Improved Approximations for Ultrametric Violation Distance

Moses Charikar[1] and Ruiquan Gao[1]

[1]Stanford University, {moses,ruiquan}@cs.stanford.edu

November 14, 2023

Abstract

We study the ULTRAMETRIC VIOLATION DISTANCE problem introduced by Cohen-Addad, Fan, Lee, and Mesmay [FOCS, 2022]. Given pairwise distances $x \in \mathbb{R}_{>0}^{\binom{[n]}{2}}$ as input, the goal is to modify the minimum number of distances so as to make it a valid ultrametric. In other words, this is the problem of fitting an ultrametric to given data, where the quality of the fit is measured by the ℓ_0 norm of the error; variants of the problem for the ℓ_∞ and ℓ_1 norms are well-studied in the literature.

Our main result is a 5-approximation algorithm for ULTRAMETRIC VIOLATION DISTANCE, improving the previous best large constant factor (≥ 1000) approximation algorithm. We give an $O(\min\{L, \log n\})$-approximation for weighted ULTRAMETRIC VIOLATION DISTANCE where the weights satisfy triangle inequality and L is the number of distinct values in the input. We also give a 16-approximation algorithm for the problem on k-partite graphs, where the input is specified on pairs of vertices that form a complete k-partite graph. All our results use a unified algorithmic framework with small modifications for the three cases.

Contents

1 Introduction

The study of ultrametrics arose in mathematics, in the work of Hensel introducing p-adic numbers (although the name was proposed later by Krasner [Kra44]). These are metric spaces where the distances between any three points i, j, k satisfy a stronger form of the triangle inequality (a.k.a., ultrametric inequality): $d(i, k) \leq \max(d(i, j), d(j, k))$. In the 1960s, ultrametrics found applications in taxonomy [Har67, Joh67, JJS67, JS71] due to the natural correspondence between the classification of objects (represented by a hierarchical clustering) and ultrametrics. (An ultrametric can be represented by the shortest path distances between leaves in a rooted tree where the root-leaf distances are equal for all leaves; the tree structure corresponds to a hierarchical clustering of the leaves). Since then, ultrametrics have been studied and used in a wide range of fields, including biology (e.g., [SS62, CE67]), physics (see the survey by [RTV86]), finance (e.g., [Man99]), and computer science.

In this work, we consider the problem of fitting an ultrametric to given measurements of distances between pairs of points – this is relevant to data-analysis settings where we have measurement noise, errors, or incomplete data. The problem of fitting ultrametrics (and tree metrics) to observations originally arose in phylogenetic analysis. It was introduced by Cavalli-Sforza and Edwards [CE67] who were interested in minimizing the ℓ_2 norm of the error, i.e., the difference between the reconstructed ultrametric/tree distances and the input. Farris [Far72] proposed the problem of minimizing the ℓ_1 norm of the error. Subsequently, these problems were shown to be NP-hard [KM86, Day87]. Since those early results, the problem of fitting ultrametrics to given data, minimizing the ℓ_p norm of the error (for various values of p) has been extensively studied in the computer science community over the past three decades (see Section 1.3). The current best-known results include a polynomial-time algorithm for the ℓ_∞ norm [FKW93], an $O(1)$-approximation for the ℓ_1-norm [CDK+21] and an $O((\log n \log \log n)^{1/p})$-approximation for ℓ_p norms, $1 < p < \infty$ [AC11].

Given the long history of the study of this family of ultrametric fitting problems, it is somewhat surprising that the natural problem of minimizing the ℓ_0 norm of the error was introduced and studied only very recently by [CFLM22]. This problem, called ULTRAMETRIC VIOLATION DISTANCE, is the focus of our work. Given pairwise distances $x \in \mathbb{R}_{>0}^{\binom{[n]}{2}}$, we study the problem of minimizing the number of entries of x we need to modify to produce a valid ultrametric. This problem is a generalization of the well-studied CORRELATION CLUSTERING problem[1], which is equivalent to instances with two distinct values in the input. Edges with the larger distance correspond to the $-$ edges and edges with the smaller distance correspond to the $+$ edges in CORRELATION CLUSTERING. An optimal ultrametric for such instances corresponds to a partition into clusters.[2] On the other hand, ULTRAMETRIC VIOLATION DISTANCE can be viewed as a collection of CORRELATION CLUSTERING problems with hierarchical structure; this viewpoint was exploited in [CFLM22].

The previous best algorithm for ULTRAMETRIC VIOLATION DISTANCE was a large constant factor approximation introduced by Cohen-Addad et al. [CFLM22][3]. Their approach involves new CORRELATION CLUSTERING algorithms whose output has additional structure, similar to the "clean" properties used by [BBC04] to obtain the first large constant factor approximation

---

[1]The input to CORRELATION CLUSTERING is a complete graph with $+$ and $-$ labels on edges, intended to represent similarity and dissimilarity of vertices. The goal is to output a disjoint partition into clusters which minimizes the number of edges that are misclassified, i.e., the number of $+$ edges that go across clusters and the number of $-$ edges that are inside clusters.

[2]Here, inter-cluster distances are set to the smaller value and intra-cluster distances are set to the larger value.

[3]the authors presented an 8,000,000 factor for their approach with unoptimized parameters, and they wrote, "it is very likely that the current approach could easily lead to a 1000-approximation but at the expense of a more tedious proof".

for CORRELATION CLUSTERING. A natural idea for improvement is to use the pivot-based approach used in all the best known algorithms for CORRELATION CLUSTERING [ACN08, CMSY15, CLN22, CLLN23]: in each round, the algorithm chooses a random unclustered vertex as the pivot and decides the set of vertices in its cluster; the algorithm proceeds until all vertices are clustered. There are two natural instantiations of this approach:

Combinatorial : put all the + neighbors of the pivot in its cluster.

LP based : compute an optimal solution for an LP relaxation of CORRELATION CLUSTERING, interpret the variables as probabilities for pairs of vertices being in the same cluster and use them for (randomly) determining cluster membership in the pivot approach.

The combinatorial version gave a 3-approximation, while the LP version initially gave a 2.5-approximation.[4] Cohen-Addad et al. [CFLM22] considered a (natural) analog of the combinatorial version for ULTRAMETRIC VIOLATION DISTANCE, where they freeze distances to the pivot and minimally modify the distances of other edges in each round, but showed a compelling negative result: even for the best choice of pivot in each round, the algorithm can only give an $\Omega(\log n)$-approximation on some instances. However, the hard instances leave open the possibility that we can incorporate the LP-based version to obtain a good constant approximation for ULTRAMETRIC VIOLATION DISTANCE. To construct the hard instances, they start with an ultrametric and carefully contaminate some edges. Here, the pivot-based approach can potentially identify the contaminated edges using an LP solution. Therefore, an interesting question to ask is

Can the pivot-based approach improve the approximation ratio for ULTRAMETRIC VIOLATION DISTANCE *by appropriately rounding the standard LP relaxation?*

Our answer to this question is positive, and we give significantly improved approximation ratios for ULTRAMETRIC VIOLATION DISTANCE and its variants, using this approach.

1.1 Our Results

Our main result is a 5-approximation for ULTRAMETRIC VIOLATION DISTANCE.

Theorem 1.1 (see Theorem 5.1). *There exists a 5-approximation algorithm for* ULTRAMETRIC VIOLATION DISTANCE.

We also study a weighted variant of ULTRAMETRIC VIOLATION DISTANCE where the weights satisfy the triangle inequality. Here, we are given an additional weight $w \in \mathbb{R}_{>0}^{\binom{[n]}{2}}$, and the objective function is the sum of $w_{i,j}$ for all edges (i, j) modified by the algorithm. We give an algorithm that (slightly) improves the previous best $O(\log n \log \log n)$-approximation [CFLM22] that also works for general weights. Interestingly, the same approximation ratio is obtained by [CFLM22] for the combinatorial pivot-based algorithm on (unweighted) ULTRAMETRIC VIOLATION DISTANCE.

Theorem 1.2 (see Theorem 6.1). *There exists an $O(\min\{L, \log n\})$-approximation algorithm for weighted* ULTRAMETRIC VIOLATION DISTANCE *where the weights satisfy the triangle inequality and L is the number of distinct values in the input.*

---

[4]later improved to a $(2.06 - \epsilon)$-approximation by [CMSY15] via a more elaborate treatment of the LP variables.

An interesting corollary of this result is a constant factor approximation for a variant of weighted CORRELATION CLUSTERING with triangle inequality constraints[5] because CORRELATION CLUSTERING problems are special cases of ULTRAMETRIC VIOLATION DISTANCE problems with $L = 2$. To the best of our knowledge, this is the first $O(1)$-approximation for this variant of CORRELATION CLUSTERING. Assuming the Unique Games Conjecture, this result also separates weighted CORRELATION CLUSTERING with and without triangle inequality constraints.

Corollary 1.3. *There exists an $O(1)$-approximation algorithm for weighted* CORRELATION CLUSTERING *where the weights satisfy the triangle inequality.*

We consider another interesting variant of (unweighted) ULTRAMETRIC VIOLATION DISTANCE with incomplete information, where the input distances are specified only on a subset of the edges. In the absence of any structure on the specified edges, a special case is unweighted CORRELATION CLUSTERING on general graphs, which does not have an $O(1)$-approximation assuming the Unique Games Conjecture [DEFI06, CKK$^+$06]. We give a 16-approximation if the specified edges form a complete k-partite graph. A similar setting has been studied in the literature on CORRELATION CLUSTERING [Ami04, AALZ12, CMSY15], culminating in a ratio of 3 on complete k-partite graphs.

Theorem 1.4 (see Theorem 7.1)**.** *There exists a 16-approximation algorithm for* ULTRAMETRIC VIOLATION DISTANCE *on complete k-partite graphs.*

1.2 Techniques

Pivot-based LP rounding algorithm. As is the case for CORRELATION CLUSTERING (e.g., [CGW05, ACN08, CMSY15]) and for fitting ultrametric with ℓ_1 objective [AC11, CDK$^+$21], a standard LP relaxation was introduced by Cohen-Addad et al. [CFLM22] to capture the objective of (weighted) ULTRAMETRIC VIOLATION DISTANCE. This LP can be viewed as a hierarchical generalization of the LP for CORRELATION CLUSTERING, in which variables $y_\ell(i, j)$ are defined for every edge $(i, j) \in \binom{[n]}{2}$ at every level $\ell \in [L]$. Variables at the same level are subject to the constraints of CORRELATION CLUSTERING (constraint (1) below), and variables for the same edge are monotonic with respect to the level (constraint (2) below). Formally, the LP relaxation is as follows, where we define $y_0(i, j) = 0$ for each edge $(i, j) \in \binom{[n]}{2}$ and define $\tilde{\ell}(i, j)$ as the input distance level of (i, j). We refer the reader to Section 2.2 for more discussion of this LP relaxation.

$$\text{minimize} \quad \sum_{\substack{i \neq j \in [n]}} w(i, j) \cdot (1 - y_{\tilde{\ell}(i,j)}(i, j) + y_{\tilde{\ell}(i,j)-1}(i, j)) \quad \text{(UMVD LP)}$$

$$\text{subject to} \quad y_\ell(i, j) \leq y_\ell(i, k) + y_\ell(k, j) \qquad \forall \ell \in [L], i, j, k \in [n] \quad (1)$$

$$y_{\ell-1}(i, j) \leq y_\ell(i, j) \qquad \forall \ell \in [L], i, j \in [n] \quad (2)$$

$$y_\ell(i, j) \in [0, 1] \qquad \forall \ell \in [L], i, j \in [n] \quad (3)$$

Previous works [AC11, CDK$^+$21, CFLM22] use this (or a very similar) LP relaxation as a hierarchy of LP solutions for the CORRELATION CLUSTERING problem. In this work, we build our algorithm based on an alternative interpretation of the relaxation, using the LP variables for an edge as a specification of a probability distribution on the distance value for that edge because we can always

---

[5]We note that this name is also used in previous works but for different settings. Previously, weighted CORRELATION CLUSTERING is defined by two weights $w_{i,j}^+$ and $w_{i,j}^-$ for each edge $(i, j) \in \binom{[n]}{2}$. The objective function of the problem is the sum of $w_{i,j}^+$ for all $-$ edges in the output and $w_{i,j}^-$ for all $+$ edges in the output. [ACN08] defines the triangle inequality constraints to be $w_{i,j}^- + w_{j,k}^- \geq w_{i,k}^-$ for any tuple (i, j, k). [ZW09]'s definition adds an additional constraint $w_{i,j}^+ + w_{j,k}^- \geq w_{i,k}^+$ for each tuple. Both settings are incomparable with ours.

find some optimal solution with $y_L(i,j) = 1$. This interpretation enables us to naturally use pivot-based algorithms to round the variables. More specifically, the algorithm starts with a recursive call involving all vertices. In each recursive call, we choose a random pivot from the vertex set of the call and decide the distance of edges that connect to the pivot according to the LP solution. To determine these distances, we consider two different known rounding schemes:

Randomized rounding: we interpret the LP variables $\{y_\ell(i,j)\}_{\ell \in [L]}$ for each edge (i,j) as a complementary cumulative distribution function (CCDF) for the distance of (i,j) and sample from the distribution; a similar method was introduced for CORRELATION CLUSTERING [ACN08] .

Deterministic rounding: we select the most probable outcome of the above distribution, breaking ties arbitrarily; a similar method can be found in [CGW05] for CORRELATION CLUSTERING.

Our algorithm uses a natural combination of these two schemes: if an edge's distribution produces a specific distance value with a high probability (close to 1), we call it a *deterministic edge* and use the deterministic rounding; otherwise, we call it a *random edge* and use the randomized rounding instead. After determining the distances of the edges connected to the pivot, we modify the distances of the other edges minimally so that the ultrametric inequalities involving the pivot are satisfied. Finally, we partition the non-pivot vertices into sets by grouping vertices with the same distance from the pivot into the same set and recursively solve the smaller instances with these vertex sets. These latter two steps can also be found in the pivot-based algorithm of [CFLM22] that gives an $O(\min\{L, \log n\})$-approximation for ULTRAMETRIC VIOLATION DISTANCE.

One key benefit of distinguishing deterministic edges and random edges is that the ultrametric inequality can be automatically satisfied in a triangle with three deterministic edges by the deterministic rounding and appropriate choices of the parameters. This property and its generalizations are frequently used in our analysis. Next, we present an example demonstrating how this property is established.

Example. Consider an ULTRAMETRIC VIOLATION DISTANCE instance with $L = 2$ and any three distinct vertices $i, j, k \in [n]$. Because we can w.l.o.g. assume $y_L(u, v) = 1$ for any distinct $u, v \in [n]$, for simplicity, we can use a, b, c to denote the only variables $y_1(i,j), y_1(i,k)$ and $y_1(j,k)$. Consider that we deterministically round the distance of (i,j) to the larger input distance if $a > \frac{2}{3}$ and deterministically round the distance of (i,j) to the smaller input distance if $a < \frac{1}{3}$. Suppose that the same rounding is done for (i,k) and (j,k). If all three edges are deterministically rounded, the only possible violation of the ultrametric inequality occurs when exactly one of the edges is rounded to the larger input distance. However, this implies that one of a, b, c is strictly greater than $\frac{2}{3}$, while the other two are strictly less than $\frac{1}{3}$, violating the triangle inequality constraint in the LP for a, b, c.

Overcoming the $\Omega(\log n)$ barrier. Our main technical contribution is an improved triple-based analysis framework to overcome the previous $\Omega(\log n)$ barrier. We start with a simple triple-based analysis and explain its pitfalls.

A simple triple-based approach: we upper bound the number of modified edges by the total number of times the distances on the edges are modified by the algorithm. The triple-based approach bounds the cost – modifications incurred for an edge (j, k) when i is chosen as a pivot – by charging the triangle (i, j, k), and the LP contributions of all three edges of a triangle are used to pay for the expected cost charged to the triangle.

This approach fails for some "bad" triangles, resulting in the LP contributions of some edges being charged $\Omega(\log n)$ times. Consider a triangle (i, j, k). Suppose in a recursive call where i is selected as the pivot, the two pivot edges (i, j) and (i, k) are both deterministic and are going to be fixed to the same distance. If the current distance of (j, k) is strictly greater than the distance (i, j) will be fixed to, then the third edge (j, k) will be modified by the algorithm in the recursive call[6]. If the LP contributions of (i, j) and (i, k) are very small (e.g., when the input distance of (i, j) and (j, k) equals the distance they are fixed in this recursive call), we are forced to charge an $\Omega(1)$ multiple of the LP contribution of (j, k) to pay for the modification of (j, k) in this triangle. At the same time, the pair j, k will be grouped together and get involved in a subsequent recursive call[7]. Again, in subsequent recursive calls, the edge (j, k) can be similarly charged by these bad triangles. As analyzed in [CFLM22], it can cause edge (j, k) to be charged by $\Omega(\log n)$ such bad triangles in expectation, resulting in an $\Omega(\log n)$ bound on the approximation ratio for the simple triple-based approach.

To address this issue, we make one key observation about these triangles:

all bad triangles have at least one edge with a high LP contribution.

For example, in the above example, the LP contribution of (j, k) must be $\Omega(1)$. This suggests a different way to upper bound the number of modified edges to handle bad triangles: *we simply upper bound the cost for edges with high LP contribution by 1.*

Thus, we focus on upper bounding the total number of modifications on edges with low LP contributions in the triple-based analysis; the simple triple-based analysis is sufficient for these edges. By appropriately choosing the threshold for low- vs. high-cost, we can prove that edges with low LP contributions are always deterministic in the algorithm, simplifying the proof. In Section 4 and 5, we refine the analysis further to obtain an approximation ratio of 5. The bad triangles belong to the "(d, d, r)-same-triangles" described there.

Comparison between our rounding scheme vs purely randomized rounding. We believe that our improved triple-based analysis framework can also be used to establish a constant approximation ratio for the purely randomized rounding scheme. It appears that this would not give a bound better than 5, and the analysis would be more tedious and less straightforward. In addition, our improved analysis framework can be directly applied to analyze our algorithm on the weighted variant of ULTRAMETRIC VIOLATION DISTANCE where the weights satisfy the triangle inequality constraints; we obtain an $O(\min\{L, \log n\})$ approximation. However, the purely randomized rounding scheme cannot be analyzed similarly to obtain this bound because its analysis for the unweighted cases heavily rely on the fact that all edges contribute equally to the objective. Example A.1 suggests that, even when $L = 2$ (the CORRELATION CLUSTERING instances), the purely randomized rounding scheme may not yield a comparable approximation in the (improved) triple-based analysis framework.

One step further: an improved algorithm for k-partite cases. Our improved triple-based analysis framework still runs into the $\Omega(\log n)$ issue (when $L = \Omega(\log n)$) for the k-partite case (a special case of the weighted case with triangle inequality) and new ideas are needed. For example, consider a triangle (i, j, u), where (i, j) and (j, u) are specified, but (i, u) is not specified. The LP contribution of (i, u) is therefore 0. In the previous analysis on complete cases, if (j, u) is a random edge, we need to use the LP contributions of both (i, j) and (i, u) to pay for the cost on (i, j)

---

[6]even if we use the purely randomized rounding scheme, this could still happen with probability $1 - o(1)$

[7]this also possibly happens with probability $1 - o(1)$ if we use the purely randomized rounding scheme.

when (i, j, u) is a bad triangle and u is chosen as the pivot. However, in the k-partite case, we can no longer rely on the LP contribution of (i, u) to pay for this type of cost. To resolve this issue, we use a (slightly) different rounding approach that treats specified edges and unspecified edges differently. First, we use a larger threshold for unspecified edges. For example, we may define a specified edge as deterministic if its distance distribution (given by the LP) puts a probability mass greater than $5/8$ on some distance value, while we define an unspecified edge as deterministic only if the probability is greater than $3/4$. Then, we employ rejection sampling on the specified random edges: with the same parameters mentioned above, if the sampled distance is in the largest quantile of the distribution, we reject the outcome and repeat the process until we accept. Using the fact that the LP variables at each level satisfy the triangle inequality constraint, the expected number of modifications on (i, j) in this triangle can be upper bounded by $O(1)$ times its LP contribution. For a formal definition of this algorithm, see Section 3, and for a detailed analysis, see Section 7 and Appendix B.7.

We note that the purely randomized rounding scheme does not give a constant approximation ratio for k-partite instances via triple-based analysis, even in the special case of CORRELATION CLUSTERING on complete k-partite graphs. Example A.1 shows the failure of fully randomized rounding for the k-partite case. This further highlights the advantage of our new rounding scheme.

1.3 Further Related Work

Fitting ultrametrics with other objectives. The problem of fitting ultrametrics for ℓ_∞ and ℓ_1 norms is well studied in the literature. Farach et al. [FKW93] give a polynomial-time exact algorithm for the ℓ_∞ objective. However, for the ℓ_p objective ($1 \le p < \infty$), its APX-hardness can be derived from the APX-hardness of CORRELATION CLUSTERING [CGW05]. There is a line of works studying its approximation algorithms [HKM05, AC11, CDK$^+$21]. The state-of-the-art algorithms are an $O(1)$-approximation algorithm for the ℓ_1 norm and $O((\log n \log \log n)^{1/p})$-approximations for ℓ_p norms ($1 < p < \infty$). Many other objectives and assumptions are also considered in the literature of fitting ultrametrics, including the maximization version [DPS$^+$13, CFLM22], the outlier deletion version [SWW17], multiplicative distortion version [DPS15] and assuming the input is in a high-dimensional Euclidean space [CKL20, CDL21].

Fitting metrics and tree metrics. Brickell et al. [BDST08] formulated the problem of fitting metrics to a given data with ℓ_p objective for $1 \le p \le \infty$, which can be solved exactly in polynomial-time via linear or convex programs. The ℓ_0 version was recently introduced and studied [GJ17, FRVB18, FGR$^+$20, CFLM22], culminating in an $O(\log n)$-approximation algorithm that runs in $O(n^3)$ time.

