Discrete Comput Geom (2013) 50:977–1032
DOI 10.1007/s00454-013-9536-7

Dimension Reduction for Finite Trees in ℓ_1

James R. Lee · Arnaud de Mesmay · Mohammad Moharrami

Received: 8 August 2012 / Revised: 27 June 2013 / Accepted: 12 August 2013 /
Published online: 10 October 2013

Abstract We show that every n-point tree metric admits a $(1+\varepsilon)$-embedding into $\ell_1^{C(\varepsilon)\log n}$, for every $\varepsilon > 0$, where $C(\varepsilon) \le O\big((\frac{1}{\varepsilon})^4 \log \frac{1}{\varepsilon}\big)$. This matches the natural volume lower bound up to a factor depending only on ε. Previously, it was unknown whether even complete binary trees on n nodes could be embedded in $\ell_1^{O(\log n)}$ with $O(1)$ distortion. For complete d-ary trees, our construction achieves $C(\varepsilon) \le O\big(\frac{1}{\varepsilon^2}\big)$.

Keywords Dimension reduction · Metric embeddings · Bi-Lipschitz distortion

1 Introduction

Let $T = (V, E)$ be a finite, connected, undirected tree, equipped with a length function on edges, $\mathsf{len} : E \to [0, \infty)$. This induces a shortest-path pseudometric,

$$d_T(u, v) = \text{length of the shortest } u\text{–}v \text{ path in } T.$$

(This is a pseudometric because we may have $d(u, v) = 0$ even for distinct $u, v \in V$.) Such a metric space (V, d_T) is called a *finite tree metric.*

J. R. Lee (✉) · M. Moharrami
University of Washington, Seattle, WA, USA
e-mail: jrl@cs.washington.edu

M. Moharrami
e-mail: mmoharrami@gmail.com

A. de Mesmay
École Normale Supérieure, Paris, France
e-mail: admesmay@gmail.com

Given two metric spaces (X, d_X) and (Y, d_Y), and a mapping $f : X \to Y$, we define the *Lipschitz constant of* f by

$$\|f\|_{\mathrm{Lip}} = \sup_{x \neq y \in X} \frac{d_Y(f(x), f(y))}{d_X(x, y)}.$$

An *L-Lipschitz* map is one for which $\|f\|_{\mathrm{Lip}} \leq L$. One defines the *distortion* of the mapping f to be $\mathsf{dist}(f) = \|f\|_{\mathrm{Lip}} \cdot \|f^{-1}\|_{\mathrm{Lip}}$, where the distortion is understood to be infinite when f is not injective. We say that (X, d_X) D-embeds into (Y, d_Y) if there is a mapping $f : X \to Y$ with $\mathsf{dist}(f) \leq D$.

Using the notation ℓ_1^k for the space $\mathbb{R}^k$ equipped with the $\|\cdot\|_1$ norm, we study the following question: How large must $k = k(n, \varepsilon)$ be so that every n-point tree metric $(1 + \varepsilon)$-embeds into ℓ_1^k?

1.1 Dimension Reduction in ℓ_1

A seminal result of Johnson and Lindenstrauss [8] implies that for every $\varepsilon > 0$, every n-point subset $X \subseteq \ell_2$ admits a $(1 + \varepsilon)$-distortion embedding into ℓ_2^k, with $k = O\left(\frac{\log n}{\varepsilon^2}\right)$. On the other hand, the known upper bounds for ℓ_1 are much weaker. Talagrand [19], following earlier results of Bourgain–Lindenstrauss–Milman [3] and Schechtman [17], showed that every n-dimensional subspace $X \subseteq \ell_1$ (and, in particular, every n-point subset) admits a $(1 + \varepsilon)$-embedding into ℓ_1^k, with $k = O\left(\frac{n \log n}{\varepsilon^2}\right)$. For n-point subsets, this was very recently improved to $k = O(n/\varepsilon^2)$ by Newman and Rabinovich [15], using the spectral sparsification techniques of Batson et al. [4].

On the other hand, Brinkman and Charikar [2] showed that there exist n-point subsets $X \subseteq \ell_1$ such that any D-embedding of X into ℓ_1^k requires $k \geq n^{\Omega(1/D^2)}$ (see also [10] for a simpler proof). Thus the exponential dimension reduction achievable in the ℓ_2 case cannot be matched for the ℓ_1 norm. More recently, it has been shown by Andoni et al. [1] that there exist n-point subsets such that any $(1 + \varepsilon)$-embedding requires dimension at least $n^{1-O(1/\log(\varepsilon^{-1}))}$. Regev [16] has given an elegant proof of both these lower bounds based on information theoretic arguments.

One can still ask about the possibility of more substantial dimension reduction for certain finite subsets of ℓ_1. Such a study was undertaken by Charikar and Sahai [5]. In particular, it is an elementary exercise to verify that every finite tree metric embeds isometrically into ℓ_1, thus the ℓ_1 dimension reduction question for trees becomes a prominent example of this type. It was shown[1] [5] that for every $\varepsilon > 0$, every n-point tree metric $(1+\varepsilon)$-embeds into ℓ_1^k with $k = O\left(\frac{\log^2 n}{\varepsilon^2}\right)$. It is quite natural to ask whether the dependence on n can be reduced to the natural volume lower bound of $\Omega(\log n)$. Indeed, it is Question 3.6 in the list "Open problems on embeddings of finite metric spaces" maintained by Matoušek [13], asked by Gupta et al.[2] As noted there, the ques-

[1] The original bound proved in [5] grew like $\log^3 n$, but this was improved using an observation of A. Gupta.

[2] Asked at the DIMACS Workshop on Discrete Metric spaces and their Algorithmic Applications (2003). The question was certainly known to others before 2003, and was asked to the first-named author by Assaf Naor earlier that year.

tion was, surprisingly, even open for the complete binary tree on n vertices. The present paper resolves this question, achieving the volume lower bound for all finite trees.

Theorem 1.1 *For every $\varepsilon > 0$ and $n \in \{1, 2, 3, \ldots\}$, the following holds. Every n-point tree metric admits a $(1+\varepsilon)$-embedding into ℓ_1^k with $k = O\big((\frac{1}{\varepsilon})^4 \log \frac{1}{\varepsilon} \log n\big)$. If the tree is a complete d-ary tree of some height, the bound improves to $k = O\big((\frac{1}{\varepsilon})^2 \log n\big)$.*

The proof for the general case is presented in Sect. 3.1. The special case of complete d-ary trees is addressed in Sect. 2. We remark that the proof also yields a randomized polynomial-time algorithm to construct the embedding.

By simple volume arguments, the $\Theta(\log n)$ factor is necessary. Regarding the dependence on ε, it is known [9] that for complete binary trees, one must have $k \geq \Omega\big(\frac{\log n}{\varepsilon^2 \log(1/\varepsilon)}\big)$, showing that, for this special case, Theorem 1.1 is tight up to a $\log(1/\varepsilon)$ factor.

1.2 Notation

For a graph $G = (V, E)$, we use the notations $V(G)$ and $E(G)$ to denote the vertex and edge sets of G, respectively. For a connected, rooted tree $T = (V, E)$ and $x, y \in V$, we use the notation P_{xy} for the unique path between x and y in T, and P_x for P_{rx}, where r is the root of T.

We use $\mathbb{N}$ for the set of positive integers $\{1, 2, 3, \ldots\}$. For $k \in \mathbb{N}$, we write $[k] = \{1, 2, \ldots, k\}$. We also use the asymptotic notation $A \lesssim B$ to denote that $A = O(B)$, and $A \asymp B$ to denote the conjunction of $A \lesssim B$ and $B \lesssim A$.

1.3 Proof Outline and Related Work

We first discuss the form that all our embeddings will take. Let $T = (V, E)$ be a finite, connected tree, and fix a root $r \in V$. For each $v \in V$, recall that P_v denotes the unique simple path from r to v. Given a labeling of edges by vectors $\lambda : E \to \mathbb{R}^k$, we can define $\varphi : V \to \mathbb{R}^k$ by

$$\varphi(x) = \sum_{e \in E(P_v)} \lambda(e). \tag{1}$$

The difficulty now lies in choosing an appropriate labeling λ. An easy observation is that if we have $\|\lambda(e)\|_1 = \mathsf{len}(e)$ for all $e \in E$ and the set $\{\lambda(e)\}_{e \in E}$ is orthogonal, then φ is an isometry. Of course, our goal is to use many fewer than $|E|$ dimensions for the embedding. We next illustrate a major probabilistic technique employed in our approach.

Re-randomization. Consider an unweighted, complete binary tree of height h. Denote the tree by $T_h = (V_h, E_h)$, let $n = 2^{h+1} - 1$ be the number of vertices, and let r denote the root of the tree. Let $\kappa \in \mathbb{N}$ be some constant which we will choose momentarily. If we assign to every edge $e \in E_h$, a label $\lambda(e) \in \mathbb{R}^\kappa$, then there is a natural mapping $\tau_\lambda : V_h \to \{0, 1\}^{\kappa h}$ given by

$$\tau_\lambda(v) = (\lambda(e_1), \lambda(e_2), \ldots, \lambda(e_k), 0, 0, \ldots, 0), \tag{2}$$

where $E(P_v) = \{e_1, e_2, \ldots, e_k\}$, and the edges are labeled in order from the root to v. Note that the preceding definition falls into the framework of (1), by extending each $\lambda(e)$ to a (κh)-dimensional vector padded with zeros, but the specification here will be easier to work with presently.

If we choose the label map $\lambda : E_h \to \{0,1\}^\kappa$ uniformly at random, the probability for the embedding τ_λ specified in (2) to have $O(1)$ distortion is at most exponentially small in n. In fact, the probability for τ_λ to be injective is already this small. This is because for two nodes $u, v \in V_h$ which are the children of the same node w, there is $\Omega(1)$ probability that $\tau_\lambda(u) = \tau_\lambda(v)$, and there are $\Omega(n)$ such independent events. In Sect. 2, we show that a judicious application of the Lovász Local Lemma [6] can be used to show that τ_λ has $O(1)$ distortion with non-zero probability. In fact, we show that this approach can handle arbitrary k-ary complete trees, with distortion $1+\varepsilon$. Unknown to us at the time of discovery, a closely related construction occurs in the context of tree codes for interactive communication [18].

Unfortunately, the use of the Local Lemma does not extend well to the more difficult setting of arbitrary trees. For the general case, we employ an idea of Schulman [18] based on *re-randomization*. To see the idea in our simple setting, consider T_h to be composed of a root r, under which lie two copies of T_{h-1}, which we call A and B, having roots r_{A} and r_{B}, respectively.

The idea is to assume that, inductively, we already have a labeling $\lambda_{h-1} : E_{h-1} \to \{0,1\}^{\kappa(h-1)}$ such that the corresponding map $\tau_{\lambda_{h-1}}$ has $O(1)$ distortion on T_{h-1}. We will then construct a random labeling $\lambda_h : E_h \to \{0,1\}^\kappa$ by using λ_{h-1} on the A-side, and $\pi(\lambda_{h-1})$ on the B-side, where π randomly alters the labeling in such a way that $\tau_{\pi(\lambda_{h-1})}$ is simply $\tau_{\lambda_{h-1}}$ composed with a random isometry of $\ell_1^{\kappa(h-1)}$. We will then argue that with positive probability (over the choice of π), τ_{λ_h} has $O(1)$ distortion,

Let $\pi_1, \pi_2, \ldots, \pi_{h-1} : \{0,1\}^\kappa \to \{0,1\}^\kappa$ be i.i.d. random mappings, where the distribution of π_1 is specified by

$$\pi_1(x_1, x_2, \ldots, x_\kappa) = (\rho_1(x_1), \rho_2(x_2), \ldots, \rho_\kappa(x_\kappa)),$$

where each ρ_i is an independent uniformly random involution $\{0,1\} \mapsto \{0,1\}$. To every edge $e \in E_{h-1}$, we can assign a height $\alpha(e) \in \{1, 2, \ldots, h-1\}$ which is its distance to the root. From a labeling $\lambda : E_{h-1} \to \{0,1\}^\kappa$, we define a random labeling $\pi(\lambda) : E_{h-1} \to \{0,1\}^\kappa$ by

$$\pi(\lambda)(e) = \pi_{\alpha(e)} \circ \lambda.$$

By a mild abuse of notation, we will consider $\pi(\lambda) : E(B) \to \{0,1\}^\kappa$.

Finally, given a labeling $\lambda_{h-1} : E_{h-1} \to \{0,1\}^\kappa$, we construct a random labeling $\lambda_h : E_h \to \{0,1\}^\kappa$ as follows:

$$\lambda_h(e) = \begin{cases} (0,0,\ldots,0) & e = (r, r_A), \\ (1,1,\ldots,1) & e = (r, r_B), \\ \lambda_{h-1}(e) & e \in E(A), \\ \pi(\lambda_{h-1})(e) & e \in E(B). \end{cases}$$

By construction, the mappings $\tau_{\lambda_h}|_{V(A)\cup\{r\}}$ and $\tau_{\lambda_h}|_{V(B)\cup\{r\}}$ have the same distortion as $\tau_{\lambda_{h-1}}$. In particular, it is easy to check that $\tau_{\pi(\lambda_{h-1})}$ is simply $\tau_{\lambda_{h-1}}$ composed with an isometry of $\{0,1\}^{\kappa(h-1)}$.

Now consider some pair $x \in V(A)$ and $y \in V(B)$. It is simple to argue that it suffices to bound the distortion for pairs with $m = d_{T_h}(r,x) = d_{T_h}(r,y)$ for $m \in \{1,2,\ldots,h\}$, so we will assume that x, y have the same height in T_h.

Observe that $\tau_{\lambda_h}(x)$ is fixed with respect to the randomness in π, thus if we write $v = \tau_{\lambda_h}(x) - \tau_{\lambda_h}(y)$, where subtraction is taken coordinate-wise, modulo 2, then v has the form

$$v \equiv \big(\underbrace{1,1,\ldots,1}_{\kappa}, b_1, b_2, \ldots, b_{\kappa(m-1)}\big)$$

where the $\{b_i\}$ are i.i.d. uniform over $\{0,1\}$. It is thus an easy consequence of Chernoff bounds that, with probability at least $1 - e^{-m\kappa/8}$, we have

$$\|\tau_{\lambda_h}(x) - \tau_{\lambda_h}(y)\|_1 = \|v\|_1 \geq \frac{\kappa \cdot d_{T_h}(x,y)}{4}.$$

Also, clearly $\|\tau_{\lambda_h}\|_{\mathrm{Lip}} \leq \kappa$.

On the other hand, the number of pairs $x \in V(A)$, $y \in V(B)$ with $m = d_{T_h}(r,x) = d_{T_h}(r,y)$ is $2^{2(m-1)}$, thus taking a union bound, we have

$$\mathbb{P}\big(\mathsf{dist}(\tau_{\lambda_h}) > \max\{4, \mathsf{dist}(\tau_{\lambda_{h-1}})\}\big) \leq \sum_{m=1}^{h} 2^{2(m-1)} e^{-m\kappa/8},$$

and the latter bound is strictly less than 1 for some $\kappa = O(1)$, showing the existence of a good map τ_{λ_h}.

This illustrates how re-randomization (applying a distribution over random isometries to one side of a tree) can be used to achieve $O(1)$ distortion for embedding T_h into $\ell_1^{O(h)}$. Unfortunately, the arguments become significantly more delicate when we handle less uniform trees. The full-blown re-randomization argument occurs in Sect. 5.

Scale Selection. The first step beyond complete binary trees would be in passing to complete d-ary trees for $d \geq 3$. The same construction as above works, but now one has to choose $\kappa \asymp \log d$. Unfortunately, if the degrees of our tree are not uniform, we have to adopt a significantly more delicate strategy. It is natural to choose a single number $\kappa(e) \in \mathbb{N}$ for every edge $e \in E$, and then put $\lambda(e) \in \frac{1}{\kappa(e)}\{0,1\}^{\kappa(e)}$ (this ensures that the analogue of the embedding τ_λ specified in (2) is 1-Lipschitz).

Observing the case of d-ary trees, one might be tempted to put

$$\kappa(e) = \left\lceil \log \frac{|T_u|}{|T_v|} \right\rceil,$$

where $e = (u,v)$ is directed away from the root, and we use T_v to denote the subtree rooted at v. If one simply takes a complete binary tree on 2^h nodes, and then connects a star of degree 2^h to every vertex, we have $\kappa(e) \asymp h$ for every edge, and thus the dimension becomes $O(h^2)$ instead of $O(h)$ as desired.

In fact, there are examples which show that it is impossible to choose $\kappa(u, v)$ to depend only on the geometry of the subtree rooted at u. These "scale selector" values have to look at the global geometry, and in particular have to encode the volume growth of the tree at many scales simultaneously. Our eventual scale selector is fairly sophisticated and impossible to describe without delving significantly into the details of the proof. For our purposes, we need to consider more general embeddings of type (1). In particular, the coordinates of our labels $\lambda(e) \in \mathbb{R}^k$ will take a range of different values, not simply a single value as for complete trees.

We do try to maintain one important, related invariant: If P_v is the sequence of edges from the root to some vertex v, then ideally for every coordinate $i \in \{1, 2, \dots, k\}$ and every value $j \in \mathbb{Z}$, there will be at most one $e \in P_v$ for which $\lambda(e)_i \in [2^j, 2^{j+1})$. Thus instead of every coordinate being "touched" at most once on the path from the root to v, every coordinate is touched at most once *at every scale* along every such path. This ensures that various scales do not interact. For technical reasons, this property is not maintained exactly, but analogous concepts arise frequently in the proof.

The restricted class of embeddings we use, along with a discussion of the invariants we maintain, are introduced in Sect. 3.2. The actual scale selectors are defined in Sect. 4.

Controlling the Topology. One of the properties that we used above for complete d-ary trees is that the depth of such a tree is $O(\log_d n)$, where n is the number of nodes in the tree. This allowed us to concatenate vectors down a root–leaf path without exceeding our desired $O(\log n)$ dimension bound. Of course, for general trees, no similar property need hold. However, there is still a bound on the *topological* depth of any n-node tree.

To explain this, let $T = (V, E)$ be a tree with root r, and define a *monotone coloring of* T to be a mapping $\chi : E \to \mathbb{N}$ such that for every $c \in \mathbb{N}$, the color class $\chi^{-1}(c)$ is a connected subset of some root–leaf path. Such colorings were used in previous works on embedding trees into Hilbert spaces [7,11,12], as well as for preivous low-dimensional embeddings into ℓ_1 [5]. The following lemma is well-known and elementary.

Lemma 1.2 *Every connected n-vertex rooted tree T admits a monotone coloring such that every root–leaf path in T contains at most $1 + \log_2 n$ colors.*

Proof For an edge $e \in E(T)$, let $\ell(e)$ denote the number of leaves beneath e in T (including, possibly, an endpoint of e). Letting $\ell(T) = \max_{e \in E} \ell(e)$, we will prove that for $\ell(T) \geq 1$, there exists a monotone coloring with at most $1+\log_2(\ell(T)) \leq 1+\log_2 n$ colors on any root–leaf path.

Suppose that r is the root of T. For an edge e, let T_e be the subtree beneath e, including the edge e itself. If r is the endpoint of edges $e_1, e_2, \dots, e_k$, we may color the edges of $T_{e_1}, T_{e_2}, \dots, T_{e_k}$ separately, since any monotone path is contained completely within exactly one of these subtrees. Thus we may assume that r is the endpoint of only one edge e_1, and then $\ell(T) = \ell(e_1)$.

Choose a leaf x in T such that each connected component of T' of $T \setminus E(P_{rx})$ has $\ell(T') \leq \ell(e_1)/2$ (this is easy to do by, e.g., ordering the leaves from left to right in a planar drawing of T). Color the edges $E(P_{rx})$ with color 1, and inductively color each non-trivial connected component T' with disjoint sets of colors from $\mathbb{N} \setminus \{1\}$. By induction, the maximum number of colors appearing on a root–leaf path in T is at most $1 + \log_2(\ell(e_1)/2) = 1 + \log_2(\ell(T))$, completing the proof. □

Instead of dealing directly with edges in our actual embedding, we will deal with color classes. This poses a number of difficulties, and one major difficulty involving vertices which occur in the middle of such classes. For dealing with these vertices, we will first preprocess our tree by embedding it into a product of a small number of new trees, each of which admits colorings of a special type. This is carried out in Sect. 3.1.

2 Warm-Up: Embedding Complete *k*-ary Trees

We first prove our main result for the special case of complete k-ary trees, with an improved dependence on ε. The main novelty is our use of the Lovász Local Lemma to analyze a simple random embedding of such trees into ℓ_1. The proof illustrates the tradeoff between concentration and the sizes of the sets $\{\{u, v\} \subseteq V : d_T(u, v) = j\}$ for each $j = 1, 2, \dots$

Theorem 2.1 *Let $T_{k,h}$ be the unweighted, complete k-ary tree of height h. For every $\varepsilon > 0$, there exists a $(1+\varepsilon)$-embedding of $T_{k,h}$ into $\ell_1^{O((h \log k)/\varepsilon^2)}$.*

In the next section, we introduce our random embedding and analyze the success probability for a single pair of vertices based on their distance. Then in Sect. 2.2, we show that with non-zero probability, the construction succeeds for all vertices. In the coming sections and later, in the proof of our main theorem, we will employ the following concentration inequality [14].

Theorem 2.2 *Let M be a non-negative number, and X_i $(1 \le i \le n)$ be independent random variables satisfying $X_i \le \mathbb{E}(X_i) + M$ for $1 \le i \le n$. Consider the sum $X = \sum_{i=1}^n X_i$ with expectation $\mathbb{E}(X) = \sum_{i=1}^n \mathbb{E}(X_i)$ and $\mathrm{Var}(X) = \sum_{i=1}^n \mathrm{Var}(X_i)$. Then we have*

$$\mathbb{P}(X - \mathbb{E}(X) \ge \lambda) \le \exp\left(\frac{-\lambda^2}{2(\mathrm{Var}(X) + M\lambda/3)}\right). \tag{3}$$

2.1 A Single Event

First $k, h \in \mathbb{N}$ and $\varepsilon > 0$. Write $T = (V, E)$ for the tree $T_{k,h}$ with root $r \in V$, and let d_T be the unweighted shortest-path metric on T. Additionally, we define

$$t = \left\lceil \frac{1}{\varepsilon} \right\rceil \tag{4}$$

and

$$m = t \lceil \log k \rceil. \tag{5}$$

Let $\{\vec{v}(1), \dots, \vec{v}(t)\}$ be the standard basis for $\mathbb{R}^t$. Let $b_1, b_2, \dots, b_m$ be chosen i.i.d. uniformly over $\{1, 2, \dots, t\}$. For the edges $e \in E$, we choose i.i.d. random labels $\lambda(e) \in \mathbb{R}^{m\times t}$, each of which has the distribution of the random vector (represented in matrix notation),

$$\frac{1}{m}\begin{pmatrix} \vec{v}(b_1) \\ \vdots \\ \vec{v}(b_m) \end{pmatrix}. \tag{6}$$

Note that for every $e \in E$, we have $\|\lambda(e)\|_1 = 1$. We now define a random mapping $g : V \to \mathbb{R}^{m(h-1)\times t}$ as follows: We put $g(r) = 0$, and otherwise

$$g(v) = \begin{pmatrix} \lambda(e_1) \\ \vdots \\ \lambda(e_j) \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{7}$$

where $e_1, e_2, \dots, e_j$ is the sequence of edges encountered on the path from the root to v. It is straightforward to check that g is 1-Lipschitz. The next observation is also immediate from the definition of g.

