Handling LP-Rounding for Hierarchical Clustering and Fitting Distances by Ultrametrics

Hyung-Chan An
Department of Computer Science and Engineering
Yonsei University
Seoul, South Korea
hyung-chan.an@yonsei.ac.kr

Mong-Jen Kao
Department of Computer Science
National Yang-Ming Chiao-Tung University
Hsinchu, Taiwan
mjkao@nycu.edu.tw

Changyeol Lee
Department of Computer Science and Engineering
Yonsei University
Seoul, South Korea
777john@yonsei.ac.kr

Mu-Ting Lee
Department of Computer Science
National Yang-Ming Chiao-Tung University
Hsinchu, Taiwan
mutinglee.cs11@nycu.edu.tw

***Abstract*—We consider the classic correlation clustering problem in the hierarchical setting. Given a complete graph $G = (V, E)$ and ℓ layers of input information, where the input of each layer consists of a non-negative weight and a labeling of the edges with either $+$ or $-$, this problem seeks to compute for each layer a partition of V such that the partition for any non-top layer subdivides the partition in the upper-layer and the weighted number of disagreements over the layers is minimized, where the disagreement of a layer is the number of $+$ edges across parts plus the number of $-$ edges within parts.**

Hierarchical correlation clustering is a natural formulation of the classic problem of fitting distances by ultrametrics, which is further known as numerical taxonomy [1]–[3] in the literature. While single-layer correlation clustering received wide attention since it was introduced in [4] and major progress evolved in the past three years [5]–[8], few is known for this problem in the hierarchical setting [9], [10]. The lack of understanding and adequate tools is reflected in the large approximation ratio known for this problem, which originates from 2021.

In this work we make both conceptual and technical contributions towards the hierarchical clustering problem. We present a simple paradigm that greatly facilitates LP-rounding in hierarchical clustering, illustrated with a delicate algorithm providing a significantly improved approximation guarantee of 25.7846 for the hierarchical correlation clustering problem.

Our techniques reveal surprising new properties and advances the current understanding for the formulation presented and subsequently used in [9]–[12] for hierarchical clustering over the past two decades. This provides a unifying interpretation on the core-technical problem in hierarchical clustering as the problem of finding cuts with prescribed properties regarding the average distance of certain cut pairs.

We further illustrate this perspective by showing that a direct application of the paradigm and techniques presented in this work gives a simple alternative to the state-of-the-art result presented in [12] for the ultrametric violation distance problem.

***Index Terms*—hierarchical correlation clustering, ultrametric embedding, correlation clustering, linear programming rounding, approximation algorithms**

I. Introduction

Clustering is among the central problems in unsupervised machine learning and data mining. For a given data set and information regarding pairwise similarity of the elements, the general objective is to come up with a partition of the elements into groups such that similar elements are clustered into the same group and dissimilar elements belong to different groups.

Correlation Clustering, among various formulations introduced towards the aforementioned objective, has been one of the most successful model since its introduction by Bansal, Blum, and Chawla in [4]. Given

This work was supported in part by the National Science and Technology Council (NSTC), Taiwan, under grants 112-2628-E-A49-017-MY3, 113-2628-E-A49-017-MY3, and 113-2634-F-A49-001-MBK. This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. RS-2021-II212068, Artificial Intelligence Innovation Hub). This work was partly supported by an IITP grant funded by the Korean Government (MSIT) (No. RS-2020-II201361, Artificial Intelligence Graduate School Program (Yonsei University)). This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2025-00563707).

2575-8454/25/$31.00 ©2025 IEEE
DOI 10.1109/FOCS63196.2025.00060

a complete graph $G = (V, E)$ and a labeling of the edges with either $+$ or $-$, the goal is to partition the vertices so as to minimize the number of disagreements between the partition computed and the input labels, namely, the number of $+$ edges clustered into different parts plus the number of $-$ edges clustered into the same part. Due to the simplicity and modularity of this formulation, correlation clustering has found vast applications in practice, e.g., finding clustering ensembles [13], duplicate detection [14], community mining [15], disambiguation tasks [16], automated labeling [17], [18], and many more.

Various algorithms with an $O(1)$-approximation guarantee exist in the literature for the correlation clustering problem, including classic results in the early 2000s [4], [19], [20], the elegant 2.06-approximation based on LP-rounding [21], and recent breakthroughs that evolved in the past three years using the Sherali-Adams hierarchy [5], [6] and a strong formulation [7], [22] known as cluster LP. Currently, the best approximation ratio is $1.437+\epsilon$, and $(24/23-\epsilon)$-approximation is NP-hard [7] for any $\epsilon > 0$.

Motivated by the large number of applications in practice, efficient approximation algorithms based on combinatorial approaches have been introduced in the literature, including linear time algorithm [20], dynamic algorithms [23], results for distributed models [24]–[26], streaming models [26]–[30], and very recent sublinear time algorithms [8], [22], [30].

Correlation Clustering in the Hierarchical Setting: In the hierarchical setting, we are given a complete graph $G = (V, E)$ and ℓ layers of input information regarding pairwise similarity of the elements, where the input information for each layer consists of a non-negative weight and a labeling of the edges with either $+$ or $-$. The goal is to produce for each layer a partition of the elements in V such that (i) the partition for any non-top layer subdivides the partition in the upper layer and (ii) the weighted disagreements over all layers is minimized.

Hierarchical correlation clustering is a natural formulation for the classic problem of fitting given distance information by ultrametrics, which is also known as numerical taxonomy in the literature [1]–[3], [9], [10], [31]. While single-layer correlation clustering was extensively studied with various types of techniques in the past two decades, the multi-layer setting remains much less understood to date. The main challenge of this problem has been in the need to produce a sequence of consistent partitioning of the elements subject to the unrelated, possibly conflicting, similarity information given for the layers.

Ailon and Charikar [9] presented both combinatorial-based and LP-rounding algorithms to obtain a $\min\{\ell + 2, O(\log n \log\log n)\}$-approximation, utilizing the pivot-based algorithm [20] and a region growing argument. In a breakthrough result for this problem, Cohen-Addad, Das, Kipouridis, Parotsidis, and Thorup [10] presented an unconventional approach to obtain the first constant factor (> 1000) approximation using the LP presented in [9] and state-of-the-art algorithms for single-layer correlation clustering. This has remained the best approximation ratio known for this problem since 2021.

Fitting Distance by Ultrametrics (Numerical Taxonomy): In the numerical taxonomy problem, we are given measured pairwise distances $\mathcal{D}\colon \binom{V}{2} \mapsto \mathbb{R}_{>0}$ for a set of elements and the goal is to produce a tree metric or an ultrametric T that spans V and minimizes the L_p-norm

$$\|T - \mathcal{D}\|_p \;:=\; \left(\sum_{\{i,j\}\in\binom{V}{2}} \mid d_T(i,j) - D(i,j) \mid^p \right)^{1/p},$$

where p is a prescribed constant with $1 \le p \le \infty$ and d_T is the distance function for T.

Since Cavalli-Sforza and Edwards introduced the numerical taxonomy problem, it has collected an extensive literature [1], [32]–[34]. While this problem was initially introduced in the L_2-norm, Farris [32] suggested using the L_1-norm in 1972. Further, it is known that for any $1 \le p \le \infty$, an algorithm that computes an ultrametric can readily be applied for computing a tree metric losing a factor of at most 3 in the approximation guarantee [10], [35]. For L_0-norm,

For the L_∞-norm, it is known that an optimal ultrametric can be computed in time proportional to the number of input distance pairs [36] and can be approximated in subquadratic time [37], [38]. For the case with general tree metrics, this problem is APX-hard and $O(1)$-approximation is known [35].

For constant p with $1 \le p < \infty$, the developments have been slower and remains much less understood to date [9], [10], [31], [39], [40]. Among them, L_1-norm in particular has been extensively studied [9], [10], [31] and a constant-factor approximation was given by [10]. For $1 < p < \infty$, $O(\log n \log\log n)$ remains the best approximation ratio [9].

When the goal is to edit the minimum number of pairwise distances so as to fit into an ultrametric, the problem is known as the ultrametric violation distance problem. This problem can be interpreted as numerical taxonomy under the L_0-norm and has been actively studied in recent years [11], [12], [41]–[45] for metric fitting, tree metric fitting, and ultrametric fitting. For the ultrametric version, the best result is a randomized 5-approximation [12]. We also note that Carmel, Das, Kipouridis and Pipis recently gave a single-pass, polynomial-time semi-streaming algorithm [45].

A. Our Result

We present a simple paradigm which greatly facilitates LP-rounding in hierarchical clustering. Our main result is a delicate algorithm for the hierarchical correlation clustering problem with a significantly improved approximation ratio compared to the previously known guarantee [10].

Theorem 1. There is a 25.7846-approximation algorithm for the hierarchical correlation clustering problem.

Our algorithm shares the same standard LP relaxation used in the literature [9]–[12] for the hierarchical clustering problems. However, we present a new property of this LP relaxation that allows us to pretend as if the objective has no negative items, intuitively speaking. Applying this property causes us to lose the multiplicative factor of up to two.

Our rounding algorithm inherits several key features from the two previous works [9], [10] with distinguishable technical characteristics, which we describe in detail in the next section. Our paradigm further reveal the core-technical problem in hierarchical clustering as the problem of finding cuts with prescribed properties regarding the average distance of a certain subset of cut pairs. To illustrate this perspective, we show that a direct application of the paradigm and techniques presented in this work leads to an alternative algorithm for the ultrametric violation distance problem that is quite simple to describe and analyze, whose performance guarantee matches the best known [12].

Corollary 2. There is a deterministic 5-approximation algorithm for the ultrametric violation distance problem.

B. Techniques and Discussion

We begin with a description on the LP formulation and an overview of the approaches introduced in [9] and [10] which handled the rounding problem in very different ways.

The LP-formulation models the clustering decisions via pairwise dissimilarity of the elements which have values within $[0, 1]$ and must satisfy the triangle inequality. Hence, it is instructive to interpret the LP-solutions as distance functions for the elements over the layers. Furthermore, the distance between any pair of elements satisfies the non-decreasing property top-down over the layers. Each label given for the element pairs over the layers corresponds to one item in the objective function with a sign being equal to the label itself, i.e., a plus label for an $\{u, v\}$ pair at the t-th layer corresponds to an item $x^{(t)}_{\{u,v\}}$ while a minus label gives an item $(1 - x^{(t)}_{\{u,v\}})$. Handling this discrepancy between signs has been the main challenge of this problem.

Following the convention in the literature, we will refer pairs labeled with $+$ to as $+$ *edges* and the rest as $-$ *edges*.

The Techniques in [9] and [10]: In [9], the hierarchical clustering is obtained in a top-down manner. This means that the decisions for the algorithm to make in each iteration is how the partition coming from the previous layer above should be subdivided, and the main challenge is to upper-bound the number of disagreements the current clustering decision will cause in all the successive layers below.

To deal with this issue, the authors in [9] distributed the overall LP value to each element and showed that, whenever a set P in the partition contains a $-$ edge $\{u, v\}$ with a distance at least $2/3$, there always exists an $r \in [0, 1/3]$ such that a ball B with radius r to be centered at either u or v will give a cut C, such that the weighted disagreements caused by C in all the successive layers below can be upper-bounded by $O(\log \log n) \cdot \log(\mathrm{Vol}(P)/\mathrm{Vol}(B)) \cdot \mathrm{Vol}(B)$, where $\mathrm{Vol}(A)$ for any $A \subseteq V$ accounts for the LP value of the edges contained within A over all the successive layers plus the LP value of the elements within A. The proof towards the existence of such a cut utilizes the famous region growing argument presented in [46] for the multicut problem. Summing up the cost over all such cuts gives a guarantee of $O(\log n \log \log n)$.

The approach presented in [10] starts from a reduction to the HIERARCHICAL CLUSTER AGREEMENT problem, in which the input for each layer is a pre-clustering of the elements into groups and the goal is to minimize the weighted symmetric difference with the input pre-clustering over the layers. The authors showed that an algorithm with an α-guarantee for the single-layer correlation clustering can readily be applied to obtain a pre-clustering for each layer with a multiplicative loss of $O(\alpha)$ in the overall guarantee.

The obtained instance for the hierarchical cluster agreement problem can be seen as an instance for the hierarchical correlation clustering problem where the intra-pre-cluster pairs act as $+$ edges and the inter-pre-cluster pairs act as $-$ edges. To handle the LP-solution for this new instance, a procedure called *LP-cleaning* is presented to further subdivide the input pre-clusters according to the LP-solution. This procedure uses a clever filtering setting to classify the elements such that, for each pre-cluster, either all the elements are made singleton pre-clusters or only a very small proportion of elements is made so. The setting guarantees that the number of "+ edges" separated in the new pre-clusters can be upper-bounded by the LP-value the fractional solution already has. Furthermore, the diameter of the new pre-clusters is guaranteed to smaller than $1/5$.

To obtain the hierarchical clustering, the authors present a brilliant approach that handles set-merging in a bottom-up manner, where the set-merging decisions are guided by the non-singleton pre-clusters computed in the above step and the structure of existing clusters coming from the previous layer. Roughly speaking, during the process, the algorithm records for each cluster a core subset which comes from a pre-cluster that has a small diameter and contains the majority of elements within the cluster. To handle the set-merging decisions for a partition coming from the previous layer, the algorithm unconditionally merges for each non-singleton pre-cluster all the clusters whose core subsets have a nonempty intersection with the pre-cluster. Then the union of the intersections becomes the core subset of the merged set. Using the properties obtained from the LP-cleaning procedure and the set-merging operation, the authors proved a set of cardinality bounds regarding the size of a cluster and its core subset via an involved induction argument.

Our Techniques: In this work we present a new paradigm that handles the LP-rounding problem for hierarchical clustering directly. Our algorithm inherits several key features from the two previous works [9], [10] with distinguishable technical characteristics.

Our algorithm uses the same LP relaxation used in previous works [9]–[12]. Our new, crucial observation is: in any optimal LP solution, the (weighted) number of $-$ edges with distance strictly smaller than one over the layers is always upper-bounded by the objective value of the LP solution itself. Hence, whenever the LP-solution pays a nonzero cost to separate a $-$ edge, the cost later incurred by that edge, if any, can readily be attributed to the cost of this LP-solution.

This suggests that we need to handle $-$ edges with distance one separately since the LP pays nothing for these edges. We will call them *forbidden edges.* Our analysis can be intuitively (but not formally) understood as defining a new instance of the problem where the forbidden edges become $-$ edges and non-forbidden edges become $+$ edges and then measuring solution costs there. This general property avoids the tricky problems in handling the discrepancy between the items with two different signs in the original objective function, greatly facilitating the task of LP-rounding and the analysis in the context of hierarchical clustering.

Our rounding algorithm consists of two components: (i) A pre-clustering algorithm which takes as input a distance function and produces a partition of the elements which guarantees bounds on both the diameters of the pre-clusters and the average distances of the non-forbidden cut edges. (ii) A delicate hierarchical clustering algorithm that handles the set-merging decisions in a bottom-up manner based on the information given by the pre-clusters and the structures of the existing partition coming from the previous layer.

For the first component, our pre-clustering algorithm is a pivot-based algorithm [9], [20] that takes an entirely different approach from the pre-clustering algorithm presented in [10] to some extent. On the other hand, our algorithm starts with a big cluster containing all the elements and iteratively subdividing the clusters until every cluster has a diameter strictly smaller than $1/3$. When this property is not yet met, an element with an eccentricity at least $1/3$ is picked, and the algorithm either makes the element a singleton cluster or it makes a cut with a ball of radius $1/3 - \epsilon$ centered at that element. This guarantees an average distance at least $1/6$ for the non-forbidden cut edges. Hence, the number of non-forbidden cut edges is bounded by a small factor to the objective value of the LP-solution. Moreover, we establish this bound using only cut edges that are not too-far-apart.

Our hierarchical clustering algorithm inherits the guided set-merging framework in [10]. Our algorithm imposes a set-merging condition that captures the elements necessary to provide a good structure for hierarchical clustering yet sufficient to yield a small constant loss in the approximation guarantee. We show that, this set-merging condition, combined with the diameter bound for the pre-clusters, leads to a geometrically-decreasing territory of the *non-core part* for any cluster in the hierarchy. This is the key to a set of substantially stronger cardinality bounds which scales with the core-parameter used in the set-merging condition.

To illustrate another use of our paradigm, we show that a direct application of our pre-clustering algorithm in a top-down manner with a radius parameter of $1/2$ yields a 5-approximation for the ultrametric violation distance problem. This provides a simple alternative algorithm to [12], which is obtained via a pivot-based randomized rounding approach top-down recursively.

Our paradigm reveals the nature of the hierarchical clustering problem as a problem of finding cuts with prescribed properties regarding the average distance for a certain subset of its cut edges. The above two results further suggest that improved approximation results would be possible if stronger properties on the cuts to be computed can be built. We believe our techniques would easily extend to other variations of hierarchical clustering problems with different objectives.

II. Problem Formulation

In the hierarchical correlation clustering problem, we are given a complete graph $G = (V, E)$ and ℓ layers of

$$\begin{aligned} \min \quad & \sum_{1\le t\le \ell} \delta_t \cdot \left(\sum_{\{u,v\}\in E_+^{(t)}} x_{\{u,v\}}^{(t)} + \sum_{\{u,v\}\in E_-^{(t)}} \left(1 - x_{\{u,v\}}^{(t)} \right) \right) & \text{LP-}(*) \\ \text{s.t.} \quad & x_{\{u,v\}}^{(t)} \le x_{\{u,p\}}^{(t)} + x_{\{p,v\}}^{(t)}, & \forall\, 1\le t\le \ell,\ u,v,p\in V, \\ & x_{\{u,v\}}^{(t+1)} \le x_{\{u,v\}}^{(t)}, & \forall\, 1\le t< \ell,\ u,v\in V. \\ & 0 \le x_{\{u,v\}}^{(t)} \le 1, & \forall\, 1\le t\le \ell,\ u,v\in V. \end{aligned}$$

Fig. 1. An LP formulation for the Hierarchical Correlation Clustering.

input information $(\delta_1, E_+^{(1)}), \ldots, (\delta_\ell, E_+^{(\ell)})$, where $\delta_t \in \mathbb{R}_{\ge 0}$ is a non-negative weight and $E_+^{(t)} \subseteq E$ is the set of edges labeled with $+$ at the t-th layer. We refer $E_+^{(t)}$ and $E_-^{(t)} := E \setminus E_+^{(t)}$ to as the $+$ *edges* and the $-$ *edges* at the t-th layer, respectively. We refer the 1-st layer to as the *bottom layer* and the ℓ-th layer to as the *top layer*.

A feasible solution to this problem is a tuple $(\mathcal{P}^{(1)}, \ldots, \mathcal{P}^{(\ell)})$, where $\mathcal{P}^{(t)}$ is a partition of V into groups such that $\mathcal{P}^{(t)}$ is a subdivision of $\mathcal{P}^{(t+1)}$ for any t with $1 \le t < \ell$. That is, for any $P \in \mathcal{P}^{(t)}$, there always exists $P' \in \mathcal{P}^{(t+1)}$ such that $P \subseteq P'$. We say that a collection of partitions $\{\mathcal{P}^{(t)}\}_{1\le t\le \ell}$ is *consistent* if it satisfies the above property.

The number of disagreements caused by $\mathcal{P}^{(t)}$ is defined to be the number of $+$ edges in $E_+^{(t)}$ that result in separated in $\mathcal{P}^{(t)}$ plus the number of $-$ edges in $E_-^{(t)}$ that are clustered into the same group in $\mathcal{P}^{(t)}$. Formally, we use

$$\begin{aligned} \#(\mathcal{P}^{(t)}) \; := \; & \sum_{P\in\mathcal{P}^{(t)}} \left| \left\{ \{p,q\} \in E_-^{(t)} \;:\; p,q \in P \right\} \right| \\ & + \sum_{\substack{P,P'\in\mathcal{P}^{(t)} \\ P\ne P'}} \left| \left\{ \{p,q\} \in E_+^{(t)} \;:\; p\in P, q \in P' \right\} \right| \end{aligned}$$

to denote the number of disagreements caused by $\mathcal{P}^{(t)}$. The goal of this problem is to compute a feasible solution $\{\mathcal{P}^{(t)}\}_{1\le t\le \ell}$ that minimizes the weighted disagreements $\sum_{1\le t\le \ell} \delta_t \cdot \#(\mathcal{P}^{(t)})$.

