Applications of Metric Embeddings in Solving Combinatorial Problems

Mohammad Moharrami

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2013

Reading Committee:

James R. Lee, Chair

Paul Beame

Daniel Grossman

Program Authorized to Offer Degree:
UW Computer Science and Engineering

University of Washington

Abstract

Applications of Metric Embeddings in Solving Combinatorial Problems

Mohammad Moharrami

Chair of the Supervisory Committee:
Associate Professor James R. Lee
Computer Science and Engineering

Metric embeddings constitute one of the fundamental tools for exploiting the underlying geometric structure of many combinatorial problems. In this dissertation we study some of the applications of metric embeddings in the field of computer science and resolve some of the previously open questions in this area. The results in this dissertation are divided into three parts.

In the first part, we study dimension reduction for tree metrics. We show that every n-point tree metric admits a $(1+\varepsilon)$ distortion embedding into $\ell_1^{C_\varepsilon \log n}$, for every $\varepsilon > 0$, where $C_\varepsilon = O\left((\frac{1}{\varepsilon})^4 \log \frac{1}{\varepsilon})\right)$. In the case of complete d-ary trees we show that this bound can be improved to $C_\varepsilon = O\left(\frac{1}{\varepsilon^2}\right)$. We also show a lower-bound for the dimension required for embedding complete d-ary trees into ℓ_1, which matches the upper bound up to a factor of $O(\log 1/\varepsilon)$.

In the second part, we construct two families of metric spaces using the graph product of [Lee and Raghavendra, DCG 2010], and use these constructions to answer two previously open questions. The first construction is used to show that for every $\alpha > 0$ and $n \in \mathbb{N}$, there exist n-point metric spaces (X, d) where every "scale" admits a Euclidean embedding with distortion at most α, but the whole space requires distortion at least $\Omega(\sqrt{\alpha \log n})$. This shows that the scale-gluing lemma [Lee, SODA 2005] is tight. Previously the matching upper bound was only known for $\alpha = O(1)$ and $\alpha = \Theta(\log n)$.

The second construction is used to answer an open problem about negative type metrics.

A metric space (X, d) is said to be of negative type if the space $(X, \sqrt{d})$ admits an isometric embedding into ℓ_2. Metrics of negative type are used to study the power of various inequalities in semi-definite programming relaxations for the Sparsest Cut problem. We exhibit a family of metric spaces $\{(X_m, d_m)\}_{m\in\mathbb{N}}$ such that $(X_m, \sqrt{d_m})$ admits constant distortion embedding into ℓ_2, yet it can not be embedded into a metric of negative type with constant distortion.

In the last part, we use a new type of random metric embedding to bound the flow and cut gap in node-capacitated planar graphs. The classical Okamura-Seymour theorem states that for an edge-capacitated, multi-commodity flow instance in which all terminals lie on a single face of a planar graph, there exists a feasible concurrent flow if and only if all cuts have capacity larger than the demand across the cut. Simple examples show that a similar theorem does not hold if the capacities are on the vertices rather than edges. Nevertheless, we show that there exists a universal constant $\delta > 0$, such that if the equivalent vertex-cut conditions are satisfied, then one can simultaneously route a δ fraction of flow for all the demands.

TABLE OF CONTENTS

Page

LIST OF FIGURES

ACKNOWLEDGMENTS

First and foremost, I would like to express my gratitude toward my advisor James R. Lee for his support and patience. I have been really lucky to have him as my mentor. He has always been great at pointing out my shortcomings and helping me to overcome them. Working with him has changed my approach to research. I would also like to thank my committee members, Paul Beame, Daniel Grossman, and Maryam Fazel.

The research in this thesis was done collaboration with Alex Jaffe, James Lee, Manor Mendel, Arnaud deMesmay and Sushast Sachdeva. Working with these collaborators have been one of the most exciting part of my research.

I have been blessed to be able to work in the CSE department which provided a great environment for research. I am thankful to many individuals in this department including: my officemates Widad and Thach; members of theory group Anna Karlin, Anup Rao, Ben, Elisa, Jessica and Punya; and our departmental staff Lindsay and Melody.

I am also grateful to have many great friends inside and outside of the CSE department who have supported me during my research. I would like to thank Hadi, Emad, James, Siu on, Siu Man, Amir, Amin, Atiye, Hamed, MohammadHossein, Mohsen, Rahman, Shirin and Siavosh. I would also like to thank Piyush Srivastava for his helpful comments on my dissertation.

Last but not least, I am indebted to my parents MohammadMehdi and Moloud, and my sister Maryam. Their faith and confidence in my abilities is what encouraged me to pursue and complete my degree.

DEDICATION

To my dear parents MohammadMehdi and Moloud

Chapter 1

INTRODUCTION

1.1 Overview

A *metric space* is a set equipped with a distance function satisfying some basic properties (see Definition 1.1 for the formal definition). Metric spaces appear naturally in many areas of computer science. For example, in genome sequencing one may consider various notions of distance such as edit distance or block distance, to measure the similarity or difference between sequences of proteins. Similarly, in network routing the distance between two nodes may be given by the round trip time between the nodes. Metric embeddings provide one of the essential tools for understanding the underlying geometric structure of such problems. A metric embedding is a map from a given metric space to another metric space[1] which approximately preserving the distance between pairs of points. This can be treated as a reduction and allows the use of existing algorithms known to perform well on these spaces.

While the main focus in metric embeddings is to preserve distances, reducing the dimension of the target metric spaces also proved to be important. Reducing the dimension allows a workaround for the exponential "curse of dimensionality," and in some cases it yields a more succinct representation of the original space. A seminal result of Johnson and Lindenstrauss [50] implies that for every $\varepsilon > 0$, every n-point subset X of ℓ_2 can be embedded into ℓ_2^k ($\mathbb{R}^k$ equipped with Euclidean norm), with $k = O(\log n/\varepsilon^2)$ while the distances are preserved up to $1+\varepsilon$ multiplicative factor. This effectively allows us to reduce the dimension of the target metric space to $\log(n)$. One may hope to achieve a similar result for reducing the dimension of a metric in other normed spaces such as ℓ_1. However, it was shown in [16] that achieving such a result for ℓ_1 metrics is impossible. In fact, the best known lower and upper bounds for dimension reduction in ℓ_1 are almost linear in the number of points in the

[1]The host metric space is usually a well-known metric space such as $\mathbb{R}^d$ with ℓ_1 (Manhattan) norm, or ℓ_2 (Euclidean) norm.

metric (see [88] and [4]). Nonetheless, this question is still open for restricted classes of ℓ_1 metrics. Perhaps one of the simplest classes of ℓ_1 metrics is the shortest path metric of trees. In the first part of this dissertation we show that it is possible to reduce the dimension of a metric that comes from a tree metric, and we provide an algorithm to find the mapping. We will discuss this result in more detail in Section 1.3.1, and provide the proof in Chapter 3.

Another place where metric spaces arise is in the analysis of approximation algorithms for some NP-hard optimization problems. For instance, we can interpret a cut as the distance function of a metric in which the points that are on different sides of the cut are "far" from each other, and the points that are on the same side of the cut are "close" to each other. Using this view in [76, 8], it became clear that the efficiency of certain mathematical programs for approximating cut problems in the graph is intimately tied to finding maps from finite metric spaces into various normed spaces such as ℓ_1 or ℓ_2.

One approach to find such maps was introduced by Rao in [91]. In this approach distances in the original space are divided into different scales and each scale is handled separately. Now, suppose we can construct a collection of maps from some finite metric space into a Euclidean space, each reflecting the geometry of a single "scale." Is there a way of gluing these mappings together to form a global mapping which reflects the entire geometry of the original space? The answers to such questions have played a fundamental role in the best-known approximation algorithms for Sparsest Cut [57, 66, 21, 5] and Graph Bandwidth [91, 57, 62], and have found applications in approximate multi-commodity max-flow/min-cut theorems in graphs [91, 57]. The second result presented in this dissertation, shows that the approaches presented in [57, 66] are optimal, disproving a conjecture stated in [66]. We will discuss the details about this result in Section 1.3.2 and present its proof in Chapter 2.

Another approach to find such maps is by first finding a map from the original metric space to an intermediary metric space, and then finding a map from that intermediary metric space to the host metric space. The intermediate step is usually chosen so that the mapping to the intermediary space can be found efficiently. Goemans and Linial suggested *negative* *type*[2] metrics as the intermediate metric space and conjectured that negative type metric spaces can be embedded into ℓ_1 such that all distances are preserved up to a constant factor. This conjecture was disproved in [58]. Subsequently, [33] and [25, 27, 26, 28] improved upon [33] to give a poly-logarithmic gap between ℓ_1 and negative type metric spaces. Despite the fact that the original conjecture does not hold, this approach has led to some breakthroughs in the design and analysis of algorithms[6, 5, 66, 21]. It is a common observation that while in all of these works a negative type metric is the intermediary metric space, in fact it is sufficient to use a more general family of metric spaces for those algorithms to obtain the same guarantees. In our second result we show a very strong quantitative separation between negative type metrics and this more general family of metric spaces. In Section 1.3.3 we will discuss this problem in more details and we present the proofs in Chapter 2.

The last result that we present in this dissertation is about the relationship between the maximum flow and the minimum cut in graphs. A classical theorem of Menger [86] states that for a single source single sink flow, the value of the maximum flow is equal to the value of the minimum cut. Equivalently, we can say: there is a flow of value f between source and sink if and only if all the cuts that separate source and sink have value at least f. One may try to generalize this result to multi-commodity flows. It follows from [76] and [46] that for graphs with n vertices, there can be a gap of $\Omega(\log n)$ between the bound that follows from cuts and the amount of flow that can be simultaneously routed in the graph. It was shown in [91] that this bound can be improved to $O(\sqrt{\log n})$ for planar graphs (and the question of whether this bound can be improved to $O(1)$ for planar graphs is still open). A seminal result in the study of multi-commodity flows is the classical Okamura-Seymour theorem [89]. which states that in a planar graph, if all sources and sinks lie on a single face of the graph, the value of maximum flow and minimum cut is exactly the same. In Chapter 5 we present a generalization of this result to node-capacitated graphs. More discussion about this problem can be found in 1.3.4.

---

[2] A metric space (X, d) is said to be of negative type if the space $(X, \sqrt{d})$ admits an isometric embedding into ℓ_2.

1.2 Basic Definitions and Notations

For standard definitions in computer science we refer to standard texts such as [55, 31]. In dealing with metric spaces, we mostly follow the notational convention from [78, Ch. 15]. In what follows, we first give some of the non-standard notations that are used in the work. We also provide some of the definition from [78, Ch. 15] and [55] for completeness.

1.2.1 Notations

In this thesis, besides the standard O, Θ, and Ω notation, we will use $g \lesssim f$, $g \asymp f$, and $g \gtrsim f$ to denote $g = O(f)$, $g = \Theta(f)$, and $g = \Omega(f)$ respectively. Moreover we may use the notation $g = \tilde{O}(f)$ if there exists a polynomial p such that $g = O\left(f \cdot p(\log f)\right)$.

For $k \in \mathbb{N}$, we write $[k] = \{1, 2, \ldots, k\}$. We will also use $\mathbb{R}^+$ to indicate the set of all positive reals. We will also use the convention $\mathbb{N} = \{1, 2, \ldots\}$.

For a graph $G = (V, E)$, we will denote by $V(G)$ and $E(G)$ vertex set and edge set of G, respectively. For a connected, rooted tree $T = (V, E)$ and $x, y \in V$, we use the notation P^T_{xy} to denote the unique path between x and y in T, and P^T_x for P^T_{rx}, where r is the root of T. In the cases that there are no ambiguity we will drop the superscript T for the ease of notation.

1.2.2 Metric Spaces

We start this section by formally defining what is a metric space.

Definition 1.1. *For a set X and a distance function $\rho : X \to [0, \infty)$, we say that (X, ρ) forms a metric space if it satisfies the following conditions:*

i) The distance function is symmetric, i.e, for all $x, y \in X$, $\rho(x, y) = \rho(y, x)$;

ii) It satisfies the triangle inequality, i,e, for all $x, y, z \in X$, $\rho(x, z) \leq \rho(x, y) + \rho(y, z)$;

iii) The distance between distinct points in X is non-zero[3], i.e, for all $x, y \in X$, $\rho(x, y) = 0$ if and only if $x = y$.

[3] We call the pair (X, ρ) a pseudo-metric if we replace this condition with $x = y \Rightarrow \rho(x, y) = 0$.

Given two metric spaces (X, ρ) and (Y, d) and a map $f : X \to Y$, we say that f is *Lipschitz* if there exists a universal constant K such that

$$\forall x, y \in X : d(f(x), f(y)) \leq K \cdot \rho(x, y). \tag{1.1}$$

We call the smallest constant K satisfying the above condition the *Lipschitz constant* of f, and denote that by $\mathrm{Lip}(f)$[4]. We now define the distortion of the map f to be:

$$\mathsf{dist}(f) = \mathrm{Lip}(f) \cdot \mathrm{Lip}(f^{-1}). \tag{1.2}$$

Map f is said to be *isometric* if $\mathsf{dist}(f) = 1$. We say that (X, ρ) D-embeds into (Y, d), if there exists a map $f : X \to Y$ with distortion at most D.

Finally, for metric spaces (X, d) and (Y, ρ) we define $c_Y(X) = \inf_{f:X\to Y} \mathsf{dist}(f)$.

Notable metric spaces

The most basic and general family of finite metric spaces that we use is the *shortest path metric* of graphs. Sometimes we will equip G with a non-negative length function $\mathsf{len} : E(G) \to [0, \infty)$, and we let d_{len} denote the shortest-path pseudo-metric on G. We refer to the pair (G, len) as a *metric graph,* and often len will be implicit, in which case we use d_G to denote the path metric. Throughout this document all the graphs are finite unless stated otherwise.

ℓ_p and ℓ_p^k metric spaces. For $p \in [1, \infty)$, and $x \in \mathbb{R}^k$, the ℓ_p norm of x is defined as

$$\|x\|_p = \left(\sum_{i=1}^{k} |x_i|^p \right)^{1/p},$$

and for $p = \infty$,

$$\|x\|_\infty = \max_{i=1}^{k} |x_i|.$$

We use ℓ_p^k to denote $\mathbb{R}^k$ equipped with $\|\cdot\|_p$ norm. Similarly, we use ℓ_p to denote the space of all infinite sequences $(x_1, x_2, \ldots)$ such that

$$\|x\|_p = \left(\sum_{i=1}^{\infty} |x_i|^p \right)^{1/p} \leq \infty.$$

---

[4]In the case that f is not Lipschitz we put $\mathrm{Lip}(f) = \infty$

In the rest of this document we use ℓ_p^k and ℓ_p as the natural metric spaces[5] associated with these spaces.

For the ease of notation, in this document we will use $c_p(X)$ to denote $c_{\ell_p}(X)$.

Negative type metrics and snowflakes. For a metric space (X, d) and a number $\alpha \in (0, 1]$, we use (X, d^α) to denote the metric space where X is equipped with the distance function $d^\alpha(x, y) = d(x, y)^\alpha$. For values $\alpha < 1$, such constructions are commonly referred to as "snowflakes." Let us call a metric space (X, d) a *D-half-snowflake* if $(X, \sqrt{d})$ D-embeds into ℓ_2. *Metrics of negative type* are 1-half-snowflakes. We will use NEG to denote the set of negative type metrics.

Cut metrics. For a set X and subset $S \subset X$, we define the cut distance d_S as

$$d_S(x, y) = |\mathbf{1}_S(x) - \mathbf{1}_S(y)|. \tag{1.3}$$

where $\mathbf{1}_S$ is the indicator function for S. We call the metric (X, d_S) a cut metric.[6] It is well known that the set of ℓ_1 metrics is contained in the cone of cut metrics.

1.2.3 Linear and Semidefinite Programs

Linear Program. Linear Programs are a mathematical model that capture a variety of optimization problems. Given $n, m \in \mathbb{R}^+$, vectors $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, and a matrix $A \in \mathbb{R}^{m \times n}$, a linear program in its standard form is written as:

minimize:	$\langle c\, , \, x \rangle$
subject to:	$Ax \geq b$
and	$x \geq 0$

where for vectors $u, v \in \mathbb{R}^n$, we use the notation $u \geq v$ to denote the coordinate-wise inequality between two vectors.

We say that a linear program is *feasible* if there exists an assignment that satisfies the constraints of the linear program, and it is *bounded* if there exists an assignment that achieves the optimal solution.

[5] the metric space on ℓ_p with distance function $\|x - y\|_p$.

[6] While we call (X, d_S) a cut metric, it is only a pseudo-metric and it does not satisfy 1.1(iii).

The *dual* of linear program written in its standard form shown above, is defined as the program:

maximize:	$\langle b\,,\,y\rangle$
subject to:	$A^T y \le c$
and	$y \ge 0$

The Strong Duality Theorem states that if a linear program is feasible and bounded, then the value of the linear program and its dual are exactly the same.

Semidefinite Program. A semidefinite matrix is a symmetric real matrix with non-negative eigenvalues. Semidefinite programs are generalization of linear programs. The standard form of semidefinite programs are as follows. Given $n, m \in \mathbb{R}^+$, vectors $C \in \mathbb{R}^{n\times n}$, $\{b_i \in \mathbb{R}\}_{i\in m}$, and matrix $\{A_i \in \mathbb{R}^{n\times n}\}_{i\in m}$, is as follows:

minimize:	$C \bullet X$
subject to:	$A_i \bullet X \ge B_i$
and	X is positive semidefinite

where we use $A \bullet B$ to denote the Frobenius inner product. While it is beyond the scope of this document, it is worth mentioning that the notation of duality is also well defined for semidefinite programs, and has many applications in combinatorial optimization.

1.3 Related Works and Background

1.3.1 Dimension Reduction for Trees

The question of reducing the dimension of a metric space has a long and rich history. Perhaps the best known result in this area is the celebrated result of Johnson and Lindenstraus [50] on dimension reduction for subsets of Euclidian space. In [50], it was shown that any finite subset of $X \in \ell_2$, can be embedded into $\ell_2^{\log|X|/\varepsilon^2}$ with distortion $1 + O(\varepsilon)$. Alon [2] also showed a matching lower bound (up to an $O(\log \frac{1}{\varepsilon})$ factor) for dimension reduction in ℓ_2.

While we know the tight bounds for dimension reduction in ℓ_2, the situation for other ℓ_p spaces (in particular ℓ_1) is not resolved yet. Following a series of works from Bourgain-Lindenstrauss-Milman [15] and Schechtman [93], Talagrand [97] showed that any finite subset of $X \in \ell_1$ can be embedded into $\ell_1^{O(|X|\log|X|/\varepsilon^2)}$ with distortion $1+\varepsilon$. Using the tools arising from spectral sparsification techniques of Batson, Spielman and Srivastava [10], Newman and Rabinovich [88] recently improved the bound from Talagrand [97] after almost two decades.

On the lower bound side Brinkman and Charikar [16] showed that for every $n \in \mathbb{N}$ there exists an n point metric $X \in \ell_1$ such that any embedding of X into ℓ_1^d with distortion D requires $d \geq n^{\Omega(1/D^2)}$. Later, Lee and Naor [64] gave a simpler proof of the same theorem. This of course implies that the same dimension reduction that is available for ℓ_2 metrics is impossible for ℓ_1 metrics. More recently, Andoni, Charikar, Neiman, and Nguyen [4], showed that that there exist n-point subsets $X \subset \ell_1$ such that any embedding of them with distortion $1+\varepsilon$ into ℓ_1^d require d to be almost linear in the size of $|X|$ ($d \geq |X|^{1-O(1/\log 1/\varepsilon)}$). Regev [92] also provided a beautiful information theoretic argument that implies both of these lower bounds.

Despite these lower bounds, one can still hope for the possibility of a better dimension reduction for certain finite subsets of ℓ_1. Such a study was undertaken by Charikar and Sahai [20]. 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 the most basic example of this type. It was shown in [20], that for every $\varepsilon > 0$, every n-point tree metric can be embedded into ℓ_1^d with distortion $(1+\varepsilon)$ with $d = O(\frac{\log^3 n}{\varepsilon^2})$ [7]. It is quite natural to ask whether the dependence on n can be improved to $\Omega(f(\varepsilon)\log n)$. It was also stated as an open question in the list "Open problems on embeddings of finite metric spaces" maintained by J. Matoušek [80], asked by Gupta, Lee, and Talwar (at the DIMACS Workshop on Discrete Metric spaces and their Algorithmic Applications (2003). The question was known to others even before 2003, and was asked by Assaf Naor earlier that year.). It is interesting to know that this question was open even for the complete binary tree on n vertices. In Chapter 3 we resolve this question[8]. We show that any n point tree metric can be embedded into ℓ_1^d with distortion $1+\varepsilon$, where $d = O(\frac{\log 1/\varepsilon}{\varepsilon^4}\log n)$. We also show that $d = O(\frac{\log n}{\varepsilon^2})$ in enough

[7]This bound was improved to $O(\frac{\log^2 n}{\varepsilon^2})$ using an observation from A. Gupta.

[8]This work is based on a collaboration with Arnaud deMesmay and James R. Lee [67].

for the special case of k-ary trees.

We also provide a lower bound on the dimension required for embedding k-ary trees [69] matching the upper bound up to a factor of $O(\log 1/\varepsilon)$. It is worth mentioning that this bound for the star graph[9] has an analog in coding theory. Let $X_n = e_1, e_2, ..., e_n$, where e_i's form the standard basis for ℓ_1^n. Then any embedding of X_n with distortion $1 + \varepsilon$ into Hamming cube $\{0,1\}^d$ requires $d = \Omega\left(\frac{\log n}{\varepsilon^2 \log 1/\varepsilon}\right)$. This bound was proved in 1977 by McEliece, Rodemich, Rumsey, and Welch [84] using the Delsarte's linear programming bound [32]. Alon's result for ℓ_2 [2] yields this bound as a special case since for $x, y \in \{0,1\}^d$, $\|x - y\|_2^2 = \|x - y\|_1$. However, the bound from [84], is not exactly comparable with the bound we prove in this dissertation. On the one hand our bound is stronger because we show a lower bound for embedding into ℓ_1^d which contains $\{0,1\}^d$. On the other hand, the lower bound from [84] corresponds to embedding of the leaves of the star graph, while our lower bound corresponds to the embedding of the whole star graph. In fact the existence of the root vertex (the vertex with degree $n - 1$) is essential to the proof presented in this dissertation.

1.3.2 Gluing over Scales

Suppose that you are given a set of maps from a finite metric space (X, ρ) into ℓ_2, each of which preserves distances for some scale of distances. Is there a way to glue these maps to produce a single map that preserves all distances? The answer to this question has played an important role in design of approximation algorithms such as Sparsest Cut [57, 66, 21, 5], Graph Bandwidth [91, 57, 62], and multi-commodity max-flow/min-cut theorems in graphs [91, 57].

Let (X, ρ) be an n-point metric space, and suppose that for every *scale* $\tau \in \mathbb{R}^+$, we are given a non-expansive mapping $\varphi_\tau : X \to \ell_2$ which satisfies the following. For every $x, y \in X$ with $\rho(x, y) \geq \tau$, we have

$$\|\varphi_\tau(x) - \varphi_\tau(y)\|_p \geq \alpha\tau.$$

---

[9] The n-node star is the simple, undirected graph, where one node has degree $n - 1$ and all other nodes have degree one.

The Gluing Lemma of [66] generalizes the approach of [57] and gives a way to combine these maps to construct an embedding of (X, d) into ℓ_2 with distortion $O(\sqrt{\alpha \log n})$. The embedding theorem of Bourgain [12], says that it is always possible to construct a map with distortion $O(\log n)$, hence we may assume that $\alpha = O(\log n)$. It was known that the bound provided in [66] is tight for $\alpha = \Theta(1)$ (see [87]) and $\alpha = \Theta(\log n)$ (see [76, 8]), but nowhere in between, and it was conjectured in [66], that it is possible to achieve distortion $O(\alpha + \sqrt{\log n})$. In Chapter 3 we show that this conjecture is false and in fact the best bound that one can achieve is $O(\sqrt{\alpha \log n})$; we also generalize this bound to other ℓ_p spaces for $p > 1$.

For the case that $p = 1$, the work of Lee [66] still implies that an embedding of (X, d) into ℓ_1 exists with distortion $O(\sqrt{\alpha \log n})$. However, the situation for the lower bound is different from other ℓ_p spaces. Lee and Naor in [71] came up with a metric that embeds well at each scale based on the 3-dimensional Heisenberg group $\mathbb{H}^3$ [10], but requires super constant distortion to be embedded into ℓ_1. It was later shown by Cheeger and Kleiner [25, 27, 26] that this metric can not be embedded into ℓ_1 with constant distortion. Building on this analysis, it was recently proved in [28] that this construction achieves an integrality gap of $(\log n)^\delta$ for some small constant $\delta > 0$. One of the corollaries of the work presented in Section 3.4 is that there exists a family of metrics such that they admit a constant distortion embedding at each scale but they can not be embedded into ℓ_1 with distortion better than $\Omega\left(\frac{(\log n)^{1/3}}{\log \log n}\right)$. Later, this bound was improved to $\Omega\left(\frac{\sqrt{\log n}}{\mathsf{poly}(\log \log n)}\right)$ by Lee and Sidiropolous [74] using a different family of metrics.

1.3.3 Sparsest Cut, Negative Type Metrics and Half-snowflakes

We start this section by recalling the definition of *Sparsest Cut Problem.* Given a graph $G = (V, E)$, a symmetric non-negative demand function $\mathsf{dem} : V \times V \to [0, \infty)$, and a capacity function $\mathsf{cap} : E \to [0, \infty)$, one defines the *sparsity* of a cut $(S, \bar{S})$ for $S \subseteq V$ as

$$\Phi_G(S; \mathsf{cap}, \mathsf{dem}) = \frac{\mathsf{cap}(S, \bar{S})}{\mathsf{dem}(S, \bar{S})},$$

[10] This construction was proposed to prove lower bounds for a different problem, however it follows from [66] that their construction admits an embedding over scales with $\alpha = O(1)$.

where $f(S,T)$ denotes sum of $f(x,y)$ over all such that $(x,y) \in (S \times T) \cap \mathsf{domain}(f)$. The value of the sparsest cut is then given by:

$$\Phi_G(\mathsf{cap}, \mathsf{dem}) = \min\{\Phi_G(S; \mathsf{cap}, \mathsf{dem}) : S \subseteq V\}. \tag{1.4}$$

Moreover, we say that an instance is *uniform* if $\mathsf{dem}(u,v) = 1$ for all $u, v \in V$.

Finding the exact value of the sparsest cut is NP-hard [82], and finding a constant approximation algorithm for it, is *Unique Game*-hard [22]. Leighton and Rao [75] suggested a linear program relaxation of this problem. It was shown in [76, 8, 46] that the ratio between the solution to the linear program and the optimal solution is precisely $\sup_{M_n} c_1(M_n)$, where M_n ranges over all metric spaces on n-points. Bourgain's embedding theorem [12] shows that this $c_1(M_n) = O(\log n)$, and in [76, 8], it was shown that this bound is tight for the shortest path metric on expander graphs.

Goemans and Linial independently proposed the following relaxation of the problem and conjectured that this it is possible to approximate the value of the sparsest cut using the following SDP:

$$\min\left\{\frac{\sum_{u,v}\mathsf{cap}(u,v)\|x_u - x_v\|_2^2}{\sum_{u,v}\mathsf{dem}(u,v)\|x_u - x_v\|_2^2} : \{x_u\}_{u\in V} \subseteq \mathbb{R}^n \text{ and } \|\cdot\|_2^2 \text{ is a metric on } \{x_u\}_{u\in V}\right\}.$$

This optimization problem can be formulated as a semidefinite program:

Minimize:	$\sum_{(u,v)\in E}\mathsf{cap}(x,y)\|x_u - x_v\|_2^2$
subject to:	$\sum_{u,v\in V}\mathsf{dem}(u,v)\|x_u - x_v\|_2^2 = 1$
$\forall u,v,w \in V:$	$\|x_u - x_v\|_2^2 \le \|x_u - x_w\|_2^2 + \|x_w - x_v\|_2^2.$

In other words, we optimize over sets of n vectors $W \subseteq \mathbb{R}^n$ which satisfy, for every $x, y, z \in W$,

$$\|x - y\|_2^2 \le \|x - z\|_2^2 + \|z - y\|_2^2.$$

Similar to the case for the linear program, the approximation guarantee of this relaxation is exactly the solution to an embedding problem; the gap is precisely the supremum of $c_1(X,d)$ over all n-point metric spaces $(X,d) \in \mathsf{NEG}$.

The Goemans-Linial SDP was used in [6], to design an polynomial time $O(\sqrt{\log n})$-approximation algorithm for the *uniform* case of Sparsest Cut. Following an earlier bound of Chawla, Gupta and Räcke [21], and building on the technique from Arora, Rao and Vazirani [6], it was shown by Alon, Lee, and Naor [5] that any n-point space of negative type can be embedded into ℓ_1 with distortion $O(\sqrt{\log n}\log\log n)$. This in fact implies an $O(\sqrt{\log n}\log\log n)$-approximation algorithm for the general case of the Sparsest Cut Problem.

Following the footsteps of [6], metrics of negative type were used to solve various other optimization problems [41, 1, 52, 19]. Given the impact of negative type metrics on the field of approximation algorithms, understanding these metrics became a fundamental question. To this end, Khot and Vishnoi [53] used the tools from Kahn, Kalai, and Linial [51] and Bourgain [14] to construct a family of negative type metrics such that $c_1(X,d) = \omega(1)$, hence disproving Goemans-Linial conjecture. Subsequently, [58, 33] gave quantitative improvements of [53].

Lee and Naor proposed a different construction based on the 3-dimensional discrete Heisenberg group (denoted $\mathbb{H}^3$) in [71]. They showed that the natural metric supported on $\mathbb{H}^3$ can be embedded into NEG with constant distortion. Following a series of works from Cheeger and Kleiner [25, 27, 26] it was shown in [28] that the restriction of $\mathbb{H}^3$ to an $n \times n \times n$ box require $O(\log^{\delta_0} n)$ distortion to be embedded into ℓ_1, for some $\delta_0 > 0$. This is in fact the best lower bound known for embedding finite negative type metrics into ℓ_1.

One common factor in all the above constructions is that in all cases it is relatively easy to show that the space is an $O(1)$-half-snowflake, and most of the effort was put into proving that the metric in question is NEG. In the case of constructions based on [53], Kolla and Lee [56] provided a short proof that the metric is $O(1)$-half-snowflake. In the case of the Heisenberg group $\mathbb{H}^3$ (equipped with the Carnot-Caratheodory metric), the classical result of Assouad [7] implies that the metric is $O(1)$-half snowflake. Indeed, the fact that one could construct an $O(1)$-half-snowflake was taken as evidence that eventually a negative type metric (1-half-snowflake) could be constructed.

Moreover, it was well-know that the algorithm from Arora, Rao and Vazirani [6] and the follow-up works do not need exact triangle inequality condition of Goemans-Linial SDP. In fact, if we replace the triangle inequality in the Goemans-Linial SDP with

$$\forall u, v \in V, \forall P_{uv} : C\|x_u - x_v\|_2^2 \leq \sum_{(a,b) \in E(P_{uv})} \|x_a - x_b\|_2^2,$$

for some constant $C > 0$,[11] all the known rounding algorithms for sparsest cut would still provide the same approximation guarantee. We call this inequality *weak triangle inequality*. Incidentally, the approximation ratio of Goemans-Linial SDP with weak triangle inequality is exactly the same minimum distortion required to embed $\frac{1}{\sqrt{C}}$-half-snowflakes into ℓ_1.

These facts lead to the following question (see the "Isometric vs. isomorphic L_2 squared" problem in [80]).

Question 1. *Can any $O(1)$-half-snowflake metrics be embedded into NEG with constant distortion?*

A positive answer to this question could extremely simplify construction of "hard" instances for Goemans-Linial SDP.

In Chapter 3 we answer this question negatively and show that the answer is negative in a very strong sense. We show that there exists a family of $O(1)$-half-snowflake, for which any embedding of them into a metric of negative type has distortion at least $\Omega((\log n)^{1/3-o(1)})$. This work is based on a collaboration with James R. Lee [68] and an improvement based on a collaboration with Sachdeva [96].

1.3.4 Vertex Separators and Okamura-Seymour Theorem

The relation between Max-flow and Min-cut has played has played a fundamental role in the design of many approximation algorithms. Perhaps the best known and most fundamental theorem that connect the value of max-flow and min-cut is the classical Max-flow Min-cut Theorem of [43]. This theorem states that given a graph with two terminals and a set of capacities on the edges, the value of maximum flow between the terminals is equal to the value of the minimum cut that separates those terminals (see Figure 1.1). One may ask whether the same bound holds in the case that there are more than one pair of terminals. For

[11] This inequality with $C = 1$ is equivalent to triangle inequality.

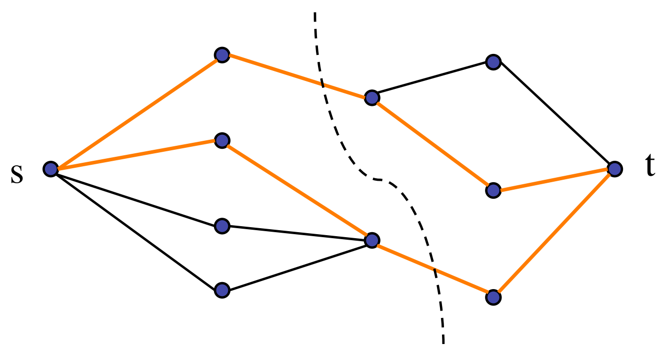


Figure 1.1: The value of the maximum flow equals to the value of the minimum cut.

an undirected graph $G = (V, E)$ together with a capacity function on edges $\mathsf{cap} : E \to [0, \infty)$, and a set of demands $\mathsf{dem} : V \times V \to [0, \infty)$. For $u, v \in V$, denote by $f_{uv} : E \to [0, \infty)$ the undirected u-v flow. The capacity constrains require that for every *feasible* flow and $e \in E$, $\sum_{u,v \in V} f_{uv}(e) \le \mathsf{cap}(e)$. Given such an instance, let $\mathsf{mcf}(G; \mathsf{cap}, \mathsf{dem})$ be the largest value δ such that one can simultaneous route $\delta \cdot \mathsf{dem}(u, v)$ units of flow between u and v for every $u, v \in V$ while not violating any of the edge capacities. This optimization describes the *maximum concurrent flow problem.*

Now, recall the definition of the sparsest cut from (1.4). It is straightforward to check that $\mathsf{mcf}_G(\mathsf{cap}, \mathsf{dem}) \le \Phi_G(S; \mathsf{cap}, \mathsf{dem})$ (see Figure 1.2). However, $\mathsf{mcf}_G(\mathsf{cap}, \mathsf{dem}) \ge \Phi_G(\mathsf{cap}, \mathsf{dem})$ does not necessarily hold. In fact, it follows from [76] and [46] that for graphs with n vertices, there can be a gap of $\Omega(\log n)$ between the bound that follows from cuts and the amount of flow that can be simultaneously routed in the graph. In other words, there are flow instances such that $\mathsf{mcf}_G(\mathsf{cap}, \mathsf{dem}) \lesssim \left(\frac{\Phi_G(\mathsf{cap}, \mathsf{dem})}{\log n}\right)$. It was shown in [91] that this gap is at most $O(\sqrt{\log n})$ for planar graphs and the question of whether this bound can be improved to $O(1)$ for planar graph is still open. For the special case where all support of dem lies on a single face of the graph, the classical Okamura-Seymour theorem [89] states that the value of maximum concurrent flow is equal to value of the sparsest cut.

We remark that Okamura-Seymour theorem has many applications beyond providing a bound on the gap between maximum concurrent flow and the sparsest cut. For instance,

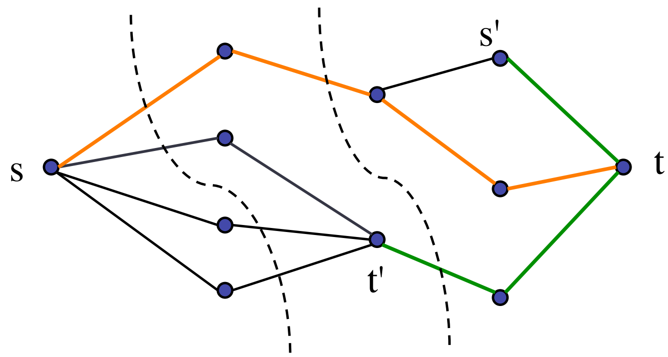


Figure 1.2: Each cut provides an upper bound on the amount of concurrent flow that can be sent across the cut in the network.

in [30] this theorem was used as a fundamental tool in solving the edge-disjoint paths problem in planar graphs with constant congestion. In fact, one of the motivations behind generalizing this theorem to the node-capacitated setting[12] is to make an step towards finding an algorithm for vertex-disjoint paths in planar graphs with constant congestion.

We study the node-capacitated networks in this dissertation. We also consider the case of submodular flows (see [29] for some of the applications of submodular flows). Note that, while one can simulate edge capacities by introducing a new vertex in the middle of an edge, it does not seem that any reduction is known by which one can simulate vertex capacities with edge capacities. Formally, we define a *vertex-capacitated flow network* by considering a function $\mathsf{cap} : V \to [0, \infty)$ assigning capacities to vertices rather than edges. The best way of to think about is as follows: If a flow of value α is sent along a path P from s to t, then it consumes $\alpha/2$ capacity at s and t and α capacity at each of the intermediate nodes of P. While, we give some justifdcation for the *boundary conditions* on s and t, we remind the readers that this particular choice does not affect any theorem in Chapter 5 which deals with approximate flow/cut gaps.

The corresponding definition of the maximum concurrent flow for node-capacitated

[12] It is a generalization because we can simulate an edge-capacitated instance with a node-capacitated instance by introducing a new vertex in the middle of each edge with the same capacity as the original edge.

graphs follows immediately; we use the notation mcf^v_G for the vertex-capacitated version. For the definition of Φ^v_G, we have to be slightly more careful. For a subset $S \subseteq V$ of the vertices, denote by $G[S]$ the induced subgraph of G on S. We define a function $\rho_S : V \times V \to \{0, \frac{1}{2}, 1\}$ by

$$\rho_S(u,v) = \begin{cases} \frac{1}{2} & |\{u,v\} \cap S| = 1 \\ 1 & u,v \in S \\ 1 & u,v \in \bar{S} \text{ and } u,v \text{ are in distinct connected components of } G[\bar{S}] \\ 0 & \text{otherwise.} \end{cases}$$

In other words, we are only given half-credit for separating u and v if exactly one of them is in the separator. Then we define

$$\Phi^v_G(S; \mathsf{cap}, \mathsf{dem}) = \frac{\sum_{v \in S} \mathsf{cap}(v)}{\sum_{u,v \in V} \mathsf{dem}(u,v) \rho_S(u,v)},$$

and $\Phi^v_G(\mathsf{cap}, \mathsf{dem}) = \min_{S \subseteq V} \Phi^v_G(S; \mathsf{cap}, \mathsf{dem})$. It is straightforward to verify that

$$\mathsf{mcf}^v_G(\mathsf{cap}, \mathsf{dem}) \leq \Phi^v_G(\mathsf{cap}, \mathsf{dem}).$$

These definitions ensure that a classical Max-flow Min-cut theorem of Menger [86] when the demand is supported on a single pair. They also allow other natural properties in the multi-commodity setting. In particular, this choice of boundary conditions in the definition ensures that for any simple path P, $\mathsf{mcf}^v_P(\mathsf{cap}, \mathsf{dem}) = \Phi^v_P(\mathsf{cap}, \mathsf{dem})$. It is an easy exercise to generalize this fact to trees (for any tree T we have $\mathsf{mcf}^v_T(\mathsf{cap}, \mathsf{dem}) = \Phi^v_T(\mathsf{cap}, \mathsf{dem})$). This bound for the trees is an essential part of the proofs presented in Chapter 5.

Unfortunately, unlike trees and paths, there is no exact vertex-capacitated analog of the Okamura-Seymour Theorem. This is demonstrated using Figure 1.3. The planar graph in Figure 1.3 has all vertices on the outer face. The capacities are specified on the vertices and the demands are given by dotted edges in the figure; all demands have value 1. It is straightforward to check that one has $\Phi^v_G(\mathsf{cap}, \mathsf{dem}) = 1$ and yet $\mathsf{mcf}^v_G(\mathsf{cap}, \mathsf{dem}) = 5/7$. The shaded nodes form a vertex cut of sparsity 1.

Nevertheless, we show that an approximate version of Okamura-Seymour Theorem does hold in the vertex-capacitated setting, answering a question posed by Chekuri and Kawarabayashi.

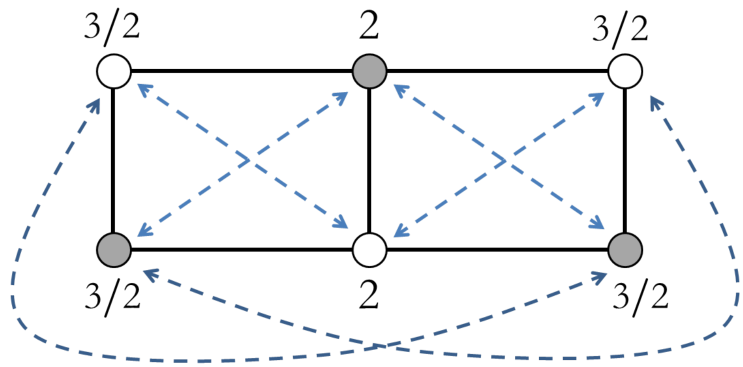


Figure 1.3: A counterexample to an exact node-capacitated Okamura-Seymour Theorem.

Theorem 1.2 (Approximate Okamura-Seymour Theorem)**.** *There exists a constant $\delta > 0$ such that the following holds. Let $G = (V, E)$ be a planar graph and let $F \subseteq V$ be any face of G. Then for any vertex capacities $\mathsf{cap} : V \to [0, \infty)$ and any demands $\mathsf{dem} : V \times V \to [0, \infty)$ supported on F, we have*

$$\mathsf{mcf}^v_G(\mathsf{cap}, \mathsf{dem}) \geq \delta \cdot \Phi^v_G(\mathsf{cap}, \mathsf{dem})\,.$$

Our result holds in the more general setting of undirected polymatroid networks which we discuss in Chapter 5.

1.4 Bibliographical Notes

The results in Chapter 2 were developed with my co-authors Lee and deMesmay [69, 63]. While Section 2.3 gives an overview of the techniques developed in these works, Section 2.4 is taken from [63], and Section 2.4 solely focus on [69].

Chapter 3 is base on my collaborations with Jaffe and Lee [68, 49]. The results and proofs on the gluing Lemma first appeared in [68]. Section 3.4 is based on [68] and an improvement from a collaboration with Sachdeva [96]. The content of Section 3.2.1 is also modification (strengthening) of the graph product presented in [65].

Contents of Chapter 4 are based on my collaboration with Lee and Mendel [67].

Chapter 2

DIMENSION REDUCTION FOR TREES IN ℓ_1

2.1 Organization

In this chapter we study the dimension reduction for tree metrics in ℓ_1. We first state all the results that we prove in Section 2.2. We give an overview of the techniques that we use in the proofs in Section 2.3. Then, in Section 2.4 we present the proofs for the upper bounds and in Section 2.5 we present the proofs for the lower bounds.

2.2 Results

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[1],

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

Such a metric space (V, d_T) is called a *finite tree metric.*

The main result that we prove in this Chapter is the following Theorem.

Theorem 2.1. *For every $\varepsilon > 0$ and $n \in \mathbb{N}$, the following holds. Every n-point tree metric admits a $(1+\varepsilon)$-embedding into ℓ_1^k with $k = O((\frac{1}{\varepsilon})^4 \log \frac{1}{\varepsilon} \log n)$.*

In the special case of d-ary trees, we prove a stronger bound.

Theorem 2.2. *Let $T_{d,h}$ be the unweighted, complete d-ary tree of height h. For every $\varepsilon > 0$, there exists a $(1+\varepsilon)$-embedding of $T_{d,h}$ into $\ell_1^{O((h \log d)/\varepsilon^2)}$.*

We remark that the proofs for these theorems also yields randomized polynomial-time algorithms to construct the embeddings.

---

[1]This is a pseudometric because we may have $d(u, v) = 0$ even for distinct $u, v \in V$.

We also prove some lower bounds on the dimension required for embedding trees into ℓ_1^k. For $n \in \mathbb{N}$, the n-star graph is a tree on n nodes such that one node has degree $n-1$ and all the other nodes have degree one. In this chapter we prove the following theorem.

Theorem 2.3. *For $\varepsilon \in (0, \frac{1}{16})$ and $n \geq 1/\varepsilon^2$, any embeddings of n-star into ℓ_1^d with distortion $1+\varepsilon$ requires dimension $d = \Omega\left(\frac{\log(n)}{\varepsilon^2 \log(1/\varepsilon)}\right)$.*

Corollary 2.4. *There exists a universal constant δ, such that, for any $\varepsilon \in (0, \delta)$, $n \geq N_\varepsilon$ and $k > 1$, any embedding of a complete k-ary tree of size n into ℓ_1^d with distortion $1+\varepsilon$ requires dimension $d = \Omega\left(\frac{\log(n)}{\varepsilon^2 \log(1/\varepsilon)}\right)$.*

2.3 Techniques

The techniques that are used in proving the lower bound for the dimension required for embedding trees into ℓ_1 are mostly elementary and the proof mostly consists of non-trivial applications of Markov inequality. In this section we mainly focus on the techniques that are used in proving the upper bound. We first discuss the form that all our embeddings will take in this Chapter. Let $T = (V, E)$ be a finite, connected tree, and fix a root $r \in V$. For each $v \in V$, recall from Section 1.2.1 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,

$$\varphi(x) = \sum_{e \in E(P_v)} \lambda(e). \tag{2.1}$$

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

$$\tau_\lambda(v) = (\lambda(e_1), \lambda(e_2), \ldots, \lambda(e_k), 0, 0, \ldots, 0), \tag{2.2}$$

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 (2.1), 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 (2.2) 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 Section 2.4.1, we show that a judicious application of the Lovász Local Lemma [37] 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 [95].

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 [95] 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

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

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

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

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 Section 2.4.5.

Scale selection. The first step beyond complete binary trees would be in passing to complete d-ary trees for $d \ge 3$. 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 (2.2) is 1-Lipschitz).

Observing the case of d-ary trees, one might be tempted to use a function based on "growth rate" to choose the scale

$$\kappa(e) = \sup_{r>0} \left\lceil \frac{\log |B(u,r)|}{r} \right\rceil, \tag{2.3}$$

where $e = (u,v)$ is directed away from the root and $B(u,r)$ denotes the ball of radius r around u. 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 (See Figure 2.1).

In fact this example implies that κ can not be local. There are also examples which show that it is impossible to choose $\kappa(u,v)$ to depend only on the geometry of the subtree rooted at u (see Figure 2.2). 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

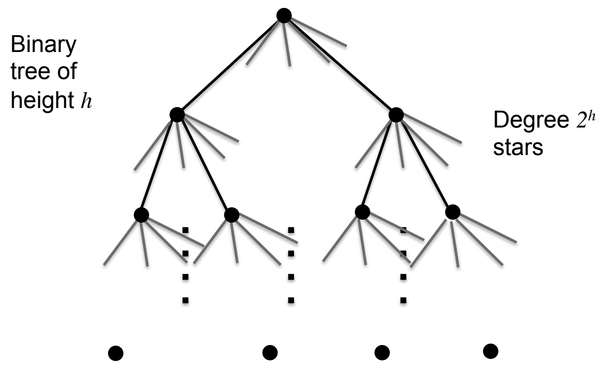


Figure 2.1: Counter example for the approach based on (2.3).

significantly into the details of the proof. For our purposes, we need to consider more general embeddings of type (2.1). 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.

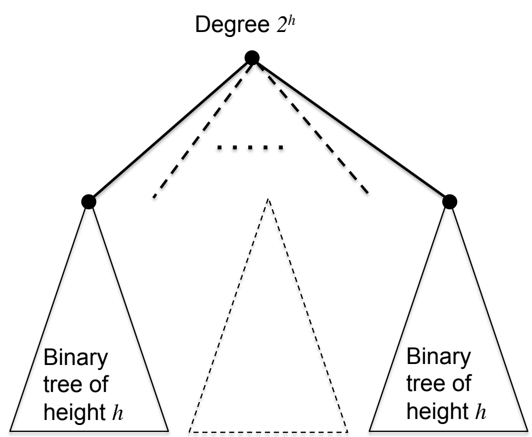


Figure 2.2: Counter example showing that $\kappa(e)$ can not only depend on the subtrees under the edges e.

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 $i \in \{1, 2, \ldots, k\}$ 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 Section 2.4.3. The actual scale selectors are defined in Section 2.4.4.

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 [77, 45, 72], as well as for preivous low-dimensional embeddings into ℓ_1 [20]. The following lemma is well-known and elementary. See Figure 2.3 for an example.

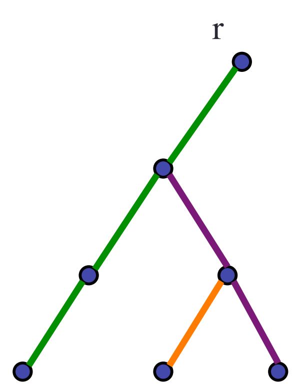


Figure 2.3: Monotone coloring of a tree.

Lemma 2.5. *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 $e_1, e_2, \ldots, e_k$, we may color the edges of $T_{e_1}, T_{e_2}, \ldots, T_{e_k}$ 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 that 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 Section 2.4.2.

2.4 Upper Bounds

We first prove Theorem 2.2, in Section 2.4.1. Then in Sections 2.4.2 and 2.4.4 we build the necessary tools to construct and analyze the embedding corresponding to Theorem 2.1, and in Section 2.4.5 we complete he proof by describing the embedding and analyzing its distortion.

2.4.1 Embedding of 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 $j = 1, 2, \ldots$

Theorem 2.6. *[Restatement of Theorem 2.2] 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 Section 2.4.1, 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 from [83].

Theorem 2.7. *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,*

$$\mathbb{P}(X - \mathbb{E}(X) \ge \lambda) \le \exp\left(\frac{-\lambda^2}{2(\mathrm{Var}(X) + M\lambda/3)}\right). \tag{2.4}$$

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,

$$t = \left\lceil \frac{1}{\varepsilon} \right\rceil, \tag{2.5}$$

and

$$m = t\lceil \log k \rceil. \tag{2.6}$$

Let $\{\vec{v}(1), \ldots, \vec{v}(t)\}$, be the standard basis for $\mathbb{R}^t$. Let $b_1, b_2, \ldots, b_m$ be chosen i.i.d. uniformly over $\{1, 2, \ldots, t\}$. 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),

$$\frac{1}{m}\begin{pmatrix} \vec{v}(b_1) \\ \vdots \\ \vec{v}(b_m) \end{pmatrix}. \tag{2.7}$$

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,

$$g(v) = \begin{pmatrix} \lambda(e_1) \\ \vdots \\ \lambda(e_j) \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{2.8}$$

where $e_1, e_2, \ldots, e_j$ 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 2.8. *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 2.9. *For $C \geq 10$, and $x, y \in V$,*

$$\mathbb{P}\Big[\|g(x) - g(y)\|_1 \leq (1 - C\varepsilon) d_T(x,y)\Big] \leq k^{-C d_T(x,y)/2}. \tag{2.9}$$

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 $a_{ij} = (i-1)m + j$, and then

$$X_{ij} = \|g(x)[a_{ij}] - g(r')[a_{ij}]\|_1 + \|g(y)[a_{ij}] - g(r')[a_{ij}]\|_1 - \|g(x)[a_{ij}] - g(y)[a_{ij}]\|_1. \tag{2.10}$$

Using the definition of g in (2.8), we can write

$$\begin{aligned}\|g(x)-g(y)\|_1 &= \sum_{i\in[h-1],j\in[m]} \left(\|g(x)[a_{ij}]-g(r')[a_{ij}]\|_1 + \|g(y)[a_{ij}]-g(r')[a_{ij}]\|_1 - X_{ij}\right)\\ &= \|g(x)-g(r')\|_1 + \|g(y)-g(r')\|_1 - \sum_{i\in[h-1],j\in[m]} X_{ij}\\ &\overset{(2.8)}{=} d_T(x,r') + d_T(y,r') - \sum_{i\in[h-1],j\in[m]} X_{ij}\\ &= d_T(x,y) - \sum_{i\in[h-1],j\in[m]} X_{ij}\,.\end{aligned}$$

We will prove the lemma by arguing that,

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

First, observe that if $i \le d_T(r,r')$ then $X_{ij}=0$ for all $j\in[m]$ since all three terms in (2.10) are zero. Furthermore, if $i \ge \min(d_T(r,x),d_T(r,y))+1$, then again $X_{ij}=0$ for all $j\in[m]$, since in this case one of the first two terms of (2.10) are 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,

$$|R| = \min(d_T(r,x),d_T(r,y)) - d_T(r,r') \le \frac{d_T(x,y)}{2}\,. \tag{2.11}$$

We continue the proof by bounding the maximum of the X_{ij} variables. Since, for every a, we have

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

we conclude that,

$$\max\left\{X_{ij} : i\in[h-1], j\in[m]\right\} \le \frac{2}{m}. \tag{2.12}$$

For $i\in R$ and $j\in[m]$, using (2.7) and (2.8), we see that $(g(x)[a_{ij}]-g(r')[a_{ij}]) = \frac{1}{m}\vec{v}(\alpha)$ and $g(y)[a_{ij}]-g(r')[a_{ij}] = \frac{1}{m}\vec{v}(\beta)$, 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 (2.12),

$$\mathbb{E}[X_{ij}] \leq \frac{2}{tm}, \tag{2.13}$$

and

$$\mathrm{Var}(X_{ij}) \leq \frac{4}{tm^2}. \tag{2.14}$$

Using (2.11), we have

$$\sum_{i=1}^{h-1}\sum_{j=1}^{m} \mathbb{E}[X_{ij}] = \sum_{i\in R}\sum_{j\in[m]} \mathbb{E}[X_{ij}] \overset{(2.13)}{\leq} \sum_{i\in R} \frac{2}{t} \overset{(2.11)}{\leq} \frac{d_T(x,y)}{t}, \tag{2.15}$$

and

$$\sum_{i=1}^{h-1}\sum_{j=1}^{m} \mathrm{Var}(X_{ij}) = \sum_{i\in R}\sum_{j\in[m]} \mathrm{Var}(X_{ij}) \overset{(2.14)}{\leq} \sum_{i\in R} \frac{4}{tm} \overset{(2.11)}{\leq} \frac{2\, d_T(x,y)}{tm}. \tag{2.16}$$

We now apply Theorem 2.7 to complete the proof:

$$\begin{aligned}
&\mathbb{P}\left[\sum_{i\in[h-1],j\in[m]} X_{ij} \geq C\left(\frac{d_T(x,y)}{t}\right)\right] \\
&\qquad = \mathbb{P}\left[\sum_{i\in[h-1],j\in[m]} X_{ij} - \frac{d_T(x,y)}{t} \geq (C-1)\left(\frac{d_T(x,y)}{t}\right)\right] \\
&\qquad \overset{(2.15)}{\leq} \mathbb{P}\left(\sum_{i\in[h-1],j\in[m]} X_{ij} - \mathbb{E}\left[\sum_{i\in[h-1],j\in[m]} X_{ij}\right] \geq (C-1)\left(\frac{d_T(x,y)}{t}\right)\right) \\
&\qquad \leq \exp\left(\frac{-((C-1)d_T(x,y)/t)^2}{2\left(\sum_{i\in[h-1],j\in[m]} \mathrm{Var}(X_{ij}) + (C-1)(d_T(x,y)/t)(\frac{2}{m})/3\right)}\right) \\
&\qquad \overset{(2.16)}{\leq} \exp\left(\frac{-((C-1)d_T(x,y)/t)^2}{2\left(2\, d_T(x,y)/(tm) + (C-1)(d_T(x,y)/t)(\frac{2}{m})/3\right)}\right) \\
&\qquad = \exp\left(\frac{-(C-1)^2}{4\,(1+(C-1)/3)} \cdot \frac{m}{t} \cdot d_T(x,y)\right).
\end{aligned}$$

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,

$$\begin{aligned}
\mathbb{P}\left[\sum_{i\in[h-1],j\in[m]} X_{ij} \geq C\varepsilon d_T(x,y)\right] &\overset{(2.5)}{\leq} \mathbb{P}\left[\sum_{i\in[h-1],j\in[m]} X_{ij} \geq C\left(\frac{d_T(x,y)}{t}\right)\right] \\
&\leq \exp\left(-\frac{Cm}{2t} d_T(x,y)\right) \\
&\overset{(2.6)}{\leq} k^{-C d_T(x,y)/2}
\end{aligned}$$

completing the proof. □

The Local Lemma Argument

We first give the statement of the Lovász Local Lemma [37] and then use it in conjunction with Lemma 2.9 to complete the proof of Theorem 2.6.