On the other hand, Cavalli-Sforza and Edwards [CE67] introduced the tree fitting problem. If there is a tree metric that exactly fits the data, the structure can be found in polynomial time [WSSB77]. However, if there is no such tree metric, the problem for any ℓ_p objective ($0 \le p \le \infty$) is APX-hard [ABF$^+$98, CGW05, Kip23]. Recent work of Kipouridis [Kip23] shows that any ρ-approximation of Ultrametric Violation Distance can be converted to a 6ρ-approximation for tree metric fiiting with ℓ_0 objective. Combined with our result, there is automatically a 30-approximation. Furthermore, Agarwala et al. [ABF$^+$98] and Cohen-Addad et al. [CDK$^+$21] give reductions from tree metrics to ultrametrics with ℓ_p objectives ($1 \le p \le \infty$), only losing a constant factor. The state-of-the-art algorithms are then a 3-approximation for ℓ_∞ norm, an $O(1)$-approximation algorithm for ℓ_1 norm, and $O((\log n \log \log n)^{1/p})$-approximations for ℓ_p norms ($1 < p < \infty$).

Correlation clustering. As mentioned before, CORRELATION CLUSTERING has been extensively studied. Bansal, Blum and Chawla [BBC04] introduced the problem and gave a constant factor approximation. Later, it was improved by [CGW05, ACN08, CMSY15, CLN22, CLLN23]. The current best algorithm is a 1.73-approximation based on the Sherali-Adams hierarchy. Furthermore, CORRELATION CLUSTERING has been proven to be APX-hard [CGW05]. Various variants of CORRELATION CLUSTERING have also been studied. Among those, works regarding weighted cases include the general version [DEFI06, CGW05], the version with probability constraints [GMT07, ACN08, CMSY15] and the version for asymmetric error costs [JKMM20].

1.4 Organization

In Section 2, we present the problem formulations and discuss how the linear program captures the objective of (weighted) ULTRAMETRIC VIOLATION DISTANCE. In Section 3, we formally describe a unified algorithm for all three cases that we study. In Section 4, we present our analytical framework. Specifically, in Section 4.2, we present our improved triple-based analysis for ULTRAMETRIC VIOLATION DISTANCE and establish two key lemmas that are used for the three cases. In Sections 5, 6 and 7, we respectively prove the approximation ratios for the three cases.

2 Preliminaries

We use $[n]$ to denote the set $\{1, 2, \cdots, n\}$. We use $\mathbb{R}_{\geq 0}$ and $\mathbb{R}_{>0}$ to denote non-negative and positive real numbers. For any set $S \subseteq [n]$, we use $\binom{S}{2}$ to denote the set of pairs of distinct elements in S, i.e., $\binom{S}{2} = \{(i,j) : i, j \in S, i < j\}$. We use x^+ to denote the positive part $\max(x, 0)$. We use a tuple $t = (i, j, k)$ to denote a triangle consisting of three distinct vertices i, j, k and all three edges between them.

2.1 The Ultrametric Violation Distance Problem

The ULTRAMETRIC VIOLATION DISTANCE problem takes as input two disjoint sets of edges $E, E_\varnothing$ such that $E \cup E_\varnothing = \binom{[n]}{2}$ and distances $x_{\text{in}} \in \mathbb{R}_{>0}^E$ that are specified on the edges in E. The goal is to find some ultrametric $x \in \mathbb{R}_{>0}^{\binom{[n]}{2}}$ such that its disagreement with the input $\sum_{(i,j)\in E} \mathbf{1}(x(i,j) \neq x_{\text{in}}(i,j))$ is minimized. We use L to denote the number of distinct distances in x_{in} and use $d_1 > d_2 > \cdots > d_L$ to denote the distinct distances in the input. For each edge $(i, j) \in E$, we use $\tilde{\ell}(i,j)$ to denote the input distance *level* of $x_{\text{in}}(i,j)$, which satisfies $x_{\text{in}}(i,j) = d_{\tilde{\ell}(i,j)}$.

The weighted version of ULTRAMETRIC VIOLATION DISTANCE takes input additional weights $w \in \mathbb{R}_{\geq 0}^{\binom{[n]}{2}}$. In addition, in the weighted version, because of the existence of zero weights, we can w.l.o.g. assume the inputs are specified on all the edges, i.e., $E = \binom{[n]}{2}$. The goal is to minimize $\sum_{(i,j)\in E} w(i,j) \cdot \mathbf{1}(x(i,j) \neq x_{\text{in}}(i,j))$. In the weighted version with triangle inequality constraints, we assume $\forall i, j, k \in [n], w(i,j) + w(j,k) \geq w(i,j)$.

Moreover, the unweighted ULTRAMETRIC VIOLATION DISTANCE on complete k-partite graphs can be viewed as a special case of the weighted version with triangle inequality constraints: we can set $w(i,j) = 1$ for $(i,j) \in E$ and set $w(i,j) = 0$ for $(i,j) \in E_\varnothing$ in the weighted version, and the complete k-partite graph implies that w satisfies the triangle inequality constraint.

2.2 LP Definitions for Ultrametric Violation Distance

We can formulate a lower bound for (weighted) ULTRAMETRIC VIOLATION DISTANCE by an integer linear program. For every edge $(i,j) \in \binom{[n]}{2}$ and every level $\ell \in [L]$, the integer program has a variable $y_\ell(i,j) \in \{0,1\}$ characterizing whether the distance of (i,j) is $\geq d_\ell$: $y_\ell(i,j) = 1$ if the output x satisfies $x(i,j) \geq d_\ell$; $y_\ell(i,j) = 0$ if the output x satisfies $x(i,j) < d_\ell$. We shall use $y_\ell(i,j)$ and $y_\ell(j,i)$ to denote the same variable for each $\ell \in [L], (i,j) \in \binom{[n]}{2}$. For convenience, we define $y_\ell(i,j) \stackrel{\text{def}}{=} 0$ for $\ell = 0$ or $i = j$, and use $\Delta y_\ell(i,j) \stackrel{\text{def}}{=} y_\ell(i,j) - y_{\ell-1}(i,j)$. To guarantee that we can recover some $x \in \mathbb{R}_{>0}^{\binom{[n]}{2}}$ from the variables, we can use constraints requiring that $y_\ell(i,j)$ is increasing with respect to ℓ for every (i,j) (constraint (2)). To further guarantee that the recovered x forms an ultrametric, we can use triangle inequality constraints on every level (constraint (1)) to ensure that the output forms a correlation clustering solution on every level. The objective of (weighted) ULTRAMETRIC VIOLATION DISTANCE is to minimize the sum of weights of disagreement between the input and the output ultrametric. Note that for each $(i,j) \in \binom{[n]}{2}$ and $\ell \in [L]$, $x(i,j) = d_\ell$ if and only if $y_\ell(i,j) = 1$ and $y_{\ell-1}(i,j) = 0$. Recall that we define $\tilde{\ell}(i,j)$ as the distance level of $x_{\text{in}}(i,j)$. Therefore, we can write the objective as $\sum_{(i,j) \in E} w(i,j) \cdot (1 - \Delta y_{\tilde{\ell}(i,j)}(i,j))$.

By relaxing the integer program, we can obtain the LP relaxation (UMVD LP), which is equivalent to the standard LP definition introduced by [CFLM22] for the weighted ULTRAMETRIC VIO-LATION DISTANCE. To ensure convenient access, we are restating the LP as follows:

$$
\begin{aligned}
\text{minimize} \quad & \sum_{i \neq j \in [n]} w(i,j) \cdot (1 - y_{\tilde{\ell}(i,j)}(i,j) + y_{\tilde{\ell}(i,j)-1}(i,j)) \\
\text{subject to} \quad & y_\ell(i,j) \leq y_\ell(i,k) + y_\ell(k,j) && \forall \ell \in [L], i,j,k \in [n] \\
& y_{\ell-1}(i,j) \leq y_\ell(i,j) && \forall \ell \in [L], i,j \in [n] \\
& y_\ell(i,j) \in [0,1] && \forall \ell \in [L], i,j \in [n]
\end{aligned}
$$

It can be observed that any optimal solution of (UMVD LP) can be easily modified to an optimal solution satisfying $y_L(i,j) = 1$ for any $(i,j) \in E$ because these variables have non-positive coefficients in the objective of (UMVD LP). Hence, we can w.l.o.g. assume $y_L(i,j) = 1$ for any $(i,j) \in E$ in the rest of the paper.

3 Pivot-based Algorithm

In this section, we present our pivot-based LP rounding algorithm for (weighted) ULTRAMETRIC VI-OLATION DISTANCE (Algorithm 1). The algorithmic framework follows the pivot-based algorithms of [AC11, CFLM22] and utilizes the optimal LP solutions of (UMVD LP). More specifically, the algorithm is recursive, parameterized by two constants $\alpha \in [0, 0.5]$ and $\beta \in [0, 1]$, and has access to all distinct distances $d_1 > d_2 > \cdots > d_L$ of the input and an optimal solution y^* satisfying $y_L^*(i,j) = 1$ for each $(i,j) \in \binom{[n]}{2}$. Each recursive call takes as input a subset of vertices $V \subseteq [n]$, the distances x between vertices in V, and an upper bound level $u \in [L]$ representing that the output distances of the call should be upper bounded by d_u. To run the algorithm, we call LP-UMVD-PIVOT($[n], x_{\text{in}}, 1$) and use the distances it returns as the output ultrametric.

In each recursive call, we randomly select a pivot vertex $i \in V$ and compute a truncated LP solution y in which all $y_\ell(i,j)$s are set to 0 for any $\ell < u$. According to the constraints of (UMVD LP), the truncated LP solution satisfies the following lemma, which enables us to view $\Delta y_\ell(i,j)$s as a probability distribution of the distance of each edge (i,j).

Algorithm 1: LP-UMVD-PIVOT(V, x, u)

> **Input**   : $V \subseteq [n]$, $u \in [L]$, $x \in \mathbb{R}_{\geq 0}^{\binom{V}{2}}$
>
> **Output**   : $x' \in \mathbb{R}_{\geq 0}^{\binom{V}{2}}$
>
> **Parameter:** $\alpha \in [\frac{1}{2}, 1]$

1 **if** $|V| \leq 2$ **or** $u = L$ **then**
2 $\quad$ **return** x

3 $i \leftarrow$ a vertex in V uniformly at random
4 $y \leftarrow y^*$
5 **for** $i, j \in \binom{V}{2}$ **do**
6 $\quad$ **for** $\ell \in [u-1]$ **do**
7 $\quad\quad$ $y_\ell(i,j) \leftarrow 0$

8 **for** $j \in V \setminus \{i\}$ **do**
9 $\quad$ $\ell^* \leftarrow \arg\max_{\ell \in [L]} \ \Delta y_\ell(i,j)$ $\qquad\qquad\qquad$ `// break ties arbitrarily`
10 $\quad$ **if** $(i,j) \in E$ **then**
11 $\quad\quad$ **if** $\Delta y_{\ell^*}(i,j) > 1 - \alpha$ **then**
12 $\quad\quad\quad$ $x'(i,j) \leftarrow d_{\ell^*}$ $\qquad\qquad\qquad$ `// deterministic edges in E`
13 $\quad\quad$ **else**
14 $\quad\quad\quad$ $x'(i,j) \leftarrow$ sampled according to Eqn. (4) $\qquad$ `// random edges in E`
15 $\quad$ **else**
16 $\quad\quad$ **if** $\Delta y_{\ell^*}(i,j) > 1 - \alpha\beta$ **then**
17 $\quad\quad\quad$ $x'(i,j) \leftarrow d_{\ell^*}$ $\qquad\qquad\qquad$ `// deterministic edges in E_∅`
18 $\quad\quad$ **else**
19 $\quad\quad\quad$ $x'(i,j) \leftarrow$ sampled according to Eqn. (5) $\qquad$ `// random edges in E_∅`

20 **for** $(j,k) \in \binom{V \setminus \{i\}}{2}$ **do**
21 $\quad$ **if** $x'(i,j) = x'(i,k)$ **then**
22 $\quad\quad$ $x'(j,k) \leftarrow \min\{x(j,k), x'(i,j)\}$
23 $\quad$ **else**
24 $\quad\quad$ $x'(j,k) \leftarrow \max\{x'(i,j), x'(i,k)\}$

25 **for** $\ell \in \{u, u+1, \cdots, L\}$ **do**
26 $\quad$ $V_\ell \leftarrow \{j \in V : x'(i,j) = d_\ell\}$
27 $\quad$ $x'|_{\binom{V_\ell}{2}} \leftarrow$ LP-UMVD-PIVOT$\left(V_\ell, x'|_{\binom{V_\ell}{2}}, \ell\right)$ $\quad$ `// x'|_(V_ℓ/2) is x' restricted to (V_ℓ/2)`
28 **return** x'

Lemma 3.1. *In each recursive call, the truncated LP solution y is a feasible solution of* (UMVD LP) *for the vertices in V and satisfies $y_L(i,j) = 1$ for any $(i,j) \in \binom{V}{2}$. In particular, $\sum_{\ell \in [L]} \Delta y_\ell(i,j) = 1$.*

We call edges having the pivot i as an endpoint *pivot edges* and call the rest of the edges *non-pivot edges*. Then, we determine the distances on the pivot edges as follows. For each non-pivot vertex $j \in V \setminus \{i\}$, we pick a *dominant level* $\ell^*(y, i, j) \in \arg\max_{\ell \in [L]} \ \Delta y_\ell(i,j)$. If the dominant level is not unique, we break ties arbitrarily. With the dominant level, we divide the edges into the following four classes and set the distances for the pivot edges as follows. Also, see Figure 1 for an
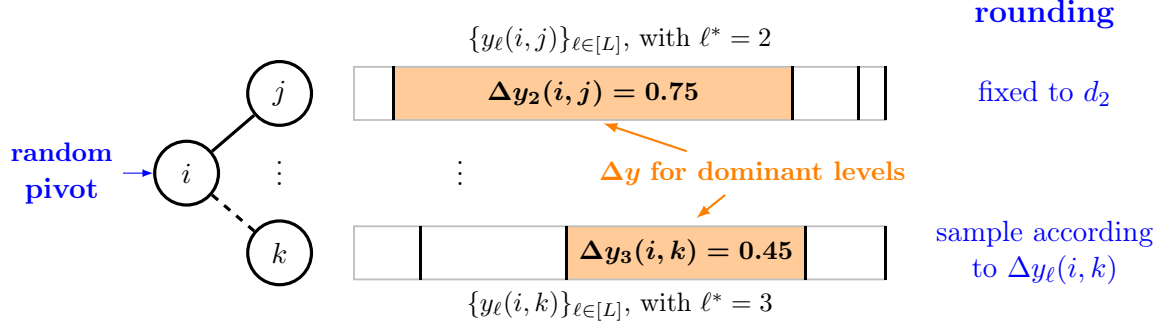
9

Figure 1: Example of our rounding process on complete graphs, where we choose $\alpha = 0.4$ and $\beta = 0$.

intuitive example of how the rounding process works on complete graphs under $\beta = 0$.

Random edges in E. $(i, j) \in E$ falls into this class if $\Delta y_{\ell^*(y,i,j)}(i,j) \leq 1 - \alpha$. When it is a pivot edge, its distance $x'(i,j)$ is random and follows the distribution (technically, the CCDF):

$$\forall \ell \in [L], \quad \Pr[x'(i,j) \geq d_\ell] = \frac{1}{1 - \alpha\beta} \cdot (y_\ell(i,j) - \alpha\beta)^+ . \tag{4}$$

Deterministic edges in E. $(i, j) \in E$ falls into this class $\Delta y_{\ell^*(y,i,j)}(i,j) > 1 - \alpha$. When it is a pivot edge, its distance $x'(i,j)$ is fixed to its dominant distance level $d_{\ell^*(y,i,j)}$.

Random edges in $E_\varnothing$. $(i, j) \in E_\varnothing$ falls into this class if $\Delta y_{\ell^*(y,i,j)}(i,j) \leq 1 - \alpha\beta$. When it is a pivot edge, its distance $x'(i,j)$ is random and follows the distribution (i.e., the CCDF):

$$\forall \ell \in [L], \quad \Pr[x'(i,j) \geq d_\ell] = y_\ell(i,j) . \tag{5}$$

Deterministic edges in $E_\varnothing$. $(i, j) \in E_\varnothing$ falls into this class if $\Delta y_{\ell^*(y,i,j)}(i,j) > 1 - \alpha\beta$. When it is a pivot edge, its distance $x'(i,j)$ is fixed to its dominant distance level $d_{\ell^*(y,i,j)}$.

Remark 3.2. When $\beta = 0$, Eqn. (4) becomes $\forall \ell \in [L]$, $\Pr[x'(i,j) \geq d_\ell] = y_\ell(i,j)$, which is an analog of the purely randomized scheme introduced by [ACN08].

After determining the pivot edges, we partition the non-pivot vertices according to the distances between them and the pivot. Suppose V_ℓ denotes the set of non-pivot vertices $j \in V$ such that $x'(i,j) = d_\ell$. For any non-pivot edge (j,k) whose endpoints j, k are partitioned into the same set V_ℓ, we minimally fix its distance by setting the distance to $\min(x(j,k), d_\ell)$. For any non-pivot edge (j,k) whose endpoints j, k are partitioned into different sets, say $j \in V_\ell, k \in V_{\ell'}$, we minimally fix its distance by setting the distance to $\max(d_\ell, d_{\ell'})$. Note that after this step, we can guarantee that the ultrametric inequality is satisfied for every triangle that involves i. Finally, in the recursive call, for each set V_ℓ, we call our algorithm on V_ℓ with an upper bound of d_ℓ on the output distance and modify the distances between vertices in V_ℓ to the output of the call.

Compared to the pivot-based algorithms of [AC11, CFLM22], our algorithm differs only in how the distances on the edges incident to the pivot are determined. Similar to their analysis, we can show that the algorithm outputs an ultrametric. For the sake of completeness, we present the proof in Appendix B.1.

Lemma 3.3. *Algorithm 1 outputs an ultrametric in polynomial time.*

10

4   Analytical Framework

In this section, we present the analytical framework for Algorithm 1 that will be used to obtain all the approximation ratios.

4.1   Definitions and Basic Properties

Notice that the recursive calls in the algorithm form a tree structure. We use the *root call* to refer to the recursive call LP-UMVD-PIVOT$([n], x_{\text{in}}, 1)$ at the beginning of the algorithm. For any recursive call, we use *child call*s to refer to the recursive calls on line 27 and use *parent call* for vice versa. If a recursive call does not have child calls, we say it is a *leaf call*.

With an abuse of notation, for any LP solution y, we use $y_{\ell^*(u,v)}(i,j)$ to denote $y_{\ell^*(y,u,v)}(i,j)$. Let $c^*(i,j)$ denote the LP cost of edge (i,j) in y^*, i.e., $c^*(i,j) \overset{\text{def}}{=} 1 - \Delta y^*_{\tilde{\ell}(i,j)}(i,j)$. With this definition, we can rewrite the LP objective (UMVD LP) and thus lower bound the total number of distances modified in the optimal solution OPT as follows:

$$\text{OPT} \geq \sum_{(i,j) \in E} w(i,j) \cdot c^*(i,j) \ . \tag{6}$$

Based on the LP costs, we shall classify the edges in E into the following two classes:

low-cost edges include all edges $(i,j) \in E$ satisfying $c^*(i,j) < \alpha$, and

high-cost edges include all edges $(i,j) \in E$ satisfying $c^*(i,j) \geq \alpha$.

We shall use E_{L} and E_{H} to denote the set of low-cost and high-cost edges. In the optimal solution of the LP, low-cost edges satisfy

$$\Delta y^*_{\ell^*(i,j)}(i,j) \geq \Delta y^*_{\tilde{\ell}(i,j)}(i,j) = 1 - c^*(i,j) > 1 - \alpha \ ,$$

and thus are deterministic in the root call. However, this may not be the case for high-cost edges. We shall subdivide the high-cost edges into the following two classes based on whether they are deterministic in the root call:

initially deterministic edges include all edges $(i,j) \in E_{\mathsf{H}}$ such that $\Delta y^*_{\ell^*(i,j)}(i,j) > 1 - \alpha$,

initially random edges include all edges $(i,j) \in E_{\mathsf{H}}$ such that $\Delta y^*_{\ell^*(i,j)}(i,j) \leq 1 - \alpha$.

We shall use E_{HD} and E_{HR} to denote the set of high-cost edges that are initially deterministic and that are initially random. For the high-cost edges that are initially deterministic, we have the following tighter lower bound for their LP costs. We defer the proof to Appendix B.2.

Lemma 4.1. *For any edge $(i,j) \in E_{\mathsf{HD}}$, $c^*(i,j) > 1 - \alpha$.*

Next, we present two basic properties of Algorithm 1. The first property claims that the increments of y on the dominant levels $\Delta y_{\ell^*(i,j)}$ are increasing from the root call to the leaf calls. We defer the proof to Appendix B.3.

Lemma 4.2. *Suppose y is the truncated LP solution in a recursive call with vertex set V. Suppose y' is the truncated LP solution in one of its child calls with vertex set $V' \subseteq V$. For each edge $(i,j) \in \binom{V'}{2}$, $\Delta y_{\ell^*(i,j)}(i,j) \leq \Delta y'_{\ell^*(i,j)}(i,j)$. In particular, for any recursive call with vertex set V and truncated LP solution y, $\forall (i,j) \in \binom{V}{2}, \Delta y_{\ell^*(i,j)}(i,j) \geq \Delta y^*_{\ell^*(i,j)}(i,j)$.*

11

From the above lemma and the fact that any edge $(i,j) \in E_L \cup E_{HD}$ satisfies $\Delta y^*_{\ell^*(i,j)}(i,j) > 1 - \alpha$, we obtain the following corollary.