Observation 2.3 For any $v \in V$ and $u \in V(P_v)$, we have $d_T(u, v) = \|g(u)-g(v)\|_1$.

For $m, n \in \mathbb{N}$, and $A \in \mathbb{R}^{m\times n}$, we use the notation $A[i] \in \mathbb{R}^n$ to refer to the ith row of A. We now bound the probability that a given pair of vertices experiences a large contraction.

Lemma 2.4 *For $C \geq 10$, and $x, y \in V$,*

$$\mathbb{P}[\|g(x) - g(y)\|_1 \leq (1 - C\varepsilon)d_T(x, y)] \leq k^{-Cd_T(x,y)/2}. \tag{8}$$

Proof Fix $x, y \in V$, and let r' denote their lowest common ancestor. We define a family of random variables $\{X_{ij}\}_{i\in[h-1], j\in[m]}$ by setting $\ell_{ij} = (i-1)m + j$, and then

$$\begin{aligned} X_{ij} = {} & \|g(x)[\ell_{ij}] - g(r')[\ell_{ij}]\|_1 + \|g(y)[\ell_{ij}] - g(r')[\ell_{ij}]\|_1 \\ & -\|g(x)[\ell_{ij}] - g(y)[\ell_{ij}]\|_1. \end{aligned} \tag{9}$$

Observe that if $i \leq d_T(r, r')$ then $X_{ij} = 0$ for all $j \in [m]$ since all three terms in (9) are zero. Furthermore, if $i \geq \min(d_T(r, x), d_T(r, y)) + 1$, then again $X_{ij} = 0$ for all $j \in [m]$, since in this case one of the first two terms of (9) is zero, and the other is equal to the last. Thus if

$$R = [h-1] \cap [d_T(r, r') + 1, \min(d_T(r, x), d_T(r, y))],$$

then $i \notin R \implies X_{ij} = 0$ for all $j \in [m]$, and additionally we have the estimate

$$|R| = \min(d_T(r,x), d_T(r,y)) - d_T(r,r') \leq \frac{d_T(x,y)}{2}. \tag{10}$$

Now, using the definition of g in (7), we can write

$$\begin{aligned}
&\|g(x)-g(y)\|_1 \\
&\quad = \sum_{i\in[h-1],j\in[m]} \big(\|g(x)[\ell_{ij}] - g(r')[\ell_{ij}]\|_1 + \|g(y)[\ell_{ij}] - g(r')[\ell_{ij}]\|_1 - X_{ij}\big) \\
&\quad = \|g(x)-g(r')\|_1 + \|g(y)-g(r')\|_1 - \sum_{i\in[h-1],j\in[m]} X_{ij} \\
&\quad \overset{(2.3)}{=} d_T(x,r') + d_T(y,r') - \sum_{i\in[h-1],j\in[m]} X_{ij} \\
&\quad = d_T(x,y) - \sum_{i\in[h-1],j\in[m]} X_{ij}.
\end{aligned}$$

We will prove the lemma by arguing that

$$\mathbb{P}\Big[\sum_{i\in[h-1],j\in[m]} X_{ij} \leq C\varepsilon d_T(x,y)\Big] \leq k^{-Cd_T(x,y)/2}.$$

We start the proof by first bounding the maximum of the X_{ij} variables. Since, for every ℓ, we have

$$\|g(x)[\ell]-g(r')[\ell]\|_1,\ \|g(y)[\ell]-g(r')[\ell]\|_1 \in \Big\{0, \frac{1}{m}\Big\},$$

we conclude that

$$\max\big\{X_{ij} : i \in [h-1], j \in [m]\big\} \leq \frac{2}{m}. \tag{11}$$

For $i \in R$ and $j \in [m]$, using (6) and (7), we see that $(g(x)[\ell_{ij}] - g(r')[\ell_{ij}]) = \frac{1}{m}\vec{v}(\alpha)$ and $g(y)[\ell_{ij}] - g(r')[\ell_{ij}] = \frac{1}{m}\vec{v}(\beta)$, where α and β are i.i.d. uniform over $\{1,\dots,t\}$. Hence, for $i \in R$ and $j \in [m]$, we have

$$\mathbb{P}[X_{ij} \neq 0] = \frac{1}{t}.$$

We can thus bound the expected value and variance of X_{ij} for $i \in R$ and $j \in [m]$ using (11),

$$\mathbb{E}[X_{ij}] \leq \frac{2}{tm} \tag{12}$$

and

$$\mathrm{Var}(X_{ij}) \leq \frac{4}{tm^2}. \tag{13}$$

Using (10), we have

$$\sum_{i=1}^{h-1}\sum_{j=1}^{m}\mathbb{E}[X_{ij}] = \sum_{i\in R}\sum_{j\in[m]}\mathbb{E}[X_{ij}] \overset{(12)}{\leq} \sum_{i\in R}\frac{2}{t} \overset{(10)}{\leq} \frac{d_T(x,y)}{t} \tag{14}$$

and

$$\sum_{i=1}^{h-1}\sum_{j=1}^{m}\mathrm{Var}(X_{ij}) = \textstyle\sum_{i\in R}\sum_{j\in[m]}\mathrm{Var}(X_{ij}) \overset{(13)}{\leq} \sum_{i\in R}\frac{4}{tm} \overset{(10)}{\leq} \frac{2\,d_T(x,y)}{tm}. \tag{15}$$

We now apply Theorem 2.2 to complete the proof:

$$\begin{aligned}
&\mathbb{P}\Big[\sum_{i\in[h-1],j\in[m]} X_{ij} \geq C\Big(\frac{d_T(x,y)}{t}\Big)\Big] \\
&\quad= \mathbb{P}\Big[\sum_{i\in[h-1],j\in[m]} X_{ij} - \frac{d_T(x,y)}{t} \geq (C-1)\Big(\frac{d_T(x,y)}{t}\Big)\Big] \\
&\quad\overset{(14)}{\leq} \mathbb{P}\Big(\sum_{i\in[h-1],j\in[m]} X_{ij} - \mathbb{E}\Big[\sum_{i\in[h-1],j\in[m]} X_{ij}\Big] \geq (C-1)\Big(\frac{d_T(x,y)}{t}\Big)\Big) \\
&\quad\leq \exp\Big(\frac{-((C-1)d_T(x,y)/t)^2}{2\big(\sum_{i\in[h-1],j\in[m]}\mathrm{Var}(X_{ij}) + (C-1)(d_T(x,y)/t)(\frac{2}{m})/3\big)}\Big) \\
&\quad\overset{(15)}{\leq} \exp\Big(\frac{-((C-1)d_T(x,y)/t)^2}{2\big(2\,d_T(x,y)/(tm) + (C-1)(d_T(x,y)/t)(\frac{2}{m})/3\big)}\Big) \\
&\quad= \exp\Big(\frac{-(C-1)^2}{4\big(1+(C-1)/3\big)}\cdot\frac{m}{t}\cdot d_T(x,y)\Big).
\end{aligned}$$

An elementary calculation shows that for $C \geq 10$, we have $\frac{(C-1)^2}{4(1+(C-1)/3)} \geq \frac{C}{2}$. Hence,

$$\begin{aligned}
&\mathbb{P}\Big[\sum_{i\in[h-1],j\in[m]} X_{ij} \geq C\varepsilon d_T(x,y)\Big] \\
&\quad\overset{(14)}{\leq} \mathbb{P}\Big[\sum_{i\in[h-1],j\in[m]} X_{ij} \geq C\Big(\frac{d_T(x,y)}{t}\Big)\Big] \leq \exp\Big(-\frac{Cm}{2t}d_T(x,y)\Big) \\
&\quad\overset{(5)}{\leq} k^{-Cd_T(x,y)/2}
\end{aligned}$$

completing the proof. □

2.2 The Local Lemma Argument

We first give the statement of the Lovász Local Lemma [6] and then use it in conjunction with Lemma 2.4 to complete the proof of Theorem 2.1.

Theorem 2.5 *Let $\mathcal{A}$ be a finite set of events in some probability space. For $A \in \mathcal{A}$, let $\Gamma(A) \subseteq \mathcal{A}$ be such that A is independent from the collection of events $\mathcal{A}\setminus(\{A\}\cup\Gamma(A))$. If there exists an assignment $x : \mathcal{A} \to (0, 1)$ such that for all $A \in \mathcal{A}$, we have*

$$\mathbb{P}(A) \le x(A) \prod_{B\in\Gamma(A)} (1 - x(B)),$$

then the probability that none of the events in $\mathcal{A}$ occur is at least $\prod_{A\in\mathcal{A}}(1-x(A)) > 0$.

Proof of Theorem 2.1 We may assume that $k \ge 2$. We will use Theorem 2.5 and Lemma 2.4 to show that with non-zero probability the following inequality holds for all $u, v \in V$

$$\|g(u) - g(v)\|_1 \le (1 - 14\varepsilon)\, d_T(u, v).$$

For $u, v \in V$, let $\mathcal{E}_{uv}$ be the event $\big\{\|g(u) - g(v)\|_1 \le (1 - 14\varepsilon)\, d_T(u, v)\big\}$. Now, for $u, v \in V$, define

$$x_{uv} = k^{-3d_T(u,v)}.$$

Observe that for vertices $u, v \in V$ and a subset $V' \subseteq V$, the event $\mathcal{E}_{uv}$ is mutually independent of the family $\{\mathcal{E}_{u'v'} : u', v' \in V'\}$ whenever the induced subgraph of T spanned by V' contains no edges from P_{uv}. Thus using Theorem 2.5, it is sufficient to show that for all $u, v \in V$,

$$\mathbb{P}(\mathcal{E}_{uv}) \le x_{uv} \prod_{\substack{s,t\in V:\\ E(P_{st})\cap E(P_{uv})\neq\emptyset}} (1 - x_{st}). \tag{16}$$

Indeed, this will complete the proof of Theorem 2.1.

To this end, fix $u, v \in V$. For $e \in E$ and $i \in \mathbb{N}$, we define the set

$$S_{e,i} = \big\{(s, t) : s, t \in V, d_T(s, t) = i, \text{ and } e \in E(P_{st})\big\}.$$

Since T is a k-ary tree,

$$|S_{e,i}| \le \sum_{j=1}^{i} k^{j-1} \cdot k^{i-j} = i \cdot k^{i-1} \le k^{2i}. \tag{17}$$

Thus we can write

$$
\begin{aligned}
& x_{uv} \prod_{\substack{s,t \in V: \\ E(P_{st}) \cap E(P_{uv}) \neq \emptyset}} (1 - x_{st}) \\
& = x_{uv} \prod_{e \in E(P_{uv})} \prod_{i \in \mathbb{N}} \prod_{(s,t) \in S_{e,i}} \big(1 - x_{st}\big) = k^{-3d_T(u,v)} \prod_{e \in E(P_{uv})} \prod_{i \in \mathbb{N}} \prod_{(s,t) \in S_{e,i}} \big(1 - k^{-3i}\big) \\
& \overset{(17)}{\geq} k^{-3d_T(u,v)} \prod_{e \in E(P_{uv})} \prod_{i \in \mathbb{N}} \big(1 - k^{-3i}\big)^{k^{2i}} \geq k^{-3d_T(u,v)} \prod_{e \in E(P_{uv})} \prod_{i \in \mathbb{N}} \big(1 - k^{2i}(k^{-3i})\big) \\
& = k^{-3d_T(u,v)} \prod_{e \in E(P_{uv})} \prod_{i \in \mathbb{N}} \big(1 - \frac{1}{k^i}\big).
\end{aligned}
$$

For $x \in [0, \frac{1}{2}]$, we have $e^{-2x} \leq 1 - x$, and since $k \geq 2$, we have $k^{-i} \leq \frac{1}{2}$ for all $i \in \mathbb{N}$, hence

$$
\begin{aligned}
& x_{uv} \prod_{\substack{s,t \in V: \\ E(P_{st}) \cap E(P_{uv}) \neq \emptyset}} (1 - x_{st}) \\
& \geq k^{-3d_T(u,v)} \prod_{e \in E(P_{uv})} \prod_{i \in \mathbb{N}} \exp\big(\frac{-2}{k^i}\big) = k^{-3d_T(u,v)} \prod_{e \in E(P_{uv})} \exp\big(-2 \sum_{i \in \mathbb{N}} \frac{1}{k^i}\big) \\
& = k^{-3d_T(u,v)} \prod_{e \in E(P_{uv})} \exp\big(\frac{-2/k}{1 - 1/k}\big) \geq k^{-3d_T(u,v)} \prod_{e \in E(P_{uv})} \exp\big(\frac{-4}{k}\big) \\
& = k^{-3d_T(u,v)} \exp\big(\frac{-4\, d_T(u,v)}{k}\big).
\end{aligned}
$$

Since $k \geq 2$, we conclude that

$$
x_{uv} \prod_{\substack{s,t \in V: \\ E(P_{st}) \cap E(P_{uv}) \neq \emptyset}} (1 - x_{st}) \geq k^{-7d_T(u,v)}.
$$

On the other hand, Lemma 2.4 applied with $C = 14$ gives

$$
\mathbb{P}[\|g(u) - g(v)\|_1 \leq (1 - 14\varepsilon) d_T(u,v)] \leq k^{-7d_T(u,v)},
$$

yielding (16), and completing the proof. □

3 Colors and Scales

In the present section, we develop some tools for our eventual embedding. The proof of our main theorem appears in the next section, but relies on a key theorem which is only proved in Sect. 5.

3.1 Monotone Colorings

Let $T = (V, E)$ be a metric tree rooted at a vertex $r \in V$. Recall that such a tree T is equipped with a length $\mathsf{len}\colon E \to [0, \infty)$. We extend this to subsets of edges $S \subseteq E$ via $\mathsf{len}(S) = \sum_{e \in S} \mathsf{len}(e)$. We recall that a *monotone coloring* is a mapping $\chi : E \to \mathbb{N}$ such that each color class $\chi^{-1}(c) = \{e \in E : \chi(e) = c\}$ is a connected subset of some root–leaf path. For a set of edges $S \subseteq E$, we write $\chi(S)$ for the set of colors occurring in S. We define the *multiplicity of* χ by

$$M(\chi) = \max_{v \in V} |\chi(P_v)|.$$

Given such a coloring χ and $c \in \mathbb{N}$, we define

$$\mathsf{len}_\chi(c) = \mathsf{len}(\chi^{-1}(c)),$$

and $\mathsf{len}_\chi(S) = \sum_{c \in S} \mathsf{len}_\chi(c)$ if $S \subseteq \mathbb{N}$.

For every $\delta \in [0, 1]$ and $x, y \in V$, we define the set of colors

$$C_\chi(x, y; \delta) = \left\{c : \mathsf{len}(P_{xy} \cap \chi^{-1}(c)) \le \delta \cdot \mathsf{len}_\chi(c)\right\} \cap (\chi(P_x) \triangle \chi(P_y)).$$

This is the set of colors c which occur in only one of P_x and P_y, and for which the contribution to P_{xy} is significantly smaller than $\mathsf{len}_\chi(c)$. We also put

$$\rho_\chi(x, y; \delta) = \mathsf{len}_\chi(C(x, y; \delta)). \tag{18}$$

We now state a key theorem that will be proved in Sect. 5.

Theorem 3.1 *For every $\varepsilon, \delta > 0$, there is a value $C(\varepsilon, \delta) = O\big((\frac{1}{\varepsilon} + \log\log \frac{1}{\delta})^3 \log \frac{1}{\varepsilon}\big)$ such that the following holds. For any metric tree $T = (V, E)$ and any monotone coloring $\chi : E \to \mathbb{N}$, there exists a mapping $F : V \to \ell_1^{C(\varepsilon,\delta)(\log n + M(\chi))}$ such that for all $x, y \in V$,*

$$(1 - \varepsilon)\, d_T(x, y) - \delta\, \rho_\chi(x, y; \delta) \le \|F(x) - F(y)\|_1 \le d_T(x, y). \tag{19}$$

The problem one now confronts is whether the loss in the $\rho_\chi(x, y; \delta)$ term can be tolerated. In general, we do not have a way to do this, so we first embed our tree into a product of a small number of trees in a way that allows us to control the corresponding ρ-terms.

Lemma 3.2 *For every $\varepsilon \in (0, 1)$, there is a number $k \asymp \frac{1}{\varepsilon}$ such that the following holds. For every metric tree $T = (V, E)$ and monotone coloring $\chi : E \to \mathbb{N}$, there exist k metric trees $T_1, T_2, \ldots, T_k$ with monotone colorings $\{\chi_i : E(T_i) \to \mathbb{N}\}_{i=1}^k$ and mappings $\{f_i : V \to V(T_i)\}_{i=1}^k$ such that $M(\chi_i) \le M(\chi)$, and $|V(T_i)| \le |V|$ for all $i \in [k]$, and the following conditions hold for all $x, y \in V$:*

(a) *We have*

$$\frac{1}{k}\sum_{i=1}^{k} d_{T_i}(f_i(x), f_i(y)) \geq (1-\varepsilon)\, d_T(x,y). \tag{20}$$

(b) *For all $i \in [k]$, we have*

$$d_{T_i}(f_i(x), f_i(y)) \leq (1+\varepsilon)\, d_T(x,y). \tag{21}$$

(c) *There exists a number $j \in [k]$ such that*

$$\varepsilon\, d_T(x,y) \geq \frac{2^{-(k+1)}}{k}\sum_{\substack{i=1\\ i\neq j}}^{k} \rho_{\chi_i}(f_i(x), f_i(y); 2^{-(k+1)}). \tag{22}$$

Using Lemma 3.2 in conjunction with Theorem 3.1, we can now prove the main theorem (Theorem 1.1).

Proof of Theorem 1.1 Let $\varepsilon > 0$ be given, let $T = (V, E)$ be an n-vertex metric tree. Let $\chi : E \to \mathbb{N}$ be a monotone coloring with $M(\chi) \leq O(\log n)$, which exists by Lemma 1.2. Apply Lemma 3.2 to obtain metric trees $T_1, \ldots, T_k$ with corresponding monotone colorings $\chi_1, \ldots, \chi_k$ and mappings $f_i : V \to V(T_i)$. Observe that $M(\chi_i) \leq O(\log n)$ for each $i \in [k]$.

Let $F_i : V(T_i) \to \ell_1^{C(\varepsilon)\log n}$ be the mapping obtained by applying Theorem 3.1 to T_i and χ_i, for each $i \in [k]$, with $\delta = 2^{-(k+1)}$, where $C(\varepsilon) = O\big(\frac{1}{\varepsilon^3}(\log \frac{1}{\varepsilon})\big)$. Finally, we put

$$F = \frac{1}{k}\big((F_1 \circ f_1) \oplus (F_2 \circ f_2) \oplus \cdots \oplus (F_k \circ f_k)\big)$$

so that $F : V \to \ell^{O((\frac{1}{\varepsilon})^4 \log \frac{1}{\varepsilon} \cdot \log n)}$. We will prove that F is a $(1+O(\varepsilon))$-embedding, completing the proof.

First, observe that each F_i is 1-Lipschitz (Theorem 3.1). In conjunction with condition (b) of Lemma 3.2 which says that $\|f_i\|_{\mathrm{Lip}} \leq 1+\varepsilon$ for each $i \in [k]$, we have $\|F\|_{\mathrm{Lip}} \leq 1+\varepsilon$.

For the other side, fix $x, y \in V$ and let $j \in [k]$ be the number guaranteed in condition (c) of Lemma 3.2. Then we have

$$\begin{aligned}
&\|F(x) - F(y)\|_1\\
&\quad = \frac{1}{k}\sum_{i=1}^{k} \|(F_i \circ f_i)(x) - (F_i \circ f_i)(y)\|_1\\
&\quad \overset{(19)}{\geq} \frac{1}{k}\sum_{i\neq j}\Big((1-\varepsilon)\, d_{T_i}(f_i(x), f_i(y)) - 2^{-(k+1)}\rho_{\chi_i}(f_i(x), f_i(y); 2^{-(k+1)})\Big)
\end{aligned}$$

$$
\begin{aligned}
&\overset{(22)}{\geq} \Big(\frac{1}{k}\sum_{i\neq j}(1-\varepsilon)\, d_{T_i}(f_i(x), f_i(y))\Big) - \varepsilon\, d_T(x,y) \\
&\geq \Big(\frac{1}{k}\sum_{i=1}^{k}(1-\varepsilon)\, d_{T_i}(f_i(x), f_i(y))\Big) - \frac{1}{k}\, d_{T_j}(f_j(x), f_j(y)) - \varepsilon\, d_T(x,y) \\
&\overset{(21)}{\geq} \Big(\frac{1}{k}\sum_{i=1}^{k}(1-\varepsilon)\, d_{T_i}(f_i(x), f_i(y))\Big) - \frac{1+\varepsilon}{k}\, d_T(x,y) - \varepsilon\, d_T(x,y) \\
&\overset{(20)}{\geq} (1-\varepsilon)^2\, d_T(x,y) - \frac{1+\varepsilon}{k}\, d_T(x,y) - \varepsilon\, d_T(x,y) \\
&\geq (1-O(\varepsilon))\, d_T(x,y),
\end{aligned}
$$

where in the final line we have used $k \asymp \frac{1}{\varepsilon}$, completing the proof. □

We now move on to the proof of Lemma 3.2. We begin by proving an analogous statement for the half line $[0,\infty)$. An $\mathbb{R}$-*star* is a metric space formed as follows: Given a sequence $\{a_i\}_{i=1}^{\infty}$ of positive numbers, one takes the disjoint union of the intervals $\{[0,a_1],[0,a_2],\ldots\}$, and then identifies the 0 point in each, which is canonically called the *root of the* $\mathbb{R}$-*star*. An $\mathbb{R}$-star S carries the natural induced length metric d_S. We refer to the associated intervals as *branches*, and the *length of a branch* is the associated number a_i. Finally, if S is an $\mathbb{R}$-star, and $x \in S\setminus\{0\}$, we use $\ell(x)$ to denote the length of the branch containing x. We put $\ell(0)=0$.