Theorem 2.10. *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*

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

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 2.6. We may assume that $k \geq 2$. We will use Theorem 2.10 and Lemma 2.9 to show that with non-zero probability the following inequality holds for all $u, v \in V$,

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

For $u, v \in V$, let $\mathcal{E}_{uv}$, be the event $\{\|g(u) - g(v)\|_1 \leq (1 - 14\varepsilon)\, d_T(u,v)\}$. 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 2.10, it is sufficient to show that for all $u, v \in V$,

$$\mathbb{P}(\mathcal{E}_{uv}) \leq x_{uv} \prod_{\substack{s,t \in V: \\ E(P_{st}) \cap E(P_{uv}) \neq \emptyset}} (1 - x_{st})\,. \tag{2.17}$$

Indeed, this will complete the proof of Theorem 2.6.

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

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

Since T is a k-ary tree,

$$|S_{e,i}| \le \sum_{j=1}^{i} k^{j-1} \cdot k^{i-j} = i \cdot k^{i-1} \le k^{2i}. \tag{2.18}$$

Thus we can write,

$$\begin{aligned}
x_{uv} \prod_{\substack{s,t\in V:\\ E(P_{st})\cap E(P_{uv})\neq\emptyset}} (1-x_{st}) &= x_{uv} \prod_{e\in E(P_{uv})} \prod_{i\in\mathbb{N}} \prod_{(s,t)\in S_{e,i}} (1-x_{st}) \\
&= k^{-3d_T(u,v)} \prod_{e\in E(P_{uv})} \prod_{i\in\mathbb{N}} \prod_{(s,t)\in S_{e,i}} \left(1-k^{-3i}\right) \\
&\overset{(2.18)}{\ge} k^{-3d_T(u,v)} \prod_{e\in E(P_{uv})} \prod_{i\in\mathbb{N}} \left(1-k^{-3i}\right)^{k^{2i}} \\
&\ge k^{-3d_T(u,v)} \prod_{e\in E(P_{uv})} \prod_{i\in\mathbb{N}} \left(1-k^{2i}(k^{-3i})\right) \\
&= k^{-3d_T(u,v)} \prod_{e\in E(P_{uv})} \prod_{i\in\mathbb{N}} \left(1-\frac{1}{k^i}\right).
\end{aligned}$$

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

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

Since $k \ge 2$, we conclude that,

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

On the other hand, Lemma 2.9 applied with $C = 14$ gives,

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

yielding (2.17), and completing the proof.

□

2.4.2 Colors and Scales

In the present section, we develop some tools for our eventual embedding. The proof of our main theorem appears in here, but relies on a key theorem which is only proved in Section 2.4.5.

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 $\mathsf{len} : E \to [0, \infty)$. 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

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

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,

$$\rho_\chi(x, y; \delta) = \mathsf{len}_\chi(C(x, y; \delta)) \,. \tag{2.19}$$

See Figure 2.4 for an example.

We now state a key theorem that will be proved in Section 2.4.5.

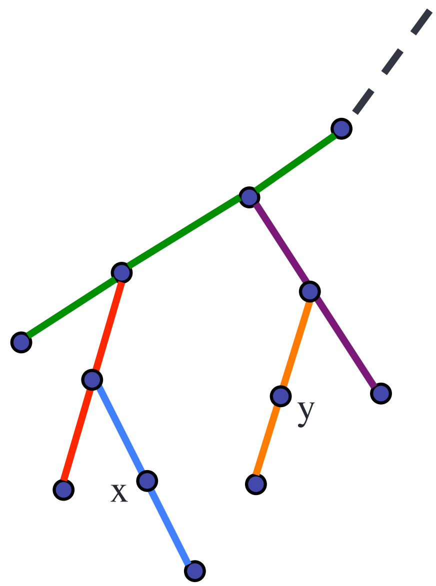


Figure 2.4: The graph distance between x and y is 5, but $\mathsf{len}_\chi(C(x,y;1/2)) = 8$ getting contribution from Red, Purple, Blue and Orange color classes

Theorem 2.11. *For every $\varepsilon, \delta > 0$, there is a value $C(\varepsilon,\delta) = O((\frac{1}{\varepsilon} + \log\log\frac{1}{\delta})^3 \log\frac{1}{\varepsilon})$ 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 $x, y \in V$,*

$$(1-\varepsilon)\, d_T(x,y) - \delta\, \rho_\chi(x,y;\delta) \le \|F(x) - F(y)\|_1 \le d_T(x,y)\,. \tag{2.20}$$

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 2.12. *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,

$$\frac{1}{k}\sum_{i=1}^{k} d_{T_i}(f_i(x), f_i(y)) \geq (1-\varepsilon)\, d_T(x,y)\,. \tag{2.21}$$

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

$$d_{T_i}(f_i(x), f_i(y)) \leq (1+\varepsilon)\, d_T(x,y)\,. \tag{2.22}$$

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

$$\varepsilon\, d_T(x,y) \geq \frac{2^{-(k+1)}}{k} \sum_{\substack{i=1 \\ i\neq j}}^{k} \rho_{\chi_i}(f_i(x), f_i(y); 2^{-(k+1)}) \tag{2.23}$$

Using Lemma 2.12 in conjunction with Theorem 2.11, we can now prove the main theorem (Theorem 2.1).

Proof of Theorem 2.1. 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 $M(\chi) \leq 1 + \log n$, which exists by Lemma 2.5. Apply Lemma 2.12 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 $M(\chi_i) \leq 1 + \log n$ 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 2.11 to T_i and χ_i, for each $i \in [k]$, with $\delta = 2^{-(k+1)}$, where $C(\varepsilon) = O(\frac{1}{\varepsilon^3}(\log \frac{1}{\varepsilon}))$. Finally, we put

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

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 2.11). In conjunction with condition (b) of Lemma 2.12 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 2.12. Then we have,

$$\begin{aligned}
\|F(x)-F(y)\|_1 &= \frac{1}{k}\sum_{i=1}^{k}\|(F_i\circ f_i)(x)-(F_i\circ f_i)(y)\|_1 \\
&\overset{(2.20)}{\geq} \frac{1}{k}\sum_{i\neq j}\Big((1-\varepsilon)\,d_{T_i}(f_i(x),f_i(y)) - 2^{-(k+1)}\rho_{\chi_i}(f_i(x),f_i(y);2^{-(k+1)})\Big) \\
&\overset{(2.23)}{\geq} \left(\frac{1}{k}\sum_{i\neq j}(1-\varepsilon)\,d_{T_i}(f_i(x),f_i(y))\right) - \varepsilon\, d_T(x,y) \\
&\geq \left(\frac{1}{k}\sum_{i=1}^{k}(1-\varepsilon)\,d_{T_i}(f_i(x),f_i(y))\right) - \frac{1}{k}\,d_{T_j}(f_j(x),f_j(y)) - \varepsilon\, d_T(x,y) \\
&\overset{(2.22)}{\geq} \left(\frac{1}{k}\sum_{i=1}^{k}(1-\varepsilon)\,d_{T_i}(f_i(x),f_i(y))\right) - \frac{1+\varepsilon}{k}\,d_T(x,y) - \varepsilon\, d_T(x,y) \\
&\overset{(2.21)}{\geq} (1-\varepsilon)^2\,d_T(x,y) - \frac{1+\varepsilon}{k}\,d_T(x,y) - \varepsilon\, d_T(x,y) \\
&\geq (1-O(\varepsilon))\,d_T(x,y)
\end{aligned}$$

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 2.12. 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 2.13. *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 $x, y \in [0, \infty)$, $\frac{1}{k}\sum_{i=1}^{k} d_{S_i}(f_i(x), f_i(y)) \geq \left(1 - \frac{7}{k}\right)|x - y|$.

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*

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

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:

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

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

$$\|f_i\|_{\mathrm{Lip}} \leq (1 - 2^{1-k})^{-1} \leq 1 + 2^{-k+1} \tag{2.24}$$

yielding condition (iii).

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

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

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

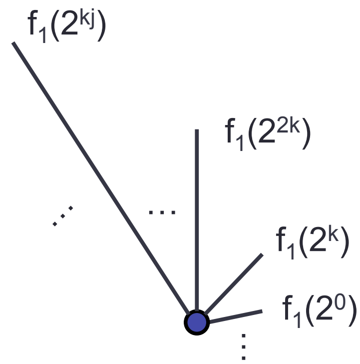


Figure 2.5: Image of the real line in S_1.

$$
\begin{aligned}
\sum_{i=1}^{k} d_{S_i}(f_i(x), f_i(0)) &\geq \sum_{l=\lfloor \log_2 x \rfloor - k+2}^{\lfloor \log_2 x \rfloor} x - 2^l \\
&= (k-1)x - \sum_{l=\lfloor \log_2 x \rfloor - k+2}^{\lfloor \log_2 x \rfloor} 2^l \\
&\geq (k-1)x - 2^{\lfloor \log_2 x \rfloor + 1} \\
&\geq (k-3)x. \qquad (2.25)
\end{aligned}
$$

Now fix $x, y \in [0, \infty)$ with $x \leq y$. If $x \leq y/2$, then we can use the triangle inequality, together with (2.24) and (2.25) to write,

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

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

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

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

$$\frac{1}{k} \sum_{i=1}^{k} d_{S_i}(f_i(x), f_i(y)) \geq \frac{1}{k} \sum_{l=0}^{k-3} (y-x) \geq \frac{k-2}{k}(y-x), \tag{2.26}$$

yielding condition (ii).

It is also straightforward to check that

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

which verifies condition (iv).

To verify condition (v), note that for $x \in [0, \infty)$, the inequality $d_{S_i}(f_i(x), f_i(0)) \leq x/2$ 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),

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

In the case that $y/2 < x \leq y$, $f_i(x)$ 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 2.12.

Proof of Lemma 2.12. We put $k = \lceil 7/\varepsilon \rceil$ and prove the following stronger statement by induction on $|V|$: There exist metric trees $T_1, T_2, \ldots, T_k$ 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 $v \in V$,

$$|\chi_i(P_{f_i(v)})| \leq |\chi(P_v)|\,. \tag{2.27}$$

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 $T_1', T_2', \ldots, T_k'$ with colorings $\chi_i' : E(T_i') \to \mathbb{N}$ and mappings $f_i' : V(T') \to V(T_i')$.

Now, let $S_1, \ldots, S_k$ and $\{g_i : [0, \infty) \to S_i\}$ be the $\mathbb{R}$-stars and mappings guaranteed by Lemma 2.13. 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 straight forward to verify that (2.27) 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 (2.27) 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 2.13 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 $i \neq j$,

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

Since there are at most two colors on the path between x and y in any T_i, by condition (v) of Lemma 2.13, there are at most four values of $i \in [k] \setminus \{j\}$ such that

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

hence

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

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,

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

Towards verifying condition (c), note that by condition (v) from Lemma 2.13, there are at most two values of i, such that

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

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

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

Now we use condition (iv) from Lemma 2.13 to conclude,

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

completing the proof. □

2.4.3 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 2.14. *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*

$$\sum_{i\in\mathbb{Z}} 2^i |f_i(x) - f_i(y)| < \infty \,. \tag{2.28}$$

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

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

where

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

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,

$$F_i(x) = \sum_{j\in\mathbb{Z}} 2^{jk+i}(f_{jk+i}(x) - f_{jk+i}(x_0)) \,. \tag{2.29}$$

It is easy to see that (2.28) 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,

$$|F_i(x) - F_i(y)| \ge (1-\varepsilon)\sum_{j\in\mathbb{Z}} \left(2^{jk+i}|f_{jk+i}(x) - f_{jk+i}(y)|\right) - 2\zeta_i(x,y). \tag{2.30}$$

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,

$$|F_i(x) - F_i(y)| = \left| \sum_{j\in S} 2^{jk+i}(f_{jk+i}(x) - f_{jk+i}(y)) + \sum_{j\in T} 2^{jk+i}(f_{jk+i}(x) - f_{jk+i}(y)) \right| \,.$$

If $S \cup T = \emptyset$, then (2.30) is immediate. Now, suppose that $S \neq \emptyset$, and let $c = i + k \cdot \max(S)$. Observe that $\max(S)$ exists by (2.28).

We then have,

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

On the other hand,

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

Therefore,

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

completing the verification of (2.30) 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 (2.30) vacuous since the right-hand side is non-positive.

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

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

This completes the proof. □

In Section 2.4.5, we will require the following straightforward corollary.

Corollary 2.15. *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 $x, y \in X$,

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

where

$$\zeta(x,y) = \sum_{k=1}^{m} \sum_{\substack{i : \exists j < i \\ f_j(x)_k - f_j(y)_k \neq 0}} 2^i (|f_i(x)_k - f_i(y)_k| - \lfloor |f_i(x)_k - f_i(y)_k| \rfloor), \tag{2.31}$$

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

2.4.4 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 Section 2.4.5 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 parent in the tree.

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 $T(c_0) = T$.

Before we describe the scale selector that we use in this Section, we give the intuition behind our choice of seemingly complicated scale selector.

Intuition

Here, we explain why the scale selector that we introduce in this section is a natural generalization of the scale selector for k-ary trees from Section 2.4.1. Note that in the embedding of k-ary trees in Section 2.4.1, one can assume that each edge has unique color, and each vertex has scale $\Theta(\frac{1}{\log k})$. Unfortunately trivial generalizations, such as setting the scale to $\log(\deg(v))$ or $\log\left(\frac{|V(T(v))|}{|V(T(\rho(v)))|}\right)$ do not result in a low distortion embedding (as was explained in Section 2.3).

For the sake of simplicity we start by generalizing the k-ary trees to the case that each edge has unit weight and the total depth of the tree is logarithmic. In this case we can still assume that each edge has a unique color and we assign each vertex a single scale rather than a set of scales. Our approach is based on the following principle: Assign the smallest possible scale 2^{-i} to each vertex v so that the total number of vertices on the path from root to v has roughly $(\log |V(T)| - \log |V(T(v))|)/2^i$ vertices at scale 2^{-i}.

On the one hand, this condition allows us to assign 2^i coordinates to the edges at scale $O(2^{-i})$ while preserving the following condition: For any scale the set of coordinates assigned to different edges of any root leaf path is disjoint. Moreover, we do not assign all the small scales to the vertices close to the root. In fact, for any vertex $v \in V(T)$, there will be roughly $\log |V(T(v))|$ coordinates at scale 2^{-i} untouched in the path from root to v and those coordinates can be used for embedding of the subtree under v.

On the other hand, the fact that we assign the smallest possible scale to each edge allows us to derive a concentration bound similar to Theorem 2.9 from Section 2.4.1. However, unlike the k-ary case a direct application of Theorem 2.7 does not yield the desired bound. In particular, we cannot show that the distance between any two points is concentrated around its mean value, however for any two vertices $u, v \in V(T)$, it is possible to show that there are $(1 - \varepsilon)d_T(u, v)$ edges on the path between u and v such that the contribution of those edges to the distance of u and v is concentrated. This can be done by showing that the scales of most of the edges on the path between u and v are "small," using a simpler version of Lemma 2.26.

We can simply extend this approach to the case that edges are weighted and the depth of the tree is logarithmic, by assigning the scales as follows: Assign the smallest possible scale 2^i for $i \in \mathbb{Z}$ to each vertex v so that for edges on the path from root to v at scale 2^i, $\sum \mathsf{len}(e)/2^i$ is roughly $\log |V(T)| - \log |V(T(v))|$. Indeed, all the arguments from the unweighted case can be generalized to the weighted case with little effort.

The main challenge for extending this approach to the trees with large depth is that we cannot assign a fraction of a coordinate to a vertex (or edge). In fact, this approach fails to produce a low dimensional embedding even for a simple path. To overcome this problem we use Lemma 2.5 to bound the topological depth of the tree. However, this raises another challenge. Namely, each color class contains more than one vertex and it is not clear how to assign a single scale to vertices inside a color class. Our final scale selector provides a way to get around this problem by finding a consistent way to assign sets of scales to vertices in each color class, while following the principle that the smallest possible set of scales is assigned to each vertex.

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,

$$\kappa(c) = \left\lfloor \log_2 \frac{|E(T(\rho(c)))|}{|E(T(c))|} \right\rfloor + 1. \tag{2.32}$$

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

$$\varphi(c) = \kappa(c) + \varphi(\rho(c)), \tag{2.33}$$

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

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

$$\varphi(c) = \sum_{c' \in \chi(E(P_{v_c})) \cup \{c\}} \kappa(c') \leq \sum_{c' \in \chi(E(P_{v_c})) \cup \{c\}} \left(1 + \log_2 \frac{|E(T(\rho(c')))|}{|E(T(c'))|}\right) \leq M(\chi) + \log_2 |E|. \tag{2.34}$$

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

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

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

$$\tau_i(v) = \min\left(\underbrace{\left\lceil \frac{d_T(v, v_c) - \min\left(d_T(v, v_c), \sum_{j=-\infty}^{i-1} 2^j \tau_j(v)\right)}{2^i} \right\rceil}_{(A)}, \underbrace{\varphi(c) - \sum_{c' \in \chi(E(P_v))} \tau_i(v_{c'})}_{(B)} \right). \tag{2.35}$$

The value of $\tau_i(v)$ will be used in Section 2.4.5 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

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

For $v \in V \setminus \{r\}$, let $c = \chi(v, p(v))$, for each $i \in \mathbb{Z}$, part (B) of (2.35) for $\tau_i(v_c)$ implies that

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

Hence,

$$\begin{aligned} \varphi(c) - \sum_{c' \in \chi(E(P_v))} \tau_i(v_{c'}) &= \varphi(c) - \tau_i(v_c) - \sum_{c' \in \chi(E(P_{v_c}))} \tau_i(v_{c'}) \\ &\geq \varphi(c) - \varphi(\rho(c)) \\ &= \kappa(c) \\ &\geq 1. \end{aligned} \tag{2.36}$$

Therefore, part (B) of (2.35) is always positive, so if $\tau_k(v) = 0$ for some $k \geq \left\lfloor \log_2 \left(\frac{m(T)}{M(\chi) + \log_2 |E|} \right) \right\rfloor$, 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 2.16. *For $v \in V$ and $k \geq \left\lfloor \log_2 \left(\frac{m(T)}{M(\chi) + \log_2 |E|} \right) \right\rfloor$, if $\tau_k(v) = 0$ then for all $i \geq k$, $\tau_i(v) = 0$.*

Comparing part (A) of (2.35) for $\tau_i(v)$ and $\tau_{i+1}(v)$ also allows us to observe the following.

Observation 2.17. *For $v \in V$ and $k \geq \left\lfloor \log_2 \left(\frac{m(T)}{M(\chi) + \log_2 |E|} \right) \right\rfloor$, if part (A) in (2.35) for $\tau_k(v)$ is less than or equal to part (B) then for all $i > k$, $\tau_i(v) = 0$.*

Properties of the Scale Selector Maps

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

Lemma 2.18. *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 $k = \left\lceil \log_2 \left(\frac{d_T(v, v_c)}{\kappa(c)} \right) \right\rceil$. 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 (2.35).

If $\tau_k(v) = 0$, then by Observation 2.16, 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 (2.35), in which case

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

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 (2.35), we see that $\tau_{k+1}(v) = 0$, and by Observation 2.16, $\tau_i(v) = 0$ for $i > k$. Alternately, $\tau_k(v)$ is determined by part (A) of (2.35), and by Observation 2.17 $\tau_i(v) = 0$ 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 2.19. *For any vertex $v \in V$ with $c = \chi(v, p(v))$, we have*

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

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

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

By Lemma 2.18, the maximum exists.

We have $\tau_{k+1}(v) = 0$, and thus inequality (2.36) implies that part (A) of (2.35) specifies $\tau_{k+1}(v)$, yielding

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

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 2.18 implies that $2^k < 2\, d_T(v, v_c)$, hence,

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

□

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. See Figure 2.6 for an example.

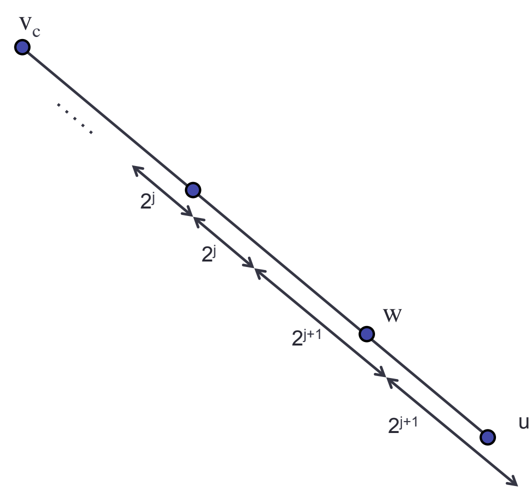


Figure 2.6: The goal of (2.35) is to cover the distance between v_c and w with smallest scales possible. In this figure $\tau_j(w) = 2$ and $\tau_{j+1}(w) = 1$, while $\tau_j(u) = 2$ and $\tau_{j+1}(u) = 2$.

Lemma 2.20. *Let $u, w \in V$ be such that $c = \chi(w, p(w)) = \chi(u, p(u))$, and $d_T(w, v_c) \leq d_T(u, v_c)$. Then for all $i \in \mathbb{Z}$, we have*

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

Proof. First let k be the smallest integer for which,

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

This k exists since, by (2.36), the right hand side is always positive, while by Lemma 2.18, the left hand side must be zero for some $k \in \mathbb{Z}$ large enough.

For $i > k$, by Observation 2.17 we have, $\tau_i(w) = 0$. Therefore, for $i > k$, we have $\tau_i(u) \geq \tau_i(w)$. We now use induction on i to show that for $i < k$, $\tau_i(u) = \tau_i(w)$, and for $i = k$, $\tau_k(u) \geq \tau_k(w)$. 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 (2.35) for $\tau_{k-1}(w)$ is an integer strictly less than part (A), hence

$$\begin{aligned}
\sum_{j=-\infty}^{k-1} 2^j \tau_j(w) &= 2^{k-1}\tau_{k-1}(w) + \sum_{j=-\infty}^{k-2} 2^j \tau_j(w) \\
&\leq 2^{k-1}\left(\left\lceil \frac{d_T(w, v_c) - \sum_{j=-\infty}^{k-2} 2^j \tau_j(w)}{2^{k-1}} \right\rceil - 1\right) + \sum_{j=-\infty}^{k-2} 2^j \tau_j(w) \\
&< 2^{k-1}\left(\frac{d_T(w, v_c) - \sum_{j=-\infty}^{k-2} 2^j \tau_j(w)}{2^{k-1}}\right) + \sum_{j=-\infty}^{k-2} 2^j \tau_j(w) \\
&\leq d_T(w, v_c)\,.
\end{aligned} \tag{2.37}$$

For $\left\lfloor \log_2\left(\frac{m(T)}{M(\chi)+\log_2|E|}\right)\right\rfloor \leq i \leq k$, by (2.37), and as $d_T(u, v_c) \geq d_T(w, v_c)$, we have

$$\min\left(d_T(w, v_c), \sum_{j=-\infty}^{i-1} 2^j \tau_j(w)\right) = \sum_{j=-\infty}^{i-1} 2^j \tau_j(w) = \min\left(d_T(u, v_c), \sum_{j=-\infty}^{i-1} 2^j \tau_j(w)\right). \tag{2.38}$$

By our induction hypothesis for all $j < i$, $\tau_j(w) = \tau_j(u)$, so using (2.38) we can write,

$$d_T(w, v_c) - \min\left(d_T(w, v_c), \sum_{j=-\infty}^{i-1} 2^j \tau_j(w)\right) \leq d_T(u, v_c) - \min\left(d_T(u, v_c), \sum_{j=-\infty}^{i-1} 2^j \tau_j(u)\right). \tag{2.39}$$

Since $\chi(w,p(w))=\chi(u,p(u))$, for all $i\in\mathbb{Z}$ part (B) of (2.35) is identical for $\tau_i(u)$ and $\tau_i(w)$. Therefore, using (2.39), 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 (2.35) 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 (2.35) is identical for $\tau_k(u)$ and $\tau_k(w)$, and inequality (2.39) implies that part (A) of (2.35) for $\tau_k(u)$ is at least as large as part (A) of (2.35) for $\tau_k(w)$, completing the proof. □

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

Lemma 2.21. *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 2.16, $\tau_k(w)=0$. Letting $c=\chi(u,p(u))$, by Lemma 2.20 we have $d_T(v_c,u)\ge d_T(v_c,w)$. Using Lemma 2.20 again, we can conclude that for all $i\in\mathbb{Z}$, $\tau_i(u)\ge\tau_i(w)$. Since $\tau_{k-1}(w)=0$, inequality (2.36) implies that part (A) of (2.35) specifies $\tau_{k-1}(w)$. Therefore,

$$\begin{aligned} d_T(w,v_c) &\le \sum_{i=-\infty}^{k-2}2^i\tau_i(w)\\ &\le \sum_{i=-\infty}^{k-2}2^i\tau_i(u)\\ &= \left(\sum_{i=-\infty}^{k-1}2^i\tau_i(u)\right)-2^{k-1}\tau_{k-1}(u). \end{aligned} \tag{2.40}$$

Since $\tau_k(u)>0$, using part (A) of (2.35), we can write

$$d_T(u,v_c)>\sum_{i=-\infty}^{k-1}2^i\tau_i(u). \tag{2.41}$$

Observation 2.16 implies that $\tau_{k-1}(u)\neq 0$, thus $\tau_{k-1}(u)\ge 1$, and using (2.40) and (2.41), 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 2.22. *For any $k \in \mathbb{N} \cup \{0\}$, and any color $c \in \chi(E)$, we have*

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

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

$$\begin{aligned} \varphi(c_m) - \varphi(c_1) &= \sum_{i=1}^{m-1} \kappa(c_i) \\ &\geq \sum_{i=1}^{m-1} \log_2\left(\frac{|E(T(c_{i+1}))|}{|E(T(c_i))|}\right) \\ &\geq \log_2\left(\frac{|E(T(c))|}{|E(T(c'))|}\right). \end{aligned} \tag{2.42}$$

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

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

∎

The following two corollaries are immediate from Lemma 2.22.

Corollary 2.23. *For any $k \in \mathbb{N}$, and any color $c \in \chi(E)$, we have*

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

Corollary 2.24. *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 2.21. The assumption is more general, and the conclusion is correspondingly weaker. This result is used primarily to enable the proof of Lemma 2.26.

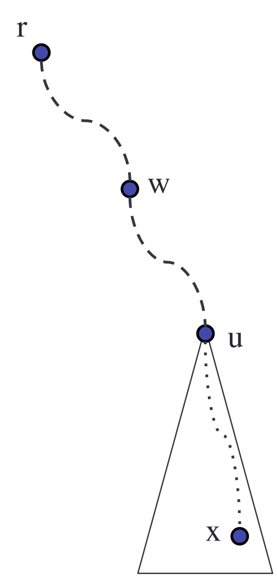


Figure 2.7: Relative position of vertices in the tree for Lemma 2.25.

Lemma 2.25. *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*

$$2^k \geq \left(\frac{6\, d_T(x, w)}{\varphi(\chi(u, p(u))) - \varphi(\chi(w, p(w)))} \right), \tag{2.43}$$

we have $\tau_k(x) = 0$ (See Figure 2.7 for relative position of vertices in the tree).

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,

$$k \geq \log_2\left(\frac{6\,d_T(x,w)}{\varphi(\chi(u,p(u)))-\varphi(\chi(w,p(w)))}\right), \tag{2.44}$$

then part (A) of (2.35) 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 (2.44) implies

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

By Lemma 2.19, we have

$$\sum_{i=1}^{m-2} 2^{k-1}\tau_{k-1}(y_i) \leq \sum_{i=1}^{m-2}\sum_{j=-\infty}^{\infty} 2^j\tau_j(y_i) \leq \sum_{i=1}^{m-2} 3\,d_T(y_i,y_{i+1}) = 3\,d_T(y_1,y_{m-1}). \tag{2.45}$$

Now, using (2.43) gives

$$\begin{aligned}\varphi(c_1)-\varphi(c_m) &\geq \varphi(\chi(u,p(u)))-\varphi(\chi(w,p(w)))\\ &\geq \frac{6\,d_T(x,w)}{2^k}\\ &\geq \frac{6\,d_T(x,y_{m-1})}{2^k}.\end{aligned} \tag{2.46}$$

Using the above inequality and (2.45), we can write

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

First, note that $c_m=\chi(y_{m-1},p(y_{m-1}))$. Now, we use part (B) of (2.35) for $\tau_k(y_{m-1})$ to write

$$\begin{aligned}d_T(x,y_1) &\leq \frac{2^{k-1}}{3}\left(\varphi(c_1)-\left(\tau_{k-1}(y_{m-1})+\sum_{c'\in\chi(E(P_{y_{m-1}}))}\tau_{k-1}(v_{c'})\right)-\sum_{i=1}^{m-2}\tau_{k-1}(y_i)\right)\\ &\leq \frac{2^{k-1}}{3}\left(\varphi(c_1)-\sum_{c'\in\chi(E(P_x))}\tau_{k-1}(v_{c'})\right)\\ &\leq 2^{k-1}\left(\varphi(\chi(x,p(x)))-\sum_{c'\in\chi(E(P_x))}\tau_{k-1}(v_{c'})\right).\end{aligned} \tag{2.47}$$

Therefore, either part (A) of (2.35) specifies $\tau_{k-1}(x)$ in which case by Observation 2.17, $\tau_i(v) = 0$ for $i \geq k$, or part (B) of (2.35) specifies $\tau_{k-1}(x)$ in which case by (2.47) we have,

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

and part (A) of (2.35) is zero for $i \geq k$. □

In Section 2.4.5, 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 2.26 is used in the proof of Lemma 2.41.

Lemma 2.26. *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,*

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

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 of $P_{vr'}$ 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 2.18 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 (2.33) 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),

$$\begin{aligned}\kappa(c_{j+1}) > \Big(&\big(\varphi(c_{b_i+1}) - \varphi(c_{a_i+1})\big) + \big(\varphi(c_{a_i+1}) - \varphi(\chi(r', p(r'))\big) + 2(\varphi(c_{j+2}) - \varphi(c_{b_i+1}))\Big)\\ &- \Big(\varphi(c_{j+2}) - \varphi(c_{b_i+1})\Big)\\ = \varphi&(c_{j+2}) - \varphi(\chi(r', p(r'))). \end{aligned} \tag{2.48}$$

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 (2.48), 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 (2.33),

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

and by (2.48),

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

Hence Lemma 2.18 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 2.25 to show that the assignment $u = y_j$ and $u' = y_{b_i}$ satisfies the conditions of the lemma. We have

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

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 2.25 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 $C = 96$. □

2.4.5 Embedding

We now present a proof of Theorem 2.11, thereby completing the proof of Theorem 2.1. 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 Section 2.4.4. 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.

The Construction

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

$$t = \lceil \varepsilon^{-1} + \log\lceil \log_2 1/\delta \rceil \rceil, \tag{2.49}$$

and

$$m = \lceil t^2 (M(\chi) + \log_2 |E|) \rceil. \tag{2.50}$$

(See Lemma 2.41 for the relation between ε and δ, and the parameters of Theorem 2.11). 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 $\beta \leq m$ since

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

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

$$\Delta_i(v)[j] = \begin{cases} \left(\frac{2^i}{t^2}, 0, 0 \ldots, 0\right) & \text{if } \alpha < j \leq \beta, \\ \left(d_T(v_c, v) - \left(\left(\sum_{\ell=-\infty}^{i-1} 2^\ell \tau_\ell(v)\right) + (\beta - \alpha)\frac{2^i}{t^2}\right), 0, 0 \ldots, 0\right) & \\ & \text{if } j = \beta + 1 \text{ and } \beta - \alpha < t^2 \tau_i(v), \\ (0, 0 \ldots, 0) & \text{otherwise.} \end{cases} \tag{2.51}$$

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 (2.35), and in that case $\sum_{\ell \leq i} 2^\ell \tau_\ell(v) > d(v, v_c)$.

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

Observation 2.27. *For $v \in V$ and $i \in \mathbb{Z}$, each row in $\Delta_i(v)$ has at most one non-zero coordinate.*

Observation 2.28. *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 Section 2.4.5.

Lemma 2.29. *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 2.30. *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 2.31. *For $c \in \chi(E)$, and $u, w \in V(\gamma_c) \setminus \{v_c\}$, we have*

$$d_T(w, u) = \sum_{i \in \mathbb{Z}} \|\Delta_i(u) - \Delta_i(w)\|_1, \tag{2.52}$$

and

$$d_T(v_c, u) = \sum_{i \in \mathbb{Z}} \|\Delta_i(u)\|_1. \tag{2.53}$$

Lemma 2.32. *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,

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

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,

$$f_{i,c}(x) = \begin{cases} 0 & \text{if } x = v_c, \\ \Delta_i(x) & \text{if } x \in V(\gamma_c) \setminus \{v_c\}, \\ \Delta_i(v_{c'}) + \Pi_{i,c'}(f_{i,c'}(x)) & \text{if } x \in V(T(c')) \setminus \{v_{c'}\} \text{ for some } c' \in \rho^{-1}(c). \end{cases} \tag{2.54}$$

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 2.11.

Lemma 2.33. *There exists a universal constant C such that the following holds with non-zero probability: For all $x, y \in V$,*

$$(1 - C\varepsilon)\, d_T(x,y) - \delta\, \rho_\chi(x,y;\delta) \le \sum_{i \in \mathbb{Z}} \|f_i(x) - f_i(y)\|_1 \le d_T(x,y)\,. \tag{2.55}$$

We will prove Lemma 2.33 in Section 2.4.5. We first make two observations, and then use them to prove Theorem 2.11. Our first observation is immediate from Observation 2.27 and Observation 2.28, since in the third case of (2.54), by Observation 2.28, $\Delta_i(v_c')$ and $\Pi_{i,c'}(f_{i,c'}(x))$ must be supported on disjoint sets of rows.

Observation 2.34. *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 2.28 and Lemma 2.31 also imply the following.

Observation 2.35. *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 2.15, we now prove Theorem 2.11.

Proof of Theorem 2.11. By Lemma 2.33, 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 2.15 to the family given by $\left\{f_i = \frac{t^2 g_i}{2^i}\right\}_{i\in\mathbb{Z}}$ to arrive at an embedding $F : V \to \ell_1^{tm\left(2+\lceil \log \frac{1}{\varepsilon}\rceil\right)}$ such that $G = F/t^2$ satisfies,

$$d_T(x,y) \geq \|G(x) - G(y)\|_1 \geq (1 - O(\varepsilon))d_T(x,y) - \delta\rho_\chi(x,y;\delta). \tag{2.56}$$

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

To achieve (2.56), 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 (2.31). Recalling this definition, we now restate ζ in terms of our explicit family $\left\{f_i = \frac{t^2 g_i}{2^i}\right\}_{i\in\mathbb{Z}}$. We have,

$$\frac{\zeta(x,y)}{t^2} = \sum_{(k_1,k_2)\in[m]\times[t]} \sum_{\substack{i:\exists j<i \\ g_j(x)[k_1,k_2]\neq g_j(y)[k_1,k_2]}} h_i(x,y;k_1,k_2)\,, \tag{2.57}$$

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,

$$d_T(x,y) = \sum_{c\in\chi(E(P_{xy}))} \mathsf{len}(E(\lambda_c)). \tag{2.58}$$

If we look at a single term in (2.57), we have

$$h_i(x,y;k_1,k_2) < \frac{2^i}{t^2}. \tag{2.59}$$

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

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

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 2.29, 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

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

By Lemma 2.18, 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

$$\|\Delta_i(u_c)\|_1 = \|g_i(u_c) - g_i(v_c)\|_1 = 0. \tag{2.60}$$

For $i < 1 + \log_2(d_T(u_c, v_c))$, by (2.59) and Lemma 2.29 we can bound the contribution of λ_c to $\zeta(x, y)$ by,

$$\sum_{j\in\mathbb{Z}} \frac{2^j|S_j(u_c, v_c)|}{t^2} < \frac{2^i}{t^2} < \frac{2d_T(u_c, v_c)}{t^2} \leq \varepsilon d_T(u_c, v_c). \tag{2.61}$$

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 (2.61),

$$\zeta(x, y) < \sum_{c\in\chi(E(P_{xy}))} \varepsilon \mathsf{len}(E(\lambda_c)) \overset{(2.58)}{\leq} \varepsilon d_T(x, y).$$

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 2.29, 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

$$\begin{aligned}
\zeta(x, y) &< \sum_{i\in\mathbb{Z}} \frac{2^i|S_i(u, w)|}{t^2} + \sum_{c\in\chi(E(P_x))\triangle\chi(E(P_y))} \sum_{i\in\mathbb{Z}} \frac{2^i|S_i(u_c, v_c)|}{t^2} \\
&\overset{(2.58)}{\leq} 4\varepsilon d_T(x, y) + \sum_{c\in\chi(E(P_x))\triangle\chi(E(P_y))} \varepsilon\, \mathsf{len}(E(\lambda_c)) \\
&\leq 5\varepsilon d_T(x, y).
\end{aligned}$$

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 2.18 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 2.21 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 (2.105), $i \leq m_w$ 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 2.20 and as $\tau_{m_w-1}(u) \neq 0$, we have $m_u \leq m_w \leq m_u + 1$. Observation 2.17, 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 2.17,

$$\tau_j(w) + \sum_{c' \in \chi(E(P_w))} \tau_j(v_{c'}) = \tau_j(u) + \sum_{c' \in \chi(E(P_u))} \tau_j(v_{c'}) = \varphi(c). \tag{2.62}$$

therefore, by Observation 2.28 for $j < m_u$ and $k \in [t^2\varphi(c)]$

$$\|(g_j(x) - g_j(u))[k]\|_1 = \|(g_j(y) - g_j(w))[k]\|_1 = 0, \tag{2.63}$$

and by Observation 2.28 and part (B) of (2.35), 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 (2.54) and (2.105), we have $\tau_j(u_{c'}) \neq 0$. Inequality (2.63) implies $j \geq m_u$, and finally by Lemma 2.18,

$$d_T(x,y) \geq d_T(u_{c'}, v_{c'}) > 2^{j-1} \geq 2^{m_u-1} \geq 2^{m_w-2} \geq 2^{i-2}. \tag{2.64}$$

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 2.32,

$$d_T(x,y) \geq d_T(u,w) \geq 2^{m_u-1} \geq 2^{m_w-2} \geq 2^{i-2}. \tag{2.65}$$

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

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

□

Properties of the Δ_i Maps

We now present proofs of Lemmas 2.29–2.32.

Proof of Lemma 2.29. For a fixed $i \in \mathbb{Z}$, by (2.105) 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 (2.105), 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

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

By part (A) of (2.35), for $i' > i$ we have $\tau_{i'}(v) = 0$, thus $\|\Delta_{i'}(v)\|_1 = 0$ and the proof is complete. □

Proof of Lemma 2.30. 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 (2.35) for $\tau_\nu(w)$ is less that or equal to part (B). This ν exists since, by (2.36), part (B) of (2.35) is always positive, while by Lemma 2.18, part (A) of (2.35) must be zero for some $\nu \in \mathbb{Z}$ large enough. First we analyze the case when $i < \nu$.

Observation 2.17 implies that part (B) of (2.35) specifies the value of $\tau_i(w)$. By Lemma 2.20 $\tau_i(u) \ge \tau_i(w)$, 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 (2.105) 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 (2.105), it is easy to verify that for all $(j, k) \in [m] \times [t]$,

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

Finally, in the case that $i > \nu$, by Observation 2.17, we have $\tau_i(w) = 0$, and $\Delta_i(w)[j, k] = 0$. □

Proof of Lemma 2.31. For all $i \in \mathbb{Z}$, recalling the definition α and β in (2.105) 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 2.19, 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 2.30 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 2.32. Without loss of generality suppose that $d_T(v_c, u) \geq d_T(v_c, w)$. We have,

$$\begin{aligned} d_T(u, w) &= \sum_{h\in\mathbb{Z}} \|\Delta_h(u) - \Delta_h(w)\|_1 \\ &\geq \sum_{h=j}^{i} \|\Delta_h(u) - \Delta_h(w)\|_1 \\ &\geq \|\Delta_i(u) - \Delta_i(w)\|_1 + \|\Delta_j(u) - \Delta_j(w)\|_1 . \end{aligned} \tag{2.66}$$

By Lemma 2.20 we have $\tau_j(w) \leq \tau_j(u)$. In the definition of $\tau_j(w)$, if part (B) of (2.35) is less than part (A), then by (2.105), 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 $\|\Delta_j(w)[h]\|_1 = \frac{2^i}{t^2}$. And, by Lemma 2.30, and Observation 2.28 for $k \in \mathbb{Z}$, $\Delta_j(w) = \Delta_j(u)$. Hence, part (A) of (2.35) must specify the value of $\tau_j(w)$. Observation 2.17 implies that $\tau_i(w) = 0$ and by (2.105), we have $\|\Delta_i(w)\|_1 = 0$.

By (2.105), since $\|\Delta_i(u)[k]\|_1 > 0$, and α from (2.105) 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 (2.105) implies that for all $k < h \leq t^2\lfloor 1 + \frac{k}{t^2} \rfloor$, we have $\|\Delta_j(w)[h]\|_1 = 0$. The same argument also shows that,

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

Hence by (2.66),

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

□

The Probabilistic Analysis

We are thus left to prove Lemma 2.33. 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 2.36. *For any $C \geq 8$, 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*

$$\mathbb{P}\Bigg[\exists\, x \in V(P_{w\,v_a}) : \sum_{i\in\mathbb{Z}} \|f_{i,a}(x) - f_{i,a}(u)\|_1 \leq \tag{2.67}$$

$$(1 - C\varepsilon)\; d_T(u, v_c) + \sum_{i\in\mathbb{Z}} \|f_{i,a}(v_c) - f_{i,a}(x)\|_1 \mid \{f_{i,c'}\}_{c'\in\rho^{-1}(a)}\Bigg]$$

$$\leq \frac{1}{\lceil \log_2 1/\delta \rceil} \exp\Big(-(C/(\varepsilon 2^{\beta+2}))\, d_T(u, v_c)\Big), \tag{2.68}$$

where $\beta = \max\{i : \exists y \in P_{u\,v_c} \backslash \{v_c\}, \tau_i(y) \neq 0\}$. (See Figure 2.8 for position of vertices in the tree.)

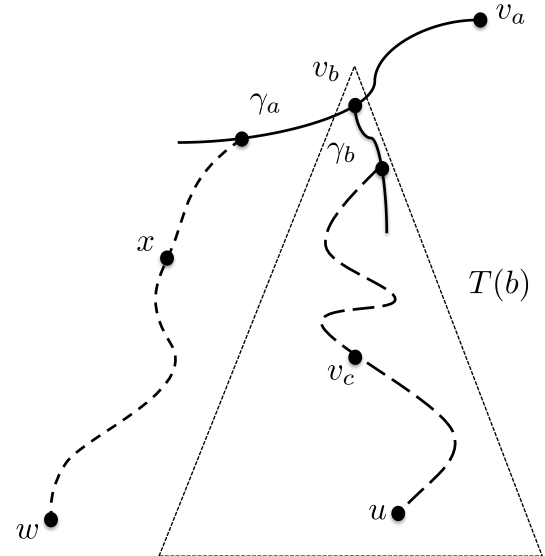


Figure 2.8: Position of vertices corresponding to the statement of Lemma 2.36.

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

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

We have,

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

By Observation 2.35, we have $d_T(u, v_c) = \sum_{i\in\mathbb{Z}} \|f_{i,a}(u) - f_{i,a}(v_c)\|_1$, therefore

$$d_T(u, v_c) - \sum_{i\in\mathbb{Z}, j\in[m]} s_{ij}(z) = \sum_{i\in\mathbb{Z}} \|f_{i,a}(z) - f_{i,a}(u)\|_1 - \sum_{i\in\mathbb{Z}} \|f_{i,a}(z) - f_{i,a}(v_c)\|_1. \tag{2.69}$$

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

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

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 $z \in P_{w\,v_a}$,

$$s_{ij}(z) \leq 2\|f_{i,a}(v_c)[j] - f_{i,a}(u)[j]\|_1, \tag{2.70}$$

Hence for all $i \in \mathbb{Z}$ and $j \in [m]$ by Observation 2.34,

$$s_{ij}(z) \leq 2\|f_{i,a}(v_c)[j] - f_{i,a}(u)[j]\|_1 \leq \frac{2^{\beta+1}}{t^2}. \tag{2.71}$$

First note that, if z is on the path between v_b and v_a then by Observation 2.35, $s_{ij}(z) = 0$. Observation 2.28 and (2.105) 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 2.30, 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

$$X_{ij} = \begin{cases} 0 & \text{if } s_{ij}(w) = 0, \\ 2\|f_{i,a}(u)[j] - f_{i,a}(v_c)[j]\|_1 & \text{if } s_{ij}(w) \neq 0. \end{cases} \tag{2.72}$$

Note that since the re-randomization in (2.54) 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 (2.70), for all $z \in P_{w\,v_a}$, we have $s_{ij}(z) \leq X_{ij}$, and thus

$$\mathbb{P}_{\mathcal{E}}\left[\exists\, x \in V(P_{w\,v_a}) : \sum_{i\in\mathbb{Z}, j\in[m]} s_{ij}(x) \geq C\varepsilon d_T(u, v_c)\right] \leq \mathbb{P}_{\mathcal{E}}\left[\sum_{i\in\mathbb{Z}, j\in[m]} X_{ij} \geq C\varepsilon d_T(u, v_c)\right]. \tag{2.73}$$

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 (2.54) and Observation 2.34, this happens at most with probability $\frac{1}{t}$, hence for $j \in [m]$, and $i \in \mathbb{Z}$,

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

This yields,

$$\mathbb{E}[X_{ij} \mid \mathcal{E}] \leq \frac{1}{t}\left(2\|f_{i,a}(u)[j] - f_{i,a}(v_c)[j]\|_1\right). \tag{2.74}$$

Now we use (2.71) to write

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

and use Observation 2.35 in conjunction with (2.74) to conclude that

$$\mathbb{E}\left[\sum_{i\in\mathbb{Z},j\in[m]} X_{ij} \mid \mathcal{E}\right] \leq \sum_{i\in\mathbb{Z},j\in[m]} \frac{2}{t}\|f_i(v_c)[j] - f_i(u)[j]\|_1 = \frac{2}{t}d_T(v_c,u), \tag{2.75}$$

and

$$\sum_{i\in\mathbb{Z},j\in[m]} \operatorname{Var}(X_{ij} \mid \mathcal{E}) \leq \sum_{i\in\mathbb{Z},j\in[m]} \frac{2^{\beta+2}}{t^3}\|f_i(v_c)[j] - f_i(u)[j]\|_1 = \frac{2^{\beta+2}}{t^3}d_T(v_c,u). \tag{2.76}$$

Define $M = \max\{X_{ij} - \mathbb{E}[X_{ij} \mid \mathcal{E}] : i \in \mathbb{Z}, j \in [m]\}$. We now apply Theorem 2.7 to complete the proof:

$$\begin{aligned}
&\mathbb{P}_{\mathcal{E}}\left[\sum_{i\in\mathbb{Z},j\in[m]} X_{ij} \geq C\left(\frac{d_T(u,v_c)}{t}\right)\right] \\
&\qquad = \mathbb{P}_{\mathcal{E}}\left[\sum_{i\in\mathbb{Z},j\in[m]} X_{ij} - \frac{2d_T(u,v_c)}{t} \geq (C-2)\left(\frac{d_T(u,v_c)}{t}\right)\right] \\
&\qquad \overset{(2.75)}{\leq} \mathbb{P}_{\mathcal{E}}\left[\sum_{i\in\mathbb{Z},j\in[m]} X_{ij} - \mathbb{E}\left[\sum_{i\in\mathbb{Z},j\in[m]} X_{ij} \mid \mathcal{E}\right] \geq (C-2)\left(\frac{d_T(u,v_c)}{t}\right)\right] \\
&\qquad \leq \exp\left(\frac{-((C-2)d_T(u,v_c)/t)^2}{2\left(\sum_{i\in\mathbb{Z},j\in[m]} \operatorname{Var}(X_{ij} \mid \mathcal{E}) + (C-2)(d_T(u,v_c)/t)M/3\right)}\right).
\end{aligned}$$

Since $\mathbb{E}[X_{ij} \mid \mathcal{E}] \geq 0$, (2.71) implies $M \leq \frac{2^{\beta+1}}{t^2}$. Now, we can plug in this bound and (2.76) to write,

$$\mathbb{P}_{\mathcal{E}}\left[\sum_{i\in\mathbb{Z},j\in[m]} X_{ij} \geq C\left(\frac{d_T(u,v_c)}{t}\right)\right]$$

$$\leq \exp\left(\frac{-((C-2)d_T(u,v_c)/t)^2}{2\left(\frac{2^{\beta+2}}{t^3}d_T(u,v_c)+(C-2)(d_T(u,v_c)/t)(2^{\beta+1}/t^2)/3\right)}\right)$$
$$= \exp\left(\frac{-t(C-2)^2 d_T(u,v_c)}{2\left(2^{\beta+2}+(C-2)(2^{\beta+1})/3\right)}\right)$$
$$= \exp\left(\frac{-(C-2)^2}{(C-2)/3+2}\left(\frac{t d_T(u,v_c)}{2^{\beta+2}}\right)\right).$$

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

$$\mathbb{P}_{\mathcal{E}}\left[\sum_{i\in\mathbb{Z},j\in[m]} X_{ij} \geq C\left(\frac{d_T(u,v_c)}{t}\right)\right]$$

$$< \exp\left(-(Ct/2^{\beta+2})\, d_T(u,v_c)\right)$$
$$\stackrel{(2.49)}{\leq} \exp\left(-C\left(\frac{1}{\varepsilon}+\log\left\lceil\log_2\frac{1}{\delta}\right\rceil\right)\left(\frac{1}{2^{\beta+2}}\right)\, d_T(u,v_c)\right)$$
$$= \left(\frac{1}{\lceil\log_2(1/\delta)\rceil}\right)^{\frac{Cd_T(u,v_c)}{2^{\beta+2}}} \cdot \exp\left(-C\left(\frac{1}{\varepsilon}\right)\left(\frac{1}{2^{\beta+2}}\right)\, d_T(u,v_c)\right).$$

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 2.18 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,

$$\mathbb{P}_{\mathcal{E}}\left[\exists\, x \in V(P_{w\,v_a}) : \sum_{i\in\mathbb{Z}} \|f_{i,a}(x)-f_{i,a}(u)\|_1 \leq (1-C\varepsilon)\; d_T(u,v_c) + \sum_{i\in\mathbb{Z}} \|f_{i,a}(v_c)-f_{i,a}(x)\|_1\right]$$

$$\stackrel{(2.69)}{\leq} \mathbb{P}_{\mathcal{E}}\left[\exists\, x \in V(P_{w\,v_c}) : \sum_{i\in\mathbb{Z},j\in[m]} s_{ij}(x) \geq C\varepsilon d_T(u,v_c)\right]$$
$$\stackrel{(2.73)}{\leq} \mathbb{P}_{\mathcal{E}}\left[\sum_{i\in\mathbb{Z},j\in[m]} X_{ij} \geq C\varepsilon\left(d_T(u,v_c)\right)\right]$$
$$\stackrel{(2.49)}{\leq} \mathbb{P}_{\mathcal{E}}\left[\sum_{i\in\mathbb{Z},j\in[m]} X_{ij} \geq C\left(\frac{d_T(u,v_c)}{t}\right)\right]$$
$$< \left(\frac{1}{\lceil\log_2(1/\delta)\rceil}\right)\cdot \exp\left(-C\left(\frac{1}{\varepsilon 2^{\beta+2}}\right)\, d_T(u,v_c)\right),$$

completing the proof. □

The Γ_a mappings. Before proving Lemma 2.33, 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 2.36. 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

$$2^k < \frac{d_T(u,w)}{\varepsilon(\varphi(\chi(u,p(u))) - \varphi(a))}. \tag{2.77}$$

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 2.37. *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

$$2^k < \frac{d_T(u, p(\Gamma_a(u)))}{\varepsilon(\varphi(\chi(u,p(u))) - \varphi(a))}. \tag{2.78}$$

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 (2.78) 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

$$d_T(u, \Gamma_a(u)) \leq d_T(u, p(\Gamma_a(u))). \tag{2.79}$$

Recalling that $p(w) = \Gamma_a(u)$, by Lemma 2.20 for all $k \in \mathbb{Z}$, $\tau_k(\Gamma_a(u)) \leq \tau_k(w)$, 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 (2.77) implies

$$2^k < \frac{d_T(u, \Gamma_a(u))}{\varepsilon(\varphi(\chi(u,p(u))) - \varphi(a))} \overset{(2.79)}{\leq} \frac{d_T(u, p(\Gamma_a(u)))}{\varepsilon(\varphi(\chi(u,p(u)) - \varphi(a))}. \tag{2.80}$$

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 (2.77),

$$2^k < \frac{d_T(u, \Gamma_a(u))}{\varepsilon(\varphi(\chi(u, p(u))) - \varphi(a))} \overset{(2.79)}{\leq} \frac{d_T(u, p(\Gamma_a(u)))}{\varepsilon(\varphi(\chi(u, p(u))) - \varphi(a))}, \tag{2.81}$$

completing the proof. □

Lemma 2.38. *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

$$d_T(u, \Gamma_a(u)) \leq d_T(u, \Gamma_a(w)). \tag{2.82}$$

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 (2.77),

$$2^k < \frac{d_T(u, \Gamma_a(u))}{\varepsilon(\varphi(\chi(u, p(u))) - \varphi(a))} \overset{(2.82)}{\leq} \frac{d_T(u, \Gamma_a(w))}{\varepsilon(\varphi(\chi(u, p(u))) - \varphi(a))}. \tag{2.83}$$

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 (2.77),

$$2^k < \frac{d_T(w, \Gamma_a(w))}{\varepsilon(\varphi(\chi(w, p(w))) - \varphi(a))} \leq \frac{d_T(u, \Gamma_a(w))}{\varepsilon(\varphi(\chi(w, p(w))) - \varphi(a))} = \frac{d_T(u, \Gamma_a(w))}{\varepsilon(\varphi(\chi(u, p(u))) - \varphi(a))}. \tag{2.84}$$

We have,

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

Hence, by (2.83) and (2.84), 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

$$2^k < \frac{d_T(u, p(\Gamma_a(w)))}{\varepsilon(\varphi(\chi(u, p(u))) - \varphi(a))}. \tag{2.85}$$

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 $\Gamma_a(u)$. □

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

$$R_a(c) = \bigcup_{i=0}^{\lceil \log_2 \frac{1}{\delta} \rceil - 1} \Big\{ u \in V(\gamma_c) : u \text{ is the furthest vertex} \\ \text{from } v_c \text{ s.t. } \Gamma_a(u) \neq u \text{ and } d(u, v_c) \leq 2^{-i}\,\mathsf{len}(\gamma_c) \Big\}. \quad (2.86)$$

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

Lemma 2.39. *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

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

If $i \leq \lceil \log_2 \frac{1}{\delta} \rceil - 1$, then (2.86) implies that either $u \in R_a(c)$, or there exists a $w \in R_a(c)$ such that

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

On the other hand, if $i > \lceil \log_2 \frac{1}{\delta} \rceil - 1$, then (2.86) implies that either $u \in R_a(c)$, or that there exists a $w \in R_a(c)$, such that

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

completing the proof. □

The following lemma, in conjunction with Lemma 2.39, reduces the number of vertices in $V(\gamma_c)$ that we need to analyze using Lemma 2.36.

Lemma 2.40. *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 $h \geq 0$, if $d(f(x), f(y)) \geq d_T(x, y) - h$ then $d(f(x'), f(y')) \geq d_T(x', y') - h$.*

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 2.33.

Lemma 2.41. *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*

$$(1-C\varepsilon)\, d_T(x,y) - \delta\, \rho_\chi(x,y;\delta) \le \sum_{i\in\mathbb{Z}} \|f_{i,c'}(x) - f_{i,c'}(y)\|_1 \le d_T(x,y)\,. \tag{2.87}$$

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

$$(1-C\varepsilon)\, d_T(x,y) - \delta\, \rho_\chi(x,y;\delta) \le \sum_{i\in\mathbb{Z}} \|f_{i,c}(x) - f_{i,c}(y)\|_1 \le d_T(x,y)\,. \tag{2.88}$$

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

$$d_T(x,y) \ge \sum_{i\in\mathbb{Z}} \|f_{i,c'}(x) - f_{i,c'}(y)\| \ge (1-C\varepsilon) d_T(x,y) - \delta\rho_\chi(x,y;\delta)\,. \tag{2.89}$$

We will prove the lemma by showing that, conditioned on $\mathcal{E}$, (2.88) 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

$$\mu(x,y) \ge \varphi(c)\,. \tag{2.90}$$

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 (2.88) 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 2.36. And let $\varepsilon_2 = 2 \cdot C'\varepsilon$, where C' is the constant from Lemma 2.26.

