

# Trees and Markov convexity

James R. Lee\*

Assaf Naor<sup>†</sup>

Yuval Peres<sup>‡</sup>

## Abstract

We show that an infinite weighted tree admits a bi-Lipschitz embedding into Hilbert space if and only if it does not contain arbitrarily large complete binary trees with uniformly bounded distortion. We also introduce a new metric invariant called Markov convexity, and show how it can be used to compute the Euclidean distortion of any metric tree up to universal factors.

## 1 Introduction

Given two metric spaces  $(X, d_X)$ ,  $(Y, d_Y)$ , and a mapping  $f : X \rightarrow Y$ , we denote the Lipschitz constant of  $f$  by  $\|f\|_{\text{Lip}}$ . If  $f$  is injective then the (bi-Lipschitz) distortion of  $f$  is defined as  $\text{dist}(f) = \|f\|_{\text{Lip}} \cdot \|f^{-1}\|_{\text{Lip}}$ . The smallest distortion with which  $X$  embeds into  $Y$  is denoted  $c_Y(X)$ , i.e.  $c_Y(X) = \inf\{\text{dist}(f) : f : X \hookrightarrow Y\}$ . When  $Y = L_p$  for some  $p \geq 1$  we use the shorter notation  $c_p(X) = c_{L_p}(X)$ . The parameter  $c_2(X)$  is known in the literature as the Euclidean distortion of  $X$ .

The ubiquitous problem of embedding metric spaces into “simpler” spaces occurs in various aspects of functional analysis, Riemannian geometry, group theory, and computer science. In most cases low distortion embeddings are used to “simplify” a geometric object by representing it as a subset of a better understood geometry. In other cases, embeddings are used to characterize important invariants such as various notions of dimensionality in metric spaces, and superreflexivity, type and cotype of normed spaces. More recently, striking applications of bi-Lipschitz embeddings were found in computer science, where the information obtained from concrete geometric representations of finite spaces is used to design efficient approximation algorithms and data structures.

The present paper is devoted to the study of the Euclidean (and  $L_p$ ) distortion of trees. In what follows by a *metric tree* we mean the shortest path metric induced on the vertices of a weighted graph-theoretical tree  $T = (V, E)$ . In fact, all of our results will hold true for arbitrary subsets of metric trees, which are characterized among all metric spaces by the well known four point condition: For every four points  $x, y, u, v$  two of the three numbers  $\{d(x, y) + d(u, v), d(x, u) + d(y, v), d(x, v) + d(y, u)\}$  are the same, and that number is at least as large as the third (see [10]). But, because our statements are asymptotic in nature, this does not increase the generality of our results, since Gupta [16] proved that any finite subset of a metric tree is bi-Lipschitz equivalent to a metric tree with distortion at most 8. The  $\mathbb{R}$ -tree corresponding to a tree  $T$  is the one-dimensional simplicial complex induced by  $T$ , i.e. the path metric obtained by replacing each edge in  $T$  by an interval whose length is the weight of the edge. The  $\mathbb{R}$ -tree corresponding to  $T$  will be denoted

---

\*Computer Science and Engineering, University of Washington. This research was conducted while the author was at U. C. Berkeley and the Institute for Advanced Study. [jrl@cs.washington.edu](mailto:jrl@cs.washington.edu)

<sup>†</sup>Microsoft Research. [anaor@microsoft.com](mailto:anaor@microsoft.com)

<sup>‡</sup>Microsoft Research and U. C. Berkeley. [peres@stat.berkeley.edu](mailto:peres@stat.berkeley.edu)

$T_{\mathbb{R}}$ . In what follows, when we refer to an  $\mathbb{R}$ -tree we mean an  $\mathbb{R}$ -tree corresponding to some metric tree. We will see later that for every metric tree  $T$  and every  $p \geq 1$ ,  $c_p(T)$  has the same order of magnitude as  $c_p(T_{\mathbb{R}})$ , so in most cases the distinction between metric trees and  $\mathbb{R}$ -trees will not be important, though in a few instances we will need to distinguish the two notions.

Let  $B_k$  denote the complete binary tree of depth  $k$  (with unit edge weights). In a famous paper [6] Bourgain proved that the Euclidean distortion of  $B_k$  is  $\Theta(\sqrt{\log k})$ . Moreover, he showed that a Banach space  $Y$  is superreflexive (i.e. admits an equivalent uniformly convex norm) if and only if  $\lim_{k \rightarrow \infty} c_Y(B_k) = \infty$ . This remarkable characterization of a linear property of Banach spaces in terms of their metric structure sparked a considerable amount of work on problems of a similar flavor (see the introduction of [30] for more information on this topic). Among the corollaries of Bourgain’s work is the following dichotomy: For a Banach space  $Y$  either  $c_Y(B_k) = 1$  for all  $k$ , or there exists  $\alpha > 0$  such that  $c_Y(B_k) = \Omega((\log k)^\alpha)$  (similar phenomena are known to hold in a few other cases—see [7, 30]). Moreover, Bourgain used his theorem to solve a question posed by Gromov, showing that the hyperbolic plane does not admit a bi-Lipschitz embedding into Hilbert space. Similar applications of Bourgain’s theorem to prove that certain metric spaces do not embed into Hilbert space were obtained by Benjamini and Schramm [5] in the case of graphs with positive Cheeger constant, and by Leuzinger [21] in the case of certain Tits buildings.

The bi-Lipschitz structure of trees has been studied extensively. Trees are the “building blocks” of hyperbolic geometry, and embeddings of certain non-positively curved spaces into products of trees are used in several contexts (see for example [9, 8, 31, 20]). Similar results (known as “probabilistic embeddings into trees”) are a powerful tool in computer science (see for example [4, 12]). We refer to [10, 27, 18] for other results on the Lipschitz structure of trees. In spite of these applications, and the vast amount of work on trees in the Lipschitz category, the following natural problem remained open: When does an infinite metric tree embed with finite distortion into Hilbert space? One of the main results of this paper is the following answer to this question.

**Theorem 1.1.** *Let  $T = (V, E)$  be an infinite metric tree. Then the following conditions are equivalent.*

1.  $c_2(T) = \infty$ .
2.  $\sup_{k \in \mathbb{N}} c_T(B_k) < \infty$ .
3. For every  $k \in \mathbb{N}$ ,  $c_T(B_k) = 1$ .

In other words, a metric tree admits a bi-Lipschitz embedding into Hilbert space if and only if it does not bi-Lipschitzly contain arbitrarily large complete binary trees. Thus there is a unique obstruction to a tree being non-Euclidean. Similar “unique obstruction” results are known only in very few cases: As we mentioned above, Bourgain [6] proved that complete binary trees are the unique obstruction to a Banach space being superreflexive; Bourgain, Milman and Wolfson [7] showed that Hamming cubes are the unique obstruction to a metric space having non-trivial type; Mendel and Naor [30] showed that  $\ell_\infty^n$  integer grids are the unique obstruction to a metric space having finite cotype; Thomassen [35] proved that certain transient graphs must contain transient trees, and Benjamini and Schramm [5] proved that a graph with positive Cheeger constant must contain a tree with positive Cheeger constant. Another result in the spirit of Theorem 1.1 is the tree Szemerédi theorem of Furstenberg and Weiss [14]: A subset of positive density in the infinite complete binary tree must contain arbitrarily large copies of complete binary trees.

It is not surprising that the Theorem 1.1 is a “local” result, in the sense that it deals with finite subsets of the metric tree  $T$ . Indeed, it is well known that a metric space embeds into Hilbert space if and only if all of its finite subsets do. It is thus natural to expect characterizations in the spirit of Theorem 1.1 to be local. Let us say that a metric space  $X$  is *finitely representable in a metric space*  $Y$  if there exists a constant  $D \geq 1$  such that for every finite subset  $F \subseteq X$  we have  $c_Y(F) \leq D$  (this is an obvious adaptation of standard terminology from Banach space theory). Thus, denoting by  $B_\infty$  the infinite unweighted complete binary tree, Theorem 1.1 can be rephrased as follows: A metric tree  $T$  admits a bi-Lipschitz embedding into Hilbert space if and only if  $B_\infty$  is not finitely representable in  $T$ . The following section contains optimal quantitative versions of this result, and explains the ingredients of its proof.

## 1.1 Markov convexity and quantitative bounds

Quantitative bounds on the Euclidean distortion of trees were obtained in [24, 28, 25, 17]. In particular, Matoušek proved in [28] that for any  $n$ -point metric tree  $T$  we have  $c_2(T) = O(\sqrt{\log \log n})$ . This result cannot be improved due to Bourgain’s lower bound for the complete binary tree. Gupta, Krauthgamer and Lee [17] obtained upper bounds on the Euclidean distortion of trees in terms of their doubling constant; in particular, they showed that every doubling tree admits a bi-Lipschitz embedding into a finite-dimensional Euclidean space. We present a new simpler proof of this fact in Section 2.3, where we also recall the definition of the doubling constant.

We shall now state an optimal quantitative version of Theorem 1.1. Given a metric space  $(X, d_X)$ ,  $k \in \mathbb{N}$  and  $c > 1$ , we denote

$$\mathcal{B}_X(c) = \max \{k \in \mathbb{N} : c_X(B_k) < c\}.$$

In what follows we write  $A \lesssim B$  to mean  $A = O(B)$ . If  $A \lesssim B$  and  $B \lesssim A$  then we write  $A \approx B$ .

**Theorem 1.2.** *Let  $T$  be an arbitrary metric tree. Then for every  $c > 1$ ,*

$$\frac{1}{c} \sqrt{\log \mathcal{B}_T(c)} \lesssim c_2(T) \lesssim \sqrt{\frac{c}{c-1}} \cdot \mathcal{B}_T(c).$$

The lower bound in Theorem 1.2 is simply Bourgain’s lower bound, and is therefore optimal. Somewhat surprisingly, the upper bound in Theorem 1.2 cannot be improved. The construction of a family of trees exhibiting this, which we call the Cantor trees, is presented in Section 3.3.3.

It follows that in order to obtain tight bounds on the Euclidean distortion of a given metric tree  $T$  we need an invariant which is more refined than the size of the biggest binary tree contained in  $T$ . This is achieved via the following definition. Let  $\{X_t\}_{t=0}^\infty$  be a Markov chain on a state space  $\Omega$ . Given an integer  $k \geq 0$  we denote by  $\{\tilde{X}_t(k)\}_{t=0}^\infty$  the process which equals  $X_t$  for time  $t \leq k$ , and evolves independently (with respect to the same transition probabilities) for time  $t > k$ . Observe that for  $k < 0$ ,  $\tilde{X}_t(k)$  and  $X_t$  evolve independently for all  $t \geq 0$ .

**Definition 1.3.** *Let  $(X, d)$  be a metric space and  $p > 0$ . We shall say that  $(X, d)$  is Markov  $p$ -convex with constant  $\Pi$  if for every Markov chain  $\{X_t\}_{t=0}^\infty$  on a state space  $\Omega$ , and every  $f : \Omega \rightarrow X$ , we have for every  $m \in \mathbb{N}$ ,*

$$\sum_{k=0}^m \sum_{t=1}^{2^m} \frac{\mathbb{E} \left[ d \left( f(X_t), f(\tilde{X}_t(t-2^k)) \right)^p \right]}{2^{kp}} \leq \Pi^p \sum_{t=1}^{2^m} \mathbb{E} [d(f(X_t), f(X_{t-1}))^p]. \quad (1)$$

The least constant  $\Pi$  above is called the Markov  $p$ -convexity constant of  $X$ , and is denoted  $\Pi_p(X)$ . We shall say that  $X$  is Markov  $p$ -convex if  $\Pi_p(X) < \infty$ .

To understand this notion, recall that the chains  $X_t$  and  $\tilde{X}_t(t - 2^k)$  run together for the first  $t - 2^k$  steps, and then evolve independently for the remaining  $2^k$  steps. Thus the left hand side in (1) is measuring the sum, over many “dyadic scales”  $k \in \{0, 1, 2, \dots\}$  of the average of the  $p$ th power of the normalized “drift” of the chain in  $X$  with respect to scale  $k$ . We will say that  $X$  has *non-trivial Markov convexity* if  $X$  is Markov  $p$ -convex for some  $p < \infty$ . We note that  $L_2$  is Markov 2-convex. More generally, the name comes from the fact that if  $X$  is a Banach space which admits an equivalent uniformly convex norm whose modulus of convexity is of power type  $p$ , then  $X$  is also Markov  $p$ -convex. These results are proved in Section 3.1.

In Bourgain’s paper [6] there is an implicit “non-linear” notion of uniform convexity related to the presence of complete binary trees. For the results in this paper, we require the above “Markov variant,” analogous to Ball’s notion of Markov type [2]. The search for Poincaré-type inequalities on metric spaces which are analogs of classical Banach space invariants have been fruitfully investigated by several authors—we refer to the papers [11, 15, 7, 34, 2, 32, 31, 30] for a discussion of this research direction, to which Definition 1.3 belongs. The following theorem shows that Markov convexity determines the Euclidean distortion of a tree, up to universal factors.

**Theorem 1.4.** *Let  $T$  be a metric tree. Then  $c_2(T) \approx \Pi_2(T_{\mathbb{R}})$ .*

Recall that  $T_{\mathbb{R}}$  denotes the  $\mathbb{R}$ -tree corresponding to  $T$ . See Remark 3.3 for a discussion of why we have to pass to  $\mathbb{R}$ -trees in Theorem 1.4.

We also obtain a combinatorial way to compute the Euclidean distortion of any tree. Let  $T = (V, E)$  be a metric tree, and let  $\chi : E \rightarrow \mathbb{Z}$  be an edge coloring. We call  $\chi$  a monotone coloring if all of its color classes are paths contained in a root-leaf path (such paths are called monotone paths in what follows). For  $\delta \in (0, 1)$ , the coloring  $\chi$  is called  $\delta$ -strong if it is monotone and for every  $u, v \in V$  at least half of the length of the path joining  $u$  and  $v$  can be covered by color classes of length at least  $\delta d_T(u, v)$ . We define  $\delta^*(T)$  to be the largest  $\delta$  for which  $T$  admits a  $\delta$ -strong coloring. The following theorem shows that  $\delta^*(T)$  determines the Euclidean distortion of  $T$ .

**Theorem 1.5.** *Let  $T$  be a metric tree. Then*

$$c_2(T) \approx \sqrt{1 + \log \left( \frac{1}{\delta^*(T)} \right)}.$$

The upper bound on  $c_2(T)$  in Theorem 1.5 continues a theme that also appeared in [24, 28, 17]: Certain edge colorings can be used to construct embeddings into  $L_2$ . Specifically, our proof draws on ideas from Matoušek’s embedding [28]. But, Matoušek’s argument requires colorings with a small number of colors, the existence of which depends only on the topology of  $T$  and does not take into account the edge lengths. Our argument for the upper bound, which is contained in Theorem 2.6, builds on Matoušek’s proof while taking the metric into consideration, and is therefore more involved.

The lower bound on  $c_2(T)$  in Theorem 1.5 goes through Theorem 1.4. We construct a special coloring of  $T$ , and show that if the coloring is not  $\delta$ -strong, then we can construct a Markov chain on  $T$  which wanders too quickly for  $T$  to have a small Markov 2-convexity constant. This is done by locating a special type of subtree of  $T$ , which we call a weak prototype—see Section 3.3

for the definition, where it is shown that weak prototypes cannot have good Markov convexity properties. This “reconstruction paradigm” is inspired by a result of [17] which shows that if a certain procedure fails to produce a good coloring, then the tree under consideration must have a large doubling constant. Our approach is able to pick out significantly more delicate sub-structures (e.g. embedded complete binary trees or the aforementioned “weak prototypes”). A key difficulty that arises in our setting involves choosing the “scale” at which the required weak prototype embeds into  $T$ . This “scale selection” argument is a central part of our proof of Theorem 1.2, Theorem 1.4, and Theorem 1.5—we refer to Section 2.2 and Section 4 for the details.

We remark that all of our results can be applied to compute the  $L_p$  distortion of trees. Namely, we show that for every  $p, c > 1$  and every metric tree  $T$ ,

$$\frac{1}{c} (\log \mathcal{B}_T(c))^{\min\{\frac{1}{p}, \frac{1}{2}\}} \lesssim c_p(T) \lesssim \left( \frac{c}{c-1} \cdot \mathcal{B}_T(c) \right)^{\min\{\frac{1}{p}, \frac{1}{2}\}}. \quad (2)$$

and

$$c_p(T) \approx \Pi_{\max\{p, 2\}}(T_{\mathbb{R}}) \approx \left[ \log \left( \frac{2}{\delta^*(T)} \right) \right]^{\min\{\frac{1}{p}, \frac{1}{2}\}}, \quad (3)$$

where the implied constants may depend only on  $p$ ; see Theorem 4.1.

The use of Markov convexity as a metric invariant, and thus a tool for proving distortion lower bounds, is not limited to the case of trees. In Section 3.2 we investigate classes of spaces which can be shown not to embed into  $L_2$  by analyzing their Markov convexity. In particular, we prove a lower bound on the Euclidean distortion of balls of finitely generated groups (equipped with the word metric) which admit a bounded non-constant harmonic function. We also bound from below the Euclidean distortion of the lamplighter group over the cycle (see Section 3.2 for the definition). In a future paper, which will be devoted to embeddings of the lamplighter group, we use the methods of [31] to show that this group has Markov type 2 in the sense of Ball [2]. Thus, Markov convexity is the only known invariant which demonstrates that this group does not well-embed into Hilbert space.

Our results, specifically Theorem 1.5, have algorithmic implications. Given an  $n$ -point metric space  $X$ , the problem of efficiently computing its distortion in a class of metric spaces up to a small factor has attracted a lot of attention in recent years, and is known as the “relative embedding” problem. We refer to [1] and the references therein for a discussion of this topic, and also for some hardness results. The Euclidean distortion of an  $n$ -point metric space can be computed in polynomial time, since this problem can be cast as a semidefinite program [23]. Hence Theorem 1.5 yields a polynomial time algorithm for estimating the parameter  $\log \left( \frac{1}{\delta^*(T)} \right)$  up to a constant factor for any tree  $T$ . In conjunction with (3), this gives a polynomial time algorithm which computes the  $L_p$  distortion of any tree up to a universal constant factor. Note that it not known whether the  $L_p$  distortion of a general finite metric space can be approximated efficiently.

## 1.2 Some open problems

We end this introduction by stating some interesting open problems that arise from our work.

**Problem 1.** In Section 3.1 we show that every  $p$ -uniformly convex Banach space is Markov  $p$ -convex. We also show that if  $X$  is a Banach lattice which is Markov  $p$ -convex then it is also

$q$ -uniformly convex for every  $q > p$ . The relation between Markov  $p$ -convexity and uniform  $p$ -convexity for general Banach spaces is unclear.

**Problem 2.** One corollary of our results is that if a metric tree is not Markov  $p$ -convex for any  $p < \infty$  then it contains arbitrarily large complete binary trees with distortion arbitrarily close to 1. It is possible that this holds true for arbitrary metric spaces, and not just metric trees. If this is the case, then it would correspond to known results in Banach space theory, and would complement the existing theory of metric type and cotype.

**Problem 3.** It would be interesting to investigate other “unique obstruction” results of the type described here. In particular, can one classify the obstructions to a planar graph being embeddable in  $L_2$ ? Another interesting generalization would be to classify the subsets of  $\mathbb{H}^2$ —the hyperbolic plane—which embed into  $L_2$ ; it seems plausible that complete binary trees are the only obstruction in this case, just as for tree metrics. In a similar vein, it might be the case that the only subsets of a product of trees which do not admit a bi-Lipschitz embedding into  $L_2$  are those that contain arbitrarily large bi-Lipschitz copies of complete binary trees. If true, then in combination with the result of [8], this would imply the same result for the hyperbolic plane.

**Problem 4.** In Section 3.2 we give lower bounds on the Euclidean distortion of the lamplighter group over the  $n$ -cycle. We do not know what is the correct asymptotic behavior of this distortion. It is also unknown whether or not these groups embed into  $L_1$  with uniformly bounded distortion.

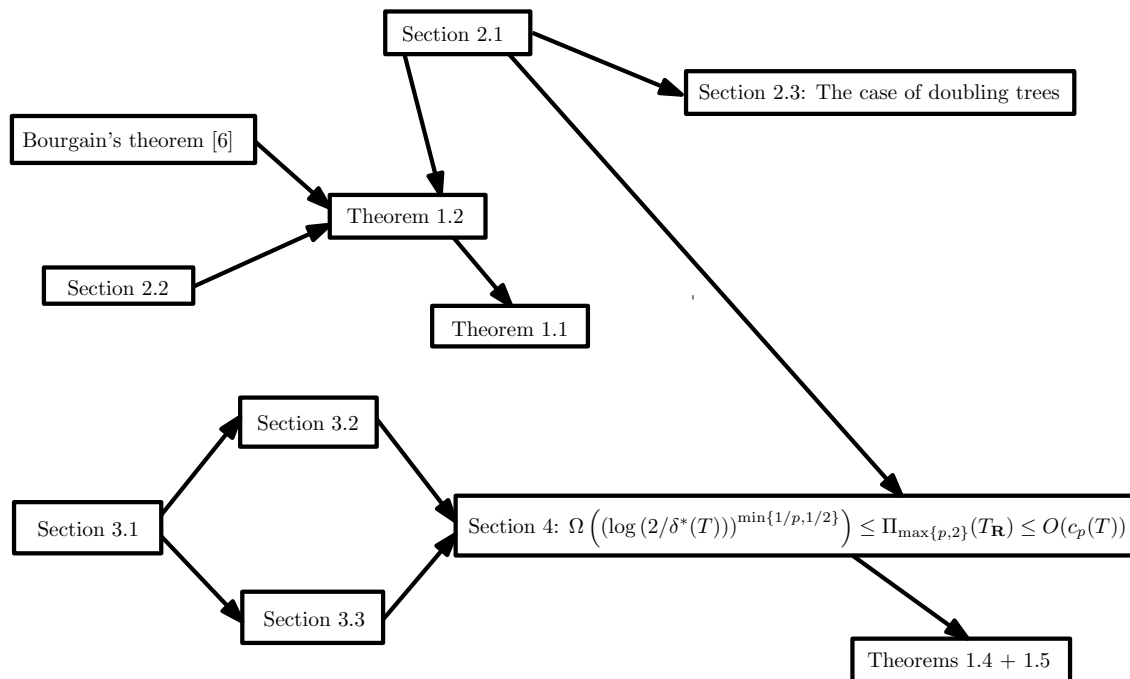


Figure 1: A schematic description of the implications between the sections in this paper.

