c\cdot r)=\varnothing\}}\cdot\tilde d_{xy}\cdot|\mathrm{Ball}_{\tilde d}(x,c\cdot r)|\Big]\\
&\qquad\qquad + \mathbb{E}\Big[\mathbf{1}_{\{\mathrm{Ball}_{\tilde d}^{S_p}(y,c\cdot r)=\varnothing\}}\cdot\tilde d_{xy}\cdot|\mathrm{Ball}_{\tilde d}(y,c\cdot r)|\Big]\Bigg)\\
&= \sum_{xy\in E^+}\frac{2}{c\cdot r\cdot q(\varepsilon)}\cdot\tilde d_{xy} = {}^{11845}\!/\!\varepsilon^6\cdot\mathsf{OPT}_1
\end{aligned}
$$

where in the last line we have conditioned on S_d, S_r, applied Corollary 2 in the same way as in the proof of Claim 27, and then finished off with Lemma 3. ☐

Bounding S_3. Finally, we bound $\mathbb{E}[|S_3|]$. Let $uv\in S_3$. Then the only way that u,v can be clustered together is if u,v are clustered by the same pivot w, where $w\in N_u^+\cap N_v^+$ and $\tilde d_{uw},\tilde d_{vw}\leqslant c\cdot r$. But then $1-\tilde d_{uv}\geqslant 1-2\delta cr$, so we have

$$
\mathbb{E}[|S_3|]\leqslant\frac{1}{1-2\delta cr}\cdot\mathbb{E}\left[\sum_{u\in V_0}\sum_{v\in N_u^-\cap V_0}(1-\tilde d_{uv})\right] = {}^{1129}\!/\!\varepsilon^2\cdot\mathsf{OPT}_1, \tag{38}
$$

where the last equality is by Lemma 4.

Continuing from line (36), and substituting in from lines (37) and (38) and Claims 27 and 28,

$$
\mathbb{E}[|E_{\mathsf{ALG}}|]\leqslant({}^1\!/\!2+{}^{2962}\!/\!\varepsilon^2+{}^2\!/\!\varepsilon^2+{}^{11845}\!/\!\varepsilon^6+{}^{1129}\!/\!\varepsilon^2)\cdot\mathsf{OPT}_1.
$$

☐

B.3 Cost between the Pre-clustering phase and the Pivot phase

In the next lemma, Lemma 21, we bound the cost of disagreeing positive edges where one endpoint is eligible and the other is ineligible. Recall a vertex u is ineligible ($u \in \overline{V_0}$) when its positive neighborhood is *not* sampled by S_d; in this case, the distances $\tilde{d}_{uv}$ are not meaningful, so we will not be able to charge the cost of these disagreements to $\tilde{d}$. Instead, we will use two different surrogates for optimal: bad triangles (Definition 10) and the fractional cost of the (true) correlation metric d (Definition 2). These will be useful for charging because every clustering must make at least one disagreement on each bad triangle, and the fractional cost of d –restricted to positive edges – is $O(1)$ approximate to optimal for the ℓ_1-norm objective (Lemma 2).

Lemma 21. *The expected cost of disagreements $uv \in E^+$ with u ineligible and v eligible is*

$$\mathbb{E}\Big[\sum_{u \in \overline{V_0}} |N_u^+ \cap V_0|\Big] \leqslant O(1/\varepsilon^4) \cdot \mathit{OPT}_1.$$

Proof of Lemma 21. First note that the second bound in the statement of the lemma follows from the first bound, since $v \succ u$ implies v is pre-clustered, so $v \in V_0$.

Define $D^+(u)$ to be the fractional cost of the positive edges incident to u with respect to the (actual) correlation metric d, that is, $D^+(u) := \sum_{v \in N_u^+} d_{uv}$. It suffices to bound

$$\mathbb{E}\Big[\sum_{u \in \overline{V_0}:\ D^+(u) > 1/4} |N_u^+|\Big] + \mathbb{E}\Big[\sum_{u \in \overline{V_0}:\ D^+(u) \leqslant 1/4} |N_u^+ \cap V_0|\Big],$$

where the expectation is taken over the randomness of the sample S_d. Bounding the first term is straightforward. We use Corollary 2 to see that $\mathbb{P}[u \in \overline{V_0}] = \prod_{v \in N_u^+} \mathbb{P}(v \notin S_d) = (1 - q(\varepsilon))^{|N_u^+|}$. Then, we charge the cost of $|N_u^+|$ to $D^+(u)$, and bound the fractional cost of positive edges with Lemma 2.

$$\begin{aligned}\mathbb{E}\Big[\sum_{u \in \overline{V_0}:\ D^+(u) > 1/4} |N_u^+|\Big] &\leqslant 4 \cdot \mathbb{E}\left[\sum_{u \in V} |N_u^+| \cdot D^+(u) \cdot \mathbf{1}_{\{u \in \overline{V_0}\}}\right] \\ &= 4 \cdot \sum_{u \in V} D^+(u) \cdot |N_u^+| \cdot (1 - q(\varepsilon))^{|N_u^+|} \leqslant 8/\varepsilon^2 \cdot \sum_{u \in V} D^+(u) \leqslant 24/\varepsilon^2 \cdot \mathsf{OPT}_1.\end{aligned}$$

To bound the second term, fix $u \in \overline{V_0}$. Suppose that u is part of a perfect clique, i.e., u is incident to no bad triangles. Then if $w \in N_v^+$ and $v \in N_u^+$, then $w \in N_u^+$. If this is the case, then $v \in \overline{V_0}$ for each $v \in N_u^+$; for, if $N_v^+ \cap S_d \neq \varnothing$, then $N_u^+ \cap S_d \neq \varnothing$, since $N_u^+ = N_v^+$. Thus, none of u's positive neighbors v will be in V_0, and therefore u will not contribute to the second sum above.

Fix an optimal clustering $\mathcal{C}_{\mathsf{OPT}}$ for the ℓ_1-norm objective, and let E^* be the disagreements with respect to $\mathcal{C}_{\mathsf{OPT}}$. By the above, we may assume that the second sum above is further restricted to u such that u is incident to a bad triangle. Thus we may arbitrarily map each u to a bad triangle $T(u)$ that is incident to u. In turn, since $\mathcal{C}_{\mathsf{OPT}}$ must make a disagreement on $T(u)$, we may map u to $e(u) \in E^*$ such that $e(u)$ is on $T(u)$.

It now suffices to bound $\mathbb{E}[\sum_{u \in \overline{V_0}:\ D^+(u) \leqslant 1/4} |N_u^+|]$ (where the sum is restricted to u that is incident to a bad triangle, though we omit this for ease). The idea is to charge the cost of $|N_u^+|$ to the disagreement e_u. We have

$$\mathbb{E}\left[\sum_{\substack{u \in \overline{V_0}: \\ D^+(u) \leqslant 1/4}} |N_u^+|\right] = \sum_{e \in E^*} \sum_{\substack{u:\ e = e(u): \\ D^+(u) \leqslant 1/4}} \mathbb{E}\left[|N_u^+| \cdot \mathbf{1}_{\{u \in \overline{V_0}\}}\right] = \sum_{e \in E^*} \sum_{\substack{u:\ e = e(u), \\ D^+(u) \leqslant 1/4}} |N_u^+| \cdot (1 - q(\varepsilon))^{|N_u^+|},$$

where again we use Corollary 2 to see that $\mathbb{P}(u \in \overline{V_0}) = (1 - q(\varepsilon))^{|N_u^+|}$. We now bound the inner sum. Fix $e = ab \in E^*$ such that e is on a bad triangle T. Now consider u such that $D^+(u) \leqslant 1/4$ and $T = T(u)$. By definition, u is on $T(u)$. Since T is a bad triangle, $u \in N_a^+$ or $u \in N_b^+$ (or both) (where we may have $u = a$ or $u = b$, due to self-loops).

Consider $u \in N_a^+$. Since $D^+(u) \leqslant 1/4$, we have $d_{ua} \leqslant 1/4$. It is then straightforward to see from $d_{ua} = 1 - {|N_u^+ \cap N_a^+|}/{|N_u^+ \cup N_a^+|}$ that $|N_u^+| \leqslant 4/3 \cdot |N_a^+|$ and $|N_a^+| \leqslant 4/3 \cdot |N_u^+|$. The analogous conclusion holds when $u \in N_b^+$. Continuing from the above,

$$\begin{aligned}
&\sum_{e \in E^*} \sum_{\substack{u:\, e=e(u),\\ D^+(u) \leqslant 1/4}} |N_u^+| \cdot (1-q(\varepsilon))^{|N_u^+|} \\
\leqslant &\sum_{ab \in E^*} \Big(\sum_{u \in N_a^+:\, d_{ua} \leqslant 1/4} |N_u^+| \cdot (1 - {\varepsilon^2}/{2})^{|N_u^+|} + \sum_{u \in N_b^+:\, d_{ub} \leqslant 1/4} |N_u^+| \cdot (1 - {\varepsilon^2}/{2})^{|N_u^+|} \Big) \\
\leqslant &\sum_{ab \in E^*} \Big(\sum_{u \in N_a^+} \frac{4}{3} \cdot |N_a^+| \cdot (1 - {\varepsilon^2}/{2})^{\frac{3}{4} \cdot |N_a^+|} + \sum_{u \in N_b^+} \frac{4}{3} \cdot |N_b^+| \cdot (1 - {\varepsilon^2}/{2})^{\frac{3}{4} \cdot |N_b^+|} \Big) \\
\leqslant &\frac{4}{3} \cdot \sum_{ab \in E^*} \Big(|N_a^+|^2 \cdot (1 - {\varepsilon^2}/{2})^{\frac{3}{4} \cdot |N_a^+|} + |N_b^+|^2 \cdot (1 - {\varepsilon^2}/{2})^{\frac{3}{4} \cdot |N_b^+|} \Big) = 4/3 \cdot \sum_{e \in E^*} {16}/{\varepsilon^4} = {64}/{3\varepsilon^4} \cdot \mathsf{OPT}_1.
\end{aligned}$$

Note in the last line we use the fact that $n^2 \cdot (1-x)^{n \cdot 3/4} \leqslant 2/x^2$. This concludes the proof. □

B.4 Proof of item 3 for Theorem 1

In this subsection, we show that the cases considered in the preceding lemmas are exhaustive, and thus that we can prove item 3 of Theorem 1.

