Vertex Cuts, Random Walks, and Dimension Reduction in Series-Parallel Graphs

Bo Brinkman*
Miami University

Adriana Karagiozova†
Princeton University

James R. Lee‡
University of Washington

ABSTRACT

We consider questions about vertex cuts in graphs, random walks in metric spaces, and dimension reduction in L_1 and L_2; these topics are intimately connected because they can each be reduced to the existence of various families of real-valued Lipschitz maps on certain metric spaces. We view these issues through the lens of shortest-path metrics on series-parallel graphs, and we discuss the implications for a variety of well-known open problems. Our main results follow.

—Every n-point series-parallel metric embeds into ℓ_1^{dom} with $O(\sqrt{\log n})$ distortion, matching a lower bound of Newman and Rabinovich. Our embeddings yield an $O(\sqrt{\log n})$ approximation algorithm for vertex sparsest cut in such graphs, as well as an $O(\sqrt{\log k})$ approximate max-flow/min-vertex-cut theorem for series-parallel instances with k terminals, improving over the $O(\log n)$ and $O(\log k)$ bounds for general graphs.

—Every n-point series-parallel metric embeds with distortion D into ℓ_1^d with $d = n^{1/\Omega(D^2)}$, matching the dimension reduction lower bound of Brinkman and Charikar.

—There exists a constant $C > 0$ such that if (X, d) is a series-parallel metric then for every stationary, reversible Markov chain $\{Z_t\}_{t=0}^{\infty}$ on X, we have for all $t \geq 0$,

$$\mathbb{E}\left[d(Z_t, Z_0)^2\right] \leq Ct \cdot \mathbb{E}\left[d(Z_0, Z_1)^2\right].$$

More generally, we show that series-parallel metrics have Markov type 2. This generalizes a result of Naor, Peres, Schramm, and Sheffield for trees.

*brinkmwj@muohio.edu

†Supported by NSF ITR grant CCR-0205594, NSF CAREER award CCR-0237113, MSPA-MCS award 0528414, and Moses Charikar's Alfred P. Sloan Fellowship.
karagioz@cs.princeton.edu

‡Supported by NSF CAREER award CCF-0644037.
jrl@cs.washington.edu

STOC'07, June 11–13, 2007, San Diego, California, USA.

Categories and Subject Descriptors

F.2 [**Theory of Computation**]: Analysis of Algorithms; G.3 [**Mathematics of Computing**]: Probability and Statistics

General Terms

Algorithms, Theory

Keywords

Metric embeddings, approximation algorithms, dimension reduction

1. INTRODUCTION

In the past 15 years, low-distortion metric embeddings—following the initial work of Linial, London, and Rabinovich [20]—have become an integral part of theoretical computer science, and the geometry of metric spaces seems to lie at the heart of some of the most important open problems in approximation algorithms (for some recent examples see, e.g. [3, 2, 1, 10, 8] and the discussions therein). For background on the field of metric embeddings and their applications in computer science, we refer to Matoušek's book [22, Ch. 15] and Indyk's survey [15].

In the present work, we consider phenomena related to finding sparse vertex cuts in graphs, the nature of dimension reduction in L_1 and L_2, and random walks on metric spaces. These seemingly disparate topics are related by a common theme: The desire to compute faithful 1-dimensional representations of a given metric space. Stated differently, our study concerns the existence of certain families of real-valued Lipschitz maps.

We first present some notation. If (X, d) is a metric space and $s \geq 1$, we use $c_1^{\mathrm{dom}(s)}(X)$ to denote the minimum distortion of any 1-Lipschitz embedding from X into L_1, subject to the constraint that every coordinate of the embedding is s-Lipschitz. In other words, we require a random mapping $F : X \to \mathbb{R}$ such that for every $x, y \in X$, we have

1. $\Pr\left(|F(x) - F(y)| \leq s \cdot d(x, y)\right) = 1$, and
2. $\frac{d(x,y)}{D} \leq \mathbb{E}|F(x) - F(y)| \leq d(x, y)$.

Matoušek and Rabinovich [23] initiated the study of this parameter in the case $s = 1$, based on the fact that Bourgain's embedding applied to an n-point metric space satisfies these constraints for $s = 1$ and $D = O(\log n)$ [5]. For this special case, we denote $c_1^{\mathrm{dom}}(X) = c_1^{\mathrm{dom}(1)}(X)$.

Applications of ℓ_1^{dom} embeddings. To understand the importance of studying this parameter, we mention the following results. The first two are algorithmic reductions due to Feige, Hajiaghayi, and Lee [10] which show that these types of embeddings can be used to find sparse vertex cuts in graphs.

THEOREM 1.1. **(Sparse vertex cuts and multi-flows)** *If $\mathcal{F}$ is a family of graphs such that for every n-point shortest-path metric (X, d) arising from assigning edge weights to the graphs in $\mathcal{F}$, we have $c_1^{\text{dom}}(X) \leq D(n)$, then there is an $O(D(n))$ approximation to the Vertex Sparsest Cut Problem on arbitrary n-node graphs from $\mathcal{F}$.*

Furthermore, the gap between the vertex sparsest cut and the maximum vertex-capacitated multi-commodity flow on any instance coming from $\mathcal{F}$ is at most $O(D(k))$ where k is the number of flow types.