## 2 Distortion bounds via the containment of binary trees

The purpose of this section is to prove the following theorem which, when combined with Bourgain's lower bound for binary trees [6], implies Theorem 1.1 and Theorem 1.2.

**Theorem 2.1.** *Let  $T$  be an arbitrary metric tree and  $p \geq 1$ . Then for every  $c > 1$ ,*

$$c_p(T) \leq 130 \left( \frac{c}{c-1} \cdot \mathcal{B}_T(c) \right)^{\min\{\frac{1}{p}, \frac{1}{2}\}}.$$

In Section 3.3.3 we will show that the asymptotic dependence on  $\mathcal{B}_T(c)$  in the upper bound of Theorem 2.1 cannot be improved.

### 2.1 Coloring based upper bounds

We begin with some definitions and notation. Let  $T = (V, E)$  be a *finite* graph-theoretic tree with positive edge lengths  $\ell : E \rightarrow (0, \infty)$ , and let  $d_T$  be the induced path metric on  $V$ . We also fix some arbitrary root  $r \in T$ . A *monotone path* in  $T$  is a connected subset (also called a *segment* in what follows) of some root-leaf path. By an *edge-coloring of  $T$* , we mean a map  $\chi : E \rightarrow \mathbb{Z}$ . We say that the coloring is *monotone* if for every  $m \in \mathbb{Z}$  the color class  $\chi^{-1}(m)$  is a monotone path. For  $u, v \in V$  we let  $P(u, v) \subseteq E$  denote the unique path from  $u$  to  $v$ , and set  $P(v) = P(v, r)$ . Given an edge coloring  $\chi : E \rightarrow \mathbb{Z}$ ,  $k \in \chi(E)$ , and  $u, v \in V$ , we write

$$\ell_k^\chi(u, v) := \sum_{\substack{e \in P(u, v) \\ \chi(e) = k}} \ell(e).$$

We also set  $\ell_k^\chi(v) = \ell_k^\chi(v, r)$ . Finally, given  $u, v \in V$  we let  $\mathbf{lca}(u, v)$  denote the *least common ancestor* of  $u$  and  $v$  in  $T$ .

**Definition 2.2** ( $\varepsilon$ -good coloring). *We say that a coloring  $\chi : E \rightarrow \mathbb{Z}$  is  $\varepsilon$ -good if it is monotone, and for every  $u, v \in T$ , the unique path from  $u$  to  $v$  contains a monochromatic segment of length at least  $\varepsilon \cdot d_T(u, v)$ . We define  $\varepsilon^*(T)$  to be the largest  $\varepsilon$  for which  $T$  admits an  $\varepsilon$ -good coloring.*

The following simple lemma will not be used in the proof of Theorem 2.1, but we include it since it illustrates the relation between colorings and embeddings, and it will be used eventually in Section 2.3.

**Lemma 2.3.** *For every weighted tree  $T$  and  $p \geq 1$ ,*

$$c_p(T) \leq \frac{2^{1/p}}{\varepsilon^*(T)}.$$

*Proof.* Fix  $\varepsilon < \varepsilon^*(T)$  and let  $\chi : E \rightarrow \mathbb{Z}$  be an  $\varepsilon$ -good coloring. Let  $\{e_k\}_{k \in \mathbb{Z}}$  be the standard basis of  $\ell_p = \ell_p(\mathbb{Z})$ . Define  $f : V \rightarrow \ell_p$  by

$$f(v) = \sum_{k \in \mathbb{Z}} \ell_k^\chi(v) e_k.$$

(Recall that  $\ell_k(v)$  is the distance that the segment colored  $k$  contributes to the path joining  $v$  to the root).

Fix  $u, v \in V$  and write  $w = \mathbf{lca}(u, v)$ . The fact that the coloring  $\chi$  is monotone implies that  $\chi(P(u, w)) \cap \chi(P(v, w)) = \emptyset$ . Thus

$$d_T(u, v) = \sum_{e \in P(u, w)} \ell(e) + \sum_{e \in P(v, w)} \ell(e) = \sum_{k \in \mathbb{Z}} |\ell_k^\chi(u) - \ell_k^\chi(v)| = \|f(u) - f(v)\|_1 \geq \|f(u) - f(v)\|_p.$$

On the other hand, since  $\chi$  is  $\varepsilon$ -good, there are  $a, b \in \mathbb{Z}$  such that  $\ell_a^\chi(u, w) \geq \varepsilon d_T(u, w)$  and  $\ell_b^\chi(v, w) \geq \varepsilon d_T(v, w)$ . It follows that

$$\|f(u) - f(v)\|_p \geq ([\ell_a^\chi(u, w)]^p + [\ell_b^\chi(v, w)]^p)^{1/p} \geq \frac{\varepsilon}{2^{1/p}} [d_T(u, w) + d_T(v, w)] = \frac{\varepsilon}{2^{1/p}} d_T(u, v). \quad (4)$$

□

To get tighter control on the Euclidean distortion of trees we introduce the notion of  $\delta$ -strong colorings.

**Definition 2.4** ( $\delta$ -strong coloring). *We say that a coloring  $\chi : E \rightarrow \mathbb{Z}$  is  $\delta$ -strong if it is monotone, and for every  $u, v \in V$*

$$\sum_{k \in \mathbb{Z}} \ell_k^\chi(u, v) \cdot \mathbf{1}_{\{\ell_k^\chi(u, v) \geq \delta d_T(u, v)\}} \geq \frac{1}{2} d_T(u, v).$$

*In words, we demand that at least half of the shortest path joining  $u$  and  $v$  is covered by color classes of length at least  $\delta d_T(u, v)$ . As before, we define  $\delta^*(T)$  to be the largest  $\delta$  for which  $T$  admits an  $\delta$ -strong coloring.*

The notions of  $\delta$ -strong colorings and  $\varepsilon$ -good colorings are related via the following simple lemma.

**Lemma 2.5.** *Every weighted tree  $T$  satisfies  $\delta^*(T) \geq 2^{-3/\varepsilon^*(T)}$ .*

*Proof.* Let  $\chi$  be an  $\varepsilon$ -good coloring of  $T$ . We will prove that it is also  $2^{-3/\varepsilon}$ -strong. In fact, we shall show that for every  $\alpha \in (0, 1]$  and  $u, v \in V$ , the total length of the monochromatic segments of length at least  $\alpha d_T(u, v)$  on the path  $P(u, v)$  satisfies

$$\sum_{k \in \mathbb{Z}} \ell_k^\chi(u, v) \cdot \mathbf{1}_{\{\ell_k^\chi(u, v) \geq \alpha d_T(u, v)\}} \geq \left(1 - \left(\frac{\alpha}{\varepsilon}\right)^{\varepsilon/2}\right) d_T(u, v). \quad (5)$$

Choosing  $\alpha = 2^{-3/\varepsilon}$  in (5), and using the fact that  $2^{1/\varepsilon} \geq \frac{1}{\varepsilon}$ , we deduce that  $\chi$  is  $2^{-3/\varepsilon}$ -strong. The proof of (5) is by induction on  $d_T(u, v)$ . If  $d_T(u, v)$  is minimal then  $P(u, v)$  is an edge, and hence monochromatic, so that the assertion holds trivially. In general, since the coloring  $\chi$  is  $\varepsilon$ -good, there are two vertices on the path  $P(u, v)$  such that the segment  $P(a, b)$  is monochromatic and  $d_T(a, b) \geq \varepsilon d_T(u, v)$ . Without loss of generality we assume that  $d_T(a, u) < d_T(b, v)$ . If  $\varepsilon < \alpha$  then there is nothing to prove, so assume that  $\varepsilon \geq \alpha$ . Denoting  $A = d_T(u, a)$ ,  $B = d_T(b, v)$  and  $D = d_T(u, v)$ , we apply the inductive hypothesis to the paths joining  $u$  and  $a$  and  $b$  and  $v$ , to get



that

$$\begin{aligned}
\sum_{k \in \mathbb{Z}} \ell_k^\chi(u, v) \cdot \mathbf{1}_{\{\ell_k^\chi(u, v) \geq \alpha d_T(u, v)\}} &\geq d_T(a, b) + \left(1 - \left(\frac{D\alpha}{A\varepsilon}\right)^{\varepsilon/2}\right) A + \left(1 - \left(\frac{D\alpha}{B\varepsilon}\right)^{\varepsilon/2}\right) B \\
&= D - \left(\frac{D\alpha}{\varepsilon}\right)^{\varepsilon/2} \left(A^{1-\varepsilon/2} + B^{1-\varepsilon/2}\right) \\
&\geq D - 2 \left(\frac{D\alpha}{\varepsilon}\right)^{\varepsilon/2} \left(\frac{A+B}{2}\right)^{1-\varepsilon/2} \tag{6}
\end{aligned}$$

$$\geq D - 2 \left(\frac{D\alpha}{\varepsilon}\right)^{\varepsilon/2} \left(\frac{(1-\varepsilon)D}{2}\right)^{1-\varepsilon/2} \tag{7}$$

$$\geq \left(1 - \left(\frac{\alpha}{\varepsilon}\right)^{\varepsilon/2}\right) D, \tag{8}$$

Where in (6) we used the concavity of the function  $t \mapsto t^{1-\varepsilon/2}$ , in (7) we used the fact that  $D = A+B+d_T(a, b) \geq A+B+\varepsilon D$ , and in (8) we used the elementary inequality  $2^{\varepsilon/2}(1-\varepsilon)^{1-\varepsilon/2} \leq 1$ , which is valid for every  $\varepsilon \in [0, 1]$ .  $\square$

In [28], Matoušek proves that if  $\chi$  is a monotone edge-coloring of  $T$  such that every root-leaf path contains at most  $M$  distinct color classes, then  $c_p(T) \leq O\left((\log M)^{\min\{\frac{1}{p}, \frac{1}{2}\}}\right)$ . Clearly any such coloring is  $1/(2M)$ -strong. The next theorem generalizes Matoušek's result along these lines.

We suggest that the reader skip this rather technical proof upon a first reading. In particular, the much simpler Lemma 2.3 suffices for the proof of Theorem 1.1, although it does not give the optimal quantitative bounds of Theorem 1.2.

**Theorem 2.6.** *For every weighted tree  $T = (V, E)$  and  $p \geq 1$ ,*

$$c_p(T) \leq 4 \left[ \log \left( \frac{2}{\delta^*(T)} \right) \right]^{\min\{\frac{1}{p}, \frac{1}{2}\}}.$$

*Proof.* We may assume that  $p \in [2, \infty)$ , since if  $p \in [1, 2)$  the required result follows by embedding  $T$  into  $\ell_2 \subseteq L_p$ . Fix  $\delta < \min\{\delta^*(T), 1/2\}$  and let  $\chi : E \rightarrow \mathbb{Z}$  be a  $\delta$ -strong coloring. Let  $\{e_k\}_{k \in \mathbb{Z}}$  be as in Lemma 2.3. For  $v \in V$  we denote by  $(k_1(v), \dots, k_{m_v}(v))$  the sequence of color classes encountered on the path from the root to  $v$ . We shall denote by  $d_j(v)$  the distance that the color class  $k_j(v)$  contributes to the path from the root to  $v$ , i.e.

$$d_j(v) = \sum_{\substack{e \in P(v) \\ \chi(e) = k_j(v)}} \ell(e).$$

Now let

$$s_i(v) = \sum_{j=i}^{m_v} \max \left\{ 0, d_j(v) - \frac{\delta}{2} \sum_{h=i}^j d_h(v) \right\},$$

and define  $f : V \rightarrow \ell_p(\mathbb{Z})$  by

$$f(v) = \sum_{i=1}^{m_v} [d_i(v)]^{1/p} [s_i(v)]^{(p-1)/p} e_{k_i(v)}.$$

We will break the proof the proof of the fact that  $f$  satisfies the required distortion bound into several claims.

**Claim 2.7.** For all  $v \in V$  and  $j \in \{1, \dots, m_v\}$ ,

$$s_i(v) \geq \frac{1}{4} \sum_{j=i}^{m_v} d_j(v).$$

*Proof.* This is where the fact that  $\chi$  is a  $\delta$ -strong coloring comes in. Indeed,

$$s_i(v) = \sum_{j=i}^{m_v} \max \left\{ 0, d_j(v) - \frac{\delta}{2} \sum_{h=i}^j d_h(v) \right\} \geq \sum_{\substack{j \in \{i, \dots, m_v\} \\ d_j(v) \geq \delta \sum_{h=i}^{m_v} d_h(v)}} \frac{d_j(v)}{2} \geq \frac{1}{4} \sum_{j=i}^{m_v} d_j(v).$$

□

**Claim 2.8.**  $\|f\|_{\text{Lip}} \leq [5 \log(3/\delta)]^{1/p}$ .

*Proof.* We need to show that for every edge  $(u, v) \in E$ ,  $\|f(u) - f(v)\|_p \leq 10[\log(1/\delta)]^{1/p}$ . Assume that  $v$  is further than  $u$  from the root of  $T$ . In this case  $k_1(u) = k_1(v), \dots, k_{m_u}(u) = k_{m_u}(v)$  and  $m_v \in \{m_u, m_u + 1\}$ . If  $m_v = m_u + 1$  then we denote for the sake of simplicity  $d_{m_v}(u) = s_{m_v}(u) = 0$ . With this notation we have that

$$\begin{aligned} \|f(u) - f(v)\|_p^p &= \sum_{i=1}^{m_v} \left| [d_i(u)]^{1/p} [s_i(u)]^{(p-1)/p} - [d_i(v)]^{1/p} [s_i(v)]^{(p-1)/p} \right|^p \\ &= \sum_{i=1}^{m_v-1} d_i(v) \left| [s_i(u)]^{(p-1)/p} - [s_i(v)]^{(p-1)/p} \right|^p + \\ &\quad \left| [d_{m_v}(u)]^{1/p} [s_{m_v}(u)]^{(p-1)/p} - [d_{m_v}(v)]^{1/p} [s_{m_v}(v)]^{(p-1)/p} \right|^p. \end{aligned}$$

Note that by our definitions,  $s_{m_v}(u) = d_{m_v}(u)$  and  $s_{m_v}(v) = d_{m_v}(v)$ . Thus

$$\begin{aligned} \|f(u) - f(v)\|_p^p &= \sum_{i=1}^{m_v-1} d_i(v) \left| [s_i(u)]^{(p-1)/p} - [s_i(v)]^{(p-1)/p} \right|^p + |d_{m_v}(u) - d_{m_v}(v)|^p \\ &\leq \sum_{i=1}^{m_v-1} d_i(v) \left| [s_i(u)]^{(p-1)/p} - [s_i(v)]^{(p-1)/p} \right|^p + [d_T(u, v)]^p. \end{aligned} \quad (9)$$

Observe that for all  $i \in \{1, \dots, m_v - 1\}$ ,  $s_i(v) \geq s_i(u)$ . Thus

$$\left| [s_i(u)]^{(p-1)/p} - [s_i(v)]^{(p-1)/p} \right| \leq \frac{|s_i(u) - s_i(v)|}{[s_i(v)]^{1/p}}, \quad (10)$$

where we used the elementary inequality  $y^\alpha - x^\alpha \leq \frac{y-x}{y^{1-\alpha}}$ , which is valid for all  $y \geq x > 0$  and  $\alpha \in (0, 1)$ .

Observe that for every  $i \leq m_v - 1$ ,

$$s_i(v) - s_i(u) = \max \left\{ 0, d_{m_v}(v) - \frac{\delta}{2} \sum_{h=i}^{m_v} d_h(v) \right\} - \max \left\{ 0, d_{m_v}(u) - \frac{\delta}{2} \sum_{h=i}^{m_v} d_h(u) \right\} \leq d_T(u, v). \quad (11)$$

Thus, combining (10) and (11) we see that

$$\begin{aligned}
\sum_{i=1}^{m_v-1} d_i(v) \left| [s_i(u)]^{(p-1)/p} - [s_i(v)]^{(p-1)/p} \right|^p &\leq \sum_{i=1}^{m_v-1} d_i(v) \cdot \frac{|s_i(u) - s_i(v)|^p}{s_i(v)} \\
&\leq [d_T(u, v)]^p \cdot \sum_{\substack{i \in \{1, \dots, m_v-1\} \\ s_i(u) \neq s_i(v)}} \frac{d_i(v)}{s_i(v)} \\
&\leq 4[d_T(u, v)]^p \cdot \sum_{\substack{i \in \{1, \dots, m_v-1\} \\ s_i(u) \neq s_i(v)}} \frac{d_i(v)}{\sum_{j=i}^{m_v} d_j(v)}, \quad (12)
\end{aligned}$$

where in the last line we used Claim 2.7.

Observe that for every  $x_1, \dots, x_k > 0$ ,

$$\sum_{i=1}^k \frac{x_i}{x_i + x_{i+1} + \dots + x_k + 1} \leq \sum_{i=k}^1 \int_{x_{i+1} + \dots + x_k}^{x_i + \dots + x_k} \frac{dt}{t+1} = \int_0^{x_1 + \dots + x_k} \frac{dt}{t+1} = \log(x_1 + \dots + x_k + 1)$$

Thus

$$\begin{aligned}
\sum_{\substack{i \in \{1, \dots, m_v-1\} \\ s_i(u) \neq s_i(v)}} \frac{d_i(v)}{\sum_{j=i}^{m_v} d_j(v)} &= \sum_{\substack{i \in \{1, \dots, m_v-1\} \\ s_i(u) \neq s_i(v)}} \frac{d_i(v)/d_{m_v}(v)}{\left( \sum_{j=i}^{m_v-1} d_j(v)/d_{m_v}(v) \right) + 1} \\
&\leq \log \left( 1 + \frac{1}{d_{m_v}(v)} \sum_{\substack{i \in \{1, \dots, m_v-1\} \\ s_i(u) \neq s_i(v)}} d_i(v) \right). \quad (13)
\end{aligned}$$

Let  $i$  be the smallest integer in  $\{1, \dots, m_v - 1\}$  such that  $s_i(u) \neq s_i(v)$ . Then by the definition of  $s_i(\cdot)$ ,

$$d_{m_v}(v) > \frac{\delta}{2} \sum_{h=i}^{m_v} d_h(v).$$

It follows that

$$\log \left( 1 + \frac{1}{d_{m_v}(v)} \sum_{\substack{i \in \{1, \dots, m_v-1\} \\ s_i(u) \neq s_i(v)}} d_i(v) \right) \leq \log \left( 1 + \frac{1}{d_{m_v}(v)} \sum_{h=i}^{m_v} d_h(v) \right) \leq \log \left( 1 + \frac{2}{\delta} \right).$$

Plugging this bound into (13), and using (12) and (9), we get that

$$\|f(u) - f(v)\|_p \leq \left[ 4 \log \left( 1 + \frac{2}{\delta} \right) + 1 \right]^{1/p} \cdot d_T(u, v) \leq [5 \log(3/\delta)]^{1/p} \cdot d_T(u, v).$$

□

Our final claim bounds  $\|f^{-1}\|_{\text{Lip}}$ .

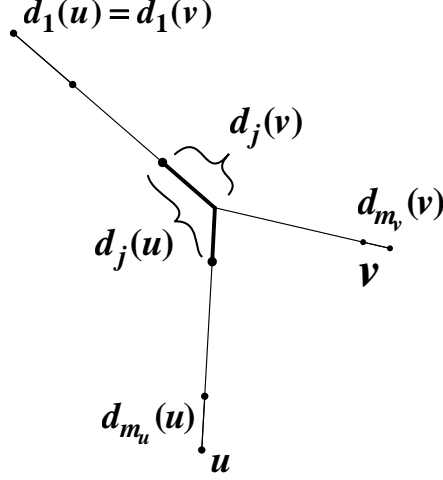


Figure 2: A schematic description of the location of  $u$  and  $v$  in the tree  $T$ . The bold segment corresponds to the color class  $k_j(u) = k_j(v)$ .

**Claim 2.9.** *The embedding  $f$  is invertible, and  $\|f^{-1}\|_{\text{Lip}} \leq 48$ .*

*Proof.* Fix  $u, v \in V$ ,  $u \neq v$ , and let  $j$  be the integer satisfying  $k_1(u) = k_1(v), \dots, k_j(u) = k_j(v)$  and  $k_{j+1}(u) \neq k_{j+1}(v)$ . It follows that  $d_1(u) = d_1(v), \dots, d_{j-1}(u) = d_{j-1}(v)$ , and we may assume without loss of generality that  $d_j(u) \geq d_j(v)$ . With this notation (see Figure 2.1 below),

$$d_T(u, v) = d_j(u) - d_j(v) + \sum_{i=j+1}^{m_u} d_i(u) + \sum_{i=j+1}^{m_v} d_i(v). \quad (14)$$

On the other hand,

$$\begin{aligned} \|f(u) - f(v)\|_p^p &\geq \left| [d_j(u)]^{1/p} [s_j(u)]^{(p-1)/p} - [d_j(v)]^{1/p} [s_j(v)]^{(p-1)/p} \right|^p + \\ &\quad \sum_{i=j+1}^{m_u} d_i(u) [s_i(u)]^{p-1} + \sum_{i=j+1}^{m_v} d_i(v) [s_i(v)]^{p-1}. \end{aligned} \quad (15)$$

Using Claim 2.7 we see that

$$\begin{aligned} \sum_{i=j+1}^{m_u} d_i(u) [s_i(u)]^{p-1} &\geq \frac{1}{4^{p-1}} \sum_{i=j+1}^{m_u} d_i(u) \left( \sum_{h=i}^{m_u} d_h(u) \right)^{p-1} \\ &\geq \frac{1}{4^{p-1}} \sum_{i=j+1}^{m_u} \int_{d_{i+1}(u) + \dots + d_{m_u}(u)}^{d_i(u) + \dots + d_{m_u}(u)} t^{p-1} dt \\ &= \frac{1}{4^{p-1}} \int_0^{d_{j+1}(u) + \dots + d_{m_u}(u)} t^{p-1} dt \\ &= \frac{1}{p4^{p-1}} \cdot \left( \sum_{i=j+1}^{m_u} d_i(u) \right)^p. \end{aligned} \quad (16)$$

Similarly,

$$\sum_{i=j+1}^{m_v} d_i(v)[s_i(v)]^{p-1} \geq \frac{1}{p4^{p-1}} \cdot \left( \sum_{i=j+1}^{m_v} d_i(v) \right)^p. \quad (17)$$

We now consider two cases:

**Case 1.**  $\frac{d_j(u)-d_j(v)}{2} \leq \sum_{i=j+1}^{m_v} d_i(v)$ . In this case, using (14) we see that

$$\begin{aligned} [d_T(u, v)]^p &\leq 3^p \left( \sum_{i=j+1}^{m_u} d_i(u) + \sum_{i=j+1}^{m_v} d_i(v) \right)^p \\ &\leq 3^p \cdot 2^{p-1} \left( \sum_{i=j+1}^{m_u} d_i(u) \right)^p + 3^p \cdot 2^{p-1} \left( \sum_{i=j+1}^{m_v} d_i(v) \right)^p \\ &\leq p4^{p-1} \cdot 3^p \cdot 2^{p-1} \|f(u) - f(v)\|_p^p, \end{aligned}$$

where in the last inequality we plugged the bounds (16) and (17) into (15). Thus we get that

$$\|f(u) - f(v)\|_p \geq \frac{1}{48} \cdot d_T(u, v),$$

as required.