Proof of item 3 for Theorem 1. Let $\mathcal{C}_{\mathsf{ALG}}$ be the clustering output by Algorithm 1, and let $\text{cost}_1(\mathcal{C}_{\mathsf{ALG}})$ be the ℓ_1-norm of the disagreement vector of $\mathcal{C}_{\mathsf{ALG}}$. We further partition the edges in disagreement based on whether they are positive or negative, and which phase in Algorithm 1 that are clustered in: positive edges where at least one endpoint of the edge is pre-clustered (Lemmas 16 and 21), negative edges where both endpoints are pre-clustered (Lemma 17), edges where both endpoints are clustered in the Pivot phase (Lemmas 19 and 20). Note this is exhaustive for the disagreements; see Figure 4. Combining the terms from the lemmas, we see

$$\mathbb{E}[||y_{\mathcal{C}_{\mathsf{ALG}}}||_1] \leqslant O({1}/{\varepsilon^6}) \cdot \mathsf{OPT}_1.$$

□

C Tight Analysis of Algorithm 1 for $p = \infty$

As in the analysis for the ℓ_1-norm, for the ℓ_∞-norm we bound the disagreements by partitioning edges based on their label and in which phase their endpoints were clustered. However, we are able to partition edges into fewer types than for the finite ℓ_p-norms (see Figure 6, versus Figures 1 and 4), and the analysis is in some sense simpler. For the ℓ_∞-norm objective, we know there exists a lower bound of $\Omega(\log n)$ in the online-with-a-sample model (Theorem 2). Intuitively, this allows us to essentially sacrifice vertices that have either small (e.g., $O(\log n)$) positive neighborhoods or a small number of close neighbors with respect to $\bar{d}$ (recall this is the estimated correlation metric of Definition 4). On the other hand, for vertices that have large positive neighborhoods and a large number of neighbors with respect to $\bar{d}$, these sets have good concentration, and thus we can obtain good estimates for $\bar{d}$ and $|\text{Ball}_{\bar{d}}^{S_b}(u, r)|$.

As with the analysis of the ℓ_p-norm ($p \neq 1$) in Section 5, we condition on the *good event* B^c (Definition 12) occurring with high probability (Lemma 15). Here we will use even more of the events defining B^c than in the previous analysis, e.g., we will use that B^c gives that for pairs of nodes u, v with large combined positive neighborhood, $\bar{d}_{uv}$ is a good estimate of d_{uv}.

C.1 Cost of Pre-clustering phase

In this section, we bound the cost of disagreements incident to a vertex u that arise when u, or a positive neighbor of u, is pre-clustered. Recall that $s^*(v)$ is the vertex in S_p that clusters v, i.e. v's center. Also

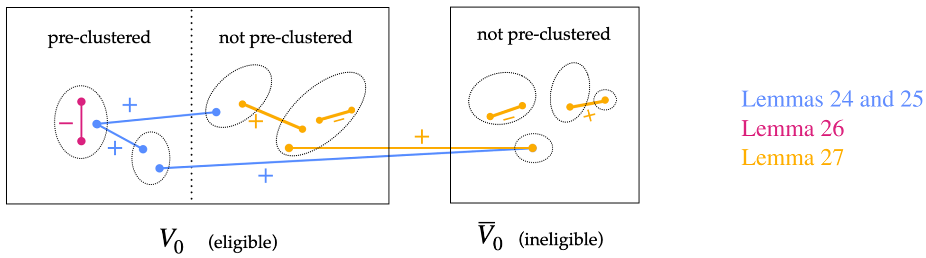

Figure 6: An overview of the cost analysis of Algorithm 1 for $p = \infty$. The edges, ovals, etc., are as in Figures 1 and 4.

recall we write $u \succ v$ if u is pre-clustered before u (with respect to the ordering on S_p, not the arrival order), or if u is pre-clustered and v is not. Recall the vertices R_1 are those isolated by $\tilde{d}$ (Definition 5). Throughout, C will be a constant chosen sufficiently large, for instance $C = 100$ is sufficient.

In Lemma 24, we bound the cost of disagreements incurred for vertex u from positive edges uv, where $v \succ u$. These disagreements are similar to those we charged in Lemma 6 for the finite ℓ_1-norms, although we do not require $u \in V_0$. Then in Lemma 25, we bound the cost of disagreements incurred for vertex u from positive edges uv, where $u \succ v$ (cf. Lemma 7). Then, Lemma 26 bounds the cost of negative edges whose endpoints were clustered together during the Pre-clustering phase.

C.1.1 Cost of positive edges

We first prove a few propositions and lemmas that will be helpful in bounding the costs of disagreements for both Lemmas 24 and 25.

We begin with Proposition 3, which says if $\mathsf{OPT}_p = 0$,[8] then Algorithm 1 finds the perfect clustering. This is necessary for our algorithm to be competitive. Its proof actually follows from the fact that Algorithm 1 has expected cost a bounded factor away from the optimal cost when $p = 1$ (i.e., item 3 of Theorem 1 is sufficient). However, we include below a more direct argument that is entirely independent of this result.

Proposition 3. *If $\mathsf{OPT}_p = 0$, then Algorithm 1 finds the perfect clustering, that is, the unique clustering with 0 disagreements.*

Proof of Proposition 3. Let G be a graph admitting a perfect clustering. Let Q be any maximal positive clique in G. We need to show that Q is precisely its own cluster in the clustering produced by Algorithm 1.

Claim 29. *If $v \in V_0$, then $N_v^+ \subseteq V_0$. In other words, if any vertex of Q is in V_0, then all of Q is contained in V_0.*

Proof of Claim 29. Recall that $V_0 = \{v \in V : N_v^+ \cap S_d \neq \varnothing\}$. So if $v \in V_0$, then there exists $w \in N_v^+$ that is in S_d. But for any $u \in N_v^+$, we also know that $w \in N_u^+$, since v and u have the same positive neighborhood. Thus u has a positive neighbor (namely w) in S_d, meaning $u \in V_0$. □

Claim 30. *If $v \in Q$ and $u \notin Q$, then $\tilde{d}_{uv} = 1$. Conversely, if $Q \subseteq V_0$, then $\tilde{d}_{uv} = 0$ for all $u, v \in Q$.*

Proof of Claim 30. If $u \notin V_0$, then $\tilde{d}_{uv} = 1$ by Fact 1. If $u \in V_0$, then $\bar{d}_{uv} = 1$, since u and v have no common positive neighbors. Thus $\tilde{d}_{uv} = 1$, since $\tilde{d}_{uv} \geqslant \bar{d}_{uv}$ always.

For the converse, note that $\bar{d}_{uv} = 0$, since $N_u^+ \cap N_v^+ \cap S_d = (N_u^+ \cup N_v^+) \cap S_d$, and $(N_u^+ \cup N_v^+) \cap S_d \neq \varnothing$. Moreover, by the first part of the claim, no negative edges have distance with respect to $\bar{d}$ strictly less than 1, so $\bar{d} = \tilde{d}$ (i.e., the rounding steps in Definition 4 do not apply). In particular, $\bar{d}_{uv} = \tilde{d}_{uv} = 0$ for $u, v \in Q$ when $Q \subseteq V_0$. □

Case 1. $Q \cap S_p \neq \varnothing$.

[8] When this holds for some p, it holds for all p.

The following subcases are exhaustive, by Claim 29.

- **Case 1a:** $Q \subseteq V_0$**.**
Let $v \in C \cap S_p$ be chosen minimally with respect to the ordering on S_p. If $v \succ v'$, then $v' \notin Q$, and then by Claim 30, $\tilde{d}_{uv'} = 1$ for every $u \in Q$, so no vertex in Q will be pre-clustered by v'. Thus, since $Q \subseteq V_0$ and $\tilde{d}_{uv} = 0$ for every $u \in Q$ (Claim 30), all of Q will be pre-clustered by v. Moreover, if $u \notin Q$, then u will not be pre-clustered by any $v \in Q$, again by Claim 30.
- **Case 1b:** $Q \cap V_0 = \varnothing$**.**
Then the vertices Q are clustered in the run of Modified Pivot on $G[\overline{V_0}]$, where in this case Modified Pivot is just the standard Pivot algorithm, which makes each maximal positive clique its own cluster.

Case 2. $Q \cap S_p = \varnothing$

In this case, no vertex in Q will be pre-clustered, because $\tilde{d}_{uv} = 1$ for every $u \notin Q$ and $v \in Q$ (Claim 30). If $Q \cap V_0 = \varnothing$, the argument is as in Case 1b. If $Q \subseteq V_0$, then Q is clustered in the run of Modified Pivot on $G[V_0 \cap V']$. Modified Pivot in this case takes the set $E_c = E^+ \cap \{uv \in E : \tilde{d}_{uv} < c \cdot r\}$ of clusterable edges as input. But by Claim 30, $E_c = E^+$, so Modified Pivot is just the standard Pivot algorithm, which will put Q in its own cluster. ☐

Given Proposition 3, we may assume in the remainder of this section that $\mathsf{OPT}_\infty \geqslant 1$.

Recall in Lemma 14 we showed with high probability $\bar{d}_{uv}$ is a good estimate of d_{uv}, so long as $|N_u^+ \cup N_v^+|$ is large. We do not obtain an analogous statement for $\tilde{d}$. In lieu of this, we show in Proposition 4 that for vertices u affected by the (second) adjustment in Definition 5 (that is, the vertices $u \in R_1$ isolated by $\tilde{d}$) we have that $|N_u^+|$ is a bounded factor away from OPT_∞ with high probability.[9] This bound is necessary, as for these vertices, the number of disagreements incident to u is $|N_u^+|$.