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

$$\sum_{k \in \mathbb{Z}} \|f_{k,c}(x) - f_{k,c}(y)\|_1 \geq d_T(x,y) - \varepsilon_1 d_T(x,y) - \varepsilon_2 d_T(\Gamma_c(x), \Gamma_c(y)) - \delta \rho_\chi(x,y;\delta). \quad (2.91)$$

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 (2.88) 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 $i \in [n]$,

$$\mathbb{P}_{\mathcal{E}}\left[X_1 \wedge \cdots \wedge X_i \mid X_1 \wedge \cdots \wedge X_{i-1} \wedge Y_i\right] = 1\,, \quad (2.92)$$

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

$$\mathbb{P}_{\mathcal{E}}\left[\overline{Y_i}\right] \leq 2^{-3(\varphi(c_i) - \varphi(c)) + 1}\,. \quad (2.93)$$

By, Lemma 2.31 and the definition of $f_{k,c}$ (2.54), 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

$$\begin{aligned}\mathbb{P}_{\mathcal{E}}[X_1 \wedge \cdots \wedge X_n] &\geq 1 - \sum_{i=2}^{n} \mathbb{P}_{\mathcal{E}}\left[\overline{Y_i}\right] \\ &\overset{(2.93)}{\geq} 1 - \sum_{i=2}^{n} 2^{-3(\varphi(c_i)-\varphi(c))+1} \\ &\overset{(2.24)}{>} 1 - 2 \cdot 2^{(2-3)} = 0,\end{aligned}$$

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

$$\sum_{k \in \mathbb{Z}} \|f_{k,c}(x) - f_{k,c}(y)\|_1 - \sum_{k \in \mathbb{Z}} \|f_{k,c}(\Gamma_c(x)) - f_{k,c}(y)\|_1 \geq (1 - \varepsilon_1/2)\, d_T(x, \Gamma_c(x))\,. \quad (2.94)$$

We now complete the proof of Lemma 2.41 by proving (2.92) and (2.93).

Proof of (2.92)**.** 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 2.31 and the definition of f_{k,c_i} (2.54), 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 (2.91) 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 2.39, there exists a vertex $z \in R_c(c_i)$ such that

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

If $d(x, v_{c_i}) \leq \delta\, \mathsf{len}(E(\gamma_{c_i}))$, then by (2.19), we have $\mathsf{len}(E(\gamma_{c_i})) = \rho_\chi(x, v_{c_i}; \delta)$, hence

$$\begin{aligned}d_T(z, \Gamma_c(z)) &\leq d_T(v_{c_i}, \Gamma_c(z)) + 2\max(\delta\, \mathsf{len}(E(\gamma_{c_i})), d_T(x, v_{c_i})) \\ &\leq d_T(v_{c_i}, \Gamma_c(z)) + 2\max(\delta\, \rho_\chi(x, v_{c_i}; \delta), d_T(x, v_{c_i})) \\ &\leq d_T(v_{c_i}, \Gamma_c(z)) + 2\,\delta\, \rho_\chi(x, v_{c_i}; \delta) + 2\, d_T(x, v_{c_i}) \\ &\leq 2\delta\, \rho_\chi(x, v_{c_i}; \delta) + 2\, d_T(x, \Gamma_c(z)). \quad (2.95)\end{aligned}$$

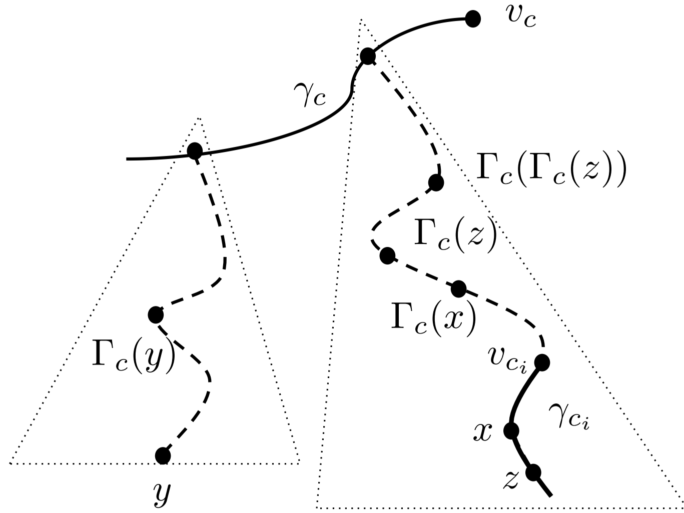


Figure 2.9: Position of vertices in the subtree $T(c)$ for Case I.

Since $z \in R_c(c_i)$, by definition we have $\Gamma_c(z) \neq z$, therefore by Lemma 2.37, $\Gamma_c(z) = v_{c'}$ 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 $\ell < i$.

We refer to Figure 2.9 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

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

We may assume that $\varepsilon_1 < 1$, otherwise there is nothing to prove. Using the preceding inequality, and applying Lemma 2.40 on pairs (z, y) and (x, y) implies that

$$\begin{aligned}\sum_{k\in\mathbb{Z}} \|f_{k,c}(x) - f_{k,c}(y)\|_1 &\geq d_T(x, y) - (\varepsilon_1/2)\, d_T(z, \Gamma_c(z)) \\ &\quad - \varepsilon_2\, d_T(\Gamma_c(\Gamma_c(z)), \Gamma_c(y)) - \varepsilon_1\, d_T(\Gamma_c(z), y) - \delta\, \rho_\chi(\Gamma_c(z), y; \delta) \\ &\overset{(2.95)}{\geq} d_T(x, y) - (\varepsilon_1/2) \Big(2\, d_T(x, \Gamma_c(z)) + 2\delta\, \rho_\chi(x, v_{c_i}; \delta)\Big) \\ &\quad - \varepsilon_2\, d_T(\Gamma_c(\Gamma_c(z)), \Gamma_c(y)) - \varepsilon_1\, d_T(\Gamma_c(z), y) - \delta\, \rho_\chi(\Gamma_c(z), y; \delta).\end{aligned}$$

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 2.37 it is easy to check that $x \in P_{z\,\Gamma_c(z)}$. Therefore, by Lemma 2.38, $d_T(\Gamma_c(\Gamma_c(z)), y) \leq d_T(\Gamma_c(z), y) \leq d_T(\Gamma_c(x), y)$, and combining this with the preceding inequality yields,

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

Recall the definition of $C(x, y; \delta)$ in (2.19). Since by Lemma 2.37, $\Gamma_c(z) = v_{c'}$ 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,

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

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

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

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 2.26. By Lemma 2.26 (with c', x, and $\frac{\varepsilon_2}{2}$ substituted for c, v, and ε, respectively, in the statement of Lemma 2.26), there exist vertices $u, u' \in \{x\} \cup \{v_a : a \in \chi(E(P_{x\,v_{c'}}))\}$ such that

$$d_T(x,u) \leq (\varepsilon_2/2)\, d_T(u',u). \tag{2.96}$$

and 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 (2.77) for $\Gamma_c(u)$, therefore

$$d_T(x,u) \leq (\varepsilon_2/2)\, d_T(u',u) \leq (\varepsilon_2/2)\, d_T(\Gamma_c(u),u). \tag{2.97}$$

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

We have,

$$\begin{aligned}\sum_{k\in\mathbb{Z}} \|f_{k,c}(x) - f_{k,c}(y)\|_1 - \sum_{k\in\mathbb{Z}} \|f_{k,c}(u) - f_{k,c}(y)\|_1 &\geq -\sum_{k\in\mathbb{Z}} \|f_{k,c}(x) - f_{k,c}(u)\|_1 \\ &\overset{(2.35)}{=} -d_T(x,u) \\ &\overset{(2.97)}{\geq} d_T(x,u) - \varepsilon_2\, d_T(u,\Gamma_c(u)) \\ &\geq d_T(x,u) - \varepsilon_2\, d_T(x,\Gamma_c(u)) \\ &= d_T(x,u) - \varepsilon_2\, d_T(\Gamma_c(x),\Gamma_c(u)).\end{aligned} \tag{2.98}$$

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,

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

Recall the definition of $C(x,y;\delta)$ in (2.19). 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 $\rho_\chi(u,y;\delta) \leq \rho_\chi(x,y;\delta)$. Now we can write,

$$\sum_{k\in\mathbb{Z}} \|f_{k,c}(u) - f_{k,c}(y)\|_1 \geq d_T(u,y) - \varepsilon_2\, d_T(\Gamma_c(u), \Gamma_c(y)) - \varepsilon_1\, d_T(u,y) - \delta\, \rho_\chi(x,y;\delta). \tag{2.99}$$

Adding (2.98) and (2.99) we can conclude that

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

completing the proof of (2.92).

Proof of (2.93)**.** We prove this inequality by first bounding the probability that (2.94) 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

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

And 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

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

Inequality (2.77) implies,

$$2^{\beta_x} < \frac{d_T(x, \Gamma_c(x))}{\varepsilon(\varphi(c_i) - \varphi(c))}. \tag{2.100}$$

By definition of R_c, for all elements $x \in R_c(c_i)$, we have $\Gamma_c(x) \neq x$. Moreover, by Lemma 2.37, $\Gamma_c(x) = v_{c'}$ 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 2.36 with $\varepsilon_1/2 = 12\varepsilon$ to write

$$\begin{aligned}\mathbb{P}_{\mathcal{E}}\Bigg[\exists y \in P_{w(x;a),v_c} : \sum_{k\in\mathbb{Z}} \|f_{k,c}(x) - f_{k,c}(y)\|_1 \\ \leq (1-\varepsilon_1/2)d_T(x,\Gamma_c(x)) + \sum_{k\in\mathbb{Z}} \|f_{k,c}(y) - f_{k,c}(\Gamma_c(x))\|_1\Bigg] \\ \leq \frac{1}{\lceil \log_2 1/\delta \rceil}\exp\left(-12\frac{d_T(x,\Gamma_c(x))}{2^{\beta_x+2}\,\varepsilon}\right) \\ \stackrel{(2.100)}{\leq} \frac{\exp(-3(\varphi(c_i)-\varphi(c)))}{\lceil \log_2 1/\delta \rceil}\,. \qquad (2.101)\end{aligned}$$

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 $R_c(c_i)$, $|R_c(c_i)| \leq \lceil \log_2 \delta^{-1} \rceil$. We also have $\varphi(c_j) \leq \varphi(c_i)$ for $j < i$, and by Corollary 2.23, $|S_x| \leq i < 2^{\varphi(c_i)-\varphi(c)+1}$. Taking a union bound over all $x \in R_c(c_i)$ and $a \in S_x$ implies,

$$\begin{aligned}\mathbb{P}_{\mathcal{E}}[\overline{Y_i}] &\stackrel{(2.101)}{\leq} \sum_{x\in R_c(c_i)} |S_x| \left(\frac{1}{\lceil \log_2 \delta^{-1} \rceil}\exp(-3(\varphi(c_i)-\varphi(c)))\right) \\ &< \left(\lceil \log_2 \delta^{-1} \rceil 2^{\varphi(c_i)-\varphi(c)+1}\right)\left(\frac{1}{\lceil \log_2 \delta^{-1} \rceil}\exp(-3(\varphi(c_i)-\varphi(c)))\right) \\ &= 2^{\varphi(c_i)-\varphi(c)+1}\exp(-3(\varphi(c_i)-\varphi(c)))\,.\end{aligned}$$

Since $\varphi(c_i) \geq \varphi(c)$, 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 (2.93). □

Finally, we present the proof of Lemma 2.33.

Proof of Lemma 2.33. Let C be the same constant as the constant in Lemma 2.41. For the sake of contradiction, suppose that

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

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

$$\mathbb{P}\left[\forall x, y \in V(T(c)),\ (1 - C\varepsilon)\, d_T(x, y) - \delta\, \rho_\chi(x, y; \delta) \le \sum_{i \in \mathbb{Z}} \|f_{i,c}(x) - f_{i,c}(y)\|_1 \le d_T(x, y)\right] = 0\,. \tag{2.102}$$

For $a \in \chi(E)$, $\kappa(a) > 0$. Hence, for all $c' \in \rho^{-1}(c)$, by (2.33), $\varphi(c') > \varphi(c)$, and by maximality of c, for all $c' \in \rho^{-1}(c)$, we have

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

But now applying Lemma 2.41 contradicts (2.102), completing the proof. □

2.5 Lower Bound

For $n \in \mathbb{N}$, the n-star graph is a tree on n nodes such that one node has degree $n-1$ and all the other nodes have degree one. In this section we prove the following theorem.

Theorem 2.42 (restatement of 2.3)**.** *For $\varepsilon \in (0, \frac{1}{16})$ and $n \ge 1/\varepsilon^2$, any embeddings of n-star into ℓ_1^d with distortion $1+\varepsilon$ requires dimension $d = \Omega\left(\frac{log(n)}{\varepsilon^2 \log(1/\varepsilon)}\right)$.*

2.5.1 Proof of Theorem 2.3

To prove Theorem 2.3 we first bound the number of almost-disjoint probability measures that we can put on a finite set, and then we use Lemma 2.43 to derive a bound on the dimension required for embedding n-star into ℓ_1 with distortion $1+\varepsilon$.

Let X be a finite set, and let S be a set of measures on the ground set X. We say that S is *ε-unrelated*, if for all distinct elements $\mu, \eta \in S$,

$$\|\mu - \eta\|_{TV} \ge \frac{1}{2}(\mu(X) + \eta(X)) - \varepsilon,$$

where $\|\cdot\|_{TV}$ is the total variation of the measure.

The following lemma is an easy corollary of a fact from [92]. We include the proof of this lemma for completeness.

Lemma 2.43. *For every $\varepsilon \in (0,1)$ and $k \in \mathbb{N}$, if there exists a map from n-star to ℓ_1^k with distortion $1+\varepsilon$, then there exists an ε-unrelated set of probability measures on $\{1, \ldots, 2k+1\}$ of size $n-1$.*

Proof. Let $f : V \to \ell_1^d$ be a map from n-star into ℓ_1^d that satisfies the condition of the lemma, and let $r \in V$ be the vertex with degree $n-1$. By translation, we may assume that $f(r) = 0$. For all other vertices $v \in V \setminus \{r\}$, we have $\|f(v)\|_1 \leq 1$. For each vertex $v \in V \setminus \{r\}$ define the measure μ_v as follows

$$\mu_v(\{i\}) = \begin{cases} \max(0, f(v)_i) & 1 \leq i \leq k \\ \max(0, -f(v)_i) & k+1 \leq i \leq 2k \\ 1 - \|f(v)\|_1 & i = 2k+1, \end{cases}$$

where we use $f(v)_i$ to denote the ith coordinate of $f(v)$. The set S has size $n-1$. To verify the pairwise distance between elements of the set, note that for all $u, v \in V \setminus \{r\}$ we have

$$\|\mu_u - \mu_v\|_{TV} = \frac{1}{2}\left(\|f(u) - f(v)\|_1 + |(1 - \|f(u)\|_1) - (1 - \|f(u)\|_1)|\right) \geq \|f(u) - f(v)\|_1.$$

Since f has distortion $1+\varepsilon$, for any two distinct elements $u, v \in S$, we have

$$\|f(u) - f(v)\|_1 \geq \left(\frac{2}{1+\varepsilon}\right) \geq 2(1-\varepsilon).$$

□

The next lemma is the final ingredient that we need to prove Theorem 2.3. Let $\mathcal{M}_k$ be the set of all measures on the set $\{1, 2, \dots, k\}$, and let $\mathcal{P}_k$ be the set of all probability measures on the set $\{1, 2, \dots, k\}$.

Lemma 2.44. *There exists a universal constant C, such that for $\varepsilon \leq 1/16$ the following statement holds. If there exists an ε-unrelated set $S \subseteq \mathcal{P}_k$, then there exists a $\frac{1}{2}$-unrelated set $T \subseteq \mathcal{P}_k$ of size at least $\frac{|S|}{14}$ such that for all $\mu \in T$, we have $|\mathrm{supp}(\mu)| \leq \lceil C\varepsilon(\varepsilon + \frac{1}{n})d \rceil$.*

We can now present the proof of Theorem 2.3.

Proof of Theorem 2.3. Suppose that there is a map from n-star to ℓ_1^d with distortion $1+\varepsilon$. Then, by Lemma 2.43 there exists a ε-unrelated set of probability measures on $\{2d+1\}$ of size $n-1$. Thus by Lemma 2.44, there must exist a $\frac{1}{2}$-unrelated set S of probability measures on $\{1, \dots, 2d+1\}$ of size $\Omega(n)$ such that all elements of S has support size at most

$$\left\lceil C \cdot \varepsilon \cdot \left(\varepsilon + \frac{1}{n-1}\right) \cdot (2d+1) \right\rceil,$$

for some universal constant C.

We now divide the problem into two cases. In the case that $C\varepsilon(\varepsilon + \frac{1}{|S|})(2d+1) < 1$, the support size of each element in S is exactly one, therefore $|S| \leq 2d+1$. Moreover, since $\varepsilon^{-2} \leq n$ we have

$$d = \Omega(|S|) = \Omega(n) = \Omega\left(\frac{\log n}{\varepsilon^2 \log(1/\varepsilon)}\right),$$

which completes the proof for this case.

In the second case, $C\varepsilon(\varepsilon + \frac{1}{|S|})(2d+1) \geq 1$. In this case since $\frac{1}{|S|} = O(\varepsilon)$, hence the support size of each element in S is bounded by $O(\varepsilon^2 d)$. There are at most $\exp(O((\varepsilon^2 \log(1/\varepsilon))d))$ different supports of size $O(\varepsilon^2 d)$ for elements in S. Moreover, the balls of radius 1/2 around points in S are disjoint and they are all inside the ball of radius 1.5 around origin. Therefore, if we fix a support of size $O(\varepsilon^2 d)$, there are at most $3^{O(\varepsilon^2 d)} = \exp(O(\varepsilon^2 d))$ different points in S that can share this support.

This in turn implies that

$$\begin{aligned} |S| &\leq \exp(O((\varepsilon^2 \log(1/\varepsilon))d)) \cdot \exp(O(\varepsilon^2 d)) \\ &\leq \exp(O(\varepsilon^2 \log(1/\varepsilon) + \varepsilon^2)d) \\ &\leq \exp(O(\varepsilon^2 \log(1/\varepsilon)d)), \end{aligned}$$

hence, $d = \Omega\left(\frac{\log |S|}{\varepsilon^2 \log(1/\varepsilon)}\right)$, completing the proof. □

Now, we need to prove Lemma 2.44. We start by stating some simple properties of total variation distance.

For a finite set S and measures $\eta, \mu : 2^S \to [0, \infty)$, we define

$$\min(\nu, \eta)(T) = \sum_{x \in T} (\min\{\mu(\{x\}), \eta(\{x\})\}).$$

For $k \in \mathbb{N}$, and measures $\mu, \eta \in \mathcal{M}_k$, we have

$$\|\mu - \eta\|_{TV} = \frac{1}{2}(\mu([k]) + \eta([k])) - \min(\mu, \eta)([k]). \tag{2.103}$$

We also use the following partial order on finite measures on the set S: $\mu \preceq \eta$, if and only if for all $T \in S$, $\mu(T) \leq \eta(T)$.

The following observation is immediate from (2.103).

Observation 2.45. *For $k \in \mathbb{N}$, and measures $\mu, \eta, \mu', \eta' \in \mathcal{M}_k$, such that $\mu' \preceq \mu$ and $\eta' \preceq \eta$, if for some $\varepsilon \geq 0$,*

$$\|\mu - \eta\|_{TV} \geq \frac{1}{2}(\mu([k]) + \eta([k])) - \varepsilon,$$

then

$$\|\mu' - \eta'\|_{TV} \geq \frac{1}{2}(\mu'([k]) + \eta'([k])) - \varepsilon.$$

The next lemma is the final ingredient that we need to prove Lemma 2.44.

Lemma 2.46. *Let $\delta \in (0,1)$ and $S \subset [0,\infty)$ such that*

$$\delta \cdot (|S|-1) \cdot \sum_{x \in S} x \geq \sum_{x,y \in S, x \neq y} \min(x,y). \tag{2.104}$$

Then there exists a set $T \subseteq S$, such that $\sum_{x \in T} x \geq \frac{1}{2} \sum_{x \in S} x$ and $|T| \leq \lceil \delta(|S|-1) \rceil$.

Proof. Let $n = |S|$, and let $a_1 \geq \ldots \geq a_n \geq 0$ be the elements of S in decreasing order. We have

$$\sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} \min(a_i, a_j) = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} a_{\max(i,j)} = \sum_{i=1}^{n} 2(i-1) a_i.$$

Let $k = \lceil \delta(|S|-1) \rceil$, we divide the above sum into two parts

$$\sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} \min(a_i, a_j) = \sum_{i=1}^{k} 2(i-1)a_i + \sum_{i=k+1}^{n} 2(i-1)a_i \geq \sum_{i=k+1}^{n} 2(i-1)a_i \geq 2k \sum_{i=k+1}^{n} a_i \geq 2\delta(|S|-1) \sum_{i=k+1}^{n} a_i.$$

Combining this inequality and (2.104) implies that $\sum_{i=k+1}^{n} a_i \leq \frac{1}{2} \sum_{x \in S} x$, therefore $\sum_{i=1}^{k} a_i \geq \frac{1}{2} \sum_{x \in S} x$. Hence the set $T = \{a_1, \ldots, a_k\}$ satisfies both conditions of the lemma.

□

Proof of Lemma 2.44. We prove this lemma by showing that each of the following statements imply the next one.

I) There exists an ε-unrelated set $S \subset \mathcal{P}_k$ of size n.

II) There exists an ε-unrelated set $S \subset \mathcal{M}_k$ of size n such that

(a) for all $\mu \in S$, $\mu([k]) \leq 1$;

(b) $\sum_{\mu \in S} \mu([k]) \geq n/4$;

(c) $\sum_{\mu \in S} |\text{supp}(\mu)| < (2\varepsilon n + 1)k$;

III) There exists an ε-unrelated set $S \subset \mathcal{M}_k$ of size at least $n/14$ such that

(a) for all $\mu \in S$, $|\text{supp}(\mu)| < 14\left(2\varepsilon + \frac{1}{n}\right)k$;

(b) for all $\mu \in S$, we have $\mu([k]) \geq 1/8$;

IV) There exists a set that satisfies all conditions of this Lemma.

To make the notation more succinct for a given set $S \subset \mathcal{M}_k$, we define,

$$\Delta_S = \sum_{\mu,\eta \in S, \mu \neq \eta} \min(\mu, \eta).$$

Note that, if for some $\varepsilon \in [0,1]$, $S \in \mathcal{P}_k$ is ε-unrelated, then (2.103) implies that

$$\Delta_S([k]) \leq \sum_{\mu,\eta \in S, \mu \neq \eta} \frac{1}{2}(\mu([k]) + \eta([k])) - \|\mu - \eta\|_{TV} \leq \sum_{\mu,\eta \in S, \mu \neq \eta} 1 - (1 - \varepsilon) = \varepsilon |S| \cdot (|S| - 1). \tag{2.105}$$

I $\Rightarrow$ II: Let $S_{\text{I}} \subset \mathcal{M}_k$ be an ε-unrelated, and let X be a random variable with state space $\{1, \ldots, k\}$ such that

$$\Pr(X = i) = \frac{\sum_{\mu \in S_{\text{I}}} \mu(\{i\})}{\sum_{\mu \in S_{\text{I}}} \mu([k])}.$$

We have

$$\mathbb{E}\left[\frac{\Delta_{S_{\text{I}}}(X)}{\sum_(\mu \in S_{\text{I}}) \mu(X)}\right] = \frac{1}{|S_{\text{I}}|} \sum_{i=1}^{k} \Delta_{S_{\text{I}}}(i) = \frac{1}{|S_{\text{I}}|} \Delta_{S_{\text{I}}}([k]) \overset{(2.105)}{\leq} \varepsilon(|S_{\text{I}}| - 1),$$

Markov's inequality implies that

$$\Pr\left(\frac{\Delta_{S_{\text{I}}}(X)}{\sum_{\mu \in S_I} \mu(X)} \leq 2\varepsilon(|S_{\text{I}}| - 1)\right) \geq \frac{1}{2}.$$

Now let

$$A = \left\{ i : \frac{\Delta(S_{\mathrm{I}})_i}{\sum_{\mu \in S_I} \mu(\{i\})} \le 2\varepsilon(|S_{\mathrm{I}}| - 1) \right\}.$$

We have

$$\sum_{\mu \in S_{\mathrm{I}}} \mu(A) = \sum_{\mu \in S_{\mathrm{I}}} \sum_{i \in A} \Pr(X = i) \cdot \sum_{\mu \in S_{\mathrm{I}}} \mu([k]) \ge \frac{1}{2} \sum_{\mu \in S_{\mathrm{I}}} \mu([k]). \tag{2.106}$$

By Lemma 2.46 for all $i \in A$, there exists a set $W_i \subseteq S_{\mathrm{I}}$ such that $|W_i| \le \lceil 2\varepsilon(|S_{\mathrm{I}}| - 1) \rceil$, and

$$\sum_{\mu \in W_i} \mu(\{i\}) \ge \frac{1}{2} \sum_{\mu \in S_{\mathrm{I}}} \mu(\{i\}). \tag{2.107}$$

For $\mu \in S_{\mathrm{I}}$, let $Y_\mu = \{i : \mu \in W_i\}$.

Let $S_{\mathrm{II}} = \{\mu \circ I_{Y_\mu}\}_{\mu \in S_{\mathrm{I}}}$. Since $\mu \circ I_{Y_\mu} \preceq \mu$, S_{II} satisfies II(a). Furthermore, Observation 2.45 implies that the set $\{\mu \circ I_{Y_\mu}\}_{\mu \in S_{\mathrm{I}}}$ is ε-unrelated. Inequality

$$\sum_{\mu \in S_I} \mu(Y_\mu) \overset{(2.107)}{\ge} \frac{1}{2} \sum_{i \in A} \sum_{\mu \in S_I} \mu(i) = \frac{1}{2} \sum_{\mu \in S_I} \mu(A) \overset{(2.106)}{\ge} |S_{\mathrm{I}}|/4$$

implies that S_{II} satisfies II(b). And, Condition II(c) also holds because

$$\sum_{\mu \in S_{\mathrm{I}}} |\mathrm{supp}(\mu \circ I_{Y_\mu})| = \sum_{i \in A} |W_i| < 2\varepsilon(|S_{\mathrm{I}}| - 1)|A| + |A| \le (2\varepsilon|S_{\mathrm{I}}| + 1)k.$$

II $\Rightarrow$ III: Suppose that S_{II} is an ε-unrelated set $S \subset \mathcal{M}_k$ of size n that satisfies all the conditions of II. We have $\max\{\mu([k])\}_{\mu \in S_{\mathrm{II}}} \le 1$, and $\sum_{\mu \in S_{\mathrm{II}}} \mu([k]) \ge |S_{\mathrm{II}}|/4$. Therefore, there exists a set $S' \subseteq S_{\mathrm{II}}$ such that for all $\mu \in S'$, $\mu([k]) \ge 1/8$, and

$$|S'| \ge \left(\frac{1/4 - 1/8}{1 - 1/8} \right) |S_{\mathrm{II}}| \ge \frac{n}{7}.$$

By Markov's inequality, there exists a set S_{III} such that $|S_{\mathrm{III}}| \ge \frac{1}{2}|S'| \ge \frac{1}{14}|S_{\mathrm{II}}|$, where for all $\mu \in S_{\mathrm{III}}$,

$$\mathrm{supp}(\mu) \le 2\frac{\sum_{\mu \in S'} |\mathrm{supp}(\mu)|}{|S'|} \le 2\frac{\sum_{\mu \in S_{\mathrm{II}}} |\mathrm{supp}(\mu)|}{|S'|} \le 2\frac{\sum_{\mu \in S_{\mathrm{II}}} |\mathrm{supp}(\mu)|}{|S_{\mathrm{II}}|/7} \overset{\mathrm{II(c)}}{\le} 14k(2\varepsilon + \frac{1}{n}).$$

The set S_{III} has size at least $\frac{n}{14}$ and by definition it satisfies conditions (a) and (b) of III.

III $\Rightarrow$ IV: Suppose S_{III} is a an ε-unrelated set $S \subset \mathcal{M}_k$ of size at least $n/14$. For each element in $\mu \in S_{\text{III}}$, let $Z_\mu \in \{1,\ldots,k\}$ be the set of $\lceil 16 \cdot \varepsilon \left(14d(2\varepsilon + \frac{1}{n})\right) \rceil$ elements of $\{1,\ldots,k\}$ that has the largest measures with respect to μ (break the ties arbitrarily). Since $\varepsilon \le \frac{1}{8}$, for all $\mu \in S'$ we have

$$\mu(Z_\mu) \ge \frac{1}{8}\left(\frac{16 \cdot \varepsilon(14k(2\varepsilon + \frac{1}{n}))}{14k(2\varepsilon + \frac{1}{n})}\right) = 2\varepsilon. \tag{2.108}$$

Let $S_{\text{IV}} = \left\{\frac{\mu \circ I_{Z_\mu}}{\mu(Z_\mu)}\right\}_{\mu \in S_{\text{III}}}$. It is easy to check that $S_{\text{IV}} \subset \mathcal{P}_k$, and has size $\frac{n}{14}$. Moreover, by our construction for all $\bar{\mu} \in S_{\text{IV}}$, $|\text{supp}(\bar{\mu})| \le \lceil 224\varepsilon(2\varepsilon + \frac{1}{n})d) \rceil$. To complete the proof we need to show S_{IV} is $\frac{1}{2}$-unrelated. Note that $\mu, \eta \in S_{\text{III}}$, by Observation 2.45 implies that

$$\min(\eta \circ I_{Z_\eta}, \mu \circ I_{Z_\mu})([k]) \overset{(2.103)}{=} \frac{\mu(Z_\mu) + \eta(Z_\mu)}{2} - \|\mu \circ I_{Z_\eta} - \mu \circ I_{Z_\eta}\|_{TV} \le \varepsilon.$$

Therefore,

$$\begin{aligned}\left\|\frac{\mu \circ I_{Z_\mu}}{\mu(Z_\mu)} - \frac{\eta \circ I_{Z_\eta}}{\eta(Z_\eta)}\right\|_{TV} &\overset{(2.103)}{=} 1 - \min\left(\frac{\mu \circ I_{Z_\mu}}{\mu(Z_\mu)}, \frac{\eta \circ I_{Z_\eta}}{\eta(Z_\eta)}\right)([k]) \\ &\overset{(2.108)}{\ge} 1 - \min\left(\frac{\mu \circ I_{Z_\mu}}{2\varepsilon}, \frac{\eta \circ I_{Z_\eta}}{2\varepsilon}\right)([k]) \overset{(2.103)}{\ge} \frac{1}{2}.\end{aligned}$$

□

2.5.2 Extension to k-ary trees

In this section we show that any lower bound on the distortion of the maps from of the star graph into ℓ_1^d also gives a lower bound for embedding of complete k-ary trees into ℓ_1^d.

Lemma 2.47. *Let T be a complete k-ary tree of height $h \ge 1$ for some $k > 1$, and let $f : T \to \ell_1^d$ be a map with distortion $(1+\varepsilon)$ for some $0 \le \varepsilon \le \frac{1}{4}$, then there exists a map from $(1 + k^{\lceil h/2 \rceil})$-star into ℓ_1^d with distortion $1 + 4\varepsilon$.*

Note that for any k-ary tree of height h, we have $k^{\lceil h/2 \rceil} = \Omega(\sqrt{|V(T)|})$. Hence, combining this lemma with Theorem 2.3 immediately gives the following corollary.

Corollary 2.48. *There exists a universal constant δ, such that, for any $\varepsilon \in (0,\delta)$, $n \ge N_\varepsilon$ and $k > 1$, any embedding of complete k-ary tree of size n into ℓ_1^d with distortion $1+\varepsilon$ requires dimension $d = \Omega\left(\frac{log(n)}{\varepsilon^2 \log(1/\varepsilon)}\right)$.*

Proof of Lemma 2.47. For the case that $h = 1$ this lemma is vacuous. Suppose now that $h > 1$. We construct a map g from $(1 + k^{\lceil h/2 \rceil})$-star into ℓ_1^d as follows.

Let S be the set of vertices at height $\lceil h/2 \rceil$ (we use the convention that height of root is zero). For any element $v \in S$, pick an arbitrary leaf l_v, of the subtree under the vertex v. We set $g(r) = 0$. For all other vertices of the star we index them with distinct elements of S, and set

$$g(v) = \frac{f(l_v) - f(v)}{h - \lceil h/2 \rceil}.$$

First note that, if the map f is non-expanding then g is also non-expanding. Moreover for two distinct elements $u, v \in S$ we have

$$\begin{aligned} 2(h - \lceil h/2 \rceil) + d_T(u, v) &= d_T(u, v) + d_T(l_v, v) + d_T(l_u, u) \\ &= d_T(l_u, l_v) \\ &\geq \|f(l_u) - f(l_v)\|_1 \\ &\geq \|f((l_u) - f(u)) - (f(l_v) - f(v))\|_1 + \|f(u) - f(v)\|_1 \\ &\geq \|f((l_u) - f(u)) - (f(l_v) - f(v))\|_1 + (1 - \varepsilon) d_T(u, v). \end{aligned}$$

Therefore,

$$\begin{aligned} \|f((l_u) - f(u)) - (f(l_v) - f(v))\|_1 &\geq 2(h - \lceil h/2 \rceil) - \varepsilon d_T(u, v) \\ &\geq 2(h - \lceil h/2 \rceil) - 2\varepsilon \lceil h/2 \rceil \\ &\geq 2(h - \lceil h/2 \rceil) - 4\varepsilon (h - \lceil h/2 \rceil) \\ &\geq (2 - 4\varepsilon)(h - \lceil h/2 \rceil). \end{aligned}$$

Since $\varepsilon \leq 1/4$, using the above inequality we can bound the distortion for the map g by $\frac{1}{1-2\varepsilon} \geq 1 + 4\varepsilon$. □

Chapter 3

EMBEDDINGS AND SCALES

3.1 Results

We divide the results presented in this chapter into two categories. We first present our results for ℓ_p, where $p > 1$ and then present the results for the case that $p = 1$. While both approaches to these two cases have some similarities the techniques and the results have some fundamental differences.

3.1.1 Lower Bounds on Embedding into ℓ_p for $p > 1$

Suppose one is given a collection of mappings from some finite metric space (X, d) into a Euclidean space, each of which reflects the geometry at some "scale" of X. Is there a non-trivial way of gluing these mappings together to form a global mapping which reflects the entire geometry of X? In this chapter, we show that the approaches of [57] and [66] are optimal, disproving a conjecture stated in [66].

The Gluing Lemma of [66] (generalizing the approach of [57]) shows that the existence of such a collection $\{\varphi_k\}$ yields a Euclidean embedding of (X, d) with distortion $O(\sqrt{\alpha \log n})$. This is known to be tight when $\alpha = \Theta(1)$ [87] and also when $\alpha = \Theta(\log n)$ [76, 8], but nowhere in between. In fact, in [66], Lee conjectured that one could achieve $O(\alpha + \sqrt{\log n})$ (this is indeed stronger, since one can always construct $\{\varphi_k\}$ with $\alpha = O(\log n)$).

In this chapter, we give a family of examples which shows that the $\sqrt{\alpha \log n}$ bound is tight for any dependence $\alpha(n) = O(\log n)$. In fact, we show more. Let $\lambda(X)$ denote the *doubling constant* of X, i.e. the smallest number λ so that every open ball in X can be covered by λ balls of half the radius. In [57], using the method of "measured descent," the authors show that (X, d) admits a Euclidean embedding with distortion $O(\sqrt{\log \lambda(X) \log n})$. (This is a special case of the Gluing Lemma since one can always find $\{\varphi_k\}$ with $\alpha = O(\log \lambda(X))$ [45]). Again, this bound was known to be tight for $\lambda(X) = \Theta(1)$ [59, 60, 45] and $\lambda(X) = n^{\Theta(1)}$

[76, 8], but nowhere in between. We provide the matching lower bound for any dependence of $\lambda(X)$ on n. We also generalize our method to give tight lower bounds on ℓ_p distortion for every fixed $p > 1$.

Theorem 3.1. *For any positive nondecreasing function $\lambda(n)$, there exists a family of n-vertex metric graphs $\{G_i\}_{i\in\mathbb{N}}$ such that $\lambda(G_i) \lesssim \lambda(n)$, and for every fixed $p > 1$,*

$$c_p(G_i) \gtrsim (\log n)^{1/q}(\log \lambda(n))^{1-1/q},$$

where $q = \max\{p, 2\}$.

3.1.2 Lower Bounds on Embedding into ℓ_1 and NEG

The main theorem that we prove on embedding into negative type metrics and ℓ_1 is the following theorem.

Theorem 3.2. *There exists an $O(1)$-half-snowflake for which any embedding into a metric of negative type incurs distortion at least $(\log n)^{1/3-o(1)}$.*

We refer the readers to Section 1.3.3 for the background and motivation behind the study of Negative Type Metrics and Snowflakes.

Our work also implies a lower bound for embedding over scales in ℓ_1. Suppose that (X, d) is an n-point metric space, and furthermore that for every value $k \in \mathbb{Z}$, there exists a 1-Lipschitz map $\varphi_k : X \to \ell_1$ such that for $x, y \in X$ satisfying $d(x, y) \geq 2^k$,

$$\|\varphi_k(x) - \varphi_k(y)\|_1 \geq \frac{2^k}{\alpha}. \tag{3.1}$$

On the one hand, we have the following.

Theorem 3.3 ([66])**.** *There exists an embedding of (X, d) into ℓ_1 with distortion $O(\sqrt{\alpha \log n})$.*

On the other hand, the work of Cheeger and Kleiner [27] shows that even if $\alpha = O(1)$, the distortion can go to infinity, and [28] gives a definite bound of $\Omega(\log n)^{\delta_0}$ for some small $\delta_0 > 0$. In this chapter we provide a new construction and that yields the following lower bound.

Theorem 3.4. *There exist n-point metric spaces (X, d) which satisfy* (3.1) *with $\alpha = O(1)$, but such that $c_1(X, d) \geq (\log n)^{1/3-o(1)}$. In fact, we even have $c_{\mathsf{NEG}}(X, d) \geq (\log n)^{1/3-o(1)}$.*

We prove the upper bound part of this theorem in Section 3.4.2 and prove the lower bound part in Section 3.4. It is worth mentioning that this results was recently improved by Lee and Sidiropoulos [74].

The k-sum embedding conjecture. In [73], it is conjectured that if a family of finite graphs $\mathcal{F}$ is such that every shortest-path metric supported on a member of $\mathcal{F}$ embeds into ℓ_1 with distortion $O(1)$, then the family $\oplus_k \mathcal{F}$ has the same property, for every $k \in \mathbb{N}$, where the $\oplus_k(\cdot)$ notation denotes the closure of $\mathcal{F}$ under the operation of taking k-sums along cliques (see [73] for a formal description). These authors refer to this as the "k-sum embedding conjecture." The conjecture is open even for $k = 2$. One of the main results of [73] is that the k-sum embedding conjecture, combined with the well-known planar embedding conjecture, implies the GNRS max-flow/min-cut conjecture in excluded-minor families [46].

Our results show that there exists an unweighted graph G whose shortest-path metric embeds into ℓ_1 with constant distortion, but that by taking repeated 2-sums of G with itself, one obtains a graph whose ℓ_1 distortion becomes arbitrarily large. This does not disprove the k-sum conjecture, because we have only considered a single shortest-path metric on G, but it does show that the proof must use something about the entire set of embeddings for metrics on G, as opposed to merely an embedding of the given metric on G.

3.2 Techniques and Overview of the Proofs

To construct the family of graphs that we use in this section we use a recursive construction of graphs based on [65]. We use $G^{\oslash k}$ to denote the following iterated graph: $G^{\oslash 0}$ is a single edge, and $G^{\oslash k+1}$ arises by replacing every edge of $G^{\oslash k}$ with a copy of G, with s and t taking the place of the endpoints of the edge.

Indeed, the base graphs used to prove lower bounds for embedding into ℓ_1 and ℓ_p for $p > 1$ are different. We first present the recursive construction that we use in our constructions which is slight generalization of $\oslash$-product from [65].

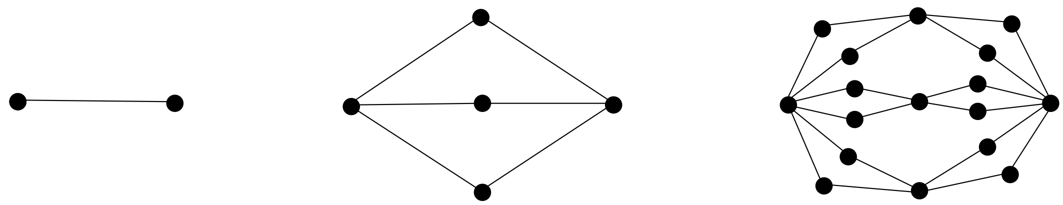


Figure 3.1: A single edge H, $H \oslash K_{2,3}$, and $H \oslash K_{2,3} \oslash K_{2,2}$.

3.2.1 Marked $\oslash$-products

An s-t graph G is a graph which has two distinguished vertices $s, t \in V(G)$. For an s-t graph, we use $s(G)$ and $t(G)$ to denote the vertices labeled s and t, respectively. We define the length of an s-t graph G as $\mathsf{len}(G) = d_{\mathsf{len}}(s,t)$. Throughout chapter, we will only be concerned with *symmetric* s-t graphs, i.e. graphs for which there is a graph automorphism which maps s to t. We assume that all s-t graphs are symmetric in the following definitions. A *marked graph* $G = (V, E)$ is one which carries an additional subset $E_M(G) \subseteq E$ of *marked edges.* Every graph is assumed to be equipped with the trivial marking $E_M(G) = E(G)$ unless a marking is otherwise specified.

Definition 3.5 (Composition of s-t graphs)**.** *Given two marked s-t graphs H and G, define $H \oslash G$ to be the s-t graph obtained by replacing each* marked *edge $(u, v) \in E_M(H)$ by a copy of G. Formally,*

- $V(H \oslash G) = V(H) \cup (E_M(H) \times (V(G) \setminus \{s(G), t(G)\}))$.
- *For every edge $(u, v) \in E(H) \setminus E_M(H)$, there is a corresponding edge in $H \oslash G$.*
- *For every edge $e = (u, v) \in E_M(H)$, there are $|E(G)|$ edges,*

$$\begin{aligned}&\Big\{\Big((e,v_1),(e,v_2)\Big) \mid (v_1,v_2) \in E(G), v_1, v_2 \notin \{s(G), t(G)\}\Big\}\\ &\quad\cup \Big\{\Big(u,(e,w)\Big) \mid (s(G),w) \in E(G)\Big\}\\ &\quad\cup \Big\{\Big((e,w),v\Big) \mid (w,t(G)) \in E(G)\Big\}\end{aligned}$$

- *The marked edges of* $H \oslash G$ *are precisely those introduced in the previous step which correspond to marked edges in* G.

- $s(H \oslash G) = s(H)$ *and* $t(H \oslash G) = t(H)$.

If H *and* G *are equipped with length functions* $\mathsf{len}_H, \mathsf{len}_G$, *respectively, we define* $\mathsf{len} = \mathsf{len}_{H \oslash G}$ *as follows. Using the preceding notation, for every edge* $e = (u, v) \in E_M(H)$,

$$\begin{aligned}
\mathsf{len}\left((e, v_1), (e, v_2)\right) &= \frac{\mathsf{len}_H(e)}{d_{\mathsf{len}_G}(s(G), t(G))} \mathsf{len}_G(v_1, v_2) \\
\mathsf{len}\left(u, (e, w)\right) &= \frac{\mathsf{len}_H(e)}{d_{\mathsf{len}_G}(s(G), t(G))} \mathsf{len}_G(s(G), w) \\
\mathsf{len}\left((e, w), v\right) &= \frac{\mathsf{len}_H(e)}{d_{\mathsf{len}_G}(s(G), t(G))} \mathsf{len}_G(w, t(G)).
\end{aligned}$$

This choice implies that $H \oslash G$ *contains an isometric copy of* $(V(H), d_{\mathsf{len}_H})$.

See Figure 3.5 for an example.

Definition 3.6 (Recursive composition)**.** *Given a marked* s-t *graph* G *and a number* $k \in \mathbb{N}$, *we define* $G^{\oslash k}$ *inductively by letting* $G^{\oslash 0}$ *be a single edge of unit length, and setting* $G^{\oslash k} = G^{\oslash k-1} \oslash G$.

The following result is straightforward.

Lemma 3.7 (Associativity of $\oslash$)**.** *For any three graphs* s-t *graphs* A, B, C, *the graphs* $(A \oslash B) \oslash C$ *and* $A \oslash (B \oslash C)$ *are isomorphic as* s-t *graphs, and isometric as metric spaces, under the natural mapping.*

Definition 3.8. *For two graphs* G, H, *a subset of vertices* $X \subseteq V(H)$ *is said to be a copy of* G *if there exists a bijection* $f : V(G) \to X$ *with distortion 1.*

Now we make the following two simple observations about copies of H and G in $H \oslash G$.

Observation 3.9. *The graph* $H \oslash G$ *contains* $|E_M(H)|$ *distinguished copies of the graph* G, *one copy corresponding to each edge in* H.

Observation 3.10. *The subset of vertices* $V(H) \subseteq V(H \oslash G)$ *form an isometric copy of* H.

3.2.2 Overview of the proof for embeddings into ℓ_p, with $p > 1$

In some sense, our lower bound examples are an interpolation between the multi-scale method of [87] and [59], and the expander Poincaré inequalities of [76, 8, 79]. We start with a vertex-transitive expander graph G on m nodes. If D is the diameter of G, then we create $D+1$ copies $G^1, G^2, \dots, G^{D+1}$ of G where $u \in G^i$ is connected to $v \in G^{i+1}$ if (u, v) is an edge in G, or if $u = v$. We then connect a vertex s to every node in G^1 and a vertex t to every node in G^{D+1} by edges of length D. This yields the graph $\vec{G}$ described in Section 3.3.2.

In Section 3.3.3, we show that whenever there is a non-contracting embedding f of $\vec{G}$ into ℓ_2, the following holds. If $\gamma = \frac{\|f(s)-f(t)\|}{d_{\vec{G}}(s,t)}$, then some edge of $\vec{G}$ gets stretched by at least $\sqrt{\gamma^2 + \Omega(\log m)^2}$, i.e. there is a "stretch increase." This is proved by combining the uniform convexity of ℓ_2 (i.e. the Pythagorean theorem), with the well-known contraction property of expander graphs mapped into Hilbert space. To convert the "average" nature of this contraction to information about a specific edge, we symmetrize the embedding over all automorphisms of G (which was chosen to be vertex-transitive).

To exploit this stretch increase recursively, we construct a graph $\vec{G}^{\oslash k}$. Now a simple induction shows that in a non-contracting embedding of $\vec{G}^{\oslash k}$, there must be an edge stretched by at least $\Omega(\sqrt{k}\log m)$. In Section 3.3.4, a similar argument will be made for ℓ_p distortion, for $p > 1$, but here we have to argue about "quadrilaterals" instead of "triangles" (in order to apply the uniform convexity inequality in ℓ_p), and it requires slightly more effort to find a good quadrilateral.

Finally, we observe that if $\widetilde{G}$ is the graph formed by adding two tails of length $3D$ hanging off s and t in $\vec{G}$, then (following the analysis of [59, 60]), one has $\log \lambda(\widetilde{G}^{\oslash k}) \lesssim \log m$. The same lower bound analysis also works for $\widetilde{G}^{\oslash k}$, so since $n = |V(\widetilde{G}^{\oslash k})| = 2^{\Theta(k \log m)}$, the lower bound is

$$\sqrt{k}\log m \approx \sqrt{\log m \log n} \gtrsim \sqrt{\log \lambda(\widetilde{G}^{\oslash k}) \log n},$$

completing the proof.

3.2.3 Overview of the proof for embeddings into ℓ_1 and NEG

The arguments and proofs for embeddings into ℓ_1 are more delicate and significantly more complicated than the case for $p > 1$. The main difference between these two case is that unlike other ℓ_p metric spaces, the shortest path between two points in ℓ_1 is not unique, hence the quadrilateral inequality that arises from uniform convexity does not hold for ℓ_1. Instead of relying on the quadrilateral inequality, we use the notion of "efficienct" embedding to prove our lower bounds[1]. We will discuss this in depth in Section 3.4.1. Our proof uses the following general framework:

i) We first construct a base graph G;

ii) Next, we show that if there is a low distortion embedding of $G^{\oslash k}$ into any metric space then there is an "efficient" embedding G;

iii) Finally, we show that there are no efficient embedding of G into ℓ_1 or NEG.

In the rest of this section we give a high level overview of each step of the proof.

Construction of the base graph

Let G be an unweighted graph with two distinguished vertices $s, t \in V(G)$.

For a parameter $m \in \mathbb{N}$, consider now the graph H_m constucted as follows. Let Q_m be the m-dimensional hypercube graph, and write $V(Q_m) = B_m \cup R_m$, where B_m and R_m denote the nodes of even and odd parity, respectively. Then Q_m is bipartite with respect to the partition (B_m, R_m). H_m is the graph which consists of $2m$ layers of the form

$$B_m^{(1)} R_m^{(1)} B_m^{(2)} R_m^{(2)} B_m^{(3)} R_m^{(3)} \cdots B_m^{(m)} R_m^{(m)}, \tag{3.2}$$

where $B_m^{(i)}$ and $R_m^{(i)}$ denote disjoint copies of B_m and R_m for $i = 1, 2, \ldots, m$, and hypercube edges are present between every pair of adjacent layers.

---

[1]in fact our techniques yield a lower bound for embedding into NEG which contains all finite ℓ_1 metrics

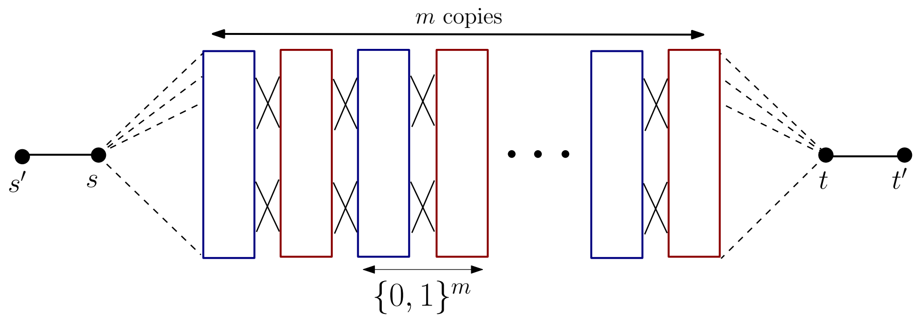


Figure 3.2: The "string of cubes" graph, which we recursively compose with itself.

We also add to H_m two nodes: s' connected to all the nodes of $B_m^{(1)}$ by edges of length m and a t' connected the nodes of $R_m^{(m)}$ by edges of length m. Finally, we add two distinguished vertices s and t and connect s to s' and t to t' by two unmarked edges of length $\frac{1}{m}$.

We call H_m the "string of cubes" graph. See Figure 3.2. Our final construction is of the form $G_{k,m} = \hat{H}_m^{\oslash k}$ for appropriate values of $k, m \in \mathbb{N}$, where $\hat{H}_m$ is a slightly modified version of H_m. We use $d_{k,m}$ to denote the shortest-path metric on $G_{k,m}$.

Differentiation and Efficiency

The analysis is based on a "differentiation"-type argument. At a very broad level, we first argue that any low-distortion embedding must be well-controlled on a small piece of our lower bound space, and then show that any well-controlled embedding is quite rigid in structure, allowing us to prove a lower bound.

Generalizations of classical differentiation theory have played a prominent role in proving the non-existence of bi-Lipschitz embeddings between various spaces, when the target space Z is sufficiently nice; see, for instance [90, 23, 71, 11, 24]. But this approach does not apply to targets like ℓ_1 which does not even guarantee differentiability for Lipschitz mappings $f : \mathbb{R} \to \ell_1$.

More recently, however, Cheeger and Kleiner [27, 25] have successfully applied weaker notions of differentiability to the study of ℓ_1 embeddings of the Heisenberg group. Subsequent papers [65, 26, 28] continue this theme. Ours is the first work to apply these techniques to

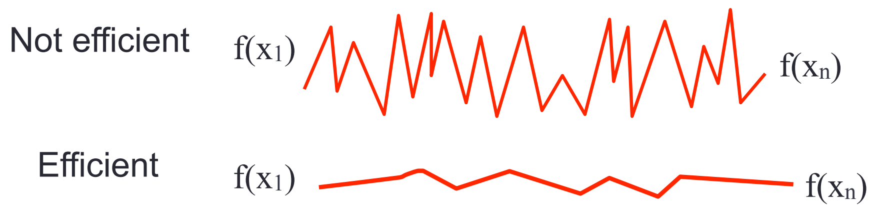


Figure 3.3: Demonsatration of of a non-efficient and an efficient embedding of line into $\mathbb{R}^2$.

prove non-embeddability into negative-type metrics.

A central role will be played by the *efficiency* of various "paths" in metric spaces, following [38, 65]. Consider a finite sequence of points equipped with a non-negative symmetric function ρ (which may *not* satisfy the triangle inequality), $\mathcal{S} = \{x_1, x_2, \ldots, x_k\}$. We say that $\mathcal{S}$ is ε-*efficient (with respect to* dist*)* if

$$\sum_{i=1}^{k-1} \rho(x_i, x_{i+1}) \leq (1+\varepsilon)\, \rho(x_1, x_k).$$

Note that if ρ is a metric, then the left-hand side is always at least $\rho(x_1, x_k)$, by the triangle inequality. See Figure 3.3 for demonstration.

The first key aspect of our approach is that we reduce the embeddability of $G^{\oslash k}$ to the study of specific types of embeddings for the base graph G. For a graph G and two nodes $s, t \in V(G)$, let $\mathcal{P}_{s,t}(G)$ denote the set of all s-t shortest-paths in G. In Section 3.4.1, we prove a quantitative variant of the following theorem, based on the "coarse differentiation" methodology of [38].

Theorem 3.11. *Let (Y, d_Y) be any metric space, and suppose that $c_Y(G^{\oslash k}) \leq D$ for all $k = 1, 2, \ldots$. Then for every $\varepsilon > 0$, there exists an embedding $f : G \to Y$ with distortion at most D, and such that for every sequence $\{x_1, x_2, \ldots, x_r\} \in \mathcal{P}_{s,t}(G)$, the sequence $\{f(x_1), f(x_2), \ldots, f(x_r)\}$ is ε-efficient in (Y, d_Y).*

A more general result was proved in [65], but in this work we obtain a quantitative improvement for the special case of iterated graphs.

Snowflake embedding: On the other hand, in the setting of half-snowflakes, we have a partial converse to the preceding theorem. If there is an embedding

$$f : (G, \sqrt{d_G}) \to \ell_2$$

with distortion D, and such that for every sequence

$$\{x_1, x_2 \ldots, x_r\} \in \mathcal{P}_{s,t}(G),$$

the sequence $\{f(x_1), f(x_2), \ldots, f(x_r)\}$ is ε-efficient with respect to the distance $\|f(x_i) - f(x_j)\|_2^2$, then for every $k = O(1/\varepsilon)$, the graph $G^{\oslash k}$ is an $O(D)$-half-snowflake. (We are actually only able to prove this for a modification of the graph $G^{\oslash k}$.)

Because of these two results, we are able to focus on a separation between embeddings of our base graph H_m into negative-type metrics and half-snowflakes, respectively, with the additional property that the embeddings are ε-efficient on s-t shortest-paths. Analyzing efficient (and approximately efficient embeddings) is the technical core of our approach, which we now address.

Poincaré Inequalities, Lower Bounds and Efficiency

To explain some of the difficulty in separating NEG metrics from half-snowflakes, we remark that for general ℓ_p embeddings, all distortion lower bounds can be proved using 1-dimensional Poincaré inequalities. However, 1-dimensional Poincaré inequalities cannot separate these two classes and a high-dimensional argument is necessary.