We recall that the Vertex Sparsest Cut Problem is simply the variant of Sparsest Cut where vertices are cut instead of edges: Given a graph $G = (V, E)$, with weights $w : V \to \mathbb{R}_+$ on the vertices and demands $\text{dem} : V \times V \to \mathbb{R}_+$, find a partition $V = A \cup S \cup B$ of the vertex set such that there are no edges between A and B, and the following ratio is minimized

$$\frac{\sum_{v \in S} w(v)}{\sum_{u \in A \cup S} \sum_{v \in B \cup S} \text{dem}(u, v)}.$$

In [10], it is shown that these embeddings can be used to round the vertex Sparsest Cut SDP—we refer to [10] for a discussion of negative type metrics.

THEOREM 1.2 (**SDP rounding for vertex cuts**, [10]). *If for every n-point metric of negative type (X, d) one has $c_1^{\text{dom}}(X) \leq D(n)$, then there is an $O(D(n))$ approximation to Vertex Sparsest Cut in general graphs.*

The next application of ℓ_1^{dom} embeddings concerns dimension reduction in L_1, via the following theorem of Brinkman and Charikar [6] (restated in our notation). The idea is that control on the Lipschitz constant of each coordinate gives one the ability to do sparse random sampling.

THEOREM 1.3 (**Dimension reduction in L_1**, [6]). *If one has $c_1^{\text{dom}(s)}(X) \leq D$ for some metric space (X, d). Then there exists a distortion $O(D)$ embedding of X into $\ell_1^{O(sD \log n)}$.*

Intrinsic dimension and subsets of L_2. There is a well-known "intrinsic" dimension reduction question for doubling subsets of L_2, apparently first asked by Lang and Plaut [17] (see also [12]): If $X \subseteq L_2$ is doubling with constant λ, does there exist an embedding of X into $\mathbb{R}^k$ with k and the distortion depending only on λ? (We refer to [14] for a discussion of doubling spaces and the doubling constant.)

We now outline how ℓ_1^{dom} embeddings might be used to give a negative answer to this question. From [12], we know that every n-point doubling metric embeds in L_2 with distortion $O(\sqrt{\log n})$. It is also easy to see that $c_1^{\text{dom}}(\mathbb{R}^k) \leq k$, since the coordinate projections are 1-Lipschitz. Thus we pose the following.

QUESTION 1. *If (X, d) is an n-point doubling metric, do we have $c_1^{\text{dom}}(X) \leq O(\sqrt{\log n})$?*

If the answer to this question is *negative*, it would yield a negative answer to the problem of dimension reduction for doubling metrics as long as the violating space has an $O(\sqrt{\log n})$-embedding into L_2 where the image is doubling (this condition does not seem particularly strong). Thus we see that both upper and lower bounds for embeddings into ℓ_1^{dom} are interesting. The preceding remark raises another question.

QUESTION 2. *If (X, d) is an n-point doubling metric, does there exist an $O(\sqrt{\log n})$-embedding into L_2 for which the image is doubling (with constant independent of n)?*

Markov type of metric spaces. Finally, we consider Ball's notion [4] of Markov type for metric spaces. A metric space (X, d) is said to have *Markov type p* if there exists a constant C such that for every stationary, reversible Markov chain $\{Z_t\}_{t=0}^{\infty}$ on $\{1, \ldots, n\}$ and every map $f : \{1, \ldots, n\} \to X$, we have

$$\mathbb{E}\, d(Z_t, Z_0)^p \leq C^p t \cdot \mathbb{E}\, d(Z_0, Z_1)^p$$

The smallest constant C for which this holds is denoted by $M_p(X)$ and called the *Markov type p-constant of X*. Although we do not yet know of any algorithmic application of Markov type, it is a notion that seems to fit in well with the questions addressed in the present paper, due to the following lemma (which is a straightforward generalization of the "potential function" approach of Naor, Peres, Schramm, and Sheffield [25]). We recall that, in general (i.e. for families of spaces which contain arbitrarily long path metrics), $p = 2$ is the best-possible Markov type.

PROPOSITION 1.4. *Let $G = (X, E)$ be a weighted graph with shortest path metric d. If there exists a map $\psi : X \to \mathbb{R}$ such that, for every path $P = \langle v_0, v_1, \ldots, v_m \rangle$ in G, one has*

$$d(v_0, v_m) \leq \max\{|\psi(v_0) - \psi(v_t)| : 0 \leq t \leq m\}$$

then X has Markov type 2, and $M_2(X) \leq O(\|\psi\|_{\text{Lip}})$.

The existence of a such a map $\psi : X \to \mathbb{R}$ seems very powerful; in particular, it suggests that the space (X, d) is "one directional" (in the sense of the geodesic structure of trees or hyperbolic spaces). The follow question seems intriguing.