Proposition 4. *Condition on the good event B^c. Fix $u \in R_1$ (the vertices isolated by $\tilde{d}$) such that $|N_u^+| \geqslant C \cdot \log n/\varepsilon^2$. Then $|N_u^+| \leqslant O(1/\varepsilon^4 \cdot \log n) \cdot \mathsf{OPT}_\infty$.*

Proof of Proposition 4. Since the event B^c holds, $|N_u^+ \cap S_d| \geqslant \frac{\varepsilon^2}{4} \cdot |N_u^+|$. Also, when $u \in R_1$, $|N_u^- \cap S_r \cap \{v \in V : \bar{d}_{uv} \leqslant 7/10\}| \geqslant \frac{10}{3} \cdot |N_u^+ \cap S_d|$. Combining the above two bounds, we have for $u \in R_1$,

$$|N_u^+| \leqslant \frac{3}{10} \cdot \frac{4}{\varepsilon^2} \cdot |N_u^- \cap \{v \in V : \bar{d}_{uv} \leqslant 7/10\}| \leqslant \frac{3}{10} \cdot \frac{4}{\varepsilon^2} \cdot \sum_{v \in N_u^-} \frac{10}{3} \cdot (1 - \bar{d}_{uv})$$

giving the first inequality in the statement of the proposition. Then by the event B^c (which says $\bar{d}$ is a good estimate of d since $|N_u^+|$ is sufficiently large) and Theorem 4 (which says that $||D||_\infty \leqslant 8 \cdot \mathsf{OPT}_1$), then $|N_u^+| \leqslant 4/\varepsilon^2 \cdot \sum_{v \in N_u^-}(1 - \bar{d}_{uv}) \leqslant 4(C+1)/3\varepsilon^4 \cdot \log n \cdot D(u) \leqslant 32(C+1)/3\varepsilon^4 \cdot \log n \cdot \mathsf{OPT}_\infty$. ☐

In the following two lemmas, Lemmas 22 and 23, we charge the cost of the positive neighborhood of a vertex to the cost of the estimated correlation metric $\bar{d}$, which we in turn know is comparable to OPT_∞ by the good event.

The proof of Lemma 22 relies on the fact that since $\bar{d}_{uv}$ is small, the joint sampled neighborhood $|N_u^+ \cap N_v^+ \cap S_d|$ is large, and moreover $|N_u^+ \cap S_d|$ and $|N_v^+ \cap S_d|$ are similar. Since we condition on the good event, $|N_v^+|$ and $|N_u^+|$ are similar too. If both neighborhood sizes $|N_u^+|$ and $|N_v^+|$ are small, the claim automatically follows (using the fact that $\mathsf{OPT}_\infty \geqslant 1$), and if they are large, then we can upper bound them by the ℓ_∞-norm with the help of Proposition 4.

Lemma 22. *Condition on the good event B^c. Fix $u, v \in V$. Suppose $|N_u^+| \geqslant C \cdot \log n/\varepsilon^2$, $\bar{d}_{u,v} \leqslant 1/4$, and $\tilde{d}_{u,v} = 1$. Then $|N_u^+| \leqslant O(1/\varepsilon^6 \cdot \log n) \cdot \mathsf{OPT}_\infty$.*

[9] Note that even though $|N_u^+|$ and OPT_∞ are deterministic quantities, the statement is probabilistic due to the event $u \in R_1$.

Proof of Lemma 22. Given the hypotheses of the lemma, the edge uv must have been rounded up in estimating the adjusted correlation metric (Definition 5). So either $u \in R_1$ or $v \in R_1$. If $u \in R_1$, we have by Proposition 4 that $|N_u^+| \leqslant {}^{32\cdot(C+1)}/_{3\varepsilon^4} \cdot \log n \cdot \mathsf{OPT}_\infty$.

Otherwise, $v \in R_1$. Since $\bar{d}_{uv} \leqslant 1/4$, we have that $|N_u^+ \cap S_d|$ and $|N_v^+ \cap S_d|$ are similar, as in the following fact. Recall we define the random variable X_w be 1 if $w \in S_d$ and 0 otherwise.

Fact 4. *For u, v with $\bar{d}_{uv} \leqslant 1/4$, we have that $|N_u^+ \cap S_d| \leqslant \frac{4}{3} \cdot |N_v^+ \cap S_d|$.*

Proof of Fact 4. We rewrite $\bar{d}_{uv}$ as

$$\bar{d}_{uv} = \frac{\sum_{w \in N_u^+ \setminus N_v^+} X_w + \sum_{w \in N_v^+ \setminus N_u^+} X_w}{\sum_{w \in N_u^+ \cup N_v^+} X_w}.$$

Fixing any realization of S_d, we can further rewrite $\bar{d}_{uv} \leqslant 1/4$ as

$$\begin{aligned} &|(N_u^+ \setminus N_v^+) \cap S_d| + |(N_v^+ \setminus N_u^+) \cap S_d| \\ &\qquad \leqslant \frac{1}{4} \cdot \big(|(N_u^+ \setminus N_v^+) \cap S_d| + |(N_v^+ \setminus N_u^+) \cap S_d| + |(N_u^+ \cap N_v^+) \cap S_d|\big), \end{aligned}$$

which is equivalent to

$$\begin{aligned} \frac{3}{4} \cdot |(N_u^+ \setminus N_v^+) \cap S_d| + \frac{3}{4} \cdot |(N_v^+ \setminus N_u^+) \cap S_d| &\leqslant \frac{1}{4} \cdot |(N_u^+ \cap N_v^+) \cap S_d| \\ |(N_u^+ \setminus N_v^+) \cap S_d| + |(N_v^+ \setminus N_u^+) \cap S_d| &\leqslant \frac{1}{3} \cdot |(N_u^+ \cap N_v^+) \cap S_d|. \end{aligned}$$

Therefore,

$$\begin{aligned} |N_u^+ \cap S_d| &= |(N_u^+ \cap N_v^+) \cap S_d| + |(N_u^+ \setminus N_v^+) \cap S_d| \\ &\leqslant |(N_u^+ \cap N_v^+) \cap S_d| + \frac{1}{3} \cdot |(N_u^+ \cap N_v^+) \cap S_d| - |(N_v^+ \setminus N_u^+) \cap S_d| \\ &= \frac{4}{3} \cdot |N_u^+ \cap N_v^+ \cap S_d| - |(N_v^+ \setminus N_u^+) \cap S_d| \\ &\leqslant \frac{4}{3} \cdot |N_v^+ \cap S_d|. \end{aligned}$$

□

Further, since B^c holds, ${}^{\varepsilon^2}/_4 \cdot |N_u^+| \leqslant |N_u^+ \cap S_d|$. Also, $|N_v^+ \cap S_d| \leqslant |N_v^+|$. Combining the above inequalities with Fact 4, we have

$$|N_u^+| \leqslant \frac{4}{\varepsilon^2} \cdot \frac{4}{3} \cdot |N_v^+|.$$

If $|N_v^+| < C \cdot \log n/\varepsilon^2$, then we are done, with $|N_u^+| \leqslant {}^{16C}/_{3\varepsilon^4} \cdot \log n$.

Otherwise, $|N_v^+| \geqslant C \cdot \log n/\varepsilon^2$, and we have by Proposition 4 that $|N_v^+| \leqslant {}^{32(C+1)}/_{3\varepsilon^4} \cdot \log n \cdot \mathsf{OPT}_\infty$, so $|N_u^+| \leqslant {}^{512(C+1)}/_{9\varepsilon^6} \cdot \log n \cdot \mathsf{OPT}_\infty$. □

Lastly, Lemma 23 shows that when conditioning on the good event, the number of nodes far from u with respect to $\tilde{d}$ is upper bounded by $\log n \cdot O(\mathsf{OPT}_\infty)$. The proof follows automatically by Proposition 4 if $u \in R_1$. If $u \notin R_1$, we can apply Lemma 22 if there is a v that satisfies the conditions of that lemma. Otherwise we are able to use the the good event, and conclude that $\bar{D}(u)$ is upper bounded by $O\big({}^1/_{\varepsilon^2} \cdot \log n\big) \cdot \mathrm{OPT}_\infty$.

Lemma 23. *Condition on the good event B^c. Fix $u \in V$. Suppose $|N_u^+| \geqslant C \cdot \log n/\varepsilon^2$, and let $s > 0$ be a constant. Then $|\{v \in N_u^+ : \tilde{d}_{uv} \geqslant s\}| \leqslant O\left({}^1/_{\varepsilon^6} \cdot \log n\right) \cdot \mathsf{OPT}_\infty$.*

Proof of Lemma 23. We consider two cases.

Case 1. $u \in R_1$.

We use the bound $|\{v \in N_u^+ : \tilde{d}_{uv} \geqslant s\}| \leqslant |N_u^+|$. Since $|N_u^+| \geqslant C \cdot \log n/\varepsilon^2$, we may apply Proposition 4, and obtain

$$|\{v \in N_u^+ : \tilde{d}_{uv} \geqslant s\}| \leqslant |N_u^+| \leqslant {}^{(32\cdot(C+1)\cdot\log n)}\!/\!{}_{3\varepsilon^4} \cdot \mathsf{OPT}_\infty.$$

Case 2. $u \notin R_1$.

We consider two subcases.