Let (X, d) be a finite metric space and fix $p \geq 1$. For any two symmetric weight functions $\omega_1, \omega_2 : X \times X \to \mathbb{R}_+$, we say that X satisfies an ℓ_p-Poincaré inequality with weights ω_1 and ω_2 if for every mapping $f : X \to \mathbb{R}$,

$$\sum_{x,y \in X} \omega_1(x, y)|f(x) - f(y)|^p \leq \sum_{x,y \in X} \omega_2(x, y)|f(x) - f(y)|^p.$$

In this case, by integrating, we obtain that for every $f : X \to \ell_p$,

$$\sum_{x,y \in X} \omega_1(x, y)\|f(x) - f(y)\|_p^p \leq \sum_{x,y \in X} \omega_2(x, y)\|f(x) - f(y)\|_p^p.$$

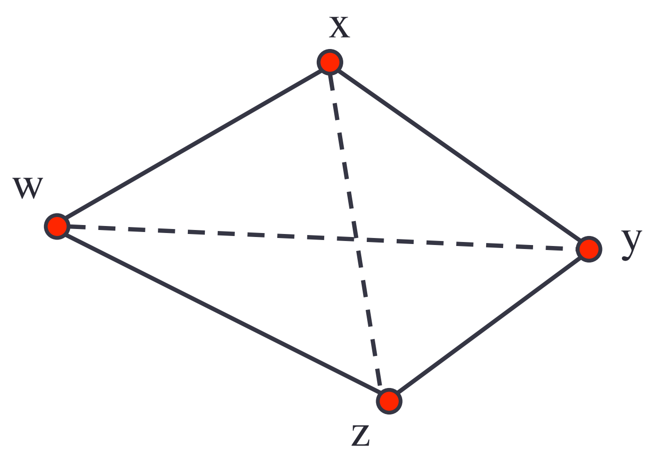


Figure 3.4: An example of Poincaré inequalities is quadrilaterals inequality $\|x-z\|_2^2 + \|y-w\|_2^2 \leq (\|x-y\|_2^2 + \|y-z\|_2^2 + \|z-w\|_2^2 + \|w-x\|_2^2)$.

Finally, if $f : X \to \ell_p$ has finite distortion, the preceding inequality says that

$$\frac{1}{\|f^{-1}\|_{\mathrm{Lip}}^p} \sum_{x,y \in X} \omega_1(x,y) d(x,y)^p \leq \|f\|_{\mathrm{Lip}}^p \sum_{x,y\in X} \omega_2(x,y) d(x,y)^p,$$

which implies that the distortion of f satisfies

$$\mathsf{dist}(f) = \|f\|_{\mathrm{Lip}} \cdot \|f^{-1}\|_{\mathrm{Lip}} \geq \left(\frac{\sum_{x,y\in X} \omega_1(x,y) d(x,y)^p}{\sum_{x,y\in X} \omega_2(x,y) d(x,y)^p} \right)^{1/p}.$$

Thus every ℓ_p-Poincaré inequality which (X,d) satisfies yields a lower bound on the distortion required to embed (X,d) into ℓ_p. It is well-known that a simple duality argument implies that this method is *universal* for proving distortion lower bounds (see [78, Ch. 15]). Thus even proving lower bounds on embedding (X,d) into infinite-dimensional ℓ_p spaces can be done entirely in the setting of 1-dimensional maps.

On the other hand, it is clear that ℓ_2-Poincaré inequalities for the space $(X, \sqrt{d})$ cannot provide a separation between half-snowflakes and NEG embeddings. In order to separate these two classes, for mappings $f : X \to \ell_2$, one has to use the high-dimensional structure of the image $f(X)$. We do this using the "differentiation" methodology discussed in the next section, since we are able to get very strong local control on f, allowing us to give a rigid interpretation of the high-dimensional structure of f when it is an NEG embedding.

A lower bound for efficient embeddings into NEG: Poincaré boosting. Recall that H_m consists of m hypercubes $C_1, C_2, \ldots, C_m$ strung together as in (3.2). First, we recall the classical Poincaré inequality of Enflo [35] for the discrete m-cube. For any $f : Q_m \to \mathbb{R}$, we have

$$\mathbb{E}_{x \in Q_m} |f(x) - f(\bar{x})|^2 \le \sum_{i=1}^{m} \mathbb{E}_{x \in Q_m} |f(x) - f(x \oplus e_i)|^2, \tag{3.3}$$

where we use $\bar{x}$ to denote x with all coordinates flipped, and $x \oplus e_i$ to denote x with the ith coordinate flipped. By integrating, we easily conclude that for any $f : Q_m \to \ell_2$,

$$\mathbb{E}_{x \in Q_m} \|f(x) - f(\bar{x})\|^2 \le \sum_{i=1}^{m} \mathbb{E}_{x \in Q_m} \|f(x) - f(x \oplus e_i)\|^2. \tag{3.4}$$

Obviously, this inequality does not yield any lower bound on the distortion for embedding Q_m into NEG (since it embeds into ℓ_1, and hence NEG isometrically).

But suppose we are given an embedding $g : H_m \to \ell_2$ which is an isometric negative-type embedding, in the sense that $\|g(x) - g(y)\|^2 = d_{H_m}(x, y)$ for all $x, y \in H_m$. Now, for each i, let $g_i = g|_{C_i} : Q_m \to \ell_2$ be the restriction of g to the ith copy of Q_m in H_m, where we think of all the maps $\{g_i\}_{i=1}^m$ as having the same domain. If we simply applied (3.4) to each $f = g_i$ and summed the resulting inequalities, we would again achieve no non-trivial lower bound.

Instead, we apply (3.4) to the mapping $f = g_1 + g_2 + \cdots + g_m$. By the strictness of the property that g is an isometry (when the range is considered with the squared norm $\|\cdot\|^2$), all the vectors $\{g_i(x) - g_i(\bar{x})\}_{i=1}^m$ are colinear, and one concludes that,

$$\|f(x) - f(\bar{x})\|^2 = \left(\sum_{j=1}^{m} \|g_j(x) - g_j(\bar{x})\| \right)^2. \tag{3.5}$$

On the other hand, if we (by abuse of notation) consider a shortest-path in H_m of the form

$$x \;-\; x \oplus e_i \;-\; x \;-\; x \oplus e_i \;-\; x \;-\; x \oplus e_i \;-\; \cdots$$

(where the elements of the path lie in the respective sets $B_m^{(1)}, R_m^{(1)}, B_m^{(2)}, R_m^{(2)}, \ldots$), then by the fact that g is an isometry, for any pairs of adjacent nodes x, x' and y, y' in such a path, we have $g(x) - g(x')$ and $g(y) - g(y')$ being orthogonal. This implies that for every $x \in Q_m$ and $i \in [m]$, we have for every $j, k \in [m]$, the property that $g_j(x) - g_j(x \oplus e_i)$ and

$g_k(x) - g_k(x \oplus e_i)$ are orthogonal (actually, this only holds for j and k of the same parity, but ignore this small issue).

We conclude that if $f = g_1 + g_2 + \cdots g_m$, then

$$\|f(x) - f(x \oplus e_i)\|^2 = \sum_{j=1}^{m} \|g_j(x) - g_j(x \oplus e_i)\|^2. \tag{3.6}$$

From (3.5), (3.6), and (3.4), we get a "boosted" Poincaré inequality of the form

$$\mathbb{E}_{x \in Q_m} \left(\sum_{j=1}^{m} \|g_j(x) - g_j(\bar{x})\| \right)^2 \leq \sum_{j=1}^{m} \sum_{i=1}^{m} \mathbb{E}_{x \in Q_m} \|g_j(x) - g_j(x \oplus e_i)\|^2. \tag{3.7}$$

Notice that to obtain this inequality, we needed to use the fully high-dimensional version (3.4) instead of (3.3), because we used the high-dimensional relationship between the various maps $\{g_i\}_{i=1}^m$.

Now, if we were told in advance that (3.7) holds, and also that each g_j has distortion at most D (again, when the range is considered with the squared distance $\|\cdot\|^2$), it would immediately yield a lower bound of $D \geq m$. (Assuming each g_j is 1-Lipschitz, the right-hand side is at most m^2, while the left-hand side is at least m^3/D.)

Of course, we started with the assumption that g was isometric, so this simply proves that H_m does not admit an isometric negative-type embedding. But now the main point is that every aspect of the preceding argument is *robust.* In Section 3.4.8, we prove a stable version of (3.5) using a distortion bound for g, and a stable version of (3.6) using the assumption that g is ε-efficient on a large fraction of s-t shortest-paths in H_m.

Combining all this together in Section 3.4.8 shows that H_m does not admit a low-distortion negative-type metric which is ε-efficient on most s-t paths (for ε small enough). Combined with a differentiation theorem like Theorem 3.11, this shows that the iterated graph $G_{k,m}$ does not admit a low-distortion negative-type metric for k, m large enough.

An upper bound for efficient embeddings into half-snowflakes. The preceding discussion yields only part of the separation between half-snowflakes and negative-type metrics. For the other side, using the "snowflake embedding" method mentioned earlier for the iterated graph $G_{k,m}$, it suffices to construct an embedding $f : (H_m, \sqrt{d_{H_m}}) \to \ell_2$ which has small distortion, and such that every s-t shortest-path in H_m is mapped 0-efficiently, when the range is considered with the squared distance $\|\cdot\|^2$. To illustrate how this is done, we will argue for the path metric P on the points $\{1, 2, \ldots, n\}$. In the actual construction, the following argument is carried out for all the shortest s-t paths simultaneously.

Suppose that $f : P \to \ell_2$ satisfies, for all $x, y \in P$,

$$\frac{|x-y|}{D} \leq \|f(x) - f(y)\|^2 \leq |x-y|. \tag{3.8}$$

Let $v_0 = \frac{f(n)-f(1)}{\|f(n)-f(1)\|}$, and put $\alpha_i = \langle v_0, f(i) \rangle$. We will also need to make the assumption that

$$\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_n, \tag{3.9}$$

which will be satisfied in our constructions.

Now, write every point $i \in P$ in the form

$$f(i) = \alpha_i v_0 + v_i,$$

where $\langle v_i, v_0 \rangle = 0$. Consider the mapping $g(i) = \alpha_i v_0 + \delta v_i$, for some $\delta \in [0, 1]$.

In this case, we have

$$\begin{aligned}\sum_{i=1}^{n-1} \|g(i+1) - g(i)\|^2 &= \sum_{i=1}^{n-1} (\alpha_{i+1} - \alpha_i)^2 + \delta^2 \|v_{i+1} - v_i\|^2 \\ &\leq \sum_{i=1}^{n-1} (\alpha_{i+1} - \alpha_i)^2 + \delta^2 n,\end{aligned}$$

where in the last inequality we have used (3.8).

Note that $\|f(n) - f(1)\| = \sum_{i+1} |\alpha_{i+1} - \alpha_i|$ by (3.9), and we have $|\alpha_{i+1} - \alpha_i| \leq 1$ by (3.8), hence

$$\sum_{i=1}^{n-1} (\alpha_{i+1} - \alpha_i)^2 \leq \|f(n) - f(1)\|.$$

On the other hand, $\|g(n) - g(1)\|^2 = \|f(n) - f(1)\|^2 \geq n/D$. It follows that for some value $\delta \gtrsim 1/\sqrt{D}$, we will have

$$\|g(n) - g(1)\|^2 = \sum_{i=1}^{n-1} \|g(i+1) - g(i)\|^2,$$

i.e. the image will of P will be 0-efficient. This gives us a general way to obtain 0-efficient embeddings for half-snowflakes, which would not work for negative-type embeddings (because in the process of decreasing δ, triangle inequalities that may have been satisfied in the image could become violated). In Section 3.4.6 we give the details of this embedding and prove an upper bound for the distortion.

3.3 Glueing over scales

3.3.1 Additional Definitions and Notations

In this section, we use $\mathsf{Aut}(G)$ to denote the group of automorphisms of G.

For $x \in X$, $r \in \mathbb{R}_+$, we define the open ball $B(x,r) = \{y \in X : d(x,y) < r\}$. Recall that the *doubling constant* of a metric space (X,d) is the infimum over all values λ such that every ball in X can be covered by λ balls of half the radius. We use $\lambda(X,d)$ to denote this value.

3.3.2 Metric Construction

Construction. In some sense, our lower bound examples are an interpolation between the multi-scale method of [87] and [59], and the expander Poincaré inequalities of [76, 8, 79]. We start with a vertex-transitive expander graph G on m nodes. If D is the diameter of G, then we create $D+1$ copies $G^1, G^2, \ldots, G^{D+1}$ of G where $u \in G^i$ is connected to $v \in G^{i+1}$ if (u,v) is an edge in G, or if $u = v$. We then connect a vertex s to every node in G^1 and a vertex t to every node in G^{D+1} by edges of length D. We call this graph $\vec{G}$.

A doubling version, following Laakso. Let $\vec{G}$ be a s-t graph as in described above, with $D = \operatorname{diam}(G)$, and let $s' = s(\vec{G}), t' = t(\vec{G})$. Consider a new metric s-t graph $\widetilde{G}$, which has two new vertices s, t and two new edges $(s,s'), (t',t)$ with $\mathsf{len}(s,s') = \mathsf{len}(t',t) = 3D$.

Claim 3.12. *For any graph G with $|V(G)| = m$, and any $k \in \mathbb{N}$, we have $\log \lambda(\widetilde{G}^{\oslash k}) \lesssim \log m$.*

The proof of the claim is similar to [59, 60], and follows from the following three Observations and Lemmas.

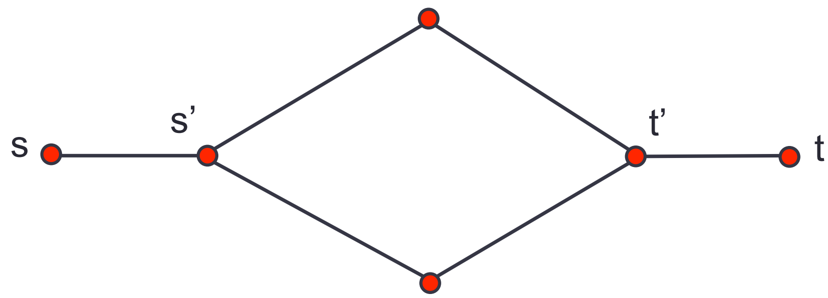


Figure 3.5: Base of Laakso Graph.

We define $\text{tri}(G) = \max_{v \in V(G)}(d_{\mathsf{len}}(s,v) + d_{\mathsf{len}}(v,t))$. For any graph G, we have $\mathsf{len}(\widetilde{G}) = d(s,t) = 9D$, and it is not hard to verify that $\text{tri}(\widetilde{G}^{\oslash k}) \leq \mathsf{len}(\widetilde{G}^{\oslash k})(1 + \frac{1}{9D-1})$. For convenience, let G_0 be the top-level copy of $\widetilde{G}$ in $\widetilde{G}^{\oslash k}$, and H be the graph $\widetilde{G}^{\oslash k-1}$. Then for any $e \in E(G_0)$, we refer to the copy of H along edge e as H_e.

Observation 3.13. *If $r > \frac{\text{tri}(\widetilde{G}^{\oslash k})}{3}$, then the ball $B(x,r)$ in $\widetilde{G}^{\oslash k}$ may be covered by at most $|V(\widetilde{G})|$ balls of radius $r/2$.*

Proof. For any $e \in E(G_0)$, we have $r > \frac{\mathsf{len}(e)}{\mathsf{len}(H)}\text{tri}(H)$, so every point in H_e is less than $r/2$ from an endpoint of e. Thus all of $\widetilde{G}^{\oslash k}$ is covered by placing balls of radius $\frac{\text{tri}(\widetilde{G}^{\oslash k})}{6}$ around each vertex of $\widetilde{G}$. □

Lemma 3.14. *If $s \in B(x,r)$, then one can cover the ball $B(x,r)$ in $\widetilde{G}^{\oslash k}$ with at most $|E(\widetilde{G})||V(\widetilde{G})|$ balls of radius $r/2$.*

Proof. First consider the case in which $r > \frac{\mathsf{len}(\widetilde{G}^{\oslash k})}{6}$. Then for any edge e in $\widetilde{G}^{\oslash k}$, we have $r > \frac{\mathsf{len}(e)}{\mathsf{len}(H)} \cdot \frac{\text{tri}(H)}{3}$. Thus by Observation 3.13, we may cover H_e by $|V(\widetilde{G})|$ balls of radius $r/2$. This gives a covering of all of $\widetilde{G}^{\oslash k}$ by at most $|E(\widetilde{G})||V(\widetilde{G})|$ balls of radius $r/2$.

Otherwise, assume $\frac{\mathsf{len}(\widetilde{G}^{\oslash k})}{6} \geq r$. Since $s \in B(x,r)$, but $2r \leq \frac{\mathsf{len}(\widetilde{G}^{\oslash k})}{3}$, the ball must be completely contained inside $H_{(s,s')}$. By induction, we can find a sufficient cover of this smaller graph. □

Lemma 3.15. *We can cover any ball $B(x,r)$ in $\widetilde{G}^{\oslash k}$ with at most $2|V(\widetilde{G})||E(\widetilde{G})|^2$ balls of radius $r/2$.*

Proof. We prove this lemma using induction. For $\widetilde{G}^{\oslash 0}$, the claim holds trivially. Next, if any H_e contains all of $B(x,r)$, then by induction we are done. Otherwise, for each H_e containing x, $B(x,r)$ contains an endpoint of e. Then by Lemma 3.14, we may cover H_e by at most $|E(\widetilde{G})||V(\widetilde{G})|$ balls of radius $r/2$. For all other edges $e' = (u,v)$, $x \notin H_{e'}$, so we have:

$$V(H_{e'}) \cap B(x,r) \subseteq B(v, \max(0, r - d(x,v))) \cup B(u, \max(0, r - d(x,u))).$$

Thus, using Lemma 3.14 on both of the above balls, we may cover $V(H_{e'}) \cap B(x,r)$ by at most $2|E(\widetilde{G})||V(\widetilde{G})|$ balls of radius $r/2$. Hence, in total, we need at most $2|V(\widetilde{G})||E(\widetilde{G})|^2$ balls of radius $r/2$ to cover all of $B(x,r)$.

□

Proof of Claim 3.12. First note that $|V(\widetilde{G})| = m(D+1) + 2 \lesssim m^2$. By Lemma 3.15, we have

$$\lambda(\widetilde{G}^{\oslash k}) \leq 2|V(\widetilde{G})||E(\widetilde{G})|^2 \leq 2|V(\widetilde{G})|^5 \lesssim m^{10}.$$

Hence $\log \lambda(\widetilde{G}^{\oslash k}) \lesssim \log m$. □

3.3.3 Lower Bound

For any $\pi \in \mathsf{Aut}(G)$, we define a corresponding automorphism $\tilde{\pi}$ of $\tilde{G}$ by $\tilde{\pi}(s) = s$, $\tilde{\pi}(t) = t$, $\tilde{\pi}(s') = s'$, $\tilde{\pi}(t') = t'$, and $\tilde{\pi}(v^{(i)}) = \pi(v)^{(i)}$ for $v \in V, i \in [D+1]$.

Lemma 3.16. *Let G be a vertex transitive graph. Let $f : V(\widetilde{G}) \to \ell_2$ be an injective mapping and define $\bar{f} : V(\widetilde{G}) \to \ell_2$ by*

$$\bar{f}(x) = \frac{1}{\sqrt{|\mathsf{Aut}(G)|}} \Big(f(\widetilde{\pi} x) \Big)_{\pi \in \mathsf{Aut}(G)}.$$

Let β be such that for every $i \in [D+1]$ there exists a vertical edge $(u^{(i)}, v^{(i)})$ with $\|\bar{f}(u^{(i)}) - \bar{f}(v^{(i)})\| \geq \beta$. Then there exists a horizontal edge $(x,y) \in E(\widetilde{G})$ such that

$$\frac{\|\bar{f}(x) - \bar{f}(y)\|^2}{d_{\widetilde{G}}(x,y)^2} \geq \frac{\bar{\|} f(s) - \bar{f}(t)\|^2}{d_{\widetilde{G}}(s,t)^2} + \frac{\beta^2}{36} \tag{3.10}$$

Proof. Let $D = \mathrm{diam}(G)$. We first observe four facts about $\bar{f}$.

(F1) $\|\bar{f}(s) - \bar{f}(t)\| = \|\bar{f}(s) - \bar{f}(t)\|$

(F2) For all $u, v \in V$,

$$\begin{aligned} \|\bar{f}(s) - \bar{f}(v^{(1)})\| &= \|\bar{f}(s) - \bar{f}(u^{(1)})\|, \\ \|\bar{f}(t) - \bar{f}(v^{(D+1)})\| &= \|\bar{f}(t) - \bar{f}(u^{(D+1)})\|. \end{aligned}$$

(F3) For every $u, v \in V$, $i \in [D]$,

$$\|\bar{f}(v^{(i)}) - \bar{f}(v^{(i+1)})\| = \|\bar{f}(u^{(i)}) - \bar{f}(u^{(i+1)})\|.$$

(F4) For every pair of vertices $u, v \in V$ and $i \in [D+1]$,

$$\langle \bar{f}(s) - \bar{f}(t), \bar{f}(u^{(i)}) - \bar{f}(v^{(i)}) \rangle = 0.$$

Let $z = \frac{\bar{f}(s)-\bar{f}(t)}{\|\bar{f}(s)-\bar{f}(t)\|}$. Fix some $r \in V$ and let $\rho_0 = |\langle z, \bar{f}(s) - \bar{f}(r^{(1)})\rangle|$, $\rho_i = |\langle z, \bar{f}(r^{(i)}) - \bar{f}(r^{(i+1)})\rangle|$ for $i = 1, 2, \ldots, D$ and $\rho_{D+1} = |\langle z, \bar{f}(t) - \bar{f}(r^{(D+1)})\rangle|$. Note that, by (F2) and (F3) above, the values $\{\rho_i\}$ do not depend on the representative $r \in V$. In this case, we have

$$\sum_{i=0}^{D+1} \rho_i \geq \|\bar{f}(s) - \bar{f}(t)\| = 9\gamma D, \tag{3.11}$$

where we put $\gamma = \frac{\|\bar{f}(s)-\bar{f}(t)\|}{d_{\widetilde{G}}(s,t)}$. Note that $\gamma > 0$ since f is injective.

Recalling that $d_{\widetilde{G}}(s,t) = 9D$ and $d_{\widetilde{G}}(s, r^{(1)}) = 4D$, observe that if $\rho_0^2 \geq \left(1 + \frac{\beta^2}{36\gamma^2}\right)(4\gamma D)^2$, then

$$\max\left(\frac{\|\bar{f}(s) - \bar{f}(s')\|^2}{d_{\widetilde{G}}(s,s')^2}, \frac{\|\bar{f}(s') - \bar{f}(r^{(1)})\|^2}{d_{\widetilde{G}}(s', r^{(1)})^2}\right) \geq \gamma^2 + \frac{\beta^2}{36},$$

verifying (3.10). The symmetric argument holds for ρ_{D+1}, thus we may assume that

$$\rho_0, \rho_{D+1} \leq 4\gamma D\sqrt{1 + \frac{\beta^2}{36\gamma^2}} \leq 4\gamma D\left(1 + \frac{\beta^2}{72\gamma^2}\right).$$

In this case, by (3.11), there must exist an index $j \in [D]$ such that

$$\rho_j \geq \left(1 - \frac{8\beta^2}{72\gamma^2}\right)\gamma = \left(1 - \frac{\beta^2}{9\gamma^2}\right)\gamma.$$

Now, consider a vertical edge $(u^{(j+1)}, v^{(j+1)})$ with $\|\bar{f}(u^{(j)}) - \bar{f}(v^{(j)})\| \geq \beta$, and $u' = \bar{f}(u^{(j)}) + \rho_j z$. From (F4), we have

$$\begin{aligned}\max(\|\bar{f}(u^{(j)}) - \bar{f}(u^{(j+1)})\|^2, \|\bar{f}(u^{(j)}) - \bar{f}(v^{(j+1)})\|^2) = \\ \|\bar{f}(u^{(j)}) - u'\|^2 + \max(\|u' - \bar{f}(v^{(j+1))}\|^2, \|u' - \bar{f}(u^{(j+1)}\|^2)) \\ \geq \rho_j^2 + \frac{\beta^2}{4} \\ \geq \left(1 - \frac{2\beta^2}{9\gamma^2}\right)\gamma^2 + \frac{\beta^2}{4} \\ \geq \gamma^2 + \frac{\beta^2}{36},\end{aligned}$$

again verifying (3.10) for one of the two edges $(u^{(j)}, v^{(j+1)})$ or $(u^{(j)}, u^{(j+1)})$. □

The following lemma is well-known, and follows from the variational characterization of eigenvalues (see, e.g. [78, Ch. 15]).

Lemma 3.17. *If $G = (V, E)$ is a d-regular graph with second Laplacian eigenvalue $\mu_2(G)$, then for any mapping $f : V \to \ell_2$, we have*

$$\mathbb{E}_{x,y\in V} \|f(x) - f(y)\|^2 \lesssim \frac{d}{\mu_2(G)} \mathbb{E}_{(x,y)\in E} \|f(x) - f(y)\|^2 \tag{3.12}$$

The next lemma shows that when we use an expander graph, we get a significant increase in stretch for edges of $\widetilde{G}$.

Lemma 3.18. *Let $G = (V, E)$ be a d-regular vertex-transitive graph with $m = |V|$ and $\mu_2 = \mu_2(G)$. If $f : V(\widetilde{G}) \to \ell_2$ is any non-contractive mapping, then there exists a horizontal edge $(x, y) \in E(\widetilde{G})$ with*

$$\frac{\|f(x) - f(y)\|^2}{d_{\widetilde{G}}(x, y)^2} \geq \frac{\|f(s) - f(t)\|^2}{d_{\widetilde{G}}(s, t)^2} + \Omega\left(\frac{\mu_2}{d}(\log_d m)^2\right). \tag{3.13}$$

Proof. We need only prove the existence of an $(x, y) \in E(\widetilde{G})$ such that (3.13) is satisfied for $\bar{f}$ (as defined in Lemma 3.16), as this implies it is also satisfied for f (possibly for some other edge (x, y)).

Consider any layer $G^{(i)}$ in $\widetilde{G}$, for $i \in [D+1]$. Applying (3.12) and using the fact that f is non-contracting, we have

$$\begin{aligned}
\mathbb{E}_{(u,v)\in E} \|\bar{f}(u^{(i)}) - \bar{f}(v^{(i)})\|^2 &= \mathbb{E}_{(u,v)\in E} \|f(u^{(i)}) - f(v^{(i)})\|^2 \\
&\gtrsim \frac{\mu_2}{d} \mathbb{E}_{u,v\in V} \|f(u^{(i)}) - f(v^{(i)})\|^2 \\
&\geq \frac{\mu_2}{d} \mathbb{E}_{u,v\in V} d_G(u,v)^2 \\
&\gtrsim \frac{\mu_2}{d} (\log_d m)^2.
\end{aligned}$$

In particular, in every layer $i \in [D+1]$, at least one vertical edge $(u^{(i)}, v^{(i)})$ has $\|\bar{f}(u^{(i)}) - \bar{f}(v^{(i)})\| \gtrsim \sqrt{\frac{\mu_2}{d}} \log_d m$. Therefore the desired result follows from Lemma 3.16. □

We now to come our main theorem.

Theorem 3.19. *If $G = (V, E)$ is a d-regular, m-vertex, vertex-transitive graph with $\mu_2 = \mu_2(G)$, then*

$$c_2(\widetilde{G}^{\oslash k}) \gtrsim \sqrt{\frac{\mu_2 k}{d}} \log_d m.$$

Proof. Let $f : V(\widetilde{G}^{\oslash k}) \to \ell_2$ be any non-contracting embedding. The theorem follows almost immediately by induction: Consider the top level copy of $\widetilde{G}$ in $\widetilde{G}^{\oslash k}$, and call it G_0. Let $(x, y) \in E(G_0)$ be the horizontal edge for which $\|f(x) - f(y)\|$ is longest. Clearly this edge spans a copy of $\widetilde{G}^{\oslash k-1}$, which we call G_1. By induction and an application of Lemma 3.18, there exists a (universal) constant $c > 0$ and an edge $(u, v) \in E(G_1)$ such that

$$\begin{aligned}
\frac{\|f(u) - f(v)\|^2}{d_{\widetilde{G}^{\oslash k}}(u,v)^2} &\geq \frac{c\mu_2(k-1)}{d}(\log_d m)^2 + \frac{\|f(x) - f(y)\|^2}{d_{\widetilde{G}^{\oslash k}}(x,y)^2} \\
&\geq \frac{c\mu_2(k-1)}{d}(\log_d m)^2 + \frac{c\mu_2}{d}(\log_d m)^2 + \frac{\|f(s) - f(t)\|^2}{d_{\widetilde{G}^{\oslash k}}(s,t)},
\end{aligned}$$

completing the proof. □

Corollary 3.20. *If $G = (V, E)$ is an $O(1)$-regular m-vertex, vertex-transitive graph with $\mu_2 = \Omega(1)$, then*

$$c_2(\widetilde{G}^{\oslash k}) \gtrsim \sqrt{k} \log m \approx \sqrt{\log m \log N},$$

where $N = |V(\widetilde{G}^{\oslash k})| = 2^{\Theta(k \log m)}$.

3.3.4 Extension to other ℓ_p Spaces

Our previous lower bound dealt only with ℓ_2. We now prove the following.

Theorem 3.21. *If $G = (V, E)$ is an $O(1)$-regular m-vertex, vertex-transitive graph with $\mu_2 = \Omega(1)$, for any $p > 1$, there exists a constant $C(p)$ such that*

$$c_p(\widetilde{G}^{\oslash k}) \gtrsim C(p) k^{1/q} \log m \approx C(p)(\log m)^{1-1/q}(\log N)^{1/q}$$

were $N = |V(\widetilde{G}^{\oslash k})|$ and $q = \max\{p, 2\}$.

The only changes required are to Lemma 3.17 and Lemma 3.16 (which uses orthogonality). The first can be replaced by Matoušek's [79] Poincaré inequality: If $G = (V, E)$ is an $O(1)$-regular expander graph with $\mu_2 = \Omega(1)$, then for any $p \in [1, \infty)$ and $f : V \to \ell_p$,

$$\mathbb{E}_{x,y \in V} \|f(x) - f(y)\|_p^p \leq O(2p)^p \, \mathbb{E}_{(x,y)\in E} \|f(x) - f(y)\|_p^p.$$

Generalizing Lemma 3.16 is more involved.

Lemma 3.22. *Let G be a vertex transitive graph, and suppose $p > 1$. If $q = \max\{p, 2\}$, then there exists a constant $K(p) > 0$ such that the following holds. Let $f : V(\widetilde{G}) \to \ell_p$ be an injective mapping and define $\bar{f} : V(\widetilde{G}) \to \ell_p$ by*

$$\bar{f}(x) = \frac{1}{|\mathsf{Aut}(G)|^{1/p}} \Big(f(\widetilde{\pi} x) \Big)_{\pi \in \mathsf{Aut}(G)}.$$

Suppose that β is such that for every $i \in [D+1]$, there exists a vertical edge $(u^{(i)}, v^{(i)})$ which satisfies $\|\bar{f}(u^{(i)}) - \bar{f}(v^{(i)})\|_p \geq \beta$. Then there exists a horizontal edge $(x, y) \in E(\widetilde{G})$ such that

$$\frac{\|\bar{f}(x) - \bar{f}(y)\|_p^q}{d_{\widetilde{G}}(x, y)^q} \geq \frac{\|f(s) - f(t)\|_p^q}{d_{\widetilde{G}}(s, t)^q} + K(p)\beta^q. \tag{3.14}$$

Proof. Let $D = \mathrm{diam}(G)$. For simplicity, we assume that D is even in what follows. We first observe three facts about $\bar{f}$.

(F1) $\|\bar{f}(s) - \bar{f}(t)\|_p = \|f(s) - f(t)\|_p$

(F2) For all $u, v \in V$,

$$\begin{aligned} \|\bar{f}(s) - \bar{f}(v^{(1)})\|_p &= \|\bar{f}(s) - \bar{f}(u^{(1)})\|_p, \\ \|\bar{f}(t) - \bar{f}(v^{(D+1)})\|_p &= \|\bar{f}(t) - \bar{f}(u^{(D+1)})\|_p. \end{aligned}$$

(F3) For every $u, v \in V$, $i \in [D]$,

$$\|\bar{f}(v^{(i)}) - \bar{f}(v^{(i+1)})\|_p = \|\bar{f}(u^{(i)}) - \bar{f}(u^{(i+1)})\|_p.$$

Fix some $r \in V$ and let $\rho_0 = \|\bar{f}(s) - \bar{f}(r^{(1)})\|_p$, $\rho_i = \|\bar{f}(r^{(2i-1)}) - \bar{f}(r^{(2i+1)})\|_p$ for $i = 1, \ldots, D/2$, $\rho_{D/2+1} = \|\bar{f}(t) - \bar{f}(r^{(D+1)})\|_p$. Also let $\rho_{i,1} = \|\bar{f}(r^{(2i-1)}) - \bar{f}(r^{(2i)})\|_p$ and $\rho_{i,2} = \|\bar{f}(r^{(2i)}) - \bar{f}(r^{(2i+1)})\|_p$ for $i = 1, \ldots, D/2$.

Note that, by (F2) and (F3) above, the values $\{\rho_i\}$ do not depend on the representative $r \in V$. In this case, we have

$$\sum_{i=0}^{D/2+1} \rho_i \geq \|\bar{f}(s) - \bar{f}(t)\|_p = 9\gamma D, \tag{3.15}$$

where we put $\gamma = \frac{\|f(s)-f(t)\|_p}{d_{\widetilde{G}}(s,t)}$. Note that $\gamma > 0$ since f is injective.

Let $\delta = \delta(p)$ be a constant to be chosen shortly. Recalling that $d_{\widetilde{G}}(s,t) = 9D$ and $d_{\widetilde{G}}(s, r^{(1)}) = 4D$, observe that if $\rho_0^q \geq \left(1 + \delta\frac{\beta^q}{\gamma^q}\right)(4\gamma D)^q$, then

$$\max\left(\frac{\|\bar{f}(s) - \bar{f}(s'))\|_p^q}{d_{\widetilde{G}}(s,s')^q}, \frac{\|\bar{f}(s') - \bar{f}(r^{(1)})\|_p^q}{d_{\widetilde{G}}(s', r^{(1)})^q}\right) \geq \gamma^q + \delta\beta^q,$$

verifying (3.14). The symmetric argument holds for $\rho_{D/2+1}$, thus we may assume that

$$\rho_0, \rho_{D/2+1} \leq 4\gamma D\left(1 + \delta\frac{\beta^q}{\gamma^q}\right)^{1/q} \leq 4\gamma D\left(1 + \delta\frac{\beta^q}{\gamma^q}\right).$$

Similarly, we may assume that $\rho_{i,1}, \rho_{i,2} \leq \gamma\left(1 + \delta\frac{\beta^q}{\gamma^q}\right)^{1/q}$ for every $i \in [D/2]$.

In this case, by (3.15), there must exist an index $j \in \{1, 2, \ldots, D/2\}$ such that

$$\rho_j \geq \left(1 - 8\delta\frac{\beta^q}{\gamma^q}\right) 2\gamma.$$

Now, consider a vertical edge $(u^{(2j)}, v^{(2j)})$ with $\|f(u^{(2j)}) - f(v^{(2j)})\|_p \geq \beta$. Also consider the vertices $v^{(2j-1)}$ and $v^{(2j+1)}$. We now replace the use of orthogonality ((F4) in Lemma 3.16) with the following well-known 4-point inequalities in ℓ_p spaces (see [?, App. A]). If $1 < p \leq 2$, then for every $u, v, w, x \in \ell_p$,

$$\|u - w\|_p^2 + (p-1)\|x - v\|_p^2 \leq \|u - v\|_p^2 + \|v - w\|_p^2 + \|x - w\|_p^2 + \|u - x\|_p^2.$$

On the other hand, if $p \geq 2$, then for every $u, v, w, x \in \ell_p$,

$$\|u-w\|_p^p + \|x-v\|_p^p \leq 2^{p-2}\left(\|u-v\|_p^p + \|v-w\|_p^p + \|x-w\|_p^p + \|u-x\|_p^p\right).$$

We apply one of these two inequalities with $x = f(u^{(2j)}), v = f(v^{(2j)}), u = f(v^{(2j-1)}), w = f(v^{(2j+1)})$. In the case $p \geq 2$, we conclude that

$$\begin{aligned} \|f(u^{(2j)})-f(v^{(2j-1)})\|_p^p+\|f(u^{(2j)})-f(v^{(2j+1)})\|_p^p &\geq 2^{-p+2}\rho_j^p+2^{-q+2}\beta^p-\rho_{j,1}^p-\rho_{j,2}^p \\ &\geq 2\gamma^p + 2^{-p+2}\beta^p - 34\delta p\beta^p. \end{aligned}$$

Thus choosing $\delta = \frac{2^{1-p}}{34p}$ yields the desired result for one of $(u^{(2j)}, v^{(2j-1)})$ or $(u^{(2j)}, v^{(2j+1)})$.

In the case $1 \leq p \leq 2$, we conclude that

$$\|f(u^{(2j)}) - f(v^{(2j-1)})\|_p^2+\|f(u^{(2j)}) - f(v^{(2j+1)})\|_p^2 \geq \rho_j^2 + (p-1)\beta^2 - \rho_{j,1}^2 - \rho_{j,2}^2.$$

A similar choice of δ again yields the desired result. □

3.4 Snowflakes, ℓ_1 and Metrics of Negative Type

3.4.1 ℓ_1 Metrics on Finite Sets and Cut Measures

For a thorough discussion on the connections between ℓ_1 metrics and cut measures; see [34]. Here we recall a couple simple facts. If (X, d) is a finite pseudometric space, and $f : X \to \ell_1$, then there exists a measure μ on 2^X such that for all $x, y \in X$,

$$\|f(x) - f(y)\|_1 = \sum_{S \subseteq X} |\mathbf{1}_S(x) - \mathbf{1}_S(y)|\, \mu(S).$$

We refer to any such measure μ on subsets of X as a *cut measure.*

Given a cut measure μ on X, we use d_μ to denote the pseudometric which assigns, to every $x, y \in X$, the value

$$d_\mu(x,y) = \mu\left(\{S : \mathbf{1}_S(x) \neq \mathbf{1}_S(y)\}\right).$$

Conversely, to every such metric d_μ, one can associate a mapping $f_\mu : X \to \ell_1$ such that for all $x, y \in X$,

$$\|f_\mu(x) - f_\mu(y)\|_1 = d_\mu(x,y).$$

For cut measures μ and η, we say that $\mu \cong \nu$ if for all $S \subset X$, $\mu(S) + \mu(\bar{S}) = \nu(S) + \nu(\bar{S})$.

Coarse differentiation

Let G be an s-t graph, (X,d) a metric space, and consider a mapping $f : V(G) \to X$. Recalling that $\mathcal{P}_{s,t}(G)$ is the set of s-t shortest-paths in G, let μ be a probability measure on $\mathcal{P}_{s,t} = \mathcal{P}_{s,t}(G)$. We say that f is *ε-efficient with respect to* μ if it satisfies

$$\mathbb{E}_{\gamma\sim\mu} \sum_{uv\in\gamma} d(f(u), f(v)) \le (1+\varepsilon)\, d(f(s), f(t)).$$

For a marked s-t graph G, we define its *marked length* by

$$\mathsf{len}_M(G) = \min_{\gamma\in\mathcal{P}_{s,t}} \sum_{uv\in\gamma:(u,v)\in E_M(G)} \mathsf{len}_G(u,v).$$

Theorem 3.23. *Let G be a marked s-t graph. Then for any $D \ge 1$ and $\varepsilon \ge 2D\left(1 - \frac{\mathsf{len}_M(G)}{\mathsf{len}(G)}\right)$, there exists a $k = O(\frac{1}{\varepsilon}\log D)$ such that the following holds. For every metric space (X,d) and mapping $f : V(G^{\oslash k}) \to X$ with distortion D, there exists a copy of G in $G^{\oslash k}$ such that $f|_G$ is ε-efficient with respect to μ.*

Proof. Assume, without loss of generality, that f is 1-Lipschitz. We claim that if f is not ε-efficient with respect to μ on any copy of G in $G^{\oslash k}$, then

$$\mathbb{E}_{\gamma\sim\mu^{\oslash k}} \sum_{uv\in\gamma} d(f(u), f(v)) \ge \left(1+\frac{\varepsilon}{2}\right)^k d\left(f(s(G^{\oslash k})), f(t(G^{\oslash k}))\right), \tag{3.16}$$

where $\mu^{\oslash k}$ is the natural iterated measure on $s(G^{\oslash k})$-$t(G^{\oslash k})$ shortest-paths in $G^{\oslash k}$. We prove this by induction on k, where the case $k = 0$ is trivial.

Write $G^{\oslash k+1} = G \oslash G^{\oslash k}$. We have one copy of $G^{\oslash k}$ in $G^{\oslash k+1}$ for every marked edge $e = (u,v) \in E_M(G)$. Denoting this copy by H_e, by the induction hypothesis, we have

$$\mathbb{E}_{\gamma\sim(\mu^{\oslash k})_e} \sum_{uv\in\gamma} d(f(u), f(v)) \ge \left(1+\frac{\varepsilon}{2}\right)^k d\left(f(s(H_e)), f(t(H_e))\right), \tag{3.17}$$

where we use $(\mu^{\oslash k})_e$ to denote the measure on H_e. Denote now the outer copy of G by G_0, and observe that if it is not mapped ε-efficiently by f, then

$$\begin{aligned}\mathbb{E}_{\gamma\sim\mu} \sum_{uv\in\gamma} d(f(u), f(v)) &\ge (1+\varepsilon)\, d(f(s(G_0)), f(t(G_0))) \\ &= (1+\varepsilon)\, d\left(f(s(G^{\oslash k+1})), f(t(G^{\oslash k+1}))\right),\end{aligned}$$

where the distribution μ here is over $s(G_0)$-$t(G_0)$ paths in $G_0 \cong G$.

Consider now a term of the form $\sum_{uv\in\gamma} d(f(u), f(v))$ on the left-hand side. Since f is 1-Lipschitz, we have

$$\begin{aligned}\sum_{uv\in\gamma:(u,v)\in E_M(G_0)} d(f(u), f(v)) &\geq \sum_{uv\in\gamma} d(f(u), f(v)) - D\left(1 - \frac{\mathsf{len}_M(G_0)}{\mathsf{len}(G_0)}\right) d_G(s(G_0), t(G_0)) \\ &\geq \sum_{uv\in\gamma} d(f(u), f(v)) - \frac{\varepsilon}{2}\, d(f(s(G_0)), f(t(G_0))) \\ &\geq \left(1 + \frac{\varepsilon}{2}\right) d(f(s(G_0)), f(t(G_0))).\end{aligned}$$

But now since each marked edge is replaced by a copy of $G^{\oslash k}$ in G, every term in the sum $\sum_{uv\in\gamma:(u,v)\in E_M} d(f(u), f(v))$ corresponds to the right-hand side of an instance of (3.17), yielding

$$\mathbb{E}_{\gamma\sim\mu^{\oslash k+1}} \sum_{uv\in\gamma} d(f(u), f(v)) \geq \left(1 + \frac{\varepsilon}{2}\right)^{k+1} d\left(f(s(G^{\oslash k+1})), f(t(G^{\oslash k+1}))\right). \tag{3.18}$$

This preceding line completes our proof of (3.16) by induction. Now, combining the triangle inequality and the fact that f is 1-Lipschitz, the left-hand side of (3.16) is at most $\mathsf{len}(G^{\oslash k})$, while the right-hand side is at least

$$\frac{\mathsf{len}(G^{\oslash k})}{D}\left(1 + \frac{\varepsilon}{2}\right)^{k},$$

yielding a contradiction for some $k \asymp \frac{\log D}{\varepsilon}$. □

3.4.2 The multi-scale Construction

We now describe our main construction, and then we prove some of its basic properties including the fact that it admits an ℓ_1 embedding at every scale. This latter fact represents half of the proof of Theorem 3.29, and will be used in Section 3.4.6 to prove that the construction is an $O(1)$-half-snowflake.

3.4.3 Metric Construction

For a parameter $m \in \mathbb{N}$, consider the graph H_m constructed as follows. Let Q_m be the m-dimensional hypercube graph, with the vertex set identified with $\{0,1\}^m$ and write $V(Q_m) =$

$B_m \cup R_m$, where B_m and R_m denote the nodes of even and odd parity, respectively. Then Q_m is bipartite with respect to the partition (B_m, R_m). We define a graph H_m as follows. First, H_m contains $2m$ layers of the form

$$B_m^{(1)} R_m^{(1)} B_m^{(2)} R_m^{(2)} B_m^{(3)} R_m^{(3)} \cdots B_m^{(m)} R_m^{(m)}, \tag{3.19}$$

where $B_m^{(i)}$ and $R_m^{(i)}$ denote disjoint copies of B_m and R_m for $i = 1, 2, \ldots, m$, and hypercube edges are present between every pair of adjacent layers.

We also add to H_m two nodes: s' connected to all the nodes of $B_m^{(1)}$ by edges of length m and t' connected to the nodes of $R_m^{(m)}$ by edges of length m. Finally, we add two distinguished vertices s and t and connect s to s' and t to t' by two edges of length $\frac{1}{m}$. All other edges of H_m are given length one. In this way, H_m becomes an s-t graph. See Figure 3.2.

Furthermore, we consider H_m as a marked graph with $E_M(H_m) = E(H_m) \backslash \{(s, s'), (t, t')\}$. We use $[Q_m]_i$ to denote the i-th copy of Q_m in H_m, i.e. $V(Q_m) = B_m^{(i)} \cup R_m^{(i)}$. For a vertex $x \in V(Q_m)$, we use $[x]_i$ to denote the corresponding vertex in $[Q_m]_i$. Even though H_m is an undirected graph, we will sometimes consider ordered pairs $\vec{e} = (u, v)$.

Let $\tau \in \mathbb{N}$ be a sufficiently large constant (see (3.42) and (3.49) for the required lower bound on τ). We construct $\hat{H}_m$ from H_m by replacing each marked edge $e \in E_M(H_m)$ with a path of length $\tau m \mathsf{len}(e)$, all of whose edges are marked, and have length $\frac{1}{\tau m}$. Our final construction is of the form $\hat{H}_m^{\oslash k}$ for appropriate choices of $m, k \in \mathbb{N}$.

We equip these graphs with the shortest-path metric, which we denote $d_{m,k}$. There is a natural injection $V(H_m^{\oslash k}) \to V(\hat{H}_m^{\oslash k})$, which will allow us to also write $d_{m,k}(x, y)$ for $x, y \in V(H_m^{\oslash k})$.

The main reason behind using $\hat{H}_m$, instead of using H_m is the way that the analysis work in Section 3.4.7. Having unmarked edges $s - s'$ and $t - t'$, implies that for all marked edges $(u, v) \in E_M(\hat{H}_m)$,

$$\max_{a \in \{s,t\}, b \in \{u,v\}} d_{\hat{H}_m}(a, b) \geq \tau \mathsf{len}_{\hat{H}_m}(u, v). \tag{3.20}$$

This inequality provides a separation between different copies of $\hat{H}_m$ in $\hat{H}_m^{\oslash k}$. This separation is one of the ingredients used in the proof of Theorem 3.37 and Lemma 3.41. However, all the proofs in the current section and Section 3.4.8 can be modified to work with $H_m^{\oslash k}$.

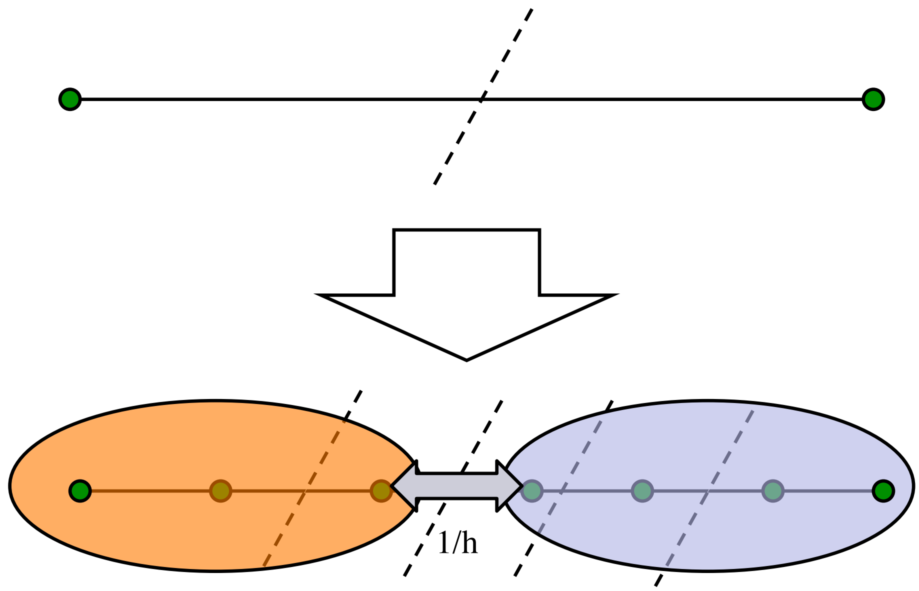


Figure 3.6: Random extension of a cut on an edge.

Extending Cut Measures

Before we describe the embedding, we need to introduce some of the tools that we need. In this section we discuss how one can randomly extend cuts from graph metrics to their subdivisions.

Randomly extending to a subdivision. Let G be a metric graph, and for $h \in \mathbb{N}$, let G_h denote the metric graph where every edge $e \in E(G)$ is replaced by a path of h edges, each of length $\mathsf{len}(e)/h$. Orient each $e \in E(G)$, and denote the new path between the endpoint of $e = (u, v)$ by $\{u = P_e(0), P_e(1), \ldots, P_e(h) = v\}$.

Given any subset $S \subseteq V(G)$, we define a random subset $\mathsf{ext}_h(S) \subseteq V(G_h)$ as follows. Let $\{X_e\}_{e \in E(G)}$ be a family of i.i.d. uniform $[0, 1]$ random variables, and put

$$\mathsf{ext}_h(S) = S \cup \bigcup_{(u,v) \in E(G)} \left\{ P_e(i) : X_e \leq \frac{h-i}{h} \mathbf{1}_S(u) + \frac{i}{h} \mathbf{1}_S(v) \right\}. \tag{3.21}$$

It is easy to check that the distribution of $\mathsf{ext}_h(S)$ does not depend on the orientation chosen for the edges of G. The preceding operation corresponds to taking a cut $S \subseteq V(G)$ in the original graph, and extending it to G_h in the following way: For every edge $(u, v) \in E(G)$

that is cut by S, we cut the new path from u to v in G_h uniformly at random (see Figure 3.6).

Now, given a cut measure μ, we define the extension $\mathcal{E}_h\mu$ to be the cut measure on G_h defined by

$$\mathcal{E}_h\mu(S) = \mu(S \cap V(G)) \cdot \mathbb{P}(S = \mathsf{ext}_h(S \cap V(G))),$$

for every subset $S \subseteq V(G_h)$. By abuse of notation, given a mapping $f : V(G) \to \ell_1$, we will use $\mathcal{E}_h f : V(G_h) \to \ell_1$ to denote the mapping which arises from constructing a cut measure μ_f from f, applying $\mathcal{E}_h$, and then passing back to an ℓ_1-valued mapping. (Recall Section 3.4.1.)

The following observation is immediate from (3.21).

Observation 3.24. *For any graph G, $h \in \mathbb{N}$, and $(x, y) \in E(G)$, if vertices $u, v \in V(G_h)$ are on the path that replaced the edge $(x, y) \in E(G)$, then for every $f : V(G) \to \ell_1$, we have*

$$\|\mathcal{E}_s f(u) - \mathcal{E}_s f(v)\|_1 = \frac{d_{G_h}(u, v)}{d_{G_h}(x, y)} \|f(x) - f(y)\|_1.$$

Lemma 3.25. *For any graph G and $h \in \mathbb{N}$, the following holds. For every $f : V(G) \to \ell_1$, we have*

$$\mathsf{dist}(\mathcal{E}_s f) \leq 5 \cdot \mathsf{dist}(f).$$

Proof. We may assume that $\|f\|_{\mathrm{Lip}} \leq 1$. One easily verifies that in this case, $\|\mathcal{E}_h f\|_{\mathrm{Lip}} \leq 1$ as well. We will identify $V(G)$ with the natural subset of $V(G_h)$. Consider an edge $(x, y) \in E(G)$, and let P be the path of length h between x and y in G_h. For $u, v \in P$, by Observation 3.24 we have

$$\|\mathcal{E}_s f(u) - \mathcal{E}_s f(v)\|_1 = \|f(x) - f(y)\|_1 \frac{d_{G_h}(u, v)}{d_{G_h}(x, y)} \geq \frac{d_{G_h}(u, v)}{\mathsf{dist}(f)}. \tag{3.22}$$

Now, consider the case when $u, v \in V(G_h)$ do not lie on the same G-edge in G_h. Let $u', v' \in V(G)$ be such that $d_{G_h}(u, u') = d_{G_h}(u, V(G))$ and $d_{G_h}(v, v') = d_{G_h}(v, V(G))$. By the triangle inequality,

$$\|\mathcal{E}_s f(u) - \mathcal{E}_s f(v)\|_1 \geq \|\mathcal{E}_s f(u') - \mathcal{E}_s f(v')\|_1 - \|\mathcal{E}_s f(u) - \mathcal{E}_s f(u')\|_1 - \|\mathcal{E}_s f(v) - \mathcal{E}_s f(v')\|_1. \tag{3.23}$$

On the other hand, since $d_{G_h}(u,u') = d_{G_h}(u,V(G))$, for any given cut $S \subseteq V(G)$, we have

$$\begin{aligned}
\|\mathcal{E}_s\mathbf{1}_S(u) - \mathcal{E}_s\mathbf{1}_S(v)\|_1 &= \Pr[\mathcal{E}_s\mathbf{1}_S(v) = \mathcal{E}_s\mathbf{1}_S(u')]\|\mathcal{E}_s\mathbf{1}_S(u) - \mathcal{E}_s\mathbf{1}_S(u')\|_1 \\
&\quad + \Pr[\mathcal{E}_s\mathbf{1}_S(v) \neq \mathcal{E}_s\mathbf{1}_S(u')](1 - \|\mathcal{E}_s\mathbf{1}_S(u) - \mathcal{E}_s\mathbf{1}_S(u')\|_1) \\
&\geq \Pr[\mathcal{E}_s\mathbf{1}_S(v) = \mathcal{E}_s\mathbf{1}_S(u')]\|\mathcal{E}_s\mathbf{1}_S(u) - \mathcal{E}_s\mathbf{1}_S(u')\|_1 \\
&\quad + \Pr[\mathcal{E}_s\mathbf{1}_S(v) \neq \mathcal{E}_s\mathbf{1}_S(u')]\|\mathcal{E}_s\mathbf{1}_S(u) - \mathcal{E}_s\mathbf{1}_S(u')\|_1 \\
&\geq \|\mathcal{E}_s\mathbf{1}_S(u) - \mathcal{E}_s\mathbf{1}_S(u')\|_1,
\end{aligned}$$

and the same holds for v and v'. Therefore,

$$4\,\|\mathcal{E}_s f(u) - \mathcal{E}_s f(v)\|_1 \geq 2\,\|\mathcal{E}_s f(u) - \mathcal{E}_s f(u')\|_1 + 2\,\|\mathcal{E}_s f(v) - \mathcal{E}_s f(v')\|_1 \,. \tag{3.24}$$

Averaging (3.23) and (3.24) yields,

$$\begin{aligned}
\frac{5}{2}\,\|\mathcal{E}_s f(u) - \mathcal{E}_s f(v)\|_1 &\geq \frac{1}{2}\left(\|\mathcal{E}_s f(u') - \mathcal{E}_s f(v')\|_1 + \|\mathcal{E}_s f(u) - \mathcal{E}_s f(u')\|_1 + \|\mathcal{E}_s f(v) - \mathcal{E}_s f(v')\|_1\right) \\
&\overset{3.24}{\geq} \frac{1}{2\,\mathsf{dist}(f)}\left(d_G(u,u') + d_G(u',v') + d_G(v',v)\right) \\
&\geq \frac{1}{2\,\mathsf{dist}(f)} d_G(u,v).
\end{aligned}$$

□

Restricting a cut measure. For a cut measure μ on 2^X and $S \subseteq X$, we define the restriction of μ to S as follows. For $A \subseteq S$, we put

$$\mu|_S(A) = \sum_{B \subseteq X \setminus S} \mu(A \cup B). \tag{3.25}$$

Note that, if $f : X \to \ell_1$ is the embedding corresponding to measure μ then $f|_S$ is the ℓ_1 embedding corresponding to $\mu|_S$.

3.4.4 An Embedding for the Base Graph

In this section we present a constant distortion embedding for the base graph $\hat{H}_m$. We will use this embedding again in Section 3.4.6 as the building block for constructing a constant distortion embedding of $(\hat{H}_m^{\oslash k}, \sqrt{d_{m,k}})$ into ℓ_2.

ℓ_1 embedding for the base graph. We now exhibit an ℓ_1 embedding of $\hat{H}_m$. We begin by defining a cut measure μ_m on H_m, which is the sum of the following three measures.