QUESTION 3. *If $G = (X, E)$ is a weighted graph metric and $\psi : X \to \mathbb{R}$ satisfies the condition of Proposition 1.4, does there exist a constant $C = C(\|\psi\|_{\text{Lip}})$ such that $c_1(X) \leq C$?*

1.1 Results and related work

In general, little is known about the construction of ℓ_1^{dom} embeddings. Bourgain's theorem [5] shows that every n-point metric space (X, d) has $c_1^{\text{dom}}(X) \leq O(\log n)$. Furthermore, it is easy to see (via Cauchy-Schwartz) that $c_1^{\text{dom}}(X) \geq c_2(X)$ where $c_2(X)$ represents the distortion required to embed X into L_2. In this paper, we study the previously mentioned notions, in particular ℓ_1^{dom} embeddings, through the lens of shortest-path metrics on series-parallel graphs.

Such graphs and their associated metrics lie at the precipice of a number of embedding questions. This is due largely to the fact that they can exhibit non-trivial *multi-scale* structure. For instance, they embed uniformly into L_1 [13], but not into L_2 [26], though they do admit L_2 embeddings better than the worst case [27]. They exhibit the worst-possible distortion for embedding n-point metric spaces into distributions over dominating trees [13], and although they embed in L_1, the required dimension of the embedding is very large [6]. Furthermore, there are interesting examples of doubling, series-parallel metrics that, e.g. disprove Assouad's conjecture for doubling metrics [16]. We give tight bounds on $\ell_1^{\mathrm{dom}(s)}$ embeddings for the class of series-parallel metrics.

THEOREM 1.5. *Let (X, d) be any n-point series-parallel metric, then for every $s \geq 1$,*

$$c_1^{\mathrm{dom}(s)}(X) = O\left(\sqrt{\frac{\log n}{\log s}}\right),$$

and this bound is tight for every $s \geq 1$.

Vertex Sparsest Cut. Our embeddings yield an $O(\sqrt{\log n})$-approximation algorithm for vertex sparsest cut in such graphs, as well as an $O(\sqrt{\log k})$-approximate max-flow/min-vertex-cut theorem for series-parallel instances with k terminals, improving over the $O(\log n)$ and $O(\log k)$ bounds for general graphs [10]. Our study of such graphs is partially motivated by the following open problem.

QUESTION 4. *Is it true that $c_1^{\mathrm{dom}(s)}(X) = o(\log n)$ when (X, d) is an n-point planar metric, or an n-point metric space of negative type?*

Dimension reduction in L_1. We show that every n-point series-parallel metric embeds with distortion D into ℓ_1^d with $d = n^{1/\Omega(D^2)}$, matching the dimension reduction lower bound of Brinkman and Charikar [6] (see also [18]). Given that the only dimension reduction lower bounds for L_1 are based on series-parallel graphs, the next question is quite natural.

QUESTION 5. *Does every n-point subset of L_1 admit a D-embedding into ℓ_1^d with $d = n^{1/\Omega(D^2)}$?*

We remark that for, say, constant values of D, it is not known how to reduce the dimension past $d = O(n)$ for general n-point subsets. For $D = O(\sqrt{\log n}\log\log n)$, the results of [2] can achieve $d = O(\log n)$.

Markov type 2. We show that series-parallel metrics have Markov type 2 (with uniformly bounded type 2 constant). This generalizes the result of [25] for trees. We recall that the authors of that paper asked whether every doubling metric or every planar metric has Markov type 2 with uniformly bounded constant.

1.2 Techniques & proof overview

Structure theorem. Our approach begins with a basic structure theorem for metrics on series-parallel graphs. We show that every such metric embeds into the product of two (perhaps slightly larger) metric spaces called *bundle trees*. These are shortest path metrics on a graph whose 2-connected components are *bundles*. A bundle is a series-parallel metric with two distinguished points s, t such that every simple s-t path has the same length.

A related structure theorem was given by Gupta et. al. [13], but their embedding is into a *distribution* over bundle trees, and there is no bound on the variance of the Lipschitz constant (they only a require a bound which holds in expectation). To overcome this, we combine the techniques of [13] with the low-variance approach of Charikar and Sahai [7] for outerplanar graphs. This is carried out in Section 4. The structure theorem reduces problems for general series-parallel metrics, to the cases of trees and series-parallel bundles.

Distributions over Lipschitz maps and dimension reduction in L_1. Recall that we are trying to construct L_1 embeddings where we have some control on the Lipschitz constant of every coordinate. Consider first the problem of "converting" an ℓ_2 embedding into an ℓ_1^{dom} embedding. Given an embedding $f : X \to \ell_2^n$, we can think of the latter space as having a basis of n i.i.d. ± 1 Bernoulli random variables $\{\varepsilon_i\}_{i=1}^n$ (with inner product $\langle x, y\rangle = \mathbb{E}(xy)$). Thus for every $x \in X$, $f(x) = \sum_{i=1}^n c_i(x)\varepsilon_i$ for some coefficients $\{c_i(x)\}$.

In this sense, we can think about the embedding as a random walk; to get an embedding into $\ell_1^{\mathrm{dom}(s)}$, we have to truncate the walk if it starts to wander too far. Unfortunately, the walks corresponding to distinct points are correlated, and any straightforward approach ends up destroying all the variance in the embedding very quickly (roughly because every point is participating in $n-1$ different random walks, one for every other $y \in X$, and we have to "stop" if any of these walks becomes violated).