Corollary 4.3. *In any recursive call, any low-cost (or initially deterministic high-cost) edge is deterministic.*

Because we choose $\alpha \le \frac{1}{2}$, we can give the second property that low-cost edges are modified only when they are non-pivot edges. We defer the proof to Appendix B.4.

Lemma 4.4. *In any recursive call with vertex set V and truncated LP solution y, for any edge $(i,j) \in E_L \cap \binom{V}{2}$, the input distance satisfies $x(i,j) = d_{\ell^*(y,i,j)}$. In particular, when an edge $(i,j) \in E_L$ appears as a pivot edge, its distance $x(i,j)$ is not modified in the recursive call.*

4.2 Towards the Approximation Ratios: Improved Triple-based Analysis for Ultrametric Violation Distance

In this subsection, we develop two key lemmas, 4.5 and 4.6, to prove the approximation ratios for the three cases and any constant $\alpha \in (0, \frac{1}{2}]$. By contrast to the triple-based analysis of the previous works (e.g., [ACN08, CMSY15]), we separately upper bound the cost of the algorithm on different classes of edges: for the high-cost edges, we upper bound the algorithm's cost by 1 and further by a constant multiple of its LP cost; for low-cost edges, we shall utilize the previous triple-based analysis. In the triple-based analysis, Lemma 4.5 establishes the approximation ratio given a scheme charging the non-pivot edges so that: (1) in each triangle $t = (i,j,k)$ the expected number of modifications on low-cost edges in t can be upper bounded by the expected total charges in t; and (2) the expected total charges on each edge can be upper bounded. In addition, Lemma 4.6 establishes an upper bound for the expected total charges on each edge by the probability that the edge can no longer be charged in the recursive algorithm (conditioned on any possible charging value).

Let ALG be the total number of distances modified by the algorithm. Because of the definition of high-cost edges and Lemma 4.1, we can upper bound its expectation as follows:

$$
\begin{aligned}
\mathbb{E}[\text{ALG}] &= \sum_{(i,j) \in E} w(i,j) \cdot \Pr[(i,j) \text{ is modified}] \\
&\le \sum_{(i,j) \in E_L} w(i,j) \cdot \Pr[(i,j) \text{ is modified}] + \sum_{(i,j) \in E_{HD}} w(i,j) + \sum_{(i,j) \in E_{HR}} w(i,j) \\
&\le \sum_{(i,j) \in E_L} w(i,j) \cdot \mathbb{E}[\#\text{times } (i,j) \text{ is modified}] + \frac{1}{1-\alpha} \cdot \sum_{(i,j) \in E_{HD}} w(i,j) \cdot c^*(i,j) \\
&\qquad\qquad + \frac{1}{\alpha} \cdot \sum_{(i,j) \in E_{HR}} w(i,j) \cdot c^*(i,j) .
\end{aligned}
\tag{7}
$$

Recall the lower bound for the optimal solution (6). To prove the approximation ratio, we only need to upper bound the number of modifications the algorithm makes on those low-cost edges by the LP value of y^*. Because of Lemma 4.4, low-cost edges are only modified when they appear as non-pivot edges in the recursive calls. Hence, we only need to consider the modifications on non-pivot edges.

For each triangle $t = (i,j,k)$, let $\mathcal{A}_t$ denote the event that i,j,k appear in the same recursive call where one of them is chosen as a pivot. For each triangle t and each edge $(i,j) \in t$, we define $M_{i,j,t} = 1$ if $\mathcal{A}_t$ happens, $(i,j) \in E_L$, and (i,j) is modified in the event $\mathcal{A}_t$. Otherwise, we define $M_{i,j,t} = 0$. With this definition, we can rewrite the total number of modifications on low-cost edges

as

$$\sum_{(i,j)\in E_\mathsf{L}} w(i,j)\cdot (\#\text{times }(i,j)\text{ is modified}) = \sum_{(i,j)\in E_\mathsf{L}} w(i,j)\cdot \sum_{t:i,j\in t} M_{i,j,t}\ . \tag{8}$$

The charging scheme. To upper bound the sum of $M_{i,j,t}$ by the LP value of the optimal solution y^*, we introduce intermediate random variables $B_{i,j,t}$, which specify how much we charge the non-pivot edge in $\mathcal{A}_t$. More specifically, for any triangle $t \in \binom{[n]}{3}$ and any edge $(i,j) \in t \cap E$, we will charge (i,j) in triangle t if both of the following conditions are satisfied:

- the event $\mathcal{A}_t$ happens (i.e., three vertices of t appear in the same recursive call where one of them is chosen as a pivot), and

- the edge (i,j) is the non-pivot edge in the event $\mathcal{A}_t$.

When both conditions are satisfied, we assign a non-negative value to $B_{i,j,t}$ and use $B_{i,j,t}\cdot w(i,j)\cdot c^*(i,j)$ to denote how much (i,j) is charged in triangle t. Note that since the event $\mathcal{A}_t$ can only happen in at most one recursive call, $B_{i,j,t}$ cannot be assigned twice and is thus well defined. Otherwise, we assign 0 to $B_{i,j,t}$ for convenience. In the analysis, we want two properties for this charging scheme:

1. we can upper bound the expectation of expression (8) by the expected sum of $B_{i,j,t}\cdot w(i,j)\cdot c^*(i,j)$, and

2. fixing an edge (i,j), we can upper bound the expectation of the sum of $B_{i,j,t}$ over all $t \ni i,j$.

Next, we present the key Lemma 4.5 in this section: given a charging scheme (of $B_{i,j,t}$s) satisfying the above two properties, we can establish an upper bound on the approximation ratio of Algorithm 1.

Lemma 4.5. *Suppose there exists a charging scheme such that*

1. for any triangle $t \in \binom{[n]}{3}$,

$$\sum_{(i,j)\in t} w(i,j)\cdot \mathbb{E}[M_{i,j,t}|\mathcal{A}_t] \le \sum_{(i,j)\in t\cap E} \mathbb{E}[B_{i,j,t}|\mathcal{A}_t]\cdot w(i,j)\cdot c^*(i,j)\ ; \tag{9}$$

2. for any edge $(i,j) \in E$, the total charges on (i,j) satisfies

$$\sum_{t:i,j\in t} \mathbb{E}[B_{i,j,t}] \le \begin{cases} \overline{B}_\mathsf{L} & \text{if } (i,j) \in E_\mathsf{L}, \\ \overline{B}_\mathsf{HD} & \text{if } (i,j) \in E_\mathsf{HD}, \\ \overline{B}_\mathsf{HR} & \text{if } (i,j) \in E_\mathsf{HR}. \end{cases} \tag{10}$$

Then, Algorithm 1 is a $\max\{\overline{B}_\mathsf{L},\ \overline{B}_\mathsf{HD} + \frac{1}{1-\alpha},\ \overline{B}_\mathsf{HR} + \frac{1}{\alpha}\}$-approximation.

Proof. According to the definition, when $\mathcal{A}_t$ does not happen, $M_{i,j,t} = B_{i,j,t} = 0$ for any $(i,j) \in t$. Therefore, $\mathbb{E}[M_{i,j,t}] = \mathbb{E}[M_{i,j,t}|\mathcal{A}_t]\cdot \Pr[\mathcal{A}_t]$ and $\mathbb{E}[B_{i,j,t}] = \mathbb{E}[B_{i,j,t}|\mathcal{A}_t]\cdot \Pr[\mathcal{A}_t]$. By multiplying $\Pr[\mathcal{A}_t]$ on both sides of Eqn. (9) and summing over all possible $t \in \binom{[n]}{3}$, we have

$$\sum_{(i,j)\in E_\mathsf{L}} w(i,j)\cdot \sum_{t:i,j\in t} \mathbb{E}[M_{i,j,t}] \le \sum_{(i,j)\in E} \left(\sum_{t:i,j\in t} \mathbb{E}[B_{i,j,t}]\right)\cdot w(i,j)\cdot c^*(i,j)\ .$$

13

According to Eqn. (6), (7), (8), and (10),

$$\mathbb{E}[ALG] \leq \overline{B}_{\mathsf{L}} \sum_{(i,j)\in E_{\mathsf{L}}} w(i,j) \cdot c^*(i,j) + \overline{B}_{\mathsf{HD}} \sum_{(i,j)\in E_{\mathsf{HD}}} w(i,j) \cdot c^*(i,j) + \overline{B}_{\mathsf{HR}} \sum_{(i,j)\in E_{\mathsf{HR}}} w(i,j) \cdot c^*(i,j)$$

$$+ \frac{1}{1-\alpha} \sum_{(i,j)\in E_{\mathsf{HD}}} w(i,j) \cdot c^*(i,j) + \frac{1}{\alpha} \sum_{(i,j)\in E_{\mathsf{HR}}} w(i,j) \cdot c^*(i,j)$$

$$\leq \max\left\{\overline{B}_{\mathsf{L}}, \ \overline{B}_{\mathsf{HD}} + \frac{1}{1-\alpha}, \ \overline{B}_{\mathsf{HR}} + \frac{1}{\alpha}\right\} \sum_{(i,j)\in E} w(i,j) \cdot c^*(i,j)$$

$$\leq \max\left\{\overline{B}_{\mathsf{L}}, \ \overline{B}_{\mathsf{HD}} + \frac{1}{1-\alpha}, \ \overline{B}_{\mathsf{HR}} + \frac{1}{\alpha}\right\} \cdot \mathrm{OPT}$$

Therefore, Algorithm 1 is a $\max\{\overline{B}_{\mathsf{L}}, \ \overline{B}_{\mathsf{HD}} + \frac{1}{1-\alpha}, \ \overline{B}_{\mathsf{HR}} + \frac{1}{\alpha}\}$-approximation. $\qquad\square$

Bounding the total charges. Next, we present a method to upper bound the total charge $\sum_{t:(i,j)\in t} \mathbb{E}[B_{i,j,t}]$ on each edge by lower bounding the probability that an edge can no longer be charged after each time it is charged. Similar analysis can be found in previous works [ACN08, CMSY15] on Correlation Clustering. In Algorithm 1, when the endpoints of an edge are partitioned into different sets on line 26 in a recursive call, the edge will no longer be charged in its child calls. Then, the key lemma for the upper bound can be formulated as follows.

Lemma 4.6. *Fix an edge $(i,j) \in E$ and any $q > 0$. Let $\mathcal{R}$ be a recursive call, $b > 0$ be a positive number, and t be a triangle containing i, j such that $\mathcal{A}_t$ happens in $\mathcal{R}$ and (i,j) is a non-pivot edge charged by $B_{i,j,t} = b$. If for any valid choice of $\mathcal{R}, b, t$ the probability that i and j are partitioned into different sets on line 26 of Algorithm 1 is at least $q \cdot b$, then the expected total charges on edge (i,j) satisfy*

$$\sum_{t:(i,j)\in t} \mathbb{E}[B_{i,j,t}] \leq q^{-1} \ .$$

Proof. We consider a stronger argument on the calls: *for any call LP-UMVD-Pivot(V, x, u) such that $i, j \in V$, $\sum_{t\in\binom{V}{3}:(i,j)\in t} \mathbb{E}[B_{i,j,t}] \leq q^{-1}$.* We shall prove it by induction in a bottom-up way. The base case is when $|V| = 2$ or $u = L$. Because the algorithm returns the input distance at the beginning of the call, according to the definition, $B_{i,j,t} = 0$ for any $i, j \in t \in \binom{V}{3}$.

Consider any call LP-UMVD-Pivot(V, x, u) such that $|V| \geq 3$ and $u < L$. Consider any $(i,j) \in \binom{V}{2}$. Suppose that the stronger argument holds for any LP-UMVD-Pivot(V', x', u') satisfying $|V'| < |V|$. If i or j is chosen as the pivot vertex in the call, $B_{i,j,t} = 0$ for any $i, j \in t \in \binom{V}{3}$ because (i,j) is not a non-pivot edge and i, j do not simultaneously appear in any of its children calls. We prove the stronger argument for this case. Otherwise, suppose that the pivot vertex is $k \neq i, j$. Let $t = (i,j,k)$. In this case, the expected total charge of (i,j) equals the sum of its expected charge in triangle t and its expected total charge in the child call involving both i, j (if exists):

$$\sum_{t'\in\binom{V}{3}:i,j\in t'} \mathbb{E}[B_{i,j,t'}] = \mathbb{E}[B_{i,j,t}] + \sum_{\ell\geq u} \Pr[i, j \in V_\ell] \cdot \mathbb{E}\left[\sum_{t'\in\binom{V_\ell}{3}:i,j\in t'} B_{i,j,t'} \ \Big| \ i, j \in V_\ell\right]$$

For any $b > 0$, conditioning on $B_{i,j,t} = b$, we have $\sum_{\ell\geq u} \Pr[i, j \in V_\ell] \leq 1 - q \cdot b$ according to the assumption of the lemma. According to the induction hypothesis, (conditioning on $B_{i,j,t} = b$) $\mathbb{E}\left[\sum_{t'\in\binom{V_\ell}{3}:i,j\in t'} B_{i,j,t'} \big| i, j \in V_\ell\right] \leq q^{-1}$ for any possible $V_\ell \subset V$. Therefore, the stronger argument

14

holds in this case because

$$\sum_{t' \in \binom{V}{3} : i, j \in t'} \mathbb{E}[B_{i,j,t'}] \leq \sum_{b \geq 0} \Pr[B_{i,j,t} = b] \cdot \left(b + (1 - q \cdot b) \cdot q^{-1} \right)$$

$$= \sum_{b \geq 0} \Pr[B_{i,j,t} = b] \cdot q^{-1} = q^{-1} .$$

$\square$

5 5-Approximation Analysis for Complete Graphs

In this section, we prove that Algorithm 1 is a 5-approximation algorithm for unweighted ULTRA-METRIC VIOLATION DISTANCE on complete graphs by setting $\beta = 0$ and an appropriate choice of α. More specifically, we will prove the following theorem:

Theorem 5.1. If $\alpha \in \left[\frac{3-\sqrt{5}}{2}, 0.5 \right], \beta = 0$, Algorithm 1 is a randomized polynomial-time $\max\{\frac{3}{1-\alpha}, \frac{2}{\alpha}\}$-approximation algorithm for ULTRAMETRIC VIOLATION DISTANCE on complete graphs. In particular, with $\alpha = 0.4$, it is a randomized polynomial-time 5-approximation for ULTRAMETRIC VIOLA-TION DISTANCE on complete graphs.

Recall that we have $E = \binom{[n]}{2}, E_\varnothing = \emptyset$ and $\forall (i,j) \in E, w(i,j) = 1$ in this case. For convenience, in the rest of this section, we assume $\beta = 0$. Upon this assumption, the CCDF function (4) for random edges in E becomes:

$$\forall \ell \in [L], \quad \Pr[x'(i,j) \geq d_\ell] = y_\ell(i,j) . \tag{11}$$

To prove Theorem 5.1, we shall use Lemma 4.5, 4.6 and the following Lemma 5.2, which presents a charging scheme that meets the two conditions we need for the approximation ratio.

Lemma 5.2. If $\alpha \in \left[\frac{3-\sqrt{5}}{2}, 0.5 \right]$, there exists a charging scheme on complete graphs such that

1. for any triangle $t \in \binom{[n]}{3}$,

$$\sum_{(i,j) \in t} \mathbb{E}[M_{i,j,t} | \mathcal{A}_t] \leq \sum_{(i,j) \in t} \mathbb{E}[B_{i,j,t} | \mathcal{A}_t] \cdot c^*(i,j) .$$

2. for any $b > 0$, $(i,j) \in E$ and any $t \ni i, j$, in any recursive call, conditioning on $\mathcal{A}_t$ and that (i,j) is a non-pivot edge charged by $B_{i,j,t} = b$, the probability that i and j are partitioned into different sets on line 26 of Algorithm 1 is at least $q_{i,j} \cdot b$, where

$$q_{i,j} = \begin{cases} \min\{\frac{1-\alpha}{3}, \frac{\alpha}{2}\} & \text{if } (i,j) \in E_\mathsf{L} \\ \frac{1-\alpha}{2} & \text{if } (i,j) \in E_\mathsf{HD} \\ \alpha & \text{if } (i,j) \in E_\mathsf{HR} \end{cases}$$

Assuming the correctness of the above lemma, we can obtain the proof of Theorem 5.1.

15

Proof of Theorem 5.1. Because of Lemma 4.6 and the second bullet of Lemma 5.2, we can present the second condition of Lemma 4.5 as follows:

$$\forall (i,j) \in E, \quad \sum_{t:i,j\in t} \mathbb{E}[B_{i,j,t}] \leq \begin{cases} \max\{\frac{3}{1-\alpha}, \frac{2}{\alpha}\} & \text{if } (i,j) \in E_{\mathsf{L}} , \\ \frac{2}{1-\alpha} & \text{if } (i,j) \in E_{\mathsf{HD}} , \\ \frac{1}{\alpha} & \text{if } (i,j) \in E_{\mathsf{HR}} . \end{cases}$$

Because the first bullet of Lemma 5.2 meets that of Lemma 4.5, if $\alpha \in \left[\frac{3-\sqrt{5}}{2}, 0.5\right]$, Algorithm 1 is a $\max\{\frac{3}{1-\alpha}, \frac{2}{\alpha}\}$-approximation for ULTRAMETRIC VIOLATION DISTANCE. $\qquad\square$

In the rest of this section, we present the charging scheme and prove Lemma 5.2. The charging scheme defines $B_{i,j,t}$ differently for different classes of triangles. Recall that given the truncated solution y, we say an edge is deterministic if $\Delta y_{\ell^*(i,j)}(i,j) > 1 - \alpha$, and otherwise, we say it is random. The triangles are classified according to the number of deterministic edges and the dominant levels of the deterministic edges in the triangles:

(d, d, d)**-triangles** have three deterministic edges.

(d, d, r)**-same-triangles** have two deterministic edges and one random edge, and the dominant levels of the deterministic edges are the same.

(d, d, r)**-diff-triangles** have two deterministic edges and one random edge, and the dominant levels of the deterministic edges are different.

(d, r, r)**-triangles** have one deterministic edge and two random edges.

(r, r, r)**-triangles** have three random edges.

Next, we present three useful lemmas. In the discussion of different classes, we have the following observation: the probability that a low-cost edge is modified is highly related to the values of $y_{\ell^*(i,j)-1}(i,j)$ and $1 - \Delta y_{\ell^*(i,j)}(i,j)$ on each edge $(i,j) \in t$ of the triangle. The first two lemmas give upper bounds for these terms based on the LP contributions $c^*(i,j)$ of the edges and will be extensively used to prove the first bullet of Lemma 5.2. We note that the proofs of these two lemmas are independent of E and $w(i,j)$ and can work for all three cases.

Lemma 5.3. *In any recursive call with vertex set V and truncated LP solution y, for any edge $(i,j) \in \binom{V}{2}$, $1 - \Delta y_{\ell^*(i,j)}(i,j)$ can be upper bounded by*

$$\begin{cases} c^*(i,j) & \text{if } (i,j) \in E_{\mathsf{L}} \cup E_{\mathsf{HR}} , \\ \frac{\alpha}{1-\alpha} \cdot c^*(i,j) & \text{if } (i,j) \in E_{\mathsf{HD}} . \end{cases}$$

In particular, $y_{\ell^*(i,j)-1}(i,j)$ and $1 - y_{\ell^*(i,j)}(i,j)$ satisfy the same upper bound.

Proof. For any $(i,j) \in \binom{[n]}{2}$, because of Lemma 4.2, we have

$$1 - \Delta y_{\ell^*(i,j)}(i,j) \leq 1 - \Delta y^*_{\ell^*(i,j)}(i,j) = c^*(i,j) .$$

In particular, for any $(i,j) \in E_{\mathsf{HD}}$, we have

$$\begin{aligned} 1 - \Delta y_{\ell^*(i,j)}(i,j) &\leq 1 - \Delta y^*_{\ell^*(i,j)}(i,j) && \text{(Lemma 4.2)} \\ &< \alpha && ((i,j) \text{ is initially deterministic}) \\ &< \frac{\alpha}{1-\alpha} \cdot c^*(i,j) . && \text{(Lemma 4.1)} \end{aligned}$$

In particular, the second lemma gives an improved bound for a certain type of edge under some conditions, which helps us to refine the analysis.

16

Lemma 5.4. *When an edge $(i, j) \in E_{\mathsf{HR}}$ appears as deterministic in some recursive call with truncated LP solution y, $y_{\ell^*(i,j)-1}(i, j) = 0$.*

Proof. Because $(i, j) \in E_{\mathsf{HR}}$, $\forall \ell \in [L], \Delta y_\ell^*(i, j) \leq 1-\alpha$. Consider any call LP-UMVD-PIVOT(V, x, u) with truncated LP solution y. For any $\ell > u$, $\Delta y_\ell(i, j) = \Delta y_\ell^*(i, j) \leq 1 - \alpha$. Since (i, j) is deterministic in the call and $y_\ell(i, j) = 0$ for $\ell < u$, the dominant level $\ell^*(y, i, j)$ can only be u. Therefore, $y_{\ell^*(i,j)-1}(i, j) = 0$. $\qquad\square$

In particular, because $1 - \Delta y_{\ell^*(i,j)}(i, j) = y_{\ell^*(i,j)-1}(i, j) + 1 - y_{\ell^*(i,j)}(i, j)$ and $\forall \ell \in [L], y_\ell(i, j) \in [0, 1]$, $y_{\ell^*(i,j)-1}(i, j)$ and $1 - y_{\ell^*(i,j)}(i, j)$ satisfy the same upper bound. $\qquad\square$

The third lemma gives lower bounds for the probabilities that the endpoints of a non-pivot edge are partitioned into different sets on line 26 of the algorithm, which will be used for the second bullet of Lemma 5.2. The proof of this lemma follows on the fact that the distance on a random edge can only equal a fixed value with probability at most $1 - \alpha$, i.e., $\Delta y_\ell(i, j) \leq 1 - \alpha$ for any random (i, j) and $\ell \in [L]$.