- If there exists $v^* \in N_u^+ \cap R_1 \cap \{v \in V : \bar{d}_{uv} \leqslant 1/4\}$, then $\bar{d}_{uv^*} \leqslant 1/4$ but $\tilde{d}_{uv^*} = 1$ (because $v^* \in R_1$). So we may apply Lemma 22 and obtain that

$$|\{v \in N_u^+ : \tilde{d}_{uv} \geqslant s\}| \leqslant |N_u^+| \leqslant {}^{(512\cdot(C+1)\cdot\log n)}\!/\!{}_{9\varepsilon^6} \cdot \mathsf{OPT}_\infty.$$

- Otherwise, $N_u^+ \cap R_1 \cap \{v \in V : \bar{d}_{uv} \leqslant 1/4\} = \varnothing$. Then

$$|\{v \in N_u^+ : \tilde{d}_{uv} \geqslant s\}| \leqslant \sum_{v \in N_u^+ \setminus R_1} {}^{1}\!/\!{}_{s} \cdot \bar{d}_{uv} + \sum_{v \in N_u^+ \cap R_1} 4 \cdot \bar{d}_{uv} \leqslant {}^{(8(C+1)(4+1/s)\cdot\log n)}\!/\!{}_{3\varepsilon^2} \cdot \mathsf{OPT}_\infty.$$

  In the first inequality we have used that if $u, v \notin R_1$ and $uv \in E^+$, then $\bar{d}_{uv} = \tilde{d}_{uv}$, along with the fact that $\tilde{d}_{uv} \geqslant s$ by assumption. In the second inequality, we have used the fact that $|N_u^+| \geqslant C \cdot \log n/\varepsilon^2$ and applied Lemma 14.

□

Equipped with all the necessary lemmas, we are ready to prove Lemmas 24 and 25.

Lemma 24. *Condition on B^c. Fix $u \in V$. Then $|\{v \in N_u^+ \mid v \succ u\}| \leqslant O\left({}^{1}\!/\!{}_{\varepsilon^6} \cdot \log n\right) \cdot \mathit{OPT}_\infty$.*

Proof of Lemma 24. We may assume $\mathsf{OPT}_\infty \geqslant 1$ per Proposition 3. Take $t := r/(2\delta)$. We have

$$\{v \in N_u^+ \mid v \succ u\} = \underbrace{\{v \in N_u^+ \mid v \succ u, \tilde{d}_{uv} \leqslant t\}}_{E_1(u)} + \underbrace{\{v \in N_u^+ \mid v \succ u, \tilde{d}_{uv} > t\}}_{E_2(u)}$$

If $|N_u^+| < C \cdot \log n/\varepsilon^2$, then we are done, so we assume for the remainder of the proof that $|N_u^+| \geqslant C \cdot \log n/\varepsilon^2$. In this case, since B^c holds, $|N_u^+ \cap S_d| \geqslant C \cdot \log n/4 \geqslant 1$, so $u \in V_0$.

Bounding $|E_2(u)|$. Applying Lemma 23 with $s = t = {}^{r}\!/\!{}_{(2\delta)}$, we have

$$|E_2(u)| \leqslant {}^{(512(C+1)\cdot\log n)}\!/\!{}_{9\varepsilon^6} \cdot \mathsf{OPT}_\infty. \tag{39}$$

Bounding $|E_1(u)|$. We case based on $|\mathrm{Ball}_{\bar{d}}(u,t)|$.

Case 1. $|\mathrm{Ball}_{\bar{d}}(u,t)| \leqslant C' \cdot \log_{1/(1-\varepsilon^2/2)} n$. We know that $\bar{d} \leqslant \tilde{d}$, so $E_1(u) \subseteq \mathrm{Ball}_{\bar{d}}(u,t)$. Thus

$$|E_1(u)| \leqslant C' \cdot \log_{1/(1-\varepsilon^2/2)} n \leqslant C' \cdot \log_{1/(1-\varepsilon^2/2)} n \cdot \mathsf{OPT}_\infty,$$

where the second inequality is from Proposition 3.

Case 2. $|\mathrm{Ball}_{\bar{d}}(u,t)| > C' \cdot \log_{1/(1-\varepsilon^2/2)} n$. We use the next two claims to bound $|E_1(u)|$. Define

$$B(u) := \bigcup_{v \in E_1(u)} \mathrm{Ball}_{\tilde{d}}(s^*(v), r).$$

Claim 31. *For every $b \in B(u)$, $\tilde{d}_{ub} > 1/10$ and $1 - \tilde{d}_{ub} > 3/10$.*

Proof of Claim 31. First we lower bound $\tilde{d}_{ub}$ for $b \in B(u)$. Since $b \in B(u)$, there exists $v \in E_1(u)$ such that $\tilde{d}_{b\,s*(v)} \leqslant r$. On the other hand, since $u \in V_0$ and v is clustered before u, $\tilde{d}_{u\,s*(v)} > c \cdot r$. By the approximate triangle inequality (Lemma 1), $\tilde{d}_{ub} \geqslant \frac{1}{\delta} \cdot \tilde{d}_{u\,s*(v)} - \tilde{d}_{b\,s*(v)}$, and rearranging and substituting in c, r, δ, we see $\tilde{d}_{ub} \geqslant c \cdot r/\delta - r > 1/10$.

Next we upper bound $\tilde{d}_{ub}$ for $b \in B(u)$. Let v and $s*(v)$ be as before. We have $\tilde{d}_{b\,s*(v)} \leqslant r$, $\tilde{d}_{v\,s*(v)} \leqslant c \cdot r$, and $\tilde{d}_{uv} \leqslant t$. By Lemma 1 and substituting in c, r, δ, we have $\tilde{d}_{ub} \leqslant \delta \cdot [\tilde{d}_{b\,s*(v)} + \delta(\tilde{d}_{v\,s*(v)} + \tilde{d}_{vu})] \leqslant \delta \cdot r + \delta^2 \cdot c \cdot r + \delta^2 \cdot t < 7/10$. ☐

Claim 32. *At least one of the following holds:* $|E_1(u)| \leqslant \frac{8}{\varepsilon^2} \cdot |B(u)|$ *or* $|E_1(u)| \leqslant O\left(1/\varepsilon^6 \cdot \log n\right) \cdot$ *OPT*$_\infty$.

Proof of Claim 32. Since B^c holds, there exists $w \in \mathrm{Ball}_{\tilde{d}}^{S_p}(u, t)$. We consider two cases:

Case 1. $\tilde{d}_{uw} = \bar{d}_{uw}$. In this case, $\tilde{d}_{uw} \leqslant t \leqslant c \cdot r$. So since $w \in S_p$ and $u \in V_0$, we know that u is pre-clustered (but not necessarily by w). So $s*(u)$ exists. For every $v \in E_1(u)$, we have that $\tilde{d}_{uv} \leqslant t$, and also $\tilde{d}_{uw} \leqslant t$ (by assumption of the case). So by the approximate triangle inequality (Lemma 1), $\tilde{d}_{vw} \leqslant \delta \cdot (\tilde{d}_{vu} + \tilde{d}_{uw}) \leqslant \delta \cdot 2t = r$ for every $v \in E_1(u)$, which means that

$$|E_1(u)| \leqslant |\mathrm{Ball}_{\tilde{d}}(w, r)|.$$

If $|\mathrm{Ball}_{\tilde{d}}(w, r)| < C \cdot \log n / q(\varepsilon)$, then we are done. So assume $|\mathrm{Ball}_{\tilde{d}}(w, r)| \geqslant C \cdot \log n / q(\varepsilon)$. Then, since B^c holds, we have that $|\mathrm{Ball}_{\tilde{d}}(w, r)| \leqslant 4 \cdot |\mathrm{Ball}_{\tilde{d}}^{S_b}(w, r)|/q(\varepsilon)$. We also have $|\mathrm{Ball}_{\tilde{d}}^{S_b}(w, r)| \leqslant |\mathrm{Ball}_{\tilde{d}}^{S_b}(s*(w), r)|$, since $\tilde{d}_{uw} \leqslant c \cdot r$ and $w \in S_p$. Moreover, since for every $v \in E_1(u)$, u is clustered after v (this by definition implies v pre-clustered, thus that $v \in V_0$), we have that $|\mathrm{Ball}_{\tilde{d}}^{S_b}(s*(u), r)| \leqslant |\mathrm{Ball}_{\tilde{d}}^{S_b}(s*(v), r)|$ for every $v \in E_1(u)$.

We know that for any $v \in E_1(u)$, $|\mathrm{Ball}_{\tilde{d}}(s*(v), r)| \leqslant |B(u)|$. So putting it all together,

$$|E_1(u)| \leqslant 4 \cdot |\mathrm{Ball}_{\tilde{d}}^{S_b}(s*(v), r)|/q(\varepsilon) \leqslant 8/\varepsilon^2 \cdot |B(u)|.$$

Case 2. $\tilde{d}_{uw} = 1$. We have $\bar{d}_{uw} \leqslant t \leqslant 1/4$, so we may apply Lemma 22 and obtain that $|E_1(u)| \leqslant |N_u^+| \leqslant (512(C+1) \cdot \log n)/9\varepsilon^6 \cdot \mathsf{OPT}_\infty$. ☐

We are done if the second condition of Claim 32 holds, so we assume that the first condition holds. Then we use Claim 31 to see that

$$|E_1(u)| \leqslant \frac{2}{q(\varepsilon)} \cdot |B(u)| \leqslant \frac{2}{q(\varepsilon)} \cdot \Big(\sum_{b \in B(u) \cap N_u^+} 10 \cdot \bar{d}_{ub} + \sum_{b \in B(u) \cap N_u^-} \frac{10}{3} \cdot (1 - \bar{d}_{ub}) \Big)$$

where we have used that since $\tilde{d}_{ub} \neq 1$ for $b \in B(u)$, we have that $\tilde{d}_{ub} = \bar{d}_{ub}$. As a consequence of conditioning on B^c, we can bound $\bar{d}$ by OPT_∞ (see Lemma 14), so

$$|E_1(u)| \leqslant (320 \cdot (C+1) \cdot \log n)/3\varepsilon^4 \cdot \mathsf{OPT}_\infty. \tag{40}$$

Combining Equations (39) and (40).

$$\left|\{v \in N_u^+ \mid v \succ u\}\right| \leqslant (1472 \cdot (C+1) \cdot \log n)/9\varepsilon^6 \cdot \mathsf{OPT}_\infty.$$

☐

Next, we bound the cost of the positive disagreements incident to u that arise when u is pre-clustered, and either $v \in N_u^+$ is pre-clustered later or not pre-clustered at all, i.e., $u \succ v$.