**Case 2.**  $\frac{d_j(u)-d_j(v)}{2} > \sum_{i=j+1}^{m_v} d_i(v)$ . In this case we observe that

$$s_j(u) = \sum_{i=j}^{m_u} \max \left\{ 0, d_i(u) - \frac{\delta}{2} \sum_{h=j}^i d_h(u) \right\} \geq \left( 1 - \frac{\delta}{2} \right) d_j(u),$$

and similarly

$$s_j(v) \leq \left( 1 - \frac{\delta}{2} \right) d_j(v) + \sum_{i=j+1}^{m_v} d_i(v) \leq \left( 1 - \frac{\delta}{2} \right) d_j(v) + \frac{d_j(u) - d_j(v)}{2}.$$

Thus

$$\begin{aligned} [d_j(u)]^{1/p} [s_j(u)]^{(p-1)/p} - [d_j(v)]^{1/p} [s_j(v)]^{(p-1)/p} &\geq \left( 1 - \frac{\delta}{2} \right)^{(p-1)/p} d_j(u) - \\ &\quad \left( 1 - \frac{\delta}{2} \right)^{(p-1)/p} d_j(v) \cdot \left( 1 + \frac{d_j(u) - d_j(v)}{(2-\delta)d_j(v)} \right)^{(p-1)/p} \\ &\geq \left( 1 - \frac{\delta}{2} \right)^{(p-1)/p} d_j(u) - \left( 1 - \frac{\delta}{2} \right)^{(p-1)/p} d_j(v) \cdot \left( 1 + \frac{d_j(u) - d_j(v)}{(2-\delta)d_j(v)} \right) \\ &= \left( 1 - \frac{\delta}{2} \right)^{(p-1)/p} \cdot \frac{1-\delta}{2-\delta} \cdot [d_j(u) - d_j(v)] \\ &\geq \frac{d_j(u) - d_j(v)}{4}. \end{aligned}$$

where we used the fact that  $\delta \leq \frac{1}{2}$ . Using (15) and the bounds (16) and (17), it follows that

$$\begin{aligned} \|f(u) - f(v)\|_p^p &\geq \frac{1}{4^p} [d_j(u) - d_j(v)]^p + \frac{1}{p4^{p-1}} \cdot \left( \sum_{i=j+1}^{m_u} d_i(u) \right)^p + \frac{1}{p4^{p-1}} \cdot \left( \sum_{i=j+1}^{m_v} d_i(v) \right)^p \\ &\geq \frac{1}{p4^p \cdot 3^{p-1}} \left( d_j(u) - d_j(v) + \sum_{i=j+1}^{m_u} d_i(u) + \sum_{i=j+1}^{m_v} d_i(v) \right)^p \\ &\geq \frac{1}{24^p} \cdot [d_T(u, v)]^p. \end{aligned}$$

□

Claim 2.9, together with Claim 2.8, concludes the proof of Theorem 2.6. □

## 2.2 Relating coloring bounds to the containment of large binary trees

The following theorem, in conjunction with Theorem 2.6 and Lemma 2.5, implies Theorem 2.1. If one is concerned with simply giving *some* upper bound on  $c_p(T)$  in terms of  $\mathcal{B}_T(c)$ , then it suffices to combine the following theorem with Lemma 2.3.

**Theorem 2.10.** *For every weighted tree  $T = (V, E)$  and every  $c > 1$ ,*

$$\mathcal{B}_T(c) \geq \frac{c-1}{250c} \cdot \frac{1}{\varepsilon^*(T)}.$$

*Proof.* We start by introducing some notation. For a vertex  $v \in V$  we denote by  $\mathcal{C}(v)$  the set of all children of  $v$  in  $T$ . Given  $u \in \mathcal{C}(v)$  we denote by  $T_u = (V_u, E_u)$  the subtree rooted at  $u$ . We also let  $F_u$  denote the tree  $F_u = (V_u \cup \{v\}, E_u \cup \{(v, u)\})$ , i.e.  $F_u$  is  $T_u$  plus the “incoming” edge  $(v, u)$ .

Recall that  $B_k = (V_k, E_k)$  is the complete binary tree of height  $k$ . Let  $r_k$  be the root of  $B_k$ , and define an auxiliary tree  $M_k$  by  $M_k = (V_k \cup \{m_k\}, E_k \cup \{(m_k, r_k)\})$  (i.e.  $M_k$  is  $B_k$  with an extra incoming edge). Given a connected subtree  $H$  of  $T$  rooted at  $r_H$ , we shall say that  $H$  *admits a copy of  $M_k$  at scale  $j$*  if there exists a one-to-one mapping  $f : M_k \rightarrow H$  such that

1.  $f(m_k) = r_H$
2.  $\|f\|_{\text{Lip}} \leq \frac{9c}{c-1} \cdot 4^j$  and  $\|f^{-1}\|_{\text{Lip}} \leq \frac{c-1}{9 \cdot 4^j}$  (thus in particular  $\text{dist}(f) \leq c$ ).

We define

$$\mu_j(H) = \max \{k \in \mathbb{N} : H \text{ admits a copy of } M_k \text{ at scale } j\},$$

or  $\mu_j(H) = -1$  if no such  $k$  exists.

We shall now define a function  $g : V \rightarrow \mathbb{Z} \cup \{\infty\}$  and a coloring  $\chi : E \rightarrow \mathbb{Z}$ . These mappings will be constructed by induction as follows. We start by setting  $g(r) = \infty$ . Assume inductively that the construction is done so that whenever  $v \in V$  is such that  $g(v)$  is defined, if  $u$  is a vertex on the path  $P(v)$  then  $g(u)$  has already been defined, and for every edge  $e \in E$  incident with  $v$ ,  $\chi(e)$  has been defined.

Let  $v \in V$  be a vertex closest to the root  $r$  for which  $g(v)$  hasn't yet been defined. Then, by our assumption, for every  $e \in P(v)$ ,  $\chi(e)$  has been defined, and for every vertex  $u$  other than  $v$  lying

on the path  $P(v)$ ,  $g(u)$  has been defined. Let  $\beta_\chi(v) \subseteq V$  denote the set of *breakpoints* of  $\chi$  in  $P(v)$ , i.e. the set of vertices on  $P(v)$  for which the incoming and outgoing edges have distinct colors (for convenience, in what follows we shall also consider the root  $r$  as a breakpoint of  $\chi$ ). We define

$$g(v) = \max \left\{ j \in \mathbb{Z} : \forall u \in \beta_\chi(v), d_T(u, v) \geq 4^{\min\{g(u), j\}} \right\}.$$

Having defined  $g(v)$  we choose one of its children  $w \in \mathcal{C}(v)$  for which

$$\mu_{g(v)}(F_w) = \max_{z \in \mathcal{C}(v)} \mu_{g(v)}(F_z).$$

Letting  $u$  be the father of  $v$  on the path  $P(v)$ , we set  $\chi(v, w) = \chi(u, v)$ , and we assign arbitrary new (i.e. which haven't been used before) distinct colors to each of the edges  $\{(v, z)\}_{z \in \mathcal{C}(v) \setminus \{w\}}$ . In other words, given the "scale"  $j = g(v)$  we order the children of  $v$  according to the size of the copy of  $M_k$  which they admit beneath them at scale  $j$ . We then continue coloring with the color  $\chi(u, v)$  the path  $P(v)$  along the edge joining  $v$  and its child which admits the largest  $M_k$  at scale  $j$ , and color the remaining edges incident with  $v$  by arbitrary distinct new colors.

This definition clearly results in a monotone coloring  $\chi$ . To motivate this somewhat complicated construction, we shall now prove some of the crucial properties of  $\chi$  and  $g$  which will be used later.

**Claim 2.11.** *Let  $P$  be any monotone path in  $T$ , and let  $(b_1, b_2, \dots, b_m)$  be the sequence of breakpoints along  $P$  ordered down the tree (i.e. in increasing distance from the root). Assume that  $j \in \mathbb{Z}$  satisfies for every  $i \in \{2, \dots, m\}$ ,  $d_T(b_i, b_{i-1}) \leq 4^j$ , and assume also that  $d_T(b_1, b_m) \geq \frac{30c}{c-1} \cdot 4^j$ . Then there exists a subsequence of the indices  $1 \leq i_1 < i_2 < \dots < i_k \leq m$  such that*

1.  $k \geq \frac{c-1}{20c \cdot 4^j} \cdot d_T(b_1, b_m)$ .
2. For every  $s \in \{1, \dots, k\}$  we have  $g(b_{i_s}) = j$ .
3. For every  $s \in \{1, \dots, k-1\}$  we have  $\frac{9}{c-1} \cdot 4^j \leq d_T(b_{i_{s+1}}, b_{i_s}) \leq \frac{9c}{c-1} \cdot 4^j$ .

*Proof.* We shall show that if  $i \in \{1, \dots, m\}$  is such that  $d_T(b_i, b_m) > \frac{4^{j+1}}{3}$  then there exists an index  $t \in \{1, \dots, m\}$  with  $g(b_t) = j$  and  $d_T(b_t, b_i) \leq 4^{j+1}$ . Assuming this fact for the moment, we conclude the proof as follows. Let  $i_1$  be the smallest integer in  $\{2, \dots, m\}$  such that  $g(b_{i_1}) = j$ . Then  $d_T(b_{i_1}, b_1) \leq 4^{j+1}$ . Assuming we defined  $i_1 < i_2 < \dots < i_s$ , if  $d_T(b_{i_s}, b_m) \leq \frac{9c}{c-1} \cdot 4^j$  we stop the construction, and otherwise we let  $t$  be the smallest integer bigger than  $i_s$  such that  $d_T(b_t, b_{i_s}) \geq \frac{4c+5}{c-1} \cdot 4^j$ . Since  $d_T(b_{t-1}, b_{i_s}) < \frac{4c+5}{c-1} \cdot 4^j$ , we know that  $d_T(b_t, b_{i_s}) < \frac{4c+5}{c-1} \cdot 4^j + 4^j$ . Thus  $d_T(b_t, b_m) > d_T(b_{i_s}, b_m) - \frac{4c+5}{c-1} \cdot 4^j - 4^j > \frac{4^{j+1}}{3}$  (because we are assuming that  $d_T(b_{i_s}, b_m) < \frac{9c}{c-1} \cdot 4^j$ ). So, there exists  $i_{s+1} \in \{1, \dots, m\}$  such that  $g(b_{i_{s+1}}) = j$  and  $d_T(b_{i_{s+1}}, b_t) \leq 4^{j+1}$ . Since by construction  $d_T(b_t, b_{i_s}) \geq \frac{4c+5}{c-1} \cdot 4^j > 4^{j+1}$  we deduce that  $i_{s+1} > i_s$  and

$$\frac{9}{c-1} \cdot 4^j \leq d_T(b_t, b_{i_s}) - d_T(b_{i_{s+1}}, b_t) \leq d_T(b_{i_{s+1}}, b_{i_s}) \leq d_T(b_{i_{s+1}}, b_t) + d_T(b_t, b_{i_s}) \leq \frac{9c}{c-1} \cdot 4^j.$$

This construction terminates after  $k$  steps, in which case we have that

$$d_T(b_1, b_m) = d_T(b_1, b_{i_1}) + \sum_{s=1}^{k-1} d_T(b_{i_s}, b_{i_{s+1}}) + d_T(b_{i_k}, b_m) \leq 4^{j+1} + (k-1) \cdot \frac{9c}{c-1} \cdot 4^j + \frac{9c}{c-1} \cdot 4^j.$$

Since  $d_T(b_1, b_m) \geq \frac{30c}{c-1} \cdot 4^j$ , this implies the required result.

It remains to show that if  $i \in \{1, \dots, m\}$  is such that  $d_T(b_i, b_m) > \frac{4^{j+1}}{3}$  then there exists  $t \in \{1, \dots, m\}$  with  $g(b_t) = j$  and  $d_T(b_t, b_i) \leq 4^{j+1}$ . We first claim that for every  $i \in \{1, \dots, m\}$  there is a breakpoint  $w \in \beta_\chi(b_i)$  with  $g(w) \geq j$  and  $d_T(w, b_i) < \frac{4^{j+1}}{3}$ . Indeed, if  $g(b_i) \geq j$  then there is nothing to prove, so assume that  $g(b_i) < j$ . By the definition of  $g$  there exists a breakpoint  $w_1 \in \beta_\chi(b_i)$  such that

$$4^{\min\{g(w_1), g(b_i)\}} \leq d_T(w_1, b_i) < 4^{\min\{g(w_1), g(b_i)+1\}}.$$

Thus necessarily  $g(w_1) \geq g(b_i) + 1$  and  $d_T(w_1, b_i) < 4^{g(b_i)+1} < 4^j$ . If  $g(b_i) + 1 \geq j$  then we are done by taking  $w = w_1$ . Otherwise, continuing in this manner we find a breakpoint  $w_2 \in \beta_\chi(w_1) \subseteq \beta_\chi(b_i)$  with  $g(w_2) \geq g(w_1) + 1 \geq g(b_i) + 2$  and  $d_T(w_2, w_1) < 4^{g(w_1)+1}$ . This procedure terminates when we find a sequence  $b_i = w_0, w_1, w_2, \dots, w_t$  with  $g(w_t) \geq j$ ,  $g(w_{t-1}) < j$ , and for every  $0 \leq s \leq t-1$ ,  $g(w_{s+1}) \geq g(w_s) + 1$  and  $d_T(w_{s+1}, w_s) < 4^{g(w_s)+1}$ . Thus

$$d_T(b_i, w_t) = \sum_{s=0}^{t-1} d_T(w_{s+1}, w_s) < \sum_{s=0}^{t-1} 4^{g(w_s)+1} < \sum_{s=-\infty}^j 4^s = \frac{4^{j+1}}{3}.$$

Now, assume that  $d_T(b_i, b_m) > \frac{4^{j+1}}{3}$ . Let  $s$  be the largest integer in  $\{i+1, \dots, m\}$  such that  $d_T(b_s, b_i) \leq \frac{4^{j+1}}{3}$  (such an integer  $s$  exists since  $d_T(b_i, b_{i+1}) \leq 4^j$ ). Then  $\frac{4^{j+1}}{3} < d_T(b_{s+1}, b_i) \leq \frac{4^{j+1}}{3} + 4^j$ . By the previous argument there is a break point  $w \in \beta_\chi(b_{s+1})$  with  $g(w) \geq j$  and  $d_T(w, b_{s+1}) < \frac{4^{j+1}}{3}$ . This implies that  $w = b_t$  for some  $t \in \{i+1, \dots, s+1\}$ , and  $d_T(b_i, b_t) \leq \frac{4^{j+1}}{3} + 4^j$ .

We proved that as long as  $b_i$  satisfies  $d_T(b_i, b_m) > \frac{4^{j+1}}{3}$ , there are  $1 \leq t \leq i \leq s \leq m$  such that  $g(b_s) \geq j$ ,  $g(b_t) \geq j$ , and  $d_T(b_t, b_i) \leq \frac{4^{j+1}}{3}$ ,  $d_T(b_s, b_i) \leq \frac{4^{j+1}}{3} + 4^j$ . Thus, by the definition of  $g$ ,

$$4^{\min\{g(b_s), g(b_t)\}} \leq d_T(b_s, b_t) \leq \frac{2 \cdot 4^{j+1}}{3} + 4^j < 4^{j+1}.$$

It follows that either  $g(b_s) = j$  or  $g(b_t) = j$ , as required.  $\square$

To conclude the proof of Theorem 2.10 we may assume that  $\varepsilon^*(T) < \frac{c-1}{240c}$ , since otherwise the assertion of Theorem 2.10 is trivial. Fix  $\frac{c-1}{240c} > \varepsilon > \varepsilon^*(T)$ . By the definition of  $\varepsilon^*(T)$ , the coloring  $\chi$  constructed above is not  $\varepsilon$ -good. Thus, there exist two vertices  $u, v \in V$  such that the path  $P(u, v)$  does not contain a monochromatic segment of length at least  $\varepsilon d_T(u, v)$ . We may assume without loss of generality that  $u$  is an ancestor of  $v$ , and let  $(b_1, b_2, \dots, b_m)$  be the sequence of breakpoints along this path, enumerated down the tree (i.e. from  $u$  to  $v$ , not necessarily including  $u$  or  $v$ ). Denoting  $D = d_T(u, v)$  we have that  $d_T(u, b_1), d_T(v, b_m), d_T(b_i, b_{i+1}) \leq \varepsilon D$  for all  $i \in \{1, \dots, m-1\}$ . Fix  $j \in \mathbb{Z}$  such that  $\varepsilon D \leq 4^j \leq 4\varepsilon D$ . This choice implies that  $d_T(b_i, b_{i+1}) \leq 4^j$  and  $d_T(b_1, b_m) \geq (1-2\varepsilon)D \geq \frac{1-2\varepsilon}{4\varepsilon} \cdot 4^j \geq \frac{30c}{c-1} \cdot 4^j$ . By Claim 2.11 there is an integer  $k \geq \frac{(c-1)(1-2\varepsilon)D}{20c \cdot 4^j} \geq \frac{c-1}{250c} \cdot \frac{1}{\varepsilon} + 2$  (using the upper bound on  $\varepsilon$ ) and a sequence of breakpoints  $s_1, \dots, s_k$  on the path  $P(u, v)$  (ordered down the tree) such that  $g(s_1) = \dots = g(s_k) = j$  and for  $i \in \{1, \dots, k-1\}$ ,  $\frac{9}{c-1} \cdot 4^j \leq d_T(s_i, s_{i+1}) \leq \frac{9c}{c-1} \cdot 4^j$ .

The proof of Theorem 2.10 will be complete once we show that  $\mathcal{B}_T(c) \geq k-2$ . For  $i \in \{1, \dots, k\}$  let  $t_i$  be the child of  $s_i$  along the path  $P(u, v)$ . We will prove by reverse induction on  $i \in \{1, \dots, k-1\}$



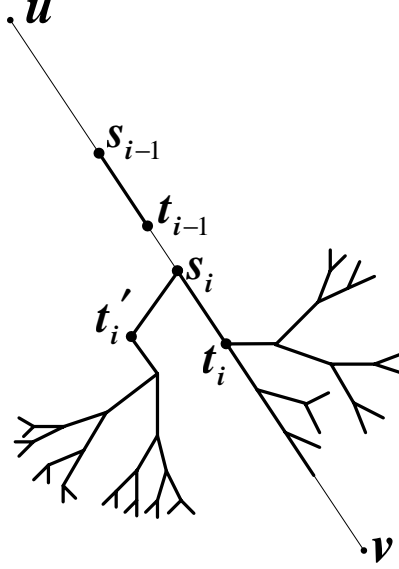


Figure 3: A schematic description of the gluing procedure in the inductive step. Because  $s_i$  was a breakpoint it must have two copies of  $M_{k-i-1}$  at scale  $j$  below it.

that  $\mu_j(F_{t_i}) \geq k - i - 1$ , implying the required result. The base case is true, i.e.  $\mu_j(F_{t_{k-1}}) \geq 0$ , since the pair  $(s_{k-1}, s_k)$  constitutes a copy of  $M_0$  at scale  $j$ .

Assuming that  $\mu_j(F_{t_i}) \geq k - i - 1$  we shall prove that  $\mu_j(F_{t_{i-1}}) \geq k - i$ . Since  $s_i$  was a breakpoint, the construction of  $\chi$  implies that there must be a child  $t'_i$  of  $s_i$ , other than  $t_i$ , for which  $\mu_j(F_{t'_i}) > \mu_j(F_{t_i}) \geq k - i - 1$ . Thus, there exist one to one mappings  $f, f' : M_{k-i-1} \rightarrow T$  such that  $f(m_{k-i-1}) = f'(m_{k-i-1}) = s_i$ ,  $f(M_{k-i-1}) \subseteq F_{t_i}$ ,  $f'(M_{k-i-1}) \subseteq F_{t'_i}$ ,  $\|f\|_{\text{Lip}}, \|f'\|_{\text{Lip}} \leq \frac{9c}{c-1} \cdot 4^j$ , and  $\|f^{-1}\|_{\text{Lip}}, \|(f')^{-1}\|_{\text{Lip}} \leq \frac{c-1}{9 \cdot 4^j}$ . Thinking of  $M_{k-i}$  as two disjoint copies of  $M_{k-i-1}$ , joined at the root  $m_{k-i}$ , we may “glue”  $f$  and  $f'$  to an embedding  $\bar{f}$  of  $M_{k-i}$  by setting  $\bar{f}(m_{k-i}) = s_{i-1}$ . Since  $\frac{9}{c-1} \cdot 4^j \leq d_T(s_i, s_{i-1}) \leq \frac{9c}{c-1} \cdot 4^j$ , this results in an embedding at scale  $j$  of  $M_{k-i}$  into  $F_{t_{i-1}}$ , as required (see Figure 3).  $\square$

### 2.3 Embedding into finite-dimensional spaces

We recall that the *doubling constant*  $\lambda(X)$  of a metric space  $X$  is the infimal value of  $\lambda$  for which every ball in  $X$  can be covered by  $\lambda$  balls of half the radius. If  $S \subseteq X$  is a  $\delta$ -separated set in  $X$ , then a standard observation is that  $|S| \leq \lambda(X)^{O(\text{diam}(S)/\delta)}$ . This section is devoted to a simpler proof of the following theorem of Gupta, Krauthgamer, and Lee originally proved in [17]. (We stress that the only results we need for this section are Lemma 2.3 and Theorem 2.10.)