We use the LP formulation in Fig. 1 from [9], [10] for the hierarchical correlation clustering problem. In this formulation, for each $1 \le t \le \ell$ and $\{u,v\} \in \binom{V}{2}$, we use an indicator variable $x_{\{u,v\}}^{(t)} \in \{0,1\}$ to denote the clustering decision for u and v at the t-th layer, i.e., $x_{\{u,v\}}^{(t)} = 0$ if and only if $u, v \in Q$ for some $Q \in \mathcal{P}^{(t)}$ and $x_{\{u,v\}}^{(t)} = 1$ otherwise.

Since the triangle inequality is satisfied, we will interpret $x^{(t)}$ as a distance function defined on the elements at the t-th layer. Moreover, for each $\{u,v\} \in \binom{V}{2}$, the hierarchical constraint requires that $x_{\{u,v\}}^{(t)}$ is non-increasing bottom-up over the layers. In the rest of this paper we will implicitly assume that $x_{\{u,u\}}^{(t)} = 0$ for any $u \in V$.

Notation: We use the following notation. For any $S \subseteq V$, we use $\overline{S}$ to denote $V \setminus S$. Let $x^{(1)}, \ldots, x^{(\ell)}$ be a feasible solution for LP-$(*)$. For any $1 \le t \le \ell$, $P, Q \subseteq V$, and $r \in \mathbb{R}_{\ge 0}$, we use

$$\mathrm{Ball}_{<r}^{(t)}(P,Q) \; := \; \left\{ v \in Q \; : \; \min_{u\in P} x_{\{v,u\}}^{(t)} < r \right\}$$

to denote the set of elements in Q that are at a distance of strictly less than r from some element in P in the t-th layer. We use

$$\mathrm{diam}^{(t)}(Q) \; := \; \max_{u,v\in Q} x_{\{u,v\}}^{(t)}$$

to denote the diameter of the set Q with respect to $x^{(t)}$.

When an arbitrary distance function x is referenced, we use $\mathrm{Ball}_{<r}^{(x)}(P,Q)$ and $\mathrm{diam}^{(x)}(Q)$ to denote the same concept with respect to the distance function x.

III. LP-Rounding Algorithm

Solve the LP-$(*)$ in Fig. 1 for an optimal solution $\tilde{x}$. For any $1 \le t \le \ell$, define

$$F^{(t)} \; := \; \left\{ \{p,q\} \in E_-^{(t)} \; : \; \tilde{x}_{\{p,q\}}^{(t)} = 1 \right\}$$

to be the set of $-$ edges with distance one at the t-th layer. We refer these edges to as *forbidden edges* since the LP solution pays no cost to separate them.

Our rounding algorithm consists of two parts. The first part is a pre-clustering algorithm that takes as input a distance function x and produces a partition $\mathcal{Q}$ with the following two properties.

1) For each $Q \in \mathcal{Q}$, the diameter of Q with respect to x is strictly smaller than $1/3$.

Algorithm 1 Hierarchical-clustering($\{\tilde{x}^{(t)}\}_{1\le t\le \ell}$)

$\mathcal{P}^{(0)} \leftarrow \{\{v\} \colon v \in V\}$ and $\Delta^{(0)}(P) \leftarrow P$ for all $P \in \mathcal{P}^{(0)}$.
for $t = 1$ to ℓ **do**
    $\mathcal{P}^{(t)} \leftarrow \emptyset$.
    $\mathcal{Q}^{(t)} \leftarrow$ Pre-clustering($\tilde{x}^{(t)}$). `// Pre-clustering from Algorithm 2.`
    for all $Q \in \mathcal{Q}^{(t)}$ **do**
        Let $\mathrm{Candi}^{(t)}(Q)$ be the set containing all the sets $P \in \mathcal{P}^{(t-1)}$ such that

$$\Delta^{(t-1)}(P) \cap P \cap Q \neq \emptyset \quad \text{and} \quad \left| \mathrm{Ball}^{(t)}_{<2/3}\left(P \cap Q,\ P \cap \overline{Q} \right) \right| < \alpha \cdot |P \cap Q|.$$

        if $\mathrm{Candi}^{(t)}(Q) \neq \emptyset$ **then** `// Merge all the sets in` $\mathrm{Candi}^{(t)}(Q)$.
            Let $P_Q := \bigcup_{P \in \mathrm{Candi}^{(t)}(Q)} P$.
            Add P_Q to $\mathcal{P}^{(t)}$ and set $\Delta^{(t)}(P_Q) \leftarrow Q$.
        end if
    end for
    for all $P \in \mathcal{P}^{(t-1)} \setminus \bigcup_{Q \in \mathcal{Q}^{(t)}} \mathrm{Candi}^{(t)}(Q)$ **do** `// Carry the unmerged sets over to` $\mathcal{P}^{(t)}$
        Add P to $\mathcal{P}^{(t)}$ and set $\Delta^{(t)}(P) \leftarrow \Delta^{(t-1)}(P)$.
    end for
end for
return $\{\mathcal{P}^{(t)}\}_{1\le t\le \ell}$.

2) The *not-too-far-apart edges* separated by $\mathcal{Q}$ have a large average distance. In particular, those with a distance at most $5/6$ already have an average distance at least $1/6$.

We describe this algorithm later in this section.

The second part is a hierarchical clustering algorithm that outputs a consistent partitioning $\{\mathcal{P}^{(t)}\}_{1\le t\le \ell}$, where each set P in $\mathcal{P}^{(t)}$ is further associated with a gluer set denoted $\Delta^{(t)}(P)$.

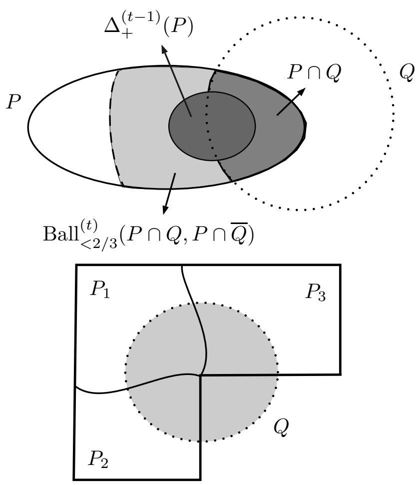


Fig. 2. The setting for intersection requirement (1) between $P \in \mathcal{P}^{(t-1)}$ and $Q \in \mathcal{Q}^{(t)}$ and the set-merging operation for the sets in $\mathrm{Candi}^{(t)}(Q)$ with Q being the gluer set.

Let $\mathcal{P}^{(0)} := \{\{v\}\}_{v\in V}$ be the initial singleton clustering and define $\Delta^{(0)}(P) := P$ for all $P \in \mathcal{P}^{(0)}$. For the t-th layer, where $t = 1, 2, \ldots, \ell$ in order, the algorithm first applies the pre-clustering algorithm on $\tilde{x}^{(t)}$ to obtain $\mathcal{Q}^{(t)}$ and iterates over all $Q \in \mathcal{Q}^{(t)}$. For each Q, the algorithm collects all the sets $P \in \mathcal{P}^{(t-1)}$ that satisfies the following intersection requirements with Q

$$\Delta_+^{(t-1)}(P) \cap Q \neq \emptyset \quad \text{and} \\ \left| \mathrm{Ball}^{(t)}_{<\frac{2}{3}}\left(P \cap Q,\ P \cap \overline{Q} \right) \right| \ < \ \alpha \cdot |P \cap Q| \, , \tag{1}$$

where $\Delta_+^{(t-1)}(P) := \Delta^{(t-1)}(P) \cap P$ will be referred to as the *core* of P and $\alpha := 0.3936$. Note that $\alpha < 1/2$. Refer to Fig. 2 for an illustration on this condition.

Let $\mathrm{Candi}^{(t)}(Q)$ denote the sets collected for Q. The algorithm merges all the sets in $\mathrm{Candi}^{(t)}(Q)$, if it is nonempty, and sets Q to be the gluer set of the merged set. When all the $Q \in \mathcal{Q}^{(t)}$ are considered, the algorithm carries all the unmerged sets in $\mathcal{P}^{(t-1)}$ over to $\mathcal{P}^{(t)}$ with their gluer sets unchanged. Refer to Algorithm 1 for a pseudo-code of this algorithm.

Consider the partition $\mathcal{P}^{(t)}$ computed for any $1 \le t \le \ell$. We refer the sets $\{P_Q\}_{Q\colon \mathrm{Candi}^{(t)}(Q)\neq\emptyset}$ to as *newly-created* at the t-th layer in the rest of this paper as they are formed as a result of merging the sets in $\mathrm{Candi}^{(t)}(Q)$ for some $Q \in \mathcal{Q}^{(t)}$. On the contrary, the unmerged sets carried over from $\mathcal{P}^{(t-1)}$ are referred to as *previously-formed*.

Since the distance of any edge is non-increasing bottom-up over the layers, it follows that the diameter of any $Q \in \mathcal{Q}^{(t')}$ at the t-th layer is also strictly smaller than $1/3$ for any $t \ge t'$. Hence, it follows by construction that $\mathrm{diam}^{(t)}(\Delta_+^{(t)}(P)) < \frac{1}{3}$ for any $P \in \mathcal{P}^{(t)}$ and $1 \le t \le \ell$.

Algorithm 2 Pre-clustering(x)

Let $\mathcal{Q} \leftarrow \{V\}$.
while there exists $Q \in \mathcal{Q}$ with $\operatorname{diam}^{(x)}(Q) \geq 1/3$ **do**
  Pick (v, Q) such that $v \in Q \in \mathcal{Q}$ and $\max_{u \in Q} x_{\{u,v\}} \geq 1/3$.
  $\mathcal{Q}' \leftarrow$ ONE-THIRD-REFINE-CUT(Q, v, x).
  Replace Q with the sets in $\mathcal{Q}'$ in $\mathcal{Q}$.
end while
return $\mathcal{Q}$.

procedure ONE-THIRD-REFINE-CUT(Q, v, x)
  if Condition (2) is satisfied for (Q, v) **then**
    return $\{\ \{v\},\ Q \setminus \{v\}\ \}$. `// make` v `a singleton`
  else
    return $\{\ \operatorname{Ball}^{(x)}_{<1/3}(v, Q),\ \ Q \setminus \operatorname{Ball}^{(x)}_{<1/3}(v, Q)\ \}$. `// cut at` $1/3 - \epsilon$
  end if
end procedure

The following lemma shows that the candidates to be merged for each $Q \in \mathcal{Q}^{(t)}$ is unambiguous, and hence Algorithm 1 is well-defined.

Lemma 3. $\operatorname{Candi}^{(t)}(Q) \cap \operatorname{Candi}^{(t)}(Q') = \emptyset$ for any $Q, Q' \in \mathcal{Q}^{(t)}$ with $Q \neq Q'$.

Proof. Suppose that $P \in \operatorname{Candi}^{(t)}(Q) \cap \operatorname{Candi}^{(t)}(Q')$ for some $P \in \mathcal{P}^{(t-1)}$ and $Q, Q' \in \mathcal{Q}^{(t)}$. Let $p \in \Delta_+^{(t-1)}(P) \cap Q$ and $q \in \Delta_+^{(t-1)}(P) \cap Q'$ be two elements. We have

$$\tilde{x}^{(t)}_{\{p,q\}} \ \leq \ \tilde{x}^{(t-1)}_{\{p,q\}} \ < \ \frac{1}{3}$$

by the non-increasing property of the distance function bottom-up over the layers. Then Condition (1) and the diameter bounds of Q, Q' at the t-th layer imply that

$$P \cap Q \subseteq \operatorname{Ball}^{(t)}_{<2/3}(P \cap Q', P \cap \overline{Q'}),$$

and

$$P \cap Q' \subseteq \operatorname{Ball}^{(t)}_{<2/3}(P \cap Q, P \cap \overline{Q}),$$

and hence $|P \cap Q| \ < \ \alpha \cdot |P \cap Q'| \ < \ \alpha^2 \cdot |P \cap Q|$, a contradiction. ☐

Below we describe the pre-clustering algorithm (Algorithm 2). The algorithm takes as input a distance function x, starts with one big set $\mathcal{Q} := \{V\}$, and refines it repeatedly until $\operatorname{diam}^{(x)}(Q) < 1/3$ for all $Q \in \mathcal{Q}$. In each refining iteration, it picks a $Q \in \mathcal{Q}$ and a vertex $v \in Q$ such that $\max_{u \in Q} x_{\{u,v\}} \geq 1/3$. If

$$\sum_{q \in \operatorname{Ball}^{(x)}_{<1/3}(v,Q)} x_{\{v,q\}} \geq \frac{1}{3} \cdot |\operatorname{Ball}^{(x)}_{<1/3}(v, Q)| - \frac{1}{6} \cdot |\operatorname{Ball}^{(x)}_{<1/2}(v, Q)| - \frac{1}{6}, \tag{2}$$

then the algorithm makes v a singleton pre-cluster by replacing Q with $\{v\}$ and $Q \setminus \{v\}$. Otherwise, Q is replaced with $\operatorname{Ball}^{(x)}_{<1/3}(v, Q)$ and $Q \setminus \operatorname{Ball}^{(x)}_{<1/3}(v, Q)$. We make a note that in (2) we use the implicit assumption that $x_{\{u,u\}} = 0$ for any $u \in V$ in the distance function x.

This concludes our rounding algorithm for the hierarchical correlation clustering problem.

IV. Overview of the Analysis

Let $\{\mathcal{P}^{(t)}\}_{1 \leq t \leq \ell}$ be the output of Algorithm 1 and $\#(\mathcal{P}^{(t)})$ be the number of disagreements caused by $\mathcal{P}^{(t)}$.

Define $NF_-^{(t)} := E_-^{(t)} \setminus F^{(t)}$ to be the set of non-forbidden $-$ edges at the t-th layer. We have that

$$\#(\mathcal{P}^{(t)}) \ \leq \ \sum_{P \in \mathcal{P}^{(t)}} \big(\ \#_{\mathrm{F}}(P) \ + \ \#_{\mathrm{NF}_-}(P)\ \big) + \sum_{\substack{P, P' \in \mathcal{P}^{(t)}, \\ P \neq P'}} \#_{\mathrm{NF}}(P, P'), \tag{3}$$

where

- $\#_{\mathrm{F}}(P) := |\{\ \{i,j\} \in F^{(t)} \colon i, j \in P\ \}|$ is the number of forbidden edges clustered within P,
- $\#_{\mathrm{NF}_-}(P) := |\{\ \{i,j\} \in NF_-^{(t)} \colon i, j \in P\ \}|$ is the number of non-forbidden $-$ edges clustered within P, and
- $\#_{\mathrm{NF}}(P, P') := |\{\ \{i,j\} \notin F^{(t)} \colon i \in P, j \in P'\ \}|$ is the number of non-forbidden edges between P and P'.

Recall that $\tilde{x}$ is an optimal solution to LP-$(*)$. To bound the weighted disagreements, we use a rather surprising property, proved in Section V-D.

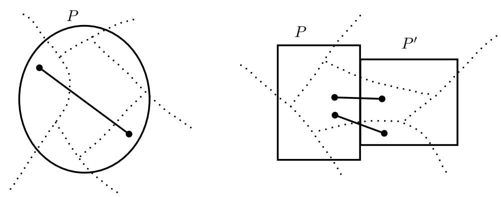


Fig. 3. Two types of disagreements we will focus on – (a) Forbidden edges clustered into the same part P. (b) Non-forbidden edges across different parts P and P'.

Lemma 4 (Section V-D)**.**

$$\sum_{1\le t\le \ell} \delta_t \cdot \left(\sum_{\{u,v\}\in E_+^{(t)}} \tilde{x}_{u,v}^{(t)} + \sum_{\{u,v\}\in E_-^{(t)}} \left(1-\tilde{x}_{u,v}^{(t)}\right) \right) \ge \sum_{1\le t\le \ell} \delta_t \cdot \left| NF_-^{(t)} \right| .$$

It follows from Lemma 4 that the weighted disagreements caused by the edges in $NF_-^{(t)}$, if any, can readily be attributed to the cost of the optimal LP solution. Furthermore, they can be treated as if they were + edges when necessary.

Next we bound $\sum_P \#_{\mathrm{F}}(P)$ and $\sum_{P\ne P'} \#_{\mathrm{NF}}(P,P')$ in terms of $|\,\mathrm{NF}(\mathcal{Q}^{(t)})|$, where we use

$$\mathrm{NF}(\mathcal{Q}^{(t)}) := \left\{ \{i,j\} \notin F^{(t)} : \{i,j\} \text{ separated in } \mathcal{Q}^{(t)} \right\}$$

to denote the set of non-forbidden edges that are separated in $\mathcal{Q}^{(t)}$. As for $\#_{\mathrm{F}}(P)$, we prove the following lemma in Section V-B and A.

Lemma 5 (Section V-B, Section A)**.** For $\alpha := 0.3936$ and any $P \in \mathcal{P}^{(t)}$, we have

$$\#_{\mathrm{F}}(P) \le \frac{(2-\alpha)(1+\alpha)^2}{2(1-\alpha)^2} \cdot \beta \cdot \left| \mathrm{NF}(\mathcal{Q}^{(t)}, P) \right|,$$

where $\beta := 0.8346$ and $\mathrm{NF}(\mathcal{Q}^{(t)}, P) := \left\{ \{i,j\} \in \mathrm{NF}(\mathcal{Q}^{(t)}) : i,j \in P \right\}$ denotes the set of edges in $\mathrm{NF}(\mathcal{Q}^{(t)})$ that reside within P.

For $\#_{\mathrm{NF}}(P,P')$, we prove the following lemma.

Lemma 6 (Section V-B)**.** For any $P, P' \in \mathcal{P}^{(t)}$ with $P \ne P'$, we have

$$\#_{\mathrm{NF}}(P,P') \le \max\left\{ \frac{1}{1-\alpha}, \frac{1+\alpha}{\alpha} \right\} \cdot \left| \mathrm{NF}(\mathcal{Q}^{(t)}, P, P') \right|,$$

where $\mathrm{NF}(\mathcal{Q}^{(t)}, P, P') := \left\{ \{i,j\} \in \mathrm{NF}(\mathcal{Q}^{(t)}) \colon\ i \in P,\ j \in P' \right\}$ denotes the set of edges in $\mathrm{NF}(\mathcal{Q}^{(t)})$ that are between P and P'.

Lemma 5 and Lemma 6 bound $\sum_P \#_{\mathrm{F}}(P) + \sum_{P\ne P'} \#_{\mathrm{NF}}(P,P')$ in terms of $|\,\mathrm{NF}(\mathcal{Q}^{(t)})|$. To further bound $|\,\mathrm{NF}(\mathcal{Q}^{(t)})|$, we show that the non-forbidden edges separated in $\mathcal{Q}^{(t)}$ have an average distance at least $1/6$ via a stronger statement.

Lemma 7 (Section V-C)**.** Consider line 4 in Algorithm 2 with input distance function x. Let v be the pivot chosen in that iteration and (Q_1, Q_2) with $v \in Q_1$ be the pair returned by the procedure ONE-THIRD-REFINE-CUT. Then

$$\sum_{\substack{\{i,j\}\in \mathrm{NF}(Q_1,Q_2),\\ i\in Q_1,\\ j\in \mathrm{Ball}^{(x)}_{<1/2}(v,Q_2)}} \left(\min\left\{ x_{\{v,j\}}, \frac{1}{3} \right\} - x_{\{v,i\}} \right) \ge \frac{1}{6} \cdot \left| \left\{ \substack{\{i,j\}\in \mathrm{NF}(Q_1,Q_2),\\ i\in Q_1,\\ j\in \mathrm{Ball}^{(x)}_{<1/2}(v,Q_2)} \right\} \right|,$$

where $\mathrm{NF}(Q_1, Q_2)$ denotes the set of non-forbidden edges between Q_1 and Q_2.

Since $|x_{\{v,i\}} - x_{\{v,j\}}|$ is a lower-bound for $x_{\{i,j\}}$ for any $i, j \in Q_1 \cup Q_2$ by the triangle inequality, Lemma 7 guarantees an average distance at least $1/6$ for the edges in $\mathrm{NF}(Q_1, Q_2)$. Moreover, although the actual distance of such edges can be much larger than the average, the statement ensures that only a reasonably small amount of it is charged to establish the bound.

Using Lemma 7, we bound $\left|\mathrm{NF}(\mathcal{Q}^{(t)}) \cap E_+^{(t)}\right|$ in terms of the objective value of $\tilde{x}^{(t)}$. Combining all the above with Inequality (3), we obtain the following lemma in Section V-E.

Lemma 8 (Section V-E)**.**

$$\sum_{1\le t\le \ell} \delta_t \cdot \#(\mathcal{P}^{(t)}) \le (7c(\alpha)+1) \cdot \sum_{1\le t\le \ell} \delta_t \cdot \left(\sum_{\{u,v\}\in E_+^{(t)}} \tilde{x}_{\{u,v\}}^{(t)} + \sum_{\{u,v\}\notin E_+^{(t)}} \left(1-\tilde{x}_{\{u,v\}}^{(t)}\right) \right),$$

where $c(\alpha) := \max\left\{ \frac{\beta(2-\alpha)(1+\alpha)^2}{2(1-\alpha)^2}, \frac{1}{1-\alpha}, \frac{1+\alpha}{\alpha} \right\} \approx 3.5406$ for $\alpha := 0.3936$, and $\beta := 0.8346$.