Lemma 25. *Condition on* B^c. *Fix* $u \in V$. *Then* $|\{v \in N_u^+ \mid u \succ v\}| \leqslant O\left(1/\varepsilon^6 \cdot \log n\right) \cdot$ *OPT*$_\infty$.

We note that the proof of Lemma 25 is very similar to that of Lemma 24.

Proof of Lemma 25. We may assume $\mathsf{OPT}_\infty \geqslant 1$ per Proposition 3. Take $t = r/(2\delta)$. We partition as in Lemma 24,

$$\{v \in N_u^+ \mid u \succ v\} = \underbrace{\{v \in N_u^+ \mid u \succ v, \tilde{d}_{uv} \leqslant t\}}_{E_1(u)} + \underbrace{\{v \in N_u^+ \mid u \succ v, \tilde{d}_{uv} > t\}}_{E_2(u)}$$

If $|N_u^+| < C \cdot \log n/\varepsilon^2$, then we are done, so we assume for the remainder of the proof that $|N_u^+| \geqslant C \cdot \log n/\varepsilon^2$. Note in this case that, since B^c holds, $|N_u^+ \cap S_d| \geqslant C \cdot \log n/4 \geqslant 1$, so $u \in V_0$.

Bounding $|E_2(u)|$. By applying Lemma 23 with $s = t$, we have that

$$|E_2(u)| \leqslant {}^{(512 \cdot (C+1) \cdot \log n)}\!/\!_{9\varepsilon^6} \cdot \mathsf{OPT}_\infty. \tag{41}$$

Bounding $|E_1(u)|$. First note that for any v that is clustered after u, we have $\tilde{d}_{v\, s*(u)} > c \cdot r$. (Note we are using here that by Fact 1, it is not possible that $v \notin V_0$, since $\tilde{d}_{uv} \leqslant t < 1$.) Define

$$W(u) := \{v \in N_u^+ \mid \tilde{d}_{uv} \leqslant t, \tilde{d}_{v\, s*(u)} > c \cdot r\}.$$

So to bound $|E_1(u)|$, it suffices to bound $|W(u)|$.

Case 1. $|Ball_{\bar{d}}(u, t)| \leqslant C' \cdot \log_{1/(1-q(\varepsilon))} n$.

We know that $\bar{d} \leqslant \tilde{d}$, so $W(u) \subseteq \mathrm{Ball}_{\bar{d}}(u, t)$. Thus $|W(u)| \leqslant C' \cdot \log_{1/(1-q(\varepsilon))} n$, concluding the case.

Case 2. $|Ball_{\bar{d}}(u, t)| > C' \cdot \log_{1/(1-q(\varepsilon))} n$.

Define

$$B(u) := \mathrm{Ball}_{\tilde{d}}(s^*(u), r).$$

Claim 33. *For every $b \in B(u)$, $\tilde{d}_{ub} \geqslant c \cdot r/\delta^2 - t/\delta - r$ and $1 - \tilde{d}_{ub} \geqslant 1 - \delta \cdot (c \cdot r + r)$.*

Proof of Claim 33. Let $b \in B(u)$. Let w be an arbitrary vertex in $W(u)$ (if $W(u) = \varnothing$, then $|E_1(u)| = 0$ and we are done). By the approximate triangle inequality (Lemma 1), we have

$$\tilde{d}_{ub} \geqslant {}^1\!/\!_\delta \cdot \left({}^1\!/\!_\delta \cdot \tilde{d}_{w\, s*(u)} - \tilde{d}_{uw}\right) - \tilde{d}_{b\, s*(u)} \geqslant c \cdot r/\delta^2 - t/\delta - r$$

and also

$$\tilde{d}_{ub} \leqslant \delta \cdot (\tilde{d}_{u\, s*(u)} + \tilde{d}_{b\, s*(u)}) \leqslant \delta \cdot (c \cdot r + r)$$

which concludes the proof of the claim.

□

Claim 34. *At least one of the following holds:*

- $|W(u)| \leqslant O\left({}^1\!/\!_{\varepsilon^2}\right) \cdot |B(u)|$
- $|W(u)| \leqslant O\left({}^1\!/\!_{\varepsilon^6} \cdot \log n\right) \cdot \mathit{OPT}_\infty$.

Proof of Claim 34. Since B^c holds, there exists $z(u) \in \mathrm{Ball}_{\bar{d}}^{S_p}(u, t)$. We consider two cases:

Case 1. $\tilde{d}_{u\, z(u)} = \bar{d}_{u\, z(u)}$.

For every $v \in W(u)$, we have that $\tilde{d}_{uv} \leqslant t$ (by definition of $W(u)$), and also $\tilde{d}_{u\,z(u)} \leqslant t$ (by assumption of the case). So by the approximate triangle inequality (Lemma 1),

$$\tilde{d}_{v\,z(u)} \leqslant \delta \cdot (\tilde{d}_{vu} + \tilde{d}_{u\,z(u)}) \leqslant \delta \cdot 2t = r$$

for every $v \in W(u)$, which means that

$$|W(u)| \leqslant |\mathrm{Ball}_{\tilde{d}}(z(u), r)|.$$

If $|\mathrm{Ball}_{\tilde{d}}(z(u), r)| < C \cdot \log n / q(\varepsilon)$, then we are done. So assume $|\mathrm{Ball}_{\tilde{d}}(z(u), r)| \geqslant C \cdot \log n / q(\varepsilon)$. Then, since B^c holds, we have that $|\mathrm{Ball}_{\tilde{d}}(z(u), r)| \leqslant 4 \cdot |\mathrm{Ball}^{S_b}_{\tilde{d}}(z(u), r)|/q(\varepsilon)$. We have $|\mathrm{Ball}^{S_b}_{\tilde{d}}(z(u), r)| \leqslant |\mathrm{Ball}^{S_b}_{\tilde{d}}(s^*(u), r)|$, since $\tilde{d}_{u\,z(u)} \leqslant t \leqslant c \cdot r$ and $z(u) \in S_p$.

So putting it all together,

$$|W(u)| \leqslant 4 \cdot |\mathrm{Ball}^{S_b}_{\tilde{d}}(s^*(u), r)|/q(\varepsilon) \leqslant {}^{8}\!/\!_{\varepsilon^2} \cdot |B(u)|.$$

Case 2. $\tilde{d}_{u\,z(u)} = 1$.

We have $\bar{d}_{u\,z(u)} \leqslant t \leqslant 1/4$, so we may apply Lemma 22 and obtain that

$$|W(u)| \leqslant |N_u^+| \leqslant {}^{(512 \cdot (C+1) \cdot \log n)}\!/\!_{9\varepsilon^6} \cdot \mathsf{OPT}_\infty.$$

□

Now we see that the claims imply the sought bound on $|E_1(u)|$. We are done if the second bullet of Claim 34 holds, so we assume that the first bullet holds. We then have, using Claim 33, that

$$\begin{aligned}|E_1(u)| = |W(u)| &\leqslant \frac{2}{q(\varepsilon)} \cdot |B(u)| \\ &\leqslant \frac{2}{q(\varepsilon)} \cdot \left(\sum_{b \in B(u) \cap N_u^-} \frac{1}{1 - \delta \cdot (c \cdot r + r)} \cdot (1 - \bar{d}_{ub}) + \sum_{b \in B(u) \cap N_u^+} \frac{1}{c \cdot r/\delta^2 - t/\delta - r} \cdot \bar{d}_{ub} \right)\end{aligned}$$

where we have used that since $\tilde{d}_{ub} \neq 1$ for $b \in B(u)$, we have that $\tilde{d}_{ub} = \bar{d}_{ub}$. So by Lemma 14 and substituting in $\delta = 10/7$, $c = 2\delta^2 + \delta$, and $r = \frac{1}{2c\delta^2}$, we have that

$$|E_1(u)| \leqslant {}^{(512(C+1) \cdot \log n)}\!/\!_{3\varepsilon^4} \cdot \mathsf{OPT}_\infty. \tag{42}$$

Combining the bounds on $|E_1(u)|$ and $|E_2(u)|$ from Equations (42) and (41), we see that

$$\left|\{v \in N_u^+ \mid u \succ v\}\right| \leqslant {}^{(2048(C+1) \cdot \log n)}\!/\!_{9\varepsilon^6} \cdot \mathsf{OPT}_\infty.$$

□

C.1.2 Cost of negative edges

Next we bound the cost of the negative disagreements adjacent to u that are incurred during the Pre-clustering phase. Note that the only way a negative disagreement can arise during the Pre-clustering phase is if the same center in S_p clusters its endpoints.

Lemma 26. *Fix u pre-clustered. Then* $\left|\{v \in N_u^- \mid v \text{ clustered with } u\}\right| \leqslant O({}^{1}\!/\!_{\varepsilon^4} \cdot \log n) \cdot \mathit{OPT}_\infty$.

Proof of Lemma 26. We may assume $\mathsf{OPT}_\infty \geqslant 1$ per Proposition 3. Let $v \in N_u^-$ be such that u, v are clustered together during the Pre-clustering phase. Then we may set $s^* := s^*(u) = s^*(v)$. Thus we have $\tilde{d}_{us^*} \leqslant c \cdot r$ and $\tilde{d}_{vs^*} \leqslant c \cdot r$. So by the approximate triangle inequality (Lemma 1), $\tilde{d}_{uv} \leqslant 2\delta c r$. Since $\tilde{d}_{uv} < 1$, $\bar{d}_{uv} = \tilde{d}_{uv}$ and $u \notin R_1$. We consider two cases.

Case 1. $|N_u^+| \geqslant C \cdot \log n / \varepsilon^2$.