**Theorem 2.12** ([17]). *A tree metric  $T$  embeds into a finite-dimensional Euclidean space if and only if  $\lambda(T) < \infty$ . In other words, every doubling tree  $T$  admits a  $D$ -embedding into  $\mathbb{R}^k$  with  $D, k$  depending only on  $\lambda(T)$ .*

Let  $T = (V, E)$  be a weighted, rooted tree. Note that the “only if” part of Theorem 2.12 is trivial. In order to prove the remaining implication we need a coloring notion weaker than  $\varepsilon$ -good. Let  $\chi : E \rightarrow \mathbb{Z}$  be a coloring of the edges of  $T$  which is *not necessarily monotone*. We will say

that  $\chi$  is  $\varepsilon$ -reasonable if the following holds for every  $u, v \in V$ . Let  $w = \mathbf{lca}(u, v)$ , and recall that  $P(w, u), P(w, v)$  denote the paths from  $w$  to  $u$  and  $v$ , respectively. Then there should exist a color  $c \in \mathbb{Z}$  for which

$$\left| \sum_{e \in P(w, u): \chi(e)=c} \ell(e) - \sum_{e \in P(w, v): \chi(e)=c} \ell(e) \right| \geq \varepsilon d_T(u, v). \quad (18)$$

Since a reasonable coloring is not necessarily monotone, it is possible to construct such colorings where  $\chi^{-1}(E)$  is finite even though  $T$  might be infinite. The number of colors used, i.e.  $|\chi^{-1}(E)|$ , controls the dimension of the embedding from Lemma 2.3.

**Lemma 2.13.** *Let  $T = (V, E)$  be a weighted tree, and suppose that  $T$  admits an  $\varepsilon$ -reasonable coloring for some  $\varepsilon > 0$ . Then  $T$  embeds in  $\mathbb{R}^k$  (equipped, e.g. with the  $L_2$  norm) with distortion  $O(1/\varepsilon)$ , and  $k = |\chi^{-1}(E)|$ .*

*Proof.* Let  $\chi : E \rightarrow \mathbb{Z}$  be an  $\varepsilon$ -reasonable coloring of  $T$ . We use the embedding  $f : V \rightarrow \ell_2$  of Lemma 2.3. In particular, it is easy to check that the definition of the embedding does not require  $\chi$  to be monotone. Observe that  $\text{Im}(f)$  lies naturally in  $\text{span}\{e_k : k \in \chi^{-1}(E)\}$ , and thus we may assume that  $f : V \rightarrow \mathbb{R}^k$  with  $k = |\chi^{-1}(E)|$ .

From the proof of Lemma 2.3, we conclude that  $\|f\|_{\text{Lip}} \leq 1$ , and thus we need only consider  $\|f^{-1}\|_{\text{Lip}}$ . But it is easy to see that condition (18) suffices to obtain a similar lower bound in equation (4) of Lemma 2.3.  $\square$

We note that the dependence of  $k$  on  $|\chi^{-1}(E)|$  in the above lemma can be improved to  $k = O(\log |\chi^{-1}(E)|)$  using a “nearly-orthogonal” set of vectors instead of the orthonormal set  $\{e_k\}_{k \in \mathbb{Z}}$ . We refer to [17] for details.

Now, clearly  $\mathcal{B}_T(2) \leq O(\log \lambda(T))$  since  $\lambda(B_m) = 2^{\Theta(m)}$ , hence  $\varepsilon^*(T) \geq 1/O(\log \lambda(T))$  using Theorem 2.10. In light of Lemma 2.13 and the preceding remark, the following result completes the proof of Theorem 2.12. (Note that we can assume  $T$  finite by compactness—a tree embeds into a finite-dimensional Euclidean space if and only if every finite subset embeds with uniformly bounded distortion).

**Theorem 2.14.** *Let  $T = (V, E)$  be a finite, weighted tree. If  $T$  admits an  $\varepsilon$ -good coloring, then it also admits an  $O(\varepsilon)$ -reasonable coloring with only  $\lambda(T)^{(1/\varepsilon)^{O(1/\varepsilon)}}$  colors.*

*Proof.* We will say that a monotone coloring  $\chi : E \rightarrow \mathbb{Z}$  is *regular* if the following holds: For every maximal monochromatic segment  $s = \{e_1, e_2, \dots, e_k\} \subseteq E$  (with edges ordered down the tree), and for every  $1 \leq i \leq k$ , we have  $\ell(e_{i+1}) \leq 2 \sum_{j=1}^i \ell(e_j)$ .

**Lemma 2.15.** *If a finite tree  $T$  admits an  $\varepsilon$ -good coloring, then  $T$  admits an  $O(\varepsilon)$ -good regular coloring.*

*Proof.* Let  $T = (V, E)$  be a rooted tree, and let  $\chi_0 : E \rightarrow \mathbb{Z}$  be an  $\varepsilon$ -good coloring of  $T$ . Suppose that some monochromatic segment  $s = \{e_1, e_2, \dots, e_k\} \subseteq E$  violates the regularity condition. Let  $i \in [k]$  be the smallest index for which  $\ell(e_{i+1}) > 2 \sum_{j=1}^i \ell(e_j)$ . We derive a new coloring  $\chi_1 : E \rightarrow \mathbb{Z}$  by coloring the edges  $e_1, \dots, e_i$  with a new unused color  $c \in \mathbb{Z}$ , i.e.  $\chi_1(e) = c$  if  $e = e_j$  for some  $1 \leq j \leq i$  and  $\chi_1(e) = \chi_0(e)$  otherwise. Continue this process inductively until the resulting coloring  $\chi' : E \rightarrow \mathbb{Z}$  is regular. This process terminates because  $T$  is finite. It remains to show that  $\chi'$  is  $O(\varepsilon)$ -good.

To this end, let  $s = \{e_1, \dots, e_k\} \subseteq E$  be a maximal monochromatic segment according to  $\chi_0$ , and let  $s_1, s_2, \dots, s_m \subseteq s$  be the maximal monochromatic segments of  $s$  according to  $\chi'$ , ordered down the tree. By construction, we have

$$\ell(s_m) \geq 2\ell(s_{m-1}) \geq \ell(s_{m-1}) + 2\ell(s_{m-2}) \geq \dots \geq \ell(s_1) + \dots + \ell(s_{m-1}),$$

hence  $\ell(s_m) \geq \frac{1}{2}\ell(s)$ . It follows that  $\chi'$  is a regular  $\varepsilon/2$ -good coloring of  $T$ .  $\square$

Let  $T$  be a rooted tree, and let  $\chi : E \rightarrow \mathbb{Z}$  be an  $\varepsilon$ -good coloring of  $T$ . Using the preceding lemma, we may assume that  $\chi$  is regular. Let  $\mathcal{C}$  be the set of color classes. We will think of segments  $s \in \mathcal{C}$  sometimes as a subset of edges and sometimes a subset of vertices (the endpoints and internal vertices of the segments), depending on the context. In everything that follows, we will assume that for  $s \neq s' \in \mathcal{C}$ , we have  $\text{diam}(s) \neq \text{diam}(s')$ . This is without loss of generality by applying arbitrarily small perturbations to  $T$ . (Alternatively, we could fix a total order on segments of equal diameter, but this would add unnecessary notation to the proof.)

For every segment  $s \in \mathcal{C}$ , we define  $p(s)$  as the vertex of  $s$  which is closest to the root. For every  $s_0 \in \mathcal{C}$  and  $K > 0$  we define a relative length function

$$\text{length}_{s_0}(s; K) = \max \left\{ \text{diam}(P(p(s), x)) : x \in s \cap B_T(p(s_0), K \cdot \text{diam}(s_0)) \right\}, \quad (19)$$

where we take  $s \in \mathcal{C}$ , and we set  $\text{length}_{s_0}(s; K) = 0$  in case the maximum is empty. In words, this is how long the segment  $s \in \mathcal{C}$  “looks” from  $p(s_0)$ , where the “view” is restricted to a ball of radius  $K \cdot \text{diam}(s_0)$ . It is important to note that even when  $s \not\subseteq B_T(p(s_0), K \cdot \text{diam}(s_0))$ , one might have  $0 < \text{length}_{s_0}(s; K) \ll K \cdot \text{diam}(s_0)$  since  $T$  is not necessarily an  $\mathbb{R}$ -tree.

Now we define carefully a *directed* graph  $G_{\mathcal{C}} = (\mathcal{C}, E_{\mathcal{C}})$ . The adjacency relationship on  $G_{\mathcal{C}}$  will be the key in producing an  $O(\varepsilon)$ -reasonable coloring. We put

$$(s, s') \in E_{\mathcal{C}} \iff \text{length}_s(s'; K) > \text{diam}(s), \quad (20)$$

for some constant  $K \geq 6$  to be chosen later. Observe, in particular, that  $(s, s') \in E_{\mathcal{C}} \implies \text{diam}(s') > \text{diam}(s)$ . We will argue that the undirected graph  $\hat{G}_{\mathcal{C}}$  which results from ignoring the edge directions in  $G_{\mathcal{C}}$  has its chromatic number bounded solely by a function of  $\lambda(T)$ . We accomplish this with the following sequence of lemmas. (This step is non-trivial since  $\hat{G}_{\mathcal{C}}$  does not have bounded degree.)

**Lemma 2.16.** *For every  $s \in \mathcal{C}$ , the out-degree is bounded, i.e.*

$$|\{s' \in \mathcal{C} : (s, s') \in E_{\mathcal{C}}\}| \leq \lambda(T)^{O(K)}.$$

*Proof.* Fix  $s \in \mathcal{C}$ . For every  $s' \in \mathcal{C}$  with  $(s, s') \in E_{\mathcal{C}}$ , let  $x_{s'} \in s'$  be the node achieving the maximum in (19). If the maximum does not exist then  $\text{length}_s(s'; K) = 0$ , hence  $(s, s') \notin E_{\mathcal{C}}$ . By definition,  $d_T(p(s), x_{s'}) \leq K \cdot \text{diam}(s)$ . Furthermore,  $d_T(p(s'), x_{s'}) = \text{length}_s(s'; K) > \text{diam}(s)$ . It follows that the set  $X_s = \{x_{s'} : (s, s') \in E_{\mathcal{C}}\}$  is  $\text{diam}(s)$ -separated. Since  $X_s \subseteq B_T(p(s), K \cdot \text{diam}(s))$ , the doubling property implies that

$$|\{s' \in \mathcal{C} : (s, s') \in E_{\mathcal{C}}\}| = |X_s| \leq \lambda(T)^{O(K)}.$$

$\square$

For any undirected graph  $G = (V_G, E_G)$  and  $v \in V_G$ , we define  $N(v)$  to be the set of neighbors of  $v$  in  $G$ , we let  $\deg(v) = |N(v)|$  and  $\deg_S(v) = |N(v) \cap S|$  for  $S \subseteq V_G$ . The next result is well-known.

**Lemma 2.17.** *Let  $G = (V_G, E_G)$  be any finite, undirected graph. Let  $k \in \mathbb{N}$  and let  $\pi : V_G \rightarrow \{1, 2, \dots, n\}$  be a permutation. We denote  $\pi_j = \{v \in V_G : \pi(v) \leq j\}$ . If, for every  $j = 1, 2, \dots, n$ , we have*

$$\deg_{\pi_{j-1}}(\pi^{-1}(j)) \leq k,$$

*then the chromatic number of  $G$  is at most  $k + 1$ .*

*Proof.* The proof follows by inductively coloring the elements  $\pi^{-1}(1), \pi^{-1}(2), \dots, \pi^{-1}(n)$  in order. If we have a palette of  $k + 1$  colors, then since  $\deg_{\pi_{j-1}}(\pi^{-1}(j)) \leq k$ , we can always choose a new color for  $\pi^{-1}(j)$  that doesn't conflict with any already colored vertex in  $\pi_{j-1}$ .  $\square$

**Corollary 2.18.** *If  $\hat{G}_C$  is the undirected version of  $G_C$ , then the chromatic number of  $\hat{G}_C$  is bounded by  $\lambda(T)^{O(K)}$ .*

*Proof.* Let  $\pi : \mathcal{C} \rightarrow \{1, 2, \dots, |\mathcal{C}|\}$  be any permutation for which  $\text{diam}(\pi(j)) \geq \text{diam}(\pi(j + 1))$  for  $1 \leq j \leq |\mathcal{C}| - 1$  (i.e. the diameters of the segments decrease monotonically). Then combining Lemmas 2.16 and the fact that  $(s, s') \in E_C \implies \text{diam}(s') > \text{diam}(s)$  shows that for  $j = 1, 2, \dots, n$ ,

$$\deg_{\pi_{j-1}}(\pi^{-1}(j)) \leq \lambda(T)^{O(K)}.$$

Applying Lemma 2.17 completes the proof.  $\square$

Now let  $\chi_C : \mathcal{C} \rightarrow [k]$  be a proper coloring of  $\hat{G}_C$  using only  $k = \lambda(T)^{O(K)}$  colors. We are done as soon as we show that  $\chi_C$  is an  $O(\varepsilon)$ -reasonable edge-coloring of  $T$  (where we consider  $\chi_C$  as a coloring of  $E$  in the obvious way) for some choice of  $K \leq (1/\varepsilon)^{O(1/\varepsilon)}$ .

**Lemma 2.19.** *Suppose that for  $s \neq s' \in \mathcal{C}$ , we have*

$$\text{diam}(s \cap P(u, v)), \text{diam}(s' \cap P(u, v)) \geq \frac{d_T(u, v)}{K/2 - 1},$$

*where  $u, v \in T$ . Then  $\chi_C(s) \neq \chi_C(s')$ .*

*Proof.* Assume that  $\text{diam}(s') > \text{diam}(s)$ , and let  $x$  be the bottom-most point of  $s' \cap P(u, v)$ . Then

$$d_T(p(s), x) \leq \text{diam}(s) + d_T(u, v) \leq (1 + (K/2 - 1)) \text{diam}(s) \leq \frac{K}{2} \cdot \text{diam}(s). \quad (21)$$

In this case,  $\text{length}_s(s'; K) \geq \text{diam}(P(p(s'), x))$ , hence if  $\text{diam}(P(p(s'), x)) > \text{diam}(s)$ , we have  $(s, s') \in E_C$ , which finishes the proof of the lemma.

So we may assume that  $\text{diam}(P(p(s'), x)) \leq \text{diam}(s)$ . We claim that in this case,  $\text{length}_s(s'; K) > \text{diam}(s)$  using the regularity of  $\chi$ . Let  $y \in s'$  be such that  $d_T(x, y) \leq \frac{K}{2} \text{diam}(s)$ , and for which  $d_T(p(s'), y)$  is maximal. If  $d_T(p(s'), y) > \text{diam}(s)$ , then we are done since by (21), we have  $d_T(p(s), y) \leq K \cdot \text{diam}(s)$ , implying  $\text{length}_s(s'; K) \geq d_T(p(s'), y) > \text{diam}(s)$ . Hence we may assume that  $d_T(p(s'), y) \leq \text{diam}(s)$ . In this case, since  $\text{diam}(s') > \text{diam}(s)$ , there exists an edge  $(y, y')$  with  $y' \in s'$  and  $d_T(p(s'), y') > \frac{K}{2} \cdot \text{diam}(s) > 3 \cdot \text{diam}(s)$ . But this implies that  $\ell(y, y') > 2 \cdot d_T(p(s'), y)$ , which contradicts the regularity of  $\chi$ .

It follows that  $\text{length}_s(s'; K) > \text{diam}(s)$ , which again implies  $(s, s') \in E_C$ .  $\square$

Now fix  $u, v \in V$  and  $w = \mathbf{lca}(u, v)$ , and suppose that  $d_T(w, u) \geq d_T(w, v)$ . Since  $\chi$  is an  $\varepsilon$ -good coloring, there exists a maximal monochromatic segment (with respect to  $\chi$ )  $s \subseteq E$  for which  $\text{diam}(s \cap P(w, u)) \geq \varepsilon d_T(w, u) \geq (\varepsilon/2)d_T(u, v)$ . Now set  $K = 4 \left(\frac{2}{\varepsilon}\right)^{1+2/\varepsilon}$ . Applying Lemma 2.19, we see that for any  $s' \subseteq E$  with  $\chi_C(s) = \chi_C(s')$ , we have  $\text{diam}(s' \cap P(w, v)) \leq \left(\frac{\varepsilon}{2}\right)^{1+2/\varepsilon}$ . But now line (5) of Lemma 2.5 implies that segments of this length can cover at most an  $\varepsilon/2$ -fraction of  $P(w, v)$  (since  $\chi$  is an  $\varepsilon$ -good coloring), which is at most an  $\varepsilon/4$ -fraction of  $P(u, v)$ . It follows that  $\chi_C$  is a  $\delta$ -reasonable coloring for  $\delta = \frac{\varepsilon}{2} - \frac{\varepsilon}{4} \geq \frac{\varepsilon}{4}$ , completing the proof.  $\square$

### 3 Markov convexity and distortion lower bounds

In this section we study Markov convexity, and show how it can be used to prove several distortion lower bounds. In particular, we will discuss the connection between Markov convexity and uniform convexity in Banach spaces, and we will prove that Theorem 2.1 is optimal.

#### 3.1 Markov convexity in Banach spaces

We start by showing that Hilbert space is Markov 2-convex. This has essentially been proved by Bourgain in [6]. We give the following proof here because the argument is extendable to the case of  $p \neq 2$ . We refer also to [25] for another variant of Bourgain's proof.

**Lemma 3.1.** *For every  $x_0, \dots, x_{2^m} \in L_2$ ,*

$$\sum_{i=1}^{2^m} \|x_i - x_{i-1}\|_2^2 = \frac{\|x_{2^m} - x_0\|_2^2}{2^m} + \sum_{k=1}^m \frac{1}{2^k} \sum_{j=1}^{2^{m-k}} \|x_{j2^k} - 2x_{(2j-1)2^{k-1}} + x_{(j-1)2^k}\|_2^2. \quad (22)$$

*Proof.* Let  $\mathcal{F}_n$  be the  $\sigma$ -algebra of subsets of  $[0, 1]$  generated by the intervals  $\left\{I_j^n := \left[\frac{j-1}{2^n}, \frac{j}{2^n}\right]\right\}_{j=1}^{2^n}$ . Define  $\varphi : [0, 1] \rightarrow L_2$  by  $\varphi \equiv x_j - x_{j-1}$  on  $I_j^m$ . Set  $\varphi_j = \mathbb{E}(\varphi | \mathcal{F}_j)$ , where the expectation is with respect to the Lebesgue measure on  $[0, 1]$ . In other words, for every  $j \in \{1, \dots, 2^k\}$  and  $t \in I_j^k$

$$\varphi_k(t) = \frac{1}{2^{m-k}} \sum_{\ell=2^{m-k}(j-1)+1}^{j2^{m-k}} (x_\ell - x_{\ell-1}) = \frac{x_{j2^{m-k}} - x_{(j-1)2^{m-k}}}{2^{m-k}}.$$

Since the sequence  $\{\varphi_k - \varphi_{k-1}\}_{k=1}^m$  is a martingale difference sequence, and  $\varphi_0$  is constant, the functions  $\varphi_0, \varphi_1 - \varphi_0, \varphi_2 - \varphi_1, \dots, \varphi_m - \varphi_{m-1}$  are orthogonal (in the Hilbert space  $L_2(L_2)$ ). Thus

$$\mathbb{E}\|\varphi_m\|_2^2 = \mathbb{E}\|\varphi_0\|_2^2 + \sum_{k=1}^m \mathbb{E}\|\varphi_k - \varphi_{k-1}\|_2^2.$$

This is precisely the required identity.  $\square$

We can remove the dyadic bias from (22) by averaging over shifts.

**Corollary 3.2.** For every  $x_0, \dots, x_{2^m} \in L_2$ ,

$$\sum_{i=1}^{2^m} \|x_i - x_{i-1}\|_2^2 \geq \frac{1}{2} \frac{\|x_{2^m} - x_0\|_2^2}{2^{2m}} + \frac{1}{2} \sum_{k=1}^m 2^{-2k} \sum_{t=1}^{2^m} \|x_t - 2x_{t-2^{k-1}} + x_{t-2^k}\|_2^2,$$

where, by convention,  $x_j = x_0$  for  $j \leq 0$ .

*Proof.* First, consider the sequence of length  $3 \cdot 2^m - 2$ ,

$$x_0, x_0, \dots, x_0, x_0, x_1, \dots, x_{2^m}, x_{2^m}, x_{2^m}, \dots, x_{2^m},$$

which is the original sequence with  $2^m - 1$  copies of  $x_0$  and  $x_{2^m}$  appended to the front and back, respectively. Call this sequence  $\{y_j\}_{j=1}^{3 \cdot 2^m - 2}$ . Now average the equality (22) over all  $2^{m+1}$  contiguous subsequences of length  $2^m + 1$ , i.e.  $\{y_i, y_{i+1}, \dots, y_{i+2^m}\}$  for  $i = 1, \dots, 2^{m+1}$ . By counting terms, this yields the desired result.  $\square$

**Theorem 3.3.** Hilbert space is Markov 2-convex. In fact,  $\Pi_2(L_2) \leq 4$ .

*Proof.* Let  $\{X_t\}_{t=0}^\infty$  be a Markov chain on a state space  $\Omega$ , and take  $f : \Omega \rightarrow L_2$ . By Corollary 3.2,

$$\begin{aligned} \sum_{t=1}^{2^m} \mathbb{E} \|f(X_t) - f(X_{t-1})\|_2^2 &\geq \frac{1}{2} \sum_{k=1}^m 2^{-2k} \sum_{t=1}^{2^m} \mathbb{E} \|f(X_t) - 2f(X_{t-2^{k-1}}) + f(X_{t-2^k})\|_2^2 \\ &\quad + \frac{1}{2} \frac{\mathbb{E} \|f(X_0) - f(X_{2^m})\|_2^2}{2^{2m}}, \end{aligned}$$

where by convention we set  $X_t = X_0$  for  $t \leq 0$ .