μ_{ver}: These cuts are the only cuts that separate s from t. For any $m \le k \le 3m-1$, we assign weight $\frac{1}{2}$ to the cut

$$S = \{v : v \in V(H_m), d_{H_m}(s,v) \le k + m^{-1}\}.$$

We also put weight $\frac{1}{m}$ on the cuts $\{s\}$ and $V(H_m) \setminus \{t\}$, and weight $\frac{m}{2}$ on the cuts $\{s, s'\}$ and $V(H_m) \setminus \{t, t'\}$.

μ_{hor}: These cuts are the hypercube cuts and they do not separate s from t. For $k \in [m]$ and $b \in \{0,1\}$, we put weight $\frac{1}{4}$ on the cut

$$S = \{s, t, s', t'\} \cup \Big\{[x]_i : x \in \{0,1\}^m, x_k = b, i \in [m]\Big\}\ .$$

μ_{st}: The cut measure puts weight $\frac{m}{4}$ on the single cut $\{s, t, s', t'\}$.

One can easily verify that for every edge $(u,v) \in E(H_m)$, we have

$$\mathsf{len}(u,v) = d_{\mu_{\mathsf{ver}}}(u,v) + d_{\mu_{\mathsf{hor}}}(u,v) + d_{\mathsf{st}}(u,v)\,. \tag{3.26}$$

Let $f_m : V(H_m) \to \ell_1$ be the embedding corresponding to the cut measure $\mu_{\mathsf{ver}} + \mu_{\mathsf{hor}} + \mu_{\mathsf{st}}$. We put $G = (H_m)_h$ for $h = \tau m^2$. The graph $\hat{H}_m$ is isometric to a subset of G. We put $g_m : V(\hat{H}_m) \to \ell_1$ as the restriction of the map $\mathcal{E}_h f_m$ to vertices of $\hat{H}_m$, i.e.

$$g_m = \mathcal{E}_h f_m|_{V(\hat{H}_m)}. \tag{3.27}$$

We now prove some of the properties of the map g_m.

Lemma 3.26. *For every $m \in \mathbb{N}$, $\mathsf{dist}(g_m) \asymp 1$.*

Proof. Lemma 3.25 implies that $\mathsf{dist}(g_m) \lesssim \mathsf{dist}(f_m) = \|f_m\|_{\mathrm{Lip}} \cdot \|f_m^{-1}\|_{\mathrm{Lip}}$. By Equation (3.26), $\|f_m\|_{\mathrm{Lip}} = 1$. Therefore, we only need to show that $\|f_m^{-1}\|_{\mathrm{Lip}} \lesssim 1$. We can lower bound the distance between any vertex $x \in V(H_m)$ and s by $d_{H_m}(x,s)/2$, using only

μ_{ver}. Similarly we can bound the distance from s', t', and t to other vertices. For all other pairs of vertices $[x]_i, [y]_j \in V(H_m) \setminus \{s, s', t, t'\}$ we have,

$$\begin{aligned} d_{H_m}([x]_i, [y]_j) &\leq \max(2|i-j|, d_{Q_m}(x,y)) \\ &\leq 2 \cdot (|i-j| + d_{Q_m}(x,y)) \\ &\leq 2 \cdot \Big(2\, d_{\mu_{\mathsf{ver}}}([x]_i, [y]_j) + 2\, d_{\mu_{\mathsf{hor}}}([x]_i, [y]_j)\Big) \\ &\leq 4 \cdot \|f_m([x]_i) - f_m([y]_j)\|_1, \end{aligned}$$

completing the proof of the lemma. □

The following observation also follows directly from (3.26) and Observation 3.24.

Observation 3.27. *For every $m \in \mathbb{N}$, and for every $(x,y) \in E_M(\hat{H}_m)$, $\|g_m(x) - g_m(y)\|_1 = d_{\hat{H}_m}(x,y)$.*

In Section 3.4.6 we construct a map $f : \hat{H}_m \to \ell_2$ based on g_m and then we modify the map to obtain a 0-efficient embedding from $\hat{H}_m$ to ℓ_2^2. The next lemma helps us control the rate that length of the edges change when we modify the mapping.

Lemma 3.28. *For every $m \in \mathbb{N}$, let ν_m be the corresponding cut measure of the map g_m, then for all edges $(u,v) \in E_M(\hat{H}_m)$, we have*

$$\left| \sum_{S \subseteq V(\hat{H}_m)} (\mathbf{1}_S(u) - \mathbf{1}_S(v))(\mathbf{1}_S(t) - \mathbf{1}_S(s))\, \nu_m(S) \right| = \frac{1}{2} d_{\hat{H}_m}(u,v).$$

Proof. Let $h = \tau m^2$. We have $\nu_m = \mathcal{E}_s \mu_m|_{V(\hat{H}_m)}$. Both $\mathcal{E}_s$ and restriction are linear operations, therefore

$$\nu_m = \mathcal{E}_s \mu_{\mathsf{ver}}|_{V(\hat{H}_m)} + \mathcal{E}_s \mu_{\mathsf{st}}|_{V(\hat{H}_m)} + \mathcal{E}_s \mu_{\mathsf{hor}}|_{V(\hat{H}_m)}.$$

Vertices s and t are on the same side of the cut for all the cuts in the support of μ_{st} and μ_{hor}. Moreover for all sets S in the support of μ_{ver}, we have $s \in S$ and $t \notin S$. Let

$\hat{\mu}_{\mathsf{ver}} = \mathcal{E}_s \mu_{\mathsf{hor}}|_{V(\hat{H}_m)}$, we have

$$\left| \sum_{S \subseteq V(\hat{H}_m)} (\mathbf{1}_S(u) - \mathbf{1}_S(v))(\mathbf{1}_S(t) - \mathbf{1}_S(s))\, \nu_m(S) \right| \quad = \quad \left| \sum_{S \subseteq V(\hat{H}_m)} (\mathbf{1}_S(u) - \mathbf{1}_S(v))\, \hat{\mu}_{\mathsf{ver}}(S) \right|. \tag{3.28}$$

Suppose now that u and v are on the path that replaces edge $(x, y) \in E(H_m)$. Without loss of generality we can assume that

$$d_{H_m}(s, x) \leq d_{(H_m)_h}(s, u) \leq d_{(H_m)_h}(s, v) \leq d_{H_m}(s, y).$$

For all S in the support of μ_{ver}, we have $\mathbf{1}_S(x) - \mathbf{1}_S(y) \geq 0$. This also implies that, $\mathcal{E}_s \mathbf{1}_S(u) - \mathcal{E}_s \mathbf{1}_S(v) \geq 0$. Thus, using (3.28) we can write

$$\begin{aligned} &\left| \sum_{S \subseteq V(\hat{H}_m)} (\mathbf{1}_S(u) - \mathbf{1}_S(v))(\mathbf{1}_S(t) - \mathbf{1}_S(s))\, \nu_m(S) \right| \\ &\qquad = \sum_{S \subseteq V(\hat{H}_m)} |\mathbf{1}_S(u) - \mathbf{1}_S(v)|\, \hat{\mu}_{\mathsf{ver}}(S) \\ &\qquad = \sum_{S \subseteq V(\hat{H}_m)} |\mathbf{1}_S(u) - \mathbf{1}_S(v)|\, \mathcal{E}_s \mu_{\mathsf{ver}}(S) \\ &\qquad = d_{\mathcal{E}_s \mu_{\mathsf{ver}}}(u, v). \end{aligned}$$

By definition of μ_{ver}, for all edges $(x, y) \in E(H_m)$, we have $d_{\mu_{\mathsf{ver}}}(x, y) = \frac{1}{2} d_{H_m}(x, y)$. Now using Observation 3.24, we can conclude that

$$\left| \sum_{S \subseteq V(\hat{H}_m)} (\mathbf{1}_S(u) - \mathbf{1}_S(v))(\mathbf{1}_S(t) - \mathbf{1}_S(s))\, \nu_m(S) \right| = \left(\frac{d_{(H_m)_h}(u, v)}{d_{(H_m)_h}(x, y)} \right) \frac{1}{2} d_{(H_m)_h}(x, y) = \frac{1}{2} d_{\hat{H}_m}(u, v),$$

completing the proof. □

3.4.5 An ℓ_1 Embedding for each Scale

Our main goal in this section is to obtain an ℓ_1 embedding for every scale of $(\hat{H}_m^{\oslash k}, d_{m,k})$.

Theorem 3.29. *There exists a constant $C > 0$, such that for every $m, k \in \mathbb{N}$, and $\Delta > 0$, there exists a 1-Lipschitz mapping $\varphi : V(\hat{H}_m^{\oslash k}) \to \ell_1$, such that for all $x, y \in V(\hat{H}_m^{\oslash k})$ satisfying $d_{m,k}(x, y) \geq \Delta$, we have*

$$\|\varphi(x) - \varphi(y)\|_1 \geq \frac{\Delta}{C}.$$

Proof. We can write $\hat{H}_m^{\oslash k} = \hat{H}_m^{\oslash k-1} \oslash \hat{H}_m$. Note that, all edges in $E_M(\hat{H}_m^{\oslash k-1})$ have the same length; set $\ell = \mathsf{len}(e)$ for some $e \in E_M(\hat{H}_m^{\oslash k-1})$. We first prove the following slightly stronger claim for the case that ℓ is larger than Δ.

Claim 3.30. *If $\ell \geq \Delta$ then there is a map $\psi : V(\hat{H}_m^{\oslash k}) \to \ell_1$, such that such that for all $x, y \in V(\hat{H}_m^{\oslash k})$,*

$$\|\psi(x) - \psi(y)\|_1 \gtrsim \min(d_{m,k}(x, y), \Delta).$$

Proof. We first construct the map ψ, and in then we show that it satisfies the condition of the claim.

Construction. The map ψ will be the ℓ_1 embedding corresponding to the sum of those measures.

Let $h = \tau m^2$, and put $G = (H_m)_h$. The graph $\hat{H}_m$ is isometric to a subset of G. Recall the definition μ_{ver}, μ_{hor} and μ_{st} from Section 3.4.4 and let $\hat{\mu}_{\mathsf{ver}}$, $\hat{\mu}_{\mathsf{hor}}$ and $\hat{\mu}_{\mathsf{st}}$ be the restriction of $\mathcal{E}_h \mu_{\mathsf{ver}}$, $\mathcal{E}_h \mu_{\mathsf{hor}}$ and $\mathcal{E}_h \mu_{\mathsf{st}}$ to $V(\hat{H}_m)$, respectively. The map ψ is the ℓ_1 embedding corresponding to the sum of the following two cum measures.

ν_{ver}: These cuts corresponds to taking a cut on vertices of $\hat{H}_m^{\oslash k-1}$ uniformly at random and then extending it to $V(\hat{H}_m^{\oslash k})$ the following way: For each edge $(u, v) \in E_M(\hat{H}_m^{\oslash k-1})$, if u and v are on the same side of the cut, all the vertices on the copy $\hat{H}_m$ on the edge (u, v) are also mapped to the same side. Otherwise, we pick a random cut from $\hat{\mu}_{\mathsf{ver}}$, and map the vertices of the copy of $\hat{H}_m$ on edge (u, v) to different sides of the cut according to that cut. We now present the formal definition.

Let $\lambda = \sum_{S \subset V(\hat{H}_m)} \hat{\mu}_{\mathsf{ver}}(S)$. We define a measure η on $2^{V(\hat{H}_m)}$ as follows. We put $\eta(V(\hat{H}_m)) = \eta(\emptyset) = 1$ and for all other $S \subseteq 2^{V(\hat{H}_m)} \setminus \{\emptyset, V(\hat{H}_m)\}$,

$$\eta(S) = \frac{\hat{\mu}_{\mathsf{ver}}(S) + \hat{\mu}_{\mathsf{ver}}(\bar{S})}{\lambda}, \tag{3.29}$$

and we put

$$\nu_{\mathsf{ver}}(S) = A \prod_{(u,v) \in E_M(\hat{H}_m^{\oslash k-1})} \eta(S \cap V(H_{(u,v)})), \tag{3.30}$$

where $H_{(u,v)}$ is the copy of $\hat{H}_m$ on the edge (u,v) of $\hat{H}_m^{\oslash k-1}$, and $A = \left(\frac{\lambda \ell}{2\mathsf{len}(\hat{H}_m)} \cdot 2^{-|V(\hat{H}_m^{\oslash k-1})|}\right)$, is the normalization factor.

By choosing these weights we make the distance between endpoints of edges in $\hat{H}_m^{\oslash k-1}$ at least $\frac{\ell}{2}$, i.e., for $(u,v) \in E_M(\hat{H}_m^{\oslash k-1})$, we have $d_{\nu_{\mathsf{ver}}}(u,v) = \lambda \geq \frac{\ell}{2}$. Moreover, for each marked edges (u,v), restricting the measure to the copy of $\hat{H}_m$ on (u,v) gives us the measure $\frac{\ell}{\mathsf{len}(\hat{H}_m)}\hat{\mu}_{\mathsf{ver}}$.

ν_{hor}: These cuts are the product measure of the $\hat{\mu}_{\mathsf{st}} + \hat{\mu}_{\mathsf{hor}}$ for copies of $\hat{H}_m$ on $E_M(\hat{H}_m^{\oslash k-1})$. Let $\lambda = \sum_{T \subset V(\hat{H}_m)} (\hat{\mu}_{\mathsf{hor}} + \hat{\mu}_{\mathsf{st}})(T)$. We define ν_{hor} as follows:

$$\nu_{\mathsf{hor}}(S) = \frac{\lambda \ell}{\mathsf{len}(\hat{H}_m)} \prod_{(u,v) \in E_M(\hat{H}_m^{\oslash k-1})} \frac{(\hat{\mu}_{\mathsf{st}} + \hat{\mu}_{\mathsf{hor}})(S \cap V(H_{(u,v)}))}{\lambda},$$

where $H_{(u,v)}$ is the copy of $\hat{H}_m$ on the edge (u,v) of $\hat{H}_m^{\oslash k-1}$. The measure is scaled by the factor λ for normalization. Note that, for all the cuts $(S, \bar{S})$ in the support of $\hat{\mu}_{\mathsf{st}} + \hat{\mu}_{\mathsf{hor}}$, we have $s, t \in S$. Moreover, it is easy to check that restricting the measure ν_{for} to the copy of $\hat{H}_m$ on (u,v) gives us the measure $\frac{\ell}{\mathsf{len}(\hat{H}_m)}(\hat{\mu}_{\mathsf{hor}} + \hat{\mu}_{\mathsf{st}})$.

Suppose that G' is the copy of $\hat{H}_m$ on some edge e of $\hat{H}_m^{\oslash k-1}$. It follows from the definition of ν_{hor} and ν_{ver}

$$(\nu_{\mathsf{hor}} + \nu_{\mathsf{ver}})|_{V(G')} \cong \frac{\ell}{\mathsf{len}(\hat{H}_m)}(\hat{\mu}_{\mathsf{ver}} + \hat{\mu}_{\mathsf{hor}} + \hat{\mu}_{\mathsf{st}}). \tag{3.31}$$

Analysis. The measure $\hat{\mu}_{\mathsf{ver}} + \hat{\mu}_{\mathsf{hor}} + \hat{\mu}_{\mathsf{st}}$ is the cut measure corresponding to the map g_m (see Section 3.4.4 for definition), and we have $\|g_m\|_{\mathrm{Lip}} \leq 1$. Using (3.31) and definition of ν_{hor} and ν_{ver}, it is easy to check that $\|\psi\|_{\mathrm{Lip}} \leq 1$. We now bound $\|\psi(x) - \psi(y)\|_1$ based on $d_{m,k}(x,y)$. For $x, y \in V(\hat{H}_m^{\oslash k})$ we divide the problem of bounding the distance between x and y into three cases. If vertices x and y are on the same copy of $\hat{H}_m$ on an edge $e \in \hat{H}_m^{k-1}$, then let H be the copy of $\hat{H}_m$ that contains both x and y. By (3.31),

$$\mathsf{dist}(\psi|_H) = \mathsf{dist}(g_m) \overset{3.26}{\asymp} 1. \tag{3.32}$$

hence,

$$\|\psi(x)-\psi(y)\|_1 \gtrsim d_{m,k}(x,y).$$

If x and y are on non-adjacent edges of the outer copy of $\hat{H}_m^{\oslash k-1}$, then it is easy to check that x and y are mapped to different sides in half the cuts of ν_{ver}. Moreover by our construction, the total measure of cuts in ν_{ver} is at least $\frac{\Delta}{2}$, therefore

$$\|\psi(x)-\psi(y)\|_1 \geq \frac{\Delta}{4}.$$

Suppose that x and y are on adjacent edges (u,v) and (v,w) on the outer copy of $\hat{H}_m^{\oslash k-1}$ respectively. It follows directly from the definition of ν_{ver} that,

$$\sum_{\substack{S\subseteq V(\hat{H}_m)\\ \mathbf{1}_S(v)=\mathbf{1}_S(w)}} |\mathbf{1}_S(x)-\mathbf{1}_S(v)|\,\nu_{\mathsf{ver}}(S) = \sum_{\substack{S\subseteq V(\hat{H}_m)\\ \mathbf{1}_S(v)\neq\mathbf{1}_S(w)}} |\mathbf{1}_S(x)-\mathbf{1}_S(v)|\,\nu_{\mathsf{ver}}(S).$$

Therefore,

$$\begin{aligned} d_{\nu_{\mathsf{ver}}}(x,v) &= \sum_{S\subseteq V(\hat{H}_m)} |\mathbf{1}_S(x)-\mathbf{1}_S(v)|\,\nu_{\mathsf{ver}}(S) \\ &= 2\sum_{\substack{S\subseteq V(\hat{H}_m)\\ \mathbf{1}_S(v)=\mathbf{1}_S(w)}} |\mathbf{1}_S(x)-\mathbf{1}_S(v)|\,\nu_{\mathsf{ver}}(S) \end{aligned}$$

Note that, for any cut in the support of ν_{ver}, if v and w are mapped to the same side of the cut then y is also mapped to that side, hence

$$\begin{aligned} d_{\nu_{\mathsf{ver}}}(x,v) &= 2\sum_{\substack{S\subseteq V(\hat{H}_m)\\ \mathbf{1}_S(v)=\mathbf{1}_S(w)}} |\mathbf{1}_S(x)-\mathbf{1}_S(y)|\,\nu_{\mathsf{ver}}(S) \\ &\leq 2\sum_{S\subseteq V(\hat{H}_m)} |\mathbf{1}_S(x)-\mathbf{1}_S(y)|\,\nu_{\mathsf{ver}}(S) \\ &= 2d_{\nu_{\mathsf{ver}}}(x,y) \end{aligned}$$

Now, let H be the copy of $\hat{H}_m$ on the edge (u,v). Since $\nu_{\mathsf{ver}}|_H = \frac{\ell}{\mathsf{len}(\hat{H}_m)}\hat{\mu}_{\mathsf{ver}}$, by definition of $\hat{\mu}_{\mathsf{ver}}$,

$$d_{m,k}(x,v) \overset{3.25}{\lesssim} d_{\nu_{\mathsf{ver}}}(x,v) \leq 2d_{\nu_{\mathsf{ver}}}(x,y).$$

The same argument implies that

$$d_{m,k}(v,y) \lesssim 2d_{\nu_{\mathsf{ver}}}(x,y).$$

Therefore,

$$\begin{aligned} d_{m,k}(x,y) &= d_{m,k}(x,v) + d_{m,k}(v,y) \\ &\lesssim 2d_{\nu_{\mathsf{ver}}}(x,y) + 2d_{\nu_{\mathsf{ver}}}(x,y) \\ &\leq 4\|\psi(x) - \psi(y)\|_1. \end{aligned}$$

□

We will use now this claim to complete the proof this theorem. If k is such that $\ell \geq \Delta$, then the map that is guaranteed by Claim 3.30 satisfies the condition of theorem. Suppose that $\ell < \Delta$, and let $r \in \mathbb{N}$ be the largest integer such that length of the edges in $E_M(\hat{H}_m^{\oslash r})$ is less than Δ, and let ψ be the mapping guaranteed by Claim 3.30 for $\hat{H}_m^{\oslash r}$.

Random extension of ψ. We can write $\hat{H}_m^{\oslash k} = \hat{H}_m^{\oslash r} \oslash \hat{H}_m^{k-r}$. Let H_e be the copy of $\hat{H}_m^{\oslash k-r}$ on edge $e \in E_M(\hat{H}_m^{\oslash r})$ in $\hat{H}_m^{\oslash k}$, and let $\{X_e\}_{e \in E_M(\hat{H}_m^{\oslash r})}$ be a family of i.i.d. uniform $[0,1]$ random variables. For a cut $(S, \bar{S})$ in $\hat{H}_m^{\oslash r}$, we define $\mathsf{ext}_{r,k}(.)$ and $\mathcal{E}_{r,k}$ similar to the definition of ext_h and $\mathcal{E}_s$. Let G be the outer copy of $\hat{H}_m^{\oslash r}$ in $\hat{H}_m^{\oslash k}$; we define

$$\mathsf{ext}_{r,k}(S) = S \cup \bigcup_{(u,v)\in E_M(G)} \left\{ x \in V(H_{(u,v)}) : X_{(u,v)} \leq \frac{d_{m,k}(v,x)}{d_{m,k}(u,v)} \mathbf{1}_S(u) + \frac{d_{m,k}(u,x)}{d_{m,k}(u,v)} \mathbf{1}_S(v) \right\}, \tag{3.33}$$

It is easy to check that the distribution of $\mathsf{ext}_{r,k}(S)$ does not depend on the orientation chosen for the edges of G. For a cut measures ν on $V(\hat{H}_m^{\oslash k})$ we define

$$\mathcal{E}_{r,k}\nu(S) = \nu(S \cap V(G) \cdot \mathbb{P}(S = \mathsf{ext}_{r,k}(S \cap V(G)).$$

The map φ is the ℓ_1 embedding corresponding of the cut measure $\mathcal{E}_{r,k}\varphi_r$.

Analysis. Since ψ is one Lipschitz, its extension φ is also 1-Lipschitz. Also by definition of r if x and y are on the same copy of $\hat{H}_m^{\oslash k-r}$ in $E_M(\hat{H}_m^{\oslash r})$ then $d_{m,k}(x,y) < \Delta$, hence we only need to bound the distances for the case that x and y are on different copies of $\hat{H}_m^{\oslash k-r}$ in $E_M(\hat{H}_m^{\oslash r})$. The rest of the proof closely follows the proof of Lemma 3.25.

Suppose that x and y are on different copies of $\hat{H}_m^{\oslash k-r}$ in $E_M(\hat{H}_m^{\oslash r})$. Let G be the outer copy of $\hat{H}_m^{\oslash r}$ in $\hat{H}_m^{\oslash k}$. Let $x', y' \in V(\hat{H}_m^{\oslash r})$ be such that $d_{m,k}(x, x') = d_{m,k}(u, V(G))$ and $d_{m,k}(y, y') = d_{m,k}(v, V(G))$. By the triangle inequality,

$$\|\varphi(x) - \varphi(y)\|_1 \geq \|\varphi(x') - \varphi(y')\|_1 - \|\varphi(x) - \varphi(x')\|_1 - \|\varphi(y) - \varphi(y')\|_1 . \qquad (3.34)$$

Furthermore, since $d_{m,k}(x, x') = d_{m,k}(x, V(G))$, for any given cut $S \subseteq V(G)$, we have

$$\begin{aligned}
\|\mathcal{E}_{r,k}\mathbf{1}_S(x) - \mathcal{E}_{r,k}\mathbf{1}_S(y)\|_1 &= \Pr[\mathcal{E}_{r,k}\mathbf{1}_S(y) = \mathcal{E}_{r,k}\mathbf{1}_S(x')]\|\mathcal{E}_{r,k}\mathbf{1}_S(x) - \mathcal{E}_{r,k}\mathbf{1}_S(x')\|_1 \\
&\quad + \Pr[\mathcal{E}_{r,k}\mathbf{1}_S(y) \neq \mathcal{E}_{r,k}\mathbf{1}_S(x')](1 - \|\mathcal{E}_{r,k}\mathbf{1}_S(x) - \mathcal{E}_{r,k}\mathbf{1}_S(x')\|_1) \\
&\geq \Pr[\mathcal{E}_{r,k}\mathbf{1}_S(y) = \mathcal{E}_{r,k}\mathbf{1}_S(x')]\|\mathcal{E}_{r,k}\mathbf{1}_S(x) - \mathcal{E}_{r,k}\mathbf{1}_S(x')\|_1 \\
&\quad + \Pr[\mathcal{E}_{r,k}\mathbf{1}_S(y) \neq \mathcal{E}_{r,k}\mathbf{1}_S(x')]\|\mathcal{E}_{r,k}\mathbf{1}_S(x) - \mathcal{E}_{r,k}\mathbf{1}_S(x')\|_1 \\
&\geq \|\mathcal{E}_{r,k}\mathbf{1}_S(x) - \mathcal{E}_{r,k}\mathbf{1}_S(x')\|_1,
\end{aligned}$$

and the same holds for y and y'. Therefore,

$$4\,\|\varphi(x) - \varphi(y)\|_1 \geq 2\,\|\varphi(x) - \varphi(x')\|_1 + 2\,\|\varphi(y) - \varphi(y')\|_1 . \qquad (3.35)$$

Averaging (3.34) and (3.35) yields,

$$\begin{aligned}
\frac{5}{2}\,\|\varphi(x) - \varphi(y)\|_1 &\geq \frac{1}{2}\left(\|\varphi(x') - \varphi(y')\|_1 + \|\varphi(x) - \varphi(x')\|_1 + \|\varphi(y) - \varphi(y')\|_1\right) \\
&= \frac{1}{2}\left(\|\psi(x') - \psi(y')\|_1 + \|\varphi(x) - \varphi(x')\|_1 + \|\varphi(y) - \varphi(y')\|_1\right) \\
&\overset{3.30}{\gtrsim} \frac{1}{2}\left(\min(\Delta, d_{m,k}(x', y')) + \|\varphi(x) - \varphi(x')\|_1 + \|\varphi(y) - \varphi(y')\|_1\right).
\end{aligned}$$

Let x be on the copy of $\hat{H}_m^{\oslash k-r}$ on the edge $(u, v) \in E_M(G)$. Note that $x' \in \{u, v\}$. We have $\Delta > d_{m,k}(u, v)$, hence $d_{m,k}(u, v) = \min(\Delta, d_{m,k}(u, v)) \lesssim \|\varphi(u) - \varphi(v)\|_1$. Plugging this bound in (3.33), implies that $\|\varphi(x) - \varphi(x')\|_1 \gtrsim d_{m,k}(x, x')$. The same inequality also holds for y and y'. Thus,

$$\|\varphi(x) - \varphi(y)\|_1 \gtrsim \left(\min(\Delta, d_{m,k}(x', y')) + \|\varphi(x) - \varphi(x')\|_1 + \|\varphi(y) - \varphi(y')\|_1\right) \gtrsim \min(\Delta, d_{m,k}(x, y)),$$

completing the proof. □

3.4.6 Embedding of $(X, \sqrt{d})$ *into* ℓ_2

Recall the definition of $\hat{H}_m$ from Section 3.4.3. We show that there exists a constant distortion embedding of $(\hat{H}_m^{\oslash k}, \sqrt{d_{m,k}})$ into ℓ_2. We start by presenting some of the tools that we need for constructing and analyzing the embedding in Section 3.4.6. Gives graph G, we also show how one can construct an embedding for $G^{\oslash k}$ from an embedding of G. In Section 3.4.6 we describe the embedding, and finally in Section 3.4.7 we bound the distortion of the embedding.

Notation and definitions

We first present some general tools that we need for our construction, and prove some of their properties.

Projection. For a points $x, y, z \in \ell_2$, we define

$$\pi\left(\overline{y\,z}\,;\,x\right) = y + \frac{\langle x-y\,,\,z-y\rangle}{\|y-z\|_2^2}(z-y).$$

The point $\pi\left(\overline{y\,z}\,;\,x\right)$ is the orthogonal projection of x on the line that passes through y and z.

By definition, for all $x, y, z \in \ell_2$ we have

$$\langle x-y\,,\,x-z\rangle = \langle x-y\,,\,x-\pi\left(\overline{x\,y}\,;\,z\right)\rangle. \tag{3.36}$$

Lemma 3.31. *Let* $X \subset \ell_2$ *be such that* $(X, \|\cdot\|_2^2)$ *is a metric. For* $a, b, c \in X$*, we have*

$$0 \le \langle a-\pi\left(\overline{a\,c}\,;\,b\right),\, a-c\rangle \le \|a-c\|_2^2.$$

Proof. We have

$$\langle a-b\,,\,a-c\rangle = \frac{\|b-a\|_2^2+\|a-c\|_2^2-\|b-c\|_2^2}{2} \ge 0,$$

Similarly we have $\langle b-c\,,\,a-c\rangle \ge 0$, therefore

$$\|a-c\|_2^2 = \langle a-b\,,\,a-c\rangle + \langle b-c\,,\,a-c\rangle \ge \langle a-b\,,\,a-c\rangle.$$

Hence,

$$0 \le \langle a-b\,,\,a-c\rangle \le \|a-c\|_2^2.$$

Using (3.36), we can write $0 \le \langle a-\pi\left(\overline{a\,c}\,;\,b\right),\, a-c\rangle \le \|a-c\|_2^2$, completing the proof. □

Let $X \subset \ell_2$ be such that $(X, \|\cdot\|_2^2)$, Lemma 3.31 implies that for any three points $a, b, c \in X$, $\measuredangle abc \leq \frac{\pi}{2}$. Using this fact a simple calculation yields the following observation.

Observation 3.32. *Let $X \subset \ell_2$ be such that $(X, \|\cdot\|_2^2)$ is a metric. For $a, b, c \in X$, we have*

$$\frac{\sqrt{2}}{2}\min(\|a-b\|_2, \|a-c\|_2) \leq \|\pi\left(\overline{b\,c}\,;\,a\right) - a\|_2 \leq \min(\|a-b\|_2, \|a-c\|_2).$$

Lemma 3.33. *Let $X \subset \ell_2$ be such that $(X, \|\cdot\|_2^2)$ is a metric. For any set of points $x, y, z, w \in X$, we have*

$$\|\pi\left(\overline{x\,y}\,;\,z\right) - \pi\left(\overline{x\,y}\,;\,w\right)\|_2 \leq \frac{\|z-w\|_2^2}{\|x-y\|_2}.$$

Proof. We prove this lemma using Lemma 3.31,

$$\begin{aligned}
\|z-w\|_2^2 &\geq |\langle z-w\,,\, z-\pi\left(\overline{z\,w}\,;\,y\right)\rangle - \langle z-w\,,\, z-\pi\left(\overline{z\,w}\,;\,x\right)\rangle| \\
&\overset{(3.36)}{=} |\langle z-w\,,\, z-y\rangle - \langle z-w\,,\, z-x\rangle| \\
&= |\langle z-w\,,\, x-y\rangle| \\
&\overset{(3.36)}{=} |\langle \pi\left(\overline{x\,y}\,;\,z\right) - \pi\left(\overline{x\,y}\,;\,w\right)\,,\, x-y\rangle| \\
&= \|(\pi\left(\overline{x\,y}\,;\,z\right) - \pi\left(\overline{x\,y}\,;\,w\right)\|_2 \cdot \|x-y\|_2.
\end{aligned}$$

□

Next, we show how one can obtain an embedding for $G^{\oslash k}$ from the embedding of a marked graph G.

Composition of s-t maps. Given two finite marked s-t graphs $G = (V, E)$ and H, and maps $f_G : V(G) \to \ell_2$ and $f_H : V(H) \to \ell_2$. Abusing the notation, we define $(f_G \oslash f_H) : G \oslash H \to \ell_2$ as follows. Let $\{U_e\}_{e \in E_M(G)}$ be a set of linear isometric operators such that, and $a, b \in E_M(G)$ such that $a \neq b$,

$$\mathsf{image}(U_a) \cap \mathsf{image}(U_b) = \{0\} \tag{3.37}$$

and

$$\mathsf{image}(U_a) \cap \mathsf{span}(f_G) = \{0\} \tag{3.38}$$

Since G is a finite, it is easy to check that $\{U_e\}_{e\in E_M(G)}$ exists. Now, for $x \in V(G)$ we put $(f_G \oslash f_H)(x) = f_G(x)$, and for x on the copy of graph H on edge $(u,v) \in E_M(G)$,

$$(f_G \oslash f_H)(x) = f_G(u) + \alpha\left(\|x^* - f_H(s(H))\|_2\left(\frac{f_G(v) - f_G(u)}{\|f_G(v) - f_G(u)\|_2}\right) + U_{(u,v)}\left(f_H(x) - x^*\right)\right), \tag{3.39}$$

where $x^* = \pi\left(\overline{f_H(t(H))\, f_H(s(H))}\,;\, f_H(x)\right)$, and

$$\alpha = \frac{\|(f_G(v) - f_G(u)\|_2}{\|(f_H(t(H)) - f_H(s(H))\|_2},$$

is the scaling factor.

The preceding operation is equivalent to taking a copy of f_H for each edge $(u,v) \in E_M(G)$, and then scaling it by the factor $\frac{\|(f_G(v)-f_G(u)\|_2}{\|(f_H(t(H))-f_H(s(H))\|_2}$, changing its basis, and then translating it to attach $s(H)\ t(H)$ onto u and v.

Congruency of maps. For a metric space (X,d) and maps $f,g : X \to \ell_2$, we say that f and g are congruent if for all $x,y \in X$,

$$\|f(x) - f(y)\|_2 = \|g(x) - g(y)\|_2,$$

and we use the notation $f \cong g$, to denote that f and g are conguent.

The next observation follows immediately from (3.39).

Observation 3.34. *Let G and H be two marked graphs and let $f_G : G \to \ell_2$ and $f_H : H \to \ell_2$ be the maps from G and H to ℓ_2. For $\vec{e} \in \vec{E}_M(G)$, let H_e be the copy of H on edge e in $G \oslash H$. We have*

$$(f_G \oslash f_H)|_{H_e} \cong \left(\frac{\|(f_G(\vec{e})\|_2}{\|(f_H(t(H)) - f_H(s(H))\|_2}\right) f_H.$$

Lemma 3.35. *Let G and H be two marked graphs and let f_G and f_H be the maps from G and H to ℓ_2 respectively, and let $g = f_G \oslash f_H$. For $x,y \in G \oslash H$ that are on two distinct copies of H on edges $(u_x,v_x),(u_y,v_y) \in E_M(G)$, we have*

$$\begin{aligned}\|g(x) - g(y))\|_2^2 = {} & \left\|\pi\left(\overline{g(u_y)\, g(v_y)}\,;\, g(y)\right) - \pi\left(\overline{g(u_x)\, g(v_x)}\,;\, g(x)\right)\right\|_2^2 \\ & + \|g(x) - \pi\left(\overline{g(u_x)\, g(v_x)}\,;\, g(x)\right)\|_2^2 \\ & + \|g(y) - \pi\left(\overline{(g(u_y)\, g(v_y)}\,;\, g(y)\right)\|_2^2.\end{aligned}$$

Proof. Let $x^* = \pi\left(\overline{f_H(t(H))\, f_H(s(H))}\,;\, f_H(x)\right)$, $y^* = \pi\left(\overline{f_H(t(H))\, f_H(s(H))}\,;\, f_H(y)\right)$,

$$\alpha_x = \frac{\|(f_G(v_x) - f_G(u_x)\|_2}{\|(f_H(t(H)) - f_H(s(H))\|_2},$$

and

$$\alpha_y = \frac{\|(f_G(v_y) - f_G(u_y)\|_2}{\|(f_H(t(H)) - f_H(s(H))\|_2}.$$

Moreover, let $U_{(u_x,v_x)}$ and $U_{(u_y,v_y)}$ be the operators as define in composition of s-t maps. Since $g(u_x) - g(v_x)$ is not in the image $U_{(u_x,v_x)}$, it is easy to check that

$$\pi\left(\overline{g(u_x)\, g(v_x)}\,;\, g(x)\right) = f_G(u_x) + \alpha_x\left(\|x^* - f_H(s(H))\|_2\left(\frac{f_G(v_x) - f_G(u_x)}{\|f_G(v_x) - f_G(u_x)\|_2}\right)\right),$$

Similarly,

$$\pi\left(\overline{g(u_y)\, g(v_y)}\,;\, g(y)\right) = f_G(u_y) + \alpha_y\left(\|y^* - f_H(s(H))\|_2\left(\frac{f_G(v_y) - f_G(u_y)}{\|f_G(v_y) - f_G(u_y)\|_2}\right)\right).$$

Using (3.39) we can write

$$g(x) = \pi\left(\overline{g(u_x)\, g(v_x)}\,;\, g(x)\right) + \alpha_x U_{(u_x,v_x)}\left(f_H(x) - x^*\right),$$

and

$$g(y) = \pi\left(\overline{g(u_y)\, g(v_y)}\,;\, g(y)\right) + \alpha_y U_{(u_y,v_y)}\left(f_H(y) - y^*\right).$$

Now, combining (3.37) and (3.38), we can conclude that

$$\begin{aligned}
\|g(y) - g(x)\|_2^2 &= \|\pi\left(\overline{g(u_y)\, g(v_y)}\,;\, g(y)\right) - \pi\left(\overline{g(u_x)\, g(v_x)}\,;\, g(x)\right)\|_2^2 + \\
&\quad + \|\alpha_x U_{(u_x,v_x)}\left(f_H(x) - x^*\right)\|_2^2 + \|\alpha_y U_{(u_y,v_y)}\left(f_H(y) - y^*\right)\|_2^2 \\
&= \|\pi\left(\overline{g(u_y)\, g(v_y)}\,;\, g(y)\right) - \pi\left(\overline{g(u_x)\, g(v_x)}\,;\, g(x)\right)\|_2^2 + \\
&\quad + \alpha_x\|f_H(x) - x^*\|_2^2 + \alpha_y\|f_H(y) - y^*\|_2^2 \\
&\overset{3.34}{=} \|\pi\left(\overline{g(u_y)\, g(v_y)}\,;\, g(y)\right) - \pi\left(\overline{g(u_x)\, g(v_x)}\,;\, g(x)\right)\|_2^2 \\
&\quad + \|g(x) - \pi\left(\overline{g(u_x)\, g(v_x)}\,;\, g(x)\right)\|_2^2 + \|g(y) - \pi\left(\overline{(g(u_y)\, g(v_y)}\,;\, g(y)\right)\|_2^2.
\end{aligned}$$

□

With Lemma 3.35 we can break the distance between two points in the embedding into three parts. Using Observation 3.34 and considering each of the three parts separately gives us the following observation.

Observation 3.36. *Let G and H be two marked graphs and let $f_G, g_G : G \to \ell_2$, and $f_H, g_H : H \to \ell_2$ be such that $f_G \cong g_G$ and $f_H \cong g_H$. We have*

$$f_G \oslash f_H \cong g_G \oslash g_H.$$

By using the $\oslash$-composition we can construct an embedding for $G^{\oslash k}$ based on $f : V(G) \to \ell_2$. In Section 3.4.8 we show that there are no 0-efficient embeddings of $\hat{H}_m$ into NEG, however by violating the triangle inequality condition we can take an embedding of $\hat{H}_m$ and make it 0-efficient. We achieve that by *contracting* a non-efficient embedding of $\hat{H}_m$.

Contracting embeddings of marked graphs. For a given s-t graph G and the map $f : V(G) \to \ell_2$, and $\alpha \in \mathbb{R}^+$ we define the affine operator $\mathcal{C}_a f$ as follows,

$$\mathcal{C}_\alpha f(v) = \pi\left(\overline{f(s)\, f(t)}\,;\, f(v)\right) + \alpha\left(f(v) - \pi\left(\overline{f(s)\, f(t)}\,;\, f(v)\right)\right).$$

It is easy to see that for edge $\vec{e} \in E(G)$, and $\alpha \in \mathbb{R}_+$,

$$\mathcal{C}_\alpha f(\vec{e}) = \pi\left(\overline{f(s)\, f(t)}\,;\, f(\vec{e})\right) + \alpha\left(f(\vec{e}) - \pi\left(\overline{f(s)\, f(t)}\,;\, f(\vec{e})\right)\right), \tag{3.40}$$

thus

$$\|\mathcal{C}_\alpha f(\vec{e})\|_2^2 = \left\|\pi\left(\overline{f(s)\, f(t)}\,;\, f(\vec{e})\right)\right\|_2^2 + \alpha^2\left\|f(\vec{e}) - \pi\left(\overline{f(s)\, f(t)}\,;\, f(\vec{e})\right)\right\|_2^2. \tag{3.41}$$

Embedding

Our construction is based on the embedding of a single copy of $\hat{H}_m$. Let g_m be the map from $\hat{H}_m$ to ℓ_1 as defined in Section 3.4.2, and let $\hat{\mu}_m$ be the corresponding cut measure of g_m. Moreover, since by Lemma 3.26 distortion of g_m, is bounded by a constant independent of τ, we may assume the following bound on the constant τ,

$$\tau \geq 2\mathsf{dist}(g_m)^2. \tag{3.42}$$

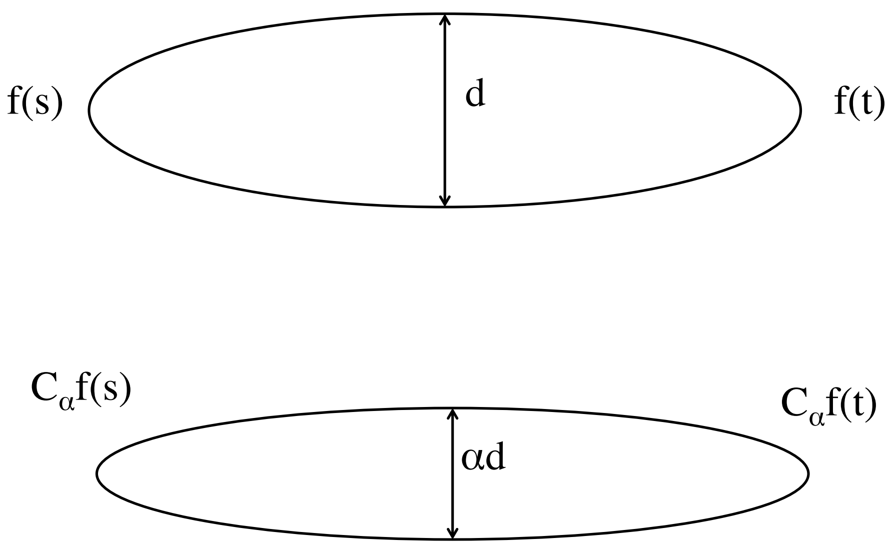


Figure 3.7: Images of f and $\mathcal{C}_\alpha f$.

We define $\tilde{g}_m : \hat{H}_m \to \ell_2$ as follows,

$$\tilde{g}_m(x) = \bigoplus_{S \subseteq V(\hat{H}_m)} \sqrt{\hat{\mu}_m(S)} \cdot \mathbf{1}_S(x). \tag{3.43}$$

It is easy to verify that for any two vertices $x, y \in V(\hat{H}_m)$ we have

$$\|g_m(x) - g_m(y)\|_1 = \|\tilde{g}_m(x) - \tilde{g}_m(y)\|_2^2. \tag{3.44}$$

For all edges $\vec{e} \in \vec{E}(\hat{H}_m)$, by Lemma 3.33

$$\left\|\pi\left(\overline{\tilde{g}_m(s)\,\tilde{g}_m(t)}\,;\,\tilde{g}_m(\vec{e})\right)\right\|_2^2 \leq \left(\frac{\|\tilde{g}_m(\vec{e})\|_2^2}{\|\tilde{g}_m(s) - \tilde{g}_m(t)\|_2}\right)^2 \leq \left(\frac{\mathsf{dist}(g_m)(\tau m)^{-1}}{\sqrt{m}}\right)^2 \overset{(3.42)}{\leq} \frac{1}{4\tau m^3}.$$

By (3.44), and definition of g_m we have $\|\tilde{g}_m(t) - \tilde{g}_m(s)\|_2^2 \geq \frac{1}{2}$, and for all $(u, v) \in E_M(\hat{H}_m)$, we have $\|\tilde{g}_m(u) - \tilde{g}_m(v)\|_2^2 \leq d_{\hat{H}_m}(u, v) = \frac{1}{\tau m}$. Plugging the bounds on $\|\tilde{g}_m(u) - \tilde{g}_m(v)\|_2^2$, $\|\tilde{g}_m(t) - \tilde{g}_m(s)\|_2^2$, and $\left\|\pi\left(\overline{\tilde{g}_m(s)\,\tilde{g}_m(t)}\,;\,\tilde{g}_m(\vec{e})\right)\right\|_2^2$ into (3.41), implies that there exist a constant

$$\frac{1}{2} < \delta \leq 1 \tag{3.45}$$

such that,

$$\mathbb{E}_{(u,v)\in E_M(\hat{H}_m)} \frac{\|\mathcal{C}_\delta \tilde{g}_m(u) - \mathcal{C}_\delta \tilde{g}_m(v)\|_2^2}{d_{\hat{H}_m}(u,v)} = \frac{\|\mathcal{C}_\delta \tilde{g}_m(t) - \mathcal{C}_\delta \tilde{g}_m(s)\|_2^2}{d_{\hat{H}_m}(s,t)}. \tag{3.46}$$

Now we use the $\oslash$-embedding and $\mathcal{C}_\delta \tilde{g}_m$ to define $f_{m,k} : (\hat{H}_m^{\oslash k}, \sqrt{d_{m,k}}) \to \ell_2$ for $m, k \in \mathbb{N}$. We put

$$f_{m,1} = \frac{\mathcal{C}_\delta \tilde{g}_m}{\|\mathcal{C}_\delta \tilde{g}_m\|_{\mathrm{Lip}}},$$

and otherwise

$$f_{m,k} = f_{m,1} \oslash f_{m,k-1}. \tag{3.47}$$

By (3.44) and Lemma 3.26, we have $\mathsf{dist}(\tilde{g}_m) = \sqrt{\mathsf{dist}(g_m)} \lesssim 1$, and by definition $\|f_{m,1}\|_{\mathrm{Lip}} = 1$. Therefore, using (3.45) we can conclude that,

$$\|{f_{m,1}}^{-1}\|_{\mathrm{Lip}} \leq \frac{1}{\delta}\mathsf{dist}(\tilde{g}_m) \lesssim 1. \tag{3.48}$$

In the next section we prove the following theorem by showing that $f_{m,k}$ has constant distortion.

Theorem 3.37. *For $k, m \in \mathbb{N}$, there exists a map $f : \left(\hat{H}_m^{\oslash k}, \sqrt{d_{m,k}}\right) \to \ell_2$ such that $\mathsf{dist}(f) \asymp 1$.*

3.4.7 Analysis of the Embedding

The goal in this section is to prove Theorem 3.37. We start the section by proving some of the properties of maps $\{f_{m,k}\}_{m,k\in\mathbb{N}}$. The first lemma that we prove in this section bounds the distance between endpoints of an edge in the image of $f_{m,1}$.

Lemma 3.38. *For $m \in \mathbb{N}$ and edge $\vec{e} \in \vec{E}_M(\hat{H}_m)$, we have*

$$\frac{\mathsf{len}_{\hat{H}_m}(e)}{\mathsf{len}(\hat{H}_m)} = \frac{\|f_{m,1}(\vec{e})\|_2^2}{\|f_{m,1}(s) - f_{m,1}(t)\|_2^2}$$

.

Proof. We have $f_{m,1} = \frac{\mathcal{C}_\delta \mathcal{C}_\delta \tilde{g}_m}{\|\tilde{g}_m\|_{\mathrm{Lip}}}$ for some $\delta \in \mathbb{R}$. By (3.43) and Observation 3.27 for any edge $(u,v) \in E_M(\hat{H}_m)$, we have

$$\|\tilde{g}_m(u) - \tilde{g}_m(v)\|_2^2 = \|g_m(u) - g_m(v)\|_1 = \mathsf{len}_{\hat{H}_m}(u,v),$$

and by (3.43) and Lemma 3.28

$$|\langle \tilde{g}_m(\vec{e}) \,,\, \tilde{g}_m(t) - \tilde{g}_m(s)\rangle| = \frac{1}{2}\mathsf{len}_{\hat{H}_m}(u,v).$$

The value of $\|\mathcal{C}_\delta f(\vec{e})\|_2$ only depends on $\|f(\vec{e})\|_2$ and $|\langle f(\vec{e}) \,,\, f(t) - f(s)\rangle|$, therefore for all edges $a, b \in E_M(\hat{H}_m)$, and $\delta \in \mathbb{R}_+$ we have $\|\mathcal{C}_\delta \tilde{g}_m(\vec{a})\|_2 = \|\mathcal{C}_\delta \tilde{g}_m(\vec{b})\|_2$. Hence, using (3.46) we can conclude that for all $\vec{e} \in E_M(\hat{H}_m)$,

$$\frac{\|\mathcal{C}_\delta \tilde{g}_m(u) - \mathcal{C}_\delta \tilde{g}_m(v)\|_2^2}{d_{\hat{H}_m}(u,v)} = \frac{\|\mathcal{C}_\delta \tilde{g}_m(t) - \mathcal{C}_\delta \tilde{g}_m(s)\|_2^2}{d_{\hat{H}_m}(s,t)},$$

completing the proof. □

The following corollary, follows directly from the above lemma and Observation 3.34.

Corollary 3.39. *For $m \in \mathbb{N}$ and $k > 1$, the following holds. Let G_0 be the outer copy of $\hat{H}_m$ in $\hat{H}_m^{\oslash k}$, and let H be the copy of $\hat{H}_m^{\oslash k-1}$ on some edge $e \in E_M(G_0)$. We have*

$$\frac{f_{m,k}|_H}{\mathsf{len}(e)} \cong \frac{f_{m,k-1}}{\mathsf{len}(\hat{H}_m^{\oslash k-1})}.$$

Recall the definition of $\hat{H}_m$ from Section 3.4.3. Since $\mathsf{dist}(f_{m,1}) \lesssim 1$ is bounded by a constant independent of τ; throughout this section we assume the following bound on τ,

$$\tau \geq (20\,\|{f_{m,1}}^{-1}\|_{\mathrm{Lip}})^2. \tag{3.49}$$

This bound on τ is used to prove Lemma 3.41. In the next lemma we bound the diameter of the image of $f_{m,k}$. We will use the above bound on τ in the proof to guarantee that diameter of "scales" form a exponential series. Then we use that to bound the diameter of the whole map. We would like to point out that a weaker bound $\tau \geq \frac{400}{m}$ is sufficient to prove Lemma 3.40.

Lemma 3.40. *For $m, k \in \mathbb{N}$, the following inequality holds*

$$\max\left\{\|f_{m,k}(u) - f_{m,k}(v)\|_2^2 : u, v \in V(\hat{H}_m^{\oslash k})\right\} \leq 5\mathsf{len}(\hat{H}_m^{\oslash k}). \tag{3.50}$$

Proof. To prove this lemma it is sufficient to show that for all $v \in V(\hat{H}_m^{\oslash k})$,

$$\|f_{m,k}(s) - f_{m,k}(v)\|_2^2 \leq \frac{5}{4} \cdot \mathsf{len}(\hat{H}_m^{\oslash k}). \tag{3.51}$$

We prove this lemma by induction. First note that for $k = 1$, we have $\|f_{m,1}\|_{\mathrm{Lip}} = 1$, which yields (3.51). Now suppose that $k > 1$, and let G_0 be the outer copy of $\hat{H}_m$ in $\hat{H}_m^{\oslash k}$, and let $x \in V(G_0)$ be the closest vertex to v. Since $f_{m,k}|_{G_0} = f_{m,1}$, if $x = v$ then the base case implies (3.51). Suppose now that v is on the copy of $\hat{H}_m^{\oslash k-1}$ on the edge $(x, y) \in E_M(G_0)$, for some $y \in V(G_0)$. By Corollary 3.39 and the induction hypothesis for the copy of $\hat{H}_m^{\oslash k-1}$ on the edge (x, y),

$$\begin{aligned}\|f_{m,k}(s) - f_{m,k}(v)\|_2^2 &\leq (\|f_{m,k}(s) - f_{m,k}(x)\|_2 + \|f_{m,k}(x) - f_{m,k}(v)\|_2)^2 \\ &\leq \left(\|f_{m,k}(s) - f_{m,k}(x)\|_2 + \sqrt{5 d_{m,k}(x, y)}\right)^2,\end{aligned}$$

Since $\|f_{m,1}\|_{\mathrm{Lip}} \leq 1$,

$$\begin{aligned}\|f_{m,k}(s) - f_{m,k}(v)\|_2^2 &\leq \left(\sqrt{\mathsf{len}(G_0)} + \sqrt{5 d_{m,k}(x, y)}\right)^2 \\ &\overset{(3.20)}{\leq} \left(\left(1 + \sqrt{\frac{5}{\tau}}\right)\sqrt{\mathsf{len}(G_0)}\right)^2 \\ &\overset{(3.49)}{\leq} \frac{5}{4}\mathsf{len}(G_0) = \frac{5}{4}\mathsf{len}(\hat{H}_m^{\oslash k}).\end{aligned}$$

□

Lemma 3.35 allows us to break the analysis of distance between points into three separate parts. We bound two of those parts using the next lemma. For the other part, we use the bound obtained by Lemma 3.41 and (3.48).

In the next lemma we show that if a vertex $v \in \hat{H}_m^{\oslash k}$ is far from s and t, then $f_{m,k}(v)$ is also far from the line that connects $f_{m,k}(s)$ and $f_{m,k}(t)$. See Figure 3.8 for relative position of $f_{m,k}(v)$, $f_{m,k}(s)$, $f_{m,k}(t)$.

Lemma 3.41. *For $k, m \in \mathbb{N}$, and for any vertex $v \in V(\hat{H}_m^{\oslash k})$, we have*

$$\min(d_{m,k}(s, v), d_{m,k}(v, t)) \asymp \left\|\left(f_{m,k}(v) - \pi\left(\overline{f_{m,k}(s)\, f_{m,k}(t)}\,;\, f_{m,k}(v)\right)\right\|_2^2. \tag{3.52}$$

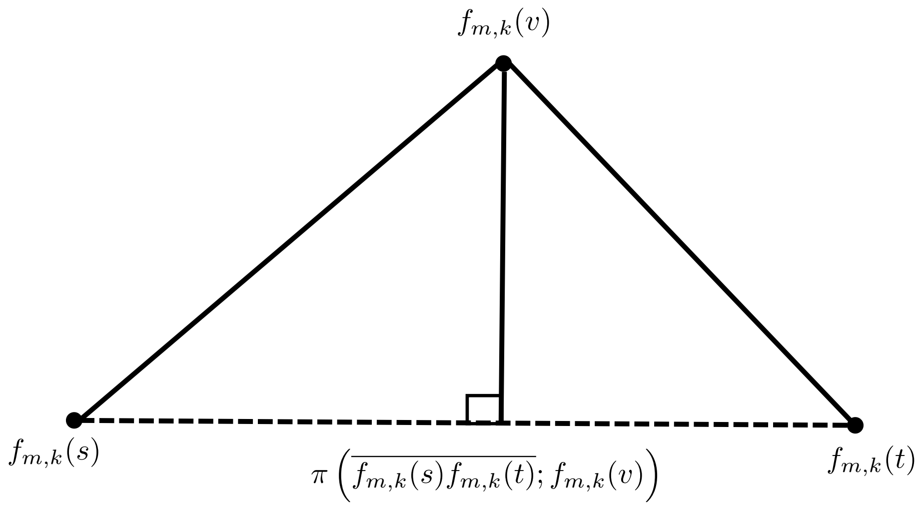


Figure 3.8: Positions of $f_{m,k}(v)$, $f_{m,k}(s)$, $f_{m,k}(t)$, and $\pi\left(\overline{f_{m,k}(s)\, f_{m,k}(t)}\,;\, f_{m,k}(v)\right)$ with respect to each other.

Proof. Let G_0 is the outer copy of $\hat{H}_m$ in $\hat{H}_m^{\oslash k}$, we first prove the lemma for vertices $v \in V(G_0)$.