Then we have by Lemma 14 that

$$\left|\left\{v \in N_u^- \mid v \text{ clustered with } u\right\}\right| \leqslant \sum_{v \in N_u^-} \frac{1}{1-2\delta cr} \cdot (1-\bar{d}_{uv}) \leqslant \frac{8 \cdot (C+1)}{3\varepsilon^2 \cdot (1-2\delta cr)} \cdot \log n \cdot \mathsf{OPT}_\infty .$$

Case 2. $|N_u^+| < C \cdot \log n / \varepsilon^2$.

Note $2\delta \cdot c \cdot r < 7/10$. Then it suffices to bound

$$|N_u^- \cap \{v \in V \mid \bar{d}_{uv} \leqslant 7/10\}|.$$

If $|N_u^- \cap \{v \in V : \bar{d}_{uv} \leqslant 7/10\}| < C \cdot \log n / q(\varepsilon)$, then we are done. So we may assume

$$|N_u^- \cap \{v \in V : \bar{d}_{uv} \leqslant 7/10\}| \geqslant C \cdot \log n / q(\varepsilon).$$

Then, since B^c holds,

$$\begin{aligned} |N_u^- \cap \{v \in V : \bar{d}_{uv} \leqslant 7/10\}| &\leqslant {}^{8}\!/\!{}_{\varepsilon^2} \cdot |N_u^- \cap S_r \cap \{v \in V : \bar{d}_{uv} \leqslant 7/10\}| \\ &\leqslant {}^{80}\!/\!{}_{3\varepsilon^2} \cdot |N_u^+ \cap S_d| \\ &\leqslant {}^{80}\!/\!{}_{3\varepsilon^2} \cdot |N_u^+| \leqslant {}^{80}\!/\!{}_{3\varepsilon^4} \cdot C \log n \cdot \mathsf{OPT}_\infty \end{aligned}$$

where in the second inequality we have used that $u \notin R_1$, and in the last we use Proposition 3. This concludes the case and the proof.

□

C.2 Cost of Pivot phase for ℓ_∞

Let $G' = (V', E')$ be the subgraph induced by the unclustered vertices. Let V_0' be defined as for the tight ℓ_1-norm bound, that is,

$$V_0' = V_0 \cap \{v \in V : \tilde{d}_{vu_i} > c \cdot r \text{ for all } u_i \in S_p\},$$

and recall that $V' = \overline{V_0} \cup V_0'$. Also recall that $\mathcal{C}_{\text{ALG}}$ is the clustering output by Algorithm 1.

Lemma 27. *Condition on the good event B^c. Fix $u \in V'$. The number of disagreements that $\mathcal{C}_{ALG}$ has in $G[V']$ incident to u is $O(\frac{1}{\varepsilon^6} \cdot \log n) \cdot OPT_\infty$.*

Proof of Lemma 27. By Proposition 3, we may assume $\mathsf{OPT}_\infty \geqslant 1$. We case on whether $u \in \overline{V_0}$ or $u \in V_0'$.

Case 1. $u \in V_0'$.

Recall that on the subgraph $G[V_0']$, we run Modified Pivot with $E_c = E^+ \cap \{uv \in E \mid \tilde{d}_{uv} < c \cdot r\}$. Let u^* be u's pivot.

Claim 35. *The number of disagreements $\mathcal{C}_{ALG}$ has in $G[V']$ incident to $u \in V_0'$ is at most $2 \cdot |N_u^+| + |N_{u^*}^+|$.*

Proof of Claim 35. To see this, observe that if u is a pivot (i.e., $u^* = u$), then u has no negative *neighbors* in its cluster, so there are at most $|N_u^+|$ disagreements in $G[V_0']$ incident to u in the worst case. On the other hand, if u is not a pivot (so $u^* \neq u$), then there may be negative disagreements in $G[V_0']$ incident to u. These are of the form $uv \in E^-$ where v's pivot is also u^*. But then $v \in N_{u^*}^+$. So the number of negative disagreements in $G[V_0']$ incident to u is at most $|N_{u^*}^+|$.

Finally, the remaining disagreements uv in $G[V']$ to account for are those with $v \in N_u^+ \cap \overline{V_0}$, thus giving an additive factor of $|N_u^+|$. □

Now we upper bound $2 \cdot |N_u^+| + |N_{u*}^+|$ for $u \in V_0'$ (thus also $u^* \in V_0'$). If $|N_u^+|, |N_{u*}^+| < C \cdot \log n/\varepsilon^2$, then we are done. So suppose $|N_u^+| \geqslant C \cdot \log n/\varepsilon^2$ (the same argument will apply for u^*). Then

$$|N_u^+| \leqslant \frac{4}{\varepsilon^2} \cdot |N_u^+ \cap S_p| \leqslant \frac{4}{\varepsilon^2} \cdot |N_u^+ \cap \{v \in V \mid \tilde{d}_{uv} > c \cdot r\}| \leqslant {}^{(2048(C+1) \cdot \log n)}/_{9 \cdot \varepsilon^6} \cdot \mathsf{OPT}_\infty,$$

where in the first inequality, we have used that B^c holds, in the second that by hypothesis $u \in V_0'$, and in the third that we may take $s = c \cdot r = 0.245$ in Lemma 23.

We conclude that the number of disagreements in $G[V']$ incident to u is bounded in the case that $u \in V_0'$ by ${}^{2048(C+1)}/_{3 \cdot \varepsilon^6} \cdot \log n \cdot \mathsf{OPT}_\infty$.

Case 2. $u \in \overline{V_0}$.

Recall that on the subgraph $G[\overline{V_0}]$, we run standard Pivot with $E_c = E^+$. Let u^* be u's pivot.

Claim 36. *The number of disagreements $\mathcal{C}_{ALG}$ has in $G[V']$ incident to $u \in \overline{V_0}$ is at most $2 \cdot |N_u^+| + |N_{u*}^+|$.*

The proof of the claim is the same as that of Claim 35 since there, we only used that $E_c \subseteq E^+$.

Now we upper bound $2 \cdot |N_u^+| + |N_{u*}^+|$ for $u \in \overline{V_0}$ (thus also $u^* \in \overline{V_0}$). Since B^c holds and $|N_u^+ \cap S_d| = 0$ (that is, $u \notin V_0$), it must be the case that $|N_u^+| \leqslant C \cdot \log n/\varepsilon^2$, and likewise for u^*. This concludes the case and the proof.

□

C.3 Proof of Item 2 of Theorem 1

We note the disagreements are all accounted for, see Figure 6. In particular, disagreements where one endpoint was pre-clustered and the other was not are accounted for by Lemmas 24 and 25. Disagreements for the entire Pivot phase are accounted for in Lemma 27.

In total, we combine the bounds from Lemmas 24, 25, 26, and 27 to see that

$$||y_{\mathcal{C}_{ALG}}||_\infty \leqslant O\left({}^{1}/_{\varepsilon^6} \cdot \log n\right) \cdot \mathsf{OPT}_\infty.$$

D Lower Bounds

In this section, we first show that any strictly online (deterministic or randomized) algorithm (i.e., an online algorithm not given any extra information, such as a sample of the nodes) has a worst-case expected competitive ratio of $\Omega(n)$ for the ℓ_∞-norm objective. We do this by showing that the lower bound instance in, e.g., [24, 22] giving a worst-case expected competitive ratio of $\Omega(n)$ for the ℓ_1-norm objective also gives a lower bound instance for the ℓ_∞-objective.

Then we show a lower bound of $\Omega(n^{1/4})$ in the random-order model for the ℓ_∞-norm objective. While the random-order model is more powerful than the strictly online setting and often allows for better competitive ratios or approximations), we show that strong lower bounds still exist in this model for ℓ_p-norm correlation clustering. This motivates our decision to study ℓ_p-norm correlation clustering in the AOS setting.

Lastly, we prove a lower bound of $\Omega(\log n/\varepsilon)$ on the expected competitive ratio of any (deterministic or randomized) algorithm in the online-with-a-sample model for the ℓ_∞-norm objective, thus proving Theorem 2. Note that the randomness here is taken over both the randomness internal to the algorithm, and the randomness of the sample given to the algorithm. The idea is to use as gadgets the instances from the lower bound in the strictly online setting.

To prove our results, we will use Yao's Min-Max Principle.

Theorem 7 (Yao's Min-Max Principle)**.** *Fix a problem with a set $\mathcal{X}$ of possible inputs and a set $\mathcal{A}$ of deterministic algorithms solving the problem. Denote the cost of algorithm $A \in \mathcal{A}$ on input $X \in \mathcal{X}$ as $c(A, X) \geqslant 0$. Further, let $\mathcal{D}$ be a distribution over $\mathcal{A}$, and let $A^* \in \mathcal{A}$ be an algorithm drawn randomly from $\mathcal{D}$ (that is, A^* is a randomized algorithm). Similarly, let $\mathcal{P}$ be a distribution over inputs $\mathcal{X}$ and let $X^* \in \mathcal{X}$ be an input drawn randomly from $\mathcal{P}$. Then*

$$\max_{X \in \mathcal{X}} \mathbb{E}_{A^* \sim \mathcal{D}}[c(A^*, X)] \geqslant \min_{A \in \mathcal{A}} \mathbb{E}_{X^* \sim \mathcal{P}}[c(A, X^*)].$$

Yao's Min-Max Principle states that the expected cost of any randomized algorithm (thus also the best one) for a worst-case input is at least the expected cost of the best deterministic algorithm over any distribution of inputs. In other words, we can lower bound the worst-case expected cost of a randomized algorithm on a (deterministic) set of inputs (i.e., the left hand side of the inequality) by lower bounding the expected cost of the best deterministic algorithm on any distribution of inputs of our choosing, but ideally one that gives us a large lower bound.

D.1 Lower bound for strictly online model

We begin by recalling the $\Omega(n)$ lower bound for the ℓ_1-norm objective in the online setting, which was first given by Mathieu, Sankur, and Schudy [24], and we note it holds even for randomized algorithms (see Theorem 3.4 in [24]). We show this bound actually holds for the ℓ_∞-norm objective too. A key gadget in the lower bound is a graph consisting of two positive cliques of size $n/2$, call them V_1 and V_2, with one positive edge $(v_1, v_2) \in V_1 \times V_2$, and all other edges between the two cliques are negative (see Section 1). Arguing as in Section 1, one can conclude any *deterministic* algorithm has cost at least $\Omega(n)$, hence competitive ratio $\Omega(n)$, with respect to the ℓ_∞-norm objective.