This yields the approximation guarantee of 25.7846.

V. Bounding the Weighted Disagreements

In this section we provide the proofs for the lemmas described in the previous section.

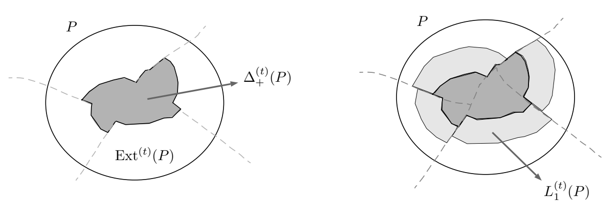
A. Cardinality Bounds for $P \in \mathcal{P}^{(t)}$

To prove Lemma 5 and Lemma 6, one of the key ingredients is a set of cardinality bounds regarding the territory of any cluster in terms of its core.

In particular, the intersection requirement in (1) leads to a decrease of the non-core territory in a geometric order for any cluster in the hierarchy.

Let $P \in \mathcal{P}^{(t)}$ be a cluster in the t-th layer. Recall that $\Delta^{(t)}(P)$ denotes the gluer set of P and $\Delta_+^{(t)}(P) := P \cap \Delta^{(t)}(P)$ is referred to as the core set of P. Additionally define

- $\mathrm{Ext}^{(t)}(P) := P \setminus \Delta_+^{(t)}(P)$ to be the extended part of P,
- $\ell(t,P)$ to be the top-most layer up to the t-th layer at which P is newly-created, and
  $L_1^{(t)}(P)$ to be the elements in the $2/3$-vicinity of $P' \cap \Delta_+^{(\ell(t,P))}(P)$ within P' at the $\ell(t,P)$-th layer over all $P' \in \mathrm{Candi}^{(\ell(t,P))}(\Delta^{(\ell(t,P))}(P))$.
  Formally,
  $$L_1^{(t)}(P) \;:=\; \bigcup_{P' \in \mathrm{Candi}^{(\ell(t,P))}(Q_P)} \mathrm{Ball}_{<2/3}^{(\ell(t,P))}\left(P' \cap Q_P, P' \cap \overline{Q_P}\right),$$
  where we use $Q_P := \Delta^{(\ell(t,P))}(P)$ to denote the gluer set of P at the $\ell(t,P)$-th layer. We note that $\ell(t,P)$ is always well-defined.
  Refer to the figure below for an illustration. Note that it follows that $|L_1^{(t)}(P)| < \alpha \cdot |\Delta_+^{(t)}(P)|$ by the merging condition in Algorithm 1.



We prove the following helper lemma regarding the cardinality of the extended part of P and the reasonably dense structure in any $2/3$-vicinity of it. The statements are proved using the intersection requirement (1) in Algorithm 1 and the diameter bound of $1/3$ for each pre-cluster.

Lemma 9. Let $P \in \mathcal{P}^{(t)}$ be a cluster. We have

$$|\,\mathrm{Ext}^{(t)}(P)| \;\le\; \min\left\{\; \frac{\alpha}{1-\alpha}\cdot|\Delta_+^{(t)}(P)|,\;\; \frac{1}{1-\alpha}\cdot|L_1^{(t)}(P)| \;\right\}.$$

Furthermore, for any nonempty $A \subseteq \mathrm{Ext}^{(t)}(P)$, there exists $K_P^{(t)}(A) \subseteq \mathrm{Ball}_{<2/3}^{(t)}(A, P \setminus A)$ such that

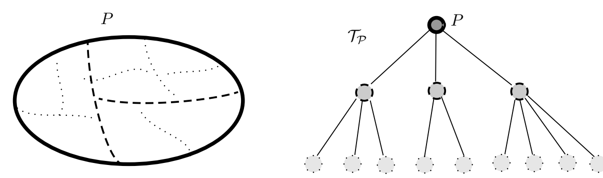
$$|A| \;\le\; \frac{\alpha}{1-\alpha}\cdot|K_P^{(t)}(A)|.$$

We prove the statements in Lemma 9 separately. Note that it suffices to prove the statements for the $\ell(t,P)$-th layer. Hence, in the following we assume that P is newly-created at the t-th layer.

Consider a tree $\mathcal{T}_P$ built to represent the sequence of set-merging processes leading to P, where each node $v \in \mathcal{T}_P$ is associated with the following two auxiliary information.

1) $H(v)$ which is a cluster newly-created at the t'-th layer for some $t' \le t$. Literally this will be the set to which the node v corresponds.
2) $\ell(v)$ which is an index of a layer at which $H(v)$ is newly-created. Refer to the construction described below.

We define $\mathcal{T}_P$ by describing a procedure to construct it. The process starts with a singleton tree with the root node r such that $H(r) := P$ and $\ell(r) := t$. In each of the iterations that follow, consider the set of current leaf nodes v in $\mathcal{T}_P$ with $\ell(v) > 1$. For each of such leaf nodes v, consider the sets contained in $\mathrm{Candi}^{(\ell(v))}(\Delta^{(\ell(v))}(v))$. For each $P' \in \mathrm{Candi}^{(\ell(v))}(\Delta^{(\ell(v))}(v))$, create a node for P', say, u, as a child node of v. Set $H(u) := P'$ and $\ell(u)$ to be the largest index between 1 and $\ell(v)$ such that P' is newly-created at the $\ell(u)$-th layer.



Proof of Lemma 9, Part I. Use a pre-order traversal on $\mathcal{T}_P$ to define a set of layers as follows. Initially, define $A_1 := \Delta_+^{(t)}(P)$ and $\mathrm{Base}_1 := A_1 \cup L_1$, where

$$L_1(P) \;:=\; \bigcup_{P' \in \mathrm{Candi}^{(t)}(\Delta^{(t)}(P))} \mathrm{Ball}_{<2/3}^{(t)}\left(P' \cap \Delta^{(t)}(P), P' \cap \overline{\Delta^{(t)}(P)}\right).$$

The traversal starts with the root node P and the initial index $i = 1$. For any vertex v encountered during the traversal, process v as follows. If $\Delta_+^{(\ell(v))}(H(v)) \subseteq \mathrm{Base}_i$, then nothing needs to be done. In this case we proceed to the next vertex directly.

On the other hand, if $\Delta_+^{(\ell(v))}(H(v)) \not\subseteq \mathrm{Base}_i$, then consider the parent node $p(v)$ of v in $\mathcal{T}_P$. Such node exists since $\Delta_+^{(\ell(r))}(H(r)) \subseteq \mathrm{Base}_1$ for the root node r. Let $Q(v) := \Delta^{(\ell(p(v)))}(H(p(v)))$ denote the gluer set of $H(p(v))$. Increase i by one and define

$$A_i \;:=\; H(v) \cap Q(v) \quad \text{and}$$
$$\mathrm{Base}_i \;:=\; \mathrm{Base}_{i-1} \,\cup\, \mathrm{Ball}_{<2/3}^{(\ell(p(v)))}\left(\,A_i,\; H(v) \cap \overline{A_i}\,\right).$$

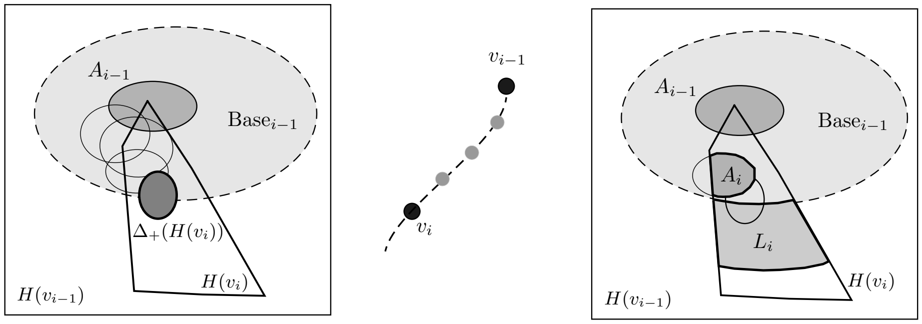


Fig. 4. From v_{i-1} on which A_{i-1} and Base_{i-1} are defined, identify the first descendant v_i whose core set is not contained within Base_{i-1}. Then A_i, L_i, and Base_i are defined accordingly.

Note that $A_i \neq \emptyset$. Also refer to Fig. 4 for an illustration for the definitions. For any $i \geq 2$, define

$$L_i := \mathrm{Base}_i \setminus \mathrm{Base}_{i-1}\,.$$

For any index $i \geq 2$, let v_i denote the specific vertex at which the sets A_i, Base_i, and L_i are defined during the pre-order traversal. For $i = 1$, define v_1 to be the root node r for consistency. Also refer to Fig. 4 for an illustration of the definitions.

We prove the following two invariant conditions regarding the sets defined during the traversal.

i. $|L_i| \leq \alpha |A_i|$ for any $i \geq 1$.
ii. $A_i \cap A_j = \emptyset$ for any $i \neq j$.

For condition (i), it suffices to consider any $i \geq 2$. By the definition of L_i we have

$$\begin{aligned} L_i &\subseteq \mathrm{Ball}^{(\ell(p(v_i)))}_{<2/3}\left(A_i,\ H(v_i) \cap \overline{A_i} \right) \\ &= \mathrm{Ball}^{(\ell(p(v_i)))}_{<2/3}\left(H(v_i) \cap Q(v_i),\ H(v_i) \cap \overline{Q(v_i)} \right), \end{aligned}$$

where we recall that $Q(v_i)$ denotes the gluer set of $H(p(v_i))$. Since $Q(v_i)$ results in the merge of $H(v_i)$, condition (1) is satisfied between $H(v_i)$ and $Q(v_i)$. Hence, $|L_i| \leq \alpha |A_i|$.

For condition (ii), consider any i, j with $1 \leq j < i$. Let v_k be the least common ancestor of v_i and v_j in $\mathcal{T}_P$. If $v_k \notin \{v_i, v_j\}$, then v_i and v_j belong to different subtrees rooted at v_k. Since the sets to which the children nodes of v_k correspond form a partition of $H(v_k)$, it follows that $H(v_i) \cap H(v_j) = \emptyset$ and this condition holds.

Now consider the other case where $v_k \in \{v_i, v_j\}$, in which v_j is a proper ancestor of v_i. Since the core of v_i is not contained within Base_{i-1}, there exists an element $q \in \Delta_+^{(\ell(v_i))}(H(v_i)) \setminus \mathrm{Base}_{i-1}$. Observe that, since A_i intersects $\Delta_+^{(\ell(v_i))}(v_i)$, by the diameter bounds of $Q(v_i)$ and $\Delta_+^{(\ell(v_i))}(v_i)$ together with the triangle inequality, we have

$$\tilde{x}^{(\ell(p(v_i)))}_{\{q,w\}} < \frac{2}{3} \quad \text{for any } w \in A_i. \tag{4}$$

Define

$$A'_j := \begin{cases} A_j, & \text{if } j > 1, \\ A_1 \cap H(v'),\ \text{where } v' \text{ is the child} & \\ \quad \text{of } v_1 \text{ such that } A_i \subseteq H(v'), & \text{otherwise.} \end{cases}$$

Note that, to prove that $A_i \cap A_j = \emptyset$, it suffices to prove the statement for A_i and A'_j. For this, we prove the following claim.

Claim.

$$\tilde{x}^{(\ell(p(v_i)))}_{\{q,u\}} \geq \frac{2}{3} \quad \text{for any } u \in A'_j.$$

Proof. Consider the case for which $j \geq 2$. We have

$$\mathrm{Base}_j \subseteq \mathrm{Base}_{i-1} \quad \text{and} \quad H(v_i) \subseteq H(v_j),$$

which shows that $q \in H(v_j) \setminus \mathrm{Base}_j$. Since $j \geq 2$, from the construction, we have that

$$\mathrm{Ball}^{(\ell(p(v_j)))}_{<2/3}\left(A_j,\ H(v_j) \cap \overline{A_j} \right) \subseteq \mathrm{Base}_j\,.$$

This implies that

$$\tilde{x}^{(\ell(p(v_i)))}_{\{q,u\}} \geq \tilde{x}^{(\ell(p(v_j)))}_{\{q,u\}} \geq \frac{2}{3} \quad \text{for any } u \in A_j,$$

where the first inequality follows from the monotonic property of the distances over the layers. Since $A'_j := A_j$ when $j > 1$, we are done with this case.

For the other case with $j = 1$, recall that v' is the child of v_1 such that $A_i \subseteq H(v')$. From the construction, we have

$$\mathrm{Ball}^{(\ell(p(v')))}_{<2/3}\left(A'_1,\ H(v_1) \cap \overline{A'_1} \right) \subseteq \mathrm{Base}_1,$$

and hence

$$\tilde{x}^{(\ell(p(v')))}_{\{q,u\}} \geq \frac{2}{3} \quad \text{for any } u \in A'_1,$$

Note that since $\ell(p(v')) = \ell(v_1) \geq \ell(p(v_i))$, the monotonicity over the layers completes the proof. $\square$

Combining the above claim with (4), we have

$$\tilde{x}^{(\ell(p(v_i)))}_{\{u,w\}} \geq \tilde{x}^{(\ell(p(v_i)))}_{\{q,u\}} - \tilde{x}^{(\ell(p(v_i)))}_{\{q,w\}} > 0.$$

Since this holds for any $w \in A_i$ and any $u \in A'_j$, we have $A_i \cap A'_j = \emptyset$ and condition (ii) follows.

We are ready to prove the statement of this lemma. It follows from the above definitions that $L_i \cap L_j = \emptyset$ for all $i \neq j$ and $|\operatorname{Ext}^{(t)}(P)| = \sum_{i \geq 1} |L_i|$. From invariant condition (i) and (ii), we obtain that

$$\begin{aligned} |\operatorname{Ext}^{(t)}(P)| \; = \; \sum_{i \geq 1} |L_i| \; &= \; |L_1| \; + \; \sum_{i \geq 2} |L_i| \\ &\leq \; |L_1| \; + \; \alpha \cdot \sum_{i \geq 2} |A_i| \\ &\leq \; |L_1| \; + \; \alpha \cdot |\operatorname{Ext}^{(t)}(P)|. \end{aligned}$$

This gives $|\operatorname{Ext}^{(t)}(P)| \leq \frac{1}{1-\alpha} \cdot |L_1|$. The first part of the lemma follows from $|L_1| \leq \alpha \cdot |\Delta_+^{(t)}(P)|$ $\square$

In the following we complete the second part of Lemma 9.

Proof of Lemma 9, Part II. Consider the tree $\mathcal{T}_P$ and the set of nodes $u \in \mathcal{T}_P$ whose core set intersects A and whose every ancestor node has its core set being disjoint with A. Formally,

$$\Delta_+^{(\ell(u))}(H(u)) \cap A \neq \emptyset \quad \text{and}$$
$$\Delta_+^{(\ell(v))}(H(v)) \cap A = \emptyset \;\text{ for any ancestor } v \text{ of } u \text{ in } \mathcal{T}_P.$$

Let $u_1, u_2, \ldots, u_k$ be the set of all such nodes and $p_1, p_2, \ldots, p_m$ be the parent nodes of $\{u_i\}_{1 \leq i \leq k}$. Note that $m \leq k$ since some nodes in $\{u_i\}_{1 \leq i \leq k}$ may share a common parent.

It follows that $A \subseteq \bigcup_{1 \leq i \leq m} \operatorname{Ext}^{(\ell(p_i))}(H(p_i))$ and hence

$$|A| \;\leq\; \frac{\alpha}{1-\alpha} \cdot \sum_{1 \leq i \leq m} \left| \Delta_+^{(\ell(p_i))}(H(p_i)) \right|$$

by the bound proved above for the first part of this lemma. Furthermore, for each p_i, there exists u_j which is a child node of p_i such that

$$\Delta_+^{(\ell(p_i))}(H(p_i)) \;\cap\; \Delta_+^{(\ell(u_j))}(H(u_j)) \;\neq\; \emptyset \quad \text{and}$$
$$A \cap \Delta_+^{(\ell(u_j))}(H(u_j)) \neq \emptyset.$$

Since $\max\{ \operatorname{diam}^{(t)}(\Delta_+^{(\ell(p_i))}(H(p_i))), \operatorname{diam}^{(t)}(\Delta_+^{(\ell(u_j))}(H(u_j))) \} < 1/3$ by the monotonic property of the distance functions over the layers and the diameter bound of the core sets, it follows that

$$\bigcup_{1 \leq i \leq m} \Delta_+^{(\ell(p_i))}(H(p_i)) \;\;\subseteq\;\; \operatorname{Ball}_{<2/3}^{(t)}(A, P \setminus A).$$

From the construction, $\operatorname{Ext}^{(\ell(p_i))}(H(p_i))$ for all $1 \leq i \leq m$ are disjoint. Hence, taking $K_P^{(t)}(A) := \bigcup_{1 \leq i \leq m} \Delta_+^{(\ell(p_i))}(H(p_i))$ completes the proof of this lemma. $\square$

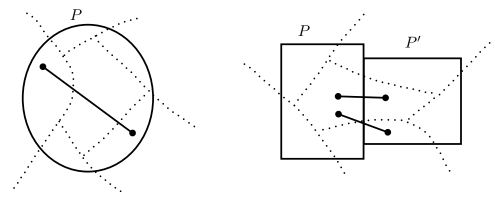


Fig. 5. Two types of disagreements to bound in this section – (a) Forbidden edges clustered into the same part P. (b) Non-forbidden edges across different parts P and P'.

B. Counting the Number of Disagreements

We count the total number of disagreements in $\mathcal{P}^{(t)}$ in terms of the number of edges in $\operatorname{NF}(\mathcal{Q}^{(t)})$ for Lemma 5 and Lemma 6. Recall that, for any $P \in \mathcal{P}^{(t)}$ and any $P' \in \mathcal{P}^{(t)}$, $P \neq P'$,

- $\#_{\mathrm{F}}(P)$ denotes the number of forbidden edges clustered into P, and
- $\#_{\mathrm{NF}}(P, P')$ denotes the number of non-forbidden edges between P and P'.

Also recall that for any cluster $P \in \mathcal{P}^{(t)}$,

- $\Delta^{(t)}(P)$ denotes the gluer set of P, $\Delta_+^{(t)}(P) := P \cap \Delta^{(t)}(P)$ is referred to as the core of P,
- $\operatorname{Ext}^{(t)}(P) := P \setminus \Delta_+^{(t)}(P)$ denotes the extended part of P, and
- $L_1^{(t)}(P)$ denotes the set of elements in the 2/3-vicinity of $P' \cap \Delta_+^{(\ell(t,P))}(P)$ within P' over all $P' \in \operatorname{Candi}^{(\ell(t,P))}(\Delta^{(\ell(t,P))}(P))$, where $\ell(t, P)$ is the index of the top-most layer up to the t-th layer at which P is newly-created.

Sketch of Lemma 5 – Forbidden edges within any P: To illustrate the ideas, we prove a weaker bound of $\frac{(2-\alpha)(1+\alpha)^2}{2(1-\alpha)^2}$ for $\#_{\mathrm{F}}(P)$ in the following. For $\frac{\beta(2-\alpha)(1+\alpha)^2}{2(1-\alpha)^2}$ with $\beta := 0.8346$, we refer the readers to Section A in the appendix for the details.

Let $P \in \mathcal{P}^{(t)}$ be a cluster. Since $\operatorname{diam}^{(t)}(\Delta_+^{(t)}(P)) < 1/3$ and the distances are non-increasing bottom-up over the layers, forbidden edges only occur between $\operatorname{Ext}^{(t)}(P)$ and P, i.e., no forbidden edges reside within $\Delta_+^{(t)}(P)$. Hence, we have

$$\begin{aligned} \#_{\mathrm{F}}(P) \leq& |\operatorname{Ext}^{(t)}(P)| \cdot \left(\frac{|\operatorname{Ext}^{(t)}(P)|}{2} + |\Delta_+^{(t)}(P)| \right) \\ \leq& \frac{1}{1-\alpha} \cdot |L_1^{(t)}(P)| \cdot \frac{2-\alpha}{2(1-\alpha)} \cdot |\Delta_+^{(t)}(P)| \quad (5) \\ \leq& \frac{2-\alpha}{2(1-\alpha)^2} \cdot \alpha \cdot |\Delta_+^{(t)}(P)|^2, \quad (6) \end{aligned}$$

where in the last two inequalities we use the bounds from Lemma 9.