Lemma 3.3 *For every $k\in\mathbb{N}$ with $k\geq 2$, there exist $\mathbb{R}$-stars $S_1,\ldots,S_k$ with mappings*

$$f_i : [0,\infty) \to S_i$$

such that the following conditions hold:

(i) *For each $i\in[k]$, $f_i(0)$ is the root of S_i.*
(ii) *For all $x,y\in[0,\infty)$, $\frac{1}{k}\sum_{i=1}^{k} d_{S_i}(f_i(x),f_i(y)) \geq (1-\frac{7}{k})|x-y|$.*
(iii) *For each $i\in[k]$, f_i is $(1+2^{-k+1})$-Lipschitz.*
(iv) *For $x\in[0,\infty)$, we have $\ell(f_i(x))\leq 2^{k-1}x$.*
(v) *For $x\in[0,\infty)$, there are at most two values of $i\in[k]$ such that*

$$d_{S_i}(f_i(0), f_i(x)) \leq 2^{-k}\,\ell(f_i(x)).$$

(vi) *For all $x,y\in[0,\infty)$, there is at most one value of $i\in[k]$ such that $f_i(x)$ and $f_i(y)$ are in different branches of S_i and*

$$2^{-k}\big(\ell(f_i(x))+\ell(f_i(y))\big) > 2\,|x-y|.$$

Proof Assume that $k\geq 2$. We first construct $\mathbb{R}$-stars $S_1,\ldots,S_k$. We will index the branches of each star by $\mathbb{Z}$. For $i\in[k]$, S_i is a star whose jth branch, for $j\in\mathbb{Z}$, has length $2^{i-1+k(j+1)}$. We will use the notation (i,j,d) to denote the point at distance d from the root on the jth branch of S_i. Observe that $(i,j,0)$ and $(i,j',0)$ describe the same point (the root of S_i) for all $j,j'\in\mathbb{N}$.

Now, we define for every $i \in [k]$, a function $f_i : [0,\infty) \to S_i$ as follows:

$$f_i(x) = \begin{cases} \big(i, j, (x - 2^{i+kj})/(1-2^{1-k})\big) & \text{for } 2^{-i}x \in [2^{kj}, 2^{k(j+1)-1}), \\ \big(i, j, 2^{i+k(j+1)} - x\big) & \text{for } 2^{-i}x \in [2^{k(j+1)-1}, 2^{k(j+1)}). \end{cases}$$

Condition (i) is immediate. It is also straightforward to verify that

$$\|f_i\|_{\mathrm{Lip}} \le (1-2^{1-k})^{-1} \le 1 + 2^{-k+1}, \tag{23}$$

yielding condition (iii).

Toward verifying condition (ii), observe that for every $x \in [0,\infty)$ and $l \in \{0, 1, \dots, k-2\}$ we have

$$d_{S_i}(f_i(x), f_i(0)) \ge (x - 2^{\lfloor \log_2 x \rfloor - l})/(1-2^{1-k}) \ge x - 2^{\lfloor \log_2 x \rfloor - l},$$

when $i = (\lfloor \log_2 x \rfloor - l) \bmod k$. Using this, we can write

$$\begin{aligned} \sum_{i=1}^{k} d_{S_i}(f_i(x), f_i(0)) &\ge \sum_{l=\lfloor \log_2 x \rfloor - k + 2}^{\lfloor \log_2 x \rfloor} x - 2^l = (k-1)x - \sum_{l=\lfloor \log_2 x \rfloor - k + 2}^{\lfloor \log_2 x \rfloor} 2^l \\ &\ge (k-1)x - 2^{\lfloor \log_2 x \rfloor + 1} \ge (k-3)x. \end{aligned} \tag{24}$$

Now fix $x, y \in [0,\infty)$ with $x \le y$. If $x \le y/2$, then we can use the triangle inequality, together with (23) and (24) to write

$$\begin{aligned} &\frac{1}{k}\sum_{i=1}^{k} d_{S_i}(f_i(x), f_i(y)) \\ &\quad \ge \frac{1}{k}\sum_{i=1}^{k} \big(d_{S_i}(f_i(y), f_i(0)) - d_{S_i}(f_i(x), f_i(0))\big) \ge (1-3/k)y - (1+2^{1-k})x \\ &\quad \ge (1-3/k)y - (1+1/k)x \ge (1-7/k)(y-x) + 4y/k - 8x/k \\ &\quad \ge (1-7/k)(y-x). \end{aligned}$$

In the case that $\frac{y}{2} \le x \le y$, for $l \in \{0, 1, \dots, k-3\}$, we have

$$d_{S_i}(f_i(x), f_i(y)) \ge (y-x)/(1-2^{1-k}) \ge y - x$$

when $i = (\lfloor \log_2 x \rfloor - l) \bmod k$. From this, we conclude that

$$\frac{1}{k}\sum_{i=1}^{k} d_{S_i}(f_i(x), f_i(y)) \ge \frac{1}{k}\sum_{l=0}^{k-3}(y-x) \ge \frac{k-2}{k}(y-x), \tag{25}$$

yielding condition (ii).

It is also straightforward to check that

$$\ell(f_i(x)) \le 2^{\lfloor \log_2 x \rfloor + k - 1} \le 2^{k-1} x,$$

which verifies condition (iv).

To verify condition (v), note that for $x \in [0, \infty)$, the inequality $d_{S_i}(f_i(x), f_i(0)) \le x/2$ can only hold for $i \bmod k \in \{\lfloor \log_2 x \rfloor, \lfloor \log_2 x \rfloor + 1\}$, hence condition (iv) implies condition (v).

Finally we verify condition (vi). We divide the problem into two cases. If $x < y/2$, then by condition (iv),

$$\ell(f_i(x)) + \ell(f_i(y)) \le 2^{k-1}(x+y) \le 2^{k-1}(2y) \le 2^{k+1}(y-x).$$

In the case that $y/2 < x \le y$, $f_i(x)$ and $f_i(y)$ can be mapped to different branches of S_i only for $i \equiv \lfloor \log_2 y \rfloor \pmod k$, yielding condition (vi). □

Finally, we move onto the proof of Lemma 3.2.

Proof of Lemma 3.2 We put $k = \lceil 7/\varepsilon \rceil$ and prove the following stronger statement by induction on $|V|$: There exist metric trees $T_1, T_2, \dots, T_k$ and monotone colorings $\chi_i : E(T_i) \to \mathbb{N}$, along with mappings $f_i : V \to V(T_i)$ satisfying the conditions of the lemma. Furthermore, each coloring χ_i satisfies the stronger condition for all $v \in V$,

$$|\chi_i(P_{f_i(v)})| \le |\chi(P_v)|. \tag{26}$$

The statement is trivial for the tree containing only a single vertex. Now suppose that we have a tree T and coloring $\chi : E \to \mathbb{N}$. Since T is connected, it is easy to see that there exists a color class $c \in \chi(E)$ with the following property. Let γ_c be the path whose edges are colored c, and let v_c be the vertex of γ_c closest to the root. Then the induced tree T' on the vertex set $(V \setminus V(\gamma_c)) \cup \{v_c\}$ is connected.

Applying the inductive hypothesis to T' and $\chi|_{E(T')}$ yields metric trees $T'_1, T'_2, \dots, T'_k$ with colorings $\chi'_i : E(T'_i) \to \mathbb{N}$ and mappings $f'_i : V(T') \to V(T'_i)$.

Now, let $S_1, \dots, S_k$ and $\{g_i : [0, \infty) \to S_i\}$ be the $\mathbb{R}$-stars and mappings guaranteed by Lemma 3.3. For each $i \in [k]$, let S'_i be the induced subgraph of S_i on the set $\{g_i(d_T(v, v_c)) : v \in V(\gamma_c)\}$, and make S'_i into a metric tree rooted at $g_i(0)$, with the length structure inherited from S_i. We now construct T_i by attaching S'_i to T'_i with the root of S'_i identified with the node $f'_i(v_c)$. The coloring χ'_i is extended to T_i by assigning to each root–leaf path in S'_i a new color. Finally, we specify functions $f_i : V \to V(T_i)$ via

$$f_i(v) = \begin{cases} f'_i(v), & v \in V(T'), \\ g_i(d_T(v_c, v)), & v \in V \setminus V(T'). \end{cases}$$

It is straightforward to verify that (26) holds for the colorings $\{\chi_i\}$ and every vertex $v \in V$. In addition, using the inductive hypothesis, we have $|V(T_i)| \le |V|$ and

$M(\chi) \le M(\chi_i)$ for every $i \in [k]$, with the latter condition following immediately from (26) and the structure of the mappings $\{f_i\}$.

We now verify that conditions (a), (b), and (c) hold. For $x, y \in V(T')$, the induction hypothesis guarantees all three conditions. If both $x, y \in V(\gamma_c) \setminus \{v_c\}$, then conditions (a) and (b) follow directly from conditions (ii) and (iii) of Lemma 3.3 applied to the maps $\{g_i\}$. To verify condition (c), let $j \in [k]$ be the single bad index from (vi). We have for all $i \ne j$,

$$\rho_{\chi_i}(f_i(x), f_i(y); 2^{-(k+1)}) \le 2^{k+1} d_T(x, y).$$

Since there are at most two colors on the path between x and y in any T_i, by condition (v) of Lemma 3.3, there are at most four values of $i \in [k] \setminus \{j\}$ such that

$$\rho_{\chi_i}(f_i(x), f_i(y); 2^{-(k+1)}) \ne 0,$$

hence

$$\frac{1}{k} \sum_{i \ne j} \rho_{\chi_i}(f_i(x), f_i(y); 2^{-(k+1)}) \le \frac{4 \cdot 2^{k+1}}{k} d_T(x, y) \le \varepsilon 2^{k+1} d_T(x, y).$$

Since $\|f_i\|_{\mathrm{Lip}}$ is determined on edges $(x, y) \in E$, and each such edge has $x, y \in V(\gamma_c)$ or $x, y \in V(T')$, we have already verified condition (b) for all $i \in [k]$ and $x, y \in V$. Finally, we verify (a) and (c) for pairs with $x \in V(T')$ and $y \in V(\gamma_c)$. We can check condition (a) using the previous two cases,

$$\begin{aligned}\frac{1}{k} \sum_{i=1}^{k} d_{T_i}(f_i(x), f_i(y)) &= \frac{1}{k} \sum_{i=1}^{k} \big(d_{T_i}(f_i(x), f_i(v_c)) + d_{T_i}(f_i(y), f_i(v_c))\big) \\ &\ge (1-\varepsilon) d_T(y, v_c) + (1-\varepsilon) d_T(x, v_c) \ge (1-\varepsilon) d_T(x, y).\end{aligned}$$

Towards verifying condition (c), note that by condition (v) from Lemma 3.3, there are at most two values of i, such that

$$\begin{aligned}&\rho_{\chi_i}(f_i(x), f_i(y); 2^{-(k+1)}) - \rho_{\chi_i}(f_i(x), f_i(v_c); 2^{-(k+1)}) \\ &\quad = \rho_{\chi_i}(f_i(y), f_i(v_c); 2^{-(k+1)}) \ne 0.\end{aligned}$$

By the induction hypothesis, there exists a number $j \in [k]$ such that

$$\varepsilon\, d_T(x, v_c) \ge \frac{2^{-(k+1)}}{k} \sum_{i \ne j} \rho_{\chi_i}(f_i(v_c), f_i(x); 2^{-(k+1)}).$$

Now we use condition (iv) from Lemma 3.3 to conclude

$$
\begin{aligned}
&\frac{2^{-(k+1)}}{k}\sum_{i\neq j}\rho_{\chi_i}(f_i(x), f_i(y); 2^{-(k+1)})\\
&\leq \frac{2^{-(k+1)}}{k}\sum_{i\neq j}\left(\rho_{\chi_i}(f_i(x), f_i(v_c); 2^{-(k+1)}) + \rho_{\chi_i}(f_i(y), f_i(v_c); 2^{-(k+1)})\right)\\
&\leq \varepsilon d_T(x, v_c) + 2\left(\frac{2^{-(k+1)}}{k}\right)(2^{k-1}d_T(y, v_c)) \leq \varepsilon\, d_T(x, v_c) + \varepsilon\, d_T(v_c, y)\\
&= \varepsilon\, d_T(x, y),
\end{aligned}
$$

completing the proof. □

3.2 Multi-scale Embeddings

We now present the basics of our multi-scale embedding approach. The next lemma is devoted to combining scales together without using too many dimensions, while controlling the distortion of the resulting map.

Lemma 3.4 *For every $\varepsilon \in (0, 1)$, the following holds. Let (X, d) be an arbitrary metric space, and consider a family of functions $\{f_i : X \to [0, 1]\}_{i\in\mathbb{Z}}$ such that for all $x, y \in X$, we have*

$$\sum_{i\in\mathbb{Z}} 2^i |f_i(x) - f_i(y)| < \infty. \tag{27}$$

Then there is a mapping $F : V \to \ell_1^{2+\lceil\log\frac{1}{\varepsilon}\rceil}$ such that for all $x, y \in X$,

$$(1-\varepsilon)\sum_{i\in\mathbb{Z}} 2^i|f_i(x) - f_i(y)| - 2\,\zeta(x, y) \leq \|F(x) - F(y)\|_1 \leq \sum_{i\in\mathbb{Z}} 2^i|f_i(x) - f_i(y)|,$$

where

$$\zeta(x, y) = \sum_{\substack{i:\exists j<i\\ f_j(x)-f_j(y)\neq 0}} 2^i(|f_i(x) - f_i(y)| - \lfloor |f_i(x) - f_i(y)|\rfloor).$$

Proof Let $k = 2 + \lceil\log 1/\varepsilon\rceil$, and fix some $x_0 \in X$. For $i \in [k]$, define $F_i : X \to \mathbb{R}$ by

$$F_i(x) = \sum_{j\in\mathbb{Z}} 2^{jk+i}(f_{jk+i}(x) - f_{jk+i}(x_0)). \tag{28}$$

It is easy to see that (27) implies absolute convergence of the preceding sum. We will consider the map $F = F_1 \oplus F_2 \oplus \cdots \oplus F_k : X \to \ell_1^k$. It is straightforward to verify that for every $x, y \in X$,

$$\|F(x) - F(y)\|_1 \le \sum_{i\in\mathbb{Z}} 2^i |f_i(x) - f_i(y)|.$$

Now, for $i \in [k]$, define

$$\zeta_i(x, y) = \sum_{\substack{j:\exists \ell < j \\ f_{\ell k+i}(x) - f_{\ell k+i}(y) \neq 0}} 2^{jk+i}(|f_{jk+i}(x) - f_{jk+i}(y)| - \lfloor |f_{jk+i}(x) - f_{jk+i}(y)| \rfloor).$$

One can easily check that $\sum_{i=1}^{k} \zeta_i(x, y) \le \zeta(x, y)$, thus showing the following for $i \in [k]$ will complete our proof of the lemma,

$$|F_i(x) - F_i(y)| \ge (1-\varepsilon) \sum_{j\in\mathbb{Z}} (2^{jk+i}|f_{jk+i}(x) - f_{jk+i}(y)|) - 2\zeta_i(x, y). \tag{29}$$

Toward this end, fix $i \in [k]$ and $x, y \in X$. Let $S = \{j \in \mathbb{Z} : |f_{jk+i}(x) - f_{jk+i}(y)| = 1\}$, and $T = \{j \in \mathbb{Z} : 0 < |f_{jk+i}(x) - f_{jk+i}(y)| < 1\}$. Clearly we then have

$$\begin{aligned}
&|F_i(x) - F_i(y)| \\
&\quad = \Big|\sum_{j\in S} 2^{jk+i}(f_{jk+i}(x) - f_{jk+i}(y)) + \sum_{j\in T} 2^{jk+i}(f_{jk+i}(x) - f_{jk+i}(y))\Big|.
\end{aligned}$$

If $S \cup T = \emptyset$, then (29) is immediate. Now, suppose that $S \neq \emptyset$, and let $c = i + k \cdot \max(S)$. Observe that $\max(S)$ exists by (27).

We then have

$$\begin{aligned}
&\sum_{j\in\mathbb{Z}} 2^{jk+i}|f_{jk+i}(x) - f_{jk+i}(y)| \\
&\quad \le 2^c + \sum_{\substack{j\in S\cup T \\ j<\max S}} 2^{kj+i} + \sum_{\substack{j\in T \\ j>\max S}} 2^{kj+i}|f_{kj+i}(x) - f_{kj+i}(y)| \\
&\quad \le 2^c + \sum_{j<\max S} 2^{kj+i} + \zeta_i(x, y) \le 2^c + 2 \cdot 2^{k(\max S - 1)+i} + \zeta_i(x, y) \\
&\quad \le 2^c(1 + 2^{1-k}) + \zeta_i(x, y) \le (1 + \varepsilon/2)2^c + \zeta_i(x, y).
\end{aligned}$$

On the other hand,

$$
\begin{aligned}
|F_i(x) - F_i(y)| &= \Big|\sum_{j\in\mathbb{Z}} 2^{kj+i}(f_{jk+i}(x) - f_{jk+i}(y))\Big| \\
&\geq 2^c - \sum_{\substack{j\in S\cup T\\ j<\max S}} 2^{kj+i} - \sum_{\substack{j\in T\\ j>\max S}} 2^{kj+i}|f_{kj+i}(x) - f_{kj+i}(y)| \\
&\geq 2^c - \sum_{j<\max S} 2^{kj+i} - \zeta_i(x,y) \geq 2^c - 2\cdot 2^{k(\max S-1)+i} - \zeta_i(x,y) \\
&\geq 2^c(1-2^{1-k}) - \zeta_i(x,y) \geq (1-\varepsilon/2)2^c - \zeta_i(x,y).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&(1-\varepsilon)\sum_{j\in\mathbb{Z}} 2^{kj+i}|f_{jk+i}(x) - f_{jk+i}(y)| \\
&\quad\leq (1-\varepsilon)((1+\varepsilon/2)2^c + \zeta_i(x,y)) \leq (1-\varepsilon/2)2^c + \zeta_i(x,y) \\
&\quad\leq |F_i(x) - F_i(y)| + 2\zeta_i(x,y),
\end{aligned}
$$

completing the verification of (29) in the case when $S \neq \emptyset$.

In the remaining case when $S = \emptyset$ and $T \neq \emptyset$, if the set T does not have a minimum element, then

$$
\sum_{j\in T} 2^{kj+i}|f_{kj+i}(x) - f_{kj+i}(y)| = \zeta_i(x,y),
$$

making (29) vacuous since the right-hand side is non-positive.

Otherwise, let $\ell = \min(T)$, and write

$$
\begin{aligned}
&|F_i(x) - F_i(y)| \\
&\quad= \Big|\sum_{j\in T} 2^{kj+i}(f_{kj+i}(x) - f_{kj+i}(y))\Big| \\
&\quad\geq 2^{\ell k+i}|f_{\ell k+i}(x) - f_{\ell k+i}(y)| - \Big|\sum_{j\in T, j>\ell} 2^{kj+i}(f_{kj+i}(x) - f_{kj+i}(y))\Big| \\
&\quad\geq 2^{\ell k+i}|f_{\ell k+i}(x) - f_{\ell k+i}(y)| - \zeta_i(x,y) \\
&\quad= \sum_{j\in\mathbb{Z}} 2^{kj+i}|f_{kj+i}(x) - f_{kj+i}(y)| - 2\,\zeta_i(x,y).
\end{aligned}
$$

This completes the proof. □

In Sect. 5, we will require the following straightforward corollary.

Corollary 3.5 *For every $\varepsilon \in (0,1)$ and $m \in \mathbb{N}$, the following holds. Let (X,d) be a metric space, and suppose we have a family of functions $\{f_i : X \to [0,1]^m\}_{i\in\mathbb{Z}}$ such* *that for all $x, y \in X$,*

$$\sum_{i\in\mathbb{Z}} 2^i \| f_i(x) - f_i(y)\|_1 < \infty.$$

Then there exists a mapping $F : V \to \ell_1^{m(2+\lceil \log \frac{1}{\varepsilon}\rceil)}$ such that for all $x, y \in X$,

$$(1-\varepsilon)\sum_{i\in\mathbb{Z}} \big(2^i \| f_i(x) - f_i(y)\|_1\big) - 2\,\zeta(x,y) \le \|F(x) - F(y)\|_1 \le \sum_{i\in\mathbb{Z}} 2^i \| f_i(x) - f_i(y)\|_1,$$

where

$$\zeta(x,y) = \sum_{k=1}^{m} \sum_{\substack{i:\exists j<i\\ f_j(x)_k - f_j(y)_k \neq 0}} 2^i \big(|f_i(x)_k - f_i(y)_k| - \lfloor |f_i(x)_k - f_i(y)_k| \rfloor\big), \tag{30}$$

and we have used the notation x_k for the k-th coordinate of $x \in \mathbb{R}^m$.

4 Scale Assignment

Let $T = (V, E)$ be a metric tree with root $r \in V$, equipped with a monotone coloring $\chi : E \to \mathbb{N}$. We will now describe a way of assigning "scales" to the vertices of T. These scale values will be used in Sect. 5 to guide our eventual embedding. The scales of a vertex will describe, roughly, the subset and magnitude of coordinates that should differ between the vertex and its neighbors. First, we fix some notation.

For every $c \in \chi(E)$, we use γ_c to denote the path in T colored c, and we use v_c to denote the vertex of γ_c which is closest to the root. We will also use the notation $T(c)$ to denote the subtree of T under the color c; formally, $T(c)$ is the induced (rooted) subtree on $\{v_c\} \cup V(T_u)$ where $u \in V$ is the child of v_c such that $\chi(v_c, u) = c$, and T_u is the subtree rooted at u.

We will write $p(v)$ for the parent of a vertex $v \in V$, and $p(r) = r$. Furthermore, we define the "parent color" of a color class by $\rho(c) = \chi(v_c, p(v_c))$ with the convention that $\chi(r, r) = c_0$, where $c_0 \in \mathbb{N} \setminus \chi(E)$ is some fixed element. Finally, we put $T(c_0) = T$.

4.1 Scale Selectors

We start by defining a function $\kappa : \chi(E) \cup \{c_0\} \to \mathbb{N}$ which describes the "branching factor" for each color class,

$$\kappa(c) = \left\lfloor \log_2 \frac{|E(T(\rho(c)))|}{|E(T(c))|} \right\rfloor + 1. \tag{31}$$

Moreover, we define $\varphi : \chi(E) \cup \{c_0\} \to \mathbb{N} \cup \{0\}$ inductively by setting $\varphi(c_0) = 0$, and

$$\varphi(c) = \kappa(c) + \varphi(\rho(c)) \tag{32}$$

for $c \in \chi(E)$.

Observe that for every color $c \in \chi(E)$, we have

$$\begin{aligned} \varphi(c) = \sum_{c' \in \chi(E(P_{v_c})) \cup \{c\}} \kappa(c') &\le \sum_{c' \in \chi(E(P_{v_c})) \cup \{c\}} \left(1 + \log_2 \frac{|E(T(\rho(c')))|}{|E(T(c'))|}\right) \\ &\le M(\chi) + \log_2 |E| \end{aligned} \tag{33}$$

Next, we use φ to inductively define our scale selectors. Let

$$m(T) = \min\left\{\mathsf{len}(e) : e \in E \text{ and } \mathsf{len}(e) > 0\right\}.$$

We now define a family of functions $\{\tau_i : V \to \mathbb{N} \cup \{0\}\}_{i \in \mathbb{Z}}$.