Case I: $v \in V(G_0)$**:** First, we show that for some contant K.

$$\min(d_{m,k}(s,v), d_{m,k}(v,t)) \leq K \left\| (f_{m,k}(v) - \pi\left(\overline{f_{m,k}(s)\, f_{m,k}(t)}\,;\, f_{m,k}(v)\right) \right\|_2^2 .$$

Recall that, $f_{m,k}|_{G_0} = f_{m,1} = \frac{\mathcal{C}_\delta \tilde{g}_m}{\|\mathcal{C}_\delta \tilde{g}_m\|_{\mathrm{Lip}}}$. By (3.44), image of $\tilde{g}_m$, equipped with $\|\cdot\|_2^2$ distance function is a metric. Therefore by Observation 3.32 for $\tilde{g}_m(s)$, $\tilde{g}_m(t)$, and $\tilde{g}_m(v)$ we have

$$\min(\|\tilde{g}_m(s) - \tilde{g}_m(v)\|_2, \|\tilde{g}_m(t) - \tilde{g}_m(v)\|_2) \leq \sqrt{2} \left\| \tilde{g}_m(v) - \pi\left(\overline{\tilde{g}_m(s)\, \tilde{g}_m(t)}\,;\, \tilde{g}_m(v)\right) \right\|_2 .$$

By (3.45) and definition of $\mathcal{C}_\delta$, we have

$$\min(\|f_{m,k}(s) - f_{m,k}(v)\|_2, \|f_{m,k}(t) - f_{m,k}(v)\|_2) \leq 2\sqrt{2} \left\| f_{m,k}(v) - \pi\left(\overline{f_{m,k}(s)\, f_{m,k}(t)}\,;\, f_{m,k}(v)\right) \right\|_2 ,$$

hence,

$$\min(d_{m,k}(s,v), d_{m,k}(v,t)) \leq 8\mathsf{dist}(f_{m,1})^2 \left\| (f_{m,k}(v) - \pi\left(\overline{f_{m,k}(s)\, f_{m,k}(t)}\,;\, f_{m,k}(v)\right) \right\|_2^2 . \tag{3.53}$$

In order to show the other side of the (3.52), first note that the point $\pi\left(\overline{f_{m,k}(s)\, f_{m,k}(t)}\,;\, f_{m,k}(v)\right)$ is the closest point to $f_{m,k}(v)$ on the line $\overline{f_{m,k}(s)\, f_{m,k}(t)}$, and

$$\left\| f_{m,k}(v) - \pi\left(\overline{f_{m,k}(s)\, f_{m,k}(t)}\,;\, f_{m,k}(v)\right)\right\|_2 \le \min(\|f_{m,k}(s) - f_{m,k}(v)\|_2, \|f_{m,k}(t) - f_{m,k}(v)\|_2)$$

$$\left\|(f_{m,k}(v) - \pi\left(\overline{f_{m,k}(s)\, f_{m,k}(t)}\,;\, f_{m,k}(v)\right)\right\|_2^2 \le \min(d_{m,k}(s,v), d_{m,k}(v,t)),$$

completing the proof for this case.

Case II: $v \in V(\hat{H}_m^{\oslash k}) \setminus V(G_0)$**:** Suppose that v is on the copy of $\hat{H}_m^{\oslash k-1}$ on the edge $(u,w) \in E_M(G_0)$ for some $w \in V(G_0)$. We will use Case I and the upper bound on $\|f_{m,k}(u) - f_{m,k}(v)\|_2$ from Lemma 3.40 to prove this case. Since π is orthogonal projection on a line,

$$\begin{aligned}\|f_{m,k}(u) - f_{m,k}(v)\|_2 \ge \Bigg| &\left\|(f_{m,k}(u) - \pi\left(\overline{f_{m,k}(s)\, f_{m,k}(t)}\,;\, f_{m,k}(u)\right)\right\|_2 \\ &- \left\|(f_{m,k}(v) - \pi\left(\overline{f_{m,k}(s)\, f_{m,k}(t)}\,;\, f_{m,k}(v)\right)\right\|_2 \Bigg|\end{aligned}$$

Using Corollary 3.39 for the copy of $\hat{H}_m^{\oslash k-1}$ on edge (u,w), we can write

$$\begin{aligned}&\left| \left\|(f_{m,k}(u) - \pi\left(\overline{f_{m,k}(s)\, f_{m,k}(t)}\,;\, f_{m,k}(u)\right)\right\|_2 - \left\|(f_{m,k}(v) - \pi\left(\overline{f_{m,k}(s)\, f_{m,k}(t)}\,;\, f_{m,k}(v)\right)\right\|_2 \right| \\ &\qquad \overset{3.40}{\le} \sqrt{5\mathsf{len}_{G_0}(u,w)} \\ &\qquad \overset{(3.20)}{\le} \sqrt{\frac{5}{\tau}\min(d_{m,k}(u,s), d_{m,k}(v,s))} \\ &\qquad \overset{(3.49)}{\le} \sqrt{\frac{\min(d_{m,k}(u,s), d_{m,k}(v,s))}{80\mathsf{dist}(f_{m,1})^2}}.\end{aligned}$$

Combining this and (3.53) for u implies that

$$\left\|(f_{m,k}(u) - \pi\left(\overline{f_{m,k}(s)\, f_{m,k}(t)}\,;\, f_{m,k}(u)\right)\right\|_2 \asymp \left\|(f_{m,k}(v) - \pi\left(\overline{f_{m,k}(s)\, f_{m,k}(t)}\,;\, f_{m,k}(v)\right)\right\|_2. \tag{3.54}$$

Triangle inequality also implies,

$$\begin{aligned}&|\min(d_{m,k}(s,u), d_{m,k}(u,t)) - \min(d_{m,k}(s,v), d_{m,k}(v,t))| \\ &\qquad \le \mathsf{len}_{G_0}(u,w) \overset{(3.20)}{\le} \frac{1}{\tau}\min(d_{m,k}(u,s), d_{m,k}(u,t)).\end{aligned}$$

therefore

$$\min(d_{m,k}(s,u), d_{m,k}(u,t)) \asymp \min(d_{m,k}(s,v), d_{m,k}(v,t)).$$

This inequality in conjunction with Case I and (3.54) completes the proof. □

We end this section by presenting the proof of Theorem 3.37.

Proof of Theorem 3.37. In order to prove this theorem, we need to show there exist universal constants $C_1, C_2 \in \mathbb{R}^+$, such that for all $k \in \mathbb{N}$, and $x, y \in V(\hat{H}_m^{\oslash k})$,

$$C_1 d_{m,k}(x,y) \le \|f_{m,k}(x) - f_{m,k}(y)\|_2^2 \le C_2 d_{m,k}(x,y). \tag{3.55}$$

We first prove this bound for the case that one of x and y is s. We do this by showing that there exists another constant $C_3 > 0$ such that,

$$C_3 d_{m,k}(x,s) \le \|f_{m,k}(x) - f_{m,k}(s)\|_2^2 \le 2 d_{m,k}(x,s), \tag{3.56}$$

and then we use this case to prove the theorem.

Let G_0 be the outer copy of $\hat{H}_m$. Since $f_{m,k}|_{G_0} = f_{m,1}$, if $x \in V(G_0)$ then the fact that both $f_{m,1}$ and ${f_{m,1}}^{-1}$ are $O(1)$-co-Lipschitz implies (3.56). Suppose now that x is on the copy of $\hat{H}_m^{\oslash k-1}$ on the edge $(u_x, v_x) \in E_M(G_0)$. By Corollary 3.39 and Lemma 3.40,

$$\|f_{m,k}(u_x) - f_{m,k}(x)\|_2^2 \le 5 d_{m,k}(u_x, v_x). \tag{3.57}$$

Using triangle inequity we can write

$$\begin{aligned}
\|f_{m,k}(x) - f_{m,k}(s)\|_2^2 &\ge (\|f_{m,k}(u_x) - f_{m,k}(s)\|_2 - \|f_{m,k}(x) - f_{m,k}(u_x)\|_2)^2 \\
&\overset{(3.57)}{\ge} \left(\frac{\sqrt{d_{m,k}(u_x, s)}}{\|{f_{m,1}}^{-1}\|_{\text{Lip}}} - \sqrt{5 d_{m,k}(u_x, v_x)} \right)^2 \\
&\overset{(3.20)}{\ge} \left(\frac{\sqrt{(1 - \frac{1}{\tau}) d_{m,k}(x,s)}}{\|{f_{m,1}}^{-1}\|_{\text{Lip}}} - \sqrt{\frac{5}{\tau} d_{m,k}(s,x)} \right)^2 \\
&\ge \left(1 - \frac{1}{\tau} - 2\sqrt{\frac{5\|{f_{m,1}}^{-1}\|_{\text{Lip}}^2}{\tau}} \right) \frac{d_{m,k}(x,s)}{\|{f_{m,1}}^{-1}\|_{\text{Lip}}^2} \\
&\overset{(3.49)}{\ge} \frac{d_{m,k}(x,s)}{2\|{f_{m,1}}^{-1}\|_{\text{Lip}}^2} \qquad (3.58) \\
&\overset{(3.48)}{\gtrsim} d_{m,k}(x,s). \qquad (3.59)
\end{aligned}$$

On the other hand since $f_{m,1}$ is 1-Lipschitz we have

$$\begin{aligned}
\|f_{m,k}(x) - f_{m,k}(s)\|_2^2 &\leq (\|f_{m,k}(u_x) - f_{m,k}(s)\|_2 + \|f_{m,k}(x) - f_{m,k}(u_x)\|_2)^2 \\
&\overset{(3.57)}{\leq} \left(\sqrt{d_{m,k}(u_x, s)} + \sqrt{5d_{m,k}(u_x, v_x)}\right)^2 \\
&\leq \left(\sqrt{(1 + \frac{1}{\tau})d_{m,k}(x, s)} + \sqrt{\frac{5}{\tau}d_{m,k}(s, x)}\right)^2 \\
&\leq \left(1 + \frac{1}{\tau} + 2\sqrt{\frac{5(\tau + 1)}{\tau^2}} + \frac{5}{\tau}\right) d_{m,k}(x, s) \\
&\overset{(3.49)}{\leq} 2d_{m,k}(x, s).
\end{aligned}$$

This inequality with (3.58) completes the proof of (3.56).

Proof of (3.55)**.** We prove this bound by induction. Similar to the previous case, for $k = 1$, the fact that $f_{m,1}$ is 1-Lipschitz, and its inverse is also $O(1)$-Lipschitz implies (3.55).

Suppose that (3.55) holds for $x, y \in V(\hat{H}_m^{\oslash k-1})$, we show that (3.55) also holds for $x, y \in V(\hat{H}_m^{\oslash k})$.

Let G_0 be the outer copy of $\hat{H}_m$ in $\hat{H}_m^{\oslash k}$. For vertices $x, y \in V(\hat{H}_m^{\oslash k})$, if x and y are both on the copy of $\hat{H}_m^{\oslash k-1}$ on some edge $(u, v) \in E_M(G_0)$, then by Corollary 3.39 and the induction hypothesis for the copy of $\hat{H}_m^{\oslash k-1}$ on (u, v), (3.55) holds.

For the rest of this section we assume that x is on the copy of $\hat{H}_m^{\oslash k-1}$ on edge $(u_x, v_x) \in E(G_0)$, and y is on the copy of $\hat{H}_m^{\oslash k-1}$ on the edge $(u_y, v_y) \in E(G_0)$, and $(u_x, v_x) \neq (u_y, v_y)$. Furthermore, we use $p_x(.)$ and $p_y(.)$ to denote $\pi\left(\overline{f_{m,k}(u_x)\, f_{m,k}(v_x)}\,;\, .\right)$ and $\pi\left(\overline{f_{m,k}(u_y)\, f_{m,k}(v_y)}\,;\, .\right)$, respectively.

Proof of the existence of C_1**.** We bound the distance between $f_{m,k}(x)$ and $f_{m,k}(y)$ by dividing it to three parts, and bounding each part separately (see Figure 3.9). By Lemma 3.35,

$$\begin{aligned}
\|f_{m,k}(x) - f_{m,k}(y)\|_2^2 \;&=\; \|f_{m,k}(x) - p_x(f_{m,k}(x))\|_2^2 + \|f_{m,k}(y) - p_y(f_{m,k}(y))\|_2^2 \\
&\quad + \|p_y(f_{m,k}(y)) - p_x(f_{m,k}(x))\|_2^2 \\
&\geq\; \max\Big(\|f_{m,k}(x) - p_x(f_{m,k}(x))\|_2^2 + \|f_{m,k}(y) - p_y(f_{m,k}(y))\|_2^2 \\
&\qquad , \|p_y(f_{m,k}(y)) - p_x(f_{m,k}(x))\|_2^2\Big). \qquad (3.60)
\end{aligned}$$

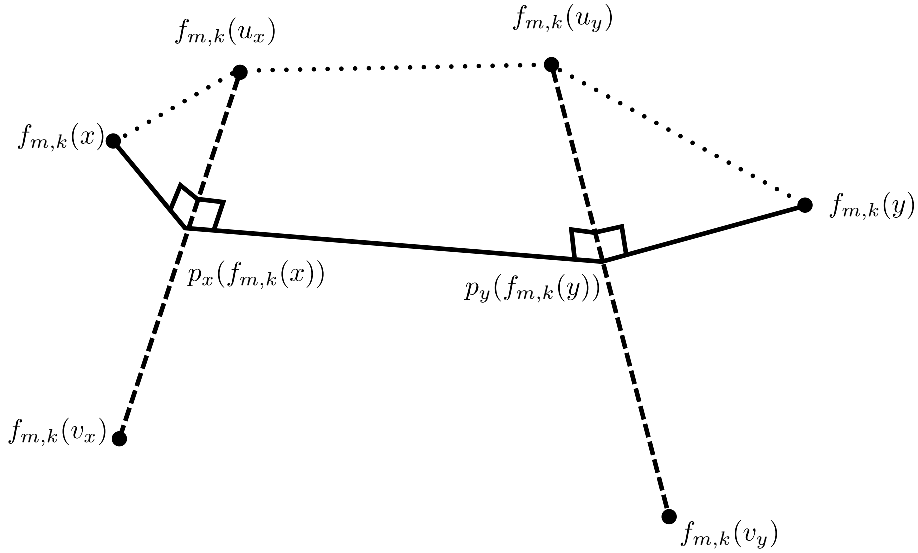


Figure 3.9: We bound $\|f_{m,k}(x)-p_x(f_{m,k}(x))\|_2$, $\|p_y(f_{m,k}(y))-f_{m,k}(y)\|_2$, and $\|p_x(f_{m,k}(x))-p_y(f_{m,k}(y))\|_2$ separately.

First, we bound $\|p_y(f_{m,k}(y)) - p_x(f_{m,k}(x))\|_2^2$ part, using triangle inequality.

$$\begin{aligned}\|p_y(f_{m,k}(y)) - p_x(f_{m,k}(x))\|_2^2 &\geq \Big(\|f_{m,k}(u_y) + f_{m,k}(u_x)\|_2 - \|p_y(f_{m,k}(y)) - f_{m,k}(u_y)\|_2 \\ &\quad -\|p_x(f_{m,k}(x)) - f_{m,k}(u_x)\|_2\Big)^2 \\ &\geq \frac{1}{2}\|f_{m,k}(u_y) - f_{m,k}(u_x)\|_2^2 - \Big(\|p_y(f_{m,k}(y)) - f_{m,k}(u_y)\|_2 \\ &\quad +\|p_x(f_{m,k}(x)) - f_{m,k}(u_x)\|_2\Big)^2.\end{aligned}$$

Since p_x and p_y are orthogonal projections,

$$\begin{aligned}\|p_y(f_{m,k}(y)) - p_x(f_{m,k}(x))\|_2^2 &\geq \frac{1}{2}\|f_{m,k}(u_y) - f_{m,k}(u_x)\|_2^2 - \Big(\|f_{m,k}(y) - f_{m,k}(u_y)\|_2 \\ &\quad +\|f_{m,k}(x) + f_{m,k}(u_x)\|_2\Big)^2 \\ &\geq \frac{1}{2}\|f_{m,k}(u_y) - f_{m,k}(u_x)\|_2^2 - 2\Big(\|f_{m,k}(y) - f_{m,k}(u_y)\|_2^2 \\ &\quad +\|f_{m,k}(x) - f_{m,k}(u_x)\|_2^2\Big).\end{aligned}$$

Using (3.56) for copies of the graph $\hat{H}_m^{\oslash k-1}$ on edges (u_x, v_x) and (u_y, v_y), we can write

$$\|p_y(f_{m,k}(y)) - p_x(f_{m,k}(x))\|_2^2 \geq \frac{1}{2}\|f_{m,k}(u_y) - f_{m,k}(u_x)\|_2^2 - C_3\Big(d_{m,k}(x, u_x) + d_{m,k}(y, u_y)\Big).$$

And finally since $f_{m,k}|_{G_0} = f_{m,1}$, we can use (3.48) to write,

$$\|p_y(f_{m,k}(u_y)) - p_x(f_{m,k}(u_x))\|_2^2 \geq \Theta(d_{m,k}(u_x, u_y)) - C_3\Big(d_{m,k}(x, u_x) + d_{m,k}(y, u_y)\Big).$$

Without loss of generality suppose that $d_{m,k}(u_y, y) \leq d_{m,k}(v_y, y)$ and $d_{m,k}(u_x, x) \leq d_{m,k}(v_x, x)$. Plugging in this bound and the bound from Lemma 3.41 into (3.60), yields the following inequality

$$\begin{aligned}
\|f_{m,k}(x) - f_{m,k}(y)\|_2^2 &\geq \max\Big(\Theta(d_{m,k}(u_x, x) + d_{m,k}(u_y, y)), \\
&\qquad \Big(\Theta\big(d_{m,k}(u_y, u_x)\big) - C_3(d_{m,k}(u_x, x) + d_{m,k}(u_y, y))\Big)\Big) \\
&\gtrsim d_{m,k}(u_x, x) + d_{m,k}(u_y, y) + d_{m,k}(u_y, u_x) \\
&\geq d_{m,k}(x, y).
\end{aligned}$$

Proof of the existence of C_2. Let G_0 be the outer copy of $\hat{H}_m$ in $\hat{H}_m^{\oslash k}$, and let P be a shortest path from x to y in $\hat{H}_m^{\oslash k}$. Moreover, let x' and y' be the closest vertices to x and y in $V(P \cap V(G_0))$. By Lemma 3.35,

$$\begin{aligned}
\|f_{m,k}(x) - f_{m,k}(y)\|_2^2 = \|f_{m,k}(x) - p_x(f_{m,k}(x))\|_2^2 + \|f_{m,k}(y) - p_y(f_{m,k}(y))\|_2^2 \\
+ \|p_y(f_{m,k}(y)) - p_x(f_{m,k}(x))\|_2^2,
\end{aligned}$$

and by Corollary 3.39 and Lemma 3.41

$$\|f_{m,k}(x) - f_{m,k}(y)\|_2^2 \lesssim d_{m,k}(x', x) + d_{m,k}(y', y) + \|p_y(f_{m,k}(y)) - p_x(f_{m,k}(x))\|_2^2. \quad (3.61)$$

On the other hand, triangle inequality implies

$$\begin{aligned}
&\|p_x(f_{m,k}(x)) - p_y(f_{m,k}(y))\|_2 \\
&\quad \leq \|p_x(f_{m,k}(x))) - f_{m,k}(x')\|_2 + \|f_{m,k}(x') - f_{m,k}(y')\|_2 + \|f_{m,k}(y') - p_y(f_{m,k}(y))\|_2 \\
&\quad \leq \|f_{m,k}(x) - f_{m,k}(x')\|_2 + \|f_{m,k}(x') - f_{m,k}(y')\|_2 + \|f_{m,k}(y') - f_{m,k}(y)\|_2.
\end{aligned}$$

Using the above bound and (3.61), we can write

$$\begin{aligned}\|f_{m,k}(x) - f_{m,k}(y)\|_2^2 &\lesssim d_{m,k}(x',x) + d_{m,k}(y',y) + 3\Big(\|f_{m,k}(x) - f_{m,k}(x')\|_2^2 \\ &\quad + \|f_{m,k}(x') - f_{m,k}(y')\|_2^2 + \|f_{m,k}(y') - f_{m,k}(y)\|_2^2\Big) \\ &\overset{(3.56)}{\leq} d_{m,k}(x',x) + d_{m,k}(y',y) + 6d_{m,k}(x,x') + 3\|f_{m,k}(x') - f_{m,k}(y')\|_2^2 \\ &\quad + 6d_{m,k}(y',y).\end{aligned}$$

Since, $f_{m,k}|_{G_0} = f_{m,1}$, and $f_{m,1}$ is 1-Lipschitz,

$$\begin{aligned}\|f_{m,k}(x) - f_{m,k}(y)\|_2^2 &\lesssim d_{m,k}(x',x) + d_{m,k}(y',y) + d_{m,k}(x,x') + d_{m,k}(x',y') + d_{m,k}(y',y) \\ &\lesssim d_{m,k}(x,y).\end{aligned}$$

□

3.4.8 Lower Bound for NEG

Our goal is now to prove the following theorem.

Theorem 3.42. *For $m \geq 1$ and $k = \lceil \sqrt{m} \log^2 m \rceil$, any* NEG*-embedding of $f : (V(H_m^{\oslash k}), d_{m,k}) \to \ell_2$ requires distortion $\gtrsim \frac{(\log N)^{1/3}}{\sqrt{\log \log N}}$, where $N = |V(H_m^{\oslash k})| \asymp 2^{O(mk)}$.*

We will use Theorem 3.23 to reduce our task to proving lower bounds on efficient embeddings. We define $e_1, e_2, \ldots, e_m$ to be the standard basis of $\mathbb{R}^m$, and for $x \in \{0,1\}^m$, we use $x \oplus e_k$ to denote coordinate-wise sum modulo 2. We define μ_m to be the uniform measure over s-t shortest paths in H_m of the form

$$(s, s', [x]_1, [x \oplus e_k]_1, [x]_2, [x \oplus e_k]_2, \ldots, [x]_m, [x \oplus e_k]_m, t', t),$$

where $k \in \{1, 2, \ldots, m\}$ and $x \in Q_m$ are chosen uniformly at random. We now state the main technical lemma of this section.

Lemma 3.43. *For any* NEG*-mapping $f : V(H_m) \to \ell_2$, if f is $O(\frac{1}{\sqrt{m}\log m})$-efficient with respect to μ_m, then* $\mathsf{dist}(f) \gtrsim m^{1/2}$.

Using the preceding lemma, Theorem 3.42 follows quickly.

Proof of Theorem 3.42. Suppose we have an NEG-mapping $f : V(\hat{H}_m^{\oslash k}) \to \ell_2$ with distortion at most $\sqrt{m}$. Then,

$$2\sqrt{m}\left(1 - \frac{\mathsf{len}_M(\hat{H}_m)}{\mathsf{len}(\hat{H}_m)}\right) = 2\sqrt{m}\left(1 - \frac{4m-1}{4m-1+2m^{-1}}\right) \lesssim \frac{1}{m\log m},$$

and $(\sqrt{m}\log m)(\log m)^{1/3} \lesssim k$. Therefore, by Theorem 3.23, there must exist an isometric copy of $(V(\hat{H}_m), d_{\hat{H}_m})$, in $\hat{H}_m^{\oslash k}$ such that $f|_{V(\hat{H}_m)}$ is $O(\frac{1}{\sqrt{m}\log m})$-efficient with respect to μ_m.

The graph $\hat{H}_m$ has an isometric copy of H_m as a subgraph. We further restrict f to $V(H_m)$. Since the NEG-mapping, $f|_{V(\hat{H}_m)}$ is $O(\frac{1}{\sqrt{m}\log m})$-efficient, $f|_{V(H_m)} : V(H_m) \to \ell_2$ is also $O(\frac{1}{\sqrt{m}\log m})$-efficient. By Lemma 3.43, any such embedding of H_m has distortion $\Omega(\sqrt{m})$. We have $\log N \lesssim m^{3/2}\log^2 m$, hence $\mathsf{dist}(f) \gtrsim m^{1/2} \gtrsim \frac{(\log N)^{1/3}}{\sqrt{\log\log N}}$. □

We now move onto the proof of Lemma 3.43. Let $\vec{E}(Q_m)$ denote the set of all ordered pairs (u, v), where $\{u, v\}$ is an edge of Q_m. In other words, the set of all undirected edges considered with both orientations. In everything that follows, for $u, v \in \ell_2$ we use $\langle u\,,\, v\rangle$ to denote the inner product on ℓ_2.

Lemma 3.44. *Let $\mathcal{F}$ be a finite set of functions of the form $f : Q_m \to \ell_2$. Then,*

$$\mathbb{E}_{f,g\in\mathcal{F}}\mathbb{E}_{x\in V(Q_m)}\left[\langle (f(x)-f(\bar{x}))\,,\,(g(x)-g(\bar{x}))\rangle\right] \le m\,\mathbb{E}_{f,g\in F}\mathbb{E}_{\vec{e}\in\vec{E}(Q_m)}\left[\langle f(\vec{e})\,,\,g(\vec{e})\rangle\right]. \tag{3.62}$$

Proof. Let $F(x) = \mathbb{E}_{f\in\mathcal{F}} f(x)$. Then,

$$\begin{aligned}
\mathbb{E}_{x\in V(Q_m)}[\|F(x)-F(\bar{x})\|_2^2] &= \mathbb{E}_{x\in V(Q_m)}[\langle F(x)-F(\bar{x})\,,\,F(x)-F(\bar{x})\rangle] \\
&= \mathbb{E}_{x\in V(Q_m)}[\langle \mathbb{E}_{f\in\mathcal{F}}(f(x)-f(\bar{x}))\,,\,\mathbb{E}_{g\in\mathcal{F}}(g(x)-g(\bar{x}))\rangle] \\
&= \mathbb{E}_{f,g\in\mathcal{F}}\mathbb{E}_{x\in V(Q_m)}[\langle f(x)-f(\bar{x})\,,\,g(x)-g(\bar{x})\rangle],
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}_{f,g\in\mathcal{F},\vec{e}\in\vec{E}(Q_m)}[\langle f(\vec{e})\,,\,g(\vec{e})\rangle] &= \mathbb{E}_{\vec{e}\in\vec{E}(Q_m)}[\langle \mathbb{E}_{f\in\mathcal{F}}\,f(\vec{e})\,,\,\mathbb{E}_{g\in\mathcal{F}}\,g(\vec{e})\rangle] \\
&= \mathbb{E}_{\vec{e}\in\vec{E}(Q_m)}[\langle F(\vec{e})\,,\,F(\vec{e})\rangle] \\
&= \mathbb{E}_{\vec{e}\in\vec{E}(Q_m)}[\|F(\vec{e})\|_2^2].
\end{aligned}$$

Therefore (3.62) follows from the inequality,

$$\mathbb{E}_{x\in V(Q_m)}[\|F(x)-F(\bar{x})\|_2^2] \le m \cdot \mathbb{E}_{\vec{e}\in \vec{E}(Q_m)}[\|F(\vec{e})\|_2^2],$$

which is simply the classical Poincaré inequality (3.3). □

The main idea in the proof of Lemma 3.43 is to apply Lemma 3.44 on some family

$$\mathcal{F} \subseteq \{f|_{[Q_m]_i} : 1 \le i \le m\}.$$

Efficiency and distortion, respectively, are used to bound the right and left-hand sides of (3.62).

Lemma 3.45. *For all $i, j \in [m]$, and for any 1-Lipschitz* NEG*-mapping $f : V(H_m) \to \ell_2$ with distortion at most D, the following inequality holds. For every $x \in V(Q_m)$,*

$$\langle f([x]_i) - f([\bar{x}]_i)\,,\, f([x]_j) - f([\bar{x}]_j)\rangle \ge \frac{m}{D} - 4\,|i-j|. \tag{3.63}$$

Proof. We will prove (3.63) using the following four inequalities, which follow directly from our assumptions.

- $\|f([x]_j) - f([x]_i)\|_2^2 \le 2\,|i-j|$ and $\|f([\bar{x}]_i) - f([\bar{x}]_j)\|_2^2 \le 2\,|i-j|$.
- $\|f([x]_i) - f([\bar{x}]_i)\|_2^2 \ge \frac{m}{D}$ and $\|f([x]_j) - f([\bar{x}]_j)\|_2^2 \ge \frac{m}{D}$.

We have,

$$\begin{aligned}
&\langle f([x]_i) - f([\bar{x}]_i)\,,\, f([x]_j) - f([\bar{x}]_j)\rangle \\
&\quad = \frac{1}{2}\left(\|f([x]_i) - f([\bar{x}]_i)\|_2^2 + \|f([x]_j) - f([\bar{x}]_j)\|_2^2 - \|f([x]_i) - f([\bar{x}]_i) - f([x]_j) + f([\bar{x}]_j)\|_2^2\right) \\
&\quad \ge \frac{m}{D} - \|f([x]_i) - f([x]_j)\|_2^2 - \|f([\bar{x}]_i) - f([\bar{x}]_j)\|_2^2 \\
&\quad \ge \frac{m}{D} - 4\,|i-j|.
\end{aligned}$$

□

Corollary 3.46. *For all $i, j \in [m]$ such that $|i-j| \le \frac{m}{8D}$, and for any 1-Lipschitz* NEG*-mapping $f : V(H_m) \to \ell_2$ with distortion at most D, the following inequality holds. For all $x \in V(Q_m)$,*

$$\langle f([x]_i) - f([\bar{x}]_i)\,,\, f([x]_j) - f([\bar{x}]_j)\rangle \ge \frac{m}{2D}.$$

Lemma 3.47. *Consider a* 1*-Lipschitz* NEG*-mapping* $f : V(H_m) \to \ell_2$*. Suppose that* $\mathsf{dist}(f) \le D$*, and* f *is* ε*-efficient with respect to* μ_m*, for some* $\varepsilon \ge \frac{4D}{m^2}$*. Then there exists an index* $p \in [m]$ *such that,*

$$\mathbb{E}_{\vec{e}\in\vec{E}(Q_m)}\mathbb{E}_{i,j\in\{p,\ldots,p+\ell\}}\langle f([\vec{e}]_i)\,,\,f([\vec{e}]_j)\rangle \lesssim \frac{D}{m}+\varepsilon, \tag{3.64}$$

where $\ell = \lfloor \frac{m}{8D} \rfloor$*.*

Proof. To prove this lemma first we examine the slack on subpaths. For $1 \le p < m-\ell$, and $x \in V(Q_m)$ consider the quantity

$$\sigma_{p,k}(x) = \sum_{r=p}^{p+\ell}\Big(\|f([x]_r)-f([x\oplus e_k]_r)\|_2^2+\|f([x\oplus e_k]_r)-f([x]_{r+1})\|_2^2\Big)-\|f([x]_p)-f([x]_{p+\ell+1})\|_2^2\,.$$

Let $s_p = \mathbb{E}_{x\in V(Q_m)}\mathbb{E}_{k\in[m]}[\sigma_{p,k}(x)]$.

Since f is ε-efficient with respect to μ_m, we have

$$\mathbb{E}_{p\in[m-\ell-1]}\,[s_p] \lesssim \varepsilon(\ell+1).$$

Therefore there must exist an index $p \in [m-\ell-1]$ such that $s_p \lesssim \varepsilon\ell$. We show that any such p satisfies (3.64).

To this end, fix $p \in [m-\ell-1]$ such that $s_p \lesssim \varepsilon\ell$, we will bound the value

$$\mathbb{E}_{\vec{e}\in\vec{E}(Q_m)}\mathbb{E}_{i,j\in\{p,\ldots,p+\ell\}}\langle f([\vec{e}]_i)\,,\,f([\vec{e}]_j)\rangle = \frac{\mathbb{E}_{\vec{e}\in\vec{E}(Q_m)}\left\|\sum_{i=p}^{p+\ell} f([\vec{e}]_i)\right\|_2^2}{(\ell+1)^2}, \tag{3.65}$$

by splitting the sum on the right hand side into two parts,

$$\left\|\sum_{i=p}^{p+\ell} f([\vec{e}]_i)\right\|_2^2 = \sum_{i=p}^{p+\ell}\|f([\vec{e}]_i)\|_2^2+\sum_{i,j=p\,:\,i\neq j}^{p+\ell}\langle f([\vec{e}]_i)\,,\,f([\vec{e}]_j)\rangle. \tag{3.66}$$

Since f is 1-Lipschitz, we have $\|f([\vec{e}]_i)\|_2^2 \le 1$, and therefore

$$\Big\|\sum_{i=p}^{p+\ell} f([\vec{e}]_i)\Big\|_2^2 \le (\ell+1)+\sum_{i,j=p:i\neq j}^{p+\ell}\langle f([\vec{e}]_i)\,,\,f([\vec{e}]_j)\rangle. \tag{3.67}$$

Thus by bounding the latter sum, we can bound (3.66), and complete the proof. For a given p and $\vec{e} \in \vec{E}(Q_m)$, let

$$\nu_{p,\vec{e}} = \max_{i\in\{p,\ldots,p+\ell\}}\sum_{\substack{j\in\{p,\ldots,p+\ell\}\\ j\neq i}}\langle f([\vec{e}]_i)\,,\,f([\vec{e}]_j)\rangle. \tag{3.68}$$

We claim that

$$s_p \geq \mathbb{E}_{\vec{e} \in \vec{E}(Q_m)} \nu_{p,\vec{e}}. \tag{3.69}$$

In this case, combining (3.69) and (3.67), we will have

$$\mathbb{E}_{\vec{e} \in \vec{E}(Q_m)} \left\| \sum_{i=p}^{p+\ell} f([\vec{e}]_i) \right\|_2^2 \lesssim (\ell+1) + (\ell+1)\mathbb{E}_{\vec{e} \in \vec{E}(Q_m)} \nu_{p,\vec{e}} \leq (\ell+1) + (\ell+1)s_p.$$

Applying (3.65) and using the fact that $s_p \lesssim \varepsilon \ell$, we will conclude that

$$\mathbb{E}_{\vec{e} \in \vec{E}(Q_m)} \mathbb{E}_{i,j \in [p,p+\ell]} \langle f([\vec{e}]_i) \, , \, f([\vec{e}]_j) \rangle \leq \frac{1+s_p}{\ell+1} \lesssim \frac{1}{\ell} + \varepsilon \lesssim \frac{D}{m} + \varepsilon,$$

completing the proof.

We now proceed to prove (3.69). Fix $\vec{e} = (x, x \oplus e_k)$ and let i_0 be the index which maximize (3.68). First note that since $\hat{H}_m$ is symmetric with respect to s and t we may assume that

$$\sum_{\substack{j \in \{p,\ldots,p+\ell\} \\ j < i_0}} \langle f([\vec{e}]_{i_0}) \, , \, f([\vec{e}]_j) \rangle \geq \sum_{\substack{j \in \{p,\ldots,p+\ell\} \\ j > i_0}} \langle f([\vec{e}]_{i_0}) \, , \, f([\vec{e}]_j) \rangle. \tag{3.70}$$

We need the following claim to prove (3.69).

Claim 3.48. *For any $j > i_0$, we have*

$$\begin{aligned} &\|f([x \oplus e_k]_{i_0}) - f([x]_j)\|_2^2 - \|f([x \oplus e_k]_{i_0}) - f([x]_{j-1})\|_2^2 \\ &\qquad \leq \|f([x]_{j-1}) - f([x \oplus e_k]_j)\|_2^2 + \|f([x \oplus e_k]_j) - f([x]_j)\|_2^2 - 2\langle f([\vec{e}_k]_{i_0}) \, , \, f([\vec{e}_k]_j) \rangle \end{aligned}$$

Proof. First note that triangle inequality implies that

$$\|f([x \oplus e_k]_{i_0}) - f([x]_j)\|_2^2 - \|f([x \oplus e_k]_{i_0}) - f([x \oplus e_k]_j)\|_2^2 \leq \|f([x]_{j-1}) - f([x \oplus e_k]_j)\|_2^2 \tag{3.71}$$

Moreover, for $u, v, w \in \ell_2$, the triangle inequality for $\|\cdot\|_2^2$ is equivalent to $\langle u - w \, , \, w - v \rangle \leq$

0, hence

$$
\begin{aligned}
\|f([x \oplus e_k]_{i_0}) - f([x]_j)\|_2^2 &= \|f([x \oplus e_k]_{i_0}) - f([x \oplus k]_{j-1})\|_2^2 + \|f([x \oplus e_k]_{j-1}) - f([x]_j)\|_2^2 \\
&\quad + 2\langle f([x \oplus e_k]_{i_0}) - f([x]_j)\,,\, f([x]_j) - f([x \oplus e_k]_j)\rangle \\
&= \|f([x \oplus e_k]_{i_0}) - f([x \oplus k]_{j-1})\|_2^2 + \|f([x \oplus e_k]_{j-1}) - f([x]_j)\|_2^2 \\
&\quad + 2\langle f([x]_{i_0}) - f([x]_j)\,,\, f([x]_j) - f([x \oplus e_k]_j)\rangle \\
&\quad + 2\langle f([x \oplus e_k]_{i_0}) - f([x]_{i_0})\,,\, f([x]_j) - f([x \oplus e_k]_j)\rangle \\
&\le \|f([x \oplus e_k]_{i_0}) - f([x \oplus k]_{j-1})\|_2^2 + \|f([x \oplus e_k]_{j-1}) - f([x]_j)\|_2^2 \\
&\quad - 2\langle f([x]_{i_0}) - f([x \oplus e_k]_{i_0})\,,\, f([x]_j) - f([x \oplus e_k]_j)\rangle
\end{aligned}
$$

This inequality in combination with inequality (3.71) completes the proof. □

Applying this claim for all $j > i_0$ and using the bound from (3.70) implies

$$
\begin{aligned}
\nu_{p,e_k} &\le \sum_{r=p}^{p+\ell} \left(\|f([x]_r) - f([x \oplus e_k]_r)\|_2^2 - \|f([x \oplus e_k]_r) - f([x]_{r+1})\|_2^2\right) - \|f([x]_p) - f([x]_{p+\ell+1})\|_2^2 \\
&\le \frac{1}{2}\left(\sum_{j=i_0+1}^{p+\ell} 2\langle f([x]_{i_0}) - f([x \oplus e_k]_{i_0})\,,\, f([x]_j) - f([x \oplus e_k]_j)\rangle\right) = \sigma_{p,k}(x),
\end{aligned}
$$

completing the proof.

□

Finally, we present the proof of Lemma 3.43 to complete this section.

Proof of Lemma 3.43. Consider a 1-Lipschitz NEG-mapping $f : V(H_m) \to \ell_2$ which is ε-efficient with respect to μ_m, where $\varepsilon = O(\frac{1}{\sqrt{m}\log m})$. Let p satisfy the conclusion of Lemma 3.47. Furthermore, let $f_i : V(Q_m) \to \ell_2$ be defined by $f_i(x) = f([x]_i)$, and put

$$
\mathcal{F} = \left\{ f_i : p \le i \le p + \left\lfloor \frac{m}{8D} \right\rfloor \right\}.
$$

An immediate application of Lemma 3.47 yields,

$$
\mathbb{E}_{f,g \in \mathcal{F}, \vec{e} \in \vec{E}(Q_m)}[f(\vec{e}) \cdot g(\vec{e})] \lesssim \frac{D}{m} + \frac{\varepsilon m}{D}.
$$

From Corollary 3.46 we have,

$$
\mathbb{E}_{f,g \in \mathcal{F}, x \in V(Q_m)}[(f(x) - f(\bar{x})) \cdot (g(x) - g(\bar{x}))] \ge \frac{m}{2D}.
$$

Applying Lemma 3.44 with the given bounds the following inequality must hold:

$$m \gtrsim \frac{\frac{m}{D}}{\frac{D}{m} + \varepsilon} \gtrsim \frac{1}{\frac{D\varepsilon}{m} + \frac{D^2}{m^2}},$$

therefore

$$D\varepsilon + \frac{D^2}{m} \gtrsim 1,$$

and $D \gtrsim m^{1/2}$. □

Chapter 4

NODE-CAPACITATED OKAMURA-SEYMOUR

4.1 Organization

In this chapter we study the relationship between Max-flow and Min-cut in planar graphs. We first state all the results that we prove in Section 4.2. Then, we give an overview of the techniques that we use in the proofs in Section 4.3. In the rest of this chapter we provide the proof of the results stated in Section 4.2.

4.2 Results

An *undirected flow network* is an undirected graph $G = (V, E)$ together with a capacity function on edges $\mathsf{cap} : E \to [0, \infty)$. A set of *demands* is specified by a symmetric mapping $\mathsf{dem} : V \times V \to [0, \infty)$. For $u, v \in V$, denote by $\varphi_{uv} : E \to [0, \infty)$ the undirected u-v flow. The (edge) capacity constrains require that for every $e \in E$, $\sum_{u,v \in V} \varphi_{uv}(e) \le \mathsf{cap}(e)$. Given such an instance, let $\mathsf{mcf}_G(\mathsf{cap}, \mathsf{dem})$ be the largest value ε such that one can simultaneous route $\varepsilon \cdot \mathsf{dem}(u, v)$ units of flow between u and v for every $u, v \in V$ while not violating any of the edge capacities. This optimization describes the *maximum concurrent flow problem.*

For two subsets $S, T \subseteq V$, let $\mathsf{cap}(S, T)$ denote the total capacity of all edges with one endpoint in S and one in T. Similarly, let $\mathsf{dem}(S, T) = \sum_{u \in S} \sum_{v \in T} \mathsf{dem}(u, v)$. To give an upper bound on mcf, we can consider cuts in G, described by subsets $S \subseteq V$. To every such subset we assign a value called the *sparsity* of the cut:

$$\Phi_G(S; \mathsf{cap}, \mathsf{dem}) = \frac{\mathsf{cap}(S, \bar{S})}{\mathsf{dem}(S, \bar{S})} .$$

It is straightforward to see that for any $S \subseteq V$, we have $\mathsf{mcf}_G(\mathsf{cap}, \mathsf{dem}) \le \Phi_G(S; \mathsf{cap}, \mathsf{dem})$. See Figure **??** The *sparsest cut* is the one which gives the best upper bound on mcf. In this vein, we define

$$\Phi_G(\mathsf{cap}, \mathsf{dem}) = \min_{S \subseteq V} \Phi_G(S; \mathsf{cap}, \mathsf{dem}) .$$

Thus we have the relationship $\mathsf{mcf}_G(\mathsf{cap}, \mathsf{dem}) \le \Phi_G(\mathsf{cap}, \mathsf{dem})$ and the flow/cut gap question asks how close this upper bound is to the truth.

To state the Okamura-Seymour theorem, we need one final piece of notation. We say that the demand function dem is *supported on a subset* $D \subseteq V$ if $\mathsf{dem}(u, v) > 0$ only when $u, v \in D$. The classical Max-flow Min-cut Theorem [43] implies that if the demand dem is supported on a two-element subset $\{s, t\} \subseteq V$, then for any capacities cap, we have $\mathsf{mcf}_G(\mathsf{cap}, \mathsf{dem}) = \Phi_G(\mathsf{cap}, \mathsf{dem})$. An extension of Hu [48] shows that if dem is supported on a 4-element subset $D \subseteq V$, the same equality holds. The Okamura-Seymour theorem states that whenever G is a planar graph and the demand is supported on a single face, there is likewise no flow/cut gap.

Theorem 4.1 ([89]). *Let $G = (V, E)$ be a planar graph, and let $F \subseteq V$ be any face of G. Then for any capacities $\mathsf{cap} : E \to [0, \infty)$ and any demands $\mathsf{dem} : V \times V \to [0, \infty)$ supported on F, we have*

$$\mathsf{mcf}_G(\mathsf{cap}, \mathsf{dem}) = \Phi_G(\mathsf{cap}, \mathsf{dem})\,.$$

Indeed, one can consider generalizations of edge-capacitated networks. A prominent example is to consider capacities on vertices.

Formally, we define a *vertex-capacitated flow network* by considering a function $\mathsf{cap} : V \to [0, \infty)$ assigning capacities to vertices instead of edges. It seems that the most elegant way to think about capacities in this setting is as follows: If a flow of value α is sent along a path P from s to t, then it consumes $\alpha/2$ capacity at s and t and α capacity at each of the intermediate nodes of P.[1] Formally, in the multi-commodity setting, the vertex capacity constrains require that for every $w \in V$, $\sum_{e \ni w} \sum_{u,v \in V} \varphi_{uv}(e) \le 2\,\mathsf{cap}(w)$. The corresponding definition of the maximum concurrent flow follows immediately; we use the notation mcf_G^v for the vertex-capacitated version. For the definition of Φ_G^v, we have to be slightly more careful. For a subset $S \subseteq V$ of the vertices, denote by $G[S]$ the induces

[1] This particular choice does not materially affect any theorem in this dissertation which deals with approximate flow/cut gaps.

subgraph of G on S. We define a function $\rho_S : V \times V \to \{0, \frac{1}{2}, 1\}$ by

$$\rho_S(u,v) = \begin{cases} \frac{1}{2} & |\{u,v\} \cap S| = 1 \\ 1 & u,v \in S \\ 1 & u,v \in \bar{S} \text{ and } u,v \text{ are in distinct connected components of } G[\bar{S}] \\ 0 & \text{otherwise.} \end{cases}$$

In other words, we are only given half-credit for separating u and v if exactly one of them is in the separator (see Figure 4.1).

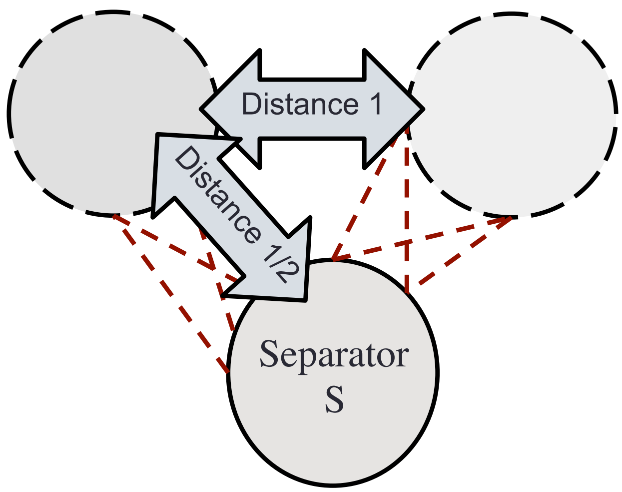


Figure 4.1: Vertex separator distance between points.

Then we define

$$\Phi^v_G(S; \mathsf{cap}, \mathsf{dem}) = \frac{\sum_{v \in S} \mathsf{cap}(v)}{\sum_{u,v \in V} \mathsf{dem}(u,v)\rho_S(u,v)},$$

and $\Phi^v_G(\mathsf{cap}, \mathsf{dem}) = \min_{S \subseteq V} \Phi^v_G(S; \mathsf{cap}, \mathsf{dem})$. It is straightforward to verify that $\mathsf{mcf}^v_G(\mathsf{cap}, \mathsf{dem}) \leq \Phi^v_G(\mathsf{cap}, \mathsf{dem})$.

These precise definitions ensure that a classical Max-flow Min-cut theorem holds when the demand is supported on a single pair (this follows from Menger's theorem [86]). They also allow other natural properties in the multi-commodity setting; it is an exercise to show that for any tree T, we have $\mathsf{mcf}^v_T(\mathsf{cap}, \mathsf{dem}) = \Phi^v_T(\mathsf{cap}, \mathsf{dem})$ for any choice of capacities and demands. Unfortunately, there is no exact vertex-capacitated analog of the Okamura-Seymour Theorem. Nevertheless, a main result of the current chapter is that an approximate version does hold in the vertex-capacitated setting, answering a question posed by Chekuri and Kawarabayashi.

Theorem 4.2 (Restatement of Theorem 1.2)**.** *There exists a constant $\varepsilon > 0$ such that the following holds. Let $G = (V, E)$ be a planar graph and let $F \subseteq V$ be any face of G. Then for any vertex capacities $\mathsf{cap} : V \to [0, \infty)$ and any demands $\mathsf{dem} : V \times V \to [0, \infty)$ supported on F, we have*

$$\mathsf{mcf}^v_G(\mathsf{cap}, \mathsf{dem}) \geq \varepsilon \cdot \Phi^v_G(\mathsf{cap}, \mathsf{dem})\,.$$

In fact, our result holds in the more general setting of undirected polymatroid networks which we discuss next.

4.2.1 Polymatroid Networks

Motivated by applications to information flow in wireless networks, Chekuri et. al. [29] introduced a generalization of vertex capacities by putting a submodular capacity function at every vertex. Recall that a function $f : 2^S \to \mathbb{R}$ over a finite set S is called *submodular* if $f(A) + f(B) \geq f(A \cap B) + f(A \cup B)$ for all $A, B \subseteq S$. Let $G = (V, E)$ be a graph and suppose that for a multi-commodity flow $\varphi = \{\varphi_{st}\}_{s,t \in V}$ in G, we use $\varphi(e) = \sum_{s,t} \varphi_{st}(e)$ to denote the total flow through the edge e. For a vertex $v \in V$, we use $E(v)$ to denote the edges incident to v. Let $\vec{\rho} = \{\rho_v : 2^{E(v)} \to [0, \infty)\}_{v \in V}$ be a collection of monotone, submodular functions called *polymatroid capacities.* A flow φ is *feasible with respect to $\vec{\rho}$* if for every $v \in V$ and every subset $S \subseteq E(v)$, we have $\sum_{e \in S} \varphi(e) \leq \rho_v(S)$.

Given a demand function $\mathsf{dem} : V \times V \to [0, \infty)$, we can define the *maximum concurrent flow value* of a polymatroid network by $\mathsf{mcf}_G(\vec{\rho}, \mathsf{dem})$ as the maximum $\varepsilon > 0$ such that one can route an ε-fraction of all demands simultaneously using a flow that is feasible with respect to $\vec{\rho}$.

The corresponding notion of a sparse cut is now a little trickier. For every subset of edges $S \subseteq E$, we can define the cut semi-metric $\sigma_S : V \times V \to \{0, 1\}$ on V by $\sigma_S(x, y) = 0$

if and only if there exists a path from x to y in the graph $G(V, E \setminus S)$. Following [29], we call a map $g : S \to V$ *valid* if it maps every edge in S to one of its two endpoints in V. We can then define the *capacity of a set* $S \subseteq E$ by

$$\nu_{\vec{\rho}}(S) = \min_{\substack{g:S\to V \\ \text{valid}}} \sum_{v\in V} \rho_v(g^{-1}(v)) \,.$$

Finally, we define the *sparsity of* S by

$$\Phi_G(S; \vec{\rho}, \mathsf{dem}) = \frac{\nu_{\vec{\rho}}(S)}{\sum_{u,v\in V} \mathsf{dem}(u,v)\sigma_S(u,v)} \,,$$

and define $\Phi_G(\vec{\rho}, \mathsf{dem}) = \min_{S\subseteq V} \Phi_G(S; \vec{\rho}, \mathsf{dem})$. It is not too difficult to see that, again,

$$\mathsf{mcf}_G(\vec{\rho}, \mathsf{dem}) \leq \Phi_G(\vec{\rho}, \mathsf{dem}) \,.$$

In [29], it is proved that when dem is supported on a single pair, we have

$$\Phi_G(\vec{\rho}, \mathsf{dem}) \leq 2 \cdot \mathsf{mcf}_G(\vec{\rho}, \mathsf{dem}) \,.$$

Unfortunately, the factor 2 is necessary, and owes itself to a slight defect in the notion of undirected polymatroid networks. If one were to say that a flow only consumes half the capacity of an edge if it originates at an endpoint (as in the vertex-capacitated case described above), then we would obtain an exact single-commodity max-flow/min-cut theorem in this setting. Indeed, for directed polymatroid networks, such a result is classical [47, 61]. Since we are concerned here with approximate flow/cut gaps, this will not be an issue, and we follow [29]. We obtain an Okamura-Seymour theorem for polymatroid networks as well, answering a question posed to us by Chandra Chekuri.

Theorem 4.3 (Polymatroid Okamura-Seymour Theorem)**.** *There exists a constant $\varepsilon > 0$ such that the following holds. Let $G = (V, E)$ be a planar graph and let $F \subseteq V$ be any face of G. Then for any polymatroid capacities $\vec{\rho}$ and any demands* $\mathsf{dem} : V \times V \to [0, \infty)$ *supported on F, we have*

$$\mathsf{mcf}_G(\vec{\rho}, \mathsf{dem}) \geq \varepsilon \cdot \Phi_G(\vec{\rho}, \mathsf{dem}) \,.$$

Theorem 4.2 is a special case of Theorem 4.3. Indeed, vertex capacity $\mathsf{cap} : V \to [0, \infty)$, is (up to a factor of 2) equivalent to vertex polymatroid capacity $\vec{\rho}_v(\emptyset) = 0$ and $\vec{\rho}_v(S) = \mathsf{cap}(v)$ for $\emptyset \neq S \subseteq E(v)$. With this definition of $\vec{\rho}$, it is immediate to check that $\mathsf{mcf}_G(\vec{\rho}, \mathsf{dem}) \leq \mathsf{mcf}^v_G(\mathsf{cap}, \mathsf{dem}) \leq 2\mathsf{mcf}_G(\vec{\rho}, \mathsf{dem})$ and $\Phi_G(\vec{\rho}, \mathsf{dem}) \leq \Phi^v_G(\mathsf{cap}, \mathsf{dem}) \leq 2\Phi_G(\vec{\rho}, \mathsf{dem})$.

4.3 Overview of the Proofs and Techniques

4.3.1 Embeddings and Flow/Cut Gaps

Our main tools in proving Theorems 4.2 and 4.3 are various embeddings of metric spaces. To this end, we first recall known results in the edge and vertex-capacitated settings. In the next section, we discuss the new types of embeddings we need to handle vertex-capacitated and polymatroid networks.

A *metric graph* $G = (V, E, \mathsf{len})$ is an undirected graph equipped with a non-negative length function on edges $\mathsf{len} : E \to [0, \infty)$. We extend the length function to paths $P \subseteq E$ by setting $\mathsf{len}(P) = \sum_{e \in P} \mathsf{len}(e)$. Associated to every such length is the shortest-path pseudo-metric on G defined by $d_{\mathsf{len}}(u, v) = \min_P \mathsf{len}(P)$ where the minimum is over all u-v paths P in G. We say that a pseudo-metric d on V is *supported on the graph* G if $d = d_{\mathsf{len}}$ for some length function on E. In many situations we will only be considering a single length function on G at a time, and then we write d_G instead of d_{len}.

We will consider embeddings of such graph metrics into various other spaces. Given two metric spaces (X, d_X) and (Y, d_Y) and a function $f : X \to Y$, we define the *Lipschitz constant of f* by

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

If $\|f\|_{\mathrm{Lip}} \le L$, we say that f is *L-Lipschitz.*

We define the *distortion* of the map f by $\mathsf{dist}(f) = \|f\|_{\mathrm{Lip}} \cdot \|f^{-1}\|_{\mathrm{Lip}}$. The L_1 *distortion of a metric space* (X, d_X), written $c_1(X, d_X)$, denotes the infimum of $\mathsf{dist}(f)$ over all maps $f : X \to L_1$. The next theorem gives a tight relationship between flow/cut gaps in graphs and L_1 embeddings of the metric supported on them. It follows from [76] and [46].

Theorem 4.4. *Consider any graph $G = (V, E)$ and any subset $D \subseteq V$. Let*

$$K_1(G, D) = \sup_{\mathsf{cap},\mathsf{dem}} \frac{\Phi_G(\mathsf{cap}, \mathsf{dem})}{\mathsf{mcf}_G(\mathsf{cap}, \mathsf{dem})}\,,$$

where the supremum is over all capacity functions $\mathsf{cap} : E \to [0, \infty)$ and all demand functions supported on D. Let

$$K_2(G, D) = \sup_d \left[\inf_{f:(V,d)\to L_1} \mathsf{dist}(f|_D) \right],$$

where the supremum is over all metrics d supported on G and the infimum is over all 1-Lipschitz mappings $f : V \to L_1$. Then $K_1(G, D) = K_2(G, D)$.

In particular, the Okamura-Seymour Theorem (Thm. 4.1) can be restated as the following fact about embeddings of planar graphs: For any metric planar graph $G = (V, E)$ and any face $F \subseteq V$, there exists a 1-Lipschitz mapping $f : V \to L_1$ such that $\mathsf{dist}(f|_F) = 1$.

Vertex-capacitated flows and ℓ_1^{dom} embeddings. Unfortunately, L_1 embeddings are not sufficient for the study of vertex-capacitated flow/cut gaps; we refer to [42] for some examples. Instead, [42] uses a stronger notion of embedding. For simplicity, we discuss such embeddings only for finite metric spaces. An ℓ_1^{dom} *embedding of a finite pseudometric space* (X, d) is a random 1-Lipschitz mapping $\Lambda : X \to \mathbb{R}$. One then defines

$$\mathsf{dist}(\Lambda) = \max_{x,y \in X} \frac{\mathbb{E}\,|\Lambda(x) - \Lambda(y)|}{d(x,y)}\,,$$

and writes $c_1^{\mathrm{dom}}(X, d)$ for the infimum of $\mathsf{dist}(\Lambda)$ over all such random mappings $\Lambda : X \to \mathbb{R}$. It is straightforward to verify that $c_1(X, d) \leq c_1^{\mathrm{dom}}(X)$ and there are many interesting cases when this inequality is strict (see [42, 17]). Such embeddings were initially studied by Matousek and Rabinovich [81]. It was shown in [42] that they can be used to bound vertex-capacitated flow/cut caps, and [29] extended this to undirected polymatroid networks.