Let us extend this lower bound to *randomized* algorithms using Yao's Principle. We define $\mathsf{CR}_\infty(A, X)$ to be the *competitive ratio* of algorithm A on instance X for the ℓ_∞-norm objective; we will take the cost c in Theorem 7 to be CR_∞. Note that, because we are in the online setting, instance X consists of both an underlying input graph *and* an order of arrival for vertices. Let G_1 denote the instance described above, with v_1 and v_2 arriving first and the remaining vertices arriving in arbitrary order. Let G_2 denote a positive clique on n nodes (with order irrelevant due to the symmetry of this instance). Consider the distribution $\mathcal{P}$ over inputs G_1 and G_2, where G_1 and G_2 each have probability 1/2 of being sampled.

Let $\mathcal{A}$ be the family of all deterministic algorithms for correlation clustering in the strictly online model. It is sufficient to partition these based on whether or not they place v_1, v_2 together. If $A \in \mathcal{A}$ puts v_1 and v_2 in the same cluster, and the remainder of network G_1 arrives online, the competitive ratio $\mathsf{CR}_\infty(A, G_1)$ is at least $n/2$, since the ℓ_∞-norm (ℓ_1-norm) cost of A on G_1 is at least $n/2$, regardless of the remaining decisions A makes, whereas the optimal (offline) cost for the ℓ_∞-norm objective on G_1 is 1; clustering v_1 and v_2 together is immediately a poor decision for G_1. On the other hand, if $A' \in \mathcal{A}$ puts v_1 and v_2 in different clusters, and instead the remainder of network G_2 arrives online, the competitive ratio $\mathsf{CR}_\infty(A', G_2)$ of A' on G_2 is ∞, since the optimal solution for the ℓ_∞-norm objective is 0.

Now we apply Yao's Principle to obtain the sought lower bound for *randomized* algorithms, where $\mathcal{D}$ is a distribution over $\mathcal{A}$ and $\mathcal{P}$ is a distribution over $\mathcal{X}$:

$$\max_{X \in \mathcal{X}} \mathbb{E}_{A^* \sim \mathcal{D}}[\mathsf{CR}_\infty(A^*, X)] \geqslant \min_{A \in \mathcal{A}} \mathbb{E}_{X^* \sim \mathcal{P}}[\mathsf{CR}_\infty(A, X^*)] \overset{\star}{\geqslant} \min\left\{\frac{1}{2} \cdot \frac{n}{2}, \infty\right\} = n/4 \tag{43}$$

where in ($\star$) we have partitioned $A \in \mathcal{A}$ based on whether v_1, v_2 are in the same or different clusters as in the preceding paragraph, and used the definition of our chosen distribution $\mathcal{P}$.

D.2 Lower bound for random-order model

In the random-order model, instead of assuming that the input arrives in the worst possible order (as in the strictly online model), we assume the entire input is fixed in advance, and its elements arrive in a uniformly random order.

We first construct a distribution $\mathcal{P}$ over inputs $\mathcal{X}$. Consider $n^{1-\delta}$ gadgets, $H_1, \ldots, H_{n^{1-\delta}}$. For each gadget independently, H_i is a copy of G_1 on n^δ nodes with probability $1/2$, and is a copy of G_2 on n^δ nodes with probability $1/2$. For the gadgets H_i that are a copy of G_1, we denote the endpoints of the lone positive edge between the two cliques in H_i as $v_{i,1}$ and $v_{i,2}$. Within a gadget, nodes arrive uniformly at random. This fully defines the distribution $\mathcal{P}$.

Lemma 28. *Given $\delta \leqslant 1/3$, every* deterministic *algorithm in the random-order model has expected competitive ratio at least $\Omega(n^\delta)$ for the ℓ_∞-norm objective.*

Proof of Lemma 28. The optimal offline algorithm has cost at least 1 for the ℓ_∞-norm objective for any instance in the support of $\mathcal{P}$.

Let $\mathsf{CR}_\infty(A, X)$ be the ℓ_∞-norm cost of algorithm A on instance X. We lower bound

$$\min_{A \in \mathcal{A}} \mathbb{E}_{X \in \mathcal{P}} \Big[\mathsf{CR}_\infty(A, X) \Big],$$

that is, the expected cost of the best deterministic algorithm.

Define E^i to be the event that H_i is a copy of G_1, and $v_{i,1}$ and $v_{i,2}$ are the first two nodes to arrive in H_i. Let $\mathcal{A}$ be the family of deterministic algorithms for the problem. We will show in the following claim that with good probability, there is some gadget H_i where event E^i occurs. Then we will argue one can reduce to the lower bound for the strictly online model on gadget H_i.

Claim 37. *For $\delta \leqslant 1/3$ we see that the probability there is some gadget where event E^i occurs is high, with*

$$\mathbb{P}[\cup_{i \in [n^{1-\delta}]} E^i] \geqslant 1 - 1/n.$$

Proof of Claim 37. Since the choice of G_1 and G_2 are each with probability 1/2, and the order nodes arrive in within a gadget is independent of whether a gadget is a copy of G_1 or G_2, we see that

$$\begin{aligned} \mathbb{P}[E^i] &= \frac{1}{2} \cdot \mathbb{P}[v_{i,1}, v_{i,2} \text{ arrive first in } H_i] \\ &\geqslant \frac{1}{2} \cdot \frac{1}{n^{2\delta}}. \end{aligned}$$

Let $\overline{E^i}$ denote the complement of the event E^i, so $\mathbb{P}[\overline{E^i}] \leqslant 1 - \frac{1}{2 \cdot n^{2\delta}}$. Also note that the events $\{E^i\}_{i \in [n^{1-\delta}]}$ are all independent. Then,

$$\begin{aligned} \mathbb{P}[\cup_{i \in [n^{1-\delta}]} E^i] = 1 - \prod_{i \in [n^{1-\delta}]} \mathbb{P}[\overline{E^i}] &= 1 - \big(\mathbb{P}[\overline{E^i}]\big)^{n^{1-\delta}} \\ &\geqslant 1 - \big(1 - \frac{1}{2 \cdot n^{2\delta}}\big)^{n^{1-\delta}} \\ &\geqslant 1 - \frac{1}{\sqrt{e}}, \end{aligned}$$

where in the last line we use that $\delta \leqslant 1/3$. □

In what follows, let H_{i*} be a gadget such that event E^{i^*} occurs. We use the claim to see the following string of inequalities:

$$\begin{aligned} \min_{A \in \mathcal{A}} \mathbb{E}_{X* \sim \mathcal{P}} \Big[\mathsf{CR}_\infty(A, X^*) \Big] &= \min_{A \in \mathcal{A}} \Big(\mathbb{P}[\cup_{i \in [n^{1-\delta}]} E^i] \cdot \mathbb{E}_{X* \sim \mathcal{P}} \Big[\mathsf{CR}_\infty(A, X^*) \mid \cup_{i \in [n^{1-\delta}]} E^i \Big] \\ &\quad + \mathbb{P}[\cap_{i \in [n^{1-\delta}]} \overline{E^i}] \cdot \mathbb{E}_{X* \sim \mathcal{P}} \Big[\mathsf{CR}_\infty(A, X^*) \mid \cap_{i \in [n^{1-\delta}]} \overline{E^i} \Big] \Big) \\ &\geqslant \Big(1 - \frac{1}{\sqrt{e}}\Big) \cdot \min_{A \in \mathcal{A}} \mathbb{E}_{X* \sim \mathcal{P}} \Big[\mathsf{CR}_\infty(A, X^*) \mid \cup_{i \in [n^{1-\delta}]} E^i \Big] \\ &\geqslant \Big(1 - \frac{1}{\sqrt{e}}\Big) \cdot \min_{A \in \mathcal{A}} \mathbb{E}_{X* \sim \mathcal{P}} \Big[\mathsf{CR}_\infty(A, H_{i*}) \mid E^{i^*} \Big] \\ &\geqslant \Big(1 - \frac{1}{\sqrt{e}}\Big) \cdot \frac{n^\delta}{4}. \end{aligned} \tag{44}$$

where in the last inequality, we argue as in subsection D.1, and in particular recover the lower bound in Equation (43), since H_{i*} arrives fully online. We note the argument for the strictly online lower bound holds because the ordering only needed to be worst-case to ensure v_1 and v_2 (or here, $v_{i,1}$ and $v_{i,2}$) arrive before the rest of the vertices, then one can consider a random order on the rest of the vertices and the result still holds. □

Theorem 8. *For any $\delta \leqslant 1/3$, any randomized algorithm in the random-order model has a worst-case expected cost of at least $\Omega\left(n^\delta\right) \cdot \mathsf{OPT}_\infty$ for the ℓ_∞-norm.*

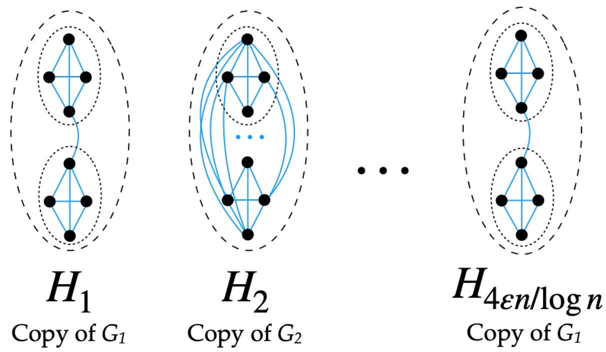

Figure 7: An instance from a distribution $\mathcal{P}$ that gives the lower bound in Lemma 29. The larger ovals with the dashed border represent the $4\varepsilon n/\log n$ gadgets. Each gadget is either a positive clique, or two equally sized positive cliques with exactly one positive edge between them. Positive edges are drawn while negative edges are not.