For $v \in V$, let $c = \chi(v, p(v))$, and put

$$\tau_i(v) = 0 \quad \text{for } i < \left\lfloor \log_2 \left(\frac{m(T)}{M(\chi) + \log_2 |E|}\right)\right\rfloor,$$

and otherwise,

$$\tau_i(v) = \min\Bigg(\underbrace{\left\lceil \frac{d_T(v, v_c) - \min\left(d_T(v, v_c), \sum_{j=-\infty}^{i-1} 2^j \tau_j(v)\right)}{2^i}\right\rceil}_{\text{(A)}}, \underbrace{\varphi(c) - \sum_{c' \in \chi(E(P_v))} \tau_i(v_{c'})}_{\text{(B)}}\Bigg). \tag{34}$$

The value of $\tau_i(v)$ will be used in Sect. 5 to determine how many coordinates of magnitude $\asymp 2^i$ change as the embedding proceeds from v_c to v. In this definition, we try to cover the distance from root to v with the smallest scales possible while satisfying the inequality

$$\varphi(c) \ge \tau_i(v) + \sum_{c' \in \chi(E(P_v))} \tau_i(v_{c'}).$$

For $v \in V \setminus \{r\}$, let $c = \chi(v, p(v))$, for each $i \in \mathbb{Z}$, part (B) of (34) for $\tau_i(v_c)$ implies that

$$\tau_i(v_c) \le \varphi(\rho(c)) - \sum_{c' \in \chi(E(P_{v_c}))} \tau_i(v_{c'}).$$

Hence,

$$\begin{aligned}\varphi(c) &- \sum_{c' \in \chi(E(P_v))} \tau_i(v_{c'}) \\ &= \varphi(c) - \tau_i(v_c) - \sum_{c' \in \chi(E(P_{v_c}))} \tau_i(v_{c'}) \geq \varphi(c) - \varphi(\rho(c)) = \kappa(c) \geq 1. \end{aligned} \tag{35}$$

Therefore, part (B) of (34) is always positive, so if $\tau_k(v) = 0$ for some $k \geq \lfloor \log_2 (\frac{m(T)}{M(\chi)+\log_2 |E|}) \rfloor$, then $\tau_k(v)$ is defined by part (A) of (18). Hence $\sum_{j=-\infty}^{k-1} 2^j \tau_j(v) \geq d_T(v, v_c)$ and the following observation is immediate.

Observation 4.1 For $v \in V$ and $k \geq \lfloor \log_2 (\frac{m(T)}{M(\chi)+\log_2 |E|}) \rfloor$, if $\tau_k(v) = 0$ then for all $i \geq k$, $\tau_i(v) = 0$.

Comparing part (A) of (34) for $\tau_i(v)$ and $\tau_{i+1}(v)$ also allows us to observe the following.

Observation 4.2 For $v \in V$ and $k \geq \lfloor \log_2 (\frac{m(T)}{M(\chi)+\log_2 |E|}) \rfloor$, if part (A) in (34) for $\tau_k(v)$ is less than or equal to part (B) then for all $i > k$, $\tau_i(v) = 0$.

4.2 Properties of the Scale Selector Maps

We now prove some key properties of the maps κ, φ, and $\{\tau_i\}$.

Lemma 4.3 *For every vertex $v \in V$ with $c = \chi(v, p(v))$, the following holds. For all $i \in \mathbb{Z}$ with $\frac{d_T(v,v_c)}{\kappa(c)} \leq 2^{i-1}$, we have $\tau_i(v) = 0$.*

Proof If $d_T(v, v_c) = 0$, the lemma is vacuous. Suppose now that $d_T(v, v_c) > 0$, and let $k = \lceil \log_2 (\frac{d_T(v,v_c)}{\kappa(c)}) \rceil$. We have $d_T(v, v_c) \geq m(T)$ and $\kappa(c) \leq \log_2 |E| + 1$, therefore

$$k \geq \left\lfloor \log_2 \left(\frac{m(T)}{M(\chi) + \log_2 |E|} \right) \right\rfloor.$$

It follows that for $i \geq k$, $\tau_i(v)$ is given by (34).

If $\tau_k(v) = 0$, then by Observation 4.1, for all $i \geq k$, $\tau_i(v) = 0$.

On the other hand if $\tau_k(v) \neq 0$ then either it is determined by part (B) of (34), in which case

$$\begin{aligned}\tau_k(v) = \varphi(c) - \sum_{c' \in \chi(E(P_v))} \tau_k(v_{c'}) &= \varphi(c) - \tau_k(v_c) - \sum_{c' \in \chi(E(P_{v_c}))} \tau_k(v_{c'}) \\ &\geq \varphi(c) - \varphi(\rho(c)) = \kappa(c),\end{aligned}$$

implying that

$$\sum_{j=-\infty}^{k} 2^j \tau_j(v) \geq \kappa(c) 2^k \geq d_T(v, v_c).$$

Examining part (A) of (34), we see that $\tau_{k+1}(v) = 0$, and by Observation 4.1, $\tau_i(v) = 0$ for $i > k$. Alternately, $\tau_k(v)$ is determined by part (A) of (34), and by Observation 4.2 $\tau_i(v) = 0$ for $i > k$, completing the proof. □

The next lemma shows how the values $\{\tau_i(v)\}$ track the distance from v_c to v.

Lemma 4.4 *For any vertex $v \in V$ with $c = \chi(v, p(v))$, we have*

$$d_T(v, v_c) \le \sum_{i=-\infty}^{\infty} 2^i \tau_i(v) \le 3\, d_T(v, v_c).$$

Proof If $d_T(v, v_c) = 0$, the lemma is vacuous. Suppose now that $d_T(v, v_c) > 0$, and let

$$k = \max\{i : \tau_i(v) \ne 0\}.$$

By Lemma 4.3, the maximum exists.

We have $\tau_{k+1}(v) = 0$, and thus inequality (35) implies that part (A) of (34) specifies $\tau_{k+1}(v)$, yielding

$$d_T(v, v_c) \le \sum_{i=-\infty}^{k} 2^i \tau_i(v) = \sum_{i=-\infty}^{\infty} 2^i \tau_i(v).$$

On the other hand, since $\tau_k(v) > 0$, we must have $d_T(v, v_c) > \sum_{i=-\infty}^{k-1} 2^i \tau_i(v)$, and Lemma 4.3 implies that $2^k < 2\, d_T(v, v_c)$, hence,

$$\begin{aligned}
\sum_{i=-\infty}^{k} 2^i \tau_i(v) &\le \sum_{i=-\infty}^{k-1} 2^i \tau_i(v) + 2^k \left\lceil \frac{d_T(v, v_c) - \sum_{i=-\infty}^{k-1} 2^i \tau_i(v)}{2^k} \right\rceil \\
&< \sum_{i=-\infty}^{k-1} 2^i \tau_i(v) + 2^k \left(\frac{d_T(v, v_c) - \sum_{i=-\infty}^{k-1} 2^i \tau_i(v)}{2^k} + 1 \right) \\
&= \sum_{i=-\infty}^{k-1} 2^i \tau_i(v) + 2^k + \left(d_T(v, v_c) - \sum_{i=-\infty}^{k-1} 2^i \tau_i(v) \right) \\
&\le d_T(v, v_c) + 2^k < 3\, d_T(v, v_c).
\end{aligned}$$

□

The following lemma shows that for any color $c \in \chi(E)$ the value of τ_i does not decrease as we move further from v_c in γ_c.

Lemma 4.5 *Let $u, w \in V$ be such that $c = \chi(w, p(w)) = \chi(u, p(u))$, and $d_T(w, v_c) \le d_T(u, v_c)$. Then for all $i \in \mathbb{Z}$, we have*

$$\tau_i(w) \le \tau_i(u).$$

Proof First let k be the smallest integer for which

$$\left\lceil \frac{d_T(w, v_c) - \min\left(d_T(w, v_c), \sum_{j=-\infty}^{k-1} 2^j \tau_j(w)\right)}{2^k} \right\rceil \le \varphi(c) - \sum_{c' \in \chi(E(P_w))} \tau_k(v_{c'}).$$

This k exists since, by (35), the right-hand side is always positive, while by Lemma 4.3, the left-hand side must be zero for some $k \in \mathbb{Z}$ large enough.

For $i > k$, by Observation 4.2 we have, $\tau_i(w) = 0$. Therefore, for $i > k$, we have $\tau_i(u) \ge \tau_i(w)$. We now use induction on i to show that for $i < k$, $\tau_i(u) = \tau_i(w)$, and for $i = k$, $\tau_k(u) \ge \tau_k(w)$. Recall that, for $i < \left\lfloor \log_2\left(\frac{m(T)}{M(\chi)+\log_2|E|}\right)\right\rfloor$, we have $\tau_i(w) = \tau_i(u) = 0$, which gives us the base case of the induction.

Now, by definition of k, part (B) of (34) for $\tau_{k-1}(w)$ is an integer strictly less than part (A), hence

$$\begin{aligned} \sum_{j=-\infty}^{k-1} 2^j \tau_j(w) &= 2^{k-1}\tau_{k-1}(w) + \sum_{j=-\infty}^{k-2} 2^j \tau_j(w) \\ &\le 2^{k-1}\left(\left\lceil \frac{d_T(w, v_c) - \sum_{j=-\infty}^{k-2} 2^j \tau_j(w)}{2^{k-1}} \right\rceil - 1\right) + \sum_{j=-\infty}^{k-2} 2^j \tau_j(w) \\ &< 2^{k-1}\left(\frac{d_T(w, v_c) - \sum_{j=-\infty}^{k-2} 2^j \tau_j(w)}{2^{k-1}}\right) + \sum_{j=-\infty}^{k-2} 2^j \tau_j(w) \\ &\le d_T(w, v_c). \end{aligned} \tag{36}$$

For $\left\lfloor \log_2\left(\frac{m(T)}{M(\chi)+\log_2|E|}\right)\right\rfloor \le i \le k$, by (36), and as $d_T(u, v_c) \ge d_T(w, v_c)$, we have

$$\begin{aligned} &\min\left(d_T(w, v_c), \sum_{j=-\infty}^{i-1} 2^j \tau_j(w)\right) \\ &\quad = \sum_{j=-\infty}^{i-1} 2^j \tau_j(w) = \min\left(d_T(u, v_c), \sum_{j=-\infty}^{i-1} 2^j \tau_j(w)\right). \end{aligned} \tag{37}$$

By our induction hypothesis for all $j < i$, $\tau_j(w) = \tau_j(u)$, so using (37) we can write

$$\begin{aligned} &d_T(w, v_c) - \min\left(d_T(w, v_c), \sum_{j=-\infty}^{i-1} 2^j \tau_j(w)\right) \\ &\quad \le d_T(u, v_c) - \min\left(d_T(u, v_c), \sum_{j=-\infty}^{i-1} 2^j \tau_j(u)\right). \end{aligned} \tag{38}$$

Since $\chi(w, p(w)) = \chi(u, p(u))$, for all $i \in \mathbb{Z}$ part (B) of (34) is identical for $\tau_i(u)$ and $\tau_i(w)$. Therefore, using (38), and the definition of k, for all $\left\lfloor \log_2\left(\frac{m(T)}{M(\chi)+\log_2|E|}\right)\right\rfloor \le$

$i < k$, part (B) of (34) specifies $\tau_i(u)$ and $\tau_i(w)$, hence

$$\tau_i(u) = \tau_i(w) = \varphi(c) - \sum_{c' \in \chi(E(P_w))} \tau_i(v_{c'}).$$

For the case that $i = k$, part (B) of (34) is identical for $\tau_k(u)$ and $\tau_k(w)$, and inequality (38) implies that part (A) of (34) for $\tau_k(u)$ is at least as large as part (A) of (34) for $\tau_k(w)$, completing the proof. □

The next lemma bounds the distance between two vertices in the graph based on $\{\tau_i\}$.

Lemma 4.6 *Let $k > \left\lfloor \log_2 \left(\frac{m(T)}{M(\chi)+\log_2 |E|}\right)\right\rfloor$ be an integer. For any two vertices w and u such that $\tau_k(u) \neq 0$, $\tau_{k-1}(w) = 0$ and $\chi(w, p(w)) = \chi(u, p(u))$, we have*

$$d_T(u, w) > 2^{k-1}.$$

Proof By Observation 4.1, $\tau_k(w) = 0$. Letting $c = \chi(u, p(u))$, by Lemma 4.5 we have $d_T(v_c, u) \geq d_T(v_c, w)$. Using Lemma 4.5 again, we can conclude that for all $i \in \mathbb{Z}$, $\tau_i(u) \geq \tau_i(w)$. Since $\tau_{k-1}(w) = 0$, inequality (35) implies that part (A) of (34) specifies $\tau_{k-1}(w)$. Therefore,

$$\begin{aligned} d_T(w, v_c) &\leq \sum_{i=-\infty}^{k-2} 2^i \tau_i(w) \leq \sum_{i=-\infty}^{k-2} 2^i \tau_i(u) \\ &= \left(\sum_{i=-\infty}^{k-1} 2^i \tau_i(u) \right) - 2^{k-1} \tau_{k-1}(u). \end{aligned} \tag{39}$$

Since $\tau_k(u) > 0$, using part (A) of (34), we can write

$$d_T(u, v_c) > \sum_{i=-\infty}^{k-1} 2^i \tau_i(u). \tag{40}$$

Observation 4.1 implies that $\tau_{k-1}(u) \neq 0$, thus $\tau_{k-1}(u) \geq 1$, and using (39) and (40), we have

$$d_T(w, u) = d_T(u, v_c) - d_T(w, v_c) > 2^{k-1},$$

completing the proof. □

The next lemma and the following two corollaries bound the number of colors c in the tree which have a small value of $\varphi(c)$.

Lemma 4.7 *For any $k \in \mathbb{N} \cup \{0\}$, and any color $c \in \chi(E)$, we have*

$$\#\left\{c' \in \chi(E(T(c))) : \varphi(c') - \varphi(c) = k\right\} \leq 2^k.$$

Proof We start the proof by comparing the size of the subtrees $T(c')$ and $T(c)$ for $c' \in \chi(E(T(c)))$.

For a given color $c' \in \chi(E(T(c)))$, we define the sequence $\{c_i\}_{i\in\mathbb{N}}$ as follows. We put $c_1 = c'$ and for $i > 1$ we put $c_i = \rho(c_{i-1})$. Suppose now that $c_m = c$, we have

$$\varphi(c_m) - \varphi(c_1) = \sum_{i=1}^{m-1} \kappa(c_i) \geq \sum_{i=1}^{m-1} \log_2\left(\frac{|E(T(c_{i+1}))|}{|E(T(c_i))|}\right) \geq \log_2\left(\frac{|E(T(c))|}{|E(T(c'))|}\right). \quad (41)$$

This inequality implies that

$$|E(T(c))| \leq 2^{\varphi(c')-\varphi(c)}|E(T(c'))|.$$

It is easy to check that for colors $a, b \in \chi(E(T(c)))$ such that $\varphi(a) = \varphi(b)$, subtrees $T(a)$ and $T(b)$ are edge disjoint. Therefore, for $k \in \mathbb{N} \cup \{0\}$, summing over all the colors c' such that $\varphi(c') - \varphi(c) = k$ gives

$$\begin{aligned}
&\#\{c' \in \chi(E(T(c))) : \varphi(c') - \varphi(c) = k\} \\
&\quad \leq \sum_{\substack{c'\in\chi(E(T(c)))\\ \varphi(c')-\varphi(c)=k}} \frac{2^k\,|E(T(c'))|}{|E(T(c))|} = 2^k \sum_{\substack{c'\in\chi(E(T(c)))\\ \varphi(c')-\varphi(c)=k}} \frac{|E(T(c'))|}{|E(T(c))|} \leq 2^k.
\end{aligned}$$

□

The following two corollaries are immediate from Lemma 4.7.

Corollary 4.8 *For any $k \in \mathbb{N}$, and any color $c \in \chi(E)$, we have*

$$\#\{c' \in \chi(E(T(c))) : \varphi(c') - \varphi(c) \leq k\} < 2^{k+1}.$$

Corollary 4.9 *For any color $c \in \chi(E)$, and constant $C \geq 2$, we have*

$$\sum_{c'\in\chi(E(T(c)))\setminus\{c\}} 2^{-C(\varphi(c')-\varphi(c))} < 2^{2-C}.$$

The next lemma is similar to Lemma 4.6. The assumption is more general, and the conclusion is correspondingly weaker. This result is used primarily to enable the proof of Lemma 4.11.

Lemma 4.10 *Let $u \in V$ and $w \in V(P_u)$ be such that $\varphi(\chi(u, p(u))) > \varphi(\chi(w, p(w)))$. For all vertices $x \in V(T_u)$, and $k \in \mathbb{Z}$ with*

$$2^k \geq \left(\frac{6\, d_T(x, w)}{\varphi(\chi(u, p(u))) - \varphi(\chi(w, p(w)))}\right), \quad (42)$$

we have $\tau_k(x) = 0$.

Proof In the case that $d_T(x, w) = 0$, this lemma is vacuous. Suppose now that $d_T(x, w) > 0$. Let $c_1, \ldots, c_m$ be the set of colors that appear on the path $P_{x\, p(w)}$, in order from x to $p(w)$, and for $i \in [m]$, let $y_i = v_{c_i}$. We prove this lemma by showing that if

$$k \geq \log_2 \left(\frac{6\, d_T(x, w)}{\varphi(\chi(u, p(u))) - \varphi(\chi(w, p(w)))} \right), \tag{43}$$

then part (A) of (34) for $\tau_k(x)$ is zero. First note that $\varphi(\chi(u, p(u))) - \varphi(\chi(w, p(w))) \leq M(\chi) + \log_2 |E|$ and $d_T(x, w) \geq m(T)$, hence (43) implies

$$k \geq \left\lfloor \log_2 \left(\frac{m(T)}{M(\chi) + \log_2 |E|} \right) \right\rfloor.$$

By Lemma 4.4, we have

$$\sum_{i=1}^{m-2} 2^{k-1} \tau_{k-1}(y_i) \leq \sum_{i=1}^{m-2} \sum_{j=-\infty}^{\infty} 2^j \tau_j(y_i) \leq \sum_{i=1}^{m-2} 3\, d_T(y_i, y_{i+1}) = 3\, d_T(y_1, y_{m-1}). \tag{44}$$

Now, using (42) gives

$$\begin{aligned} \varphi(c_1) - \varphi(c_m) &\geq \varphi(\chi(u, p(u))) - \varphi(\chi(w, p(w))) \\ &\geq \frac{6\, d_T(x, w)}{2^k} \geq \frac{6\, d_T(x, y_{m-1})}{2^k}. \end{aligned} \tag{45}$$

Using the above inequality and (44), we can write

$$\begin{aligned} d_T(x, y_1) &= d_T(x, y_{m-1}) - d_T(y_1, y_{m-1}) \\ &\leq \frac{2^{k-1}}{3} \Bigl(\varphi(c_1) - \varphi(c_m) - \sum_{i=1}^{m-2} \tau_{k-1}(y_i) \Bigr). \end{aligned}$$

First, note that $c_m = \chi(y_{m-1}, p(y_{m-1}))$. Now, we use part (B) of (34) for $\tau_k(y_{m-1})$ to write

$$\begin{aligned} d_T(x, y_1) &\leq \frac{2^{k-1}}{3} \Bigl(\varphi(c_1) - \bigl(\tau_{k-1}(y_{m-1}) + \sum_{c' \in \chi(E(P_{y_{m-1}}))} \tau_{k-1}(v_{c'}) \bigr) - \sum_{i=1}^{m-2} \tau_{k-1}(y_i) \Bigr) \\ &\leq \frac{2^{k-1}}{3} \Bigl(\varphi(c_1) - \sum_{c' \in \chi(E(P_x))} \tau_{k-1}(v_{c'}) \Bigr) \\ &\leq 2^{k-1} \Bigl(\varphi(\chi(x, p(x))) - \sum_{c' \in \chi(E(P_x))} \tau_{k-1}(v_{c'}) \Bigr). \end{aligned} \tag{46}$$

Therefore, either part (A) of (34) specifies $\tau_{k-1}(x)$ in which case by Observation 4.2, $\tau_i(v) = 0$ for $i \geq k$, or part (B) of (34) specifies $\tau_{k-1}(x)$ in which case by (46) we have

$$\tau_{k-1}(x)2^{k-1} \geq d_T(x, y_1),$$

and part (A) of (34) is zero for $i \geq k$. □

In Sect. 5, we give the description of our embedding and analyze its distortion. In the analysis of the embedding, for a given pair of vertices $x, y \in V$, we divide the path between x and y into subpaths and for each subpath we show that either the contribution of that subpath to the distance between x and y in the embedding is "large" through a concentration of measure argument, or we use the following lemma to show that the length of the subpath is "small," compared to the distance between x and y. The complete argument is somewhat more delicate and one can find the details of how Lemma 4.11 is used in the proof of Lemma 5.15.

Lemma 4.11 *There exists a constant $C > 0$ such that the following holds. For any $c \in \chi(E)$ and $v \in V(T(c))$ with $v \neq v_c$ and for any $\varepsilon \in (0, \frac{1}{2}]$, there are vertices $u, u' \in V$ with $u \neq u'$ and $d_T(u, v) \leq \varepsilon\, d_T(u, u')$, and such that*

$$u, u' \in \{v_a : a \in \chi(E(P_{v\, v_c}))\} \cup \{v\}.$$

Furthermore, for all vertices $x \in V(P_{u'u}) \setminus \{u'\}$, for all $k \in \mathbb{Z}$,

$$\tau_k(x) \neq 0 \implies 2^k < \left(\frac{C d_T(u, u')}{\varepsilon(\varphi(\chi(u, p(u))) - \varphi(\chi(v_c, p(v_c))))}\right).$$

Proof Let $r' = v_c$, and let $c_1, \ldots, c_m$ be the set of colors that appear on the path $P_{vr'}$ in order from v to r', and put $c_{m+1} = \chi(r', p(r'))$. We define $y_0 = v$, and for $i \in [m]$, $y_i = v_{c_i}$. Note that $\{y_0, \ldots, y_m\} = \{v\} \cup \{v_a : a \in \chi(E(P_{v\, v_c}))\}$, and for $i \leq m$, $\chi(y_i, p(y_i)) = c_{i+1}$. We give a constructive proof for the lemma.

For $i \in \mathbb{N}$, we construct a sequence $(a_i, b_i) \in \mathbb{N} \times \mathbb{N}$, the idea being that $P_{y_{a_i}, y_{b_i}}$ is a nonempty subpath $P_{vr'}$ such that for different values of i, these subpaths are edge disjoint. At each step of the construction either we can use (a_i, b_i) to find u and u' such that they satisfy the properties of this lemma, or we find (a_{i+1}, b_{i+1}) such that $b_{i+1} < b_i$. The last condition guarantees that we can always find u and u' that satisfy the conditions of this lemma.

We start with $a_1 = m$ and $b_1 = m - 1$. If $d_T(v, y_{b_1}) \leq \varepsilon d_T(y_{a_1}, y_{b_1})$ then

$$\left(\frac{2d_T(y_m, y_{m-1})}{\varphi(\chi(y_{m-1}, p(y_{m-1}))) - \varphi(\chi(r', p(r')))}\right) = \frac{2d_T(y_{a_1}, y_{b_1})}{\kappa(c)}$$

and by Lemma 4.3 the assignment $u' = y_{a_1}$ and $u = y_{b_1}$ satisfies the conditions of this lemma if $C \geq 1$. Otherwise, for $i \geq 1$, we choose (a_{i+1}, b_{i+1}) based on (a_i, b_i), and construct the rest of the sequence preserving the following three properties:

(i) $\varphi(c_{b_i+1}) - \varphi(c_{a_i+1}) \geq \varphi(c_{a_i+1}) - \varphi(\chi(r', p(r')))$;
(ii) $d_T(y_{b_i}, v) \geq \varepsilon d_T(y_{b_i}, y_{a_i})$;
(iii) $a_i > b_i$.