Theorem 4.5 ([29]). *Consider a graph $G = (V, E)$ and a subset $D \subseteq V$. Suppose there is a constant $K \geq 1$ such that for every metric d supported on G, we have $c_1^{\mathrm{dom}}(D, d) \leq K$. Then for every set of polymatroid capacities $\vec{\rho}$ on G and every $\mathsf{dem} : V \times V \to [0, \infty)$ supported on D, we have*

$$\mathsf{mcf}_G(\vec{\rho}, \mathsf{dem}) \geq \frac{1}{2K} \Phi_G(\vec{\rho}, \mathsf{dem})\,.$$

Despite the power of the preceding theorem, it is insufficient for proving our main results. Since $c_1^{\mathrm{dom}}(X, d)$ is at least the Euclidean distortion of (X, d), Bourgain's lower bound on the Euclidean distortion of trees [13] implies that there are n-point tree metrics (T_n, d_n) with $c_1^{\mathrm{dom}}(T_n, d_n) = \Omega(\sqrt{\log\log n})$. In the next section, we introduce a new notion of embedding that is sufficient for proving vertex-capacitiated and polymatroid versions of the Okamura-Seymour theorem.

4.3.2 Length Functions, Star-shaped Embeddings, and Single-scale Gradients

We first setup a polymatroid embedding problem which follows from the duality theorem of [29]. Fix a finite ground set S. Given a function $\rho : \{0,1\}^S \to \{0,1\}$, we define its *Lovász extension* $\hat{\rho} : [0,\infty)^S \to [0,\infty)$ by

$$\hat{\rho}(z) = \int_0^\infty \rho(z^\theta) d\theta \,,$$

where $z^\theta \in \{0,1\}^S$ has $(z^\theta)_i = 1$ whenever $z_i \geq \theta$. Observe that for a constant $\alpha > 0$, we have $\hat{\rho}(\alpha \cdot z) = \alpha \cdot \hat{\rho}(z)$. We will associate 2^S and $\{0,1\}^S$ via the mapping which sends a subset $A \subseteq S$ to its characteristic function $\mathbf{1}_A \in \{0,1\}^S$. Likewise, we will associate functions $S \to [0,\infty)$ with elements of $[0,\infty)^S$.

In the rest of this section, we will consider families of functions $\mathcal{F} = \{\ell_v : E(v) \to [0,\infty)\}_{v \in V}$ associated to a graph $G = (V,E)$. Given a length function $\mathsf{len} : E \to [0,\infty)$, we say that $\mathcal{F}$ *is adapted to* len if for every edge $e = \{u,v\} \in E$, we have

$$\mathsf{len}(e) \leq \ell_u(e) + \ell_v(e) \,.$$

Theorem 4.6 (Duality Theorem, [29, Sec. 3])**.** *For any graph $G = (V,E)$ the following holds. For any polymatroid capacities $\vec{\rho} = \{\rho_v : v \in V\}$ and any demands $\mathsf{dem} : V \times V \to [0,\infty)$,*

$$\mathsf{mcf}_G(\vec{\rho}, \mathsf{dem}) = \min_{\mathsf{len}, \{\ell_v\}} \left[\frac{\sum_{v \in V} \hat{\rho}_v(\ell_v)}{\sum_{u,v \in V} \mathsf{dem}(u,v) d_{\mathsf{len}}(u,v)} \right], \tag{4.1}$$

where the minimum is over all length functions $\mathsf{len} : E \to [0,\infty)$ on G and all families $\{\ell_v : E(v) \to [0,\infty)\}_{v \in V}$ adapted to len.

The preceding theorem shows that to prove flow/cut gaps, it suffices to find for every given length function len and any $\{\ell_v\}_{v \in V}$ adapted to len, a set $S \subseteq E$ for which

$$\Phi_G(S; \vec{\rho}, \mathsf{dem}) \leq C \cdot \frac{\sum_{v \in V} \hat{\rho}_v(\ell_v)}{\sum_{u,v \in V} \mathsf{dem}(u,v) d_{\mathsf{len}}(u,v)}$$

for some constant $C > 0$. This gives rise to an embedding problem with differs from the classical one in a way which we now describe informally.

In the case of edge-capacitated flows and L_1 embeddings, to satisfy the Lipschitz property, it suffices to consider the stretch of each edge separately. For vertex-capacitated flows, and more generally polymatroid networks, we must *coordinate* the stretch of the edges adjacent to a vertex. In essence, a vertex has to "pay" in the corresponding "Lipschitz constant" if *any* of its adjacent edges is stretched. Thus we should try as much as possible to stretch the edges adjacent to a vertex simultaneously.

This makes some standard techniques (e.g. random embeddings into trees as in [46]) inappropriate for our study (although some of the principles in [46] will prove invaluable). Certainly ℓ_1^{dom} embeddings achieve this coordination because they are (by definition) Lipschitz in every coordinate, but as we mentioned earlier, they are insufficient for proving our main theorems.

To satisfy this goal, we must pay careful attention to the image of the edges in our embeddings. On the other hand, to overcome the limitations of ℓ_1^{dom}, we will increase our target spaces to include general metric trees.

Star-shaped mappings. Say that a graph H is *star-shaped* if H is the subdivision of some star graph. Suppose that $G = (V, E)$ is a graph, T is a tree, and $\lambda : V \to V(T)$ is an arbitrary map. For every $u, v \in V(T)$, let $P_{uv} \subseteq V(T)$ be the unique simple path between u and v in T. We say that λ is a *star-shaped mapping* if, for every $u \in V(T)$, the induced graph on

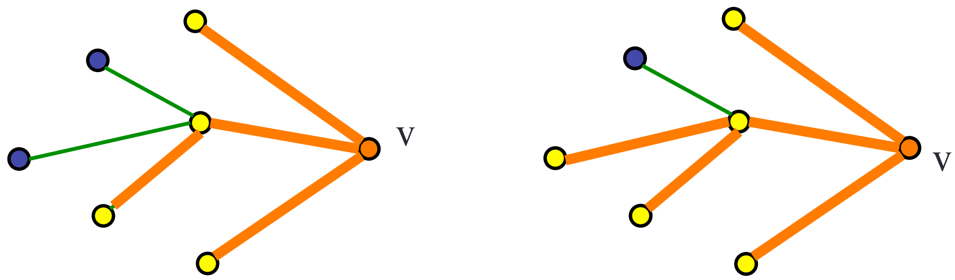
$$\{P_{uv} : v \in V(T), E(\lambda^{-1}(u), \lambda^{-1}(v)) \neq \emptyset\}$$

is star-shaped. In other words, if we consider the paths in T which correspond to edges in G, then all such paths emanating from the same vertex in T should form a star-shaped subgraph. See Figure 4.2 for an example.

In addition to controlling the *shape* of a mapping, we need to control the lengths of the "arms" of the star simultaneously. Fortunately (and this property will be crucial to the approach of Section 4.6), we will only need to bound the stretch over single scales.

Single-scale ℓ_∞ gradients. If we are given a metric graph $G = (V, E, \mathsf{len})$ and a mapping $f : V \to (X, d_X)$ into a metric space (X, d_X), we make the following definition: For any $\tau > 0$,

$$|\nabla_\tau f(u)|_\infty = \sup \left\{ \frac{d_X(f(u), f(v))}{\mathsf{len}(u, v)} : \{u, v\} \in E \text{ and } \mathsf{len}(u, v) \in [\tau, 2\tau] \right\} .$$



(a) The induced graph on $\{P_{uv} : v \in V(T), E(\lambda^{-1}(u), \lambda^{-1}(v)) \neq \emptyset\}$ is star-shaped.

(b) The induced graph on $\{P_{uv} : v \in V(T), E(\lambda^{-1}(u), \lambda^{-1}(v)) \neq \emptyset\}$ has two vertices with degree three, hence it is not star-shaped.

Figure 4.2: Star-shaped embedding.

In Section 4.4, we prove the following theorem which shows how such mappings can be used for polymatroid flow/cut gaps.

Theorem 4.7 (Main rounding theorem)**.** *Let $G = (V, E, \mathsf{len})$ be a metric graph and suppose there exists a random metric tree T and a random star-shaped mapping $F : V \to V(T)$ such that for some $K \geq 1$,*

$$\max_{v \in V} \sup_{\tau > 0} \mathbb{E}\, |\nabla_\tau F(v)|_\infty \leq K\,. \tag{4.2}$$

Then for any family of functions $\{\ell_v : E(v) \to [0, \infty)\}_{v \in V}$ adapted to len, and for any polymatroid capacities $\vec{\rho} = \{\rho_v\}_{v \in V}$ and demands $\mathsf{dem} : V \times V \to [0, \infty)$, we have

$$\Phi_G(\vec{\rho}, \mathsf{dem}) \leq \frac{64K \sum_{v \in V} \hat{\rho}_v(\ell_v)}{\sum_{u,v \in V} \mathsf{dem}(u, v) \cdot \mathbb{E}\left[d_T(F(u), F(v))\right]}\,. \tag{4.3}$$

4.3.3 The Embedding Theorem

In light of Theorem 4.7, we are able to prove Theorem 4.3 by constructing appropriate random embeddings into trees. In the present section we state our main embedding theorem and give an outline of its proof.

Theorem 4.8. *There exist constants $K, L \geq 1$ such that the following holds. If $G = (V, E)$ is a metric planar graph, and $F \subseteq V$ is any face of G, then there exists a random tree T*

and random star-shaped mapping $\Lambda : V \to V(T)$ such that the following conditions hold.

i) For every $u \in V$ and $\tau > 0$, we have $\mathbb{E}\,|\nabla_\tau \Lambda(u)|_\infty \le K$.

ii) For every $u, v \in F$,

$$\mathbb{E}\left[d_T(\Lambda(u), \Lambda(v))\right] \ge \frac{d_G(u,v)}{L}\,. \tag{4.4}$$

Combined with the rounding theorem (Theorem 4.7) and duality (Theorem 4.6), this immediately yields Theorem 4.3 and, in particular, a vertex-capacitated Okamura-Seymour theorem (Theorem 4.2).

Proof of Theorem 4.3. Fix a planar graph $G = (V, E)$, a face $F \subseteq V$ of G, demands $\mathsf{dem} : V \times V \to [0, \infty)$ supported on F, and polymatroid capacities $\vec{\rho}$. By Theorem 4.7, there exists a length function $\mathsf{len} : E \to [0, \infty)$ and a family $\{\ell_v : E(v) \to [0, \infty)\}_{v \in V}$ adapted to len such that

$$\mathsf{mcf}_G(\vec{\rho}, \mathsf{dem}) = \frac{\sum_{v \in V} \hat{\rho}_v(\ell_v)}{\sum_{u,v \in V} \mathsf{dem}(u,v) d_{\mathsf{len}}(u,v)}\,.$$

Consider the metric planar graph $G = (V, E, \mathsf{len})$. By Theorem 4.8 there exist a random tree T and a random star-shaped embedding $\Lambda : V \to T$ satisfying (4.2) with $K = 1$, and (4.4) with some universal constant $L > 0$. Applying Theorem 4.7 with Λ, we conclude

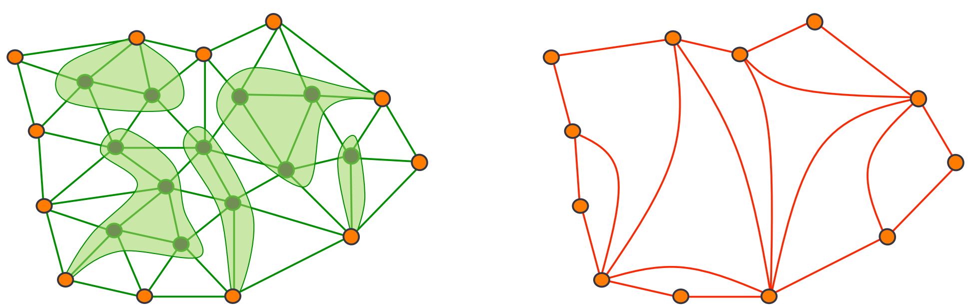
$$\Phi_G(\vec{\rho}, \mathsf{dem}) \overset{(4.3)\wedge(4.4)}{\le} \frac{64KL \sum_{v \in V} \hat{\rho}(\ell_v)}{\sum_{u,v \in V} \mathsf{dem}(u,v) d_G(u,v)} \overset{(4.1)}{=} 64KL \cdot \mathsf{mcf}_G(\vec{\rho}, \mathsf{dem}). \qquad \square$$

We now give a brief outline of the proof of Theorem 4.8.

First step: Outerplanar graphs into random trees. Theorem 4.8 is proved in two main steps. First, in Section 4.5, we prove it for the special case of outerplanar graphs; this is precisely the situation where the face F satisfies $F = V$ in Theorem 4.8. It is known that outerplanar graph metrics embed into distributions over dominating trees [46], but this is not sufficient for our purposes; these maps are not star-shaped and do not satisfy the gradient conditions. Instead our proof is inspired by the result of Charikar and Sahai [20] stating that every outerplanar graph metric can be embedded into the product of two trees with $O(1)$ distortion. In particular, each of these two embeddings must be $O(1)$-Lipschitz, so one hopes that the star-shaped and gradient properties might be achievable with their techniques.

Indeed, by following their basic induction and using a heavily modified variant of their embedding, we are able to obtain the desired result. Unfortunately, for this purpose we are not able to obtain a product of two trees; instead we need an entire distribution, but this suffices in light of Theorem 4.7.

Second step: Retracting onto a face. The second step follows the approach of [36] for proving that face metrics (i.e. those metrics arising from taking the shortest-path metric on a planar graph restricted to a face) embed into distributions over dominating trees; this result was originally proved in [73] via a different method. In [36], the authors randomly retract a planar graph $G = (V, E)$ onto a prescribed face $F \subseteq V$ in such a way that edges are not stretched too much in expectation. (See Figure 4.3)



(a) Partition the graph such that each partition has a vertex on the restricted face.

(b) retract the vertices in each partition to the vertex onto vertex from the restricted face.

Figure 4.3: Retracting onto a restricted face.

Their embedding has the rather convenient property (not shared by previous random retractions) that stars are mapped to stars, satisfying our star-shaped ambitions. Thus we are left to wrestle with the ℓ_∞ gradient issue. By using stronger properties of known random partitioning schemes for planar graphs [54]—specifically the fact that such partitions are "padded" in the language of [45, 57]—we are able to show that all single-scale ℓ_∞ gradients are $O(1)$ in expectation under the random retraction. We remark that this mapping does *not* preserve global ℓ_∞ gradients in expectation, and this is the main reason we have introduced the single-scale definition. This pushes some non-trivial work to the rounding theorem in Section 4.4 which must now show that all the scales can be rounded simultaneously.

4.3.4 Additional Definitions and Notations

Here we review some additional definitions before diving into the proofs. We deal exclusively with finite graphs $G = (V, E)$ which are free of loops and parallel edges. Note that $d_G(x, y) = 0$ may occur even when $x \neq y$, and if G is disconnected, there will be pairs $x, y \in V$ with $d_G(x, y) = \infty$. We allow both possibilities throughout this chapter. An important point is that *all length functions in this chapter are assumed to be reduced,* i.e. they satisfy the property that for every $e = (u, v) \in E$, $\mathsf{len}(e) = d_G(u, v)$. For $v \in V$ and $R \geq 0$, we write $B_G(v, R) = \{u \in V : d_G(u, v) \leq R\}$.

In the present chapter, paths in graphs are always simple, i.e., no vertex appears twice. Given a metric graph G, we extend the length function to paths $P \subseteq E$ by setting $\mathsf{len}(P) = \sum_{e \in P} \mathsf{len}(e)$. We recall that for a subset $S \subseteq V$, $G[S]$ represents the induced graph on S. For a pair of subsets $S, T \subseteq V$, we use the notations $E(S, T) = \{(u, v) \in E : u \in S, v \in T\}$ and $E(S) = E(S, S)$, and if $v \in V$, we write $E(v) = E(\{v\}, V \setminus \{v\})$.

Given a set X, a *random map* $F : X \to Y$ is shorthand for some probability space (Ω, μ) and a distribution over mappings $\{F_\omega : X \to Y_\omega\}_{\omega \in \Omega}$. Note that both F and Y are random variables. In all our constructions, X and Y_ω are finite sets. When no confusion arises, probabilistic expressions containing F and Y should be understood as been taken over the probability space (Ω, μ). When we refer to a property of Y or F, it should be understood that this property holds *for all* Y_ω and $F_\omega : X \to Y_\omega$, $\omega \in \Omega$.

4.4 Polymatroid Networks and Embeddings

Our primary goal in the present section is to prove Theorem 4.7 which shows that random tree embeddings can be used to bound flow/cut gaps in polymatroid networks. We start in Section 4.4.1 by showing that a fixed "thin" mapping into a tree can be use for rounding. In Section 4.4.2, we prove the crucial property that every star-shaped mapping into a tree can be converted to a random thin map. Finally in Section 4.4.3, we combine these results with a multi-scale analysis to show that a suitable distribution over star-shaped mappings into random trees suffices for rounding.

4.4.1 Thin-star Tree Rounding

Consider a graph G, a connected tree T, and a map $f : V(G) \to V(T)$. For every pair $u, v \in V$, let P_{uv} denote the unique simple path from $f(u)$ to $f(v)$ in T. We say that f is Δ-*thin* if, for every $u \in V(G)$, the induced graph on $\bigcup_{v:\{u,v\}\in E(G)} P_{uv}$ can be covered by Δ simple paths in T emanating from $f(u)$. The next lemma gives a generalization of line-embedding rounding [42, 29] to arbitrary thin maps into trees.

Lemma 4.9. *Let $G = (V, E)$ be a graph, T a connected metric tree, and let $f : V \to V(T)$ be a Δ-thin map. Suppose that the set of functions $\{\ell_v : E(v) \to [0, \infty)\}_{v\in V}$ is such that $d_T(f(u), f(v)) \le \ell_u(e) + \ell_v(e)$ for every edge $e = \{u, v\} \in E$.*

Then for any polymatroid capacities $\vec{\rho} = \{\rho_v\}_{v\in V}$ and demands $\mathsf{dem} : V \times V \to [0, \infty)$, there exists a subset of edges $S \subseteq E$ such that

$$\Phi_G(S; \vec{\rho}, \mathsf{dem}) \le \frac{\Delta \sum_{v\in V} \hat{\rho}_v(\ell_v)}{\sum_{u,v\in V} \mathsf{dem}(u, v) \cdot d_T(f(u), f(v))}\,.$$

Proof. For every edge $\{u, v\} \in E$, let P_{uv} denote the unique simple path between $f(u)$ and $f(v)$ in T. For every $a \in E(T)$, we define the subset $S(a) \subseteq E$ by

$$S(a) = \Big\{\{u, v\} \in E : a \in E(P_{uv})\Big\}\,.$$

Observe that if $a \in E(P_{xy})$ for some $x, y \in V$, then $\sigma_{S(a)}(x, y) = 1$. Thus we have, for any $x, y \in V$,

$$\sum_{a\in E(T)} \mathsf{len}_T(a) \cdot \sigma_{S(a)}(x, y) \ge d_T(f(x), f(y))\,. \tag{4.5}$$

Next, we give an upper bound on $\nu_{\vec{\rho}}(S(a))$ for every $a \in E(T)$. First, arbitrarily orient the edges of $E(T)$. Fix $a = (x, y) \in E(T)$ according to this orientation. Consider any $\lambda \in [0, \mathsf{len}_T(a)]$. For an edge $e \in S(a)$, choose the orientation $e = (u, v)$ such that P_{uv} traverses a in the order (x, y). We will assign the edge e to the vertex u if

$$d_T(f(u), x) + \lambda \le \ell_u(e)\,, \tag{4.6}$$

and otherwise assign e to the vertex v. This gives, for every $\lambda \in [0, \mathsf{len}_T(a)]$, a valid assignment $g_{a,\lambda} : S(a) \to V$. Integrating yields

$$\mathsf{len}_T(a) \cdot \nu_{\vec{\rho}}(S(a)) \leq \int_0^{\mathsf{len}_T(a)} \left(\sum_{v \in V} \rho_v(g_{a,\lambda}^{-1}(v)) \right) d\lambda \,. \tag{4.7}$$

Our next goal is to show that, for every $v \in V$, we have

$$\sum_{a \in E(T)} \int_0^{\mathsf{len}_T(a)} \rho_v(g_{a,\lambda}^{-1}(v)) \, d\lambda \leq \Delta \hat{\rho}_v(\ell_v) \,. \tag{4.8}$$

To this end, fix $v \in V$. Since f is Δ-thin, there are $k \leq \Delta$ paths $P_1, P_2, \ldots, P_k$ in T emanating from $f(v) \in V(T)$ such that the following holds: If $S(a)$ contains an edge with endpoint v, then $a \in E(P_i)$ for some $i \in \{1, 2, \ldots, k\}$. Thus we can write

$$\sum_{a \in E(T)} \int_0^{\mathsf{len}_T(a)} \rho_v(g_{a,\lambda}^{-1}(v)) \, d\lambda \leq \sum_{i=1}^{k} \sum_{a \in E(P_i)} \int_0^{\mathsf{len}_T(a)} \rho_v(g_{a,\lambda}^{-1}(v)) \, d\lambda \,, \tag{4.9}$$

and it suffices to bound each term of the latter sum separately.

To this end, fix $i \in \{1, 2, \ldots, k\}$. For $\theta \in [0, \mathsf{len}(P_i)]$, let

$$S_v(\theta) = \Big\{ \{u, v\} \in E : f(u) \in V(P_i) \text{ and } \ell_v(\{u, v\}) \geq \theta \Big\} \,.$$

By the assignment rule (4.6), the fact that $d_T(f(u), f(v)) \leq \ell_u(\{u, v\}) + \ell_v(\{u, v\})$ for every $\{u, v\} \in E$, and monotonicity of ρ_v, we have

$$\begin{aligned} \sum_{a \in E(P_i)} \int_0^{\mathsf{len}_T(a)} \rho_v(g_{a,\lambda}^{-1}(v)) \, d\lambda &\leq \int_0^{\infty} \rho_v(S_v(\theta)) \, d\theta \\ &\leq \int_0^{\infty} \rho_v(\ell_v^{\theta}) \, d\theta \\ &= \hat{\rho}_v(\ell_v) \,, \end{aligned}$$

where in the final line we have used the definition of the Lovász extension $\hat{\rho}_v$ and the notation: $\ell_v^{\theta}(\{u, v\}) = 1$ if $\ell_v(\{u, v\}) \geq \theta$ and $\ell_v^{\theta}(\{u, v\}) = 0$ otherwise. Combining this with (4.9) yields (4.8).

Now interchanging sums and integrals in (4.8) and summing (4.7) over $a \in E(T)$ yields

$$\sum_{a \in E(T)} \mathsf{len}_T(a) \cdot \nu_{\vec{\rho}}(S(a)) \leq \Delta \sum_{v \in V} \hat{\rho}_v(\ell_v) \,.$$

Using this in conjunction with (4.5), we have

$$\begin{aligned}\min_{S\subseteq E}\Phi_G(S;\vec{\rho},\mathsf{dem}) &\le \min_{a\in E(T)}\frac{\nu_{\vec{\rho}}(S(a))}{\sum_{u,v\in V}\mathsf{dem}(u,v)\sigma_{S(a)}(u,v)}\\ &\le \frac{\sum_{a\in E(T)}\mathsf{len}_T(a)\cdot\nu_{\vec{\rho}}(S(a))}{\sum_{a\in E(T)}\mathsf{len}_T(a)\sum_{u,v\in V}\mathsf{dem}(u,v)\sigma_{S(a)}(u,v)}\\ &\le \frac{\Delta\sum_{v\in V}\hat{\rho}_v(\ell_v)}{\sum_{u,v\in V}\mathsf{dem}(u,v)d_T(f(u),f(v))},\end{aligned}$$

completing the proof. □

4.4.2 Random Thinning

Next we show how an arbitrary star-shaped map into a tree can be converted into a random 4-thin map.

Lemma 4.10. *Let $G=(V,E)$ be a graph, T a connected metric tree, and let $f:V\to V(T)$ be a 1-Lipschitz star-shaped map. Then there exists a random connected metric tree T' and a random 4-thin map $F:V\to V(T')$ satisfying the following conditions:*

i) F is 1-Lipschitz with probability one.

ii) For every $u,v\in V$, we have

$$\mathbb{E}\,d_{T'}(F(u),F(v))\ge\frac{1}{2}\,d_T(f(u),f(v))\,.$$

Proof. In what follows, for a tree T, we will use the notation P^T_{xy} to denote the unique simple path between $x,y\in V(T)$.

We will proceed by constructing a random metric tree T' and map $\Phi:V(T)\to V(T')$ and then showing that $F=\Phi\circ f$ satisfies the conclusion of the lemma. Suppose that the tree T is rooted at some fixed vertex. For a vertex $x\in V(T)$, let T_x be the subtree rooted at x. A *vertical path* in a rooted tree is a path P in which every $u,v\in P$ have ancestor-descendant relationship.

Claim 4.11. *For every $x\in V(T)$, there exists a random metric tree T'_x, and a random 1-Lipschitz map $\Phi_x:V(T_x)\to V(T'_x)$ which satisfies the following conditions:*

i) Φ_x *maps every vertical path in* T_x *isometrically to a vertical path in* T'_x*.*

ii) For all $v \in V(T_x)$*,* $\Phi_x|_{V(T_v)} = \Phi_v$*.*

iii) For all $v \in V(T_x)$*, the set of vertices* $\{\Phi_x(u) : u \in V(T_v), E(f^{-1}(u), f^{-1}(v)) \neq \emptyset\}$ *can be covered by at most* two *vertical paths emanating from* $\Phi_x(v)$ *in* T'_x*.*

iv) For every $u, v \in V(T_x)$*, we have*

$$\mathbb{E}\left[d_{T'}(\Phi_x(u), \Phi_x(v))\right] \geq \frac{1}{2}\, d_T(u, v)\,.$$

Proof. We construct the map Φ_x by induction on the height of x in T. When x is a leaf, the statement is partically vacuous. The inductive step is carried in two steps: In the first step we construct a tree $\tilde{T}_x$ and a map $\tilde{\Phi} : V(T_x) \to V(\tilde{T}_x)$ as follows: Let $u_1, \ldots, u_m$, be the children of x in T, and let $T'_1, T'_2, \ldots, T'_m$ be the random trees, and $\Phi_1, \Phi_2, \ldots, \Phi_m$ be the randoms maps resulting from applying the claim inductively to each T_{u_i}. We construct the graph $\tilde{T}_x$ by replacing each T_{u_i} with T'_i in T_x. We put $\tilde{\Phi}(x)$ as the root of $\tilde{T}_x$, and for $v \in V(T_{u_i})$, we put $\tilde{\Phi}(v) = \Phi_i(v)$. We also define $\tilde{f} = \tilde{\Phi} \circ f$.

Let $S = \left\{P^{\tilde{T}_x}_{\tilde{\Phi}(x)\tilde{v}} : \tilde{v} \in \tilde{T}_x, E(\tilde{f}^{-1}(x), \tilde{f}^{-1}(\tilde{v})) \neq \emptyset\right\}$, and let H be the subgraph of $\tilde{T}_x$ induced by S. The map f is a star-shaped, and each map Φ_i, maps root leaf paths in T_{u_i} to root leaf paths in T'_i, therefore the set S is star-shaped. Hence, there exists $k \leq m$ edge disjoint paths $P_1, P_2, \ldots, P_k$ in $\tilde{T}_x$ emanating from $\tilde{\Phi}(x)$ that cover S.

Let $\tilde{T}_1, \ldots, \tilde{T}_n$ be the connected components of $\tilde{T}_x$ after removing the edges of H, and let $\tilde{v}_j \in S$ be the root of the tree $\tilde{T}_j$.

We define the random tree T'_x, and the random map $\Phi_x : T_x \to T'_x$ as follows. Consider a root r connected to two paths B_1 and B_2. We define $\Phi_x(x) = r$. Moreover, we map each path $P_1, P_2, \ldots, P_k$ isometrically and independently at random to one of the two paths B_1, and B_2 (see Figure 4.4 for an example). Then we complete the construction by gluing root of each tree $\tilde{T}_j$ to $\Phi_x(\tilde{v}_j)$ (the image of $\tilde{v}_j$ in either B_1 or B_2).

It is straightforward to check that Φ satisfies Claim 4.11(i). Using the inductive hypothesis it is sufficient to check Claim 4.11(ii) with respect to the children of x in T, and for them the claim is easily seen to be true.

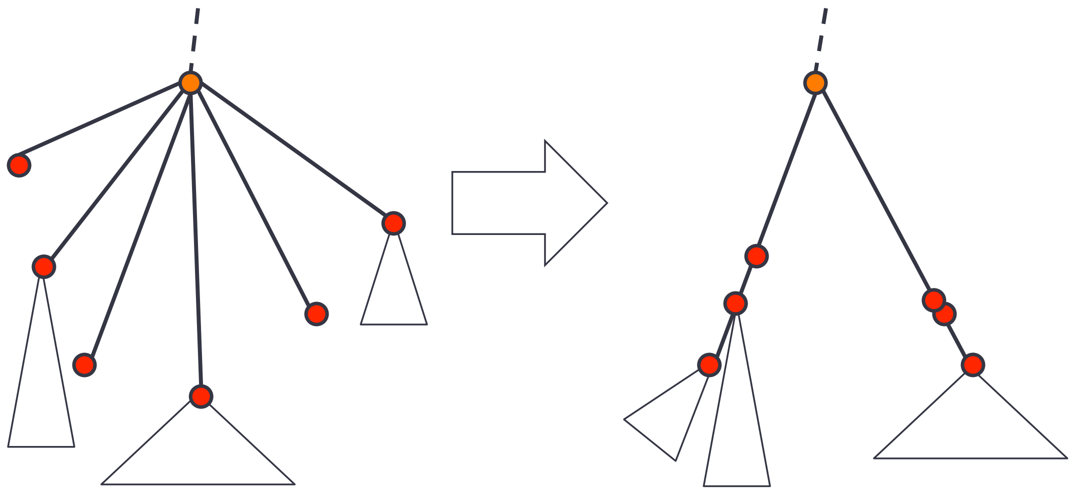


Figure 4.4: Merging branches of the tree.

To verify Claim 4.11(iii), first note that for $v \in T_x \setminus \{x\}$ this condition holds by our inductive construction, and Claim 4.11(ii). Moreover, For the case that $v = x$, all the vertices $u \in V(T_x)$ such that $E(f^{-1}(x), f^{-1}(u)) \neq \emptyset$ are mapped to paths B_1 and B_2, therefore Claim 4.11(iii) holds for all $v \in V(T_x)$.

To verify Claim 4.11(iv), first note that for $u, v \in V(T_x)$, if $u, v \in V(T_{u_i})$ for some i, then $d_{T'_i}(\Phi_{u_i}(u), \Phi_{u_i}(v)) = d_{T'_x}(\Phi_x(u), \Phi_x(v))$ and by our inductive construction

$$\frac{1}{2} d_T(u, v) \leq d_{T'_x}(\Phi_x(u), \Phi_x(v)) .$$

Moreover if u and v do not belong to the same subtree rooted at one of x's children, then with probability 1/2, $d_T(u, v) = d_{T'_x}(\Phi_x(u), \Phi_x(v))$. Therefore Claim 4.11(iv) holds for all $u, v \in V(T_x)$, completing the proof of the claim. □

The map f is star-shaped, therefore there must exist k edge disjoint paths $P_1, P_2, \ldots, P_k$ emanating from v that cover $\{f(u) : (u, v) \in E(G)\}$ in T. Moreover, since these paths are edge disjoint, at most one of these paths is not contained in $T_{f(v)}$. Without loss of generality assume that $P_1, P_2, \ldots, P_{k-1}$ are contained in $T_{f(v)}$. Let T' and Φ to be the tree and the map resulting from applying claim to the root of T. By our construction we can cover the image of the paths $P_1, P_2, \ldots, P_{k-1}$ by at most two paths emanating from $\Phi(f(v))$ in T'.

Finally, every path in a rooted tree is a union of at most two vertical paths and each vertical path in T is mapped to a vertical path in T', the image of the path P_k can be covered by at most two paths emanating from $\Phi(f(v))$ in T', completing the proof of Lemma 4.10. $\square$

The next result follows from Lemma 4.10 and Lemma 4.9.

Corollary 4.12. *Let $G = (V, E)$ be a graph, T a connected metric tree, and $f : V \to V(T)$ a star-shaped mapping. Suppose that the set of functions $\{\ell_v : E(v) \to [0, \infty)\}_{v \in V}$ is such that $d_T(f(u), f(v)) \leq \ell_u(e) + \ell_v(e)$ for every edge $e = \{u, v\} \in E$.*

Then for any polymatroid capacities $\vec{\rho} = \{\rho_v\}_{v \in V}$ and demands $\mathsf{dem} : V \times V \to [0, \infty)$, there exists a subset of edges $S \subseteq E$ such that

$$\Phi_G(S; \vec{\rho}, \mathsf{dem}) \leq \frac{8 \sum_{v \in V} \hat{\rho}_v(\ell_v)}{\sum_{u,v \in V} \mathsf{dem}(u, v) \cdot d_T(f(u), f(v))} .$$

4.4.3 Rounding Random Star-shaped Embeddings

Finally, we are ready prove the main result of this section connecting embeddings to polymatroid flow/cut gaps. We restate Theorem 4.7 here for the sake of the reader

Theorem 4.13. *Let $G = (V, E, \mathsf{len})$ be a metric graph and suppose there exists a random connected metric tree T and a random star-shaped mapping $F : V \to V(T)$ such that for some $K \geq 1$,*

$$\max_{v \in V} \sup_{\tau > 0} \mathbb{E}\, |\nabla_\tau F(v)|_\infty \leq K .$$

Then for any set of functions $\{\ell_v : E(v) \to [0, \infty)\}_{v \in V}$ that is adapted to len, and for any polymatroid capacities $\vec{\rho} = \{\rho_v\}_{v \in V}$ and demands $\mathsf{dem} : V \times V \to [0, \infty)$, there exists a subset of edges $S \subseteq E$ such that

$$\Phi_{\vec{\rho}, \mathsf{dem}}(S) \leq \frac{64K \sum_{v \in V} \hat{\rho}_v(\ell_v)}{\sum_{u,v \in V} \mathsf{dem}(u, v) \cdot \mathbb{E}\left[d_T(F(u), F(v))\right]} .$$

Proof. Using the fact that $\hat{\rho}_v$ is monotone, we may first scale $\{\ell_v(e) : v \in V\,, e \in E(v)\}$ down and assume that for $\{u, v\} \in E$, we have $\mathsf{len}(\{u, v\}) = \ell_u(\{u, v\}) + \ell_v(\{u, v\})$. Next, by rounding all the length functions up, we may assume that $\{\ell_v(e) : v \in V\,, e \in E(v)\}$ are dyadic:

$$\{\ell_v(e) : v \in V, e \in E(v)\} \subseteq \{2^h : h \in \mathbb{Z}\}, \tag{4.10}$$

and that for $\{u,v\} \in E$, we have

$$\mathsf{len}(\{u,v\}) \geq \frac{1}{2}\left(\ell_u(\{u,v\}) + \ell_v(\{u,v\})\right). \tag{4.11}$$

Now define the random functions $\{\tilde{\ell}_v : E(v) \to [0,\infty)\}_{v\in V}$ by

$$\tilde{\ell}_v(\{u,v\}) = \begin{cases} 0 & \text{if } \ell_v(\{u,v\}) < \ell_u(\{u,v\}) \\ 2\ell_v(\{u,v\}) \cdot \frac{d_T(F(u),F(v))}{\mathsf{len}(u,v)} & \text{otherwise.} \end{cases}$$

Then, by definition, we have $d_T(F(u),F(v)) \leq \tilde{\ell}_u(\{u,v\}) + \tilde{\ell}_v(\{u,v\})$ for every $\{u,v\} \in E$ since $\{\ell_v\}$ is adapted to len.

We define a new family $\{\hat{\ell}_v\}$ by $\hat{\ell}_v(e) = \sup\{\tilde{\ell}_v(e') : \ell_v(e') \leq \ell_v(e)\}$. Observe that $\hat{\ell}_v \geq \tilde{\ell}_v$ pointwise, thus by monotonicity, $\hat{\rho}_v(\hat{\ell}_v) \geq \hat{\rho}_v(\tilde{\ell}_v)$. Additionally, we have $\ell_v(e) \leq \ell_v(e')$ if and only if $\hat{\ell}_v(e) \leq \hat{\ell}_v(e')$. Thus the collections of edge sets $\{\ell_v^\theta : \theta \in [0,\infty)\}$ and $\{\hat{\ell}_v^\theta : \theta \in [0,\infty)\}$ are identical.

Enumerate the set of values $\{\ell_v(e) : e \in E(v)\} \cup \{0\}$ by $0 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_k$ so that

$$\hat{\rho}_v(\ell_v) = \sum_{i=0}^{k-1} (\tau_{i+1} - \tau_i)\rho_v(\ell_v^{\tau_i}) \geq \frac{1}{2}\sum_{i=0}^{k-1} \tau_{i+1}\rho_v(\ell_v^{\tau_i}), \tag{4.12}$$

where the latter inequality holds since $\tau_{i+1} \geq 2\tau_i$ by (4.10).

For $i = 1,2,\ldots,k$, we can likewise set $\hat{\tau}_i = \max\{\hat{\ell}_v(e) : \ell_v(e) = \tau_i\}$. By construction, we have $0 = \hat{\tau}_0 \leq \hat{\tau}_1 \leq \hat{\tau}_2 \leq \cdots \leq \hat{\tau}_k$, and

$$\hat{\rho}_v(\hat{\ell}_v) = \sum_{i=0}^{k-1} (\hat{\tau}_{i+1} - \hat{\tau}_i)\rho_v(\ell_v^{\tau_i}).$$

Finally, define $\tilde{\tau}_i = \max\{\tilde{\ell}_v(e) : \ell_v(e) = \tau_i\}$. Observe that if $\tilde{\tau}_{i+1} \neq \hat{\tau}_{i+1}$, then $\hat{\tau}_{i+1} = \hat{\tau}_i$, thus we can write

$$\hat{\rho}_v(\hat{\ell}_v) \leq \sum_{i=0}^{k-1} \tilde{\tau}_{i+1}\rho_v(\ell_v^{\tau_i}). \tag{4.13}$$

Using the definition of $\tilde{\ell}_v$, we have

$$\tilde{\tau}_i = \max\left(\{0\} \cup \left\{\tilde{\ell}_v(e) : \ell_v(e) = \tau_i \text{ and } \ell_v(e) \geq \frac{1}{2}\mathsf{len}(e)\right\}\right).$$

Furthermore, by (4.11), if $\ell_v(e) = \tau_i$, then $\mathsf{len}(e) \geq \frac{1}{2}\tau_i$. Thus,

$$\mathbb{E}[\tilde{\tau}_i] \leq \left(\mathbb{E}|\nabla_{\tau_i/2}F(v)|_\infty + \mathbb{E}|\nabla_{\tau_i}F(v)|_\infty\right)\tau_i \leq 2K\tau_i.$$

Using (4.13) and (4.12), this implies

$$\mathbb{E}[\hat{\rho}_v(\hat{\ell}_v)] \le \sum_{i=0}^{k-1} \mathbb{E}[\tilde{\tau}_{i+1}]\rho_v(\ell_v^{\tau_i}) \le 2K \sum_{i=0}^{k-1} \tau_{i+1}\rho_v(\ell_v^{\tau_i}) \le 4K\hat{\rho}_v(\ell_v)\,.$$

Applying Corollary 4.12 completes the proof. □

4.5 Star-shaped Embeddings of Outerplanar Graphs into Trees

Our goal is now to prove that every metric outerplanar graph admits a random Lipschitz, star-shaped embedding into a random tree.

Theorem 4.14. *There is a constant $K \ge 1$ such that the following holds. Let $G = (V, E)$ be a metric outerplanar graph. Then there is a random metric tree T and a random 1-Lipschitz, star-shaped mapping $F : V \to V(T)$ such that for every $u, v \in V$, $\mathbb{E}[d_T(F(u), F(v))] \ge d_G(u, v)/K$.*

We begin by setting up the notations and definitions needed to prove Theorem 4.14.

4.5.1 More Notations and Definitions

For a graph $G = (V, E)$, and $v \in V$, we use the notation $N_G(v) = \{u : (u, v) \in E\}$ to denote the set neighbors of v in the graph G. For a path P, we define the cycle $C(P, \ell)$ as the cycle obtained by connecting the endpoints of P with an edge of length ℓ. The length of the cycle C, is given by $\mathsf{len}(C) = \mathsf{len}(P) + \ell$. In this section, it is helpful to think of cycles as continuous cycles and $V(P) \subseteq C$ as points on the cycle.

For a cycle C and a point p on the cycle we define $\mathsf{flat}(C, p)$ to be the path where p is one end point and $x \in C$ is mapped to the point at distance $d_C(p, x)$ from p on the path. Moreover for points $x, y \in C$ we use $d_{\mathsf{flat}(C,p)}(x, y) = |d_C(x, p) - d_C(y, p)|$ to denote the distance between x and y on the path $\mathsf{flat}(C, p)$. See Figure 4.5 for an example.

For two paths $P = (u_1, \ldots, u_m)$ and $Q = (v_1, \ldots, v_n)$ with the same length, we define the *glue*ing of P and Q as follows. We first identifying the end points u_1 with v_1, and u_m with v_n to specify the end points of the resulting path. Then we map each point $x \in V(P) \cup V(Q)$ so that the distance between x and the end points of the path is preserved.

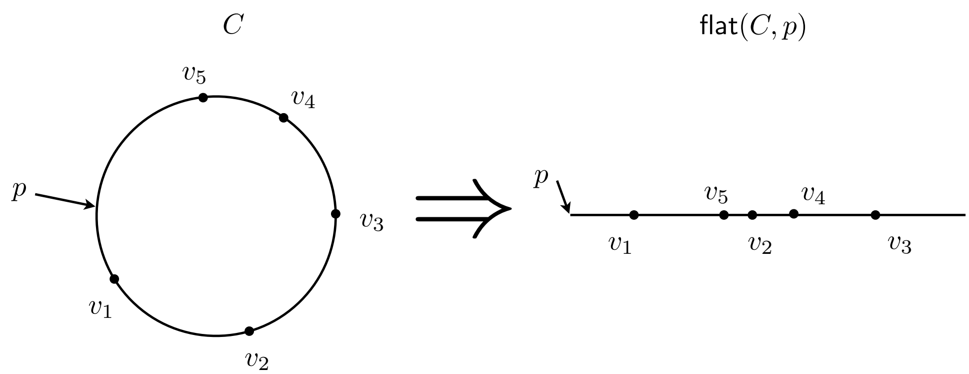


Figure 4.5: The flattening of the cycle C.

Finally, for two given trees, T_1 and T_2, and pairs of vertices $u_1, v_1 \in V(T_1)$ and $u_2, v_2 \in V(T_2)$ such that $d_{T_1}(u_1, v_1) = d_{T_2}(u_2, v_2)$, we define $\mathsf{glue}(T_1, T_2; u_1, v_1; u_2, v_2)$ as the tree resulting from gluing the trees T_1 and T_2 on the unique path between u_1 and v_1 in T_1 and u_2 and v_2 in T_2. See Figure 4.6.

4.5.2 Framework

Our approach to Theorem 4.14 employs the framework of Charikar and Sahai (see Theorem 4 in [20]). Any outerplanar graph can be constructed by considering a sequence of paths P_i, and then doing the following: Start with $G_1 = P_1$. At step i, we consider some edge $e_i = (u_i, v_i)$ on the outer face of G_i, and obtain G_{i+1} by either attaching the endpoints of P_i to u_i and v_i, or by attaching only one endpoint of P_i to either u_i or v_i.

In this section we only consider biconnected outerplanar graphs (so the endpoints of P_i are always attached to u_i and v_i), since we can simply take the embedding of biconnected components of a graph that are connected by a single vertex into trees, and glue the trees on the image of the common vertex to obtain an embedding for the whole graph.

We also use the concept of a slack structure [46]. We say that an outerplanar graph has an α-slack structure if it can be built out of paths P_i such that the length of any path P_i which attaches to both endpoints of an edge e_i is at least α times the length of e_i. The following lemma is a straightforward generalization of a fact from [46], where it is proved

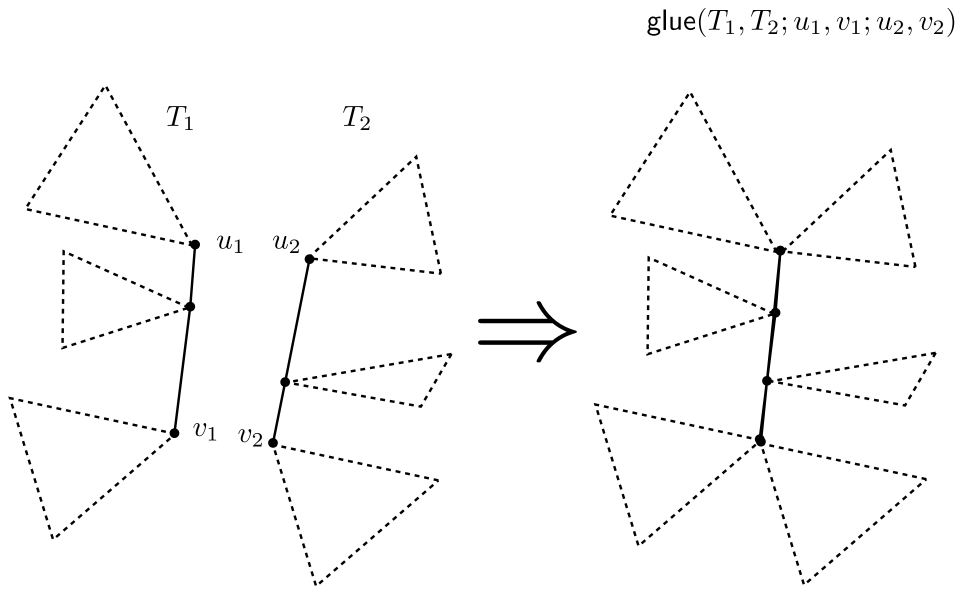


Figure 4.6: Gluing of the trees T_1 and T_2.

for $\alpha = 2$.

Lemma 4.15 ([46]). *Consider any $\alpha \geq 1$. Given an outerplanar metric graph $G = (V, E, \mathsf{len}_G)$, there is an outerplanar metric graph $H = (V, E', \mathsf{len}_H)$ with $E' \subseteq E$, and such that H has an α-slack structure. Furthermore, $d_G \geq d_H \geq (1/\alpha)d_G$, and for every $(u, v) \in E'$,*

$$d_H(u, v) = \mathsf{len}_H(u, v). \tag{4.14}$$

Thus, by incurring distortion at most α, we may assume that the outerplanar graph G has an α-slack structure. We will build our embedding inductively based on the sequence of the paths $P_1 \ldots, P_m$ provided by Lemma 4.15.

Random extension. Given an embedding of a metric graph G into a random metric tree T, $F : V(G) \to V(T)$ and a new path P attached to the points $u, v \in V(G)$, we extend the embedding of G to an embedding for $\hat{G} = G \cup P$ into a random tree $\hat{T}$, using the following operation. Let $C = C(P, d_T(F(u), F(v)))$. To extend the tree T, we choose two *anchor* points $p, q \in C$, and map the vertices of C onto two paths $L = \mathsf{flat}(C, p)$ and $R = \mathsf{flat}(C, q)$.

We put $\hat{T} = \mathsf{glue}(T, L; F(u), F(v); u, v)$ with probability $1/2$ and $\hat{T} = \mathsf{glue}(T, R; F(u), F(v); u, v)$ with probability $1/2$. This specifies a random mapping $\hat{F} : V(G) \to V(\hat{T})$. Since it will be clear from context which vertices we are gluing onto, we will use the notations $\mathsf{glue}(T, L)$ and $\mathsf{glue}(T, R)$ without specifying the vertices. Note that the gluing can be done if and only if $d_L(u, v) = d_R(u, v) = d_C(u, v) = d_T(F(u), F(v))$. Moreover, If the map $F : V(G) \to V(T)$ is 1-Lipschitz, then so is the extension $\hat{F}$.

A significant difference between our construction and that of [20] is in the way we choose the anchor points. For our purposes, it is not enough to simply look at the α-slack graph; we need to use the structure of the original graph when we choose the anchor points in order to maintain the star-shaped property. The algorithm of [20] is able to construct an embedding using only two trees, while we embed the graph into a distribution over trees. In the next section, we prove a distortion bound for this embedding based on the distance between the anchor points in the cycle.

4.5.3 Distortion Bound

Before we can state the main lemma of this section, we need the following definition. For a cycle C, and points $u, v \in C$ we say that a pair of points $p, q \in C$ is (α, β)-*apart with respect to another pair* $\{u, v\}$ if $d_C(p, q) = \alpha \mathsf{len}(C)$ and for $a \in \{u, v\}$ and $b \in \{p, q\}$:

$$\beta \,\mathsf{len}(C) \leq d_C(a, b) \leq \left(\frac{1}{2} - \beta\right) \mathsf{len}(C).$$

We now state a lemma capturing our main inductive step.

Lemma 4.16. *Let G be a graph, T be random metric tree, and let $F : V(G) \to V(T)$ be a random* 1*-Lipschitz map such that $\mathbb{E}[d_T(F(x), F(y)] \geq d_G(x, y)/6$ for every $x, y \in V(G)$. Let $\hat{G}$ be a graph constructed by attaching a path P with*

$$\mathsf{len}(P) \geq 160 \cdot d_G(u, v) \tag{4.15}$$

onto a pair of vertices $u, v \in V(G)$. Let $C = C(P, d_T(F(u), F(v)))$, and $p, q \in C$ be any pair of points that are $(1/6, 1/16)$-apart with respect to $\{u, v\}$ in C, and let $\hat{T}$ be the random extension of T by C with respect to the anchor points p and q. The embedding $\hat{F} : V(\hat{G}) \to V(\hat{T})$ is also 1-Lipschitz and such that for all $x, y \in V(\hat{G})$, $\mathbb{E}[d_{\hat{T}}(\hat{F}(x), \hat{F}(y)] \geq d_{\hat{G}}(x, y)/6$.

We will use mainly the following two properties of (α, β)-apart pairs in the proof of Lemma 4.16.

Observation 4.17. *For any $\beta \in [0, 1/2]$, the following holds. Suppose C is a cycle and $a, b \in C$ are such that*

$$\beta \,\mathsf{len}(C) \le d_C(a, b) \le \left(\frac{1}{2} - \beta\right) \mathsf{len}(C).$$

Then for any $x, y \in C$ with $\max\{d_C(x, a), d_C(y, a)\} \le \beta \mathsf{len}(C)$, *we have*

$$d_{\mathsf{flat}(C,b)}(x, y) = d_C(x, y).$$

Lemma 4.18. *Let C be a cycle. For $\alpha \in [0, 1/4]$ and $p, q \in C$ such that $d_C(p, q) = \alpha \mathsf{len}(C)$, the following holds. For any pair of vertices $x, y \in C$,*

$$d_{\mathsf{flat}(C,p)}(x, y) + d_{\mathsf{flat}(C,q)}(x, y) \ge 4\alpha d_C(x, y). \tag{4.16}$$

Proof. We divide the analysis into two cases. If neither p or its antipodal point $\bar{p}$ (the point at distance $\mathsf{len}(C)/2$ from p on the cycle) lie on a shortest path between x and y then a simple application of Observation 4.17 implies that

$$d_{\mathsf{flat}(C,p)}(x, y) + d_{\mathsf{flat}(C,q)}(x, y) \ge d_{\mathsf{flat}(C,p)}(x, y) = d_C(x, y) \ge 4\alpha d_C(x, y),$$

and an analogous inequality holds if neither q or it antipodal point $\bar{q}$ lie on a shortest path between x and y.

Next, suppose that $p' \in \{p, \bar{p}\}$ and $q' \in \{q, \bar{q}\}$ lie on the same shortest path between x and y. It is easy to check that $\mathsf{flat}(C, p)$ is isometric to $\mathsf{flat}(C, p')$ and $\mathsf{flat}(C, q)$ is isometric to $\mathsf{flat}(C, q')$. We have $d_C(p', q') \in \{\alpha \mathsf{len}(C), (\frac{1}{2} - \alpha)\mathsf{len}(C)\}$ and $\alpha \in [0, 1/4]$, therefore $d_C(p', q') \ge \alpha \mathsf{len}(C)$. Since both p' and q' are on the same shortest path between x and y we can write,

$$\begin{aligned}
d_{\mathsf{flat}(C,p)}(x, y) + d_{\mathsf{flat}(C,q)}(x, y) &= d_{\mathsf{flat}(C,p')}(x, y) + d_{\mathsf{flat}(C,q')}(x, y) \\
&= |d_C(p', x) - d_C(p', y)| + |d_C(q', y) - d_C(q', x)| \\
&\ge |d_C(p', x) - d_C(q', x) + d_C(q', y) - d_C(p', y)| \\
&= |d_C(p', x) - d_C(q', x)| + |d_C(q', y) - d_C(p', y)| \\
&= 2d_C(p', q') \ge 2\alpha \mathsf{len}(C) \ge 4\alpha d_C(x, y). \square
\end{aligned}$$

Proof of Lemma 4.16. Since F is 1-Lipschitz and the random extension preserves the 1-Lipschitz condition, $\hat{F}$ is also 1-Lipschitz. We divide the analysis of the expected contraction of the pairs $x, y \in V(\hat{G})$ into three cases.

Case I. $x, y \in V(G)$**:** In this case,

$$\mathbb{E}[d_{\hat{T}}(\hat{F}(x), \hat{F}(y))] = \mathbb{E}[d_T(F(x), F(y))] \geq \frac{1}{6} d_G(x,y) = \frac{1}{6} d_{\hat{G}}(x,y).$$

Case II. $x \in V(P)$ **and** $y \in V(G)$**:** Observe that a shortest path in $\hat{G}$ connecting x to y must pass through either u or v. Suppose, without loss of generality that $d_{\hat{G}}(x,y) = d_P(x,u) + d_G(u,y)$. Let $w \in V(T)$ be the closest vertex to $F(y)$ (with respect to d_T) from the unique path connecting $F(u)$ and $F(v)$ in T. In this case by (4.15), we have

$$d_T(F(u), w) \leq d_T(F(u), F(v)) \leq d_G(u,v) \leq \mathsf{len}(C)/160.$$

Suppose first that $d_{\hat{G}}(u,x) \leq \mathsf{len}(C)/16$. Since p and q are $(1/6, 1/16)$-apart, Observation 4.17 implies that

$$d_{\hat{T}}(\hat{F}(x), \hat{F}(y)) = d_C(x,u) + d_T(F(u), w) + d_T(w, F(y)) = d_{\hat{G}}(x,u) + d_T(F(u), F(y)). \quad (4.17)$$

Suppose next that $d_{\hat{G}}(u,x) > \mathsf{len}(C)/16$. Denote by $w' \in C$ the point with $d_C(u,w') = d_T(F(u), w)$. Lemma 4.18 implies

$$\begin{aligned}
\mathbb{E}[d_{\hat{T}}(\hat{F}(x), \hat{F}(y)) \mid F] &= \mathbb{E}[d_{\hat{T}}(\hat{F}(x), w) \mid F] + d_T(w, F(y)) \\
&\overset{(4.16)}{\geq} \frac{1}{3} d_C(x, w') + d_T(w, F(y)) \\
&\geq \frac{1}{3}(d_C(u,x) - d_C(u,w')) + (d_T(F(u), F(y)) - d_T(F(u), w)) \\
&\geq \frac{1}{3}(d_C(u,x) - d_C(u,v)) + (d_T(F(u), F(y)) - d_T(F(u), F(v))).
\end{aligned}$$

Since $d_{\hat{G}}(u,x) \leq d_C(u,x) + d_{\hat{G}}(u,v) - d_C(u,v)$, we have

$$\begin{aligned}
\mathbb{E}[d_{\hat{T}}(\hat{F}(x), \hat{F}(y)) \mid F] &\geq \frac{1}{3}(d_{\hat{G}}(u,x) - d_{\hat{G}}(u,v)) + (d_T(F(u), F(y)) - d_{\hat{G}}(u,v)) \\
&\geq \frac{1}{3} d_{\hat{G}}(u,x) - \frac{4}{3} d_{\hat{G}}(u,v) + d_T(F(u), F(y)) \\
&\overset{(4.15)}{\geq} \frac{1}{3} d_{\hat{G}}(u,x) - \frac{16 \cdot 4}{3 \cdot 160} d_{\hat{G}}(u,x) + d_T(F(u), F(y)) \\
&\geq \frac{1}{6} d_{\hat{G}}(u,x) + d_T(F(u), F(y)). \qquad (4.18)
\end{aligned}$$

Putting (4.17) and (4.18) together, we can conclude that

$$\mathbb{E}\left[d_{\hat{T}}(\hat{F}(x),\hat{F}(y))\right] \geq \frac{1}{6}d_{\hat{G}}(u,x)+\mathbb{E}\left[d_T(F(u),F(y))\right] \geq \frac{1}{6}(d_{\hat{G}}(x,u)+d_{\hat{G}}(u,y)) = \frac{1}{6}d_{\hat{G}}(x,y).$$

Case III. $x, y \in V(P)$**:** In this case we divide the problem into three cases again. If no shortest path between x and y in the graph $\hat{G}$ goes through the edge (u,v), then

$$d_C(x,y) \geq \min\{d_{\hat{G}}(x,y), \mathsf{len}(P) - d_{\hat{G}}(x,y)\}.$$

Moreover, (4.15) implies that $d_{\hat{G}}(x,y) \leq \frac{1}{2}(\mathsf{len}(P) + d_{\hat{G}}(u,v)) \leq \frac{81}{160}\mathsf{len}(P)$. Thus,

$$d_C(x,y) \geq \min\Big\{d_{\hat{G}}(x,y), \frac{160}{81}d_{\hat{G}}(x,y) - d_{\hat{G}}(x,y)\Big\} \geq \frac{79}{81}d_{\hat{G}}(x,y) \geq \frac{1}{2}d_{\hat{G}}(x,y).$$

Hence, using Lemma 4.18 we can conclude that

$$\mathbb{E}[d_{\hat{T}}(\hat{F}(x),\hat{F}(y)) \mid F] \geq \frac{1}{3}d_C(x,y) \geq \frac{1}{6}d_{\hat{G}}(x,y).$$

If a shortest path between x and y in the graph $\hat{G}$ passes through the edge (u,v) and $d_{\hat{G}}(x,y) > \mathsf{len}(P)/16$, then by Lemma 4.18

$$\begin{aligned}\mathbb{E}[d_{\hat{T}}(\hat{F}(x),\hat{F}(y)) \mid F] &\geq \frac{1}{3}d_C(x,y) \geq \frac{1}{3}(d_{\hat{G}}(x,y) - d_{\hat{G}}(u,v))\\ &\overset{(4.15)}{>} \frac{1}{3}(1-16/160)d_{\hat{G}}(x,y) > \frac{1}{6}d_{\hat{G}}(x,y).\end{aligned}$$

Finally, we consider the case where a shortest path between x and y in the graph $\hat{G}$ passes through the edge (u,v) and $d_{\hat{G}}(x,y) \leq \mathsf{len}(P)/16$. Suppose that $d_{\hat{G}}(u,x) \leq d_{\hat{G}}(v,x)$. It is easy to check that we also have $d_C(u,x) \leq d_C(v,x)$, and that the shortest path between x and y in C also passes through the edge (u,v). Hence, by Observation 4.17,

$$d_{\hat{T}}(\hat{F}(x),\hat{F}(y)) = d_C(x,y) = d_{\hat{G}}(x,u) + d_T(F(u),F(v)) + d_{\hat{G}}(v,y).$$

Thus,

$$\begin{aligned}\mathbb{E}\left[d_{\hat{T}}(\hat{F}(x),\hat{F}(y))\right] &\geq \mathbb{E}\left[d_{\hat{G}}(x,u) + d_T(F(u),F(v)) + d_{\hat{G}}(v,y)\right]\\ &= d_{\hat{G}}(x,u) + d_{\hat{G}}(y,v) + \mathbb{E}\left[d_T(F(u),F(v))\right]\\ &\geq d_{\hat{G}}(x,u) + d_{\hat{G}}(y,v) + \frac{1}{6}d_{\hat{G}}(u,v)\\ &\geq \frac{1}{6}d_{\hat{G}}(x,y),\end{aligned}$$

completing the proof. $\square$

4.5.4 The Star-shaped Property

To prove Theorem 4.14, we need to choose the anchor points such that the resulting map is star-shaped. The following lemma and its corollary provide the main tools necessary to achieve this.