Observe that for every two i.i.d. random vectors  $Z, Z' \in L_2$ , and every constant  $a \in L_2$ ,  $\mathbb{E} \|Z - Z'\|_2^2 \leq 2\mathbb{E} \|Z - a\|_2^2$ . Thus, using the fact that conditioned on  $\mathcal{X} = (X_0, \dots, X_{t-2^{k-1}})$  the random vectors  $f(X_t)$  and  $f(\tilde{X}_t(t - 2^{k-1}))$  are i.i.d., we see that

$$\begin{aligned} \mathbb{E} \|f(X_t) - f(\tilde{X}_t(t - 2^{k-1}))\|_2^2 &= \mathbb{E} \left( \mathbb{E} \left( \|f(X_t) - f(\tilde{X}_t(t - 2^{k-1}))\|_2^2 \middle| \mathcal{X} \right) \right) \\ &\leq 2 \mathbb{E} \|f(X_t) - 2f(X_{t-2^{k-1}}) + f(X_{t-2^k})\|_2^2. \end{aligned}$$

Likewise,  $\mathbb{E} \|f(X_0) - f(\tilde{X}_{2^m}(0))\|_2^2 \leq 2 \mathbb{E} \|f(X_0) - f(X_{2^m})\|_2^2$ . It follows that

$$\begin{aligned} \sum_{t=1}^{2^m} \mathbb{E} \|f(X_t) - f(X_{t-1})\|_2^2 &\geq \frac{1}{4} \sum_{k=1}^{m+1} 2^{-2k} \sum_{t=1}^{2^m} \mathbb{E} \|f(X_t) - f(\tilde{X}_t(t - 2^{k-1}))\|_2^2, \\ &= \frac{1}{16} \sum_{k=0}^m 2^{-2k} \sum_{t=1}^{2^m} \mathbb{E} \|f(X_t) - f(\tilde{X}_t(t - 2^k))\|_2^2, \end{aligned}$$

completing the proof.  $\square$

**Remark 3.1.** The above argument can be generalized to prove that  $p$ -convex Banach spaces are Markov  $p$ -convex. Recall that a Banach space  $X$  is said to be  $p$ -convex with constant  $K$  (see [3]) if for every  $x, y \in X$ ,

$$2\|x\|^p + \frac{2}{K^p} \|y\|^p \leq \|x + y\|^p + \|x - y\|^p.$$

The least such constant  $K$  is denoted  $K_p(X)$ .

We claim that for every Banach space  $X$ ,

$$\Pi_p(X) \leq 2^{(p-1)/p} (2^{p-1} - 1)^{1/p} \cdot K_p(X) \leq 4K_p(X).$$

Indeed, repeating the argument of Lemma 3.1, we replace the use of orthogonality by Pisier's inequality [33] to get that (see the argument in [2] for the constant used below),

$$(2^{p-1} - 1)[K_p(X)]^p \mathbb{E} \|\varphi_m\|_X^p \geq \mathbb{E} \|\varphi_0\|_X^p + \sum_{k=1}^m \mathbb{E} \|\varphi_k - \varphi_{k-1}\|_X^p.$$

As in the proof of Theorem 3.3 (and using the notation there) this shows that

$$\begin{aligned} (2^{p-1} - 1)[K_p(X)]^p \sum_{t=1}^{2^m} \mathbb{E} \|f(X_t) - f(X_{t-1})\|_X^p &\geq \\ \frac{1}{2} \sum_{k=1}^m 2^{-pk} \sum_{t=1}^{2^m} \mathbb{E} \|f(X_t) - 2f(X_{t-2^{k-1}}) + f(X_{t-2^k})\|_X^p &+ \frac{1}{2} \frac{\mathbb{E} \|f(X_0) - f(X_{2^m})\|_X^p}{2^{pm}}. \end{aligned}$$

Since for every two i.i.d. random vectors  $Z, Z' \in X$ , and every constant  $a \in X$ , we have that  $\mathbb{E} \|Z - Z'\|_X^p \leq 2^{p-1} \mathbb{E} \|Z - a\|_X^p$  (this fact follows from a straightforward interpolation argument), we conclude exactly as in the proof of Theorem 3.3.

We mention some partial converses to Remark 3.1.

**Corollary 3.4.** *Let  $X$  be an infinite dimensional Banach space. Then  $\Pi_p(X) < \infty$  implies that  $X$  is superreflexive and has cotype  $q$  for every  $q > p$ .*

*Proof.* Let  $q_X = \inf\{q : X \text{ has cotype } q\}$ . By the Maurey-Pisier [29] theorem,  $X$  contains copies of  $\ell_{q_X}^n$  with distortion uniformly bounded in  $n$ . By Bourgain's embedding of trees into  $\ell_{q_X}$  [6], this implies that  $c_X(B_m) = O((\log m)^{1/q_X})$ . From Bourgain's lower bound [6], or alternatively Claim 3.7 below, we deduce that  $q_X \leq p$ , as required. The fact that  $X$  is superreflexive follows from Bourgain's characterization of superreflexivity [6].  $\square$

**Corollary 3.5.** *Let  $X$  be a Banach lattice with  $\Pi_p(X) < \infty$ . Then for every  $q > p$ ,  $X$  admits a  $q$ -convex equivalent norm.*

*Proof.* This is a direct consequence of a theorem of Figiel [13] (see [22], page 100) which says that a Banach lattice with cotype  $q$  and non-trivial type can be renormed to be  $q$ -convex ( $X$  has non-trivial type since it is superreflexive).  $\square$

## 3.2 Distortion lower bounds

We can now use the discrepancy between  $\Pi_p(X)$  and  $\Pi_p(Y)$  to prove distortion lower bounds for embeddings between the two spaces.

**Lemma 3.6.** *Let  $(X, d_X), (Y, d_Y)$  be metric spaces, then for every  $p < \infty$ , we have*

$$c_Y(X) \geq \frac{\Pi_p(X)}{\Pi_p(Y)}.$$

*Proof.* Fix  $\Pi > \Pi_p(Y)$ . Let  $g : X \rightarrow Y$  be a bi-Lipschitz map, let  $\{X_t\}_{t=0}^\infty$  be a Markov chain with state space  $\Omega$ , and let  $f : \Omega \rightarrow X$ . Then

$$\begin{aligned}
& \sum_{k=0}^m \sum_{t=1}^{2^m} \frac{\mathbb{E} \left[ d_X \left( f(X_t), f(\tilde{X}_t(t-2^k)) \right)^p \right]}{2^{kp}} \\
& \leq \|g^{-1}\|_{\text{Lip}}^p \cdot \sum_{k=0}^m \sum_{t=1}^{2^m} \frac{\mathbb{E} \left[ d_Y \left( g(f(X_t)), g(f(\tilde{X}_t(t-2^k))) \right)^p \right]}{2^{kp}} \\
& \leq \|g^{-1}\|_{\text{Lip}}^p \cdot \Pi^p \cdot \sum_{t=1}^{2^m} \mathbb{E} [d_Y(g(f(X_t)), g(f(X_{t-1})))^p] \\
& \leq \|g\|_{\text{Lip}}^p \cdot \|g^{-1}\|_{\text{Lip}}^p \cdot \Pi^p \cdot \sum_{t=1}^{2^m} \mathbb{E} [d_Y(f(X_t), f(X_{t-1}))^p].
\end{aligned}$$

It follows that  $\Pi_p(X) \leq c_Y(X) \cdot \Pi_p(Y)$ , as required.  $\square$

As a warm up to the more involved lower bounds that will follow, we show how Markov convexity can be used to prove Bourgain's theorem for complete binary trees.

**Claim 3.7.** *For every  $m \in \mathbb{N}$ , we have  $\Pi_p(B_{2^m}) \geq 2^{1-\frac{2}{p}} \cdot m^{\frac{1}{p}}$ .*

*Proof.* Let  $\{X_t\}_{t=0}^\infty$  be the forward random walk on  $B_{2^m}$  (which goes left/right with probability  $\frac{1}{2}$ ), starting from the root, with the leaves as absorbing states. Then

$$\sum_{t=1}^{2^m} \mathbb{E} [d_{B_{2^m}}(X_t, X_{t-1})^p] \leq 2^m.$$

Moreover, in the forward random walk, after splitting at time  $r \leq 2^m$  with probability at least  $\frac{1}{2}$  two independent walks will accumulate distance which is at least twice the number of steps (until a leaf is encountered). Thus

$$\sum_{k=0}^m \sum_{t=1}^{2^m} \frac{\mathbb{E} \left[ d_{B_{2^m}} \left( X_t, \tilde{X}_t(t-2^k) \right)^p \right]}{2^{kp}} \geq \sum_{k=0}^m \sum_{t=1}^{2^m-2^k} \frac{1}{2^{kp}} \cdot \frac{1}{2} \cdot 2^{(k+1)p} \geq 2^{p-2} \cdot m \cdot 2^m.$$

The claim follows.  $\square$

Since  $L_p$  is Markov  $\max\{2, p\}$ -convex for every  $p > 1$ , combining Claim 3.7 with Lemma 3.6 recovers Bourgain's result [6], i.e. for every  $p > 1$ , we have  $c_p(B_k) \geq \Omega \left( (\log k)^{\min\{\frac{1}{2}, \frac{1}{p}\}} \right)$ .

**Simple random walks with positive speed.** In fact, the proof of Claim 3.7 applies in more general situations where a random walk has positive speed. We consider some examples.

Let  $G = (V, E)$  be an infinite, vertex-transitive graph of bounded degree. Let  $\{X_t\}_{t=0}^\infty$  be a simple random walk on  $G$  starting from an arbitrary vertex. Denote by  $d_G$  the shortest path metric on  $G$ . One defines the *speed of the random walk* as the limit

$$\lim_{t \rightarrow \infty} \frac{\mathbb{E} d_G(X_0, X_t)}{t}.$$

Subadditivity implies that the limit above always exists.



**Lemma 3.8.** *If the speed of the simple random walk on a vertex-transitive graph  $G$  is at least  $s > 0$ , then  $\Pi_p(B_G(R)) = \Omega((\log R)^{1/p})$ , where  $B_G(R)$  denotes the ball of radius  $R$  in  $G$ . In particular for every  $p > 1$ ,*

$$c_p(B_G(R)) = \Omega\left((\log R)^{\min\{\frac{1}{2}, \frac{1}{p}\}}\right).$$

*Proof.* The proof is similar to that of Claim 3.7. One simply observes that for two independent simple random walks  $X_t, \tilde{X}_t$  started at the same point, we have

$$\lim_{t \rightarrow \infty} \frac{\mathbb{E} d_G(X_t, \tilde{X}_t)}{t} = \lim_{t \rightarrow \infty} \frac{\mathbb{E} d_G(X_0, X_{2t})}{t} \geq 2s > 0.$$

In particular, for  $t$  large enough, with constant probability we have  $d_G(X_t, \tilde{X}_t) = \Omega(t)$ .  $\square$

As an application, consider the lamplighter group over  $\mathbb{Z}^d$ . This is a group with elements  $(f, x)$ , where  $x \in \mathbb{Z}^d$ , and  $f : \mathbb{Z}^d \rightarrow \{0, 1\}$  with  $f(y) = 0$  for all but finitely many  $y \in \mathbb{Z}^d$ . Traditionally, one imagines a lamp placed at every element of  $\mathbb{Z}^d$ , where each lamp can either be on or off. In the pair  $(f, x)$ ,  $f$  denotes the settings of all the lamps, and  $x$  denotes the position of the lamplighter. Accordingly, the generating set consists of two types of moves.

1. The lamplighter can move to an adjacent vertex in  $\mathbb{Z}^d$ , i.e.  $(f, x) \mapsto (f, x')$  where  $x'$  is adjacent to  $x$  in the standard Cayley graph of  $\mathbb{Z}^d$  or
2. The lamplighter can turn on/off the lamp at  $x$ , i.e.  $(f, x) \mapsto (f', x)$  where  $f'(y) = f(y)$  for  $y \neq x$  and  $f'(x) = 1 - f(x)$ .

We will use  $L(\mathbb{Z}^d)$  to denote the associated group as well as the Cayley graph with the described generators. A result of Kaĭmanovich and Vershik [19] shows that the simple random walk on  $L(\mathbb{Z}^d)$  has positive speed for  $d > 2$ . Using Remark 3.1 and Lemma 3.8, we conclude:

**Corollary 3.9.** *For  $d > 2$ , the word metric on  $L(\mathbb{Z}^d)$  does not embed into any  $p$ -convex Banach space. In particular, if  $B_{L(\mathbb{Z}^d)}(R)$  denotes a ball of radius  $R$ , then for  $p > 1$ ,*

$$c_p\left(B_{L(\mathbb{Z}^d)}(R)\right) \geq \Omega\left((\log R)^{\min\{\frac{1}{2}, \frac{1}{p}\}}\right).$$

We remark that, by a theorem of Varopoulos [36], the simple random walk on the Cayley graph of a finitely-generated group has positive speed if and only if there exists a bounded, non-constant, harmonic function on the graph.

Finally, we consider the finite lamplighter groups over  $\mathbb{Z}_N = \mathbb{Z}/(N\mathbb{Z})$ , which we denote by  $L(\mathbb{Z}_N)$ . In this case, the simple random walk on  $L(\mathbb{Z}_N)$  does not have positive speed, but it is still possible to prove a distortion lower bound because the Markov chains in the definition of Markov convexity need not be reversible. In particular, consider the chain  $\{X_t\}_{t=0}^\infty$  defined as follows.  $X_0 = (f, 0)$  where  $f \equiv 0$ , i.e. all lamps are turned off. If  $X_t = (f, i)$ , then with probability  $\frac{1}{2}$ , we put  $X_{t+1} = (f, i+1)$ , and with probability  $\frac{1}{2}$  we put  $X_{t+1} = (f', i+1)$  where  $f'(i+1) = 1 - f(i+1)$ . Arguing essentially exactly as in Claim 3.7 for times  $t \leq N$ , we have the following.

**Proposition 3.10.** *For every  $p < \infty$ , we have  $\Pi_p(L(\mathbb{Z}_N)) \geq \Omega\left((\log N)^{\frac{1}{p}}\right)$ . In particular,  $c_p(L(\mathbb{Z}_N)) \geq \Omega\left((\log N)^{\min\{\frac{1}{2}, \frac{1}{p}\}}\right)$ .*

Proposition 3.10 can also be proved by exhibiting an embedding of a complete binary tree of depth  $\Theta(N)$  into  $L(\mathbb{Z}_N)$ —see [26].

### 3.3 Weak prototypes and Markov convexity

In this section we study the Markov convexity properties of a special class of trees called weak prototypes. These trees will play a central role in Section 4, where Theorem 1.4 is proved. We begin with some definitions (we continue using here the notation of Section 2.1).

In what follows, by a path metric  $P = (u_1, \dots, u_m)$  we simply mean a graph theoretical path from  $u_1$  to  $u_m$  with edges  $(u_1, u_2), (u_2, u_3), \dots, (u_{m-1}, u_m)$  and edge weights  $\{\ell(u_j, u_{j+1})\}_{j=1}^{m-1} \subseteq [0, \infty)$ . The length of  $P$ , denoted  $\ell(P)$ , is given by  $\ell(P) = \sum_{j=1}^{m-1} \ell(u_j, u_{j+1})$ . Given a monotone path  $P$  in  $T$ , and a set of vertices  $(v_1, \dots, v_m)$  on  $P$ , ordered down the tree and not necessarily containing all the vertices of  $T$  lying on  $P$ , we will call the path metric on  $(v_1, \dots, v_m)$  with the edge weights  $\{d_T(v_j, v_{j+1})\}_{j=1}^{m-1}$ , the path metric induced by  $T$  on  $(v_1, \dots, v_m)$ .

Given a path metric  $P = (u_1, \dots, u_m)$  and  $\varepsilon, \delta \in (0, 1)$  we shall say the path  $P$  is  $(\varepsilon, \delta)$ -weak if at least an  $\varepsilon$ -fraction of the length of  $P$  is composed of edges of length at most  $\delta\ell(P)$ , i.e.

$$\sum_{\substack{j \in \{1, \dots, m-1\} \\ \ell(u_j, u_{j+1}) \leq \delta\ell(P)}} \ell(u_j, u_{j+1}) \geq \varepsilon\ell(P) = \varepsilon \sum_{j=1}^{m-1} \ell(u_j, u_{j+1}).$$

A monotone path  $P(u, v)$  in  $T$  will be called *degree-2  $(\varepsilon, \delta)$ -weak* if the following condition holds true. Let  $(u_1, \dots, u_m)$  be the vertices of  $T$  on  $P$ , ordered down the tree, who have at least two children in  $T$ . Then we require that the path metric induced by  $T$  on  $(u, u_1, \dots, u_m, v)$  is  $(\varepsilon, \delta)$ -weak. In other words, call a monotone path  $P$  in  $T$  a *strait* if every vertex on  $P$  has exactly one child, except possibly for the initial and final vertices. Then  $P(u, v)$  is degree-2  $(\varepsilon, \delta)$ -weak if at least an  $\varepsilon$ -fraction of the length of  $P(u, v)$  is composed of maximal straits of length at most  $\delta d_T(u, v)$ .

**Definition 3.11.** Fix  $\varepsilon, \delta, R > 0$ . A tree  $T = (V, E)$  with edge lengths  $\ell : E \rightarrow (0, \infty)$  is called an  $(\varepsilon, \delta)$ -weak prototype with height ratio  $R$  if the following conditions are satisfied.

- Every non-leaf vertex of  $T$  has exactly one or two children.
- Every root-leaf path of  $T$  is degree-2  $(\varepsilon, \delta)$ -weak.
- If  $h$  is the length of the shortest root-leaf path in  $T$  and  $h'$  is the length of the longest root-leaf path in  $T$ , then  $h'/h \leq R$ .

#### 3.3.1 Markov convexity for unweighted weak prototypes

First, we will prove a lower bound on the Markov convexity constants of a special class of unweighted weak prototypes. Later, we will show that every weak prototype can be approximated by a weak prototype satisfying these conditions.

**Theorem 3.12.** Let  $(T, d_T)$  be an unweighted  $(\varepsilon, \delta)$ -weak prototype with height ratio 1 and height  $2^m$  for some  $m \in \mathbb{N}$ . Then for every  $p \geq 1$ , we have

$$\Pi_p(T) \geq \left( \frac{\varepsilon}{4} [\log_2(\varepsilon/\delta) - 4] \right)^{1/p}.$$

*Proof.* Let  $r$  be the root of  $T$ . Let  $\{X_t\}_{t=0}^\infty$  be the Markov chain on  $T$  defined as follows. Initially,  $X_0 = r$ . If  $X_t$  is a leaf node, then  $X_{t+1} = X_t$ , and otherwise  $X_{t+1}$  is a uniformly random child of  $X_t$ .

First, we have  $d_T(X_{t-1}, X_t) \leq 1$  for every  $t \geq 1$ . Thus it suffices to show that

$$\frac{1}{2^m} \sum_{k=0}^m \sum_{t=1}^{2^m} \frac{\mathbb{E} \left[ d_T \left( X_t, \tilde{X}_t(t-2^k) \right) \right]^p}{2^{kp}} \geq \frac{\varepsilon}{4} \left[ \log_2 \left( \frac{\varepsilon}{\delta} \right) - 4 \right].$$

Recall that a monotone path  $P$  in  $T$  is a *strait* if every node of  $P$  has exactly one child, except possibly for the initial and final nodes. Additionally, say that a node  $v \in T$  is a *branch point* if  $v$  has at least two children. Clearly the edges of every root-leaf path partition into maximal straights with branch points at the ends (except for the root and leaves). Let  $\mathcal{B}_k(t)$  be the event that the set  $\{X_t, X_{t+1}, \dots, X_{t+2^k-1}\}$  contains a branch point. Observe that whenever  $2^{k-1} \geq \delta 2^m$ , we have

$$\sum_{t=0}^{2^m-1} \Pr[\mathcal{B}_k(t)] \geq \sum_{t=0}^{2^m-1} \Pr \left[ X_t \text{ falls in a maximal strait of length at most } 2^{k-1} \right] \geq \varepsilon 2^m, \quad (23)$$

since every root-leaf path of  $T$  is degree-2  $(\varepsilon, \delta)$ -weak. Furthermore, if  $k \leq m$ , and  $t \leq 2^m - 2^k$ , then

$$\mathcal{B}_k(t) \text{ occurs} \implies \Pr \left[ d_T(X_{t+2^k}, \tilde{X}_{t+2^k}(t)) \geq 2^k \right] \geq \frac{1}{2}, \quad (24)$$

since upon hitting a branch point, the two chains will diverge with probability at least  $\frac{1}{2}$  for at least  $2^{k-1}$  additional steps.