Proof of Theorem 8. Let $\mathcal{D}$ be any distribution over $\mathcal{A}$, the family of deterministic algorithms for the problem. Let $A^* \sim \mathcal{D}$. Then the worst-case expected cost of A^* over inputs $\mathcal{X}$ is

$$\max_{X\in\mathcal{X}} \mathbb{E}_{A^*}[\mathsf{CR}_\infty(A^*, X)] = \max_{X\in\mathcal{X}} \mathbb{E}_{A^*}[\mathsf{CR}_\infty(A^*, X)] \geqslant \min_{A\in\mathcal{A}} \mathbb{E}_{X^*\sim\mathcal{P}}[\mathsf{CR}_\infty(A, X^*)] = \Omega(n^\delta),$$

where the inequality is by Yao's Min-Max Principle (Theorem 7) and the final bound is by the lower bound on the same expression (44) in the proof of Lemma 28. □

D.3 Lower bound for online-with-a-sample model

We now prove Theorem 2, specifically, the $\Omega(\log n/\varepsilon)$ lower bound on the competitive ratio for the ℓ_∞-norm objective in the online-with-a-sample model. The idea is to split the n vertices into several gadgets, where each gadget is a (smaller) copy of G_1 or G_2, and there are no positive edges between gadgets. The key to "reducing" to the lower bound in the strictly online setting (subsection D.1) is to show that, with constant probability, there is at least one gadget that the sample S does not hit; thus, this gadget is revealed in a strictly online fashion, so the algorithm incurs high cost. Note that for the ℓ_∞-norm, the optimal objective value does *not* increase with the number of gadgets. So in setting the gadget size to maximize our lower bound, the only tradeoff is that between the probability of not hitting at least one gadget (high if there are many gadgets) and the cost incurred on a gadget that is not hit (high if there are few gadgets).

We first construct a distribution $\mathcal{P}$ over inputs $\mathcal{X}$. See Figure 7 for an illustration of the construction. Consider ${}^{4\varepsilon n}/_{\log n}$ gadgets H_i for $i \in \left[{}^{4\varepsilon n}/_{\log n}\right]$. For each gadget independently, H_i is a copy of G_1 on ${}^{\log n}/_{4\varepsilon}$ vertices with probability 1/2, and is a copy of G_2 (again on ${}^{\log n}/_{4\varepsilon}$ vertices) with probability 1/2. The order in which gadgets arrive does not matter, but we assume the worst-case order for each individual gadget as in subsection D.1. Further, let $\mathcal{A}$ be the family of deterministic algorithms for the problem. Note that in order to consider deterministic algorithms, one must consider the random sample S to be part of the input. Algorithm $A \in \mathcal{A}$ receives upfront a realization of the random sample S once this has been fixed, and all of its decisions are deterministic henceforth.

Lemma 29. *Given* ${}^{1}/_{n^{1/4}} \leqslant \varepsilon \leqslant {}^{3}/_{4}$, *every* deterministic *algorithm in the online-with-a-sample model has worst-case expected competitive ratio at least* $\Omega(\log n/\varepsilon)$ *for the* ℓ_∞*-norm objective.*

Proof of Lemma 29. First, it is clear that the optimal offline algorithm has cost at most 1 for the ℓ_∞-norm objective for any instance in the support of $\mathcal{P}$.

Let $\mathsf{CR}_\infty(A, S, X)$ be the ℓ_∞-norm cost of algorithm A given sample S on instance X. We lower bound

$$\min_{A\in\mathcal{A}} \max_{X\in\mathcal{X}} \mathbb{E}_S\Big[\mathsf{CR}_\infty(A, S, X)\Big],$$

that is, the worst-case expected cost of the best deterministic algorithm, where the expectation is only over the random sample S.

First we will lower bound the probability that for any fixed instance X in the support of $\mathcal{P}$, at least one gadget of X is not sampled by S (meaning *none* of its vertices are in S).

Claim 38. *For any instance X in the support of $\mathcal{P}$, and for $\frac{1}{n^{1/4}} \leqslant \varepsilon \leqslant \frac{3}{4}$, there is some gadget that is not sampled with probability at least 0.99.*

Proof of Claim 38. Fix an instance X in the support of $\mathcal{P}$. Recall X contains ${}^{4\varepsilon n}/_{\log n}$ gadgets H_i of size ${}^{\log n}/_{4\varepsilon}$ each. The probability that not all gadgets are sampled by S (where we use fact that nodes in X are in S independently and with probability ε) is

$$\begin{aligned}\mathbb{P}[\text{not all gadgets in } X \text{ sampled}] &= 1 - \prod_{i=1}^{4\varepsilon n/\log n} \mathbb{P}[\text{gadget } H_i \text{ sampled}] \\ &= 1 - \prod_{i=1}^{4\varepsilon n/\log n} \left(1-(1-\varepsilon)^{|H_i|}\right) = 1 - \left(1-(1-\varepsilon)^{\log n/4\varepsilon}\right)^{4\varepsilon n/\log n}.\end{aligned}$$

The we continue lower bounding the above using the fact that $1-x \geqslant e^{-2x}$ for $0 \leqslant x \leqslant 3/4$:

$$\begin{aligned}\mathbb{P}[\text{not all gadgets in } X \text{ sampled}] &= 1 - \left(1-(1-\varepsilon)^{\log n/4\varepsilon}\right)^{4\varepsilon n/\log n} \geqslant 1 - \left(1-e^{-\log n/2}\right)^{4\varepsilon n/\log n} \\ &= 1 - \left(1-1/\sqrt{n}\right)^{4\varepsilon n/\log n} \geqslant 1 - e^{-4\varepsilon\sqrt{n}/\log n} \\ &\geqslant 1 - e^{-4n^{1/4}/\log n} \geqslant 0.99,\end{aligned}$$

where the penultimate inequality holds using $\varepsilon \geqslant 1/n^{1/4}$. ☐

Now we are ready for the lower bound. For shorthand, let $E_{X*,S}$ be the event the sample S hits every gadget in X^*, and let $\overline{E}_{X*,S}$ be its complement. Let H^* be a gadget that is *not* sampled by S in the case $\overline{E}_{X*,S}$ holds (i.e., $E_{X*,S}$ does not hold). For any $A \in \mathcal{A}$, let $\mathsf{cost}(A, H^*)$ denote the ℓ_∞-norm objective cost of A restricted to the gadget H^*. Note that by definition, H^* arrives *fully online*. Let (X^*, S) be drawn from the joint distribution of X^* and S. Note that X^* and S are independent. Then

$$\begin{aligned}\min_{A\in\mathcal{A}}\max_{X\in\mathcal{X}}\mathbb{E}_S\Big[\mathsf{CR}_\infty(A,S,X)\Big] &\geqslant \min_{A\in\mathcal{A}}\mathbb{E}_{X*\sim\mathcal{P}}\mathbb{E}_S\Big[\mathsf{CR}_\infty(A,S,X^*)\Big] &(45)\\ &= \min_{A\in\mathcal{A}}\mathbb{E}_{(X*,S)}\Big[\mathsf{CR}_\infty(A,S,X^*)\Big] &(46)\\ &\geqslant \min_{A\in\mathcal{A}}\mathbb{E}_{(X*,S)}\Big[\mathsf{CR}_\infty(A,S,X^*) \mid \overline{E}_{X*,S}\Big]\cdot\mathbb{P}[\overline{E}_{X*,S}] \\ &\geqslant \mathbb{P}[\overline{E}_{X*,S}]\cdot\min_{A\in\mathcal{A}}\mathbb{E}_{(X*,S)}\Big[\mathsf{cost}(A,H^*) \mid \overline{E}_{X*,S}\Big]\cdot &(47)\\ &= \mathbb{P}[\overline{E}_{X*,S}]\cdot\min_{A\in\mathcal{A}}\Big(\frac{1}{2}\cdot\mathsf{cost}(A,G_1)+\frac{1}{2}\cdot\mathsf{cost}(A,G_2)\Big) &(48)\\ &\geqslant \mathbb{P}[\overline{E}_{X*,S}]\cdot\frac{\log n}{16\varepsilon} &(49)\\ &\geqslant 0.99\cdot\frac{\log n}{16\varepsilon}\end{aligned}$$

Line (46) holds since the sampling procedure is independent of the instance, so we can swap the expectations by Fubini-Tonelli. In line (47) we have used that the optimal solution for any instance $X^* \sim \mathcal{P}$ has cost at most 1. Finally, in lines (48) and (49), we argue as in subsection D.1, and in particular recover the lower bound in (43), since H^* arrives fully online. In the last line, we have applied Claim 38. This concludes the proof of the lemma. ☐

Proof of Theorem 2. Let $\mathcal{D}$ be any distribution over $\mathcal{A}$, the family of deterministic algorithms for the problem. Let $A^* \sim \mathcal{D}$. Then the worst-case expected cost of A^* over inputs $\mathcal{X}$ is

$$\begin{aligned}\max_{X \in \mathcal{X}} \mathbb{E}_{(A^*,S)}[\mathsf{CR}_\infty(A^*, X, S)] = \max_{X \in \mathcal{X}} \mathbb{E}_{A^*}\mathbb{E}_S[\mathsf{CR}_\infty(A^*, X, S)] &\geqslant \min_{A \in \mathcal{A}} \mathbb{E}_{X^* \sim \mathcal{P}}\mathbb{E}_S[\mathsf{CR}_\infty(A, X^*, S)] \\ &= \Omega(\log n/\varepsilon),\end{aligned}$$

where the inequality is by Yao's Min-Max Principle (Theorem 7) and the final bound is by the lower bound on the same expression (45) given in the proof of Lemma 29.

The lower bound of $\Omega(1/\varepsilon)$ for the ℓ_1-norm in the statement of the theorem is already proven in Theorem 1 of [22].

□