Let $j \in \{0, \ldots, m\}$ be the maximum integer such that $\varepsilon d_T(y_j, y_{b_i}) \geq d_T(v, y_j)$. Note that $j < b_i$, and the maximum always exists because $y_0 = v$. We will now split the proof into three cases.

Case I: $\varphi(c_{j+2}) - \varphi(c_{b_i+1}) \geq 2(\varphi(c_{b_i+1}) - \varphi(c_{a_i+1}))$.

In this case by condition (iii), $\varphi(c_{b_i+1}) - \varphi(c_{a_i+1}) > 0$. Hence $j+1 < b_i$, and we can preserve conditions (i), (ii) and (iii) with

$$(a_{i+1}, b_{i+1}) = (b_i, j+1).$$

Case II: $\varphi(c_{j+2}) - \varphi(c_{b_i+1}) < 2(\varphi(c_{b_i+1}) - \varphi(c_{a_i+1}))$ and $\varphi(c_{j+1}) - \varphi(c_{b_i+1}) \geq 6(\varphi(c_{b_i+1}) - \varphi(c_{a_i+1}))$.

In this case by (32) we have

$$\kappa(c_{j+1}) = \varphi(c_{j+1}) - \varphi(c_{j+2}) = (\varphi(c_{j+1}) - \varphi(c_{b_i+1})) - (\varphi(c_{j+2}) - \varphi(c_{b_i+1})).$$

Using the conditions of this case, we write

$$\begin{aligned}
\kappa(c_{j+1}) &= (\varphi(c_{j+1}) - \varphi(c_{b_i+1})) - (\varphi(c_{j+2}) - \varphi(c_{b_i+1}))\\
&\geq 6(\varphi(c_{b_i+1}) - \varphi(c_{a_i+1})) - (\varphi(c_{j+2}) - \varphi(c_{b_i+1}))\\
&= \big(2(\varphi(c_{b_i+1}) - \varphi(c_{a_i+1})) + 4(\varphi(c_{b_i+1}) - \varphi(c_{a_i+1}))\big) - \big(\varphi(c_{j+2}) - \varphi(c_{b_i+1})\big)\\
&> \big(2(\varphi(c_{b_i+1}) - \varphi(c_{a_i+1})) + 2(\varphi(c_{j+2}) - \varphi(c_{b_i+1}))\big) - \big(\varphi(c_{j+2}) - \varphi(c_{b_i+1})\big),
\end{aligned}$$

and by condition (i),

$$\begin{aligned}
\kappa(c_{j+1}) > \Big(&\big(\varphi(c_{b_i+1}) - \varphi(c_{a_i+1})\big) + \big(\varphi(c_{a_i+1}) - \varphi(\chi(r', p(r')))\big)\\
&+ 2(\varphi(c_{j+2}) - \varphi(c_{b_i+1}))\Big) - \big(\varphi(c_{j+2}) - \varphi(c_{b_i+1})\big)\\
= \ &\varphi(c_{j+2}) - \varphi(\chi(r', p(r'))).
\end{aligned} \tag{47}$$

Thus if $d_T(y_{j+1}, v) \geq \varepsilon\, d_T(y_j, y_{j+1})$, then $(a_{i+1}, b_{i+1}) = (j+1, j)$, satisfies condition (i) by (47), and it is also easy to verify that it satisfies conditions (ii) and (iii). If $d_T(y_{j+1}, v) < \varepsilon\, d_T(y_j, y_{j+1})$, then by (32),

$$\varphi(\chi(y_j, p(y_j))) = \varphi(c_{j+1}) = \kappa(c_{j+1}) + \varphi(c_{j+2})$$

and by (47),

$$\begin{aligned}
&\Big(\frac{2d_T(y_j, y_{j+1})}{(\varphi(\chi(y_j, p(y_j))) - \varphi(\chi(r', p(r'))))}\Big)\\
&\quad = \Big(\frac{2d_T(y_j, y_{j+1})}{\kappa(c_{j+1}) + \varphi(c_{j+2}) - \varphi(\chi(r', p(r')))}\Big) > \frac{d_T(y_j, y_{j+1})}{\kappa(c_{j+1})}.
\end{aligned}$$

Hence Lemma 4.3 implies that the assignment $u' = y_{j+1}$ and $u = y_j$ satisfies the conditions of this lemma if $C \geq 2$.

Case III: $\varphi(c_{j+1}) - \varphi(c_{b_i+1}) < 6(\varphi(c_{b_i+1}) - \varphi(c_{a_i+1}))$.

In this case we use Lemma 4.10 to show that the assignment $u = y_j$ and $u' = y_{b_i}$ satisfies the conditions of the lemma. We have

$$\begin{aligned}
&\varphi(\chi(y_j, p(y_j))) - \varphi(\chi(r', p(r')))\\
&\quad = \varphi(c_{j+1}) - \varphi(\chi(r', p(r')))\\
&\quad = (\varphi(c_{j+1} - \varphi(c_{b_i+1})) + (\varphi(c_{b_i+1}) - \varphi(c_{a_i+1})) + (\varphi(c_{a_i+1}) - \varphi(\chi(r', p(r'))))\\
&\quad < 6(\varphi(c_{b_i+1}) - \varphi(c_{a_i+1})) + (\varphi(c_{b_i+1}) - \varphi(c_{a_i+1})) + (\varphi(c_{a_i+1}) - \varphi(\chi(r', p(r')))),
\end{aligned}$$

and by condition (i),

$$\varphi(\chi(y_j, p(y_j))) - \varphi(\chi(r', p(r'))) < 8(\varphi(c_{b_i+1}) - \varphi(c_{a_i+1})).$$

Condition (ii) and the definition of y_j imply that

$$d_T(y_j, y_{b_i}) \geq (1-\varepsilon) d_T(v, y_{b_i}) \geq \varepsilon(1-\varepsilon) d_T(y_{a_i}, y_{b_i}) \geq \frac{\varepsilon}{2} d_T(y_{a_i}, y_{b_i}).$$

Hence,

$$\left(\frac{6(\frac{2}{\varepsilon}) d_T(y_j, y_{b_i})}{\frac{1}{8}(\varphi(\chi(y_j, p(y_j))) - \varphi(\chi(r', p(r'))))}\right) \geq \left(\frac{6 d_T(y_{b_i}, y_{a_i})}{\varphi(c_{b_i+1}) - \varphi(c_{a_i+1})}\right),$$

and by applying Lemma 4.10 with $u = y_{b_i}$ and $w = y_{a_i}$, we can conclude that the assignment $u = y_j$ and $u' = y_{b_i}$ satisfies the conditions of this lemma with $C = 96$. □

5 The Embedding

We now present a proof of Theorem 3.1, thereby completing the proof of Theorem 1.1. We first introduce a random embedding of the tree T into ℓ_1, and then show that, for a suitable choice of parameters, with non-zero probability our construction satisfies the conditions of the theorem.

Notation: We use the notations and definitions introduced in Sect. 4. Moreover, in this section, for $c \in \chi(E) \cup \{\chi(r, p(r))\}$, we use $\rho^{-1}(c)$ to denote the set of colors $c' \in \chi(E)$ such that $\rho(c') = c$, i.e. the colors of the "children" of c. For $m, n \in \mathbb{N}$, and $A \in \mathbb{R}^{m\times n}$, we use the notation $A[i]$ to refer to the ith row of A and $A[i, j]$ to refer to the jth element in the ith row.

5.1 The Construction

Fix $\delta, \varepsilon \in (0, \frac{1}{2}]$, and let

$$t = \lceil \varepsilon^{-1} + \log\lceil \log_2 1/\delta \rceil\rceil \tag{48}$$

and

$$m = \lceil t^2 (M(\chi) + \log_2 |E|)\rceil. \tag{49}$$

(See Lemma 5.15 for the relation between ε and δ, and the parameters of Theorem 3.1.) For $i \in \mathbb{Z}$, we first define the map $\Delta_i : V \to \mathbb{R}^{m\times t}$, and then we use it to construct our final embedding.

For a vertex $v \in V$ and $c = \chi(v, p(v))$, let $\alpha = \sum_{c' \in \chi(E(P_v))} t^2 \tau_i(v_{c'})$, and

$$\beta = \alpha + \min\left(t^2 \tau_i(v), \left\lfloor \frac{d_T(v_c, v) - \sum_{\ell=-\infty}^{i-1} 2^\ell \tau_\ell(v)}{2^i/t^2} \right\rfloor\right).$$

Note that $\beta \le m$ since

$$\tau_i(v) + \sum_{c' \in \chi(E(P_v))} \tau_i(v_c') \le \varphi(c) \le M(\chi) + \log_2 |E|.$$

For $j \in [m]$, we define

$$\Delta_i(v)[j] = \begin{cases} \left(\frac{2^i}{t^2}, 0, 0 \ldots, 0\right) & \text{if } \alpha < j \le \beta, \\ \left(d_T(v_c, v) - \left(\left(\sum_{\ell=-\infty}^{i-1} 2^\ell \tau_\ell(v)\right) + (\beta - \alpha)\frac{2^i}{t^2}\right), 0, 0 \ldots, 0\right) \\ \quad \text{if } j = \beta + 1 \text{ and } \beta - \alpha < t^2 \tau_i(v), \\ (0, 0 \ldots, 0) & \text{otherwise.} \end{cases} \tag{50}$$

Observe that the scale selector τ_i chooses the scales in this definition, and for $v \in V$ and $i \in \mathbb{Z}$, $\Delta_i(v) = 0$ when $\tau_i(v) = 0$. Also note that the second case in the definition only occurs when $\tau_i(v)$ is specified by part (A) of (34), and in that case $\sum_{\ell \le i} 2^\ell \tau_\ell(v) > d(v, v_c)$.

Now, we present some key properties of the map $\Delta_i(v)$. The following two observations follow immediately from the definitions.

Observation 5.1 For $v \in V$ and $i \in \mathbb{Z}$, each row in $\Delta_i(v)$ has at most one non-zero coordinate.

Observation 5.2 For $v \in V$ and $i \in \mathbb{Z}$, let $\alpha = \sum_{c' \in \chi(E(P_v))} t^2 \tau_i(v_{c'})$. For $j \notin (\alpha, \alpha + t^2 \tau_i(v)]$, we have

$$\Delta_i(v)[j] = (0, \ldots, 0).$$

Proofs of the next four lemmas will be presented in Sect. 5.2.

Lemma 5.3 *For $v \in V$, there is at most one $i \in \mathbb{Z}$ and at most one couple $(j,k) \in [m] \times [t]$ such that $\Delta_i(v)[j,k] \notin \{0, \frac{2^i}{t^2}\}$.*

Lemma 5.4 *Let $c \in \chi(E)$, and $u, w \in V(\gamma_c) \backslash \{v_c\}$ be such that $d_T(w, v_c) \leq d_T(u, v_c)$. For all $i \in \mathbb{Z}$ and $(j,k) \in [m] \times [t]$, we have*

$$\Delta_i(w)[j,k] \leq \Delta_i(u)[j,k].$$

Lemma 5.5 *For $c \in \chi(E)$, and $u, w \in V(\gamma_c) \setminus \{v_c\}$, we have*

$$d_T(w,u) = \sum_{i \in \mathbb{Z}} \|\Delta_i(u) - \Delta_i(w)\|_1 \tag{51}$$

and

$$d_T(v_c, u) = \sum_{i \in \mathbb{Z}} \|\Delta_i(u)\|_1. \tag{52}$$

Lemma 5.6 *For $c \in \chi(E), u, w \in V(\gamma_c) \setminus \{v_c\}, i > j$ and $k \in [m]$, if both $\|\Delta_i(u)[k] - \Delta_i(w)[k]\|_1 \neq 0$, and $\|\Delta_j(u)[k] - \Delta_j(w)[k]\|_1 \neq 0$, then $d_T(u,w) \geq 2^{j-1}$.*

Re-randomization. For $t \in \mathbb{N}$, let $\pi_t : \mathbb{R}^t \to \mathbb{R}^t$ be a random mapping obtained by uniformly permuting the coordinates in $\mathbb{R}^t$. Let $\{\sigma_i\}_{i \in [m]}$ be a sequence of i.i.d. random variables with the same distribution as π_t. We define the random variable $\pi_{t,m} : \mathbb{R}^{m \times t} \to \mathbb{R}^{m \times t}$ as follows:

$$\pi_{t,m} \begin{pmatrix} r_1 \\ \vdots \\ r_m \end{pmatrix} = \begin{pmatrix} \sigma_1(r_1) \\ \vdots \\ \sigma_m(r_m) \end{pmatrix}.$$

The Construction. We now use re-randomization to construct our final embedding. For $c \in \chi(E)$, and $i \in \mathbb{Z}$, the map $f_{i,c} : V(T(c)) \to \mathbb{R}^{m \times t}$ will represent an embedding of the subtree $T(c)$ at scale $2^i/t^2$. Recall that

$$V(T(c)) = V(\gamma_c) \cup \Big(\bigcup_{c' \in \rho^{-1}(c)} V(T(c')) \setminus \{v_{c'}\} \Big).$$

Let $\{\Pi_{i,c'} : i \in \mathbb{Z}, c' \in \rho^{-1}(c)\}$ be a sequence of i.i.d. random variables which each have the distribution of $\pi_{t,m}$. We define $f_{i,c} : V(T(c)) \to \mathbb{R}^{m \times t}$ as follows:

$$f_{i,c}(x) = \begin{cases} 0 & \text{if } x = v_c, \\ \Delta_i(x) & \text{if } x \in V(\gamma_c) \setminus \{v_c\}, \\ \Delta_i(v_{c'}) + \Pi_{i,c'}(f_{i,c'}(x)) & \text{if } x \in V(T(c')) \setminus \{v_{c'}\} \text{ for some } c' \in \rho^{-1}(c). \end{cases} \tag{53}$$

Re-randomization permutes the elements within each row, and the permutations are independent for different subtrees, scales, and rows. Finally, we define $f_i = f_{i,c_0}$, where $c_0 = \chi(r, p(r))$. We use the following lemma to prove Theorem 3.1.

Lemma 5.7 *There exists a universal constant C such that the following holds with non-zero probability: For all $x, y \in V$,*

$$(1 - C\varepsilon)\, d_T(x, y) - \delta\, \rho_\chi(x, y; \delta) \le \sum_{i\in\mathbb{Z}} \|f_i(x) - f_i(y)\|_1 \le d_T(x, y). \quad (54)$$

We will prove Lemma 5.7 in Sect. 5.3. We first make two observations, and then use them to prove Theorem 3.1. Our first observation is immediate from Observations 5.1 and 5.2, since in the third case of (53), by Observation 5.2, $\Delta_i(v'_c)$ and $\Pi_{i,c'}(f_{i,c'}(x))$ must be supported on disjoint sets of rows.

Observation 5.8 For any $v \in V$ and for any row $j \in [m]$, there is at most one non-zero coordinate in $f_i(v)[j]$.

Observation 5.2 and Lemma 5.5 also imply the following.

Observation 5.9 For any $v \in V$ and $u \in P_v$, we have

$$d_T(u, v) = \sum_{i\in\mathbb{Z}} \|f_i(u) - f_i(v)\|_1.$$

Using these, together with Corollary 3.5, we now prove Theorem 3.1.

Proof of Theorem 3.1 By Lemma 5.7, there exists a choice of mappings $\{g_i\}_{i\in\mathbb{Z}}$ such that for all $x, y \in V$,

$$d_T(x, y) \ge \sum_{i\in\mathbb{Z}} \|g_i(x) - g_i(y)\| \ge (1 - O(\varepsilon)) d_T(x, y) - \delta\rho_\chi(x, y; \delta).$$

We will apply Corollary 3.5 to the family given by $\big\{f_i = \frac{t^2 g_i}{2^i}\big\}_{i\in\mathbb{Z}}$ to arrive at an embedding $F : V \to \ell_1^{tm(2+\lceil \log \frac{1}{\varepsilon}\rceil)}$ such that $G = F/t^2$ satisfies

$$d_T(x, y) \ge \|G(x) - G(y)\|_1 \ge (1 - O(\varepsilon)) d_T(x, y) - \delta\rho_\chi(x, y; \delta). \quad (55)$$

Observe that the codomain of f_i is $\mathbb{R}^{m\times t}$, where $mt = \Theta\big((\frac{1}{\varepsilon} + \log\log(\frac{1}{\delta}))^3 \log n\big)$, and the codomain of G is $\mathbb{R}^d$, where $d = \Theta\big(\log \frac{1}{\varepsilon}(\frac{1}{\varepsilon} + \log\log(\frac{1}{\delta}))^3 \log n\big)$.

To achieve (55), we need only show that for every $x, y \in V$, we have $\frac{\zeta(x,y)}{t^2} \lesssim \varepsilon d_T(x, y)$, where $\zeta(x, y)$ is defined in (30). Recalling this definition, we now restate ζ in terms of our explicit family $\big\{f_i = \frac{t^2 g_i}{2^i}\big\}_{i\in\mathbb{Z}}$. We have

$$\frac{\zeta(x, y)}{t^2} = \sum_{(k_1,k_2)\in[m]\times[t]} \sum_{\substack{i:\exists j<i\\ g_j(x)[k_1,k_2]\ne g_j(y)[k_1,k_2]}} h_i(x, y; k_1, k_2)\,, \quad (56)$$

where

$$h_i(x, y; k_1, k_2) = \frac{2^i}{t^2}\left(\frac{t^2}{2^i}\left|g_i(x)[k_1, k_2] - g_i(y)[k_1, k_2]\right| - \left\lfloor\left|\frac{t^2}{2^i}g_i(x)[k_1, k_2] - \frac{t^2}{2^i}g_i(y)[k_1, k_2]\right|\right\rfloor\right).$$

Fix $x, y \in V$. For $c \in \chi(E(P_{xy}))$, let λ_c be the induced subgraph on $V(P_{xy}) \cap V(\gamma_c)$, i.e. the subpath of P_{xy} where all edges are colored by color c. We have

$$d_T(x, y) = \sum_{c \in \chi(E(P_{xy}))} \mathsf{len}(E(\lambda_c)). \tag{57}$$

If we look at a single term in (56), we have

$$h_i(x, y; k_1, k_2) < \frac{2^i}{t^2}. \tag{58}$$

For $u, v \in P_{xy}$, let

$$S_i(u, v) = \big\{(k_1, k_2) \in [m] \times [t] : h_i(u, v; k_1, k_2) \neq 0 \text{ and } \exists j < i : g_j(x)[k_1, k_2] \neq g_j(y)[k_1, k_2]\big\}.$$

Now, notice that if $\frac{t^2}{2^i}(g_i(x)[k_1, k_2] - g_i(y)[k_1, k_2])$ is fractional, then there must exist a subpath λ_c, for a color $c \in \chi(E(P_{xy}))$, with endpoints u_c and w_c such that $\frac{t^2}{2^i}(g_i(u_c)[k_1, k_2] - g_i(w_c)[k_1, k_2])$ is fractional too. Hence we have

$$\zeta(x, y) < \sum_{c \in \chi(E(P_{xy}))} \sum_{i \in \mathbb{Z}} \frac{2^i |S_i(u_c, w_c)|}{t^2}.$$

We call $\sum_{i \in \mathbb{Z}} \frac{2^i |S_i(u_c, w_c)|}{t^2}$ the contribution of λ_c for each color $c \in \chi(E(P_{xy}))$.

We divide the analysis of the paths λ_c for $c \in \chi(E(P_{xy}))$ into two cases. For $c \in \chi(E(P_x)) \triangle \chi(E(P_y))$, the vertex v_c is one endpoint of the path λ_c. Let u_c be the other. By Lemma 5.3, there is at most one $i \in \mathbb{Z}$ and $(k_1, k_2) \in [m] \times [t]$ such that $h_i(u_c, v_c; k_1, k_2) \neq 0$, and

$$\Big|\bigcup_{i \in \mathbb{Z}} S_i(u_c, v_c)\Big| \leq 1.$$

By Lemma 4.3, for all $i \in \mathbb{Z}$ with $d_T(u_c, v_c) \leq 2^{i-1}$, we have $\tau_i(u_c) = 0$ and

$$\|\Delta_i(u_c)\|_1 = \|g_i(u_c) - g_i(v_c)\|_1 = 0. \tag{59}$$

For $i < 1 + \log_2(d_T(u_c, v_c))$, by (58) and Lemma 5.3 we can bound the contribution of λ_c to $\zeta(x, y)$ by

$$\sum_{j\in\mathbb{Z}} \frac{2^j|S_j(u_c, v_c)|}{t^2} < \frac{2^i}{t^2} < \frac{2d_T(u_c, v_c)}{t^2} \leq \varepsilon d_T(u_c, v_c). \tag{60}$$

Note that there is at most one color in $\chi(E(P_{xy})) \setminus (\chi(E(P_x))\triangle\chi(E(P_y)))$. If no such color exists, then by (60),

$$\zeta(x, y) < \sum_{c\in\chi(E(P_{xy}))} \varepsilon\, \mathsf{len}(E(\lambda_c)) \overset{(57)}{\leq} \varepsilon d_T(x, y).$$

Suppose now that $\{c\} = \chi(E(P_{xy})) \setminus (\chi(E(P_x))\triangle\chi(E(P_y)))$. Let $u, w \in V(\lambda_c)$ be the closest vertices to x and y, respectively. For $i \in \mathbb{Z}$ we will show that if $h_i(u, w; k_1, k_2) \neq 0$, then either $d_T(x, y) \geq 2^{i-2}$, or for all $j < i$, we have $(g_j(x) - g_j(y))[k_1, k_2] = 0$. Then, by Lemma 5.3, there are at most two elements in $g_i(u) - g_i(w)$ that are not in $\{0, \frac{2^i}{t^2}, -\frac{2^i}{t^2}\}$, therefore we can conclude

$$\begin{aligned}\zeta(x, y) &< \sum_{i\in\mathbb{Z}} \frac{2^i|S_i(u, w)|}{t^2} + \sum_{c\in\chi(E(P_x))\triangle\chi(E(P_y))}\sum_{i\in\mathbb{Z}} \frac{2^i|S_i(u_c, v_c)|}{t^2}\\ &\overset{(57)}{\leq} 4\varepsilon d_T(x, y) + \sum_{c\in\chi(E(P_x))\triangle\chi(E(P_y))} \varepsilon\, \mathsf{len}(E(\lambda_c))\\ &\leq 5\varepsilon d_T(x, y).\end{aligned}$$

Without loss of generality suppose that $d_T(u, v_c) \leq d_T(w, v_c)$. If $d_T(w, v_c) = 0$ then the contribution of λ_c to $\zeta(x, y)$ is zero. Suppose now that $d_T(w, v_c) > 0$, and let $m_w = \max\{i : \tau_i(w) \neq 0\}$. By Lemma 4.3 the maximum always exists.

We will now split the rest of the proof into two cases.

Case 1: $\tau_{m_w-1}(u) = 0$.