Lemma 5.5. *Consider any recursive call LP-UMVD-PIVOT(V, x, u) with $|V| > 2$. Suppose i is the pivot vertex of the call. For any $(j, k) \in \binom{V \setminus \{i\}}{2}$, the probability that j, k are partitioned into different sets on line 26 can be lower bounded by*

- *1 if both (i, j) and (i, k) are deterministic in the call but they have different dominant levels, or*

- *α if at least one of (i, j) and (i, k) is random.*

Proof. For the first bullet, according to the algorithm, the distances of (i, j) and (i, k) are deterministically different, and thus j, k are partitioned into different sets with probability 1.

For the second bullet, suppose that the truncated LP solution in the call is y. W.l.o.g., we assume that (i, j) is random. Note that this implies $\forall \ell \in [L], \Delta y_\ell(i, j) \leq \Delta y_{\ell^*(i,j)}(i, j) \leq 1 - \alpha$. According to the CCDF (11), the probability that j and k are partitioned into different sets on line 26 is

$$1 - \sum_{\ell \in [L]} \Pr[j, k \in V_\ell] = 1 - \sum_{\ell \in [L]} \Pr[x'(i, j) = d_\ell] \cdot \Pr[x'(i, k) = d_\ell]$$

$$= 1 - \sum_{\ell \in [L]} \Delta y_\ell(i, j) \cdot \Pr[x'(i, k) = d_\ell]$$

$$\geq 1 - \sum_{\ell \in [L]} (1 - \alpha) \cdot \Pr[x'(i, k) = d_\ell]$$

$$= 1 - (1 - \alpha) = \alpha .$$

$\qquad\square$

Next, we present the charging scheme and prove Lemma 5.2 for each class of triangles. Due to Corollary 4.3, there is no low-cost edge in (r, r, r)-triangles. Therefore, we only need to discuss the remaining four classes of triangles. Consider any triangle $t = (i, j, k)$ and any recursive call with a vertex set involving i, j, k and truncated LP solution y. For convenience, we will condition on $\mathcal{A}_t$ in the rest of the analysis in this section, and all the expectations will be automatically conditioned on $\mathcal{A}_t$. Note that the two pivot edges in (d, r, r)-triangles and (d, d, r)-diff-triangles always satisfy one of the conditions in Lemma 5.5. The endpoints of the non-pivot edges in these triangles will

17

be partitioned into different sets with a constant probability (under the choice of $\alpha \in [\frac{3-\sqrt{5}}{2}, 0.5]$). Hence, it should not be surprising to prove Lemma 5.2 for these triangles, and we shall defer their proofs to Appendix B.5.

Recall our charging scheme. Conditioning on $\mathcal{A}_t$, we charge an edge in t only when the edge is non-pivot. If an edge (i', j') is not charged in t, we automatically assign 0 to $B_{i',j',t}$. For convenience, we will only specify $B_{i',j',t}$ when (i', j') is the non-pivot edge in the rest of the analysis.

(d, d, d)-**triangles.** Because all three edges in the triangle are deterministic, we have $\Delta y_{\ell^*(i,j)}(i, j)$, $\Delta y_{\ell^*(i,k)}(i, k)$, $\Delta y_{\ell^*(j,k)}(j, k) > 1 - \alpha$. Note that this implies that

$$\forall i' \neq j' \in t, \quad y_{\ell^*(i',j')-1}(i', j') < \alpha, \ y_{\ell^*(i',j')}(i', j') > 1 - \alpha . \tag{12}$$

Note that the distance of any pivot edge $(u, v) \in t$ is set to its dominant distance level $d_{\ell^*(y,u,v)}$. Because of Lemma 4.4, the input distance of any low-cost edge $(u, v) \in t$ is also $d_{\ell^*(y,u,v)}$. Therefore, a low-cost edge is modified only when it is non-pivot in the call and the ultrametric inequality is violated on the three dominant distance levels, i.e., there exists a permutation (i', j', k') of i, j, k such that $\ell^*(y, i', j') < \min\{\ell^*(y, j', k'), \ell^*(y, k', i')\}$ (equivalently, $d_{\ell^*(y,i',j')} > \max\{d_{\ell^*(y,j',k')}, d_{\ell^*(y,k',i')}\}$).

If there is no violation of the ultrametric inequality on these three deterministic levels, $M_{i,j,t} = M_{i,k,t} = M_{j,k,t} = 0$. For this case, we define $B_{i,j,t}, B_{i,k,t}, B_{j,k,t} \stackrel{\text{def}}{=} 0$, and Lemma 5.2 then clearly holds.

Otherwise, w.l.o.g., we assume that $\ell^*(y, j, k) < \min\{\ell^*(y, i, j), \ell^*(y, i, k)\}$. In this case, for any permutation (i', j', k') of i, j, k,

$$M_{i',j',t} = \mathbf{1}(k' \text{ is the pivot vertex}) \cdot \mathbf{1}((i', j') \in E_\mathsf{L}).$$

Because each vertex in t is chosen as the pivot with equal probability $1/3$ conditioning on $\mathcal{A}_t$, in this case,

$$\sum_{(i',j')\in t} \mathbb{E}[M_{i',j',t}] = \frac{1}{3} \cdot \Big(\mathbf{1}((i, j) \in E_\mathsf{L}) + \mathbf{1}((i, k) \in E_\mathsf{L}) + \mathbf{1}((j, k) \in E_\mathsf{L}) \Big) .$$

Because y is feasible in (UMVD LP) and because of Lemma 5.3,

$$
\begin{aligned}
c^*(i, j) + c^*(i, k) &\geq y_{\ell^*(i,j)-1}(i, j) + y_{\ell^*(i,k)-1}(i, k) & \text{(Lemma 5.3)} \\
&\geq y_{\ell^*(j,k)}(i, j) + y_{\ell^*(j,k)}(i, k) & \text{(LP constraint (2))} \\
&\geq y_{\ell^*(j,k)}(j, k) > 1 - \alpha . & \text{(LP constraint (1) \& Eqn. (12))}
\end{aligned}
$$

Remark 5.6. If $\alpha \leq \frac{1}{3}$, $y_{\ell^*(i,j)-1}(i, j) + y_{\ell^*(i,k)-1}(i, k) \leq 2\alpha \leq 1 - \alpha$ and one can easily find a contradiction. This implies no violation of the ultrametric inequality in this type of triangle if the parameter α is less than $\frac{1}{3}$.

Because of Lemma 5.4, Eqn. (12) and $\alpha \leq \frac{1}{2}$, if $(i, j) \in E_\mathsf{HR}$ or $(i, k) \in E_\mathsf{HR}$, $y_{\ell^*(i,j)-1}(i, j) + y_{\ell^*(i,k)-1}(i, k) < \alpha \leq 1 - \alpha$, which violates the above inequality. Therefore, edges $(i, j), (i, k) \notin E_\mathsf{HR}$. Based on this fact, when the corresponding edge is non-pivot in the recursive call, we shall define $B_{i,j,k}, B_{i,k,t}, B_{j,k,t}$ as follows:

$$
B_{i,j,t} \stackrel{\text{def}}{=} \begin{cases} \frac{3}{1-\alpha} & \text{if } (i, j) \in E_\mathsf{L} \\ \frac{2}{1-\alpha} & \text{if } (i, j) \in E_\mathsf{HD} \end{cases} , \quad
B_{i,k,t} \stackrel{\text{def}}{=} \begin{cases} \frac{3}{1-\alpha} & \text{if } (i, k) \in E_\mathsf{L} \\ \frac{2}{1-\alpha} & \text{if } (i, k) \in E_\mathsf{HD} \end{cases} , \quad
B_{j,k,t} \stackrel{\text{def}}{=} 0 ,
$$

which implies the first bullet of Lemma 5.2 for this case:

$$\sum_{(i',j')\in t} \mathbb{E}[B_{i',j',t}] \cdot c^*(i',j') = \frac{2 + \mathbf{1}((i,j)\in E_L)}{3(1-\alpha)} \cdot c^*(i,j) + \frac{2 + \mathbf{1}((i,k)\in E_L)}{3(1-\alpha)} \cdot c^*(i,k)$$

$$\geq \frac{\mathbf{1}((i,j)\in E_L) + \mathbf{1}((i,k)\in E_L) + \mathbf{1}((j,k)\in E_L)}{3(1-\alpha)} \cdot \big(c^*(i,j) + c^*(i,k)\big)$$

$$> \frac{\mathbf{1}((i,j)\in E_L) + \mathbf{1}((i,k)\in E_L) + \mathbf{1}((j,k)\in E_L)}{3} = \sum_{(i',j')\in t} \mathbb{E}[M_{i',j',t}] ,$$

Further, when the edge (i,j) (or (i,k)) is non-pivot, two pivot edges in the triangle are both deterministic but have different dominant levels. Because of Lemma 5.5, the probability i,j (or i,k) are then partitioned into different sets on line 26 is at least 1, which is at least

$$\begin{cases} \frac{1-\alpha}{3} \cdot B_{i,j,t} & \text{if } (i,j) \in E_L \\ \frac{1-\alpha}{2} \cdot B_{i,j,t} & \text{if } (i,j) \in E_{HD} \end{cases} \quad \text{and} \quad \begin{cases} \frac{1-\alpha}{3} \cdot B_{i,k,t} & \text{if } (i,k) \in E_L \\ \frac{1-\alpha}{2} \cdot B_{i,k,t} & \text{if } (i,k) \in E_{HD} \end{cases} .$$

This definition satisfies the second bullet of Lemma 5.2 for this case.

(d,d,r)**-same-triangles.** W.l.o.g., we assume (i,j) and (i,k) are the deterministic edges. Because of Lemma 4.4 and $\ell^*(y,i,j) = \ell^*(y,i,k)$, the input distances of edges (i,j) and (i,k) are both $d_{\ell^*(y,i,j)}$. Therefore, $M_{i,j,t}$ (or $M_{i,k,t}$) equals 1 only when the pivot is k (resp., j), it is low-cost, and the random distance $x'(j,k) > d_{\ell^*(y,i,j)}$, i.e.,

$$M_{i,j,t} = \mathbf{1}(k \text{ is the pivot vertex}) \cdot \mathbf{1}((i,j)\in E_L) \cdot \mathbf{1}(x'(j,k) > d_{\ell^*(y,i,j)}) ,$$
$$M_{i,k,t} = \mathbf{1}(j \text{ is the pivot vertex}) \cdot \mathbf{1}((i,k)\in E_L) \cdot \mathbf{1}(x'(j,k) > d_{\ell^*(y,i,j)}) .$$

According to the CCDF (11) of the random distance, the probability that $x'(j,k) > d_{\ell^*(y,i,j)}$ is

$$y_{\ell^*(i,j)-1}(j,k) \leq y_{\ell^*(i,j)-1}(i,j) + y_{\ell^*(i,j)-1}(i,k) \qquad\qquad (\text{LP constraint } (1))$$

$$= y_{\ell^*(i,j)-1}(i,j) + y_{\ell^*(i,k)-1}(i,k) . \qquad\qquad (13)$$

Because each vertex in t is chosen as the pivot with equal probability $1/3$ conditioning on $\mathcal{A}_t$, in this class of triangles,

$$\sum_{(i',j')\in t} \mathbb{E}[M_{i',j',t}] \leq \frac{2(y_{\ell^*(i,j)-1}(i,j) + y_{\ell^*(i,k)-1}(i,k))}{3} .$$

Accordingly, when the corresponding edge is non-pivot, we define $B_{i,j,t}, B_{i,k,t}, B_{j,k,t}$ as follows:

$$B_{i,j,t} \stackrel{\text{def}}{=} \begin{cases} 2 & \text{if } (i,j) \in E_L \\ \frac{2\alpha}{1-\alpha} & \text{if } (i,j) \in E_{HD} \\ 0 & \text{if } (i,j) \in E_{HR} \end{cases} , \quad B_{i,k,t} \stackrel{\text{def}}{=} \begin{cases} 2 & \text{if } (i,k) \in E_L \\ \frac{2\alpha}{1-\alpha} & \text{if } (i,k) \in E_{HD} \\ 0 & \text{if } (i,k) \in E_{HR} \end{cases} , \quad B_{j,k,t} \stackrel{\text{def}}{=} 0$$

Because of Lemma 5.3 and 5.4, this definition implies the first bullet of Lemma 5.2:

$$\sum_{(i',j')\in t} \mathbb{E}[B_{i',j',t}] \cdot c^*(i',j') \geq \frac{2}{3} \cdot y_{\ell^*(i,j)-1}(i,j) + \frac{2}{3} \cdot y_{\ell^*(i,k)-1}(i,k) \geq \sum_{(i',j')\in t} \mathbb{E}[M_{i',j',t}] .$$

19

Further, when the edge (i,j) (or (i,k)) is non-pivot, one of the pivot edges in the triangle is random. Because of Lemma 5.5, the probability i,j (or i,k) are then partitioned into different sets on line 26 is at least α, which is at least

$$\begin{cases} \frac{\alpha}{2} \cdot B_{i,j,t} & \text{if } (i,j) \in E_{\mathsf{L}} \\ \frac{1-\alpha}{2} \cdot B_{i,j,t} & \text{if } (i,j) \in E_{\mathsf{HD}} \\ 1 \cdot B_{i,j,t} & \text{if } (i,j) \in E_{\mathsf{HR}} \end{cases} \quad \text{and} \quad \begin{cases} \frac{\alpha}{2} \cdot B_{i,k,t} & \text{if } (i,k) \in E_{\mathsf{L}} \\ \frac{1-\alpha}{2} \cdot B_{i,k,t} & \text{if } (i,k) \in E_{\mathsf{HD}} \\ 1 \cdot B_{i,k,t} & \text{if } (i,k) \in E_{\mathsf{HR}} \end{cases} .$$

This definition satisfies the second bullet of Lemma 5.2 for this class of triangles.

6 $O(\min\{L, \log n\})$-Approximation Analysis for Weighted Instances with Triangle Inequality Constraints

In this section, we prove that Algorithm 1 is an $O(\min\{L, \log n\})$-approximation algorithm for weighted ULTRAMETRIC VIOLATION DISTANCE with triangle inequality constraints by setting $\beta = 0$ and an appropriate choice of α. More specifically, we will prove the following theorem:

Theorem 6.1. If $\alpha = \frac{1}{3}, \beta = 0$, Algorithm 1 is a randomized polynomial-time $O(\min\{L, \log n\})$-approximation algorithm for weighted ULTRAMETRIC VIOLATION DISTANCE where weights satisfy the triangle inequality.

In Section 2, we have w.l.o.g. assumed that $E = \binom{[n]}{2}$ for weighted instances. Because of Lemma 4.5, the proof of Theorem 6.1 directly follows the following lemma.

Lemma 6.2. If $\alpha = \frac{1}{3}, \beta = 0$, there exists a charging scheme on weighted cases with triangle inequality constraints such that

1. for any triangle $t \in \binom{[n]}{3}$,

$$\sum_{(i,j) \in t} w(i,j) \cdot \mathbb{E}[M_{i,j,t} | \mathcal{A}_t] \leq \sum_{(i,j) \in t} \mathbb{E}[B_{i,j,t} | \mathcal{A}_t] \cdot w(i,j) \cdot c^*(i,j) .$$

2. for any edge $(i,j) \in \binom{[n]}{2}$, $\sum_{t:i,j \in t} \mathbb{E}[B_{i,j,t}] = O(\min\{L, \log n\})$.

In the rest of this section, we will prove Lemma 6.2. For convenience, we assume $\alpha = \frac{1}{3}$ and $\beta = 0$. Because $\beta = 0$, we can still use the lemmas and inequalities established in Section 5, including Lemma 5.3, 5.4, 5.5 and inequalities (12), (13).

We classify the charge $B_{i,j,t}$ into two types. Let $B_{i,j,t}^{(1)}$ and $B_{i,j,t}^{(2)}$ denote charges of the two types, where $B_{i,j,t}^{(1)} = B_{i,j,t}, B_{i,j,t}^{(2)} = 0$ if the charge is classified to the first type and $B_{i,j,t}^{(1)} = 0, B_{i,j,t}^{(2)} = B_{i,j,t}$ if the charge is classified as the second type. With the classification, the total charge on (i,j) can be rewritten as the sum of its first-type charges and its second-type charges, i.e., $\sum_{t:i,j \in t} B_{i,j,t}^{(1)} + \sum_{t:i,j \in t} B_{i,j,t}^{(2)}$. Therefore, we can upper bound for total charges of different types to upper bound the total charges:

$$\sum_{t:i,j \in t} \mathbb{E}[B_{i,j,t}^{(1)}] = O(1) \quad \text{and} \quad \sum_{t:i,j \in t} \mathbb{E}[B_{i,j,t}^{(2)}] = O(\min\{L, \log n\}) .$$

For the first type, we will use the same method in Section 5 by showing the condition in Lemma 4.6 with $q = \Omega(1)$. For the second type, we will use the technique of [CFLM22] to analyze the $O(\log\{L, \log n\})$ approximation ratio of their pivot-based algorithm.

Similarly, as in Section 5, we prove the lemma by presenting the charging scheme for each class of triangles. Consider any triangle $t = (i, j, k)$ and any recursive call with a vertex set involving all of i, j, k. Suppose that the recursive call has an upper bound level u and a truncated LP solution y. For convenience, we will condition on $\mathcal{A}_t$ in the rest of this section, and all expectations will be automatically conditioned on $\mathcal{A}_t$. Further, we similarly defer the proofs for (d, d, r)-diff-triangles and (d, r, r)-triangles to Appendix B.6.

(d, d, d)**-triangles.** Note that the distance of any pivot edge $(u, v) \in t$ is set to its dominant distance level $d_{\ell^*(y, u, v)}$. Because of Lemma 4.4, any low-cost edge (u, v) has an input distance $x(u, v)$ equal to the dominant distance level $d_{\ell^*(y, u, v)}$. If the non-pivot edge is low-cost, it is modified only when the ultrametric inequality is violated at the three dominant distance levels, i.e., there exists a permutation (i', j', k') of i, j, k such that $\ell^*(y, i', j') < \min\{\ell^*(y, i', k'), \ell^*(y, j', k')\}$. However, $\ell^*(y, i', j') < \min\{\ell^*(y, i', k'), \ell^*(y, j', k')\}$ implies

$$
\begin{align}
\alpha &> 1 - y_{\ell^*(i', j')}(i', j') & \text{(Eqn. (12))} \\
&\geq 1 - y_{\ell^*(i', j')}(i', k') - y_{\ell^*(i', j')}(j', k') & \text{(LP constraint (1))} \\
&\geq 1 - y_{\ell^*(i', k')-1}(i', k') - y_{\ell^*(j', k')-1}(j', k') & \text{(LP constraint (2))} \\
&> 1 - 2\alpha , & \text{(Eqn. (12))}
\end{align}
$$

which violates our choice of parameter $\alpha = \frac{1}{3}$. Therefore, the low-cost non-pivot edges cannot be modified, and we have $M_{i,j,t} = M_{i,k,t} = M_{j,k,t} = 0$ in this case. Accordingly, by always defining $B_{i,j,t}, B_{i,k,t}, B_{j,k,t} \stackrel{\text{def}}{=} 0$ and classifying the charges into the first type, we show the first bullet of Lemma 6.2 and the condition in Lemma 4.6 for this class of triangles.

(d, d, r)**-same-triangles.** W.l.o.g. we assume (i, j) and (i, k) are deterministic. Because of Corollary 4.3, we have $(j, k) \in E_H$, $M_{j,k,t} = 0$, and $c^*(j, k) \geq \alpha$. Recall that we have shown for this class of triangles in Section 5:

$$
\begin{align}
M_{i,j,t} &= \mathbf{1}(k \text{ is the pivot vertex}) \cdot \mathbf{1}((i, j) \in E_L) \cdot \mathbf{1}(x'(j, k) > d_{\ell^*(y, i, j)}) , \\
M_{i,k,t} &= \mathbf{1}(j \text{ is the pivot vertex}) \cdot \mathbf{1}((i, k) \in E_L) \cdot \mathbf{1}(x'(j, k) > d_{\ell^*(y, i, j)}) ,
\end{align}
$$

and the probability that $x'(j, k) > d_{\ell^*(y, i, j)}$ is at most $y_{\ell^*(i, j)-1}(i, j) + y_{\ell^*(i, k)-1}(i, k)$. Because of Lemma 5.3, $y_{\ell^*(i, j)-1}(i, j) \leq c^*(i, j)$, $y_{\ell^*(i, k)-1}(i, k) \leq c^*(i, k)$. Because edges (i, j) and (i, k) are deterministic and Eqn. (12), $y_{\ell^*(i, j)-1}(i, j), y_{\ell^*(i, k)-1}(i, k) < \alpha \leq c^*(j, k)$. Therefore, in this class of

21

triangles, we have

$$\sum_{(i',j')\in t} w(i',j') \cdot \mathbb{E}[M_{i',j',t}] \leq \frac{y_{\ell^*(i,j)-1}(i,j) + y_{\ell^*(i,k)-1}(i,k)}{3} \cdot \big(w(i,j) + w(i,k)\big)$$

$$\leq \frac{w(i,j) \cdot c^*(i,j) + w(i,k) \cdot c^*(i,k)}{3} + \frac{y_{\ell^*(i,j)-1}(i,j) \cdot w(i,k)}{3}$$

$$+ \frac{y_{\ell^*(i,k)-1}(i,k) \cdot w(i,j)}{3}$$

$$\leq \frac{w(i,j) \cdot c^*(i,j) + w(i,k) \cdot c^*(i,k)}{3} + \frac{y_{\ell^*(i,j)-1}(i,j) \cdot (w(i,j) + w(j,k))}{3}$$

$$+ \frac{y_{\ell^*(i,k)-1}(i,k) \cdot (w(i,k) + w(j,k))}{3} \qquad \text{(triangle inequality for } w\text{)}$$

$$\leq \frac{w(i,j) \cdot c^*(i,j) + w(i,k) \cdot c^*(i,k)}{3} + \frac{c^*(i,j) \cdot w(i,j) + c^*(j,k) \cdot w(j,k)}{3}$$

$$+ \frac{c^*(i,k) \cdot w(i,k) + c^*(j,k) \cdot w(j,k)}{3}$$

$$= \frac{2\big(w(i,j) \cdot c^*(i,j) + w(j,k) \cdot c^*(j,k) + w(i,k) \cdot c^*(i,k)\big)}{3} \ .$$

On the other hand, if $\ell^*(y,i,j) = u$, according to the definition of the truncated LP solution, $y_{\ell^*(i,j)-1}(i,j) = 0$. In this case, $\sum_{(i',j')\in t} w(i',j') \cdot \mathbb{E}[M_{i',j',t}] = 0$. Accordingly, when the corresponding edge is non-pivot, we can define $B_{i,j,t}, B_{i,k,t}, B_{j,k,t} \stackrel{\text{def}}{=} 2 \cdot \mathbf{1}(\ell^*(y,i,j) > u)$ to prove the first bullet of Lemma 6.2. Furthermore, we classify the charges on (i,j) and (i,k) into the first type, while we classify the charge on (j,k) into the second type. Note that in this class of triangles, when (i,j) or (i,k) is non-pivot, there is a random pivot edge. Because of Lemma 5.5, when being non-pivot, i,j (or i,k) will be partitioned into different sets on line 26 with probability at least $\alpha = \frac{1}{3}$, which is at least $\Omega(1) \cdot B_{i,j,t}$ (resp., $\Omega(1) \cdot B_{i,k,t}$). Therefore, this definition for (i,j) and (i,k) satisfies the condition of Lemma 4.6 with $q = \Omega(1)$. Next, we will upper bound the expected total charge of the second type on the random edges in this class of triangles.