Thus we have to redesign our embeddings so they allow for "local" truncation rules that still allow the image of x to vary significantly (within the standard deviation). In order to get a tight result for the whole range $s \in [1, \infty)$, we also need to have a method for slowly increasing the variance (as in Corollary 3.2, we can think of this as losing control on the p^{th} moments as p decays from 2 to 1, much as in the lower bound proof of [18]).

In Section 3, we give ℓ_1^{dom} embeddings for bundles and trees, and a reduction of bundle trees to these two cases. arbitrary L_1 embeddings, this reduction is non-trivial). For trees, we use a simple edge-by-edge random walk based on the standard caterpillar decomposition approach (see [21, 12, 19, 11]). The main technical component of the proof is the embedding for series-parallel bundles. We use a local truncation strategy given by the tree-like nature of their construction sequence (see Section 2.1). We are careful to main two properties: monotonicity (Lemma 3.5) and unbiasedness (Lemma 3.6) that allow us to view the stretch of an edge as a martingale—we then bound the effect of truncation using Doob's maximal inequality. We suspect that these two properties will play an essential role in future ℓ_1^{dom} embeddings, as well as in the proof of lower bounds.

Markov type 2. To prove that series-parallel metrics have Markov type 2, there are three steps, which occur in Section 5. The first is a straight-forward generalization of the result of [25] which reduces the problem to the construction of appropriate Lipschitz maps on our space. The second step is the structure theorem of Section 4, which allows us to pass to a bundle tree. The final step is to prove that bundle trees themselves have Markov type 2.

2. PRELIMINARIES

We briefly review some general facts and notation. If $(X, d_X), (Y, d_Y)$ are metric spaces, and $f : X \to Y$, then we write $\|f\|_{\mathrm{Lip}} = \sup_{x \neq y \in X} \frac{d_Y(f(x),f(y))}{d_X(x,y)}$. If f is injective, then the *distortion of* f is $\|f\|_{\mathrm{Lip}} \cdot \|f^{-1}\|_{\mathrm{Lip}}$. We will often encounter the product metric space $X \times Y$ which we always equip with the ℓ_1 distance $d_{X\times Y}((x,y),(x',y')) = d_X(x,x') + d_Y(y,y')$. If Z is a real-valued random variable, we write $\|Z\|_p = (\mathbb{E}|Z|^p)^{\frac{1}{p}}$, for $p < \infty$, and $\|Z\|_\infty = \inf\{C : \Pr(Z \le C) = 1\}$.

If G is a graph, we use $V(G)$ and $E(G)$ to denote its vertex and edge set, respectively. For weighted graphs, we use $\mathsf{len}(e)$ to denote the length of an edge, and for a path P consisting of edges $e_1, \dots, e_m$, we write $\mathsf{len}(P)$ for $\mathsf{len}(e_1) + \cdots \mathsf{len}(e_m)$. We use the notation (u,v) throughout the paper for an *undirected* edge (thus the ordering does not matter). If E_1, E_2 are two expressions, we sometimes write $E_1 \lesssim E_2$ for $E_1 = O(E_2)$, and $E_1 \approx E_2$ if both $E_1 \lesssim E_2$ and $E_2 \lesssim E_1$.

2.1 Series-parallel metrics

In this section, we give a quick review of series-parallel graphs and their shortest-path metrics. For more information, we refer to [13].

Series-parallel graphs, metrics, and bundles. A *series-parallel graph* (SP graph) $G = (V,E)$ has *endpoints* $s,t \in V$. The class of SP graphs can be constructed inductively as follows. A single edge (s,t) is an SP graph with endpoints s,t. Given two SP graphs H and H' with endpoints $s,t \in V(H)$ and $s',t' \in V(H')$, we can form a new SP graph by taking the disjoint union of H and H' and then (1) identifying t with s' (called series composition) or (2) identifying s with s' and t with t' (called parallel composition). It is well-known that the class of treewidth-2 graphs is precisely the class of graphs whose 2-connected components are SP graphs.

A *series-parallel metric space* (SP metric) is a metric space (X,d) which arises from the shortest-path distance on an SP graph $G = (V,E)$, where each edge $e \in E$ is assigned a non-negative length $\mathsf{len}(e)$. We say that (X,d) has an associated (weighted) SP graph $G = (V,E)$ (where the edge weights are naturally given by $d(u,v)$ for $(u,v) \in E$). We will sometimes refer to the endpoints of G as the endpoints of X. An *SP bundle metric* is an SP metric space (X,d) with endpoints $s,t \in X$ such that every simple s-t path in the associated graph G has the same length, and such that G is 2-edge-connected.