We conclude that when  $2^{k-1} \geq \delta 2^m$  and  $k \leq m$ ,

$$\begin{aligned} \sum_{t=1}^{2^m} \mathbb{E} \left[ d \left( X_t, \tilde{X}_t(t-2^k) \right) \right]^p &\geq \sum_{t=0}^{2^m-2^k} \mathbb{E} \left[ d \left( X_{t+2^k}, \tilde{X}_{t+2^k}(t) \right) \right]^p \\ &\geq \sum_{t=0}^{2^m-2^k} \frac{1}{2} \cdot 2^{kp} \cdot \Pr[\mathcal{B}_k(t)] \end{aligned} \quad (25)$$

$$\geq 2^{kp-1} \cdot (\varepsilon 2^m - 2^k), \quad (26)$$

where in (25) we used (24), and in (26) we used (23) along with a correction term for boundary values of  $k$ . Therefore,

$$\begin{aligned} \frac{1}{2^m} \sum_{k=0}^m \sum_{t=1}^{2^m} \frac{\mathbb{E} \left[ d_T \left( X_t, \tilde{X}_t(t-2^k) \right) \right]^p}{2^{kp}} &\geq 2^{-(m+1)} \sum_{k \geq 1 + \log_2(\delta 2^m)}^m \max \left\{ 0, \varepsilon 2^m - 2^k \right\} \\ &\geq \frac{\varepsilon}{4} \left[ \log \left( \frac{\varepsilon}{\delta} \right) - 4 \right]. \end{aligned}$$

The proof of Theorem 3.12 is complete.  $\square$

### 3.3.2 Distortion bounds for weak prototypes

In this section, we show how pass from a finite tree  $T$  to a more well-behaved tree  $\tilde{T}$  such that  $c_p(\tilde{T}) = O(1) \cdot c_p(T)$  for every  $p \in [1, \infty)$ . We use this transformation to prove distortion lower bounds for arbitrary weak prototypes.

**Lemma 3.13.** *Let  $(T, d_T)$  be a finite, graph-theoretic metric tree, and let  $T_{\mathbb{R}}$  be the  $\mathbb{R}$ -tree that results from replacing every edge of  $e \in E(T)$  by a closed interval whose length is  $\text{length}(e)$ . Then for every  $p \in [1, \infty)$ , we have  $c_p(T_{\mathbb{R}}) \leq 5c_p(T)$ .*

*Proof.* Fix a root  $r$  of  $T$  (and, in particular, an orientation of the edges). Let  $f : T \rightarrow L_p$  be an embedding of  $T$ . Let  $\{\beta_{uv}\}_{uv \in E(T)} \subseteq L_p$  be a system of disjointly supported unit vectors each of whose support is also disjoint from the support of  $\text{Im}(f)$ . Denote a point  $x \in T_{\mathbb{R}}$  by  $x = (u, v, \eta)$ , where  $uv \in E(T)$ , we have  $d_T(r, u) \leq d_T(r, x) \leq d_T(r, v)$ , and  $d_T(x, u) = \eta \cdot d_T(u, v)$  for  $\eta \in [0, 1]$ . Assume that  $\|f\|_{\text{Lip}} = 1$ . We define an embedding  $g : T_{\mathbb{R}} \rightarrow L_p$  by

$$g(u, v, \eta) = (1 - \eta)f(u) + \eta f(v) + \eta d_T(u, v)\beta_{uv}.$$

Fix  $(u, v, \eta), (u', v', \eta') \in T_{\mathbb{R}}$ . If  $u'$  is not a descendant of  $u$  or vice-versa then

$$d_{T_{\mathbb{R}}}((u, v, \eta), (u', v', \eta')) = \eta d_T(u, v) + \eta' d_T(u', v') + d_T(u, u').$$

Thus

$$\begin{aligned} & \|g(u, v, \eta) - g(u', v', \eta')\|_p \\ & \leq \|g(u, v, \eta) - g(u, v, 0)\|_p + \|g(u, v, 0) - g(u', v', 0)\|_p + \|g(u', v', 0) - g(u', v', \eta')\|_p \\ & = \|\eta(f(v) - f(u)) + \eta d_T(u, v)\beta_{uv}\|_p + \|(1 - \eta)(f(u) - f(u'))\|_p \\ & \quad + \|\eta'(f(v') - f(u')) + \eta' d_T(u', v')\beta_{u'v'}\|_p \\ & = \eta (\|f(u) - f(v)\|_p^p + d_T(u, v)^p)^{1/p} + (1 - \eta)\|f(u) - f(u')\|_p \\ & \quad + \eta' (\|f(u') - f(v')\|_p^p + d_T(u', v')^p)^{1/p} \\ & \leq 2^{1/p}\eta d_T(u, v) + 2^{1/p}\eta' d_T(u', v') + (1 - \eta)d_T(u, u') \\ & \leq 2^{1/p} \cdot d_{T_{\mathbb{R}}}((u, v, \eta), (u', v', \eta')). \end{aligned}$$

If  $u'$  is a strict descendant of  $u$  then

$$d_{T_{\mathbb{R}}}((u, v, \eta), (u', v', \eta')) = (1 - \eta)d_T(u, v) + \eta' d_T(u', v') + d_T(v, u'), \quad (27)$$

and a similar reasoning shows that  $\|g(u, v, \eta) - g(u', v', \eta')\|_p \leq 2^{1/p} \cdot d_{T_{\mathbb{R}}}((u, v, \eta), (u', v', \eta'))$ . The case of  $u = u'$  is even simpler, so we have shown that  $\|g\|_{\text{Lip}} \leq 2^{1/p}\|f\|_{\text{Lip}}$ .

On the other hand, we will now show that  $\|g^{-1}\|_{\text{Lip}} \leq \frac{5}{2^{1/p}}$ . Assume first of all that  $u'$  is not a

descendant of  $u$  or vice-versa. Then

$$\begin{aligned}
& \|g(u, v, \eta) - g(u', v', \eta')\|_p^p \\
& \geq [\eta d_T(u, v)]^p + [\eta' d_T(u', v')]^p + \left| \|f(u) - f(u')\|_p - \eta \|f(u) - f(v)\|_p - \eta' \|f(u') - f(v')\|_p \right|^p \\
& \geq [\eta d_T(u, v)]^p + [\eta' d_T(u', v')]^p + \left| \frac{d_T(u, u')}{\|f^{-1}\|_{\text{Lip}}} - \eta d_T(u, v) - \eta' d_T(u', v') \right|^p \\
& \geq \frac{5}{2} \left[ \frac{1}{5} \left( \frac{d_T(u, u')}{\|f^{-1}\|_{\text{Lip}}} + \eta d_T(u, v) + \eta' d_T(u', v') \right) \right]^p \\
& \geq 2 \left[ \frac{1}{5} \left( \frac{d_{T_{\mathbb{R}}}((u, v, \eta), (u', v', \eta'))}{\|f^{-1}\|_{\text{Lip}}} \right) \right]^p,
\end{aligned} \tag{28}$$

Where in (28) we used the convexity of the function  $a \mapsto |a|^p$ , which implies that for all  $a, b, c \in \mathbb{R}$  we have  $|a|^p + |b|^p + |c|^p \geq \frac{5}{2} \left| \frac{2}{5}a + \frac{2}{5}b + \frac{1}{5}c \right|^p$ .

If  $u'$  is a strict descendant of  $u$  then  $d_{T_{\mathbb{R}}}((u, v, \eta), (u', v', \eta'))$  is given by (27). Denote this distance by  $D$ , and for the sake of simplicity write  $L = \|f^{-1}\|_{\text{Lip}}$ . Since

$$\|g(u, v, \eta) - g(u', v', \eta')\|_p^p \geq [\eta d_T(u, v)]^p + [\eta' d_T(u', v')]^p \geq 2^{1-p} [\eta d_T(u, v) + \eta' d_T(u', v')]^p,$$

we may assume that  $\eta d_T(u, v) + \eta' d_T(u', v') < \frac{2D}{5L}$ . In this case

$$\begin{aligned}
\|g(u, v, \eta) - g(u', v', \eta')\|_p & \geq \left\| (1 - \eta)f(u) + \eta f(v) - (1 - \eta')f(u') - \eta' f(v') \right\|_p \\
& \geq \|f(u) - f(u')\|_p - \eta \|f(u) - f(v)\|_p - \eta' \|f(u') - f(v')\|_p \\
& \geq \frac{d_T(u, u')}{L} - \eta d_T(u, v) - \eta' d_T(u', v') \\
& \geq \frac{D - \eta' d_T(u', v')}{L} - \frac{2D}{5L} \\
& \geq \frac{D}{L} - \frac{D}{5L^2} - \frac{2D}{5L} \\
& \geq \frac{D}{5L}.
\end{aligned}$$

The remaining case is when  $u = u'$  and  $v = v'$ . But then  $\|g(u, v, \eta) - g(u', v', \eta')\|_p = |\eta - \eta'| \cdot \|f(u) - f(v)\|_p$ , and the required lower bound is trivial.

We have thus proved that  $\|g\|_{\text{Lip}} \cdot \|g^{-1}\|_{\text{Lip}} \leq 5 \|f\|_{\text{Lip}} \cdot \|f^{-1}\|_{\text{Lip}}$ , as required.  $\square$

**Remark 3.2.** The above lemma does *not* hold if we allow ‘‘Steiner’’ nodes in the tree  $T$ . To observe this, consider the subset  $L \subseteq B_m$  of leaves of a complete binary tree of height  $m$ , and let  $r$  be the root of  $B_m$ . Then it is not difficult to see that  $c_2(L \cup \{r\}) \leq O(1)$  (independent of  $m$ ), while  $c_2(B_m) \rightarrow \infty$  by Bourgain’s theorem for  $B_m$  [6].

We now replace any weak prototype by an ‘‘equivalent’’ prototype with height ratio 1.

**Lemma 3.14.** *Let  $(T, d_T)$  be any finite metric tree. Then there exists a finite, unweighted metric tree  $(\tilde{T}, d_{\tilde{T}})$  with height  $2^m$  for some  $m \in \mathbb{N}$  such that  $c_p(\tilde{T}) \leq O(1) \cdot c_p(T)$  for any  $p \in [1, \infty)$ . Furthermore, if  $T$  is an  $(\varepsilon, \delta)$ -weak prototype with height ratio  $R$ , then  $\tilde{T}$  is an unweighted  $(\frac{\varepsilon}{2R}, \delta)$ -weak prototype with height ratio 1.*

*Proof.* Fix a root  $r$  of  $T$ . Since  $T$  is finite, by rescaling and paying an arbitrarily small distortion, we may assume that all edge lengths are integral. For every node  $x \in T$ , Let  $m = \lceil \log_2 \max_{x \in T} d_T(x, r) \rceil$ . We now define a tree  $T'$  as follows. For every leaf  $\ell \in T$ , define a new node  $\tilde{\ell}$ , and create a new edge  $(\ell, \tilde{\ell})$  of length  $2^m - d_T(r, \ell)$ . Thus the length of every root-leaf path in  $T'$  is exactly  $2^m$ . To see that  $c_p(T') = \Theta(1) \cdot c_p(T)$ , let  $f : T \rightarrow L_p$  be an embedding of  $T$ , and let  $\{\beta_\ell\} \subseteq L_p$  be a system of disjointly supported vectors each of whose support is disjoint to  $\text{Im}(f)$ . One can extend the embedding by defining  $f(\tilde{\ell}) = f(\ell) + d_{T'}(\ell, \tilde{\ell}) \cdot \beta_\ell$  so that  $c_p(T') \leq O(1) \cdot c_p(T)$ . Observe that if  $T$  had height ratio  $R$ , then the length of any root-leaf path in  $T'$  has increased by at most a factor  $2R$  over its previous length in  $T$ .

We pass from  $T'$  to  $T'_\mathbb{R}$  using Lemma 3.13, and then to  $\tilde{T}$  by simply taking the vertex set of  $\tilde{T}$  to be  $V(\tilde{T}) = \{v \in T' : d_{T'}(v, r) \in \mathbb{N}\}$ . We define  $d_{\tilde{T}}$  as the *unweighted* shortest path metric on  $\tilde{T}$ . then  $(\tilde{T}, d_{\tilde{T}})$  embeds isometrically into  $T'_\mathbb{R}$ . Hence  $c_p(\tilde{T}) = \Theta(1) \cdot c_p(T'_\mathbb{R}) = \Theta(1) \cdot c_p(T') = \Theta(1) \cdot c_p(T)$ . Furthermore, every root-leaf path in  $\tilde{T}$  has length precisely  $2^m$ . Finally, observe that if  $T$  was  $(\varepsilon, \delta)$ -weak with height ratio  $R$ , then  $\tilde{T}$  is an unweighted  $(\varepsilon/(2R), \delta)$ -weak prototype (because some root-leaf path from  $T$  might have increased by a factor of at most  $2R$ ) with height ratio 1.  $\square$

The following corollary follows from Theorem 3.12 and Lemma 3.14.

**Corollary 3.15.** *Let  $(T, d_T)$  be an  $(\varepsilon, \delta)$ -weak prototype with height ratio  $R$ , then for any  $p > 1$ ,*

$$c_p(T, d_T) \geq \Omega(1) \cdot c_p(\tilde{T}, d_{\tilde{T}}) \geq \Omega(1) \cdot \Pi_q(\tilde{T}, d_{\tilde{T}}) \geq \Omega(1) \cdot \left( \frac{\varepsilon}{R} \log \left( \frac{\varepsilon}{\delta R} \right) \right)^{1/q},$$

where  $q = \max\{2, p\}$  and  $\tilde{T}$  is the associated unweighted prototype from Lemma 3.14.

The corollary follows by applying Theorem 3.12 to  $\tilde{T}$  and using the relationship between Markov convexity and distortion from Lemma 3.6, along with the known Markov convexity of  $L_p$  spaces (Remark 3.1).

### 3.3.3 The Cantor trees

Recall that in Theorem 2.1, we showed that for any tree  $T$  and  $p \geq 1$ , we have, for every  $c > 1$ ,

$$c_p(T) \leq O(1) \left( \frac{c}{c-1} \cdot \mathcal{B}_T(c) \right)^{\min\{\frac{1}{p}, \frac{1}{2}\}}.$$

Here, we will show that this dependence on  $\mathcal{B}_T(c)$  cannot be improved by exhibiting a family  $\{C_i\}_{i=0}^\infty$  of metric trees with  $|C_i| \rightarrow \infty$  and such that for any fixed  $c > 1$ ,

$$c_p(C_i) \geq \Omega(1) \cdot \Pi_{\max\{2, p\}}(C_i) \geq \Omega(1) \cdot \mathcal{B}_{C_i}(c)^{\min\{\frac{1}{p}, \frac{1}{2}\}} \quad (29)$$

Let  $T$  be a rooted (unweighted) graph-theoretic tree. For any root leaf path  $P = \{v_0, v_1, \dots, v_m\}$ , we define the *downward degree sequence*  $d_\downarrow(P) = \{d_\downarrow(v_0), d_\downarrow(v_1), \dots, d_\downarrow(v_m)\}$  where  $d_\downarrow(v)$  is the number of children of  $v$  in  $T$ . We will say that  $T$  is a *spherically symmetric tree (SST)* if for any pair of root-leaf paths  $P, P'$  we have  $d_\downarrow(P) = d_\downarrow(P')$ . Clearly any such tree can be completely specified by giving the degree sequence of a root-leaf path (see Figure 4).

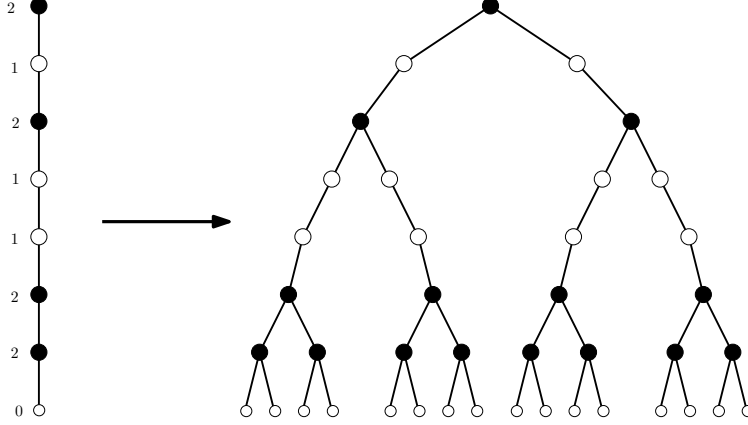


Figure 4: A downward degree sequence and the corresponding SST.

**Definition of the Cantor trees.** We now describe a family of downward degree sequences inductively. For two sequences  $S, S'$  we define  $S \otimes S'$  as their concatenation. For every  $i \in \mathbb{N}$ , we use  $\text{ones}(i) = \otimes^i \{1\}$  to denote a sequence of  $i$  ones. Now define  $S_0 = \{2\}$  and inductively

$$S_{i+1} = S_i \otimes \text{ones}(2^i - 1) \otimes S_i.$$

Hence the first few sequences are  $\{\{2\}, \{22\}, \{22122\}, \{2212211122122\}, \dots\}$ . To make these proper downward degree sequences, we define  $\tilde{S}_i$  to be  $S_i$  except with the last element changed from 2 to 0. Finally, we let  $C_i$  be the unique SST with downward degree sequence  $\tilde{S}_i$ . We call these *Cantor trees* because the patterns of 2's resemble finite approximations to the middle-thirds Cantor set. It is clear that  $\text{length}(S_i) = 2 \cdot \text{length}(S_{i-1}) + 2^{i-1} - 1 = i \cdot 2^{i-1} + 1$ , and that  $\log \log |C_i| = \Theta(i)$ . The next two lemmas are somewhat less obvious.

**Lemma 3.16.** *For every  $i \geq 1$ , the tree  $C_i$  is a  $(\frac{1}{2}, 2^{-i/3})$ -weak prototype.*

*Proof.* We need to show that every root-leaf path in  $C_i$  is degree-2  $(\frac{1}{2}, 2^{-i/3})$ -weak. Fix any such path  $P$ . It is easy to see that the maximal straits in  $P$  are given by consecutive sequences of 1's in the downward degree sequence of  $C_i$ : A sequence of  $k$  consecutive 1's refers to a strait of length  $k + 1$ . Therefore for every  $j \leq i - 1$ , there are  $2^{i-1-j}$  disjoint maximal straits of length  $2^j$  in  $P$ . The question becomes how small we need to choose  $m$  before

$$\frac{1}{2} \text{height}(C_i) = \frac{1}{2} (i \cdot 2^{i-1} + 2) \leq \sum_{j=m}^{i-1} 2^{i-1-j} \cdot 2^j = 2^{i-1} (i - m).$$

Clearly we must have  $m < i/2$ , hence  $C_i$  is a  $(\frac{1}{2}, \delta)$ -weak prototype for

$$\delta = \frac{2^{i/2}}{i \cdot 2^{i-1} + 1} \leq 2^{-i/3}.$$

□

Combining this with Theorem 3.12 yields the following.

**Corollary 3.17.** *For every  $p < \infty$ ,  $\Pi_p(C_i) \geq \Omega(i^{1/p})$ .*

The following claim completes the proof of (29).

**Claim 3.18.** *For every fixed  $c > 1$ ,  $\mathcal{B}_{C_i}(c) \leq O(i)$  as  $i \rightarrow \infty$ .*

*Proof.* The idea of the proof is simple: If the edges of  $B_m$  are mapped far apart in  $C_i$ , then we can use the diameter of  $C_i$  to upper bound the size of  $m$ . Otherwise, if the edges are mapped close together, then essentially the entire image of  $B_m$  must lie inside some copy of  $C_{i-1}$  in  $C_i$ . This is because there is a “buffer” of length  $2^{i-1}$  between copies of  $C_{i-1}$  in  $C_i$  which contains no branch points. An edge of  $B_m$  must stretch over this buffer if the image of  $B_m$  spans multiple copies of  $C_{i-1}$ .

For our induction, it will be easier to bound  $\mathcal{B}_{\tilde{C}_i}(c)$  for a slightly different family of trees  $\tilde{C}_i$ . Let  $\tilde{C}_i$  be the tree  $C_i$  with the following two additions:

1. We append a path  $H_i$  of length  $2^{i-2}$  to the root of  $C_i$ .
2. We append a path of length  $2^{i-2}$  to every leaf of  $C_i$ . We will use  $\mathcal{L} = \{L_i\}$  to refer to this family of paths.

Clearly  $\mathcal{B}_{C_i}(c) \leq \mathcal{B}_{\tilde{C}_i}(c)$ .

We may assume that  $i \geq 1$  is sufficiently large with respect to  $c$ . Let  $f : B_m \rightarrow \tilde{C}_i$  be a bi-Lipschitz embedding of  $B_m$  into  $\tilde{C}_i$  with distortion  $c = \|f\|_{\text{Lip}} \cdot \|f^{-1}\|_{\text{Lip}}$ . Assume, for the sake of contradiction, that  $m \geq 256 i \cdot c \log(c + 1)$ .

Clearly

$$\text{diam}(\tilde{C}_i) \geq \max_{u,v \in B_m} d_{\tilde{C}_i}(f(u), f(v)) \geq \frac{2m}{\|f^{-1}\|_{\text{Lip}}} \geq \frac{2m\|f\|_{\text{Lip}}}{c}.$$

Since  $\text{diam}(\tilde{C}_i) \leq i \cdot 2^{i+2}$ , we conclude that

$$\max_{uv \in E(B_m)} d_{\tilde{C}_i}(f(u), f(v)) = \|f\|_{\text{Lip}} \leq \frac{i \cdot 2^{i+2} \cdot c}{m} \leq \frac{2^{i-6}}{\log(c + 1)}, \quad (30)$$

where  $E(B_m)$  is the set of edges in  $B_m$ .

We will now show that (30) implies that  $f(B_m)$  is contained completely inside an isometric copy of  $\tilde{C}_{i-1}$ . By induction, this will be a contradiction and finish the proof. Let us consider a “top-down” decomposition of  $\tilde{C}_i$  into disjoint pieces. From the root downward, we see  $H_i$ , then a copy of  $C_i$ , then the family of paths  $\mathcal{L}$ . If we also break  $C_i$  into constituent pieces, we see:

1.  $H_i$ ,
2. a copy  $C_{i-1}$ ,
3. a family of paths  $\mathcal{P}$  of length  $2^{i-1}$  connected to the leaves of (2),
4. copies of  $C_{i-1}$  connected to every endpoint of the paths from (3),
5. the family of paths  $\mathcal{L}$  connected to the leaves of the copies of  $C_i$  from (4).