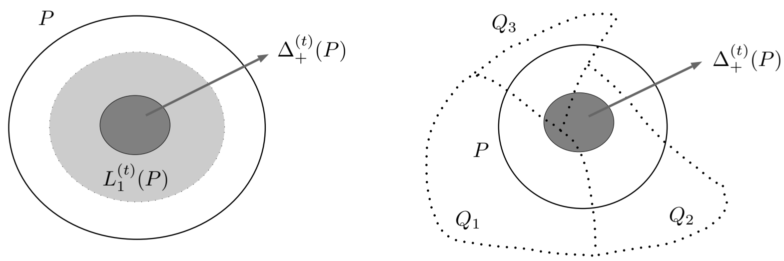


Fig. 6. Two cases for the cluster P. (a) P is newly-created at the t-th layer. In this case, the edges between $L_1^{(t)}(P)$ and $\Delta_+^{(t)}(P)$ must be non-forbidden and reside between different pre-clusters within P. (b) P is created at a layer lower than t. In this case, $P \notin \mathrm{Candi}^{(t)}(Q)$ for any pre-cluster Q that intersects $\Delta_+^{(t)}(P)$.

We have two cases to consider. If P is a newly-formed cluster at the t-th layer, then any edge between $\Delta_+^{(t)}(P)$ and $L_1^{(t)}(P)$ crosses different pre-clusters and is non-forbidden by the way $L_1^{(t)}(P)$ is defined. Hence, these edges are contained within $\mathrm{NF}(\mathcal{Q}^{(t)}, P)$ and we have $|L_1^{(t)}(P)| \cdot |\Delta_+^{(t)}(P)| \le \left|\mathrm{NF}(\mathcal{Q}^{(t)}, P)\right|$. It follows from (5) that

$$\#_{\mathrm{F}}(P) \;\le\; \frac{2-\alpha}{2(1-\alpha)^2} \cdot \left|\mathrm{NF}(\mathcal{Q}^{(t)}, P)\right|.$$

Second, if P is a previously-formed cluster at a lower layer, then consider the set of pre-clusters in $\mathcal{Q}^{(t)}$ that intersect the core set $\Delta_+^{(t)}(P)$. Let $Q_1, \ldots, Q_k$ denote these pre-clusters and assume W.L.O.G. that $|Q_1 \cap \Delta_+^{(t)}(P)| = \max_{1\le j\le k} |Q_j \cap \Delta_+^{(t)}(P)|$. Since $P \notin \mathrm{Candi}^{(t)}(Q_1)$, we have

$$\begin{aligned} B_1 \;:=&\; \left|\, \mathrm{Ball}^{(t)}_{<\frac{2}{3}}\left(\, P \cap Q_1,\; P \cap \overline{Q_1}\,\right) \right| \\ \ge&\; \alpha \cdot |P \cap Q_1|\,. \end{aligned} \tag{7}$$

We have two subcases to consider regarding the relative size of $|Q_j \cap \Delta_+^{(t)}(P)|$ for all j.

- If $\sum_{2\le j\le k} |Q_j \cap \Delta_+^{(t)}(P)| < \alpha \cdot |Q_1 \cap \Delta_+^{(t)}(P)|$, then

$$\begin{aligned} |\Delta_+^{(t)}(P)|^2 \;&\le\; (1+\alpha)^2 \cdot |Q_1 \cap \Delta_+^{(t)}(P)|^2 \\ &\le\; \frac{(1+\alpha)^2}{\alpha} \cdot |Q_1 \cap \Delta_+^{(t)}(P)| \cdot B_1 \end{aligned}$$

  by Condition (7). Since the edges between $Q_1 \cap \Delta_+^{(t)}(P)$ and $\mathrm{Ball}^{(t)}_{<\frac{2}{3}}\left(\, P \cap Q_1,\; P \cap \overline{Q_1}\,\right)$ are non-forbidden, reside within P, and cross different pre-clusters, they are contained within $\mathrm{NF}(\mathcal{Q}^{(t)}, P)$. By (6) we have

$$\#_{\mathrm{F}}(P) \;\le\; \frac{2-\alpha}{2(1-\alpha)^2} \cdot (1+\alpha)^2 \cdot \left|\mathrm{NF}(\mathcal{Q}^{(t)}, P)\right|.$$

- If $\sum_{2\le j\le k} |Q_j \cap \Delta_+^{(t)}(P)| \ge \alpha \cdot |Q_1 \cap \Delta_+^{(t)}(P)|$, since $\alpha \le 1/2$, it follows that $Q_1, \ldots, Q_k$ can be partitioned into two groups $\mathcal{G}_1$ and $\mathcal{G}_2$ such that[1]

$$\begin{aligned} \alpha \cdot \sum_{Q\in\mathcal{G}_1} |Q \cap \Delta_+^{(t)}(P)| \le& \sum_{Q\in\mathcal{G}_2} |Q \cap \Delta_+^{(t)}(P)| \\ \le& \sum_{Q\in\mathcal{G}_1} |Q \cap \Delta_+^{(t)}(P)|. \end{aligned}$$

  Define $G_1 := \sum_{Q\in\mathcal{G}_1} |Q \cap \Delta_+^{(t)}(P)|$ and $G_2 := \sum_{Q\in\mathcal{G}_2} |Q \cap \Delta_+^{(t)}(P)|$ for short. We have

$$\begin{aligned} |\Delta_+^{(t)}(P)|^2 \;&=\; (G_1+G_2)^2 \\ &=\; \left(\frac{G_1}{G_2} + \frac{G_2}{G_1} + 2\right) \cdot G_1 \cdot G_2 \\ &\le\; \left(\frac{1}{\alpha} + \alpha + 2\right) \cdot G_1 \cdot G_2 \\ &=\; \frac{(1+\alpha)^2}{\alpha} \cdot G_1 \cdot G_2, \end{aligned}$$

  where the last inequality follows since the function $f(x) = x + 1/x$ attains its maximum value at $x = \alpha$ within the interval $[\alpha, 1]$. Since the edges counted between G_1 and G_2 are contained within $\mathrm{NF}(\mathcal{Q}^{(t)}, P)$, again we have

$$\#_{\mathrm{F}}(P) \;\le\; \frac{2-\alpha}{2(1-\alpha)^2} \cdot (1+\alpha)^2 \cdot \left|\mathrm{NF}(\mathcal{Q}^{(t)}, P)\right|.$$

We provide the details for the improved bound $\frac{\beta(2-\alpha)(1+\alpha)^2}{2(1-\alpha)^2}$ with $\beta := 0.8346$ in Section A in the appendix for further reference.

Proof of Lemma 6 – Non-forbidden edges across P and P': This type of disagreements consists of two different types, namely, whether or not they reside within the same pre-cluster.

[1]One of such ways is to consider Q_j in non-ascending order of $|Q_j \cap \Delta_+^{(t)}(P)|$ for all $1 \le j \le k$, and assign each Q_j considered to the group that has a smaller intersection with $\Delta^{(t)}(P)$ in size.

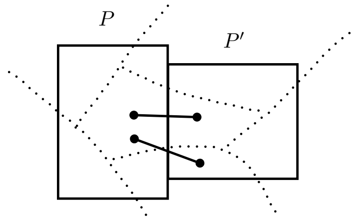


Fig. 7. Two types of non-forbidden edges across different parts P and P' – Whether or not they reside within the same pre-cluster.

First, the number of non-forbidden edges between P and P' that are not within the same pre-cluster is at most $\left|\mathrm{NF}(\mathcal{Q}^{(t)}, P, P')\right|$, where

$$\begin{aligned} &\mathrm{NF}(\mathcal{Q}^{(t)}, P, P') \\ &\quad := \{\{i,j\} \in \mathrm{NF}(\mathcal{Q}^{(t)}) \colon i \in P, j \in P'\} \end{aligned} \tag{8}$$

denotes the set of non-forbidden edges in $\mathrm{NF}(\mathcal{Q}^{(t)})$ that resides between P and P'.

Consider the set of non-forbidden edges that resides between P and P' and that belongs to the same Q for some $Q \in \mathcal{Q}^{(t)}$. To count the number of such edges, fix a $Q \in \mathcal{Q}^{(t)}$ with $P \cap Q \neq \emptyset$ and $P' \cap Q \neq \emptyset$. By the design of Algorithm 1, at most one of P and P' can be newly-created at this layer and have Q being its gluer set.

Without loss of generality, assume that P is either previously-formed or newly-created with a gluer set other than Q. In the following, for this $\{P, P'\}$ pair, we fix P and count the number of non-forbidden edges that reside in Q and that have with one end in P and the other end in P'. We have two cases to consider.

If $Q \cap \Delta_+^{(t)}(P) = \emptyset$, then there exists $K_P^{(t)}(P \cap Q) \subseteq \mathrm{Ball}_{<2/3}^{(t)}(P \cap Q, P \setminus Q)$ such that

$$|P \cap Q| \;\leq\; \frac{\alpha}{1-\alpha} \cdot |K_P^{(t)}(P \cap Q)|$$

by Lemma 9.

Hence,

$$|P \cap Q| \cdot |P' \cap Q| \;\leq\; \frac{\alpha}{1-\alpha} \cdot |K_P^{(t)}(P \cap Q)| \cdot |P' \cap Q|.$$

Note that, the edges counted in the right-hand-side above reside between P and P'. Each of them has one end in $P' \cap Q$ and the other end in $P \cap Q'$ for some other pre-cluster Q'. Moreover, they are non-forbidden. It follows that

$$\begin{aligned} &|P \cap Q| \cdot |P' \cap Q| \\ &\qquad \leq\; \frac{\alpha}{1-\alpha} \cdot \left|\mathrm{NF}(\mathcal{Q}^{(t)}, P, P', Q)\right|, \end{aligned} \tag{9}$$

where

$$\begin{aligned} &\mathrm{NF}(\mathcal{Q}^{(t)}, P, P', Q) \\ &:= \{\ \{i,j\} \in \mathrm{NF}(\mathcal{Q}^{(t)}) \;:\; i \in P \setminus Q,\ j \in P' \cap Q\ \} \end{aligned}$$

denotes the set of edges in $\mathrm{NF}(\mathcal{Q}^{(t)})$ that have one end in $P \setminus Q$ and the other in $P' \cap Q$.

For the second case, suppose that $Q \cap \Delta_+^{(t)}(P) \neq \emptyset$. It follows that P must be previously-formed. Furthermore, $P \notin \mathrm{Candi}^{(t)}(Q)$. Hence, we have

$$\begin{aligned} &|P \cap Q| \cdot |P' \cap Q| \\ &\leq \frac{1}{\alpha} \cdot \left| \mathrm{Ball}_{<\frac{2}{3}}^{(t)} \left(P \cap Q,\ P \cap \overline{Q} \right) \right| \cdot |P' \cap Q| \\ &\leq \frac{1}{\alpha} \cdot \left|\mathrm{NF}(\mathcal{Q}^{(t)}, P, P', Q)\right|. \end{aligned} \tag{10}$$

Combining (9) and (10), it follows that

$$\begin{aligned} &\#_{\mathrm{NF}}(P, P', Q) \;\leq \\ &\qquad \max\left\{ \frac{\alpha}{1-\alpha},\ \frac{1}{\alpha} \right\} \cdot \left|\mathrm{NF}(\mathcal{Q}^{(t)}, P, P', Q)\right|, \end{aligned}$$

where $\#_{\mathrm{NF}}(P, P', Q)$ denotes the number of non-forbidden edges that are between P and P' and that belong to Q. Summing up over all Q with $P \cap Q \neq \emptyset$ and $P' \cap Q \neq \emptyset$ and further taking (8) into account, we obtain

$$\begin{aligned} &\#_{\mathrm{NF}}(P, P') \\ &\leq \left(\max\left\{ \frac{\alpha}{1-\alpha}, \frac{1}{\alpha} \right\} + 1 \right) \cdot \left|\mathrm{NF}(\mathcal{Q}^{(t)}, P, P')\right|. \end{aligned} \tag{11}$$

C. Average Distance of Non-Forbidden Cut Edges

Consider the execution of Algorithm 2. Let x be the input distance function. Suppose that the algorithm picks a pair (v, Q) with $v \in Q \in \mathcal{Q}$ and $\max_{u \in Q} x_{\{u,v\}} \geq 1/3$ in some iteration and let (Q_1, Q_2) with $v \in Q_1$ be the pair returned by the procedure ONE-THIRD-REFINE-CUT.

Recall that we use $\mathrm{NF}(Q_1, Q_2)$ to denote the set of non-forbidden edges between Q_1 and Q_2. For the ease of notation define

$$B_{1/3} \;:=\; \mathrm{Ball}_{<1/3}^{(x)}(v, Q), \quad B_{1/2} \;:=\; \mathrm{Ball}_{<1/2}^{(x)}(v, Q), \quad \text{and} \quad Q_2' := Q_2 \cap B_{1/2}.$$

We prove the following lemma.

Lemma 10 (Restate of Lemma 7)**.**

$$\sum_{\substack{\{i,j\} \in \mathrm{NF}(Q_1, Q_2), \\ i \in Q_1,\ j \in Q_2'}} \left(\min\left\{ x_{\{v,j\}},\ \frac{1}{3} \right\} \;-\; x_{\{v,i\}} \right) \;\geq\; \frac{1}{6} \cdot \left| \left\{ \begin{smallmatrix} \{i,j\} \in \mathrm{NF}(Q_1, Q_2), \\ i \in Q_1,\ j \in Q_2' \end{smallmatrix} \right\} \right|.$$

Proof. For any $p, q \in B_{2/3}$, define $d(p,q) := |\min\{x_{\{v,p\}}, 1/3\} - \min\{x_{\{v,q\}}, 1/3\}| - 1/6$. Since $Q_1 \subseteq B_{1/3}$, to prove the statement of this lemma, it suffices to prove that

$$\sum_{\{p,q\} \in \mathrm{NF}(Q_1, Q_2')} d(p,q) \;\geq\; 0. \tag{12}$$

From the setting of the procedure ONE-THIRD-REFINE-CUT, we have

$$(Q_1, Q_2') \;\in\; \left\{ \begin{array}{l} \mathrm{Cut}_1 \;=\; \big(\, \{v\},\; B_{1/2} \setminus \{v\} \,\big)\,, \\ \mathrm{Cut}_2 \;=\; \big(\, B_{1/3},\; B_{1/2} \setminus B_{1/3} \,\big) \end{array} \right\}.$$

Hence, to prove (12), it suffices to prove that

$$W := \max_{1 \le i \le 2} \left\{ \sum_{\{p,q\} \in \mathrm{NF}(\mathrm{Cut}_i)} d(p,q) \right\} \geq 0. \tag{13}$$

In the following we prove (13).

Let $k := |B_{1/3}|$ and $m := |B_{1/2} \setminus B_{1/3}|$. For Cut_1, we have

$$\begin{aligned} &\sum_{\{p,q\} \in \mathrm{NF}(\mathrm{Cut}_1)} d(p,q) \\ &= \sum_{q \in B_{1/3}} x_{\{v,q\}} + \frac{1}{3} \cdot |B_{1/2} \setminus B_{1/3}| - \frac{1}{6} \cdot \big(|B_{1/2}| - 1\big) \\ &= \sum_{q \in B_{1/3}} x_{\{v,q\}} + \frac{1}{6} \cdot |B_{1/2}| - \frac{1}{3} \cdot |B_{1/3}| + \frac{1}{6} \qquad (14) \\ &= \sum_{q \in B_{1/3}} x_{\{v,q\}} + \frac{1}{6} \cdot (m - k + 1). \qquad (15) \end{aligned}$$

Note that the nonnegativity of (14) is exactly tested by the procedure ONE-THIRD-REFINE-CUT.

For Cut_2, observe that any $p \in B_{1/3}$ and $q \in B_{1/2}$ always forms a non-forbidden edge. By a similar argument to the above, we have

$$\begin{aligned} &\sum_{\{p,q\} \in \mathrm{NF}(\mathrm{Cut}_2)} d(p,q) \\ &\quad= \; \frac{1}{6} \cdot |\,\mathrm{NF}(\mathrm{Cut}_2)| \;-\; m \cdot \sum_{q \in B_{1/3}} x_{\{v,q\}}. \end{aligned} \tag{16}$$

From the definition of W in (13) combined with (15) and (16), we obtain

$$\begin{aligned} W \;\geq\; & \frac{m}{m+1} \cdot \sum_{\{p,q\} \in \mathrm{NF}(\mathrm{Cut}_1)} d(p,q) \\ & + \; \frac{1}{m+1} \cdot \sum_{\{p,q\} \in \mathrm{NF}(\mathrm{Cut}_2)} d(p,q) \\ = \; & \frac{m}{m+1} \cdot \left(\sum_{q \in B_{1/3}} x_{\{v,q\}} \;+\; \frac{1}{6} \cdot (m - k + 1) \right) \\ & + \; \frac{m}{m+1} \cdot \left(\frac{1}{6} \cdot k - \sum_{q \in B_{1/3}} x_{\{v,q\}} \right) \\ = \; & \frac{m}{6(m+1)} \cdot (m+1) \;\geq\; 0, \end{aligned}$$

where in the second last equality we use the fact that $|\,\mathrm{NF}(\mathrm{Cut}_2)| = k \cdot m$. □

Recall that we define $Q_2' := Q_2 \cap B_{1/2}$. The following corollary, which is obtained by taking into accounts the edges $\{i,j\}$ with $i \in Q_1$, $j \in Q_2 \setminus Q_2'$, summarizes the guarantee for the average distance of non-forbidden cut edges.

Corollary 11.

$$\begin{aligned} & \frac{1}{6} \cdot |\mathrm{NF}(Q_1, Q_2)| \\ & \leq \sum_{\substack{\{i,j\} \in \mathrm{NF}(Q_1,Q_2), \\ i \in Q_1, j \in Q_2'}} \left(\min\left\{ x_{\{v,j\}}, \frac{1}{3} \right\} - x_{\{v,i\}} \right) \\ & \qquad + \sum_{\substack{\{i,j\} \in \mathrm{NF}(Q_1,Q_2), \\ i \in Q_1,\, j \in Q_2 \setminus Q_2', \\ \{i,j\} \in E_+^{(t)}}} x_{\{i,j\}} \\ & \qquad + \frac{1}{6} \cdot \left| \left\{ \begin{array}{c} \{i,j\} \in \mathrm{NF}(Q_1,Q_2), \\ i \in Q_1,\, j \in Q_2 \setminus Q_2', \\ \{i,j\} \in E_-^{(t)} \end{array} \right\} \right|. \end{aligned}$$

Proof. Observe that, for any $i \in Q_1$, $j \in Q_2 \setminus Q_2'$, we have $x_{\{i,j\}} \geq x_{\{v,j\}} - x_{\{v,i\}} \geq 1/6$. □

The following lemma relates the number of non-forbidden edges separated by Q_1 and Q_2 to the objective value of the input distance function in terms of the original input instance $(E_+^{(t)}, E_-^{(t)})$.

Lemma 12.

$$\begin{aligned} \sum_{\substack{\{i,j\} \in \mathrm{NF}(Q_1,Q_2), \\ \{i,j\} \in E_+^{(t)}}} x_{\{i,j\}} \;+\; & \sum_{\substack{\{i,j\} \in \mathrm{NF}(Q_1,Q_2), \\ \{i,j\} \in E_-^{(t)}}} \big(1 - x_{\{i,j\}}\big) \\ & \geq\; \frac{1}{6} \cdot \left| \left\{ \begin{array}{c} \{i,j\} \in \mathrm{NF}(Q_1,Q_2), \\ \{i,j\} \in E_+^{(t)} \end{array} \right\} \right|. \end{aligned}$$

Proof. Recall that $Q_2' := Q_2 \cap B_{1/2}$. First, we prove that

$$\sum_{\substack{\{i,j\}\in \mathrm{NF}(Q_1,Q_2),\\ \{i,j\}\in E_+^{(t)}}} x_{\{i,j\}} + \sum_{\substack{\{i,j\}\in \mathrm{NF}(Q_1,Q_2),\\ \{i,j\}\in E_-^{(t)}}} \left(1 - x_{\{i,j\}}\right) + \frac{1}{6}\cdot\left|\left\{ \substack{\{i,j\}\in \mathrm{NF}(Q_1,Q_2'),\\ \{i,j\}\in E_-^{(t)}} \right\}\right|$$
$$\geq \sum_{\substack{\{i,j\}\in \mathrm{NF}(Q_1,Q_2),\\ i\in Q_1,\ j\in Q_2'}} \left(\min\left\{ x_{\{v,j\}},\ \frac{1}{3} \right\} - x_{\{v,i\}} \right) + \sum_{\substack{\{i,j\}\in \mathrm{NF}(Q_1,Q_2),\\ i\in Q_1,\ j\in Q_2\setminus Q_2',\\ \{i,j\}\in E_+^{(t)}}} x_{\{i,j\}}. \tag{17}$$

To prove (17), consider any $\{i,j\} \in \mathrm{NF}(Q_1,Q_2)$ with $i \in Q_1$.