Series-parallel bundle construction. We can specify an alternate inductive construction sequence for SP bundle metrics. The two-point metric space $\{s,t\}$ with any value $d(s,t) \in \mathbb{R}_+$ is an SP bundle metric with the obvious associated graph. Given any SP bundle metric X with associated graph $G = (V,E)$, we can choose an edge $e = (s,t) \in E$ and consider the new graph G' specified as follows. For some $k \in \mathbb{N}$, $V(G') = V \cup \{x_1, \dots, x_k\}$ and $E(G') = E \cup \{(s,x_i),(x_i,t)\}_{i=1}^k \setminus (s,t)$. Furthermore, we construct a new metric (X',d) on $X' = V(G')$ by extending the distance function d to $X' \setminus X$ as follows: For some numbers $d_1, \dots, d_k \in \mathbb{R}_+$ with $d_i \le d(s,t)$ for every $i \in [k]$, we set $d(s,x_i) = d_i$ and $d(x_i,t) = d(s,t) - d_i$. The vertices $x_1, \dots, x_k$ are called the *children of the edge* $(s,t) \in E(G)$.

Thus for any SP bundle metric (X,d), we have a construction sequence $\mathcal{G} = \langle G_0, G_1, \dots, G_m \rangle$, where G_m is the graph associated to X. We say that the construction sequence is *ε-regular* if $\left(\frac{1}{2} - \varepsilon\right) d(s,t) \le d(s,x) \le \left(\frac{1}{2} + \varepsilon\right) d(s,t)$ whenever $x \in V(G_{j+1})$ is a child of the edge $(s,t) \in E(G_j)$.

We can naturally associate a rooted tree $T_{\mathcal{G}}$ with the construction sequence $\mathcal{G} = \langle G_0, G_1, \dots, G_m \rangle$ as follows. First, we have $V(T_{\mathcal{G}}) = V(G_m)$. Now, for any $v \notin V(G_0)$, there exists a unique number $\mathsf{lev}(v) \in \{1, \dots, m\}$ so that $v \in V(G_{\mathsf{lev}(v)})$ but $v \notin V(G_{\mathsf{lev}(v)-1})$. If $V(G_0) = \{s_0, t_0\}$ then we assign arbitrarily $\mathsf{lev}(s_0) = -1, \mathsf{lev}(t_0) = 0$. We define s_0 as the root of $T_{\mathcal{G}}$, and

$$E(T_{\mathcal{G}}) = \{(s,x) : (s,t) \in E(G_j), x \in V(G_{j+1}) \text{ is a child of } (s,t), \text{ and } \mathsf{lev}(s) > \mathsf{lev}(t)\}.$$

Observe that this is well-defined since, for any $(s,t) \in E(G_j)$, we have $\mathsf{lev}(s) \neq \mathsf{lev}(t)$. Finally, for $v \in V(T_{\mathcal{G}})$, we define $A_{\mathcal{G}}(v) \subseteq V(T_{\mathcal{G}})$ to be the set of ancestors of v in $T_{\mathcal{G}}$, and we define $\Delta_{\mathcal{G}}(v) \subseteq V(T_{\mathcal{G}})$ as the set of descendants of v in $T_{\mathcal{G}}$.

3. EMBEDDINGS

In this section, we prove Theorem 1.5. The proof proceeds over the next four sections. In Sections 3.1 and 3.2, we prove the theorem for regular SP bundle metrics. In Section 3.3, we prove the theorem for tree metrics, and in Section 3.4, we combine all these results with the Structure Theorem (Thm 4.1) to obtain Theorem 1.5.

We now discuss the consequences. The algorithmic applications of these theorems are addressed in Section 1. First, we show that the lower bound of Newman and Rabinovich [26, 23] is tight for series-parallel graphs. This follows immediately by setting $s = 1$ above.

COROLLARY 3.1. (**Dominated L_1 embeddings**) *For any n-point series-parallel metric (X,d), one has $c_1^{\mathrm{dom}}(X) = O(\sqrt{\log n})$.*

Next, we show that the L_p distortion lower bound of Lee and Naor [18] is also tight. The correspondence with L_p embeddings also shows that the upper bound of Theorem 1.5 is tight.

COROLLARY 3.2. (**L_p embeddings,** $1 < p < 2$) *There exists a constant $K \ge 1$ such that for any n-point series-parallel metric and any number $p = p(n) \in [1,2]$, we have $c_p(X) \le K\sqrt{(p-1)\log n}$.*

PROOF. Setting $s = \exp(1/(p-1))$ in Theorem 1.5, we have

$$\|F(x) - F(y)\|_p \ge \|F(x) - F(y)\|_1 \ge \frac{d(x,y)}{O(\sqrt{(p-1)\log n})}.$$

On the other hand,

$$\begin{aligned} \|F(x) - F(y)\|_p &= \|(F(x) - F(y))^p\|_1^{1/p} \\ &\le \|F(x) - F(y)\|_1^{1/p} \cdot \|F(x) - F(y)\|_\infty^{(p-1)/p} \\ &\le d(x,y)^{1/p} \cdot s^{(p-1)/p} \cdot d(x,y)^{1-1/p} \\ &\le \exp(1) \cdot d(x,y). \end{aligned}$$

□

Finally, we match the dimension reduction lower bound of Brinkman and Charikar [6]. This follows from combining Theorem 1.5 with Theorem 1.3.