Bounding the total charges of the second type. Fix an edge $(j,k) \in E$. In any recursive call with an upper bound level u, if $B^{(2)}_{j,k,t} \neq 0$, where the pivot in the call is i and $t = (i,j,k)$, we have $\ell^*(y,i,j) > u$. At the same time, the algorithm modifies the distance of (j,k) to $d_{\ell^*(y,i,j)}$ and partitions j, k to a child call with upper bound level $u' = \ell^*(y,i,j) > u$. Therefore, the recursive calls that charge (j,k) for the second type must have distinct upper bound levels u, and thus $\sum_{t:(i,j)\in t} B^{(2)}_{j,k,t} \leq O(L)$.

On the other hand, we have the following characterization for the dominant levels of each edge in each recursive call: the dominant level equals either its dominant level in y^* or the upper bound level u of the call. Its proof follows the fact that the truncated LP solution satisfies $\forall \ell < u, \Delta y_\ell(i,j) = 0$, $\Delta y_u(i,j) = y^*_u(i,j)$ and $\forall \ell > u, \Delta y_\ell(i,j) = \Delta y^*_\ell(i,j)$ in a call with upper bound level u. Furthermore, because descendant calls have higher upper bound levels, if the dominant level of edge (i,j) does not equal $\ell^*(y^*,i,j)$ in a call, it will no longer equal $\ell^*(y^*,i,j)$ in all its descendant calls. Formally, all previous arguments can be summarized by the following lemma.

Lemma 6.3. *In any recursive call with an upper bound level u, the dominant distance level of any edge (i,j) satisfies $\ell^*(y,i,j) = u$ or $\ell^*(y,i,j) = \ell^*(y^*,i,j)$. In particular, if $\ell^*(y,i,j) = u$ in any recursive call with truncated LP solution y and upper bound level u, $\ell^*(y',i,j) = u'$ in any of its descendant calls with truncated LP solution y' and upper bound u'.*

In a recursive call, we call vertices $i \neq j, k$ a bad pivot of (j, k) if $\ell^*(y, i, j) = \ell^*(y, i, k) > u$. Otherwise, we call i a good pivot. Because of Lemma 6.3, any bad pivot i of (j, k) satisfies $\ell^*(y, i, j) = \ell^*(y^*, i, j)$, and good pivots in a call never become bad in its descendant calls. Further, if the pivot i is bad in the recursive call, j, k will appear in a child call with an upper bound level $u' = \ell^*(y^*, i, j)$. Therefore, any bad pivot i' with $\ell^*(y^*, i', j) \geq \ell^*(y^*, i, j)$ will become good (or disappear) in the descendant calls. Let $\overline{B}^{(2)}(n')$ denote an upper bound for the expected total charges of the second type on (j, k) in a recursive call and its descendant calls, where there are at most n' bad pivots for the edge. It is clear that $\sum_{t:j,k\in t} \mathbb{E}[B^{(2)}_{j,k,t}] \leq \overline{B}^{(2)}(n-2)$. We claim that $\overline{B}^{(2)}(0) = 0$, and that $\overline{B}^{(2)}(n') \leq 4\ln n' + 2$ for any $n' \geq 1$ to finish proving that $\sum_{t:j,k\in t} \mathbb{E}[B^{(2)}_{j,k,t}] = O(\log n)$. We prove the claim by induction. The base case is when $n' = 0$. It is clear that we have $\overline{B}^{(2)}(0) = 0$ because there is no bad pivot and there is no charge of the second type on (j, k). For $n' \geq 1$, let n'_ℓ be the number of bad pivots i with $\ell^*(y, i, j) \geq \ell$ in the recursive. Based on our previous discussion, we have

$$\overline{B}^{(2)}(n') \leq 2 + \sum_{\ell > u} \frac{n'_\ell - n'_{\ell+1}}{n'} \cdot \overline{B}^{(2)}(n'_{\ell+1})$$

$$\leq 2 + \sum_{i \in [n']} \frac{1}{n'} \cdot \overline{B}^{(2)}(i-1)$$

$$\leq 2 + \sum_{i \in [n'-1]} \frac{4\ln i + 2}{n'}$$

$$\leq 4 - \frac{2}{n'} + \frac{4}{n'} \int_1^{n'} \ln x \, dx$$

$$= 4 - \frac{2}{n'} + \left(4\ln n' - 4 + \frac{4}{n'}\right) \leq 4\ln n' + 2 \ .$$

7 16-Approximation Analysis for k-Partite Graphs

In this section, we prove that Algorithm 1 is a 16-approximation algorithm for unweighted ULTRAMETRIC VIOLATION DISTANCE on complete k-partite graphs by appropriate choices of the parameters α, β. More specifically, we will prove the following theorem:

Theorem 7.1. *If $\alpha = \frac{3}{8}, \beta = \frac{2}{3}$, Algorithm 1 is a randomized polynomial-time 16-approximation algorithm for* ULTRAMETRIC VIOLATION DISTANCE *on complete k-partite graphs.*

We shall first prove the following analog Lemma 7.2 of Lemma 5.5. To ensure that endpoints of a non-pivot edge are partitioned into different sets on line 26 when one of the pivot edges is random, which is a frequently used property in the proof in Section 5, our choice of the parameter β follows the criteria $\alpha\beta$ and $\frac{\alpha(1-\beta)}{1-\alpha\beta}$ can both be lower bounded by some constant.

Lemma 7.2. *Consider any recursive call* LP-*UMVD*-PIVOT(V, x, u) *with $|V| > 2$. Suppose i is the pivot vertex of the call. For any $(j, k) \in \binom{V \setminus \{i\}}{2}$, the probability that j, k are partitioned into different sets on line 26 can be lower bounded by*

- *1 if both (i, j) and (i, k) are deterministic in the call but they have different dominant levels,*

- *$\alpha\beta$ if one of (i, j) and (i, k) is random in $E_\varnothing$, or*

- *$\frac{\alpha(1-\beta)}{1-\alpha\beta}$ if one of (i, j) and (i, k) is random in E.*

23

Proof. For the first bullet, according to the algorithm, the distances of the pivot edges (i, j) and (i, k) are deterministically different, and thus j, k are partitioned into different sets with probability 1 on line 26.

For the second bullet, suppose the truncated LP solution in the call is y. W.l.o.g., we assume that (i, j) is random in $E_\varnothing$. Note that this implies $\forall \ell \in [L], \Delta y_\ell(i, j) \leq \Delta y_{\ell^*(i,j)}(i, j) \leq 1 - \alpha\beta$. According to the CCDF (5), the probability that j, k are partitioned into different sets on line 26 is

$$1 - \sum_{\ell \in [L]} \Pr[j, k \in V_\ell] = 1 - \sum_{\ell \in [L]} \Pr[x'(i, j) = d_\ell] \cdot \Pr[x'(i, k) = d_\ell]$$

$$= 1 - \sum_{\ell \in [L]} \Delta y_\ell(i, j) \cdot \Pr[x'(i, k) = d_\ell]$$

$$\geq 1 - \sum_{\ell \in [L]} (1 - \alpha\beta) \cdot \Pr[x'(i, k) = d_\ell]$$

$$= 1 - (1 - \alpha\beta) = \alpha\beta \ .$$

For the third bullet, suppose the truncated LP solution in the call is y. W.l.o.g., we assume that (i, j) is random in E. Note that this implies $\forall \ell \in [L], \Delta y_\ell(i, j) \leq \Delta y_{\ell^*(i,j)}(i, j) \leq 1 - \alpha$. According to the CCDF (4), the probability that j, k are partitioned into different sets on line 26 is

$$1 - \sum_{\ell \in [L]} \Pr[j, k \in V_\ell] = 1 - \sum_{\ell \in [L]} \Pr[x'(i, j) = d_\ell] \cdot \Pr[x'(i, k) = d_\ell]$$

$$= 1 - \sum_{\ell \in [L]} \frac{(y_\ell(i, j) - \alpha\beta)^+ - (y_{\ell-1}(i, j) - \alpha\beta)^+}{1 - \alpha\beta} \cdot \Pr[x'(i, k) = d_\ell]$$

$$\geq 1 - \sum_{\ell \in [L]} \frac{\Delta y_\ell(i, j)}{1 - \alpha\beta} \cdot \Pr[x'(i, k) = d_\ell] \qquad \text{(LP constraint (2))}$$

$$\geq 1 - \sum_{\ell \in [L]} \frac{1 - \alpha}{1 - \alpha\beta} \cdot \Pr[x'(i, k) = d_\ell]$$

$$= 1 - \frac{1 - \alpha}{1 - \alpha\beta} = \frac{\alpha(1 - \beta)}{1 - \alpha\beta} \ .$$

$\square$

Next, we present the key Lemma 7.3 of this section, which gives the condition we need in Lemma 4.5 and Lemma 4.6 under the specific choices of α, β. As the proof of the lemma is quite technically involved by the proofs we have shown in the previous two sections, we defer it to Appendix B.7.

Lemma 7.3. *If $\alpha = \frac{3}{8}, \beta = \frac{2}{3}$, there exists a charging scheme such that*

1. for any triangle $t \in \binom{[n]}{3}$,

$$\sum_{(i,j) \in t} \mathbb{E}[M_{i,j,t} | \mathcal{A}_t] \leq \sum_{(i,j) \in t \cap E} \mathbb{E}[B_{i,j,t} | \mathcal{A}_t] \cdot c^*(i, j) \ . \tag{14}$$

2. for any $b > 0$, $(i, j) \in E$ and any $t \ni i, j$, in any recursive call, conditioning on $\mathcal{A}_t$ and that (i, j) is a non-pivot edge charged by $B_{i,j,t} = b$, the probability that i and j are partitioned into
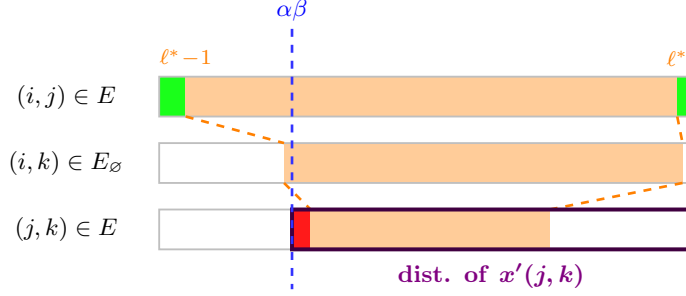
Figure 2: An example of the rounding process in (d, d, r)-same-triangles with a deterministic edge in $E_\varnothing$. The length of the red part indicates the probability that (i, j) is modified in this triangle (up to a constant factor). The total length of the green parts indicates the LP cost of (i, j).

different sets on line 26 of Algorithm 1 is at least $q_{i,j} \cdot b$, where

$$q_{i,j} = \begin{cases} 1/16 & \text{if } (i,j) \in E_\mathsf{L} \\ 1/8 & \text{if } (i,j) \in E_\mathsf{HD} \\ 3/32 & \text{if } (i,j) \in E_\mathsf{HR} \end{cases}$$

A technical highlight in this section is that we round differently for edges in E versus edges in $E_\varnothing$. This overcomes the challenge we face for (d, d, r)-same-triangles when analyzing the weighted cases with triangle inequality constraints. Specifically, in a recursive call with truncated LP solution y, when considering a (d, d, r)-same-triangle $t = (i, j, k)$, in which (j, k) is the random edge, we encounter the issue if the triangle has a deterministic edge in $E_\varnothing$, say $(i, k) \in E_\varnothing$. We can upper bound the probability that (i, j) is modified by $c^*(i, j) + c^*(i, k)$ (shown by Eqn. (13) and Lemma 5.3). However, on k-partite graphs, the LP contribution of (i, k) is 0, and we need to charge the edge (j, k) when $c^*(i, j) \ll c^*(i, k)$. As in the analysis of the weighted case, this only results in an approximation ratio of $O(\min\{L, \log n\})$. To get a constant factor approximation, we modify the algorithm so as to upper bound the probability that (i, j) is modified solely by $c^*(i, j)$. We introduce different thresholds to determine whether edges in E and $E_\varnothing$ are deterministic and round random edges in E in a (slightly) different way. As shown in Section 3, the threshold for edges in E is $1 - \alpha$ while the threshold for edges in $E_\varnothing$ is $1 - \alpha\beta$. Suppose ℓ^* is the dominant level of (i, j). According to the CCDF (4), the distance of the random edge $(j, k) \in E$ is strictly greater than d_{ℓ^*} with probability $\frac{(y_{\ell^*-1}(j,k) - \alpha\beta)^+}{1 - \alpha\beta}$. Using the LP constraint (1) and the fact that the deterministic edge (i, k) has $y_{\ell^*-1}(i, k) < \alpha\beta$, we can upper bound the probability that (i, j) is modified solely by $O(y_{\ell^*-1}(i, j))$. The detailed analysis can be found in Case 2 of Appendix B.7.2. See Figure 2 for an illustration of why our algorithm resolves this issue.

Finally, we establish the approximation ratio for k-partite graphs.

Proof of Theorem 7.1. Because of Lemma 4.6 and the second bullet of Lemma 7.3, we can present the second condition of Lemma 4.5 as follows:

$$\forall (i,j) \in E, \quad \sum_{t:i,j \in t} \mathbb{E}[B_{i,j,t}] \leq \begin{cases} 16 & \text{if } (i,j) \in E_\mathsf{L}, \\ 8 & \text{if } (i,j) \in E_\mathsf{HD}, \\ 32/3 & \text{if } (i,j) \in E_\mathsf{HR}. \end{cases}$$

25

Because the first bullet of Lemma 7.3 meets that of Lemma 4.5, if $\alpha = \frac{8}{3}, \beta = \frac{2}{3}$, Algorithm 1 is at most $\max\{16, 8 + \frac{8}{5}, \frac{32}{3} + \frac{8}{3}\} = 16$-approximation for Ultrametric Violation Distance on complete k-partite graphs. $\qquad\square$

Acknowledgment

References

[AALZ12] Nir Ailon, Noa Avigdor-Elgrabli, Edo Liberty, and Anke van Zuylen. Improved approximation algorithms for bipartite correlation clustering. *SIAM J. Comput.*, 41(5):1110–1121, 2012.

[ABF+98] Richa Agarwala, Vineet Bafna, Martin Farach, Mike Paterson, and Mikkel Thorup. On the approximability of numerical taxonomy (fitting distances by tree metrics). *SIAM Journal on Computing*, 28(3):1073–1085, 1998.

[AC11] Nir Ailon and Moses Charikar. Fitting tree metrics: Hierarchical clustering and phylogeny. *SIAM J. Comput.*, 40(5):1275–1291, 2011.

[ACN08] Nir Ailon, Moses Charikar, and Alantha Newman. Aggregating inconsistent information: Ranking and clustering. *J. ACM*, 55(5):23:1–23:27, 2008.

[Ami04] Noga Amit. *The bicluster graph editing problem*. PhD thesis, Tel Aviv University, 2004.

[BBC04] Nikhil Bansal, Avrim Blum, and Shuchi Chawla. Correlation clustering. *Machine learning*, 56:89–113, 2004.

[BDST08] Justin Brickell, Inderjit S Dhillon, Suvrit Sra, and Joel A Tropp. The metric nearness problem. *SIAM Journal on Matrix Analysis and Applications*, 30(1):375–396, 2008.

[CDK+21] Vincent Cohen-Addad, Debarati Das, Evangelos Kipouridis, Nikos Parotsidis, and Mikkel Thorup. Fitting distances by tree metrics minimizing the total error within a constant factor. In *62nd IEEE Annual Symposium on Foundations of Computer Science*, pages 468–479, 2021.

[CDL21] Vincent Cohen-Addad, Rémi De Joannis De Verclos, and Guillaume Lagarde. Improving ultrametrics embeddings through coresets. In *International Conference on Machine Learning*, pages 2060–2068. PMLR, 2021.

[CE67] Luigi L Cavalli-Sforza and Anthony WF Edwards. Phylogenetic analysis. models and estimation procedures. *American journal of human genetics*, 19(3 Pt 1):233, 1967.

[CFLM22] Vincent Cohen-Addad, Chenglin Fan, Euiwoong Lee, and Arnaud de Mesmay. Fitting metrics and ultrametrics with minimum disagreements. In *63rd IEEE Annual Symposium on Foundations of Computer Science*, pages 301–311, 2022.

[CGW05] Moses Charikar, Venkatesan Guruswami, and Anthony Wirth. Clustering with qualitative information. *J. Comput. Syst. Sci.*, 71(3):360–383, 2005.

[CKK+06] Shuchi Chawla, Robert Krauthgamer, Ravi Kumar, Yuval Rabani, and D. Sivaku-mar. On the hardness of approximating multicut and sparsest-cut. *Comput. Complex.*, 15(2):94–114, 2006.

[CKL20] Vincent Cohen-Addad, CS Karthik, and Guillaume Lagarde. On efficient low distortion ultrametric embedding. In *International Conference on Machine Learning*, pages 2078–2088. PMLR, 2020.

[CLLN23] Vincent Cohen-Addad, Euiwoong Lee, Shi Li, and Alantha Newman. Handling corre-lated rounding error via preclustering: A 1.73-approximation for correlation clustering. In *64th IEEE Annual Symposium on Foundations of Computer Science*, 2023.

[CLN22] Vincent Cohen-Addad, Euiwoong Lee, and Alantha Newman. Correlation clustering with sherali-adams. In *63rd IEEE Annual Symposium on Foundations of Computer Science*, pages 651–661, 2022.

[CMSY15] Shuchi Chawla, Konstantin Makarychev, Tselil Schramm, and Grigory Yaroslavtsev. Near optimal LP rounding algorithm for correlationclustering on complete and com-plete k-partite graphs. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, pages 219–228, 2015.

[Day87] William HE Day. Computational complexity of inferring phylogenies from dissimilarity matrices. *Bulletin of mathematical biology*, 49(4):461–467, 1987.

[DEFI06] Erik D Demaine, Dotan Emanuel, Amos Fiat, and Nicole Immorlica. Correlation clus-tering in general weighted graphs. *Theoretical Computer Science*, 361(2-3):172–187, 2006.

[DPS+13] Geet Duggal, Rob Patro, Emre Sefer, Hao Wang, Darya Filippova, Samir Khuller, and Carl Kingsford. Resolving spatial inconsistencies in chromosome conformation measure-ments. *Algorithms for Molecular Biology*, 8:1–10, 2013.

[DPS15] Marco Di Summa, David Pritchard, and Laura Sanità. Finding the closest ultrametric. *Discrete Applied Mathematics*, 180:70–80, 2015.

[Far72] James S Farris. Estimating phylogenetic trees from distance matrices. *The American Naturalist*, 106(951):645–668, 1972.

[FGR+20] Chenglin Fan, Anna C Gilbert, Benjamin Raichel, Rishi Sonthalia, and Gregory Van Buskirk. Generalized metric repair on graphs. In *17th Scandinavian Symposium and Workshops on Algorithm Theory (SWAT 2020)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020.

[FKW93] Martin Farach, Sampath Kannan, and Tandy Warnow. A robust model for finding optimal evolutionary trees. In *Proceedings of the Twenty-Fifth Annual ACM Symposium on Theory of Computing*, pages 137–145, 1993.

[FRVB18] Chenglin Fan, Benjamin Raichek, and Gregory Van Buskirk. Metric violation distance: Hardness and approximation. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 196–209. SIAM, 2018.

[GJ17]   Anna C Gilbert and Lalit Jain. If it ain't broke, don't fix it: Sparse metric repair. In *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 612–619. IEEE, 2017.

[GMT07]   Aristides Gionis, Heikki Mannila, and Panayiotis Tsaparas. Clustering aggregation. *Acm transactions on knowledge discovery from data (tkdd)*, 1(1):4–es, 2007.

[Har67]   John A Hartigan. Representation of similarity matrices by trees. *Journal of the american Statistical Association*, 62(320):1140–1158, 1967.

[HKM05]   Boulos Harb, Sampath Kannan, and Andrew McGregor. Approximating the best-fit tree under l p norms. In *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques: 8th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems*, pages 123–133. Springer, 2005.

[JJS67]   CJ Jardine, Nicholas Jardine, and Robin Sibson. The structure and construction of taxonomic hierarchies. *Mathematical Biosciences*, 1(2):173–179, 1967.