We now define a family of disjoint sub-trees of  $\tilde{C}_i$  each of which is an isometric copy of  $\tilde{C}_{i-1}$ . The first copy  $\tilde{C}_{i-1}^{(0)}$  consists of the bottom  $2^{i-3}$  nodes of  $H_i$ , the copy of  $C_{i-1}$  from (2) above, and the top  $2^{i-3}$  nodes of each path  $p \in \mathcal{P}$  (from (3)). The other copies are indexed by paths  $p \in \mathcal{P}$ . For each such path, we construct  $\tilde{C}_{i-1}^{(p)}$  using the bottom  $2^{i-3}$  nodes of  $p$ , the copy of  $C_{i-1}$  from (4) connected to the bottom of  $p$ , and the top  $2^{i-3}$  nodes of each path from  $\mathcal{L}$  connected to this copy of  $C_{i-1}$ .

We claim that there exists some  $j \in \{0\} \cup \mathcal{P}$  for which  $f(B_m) \subseteq \tilde{C}_{i-1}^{(j)}$ . We now prove the most difficult case; the other cases are similar. Suppose, for the sake of contradiction, there exist  $x, y \in B_m$  for which  $f(x) \in \tilde{C}_{i-1}^{(0)}$  and  $f(y) \in \tilde{C}_{i-1}^{(p)}$  for some  $p \in \mathcal{P}$ . By (30), every edge of  $B_m$  has length at most  $2^{i-6}$ , hence there must be some node  $z \in B_m$  for which  $f(z)$  lies in the middle  $2^{i-3}$  nodes of  $p$ . In particular,  $|B_{\tilde{C}_i}(f(z), r)| \leq 2r + 1$  for every  $r \leq 2^{i-3}$ , since  $B_{\tilde{C}_i}(f(z), 2^{i-3}) \subseteq \mathcal{P}$ . Furthermore,  $|f(B_m) \cap B_{\tilde{C}_i}(f(z), r)| \leq (2r + 1)\|f^{-1}\|_{\text{Lip}}$ . On the other hand,  $|B_{B_m}(z, r')| \geq 2^{r'/2}$  for  $r' \leq m$ .

But we have

$$f\left(B_{B_m}\left(z, \frac{r}{\|f\|_{\text{Lip}}}\right)\right) \subseteq B_{\tilde{C}_i}(f(z), r).$$

Let  $r = \min\{2ic\|f\|_{\text{Lip}}, 2^{i-3}\}$ . Since, in particular  $r \leq m\|f\|_{\text{Lip}}$ , the above considerations yield

$$2^{\frac{r}{2\|f\|_{\text{Lip}}}} \leq (2r + 1)\|f^{-1}\|_{\text{Lip}} = \frac{(2r + 1)c}{\|f\|_{\text{Lip}}} \leq \frac{4rc}{\|f\|_{\text{Lip}}}. \quad (31)$$

Observe that the inequality  $2^B \geq 8Bc$  hold as long as  $B \geq 10 \log(c + 1)$  and  $c \geq 1$ , but it is easy to check that for  $i \geq 10$ , we have  $\frac{r}{2\|f\|_{\text{Lip}}} \geq 10 \log(c + 1)$  (using (30)), yielding a contradiction. This completes the proof.  $\square$

**Remark 3.3.** *Observe that the two point space  $A = \{x, y\}$  with, say,  $d(x, y) = 1$  is a tree metric for which  $\Pi_p(A) = O(1)$  for every  $p < \infty$ . On the other hand, it is easy to see that  $\Pi_p([0, 1]) = \infty$  for every  $p < 2$ , thus in general  $\Pi_p(T_{\mathbb{R}}) \not\approx \Pi_p(T)$  for  $p < 2$ . For  $p \geq 2$ , the relationship is less clear, though we suspect that a similar phenomenon holds in this case. A possible example for which  $\Pi_2(T_{\mathbb{R}}^{(i)}) \not\approx \Pi_2(T^{(i)})$  is when  $T^{(i)}$  is the Cantor tree  $C_i$  with every maximal strait replaced by a single long edge. Using techniques similar to Lemma 3.18, one might show that  $\Pi_2(T^{(i)}) = O(1)$  as  $i \rightarrow \infty$  while  $\Pi_2(T_{\mathbb{R}}^{(i)}) \approx \Pi_2(C_i) \rightarrow \infty$ . We do not pursue this line of reasoning further in the present work.*

## 4 Characterizing the distortion: strong colorings and Markov convexity

In this section we will continue to use the notation of Section 2.1. Moreover, unless explicitly stated otherwise, all paths will be assumed to be monotone. Many of the concepts and definitions used in this section were introduced in Section 3.3, so we suggest that the reader will be familiar with Section 3.3 before reading the present section.

The following result, which contains Theorem 1.4, is the main theorem of this section:

**Theorem 4.1** (The  $L_p$  distortion of trees). *For  $1 < p < \infty$  and every metric tree  $T = (V, E)$ ,*

$$c_p(T) = \Theta(\Pi_{\max\{p,2\}}(T_{\mathbb{R}})) = \Theta\left(\left[\log\left(\frac{2}{\delta^*(T)}\right)\right]^{\min\{\frac{1}{p}, \frac{1}{2}\}}\right),$$

where the implied constants may depend only on  $p$ .

Before proving Theorem 4.1 we make some observations. By Lemma 3.6 for every  $q > 0$  and every two metric spaces  $(X, d_X)$  and  $(Y, d_Y)$ ,  $c_Y(X) \geq \frac{\Pi_q(X)}{\Pi_q(Y)}$ . Since  $L_p$  is  $\max\{p, 2\}$  uniformly convex, Remark 3.1 implies that  $\Pi_{\max\{p,2\}}(L_p) < \infty$ . This observation, together with Theorem 2.6 and Lemma 3.13, implies that

$$\Omega(\Pi_{\max\{p,2\}}(T_{\mathbb{R}})) \leq \frac{1}{5}c_p(T_{\mathbb{R}}) \leq c_p(T) \leq O\left(\left[\log\left(\frac{2}{\delta^*(T)}\right)\right]^{\min\{\frac{1}{p}, \frac{1}{2}\}}\right).$$

Thus, using Corollary 3.15, the proof of Theorem 4.1 will be complete if show that if a metric tree  $T = (V, E)$  does not admit any  $\delta$ -strong coloring then there exists a subtree of  $T$  which is  $(\Omega(1), 2 \cdot \delta^{\Omega(1)})$ -weak prototype with height ratio  $O(1)$ . It is clearly enough to prove this for small enough  $\delta$ , so we assume in what follows that  $\delta < (140)^{-2880}$  (the proof below yields much better constants, but we chose this rough bound to simplify the ensuing exposition). The proof of this assertion is analogous to the proof of Theorem 2.10, where “strong” colorings replace “good” colorings, and weak prototypes take the place of complete binary trees. Since the structure of a weak prototype is not as cleanly recursive as that of a complete binary tree, there are some inevitable added complications. The argument will be broken down into several steps.

#### 4.1 Preliminary results on paths in trees

In what follows, given  $u, v \in V$  we shall say that a set of consecutive edges  $C \subseteq P(u, v)$  is a  $\delta$ -cluster if  $\ell(e) \leq \delta d_T(u, v)$  for every  $e \in C$ .

**Lemma 4.2.** *Fix  $\alpha \in (0, \frac{1}{2})$ ,  $\delta \in (0, 1)$ , and denote  $\tau = \frac{1}{2-4\alpha}$ . Assume that  $u, v \in V$  are such that the path  $P(u, v)$  is  $(\frac{1}{2} + \alpha, \delta)$ -weak. Then at least an  $\alpha$ -fraction of the length of  $P(u, v)$  is covered by  $\delta$ -clusters of length at least  $\tau\delta d_T(u, v)$ . Moreover, at least an  $\alpha$ -fraction of the length of  $P(u, v)$  is covered by edge-disjoint  $\delta$ -clusters of length between  $\tau\delta d_T(u, v)$  and  $(2\tau + 1)\delta d_T(u, v)$ .*

*Proof.* Fix  $u, v \in V$  and denote  $P = P(u, v)$  and  $d = d_T(u, v) = \ell(P)$ . Let  $M$  be the set of maximal  $\delta$ -clusters (with respect to inclusion) contained in  $P$ . In what follows, for a  $\delta$ -cluster  $C \subseteq P$  we write  $\ell(C) = \sum_{e \in C} \ell(e)$ . Define  $S = \{C \in M : \ell(C) < \tau\delta d\}$ . For every  $C \in S$ , since  $C$  is a maximal  $\delta$ -cluster, there is an edge  $e_C \in P \setminus C$  which is incident with an edge in  $C$ , such that  $\ell(e_C) > \delta d \geq \frac{\ell(C)}{\tau}$ . Note that for every edge  $e \in P$ ,  $|\{C \in S : e_C = e\}| \leq 2$ . Now,

$$\sum_{C \in S} \frac{\ell(C)}{\tau} \leq \sum_{C \in S} \ell(e_C) \leq 2 \sum_{\substack{e \in P \\ \ell(e) > \delta d}} \ell(e) \leq 2 \left( d - \sum_{\substack{e \in P \\ \ell(e) \leq \delta d}} \ell(e) \right) \leq 2 \left( 1 - \frac{1}{2} - \alpha \right) d = (1 - 2\alpha) d.$$

Using the fact that the path  $P$  is  $(\frac{1}{2} + \alpha, \delta)$ -weak, we see that

$$\left(\frac{1}{2} + \alpha\right) d \leq \sum_{\substack{e \in P \\ \ell(e) \leq \delta d}} \ell(e) = \sum_{\substack{C \in M \\ \ell(C) \geq \tau \delta d}} \ell(C) + \sum_{C \in S} \ell(C) \leq \sum_{\substack{C \in M \\ \ell(C) \geq \tau \delta d}} \ell(C) + (1 - 2\alpha) \tau d.$$

Recalling that  $\tau = \frac{1}{2-4\alpha}$ , we see that  $\sum_{\substack{C \in M \\ \ell(C) \geq \tau \delta d}} \ell(C) \geq \alpha d$ , as required.

The final assertion of Lemma 4.2 is simply the fact that for any weighted path  $P = (u_1, \dots, u_m)$  such that for each  $j \in \{1, \dots, m-1\}$  we have  $\ell(u_j, u_{j+1}) \leq a$ , but  $\sum_{j=1}^{m-1} \ell(u_j, u_{j+1}) \geq A$ , there are indices  $1 = p_1 < p_2 < \dots < p_k = m$  such that for all  $j \in \{1, \dots, k-1\}$  we have  $\sum_{i=p_j}^{p_{j+1}} \ell(u_i, u_{i+1}) \in [A, 2A+a]$ . Indeed, let  $p_2 > p_1$  be the first index such that  $\sum_{i=p_1}^{p_2} \ell(u_i, u_{i+1}) \geq A$ . Then  $\sum_{i=p_1}^{p_2} \ell(u_i, u_{i+1}) \leq A+a$ . Continuing inductively as long as the length of the remaining path is at least  $A$  we find  $1 = p_1 < p_2 < \dots < p_k$  such that for  $j \in \{1, \dots, k-1\}$  we have  $\sum_{i=p_j}^{p_{j+1}} \ell(u_i, u_{i+1}) \in [A, A+a]$ , and  $\sum_{i=p_k}^m \ell(u_i, u_{i+1}) < A$ . The required result follows by replacing  $p_k$  with  $m$ , which increases the length of the final segment by at most  $A$ .  $\square$

In order to proceed we need to generalize the notions of  $\varepsilon$ -good and  $\delta$ -strong colorings. A coloring  $\chi : E \rightarrow \mathbb{Z}$  will be called  $(\varepsilon, \delta)$ -strong if it is monotone, and for every  $u, v \in V$

$$\sum_{k \in \mathbb{Z}} \ell_k^\chi(u, v) \cdot \mathbf{1}_{\{\ell_k^\chi(u, v) \geq \delta d_T(u, v)\}} \geq \varepsilon d_T(u, v).$$

Note that we can always assume that  $\varepsilon \geq \delta$ . Using the terminology of Section 2.1, an  $\varepsilon$ -good coloring is the same as an  $(\varepsilon, \varepsilon)$ -strong coloring, and a  $\delta$ -strong coloring is the same as a  $(\frac{1}{2}, \delta)$ -strong coloring. Thus the following lemma is a generalization of Lemma 2.5.

**Lemma 4.3.** *Fix  $\varepsilon \in (0, \frac{1}{2}]$  and  $\delta \in (0, \varepsilon)$ . Then any  $(\varepsilon, \delta)$ -strong coloring is also a  $(\frac{\delta}{4\varepsilon})^{3/\varepsilon}$ -strong coloring.*

*Proof.* The proof is a slight modification of the proof of Lemma 2.5. Let  $\chi : E \rightarrow \mathbb{Z}$  be an  $(\varepsilon, \delta)$ -strong coloring, and denote  $\theta = \frac{\varepsilon}{2 \log(\frac{4\varepsilon}{\delta})}$ . We shall show that for every  $\alpha \in (0, 1]$  and  $u, v \in V$ , the total length of the monochromatic segments of length at least  $\alpha d_T(u, v)$  on the path  $P(u, v)$  satisfies

$$\sum_{k \in \mathbb{Z}} \ell_k^\chi(u, v) \cdot \mathbf{1}_{\{\ell_k^\chi(u, v) \geq \alpha d_T(u, v)\}} \geq \left(1 - \left(\frac{\alpha}{\delta}\right)^\theta\right) d_T(u, v). \quad (32)$$

There are points  $a_1, b_1, a_2, b_2, \dots, a_m, b_m \in V$ , ordered consecutively (from  $u$  to  $v$ ) on the path  $P(u, v)$ , such that the color classes of length at least  $\delta d_T(u, v)$  on the path  $P(u, v)$  are precisely the intervals  $\{[a_j, b_j]\}_{j=1}^m$ . Denote for the sake of simplicity  $b_0 = u$  and  $a_{m+1} = v$ , and define  $\beta > 0$  by  $\beta d_T(u, v) = \sum_{j=1}^m d_T(a_j, b_j)$ . Since the coloring is  $(\varepsilon, \delta)$ -strong, we know that  $\beta \geq \varepsilon$ . By the definition of  $\beta$  we are also assured that  $m \leq \beta/\delta$ . If  $\alpha > \delta$  then inequality (32) holds vacuously, so

we assume that  $\alpha \leq \delta$ . Arguing inductively as in the proof of Lemma 2.5 we see that

$$\begin{aligned} \sum_{k \in \mathbb{Z}} \ell_k^X(u, v) \cdot \mathbf{1}_{\{\ell_k^X(u, v) \geq \alpha d_T(u, v)\}} &\geq \sum_{j=1}^m d_T(a_j, b_j) + \sum_{j=0}^m \left(1 - \left(\frac{\alpha d_T(u, v)}{\delta d_T(b_j, a_{j+1})}\right)^\theta\right) d_T(b_j, a_{j+1}) \\ &= d_T(u, v) - \sum_{j=0}^m \left(\frac{\alpha}{\delta}\right)^\theta [d_T(u, v)]^\theta \cdot [d_T(b_j, a_{j+1})]^{1-\theta} \\ &\geq d_T(u, v) - (m+1) \left(\frac{\alpha}{\delta}\right)^\theta [d_T(u, v)]^\theta \left(\frac{1}{m+1} \sum_{j=0}^m d_T(b_j, a_{j+1})\right)^{1-\theta} \end{aligned} \quad (33)$$

$$\begin{aligned} &= d_T(u, v) - (m+1) \left(\frac{\alpha}{\delta}\right)^\theta [d_T(u, v)]^\theta \left(\frac{(1-\beta)d_T(u, v)}{m+1}\right)^{1-\theta} \\ &= \left(1 - \left(\frac{\alpha}{\delta}\right)^\theta (m+1)^\theta (1-\beta)^{1-\theta}\right) d_T(u, v) \\ &\geq \left(1 - \left(\frac{\alpha}{\delta}\right)^\theta \left(\frac{\beta}{\delta} + 1\right)^\theta (1-\beta)^{1-\theta}\right) d_T(u, v) \end{aligned} \quad (34)$$

$$\begin{aligned} &\geq \left(1 - \left(\frac{\alpha}{\delta}\right)^\theta \left(\frac{2}{\delta}\right)^\theta \beta^\theta (1-\beta)^{1-\theta}\right) d_T(u, v) \\ &\geq \left(1 - \left(\frac{\alpha}{\delta}\right)^\theta \left(\frac{2}{\delta}\right)^\theta \varepsilon^\theta (1-\varepsilon)^{1-\theta}\right) d_T(u, v) \end{aligned} \quad (35)$$

$$\geq \left(1 - \left(\frac{\alpha}{\delta}\right)^\theta\right) d_T(u, v), \quad (36)$$

where in (33) we used the concavity of the function  $t \mapsto t^{1-\theta}$ , in (34) we used the fact that  $m \leq \beta/\delta$ , in (35) we used the fact that the function  $s \mapsto s^\theta(1-s)^{1-\theta}$  is decreasing on  $[0, 1]$  and that  $\beta \geq \varepsilon \geq \theta$  (which follows from the definition of  $\theta$  and the fact that  $\varepsilon \geq \delta$ ), and in (36) we used the elementary inequality  $\left(\frac{2}{\delta}\right)^\theta \varepsilon^\theta (1-\varepsilon)^{1-\theta} \leq 1$ , which is equivalent to  $\theta \leq \frac{\log(1-\varepsilon)}{\log[(1-\varepsilon)\delta/(2\varepsilon)]}$ , and this follows from the definition of  $\theta$  since  $\varepsilon \leq \frac{1}{2}$ .  $\square$

Recall that for  $R > 0$  a subset  $N$  of a metric space  $X$  is an  $R$ -net if for every distinct  $x, y \in N$  we have  $d(x, y) \geq R$  and for every  $z \in X$  there is  $x \in N$  with  $d(x, z) < R$ . In what follows we shall use the following variant of this notion.

**Definition 4.4.** Let  $T = (V, E)$  be a tree rooted at  $r$  with edge weights  $\ell : E \rightarrow (0, \infty)$ . For  $R > 0$  we shall call a set  $N \subseteq V$  an upward  $R$ -net of  $T$  if for every  $x, y \in N$  such that  $x$  is an ancestor of  $y$  we have  $d_T(x, y) \geq R$  and for every  $v \in V$  there is  $x \in P(v) \cap N$  such that  $d_T(v, x) < R$ . In other words,  $N$  is an upward  $R$ -net of  $T$  if and only if for every  $v \in V$ ,  $N \cap P(v)$  is an  $R$ -net in  $P(v)$ .

Observe that an upward  $R$ -net in  $T$  need not be an  $R$ -net in  $T$ . However, the following easy lemma shows that upward  $R$ -nets always exist.

**Lemma 4.5.**  $T$  admits an upward  $R$ -net for every  $R > 0$ .

*Proof.* The proof is an easy induction on  $V$ . For  $|V| = 1$  the result is trivial. Assume that  $|V| > 1$  and let  $v \in V$  be a leaf of  $T$ . Let  $u \in V$  be the father of  $v$ . By the inductive hypothesis the tree  $T' = (V \setminus \{v\}, E \setminus \{(u, v)\})$  admits an upward  $R$ -net  $N'$ . Thus there exists  $x \in N \cap P(u)$  such that  $d_{T'}(x, u) = d_T(x, u) < R$ . If  $\ell(u, v) \geq R - d_T(x, u)$  define  $N = N' \cup \{v\}$ , which is clearly an upward  $R$ -net in  $T$ . Otherwise  $d_T(v, x) < R$ , and since  $x \in P(v)$ , it follows that  $N'$  is also an upward  $R$ -net in  $T$ .  $\square$

## 4.2 Construction of a special coloring and the proof of Theorem 4.1

Our basic strategy is similar to the proof of Theorem 2.10. To emphasize the similarities between the two proofs, we will use the same notation for the weight functions and the coloring that we construct (this will not cause any confusion since Section 2.2 can be read independently of the present section). As in the proof of Theorem 2.10 we will define a weight function  $\mu_j$  on subtrees of  $T$ , and a “scale selector”  $g : V \rightarrow \mathbb{Z} \cup \{\infty\}$ , which will be used to construct a coloring  $\chi$  of  $T$ . The fact that  $\chi$  is not  $\delta$ -strong will be used to find an appropriate copy of a weak prototype in  $T$ .

We begin with some notation. Let  $Q$  be a (weighted) path with initial vertex  $x$  and final vertex  $y$ , and let  $F$  be an arbitrary tree with root  $y$  (but otherwise disjoint from  $Q$ ). For  $\varepsilon, \delta \in (0, 1)$  and  $L > 0$  we define  $\rho(\varepsilon, \delta, L; Q, F)$  to be the least minimum distance from the root to a leaf in a subtree  $F' \subseteq F$  which satisfies the following three conditions:

1. Every non-leaf vertex of  $F'$  has exactly one or two children.
2. Let  $P$  be a root-leaf path in  $F'$ , and let  $\tilde{P}$  be the vertices on  $P$  which are either one of the endpoints of  $P$  or have 2 children in  $F'$ . Then the path  $Q \cup \tilde{P}$  is  $(\varepsilon, \delta)$ -weak.
3. Every path from  $x$  to a leaf of  $F'$  has length at most  $3L$ .

Next, we construct a monotone coloring  $\chi : E \rightarrow \mathbb{Z}$  and a “scale selector”  $g : V \rightarrow \mathbb{Z} \cup \{\infty\}$  in a similar way to what was done in Section 2.2. Along the way we will also construct weight functions  $\{\mu_s\}_{s \in \mathbb{Z}}$  on subtrees of  $T$ . As in Section 2.2 we start by setting  $g(r) = \infty$  and we assume inductively that the construction is done so that whenever  $v \in V$  is such that  $g(v)$  is defined, if  $u$  is a vertex on the path  $P(v)$  then  $g(u)$  has already been defined, and for every edge  $e \in E$  incident to  $v$ ,  $\chi(e)$  has been defined.