COROLLARY 3.3. (**Dimension reduction in L_1**) *Let (X,d) be any n-point series-parallel metric. Then for every $D \ge 1$, there exists a D-embedding $f : X \to \ell^d$ with $d = n^{1/\Omega(D^2)}$.*

3.1 Bundles: Basic construction

Let $D \geq 2$ be given. Let (X, d) be an n-point SP bundle metric, with an associated weighted series-parallel graph $G = (X, E)$, and endpoints $s_0, t_0 \in X$.

Let $\mathcal{G} = \langle G_0, G_1, \ldots, G_m = G \rangle$ be the construction sequence for G, and let $T_\mathcal{G}$ be the corresponding construction tree. Throughout this section, we use $\Gamma(T_\mathcal{G})$ to denote the set of nodes in $T_\mathcal{G}$ with *more than one child,* and we write $\Phi = \Phi(T_\mathcal{G}) = |\Gamma(T_\mathcal{G})|$ for the number of such nodes.

We will specify a set $R_\mathcal{G}$ and a random mapping $F : X \to \mathbb{R}$ inductively. G_0 has two nodes $s_0, t_0 \in V(G_0)$, and a single edge (s_0, t_0). We let $F(s_0) = 0$ and $F(t_0) = d(s_0, t_0)$.

Now, given a mapping $F : V(G_{k-1}) \to \mathbb{R}$, we randomly extend it to a map on $V(G_k) \supseteq V(G_{k-1})$. One forms G_k from G_{k-1} by selecting a single edge $(s,t) \in E(G_k)$, and adding child vertices $x_1, x_2, \ldots, x_r$ for some $r \geq 1$. We recall that to every child x_i is associated a set of descendent vertices $\Delta_\mathcal{G}(x_i)$. By reordering, we may assume that $|\Gamma(T_\mathcal{G}) \cap \Delta_\mathcal{G}(x_1)| \geq |\Gamma(T_\mathcal{G}) \cap \Delta_\mathcal{G}(x_i)|$ for all $1 \leq i \leq r$. We put $R_\mathcal{G} \leftarrow R_\mathcal{G} \cup \{x_2, x_3, \ldots, x_r\}$. Assume, without loss of generality, that $F(s) \leq F(t)$.

Now, we set $S = \emptyset$ if

$$|F(s) - F(t)| > d(s,t)\left(4 \cdot \exp\left(\frac{1 + \log_2 \Phi}{D^2}\right)\right)$$

Otherwise, $S = \{2, 3, \ldots, r\}$. For $i \notin S$, we define

$$F(x_i) = F(s) + \frac{d(s, x_i)}{d(s,t)}|F(s) - F(t)|. \tag{1}$$

If $i \in S$, letting $\{\varepsilon(x_i)\}_{i \in S}$ be a system of i.i.d. uniform ± 1 random variables, we define

$$F(x_i) = \begin{cases} F(s) + \frac{d(s,x_i)}{d(s,t)}|F(s) - F(t)|\left(1 + \frac{\varepsilon(x_i)}{D}\right) \\ \qquad\qquad\qquad \text{if } d(s,x_i) \leq d(t,x_i) \\ F(t) - \frac{d(t,x_i)}{d(s,t)}|F(s) - F(t)|\left(1 + \frac{\varepsilon(x_i)}{D}\right) \\ \qquad\qquad\qquad \text{if } d(t,x_i) < d(s,x_i) \end{cases} \tag{2}$$

3.2 Bundles: Analysis

We now analyze the random map $F : X \to \mathbb{R}$ defined in Section 3.1. We start with some simple observations.

Lemma 3.4. *If $x \in X$, and $A_\mathcal{G}(x)$ are the ancestors of x, then $|A_\mathcal{G}(x) \cap R_\mathcal{G}| \leq 1 + \log_2 \Phi$.*

Proof. Recall that there is a rooted tree $T_\mathcal{G}$ associated to the construction sequence $\mathcal{G}$ such that $A_\mathcal{G}(x)$ lists precisely the vertices from the root to x in $T_\mathcal{G}$. If $y \in R_\mathcal{G}$, then it was chosen because y had at least one distinct sibling z with $|\Gamma(T_\mathcal{G}) \cap \Delta_\mathcal{G}(z)| \geq |\Gamma(T_\mathcal{G}) \cap \Delta_\mathcal{G}(y)|$. Furthermore, if p is the parent of y, then $p \in \Gamma(T_\mathcal{G})$. This implies (inductively) that if $|A_\mathcal{G}(x) \cap R_\mathcal{G}| = k$, then $\Phi = |\Gamma(T_\mathcal{G})| \geq 2^{k-1}$, hence $k \leq 1 + \log_2 \Phi$. □

Lemma 3.5. (**Monotonicity**) *If $x, y \in X$ are two vertices lying along a simple s_0-t_0 path in G for which $d(x, s_0) \leq d(y, s_0)$, then $F(x) \leq F(y)$ holds.*

Proof. This follows by verifying that for any edge $(s,t) \in E(G_k)$ with $F(s) \leq F(t)$, and any child vertex $x \in V(G_{k+1})$, we have $F(s) \leq F(x) \leq F(t)$. If $F(x)$ is defined according to equation (1), this is immediate. Otherwise $F(x)$ is defined according to (2); consider, for instance, the case $d(s,x) \leq d(t,x)$. In this case, $\frac{d(s,x)}{d(x,t)} \leq \frac{1}{2}$ and since $D \geq 1$, we have $1 + \frac{1}{D} \leq 2$. It follows that $F(s) \leq F(x) \leq F(t)$. The other case is similar. □