In this case by Lemma 4.6 we have $d_T(u, w) > 2^{m_w-1}$. For $(k_1, k_2) \in [m] \times [t]$, if $h_i(u, w; k_1, k_2) \neq 0$ then by (50), $i \leq m_w$ and

$$\frac{2^i}{t^2} \leq \frac{2^{m_w}}{t^2} < \frac{2d_T(u, w)}{t^2} \leq \frac{2d_T(x, y)}{t^2} \leq \varepsilon d_T(x, y).$$

Case 2: $\tau_{m_w-1}(u) \neq 0$.

Let $m_u = \max\{i : \tau_i(u) \neq 0\}$. By Lemma 4.5 and as $\tau_{m_w-1}(u) \neq 0$, we have $m_u \leq m_w \leq m_u + 1$. Observation 4.2 implies that for all $j < m_u$,

$$\tau_j(u) + \sum_{c'\in\chi(E(P_u))} \tau_j(v_{c'}) = \varphi(c).$$

We have $m_w \geq m_u$, and by Observation 4.2,

$$\tau_j(w) + \sum_{c' \in \chi(E(P_w))} \tau_j(v_{c'}) = \tau_j(u) + \sum_{c' \in \chi(E(P_u))} \tau_j(v_{c'}) = \varphi(c). \tag{61}$$

Therefore, by Observation 5.2 for $j < m_u$ and $k \in [t^2\varphi(c)]$,

$$\|(g_j(x) - g_j(u))[k]\|_1 = \|(g_j(y) - g_j(w))[k]\|_1 = 0, \tag{62}$$

and by Observation 5.2 and part (B) of (34), for all $i \in \mathbb{Z}$, all the non-zero elements of $g_i(u) - g_i(w)$ are in the first $t^2\varphi(c)$ rows.

Suppose that there exists $k \in [m]$ such that $\|(g_i(u) - g_i(w))[k]\|_1 \neq 0$. Now, we divide the proof into two cases again.

Case 2.1: There exists a $j < i$ such that $\|(g_j(x) - g_j(u))[k]\|_1 + \|(g_j(y) - g_j(w))[k]\|_1 \neq 0$.

In this case, there must exist some $c' \in \chi(E(P_x)) \triangle \chi(E(P_y))$ such that

$$\|(g_j(v_{c'}) - g_j(u_{c'}))[k]\|_1 \neq 0.$$

By (53) and (50), we have $\tau_j(u_{c'}) \neq 0$. Inequality (62) implies $j \geq m_u$, and finally by Lemma 4.3,

$$d_T(x, y) \geq d_T(u_{c'}, v_{c'}) > 2^{j-1} \geq 2^{m_u-1} \geq 2^{m_w-2} \geq 2^{i-2}. \tag{63}$$

Case 2.2: $\|(g_j(x) - g_j(u))[k]\|_1 + \|(g_j(y) - g_j(w))[k]\|_1 = 0$ for all $j < i$.

In this case, either for all $j < i$, $\|g_j(x)[k] - g_j(y)[k]\|_1 = 0$ which implies that for $k' \in [t]$, $(k, k') \notin S_i(u, w)$, or $\|g_j(u)[k] - g_j(w)[k]\|_1 \neq 0$ for some $j < i$. If $\|g_j(u)[k] - g_j(w)[k]\|_1 \neq 0$ for some $j < i$ then by Lemma 5.6,

$$d_T(x, y) \geq d_T(u, w) \geq 2^{m_u-1} \geq 2^{m_w-2} \geq 2^{i-2}. \tag{64}$$

For $i > m_w$ we have $\|g_i(u) - g_i(w)\|_1 = 0$, therefore in both cases if $h_i(x, y; k_1, k_2) \neq 0$ either for all $j < i$, $\|g_j(x)[k] - g_j(y)[k]\|_1 = 0$ or

$$\frac{2^i}{t^2} \leq \frac{4d_T(x, y)}{t^2} \leq 2\varepsilon d_T(x, y).$$

□

5.2 Properties of the Δ_i Maps

We now present proofs of Lemmas 5.3–5.6.

Proof of Lemma 5.3 For a fixed $i \in \mathbb{Z}$, by (50) there is at most one element in $\Delta_i(v)$ that takes a value other than $\{0, \frac{2^i}{t^2}\}$.

We prove this lemma by showing that if for some $i \in \mathbb{Z}$, and $(j,k) \in [m] \times [t]$,

$$\Delta_i(v)[j,k] \notin \left\{0, \frac{2^i}{t^2}\right\},$$

then for all $i' > i$ and $(j',k') \in [m] \times [t]$, we have $\Delta_{i'}(v)[j',k'] = 0$. Let $c = \chi(v, p(v))$. Using (50), we can conclude that

$$t^2 \tau_i(v) > \left\lfloor \frac{d_T(v_c, v) - \sum_{\ell=-\infty}^{i-1} 2^\ell \tau_\ell(v)}{2^i/t^2} \right\rfloor.$$

Since the left-hand side is an integer,

$$t^2 \tau_i(v) \geq \frac{d_T(v_c, v) - \sum_{\ell=-\infty}^{i-1} 2^\ell \tau_\ell(v)}{2^i/t^2}$$

and

$$\begin{aligned}\sum_{\ell \leq i} 2^\ell \tau_\ell(v) &= 2^i \tau_i(v) + \sum_{\ell<i} 2^\ell \tau_\ell(v) \\ &\geq 2^i \left(\frac{d_T(v_c, v) - \sum_{\ell<i} 2^\ell \tau_\ell(v)}{2^i}\right) + \sum_{\ell<i} 2^\ell \tau_\ell(v) \geq d_T(v_c, v).\end{aligned}$$

By part (A) of (34), for $i' > i$ we have $\tau_{i'}(v) = 0$, thus $\|\Delta_{i'}(v)\|_1 = 0$ and the proof is complete. □

Proof of Lemma 5.4 For $i < \left\lfloor \log_2 \left(\frac{m(T)}{M(\chi) + \log_2 |E|}\right)\right\rfloor$ we have $\|\Delta_k(u)\| = \|\Delta_k(w)\|_1 = 0$.

Let ν be the minimum integer greater than $\left\lfloor \log_2 \left(\frac{m(T)}{M(\chi)+\log_2 |E|}\right)\right\rfloor - 1$ such that part (A) of (34) for $\tau_\nu(w)$ is less that or equal to part (B). This ν exists since, by (35), part (B) of (34) is always positive, while by Lemma 4.3, part (A) of (34) must be zero for some $\nu \in \mathbb{Z}$ large enough. First we analyze the case when $i < \nu$.

Observation 4.2 implies that part (B) of (34) specifies the value of $\tau_i(w)$. By Lemma 4.5 $\tau_i(u) \geq \tau_i(w)$, but the part (B) for $\tau_i(u)$ is the same as for $\tau_i(w)$, so we must have $\tau_i(u) = \tau_i(w)$, and the same reasoning holds for $\tau_\ell(w)$ for $\ell < i$. Using this and the fact that part (A) does not define $\tau_i(w)$, we have

$$2^i \tau_i(w) + \sum_{\ell<i} 2^\ell \tau_\ell(w) = 2^i \tau_i(u) + \sum_{\ell<i} 2^\ell \tau_\ell(u) < d_T(v_c, w) < d_T(v_c, u).$$

Therefore, the second case in (50) happens neither for u nor for w, and for $i < \nu$ we have $\Delta_i(u) = \Delta_i(w)$.

We now consider the case $i = \nu$. We have already shown that for $\ell < i$, $\tau_\ell(u) = \tau_\ell(w)$, and using (50), it is easy to verify that for all $(j,k) \in [m] \times [t]$,

$$\Delta_i(u)[j,k] \geq \Delta_i(w)[j,k].$$

Finally, in the case that $i > \nu$, by Observation 4.2, we have $\tau_i(w) = 0$ and $\Delta_i(w)[j,k] = 0$. □

Proof of Lemma 5.5 For all $i \in \mathbb{Z}$, recalling the definition α and β in (50) for $\Delta_i(u)$, we have

$$\beta - \alpha = \min\left(t^2\tau_i(v), \left\lfloor \frac{d_T(v_c, v) - \sum_{\ell=-\infty}^{i-1} 2^\ell \tau_\ell(v)}{2^i/t^2} \right\rfloor\right).$$

and by definition of $\Delta_i(u)$ we have

$$\|\Delta_i(u)\|_1 = \min\left(2^i\tau_i(u), d_T(u, v_c) - \sum_{j<i} 2^j \tau_j(u)\right).$$

By Lemma 4.4, we have $\sum_{i\in\mathbb{Z}} 2^i\tau_i(u) \geq d_T(u, v_c)$, therefore $d_T(v_c, u) = \sum_{i\in\mathbb{Z}} \|\Delta_i(u)\|_1$. The same argument also implies that $d_T(w, v_c) = \sum_{i\in\mathbb{Z}} \|\Delta_i(w)\|_1$.

Now, suppose that $d_T(u, v_c) \geq d(w, v_c)$. Then Lemma 5.4 implies that

$$\|\Delta_i(u) - \Delta_i(w)\|_1 = \|\Delta_i(u)\|_1 - \|\Delta_i(w)\|_1 = d_T(v_c, u) - d_T(v_c, w) = d_T(w, u).$$

□

Proof of Lemma 5.6 Without loss of generality suppose that $d_T(v_c, u) \geq d_T(v_c, w)$. We have

$$\begin{aligned} d_T(u, w) &= \sum_{h\in\mathbb{Z}} \|\Delta_h(u) - \Delta_h(w)\|_1 \geq \sum_{h=j}^{i} \|\Delta_h(u) - \Delta_h(w)\|_1 \\ &\geq \|\Delta_i(u) - \Delta_i(w)\|_1 + \|\Delta_j(u) - \Delta_j(w)\|_1. \end{aligned} \tag{65}$$

By Lemma 4.5 we have $\tau_j(w) \leq \tau_j(u)$. In the definition of $\tau_j(w)$, if part (B) of (34) is less than part (A), then by (50), for all h such that

$$\sum_{c'\in\chi(E(P_v))} t^2\tau_j(v_{c'}) < h \leq t^2\varphi(c),$$

we have $\|\Delta_j(w)[h]\|_1 = \frac{2^i}{t^2}$. By Lemma 5.4 and Observation 5.2 for $k \in \mathbb{Z}$, $\Delta_j(w) = \Delta_j(u)$. Hence, part (A) of (34) must specify the value of $\tau_j(w)$. Observation 4.2 implies that $\tau_i(w) = 0$ and by (50), we have $\|\Delta_i(w)\|_1 = 0$.

By (50), since $\|\Delta_i(u)[k]\|_1 > 0$, and α from (50) is a multiple of t^2 for all $t^2\lfloor\frac{k}{t^2}\rfloor < h < k$ we have $\|\Delta_i(u)[h]\|_1 = \frac{2^i}{t^2}$. This implies that

$$\|\Delta_i(u) - \Delta_i(w)\|_1 \geq \frac{2^i}{t^2}\left(k - 1 - t^2\left\lfloor\frac{k}{t^2}\right\rfloor\right) \geq \frac{2^j}{t^2}\left(k - 1 - t^2\left\lfloor\frac{k}{t^2}\right\rfloor\right).$$

Moreover, $\|\Delta_j(w)[k]\|_1 < \frac{2^j}{t^2}$, and (50) implies that for all $k < h \le t^2\lfloor 1 + \frac{k}{t^2}\rfloor$, we have $\|\Delta_j(w)[h]\|_1 = 0$. The same argument also shows that

$$\|\Delta_j(u) - \Delta_j(w)\|_1 \ge \frac{2^j}{t^2}\big(t^2\big\lfloor 1 + \frac{k}{t^2}\big\rfloor - k\big).$$

Hence by (65),

$$d_T(u, w) \ge \frac{t^2 - 1}{t^2} 2^j \ge 2^{j-1}.$$

□

5.3 The Probabilistic Analysis

We are thus left to prove Lemma 5.7. For $c \in \chi(E)$, we analyze the embedding for $T(c)$ by going through all $c' \in \chi(E(T(c)))$ one by one in increasing order of $\varphi(c')$. Our first lemma bounds the probability of a bad event, i.e. of a subpath not contributing enough to the distance in the embedding.

Lemma 5.10 *For any $C \ge 8$, the following holds. Consider three colors $a \in \chi(E)$, $b \in \rho^{-1}(a)$, and $c \in \chi(E(P_{u\,v_b}))$ for some $u \in V(T(b))$. Then for every $w \in V(T(a)) \setminus V(T(b))$, we have*

$$\begin{aligned}\mathbb{P}\Big[\exists x \in V(P_{w\,v_a}) : \sum_{i\in\mathbb{Z}} \|f_{i,a}(x) - f_{i,a}(u)\|_1 &\le (1 - C\varepsilon)\, d_T(u, v_c) \\ + \sum_{i\in\mathbb{Z}} \|f_{i,a}(v_c) - f_{i,a}(x)\|_1 \mid \{f_{i,c'}\}_{c'\in\rho^{-1}(a)}\Big] \\ \le \frac{1}{\lceil \log_2 1/\delta\rceil} \exp\big(-(C/(\varepsilon 2^{\beta+2}))\, d_T(u, v_c)\big), \qquad (66)\end{aligned}$$

where $\beta = \max\{i : \exists y \in P_{u\,v_c}\backslash\{v_c\}, \tau_i(y) \ne 0\}$. (See Fig. 1 for position of vertices in the tree.)

Proof Recall that $\mathbb{R}^{m\times t}$ is the codomain of $f_{i,a}$. For $i \in \mathbb{Z}$, and $j \in [m]$, and $z \in V(P_{w\,v_a})$, let

$$\begin{aligned}s_{ij}(z) = \big\| f_{i,a}(z)[j] - f_{i,a}(v_c)[j]\big\|_1 + \big\| f_{i,a}(v_c)[j] \\ - f_{i,a}(u)[j]\big\|_1 - \big\| f_{i,a}(z)[j] - f_{i,a}(u)[j]\big\|_1 .\end{aligned}$$

We have

$$\begin{aligned}\sum_{i\in\mathbb{Z}} \|f_{i,a}(u) - f_{i,a}(v_c)\|_1 + \sum_{i\in\mathbb{Z}} \|f_{i,a}(v_c) - f_{i,a}(z)\|_1 \\ = \sum_{i\in\mathbb{Z}} \|f_{i,a}(z) - f_{i,a}(u)\|_1 + \sum_{i\in\mathbb{Z}, j\in[m]} s_{ij}(z).\end{aligned}$$

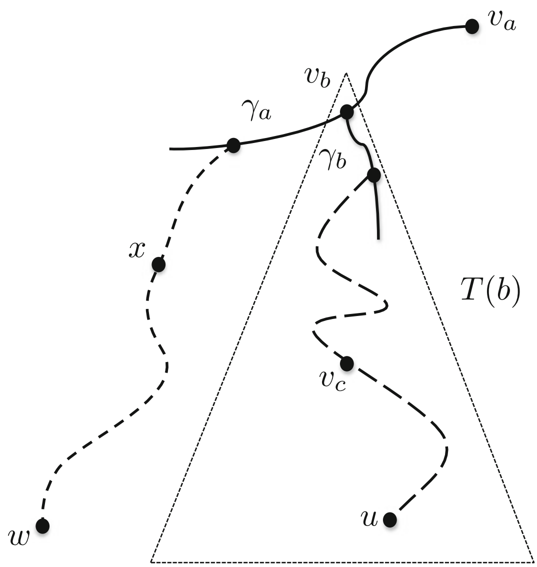

Fig. 1 Position of vertices corresponding to the statement of Lemma 5.10

By Observation 5.9, we have $d_T(u, v_c) = \sum_{i\in\mathbb{Z}} \|f_{i,a}(u) - f_{i,a}(v_c)\|_1$, therefore

$$d_T(u, v_c) - \sum_{i\in\mathbb{Z}, j\in[m]} s_{ij}(z) = \sum_{i\in\mathbb{Z}} \|f_{i,a}(z) - f_{i,a}(u)\|_1 - \sum_{i\in\mathbb{Z}} \|f_{i,a}(z) - f_{i,a}(v_c)\|_1. \tag{67}$$

Let $\mathcal{E} = \{f_{i,c'} : c' \in \rho^{-1}(a)\}$. We define $\mathbb{P}_{\mathcal{E}}[\cdot] = \mathbb{P}[\cdot \mid \mathcal{E}]$. In order to prove this theorem, we bound

$$\mathbb{P}_{\mathcal{E}}\Big[\exists x \in V(P_{w\,v_a}) : \sum_{i\in\mathbb{Z}, j\in[m]} s_{ij}(x) \geq C\varepsilon d_T(u, v_c)\Big].$$

We start by bounding the maximum of the random variables s_{ij}.

For $i > \beta$ we have $\Delta_i(u) = \Delta_i(v_c)$, hence $f_{i,a}(u) = f_{i,a}(v_c)$. Using the triangle inequality for all $i \in \mathbb{Z}$, $j \in [m]$ and $z \in P_{w\,v_a}$,

$$s_{ij}(z) \leq 2\|f_{i,a}(v_c)[j] - f_{i,a}(u)[j]\|_1, \tag{68}$$

Hence for all $i \in \mathbb{Z}$ and $j \in [m]$ by Observation 5.8,

$$s_{ij}(z) \leq 2\|f_{i,a}(v_c)[j] - f_{i,a}(u)[j]\|_1 \leq \frac{2^{\beta+1}}{t^2}. \tag{69}$$

First note that, if z is on the path between v_b and v_a then by Observation 5.9, $s_{ij}(z) = 0$. Observation 5.2 and (50) imply that if $\|f_{i,a}(u)[j] - f_{i,a}(v_c)[j]\|_1 \neq 0$ then $\|f_{i,a}(v_c)[j]\|_1 = 0$. From this, we can conclude that $s_{ij}(z) \neq 0$ if and only if there exists a $k \in [t]$ such that both $f_{i,a}(u)[j,k] - f_{i,a}(v_c)[j,k] \neq 0$ and $f_{i,a}(z)[j,k] \neq 0$. Since by Lemma 5.4, for all $i \in \mathbb{Z}$, $j \in [m]$ and $k \in [t]$, we have $f_{i,a}(w)[j,k] \geq f_{i,a}(z)[j,k]$, we conclude that for $z \in P_{w\,v_a}$ if $s_{ij}(z) \neq 0$ then $s_{ij}(w) \neq 0$.

Now, for $i \in \mathbb{Z}$ and $j \in [m]$, we define a random variable

$$X_{ij} = \begin{cases} 0 & \text{if } s_{ij}(w) = 0, \\ 2\|f_{i,a}(u)[j] - f_{i,a}(v_c)[j]\|_1 & \text{if } s_{ij}(w) \neq 0. \end{cases} \tag{70}$$

Note that since the re-randomization in (53) is performed independently on each row and at each scale, the random variables $\{X_{ij} : i \in \mathbb{Z}, j \in [m]\}$ are mutually independent. By (68), for all $z \in P_{w\,v_a}$, we have $s_{ij}(z) \leq X_{ij}$, and thus

$$\begin{aligned} &\mathbb{P}_{\mathcal{E}}\Big[\exists x \in V(P_{w\,v_a}) : \sum_{i\in\mathbb{Z}, j\in[m]} s_{ij}(x) \geq C\varepsilon d_T(u, v_c)\Big] \\ &\quad\leq \mathbb{P}_{\mathcal{E}}\Big[\sum_{i\in\mathbb{Z}, j\in[m]} X_{ij} \geq C\varepsilon d_T(u, v_c)\Big]. \end{aligned} \tag{71}$$

As before, for X_{ij} to be non-zero, it must be that $k \in [t]$ is such that $f_{i,a}(w)[j,k] \neq 0$ and $f_{i,a}(u)[j,k] - f_{i,a}(v_c)[j,k] \neq 0$. Since $w \notin V(T(b))$ with the re-randomization in (53) and Observation 5.8, this happens at most with probability $\frac{1}{t}$, hence for $j \in [m]$, and $i \in \mathbb{Z}$,

$$\begin{aligned} \mathbb{P}_{\mathcal{E}}[X_{ij} \neq 0] = \mathbb{P}_{\mathcal{E}}\big[&\|f_{i,a}(w)[j] - f_{i,a}(v_c)[j]\|_1 + \|f_{i,a}(v_c)[j] \\ &- f_{i,a}(u)[j]\|_1 - \|f_{i,a}(w)[j] - f_{i,a}(u)[j]\|_1 \neq 0\big] \leq \frac{1}{t}. \end{aligned}$$

This yields

$$\mathbb{E}[X_{ij} \mid \mathcal{E}] \leq \frac{1}{t}\big(2\|f_{i,a}(u)[j] - f_{i,a}(v_c)[j]\|_1\big). \tag{72}$$

Now we use (69) to write

$$\mathrm{Var}(X_{ij} \mid \mathcal{E}) \leq \frac{1}{t}\big(2\|f_{i,a}(u)[j] - f_{i,a}(v_c)[j]\|_1\big)^2 \leq \frac{2^{\beta+2}}{t^3}\|f_{i,a}(u)[j] - f_{i,a}(v_c)[j]\|_1,$$

and use Observation 5.9 in conjunction with (72) to conclude that

$$\mathbb{E}\Big[\sum_{i\in\mathbb{Z}, j\in[m]} X_{ij} \mid \mathcal{E}\Big] \leq \sum_{i\in\mathbb{Z}, j\in[m]} \frac{2}{t}\|f_i(v_c)[j] - f_i(u)[j]\|_1 = \frac{2}{t} d_T(v_c, u) \tag{73}$$

and

$$\sum_{i\in\mathbb{Z},\,j\in[m]} \mathrm{Var}(X_{ij} \mid \mathcal{E}) \le \sum_{i\in\mathbb{Z},\,j\in[m]} \frac{2^{\beta+2}}{t^3} \| f_i(v_c)[j] - f_i(u)[j]\|_1 = \frac{2^{\beta+2}}{t^3} d_T(v_c, u). \tag{74}$$

Define $M = \max\{X_{ij} - \mathbb{E}[X_{ij} \mid \mathcal{E}] : i \in \mathbb{Z},\, j \in [m]\}$. We now apply Theorem 2.2 to complete the proof:

$$\begin{aligned}
&\mathbb{P}_{\mathcal{E}}\Big[\sum_{i\in\mathbb{Z},\,j\in[m]} X_{ij} \ge C\big(\frac{d_T(u,v_c)}{t}\big)\Big] \\
&= \mathbb{P}_{\mathcal{E}}\Big[\sum_{i\in\mathbb{Z},\,j\in[m]} X_{ij} - \frac{2d_T(u,v_c)}{t} \ge (C-2)\big(\frac{d_T(u,v_c)}{t}\big)\Big] \\
&\overset{(73)}{\le} \mathbb{P}_{\mathcal{E}}\Big[\sum_{i\in\mathbb{Z},\,j\in[m]} X_{ij} - \mathbb{E}\Big[\sum_{i\in\mathbb{Z},\,j\in[m]} X_{ij} \mid \mathcal{E}\Big] \ge (C-2)\Big(\frac{d_T(u,v_c)}{t}\Big)\Big] \\
&\le \exp\Big(\frac{-((C-2)d_T(u,v_c)/t)^2}{2\big(\sum_{i\in\mathbb{Z},\,j\in[m]} \mathrm{Var}(X_{ij} \mid \mathcal{E}) + (C-2)(d_T(u,v_c)/t)M/3\big)}\Big).
\end{aligned}$$