[JKMM20] Jafar Jafarov, Sanchit Kalhan, Konstantin Makarychev, and Yury Makarychev. Correlation clustering with asymmetric classification errors. In *International Conference on Machine Learning*, pages 4641–4650. PMLR, 2020.

[Joh67]   Stephen C Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254, 1967.

[JS71]   Nicholas Jardine and Robin Sibson. Mathematical taxonomy. Technical report, 1971.

[Kip23]   Evangelos Kipouridis. Fitting tree metrics with minimum disagreements. In Inge Li Gørtz, Martin Farach-Colton, Simon J. Puglisi, and Grzegorz Herman, editors, *31st Annual European Symposium on Algorithms, ESA 2023, September 4-6, 2023, Amsterdam, The Netherlands*, volume 274 of *LIPIcs*, pages 70:1–70:10. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2023.

[KM86]   Mirko Křivánek and Jaroslav Morávek. Np-hard problems in hierarchical-tree clustering. *Acta informatica*, 23:311–323, 1986.

[Kra44]   Marc Krasner. Nombres semi-réels et espaces ultramétriques. *Comptes-Rendus de l'Académie des Sciences*, 2:219, 1944.

[Man99]   Rosario N Mantegna. Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, 11:193–197, 1999.

[RTV86]   Rammal Rammal, Gérard Toulouse, and Miguel Angel Virasoro. Ultrametricity for physicists. *Reviews of Modern Physics*, 58(3):765, 1986.

[SS62]   Peter HA Sneath and Robert R Sokal. Numerical taxonomy. *Nature*, 193:855–860, 1962.

[SWW17]   Anastasios Sidiropoulos, Dingkang Wang, and Yusu Wang. Metric embeddings with outliers. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 670–689. SIAM, 2017.

[WSSB77] M.S. Waterman, T.F Smith, M. Singh, and W.A. Beyer. Additive evolutionary trees. *Journal of Theoretical Biology*, 64(2):199–213, 1977.

[ZW09]   Anke van Zuylen and David P. Williamson. Deterministic pivoting algorithms for constrained ranking and clustering problems. *Math. Oper. Res.*, 34(3):594–620, 2009.

A   Examples Regarding Different Rounding Schemes

In this appendix section, we present examples demonstrating the issues with the purely randomized rounding scheme and how our rounding scheme resolves them. Our first example is an unweighted Ultrametric Violation Distance instance on complete k-partite graphs with $L = 2$, where triple-based analysis encounters issues when analyzing the purely randomized rounding scheme to a constant factor approximation.

Example A.1. *Consider an* Ultrametric Violation Distance *instance on complete k-partite graphs with $L = 2$. Consider three distinct vertices $i, j, k \in [n]$. Suppose $(i, j), (i, k) \in E, (j, k) \in E_\varnothing$, and the input satisfies $x_{in}(i, j) = d_2, x_{in}(i, k) = x_{in}(j, k) = d_1$. In the language of* Correlation Clustering*, (i, j) is a + edge while (i, k) and (j, k) are - edges. Because we assume $y_L(u, v) = 1$ for any distinct $u, v \in [n]$, we can use a, b, c to respectively denote the only variables $y_1(i, j), y_1(i, k), y_1(j, k)$ and we call them the lengths of the edges. Pick a huge $m = \omega(1)$ and a tiny $\epsilon = o(1/m)$. Suppose the LP solution gives $a = m\epsilon, b = 1 - \epsilon$ and $c = 1 - (m + 1)\epsilon$.*

The LP contributions on these three edges are respectively $m\epsilon, \epsilon, 0$ because edges in $E_\varnothing$ do not contribute to the LP objective. On the other hand, if we use the purely randomized rounding, when the pivot is k, with probability $b(1 - c) = (1 - o(1)) \cdot (m + 1)\epsilon, x'(i, k) = d_1, x'(j, k) = d_2$ and we need to modify the distance of (i, j) to d_1 accordingly. Hence, the expected cost of the algorithm in this triangle is at least $(1 - o(1)) \cdot (m + 1)\epsilon$. If we want the cost to be upper bounded by the charges on the edges, we need: either charge (i, j) by a constant multiple of its LP contribution; or charge (i, k) by a $\Omega(m)$ multiple of its LP contribution. If we want to prove a constant factor approximation, we can only place our hope on charging (i, j). However, with probability $bc = 1 - o(1), x'(i, k) = x'(j, k) = d_1$ and (i, j) will appear in another recursive call. This means that the total charge on (i, j) can be a $\Omega(1)/(1 - bc) = \omega(1)$ multiple of its LP contribution, which does not lead to a constant factor approximation analysis.

Notice that in the first example, the three edges are deterministic if α is some positive constant and they have very low LP contributions. Our second example is an arbitrary weighted Ultrametric Violation Distance instance with $L = 2$, which is more general than the k-partite cases. In this example, our rounding scheme will not modify any edge in such triangles. This example is a formal version of the example we present in the Introduction.

Example A.2. *Consider any weighted* Ultrametric Violation Distance *instance with $L = 2$ and the triangle inequality constraint. Suppose we choose parameters $\alpha = \frac{1}{3}, \beta = 0$ in Algorithm 1. Consider three distinct vertices $i, j, k \in [n]$. Because we have w.l.o.g. assumed $y_L(u, v) = 1$ for any distinct $u, v \in [n]$, we can use a, b, c to denote the only variables $y_1(i, j), y_1(i, k), y_1(j, k)$ and we call them the lengths of the edges. Consider the scenario where $(i, j), (i, k)$ and (j, k) are all deterministic, i.e., $a, b, c \in [0, \frac{1}{3}) \cup (\frac{2}{3}, 1]$. If the length of an edge is less than $\frac{1}{3}$, the dominant level is 2; otherwise, the dominant level is 1. In addition, we assume that all three edges have LP contributions less than $\frac{1}{3}$, which implies their input equals their dominant distance level. To show that our algorithm does not modify any edge in this triangle, it suffices to show that the ultrametric inequality is automatically satisfied if we set the distances to the dominant levels.*

By setting the distances to the dominant distance levels, the only possible violation of the ultrametric inequality is when two of them have dominant level 2 and the other one has dominant level

1, w.l.o.g. say, (j, k) is the only edge with dominant level 1. If this violation happens, we will have $a, b < \frac{1}{3}$ and $c > \frac{2}{3}$, violating the LP constraint (1) *by $a + b < c$.*

B Omitted Proofs

B.1 Proof of Lemma 3.3

First, we can observe that the upper bound level u automatically implies an upper bound for the input:

Lemma B.1. *In any call* LP-*UMVD*-PIVOT(V, x, u) *of our algorithm, the input x satisfies $x(i, j) \leq d_u$ for any $(i, j) \in \binom{V}{2}$.*

Proof. We prove the lemma by induction. The base case is clear for the root call because $u = 1$ and d_1 equals the largest entry in x_{in}. Suppose the lemma holds for LP-UMVD-PIVOT(V, x, u), whose random pivot equals i, and consider any of its children calls LP-UMVD-PIVOT(V_ℓ, x', ℓ). According to the definition of V_ℓ (line 26), for any $j, k \in V_\ell$, $x'(i, j) = x'(i, k) = d_\ell$. Because the algorithm sets $x'(j, k)$ to $\min\{x(j, k), x'(i, j)\}$ in this case, we have $x'(j, k) \leq d_\ell$ and we prove the lemma for this child call. $\square$

Next, we show a stronger version of the lemma by induction: each call LP-UMVD-PIVOT(V, x, u) returns an ultrametric x' satisfying $x'(i, j) \leq d_u$ for any $(i, j) \in \binom{V}{2}$. The base cases are clear when $|V| \leq 2$ or $u = L$.

Consider a call with $|V| = n_0$ ($n_0 \geq 3$) and $u < L$. Suppose we have shown the stronger version for any $|V| < n_0$. Observe that the set of return values in x' is a subset of those in x. Because of Lemma B.1, we have $x'(i, j) \leq d_u$ for any $(i, j) \in \binom{V}{2}$. Suppose i is the random pivot of the call. We discuss two cases to show that the returned x' satisfies the ultrametric inequality for any triangle in V.

The triangle involves i. Fix any distinct $j, k \in V \setminus \{i\}$. If $x'(i, j) = x'(i, k)$, j, k will appear in a child call, and thus the ultrametric inequality is satisfied because $x'(j, k) \leq x(i, j)$ according to the induction hypothesis. If $x'(i, j) \neq x'(i, k)$, the algorithm sets $x'(j, k)$ to $\max\{x'(i, j), x'(i, k)\}$ and thus the ultrametric inequality is satisfied.

The triangle does not involve i. Fix any distinct $j, k, r \in V \setminus \{i\}$. If $x'(i, j) = x'(i, k) = x'(i, r)$, the vertices appear in a child call and thus satisfy the ultrametric inequality according to the induction hypothesis. Otherwise, either one of them is strictly greater than two others or two of them are equal and strictly greater than the other one. For the first case, w.l.o.g., suppose $x'(i, r) > x'(i, j), x'(i, k)$. The algorithm then sets $x'(j, r)$ and $x'(k, r)$ to $x'(i, r)$. Therefore, $x'(j, k) = \max\{x'(i, j), x'(i, k)\}$ or j, k will appear in a child call, where we both have $x'(j, k) < x'(j, r)$ according to the induction hypothesis. For the second case, w.l.o.g., suppose $x'(i, r) < x'(i, j) = x'(i, k)$. The algorithm then sets $x'(j, r)$ and $x'(k, r)$ to $x'(i, j)$. And, j, k will appear in a child call and thus $x'(j, k) \leq x'(i, j)$ according to the induction hypothesis.

B.2 Proof of Lemma 4.1

Because (i, j) is initially deterministic, if the dominant level $\tilde{\ell}(i, j) = \ell^*(y^*, i, j)$, $c^*(i, j) = 1 - \Delta y^*_{\ell^*(i,j)}(i, j) < \alpha$, violating (i, j) is a high-cost edge. Therefore, $\tilde{\ell}(i, j) \neq \ell^*(y^*, i, j)$. Since $y_0^*(i, j) = 0$ and $y_L^*(i, j) = 1$, $\sum_{\ell=1}^{L} \Delta y_\ell^*(i, j) = 1$. Hence, $c^*(i, j) = 1 - \Delta y^*_{\tilde{\ell}(i,j)}(i, j) \geq \Delta y^*_{\ell^*(i,j)}(i, j) > 1 - \alpha$.

B.3 Proof of Lemma 4.2

Suppose u is the level in the recursive call and u' is the level in the child call. According to the algorithm, we have $u' \geq u$. Fix any edge $(i,j) \in \binom{V'}{2}$.

If $\ell^*(y,i,j) > u'$, since the algorithm sets both $y_\ell(i,j)$ and $y'_\ell(i,j)$ to $y^*_\ell(i,j)$ for any $\ell \in \{u', u'+1, \cdots, L\}$, we have $\Delta y'_{\ell^*(y,i,j)}(i,j) = \Delta y^*_{\ell^*(y,i,j)}(i,j) = \Delta y_{\ell^*(y,i,j)}(i,j)$. Therefore, according to the definition of the dominant level, $\Delta y'_{\ell^*(i,j)}(i,j) \geq \Delta y'_{\ell^*(y,i,j)}(i,j) = \Delta y_{\ell^*(i,j)}(i,j)$.

Notice that $\Delta y_\ell(i,j) \leq y^*_\ell(i,j)$ for any $\ell \in [L]$ and $\Delta y'_{u'}(i,j) = y^*_{u'}(i,j)$. On the other hand, if $\ell^*(y,i,j) \leq u'$, because of Lemma 3.1 and the LP constraint (2), we have $\Delta y'_{u'}(i,j) = y^*_{u'}(i,j) \geq y^*_{\ell^*(y,i,j)}(i,j) \geq \Delta y_{\ell^*(i,j)}(i,j)$ and thus $\Delta y'_{\ell^*(i,j)}(i,j) \geq \Delta y_{\ell^*(i,j)}(i,j)$.

In particular, in the root call, because $u = 1$, y^* equals the truncated LP solution. Therefore, for any $(i,j) \in \binom{[n]}{2}$, $\Delta y_{\ell^*(i,j)}(i,j) \geq \Delta y^*_{\ell^*(i,j)}(i,j)$.

B.4 Proof of Lemma 4.4

We prove the lemma by a top-down induction on the recursive calls. The base case is the root call LP-UMVD-PIVOT($[n], x_{\text{in}}, 1$). In the root call, because $u = 1$, the truncated LP solution y equals y^*. For any $(i,j) \in E_L$, according to the definition of low-cost edges, $\Delta y_{\tilde{\ell}(i,j)}(i,j) = 1 - c^*(i,j) > 1 - \alpha$. From Lemma 3.1 and $\alpha \leq \frac{1}{2}$, $\tilde{\ell}(i,j) = \ell^*(y,i,j)$. Since $x(i,j) = d_{\tilde{\ell}(i,j)}$ in the root call, we prove the lemma for the base case.

Consider any call LP-UMVD-PIVOT(V, x, u) and any child call LP-UMVD-PIVOT(V_ℓ, x', ℓ). Note that $\ell \geq u$ and the input distance x' of the child call is consistent with the x' in the parent call before partitioning non-pivot vertices (line 26). Assume we have proved the lemma for the parent call. Suppose the truncated LP solutions in the calls are respectively y and y' and the pivot vertex of the parent call is i. Consider any edge $(j,k) \in \binom{V_\ell}{2}$. According to the definition of V_ℓ, $x'(i,j) = x'(i,k) = d_\ell$. Because the algorithm sets $x'(j,k)$ to $\min\{x(j,k), d_\ell\}$ in the parent call, we can then prove the lemma for the child call by discussing the following two cases.

1. If $x(j,k) \geq d_\ell$, $x'(j,k) = d_\ell$. According to the induction hypothesis, $d_{\ell^*(y,j,k)} = x(j,k) \geq d_\ell$ and thus $\ell^*(y,i,j) \leq \ell$. Because of the definition of the truncated LP solutions, the LP constraint (2) and Corollary 4.3, we have $\Delta y'_\ell(i,j) = y^*_\ell(i,j) \geq y^*_{\ell^*(y,i,j)}(i,j) \geq \Delta y_{\ell^*(i,j)}(i,j) > 1 - \alpha$. Since $\sum_{\ell \in [L]} \Delta y'_\ell(j,k) = 1$ (Lemma 3.1) and $\alpha \leq \frac{1}{2}$, $\ell^*(y', j, k) = \ell$ and thus $x'(j,k) = d_{\ell^*(y',j,k)}$.

2. If $x(j,k) < d_\ell$, $x'(j,k) = x(j,k)$. According to the induction hypothesis, $d_{\ell^*(y,j,k)} = x(j,k)$ and thus $x'(j,k) = d_{\ell^*(y,i,j)}$ and $\ell^*(y,j,k) > \ell \geq u$. Because of the definition of the truncated LP solutions and Corollary 4.3, $\Delta y'_{\ell^*(y,j,k)}(j,k) = \Delta y^*_{\ell^*(y,j,k)}(j,k) = \Delta y_{\ell^*(j,k)}(j,k) > 1 - \alpha$. Since $\sum_{\ell \in [L]} \Delta y'_\ell(j,k) = 1$ (Lemma 3.1) and $\alpha \leq \frac{1}{2}$, $\ell^*(y',i,j) = \ell^*(y,i,j)$ and thus $x'(j,k) = d_{\ell^*(y',i,j)}$.

In particular, because the algorithm sets the distances of deterministic pivot edges to its dominant distance level, any edge $(i,j) \in E_L$ is not modified in any recursive call where it appears as a pivot edge.

B.5 Missing proof of Lemma 5.2

B.5.1 (d,d,r)-diff-triangles

W.l.o.g. we assume (i,j) and (i,k) are deterministic, and $\ell^*(y,i,j) < \ell^*(y,i,k)$ (equivalently, $d_{\ell^*(y,i,j)} > d_{\ell^*(y,i,k)}$). Because of Lemma 4.4, the input distances of (i,j) and (i,k) are respectively

$d_{\ell^*(y,i,j)}$ and $d_{\ell^*(y,i,k)}$. Since $d_{\ell^*(y,i,j)} > d_{\ell^*(y,i,k)}$, $M_{i,j,t}$ (or $M_{i,k,t}$) equals 1 only when the pivot is k (resp., j), it is low-cost and the random distance $x'(j,k) \neq d_{\ell^*(y,i,j)}$, i.e.,

$$M_{i,j,t} = \mathbf{1}(k \text{ is the pivot vertex}) \cdot \mathbf{1}((i,j) \in E_L) \cdot \mathbf{1}(x'(j,k) \neq d_{\ell^*(y,i,j)}) \,,$$
$$M_{i,k,t} = \mathbf{1}(j \text{ is the pivot vertex}) \cdot \mathbf{1}((i,k) \in E_L) \cdot \mathbf{1}(x'(j,k) \neq d_{\ell^*(y,i,j)}) \,.$$

According to the CCDF (11) of the random distance, because y is feasible in (UMVD LP), the probability that $x'(j,k) \neq d_{\ell^*(y,i,j)}$ is

$$
\begin{aligned}
1 - \Delta y_{\ell^*(i,j)}(j,k) &= 1 - y_{\ell^*(i,j)}(j,k) + y_{\ell^*(i,j)-1}(j,k) \\
&\leq 1 - (y_{\ell^*(i,j)}(i,j) - y_{\ell^*(i,j)}(i,k)) + y_{\ell^*(i,j)-1}(i,j) + y_{\ell^*(i,j)-1}(i,k) \\
&\leq (1 - \Delta y_{\ell^*(i,j)}(i,j)) + 2y_{\ell^*(i,k)-1}(i,k) \,.
\end{aligned}
\tag{15}
$$

Because of Corollary 4.3, $(j,k) \in E_{HR}$ and thus $c^*(i,j) \geq \alpha$. Note that the deterministic edge (i,k) satisfies $y_{\ell^*(i,k)-1}(i,k) < \alpha$ (Eqn. (12)). We can upper bound $1 - \Delta y_{\ell^*(i,j)}(j,k)$ by $1 - \Delta y_{\ell^*(i,j)}(i,j) + y_{\ell^*(i,k)-1}(i,k) + c^*(j,k)$. Therefore, in this class of triangles,

$$\sum_{(i',j') \in t} \mathbb{E}[M_{i',j',t}] \leq \frac{2\big(1 - \Delta y_{\ell^*(i,j)}(i,j) + y_{\ell^*(i,k)-1}(i,k) + c^*(j,k)\big)}{3} \,.$$

Accordingly, when the corresponding edge is non-pivot, we define $B_{i,j,t}, B_{i,k,t}, B_{j,k,t}$ as follows:

$$B_{i,j,t} \overset{\text{def}}{=} \begin{cases} 2 & \text{if } (i,j) \in E_L \\ \frac{2\alpha}{1-\alpha} & \text{if } (i,j) \in E_{HD} \,, \\ 1 & \text{if } (i,j) \in E_{HR} \end{cases} \qquad B_{i,k,t} \overset{\text{def}}{=} \begin{cases} 2 & \text{if } (i,k) \in E_L \\ \frac{2\alpha}{1-\alpha} & \text{if } (i,k) \in E_{HD} \,, \\ 0 & \text{if } (i,k) \in E_{HR} \end{cases} \qquad B_{j,k,t} \overset{\text{def}}{=} 2$$

Because of Lemma 5.3 and 5.4, this definition satisfies

$$\mathbb{E}[B_{i,j,t}] \cdot c^*(i,j) \geq \frac{2}{3} \cdot (1 - \Delta y_{\ell^*(i,j)}(i,j)) \,,$$
$$\mathbb{E}[B_{i,k,t}] \cdot c^*(i,k) \geq \frac{2}{3} \cdot y_{\ell^*(i,k)-1}(i,k) \,, \quad \mathbb{E}[B_{j,k,t}] \cdot c^*(j,k) \geq \frac{2}{3} \cdot c^*(j,k) \,,$$

and thus implies the first bullet of Lemma 5.2.

Further, when the edge (i,j) (or (i,k)) is non-pivot, one of the pivot edges in the triangle is random. Because of Lemma 5.5, the probability i,j (or i,k) are then partitioned into different sets on line 26 is at least α, which is at least

$$\begin{cases} \frac{\alpha}{2} \cdot B_{i,j,t} & \text{if } (i,j) \in E_L \\ \frac{1-\alpha}{2} \cdot B_{i,j,t} & \text{if } (i,j) \in E_{HD} \quad \text{and} \\ \alpha \cdot B_{i,j,t} & \text{if } (i,j) \in E_{HR} \end{cases} \qquad \begin{cases} \frac{\alpha}{2} \cdot B_{i,k,t} & \text{if } (i,k) \in E_L \\ \frac{1-\alpha}{2} \cdot B_{i,k,t} & \text{if } (i,k) \in E_{HD} \,. \\ 1 \cdot B_{i,k,t} & \text{if } (i,k) \in E_{HR} \end{cases}$$

When the edge (j,k) is non-pivot, two pivot edges are both deterministic but have different dominant levels. Because of Lemma 5.5, the probability i,j (or i,k) are then partitioned into different sets on line 26 is 1, which is at least $\frac{1}{2} \cdot B_{j,k,t} \geq \max\{\frac{1-\alpha}{2}, \alpha\} \cdot B_{j,k,t}$ because $\alpha \leq \frac{1}{2}$. Hence, this definition satisfies the second bullet of Lemma 5.2 for this case.

B.5.2 (d, r, r)-triangles

W.l.o.g., we assume (j, k) is the deterministic edge. Because of Corollary 4.3, $(i, j), (i, k) \in E_{\mathsf{HR}}$ and thus $M_{i,j,t} = M_{i,k,t} = 0$. If $(j, k) \notin E_{\mathsf{L}}$, $M_{j,k,t} = 0$. Accordingly, by defining $B_{i,j,t}, B_{i,k,t}, B_{j,k,t} \stackrel{\text{def}}{=} 0$, we show Lemma 5.2 for the case $(j, k) \notin E_{\mathsf{L}}$.