1) If $\{i,j\}$ is a $+$ edge in $E_+^{(t)}$, then using the triangle inequality we have $x_{\{i,j\}} \geq x_{\{v,j\}} - x_{\{v,i\}}$ and hence $x_{\{i,j\}} \geq \min\left\{x_{\{v,j\}},\ \frac{1}{3}\right\} - x_{\{v,i\}}$.
2) If $\{i,j\}$ is a $-$ edge in $E_-^{(t)}$ with $j \in Q_2'$, then applying the setting and the triangle inequality we have $x_{\{i,j\}} \leq x_{\{v,i\}} + x_{\{v,j\}} \leq 5/6$, and hence

$$\left(1 - x_{\{i,j\}}\right) + \frac{1}{6} \geq \frac{1}{3} \geq \min\left\{ x_{\{v,j\}},\ \frac{1}{3} \right\} - x_{\{v,i\}}.$$

The above compares the left-hand side of (17) with its right-hand side for all cases. Hence, we have (17). Adding $\frac{1}{6}\cdot\left|\left\{ \substack{\{i,j\}\in \mathrm{NF}(Q_1,Q_2),\\ i\in Q_1,\ j\in Q_2\setminus Q_2',\\ \{i,j\}\in E_-^{(t)}} \right\}\right|$ to both sides of (17) and applying Corollary 11, it follows that

$$\sum_{\substack{\{i,j\}\in \mathrm{NF}(Q_1,Q_2),\\ \{i,j\}\in E_+^{(t)}}} x_{\{i,j\}} + \sum_{\substack{\{i,j\}\in \mathrm{NF}(Q_1,Q_2),\\ \{i,j\}\in E_-^{(t)}}} \left(1 - x_{\{i,j\}}\right) + \frac{1}{6}\cdot\left|\left\{ \substack{\{i,j\}\in \mathrm{NF}(Q_1,Q_2),\\ \{i,j\}\in E_-^{(t)}} \right\}\right| \geq \frac{1}{6}\cdot\left|\mathrm{NF}(Q_1,Q_2)\right|,$$

and this lemma follows. □

Since Lemma 12 holds for every (Q_1,Q_2) output by the procedure ONE-THIRD-REFINE-CUT, we have the following corollary for the pre-cluster $\mathcal{Q}$ output by Algorithm 2.

Corollary 13.

$$\sum_{\substack{\{i,j\}\in \mathrm{NF}(\mathcal{Q}^{(t)}),\\ \{i,j\}\in E_+^{(t)}}} \tilde{x}^{(t)}_{\{i,j\}} + \sum_{\substack{\{i,j\}\in \mathrm{NF}(\mathcal{Q}^{(t)}),\\ \{i,j\}\in E_-^{(t)}}} \left(1 - \tilde{x}^{(t)}_{\{i,j\}}\right) + \frac{1}{6}\cdot\left|\left\{ \substack{\{i,j\}\in \mathrm{NF}(\mathcal{Q}^{(t)}),\\ \{i,j\}\in E_-^{(t)}} \right\}\right| \geq \frac{1}{6}\cdot\left|\mathrm{NF}(\mathcal{Q}^{(t)})\right|.$$

D. Relating the Objectives

We prove the following key technical lemma regarding the weighted cardinality of non-forbidden $-$ edges over the layers in any optimal LP-solution.

Lemma 14 (Restate of Lemma 4)**.**

$$\sum_{1\leq t\leq \ell} \delta_t \cdot \left(\sum_{\{u,v\}\in E^{(t)}} \tilde{x}^{(t)}_{u,v} + \sum_{\{u,v\}\in E_-^{(t)}} \left(1 - \tilde{x}^{(t)}_{u,v}\right) \right) \geq \sum_{1\leq t\leq \ell} \delta_t \cdot \left| NF_-^{(t)} \right|.$$

where we use $NF_-^{(t)} := E_-^{(t)} \setminus F^{(t)}$ to denote the set of non-forbidden $-$ edges at the t-th layer.

Fix an optimal solution $\tilde{x}$ to LP-($*$). In the following we modify the constraints in LP-($*$) step by step, while keeping the invariant that $\tilde{x}$ remains an optimal solution to the LP.

For each variable $x^{(t)}_{\{u,v\}}$ such that $\tilde{x}^{(t)}_{\{u,v\}} = 1$, replace all occurrences of $x^{(t)}_{\{u,v\}}$ with the constant 1. Note that the restriction of $\tilde{x}$ to the surviving variables continues to be an optimal solution to the LP after this modification.

There are three types of constraints in LP-($*$) other than the nonnegativity constraints, namely, $x^{(t)}_{\{u,p\}} + x^{(t)}_{\{p,v\}} \geq x^{(t)}_{\{u,v\}}$, $x^{(t)}_{\{u,v\}} \geq x^{(t+1)}_{\{u,v\}}$, and $x^{(t)}_{\{u,v\}} \leq 1$. We further modify the LP to remove *redundant* constraints, which we describe in the following.

- For each $1 \leq t \leq \ell$ and $u,v,p \in V$, remove the constraint $x^{(t)}_{\{u,p\}} + x^{(t)}_{\{p,v\}} \geq x^{(t)}_{\{u,v\}}$ if at least one variable on the left-hand side was replaced with 1.
- For each $1 \leq t \leq \ell$ and $u,v \in V$, remove the constraint $x^{(t)}_{\{u,v\}} \geq x^{(t+1)}_{\{u,v\}}$ if at least one variable was replaced with 1.
- For each $1 \leq t \leq \ell$ and $u,v \in V$, remove $x^{(t)}_{\{u,v\}} \leq 1$ if $x^{(t)}_{\{u,v\}}$ was replaced with 1.

Let $\mathrm{SV}^{(t)} := \{\{u,v\} \colon \tilde{x}^{(t)}_{\{u,v\}} < 1\}$ be the set of variables that survived in layer t, $\tilde{x}^*$ be the restriction of $\tilde{x}$ to $\{\mathrm{SV}^{(t)}\}_t$, and LP-($**$) be the LP obtained by the above procedure.

$$\min \sum_{1\le t\le \ell} \delta_t \cdot \left(\left|E_+^{(t)} \setminus \mathrm{SV}^{(t)}\right| + \sum_{\substack{\{u,v\}\in E_+^{(t)},\\ \{u,v\}\in \mathrm{SV}^{(t)}}} x_{\{u,v\}}^{(t)} + \sum_{\substack{\{u,v\}\in E_-^{(t)},\\ \{u,v\}\notin F^{(t)}}} (1 - x_{\{u,v\}}^{(t)}) \right) \qquad \text{LP-}(**)$$

$$\begin{aligned}
\text{s.t.}\quad & x_{\{u,p\}}^{(t)} + x_{\{p,v\}}^{(t)} \ge x_{\{u,v\}}^{(t)}, && \forall\, 1\le t\le \ell,\ \{u,p\},\{p,v\},\{u,v\}\in \mathrm{SV}^{(t)},\\
& x_{\{u,p\}}^{(t)} + x_{\{p,v\}}^{(t)} \ge 1, && \forall\, 1\le t\le \ell,\ \{u,p\},\{p,v\}\in \mathrm{SV}^{(t)},\ \{u,v\}\notin \mathrm{SV}^{(t)},\\
& x_{\{u,v\}}^{(t+1)} \le x_{\{u,v\}}^{(t)}, && \forall\, 1\le t< \ell,\ \{u,v\}\in \mathrm{SV}^{(t)}\cap \mathrm{SV}^{(t+1)},\\
& 0 \le x_{\{u,v\}}^{(t)} \le 1, && \forall\, 1\le t\le \ell,\ \{u,v\}\in \mathrm{SV}^{(t)}.
\end{aligned}$$

We have the following lemma.

Lemma 15. $\tilde{x}^*$ is an optimal solution for LP-$(**)$.

Proof. We claim that removing the above constraints does not change the set of feasible solutions, and hence $\tilde{x}^*$ remains an optimal solution to the resulting LP.

Consider the first type of constraints. The removed constraints are in the form of $1+x_{\{p,v\}}^{(t)} \ge x_{\{u,v\}}^{(t)}$, $1+1 \ge x_{\{u,v\}}^{(t)}$, $1 + x_{\{p,v\}}^{(t)} \ge 1$, or $1+1 \ge 1$.

- For $1 + x_{\{p,v\}}^{(t)} \ge x_{\{u,v\}}^{(t)}$, where $\{p,v\},\{u,v\} \in \mathrm{SV}^{(t)}$, the removed constraint is implied by $x_{\{p,v\}}^{(t)} \ge 0$ and $x_{\{u,v\}}^{(t)} \le 1$, which are constraints that still exist in LP-$(**)$.
- For $1 + 1 \ge x_{\{u,v\}}^{(t)}$, where $\{u,v\} \in \mathrm{SV}^{(t)}$, the removed constraint is implied by $x_{\{u,v\}}^{(t)} \le 1$, a constraint still existing in LP-$(**)$.
- For $1 + x_{\{p,v\}}^{(t)} \ge 1$ with $\{p,v\} \in \mathrm{SV}^{(t)}$, again it is implied by $x_{\{p,v\}}^{(t)} \ge 0$, which exists in LP-$(**)$.

Consider the second type of constraints, i.e., $x_{\{u,v\}}^{(t)} \ge x_{\{u,v\}}^{(t+1)}$. If $x_{\{u,v\}}^{(t+1)}$ was replaced with 1, then $\tilde{x}_{\{u,v\}}^{(t)} = 1$ and $\{u,v\} \notin \mathrm{SV}^{(t)}$. If only $x_{\{u,v\}}^{(t)}$ was replaced, then $1 \ge x_{\{u,v\}}^{(t+1)}$ is a constraint that persists in LP-$(**)$. Finally, for the third type of constraints, $x_{\{u,v\}}^{(t)} \le 1$, the claimed statement is trivial. This proves the lemma. ☐

Let us now consider the dual of LP-$(**)$, which has an objective function of the following form

$$\max \sum_{1\le t\le \ell} \delta_t \cdot \Bigg(\left|E_+^{(t)} \setminus \mathrm{SV}^{(t)}\right| + \left|NF_-^{(t)}\right| + \sum_{\substack{\{u,v\}\notin \mathrm{SV}^{(t)}\\ p:\{u,p\},\{p,v\}\in \mathrm{SV}^{(t)}}} \beta_{\{u,v\},p}^{(t)} - \sum_{\{u,v\}\in \mathrm{SV}^{(t)}} \eta_{\{u,v\}}^{(t)} \Bigg),$$

where $\{\beta_{\{u,v\},p}^{(t)}\}_{1\le t\le \ell,\ \substack{\{u,v\}\notin \mathrm{SV}^{(t)}\\ p:\{u,p\},\{p,v\}\in \mathrm{SV}^{(t)}}}$ and $\{\eta_{\{u,v\}}^{(t)}\}_{1\le t\le \ell,\ \{u,v\}\in \mathrm{SV}^{(t)}}$ are non-negative dual variables for the second set and the last set of constraints in LP-$(**)$, respectively.

Since $\tilde{x}_{\{u,v\}}^{*(t)} < 1$ for any $1 \le t \le \ell$ and $\{u,v\} \in \mathrm{SV}^{(t)}$, the complementary slackness condition states that in any optimal dual solution with $\tilde{y}^*$ which contains η^* as dual variables for the last set of constraints in LP-$(**)$, we always have that

$$\eta_{\{u,v\}}^{*(t)} = 0 \quad \text{for any } 1 \le t \le \ell \text{ and } \{u,v\} \in \mathrm{SV}^{(t)}.$$

This implies that $\sum_{1\le t\le \ell} \delta_t \cdot |NF_-^{(t)}| \le \mathrm{Val}(\tilde{y}^*) = \mathrm{Val}(\tilde{x}^*)$, where $\mathrm{Val}(\tilde{x}^*)$ and $\mathrm{Val}(\tilde{y}^*)$ denote the objective value of $\tilde{x}^*$ and $\tilde{y}^*$, and it follows that

$$\sum_{1\le t\le \ell} \delta_t \cdot \left|NF_-^{(t)}\right| \le \sum_{1\le t\le \ell} \delta_t \cdot \left(\sum_{\{u,v\}\in E_+^{(t)}} \tilde{x}_{\{u,v\}}^{(t)} + \sum_{\{u,v\}\in E_-^{(t)}} \left(1 - \tilde{x}_{\{u,v\}}^{(t)}\right) \right). \qquad (18)$$

E. Putting Things Together

Now we are ready to prove Lemma 8. Consider the statements of Lemma 5 and Lemma 6. By the definition of $\mathrm{NF}(\mathcal{Q}^{(t)}, P)$ and $\mathrm{NF}(\mathcal{Q}^{(t)}, P, P')$, we have that

$$\mathrm{NF}(\mathcal{Q}^{(t)}) = \bigcup_{P\in \mathcal{P}^{(t)}} \mathrm{NF}(\mathcal{Q}^{(t)}, P) \cup \bigcup_{P,P'\in \mathcal{P}^{(t)},\ P\ne P'} \mathrm{NF}(\mathcal{Q}^{(t)}, P, P').$$

Hence, the two lemmas give that

$$\sum_{P \in \mathcal{P}^{(t)}} \#_{\mathrm{F}}(P) \; + \sum_{\substack{P,P' \in \mathcal{P}^{(t)} \\ P \neq P'}} \#_{\mathrm{NF}}(P,P') \le \; c(\alpha) \cdot \left|\mathrm{NF}(\mathcal{Q}^{(t)})\right|, \tag{19}$$

where $c(\alpha) := \max\left\{\frac{\beta(2-\alpha)(1+\alpha)^2}{2(1-\alpha)^2}, \frac{1}{1-\alpha}, \frac{1+\alpha}{\alpha}\right\} \approx 3.5406$ for $\alpha := 0.3936$, and $\beta := 0.8346$.

Applying Inequality (19) on Inequality (3), we have

$$\begin{aligned} &\#(\mathcal{P}^{(t)}) \\ &\le \sum_{P \in \mathcal{P}^{(t)}} \#_{NF_-}(P) + \sum_{P \in \mathcal{P}^{(t)}} \#_{\mathrm{F}}(P) \\ &\qquad + \sum_{\substack{P,P' \in \mathcal{P}^{(t)} \\ P \neq P'}} \#_{\mathrm{NF}}(P,P') \\ &\le \; |NF_-^{(t)}| \; + \; c(\alpha) \cdot \left|\mathrm{NF}(\mathcal{Q}^{(t)})\right|, \end{aligned} \tag{20}$$

where we use the fact that $\#_{NF_-}(P)$ is the number of non-forbidden $-$ edges clustered within P. By Corollary 13, the R.H.S. of (20) is upper-bounded by

$$\begin{aligned} &|NF_-^{(t)}| \\ &+ \; c(\alpha) \cdot \Bigg(\; |NF_-^{(t)}| + \\ &6 \cdot \Bigg(\sum_{\substack{\{i,j\} \in \mathrm{NF}(\mathcal{Q}^{(t)}), \\ \{i,j\} \in E_+^{(t)}}} \tilde{x}^{(t)}_{\{i,j\}} + \sum_{\substack{\{i,j\} \in \mathrm{NF}(\mathcal{Q}^{(t)}), \\ \{i,j\} \in E_-^{(t)}}} \left(1 - \tilde{x}^{(t)}_{\{i,j\}}\right) \Bigg) \Bigg). \end{aligned}$$

Summing up the weighted disagreements over all layers t with $1 \le t \le \ell$ and apply (18), we obtain

$$\begin{aligned} &\sum_{1 \le t \le \ell} \delta_t \cdot \#(\mathcal{P}^{(t)}) \; \le \; (\; 7c(\alpha) + 1 \;) \cdot \\ &\sum_{1 \le t \le \ell} \delta_t \cdot \Bigg(\sum_{\{u,v\} \in E_+^{(t)}} \tilde{x}^{(t)}_{\{u,v\}} + \sum_{\{u,v\} \notin E_+^{(t)}} \left(1 - \tilde{x}^{(t)}_{\{u,v\}}\right) \Bigg). \end{aligned}$$

VI. Extension to Ultrametric Violation Distance

Recall that given a set of pairwise measured distance for a set V of elements, the goal of the ultrametric violation distance problem is to edit the minimum number of input distances so that there is a perfect fit to an ultrametric. In [11], [12] the following formulation is introduced for this problem, where $t_{\{u,v\}}$ denotes the supposed layer at which u and v are separated in the ultrametric when a perfect fit for the given distances exists.

$$\begin{aligned} \min \sum_{\substack{u,v \in V, \\ u \neq v}} & \left(\left(1 - x_{\{u,v\}}^{(t_{\{u,v\}})}\right) + x_{\{u,v\}}^{(t_{\{u,v\}}+1)} \right) \quad \text{LP-}(L_0) \\ \text{s.t.} \quad & x^{(t)}_{\{u,v\}} \; \le \; x^{(t)}_{\{u,p\}} \; + \; x^{(t)}_{\{p,v\}}, \\ & \qquad\qquad \forall\, 1 \le t \le \ell,\; u,v,p \in V, \\ & 0 \; \le \; x^{(t+1)}_{\{u,v\}} \; \le \; x^{(t)}_{\{u,v\}} \; \le \; 1, \\ & \qquad\qquad \forall\, 1 \le t < \ell,\; u,v \in V. \end{aligned}$$

Fig. 8. LP formulation for the Ultrametric Violation Distance.

As for the LP-($*$) for the hierarchical correlation clustering problem, we implicitly assume in the following that $x^{(t)}_{\{u,u\}} = 0$ for all $u \in V$ holds in any feasible solution x for LP-(L_0). Furthermore, we extend the definition such that $x^{(\ell+1)}_{\{u,v\}} := 0$ for any $u,v \in V$.

Let $\tilde{x}$ be an optimal solution to LP-(L_0). The algorithm begins with a big cluster $\mathcal{P}^{(\ell+1)} := \{V\}$ and proceeds in a top-down manner. For each iteration t with $t = \ell, \ldots, 1$, the algorithm uses $\mathcal{P}^{(t)} := \mathcal{P}^{(t+1)}$ as the initial clustering and repeats until $\mathrm{diam}^{(t)}(P) < 1/2$ holds for all $P \in \mathcal{P}^{(t)}$. If $P \in \mathcal{P}^{(t)}$ contains an edge (u,v) with distance at least $1/2$, then the cutting procedure One-Half-Refine-Cut is applied to separate this edge.

The procedure One-Half-Refine-Cut takes as input a tuple (P,v,x), where P is a set, $v \in P$ is the pivot, and x is a distance function, and tests the following condition. If

$$\sum_{q \in \mathrm{Ball}^{(x)}_{<1/2}(v,P)} x_{\{v,q\}} \ge \frac{1}{2} \cdot |\,\mathrm{Ball}^{(x)}_{<1/2}(v,P)| - \frac{1}{4} \cdot |\,\mathrm{Ball}^{(x)}_{<3/4}(v,P)| - \frac{1}{4}, \tag{21}$$

then the algorithm makes v a singleton cluster by replacing P with $\{v\}$ and $P \setminus \{v\}$. Otherwise, P is replaced with $\mathrm{Ball}^{(x)}_{<1/2}(v,P)$ and $P \setminus \mathrm{Ball}^{(x)}_{<1/2}(v,P)$.

Approximation Guarantee

We prove the following theorem for the statement in Corollary 2.