Since $\mathbb{E}[X_{ij} \mid \mathcal{E}] \ge 0$, (69) implies $M \le \frac{2^{\beta+1}}{t^2}$. Now, we can plug in this bound and (74) to write

$$\begin{aligned}
&\mathbb{P}_{\mathcal{E}}\Big[\sum_{i\in\mathbb{Z},\,j\in[m]} X_{ij} \ge C\Big(\frac{d_T(u,v_c)}{t}\Big)\Big] \\
&\quad\le \exp\Big(\frac{-((C-2)d_T(u,v_c)/t)^2}{2\big(\frac{2^{\beta+2}}{t^3} d_T(u,v_c) + (C-2)(d_T(u,v_c)/t)(2^{\beta+1}/t^2)/3\big)}\Big) \\
&\quad= \exp\Big(\frac{-t(C-2)^2 d_T(u,v_c)}{2\big(2^{\beta+2} + (C-2)(2^{\beta+1})/3\big)}\Big) \\
&\quad= \exp\Big(\frac{-(C-2)^2}{(C-2)/3+2}\Big(\frac{t d_T(u,v_c)}{2^{\beta+2}}\Big)\Big).
\end{aligned}$$

An elementary calculation shows that for $C \ge 8$, $\frac{(C-2)^2}{(C-2)/3+2} > C$, hence

$$\begin{aligned}
&\mathbb{P}_{\mathcal{E}}\Big[\sum_{i\in\mathbb{Z},\,j\in[m]} X_{ij} \ge C\Big(\frac{d_T(u,v_c)}{t}\Big)\Big] \\
&\quad< \exp\Big(-(Ct/2^{\beta+2})\, d_T(u,v_c)\Big)
\end{aligned}$$

$$\overset{(48)}{\le} \exp\Big(-C\Big(\frac{1}{\varepsilon}+\log\Big\lceil \log_2\frac{1}{\delta}\Big\rceil\Big)\Big(\frac{1}{2^{\beta+2}}\Big)\,d_T(u,v_c)\Big)$$
$$= \Big(\frac{1}{\lceil \log_2(1/\delta)\rceil}\Big)^{\frac{Cd_T(u,v_c)}{2^{\beta+2}}}\cdot\exp\Big(-C\Big(\frac{1}{\varepsilon}\Big)\Big(\frac{1}{2^{\beta+2}}\Big)\,d_T(u,v_c)\Big).$$

Since there exists a $y\in P_{u\,v_c}\backslash\{v_c\}$ such that $\tau_\beta(y)\neq 0$, and for all $c'\in\chi(E)$, $\kappa(c')\geq 1$, Lemma 4.3 implies that $d_T(u,v_c) > 2^{\beta-1}$, and for $C\geq 8$, we have $\frac{Cd_T(u,v_c)}{2^{\beta+2}} > 1$. Therefore,

$$\begin{aligned}
&\mathbb{P}_{\mathcal{E}}\Big[\exists x\in V(P_{w\,v_a}) : \sum_{i\in\mathbb{Z}}\|f_{i,a}(x)-f_{i,a}(u)\|_1\\
&\quad\le (1-C\varepsilon)\,d_T(u,v_c)+\sum_{i\in\mathbb{Z}}\|f_{i,a}(v_c)-f_{i,a}(x)\|_1\Big]\\
&\quad\overset{(67)}{\le}\mathbb{P}_{\mathcal{E}}\Big[\exists x\in V(P_{w\,v_c}) : \sum_{i\in\mathbb{Z},\,j\in[m]} s_{ij}(x)\geq C\varepsilon d_T(u,v_c)\Big]\\
&\quad\overset{(71)}{\le}\mathbb{P}_{\mathcal{E}}\Big[\sum_{i\in\mathbb{Z},\,j\in[m]} X_{ij}\geq C\varepsilon\,(d_T(u,v_c))\Big]\\
&\quad\overset{(48)}{\le}\mathbb{P}_{\mathcal{E}}\Big[\sum_{i\in\mathbb{Z},\,j\in[m]} X_{ij}\geq C\Big(\frac{d_T(u,v_c)}{t}\Big)\Big]\\
&\quad< \Big(\frac{1}{\lceil\log_2(1/\delta)\rceil}\Big)\cdot\exp\Big(-C\Big(\frac{1}{\varepsilon 2^{\beta+2}}\Big)\,d_T(u,v_c)\Big),
\end{aligned}$$

completing the proof. □

The Γ_a Mappings. Before proving Lemma 5.7, we need some more definitions. For a color $a\in\chi(E)$, we define a map $\Gamma_a : V(T(a))\to V(T(a))$ based on Lemma 5.10. For $u\in V(\gamma_a)$, we put $\Gamma_a(u)=u$. For all other vertices $u\in V(T(a))\setminus V(\gamma_a)$, there exists a unique color $b\in\rho^{-1}(a)$ such that $u\in V(T(b))$. We define $\Gamma_a(u)$ as the vertex $w\in V(P_{uv_b})$ which is closest to the root among those vertices satisfying the following condition: For all $v\in V(P_{uw})\setminus\{w\}$ and $k\in\mathbb{Z}$, $\tau_k(v)\neq 0$ implies

$$2^k < \frac{d_T(u,w)}{\varepsilon(\varphi(\chi(u,p(u)))-\varphi(a))}. \tag{75}$$

Clearly such a vertex exists, because the conditions are vacuously satisfied for $w=u$. We now prove some properties of the map Γ_a.

Lemma 5.11 *Consider any $a\in\chi(E)$ and $u\in V(T(a))$ such that $\Gamma_a(u)\neq u$. Then we have $\Gamma_a(u)=v_c$ for some $c\in\chi(E(P_{uv_a}))\setminus\{a\}$.*

Proof Let $w\in V(P_{u\,\Gamma_a(u)})$ be such that $\Gamma_a(u)=p(w)$. The vertex w always exists because $\Gamma_a(u)\in V(P_u)\setminus\{u\}$. If $\chi(w,\Gamma_a(u))\neq\chi(\Gamma_a(u),p(\Gamma_a(u)))$ then $\Gamma_a(u)$ is v_c for some $c\in\chi(E(P_{u\,v_a}))\setminus\{a\}$.

Now, for the sake of contradiction suppose that $\chi(w, \Gamma_a(u)) = \chi(\Gamma_a(u), p(\Gamma_a(u)))$. In this case, we show that for all $v \in P_{u\, p(\Gamma_a(u))} \setminus \{p(\Gamma_a(u))\}$, and $k \in \mathbb{Z}$, $\tau_k(v) \neq 0$ implies

$$2^k < \frac{d_T(u, p(\Gamma_a(u)))}{\varepsilon(\varphi(\chi(u, p(u))) - \varphi(a))}. \tag{76}$$

This is a contradiction since by definition of Γ_a, it must be that $\Gamma_a(u)$ is the closest vertex to the root satisfying this condition, yet $p(\Gamma_a(u))$ is closer to root than $\Gamma_a(u)$.

Observe that

$$V(P_{u\, p(\Gamma_a(u))}) \setminus \{p(\Gamma_a(u))\} = V(P_{u\, \Gamma_a(u)}).$$

We first verify (76) for $\Gamma_a(u)$ and $k \in \mathbb{Z}$ with $\tau_k(\Gamma_a(u)) \neq 0$. Since $\Gamma_a(u) \in V(P_u)$, we have

$$d_T(u, \Gamma_a(u)) \le d_T(u, p(\Gamma_a(u))). \tag{77}$$

Recalling that $p(w) = \Gamma_a(u)$, by Lemma 4.5 for all $k \in \mathbb{Z}$, $\tau_k(\Gamma_a(u)) \le \tau_k(w)$, therefore for all $k \in \mathbb{Z}$, with $\tau_k(\Gamma_a(u)) \neq 0$, we have $\tau_k(w) \neq 0$ as well, hence (75) implies

$$2^k < \frac{d_T(u, \Gamma_a(u))}{\varepsilon(\varphi(\chi(u, p(u))) - \varphi(a))} \overset{(77)}{\le} \frac{d_T(u, p(\Gamma_a(u)))}{\varepsilon(\varphi(\chi(u, p(u)) - \varphi(a))}. \tag{78}$$

For all other vertices, $v \in V(P_{u\Gamma_a(u)}) \setminus \{\Gamma_a(u)\}$, and $k \in \mathbb{Z}$ with $\tau_k(v) \neq 0$ by (75),

$$2^k < \frac{d_T(u, \Gamma_a(u))}{\varepsilon(\varphi(\chi(u, p(u))) - \varphi(a))} \overset{(77)}{\le} \frac{d_T(u, p(\Gamma_a(u)))}{\varepsilon(\varphi(\chi(u, p(u))) - \varphi(a))}, \tag{79}$$

completing the proof. □

Lemma 5.12 *Suppose that $a \in \chi(E)$ and $u \in V(T(a))$. For any $w \in V(P_{u\, \Gamma_a(u)})$ such that $\chi(u, p(u)) = \chi(w, p(w))$ we have $\Gamma_a(w) \in V(P_{u\, \Gamma_a(u)})$.*

Proof For the sake of contradiction, suppose that $\Gamma_a(w) \notin V(P_{u\, \Gamma_a(u)})$. Since $w \in V(P_u)$ and $\Gamma_a(w) \notin V(P_{u\, \Gamma_a(u)})$, we have $\Gamma_a(w) \in V(P_{\Gamma_a(u)})$ and

$$d_T(u, \Gamma_a(u)) \le d_T(u, \Gamma_a(w)). \tag{80}$$

Since $w \in V(P_{u\, \Gamma_a(u)})$ by assumption, for all vertices, we have $V(P_{u\, w}) \setminus \{w\} \subseteq V(P_{u\, \Gamma_a(u)}) \setminus \{\Gamma_a(u)\}$. Thus for all $v \in V(P_{u\, w}) \setminus \{w\}$ and $k \in \mathbb{Z}$ with $\tau_k(v) \neq 0$ by (75),

$$2^k < \frac{d_T(u, \Gamma_a(u))}{\varepsilon(\varphi(\chi(u, p(u))) - \varphi(a))} \overset{(80)}{\le} \frac{d_T(u, \Gamma_a(w))}{\varepsilon(\varphi(\chi(u, p(u))) - \varphi(a))}. \tag{81}$$

The fact that $w \in V(P_{u\,\Gamma_a(u)})$ also implies that $d_T(w, \Gamma_a(w))) \leq d_T(u\,\Gamma_a(w)))$. Therefore, for all vertices $v \in V(P_{w\,\Gamma_a(w)}) \setminus \{\Gamma_a(w)\}$ and $k \in \mathbb{Z}$ with $\tau_k(v) \neq 0$ by (75),

$$\begin{aligned} 2^k &< \frac{d_T(w, \Gamma_a(w))}{\varepsilon(\varphi(\chi(w, p(w))) - \varphi(a))} \leq \frac{d_T(u, \Gamma_a(w))}{\varepsilon(\varphi(\chi(w, p(w))) - \varphi(a))} \\ &= \frac{d_T(u, \Gamma_a(w))}{\varepsilon(\varphi(\chi(u, p(u))) - \varphi(a))}. \end{aligned} \tag{82}$$

We have

$$V(P_{u\,\Gamma_a(w)}) = V(P_{u\,w}) \cup \big(V(P_{w\,\Gamma_a(w)}) \setminus \{\Gamma_a(w)\}\big).$$

Hence, by (81) and (82), for all $v \in V(P_{u\,\Gamma_a(w)}) \setminus \{\Gamma_a(w)\}$ and $k \in \mathbb{Z}$, $\tau_k(v) \neq 0$ implies

$$2^k < \frac{d_T(u, p(\Gamma_a(w)))}{\varepsilon(\varphi(\chi(u, p(u))) - \varphi(a))}. \tag{83}$$

This is a contradiction to the definition of $\Gamma_a(u)$, since $\Gamma_a(u)$ must be the closest vertex to the root satisfying this condition, yet $\Gamma_a(w)$ is closer to root than $\Gamma_a(u)$. $\square$

Defining Representatives for γ_c. Now, for each $c \in \chi(E)$, we define a small set of representatives for vertices in γ_c. Later, we use these sets to bound the contraction of pairs of vertices that have one endpoint in γ_c.

For $a \in \chi(E)$ and $c \in \chi(E(T(a))) \setminus \{a\}$, we define the set $R_a(c) \subseteq V(\gamma_c)$, the set of representatives for γ_c, as follows:

$$\begin{aligned} R_a(c) = \bigcup_{i=0}^{\lceil \log_2 \frac{1}{\delta} \rceil - 1} \big\{ &u \in V(\gamma_c) : u \text{ is the furthest vertex} \\ &\text{from } v_c \text{ s.t. } \Gamma_a(u) \neq u \text{ and } d(u, v_c) \leq 2^{-i}\,\mathsf{len}(\gamma_c) \big\}. \end{aligned} \tag{84}$$

The next lemma shows when a vertex has a close representative.

Lemma 5.13 *Consider $a \in \chi(E)$ and $c \in \chi(E(T(a))) \setminus \{a\}$. For all vertices $u \in V(\gamma_c)$ with $\Gamma_a(u) \neq u$ there exists a $w \in R_a(c)$ such that*

$$d_T(u, v_c) \leq d_T(w, v_c) \leq 2 \max\big(d_T(u, v_c), \delta\,\mathsf{len}(\gamma_c)\big).$$

Proof Let $i \geq 0$ be such that

$$\frac{d_T(u, v_c)}{\mathsf{len}(\gamma_c)} \in \big(2^{-i-1}, 2^{-i}\big].$$

If $i \le \lceil \log_2 \frac{1}{\delta} \rceil - 1$, then (84) implies that either $u \in R_a(c)$, or there exists a $w \in R_a(c)$ such that

$$d_T(u, v_c) < d_T(w, v_c) \le \frac{\mathsf{len}(\gamma_c)}{2^i} \le 2\, d_T(u, v_c).$$

On the other hand, if $i > \lceil \log_2 \frac{1}{\delta} \rceil - 1$, then (84) implies that either $u \in R_a(c)$, or that there exists a $w \in R_a(c)$ such that

$$d_T(u, v_c) < d_T(w, v_c) \le \frac{\mathsf{len}(\gamma_c)}{2^{\lceil \log_2 \frac{1}{\delta} \rceil - 1}} \le 2\delta\, \mathsf{len}(\gamma_c),$$

completing the proof. □

The following lemma, in conjunction with Lemma 5.13, reduces the number of vertices in $V(\gamma_c)$ that we need to analyze using Lemma 5.10.

Lemma 5.14 *Let (X, d) be a pseudometric, and let $f : V \to X$ be a 1-Lipschitz map. For $x, y \in V$, and $x', y' \in V(P_{xy})$ and $h \ge 0$, if $d(f(x), f(y)) \ge d_T(x, y) - h$ then $d(f(x'), f(y')) \ge d_T(x', y') - h$.*

Proof Suppose without loss of generality that $d_T(x', x) \le d_T(y', x)$. Using the triangle inequality,

$$\begin{aligned} d(f(x'), f(y')) &\ge d(f(x), f(y)) - d(f(x), f(x')) - d(f(y), f(y')) \\ &\ge (d_T(x, y) - h) - d(f(x), f(x')) - d(f(y), f(y')) \\ &\ge d_T(x, y) - d_T(x, x') - d_T(y, y') - h \\ &= d_T(x', y') - h. \end{aligned}$$

□

The following lemma constitutes the inductive step of the proof of Lemma 5.7.

Lemma 5.15 *There exists a universal constant C such that for any color $c \in \chi(E) \cup \{\chi(r, p(r))\}$, the following holds. Suppose that, with non-zero probability, for all $c' \in \rho^{-1}(c)$, and for all pairs $x, y \in V(T(c'))$, we have*

$$(1 - C\varepsilon)\, d_T(x, y) - \delta\, \rho_\chi(x, y; \delta) \le \sum_{i \in \mathbb{Z}} \|f_{i,c'}(x) - f_{i,c'}(y)\|_1 \le d_T(x, y). \tag{85}$$

Then with non-zero probability for all $x, y \in V(T(c))$, we have

$$(1 - C\varepsilon)\, d_T(x, y) - \delta\, \rho_\chi(x, y; \delta) \le \sum_{i \in \mathbb{Z}} \|f_{i,c}(x) - f_{i,c}(y)\|_1 \le d_T(x, y). \tag{86}$$

Proof Let $\mathcal{E}$ denote the event that, for all $c' \in \rho^{-1}(c)$, and all $x, y \in V(T(c'))$, we have

$$d_T(x, y) \ge \sum_{i \in \mathbb{Z}} \|f_{i,c'}(x) - f_{i,c'}(y)\| \ge (1 - C\varepsilon) d_T(x, y) - \delta \rho_\chi(x, y; \delta). \tag{87}$$

We will prove the lemma by showing that, conditioned on $\mathcal{E}$, (86) holds with non-zero probability.

For $x, y \in V(T(c))$ we define

$$\mu(x, y) = \max\{\varphi(a) : a \in \chi(E) \text{ and } x, y \in V(T(a))\}.$$

Note that since $x, y \in V(T(c))$, we have

$$\mu(x, y) \geq \varphi(c). \tag{88}$$

It is easy to see that if $\mu(x, y) > \varphi(c)$, then $x, y \in V(T(c'))$ for some $c' \in \rho^{-1}(c)$. By construction, if $c' \in \rho^{-1}(c)$ and $x, y \in V(T(c'))$, then

$$\|f_{i,c}(x) - f_{i,c}(y)\| = \|f_{i,c'}(x) - f_{i,c'}(y)\|,$$

hence $\mathcal{E}$ implies that (86) holds for all such pairs. Thus in the remainder of the proof, we need only handle pairs $x, y \in V(T(c))$ with $\mu(x, y) = \varphi(c)$.

Write $\chi(E(T(c))) = \{c_1, c_2, \ldots, c_n\}$, where the colors are ordered so that $\varphi(c_j) \leq \varphi(c_{j+1})$ for $j = 1, 2, \ldots, n-1$. Let $\varepsilon_1 = 24\varepsilon$, where the constant 24 comes from Lemma 5.10. And let $\varepsilon_2 = 2 \cdot C'\varepsilon$, where C' is the constant from Lemma 4.11.

For $i \in [m]$, we define the event X_i as follows: For all $j \leq i$, all $x \in V(\gamma_{c_i})$ and $y \in V(\gamma_{c_j})$ with $\mu(x, y) = \varphi(c)$, we have

$$\begin{aligned}\sum_{k\in\mathbb{Z}} &\|f_{k,c}(x) - f_{k,c}(y)\|_1 \\ &\geq d_T(x, y) - \varepsilon_1 d_T(x, y) - \varepsilon_2 d_T(\Gamma_c(x), \Gamma_c(y)) - \delta\rho_\chi(x, y; \delta).\end{aligned} \tag{89}$$

For all pairs $x \in V(\gamma_{c_i})$ and $y \in V(\gamma_{c_j})$, the event $X_{\max(i,j)}$ implies

$$\sum_{k\in\mathbb{Z}} \|f_{k,c}(x) - f_{k,c}(y)\|_1 \geq d_T(x, y) - (\varepsilon_1 + \varepsilon_2) d_T(x, y) - \delta\rho_\chi(x, y; \delta).$$

In particular this shows that for $C = 2 \cdot C' + 24$, if the events $X_1, X_2, \ldots, X_n$ all occur, then (86) holds for all pairs $x, y \in V(T(c))$. Hence we are left to show that

$$\mathbb{P}[X_1 \wedge \cdots \wedge X_n \mid \mathcal{E}] > 0.$$

To this end, we define new events $\{Y_i : i \in [n]\}$ and we show that for every $i \in [n]$,

$$\mathbb{P}_{\mathcal{E}}\big[X_1 \wedge \cdots \wedge X_i \mid X_1 \wedge \cdots \wedge X_{i-1} \wedge Y_i\big] = 1\,, \tag{90}$$

and then we bound the probability that Y_i does not occur by

$$\mathbb{P}_{\mathcal{E}}\big[\overline{Y_i}\big] \leq 2^{-3(\varphi(c_i)-\varphi(c))+1}. \tag{91}$$

By Lemma 5.5 and the definition of $f_{k,c}$ (53), we have $\mathbb{P}_{\mathcal{E}}[X_1] = 1$. Since for all $i \in \{2, \ldots n\}$, $c_i \in \chi(E(T(c))) \setminus \{c\}$, we have

$$\mathbb{P}_{\mathcal{E}}[X_1 \wedge \cdots \wedge X_n]$$
$$\geq 1 - \sum_{i=2}^{n} \mathbb{P}_{\mathcal{E}}\big[\overline{Y_i}\big] \overset{(91)}{\geq} 1 - \sum_{i=2}^{n} 2^{-3(\varphi(c_i)-\varphi(c))+1} \overset{(4.9)}{>} 1 - 2 \cdot 2^{(2-3)} = 0,$$

which completes the proof.

For each $i \in [n]$, we define the event Y_i as follows: For all $j < i$, and all vertices $x \in R_c(c_i)$ and $y \in V(\gamma_{c_j})$ with $\mu(x, y) = \varphi(c)$, we have

$$\sum_{k\in\mathbb{Z}} \|f_{k,c}(x) - f_{k,c}(y)\|_1 - \sum_{k\in\mathbb{Z}} \|f_{k,c}(\Gamma_c(x)) - f_{k,c}(y)\|_1 \geq (1 - \varepsilon_1/2)\, d_T(x, \Gamma_c(x)). \tag{92}$$

We now complete the proof of Lemma 5.15 by proving (90) and (91).

Proof of (90). Suppose that $X_1, \ldots, X_{i-1}$ and Y_i hold. We will show that X_i holds as well. First note that for all vertices in $x, y \in V(\gamma_{c_i})$, by Lemma 5.5 and the definition of f_{k,c_i} (53), we have

$$d_T(x, y) = \sum_{k\in\mathbb{Z}} \|f_{k,c_i}(x) - f_{k,c_i}(y)\|_1 = \sum_{k\in\mathbb{Z}} \|f_{k,c}(x) - f_{k,c}(y)\|_1,$$

thus we only need to prove (89) for pairs $x \in V(\gamma_{c_i})$ and $y \in V(\gamma_{c_j})$ for $j < i$ and $\mu(x, y) = \varphi(c)$. We now divide the pairs with one endpoint in γ_{c_i} into two cases based on Γ_c.

Case I: $x \in V(\gamma_{c_i})$ with $x \neq \Gamma_c(x)$, and $y \in V(\gamma_{c_j})$ for some $j < i$, and $\mu(x, y) = \varphi(c)$.