We now define a random process with respect to a set of h numbers $\{\delta_1, \delta_2, \ldots \delta_h\}$ and a truncation point T. The process is defined inductively as follows $Y_0 = 1$, and for $1 \leq t \leq h$,

$$Y_t = \begin{cases} Y_{t-1} & Y_{t-1} > T \\ (1 + \varepsilon_t \delta_t) Y_{t-1} & \text{otherwise,} \end{cases}$$

where $\{\varepsilon_t\}_{t=1}^h$ is a system of i.i.d. uniform ± 1 random variables. We define $Y_t = Y_h$ for $t \geq h$, and denote the random variable Y_t by $\mathbf{Y}_t(\delta_1, \ldots, \delta_h; T)$. Observe that $\mathbf{Y}_t(\delta_1, \ldots, \delta_h; T)$ $= \mathbf{Y}_t(\delta_1, \ldots, \delta_t; T)$. We first argue that this process controls the distribution of edge lengths under F.

For two random variables A, B (defined on possibly different probability spaces), we use the notation $A \sim B$ to mean that A and B are equal in distribution.

Lemma 3.6 (**Edge distribution**). *If $(s,t) \in E(G_k)$ is any edge, then*

$$|F(s) - F(t)| \sim d(s,t) \cdot \mathbf{Y}_h(\delta_1, \ldots, \delta_h; T),$$

where $0 \leq \delta_i \leq \frac{1}{D}$ for $1 \leq i \leq h$, $h \leq 1 + \log_2 \Phi$, and $T = 4 \cdot \exp\left(\frac{1+\log_2 \Phi}{D^2}\right)$.

Proof. Inductively, we assume that $|F(s) - F(t)| \sim d(s,t) \cdot \mathbf{Y}_{h-1}(\delta_1, \ldots, \delta_{h-1}; T)$, and let $x \in V(G_{k+1})$ be a child of the edge $(s,t) \in E(G_k)$. If $F(x)$ is defined according to (1), then

$$\begin{aligned} |F(x) - F(s)| &\sim d(x,s) \cdot \mathbf{Y}_{h-1}(\delta_1, \ldots, \delta_{h-1}; T) \\ |F(x) - F(t)| &\sim d(x,t) \cdot \mathbf{Y}_{h-1}(\delta_1, \ldots, \delta_{h-1}; T). \end{aligned}$$

In particular, this happens if $\mathbf{Y}_{h-1}(\delta_1, \ldots, \delta_{h-1}; T) > 4 \cdot \exp\left(\frac{1+\log_2 \Phi}{D^2}\right) = T$.

Otherwise, assume that $d(x,s) \leq d(x,t)$. Then we have

$$\begin{aligned} |F(x) - F(s)| &= \frac{d(x,s)}{d(s,t)}|F(s) - F(t)|\left(1 + \frac{\varepsilon(x)}{D}\right) \\ &\sim d(x,s)\left(1 + \frac{\varepsilon(x)}{D}\right)\mathbf{Y}_{h-1}(\delta_1, \ldots, \delta_{h-1}; T) \\ &\sim d(x,s) \cdot \mathbf{Y}_h(\delta_1, \ldots, \delta_{h-1}, \tfrac{1}{D}; T). \end{aligned}$$

Additionally,

$$\begin{aligned} &|F(x) - F(t)| \\ &\quad = |F(s) - F(t)| - |F(s) - F(x)| \\ &\quad = |F(s) - F(t)|\left(1 - \frac{d(s,x)}{d(s,t)}\left(1 + \frac{\varepsilon(x)}{D}\right)\right) \\ &\quad \sim d(x,t)\left(1 - \frac{\varepsilon(x)}{D}\frac{d(s,x)}{d(x,t)}\right)\mathbf{Y}_{h-1}(\delta_1, \ldots, \delta_{h-1}; T) \\ &\quad \sim d(x,t) \cdot \mathbf{Y}_h(\delta_1, \ldots, \delta_{h-1}, \delta_h; T), \end{aligned}$$

where $\delta_h = \frac{1}{D}\frac{d(s,x)}{d(s,t)} \leq \frac{1}{D}$. The case where $d(x,s) > d(x,t)$ is identical.