Theorem 16. Let $\{\mathcal{P}^{(t)}\}_{1 \le t \le \ell}$ be the output of Algorithm 3 and $\hat{x}$ be the rounded integer distance function

Algorithm 3 Ultrametric-Violation-Distance($\{\tilde{x}\}_{1\le t\le \ell}$)

Let $\mathcal{P}^{(\ell+1)} \leftarrow \{V\}$.
for $t = \ell$ down to 1 **do**
  Let $\mathcal{P}^{(t)} \leftarrow \mathcal{P}^{(t+1)}$.
  while $\mathrm{diam}^{(t)}(P) \ge 1/2$ for some $P \in \mathcal{P}^{(t)}$ **do**
    Pick $P \in \mathcal{P}^{(t)}$ and $v \in P$ such that $\max_{u\in P} \tilde{x}^{(t)}_{\{u,v\}} \ge 1/2$.
    $\mathcal{P}' \leftarrow$ ONE-HALF-REFINE-CUT$(P, v, \tilde{x}^{(t)})$. `// Compute a refined cut`
    Replace P with the sets in $\mathcal{P}'$ in $\mathcal{P}^{(t)}$.
  end while
end for
return $\{\mathcal{P}^{(t)}\}_{1\le t\le \ell}$.

procedure ONE-HALF-REFINE-CUT(P, v, x)
  if Condition (21) is satisfied for (P, v) **then**
    return $\{\ \{v\},\ P \setminus \{v\}\ \}$. `// make` v `a singleton`
  else
    return $\{\ \mathrm{Ball}^{(x)}_{<1/2}(v, P),\ \ P \setminus \mathrm{Ball}^{(x)}_{<1/2}(v, P)\ \}$. `// cut at` $1/2 - \epsilon$
  end if
end procedure

to which $\mathcal{P}^{(t)}$ corresponds. We have

$$\sum_{\substack{u,v\in V,\\ u\neq v}} \left(\left(1 - \hat{x}^{(t_{\{u,v\}})}_{\{u,v\}} \right) + \hat{x}^{(t_{\{u,v\}}+1)}_{\{u,v\}} \right) \le 5 \cdot \sum_{\substack{u,v\in V,\\ u\neq v}} \left(\left(1 - \tilde{x}^{(t_{\{u,v\}})}_{\{u,v\}} \right) + \tilde{x}^{(t_{\{u,v\}}+1)}_{\{u,v\}} \right),$$

where $\tilde{x}$ is an optimal solution to LP-(L_0).

Observe that each edge $\{u, v\}$ contributes exactly two items in the objective value of $\tilde{x}$, namely, $\tilde{x}^{(t_{\{u,v\}}+1)}_{\{u,v\}}$ and $1 - \tilde{x}^{(t_{\{u,v\}})}_{\{u,v\}}$. We consider $\{u, v\}$ a $+$ edge for all the layers above $t_{\{u,v\}}$ and a $-$ edge for the remaining layers. In this regard, define

$$E^{(t)}_{+} := \{\ \{u, v\} : t_{\{u,v\}} < t\ \} \quad \text{and} \quad E^{(t)}_{-} := \{\ \{u, v\} : t_{\{u,v\}} \ge t\ \}.$$

Define $F := \left\{ \{u, v\} : \tilde{x}^{(t_{\{u,v\}})}_{\{u,v\}} = 1 \right\}$ to be the set of *forbidden edges* and $NF := \binom{V}{2} \setminus F$.

With exactly the same argument as in Lemma 4 (proved in Section V-D), we have the following updated version of statement for LP-(L_0).

Lemma 17. Let $\tilde{x}$ be an optimal solution to LP-(L_0). We have

$$\sum_{\substack{u,v\in V,\\ u\neq v}} \left(\left(1 - \tilde{x}^{(t_{\{u,v\}})}_{\{u,v\}} \right) + \tilde{x}^{(t_{\{u,v\}}+1)}_{\{u,v\}} \right) \ge |\ NF\ |.$$

Consider the execution of Algorithm 3 and the calls the algorithm makes to the procedure ONE-HALF-REFINE-CUT. Let k be the number of calls to the procedure and $\{(Q^{(i)}_1, Q^{(i)}_2)\}_{1\le i\le k}$ be the set of pairs returned by the procedure upon these calls. For each $(Q^{(i)}_1, Q^{(i)}_2)$, define

$$\mathrm{NExtm}\left(Q^{(i)}_1, Q^{(i)}_2\right) := \left\{ \{u, v\} : u \in Q^{(i)}_1,\ v \in Q^{(i)}_2,\ x^{(t_i)}_{\{u,v\}} < 1 \right\}$$

to be the set of edges that are separated by $Q^{(i)}_1$ and $Q^{(i)}_2$ and that have distances strictly smaller than 1 at the t_i-th layer, where we use t_i to denote the layer at which $(Q^{(i)}_1, Q^{(i)}_2)$ is separated. We will refer these edges to as *non-extreme cut edges*.

For any edge $\{u, v\}$, define

$$\#_{\{u,v\}} := \left(1 - \hat{x}^{(t_{\{u,v\}})}_{\{u,v\}} \right) + \hat{x}^{(t_{\{u,v\}}+1)}_{\{u,v\}} \quad \text{and}$$
$$\mathrm{Val}_{\{u,v\}} := \left(1 - \tilde{x}^{(t_{\{u,v\}})}_{\{u,v\}} \right) + \tilde{x}^{(t_{\{u,v\}}+1)}_{\{u,v\}}$$

to be the disagreement caused by the edge $\{u, v\}$ and the LP value the edge $\{u, v\}$ has, respectively. To upper-bound $\#_{\{u,v\}}$, let $t^*_{\{u,v\}}$ be the top-most layer at which $\{u, v\}$ is separated for the first time in the hierarchy. Define $t^*_{\{u,v\}}$ to be zero if $\{u, v\}$ is never separated in the hierarchy.

It is clear that $\#_{\{u,v\}} = 1$ only when $t^*_{\{u,v\}} \neq t_{\{u,v\}}$. Consider the following two cases.

- $t^*_{\{u,v\}} = 0$.
  In this case, $\{u, v\}$ is never separated. Then it follows from the design of Algorithm 3 that $\tilde{x}^{(t_{\{u,v\}})}_{\{u,v\}} < \frac{1}{2}$. Hence,

$$\#_{\{u,v\}} \le 2 \cdot \left(1 - \tilde{x}^{(t_{\{u,v\}})}_{\{u,v\}} \right) \le 2 \cdot \mathrm{Val}_{\{u,v\}}.$$

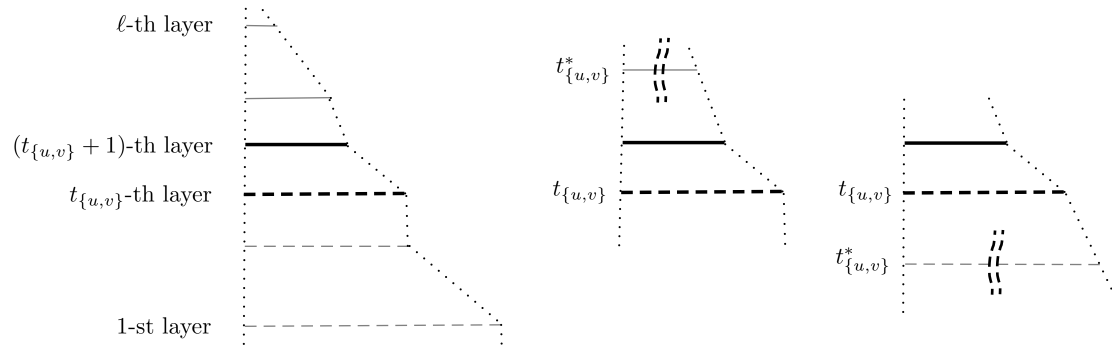


Fig. 9. (a) We consider $\{u,v\}$ a $+$ edge for all the layers above the $t_{\{u,v\}}$-th layer and a $-$ edge for the remaining layers. Moreover, $\{u,v\}$ contributes to the objective value only at the $(t_{\{u,v\}}+1)$-th and the $t_{\{u,v\}}$-th layers. (b) Two types of disagreements for the $\{u,v\}$ edge, namely, $t^*_{\{u,v\}} > t_{\{u,v\}}$ or $t^*_{\{u,v\}} < t_{\{u,v\}}$.

- $0 < t^*_{\{u,v\}} \neq t_{\{u,v\}}$.

  In this case, $\{u,v\}$ is separated by exactly one pair in $\mathcal{Q}$ and this happens at the $t^*_{\{u,v\}}$-th layer in the hierarchy. Denote this particular pair by $(Q_1^{(i)}, Q_2^{(i)})$.

  Further consider the following subcases.

  – If $\{u,v\} \notin \mathrm{NExtm}(Q_1^{(i)}, Q_2^{(i)})$ and $t^*_{\{u,v\}} > t_{\{u,v\}}$, then $\tilde{x}_{\{u,v\}}^{(t^*_{\{u,v\}})} = 1$, which implies that $\tilde{x}_{\{u,v\}}^{(t)} = 1$ for all $t \le t^*_{\{u,v\}}$, and hence

  $$\#_{\{u,v\}} \;=\; 1 \;=\; \tilde{x}_{\{u,v\}}^{(t_{\{u,v\}}+1)} \;=\; \mathrm{Val}_{\{u,v\}}\,.$$

  – If $\{u,v\} \notin \mathrm{NExtm}(Q_1^{(i)}, Q_2^{(i)})$ and $t^*_{\{u,v\}} < t_{\{u,v\}}$, then it follows from the design of Algorithm 3 that $\tilde{x}_{\{u,v\}}^{(t_{\{u,v\}})} < \frac{1}{2}$. Hence again

  $$\#_{\{u,v\}} \;\le\; 2\cdot\Big(1 - \tilde{x}_{\{u,v\}}^{(t_{\{u,v\}})}\Big) \;\le\; 2\cdot\mathrm{Val}_{\{u,v\}}\,.$$

From the above three cases, we obtain that

$$\begin{aligned}
&\sum_{u\neq v} \#_{\{u,v\}} \\
&= \Big|\Big\{\, \{u,v\} \,:\, 0 < t^*_{\{u,v\}} \neq t_{\{u,v\}} \,\Big\}\Big| \\
&\quad + \Big|\Big\{\, \{u,v\} \,:\, t^*_{\{u,v\}} = 0 \,\Big\}\Big| \\
&\le \sum_{1\le i\le k} \Big|\mathrm{NExtm}(Q_1^{(i)}, Q_2^{(i)})\Big| \\
&\quad + \; 2\cdot \sum_{\substack{\{u,v\}\,:\, t^*_{\{u,v\}}=0 \text{ or} \\ \left(0 < t^*_{\{u,v\}} \neq t_{\{u,v\}} \text{ and } \tilde{x}_{\{u,v\}}^{(t^*_{\{u,v\}})}=1\right)}} \mathrm{Val}_{\{u,v\}}\,. \qquad (22)
\end{aligned}$$

The following lemma, which is the updated version of Corollary 13 for the Algorithm 3, bounds the number of edges in $F \cap \mathrm{NExtm}(Q_1^{(i)}, Q_2^{(i)})$ in terms of the average distance of the edges in $\mathrm{NExtm}(Q_1^{(i)}, Q_2^{(i)})$. We provide the proof in Section B in the appendix for further reference.

Lemma 18 (Section B)**.** Let (Q_1, Q_2) be a pair returned by the procedure ONE-HALF-REFINE-CUT. We have that

$$\sum_{\{i,j\}\in\mathrm{NExtm}(Q_1,Q_2)} \mathrm{Val}_{\{i,j\}} + \frac{1}{4}\cdot\left|\left\{\begin{smallmatrix}\{i,j\}\in\mathrm{NExtm}(Q_1,Q_2),\\ \{i,j\}\in NF\end{smallmatrix}\right\}\right| \;\ge\; \frac{1}{4}\cdot\left|\mathrm{NExtm}(Q_1,Q_2)\right|.$$

Combining Lemma 18 with (22) and Lemma 17, we obtain that

$$\begin{aligned}
\sum_{u\neq v} \#_{\{u,v\}} \;&\le\; 4\cdot \sum_{u\neq v} \mathrm{Val}_{\{u,v\}} \;+\; |\,NF\,| \\
&\le\; 5\cdot \sum_{u\neq v} \mathrm{Val}_{\{u,v\}}\,.
\end{aligned}$$

This proves Theorem 16.

VII. CONCLUSION

In this work, we present a new paradigm that advances the current understanding for hierarchical clustering in both conceptual and technical capacities. A natural question following our results is whether the presented paradigm can be extended to other variations of hierarchical clustering problems with different objectives. The technical problem boils down to the problem of finding cuts with prescribed properties regarding the average distances for the problem considered.

Another natural question is whether we can obtain better approximation result via improving the partitioning algorithm, e.g., ONE-HALF-REFINE-CUT in Algorithm 3. The current partitioning algorithm can be interpreted as follows: sort the points by their distance from the pivot, and cut this sorted list either at distance

ϵ or $1/2 - \epsilon$. One could ask: *what if we allow cutting to happen anywhere in the list?* We believe such an algorithm which partitions the ordered list of points into two consecutive sublists may be of interest.

Appendix

A. Lemma 5 – Forbidden Edges within any P.

Let $P \in \mathcal{P}^{(t)}$ be a cluster and recall that

- $\#_{\mathrm{F}}(P)$ denotes the number of forbidden edges clustered into P,
- $\Delta^{(t)}(P)$ denotes the gluer set of P, $\Delta_+^{(t)}(P) := P \cap \Delta^{(t)}(P)$ is referred to as the core of P,
- $\mathrm{Ext}^{(t)}(P) := P \backslash \Delta_+^{(t)}(P)$ denotes the extended part of P, and
- $L_1^{(t)}(P)$ denotes the set of elements in the $2/3$-vicinity of $P' \cap \Delta_+^{(\ell(t,P))}(P)$ within P' over all $P' \in \mathrm{Candi}^{(\ell(t,P))}(\Delta^{(\ell(t,P))}(P))$, where $\ell(t,P)$ is the index of the top-most layer up to the t-th layer at which P is newly-created.

We prove the following lemma.

Lemma 19 (Restate of Lemma 5)**.** For $\alpha := 0.3936$ and any $P \in \mathcal{P}^{(t)}$, we have

$$\#_{\mathrm{F}}(P) \;\leq\; \frac{(2-\alpha)(1+\alpha)^2}{2(1-\alpha)^2} \cdot \beta \cdot \left|\mathrm{NF}(\mathcal{Q}^{(t)}, P)\right|,$$

where $\beta := 0.8346$ and $\mathrm{NF}(\mathcal{Q}^{(t)}, P) := \left\{ \{i,j\} \in \mathrm{NF}(\mathcal{Q}^{(t)}) \,:\, i,j \in P \right\}$ denotes the set of edges in $\mathrm{NF}(\mathcal{Q}^{(t)})$ residing within P.

Proof. Since $\mathrm{diam}^{(t)}(\Delta_+^{(t)}(P)) < 1/3$, forbidden edges only occur between elements in $\mathrm{Ext}^{(t)}(P)$ and that in P. Hence, we have

$$\begin{aligned}\#_{\mathrm{F}}(P) \;&\leq\; |\,\mathrm{Ext}^{(t)}(P)| \cdot \left(\frac{|\,\mathrm{Ext}^{(t)}(P)|}{2} + |\Delta_+^{(t)}(P)| \right) \\ &\leq\; \frac{1}{1-\alpha} \cdot |L_1^{(t)}(P)| \cdot \frac{2-\alpha}{2(1-\alpha)} \cdot |\Delta_+^{(t)}(P)|, \qquad (23) \\ &\leq\; \frac{2-\alpha}{2(1-\alpha)^2} \cdot \alpha \cdot |\Delta_+^{(t)}(P)|^2, \qquad (24)\end{aligned}$$

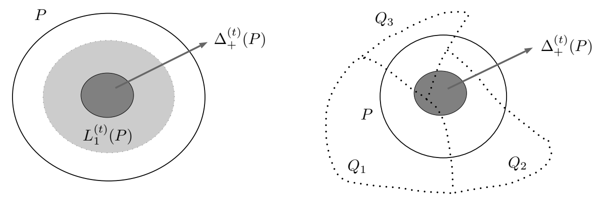
where we apply Lemma 9 in the last two inequalities.

We have two cases to consider. If P is a newly-formed cluster at the t-th layer, then any edge between $\Delta_+^{(t)}(P)$ and $L_1^{(t)}(P)$ crosses different pre-clusters and is non-forbidden by the way $L_1^{(t)}(P)$ is defined. Hence, these edges are contained within $\mathrm{NF}(\mathcal{Q}^{(t)}, P)$ and we have

$$|L_1^{(t)}(P)| \cdot |\Delta_+^{(t)}(P)| \;\leq\; \left|\mathrm{NF}(\mathcal{Q}^{(t)}, P)\right|.$$

Hence, from (23) we obtain

$$\#_{\mathrm{F}}(P) \;\leq\; \frac{2-\alpha}{2(1-\alpha)^2} \cdot \left|\mathrm{NF}(\mathcal{Q}^{(t)}, P)\right|. \qquad (25)$$



If P is a previously-formed cluster at a lower layer, then consider the set of pre-clusters in $\mathcal{Q}^{(t)}$ that intersect the core set $\Delta_+^{(t)}(P)$. Let $Q_1, \ldots, Q_k$ denote these pre-clusters and assume W.L.O.G. that $|Q_1 \cap \Delta_+^{(t)}(P)| = \max_{1 \leq j \leq k} |Q_j \cap \Delta_+^{(t)}(P)|$. Since $P \notin \mathrm{Candi}^{(t)}(Q_1)$, by Step 6 of Algorithm 1, we have

$$B_1 \;:=\; \left|\; \mathrm{Ball}^{(t)}_{<\frac{2}{3}}\left(P \cap Q_1,\; P \cap \overline{Q_1} \right) \;\right| \;\geq\; \alpha \cdot |P \cap Q_1|\,. \qquad (26)$$

We have two subcases to consider regarding the relative size of $|Q_j \cap \Delta_+^{(t)}(P)|$ for all j.

Case (i) – Imbalanced in Size: If $\sum_{2 \leq j \leq k} |Q_j \cap \Delta_+^{(t)}(P)| \;<\; \alpha \cdot |Q_1 \cap \Delta_+^{(t)}(P)|$, then

$$|\Delta_+^{(t)}(P)|^2 \;\;\leq\;\; (1+\alpha)^2 \cdot |Q_1 \cap \Delta_+^{(t)}(P)|^2. \qquad (27)$$

To bound $|Q_1 \cap \Delta_+^{(t)}(P)|^2$, further consider two subcases regarding the size of $L_1^{(t)}(P)$ and $\Delta_+^{(t)}(P)$.

1) If $|L_1^{(t)}(P)| \;\leq\; \beta \cdot \alpha |\Delta_+^{(t)}(P)|$, then Inequality (23) yields a good bound. Combined with (27), we have

$$\begin{aligned}\#_{\mathrm{F}}(P) \;&\leq\; \frac{2-\alpha}{2(1-\alpha)^2} \cdot \beta \cdot \alpha \cdot |\Delta_+^{(t)}(P)|^2 \\ &\leq\; \frac{2-\alpha}{2(1-\alpha)^2} \cdot (1+\alpha)^2 \\ &\qquad \cdot \beta \cdot \alpha \cdot |Q_1 \cap \Delta_+^{(t)}(P)|^2 \\ &\leq\; \frac{2-\alpha}{2(1-\alpha)^2} \cdot (1+\alpha)^2 \\ &\qquad \cdot \beta \cdot |Q_1 \cap \Delta_+^{(t)}(P)| \cdot B_1,\end{aligned}$$

where we use Condition (26) in the last inequality. Since the edges between $Q_1 \cap \Delta_+^{(t)}(P)$ and $\mathrm{Ball}^{(t)}_{<2/3}\left(P \cap Q_1,\; P \cap \overline{Q_1} \right)$ are non-forbidden, reside within P, and cross different pre-clusters, it follows that

$$\#_{\mathrm{F}}(P) \leq \frac{2-\alpha}{2(1-\alpha)^2}(1+\alpha)^2 \beta \left|\mathrm{NF}(\mathcal{Q}^{(t)}, P)\right|. \qquad (28)$$

2) If $|L_1^{(t)}(P)| \;\;\geq\;\; \beta \cdot \alpha |\Delta_+^{(t)}(P)|$, then a decent number of elements exist in the $2/3$-vicinity of $\Delta_+^{(t)}(P)$. Define for short the following notations.

- $\ell_1 := |Q_1 \cap L_1^{(t)}(P)|$ and $\ell_2 := \sum_{2 \leq j \leq k} |Q_j \cap L_1^{(t)}(P)|$,

- $\ell := |L_1^{(t)}(P)| - (\ell_1 + \ell_2)$,
- $G_1 := |Q_1 \cap \Delta_+^{(t)}(P)|$ and $G_2 := \sum_{2 \le j \le k} |Q_j \cap \Delta_+^{(t)}(P)|$.

Further consider two subcases regarding the relative size of $Q_1 \cap \mathrm{Ext}^{(t)}(P)$ and $L_1^{(t)}(P)$.

If $\ell_1 \ge \eta \cdot |L_1^{(t)}(P)|$, where $\eta := \frac{1-\beta}{\alpha\beta^2} \approx 0.6034$, then Q_1 contains a large number of elements in addition to those in $Q_1 \cap \Delta_+^{(t)}(P)$.