Next, we consider the case $(j, k) \in E_{\mathsf{L}}$. Only when the pivot in the recursive call is i, the low-cost edge (j, k) can be modified. Because of Lemma 4.4, the input distance of the call satisfies $x(j, k) = d_{\ell^*(y,j,k)}$. If the random distances $x'(i, j), x'(i, k)$ satisfy $x'(i, j) = x'(i, k) \geq d_{\ell^*(y,j,k)}$ or $x'(i, j) < d_{\ell^*(y,j,k)} = x'(i, k)$ or $x'(i, j) = d_{\ell^*(y,j,k)} > x'(i, k)$, edge (j, k) is not modified, i.e., $M_{j,k,t} = 0$. Therefore, according to the CCDF (11) of the random distance, in this case,

$$
\sum_{(i',j') \in t} \mathbb{E}[M_{i',j',t}] = \frac{1}{3} \cdot \left(1 - \sum_{\ell=1}^{\ell^*(y,j,k)} \Delta y_\ell(i, j) \cdot \Delta y_\ell(i, k) - \Delta y_{\ell^*(j,k)}(i, k)(1 - y_{\ell^*(j,k)}(i, j)) \right.
$$
$$
\left. - \Delta y_{\ell^*(j,k)}(i, j)(1 - y_{\ell^*(j,k)}(i, k)) \right) . \tag{16}
$$

Based on the above equality, we can further prove the following upper bound for the expected number of modifications on low-cost edges in this triangle.

Lemma B.2. *If $\alpha \in [\frac{3 - \sqrt{5}}{2}, \frac{1}{2}]$, $\sum_{(i',j') \in t} \mathbb{E}[M_{i',j',t}] \leq \frac{1}{3} \cdot (c^*(i, j) + c^*(i, k) + c^*(j, k))$.*

Proof. We shall prove the lemma by discussing the magnitude of $\ell^*(y, i, j)$, $\ell^*(y, i, k)$, and $\ell^*(y, j, k)$.

If $\ell^*(y, i, j), \ell^*(y, i, k) > \ell^*(y, j, k)$,

$$
\begin{aligned}
c^*(i, j) + c^*(i, k) &\geq y_{\ell^*(i,j)-1}(i, j) + y_{\ell^*(i,k)-1}(i, k) & \text{(Lemma 5.3)} \\
&\geq y_{\ell^*(j,k)}(i, j) + y_{\ell^*(j,k)}(i, k) & \text{(LP constraint (2))} \\
&\geq y_{\ell^*(j,k)}(j, k) & \text{(LP constraint (1))} \\
&\geq 1 - c^*(j, k) . & \text{(Lemma 5.3)}
\end{aligned}
$$

Because the right-hand side of Eqn. (16) can be trivially upper bounded by $\frac{1}{3}$, $\sum_{(i',j') \in t} \mathbb{E}[M_{i',j',t}] \leq \frac{1}{3} \cdot (c^*(i, j) + c^*(i, k) + c^*(j, k))$.

If $\ell^*(y, i, j) < \ell^*(y, j, k)$ (or $\ell^*(y, i, k) < \ell^*(y, j, k)$ in symmetry), we can upper bound Eqn. (16) by $\frac{1}{3} \cdot (1 - \Delta y_{\ell^*(i,j)}(i, j) \cdot \Delta y_{\ell^*(i,j)}(i, k))$. Because y satisfies the LP constraint (1) and (2),

$$
y_{\ell^*(i,j)}(i, j) - y_{\ell^*(i,j)}(i, k) \leq y_{\ell^*(i,j)}(j, k) \leq y_{\ell^*(j,k)-1}(j, k) ,
$$
$$
y_{\ell^*(i,j)-1}(i, j) - y_{\ell^*(i,j)-1}(i, k) \geq -y_{\ell^*(i,j)-1}(j, k) \geq -y_{\ell^*(j,k)-1}(j, k) ,
$$

and thus $\Delta y_{\ell^*(i,j)}(i, j) - \Delta y_{\ell^*(i,j)}(i, k) \leq 2y_{\ell^*(j,k)-1}(j, k)$. Because of Lemma 5.3, $\Delta y_{\ell^*(i,j)}(i, j) \geq 1 - c^*(i, j)$ and $y_{\ell^*(j,k)-1}(j, k) \leq c^*(j, k)$. Therefore, $\Delta y_{\ell^*(i,j)}(i, k) \geq 1 - c^*(i, j) - 2c^*(j, k)$. We can upper bound Eqn. (16) in this case by

$$
\begin{aligned}
\frac{1}{3} \cdot \left(1 - \Delta y_{\ell^*(i,j)}(i, j) \cdot \Delta y_{\ell^*(i,j)}(i, k) \right) &\leq \frac{1}{3} \cdot \left(1 - (1 - c^*(i, j)) \cdot (1 - c^*(i, j) - 2c^*(j, k)) \right) \tag{17} \\
&\leq \frac{1}{3} \cdot \left(2(1 - c^*(i, j)) \cdot c^*(j, k) + 1 - (1 - c^*(i, j))^2 \right) \\
&= \frac{1}{3} \cdot \left((2 - 2c^*(i, j)) \cdot c^*(j, k) + (2 - c^*(i, j)) \cdot c^*(i, j) \right) \\
&\leq \frac{2 - \alpha}{3} \cdot \left(c^*(i, j) + c^*(j, k) \right) . & (c^*(i, j) \geq \alpha)
\end{aligned}
$$

33

On the other hand, because the random edge $(i, k) \in E_{\mathsf{HR}}$, we can also upper bound Eqn. (16) by $\frac{c^*(i,k)}{3\alpha}$. Averaging the previous two upper bounds of Eqn. (16), we can get

$$\sum_{(i',j')\in t} \mathbb{E}[M_{i',j',t}] \leq \frac{1}{1+2\alpha-\alpha^2} \cdot \frac{2-\alpha}{3} \cdot \left(c^*(i,j) + c^*(j,k)\right) + \frac{2\alpha-\alpha^2}{1+2\alpha-\alpha^2} \cdot \frac{c^*(i,k)}{3\alpha}$$

$$\leq \frac{2-\alpha}{3(1+2\alpha-\alpha^2)} \cdot \left(c^*(i,j) + c^*(i,k) + c^*(j,k)\right)$$

$$\leq \frac{1}{3} \cdot \left(c^*(i,j) + c^*(i,k) + c^*(j,k)\right) . \qquad (\forall \alpha \in \left[\tfrac{3-\sqrt{5}}{2}, \tfrac{1}{2}\right], 2-\alpha \leq 1+2\alpha-\alpha^2)$$

If $\ell^*(y, i, j) = \ell^*(y, j, k)$ (or $\ell^*(y, i, k) = \ell^*(y, j, k)$ in symmetry), we can upper bound Eqn. (16) by $1 - \Delta y_{\ell^*(i,j)}(i,j) \cdot (1 - y_{\ell^*(i,j)-1}(i,k))$. Because of Lemma 5.3 and because the truncated LP solution y satisfies LP constraint (1),

$$y_{\ell^*(i,j)-1}(i,k) \leq y_{\ell^*(i,j)-1}(i,j) + y_{\ell^*(i,j)-1}(j,k) \leq c^*(i,j) + c^*(j,k) .$$

Hence, we can further upper bound Eqn. (16) by

$$\frac{1}{3} \cdot \left(1 - (1 - c^*(i,j)) \cdot (1 - c^*(i,j) - c^*(j,k))\right) ,$$

which is stronger than (17). Following the steps in the previous case, we can then similarly show $\sum_{(i',j')\in t} \mathbb{E}[M_{i',j',t}] \leq \frac{1}{3} \cdot (c^*(i,j) + c^*(i,k) + c^*(j,k))$. $\qquad \square$

Accordingly, when the corresponding edge is non-pivot, we define $B_{i,j,t}, B_{i,k,t}, B_{j,k,t}$ as follows:

$$B_{i,j,t}, \; B_{i,k,t}, \; B_{j,k,t} \stackrel{\text{def}}{=} 1 ,$$

which implies the first bullet of Lemma 5.2:

$$\sum_{(i',j')\in t} \mathbb{E}[B_{i',j',t}] \cdot c^*(i',j') = \frac{1}{3} \cdot \left(c^*(i,j) + c^*(i,k) + c^*(j,k)\right) \geq \sum_{(i',j')\in t} \mathbb{E}[M_{i',j',t}] .$$

Note that because of Lemma 5.5, when being non-pivot, endpoints of all three edges in the triangle will be partitioned into different sets on line 26 with probability at least α, which is $\alpha \cdot B_{i,j,t} \geq \frac{1-\alpha}{2} \cdot B_{i,j,t}$ (resp., for $B_{i,k,t}$ and $B_{j,k,t}$) when $\alpha \geq \frac{3-\sqrt{5}}{2}$. Therefore, this definition of $B_{i',j',t}$s satisfies the second bullet of Lemma 5.2 for this class of triangles.

B.6 Missing proof of Lemma 6.2

B.6.1 (d, d, r)-diff-triangles

W.l.o.g. we assume (i, j) and (i, k) are deterministic, and (i, j) has a lower dominant level, i.e., $\ell^*(y, i, j) < \ell^*(y, i, k)$ (equivalently, $d_{\ell^*(y,i,j)} > d_{\ell^*(y,i,k)}$). Because of Corollary 4.3, we have $(j, k) \in E_{\mathsf{L}}, M_{j,k,t} = 0$ and $c^*(j, k) \geq \alpha$. Recall that in Section 5, we have shown for this class of triangles:

$$M_{i,j,t} = \mathbf{1}(k \text{ is the pivot vertex}) \cdot \mathbf{1}((i,j) \in E_{\mathsf{L}}) \cdot \mathbf{1}(x'(j,k) \neq d_{\ell^*(y,i,j)}) ,$$

$$M_{i,k,t} = \mathbf{1}(j \text{ is the pivot vertex}) \cdot \mathbf{1}((i,k) \in E_{\mathsf{L}}) \cdot \mathbf{1}(x'(j,k) \neq d_{\ell^*(y,i,j)}) ,$$

and the probability that $x'(j,k) \neq d_{\ell^*(y,i,j)}$ is at most $(1 - \Delta y_{\ell^*(i,j)}(i,j)) + 2y_{\ell^*(i,k)-1}(i,k)$. Because of Lemma 5.3, $1 - \Delta y_{\ell^*(i,j)}(i,j) \leq c^*(i,j)$, $y_{\ell^*(i,k)-1}(i,k) \leq c^*(i,k)$. Because edges (i, j) and (i, k)

are deterministic, $1 - \Delta y_{\ell^*(i,j)}(i,j), y_{\ell^*(i,k)-1}(i,k) < \alpha \le c^*(j,k)$ (Eqn. (12)). Therefore, in this class of triangles, we have

$$\sum_{(i',j')\in t} w(i',j') \cdot \mathbb{E}[M_{i',j',t}] \le \frac{(1 - \Delta y_{\ell^*(i,j)}(i,j)) + 2y_{\ell^*(i,k)-1}(i,k)}{3} \cdot \big(w(i,j) + w(i,k)\big)$$

$$\le \frac{w(i,j) \cdot c^*(i,j) + 2w(i,k) \cdot c^*(i,k)}{3} + \frac{2y_{\ell^*(i,k)-1}(i,k) \cdot (w(i,k) + w(j,k))}{3}$$
$$+ \frac{(1 - \Delta y_{\ell^*(i,j)}(i,j)) \cdot (w(i,j) + w(j,k))}{3} \quad \text{(triangle inequality for } w)$$

$$\le \frac{w(i,j) \cdot c^*(i,j) + 2w(i,k) \cdot c^*(i,k)}{3} + \frac{2\big(c^*(i,k) \cdot w(i,k) + c^*(j,k) \cdot w(j,k)\big)}{3}$$
$$+ \frac{c^*(i,j) \cdot w(i,j) + c^*(j,k) \cdot w(j,k)}{3}$$

$$= \frac{2w(i,j) \cdot c^*(i,j) + 3w(j,k) \cdot c^*(j,k) + 4w(i,k) \cdot c^*(i,k)}{3} \; .$$

Accordingly, when the corresponding edge is non-pivot, we can define $B_{i,j,t}, B_{i,k,t}, B_{j,k,t} \overset{\text{def}}{=} 4$ to prove the first bullet of Lemma 6.2. Further, we will classify all charges into the first type, i.e., $\forall (u,v) \in t, B^{(1)}_{u,v,t} = B_{u,v,t}$. Note that in this class of triangles, either there is a random pivot edge or the two deterministic pivot edges have different dominant levels. Because of Lemma 5.5, when being non-pivot, the endpoints of all three edges in t will be partitioned into different sets on line 26 with probability at least $\alpha = \frac{1}{3}$, which is $\Omega(1) \cdot B^{(1)}_{u,v,t}$ for any $(u,v) \in t$. Therefore, this definition satisfies the condition of Lemma 4.6 with $q = \Omega(1)$.

B.6.2 (d, r, r)-triangles

W.l.o.g., we assume (j,k) is the deterministic edge. Because of Corollary 4.3, $(i,j),(i,k) \in E_{\text{HR}}$, and we have $M_{i,j,t} = M_{i,k,t} = 0$ and $c^*(i,j), c^*(i,k) \ge \alpha = \frac{1}{3}$. Therefore, in this class of triangles,

$$\sum_{(i',j')\in t} w(i',j') \cdot \mathbb{E}[M_{i',j',t}] \le \frac{w(j,k)}{3} \le \frac{w(i,j) + w(i,k)}{3} \le w(i,j) \cdot c^*(i,j) + w(i,k) \cdot c^*(i,k) \; .$$

Accordingly, when the corresponding edge is non-pivot, we can define $B_{i,j,t} \overset{\text{def}}{=} 0$ and $B_{i,k,t}, B_{j,k,t} \overset{\text{def}}{=} 3$ to prove the first bullet of Lemma 6.2. Further, we will classify all non-trivial charges into the first type, i.e., $B^{(1)}_{i,k,t} = B_{i,k,t}, B^{(1)}_{j,k,t} = B_{j,k,t}$. Because of Lemma 5.5, when being non-pivot, i, k (or j, k) will be partitioned into different sets on line 26 with probability at least $\alpha = \frac{1}{3}$, which is $\Omega(1) \cdot B^{(1)}_{i,k,t}$ (resp., for $B^{(1)}_{j,k,t}$). Therefore, this definition satisfies the condition of Lemma 4.6 with $q = \Omega(1)$.

B.7 Proof of Lemma 7.3

For convenience, we assume $\alpha = \frac{3}{8}, \beta = \frac{2}{3}$ in the proof. With this assumption, the CCDF functions (4) and (5) respectively become:

$$\text{for } (i,j) \in E: \quad \forall \ell \in [L], \quad \Pr[x'(i,j) \ge d_\ell] = \frac{(4 \cdot y_\ell(i,j) - 1)^+}{3} \; . \tag{18}$$

$$\text{for } (i,j) \in E_\varnothing: \quad \forall \ell \in [L], \quad \Pr[x'(i,j) \ge d_\ell] = y_\ell(i,j) \; . \tag{19}$$

Given the truncated LP solution y, we define an edge (i,j) to be deterministic if

- $(i,j) \in E$ and $\Delta y_{\ell^*(i,j)}(i,j) > 1 - \alpha = \frac{5}{8}$, or

- $(i,j) \in E_\varnothing$ and $\Delta y_{\ell^*(i,j)}(i,j) > 1 - \alpha\beta = \frac{3}{4}$.

Otherwise, we define it to be random. In addition, we restate Lemma 7.2 with the specific choice of α, β in the following lemma.

Lemma B.3 (Restatement of Lemma 7.2). *Consider any recursive call* LP-*UMVD*-PIVOT(V, x, u) *with* $|V| > 2$. *Suppose* i *is the pivot vertex of the call. For any* $(j,k) \in \binom{V \setminus \{i\}}{2}$, *the probability that* j, k *are partitioned into different sets on line 26 can be lower bounded by*

- 1 *if both* (i,j) *and* (i,k) *are deterministic in the call but they have different dominant levels,*

- 1/4 *if one of* (i,j) *and* (i,k) *is random in* $E_\varnothing$, *or*

- 1/6 *if one of* (i,j) *and* (i,k) *is random in* E.

Corollary B.4. *Consider any recursive call* LP-*UMVD*-PIVOT(V, x, u) *with* $|V| > 2$. *Suppose* i *is the pivot vertex of the call. For any* $(j,k) \in \binom{V \setminus \{i\}}{2}$, *the probability that* j, k *are partitioned into different sets on line 26 can be lower bounded by* $1/6$ *if at least one of* (i,j) *and* (i,k) *is random.*

Note that each triangle cannot have exactly one edge in E because E is a k-partite graph. When a triangle does not have any edge in E, no edge in the triangle is low-cost, and thus $M_{i,j,t} = M_{i,k,t} = M_{j,k,t} = 0$. By trivially defining $B_{i,j,t}, B_{i,k,t}, B_{j,k,t} \overset{\text{def}}{=} 0$, we can prove Lemma 7.3. Therefore, it suffices to only consider the triangles with at least two edges in E in the rest of the analysis.

We prove the lemma by presenting the charging scheme for each class of triangles. Consider any triangle $t = (i,j,k)$ and any recursive call with a vertex set involving all of i, j, k. Suppose the recursive call has a truncated LP solution y. For convenience, we will condition on $\mathcal{A}_t$ in the rest of this section, and all expectations will be automatically conditioned on $\mathcal{A}_t$.

B.7.1 (d, d, d)-triangles

Note that the choice of β does not influence how Algorithm 1 rounds deterministic pivot edges in E. In Section 5, when all three edges are all in E, we have shown Lemma 5.2 for this class of triangles, which gives a stronger condition of Lemma 4.6 than that of Lemma 7.3 under $\alpha = \frac{3}{8}$. Next, we only need to analyze the case when two edges are in E for this class of triangles.

W.l.o.g., we assume $(i,j), (i,k) \in E$ and $(j,k) \in E_\varnothing$. Note that the distance of any pivot edge $(i', j') \in t$ is set to its dominant distance level $d_{\ell^*(y,i',j')}$. Because of Lemma 4.4, any low-cost edge (i', j') has an input distance $x(i', j') = d_{\ell^*(y,i',j')}$. If the non-pivot edge is low-cost, it is modified only when the ultrametric inequality is violated on the three dominant distance levels, i.e., there exists a permutation (i', j', k') of i, j, k such that $\ell^*(y, i', j') < \min\{\ell^*(y, i', k'), \ell^*(y, j', k')\}$. Because $\ell^*(y, i', j') < \min\{\ell^*(y, i', k'), \ell^*(y, j', k')\}$ implies

$$
\begin{aligned}
1 - \Delta y_{\ell^*(i',j')}(i',j') &\geq 1 - y_{\ell^*(i',j')}(i',j') \\
&\geq 1 - y_{\ell^*(i',j')}(i',k') - y_{\ell^*(i',j')}(j',k') &\text{(LP constraint (1))} \\
&\geq 1 - y_{\ell^*(i',k')-1}(i',k') - y_{\ell^*(j',k')-1}(j',k') &\text{(LP constraint (2))} \\
&\geq 1 - (1 - \Delta y_{\ell^*(i',k')-1}(i',k')) - (1 - \Delta y_{\ell^*(j',k')-1}(j',k')) . &(20)
\end{aligned}
$$

However, because all three edges are deterministic in the recursive call, $1 - \Delta y_{\ell^*(i,j)}(i,j), 1 - \Delta y_{\ell^*(i,k)}(i,k) < \frac{3}{8}$ and $1 - \Delta y_{\ell^*(j,k)}(j,k) < \frac{1}{4}$. Because $\frac{3}{8} \cdot 2 + \frac{1}{4} = 1$, Inequality (20) cannot hold. Therefore, the low-cost non-pivot edges are not modified in this case, and we have $M_{i,j,t} = M_{i,k,t} = M_{j,k,t} = 0$. Accordingly, by always defining $B_{i,j,t}, B_{i,k,t}, B_{j,k,t} \overset{\text{def}}{=} 0$, we show Lemma 7.3 for this case.

B.7.2 (d, d, r)-same-triangles

W.l.o.g., we assume (i, j) and (i, k) are deterministic. This implies that $M_{j,k,t} = 0$. Because of Lemma 4.4 and $\ell^*(y, i, j) = \ell^*(y, i, k)$, the input distances of (i, j) and (i, k) equal $d_{\ell^*(y,i,j)}$. Therefore, $M_{i,j,t}$ (or $M_{i,k,t}$) equals 1 only when the pivot is k (resp., j), it is low-cost and the random distance $x'(j, k) > d_{\ell^*(y,i,j)}$, i.e.,

$$M_{i,j,t} = \mathbf{1}(k \text{ is the pivot vertex}) \cdot \mathbf{1}((i, j) \in E_\mathsf{L}) \cdot \mathbf{1}(x'(j, k) > d_{\ell^*(y,i,j)}) ,$$
$$M_{i,k,t} = \mathbf{1}(j \text{ is the pivot vertex}) \cdot \mathbf{1}((i, k) \in E_\mathsf{L}) \cdot \mathbf{1}(x'(j, k) > d_{\ell^*(y,i,j)}) .$$

Note that triangles in this class with at least two edges in E can be divided into the following three cases: (1) the random edge is in $E_\varnothing$, (2) a deterministic edge is in $E_\varnothing$, and (3) all three edges are in E. Next, we shall discuss these three cases to prove Lemma 7.3 for this class of triangles.