In this case, by Lemma 5.13, there exists a vertex $z \in R_c(c_i)$ such that

$$d(x, v_{c_i}) \leq d(z, v_{c_i}) \leq 2 \max\big(\delta\, \mathsf{len}(E(\gamma_{c_i})), d_T(x, v_{c_i})\big).$$

If $d(x, v_{c_i}) \leq \delta\, \mathsf{len}\,(E(\gamma_{c_i}))$, then by (18), we have $\mathsf{len}(E(\gamma_{c_i})) = \rho_\chi(x, v_{c_i}; \delta)$, hence

$$\begin{aligned} d_T(z, \Gamma_c(z)) &\leq d_T(v_{c_i}, \Gamma_c(z)) + 2\max(\delta\, \mathsf{len}(E(\gamma_{c_i})), d_T(x, v_{c_i})) \\ &\leq d_T(v_{c_i}, \Gamma_c(z)) + 2\max(\delta\, \rho_\chi(x, v_{c_i}; \delta), d_T(x, v_{c_i})) \\ &\leq d_T(v_{c_i}, \Gamma_c(z)) + 2\,\delta\, \rho_\chi(x, v_{c_i}; \delta) + 2\, d_T(x, v_{c_i}) \\ &\leq 2\delta\, \rho_\chi(x, v_{c_i}; \delta) + 2\, d_T(x, \Gamma_c(z)). \end{aligned} \tag{93}$$

Since $z \in R_c(c_i)$, by definition we have $\Gamma_c(z) \neq z$, therefore by Lemma 5.11, $\Gamma_c(z) = v_{c'}$ for some color $c' \in \chi(P_{z\, v_c}) \setminus \{c\}$. The function φ is non-decreasing along any root–leaf path, hence $\chi(\Gamma_c(z), p(\Gamma_c(z))) = c_\ell$ for some $\ell < i$.

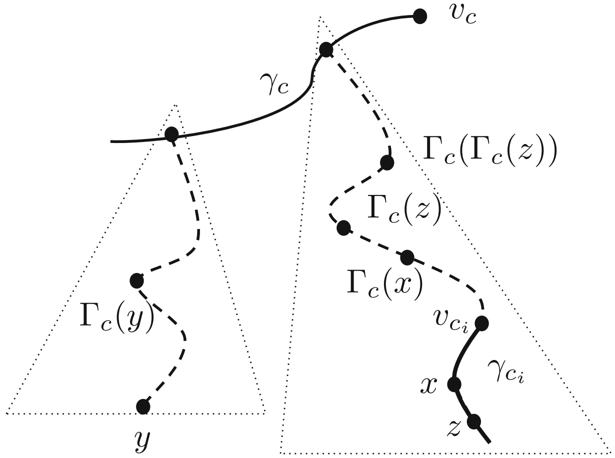

Fig. 2 Position of vertices in the subtree $T(c)$ for Case I

We refer to Fig. 2 for the relative position of the vertices referenced in the following inequalities. Using our assumption that $X_1, \ldots, X_{i-1}$ and Y_i hold, we can write

$$\begin{aligned}
&\sum_{k\in\mathbb{Z}} \|f_{k,c}(z) - f_{k,c}(y)\|_1 \\
&\overset{Y_i}{\geq} d_T(\Gamma_c(z), z) - (\varepsilon_1/2)\, d_T(z, \Gamma_c(z)) + \sum_{k\in\mathbb{Z}} \|f_{k,c}(\Gamma_c(z)) - f_{k,c}(y)\|_1 \\
&\overset{X_{\max(\ell,j)}}{\geq} d_T(\Gamma_c(z), z) - (\varepsilon_1/2)\, d_T(z, \Gamma_c(z)) + d_T(\Gamma_c(z), y) \\
&\quad -\varepsilon_2\, d_T(\Gamma_c(\Gamma_c(z)), \Gamma_c(y)) - \varepsilon_1\, d_T(\Gamma_c(z), y) - \delta\, \rho_\chi(\Gamma_c(z), y; \delta) \\
&\geq d_T(y, z) - (\varepsilon_1/2)\, d_T(z, \Gamma_c(z)) - \varepsilon_2\, d_T(\Gamma_c(\Gamma_c(z)), \Gamma_c(y)) \\
&\quad -\varepsilon_1\, d_T(\Gamma_c(z), y) - \delta\, \rho_\chi(\Gamma_c(z), y; \delta).
\end{aligned}$$

We may assume that $\varepsilon_1 < 1$, otherwise there is nothing to prove. Using the preceding inequality, and applying Lemma 5.14 on pairs (z, y) and (x, y) implies that

$$\begin{aligned}
&\sum_{k\in\mathbb{Z}} \|f_{k,c}(x) - f_{k,c}(y)\|_1 \\
&\quad \geq d_T(x, y) - (\varepsilon_1/2)\, d_T(z, \Gamma_c(z)) - \varepsilon_2\, d_T(\Gamma_c(\Gamma_c(z)), \Gamma_c(y)) \\
&\qquad -\varepsilon_1\, d_T(\Gamma_c(z), y) - \delta\, \rho_\chi(\Gamma_c(z), y; \delta) \\
&\quad \overset{(93)}{\geq} d_T(x, y) - (\varepsilon_1/2)\big(2\, d_T(x, \Gamma_c(z)) + 2\delta\, \rho_\chi(x, v_{c_i}; \delta)\big) \\
&\qquad -\varepsilon_2\, d_T(\Gamma_c(\Gamma_c(z)), \Gamma_c(y)) - \varepsilon_1\, d_T(\Gamma_c(z), y) - \delta\, \rho_\chi(\Gamma_c(z), y; \delta).
\end{aligned}$$

where in the last line we have used the fact that $\varepsilon_1 \leq 1$.

We have $\chi(x, p(x)) = \chi(z, p(z)) = c_i$. Moreover, since $\Gamma_c(z) \neq z$, using Lemma 5.11 it is easy to check that $x \in P_{z\,\Gamma_c(z)}$. Therefore, by Lemma 5.12,

$d_T(\Gamma_c(\Gamma_c(z)), y) \leq d_T(\Gamma_c(z), y) \leq d_T(\Gamma_c(x), y)$, and combining this with the preceding inequality yields

$$\sum_{k\in\mathbb{Z}} \|f_{k,c}(x) - f_{k,c}(y)\|_1 \geq d_T(x, y) - (\varepsilon_1/2)\big(2\, d_T(x, \Gamma_c(z)) + 2\delta\, \rho_\chi(x, v_{c_i}; \delta)\big)$$
$$-\varepsilon_2\, d_T(\Gamma_c(x), \Gamma_c(y)) - \varepsilon_1\, d_T(\Gamma_c(z), y) - \delta\rho_\chi(\Gamma_c(z), y; \delta).$$

Recall the definition of $C(x, y; \delta)$ in (18). Since by Lemma 5.11, $\Gamma_c(z) = v_{c'}$ for some color $c' \in \chi(P_{z\, v_c}) \setminus \{c\}$, we have $C(\Gamma_c(z), y; \delta) \subseteq C(v_{c_i}, y; \delta)$, hence $\rho_\chi(v_{c_i}, y; \delta) \geq \rho_\chi(\Gamma_c(z), y; \delta)$ and thus,

$$\begin{aligned}
&\sum_{k\in\mathbb{Z}} \|f_{k,c}(x) - f_{k,c}(y)\|_1 \\
&\quad\geq d_T(x, y) - (\varepsilon_1/2)\big(2\, d_T(x, \Gamma_c(z)) + 2\delta\, \rho_\chi(x, v_{c_i}; \delta)\big) \\
&\qquad -\varepsilon_2\, d_T(\Gamma_c(x), \Gamma_c(y)) - \varepsilon_1\, d_T(\Gamma_c(z), y) - \delta\, \rho_\chi(v_{c_i}, y; \delta) \\
&\quad\geq d_T(x, y) - \varepsilon_1\, d_T(x, \Gamma_c(z)) - \varepsilon_2\, d_T(\Gamma_c(x), \Gamma_c(y)) - \varepsilon_1 d_T(\Gamma_c(z), y) \\
&\qquad -\delta\big(\rho_\chi(v_{c_i}, y; \delta) + \varepsilon_1 \rho_\chi(x, v_{c_i}; \delta)\big) \\
&\quad\geq d_T(x, y) - \varepsilon_1\, d_T(x, \Gamma_c(z)) - \varepsilon_2\, d_T(\Gamma_c(x), \Gamma_c(y)) - \varepsilon_1 d_T(\Gamma_c(z), y) \\
&\qquad -\delta\big(\rho_\chi(x, v_{c_i}; \delta) + \rho_\chi(v_{c_i}, y; \delta)\big),
\end{aligned}$$

where in the last line we have again used that $\varepsilon_1 < 1$.

The set of colors that appear on the paths $P_{x\, v_{c_i}}$ and $P_{v_{c_i}\, y}$ are disjoint, therefore $\rho_\chi(x, y; \delta) = \rho_\chi(x, v_{c_i}; \delta) + \rho_\chi(v_{c_i}, y; \delta)$, and

$$\begin{aligned}
&\sum_{k\in\mathbb{Z}} \|f_{k,c}(x) - f_{k,c}(y)\|_1 \\
&\qquad\geq d_T(x, y) - \varepsilon_1\, d_T(x, \Gamma_c(z)) \\
&\qquad\qquad -\varepsilon_2\, d_T(\Gamma_c(x), \Gamma_c(y)) - \varepsilon_1\, d_T(\Gamma_c(z), y) - \delta\rho_\chi(x, y; \delta) \\
&\qquad= d_T(x, y) - \varepsilon_1\, d_T(x, y) - \varepsilon_2\, d_T(\Gamma_c(x), \Gamma_c(y)) - \delta\rho_\chi(x, y; \delta).
\end{aligned}$$

Case II: $x \in V(\gamma_{c_i})$ with $x = \Gamma_c(x)$, and $y \in V(\gamma_{c_j})$ for some $j < i$, and $\mu(x, y) = \varphi(c)$.

In this case, we first note that since $c = c_1$, $x \notin V(\gamma_c)$. Hence we can suppose that $x \in V(T(c'))$ for some $c' \in \rho^{-1}(c)$. Recall that $\frac{\varepsilon_2}{2} = C'\varepsilon$, where C' is the constant from Lemma 4.11. By Lemma 4.11 (with c', x, and $\frac{\varepsilon_2}{2}$ substituted for c, v, and ε, respectively, in the statement of Lemma 4.11), there exist vertices $u, u' \in \{x\} \cup \{v_a : a \in \chi(E(P_{x\, v_{c'}}))\}$ such that

$$d_T(x, u) \leq (\varepsilon_2/2)\, d_T(u', u). \tag{94}$$

For all vertices $z \in V(P_{u'u}) \setminus \{u'\}$ and for all $k \in \mathbb{Z}$,

$$\tau_k(z) \neq 0 \implies 2^k < \left(\frac{d_T(u, u')}{\varepsilon(\varphi(\chi(u, p(u))) - \varphi(\chi(v_{c'}, p(v_{c'}))))}\right).$$

We have $\chi(v_{c'}, p(v_{c'})) = c$, and this condition is exactly the same condition as (75) for $\Gamma_c(u)$, therefore

$$d_T(x, u) \le (\varepsilon_2/2)\, d_T(u', u) \le (\varepsilon_2/2)\, d_T(\Gamma_c(u), u). \tag{95}$$

Note that the assumption that $\Gamma_c(x) = x$ implies that $u \ne x$ and $u = v_a$ for some color $a \in \chi(E(P_{x\, v_{c'}}))$.

We have

$$\begin{aligned}
&\sum_{k\in\mathbb{Z}} \|f_{k,c}(x) - f_{k,c}(y)\|_1 - \sum_{k\in\mathbb{Z}} \|f_{k,c}(u) - f_{k,c}(y)\|_1 \\
&\quad \ge -\sum_{k\in\mathbb{Z}} \|f_{k,c}(x) - f_{k,c}(u)\|_1 \overset{(5.9)}{=} -d_T(x, u) \overset{(95)}{\ge} d_T(x, u) - \varepsilon_2\, d_T(u, \Gamma_c(u)) \\
&\quad \ge d_T(x, u) - \varepsilon_2\, d_T(x, \Gamma_c(u)) = d_T(x, u) - \varepsilon_2\, d_T(\Gamma_c(x), \Gamma_c(u)).
\end{aligned} \tag{96}$$

Since $u = v_a$ for some color $a \in \chi(E(P_{x\, v_{c'}}))$, $\chi(u, p(u)) = c_\ell$, for some $\ell < i$ and $X_{\max(\ell, j)}$ implies that

$$\sum_{k\in\mathbb{Z}} \|f_{k,c}(u) - f_{k,c}(y)\|_1 \ge d_T(u, y) - \varepsilon_2\, d_T(\Gamma_c(u), \Gamma_c(y)) - \varepsilon_1\, d_T(u, y) - \delta\, \rho_\chi(u, y; \delta).$$

Recall the definition of $C(x, y; \delta)$ in (18). We have $u = v_a$ for some color $a \in (E(P_{x\, v_{c'}}))$, therefore $C(u, y; \delta) \subseteq C(x, y; \delta)$, and $\rho_\chi(u, y; \delta) \le \rho_\chi(x, y; \delta)$. Now we can write

$$\begin{aligned}
&\sum_{k\in\mathbb{Z}} \|f_{k,c}(u) - f_{k,c}(y)\|_1 \\
&\quad \ge d_T(u, y) - \varepsilon_2\, d_T(\Gamma_c(u), \Gamma_c(y)) - \varepsilon_1\, d_T(u, y) - \delta\, \rho_\chi(x, y; \delta).
\end{aligned} \tag{97}$$

Adding (96) and (97) we can conclude that

$$\begin{aligned}
&\sum_{k\in\mathbb{Z}} \|f_{k,c}(x) - f_{k,c}(y)\|_1 \\
&\quad \ge d_T(u, y) + d_T(u, x) - \varepsilon_2\, (d_T(\Gamma_c(x), \Gamma_c(u)) \\
&\qquad + d_T(\Gamma_c(u), \Gamma_c(y))) - \varepsilon_1\, d_T(x, y) - \delta\, \rho_\chi(x, y; \delta) \\
&\quad \ge d_T(x, y) - \varepsilon_2\, d_T(\Gamma_c(x), \Gamma_c(y)) - \varepsilon_1\, d_T(x, y) - \delta\, \rho_\chi(x, y; \delta),
\end{aligned}$$

completing the proof of (90).

Proof of (91). We prove this inequality by first bounding the probability that (92) holds for a fixed x and all $y \in V(\gamma_{c_j})$ (for a fixed $j \in \{1, \ldots, i-1\}$) with $\mu(x, y) = \varphi(c)$. Then we use a union bound to complete the proof.

We start the proof by giving some definitions. For a vertex $x \in R_c(c_i)$, let

$$S_x = \big\{ j \in \{1, \ldots, i-1\} : \text{ there exists a } v \in V(\gamma_{c_j}) \text{ such that } \mu(x, v) = \varphi(c) \big\}.$$

For $a \in S_x$, we define $w(x;a)$ as the vertex $v \in V(\gamma_a)$ which is furthest from the root among those satisfying $\mu(x,v) = \varphi(c)$. Finally for $x \in R_c(c_i)$, we put

$$\beta_x = \max\big\{k \in \mathbb{Z} : \exists z \in P_{x\,\Gamma_c(x)} \setminus \{\Gamma_c(x)\},\ \tau_k(z) \neq 0\big\}.$$

Inequality (75) implies

$$2^{\beta_x} < \frac{d_T(x, \Gamma_c(x))}{\varepsilon(\varphi(c_i) - \varphi(c))}. \tag{98}$$

By definition of R_c, for all elements $x \in R_c(c_i)$, we have $\Gamma_c(x) \neq x$. Moreover, by Lemma 5.11, $\Gamma_c(x) = v_{c'}$ for some $c' \in \chi(E(P_{x\,v_c})) \setminus \{c\}$. Now, for $x \in R_c(c_i)$ and $a \in S_x$ we apply Lemma 5.10 with $\varepsilon_1/2 = 12\varepsilon$ to write

$$\begin{aligned}
&\mathbb{P}_{\mathcal{E}}\Big[\exists y \in P_{w(x;a),v_c} : \sum_{k\in\mathbb{Z}} \|f_{k,c}(x) - f_{k,c}(y)\|_1 \\
&\quad \le (1-\varepsilon_1/2)d_T(x,\Gamma_c(x)) + \sum_{k\in\mathbb{Z}} \|f_{k,c}(y) - f_{k,c}(\Gamma_c(x))\|_1\Big] \\
&\quad \le \frac{1}{\lceil \log_2 1/\delta\rceil} \exp\Big(-12\frac{d_T(x,\Gamma_c(x))}{2^{\beta_x+2}\varepsilon}\Big) \\
&\quad \stackrel{(98)}{\le} \frac{\exp(-3(\varphi(c_i)-\varphi(c)))}{\lceil \log_2 1/\delta\rceil}.
\end{aligned} \tag{99}$$

Note that, for all $y \in V(\gamma_{c_a})$ with $\mu(x,y) = \varphi(c)$, we have $y \in P_{w(x;a),v_c}$.

By definition of $R_c(c_i)$, $|R_c(c_i)| \le \lceil \log_2 \delta^{-1}\rceil$. We also have $\varphi(c_j) \le \varphi(c_i)$ for $j < i$, and by Corollary 4.8, $|S_x| \le i < 2^{\varphi(c_i)-\varphi(c)+1}$. Taking a union bound over all $x \in R_c(c_i)$ and $a \in S_x$ implies

$$\begin{aligned}
\mathbb{P}_{\mathcal{E}}[\overline{Y_i}] &\stackrel{(99)}{\le} \sum_{x\in R_c(c_i)} |S_x|\Big(\frac{1}{\lceil \log_2 \delta^{-1}\rceil} \exp(-3(\varphi(c_i)-\varphi(c)))\Big) \\
&< \big(\lceil \log_2 \delta^{-1}\rceil 2^{\varphi(c_i)-\varphi(c)+1}\big)\Big(\frac{1}{\lceil \log_2 \delta^{-1}\rceil} \exp(-3(\varphi(c_i)-\varphi(c)))\Big) \\
&= 2^{\varphi(c_i)-\varphi(c)+1} \exp(-3(\varphi(c_i)-\varphi(c))).
\end{aligned}$$

Since $\varphi(c_i) \ge \varphi(c)$, by an elementary calculation we conclude that

$$\mathbb{P}_{\mathcal{E}}[\overline{Y_i}] < 2 \cdot 2^{-3(\varphi(c_i)-\varphi(c))},$$

which completes the proof of (91). $\square$

Finally, we present the proof of Lemma 5.7.

Proof of Lemma 5.7 Let C be the same constant as the constant in Lemma 5.15. For the sake of contradiction, suppose that

$$\mathbb{P}\Big[\forall x, y \in V,\ (1 - C\varepsilon)\, d_T(x, y) - \delta\, \rho_\chi(x, y; \delta) \\ \leq \sum_{i \in \mathbb{Z}} \| f_i(x) - f_i(y)\|_1 \leq d_T(x, y)\Big] = 0.$$

Now let $c \in \chi(E) \cup \{\chi(r, p(r))\}$ be a color with a maximal value of $\varphi(c)$ such that

$$\mathbb{P}\Big[\forall x, y \in V(T(c)),\ (1 - C\varepsilon)\, d_T(x, y) - \delta\, \rho_\chi(x, y; \delta) \\ \leq \sum_{i \in \mathbb{Z}} \| f_{i,c}(x) - f_{i,c}(y)\|_1 \leq d_T(x, y)\Big] = 0. \tag{100}$$

For $a \in \chi(E)$, $\kappa(a) > 0$. Hence, for all $c' \in \rho^{-1}(c)$, by (32), $\varphi(c') > \varphi(c)$, and by maximality of c, for all $c' \in \rho^{-1}(c)$, we have

$$\mathbb{P}\Big[x, y \in V(T(c')),\ (1 - C\varepsilon)\, d_T(x, y) - \delta\, \rho_\chi(x, y; \delta) \\ \leq \sum_{i \in \mathbb{Z}} \| f_{i,c'}(x) - f_{i,c'}(y)\|_1 \leq d_T(x, y)\Big] > 0.$$

But now applying Lemma 5.15 contradicts (100), completing the proof. □

Acknowledgments This research was partially supported by NSF Grants CCF-0644037, CCF-0915251, and a Sloan Research Fellowship. A significant portion of this work was completed during a visit of the authors to the Institut Henri Poincaré.

References

1. Andoni, A., Charikar, M., Neiman, O., Nguyen, H.L.: Near linear lower bound for dimension reduction in l1. In: FOCS, pp. 315–323 (2011)
2. Brinkman, B., Charikar, M.: On the impossibility of dimension reduction in ℓ_1. J. ACM **52**(5), 766–788 (2005)
3. Bourgain, J., Lindenstrauss, J., Milman, V.: Approximation of zonoids by zonotopes. Acta Math. **162**(1–2), 73–141 (1989)
4. Batson, J.D., Spielman, D.A., Srivastava, N.: Twice-Ramanujan sparsifiers. SIAM J. Comput. **41**(6), 1704–1721 (2012)
5. Charikar, M., Sahai, A.: Dimension reduction in the ℓ_1 norm. In: FOCS, pp. 551–560 (2002)
6. Erdős, P., Lovász, L.: Problems and results on 3-chromatic hypergraphs and some related questions. In: Infinite and Finite Sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), vol. II, pp. 609–627. Colloquium Mathematical Society János Bolyai, vol. 10. North-Holland, Amsterdam (1975)
7. Gupta, A., Krauthgamer, R., Lee, J.R.: Bounded geometries, fractals, and low-distortion embeddings. In: FOCS, pp. 534–543 (2003)
8. Johnson, W.B., Lindenstrauss, J.: Extensions of Lipschitz mappings into a Hilbert space. In: Conference in Modern Analysis and Probability (New Haven, CT, 1982), Contemporary Mathematics, vol. 26, pp. 189–206. American Mathematical Society, Providence (1984)

9. Lee, J.R., de Mesmay, A., Moharrami, M.: Dimension reduction for finite trees in l_1. In: SODA, pp. 43–50 (2012)
10. Lee, J.R., Naor, A.: Embedding the diamond graph in L_p and dimension reduction in L_1. Geom. Funct. Anal. **14**(4), 745–747 (2004)
11. Lee, J.R., Naor, A., Peres, Y.: Trees and Markov convexity. Geom. Funct. Anal. **18**(5), 1609–1659 (2009)
12. Matoušek, J.: On embedding trees into uniformly convex Banach spaces. Israel J. Math. **114**, 221–237 (1999)
13. Matoušek, J.: Open problems on embeddings of finite metric spaces. http://kam.mff.cuni.cz/matousek/haifaop.ps (2002)
14. McDiarmid, C.: Concentration. In: Habib, M., McDiarmid, C., Ramirez-Alfonsin, J., Reed, B. (eds.) Probabilistic Methods for Algorithmic Discrete Mathematics, Algorithms and Combinations, vol. 16, pp. 195–248. Springer, Berlin (1998)
15. Newman, I., Rabinovich, Y.: On cut dimension of ℓ_1 metrics and volumes, and related sparsification techniques. CoRR. http://CoRR/abs/1002.3541 (2010)
16. Regev, O.: Entropy-based bounds on dimension reduction in L_1. Israel J. Math. http://arxiv/abs/1108.1283 (2011)
17. Schechtman, G.: More on embedding subspaces of L_p in l_r^n. Compos. Math. **61**(2), 159–169 (1987)
18. Schulman, L.J.: Coding for interactive communication. IEEE Trans. Inform. Theory 42(6, part 1), 1745–1756 (1996). Codes and complexity
19. Talagrand, M.: Embedding subspaces of L_1 into l_1^N. Proc. Am. Math. Soc. **108**(2), 363–369 (1990)