In particular, we have $\ell_1 \ge \eta \cdot |L_1^{(t)}(P)| \ge \alpha\beta\eta \cdot |\Delta_+^{(t)}(P)| \ge \alpha\beta\eta \cdot G_1$. Applying Condition (26), we obtain

$$\begin{aligned} B_1 &\ge \alpha \cdot |P \cap Q_1| \ge \alpha \cdot (\ell_1 + G_1) \\ &\ge \alpha \cdot (1 + \alpha\beta\eta) \cdot G_1. \end{aligned}$$

Following (27) and that $G_1 := |Q_1 \cap \Delta_+^{(t)}(P)|$, we obtain

$$\begin{aligned} |\Delta_+^{(t)}(P)|^2 &\le (1+\alpha)^2 \cdot G_1^2 \\ &\le \frac{(1+\alpha)^2}{\alpha \cdot (1+\alpha\beta\eta)} \cdot G_1 \cdot B_1 \\ &= \frac{(1+\alpha)^2 \cdot \beta}{\alpha} \cdot G_1 \cdot B_1, \end{aligned}$$

where in the last equality we plug in the setting of η to obtain that $\frac{1}{1+\alpha\beta\eta} = \beta$. Since $G_1 \cdot B_1 \le \left|\mathrm{NF}(\mathcal{Q}^{(t)}, P)\right|$, from (24) we have

$$\begin{aligned} \#_{\mathrm{F}}(P) \le\ & \frac{2-\alpha}{2(1-\alpha)^2} \cdot (1+\alpha)^2 \\ & \cdot \beta \cdot \left|\mathrm{NF}(\mathcal{Q}^{(t)}, P)\right|. \end{aligned} \tag{29}$$

If $\ell_1 \le \eta \cdot |L_1^{(t)}(P)|$, then a decent fraction of elements in $L_1^{(t)}(P)$ lies outside Q_1 and is ready to pair up with elements in $Q_1 \cap \Delta_+^{(t)}(P)$. We have

$$\begin{aligned} \ell_2 + \ell &\ge (1-\eta) \cdot |L_1^{(t)}(P)| \\ &\ge \alpha \cdot \beta \cdot (1-\eta) \cdot |\Delta_+^{(t)}(P)|. \end{aligned} \tag{30}$$

Let $\gamma := \frac{2\alpha\beta\cdot(1-\eta)}{1+\alpha\beta\cdot(1-\eta)} \approx 0.2305$. We have

$$\begin{aligned} & |\Delta_+^{(t)}(P)|^2 \\ &= (\, G_1 \,+\, G_2 \,) \cdot |\Delta_+^{(t)}(P)| \\ &= \gamma \cdot G_1 \cdot |\Delta_+^{(t)}(P)| \\ &\quad + (\, (1-\gamma) \cdot G_1 \,+\, G_2 \,) \cdot (\, G_1 \,+\, G_2 \,) \\ &\le \frac{\gamma}{\alpha\beta(1-\eta)} \cdot G_1 \cdot (\ell_2 + \ell) \\ &\quad + \frac{\gamma}{\alpha\beta(1-\eta)} \cdot G_1 \cdot G_2 \\ &\quad + (\, 1-\gamma \,) \cdot G_1^2 \\ &\quad + \left(2 - \gamma - \frac{\gamma}{\alpha\beta(1-\eta)}\right) \cdot G_1 G_2 + G_2^2, \end{aligned} \tag{31}$$

where in the last inequality we apply Inequality (30). Note that by the setting of γ, for the coefficient of $G_1 \cdot G_2$ in the above, we have

$$2 - \gamma - \frac{\gamma}{\alpha\beta(1-\eta)} \ \ge\ 0.$$

Hence, all the coefficients in the right-hand-side of (31) are non-negative, and it gives a valid upper-bound of $|\Delta_+^{(t)}(P)|^2$ in terms of edges counted in $G_1 \cdot (\ell_2 + \ell + G_2)$, G_1^2, $G_1 \cdot G_2$, and G_2^2. Since $G_2 < \alpha \cdot G_1$, from (31) we have

$$\begin{aligned} & |\Delta_+^{(t)}(P)|^2 \\ &\le \frac{\gamma}{\alpha\beta(1-\eta)} \cdot G_1 \cdot (\ell_2 + \ell + G_2) \\ &\quad + \left((1-\gamma) + \alpha\left(2 - \gamma - \frac{\gamma}{\alpha\beta(1-\eta)}\right) + \alpha^2\right) G_1^2 \\ &\le \frac{\gamma}{\alpha\beta(1-\eta)} \cdot G_1 \cdot (\ell_2 + \ell + G_2) \\ &\quad + \frac{1}{\alpha}\left((1+\alpha)^2 - \gamma(1+\alpha) - \frac{\gamma}{\beta(1-\eta)}\right) G_1 B_1, \end{aligned} \tag{32}$$

where in the last inequality we apply Condition (26). Since the edges counted in $G_1 \cdot (\ell_2 + \ell + G_2)$ and $G_1 \cdot B_1$ are non-forbidden, it follows that

$$\begin{aligned} \max\left\{\, G_1 \cdot (\ell_2 + \ell + G_2)\,,\ G_1 \cdot B_1 \,\right\} & \\ \le \left|\mathrm{NF}(\mathcal{Q}^{(t)}, P)\right|. & \end{aligned}$$

Combining the above with (32), we obtain

$$\begin{aligned} & \#_{\mathrm{F}}(P) \\ &\le \frac{2-\alpha}{2(1-\alpha)^2} \cdot (1+\alpha)^2 \\ &\qquad \cdot \left(1 - \frac{\gamma}{1+\alpha}\right) \cdot \left|\mathrm{NF}(\mathcal{Q}^{(t)}, P)\right|, \end{aligned} \tag{33}$$

where $\frac{\gamma}{1+\alpha} = \frac{1}{1+\alpha} \cdot \frac{2\alpha\beta^2 + 2\beta - 2}{\alpha\beta^2 + 2\beta - 1}$ by plugging in the setting for η.

Case (ii) – Balanced in Size: If $\sum_{2 \le j \le k} |Q_j \cap \Delta_+^{(t)}(P)| \ge \alpha \cdot |Q_1 \cap \Delta_+^{(t)}(P)|$, since $\alpha \le 1/2$, it follows that $Q_1, \ldots, Q_k$ can be partitioned into two groups $\mathcal{G}_1$

and $\mathcal{G}_2$ such that[2]

$$\alpha \cdot \sum_{Q \in \mathcal{G}_1} |Q \cap \Delta_+^{(t)}(P)| \leq \sum_{Q \in \mathcal{G}_2} |Q \cap \Delta_+^{(t)}(P)| \leq \sum_{Q \in \mathcal{G}_1} |Q \cap \Delta_+^{(t)}(P)|.$$

Define for short the following notations.

- $G_1 := \sum_{Q \in \mathcal{G}_1} |Q \cap \Delta_+^{(t)}(P)|$, and
- $G_2 := \sum_{Q \in \mathcal{G}_2} |Q \cap \Delta_+^{(t)}(P)|$.

We have

$$\begin{aligned} |\Delta_+^{(t)}(P)|^2 &= (G_1 + G_2)^2 \\ &= \left(\frac{G_1}{G_2} + \frac{G_2}{G_1} + 2 \right) \cdot G_1 \cdot G_2 \\ &\leq \left(\frac{1}{\alpha} + \alpha + 2 \right) \cdot G_1 \cdot G_2 \\ &= \frac{(1+\alpha)^2}{\alpha} \cdot G_1 \cdot G_2, \end{aligned} \quad (34)$$

where the last inequality follows since, within the interval $[\alpha, 1]$, the function $f(x) = x + 1/x$ attains its maximum value at $x = \alpha$.

Further consider two subcases regarding the relative size of $L_1^{(t)}(P)$ and $|\Delta_+^{(t)}(P)|$.

1) If $|L_1^{(t)}(P)| \leq \beta \cdot \alpha |\Delta_+^{(t)}(P)|$, then following Inequality (23) and (34) we have

$$\begin{aligned} \#_{\mathrm{F}}(P) &\leq \frac{2-\alpha}{2(1-\alpha)^2} \cdot \beta \cdot \alpha \cdot |\Delta_+^{(t)}(P)|^2 \\ &\leq \frac{2-\alpha}{2(1-\alpha)^2} \cdot (1+\alpha)^2 \\ &\qquad \cdot \beta \cdot \left| \mathrm{NF}(\mathcal{Q}^{(t)}, P) \right|. \end{aligned} \quad (35)$$

2) If $|L_1^{(t)}(P)| \geq \beta \cdot \alpha |\Delta_+^{(t)}(P)|$, then a decent number of edges exists between $L_1^{(t)}(P)$ and $\Delta_+^{(t)}(P)$. In this regard, define the following notations.
   - $\ell_1 := \sum_{Q \in \mathcal{G}_1} |Q \cap L_1^{(t)}(P)|$,
     $\ell_2 := \sum_{Q \in \mathcal{G}_2} |Q \cap L_1^{(t)}(P)|$,
   - $L := |L_1^{(t)}(P)|$, and $\ell := L - (\ell_1 + \ell_2)$.

   Further define $G := G_1 \cdot (\ell_2 + \ell) + G_2 \cdot (\ell_1 + \ell)$ to count edges between $L_1^{(t)}(P)$ and $\Delta_+^{(t)}(P)$. We have

$$\begin{aligned} G &\geq \left(\frac{\ell_2}{L} + \frac{\ell}{L} \right) \cdot G_1 L + \alpha \left(\frac{\ell_1}{L} + \frac{\ell}{L} \right) \cdot G_1 L \\ &\geq \alpha G_1 L \geq \alpha^2 \beta G_1 \cdot |\Delta_+^{(t)}(P)|, \end{aligned}$$

which the second inequality follows from the fact that the previous R.H.S. attains its minimum value when $\ell_1 = L$ and $\ell_2 = \ell = 0$. Since

$$G_1 \cdot |\Delta_+^{(t)}(P)| = G_1^2 + G_1 \cdot G_2 \geq 2 \cdot G_1 \cdot G_2,$$

we obtain that

$$\begin{aligned} G_1 \cdot G_2 &= \zeta \cdot G_1 G_2 + (1-\zeta) \cdot G_1 G_2 \\ &\leq \frac{\zeta}{2\alpha^2 \beta} \cdot G + (1-\zeta) \cdot G_1 G_2, \end{aligned} \quad (36)$$

where $\zeta := \frac{2\alpha^2 \cdot \beta}{1 + 2\alpha^2 \cdot \beta}$. Note that, the setting of ζ satisfies that

$$\frac{\zeta}{2\alpha^2 \beta} = 1 - \zeta.$$

Combining (36) with (34), we obtain

$$\begin{aligned} &|\Delta_+^{(t)}(P)|^2 \\ &\leq \frac{(1+\alpha)^2}{\alpha} \cdot \frac{1}{1 + 2\alpha^2 \beta} \cdot (G + G_1 \cdot G_2). \end{aligned}$$

Since G and $G_1 \cdot G_2$ count two disjoint sets of non-forbidden edges in $\mathrm{NF}(\mathcal{Q}^{(t)}, P)$, it follows from (24) that

$$\begin{aligned} \#_{\mathrm{F}}(P) &\leq \frac{2-\alpha}{2(1-\alpha)^2} \cdot (1+\alpha)^2 \\ &\qquad \cdot \frac{1}{2\alpha^2 \beta} \cdot \left| \mathrm{NF}(\mathcal{Q}^{(t)}, P) \right|. \end{aligned} \quad (37)$$

Combining Inequalities (28), (29), (33), (35), and (37), we obtain

$$\#_{\mathrm{F}}(P) \leq \frac{2-\alpha}{2(1-\alpha)^2} (1+\alpha)^2 W \cdot \left| \mathrm{NF}(\mathcal{Q}^{(t)}, P) \right|,$$

where

$$W := \max \begin{cases} \beta, \\ 1 - \frac{2\alpha\beta^2 + 2\beta - 2}{(1+\alpha)(\alpha\beta^2 + 2\beta - 1)}, \\ \frac{1}{1 + 2\alpha^2 \beta}, \end{cases}$$

which has a value of 0.8346 with the setting $\alpha := 0.3936$ and $\beta := 0.8346$. Since $W = \beta$ and $(1+\alpha)^2 \beta \geq 1$, the statement of this lemma follows. □

B. Lemma 18 – Average Distance of Non-extreme Cut Edges

Consider the procedure ONE-HALF-REFINE-CUT with input tuple (P, v, x), where x is a distance function, P is a set with $\mathrm{diam}^{(x)}(P) \geq 1/2$, and $v \in P$ is the pivot with $\max_{u \in P} x_{\{v,u\}} \geq 1/2$.

Suppose that the procedure is called at the t-th layer and (Q_1, Q_2) with $v \in Q_1$ is the pair returned by the procedure ONE-HALF-REFINE-CUT. Recall that we use

$$\mathrm{Val}_{\{u,v\}} := \left(1 - \tilde{x}_{\{u,v\}}^{(t_{\{u,v\}})} \right) + \tilde{x}_{\{u,v\}}^{(t_{\{u,v\}}+1)}$$

[2]Note that, one way is to start with two empty groups and consider Q_j in non-ascending order of $|Q_j \cap \Delta_+^{(t)}(P)|$ for all $1 \leq j \leq k$. For each Q_j considered, assign it to the group that has a smaller intersection with $\Delta^{(t)}(P)$ in size.

```
procedure ONE-HALF-REFINE-CUT(P, v, x)
    if Condition (21) is satisfied for (P, v) then
        return { {v}, P \ {v} }.
          // make v a singleton
    else
        return {Ball^(x)_{<1/2}(v, P), P\Ball^(x)_{<1/2}(v, P)}
          // cut at 1/2 − ε
    end if
end procedure
```

to denote the objective value the edge $\{u,v\}$ possesses, $\mathrm{NExtm}(Q_1,Q_2)$ to denote the set of edges with distances strictly smaller than 1 between Q_1 and Q_2, and $NF := \binom{V}{2} \setminus F$ to denote the set of edges $\{u,v\}$ with $\tilde{x}^{(t_{\{u,v\}})}_{\{u,v\}} < 1$.

In this section we prove the following lemma.

Lemma 20 (Restate of Lemma 18)**.**

$$\sum_{\{i,j\}\in \mathrm{NExtm}(Q_1,Q_2)} \mathrm{Val}_{\{i,j\}} + \frac{1}{4}\cdot \left|\left\{ \substack{\{i,j\}\in \mathrm{NExtm}(Q_1,Q_2),\\ \{i,j\}\in NF} \right\}\right| \geq \frac{1}{4}\cdot \left|\mathrm{NExtm}(Q_1,Q_2)\right| .$$

For the ease of notation define

$$B_{1/4} := \mathrm{Ball}^{(x)}_{<1/4}(v,P), \quad B_{1/2} := \mathrm{Ball}^{(x)}_{<1/2}(v,P),$$
$$B_{3/4} := \mathrm{Ball}^{(x)}_{<3/4}(v,P), \quad \text{and} \quad Q_2' := Q_2 \cap B_{3/4}.$$

To prove Lemma 20, first we bound the cardinality of $\mathrm{NExtm}(Q_1,Q_2)$ in terms of the average distance of the edges it contains. The following lemma is the updated version of Lemma 7 for the procedure ONE-HALF-REFINE-CUT.

Lemma 21.

$$\sum_{\substack{\{i,j\}\in \mathrm{NExtm}(Q_1,Q_2),\\ i\in Q_1,\, j\in Q_2'}} \left(\min\left\{ x_{\{v,j\}},\ \frac{1}{2} \right\} - x_{\{v,i\}} \right) \geq \frac{1}{4}\cdot \left|\left\{ \substack{\{i,j\}\in \mathrm{NExtm}(Q_1,Q_2),\\ i\in Q_1,\, j\in Q_2'} \right\}\right| .$$

Proof. For any $p,q \in B_{3/4}$, define $d(p,q) := |\min\{x_{\{v,p\}},1/2\} - \min\{x_{\{v,q\}},1/2\}| - 1/4$. Since $Q_1 \subseteq B_{1/2}$, to prove this lemma, it suffices to prove that

$$\sum_{\{p,q\}\in \mathrm{NExtm}(Q_1,Q_2')} d(p,q) \ \geq\ 0. \tag{38}$$

From the design of the procedure ONE-HALF-REFINE-CUT, we have

$$(Q_1,Q_2') \in \left\{ \begin{array}{l} \mathrm{Cut}_1 = \big(\ \{v\},\ B_{3/4}\setminus\{v\}\ \big), \\ \mathrm{Cut}_2 = \big(\ B_{1/2},\ B_{3/4}\setminus B_{1/2}\ \big) \end{array} \right\}.$$

Hence, to prove (38), it suffices to prove that

$$W := \max_{1\leq i\leq 2} \left\{ \sum_{\{p,q\}\in \mathrm{NExtm}(\mathrm{Cut}_i)} d(p,q) \right\} \geq 0. \tag{39}$$

In the following we prove (39).

Let $k := |B_{1/4}|$, $\ell := |B_{1/2}\setminus B_{1/4}|$, and $m := |B_{3/4}\setminus B_{1/2}|$. For Cut_1, any $q \in B_{3/4}\setminus\{v\}$ always forms a non-extreme edges with v. Hence, we have

$$\begin{aligned}
&\sum_{\{p,q\}\in \mathrm{NExtm}(\mathrm{Cut}_1)} d(p,q) \\
&= \sum_{q\in B_{1/2}} x_{\{v,q\}} + \frac{1}{2}\cdot |B_{3/4}\setminus B_{1/2}| - \frac{1}{4}\cdot \big(|B_{3/4}| - 1\big) \\
&= \sum_{q\in B_{1/2}} x_{\{v,q\}} + \frac{1}{4}\cdot |B_{3/4}| - \frac{1}{2}\cdot |B_{1/2}| + \frac{1}{4} \qquad (40) \\
&= \sum_{q\in B_{1/2}} x_{\{v,q\}} + \frac{1}{4}\cdot (m - k - \ell + 1). \qquad (41)
\end{aligned}$$

Note that the nonnegativity of (40) is exactly the condition tested by the procedure ONE-HALF-REFINE-CUT.

For Cut_2, observe that any $p \in B_{1/4}$ and $q \in B_{3/4}$ always forms a non-extreme edge. For any $p \in B_{1/2}\setminus B_{1/4}$, let $N(p)$ denote the number of elements in $B_{3/4}\setminus B_{1/2}$ that forms a non-extreme edge with p. It follows that

$$\sum_{\{p,q\}\in \mathrm{NExtm}(\mathrm{Cut}_2)} d(p,q) = \frac{1}{4}|\,\mathrm{NExtm}(\mathrm{Cut}_2)| - m\cdot \sum_{q\in B_{1/4}} x_{\{v,q\}} - \sum_{q\in B_{1/2}\setminus B_{1/4}} N(q) x_{\{v,q\}}. \tag{42}$$

From the definition of W in (39) with (41) and (42), we obtain

$$\begin{aligned}
W &\geq \frac{m}{m+1}\cdot \sum_{\{p,q\}\in \mathrm{NExtm}(\mathrm{Cut}_1)} d(p,q) + \frac{1}{m+1}\cdot \sum_{\{p,q\}\in \mathrm{NExtm}(\mathrm{Cut}_2)} d(p,q) \\
&= \frac{m}{m+1}\cdot \left(\sum_{q\in B_{1/2}} x_{\{v,q\}} + \frac{1}{4}(m-k-\ell+1) \right) \\
&\quad + \frac{1}{m+1}\cdot \left(\frac{1}{4}|\,\mathrm{NExtm}(\mathrm{Cut}_2)| - m \sum_{q\in B_{1/4}} x_{\{v,q\}} - \sum_{q\in B_{1/2}\setminus B_{1/4}} N(q) x_{\{v,q\}} \right).
\end{aligned}$$

Further plugging in $|\text{NExtm}(\text{Cut}_2)| = m \cdot k + \sum_{q \in B_{1/2} \setminus B_{1/4}} N(q)$, we obtain

$$W \geq \frac{1}{m+1} \cdot \left(\sum_{q \in B_{1/2} \setminus B_{1/4}} (m - N(q)) x_{\{v,q\}} + \frac{1}{4} m(m - \ell + 1) + \frac{1}{4} \sum_{q \in B_{1/2} \setminus B_{1/4}} N(q) \right)$$
$$\geq \frac{m}{4(m+1)} \cdot (m+1) \geq 0,$$

where in the second last inequality we use the fact that $x_{\{v,q\}} \geq 1/4$ for any $q \in B_{1/2} \setminus B_{1/4}$. ☐

Recall that t is the layer at which the procedure ONE-HALF-REFINE-CUT is called and the pair (Q_1, Q_2) with $v \in Q_1$ is separated. Also recall that $E_+^{(t)}$ and $E_-^{(t)}$ denote the set of $+$ edges and the set of $-$ edges at the t-th layer.