Case 1: the random edge is in $E_\varnothing$. According to the CCDF (19) of the random distance, the probability that $x'(j, k) > d_{\ell^*(y,i,j)}$ is

$$y_{\ell^*(i,j)-1}(j, k) \le y_{\ell^*(i,j)-1}(i, j) + y_{\ell^*(i,j)-1}(i, k) \qquad \text{(LP constraint (1))}$$
$$= y_{\ell^*(i,j)-1}(i, j) + y_{\ell^*(i,k)-1}(i, k) .$$

Therefore, in this case,

$$\sum_{(i',j')\in t} \mathbb{E}[M_{i',j',t}] \le \frac{y_{\ell^*(i,j)-1}(i, j) + y_{\ell^*(i,k)-1}(i, k)}{3} \cdot \left(\mathbf{1}((i, j) \in E_\mathsf{L}) + \mathbf{1}((i, k) \in E_\mathsf{L}) \right) .$$

Accordingly, when the corresponding edge is non-pivot, we define $B_{i,j,t}, B_{i,k,t}, B_{j,k,t}$ as follows:

$$B_{i,j,t} \stackrel{\text{def}}{=} \begin{cases} 2 & \text{if } (i, j) \in E_\mathsf{L} \\ 1 & \text{if } (i, j) \in E_\mathsf{HD} \\ 0 & \text{if } (i, j) \in E_\mathsf{HR} \end{cases} , \quad B_{i,k,t} \stackrel{\text{def}}{=} \begin{cases} 2 & \text{if } (i, k) \in E_\mathsf{L} \\ 1 & \text{if } (i, k) \in E_\mathsf{HD} \\ 0 & \text{if } (i, k) \in E_\mathsf{HR} \end{cases} , \quad B_{j,k,t} \stackrel{\text{def}}{=} 0$$

Because of Lemma 5.3 and 5.4, this definition implies the first bullet of Lemma 7.3:

$$\sum_{(i',j')\in t} \mathbb{E}[B_{i',j',t}] \cdot c^*(i', j') \ge \frac{1 + \mathbf{1}((i, j) \in E_\mathsf{L})}{3} \cdot y_{\ell^*(i,j)-1}(i, j) + \frac{1 + \mathbf{1}((i, k) \in E_\mathsf{L})}{3} \cdot y_{\ell^*(i,k)-1}(i, k)$$

$$\ge \frac{\mathbf{1}((i, j) \in E_\mathsf{L}) + \mathbf{1}((i, k) \in E_\mathsf{L})}{3} \cdot \left(y_{\ell^*(i,j)-1}(i, j) + y_{\ell^*(i,k)-1}(i, k) \right)$$

$$\ge \sum_{(i',j')\in t} \mathbb{E}[M_{i',j',t}] .$$

Further, when the edge (i, j) (or (i, k)) is non-pivot, one of the pivot edges in the triangle is random. Because of Corollary B.4, the probability that i, j (or i, k) are then partitioned into different sets on line 26 is at least $\frac{1}{6}$, which is at least

$$\begin{cases} \frac{1}{12} \cdot B_{i,j,t} & \text{if } (i, j) \in E_\mathsf{L} \\ \frac{1}{6} \cdot B_{i,j,t} & \text{if } (i, j) \in E_\mathsf{HD} \\ 1 \cdot B_{i,j,t} & \text{if } (i, j) \in E_\mathsf{HR} \end{cases} \quad \text{and} \quad \begin{cases} \frac{1}{12} \cdot B_{i,k,t} & \text{if } (i, k) \in E_\mathsf{L} \\ \frac{1}{6} \cdot B_{i,k,t} & \text{if } (i, k) \in E_\mathsf{HD} \\ 1 \cdot B_{i,k,t} & \text{if } (i, k) \in E_\mathsf{HR} \end{cases}$$

Hence, this definition satisfies the second bullet of Lemma 7.3 for this case.

Case 2: one deterministic edge is in $E_\varnothing$. W.l.o.g., we assume that the deterministic edge $(i,k) \in E_\varnothing$, which implies $M_{j,k,t} = 0$ and $y_{\ell^*(i,j)-1}(i,k) = y_{\ell^*(i,k)-1}(i,k) \le 1 - \Delta y_{\ell^*(i,k)-1}(i,k) < \frac{1}{4}$. According to the CCDF (18) of the random distance, the probability that $x'(j,k) > d_{\ell^*(y,i,j)}$ is

$$\frac{\left(4 \cdot y_{\ell^*(i,j)-1}(j,k) - 1\right)^+}{3} \le \frac{\left(4 \cdot (y_{\ell^*(i,j)-1}(i,j) + y_{\ell^*(i,j)-1}(i,k)) - 1\right)^+}{3} \qquad \text{(LP constraint (1))}$$

$$\le \frac{4 \cdot y_{\ell^*(i,j)-1}(i,j)}{3} \ .$$

Therefore, in this case,

$$\sum_{(i',j') \in t} \mathbb{E}[M_{i',j',t}] \le \frac{4 \cdot \mathbf{1}((i,j) \in E_\mathsf{L}) \cdot y_{\ell^*(i,j)-1}(i,j)}{9} \ .$$

Accordingly, when the corresponding edge is non-pivot, we define $B_{i,j,t}, B_{i,k,t}, B_{j,k,t}$ as follows:

$$B_{i,j,t} \overset{\text{def}}{=} \begin{cases} 4/3 & \text{if } (i,j) \in E_\mathsf{L} \\ 0 & \text{otherwise} \end{cases} \quad , \qquad B_{i,k,t}, B_{j,k,t} \overset{\text{def}}{=} 0 \ .$$

Because of Lemma 5.3, this definition implies the first bullet of Lemma 7.3:

$$\sum_{(i',j') \in t \cap E} \mathbb{E}[B_{i,j,t}] \cdot c^*(i,j) \ge \frac{4 \cdot \mathbf{1}((i,j) \in E_\mathsf{L}) \cdot y_{\ell^*(i,j)-1}(i,j)}{9} \ge \sum_{(i',j') \in t} \mathbb{E}[M_{i',j',t}] \ .$$

Further, when the edge (i,j) is non-pivot, one of the pivot edges in the triangle is random. Because of Corollary B.4, the probability that i,j are then partitioned into different sets on line 26 is at least $\frac{1}{6}$, which is at least $\frac{1}{8} \cdot B_{i,j,t}$. Hence, this definition satisfies the second bullet of Lemma 7.3 for this class of triangles.

Case 3: all three edges are in E. According to the CCDF (4) of the random distance, the probability that $x'(j,k) > d_{\ell^*(y,i,j)}$ is

$$\frac{\left(4 \cdot y_{\ell^*(i,j)-1}(j,k) - 1\right)^+}{3} \le \frac{4 \cdot y_{\ell^*(i,j)-1}(j,k)}{3} \le \frac{4 \cdot (y_{\ell^*(i,j)-1}(i,j) + y_{\ell^*(i,j)-1}(i,k))}{3}$$

$$\text{(LP constraint (1))}$$

$$= \frac{4 \cdot (y_{\ell^*(i,j)-1}(i,j) + y_{\ell^*(i,k)-1}(i,k))}{3} \ .$$

Therefore, in this case,

$$\sum_{(i',j') \in t} \mathbb{E}[M_{i',j',t}] \le \frac{4 \cdot (y_{\ell^*(i,j)-1}(i,j) + y_{\ell^*(i,k)-1}(i,k))}{9} \cdot \left(\mathbf{1}((i,j) \in E_\mathsf{L}) + \mathbf{1}((i,k) \in E_\mathsf{L})\right) \ .$$

Accordingly, when the corresponding edge is non-pivot, we define $B_{i,j,t}, B_{i,k,t}, B_{j,k,t}$ as follows:

$$B_{i,j,t} \overset{\text{def}}{=} \begin{cases} 8/3 & \text{if } (i,j) \in E_\mathsf{L} \\ 4/3 & \text{if } (i,j) \in E_\mathsf{HD} \\ 0 & \text{if } (i,j) \in E_\mathsf{HR} \end{cases} , \quad B_{i,k,t} \overset{\text{def}}{=} \begin{cases} 8/3 & \text{if } (i,k) \in E_\mathsf{L} \\ 4/3 & \text{if } (i,k) \in E_\mathsf{HD} \\ 0 & \text{if } (i,k) \in E_\mathsf{HR} \end{cases} , \quad B_{j,k,t} \overset{\text{def}}{=} 0$$

Because of Lemma 5.3 and 5.4, this definition implies

$$\mathbb{E}[B_{i,j,t}] \cdot c^*(i,j) \geq \frac{4 \cdot (1 + \mathbf{1}((i,j) \text{ is } \mathsf{L}))}{9} \cdot y_{\ell^*(i,j)-1}(i,j) \ ,$$

$$\mathbb{E}[B_{i,k,t}] \cdot c^*(i,k) \geq \frac{4 \cdot (1 + \mathbf{1}((i,j) \text{ is } \mathsf{L}))}{9} \cdot y_{\ell^*(i,k)-1}(i,k) \ ,$$

and thus implies the first bullet of Lemma 7.3:

$$\sum_{(i',j') \in t \cap E} \mathbb{E}[B_{i',j',t}] \cdot c^*(i',j') \geq \frac{4 \cdot (\mathbf{1}((i,j) \in E_\mathsf{L}) + \mathbf{1}((i,k) \in E_\mathsf{L}))}{9} \cdot \left(y_{\ell^*(i,j)-1}(i,j) + y_{\ell^*(i,k)-1}(i,k) \right)$$

$$\geq \sum_{(i',j') \in t} \mathbb{E}[M_{i',j',t}] \ .$$

Further, when the edge (i,j) (or (i,k)) is non-pivot, one of the pivot edges in the triangle is random. Because of Corollary B.4, the probability i,j (or i,k) are then partitioned into different sets on line 26 is at least $\frac{1}{6}$, which is at least

$$\begin{cases} \frac{1}{16} \cdot B_{i,j,t} & \text{if } (i,j) \in E_\mathsf{L} \\ \frac{1}{8} \cdot B_{i,j,t} & \text{if } (i,j) \in E_\mathsf{HD} \\ 1 \cdot B_{i,j,t} & \text{if } (i,j) \in E_\mathsf{HR} \end{cases} \quad \text{and} \quad \begin{cases} \frac{1}{16} \cdot B_{i,k,t} & \text{if } (i,k) \in E_\mathsf{L} \\ \frac{1}{8} \cdot B_{i,k,t} & \text{if } (i,k) \in E_\mathsf{HD} \\ 1 \cdot B_{i,k,t} & \text{if } (i,k) \in E_\mathsf{HR} \end{cases}$$

Hence, this definition satisfies the second bullet of Lemma 7.3 for this case.

B.7.3 (d,d,r)-diff-triangles

W.l.o.g. we assume (i,j) and (i,k) are deterministic, and (i,j) has a lower dominant level, i.e., $\ell^*(y,i,j) < \ell^*(y,i,k)$ (equivalently, $d_{\ell^*(y,i,j)} > d_{\ell^*(y,i,k)}$). This implies $M_{j,k,t} = 0$. Because of Lemma 4.4, the input distance of (i,j) (or (i,k)) equals $d_{\ell^*(y,i,j)}$ (resp. $d_{\ell^*(y,i,k)}$). Therefore, $M_{i,j,t}$ (or $M_{i,k,t}$) equals 1 only when the pivot is k (resp., j), it is low-cost and the random distance $x'(j,k) \neq d_{\ell^*(y,i,j)}$, i.e.,

$$M_{i,j,t} = \mathbf{1}(k \text{ is the pivot vertex}) \cdot \mathbf{1}((i,j) \in E_\mathsf{L}) \cdot \mathbf{1}(x'(j,k) \neq d_{\ell^*(y,i,j)}) \ ,$$

$$M_{i,k,t} = \mathbf{1}(j \text{ is the pivot vertex}) \cdot \mathbf{1}((i,k) \in E_\mathsf{L}) \cdot \mathbf{1}(x'(j,k) \neq d_{\ell^*(y,i,j)}) \ .$$

Note that triangles in this class with at least two edges in E can be divided into the following two cases: (1) the random edge is in $E_\varnothing$, (2) the random edge is in E. Next, we shall discuss these two cases to prove Lemma 7.3 for this class of triangles.

Case 1: the random edge is in $E_\varnothing$. According to the CCDF (18) of the random distance, the probability that $x'(j,k) \neq d_{\ell^*(y,i,j)}$ is

$$\begin{aligned} 1 - \Delta y_{\ell^*(i,j)}(j,k) &= 1 - y_{\ell^*(i,j)}(j,k) + y_{\ell^*(i,j)-1}(j,k) \\ &\leq 1 - (y_{\ell^*(i,j)}(i,j) - y_{\ell^*(i,j)}(i,k)) + y_{\ell^*(i,j)-1}(i,j) + y_{\ell^*(i,j)-1}(i,k) \\ &\leq (1 - \Delta y_{\ell^*(i,j)}(i,j)) + 2 y_{\ell^*(i,k)-1}(i,k) \end{aligned}$$

Therefore, in this class of triangles,

$$\sum_{(i',j') \in t} \mathbb{E}[M_{i',j',t}] \leq \frac{(1 - \Delta y_{\ell^*(i,j)}(i,j)) + 2 y_{\ell^*(i,k)-1}(i,k)}{3} \cdot \left(\mathbf{1}((i,j) \in E_\mathsf{L}) + \mathbf{1}((i,k) \in E_\mathsf{L}) \right) \ .$$

Accordingly, when the corresponding edge is non-pivot, we define $B_{i,j,t}, B_{i,k,t}, B_{j,k,t}$ as follows:

$$B_{i,j,t} \stackrel{\text{def}}{=} \begin{cases} 2 & \text{if } (i,j) \in E_{\mathsf{L}} \\ 1 & \text{if } (i,j) \in E_{\mathsf{HD}} \\ 1 & \text{if } (i,j) \in E_{\mathsf{HR}} \end{cases} , \qquad B_{i,k,t} \stackrel{\text{def}}{=} \begin{cases} 4 & \text{if } (i,k) \in E_{\mathsf{L}} \\ 2 & \text{if } (i,k) \in E_{\mathsf{HD}} \\ 0 & \text{if } (i,k) \in E_{\mathsf{HR}} \end{cases} , \qquad B_{j,k,t} \stackrel{\text{def}}{=} 0 .$$

Because of Lemma 5.3 and 5.4, this definition implies

$$c^*(i,j) \cdot \mathbb{E}[B_{i,j,t}] \geq \frac{1 + \mathbf{1}((i,j) \text{ is } \mathsf{L})}{3} \cdot (1 - \Delta y_{\ell^*(i,j)}(i,j)) ,$$

$$c^*(i,k) \cdot \mathbb{E}[B_{i,k,t}] \geq \frac{1 + \mathbf{1}((i,j) \text{ is } \mathsf{L})}{3} \cdot 2 y_{\ell^*(i,k)-1}(i,k) ,$$

and thus implies the first bullet of Lemma 7.3:

$$\sum_{(i',j') \in t \cap E} \mathbb{E}[B_{i',j',t}] \cdot c^*(i',j') \geq \frac{\mathbf{1}((i,j) \in E_{\mathsf{L}}) + \mathbf{1}((i,k) \in E_{\mathsf{L}})}{3} \cdot \left(1 - \Delta y_{\ell^*(i,j)}(i,j) + 2 y_{\ell^*(i,k)-1}(i,k)\right)$$

$$\geq \sum_{(i',j') \in t} \mathbb{E}[M_{i',j',t}] .$$

Further, when the edge (i,j) (or (i,k)) is non-pivot, one of the pivot edges in the triangle is random in $E_{\varnothing}$. Because of Lemma B.3, the probability that i,j (or i,k) are then partitioned into different sets on line 26 is at least $\frac{1}{4}$, which is at least

$$\begin{cases} \frac{1}{8} \cdot B_{i,j,t} & \text{if } (i,j) \in E_{\mathsf{L}} \\ \frac{1}{4} \cdot B_{i,j,t} & \text{if } (i,j) \in E_{\mathsf{HD}} \\ \frac{1}{4} \cdot B_{i,j,t} & \text{if } (i,j) \in E_{\mathsf{HR}} \end{cases} \quad \text{and} \quad \begin{cases} \frac{1}{16} \cdot B_{i,k,t} & \text{if } (i,k) \in E_{\mathsf{L}} \\ \frac{1}{8} \cdot B_{i,k,t} & \text{if } (i,k) \in E_{\mathsf{HD}} \\ 1 \cdot B_{i,k,t} & \text{if } (i,k) \in E_{\mathsf{HR}} \end{cases}$$

Hence, this definition satisfies the second bullet of Lemma 7.3 for this case.

Case 2: the random edge is in E. In this case, we always have $\sum_{(i',j') \in t} \mathbb{E}[M_{i',j',t}] \leq \frac{2}{3}$. Accordingly, when the corresponding edge is non-pivot, we define $B_{i,j,t}, B_{i,k,t}, B_{j,k,t}$ as follows:

$$B_{i,j,t}, B_{i,k,t} \stackrel{\text{def}}{=} 0 , \qquad B_{j,k,t} \stackrel{\text{def}}{=} 16/3 .$$

Because of Corollary 4.3, $(j,k) \in E_{\mathsf{HR}}$ and thus $c^*(j,k) \geq \frac{3}{8}$. This definition implies the first bullet of Lemma 7.3:

$$\sum_{(i',j') \in t \cap E} \mathbb{E}[B_{i',j',t}] \cdot c^*(i',j') \geq \frac{1}{3} \cdot \frac{16}{3} \cdot \frac{3}{8} = \frac{2}{3} \geq \sum_{(i',j') \in t} \mathbb{E}[M_{i',j',t}] .$$

Further, when the edge (j,k) is non-pivot, both pivot edges in the triangle are deterministic but have different dominant levels. Because of Lemma B.3, the probability that j,k are then partitioned into different sets on line 26 is 1, which is at least $\frac{3}{16} \cdot B_{j,k,t}$. Hence, this definition satisfies the second bullet of Lemma 7.3 for this case.

B.7.4 (d, r, r)**-triangles**

W.l.o.g., we assume (j, k) is the deterministic edge. Because of Corollary 4.3, $(i, j), (i, k) \notin E_{\mathsf{L}}$ and thus $M_{i,j,t} = M_{i,k,t} = 0$. If $(j, k) \notin E_{\mathsf{L}}$, no low-cost edges are modified in this triangle, and thus $M_{j,k,t} = 0$. Accordingly, by always defining $B_{i,j,t} = B_{i,k,t} = B_{j,k,t} = 0$, we show Lemma 7.3 for the case $(j, k) \notin E_{\mathsf{L}}$.

Next, we consider the case $(j, k) \in E_{\mathsf{L}}$. It is clear to upper bound $M_{i,j,t}$ by 1 and thus

$$\sum_{(i',j') \in t} \mathbb{E}[M_{i',j',t}] \leq \frac{1}{3}.$$

Since $E_{\mathsf{L}} \subseteq E$, $(j, k) \in E$. Because E is k-partite, at least one of (i, j) and (i, k) is in E. W.l.o.g., we assume $(i, j) \in E$. We shall prove Lemma 7.3 by discussing whether $(i, k) \in E$.

Case 1: $(i, k) \notin E$. Because (i, j) is random in E and because of Corollary 4.3, $(i, j) \in E_{\mathsf{HR}}$ and thus $c^*(i, j) \geq \frac{3}{8}$. Accordingly, when the corresponding edge is non-pivot, we define $B_{i,j,t}, B_{i,k,t}, B_{j,k,t}$ as follows:

$$B_{i,j,t} \stackrel{\text{def}}{=} 8/3, \qquad B_{i,k,t}, B_{j,k,t} \stackrel{\text{def}}{=} 0,$$

which implies the first bullet of Lemma 7.3 for this class of triangles:

$$\sum_{(i',j') \in t \cap E} \mathbb{E}[B_{i',j',t}] \cdot c^*(i', j') \geq \frac{1}{3} \cdot \frac{8}{3} \cdot \frac{3}{8} = \frac{1}{3} \geq \sum_{(i',j') \in t} \mathbb{E}[M_{i',j',t}].$$

Further, when the edge (i, j) is non-pivot, one of the pivot edges in the triangle is random in $E_{\varnothing}$. Because of Lemma B.3, the probability that i, j are then partitioned into different sets on line 26 is at least $\frac{1}{4}$, which equals $\frac{3}{32} \cdot B_{i,j,t}$. Because $(i, j) \in E_{\mathsf{HR}}$, this definition satisfies the second bullet of Lemma 7.3 for this case.

Case 2: $(i, k) \in E$. Because (i, j) and (i, k) are random in E and because of Corollary 4.3, $(i, j), (i, k) \in E_{\mathsf{HR}}$ and thus $c^*(i, j), c^*(i, k) \geq \frac{3}{8}$. Accordingly, when the corresponding edge is non-pivot, we define $B_{i,j,t}, B_{i,k,t}, B_{j,k,t}$ as follows:

$$B_{i,j,t}, B_{i,k,t} \stackrel{\text{def}}{=} 4/3, \qquad B_{j,k,t} \stackrel{\text{def}}{=} 0,$$

which implies the first bullet of Lemma 7.3 for this class of triangles:

$$\sum_{(i',j') \in t \cap E} \mathbb{E}[B_{i',j',t}] \cdot c^*(i', j') \geq \frac{1}{3} \cdot \frac{4}{3} \cdot \frac{3}{8} + \frac{1}{3} \cdot \frac{4}{3} \cdot \frac{3}{8} = \frac{1}{3} \geq \sum_{(i',j') \in t} \mathbb{E}[M_{i',j',t}].$$

Further, when the edge (i, j) (or (i, k)) is non-pivot, one of the pivot edges in the triangle is random in E. Because of Lemma B.3, the probability that i, j (or i, k) are then partitioned into different sets on line 26 is at least $\frac{1}{6}$, which equals $\frac{1}{8} \cdot B_{i,j,t}$ (resp., $\frac{1}{8} \cdot B_{i,k,t}$). Because $(i, j), (i, k) \in E_{\mathsf{HR}}$, this definition satisfies the second bullet of Lemma 7.3 for this case.