Lemma 4.19. *For any $\alpha \in (0, 1/6)$, $\beta \leq \frac{1}{4} - \frac{3}{2}\alpha$, and $\delta < \alpha$ the following holds. Let P be a path with endpoints u, v, and let $C = C(P, \delta \mathsf{len}(P))$. For any finite subset $S \subseteq C$, there are points $p, p' \in C$ that are $(\frac{1}{4} - \frac{3}{2}\alpha - \beta, \beta)$-apart with respect to u and v, and moreover for $q \in \{p, p'\}$,*

i) $d_C(q, u) \leq d_C(q, v)$;

ii) For $x \in V(P)$ with $d_P(x, u) \leq \left(\frac{1}{2} + \alpha\right) \mathsf{len}(P)$, we have $d_C(q, x) \leq d_C(q, u)$;

iii) For any two distinct elements $x, y \in S$, we have $d_C(q, x) \neq d_C(q, y)$.

Proof. For $\eta \in (\delta, \alpha)$, it is easy to verify that the two points $p, p' \in C$ with

$$\begin{aligned} d_C(u, p) &= \left(\frac{1}{4} + \frac{3\alpha}{2} - \eta\right) \mathsf{len}(C) \\ d_C(u, p') &= \left(\frac{1}{2} - \eta - \beta\right) \mathsf{len}(C) \end{aligned}$$

satisfy conditions (a) and (b). Condition (c) also holds for almost all $\eta \in (\delta, \alpha)$. □

Corollary 4.20. *For $\delta \leq 1/160$, the following holds. Let P be a path with endpoints u, v, and let $C = C(P, \delta \mathsf{len}(P))$. For any finite subset $S \subseteq C$, there are points $p, p' \in C$ that are $(1/6, 1/16)$-apart with respect to u and v, and moreover for $q \in \{p, p'\}$,*

i) $d_C(q, u) \leq d_C(q, v)$;

ii) For $x \in V(P)$ with $d_P(x, u) \leq \left(\frac{1}{2} + \frac{1}{160}\right) \mathsf{len}(P)$, we have $d_C(q, x) \leq d_C(q, u)$;

iii) For any two distinct elements $x, y \in S$, we have $d_C(q, x) \neq d_C(q, y)$.

The next lemma is the final ingredient that we need to prove Theorem 4.14.

Lemma 4.21. *Let $G = (V, E)$ be a biconnected outerplanar graph with outer face C, and let d_C be the path pseudometric in the induced graph $G[C]$. Consider any point $p \in C$ and edge (u, v) on the outer face, and suppose that $d_C(p, u) < d_C(p, v)$. Then one of the following two conditions hold:*

- $\forall w \in N_G(v) : d_C(p, w) \geq d_C(p, u)$;
- $\forall w \in N_G(u) : d_C(p, w) \leq d_C(p, v)$.

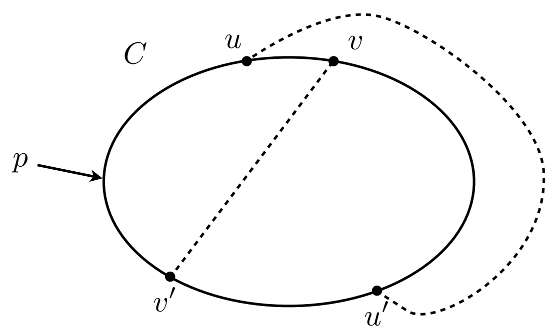


Figure 4.7: Positions of u, u', v, and v' on the cycle.

Proof. For the sake of contradiction suppose that none of the conditions of the lemma hold. Let $v' \in N_G(v)$, be such that $d_C(p, v') < d_C(p, u)$ and let $u' \in N_G(u)$ be such that $d_C(p, u') > d_C(p, v)$ (See Figure 4.7.) It is easy to check that the union of the cycle C with the edges (u, u') and (v, v') contains K_4 as minor, therefore G cannot be outerplanar, and we have arrived at a contradiction. □

We can now prove Theorem 4.14.

Proof of Theorem 4.14. Without loss of generality, we may assume that H is 2-connected, as it is trivial to construct the desired embedding for H from embeddings for each of its 2-connected components. To construct the embedding, we first use Lemma 4.15 to transform the graph into an 160-slack graph H, with a decomposition into paths $P_1, \ldots P_m$. Then, we use the random extension algorithm from Section 4.5.2 to inductively build an embedding from H into random trees. To avoid ambiguity, in what follows for $x, y \in V(G)$ and and a mapping $F : V(G) \to V(T)$, we will use the notation $P^T_{F(x)F(y)}$ to denote the unique simple path between $F(x)$ and $F(y)$ in T.

We start with the graph $H_1 = P_1$ and at step $i \geq 2$ we construct the graph H_i by attaching the path P_i to the edge $(u_i, v_i) \in E(H_{i-1})$. (Note that since H is 2-connected, the endpoints of P_i are always attached to both u_i and v_i.)

We construct the embedding for H_i from an embedding $F_{i-1} : V(H_{i-1}) \to V(T_{i-1})$ as follows. We use random extension with anchor points which are $(1/6, 1/16)$-apart with respect to u_i and v_i to extend T_{i-1} and F_{i-1} to T_i and $F_i : V(H_i) \to V(T_i)$. Moreover, we choose the anchor points so that the resulting map is injective and it maintains the following additional property:

or all $i \geq 1$ and any edge (x, y) on the outer face of H_i, there is at least one endpoint x such that for all $w \in N_G(x) \cap V(H_i)$,

$$\left(P^{T_i}_{F_i(x)F_i(w)} \subseteq P^{T_i}_{F_i(x)F_i(y)}\right) \vee \left(P^{T_i}_{F_i(x)F_i(w)} \cap P^{T_i}_{F_i(x)F_i(y)} = \{F_i(x)\}\right). \qquad (4.19)$$

We call a vertex that satisfies the above property a *good* vertex for the edge (x, y) with respect to $V(H_i)$ and F_i.

If the edge (u_i, v_i), we choose the anchor points on the cycle $C = C(P_i, d_T(u_i, v_i))$ such that they satisfy conditions (a) and (b) of Corollary 4.20, with v being a good vertex of the edge (u_i, v_i) with respect to T, and $\delta = d_T(u_i, v_i)/\mathsf{len}(P_i)$. See Figure 4.8 for an example. Note that Corollary 4.20(c) implies that we can always find such anchor points while maintaining the invariant that the map is injective.

Since we choose the anchor points to be $(1/6, 1/16)$-apart and H is 160-slack, by Lemma 4.16 this embedding is 1-Lipschitz and has constant distortion. Thus we only need to prove the following statements to complete the proof:

i) Each edge on the outer face of H_i has a good vertex with respect to $V(H_i)$ and F_i.

ii) This construction produce a star-shaped embedding.

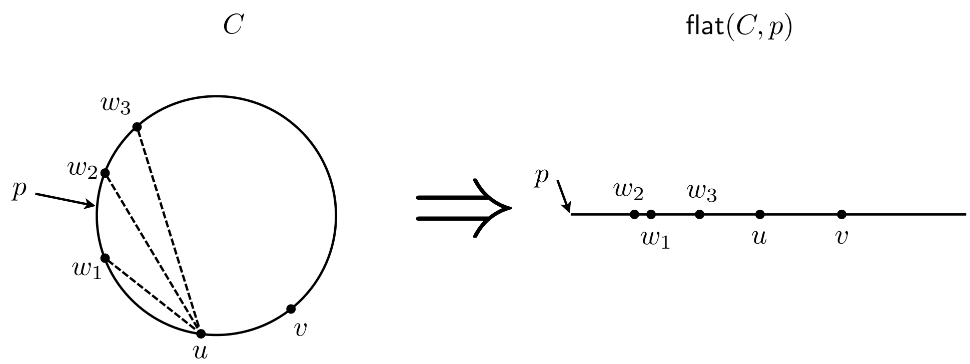


Figure 4.8: If u is not a good vertex for the edge (u, v), then conditions (a) and (b) of Corollary 4.20 imply that u lies between v and $w \in (N_G(u) \cap P_i) \setminus \{v\}$ on the path $\mathsf{flat}(C, p)$.

Proof of (i). Fix a mapping $F_{i-1} : V(H_{i-1}) \to V(T_{i-1})$ and let $F_i : V(H_i) \to V(T_i)$ be the random extension of F_{i-1}, where the anchor points are chosen such that they satisfy the conditions (a) and (b) of Corollary 4.20. We prove this claim inductively. Suppose that all edges on the outer face of H_{i-1} have a good vertex with respect to $V(H_{i-1})$ and F_{i-1}; we show that all edges on the outer face of H_i have a good vertex with respect to $V(H_i)$ and F_i. For each edge (x, y) on the outer face of H_i, we divide the analysis into two main cases.

Case I: $(x, y) \in E(H_{i-1})$**.** Suppose that x is a a good vertex for the edge (x, y) with respect to $V(H_{i-1})$ and F_{i-1}. If $x \notin \{u_i, v_i\}$ or $P^{T_{i-1}}_{F_{i-1}(x)F_{i-1}(y)} \cap P^{T_{i-1}}_{F_{i-1}(u_i)F_{i-1}(v_i)} = \{F_{i-1}(x)\}$ then the inductive hypothesis easily implies that x is also a good vertex for the edge (x, y) with respect to $V(H_i)$ and F_i.

If, on the other hand, $x \in \{u_i, v_i\}$ and $P^{T_{i-1}}_{F_{i-1}(x)F_{i-1}(y)} \cap P^{T_{i-1}}_{F_{i-1}(u_i)F_{i-1}(v_i)} \neq \{F_{i-1}(x)\}$ then, by (4.19) applied to the edge (x, y), $P^{T_{i-1}}_{F_{i-1}(u_i)F_{i-1}(v_i)}$ must be a subpath of $P^{T_{i-1}}_{F_{i-1}(x)F_{i-1}(y)}$. Therefore, by (4.19) applied to the edge (u_i, v_i), the vertex in $\{u_i, v_i\} \setminus \{x\}$ cannot be a good vertex for the edge (u_i, v_i) with respect to $V(H_{i-1})$ and F_{i-1}. Hence, the inductive hypothesis implies that x is also a good vertex for (u_i, v_i) with respect to $V(H_{i-1})$ and F_{i-1}. In this case for all $w \in N_G(x) \cap V(P_i)$ by (4.14), $d_{P_i}(x, w) = d_{H_i}(x, w) \leq (\mathsf{len}(P_i) + d_{H_i}(u_i, v_i))/2$. Hence conditions (a) and (b) of Corallary 4.20 imply that

$$P^{T_i}_{F_i(x)F_i(y)} \cap P^{T_i}_{F_i(w)F_i(x)} \subseteq T_{i-1} \cap P^{T_i}_{F_i(w)F_i(x)} = P^{T_i}_{F_i(u_i)F_i(v_i)} \cap P^{T_i}_{F_i(w)F_i(x)} = \{F_i(x)\}.$$

And, in the case that $w \in N_G(v) \cap V(H_{i-1})$ since x was a good vertex of (x, y) with respect to $V(H_{i-1})$ and F_{i-1}, then x and w also satisfy (4.19) with respect to $V(H_i)$ and F_i.

Case II: $(x, y) \in E(P_i)$**.** Lemma 4.21 implies that each edge (x, y) on the outer face of the subgraph induced by G on $V(P_i)$ (which necessarily contains $E(P_i)$) has a good vertex with respect to $V(P_i)$ and F_i. For an edge $(x, y) \in E(P_i)$ we prove the following statement. If x is a good vertex for the edge (x, y) with respect to $V(P_i)$ and F_i then x is also a good vertex for the edge (x, y) with respect to $V(H_i)$ and F_i. Let $w \in N_G(x) \cap V(H_i)$. If $w \in V(P_i)$ then the assumption that x is good for (x, y) with respect to $V(P_i)$ implies that it satisfies (4.19). Therefore, we only need to verify (4.19) for $w \in N_G(x) \cap V(H_{i-1})$.

If $x \notin \{u_i, v_i\}$ or $P^{T_i}_{F_i(x)F_i(y)} \cap P^{T_i}_{F_i(u_i)F_i(v_i)} = \{F_i(x)\}$ then for all $w \in N_G(x) \cap V(H_{i-1})$, we have $P^{T_i}_{F_i(x)F_i(y)} \cap P^{T_i}_{F_i(x)F_i(w)} \subseteq \{F_i(x)\}$ and (4.19) holds.

Now, we consider the case that $x = v_i$ and $P^{T_i}_{F_i(x)F_i(y)} \cap P^{T_i}_{F_i(u_i)F_i(v_i)} \neq \{F_i(x)\}$, where v_i is a good vertex for the edge (u_i, v_i) with respect to $V(H_{i-1})$ and F_{i-1}. In this case since x is a good vertex for the edge (x, y) with respect to $V(P_i)$ and F_i, then $P^{T_i}_{F_i(u_i)F_i(v_i)}$ must be a subpath of $P^{T_i}_{F_i(x)F_i(y)}$. Moreover, (4.19) for the edge (u_i, v_i) with respect to $V(H_{i-1})$ and F_{i-1}, implies that for $w \in N_G(v_i) \cap V(H_{i-1})$, either $P^{T_i}_{F_i(x)F_i(w)} \subseteq P^{T_i}_{F_i(u_i)F_i(v_i)}$ in which case

$$P^{T_i}_{F_i(x)F_i(w)} \subseteq P^{T_i}_{F_i(u_i)F_i(v_i)} \subseteq P^{T_i}_{F_i(x)F_i(y)},$$

or $P^{T_i}_{F_i(x)F_i(w)} \cap P^{T_i}_{F_i(u_i)F(v_i)} = \{F_i(x)\}$ in which case

$$\begin{aligned} P^{T_i}_{F_i(x)F_i(w)} \cap P^{T_i}_{F_i(x)F_i(y)} &= (P^{T_i}_{F_i(x)F_i(w)} \cap P^{T_i}_{F_i(u_i)F_i(v_i)}) \cup (P^{T_i}_{F_i(x)F_i(w)} \cap (P^{T_i}_{F_i(x)F_i(y)} \setminus P^{T_i}_{F_i(u_i)F_i(v_i)})) \\ &\subseteq \{F_i(x)\} \cup (P^{T_i}_{F_i(x)F_i(w)} \cap (P^{T_i}_{F_i(x)F_i(y)} \setminus P^{T_i}_{F_i(u_i)F_i(v_i)})), \end{aligned}$$

but the left hand side is a path whereas the right hand side is a disjoint union of a point and a path, hence $P^{T_i}_{F_i(x)F_i(w)} \cap P^{T_i}_{F_i(x)F_i(y)} = \{F_i(x)\}$.

The final case is where $x = u_i$. In this case by (4.14), $d_{P_i}(x, y) = d_{H_I}(x, y) \leq (\mathsf{len}(P) + d_G(u, v))/2$. Hence by conditions (a) and (b) of Corollary 4.20, $P^{T_i}_{F_i(x)F_i(y)} \cap P^{T_i}_{F_i(u_i)F_i(v_i)} = \{F_i(u_i)\}$. This case was already dealt with above.

Proof of (ii). Again, let $F_{i-1} : V(H_{i-1}) \to V(T_{i-1})$ a random mapping and let $F_i : V(H_i) \to V(T_i)$ be the random extension of F_{i-1}. Note that since F_i is injective, the map is star-shaped if and only if for all $x \in V(H_i)$,

$$\{P^{T_i}_{F_i(x)F_i(y)} : y \in N_G(x) \cap V(H_i)\}$$

is star-shaped.

We now prove this claim by induction. Suppose that the map F_{i-1} is star shaped; we will show that the map F_i is also star shaped. For all vertices $x \in V(H_{i-1}) \setminus \{u_i, v_i\}$, by the induction hypothesis,

$$\{P^{T_{i-1}}_{F_{i-1}(x)F_{i-1}(y)} : y \in N_G(x) \cap V(H_i)\} = \{P^{T_{i-1}}_{F_{i-1}(x)F_{i-1}(y)} : y \in N_G(x) \cap V(H_{i-1})\}$$

is star-shaped, and this set remains star-shaped after extension to F_i.

For $x \in P_i \setminus \{u_i, v_i\}$, all neighbors of x are mapped to a path in T_i, hence

$$\{P^{T_i}_{F_i(x)F_i(y)} : y \in N_G(x) \cap V(H_i)\}$$

is also star-shaped.

Next, suppose that $x = v_i$ is a good vertex for the edge (u_i, v_i) with respect to $V(H_{i-1})$ and F_{i-1}. By the induction hypothesis, $\{P^{T_{i-1}}_{F_{i-1}(v_i)F_{i-1}(y)} : y \in N_G(x) \cap V(H_{i-1})\}$ is star shaped. Furthermore, since v_i is a good vertex for (u_i, v_i) with respect to $V(H_{i-1})$ and F_{i-1}, each $y \in N_G(x) \cap V(H_{i-1})$ is either on the path $P^{T_{i-1}}_{F_{i-1}(u_i)F_{i-1}(v_i)}$, or $P^{T_{i-1}}_{F_{i-1}(u_i)F_{i-1}(v_i)} \cap P^{T_{i-1}}_{F_{i-1}(v_i)F_{i-1}(y)} = \{F_{i-1}(v_i)\}$. Therefore adding the paths $P^{T_i}_{F_i(v_i)F_i(y)}$, for $y \in N_G(v_i) \cap V(P_i)$, the set $\{P^{T_i}_{F_i(v_i)F_i(y)} : y \in N_G(x) \cap V(H_i)\}$ remains star-shaped.

Finally if $x = u_i$, then for $w \in N_G(x) \cap V(P_i)$ by (4.14), $d_{P_i}(x, w) = d_{H_i}(x, w) \leq (\mathsf{len}(P_i) + d_G(u_i, v_i))/2$. Hence combining condition (a) and (b) of Corollary 4.20 we can conclude that

$$V(T_{i-1}) \cap P^{T_i}_{F_i(w)F_i(x)} = P^{T_i}_{F_i(u_i)F_i(v_i)} \cap P^{T_i}_{F_i(w)F_i(x)} = \{F_i(x)\}.$$

In other words, all G-neighbors of x in P_i are mapped to a new branch in T_i that intersects T_{i-1} only at $F_i(x)$. Thus $\{P^{T_i}_{F_i(x)F_i(y)} : y \in N_G(x) \cap V(H_i)\}$ is also star-shaped. □

4.6 Connected Random Retractions

Our goal now is to complete the proof of Theorem 4.8 by showing that every planar graph can be randomly retracted onto a specified face in such a way that the face can itself be endowed with an outerplanar metric. Combining this with our embedding of outerplanar graphs into random trees from Section 4.5, we will be able to prove Theorem 4.8; this is done in Section 4.6.3.

In the next section, we review the notion of "padded partitions" of metric spaces. The existence of such partitions for planar graphs (due to [54]) will be one of our two central ingredients here. The other ingredient is the method of [36] for the construction of random *connected* retractions. They work with a weaker notion of random partitions, so their results (as stated in [36]) are not strong enough for us. In Section 4.6.2, we follow their proof closely but use padded partitions, allowing us to obtain the stronger conclusion we require.

4.6.1 Padded Partitions of Graphs

Random partitions are a powerful tool in the theory of embeddings of finite metric spaces; see, e.g., [9, 91, 57, 70]. A particularly powerful notion is that of a "padded" partition. We review the relevant definitions in the special setting of finite metric spaces.

Consider a metric space (X, d). We will sometimes think of a partition P of X as a map $P : X \to 2^X$ sending each $x \in X$ to the unique set in P containing it. We say that P is *τ-bounded* if $\mathrm{diam}(S) \leq \tau$ for every $S \in P$. We say that a random partition $\mathcal{P}$ is τ-bounded if this holds almost surely. A random partition $\mathcal{P}$ is *(α, τ)-padded* if it is τ-bounded and, additionally, for every $x \in X$ and $R \geq 0$, we have

$$\mathbb{P}[B_X(x, R) \nsubseteq \mathcal{P}(x)] \leq \alpha \cdot \frac{R}{\tau},$$

where $B_X(x, R) = \{y \in X : d(x, y) \leq R\}$.

The main random partitioning result we require is from [54], though it first appeared in this form later (see [91, 57, 70]).

Theorem 4.22 ([54])**.** *There exists a constant $\alpha > 0$ such that if $G = (V, E)$ is a metric planar graph, then for every $\tau > 0$, (V, d_G) admits an (α, τ)-padded random partition. Furthermore, the distribution of the partition can be sampled from in polynomial-time in the size of G.*

4.6.2 Random Retractions

We now use random partitions to construct random retractions. This was first done in [18] in the context of the 0-extension problem on graphs. Further work includes [70], which concerns the Lipschitz extension problem, and [73], where the authors are primarily concerned with randomly simplifying the topology of metric graphs. The proof of the next theorem follows from the techniques of [36] for constructing a connected retraction. We are able to obtain a stronger conclusion by using a stronger assumption about the random partitions.

Theorem 4.23. *Let $G = (V, E)$ be a metric graph and suppose that for some $\alpha \geq 2$ and every $\tau \geq 0$, (V, d_G) admits an (α, τ)-padded random partition. Then for any subset $S \subseteq V$, there exists a random mapping $F : V \to S$ such that the following properties hold.*

i) For every $x \in S$, $F(x) = x$.

ii) For every $x \in V$ and $\tau > 0$, $\mathbb{E}\,|\nabla_\tau F(x)|_\infty \leq O(\alpha \log \alpha)$.

iii) For every $x \in S$, the set $F^{-1}(x)$ is a connected subset of G.

Proof. Since we are dealing with finite graphs, without loss of generality, we may assume that $d_G(S, V \setminus S) > 1$. Let $k_0 = \lceil \log_2 \operatorname{diam}_G(V) \rceil$. First, we let $\{P_k : 1 \leq k < k_0\}$ be a sequence of independent $(\alpha, 2^k)$-padded random partitions. We define $P_{k_0} = \{V\}$. Following [36], we inductively define a sequence of random maps $\{F_k : V_k \to S : 1 \leq k \leq k_0\}$, where $V_k \subseteq V$ and $(F_{k+1})|_{V_k} = F_k$ for each $k \in \mathbb{N}$. We will define $F = F_{k_0}$.

First, we put $V_0 = S$ and $F_0(x) = x$ for $x \in S$. Now suppose that V_{k-1} and F_{k-1} are defined for some $k \geq 1$. We will use the notation $\hat{P}_k(x)$ for the connected component of $G[P_k(x)]$ containing x. We let

$$V_k = \{x \in V : \hat{P}_k(x) \cap V_{k-1} \neq \emptyset\}$$

be the set of vertices which are connected to V_{k-1} through their set in P_k.

For every $T \in P_k$ and every non-empty connected component C of $G[T \cap (V_k \setminus V_{k-1})]$, let v_C be a neighbor of C in $V_{k-1} \cap P_k(x)$, which must exist by the definition of V_k. For

$x \in C$, we define $F_k(x) = F_{k-1}(v_C)$. For all other $x \in V_k$, we must have $x \in V_{k-1}$, and we put $F_k(x) = F_{k-1}(x)$. By definition, we have $V_{k_0} = V$ and we define $F = F_{k_0}$.

By construction, property (i) holds. Property (iii) follows easily by induction: It is true for every $x \in S$ and $0 \le k \le k_0$ that $F_k^{-1}(x)$ is connected. We are thus left to verify property (ii). For every vertex $x \in V$, define $L(x) = \min\{k : x \in V_k\}$. We make the following claim.

Claim 4.24. *For every $x \in V$, $d_G(x, S) \le d_G(x, F(x)) < 2^{L(x)+1}$*

Proof. The first inequality is immediate since $F(x) \in S$. The second follows by induction: For any $0 \le k \le k_0$, we claim that if $x \in V_k$, then $d_G(x, F_k(x)) < 2^{k+1}$. This is clear for $L(x) = 0$ since $V_0 = S$ and $F_0(x) = x$ for all $x \in S$. If $L(x) = k > 0$, then $x \in V_k \setminus V_{k-1}$, hence $F_k(x) = F_{k-1}(y)$ for some $y \in P_k(x) \cap V_{k-1}$. Thus,

$$d_G(x, F_k(x)) \le \operatorname{diam}(P_k(x)) + d_G(y, F_{k-1}(y)) .$$

Since P_k is 2^k-bounded, we have $\operatorname{diam}(P_k(x)) \le 2^k$ and by induction, $d_G(y, F_{k-1}(y)) < 2^k$. It follows that $d_G(x, F_k(x)) < 2^{k+1}$, completing the proof. □

Now fix $x \in V$ and $\tau > 0$ and let $B = B_G(x, 2\tau)$. We will employ the bound

$$|\nabla_\tau F(x)|_\infty \le \frac{\operatorname{diam}_G(F(B))}{\tau} . \tag{4.20}$$

Let $\mathrm{cov}_B = \max\{L(y) : y \in B\}$ and $\mathrm{hit}_B = \min\{L(y) : y \in B\}$. Using the triangle inequality and Claim 4.24, we have

$$\operatorname{diam}_G(F(B)) \le 2^{\mathrm{cov}_B + 2} + 4\tau .$$

Let $\mathcal{E}_k$ be the event that $B \not\subseteq P_k(x)$. Observe that

$$\neg\mathcal{E}_{\mathrm{hit}_B} \implies |F(B)| = 1 \implies \operatorname{diam}_G(F(B)) = 0,$$

because in this case, every vertex in B is in the same connected component of x in $G[P_k(x) \setminus V_{k-1}]$, where $k = \mathrm{hit}_B$ (they are all connected through x). Hence,

$$\operatorname{diam}_G(F(B)) \le 4\tau + \mathbf{1}_{\{\mathcal{E}_{\mathrm{hit}_B}\}} \cdot 2^{\mathrm{cov}_B + 2} . \tag{4.21}$$

Let $\Delta = \min\{d_G(y,S) : y \in B\}$, and $m = \lceil \log_2 \max(1,\Delta) \rceil$. By Claim 4.24, we have $\mathrm{hit}_B \geq m-1$, hence

$$\mathbb{E}\left[\mathbf{1}_{\{\mathcal{E}_{\mathrm{hit}_B}\}} \cdot 2^{\mathrm{cov}_B+2}\right] = \sum_{k \geq m-1} \mathbb{P}(k = \mathrm{hit}_B, \mathcal{E}_k) \cdot \mathbb{E}\left[2^{\mathrm{cov}_B+2} \,\middle|\, \mathcal{E}_k, \mathrm{hit}_B = k\right] \quad (4.22)$$

Now, observe that, for $k \geq \mathrm{hit}_B$,

$$\neg \mathcal{E}_k \implies \mathrm{cov}_B \leq k,$$

because, in this case, $P_k(x)$ must contain a vertex $v \in V_{\mathrm{hit}_B}$ and a shortest-path from v to x, implying that $B_G(x, 2\tau) \subseteq V_k$. The padding property implies that

$$\mathbb{P}\left(\mathcal{E}_j\right) \leq \frac{2\alpha\tau}{2^j}\,.$$

In particular, for any $k \geq \max(\mathrm{hit}_B, \log_2 2\alpha\tau)$, we have

$$\mathbb{P}(\mathrm{cov}_B \geq k+j \mid \mathcal{E}_k, \mathrm{hit}_B = k) \leq \prod_{i=k+1}^{k+j-1} \mathbb{P}\left(\mathcal{E}_i\right) \leq \prod_{i=k+1}^{k+j-1} 2^{k-i} \leq 2^{-(2j-3)}\,. \quad (4.23)$$

Using this yields

$$\mathbb{E}\left[2^{\mathrm{cov}_B+2} \,\middle|\, \mathcal{E}_k, \mathrm{hit}_B = k\right] \leq 4 \cdot \max(2^k, 2\alpha\tau) \sum_{j=0}^{\infty} 2^{-(2j-3)} \cdot 2^j \leq 2^{k+6} + 128\alpha\tau\,. \quad (4.24)$$

Combining (4.22) and (4.24), we can write

$$\begin{aligned}
\mathbb{E}\left[\mathbf{1}_{\{\mathcal{E}_{\mathrm{hit}_B}\}} \cdot 2^{\mathrm{cov}_B+2}\right] &\leq \sum_{k \geq m-1} \mathbb{P}(k = \mathrm{hit}_B, \mathcal{E}_k) \cdot (2^{k+6} + 128\alpha\tau) \\
&\leq 128\alpha\tau + \sum_{k \geq m-1} \mathbb{P}(k = \mathrm{hit}_B, \mathcal{E}_k) 2^{k+6} \\
&\leq 128\alpha\tau + \sum_{k \geq m-1} \mathbb{P}(k \leq \mathrm{hit}_B) \mathbb{P}(\mathcal{E}_k) 2^{k+6}, \quad (4.25)
\end{aligned}$$

where the last inequality holds because

$$\mathbb{P}(k = \mathrm{hit}_B, \mathcal{E}_k) \leq \mathbb{P}(k-1 < \mathrm{hit}_B, \mathcal{E}_k) = \mathbb{P}(\mathrm{hit}_B > k-1) \cdot \mathbb{P}(\mathcal{E}_k \mid \mathrm{hit}_B > k-1) = \mathbb{P}(\mathrm{hit}_B > k-1) \cdot \mathbb{P}(\mathcal{E}_k).$$

Let $y \in B$ and $w \in S$ be a pair of points such that $d_G(y,w) = \Delta$. It is easy to check that if for some k, $B_G(w,\Delta) \subseteq P_k(w)$ then $\mathrm{hit}_B \leq k$, thus the properties of padded partitions imply that

$$\mathbb{P}(\mathrm{hit}_B \geq k) \leq \prod_{i<k} \mathbb{P}(B(w,\Delta) \nsubseteq P_i(w)) \leq \prod_{i<k} \min\left(1, \frac{2\alpha 2^m}{2^i}\right) \leq \min\left(1, \frac{2\alpha 2^m}{2^{k-1}}\right)$$

Combining this with (4.25), we can conclude that

$$\begin{aligned}
\mathbb{E}\left[\mathbf{1}_{\{\mathcal{E}_{\mathrm{hit}_B}\}} \cdot 2^{\mathrm{cov}_B+2}\right] &\le 128\alpha\tau + \sum_{k\ge m-1} \min\left(1, \frac{2\alpha 2^m}{2^{k-1}}\right)\left(\frac{2\alpha\tau}{2^k}\right) 2^{k+6} \\
&= 128\alpha\tau + \sum_{k\ge m-1} \min\left(1, \frac{2\alpha 2^m}{2^k}\right)(128\alpha\tau) \\
&= 128\alpha\tau + \sum_{k=m-1}^{m+\lceil\log_2\alpha\rceil} 128\alpha\tau + \sum_{k>m+\lceil\log_2\alpha\rceil} \left(\frac{2\alpha 2^m}{2^k}\right)(128\alpha\tau) \\
&\le O(\alpha\tau) + O\left(\alpha\tau\log\alpha\right) + O(\alpha\tau) \\
&\le O\left((\alpha\log\alpha)\tau\right).
\end{aligned}$$

Combining this with (4.20) and (4.21) completes the verification of property (ii). □

4.6.3 Retracting to an Outerplanar Graph

Finally, we use the random retractions of the preceding section to randomly embed every metric on the face of a planar graph into an outerplanar graph in a suitable way. This technique is also taken from [36], although again we require some stronger properties of the embedding.

Theorem 4.25. *There is a constant $K > 1$ such that for any metric planar graph $G = (V, E)$ and face $V_0 \subseteq V$, there is a random outerplanar metric graph H and a random mapping $F : V \to V(H)$ satisfying the following:*

i) For every edge $\{u, v\} \in E$, either $F(u) = F(v)$ or $\{F(u), F(v)\} \in E(H)$.

ii) For every $u, v \in V_0$, $d_H(F(u), F(v)) \ge d_G(u, v)$.

iii) For every $u \in V$ and $\tau \ge 0$, we have $\mathbb{E}\,|\nabla_\tau F(u)|_\infty \le K$.

Proof. By Theorem 4.22, we can apply Theorem 4.23 to the metric graph G with $S = V_0$. Let $F : V \to V_0$ be the random mapping guaranteed by Theorem 4.23.

We construct a metric graph H with vertex set V_0 and an edge $\{u, v\}$ of length $d_G(u, v)$ whenever there is an edge between the sets $F^{-1}(u)$ and $F^{-1}(v)$ in G. Since the sets $\{F^{-1}(u) : u \in V_0\}$ are connected, the resulting graph H is outerplanar. Also, property (i) is immediate.

Property (ii) follows because F is the identity on V_0 and the edges in H have length equal to the distance between their endpoints in (V_0, d_G). Property (iii) follows from Theorem 4.23(i). □

We can now complete the proof of Theorem 4.8.

Proof of Theorem 4.8. Let $\Lambda_1 : V \to V(H)$ be the random mapping from V onto the vertices of an outerplanar metric graph H guaranteed from Theorem 4.25. Let $\Lambda_2 : V(H) \to V(T)$ be the random mapping of H into trees from Theorem 4.14. The mapping $\Lambda = \Lambda_2 \circ \Lambda_1 : V \to V(T)$ is mapping guaranteed by the theorem. Combining the star-shaped property of Theorem 4.14 with Theorem 4.25(i) implies that Λ is star-shaped. Property (ii) of Theorem 4.8 is a consequence of Theorem 4.25(ii) and Theorem 4.14. Finally, property (i) follows from the Lipschitz condition of Theorem 4.14 and property (iii) of Theorem 4.25. □

Chapter 5

CONCLUSION AND OPEN QUESTIONS

While in this dissertation we answered some of the questions in the area of metric embeddings, there are many outstanding questions that are interesting to pursue. In this section we list some of these questions.

Dimension Reduction. In Chapter 2 we proved a tight upper bound and lower bounds for near isometric embedding of trees into ℓ_1 [1]. While the proofs are relatively simple for k-ary trees, the generalization of embeddings of k-ary trees to arbitrary trees is very technical and complicated. It would be very interesting to find a simpler proof for Theorem 2.1. A simpler proof for this theorem may yield new insights into understanding tree metrics and may allow us to answer the following question.

Question 2. *For $\varepsilon > 0$, is it possible to $(1+\varepsilon)$-embed any tree on n points into $\ell_1^{O\left(\frac{\log n}{\varepsilon^2}\right)}$?*

A more general question that is still open is the question of dimension reduction for other ℓ_p spaces. Schechtman [94] generalizes the techniques from [88] to show that for $0 < p < 2$, any n point metric in ℓ_p, $(1+\varepsilon)$-embeds into $\ell_p^{C_\varepsilon n}$, where $C_\varepsilon = \varepsilon^{-2-\frac{2}{p}}$. However, this bound for ℓ_1 is still far off from the $n^{O(\frac{1}{D^2})}$ lower bound proved by Brinkman and Charikar [16]. The situation for the lower bound is even worse. It is not clear how one can extend the current techniques to prove lower bounds for ℓ_1 to other metric spaces. In particular, the uniform convexity argument in [64] implies that the family of metrics that is used to prove the lower bounds for dimension reduction in ℓ_1 does not even embed with constant distortion into ℓ_p for $p > 1$.

Question 3. *For $p > 0$, and $D > 1$ what is the minimum dimension k such that any n-point subset of ℓ_p can be embedded into ℓ_p^k with distortion at most D?*

---

[1] For embeddings with distortion $1+\varepsilon$, the resulting bound is tight up to a factor depending only on ε

Sparsest Cut. The main question that is still open is whether there exists a polynomial time constant factor approximation algorithm for the Sparsest Cut Problem. It was shown in [3] that there are no polynomial time approximation scheme for solving the sparsest cut problem unless NP-complete problems can be solved in randomized subexponential time. Moreover if the Unique Games Conjecture is true then [22] implies that constant approximation of the value of the sparsest cut is NP-hard. However, in the light of new sub-exponential algorithms for Unique Game Problem, a major question that one can ask is whether one can approximate the value of the sparsest cut within a constant factor in subexponential time. One approach to solve such problems is the use of polynomial hierarchies. Meka [85] recently showed a lower bound on the performance of Sherali-Adams Hierarchy, however there is nothing known about the performance of other hierarchies such as the Lasserre hierarchy; in fact the exact bound is not known even for applying one round of Lasserre. It is worth noting that the added triangle inequality in the Goemans-Linial SDP can be deduced from applying only one round of the Lasserre hierarchy to the natural SDP for the Sparsest Cut problem.

Question 4. *Is there a $\delta > 0$ such that $n^{1-\delta}$ rounds of the Lasserre hierarchy for the natural SDP of the Sparsest Cut Problem has constant integrality gap?*

One may try narrowing the problem to finding a constant factor approximation algorithm for restricted families of graphs such as planar graphs. It was conjectured in [46] that even the basic linear program suggested by Leighton and Rao [75], has a constant integrality gap for excluded minor families of graphs. Rao [91] has shown that the integrality gap for these families of graphs is at most $O(\sqrt{\log n})$, where n is the number of vertices in the graph, however there has not been any significant progress for more than a decade on this problem.

Vertex Separataors and Okamura-Seymour. One of the main motivations behind studying the Okamura-Seymour theorem for vertex capacitated graphs was the approach of Chekuri, Chandra and Shepherd [30] to find edge-disjoint paths in planar graphs. The Okamura-Seymour Theorem was used as a fundamental tool in solving the edge-disjoint paths problem in their approach. In Chapter 4 we have shown that it is possible to generalize the Okamura-Seymour Theorem to the vertex-capacitated setting. The answer to the following question will pave the way to generalize the rest of the algorithms/proofs in [30] to Vertex-disjoint Path Problem.

Question 5. *If all the demands are integeral, is it possible to route a constant fraction of the flows as integral flows while violating the capacity of the edges at most by a constant fraction?*

BIBLIOGRAPHY

[1] Amit Agarwal, Moses Charikar, Konstantin Makarychev, and Yury Makarychev. $O(\sqrt{\log n})$ approximation algorithms for Min UnCut, Min 2CNF deletion, and directed cut problems. In *STOC'05: Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 573–581. ACM, New York, 2005.

[2] Noga Alon. Problems and results in extremal combinatorics. I. *Discrete Math.*, 273(1-3):31–53, 2003. EuroComb'01 (Barcelona).

[3] Christoph Ambühl, Monaldo Mastrolilli, and Ola Svensson. Inapproximability results for maximum edge biclique, minimum linear arrangement, and sparsest cut. *SIAM J. Comput.*, 40(2):567–596, April 2011.

[4] Alexandr Andoni, Moses Charikar, Ofer Neiman, and Huy L. Nguyen. Near linear lower bound for dimension reduction in l1. In *FOCS*, pages 315–323, 2011.

[5] Sanjeev Arora, James R. Lee, and Assaf Naor. Euclidean distortion and the sparsest cut. *J. Amer. Math. Soc.*, 21:1–21, 2008.

[6] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings and graph partitioning. *J. ACM*, 56(2):Art. 5, 37, 2009.

[7] Patrice Assouad. Plongements lipschitziens dans $\mathbf{R}^n$. *Bull. Soc. Math. France*, 111(4):429–448, 1983.

[8] Y. Aumann and Y. Rabani. An $O(\log k)$ approximate min-cut max-flow theorem and approximation algorithm. *SIAM J. Comput.*, 27(1):291–301, 1998.

[9] Yair Bartal. On approximating arbitrary metrices by tree metrics. In *STOC '98 (Dallas, TX)*, pages 161–168. ACM, New York, 1998.

[10] Joshua D. Batson, Daniel A. Spielman, and Nikhil Srivastava. Twice-ramanujan sparsifiers. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing*, pages 255–262, 2009.

[11] Yoav Benyamini and Joram Lindenstrauss. *Geometric nonlinear functional analysis. Vol. 1*, volume 48 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2000.

[12] J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.*, 52(1-2):46–52, 1985.

[13] J. Bourgain. The metrical interpretation of superreflexivity in Banach spaces. *Israel J. Math.*, 56(2):222–230, 1986.

[14] J. Bourgain. On the distribution of the fourier spectrum of boolean functions. *Israel J. Math.*, (131):269–276, 2002.

[15] J. Bourgain, J. Lindenstrauss, and V. Milman. Approximation of zonoids by zonotopes. *Acta Math.*, 162(1-2):73–141, 1989.

[16] Bo Brinkman and Moses Charikar. On the impossibility of dimension reduction in ℓ_1. *J. ACM*, 52(5):766–788, 2005.

[17] Bo Brinkman, Adriana Karagiozova, and James R. Lee. Vertex cuts, random walks, and dimension reduction in series-parallel graphs. In *STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pages 621–630. ACM, New York, 2007.

[18] Gruia Calinescu, Howard Karloff, and Yuval Rabani. Approximation algorithms for the 0-extension problem. In *Proceedings of the 12th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 8–16, Philadelphia, PA, 2001. SIAM.

[19] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Near-optimal algorithms for unique games (extended abstract). In *STOC'06: Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, pages 205–214. ACM, New York, 2006.

[20] Moses Charikar and Amit Sahai. Dimension reduction in the ℓ_1 norm. In *FOCS*, pages 551–560, 2002.

[21] S. Chawla, A. Gupta, and H. Räcke. An improved approximation to sparsest cut. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*, Vancouver, 2005. ACM.

[22] Shuchi Chawla, Robert Krauthgamer, Ravi Kumar, Yuval Rabani, and D. Sivakumar. On the hardness of approximating multicut and sparsest-cut. *Comput. Complexity*, 15(2):94–114, 2006.

[23] J. Cheeger. Differentiability of Lipschitz functions on metric measure spaces. *Geom. Funct. Anal.*, 9(3):428–517, 1999.

[24] J. Cheeger and B. Kleiner. On the differentiation of Lipschitz maps from metric measure spaces to Banach spaces. Preprint, 2006.

[25] Jeff Cheeger and Bruce Kleiner. Generalized differential and bi-Lipschitz nonembedding in ℓ_1. *C. R. Math. Acad. Sci. Paris*, 343(5):297–301, 2006.

[26] Jeff Cheeger and Bruce Kleiner. Differentiating maps into ℓ_1 and the geometry of bv functions. *arXiv*, 0907(3295), 2009.

[27] Jeff Cheeger and Bruce Kleiner. Differentiating maps into L^1, and the geometry of BV functions. *Ann. of Math. (2)*, 171(2):1347–1385, 2010.

[28] Jeff Cheeger, Bruce Kleiner, and Assaf Naor. A $(\log n)^{\Omega(1)}$ integrality gap for the sparsest cut sdp. In *FOCS '09: Proceedings of the 2009 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 555–564, Washington, DC, USA, 2009. IEEE Computer Society.

[29] Chandra Chekuri, Sreeram Kannan, Adnan Raja, and Pramod Viswanath. Multicommodity flows and cuts in polymatroidal networks. In *ITCS*, pages 399–408, 2012.

[30] Chandra Chekuri, Sanjeev Khanna, and F. Bruce Shepherd. Edge-disjoint paths in planar graphs with constant congestion. *SIAM J. Comput.*, 39(1):281–301, 2009.

[31] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms.* The MIT Press, 3rd edition, 2009.

[32] P. Delsarte. An algebraic approach to the association schemes of coding theory. *Philips Res. Rep. Suppl.*, (10):vi+97, 1973.

[33] Nikhil R. Devanur, Subhash A. Khot, Rishi Saket, and Nisheeth K. Vishnoi. Integrality gaps for sparsest cut and minimum linear arrangement problems. In *STOC'06: Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, pages 537–546. ACM, New York, 2006.

[34] Michel Marie Deza and Monique Laurent. *Geometry of cuts and metrics*, volume 15 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 1997.

[35] P. Enflo. On the nonexistence of uniform homeomorphisms between L_p-spaces. *Ark. Mat.*, 8:103–105, 1969.

[36] Matthias Englert, Anupam Gupta, Robert Krauthgamer, Harald Räcke, Inbal Talgam-Cohen, and Kunal Talwar. Vertex sparsifiers: New results from old techniques. In *APPROX-RANDOM*, pages 152–165, 2010.

[37] P. Erdős and L. Lovász. 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*, pages 609–627. Colloq. Math. Soc. János Bolyai, Vol. 10. North-Holland, Amsterdam, 1975.

[38] Alex Eskin, David Fisher, and Kevin Whyte. Quasi-isometries and rigidity of solvable groups. *Pure Appl. Math. Q.*, 3(4, part 1):927–947, 2007.

[39] Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. *J. Comput. System Sci.*, 69(3):485–497, 2004.

[40] Uriel Feige, Mohammad Taghi Hajiaghayi, and James R. Lee. Improved approximation algorithms for minimum-weight vertex separators [extended abstract]. In *STOC'05: Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 563–572. ACM, New York, 2005.

[41] Uriel Feige, Mohammadtaghi Hajiaghayi, and James R. Lee. Improved approximation algorithms for minimum weight vertex separators. *SIAM J. Comput.*, 38(2):629–657, 2008.

[42] Uriel Feige, Mohammadtaghi Hajiaghayi, and James R. Lee. Improved approximation algorithms for minimum weight vertex separators. *SIAM J. Comput.*, 38(2):629–657, 2008. Prelim. version in *STOC 2005.*

[43] L. R. Ford and D. R. Fulkerson. Maximal flow through a network. *Canadian Journal of Mathematics*, 8:399–404, 1956.

[44] Lance Fortnow and Salil P. Vadhan, editors. *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011*. ACM, 2011.

[45] Anupam Gupta, Robert Krauthgamer, and James R. Lee. Bounded geometries, fractals, and low-distortion embeddings. In *FOCS*, pages 534–543, 2003.

[46] Anupam Gupta, Ilan Newman, Yuri Rabinovich, and Alistair Sinclair. Cuts, trees and l_1-embeddings of graphs. *Combinatorica*, 24(2):233–269, 2004.

[47] Refael Hassin. Minimum cost flow with set-constraints. *Networks*, 12(1):1–21, 1982.

[48] T. C. Hu. Multi-commodity network flows. *J. ORSA*, 11:344–360, 1963.

[49] Alexander Jaffe, James R. Lee, and Mohammad Moharrami. On the optimality of gluing over scales. *Discrete Comput. Geom.*, 46(2):270–282, 2011.

[50] William B. Johnson and Joram Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability (New Haven, Conn., 1982)*, volume 26 of *Contemp. Math.*, pages 189–206. Amer. Math. Soc., Providence, RI, 1984.

[51] Jeff Kahn, Gil Kalai, and Nathan Linial. The influence of variables on boolean functions (extended abstract). In *FOCS*, pages 68–80, 1988.

[52] George Karakostas. A better approximation ratio for the vertex cover problem. *ACM Trans. Algorithms*, 5(4):1–8, 2009.

[53] S.A. Khot and N.K. Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into ℓ_1. In *46th Annual Symposium on Foundations of Computer Science*, pages 53–62. IEEE Computer Soc., Los Alamitos, CA, 2005.

[54] Philip N. Klein, Serge A. Plotkin, and Satish Rao. Excluded minors, network decomposition, and multicommodity flow. In *Proceedings of the 25th Annual ACM Symposium on Theory of Computing*, pages 682–690, 1993.

[55] Jon Kleinberg and Eva Tardos. *Algorithm Design*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2005.

[56] Alexandra Kolla and James R. Lee. Sparsest cut on quotients of the hypercube. Preprint, 2008.

[57] R. Krauthgamer, J. R. Lee, M. Mendel, and A. Naor. Measured descent: a new embedding method for finite metrics. *Geom. Funct. Anal.*, 15(4):839–858, 2005.

[58] Robert Krauthgamer and Yuval Rabani. Improved lower bounds for embeddings into L_1. In *SODA '06: Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, pages 1010–1017, New York, NY, USA, 2006. ACM Press.

[59] Tomi J. Laakso. Plane with A_∞-weighted metric not bi-Lipschitz embeddable to $\mathbb{R}^N$. *Bull. London Math. Soc.*, 34(6):667–676, 2002.

[60] Urs Lang and Conrad Plaut. Bilipschitz embeddings of metric spaces into space forms. *Geom. Dedicata*, 87(1-3):285–307, 2001.

[61] E. L. Lawler and C. U. Martel. Computing maximal "polymatroidal" network flows. *Math. Oper. Res.*, 7(3):334–347, 1982.

[62] J. R. Lee. Volume distortion for subsets of Euclidean spaces. *Discrete Comput. Geom.*, 41(4):590–615, 2009.

[63] J. R. Lee, Arnaud de Mesmay, and Mohammad Moharrami. Dimension reduction for finite trees in ℓ_1. In *SODA*, 2012.

[64] J. R. Lee and A. Naor. Embedding the diamond graph in L_p and dimension reduction in L_1. *Geom. Funct. Anal.*, 14(4):745–747, 2004.

[65] J. R. Lee and P. Rhagevendra. Coarse differentiation and multi-flows in planar graphs. 2007.

[66] James R. Lee. On distance scales, embeddings, and efficient relaxations of the cut cone. In *SODA '05: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 92–101, Philadelphia, PA, USA, 2005. Society for Industrial and Applied Mathematics.

[67] James R. Lee, Manor Mendel, and Mohammad Moharrami. A node-capacitated okamura-seymour theorem. *CoRR*, abs/1209.2744, 2012.

[68] James R. Lee and Moharrami Mohammad. Bilipschitz snowflakes metrics of negative type, and psd flows. In *STOC*, 2010.

[69] James R. Lee and Mohammad Moharrami. A lower bound on dimension reduction for trees in ℓ_1. 2013. arXiv:1302.6542.

[70] James R. Lee and Assaf Naor. Extending Lipschitz functions via random metric partitions. *Invent. Math.*, 160(1):59–95, 2005.

[71] James R. Lee and Assaf Naor. L_p metrics on the Heisenberg group and the goemans-linial conjecture. In *FOCS*, pages 99–108, 2006.

[72] James R. Lee, Assaf Naor, and Yuval Peres. Trees and Markov convexity. *Geom. Funct. Anal.*, 18(5):1609–1659, 2009.

[73] James R. Lee and Anastasios Sidiropoulos. On the geometry of graphs with a forbidden minor. In *STOC'09—Proceedings of the 2009 ACM International Symposium on Theory of Computing*, pages 245–254. ACM, New York, 2009.

[74] James R. Lee and Anastasios Sidiropoulos. Near-optimal distortion bounds for embedding doubling spaces into l_1. In Fortnow and Vadhan [44], pages 765–772.

[75] Tom Leighton and Satish Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *J. ACM*, 46(6):787–832, 1999.

[76] Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

[77] J. Matoušek. On embedding trees into uniformly convex Banach spaces. *Israel J. Math.*, 114:221–237, 1999.

[78] J. Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002.

[79] Jiří Matoušek. On embedding expanders into l_p spaces. *Israel J. Math.*, 102:189–197, 1997.

[80] Jiri Matoušek. Open problems on embeddings of finite metric spaces. *Available at http://kam.mff.cuni.cz/∼matousek/haifaop.ps*, 2002.

[81] Jiří Matoušek and Yuri Rabinovich. On dominated l_1 metrics. *Israel J. Math.*, 123:285–301, 2001.

[82] David W. Matula and Farhad Shahrokhi. Sparsest cuts and bottlenecks in graphs. *Discrete Appl. Math.*, 27(1-2):113–123, 1990. Computational algorithms, operations research and computer science (Burnaby, BC, 1987).

[83] Colin McDiarmid. Concentration. In *Probabilistic methods for algorithmic discrete mathematics*, volume 16 of *Algorithms Combin.*, pages 195–248. Springer, Berlin, 1998.

[84] Robert J. McEliece, Eugene R. Rodemich, Howard Rumsey, Jr., and Lloyd R. Welch. New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities. *IEEE Trans. Information Theory*, IT-23(2):157–166, 1977.

[85] Raghu Meka. A PRG for lipschitz functions of polynomials with applications to sparsest cut. 2013. arXiv:1211.1109.

[86] K. Menger. Zur allgemeinen kurventheorie. *Fund. Math.*, 10:96–116, 1927.

[87] Ilan Newman and Yuri Rabinovich. A lower bound on the distortion of embedding planar metrics into Euclidean space. *Discrete Comput. Geom.*, 29(1):77–81, 2003.

[88] Ilan Newman and Yuri Rabinovich. On cut dimension of ℓ_1 metrics and volumes, and related sparsification techniques. *CoRR*, abs/1002.3541, 2010.

[89] Haruko Okamura and P. D. Seymour. Multicommodity flows in planar graphs. *J. Combin. Theory Ser. B*, 31(1):75–81, 1981.

[90] Pierre Pansu. Métriques de Carnot-Carathéodory et quasiisométries des espaces symétriques de rang un. *Ann. of Math. (2)*, 129(1):1–60, 1989.

[91] Satish Rao. Small distortion and volume preserving embeddings for planar and Euclidean metrics. In *Proceedings of the 15th Annual Symposium on Computational Geometry*, pages 300–306, New York, 1999. ACM.

[92] Oded Regev. Entropy-based bounds on dimension reduction in L_1. 2011. arXiv:1108.1283.

[93] Gideon Schechtman. More on embedding subspaces of L_p in l_r^n. *Compositio Math.*, 61(2):159–169, 1987.

[94] Gideon Schechtman. Dimension reduction in l_p, $0 < p < 2$. 2011. arXiv:1110.2148.

[95] Leonard J. Schulman. Coding for interactive communication. *IEEE Trans. Inform. Theory*, 42(6, part 1):1745–1756, 1996. Codes and complexity.

[96] Sachdeva Sushant and Moharrami Mohammad. The power of weak vs. storng triangle inequality. 2011.

[97] Michel Talagrand. Embedding subspaces of L_1 into l_1^N. *Proc. Amer. Math. Soc.*, 108(2):363–369, 1990.