We have the following lemma.

Lemma 22.

$$\sum_{\substack{\{i,j\} \in \text{NExtm}(Q_1,Q_2), \\ \{i,j\} \in E_+^{(t)}}} x_{\{i,j\}} + \sum_{\substack{\{i,j\} \in \text{NExtm}(Q_1,Q_2), \\ \{i,j\} \in E_-^{(t)}}} \left(1 - x_{\{i,j\}}\right) + \frac{1}{4} \cdot \left| \left\{ \begin{smallmatrix} \{i,j\} \in \text{NExtm}(Q_1,Q_2'), \\ \{i,j\} \in E_-^{(t)} \end{smallmatrix} \right\} \right|$$
$$\geq \sum_{\substack{\{i,j\} \in \text{NExtm}(Q_1,Q_2), \\ i \in Q_1,\, j \in Q_2'}} \left(\min\left\{ x_{\{v,j\}}, \frac{1}{2} \right\} - x_{\{v,i\}} \right) + \sum_{\substack{\{i,j\} \in \text{NExtm}(Q_1,Q_2), \\ i \in Q_1,\, j \in Q_2 \setminus Q_2', \\ \{i,j\} \in E_+^{(t)}}} x_{\{i,j\}}.$$

Proof. To prove this lemma, we compare both sides of the inequality for each $\{i,j\} \in \text{NExtm}(Q_1, Q_2)$ with $i \in Q_1$.

1) If $\{i,j\}$ is a $+$ edge in $E_+^{(t)}$, then using the triangle inequality, we have $x_{\{i,j\}} \geq x_{\{v,j\}} - x_{\{v,i\}}$ and hence $x_{\{i,j\}} \geq \min\left\{x_{\{v,j\}}, \frac{1}{2}\right\} - x_{\{v,i\}}$.
2) If $\{i,j\}$ is a $-$ edge in $E_-^{(t)}$ with $j \in Q_2'$, then further consider the following subcases.
   a) If Q_1 is a singleton-cluster, then it follows that $x_{\{i,j\}} \leq 3/4$ and
   $$\left(1 - x_{\{i,j\}}\right) + \frac{1}{4} \geq \frac{1}{2} \geq \min\left\{ x_{\{v,j\}}, \frac{1}{2} \right\} - x_{\{v,i\}}.$$
   b) If $Q_1 = B_{1/2}$, then $j \in B_{3/4} \setminus B_{1/2}$.
      i) If $i \in B_{1/4}$, then the triangle inequality implies that
      $$1 - x_{\{i,j\}} \geq 1 - x_{\{v,i\}} - x_{\{v,j\}} \geq \frac{1}{4} - x_{\{v,i\}}.$$
      On the other hand, $\min\{x_{\{v,j\}}, \frac{1}{2}\} - x_{\{v,i\}} = \frac{1}{2} - x_{\{v,i\}}$. Hence,
      $$\left(1 - x_{\{i,j\}}\right) + \frac{1}{4} \geq \min\left\{ x_{\{v,j\}}, \frac{1}{2} \right\} - x_{\{v,i\}}.$$
      ii) If $i \in B_{1/2} \setminus B_{1/4}$, then
      $$\left(1 - x_{\{i,j\}}\right) + \frac{1}{4} \geq \frac{1}{4} = \frac{1}{2} - \frac{1}{4} \geq \min\left\{ x_{\{v,j\}}, \frac{1}{2} \right\} - x_{\{v,i\}}.$$

We have compared all $\{i,j\}$ in the above case arguments. This proves the lemma. ☐

In the following we prove Lemma 20.

Proof of Lemma 20. We have that

$$\sum_{\{i,j\} \in \text{NExtm}(Q_1,Q_2)} \text{Val}_{\{i,j\}} \geq \sum_{\substack{\{i,j\} \in \text{NExtm}(Q_1,Q_2), \\ \{i,j\} \in E_+^{(t)}}} x_{\{i,j\}} + \sum_{\substack{\{i,j\} \in \text{NExtm}(Q_1,Q_2), \\ \{i,j\} \in E_-^{(t)}}} \left(1 - x_{\{i,j\}}\right)$$

by the definition of $E_+^{(t)}$, $E_-^{(t)}$, and the non-decreasing property of $\tilde{x}_{\{i,j\}}$ over the layers. Combining the above statement with Lemma 22, we obtain that

$$\sum_{\{i,j\} \in \text{NExtm}(Q_1,Q_2)} \text{Val}_{\{i,j\}} + \frac{1}{4} \left| \left\{ \begin{smallmatrix} \{i,j\} \in \text{NExtm}(Q_1,Q_2'), \\ \{i,j\} \in E_-^{(t)} \end{smallmatrix} \right\} \right|$$
$$\geq \sum_{\substack{\{i,j\} \in \text{NExtm}(Q_1,Q_2), \\ i \in Q_1,\, j \in Q_2'}} \left(\min\left\{ x_{\{v,j\}}, \frac{1}{2} \right\} - x_{\{v,i\}} \right) + \sum_{\substack{\{i,j\} \in \text{NExtm}(Q_1,Q_2), \\ i \in Q_1,\, j \in Q_2 \setminus Q_2', \\ \{i,j\} \in E_+^{(t)}}} x_{\{i,j\}}$$
$$\geq \sum_{\substack{\{i,j\} \in \text{NExtm}(Q_1,Q_2), \\ i \in Q_1,\, j \in Q_2'}} \left(\min\left\{ x_{\{v,j\}}, \frac{1}{2} \right\} - x_{\{v,i\}} \right) + \frac{1}{4} \cdot \left| \left\{ \begin{smallmatrix} \{i,j\} \in \text{NExtm}(Q_1,Q_2), \\ i \in Q_1,\, j \in Q_2 \setminus Q_2', \\ \{i,j\} \in E_+^{(t)} \end{smallmatrix} \right\} \right|, \tag{43}$$

where in the last inequality we use the fact that $x_{\{i,j\}} \geq 1/4$ for any $\{i,j\} \in \text{NExtm}(Q_1, Q_2)$ with $i \in Q_1,\ j \in$

$Q_2 \setminus Q_2'$ by the design of the procedure ONE-HALF-REFINE-CUT.

Adding $\frac{1}{4} \cdot \left| \left\{ \substack{\{i,j\} \in \mathrm{NExtm}(Q_1, Q_2), \\ i \in Q_1,\, j \in Q_2 \setminus Q_2', \\ \{i,j\} \in E_-^{(t)}} \right\} \right|$ to both sides of (43) and combining it with Lemma 21, we obtain that

$$\sum_{\{i,j\} \in \mathrm{NExtm}(Q_1,Q_2)} \mathrm{Val}_{\{i,j\}} + \frac{1}{4} \left| \left\{ \substack{\{i,j\} \in \mathrm{NExtm}(Q_1,Q_2), \\ \{i,j\} \in E_-^{(t)}} \right\} \right| \geq \frac{1}{4} \cdot \left| \mathrm{NExtm}(Q_1, Q_2) \right| .$$

The statement of this lemma follows from the above inequality and the fact that

$$\left| \left\{ \substack{\{i,j\} \in \mathrm{NExtm}(Q_1,Q_2), \\ \{i,j\} \in NF} \right\} \right| \geq \left| \left\{ \substack{\{i,j\} \in \mathrm{NExtm}(Q_1,Q_2), \\ \{i,j\} \in E_-^{(t)}} \right\} \right| .$$

□

ACKNOWLEDGMENT

We thank the anonymous reviewers for their helpful comments.

REFERENCES

[1] L. L. Cavalli-Sforza and A. W. F. Edwards, "Phylogenetic analysis: Models and estimation procedures," *Evolution*, vol. 21, no. 3, pp. 550–570, 1967. [Online]. Available: http://www.jstor.org/stable/2406616

[2] P. H. Sneath and R. R. Sokal, "Numerical taxonomy," *Nature*, vol. 193, pp. 855–860, 1962.

[3] ——, *Numerical Taxonomy: The Principles and Practice of Numerical Classification*. W H Freeman & Co, 1963.

[4] N. Bansal, A. Blum, and S. Chawla, "Correlation clustering," *Mach. Learn.*, vol. 56, no. 1-3, pp. 89–113, 2004, (Announced at FOCS'02). [Online]. Available: https://doi.org/10.1023/B:MACH.0000033116.57574.95

[5] V. Cohen-Addad, E. Lee, and A. Newman, "Correlation clustering with sherali-adams," in *63rd IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, 2022, pp. 651–661. [Online]. Available: https://doi.org/10.1109/FOCS54457.2022.00068

[6] V. Cohen-Addad, E. Lee, S. Li, and A. Newman, "Handling correlated rounding error via preclustering: A 1.73-approximation for correlation clustering," in *64th IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, 2023, pp. 1082–1104. [Online]. Available: https://doi.org/10.1109/FOCS57990.2023.00065

[7] N. Cao, V. Cohen-Addad, E. Lee, S. Li, A. Newman, and L. Vogl, "Understanding the cluster linear program for correlation clustering," in *Proceedings of the 56th Annual ACM Symposium on Theory of Computing (STOC)*, 2024, p. 1605–1616. [Online]. Available: https://doi.org/10.1145/3618260.3649749

[8] V. Cohen-Addad, D. R. Lolck, M. Pilipczuk, M. Thorup, S. Yan, and H. Zhang, "Combinatorial correlation clustering," in *Proceedings of the 56th Annual ACM Symposium on Theory of Computing (STOC)*, 2024, pp. 1617–1628. [Online]. Available: https://doi.org/10.1145/3618260.3649712

[9] N. Ailon and M. Charikar, "Fitting tree metrics: Hierarchical clustering and phylogeny," *SIAM J. Comput.*, vol. 40, no. 5, pp. 1275–1291, 2011, (Announced at FOCS'05). [Online]. Available: https://doi.org/10.1137/100806886

[10] V. Cohen-Addad, D. Das, E. Kipouridis, N. Parotsidis, and M. Thorup, "Fitting distances by tree metrics minimizing the total error within a constant factor," *J. ACM*, vol. 71, no. 2, pp. 10:1–10:41, 2024, (Announced at FOCS'21). [Online]. Available: https://doi.org/10.1145/3639453

[11] V. Cohen-Addad, C. Fan, E. Lee, and A. de Mesmay, "Fitting metrics and ultrametrics with minimum disagreements," *SIAM J. Comput.*, vol. 54, no. 1, pp. 92–133, 2025, (Announced at FOCS'22). [Online]. Available: https://doi.org/10.1137/22m1520190

[12] M. Charikar and R. Gao, "Improved approximations for ultrametric violation distance," in *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2024, pp. 1704–1737. [Online]. Available: https://doi.org/10.1137/1.9781611977912.68

[13] F. Bonchi, A. Gionis, and A. Ukkonen, "Overlapping correlation clustering," *Knowl. Inf. Syst.*, vol. 35, no. 1, pp. 1–32, 2013. [Online]. Available: https://doi.org/10.1007/s10115-012-0522-9

[14] A. Arasu, C. Ré, and D. Suciu, "Large-scale deduplication with constraints using dedupalog," in *Proceedings of the 25th International Conference on Data Engineering, (ICDE)*, 2009, pp. 952–963. [Online]. Available: https://doi.org/10.1109/ICDE.2009.43

[15] Y. Chen, S. Sanghavi, and H. Xu, "Clustering sparse graphs," in *Proceedings of the 26th Advances in Neural Information Processing Systems (NIPS)*, 2012, pp. 2213–2221. [Online]. Available: https://proceedings.neurips.cc/paper/2012/hash/1e6e0a04d20f50967c64dac2d639a577-Abstract.html

[16] D. V. Kalashnikov, Z. Chen, S. Mehrotra, and R. Nuray-Turan, "Web people search via connection analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 11, pp. 1550–1565, 2008. [Online]. Available: https://doi.org/10.1109/TKDE.2008.78

[17] R. Agrawal, A. Halverson, K. Kenthapadi, N. Mishra, and P. Tsaparas, "Generating labels from clicks," in *Proceedings of the 2nd ACM International Conference on Web Search and Data Mining*, 2009, pp. 172–181. [Online]. Available: https://doi.org/10.1145/1498759.1498824

[18] D. Chakrabarti, R. Kumar, and K. Punera, "A graph-theoretic approach to webpage segmentation," in *Proceedings of the 17th International Conference on World Wide Web (WWW)*, 2008, pp. 377–386. [Online]. Available: https://doi.org/10.1145/1367497.1367549

[19] M. Charikar, V. Guruswami, and A. Wirth, "Clustering with qualitative information," *J. Comput. Syst. Sci.*, vol. 71, no. 3, pp. 360–383, 2005, (Announced at FOCS'03). [Online]. Available: https://doi.org/10.1016/j.jcss.2004.10.012

[20] N. Ailon, M. Charikar, and A. Newman, "Aggregating inconsistent information: Ranking and clustering," *J. ACM*, vol. 55, no. 5, pp. 23:1–23:27, 2008, (Announced at STOC'05). [Online]. Available: https://doi.org/10.1145/1411509.1411513

[21] 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 47th Annual ACM on Symposium on Theory of Computing (STOC)*, 2015, pp. 219–228. [Online]. Available: https://doi.org/10.1145/2746539.2746604

[22] N. Cao, V. Cohen-Addad, E. Lee, S. Li, D. R. Lolck, A. Newman, M. Thorup, L. Vogl, S. Yan, and H. Zhang, "Solving the correlation cluster LP in sublinear time," in *Proceedings of the 57th Annual ACM Symposium on Theory of Computing, STOC 2025*. ACM, 2025, pp. 1154–1165. [Online]. Available: https://doi.org/10.1145/3717823.3718181

[23] S. Behnezhad, M. Derakhshan, M. Hajiaghayi, C. Stein, and M. Sudan, "Fully dynamic maximal independent set with polylogarithmic update time," in *60th IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, 2019, pp. 382–405. [Online]. Available: https://doi.org/10.1109/FOCS.2019.00032

[24] N. Cao, S. Huang, and H. Su, "Breaking 3-factor approximation for correlation clustering in polylogarithmic rounds," in *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2024, pp. 4124–4154. [Online]. Available: https://doi.org/10.1137/1.9781611977912.143

[25] S. Behnezhad, M. Charikar, W. Ma, and L. Tan, "Almost 3-approximate correlation clustering in constant rounds," in *Proceedings of 63rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2022, pp. 720–731. [Online]. Available: https://doi.org/10.1109/FOCS54457.2022.00074

[26] V. Cohen-Addad, S. Lattanzi, S. Mitrovic, A. Norouzi-Fard, N. Parotsidis, and J. Tarnawski, "Correlation clustering in constant many parallel rounds," in *Proceedings of the 38th International Conference on Machine Learning (ICML)*, vol. 139, 2021, pp. 2069–2078. [Online]. Available: http://proceedings.mlr.press/v139/cohen-addad21b.html

[27] S. Behnezhad, M. Charikar, W. Ma, and L. Tan, "Single-pass streaming algorithms for correlation clustering," in *Proceedings of the 2023 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2023, pp. 819–849. [Online]. Available: https://doi.org/10.1137/1.9781611977554.ch33

[28] K. Makarychev and S. Chakrabarty, "Single-pass pivot algorithm for correlation clustering. keep it simple!" in *36th Advances in Neural Information Processing Systems (NIPS)*, 2023. [Online]. Available: http://papers.nips.cc/paper_files/paper/2023/hash/149ad6e32c08b73a3ecc3d11977fcc47-Abstract-Conference.html

[29] M. Cambus, F. Kuhn, E. Lindy, S. Pai, and J. Uitto, "A (3 + ε)-approximate correlation clustering algorithm in dynamic streams," in *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2024, pp. 2861–2880. [Online]. Available: https://doi.org/10.1137/1.9781611977912.101

[30] S. Assadi and C. Wang, "Sublinear time and space algorithms for correlation clustering via sparse-dense decompositions," in *Proceedings of the 13th Conference on Innovations in Theoretical Computer Science (ITCS)*, ser. LIPIcs, vol. 215, 2022, pp. 10:1–10:20. [Online]. Available: https://doi.org/10.4230/LIPIcs.ITCS.2022.10

[31] B. Harb, S. Kannan, and A. McGregor, "Approximating the best-fit tree under l_p norms," in *8th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX-RANDOM)*, vol. 3624, 2005, pp. 123–133. [Online]. Available: https://doi.org/10.1007/11538462_11

[32] J. S. Farris, "Estimating phylogenetic trees from distance matrices," *The American Naturalist*, vol. 106, no. 951, pp. 645–668, 1972.

[33] M. Waterman, T. Smith, M. Singh, and W. Beyer, "Additive evolutionary trees," *Journal of Theoretical Biology*, vol. 64, pp. 199–213, 1977.

[34] W. H. Day, "Computational complexity of inferring phylogenies from dissimilarity matrices," *Bulletin of Mathematical Biology*, vol. 49, pp. 461–467, 1987.

[35] R. Agarwala, V. Bafna, M. Farach, M. Paterson, and M. Thorup, "On the approximability of numerical taxonomy (fitting distances by tree metrics)," *SIAM J. Comput.*, vol. 28, no. 3, pp. 1073–1085, 1999, (Announced at SODA'96). [Online]. Available: https://doi.org/10.1137/S0097539795296334

[36] M. Farach, S. Kannan, and T. J. Warnow, "A robust model for finding optimal evolutionary trees," *Algorithmica*, vol. 13, no. 1/2, pp. 155–179, 1995, (Announced at STOC'93). [Online]. Available: https://doi.org/10.1007/BF01188585

[37] V. Cohen-Addad, R. de Joannis de Verclos, and G. Lagarde, "Improving ultrametrics embeddings through coresets," in *Proceedings of the 38th International Conference on Machine Learning (ICML)*, vol. 139, 2021, pp. 2060–2068. [Online]. Available: http://proceedings.mlr.press/v139/cohen-addad21a.html

[38] V. Cohen-Addad, Karthik C. S., and G. Lagarde, "On efficient low distortion ultrametric embedding," in *Proceedings of the 37th International Conference on Machine Learning (ICML)*, vol. 119, 2020, pp. 2078–2088. [Online]. Available: http://proceedings.mlr.press/v119/cohen-addad20a.html

[39] B. Ma, L. Wang, and L. Zhang, "Fitting distances by tree metrics with increment error," *J. Comb. Optim.*, vol. 3, no. 2-3, pp. 213–225, 1999. [Online]. Available: https://doi.org/10.1023/A:1009837726913

[40] K. Dhamdhere, "Approximating additive distortion of embeddings into line metrics," in *7th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX-RANDOM)*, vol. 3122, 2004, pp. 96–104. [Online]. Available: https://doi.org/10.1007/978-3-540-27821-4_9

[41] A. C. Gilbert and L. Jain, "If it ain't broke, don't fix it: Sparse metric repair," in *55th Annual Allerton Conference on Communication, Control, and Computing, (Allerton)*, 2017, pp. 612–619. [Online]. Available: https://doi.org/10.1109/ALLERTON.2017.8262793

[42] C. Fan, A. C. Gilbert, B. Raichel, R. Sonthalia, and G. V. Buskirk, "Generalized metric repair on graphs," in *17th Scandinavian Symposium and Workshops on Algorithm Theory (SWAT)*, vol. 162, 2020, pp. 25:1–25:22. [Online]. Available: https://doi.org/10.4230/LIPIcs.SWAT.2020.25

[43] C. Fan, B. Raichel, and G. V. Buskirk, "Metric violation distance: Hardness and approximation," *Algorithmica*, vol. 84, no. 5, pp. 1441–1465, 2022, (Announced at SODA'18). [Online]. Available: https://doi.org/10.1007/s00453-022-00940-0

[44] E. Kipouridis, "Fitting tree metrics with minimum disagreements," in *31st Annual European Symposium on Algorithms, ESA 2023*, ser. LIPIcs, I. L. Gørtz, M. Farach-Colton, S. J. Puglisi, and G. Herman, Eds., vol. 274, 2023, pp. 70:1–70:10. [Online]. Available: https://doi.org/10.4230/LIPIcs.ESA.2023.70

[45] A. Carmel, D. Das, E. Kipouridis, and E. Pipis, "Fitting tree metrics and ultrametrics in data streams," in *52nd International Colloquium on Automata, Languages, and Programming, ICALP 2025*, ser. LIPIcs, vol. 334, 2025, pp. 42:1–42:21. [Online]. Available: https://doi.org/10.4230/LIPIcs.ICALP.2025.42

[46] N. Garg, V. V. Vazirani, and M. Yannakakis, "Approximate max-flow min-(multi)cut theorems and their applications," *SIAM J. Comput.*, vol. 25, no. 2, pp. 235–251, 1996, (Announced at STOC'93). [Online]. Available: https://doi.org/10.1137/S0097539793243016