For every  $t \in \mathbb{Z}$  let  $N_t$  be an upward  $4^t$ -net of  $T$ . Since  $N_t$  is an upward  $4^t$ -net, for all  $w \in V$  we are assured that  $N_t \cap P(w) \cap B_T(w, 4^t) \neq \emptyset$ . We define  $\lambda_t(w)$  to be the point in  $N_t \cap P(w) \cap B_T(w, 2 \cdot 4^t)$  which is furthest away from  $w$ . Now let  $t(s) \in \mathbb{Z}$  be such that

$$240 \cdot \delta^{-\frac{1}{2880}} \cdot 4^s \leq 4^{t(s)} < 960 \cdot \delta^{-\frac{1}{2880}} \cdot 4^s.$$

Take  $v \in V$  which is the vertex closest to the root  $r$  for which  $g(v)$  hasn't yet been defined, and as in Section 2.2 we set

$$g(v) = \max \left\{ j \in \mathbb{Z} : \forall u \in \beta_\chi(v), d_T(u, v) \geq 4^{\min\{g(u), j\}} \right\}. \quad (37)$$

Recall that  $\beta_\chi(v)$  denotes the set of breakpoints of  $\chi$  along the path  $P(v)$ , and that by the inductive hypothesis the path  $P(v)$  has been entirely colored. Let  $F$  be a subtree of  $T$  rooted at  $v$ . We shall now define  $\mu_s(F)$ . To this end define a subset of the path  $P(v)$  by

$$Q_s(v) = \{\lambda_{t(s)}(v)\} \cup \{w \in \beta_\chi(v) : g(w) = g(v) \text{ and } \lambda_{t(s)}(w) = \lambda_{t(s)}(v)\}. \quad (38)$$

With this notation we can define

$$\mu_s(F) = \rho \left( \frac{1}{2500}, \delta^{\frac{1}{2880}}, 4^{t(s)}; Q_s(v), F \right). \quad (39)$$

In (39) we extended the definition of  $\mu_s$  to all subtrees of  $T$  rooted at  $v$ . We next choose one of the children of  $v$ ,  $w \in \mathcal{C}(v)$ , for which

$$\mu_{g(v)}(F_w) = \max_{z \in \mathcal{C}(v)} \mu_{g(v)}(F_z).$$

Observe that  $\mu_{g(v)}(F_z)$  is defined for all the children of  $v$ , since  $F_z$  is a subtree of  $T$  rooted at  $v$  (it is the subtree rooted at  $z \in \mathcal{C}(v)$  together with the incoming edge  $\{v, z\}$ ). Letting  $u$  be the father of  $v$  on the path  $P(v)$ , we set  $\chi(v, w) = \chi(u, v)$ , and we assign arbitrary new (i.e. which haven't been used before) distinct colors to each of the edges  $\{(v, z)\}_{z \in \mathcal{C}(v) \setminus \{w\}}$ .

This construction yields a monotone coloring  $\chi$ , a function  $g : V \rightarrow \mathbb{Z} \cup \{\infty\}$ , and weight function  $\{\mu_s\}_{s \in \mathbb{Z}}$  defined on subtrees of  $T$ . In particular, we note here that Claim 2.11 still holds true, since its proof only used the fact that  $g$  was defined as in (37), and this formula is identical to the one used in Section 2.2. The following lemma contains the crucial properties of the coloring  $\chi$ .

**Lemma 4.6.** *Assume that the above coloring  $\chi$  is not  $\delta$ -strong. Then there exists a sequence of vertices  $Q = (x, w_1, \dots, w_N)$ , ordered down the tree, and a number  $L > 0$ , such that if we define  $s, t \in \mathbb{Z}$  by  $4^{s-1} < \frac{1}{240} \delta^{\frac{1}{2880}} L \leq 4^s$  and  $240 \delta^{-\frac{1}{2880}} 4^s \leq 4^t < 960 \delta^{-\frac{1}{2880}} 4^s$ , then the path metric induced by  $T$  on  $Q$  has the following properties:*

1. For every  $j \in \{1, \dots, N\}$  the vertex  $w_j$  is a breakpoint of  $\chi$ .
2. For every  $j \in \{1, \dots, N\}$  we have  $g(w_j) = s$  and  $\lambda_t(w_j) = x$ .
3. The path  $Q$  is  $\left(\frac{1}{2500}, \delta^{\frac{1}{2880}}\right)$ -weak.
4. The length of  $Q$  satisfies  $\ell(Q) = d_T(x, w_N) \in \left[\frac{L}{850}, 3L\right]$ .

Before passing to the proof of Lemma 4.6 we show how it can be used to complete the proof of Theorem 4.1.

*Proof of Theorem 4.1.* With Lemma 4.6 at hand, the proof of Theorem 4.1 is similar to the final step of the proof of Theorem 2.10. Assume that  $\delta > \delta^*(T)$ . Let  $Q = (x, w_1, \dots, w_N)$  be the path constructed in Lemma 4.6, and we shall also use the same  $s, t, L$  obtained there. Observe that using the notation in (38) we may assume that  $Q = Q_s(w_N)$ . Indeed, by adding to  $Q$  any additional breakpoint  $w$  of  $\chi$  along  $P(w_N)$  with  $g(w) = g(w_N) = s$  and  $\lambda_t(w) = \lambda_t(w_N) = x$  we do not change the conclusion Lemma 4.6.

For  $i \in \{1, \dots, N-1\}$  let  $z_i$  be the child of  $w_i$  along the path  $P(x, w_N)$ , and denote for the sake of simplicity  $z_N = w_N$ . We shall prove by induction on  $i \geq 0$  that the subtree of  $T$  rooted at  $z_{N-i}$  (i.e. the tree  $T_{z_{N-i}}$ ) has a further sub-tree  $W_i$  satisfying the following properties.

1. Every path from  $x$  to a leaf of  $W_i$  has length at most  $3 \cdot 4^t$ .
2. Every non-leaf vertex of  $W_i$  has exactly one or two children.

3. Let  $P$  be a path from  $z_{N-i}$  to a leaf of  $W_i$ , and let  $\tilde{P}$  be the vertices on  $P$  which are either one of the endpoints of  $P$  or have 2 children in  $W_i$ . Then the path  $(x, w_1, \dots, w_{N-i}) \cup \tilde{P}$  is  $\left(\frac{1}{2500}, \delta^{\frac{1}{2880}}\right)$ -weak.

4. Every root-leaf path in  $W_i$  has length at least  $d_T(z_{N-i}, w_N)$ .

For  $i = 0$  we just take  $W_0$  to be the singleton  $w_N$ , and the fact that the required properties are satisfied is asserted in Lemma 4.6. Similarly, for  $i = 1$  we let  $W_1$  be the tree consisting of the single edge  $(z_{N-1}, w_N)$  which satisfies the required properties due to Lemma 4.6. Assuming the existence of  $W_i$  we proceed inductively as follows. Since  $w_{N-i}$  is a breakpoint of  $\chi$ , the construction of  $\chi$  in the proof of Theorem 2.10, and the fact that  $g(w_{N-i}) = s$ , implies that there is a child  $z'_{N-i}$  of  $w_{N-i}$  for which  $\mu_s(F_{z'_{N-i}}) > \mu_s(F_{z_{N-i}})$  (recall that for  $u \in V$  the tree  $F_u$  is the subtree rooted at  $u$  plus the edge joining  $u$  and its parent in  $T$ ). Now, since  $Q_s(w_N) = Q$  we also know that (since  $\lambda_t(w_{N-i}) = x$ )  $Q_s(w_{N-i}) = \{x, w_1, \dots, w_{N-i}\}$ . Thus by the definition on  $\mu_s$  in (39)

$$\mu_s(F_{z_{N-i}}) = \rho\left(\frac{1}{2500}, \delta^{\frac{1}{2880}}, 4^t; Q_s(w_{N-i}), F_{z_{N-i}}\right) = \rho\left(\frac{1}{2500}, \delta^{\frac{1}{2880}}, 4^t; \{x, w_1, \dots, w_{N-i}\}, F_{z_{N-i}}\right).$$

But,  $W_i$  is a subtree of  $F_{z_{N-i}}$  in which every non-leaf vertex has two children, for every path from  $x$  to a leaf of  $W_i$  the path metric induced by  $T$  on the vertices which are either  $x$ , or one of the  $w_j$ , or a leaf in  $W_i$ , or have 2 children in  $W_i$ , is  $\left(\frac{1}{2500}, \delta^{\frac{1}{2880}}\right)$ -weak, every path from  $x$  to a leaf of  $W_i$  has length at most  $3L \leq 3 \cdot 4^t$ , and the minimal distance from a root to a leaf of  $W_i$  is at least  $d_T(z_{N-i}, w_N)$ . Thus the definition of  $\rho$  implies that  $\mu_s(F_{z_{N-i}}) \geq d_T(z_{N-i}, w_N)$ . It follows that  $\mu_s(F_{z'_{N-i}}) > d_T(z_{N-i}, w_N)$ , implying the existence of a subtree  $W'_i$  of  $F_{z'_{N-i}}$ , which has the same properties as those stated above for  $W_i$ . Joining these two subtrees at  $w_{N-i}$ , and adding an edge from  $z_{N-i+1}$  to  $w_{N-i}$  we obtain a subtree  $W_{N-i+1}$  rooted at  $z_{N-i+1}$  with the desired properties. We recommend that the reader will follow the above construction using a drawing analogous to Figure 3.

The tree  $T'$  obtained by joining the edges  $(x, w_1), (w_1, z_1)$  to  $W_{N-1}$  is a subtree of  $T$  which is a  $\left(\frac{1}{2500}, \delta^{\frac{1}{2880}}\right)$ -weak prototype with height ratio at most  $\frac{3 \cdot 4^t}{d(x, w_N)} \leq \frac{8L}{L/850} = 6800$ . As explained in the discussion following Theorem 4.1, this completes the proof.  $\square$

Thus, all that remains is to prove Lemma 4.6.

*Proof of Lemma 4.6.* Since we are assuming that the coloring  $\chi$  is not  $\delta$ -strong, Lemma 4.3 implies that  $\chi$  is also not  $\left(\frac{1}{960}, \frac{1}{240}\delta^{\frac{1}{2880}}\right)$ -strong. Thus there exist two vertices  $u, v \in V$  such that more than a  $\frac{959}{960}$ -fraction of the length of the path joining  $u$  and  $v$  is covered by color classes of length less than  $\frac{1}{240}\delta^{\frac{1}{2880}} \cdot d_T(u, v)$ . Let  $(b_1, \dots, b_m)$  be the breakpoints of the coloring  $\chi$  along the path  $P(u, v)$ , ordered from  $u$  to  $v$ . We also denote  $b_0 = u$  and  $b_{m+1} = v$ . Thus

$$\sum_{j=1}^{m+1} d_T(b_{j-1}, b_j) \cdot \mathbf{1}_{\left\{d_T(b_{j-1}, b_j) < \frac{1}{240}\delta^{\frac{1}{2880}} \cdot d_T(u, v)\right\}} \geq \frac{959}{960} d_T(u, v) = \left(\frac{1}{2} + \frac{479}{960}\right) d_T(u, v).$$

Lemma 4.2 (with  $\alpha = \frac{479}{960}$  and  $\tau = \frac{1}{2-4\alpha} = 240$ ) implies that there exists a sequence of indices  $0 \leq p_1 < q_1 \leq p_2 < q_2 \leq \dots \leq p_{k-1} < q_{k-1} \leq p_k < q_k \leq m+1$  such that for every  $1 \leq i \leq k$

we have  $d_T(b_{p_i}, b_{q_i}) \in \left[ \delta^{\frac{1}{2880}} d_T(u, v), 3\delta^{\frac{1}{2880}} d_T(u, v) \right]$  and every  $p_i < j \leq q_i$  satisfies  $d_T(b_{j-1}, b_j) \leq \frac{1}{240} \delta^{\frac{1}{2880}} d_T(u, v)$ . Moreover, the total length of these “long  $\frac{1}{240} \delta^{\frac{1}{2880}}$ -clusters” is

$$\sum_{i=1}^k d_T(b_{p_i}, b_{q_i}) \geq \frac{479}{960} d_T(u, v). \quad (40)$$

It follows in particular from (40) that  $k \geq \frac{479}{3 \cdot 960} \cdot \delta^{-\frac{1}{2880}} \geq 20$  (since  $\delta < (140)^{-2880}$ ).

Denote  $L = d_T(u, v)$  and recall that  $s \in \mathbb{Z}$  is defined by  $4^{s-1} < \frac{1}{240} \delta^{\frac{1}{2880}} L \leq 4^s$ . Fix  $1 \leq i \leq k$  and apply Claim 2.11 to the path  $P(b_{p_i}, b_{q_i})$  with  $c = 2$  (which we are allowed to do by the definition of  $s$ ). It follows that there exist at least two indices  $p_i \leq j_1(i) < j_2(i) \leq q_i$  such that  $g(b_{j_1(i)}) = g(b_{j_2(i)}) = s$  and  $9 \cdot 4^s \leq d_T(b_{j_1(i)}, b_{j_2(i)}) \leq 18 \cdot 4^s$ .

Now  $t \in \mathbb{Z}$  is given by  $240\delta^{-\frac{1}{2880}} 4^s \leq 4^t < 960\delta^{-\frac{1}{2880}} 4^s$ . Note that by the definition of  $s$  this implies that  $L \leq 4^t < 16L$ . For each point  $w \in \{b_{j_1(1)}, b_{j_2(1)}, \dots, b_{j_1(k)}, b_{j_2(k)}\}$  the vertex  $\lambda_s(w)$  is in  $B_T(w, 2 \cdot 4^t) \cap N_t \cap P(w) \subseteq B_T(v, 2 \cdot 4^t + L) \cap N_t \cap P(v) \subseteq B_T(v, 3 \cdot 4^t) \cap N_t \cap P(v)$ . Since  $N_t \cap P(v)$  is  $4^t$ -separated, it follows that there are at most 4 possible vertices which could equal  $\lambda_t(w)$ . Thus there is a vertex  $x \in V$  and a subinterval  $J \subseteq \{1, \dots, k\}$  of size at least  $\frac{k}{4} - 1 \geq \frac{k}{5}$  (since  $k \geq 20$ ) such that for all  $i \in J$  we have  $\lambda_t(b_{j_1(i)}) = \lambda_t(b_{j_2(i)}) = x$ . Note that since  $x = \lambda_t(w)$  for some  $w \in \{b_{j_1(1)}, b_{j_2(1)}, \dots, b_{j_1(k)}, b_{j_2(k)}\}$ , we know that  $x$  is the point in  $N_t \cap P(w) \cap B_T(w, 2 \cdot 4^t)$  which is furthest from  $w$ . Since  $N_t$  is an upward  $4^t$ -net, there is a point  $y \in N_t \cap P(u) \cap B_T(u, 4^t)$ . So, using  $d_T(w, u) \leq L \leq 4^t$ , we see that  $y \in N_t \cap P(w) \cap B_T(w, 2 \cdot 4^t)$ . Thus  $x \in P(y) \subseteq P(u)$ , i.e.  $x$  is closer to the root than  $u$ .

Consider the path metric induced on the vertices  $Q = \{x\} \cup \{b_{j_1(i)}\}_{i \in J} \cup \{b_{j_2(i)}\}_{i \in J}$ . For simplicity of notation we enumerate it down the tree by  $Q = (x, w_1, \dots, w_N)$ . We bound the length of  $Q$  as follows. First,  $\ell(Q) = d_T(x, w_N) = d_T(x, w_1) + d_T(w_1, w_N) \leq 2 \cdot 4^t + L \leq 3L$ . On the other hand, using (40) we see that

$$\begin{aligned} \ell(Q) &\geq \sum_{i \in J} d_T(b_{j_1(i)}, b_{j_2(i)}) \geq \sum_{i \in J} 9 \cdot 4^s \geq \frac{k}{5} \cdot 9 \cdot 4^s \\ &\geq \frac{9}{5} \sum_{i=1}^k \frac{1}{240} \delta^{\frac{1}{2880}} L \geq \frac{3}{400} \sum_{i=1}^k \frac{1}{3} d_T(b_{p_i}, b_{q_i}) \geq \frac{479}{400 \cdot 960} L \geq \frac{\ell(Q)}{2500}. \end{aligned}$$

This also shows that the path  $Q$  is  $\left(\frac{1}{2500}, \delta^{\frac{1}{2880}}\right)$ -weak, completing the proof of Lemma 4.6.  $\square$

## References

- [1] M. Bădoiu, J. Chuzhoy, P. Indyk, and A. Sidiropoulos. Low-distortion embeddings of general metrics into the line. In *STOC'05: Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 225–233, New York, 2005. ACM.
- [2] K. Ball. Markov chains, Riesz transforms and Lipschitz maps. *Geom. Funct. Anal.*, 2(2):137–172, 1992.
- [3] K. Ball, E. A. Carlen, and E. H. Lieb. Sharp uniform convexity and smoothness inequalities for trace norms. *Invent. Math.*, 115(3):463–482, 1994.



- [4] Y. Bartal. On approximating arbitrary metrics by tree metrics. In *STOC '98 (Dallas, TX)*, pages 161–168. ACM, New York, 1999.
- [5] I. Benjamini and O. Schramm. Every graph with a positive Cheeger constant contains a tree with a positive Cheeger constant. *Geom. Funct. Anal.*, 7(3):403–419, 1997.
- [6] J. Bourgain. The metrical interpretation of superreflexivity in Banach spaces. *Israel J. Math.*, 56(2):222–230, 1986.
- [7] J. Bourgain, V. Milman, and H. Wolfson. On type of metric spaces. *Trans. Amer. Math. Soc.*, 294(1):295–317, 1986.
- [8] S. Buyalo and V. Schroeder. Embedding of hyperbolic spaces in the product of trees. *Geom. Dedicata*, 113:75–93, 2005.
- [9] A. N. Dranishnikov. On hypersphericity of manifolds with finite asymptotic dimension. *Trans. Amer. Math. Soc.*, 355(1):155–167 (electronic), 2003.
- [10] A. W. M. Dress. Trees, tight extensions of metric spaces, and the cohomological dimension of certain groups: a note on combinatorial properties of metric spaces. *Adv. in Math.*, 53(3):321–402, 1984.
- [11] P. Enflo. On infinite-dimensional topological groups. In *Séminaire sur la Géométrie des Espaces de Banach (1977–1978)*, pages Exp. No. 10–11, 11. École Polytech., Palaiseau, 1978.
- [12] J. Fakcharoenphol, S. Rao, and K. Talwar. A tight bound on approximating arbitrary metrics by tree metrics. In *Proceedings of the 35th Annual ACM Symposium on Theory of Computing*, pages 448–455, 2003.
- [13] T. Figiel. On the moduli of convexity and smoothness. *Studia Math.*, 56(2):121–155, 1976.
- [14] H. Furstenberg and B. Weiss. Markov processes and Ramsey theory for trees. *Combin. Probab. Comput.*, 12(5-6):547–563, 2003. Special issue on Ramsey theory.
- [15] M. Gromov. Filling Riemannian manifolds. *J. Differential Geom.*, 18(1):1–147, 1983.
- [16] A. Gupta. Steiner points in tree metrics don't (really) help. In *Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms (Washington, DC, 2001)*, pages 220–227, Philadelphia, PA, 2001. SIAM.
- [17] A. Gupta, R. Krauthgamer, and J. R. Lee. Bounded geometries, fractals, and low-distortion embeddings. In *44th Symposium on Foundations of Computer Science*, pages 534–543, 2003.
- [18] W. B. Johnson, J. Lindenstrauss, D. Preiss, and G. Schechtman. Lipschitz quotients from metric trees and from Banach spaces containing  $l_1$ . *J. Funct. Anal.*, 194(2):332–346, 2002.
- [19] V. A. Kaĭmanovich and A. M. Vershik. Random walks on discrete groups: boundary and entropy. *Ann. Probab.*, 11(3):457–490, 1983.
- [20] U. Lang and T. Schlichenmaier. Nagata dimension, quasisymmetric embeddings, and Lipschitz extensions. *Internat. Math. Research Notices*, 58:3625–3655, 2005.

- [21] E. Leuzinger. Bi-Lipschitz embeddings of trees into Euclidean buildings. *Geom. Dedicata*, 102:109–125, 2003.
- [22] J. Lindenstrauss and L. Tzafriri. *Classical Banach spaces. II*, volume 97 of *Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas]*. Springer-Verlag, Berlin, 1979. Function spaces.
- [23] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.
- [24] N. Linial, A. Magen, and M. E. Saks. Low distortion Euclidean embeddings of trees. *Israel J. Math.*, 106:339–348, 1998.
- [25] N. Linial and M. Saks. The Euclidean distortion of complete binary trees. *Discrete Comput. Geom.*, 29(1):19–21, 2003.
- [26] R. Lyons, R. Pemantle, and Y. Peres. Random walks on the lamplighter group. *Ann. Probab.*, 24(4):1993–2006, 1996.
- [27] J. Matoušek. Extension of Lipschitz mappings on metric trees. *Comment. Math. Univ. Carolin.*, 31(1):99–104, 1990.
- [28] J. Matoušek. On embedding trees into uniformly convex Banach spaces. *Israel J. Math.*, 114:221–237, 1999.
- [29] B. Maurey and G. Pisier. Séries de variables aléatoires vectorielles indépendantes et propriétés géométriques des espaces de Banach. *Studia Math.*, 58(1):45–90, 1976.
- [30] M. Mendel and A. Naor. Metric cotype. To appear in the *Annals of Mathematics*, 2005.
- [31] A. Naor, Y. Peres, O. Schramm, and S. Sheffield. Markov chains in smooth Banach spaces and Gromov hyperbolic metric spaces. *Duke Math. J.*, 134(1):165–197, 2006.
- [32] A. Naor and G. Schechtman. Remarks on non linear type and Pisier’s inequality. *J. Reine Angew. Math.*, 552:213–236, 2002.
- [33] G. Pisier. Martingales with values in uniformly convex spaces. *Israel J. Math.*, 20(3-4):326–350, 1975.
- [34] G. Pisier. Probabilistic methods in the geometry of Banach spaces. In *Probability and analysis (Varenna, 1985)*, volume 1206 of *Lecture Notes in Math.*, pages 167–241. Springer, Berlin, 1986.
- [35] C. Thomassen. Isoperimetric inequalities and transient random walks on graphs. *Ann. Probab.*, 20(3):1592–1600, 1992.
- [36] N. T. Varopoulos. Long range estimates for Markov chains. *Bull. Sci. Math. (2)*, 109(3):225–252, 1985.