Finally, the preceding proof shows that

$$h \leq |A_\mathcal{G}(x) \cap R_\mathcal{G}| \leq 1 + \log_2 \Phi$$

by Lemma 3.4. □

Corollary 3.7. *If $x, y \in X$ lie on a simple s_0-t_0 path in G, then $\mathbb{E}\,|F(x) - F(y)| = d(x, y)$. In particular, for every $x, y \in X$, we have $\mathbb{E}\,|F(x) - F(y)| \le d(x, y)$.*

Proof. If x, y lie on a simple s_0-t_0 path $x = v_0, \ldots, v_q = y$ then, because (X, d) is an SP bundle, this is also a shortest path. Applying Lemma 3.5 and using linearity of expectation yields

$$\mathbb{E}\,|F(x) - F(y)| = \sum_{i=0}^{q-1} \mathbb{E}\,|F(v_i) - F(v_{i+1})|.$$

For any edges $(u, v) \in E(G) = E(G^m)$, applying Lemma 3.6, we see that

$$|F(u) - F(v)| \sim d(u, v) \cdot \mathbf{Y}_h(\delta_1, \ldots, \delta_h; T)$$

for some choice of parameters $\{\delta_i\}, T$. But the process $\{\mathbf{Y}_t(\delta_1, \ldots, \delta_h; T)\}_{t \ge 0}$ is easily seen to be a martingale, hence $\mathbb{E}\,\mathbf{Y}_h(\delta_1, \ldots, \delta_h; T) = \mathbb{E}\,\mathbf{Y}_0(\delta_1, \ldots, \delta_h; T) = 1$. Since (v_i, v_{i+1}) is an edge for $0 \le i \le q-1$, we conclude that $\mathbb{E}\,|F(x) - F(y)| = \sum_{i=0}^{q-1} d(v_i, v_{i+1}) = d(x, y)$. The final remark of the corollary follows from the fact that every edge in G lies along a simple s_0-t_0 path. $\square$

We come now to our main probabilistic lemma. Let $\mathbf{Y}^*_t(\delta_1, \ldots, \delta_h; T) = \max_{0 \le i \le t} \mathbf{Y}_t(\delta_1, \ldots, \delta_h; T)$.

Lemma 3.8. *Suppose that for every $i \in [h]$, $\delta_i \in [0, \frac{1}{2}]$, and $T \ge 4 \cdot e^{\sum_{i=1}^h \delta_i^2}$. Then,*

$$\Pr\left(\mathbf{Y}^*_h(\delta_1, \ldots, \delta_h; T) < T\right) \ge \frac{1}{2} \text{ and}$$
$$\mathbb{E}\left[\mathbf{Y}_h(\delta_1, \ldots, \delta_h; T) \,\middle|\, \mathbf{Y}^*_h(\delta_1, \ldots, \delta_h; T) < T\right] \ge \frac{1}{2}.$$

Proof. We begin by setting $\mathbf{Y}_t = \mathbf{Y}_t(\delta_1, \ldots, \delta_h; T)$ and $\mathbf{Y}^*_t = \mathbf{Y}^*_t(\delta_1, \ldots, \delta_h; T)$ for $t \ge 0$. Since $\{\mathbf{Y}_t\}_{t \ge 0}$ is a martingale, Doob's maximal inequality (see e.g. [9, §4.4]) yields

$$\Pr(\mathbf{Y}^*_h \ge T) \le \frac{\mathbb{E}\,\mathbf{Y}_h}{T} = \frac{1}{T} \le \frac{1}{2},$$

proving the first claim of the lemma.

Next, we note that $\mathbb{E}\,\mathbf{Y}_h^2 \le \prod_{i=1}^h (1 + \delta_i^2)$, since $\mathbb{E}\,\mathbf{Y}_t^2 \le (1 + \delta_t^2)\mathbb{E}\,\mathbf{Y}_{t-1}^2$. Moreover, $\{\mathbf{Y}_t^2\}_{t \ge 0}$ is a sub-martingale; employing Doob's inequality again,

$$\Pr\left(\mathbf{Y}^*_h \ge T\right) \le \frac{\mathbb{E}\,\mathbf{Y}_h^2}{T^2} \le \frac{\prod_{i=1}^h (1 + \delta_i^2)}{T^2} \le \frac{e^{\sum_{i=1}^h \delta_i^2}}{T^2}. \quad (3)$$

Thus we have

$$\begin{aligned}
\mathbb{E}\left[\mathbf{Y}_h \mid \mathbf{Y}^*_h < T\right] &\ge \mathbb{E}\left[\mathbf{Y}_h\right] - \Pr\left(\mathbf{Y}^*_h \ge T\right) \cdot \mathbb{E}\left[\mathbf{Y}_h \mid \mathbf{Y}^*_h \ge T\right] \\
&\ge 1 - \frac{e^{\sum_{i=1}^h \delta_i^2}}{T^2} \cdot \mathbb{E}[\mathbf{Y}_h \mid \mathbf{Y}^*_h \ge T] \\
&\ge 1 - \frac{e^{\sum_{i=1}^h \delta_i^2}}{T^2} \cdot (2T) \\
&\ge 1 - \frac{2 \cdot e^{\sum_{i=1}^h \delta_i^2}}{T} \\
&\ge \frac{1}{2}.
\end{aligned}$$

where in the penultimate line, we have used the fact that $\Pr(\mathbf{Y}_h < 2T) = 1$. $\square$

We now show a lower bound on $\mathbb{E}\,|F(x) - F(y)|$ for $x, y \in X$.

Lemma 3.9 **(Lower bound for regular bundles)**. *If the construction sequence $\mathcal{G}$ is $\frac{1}{4}$-regular, then for any $x, y \in X$, we have $\mathbb{E}\,|F(x) - F(y)| \ge \frac{d(x,y)}{128D}$.*

Proof of Lemma 3.9. If $x, y \in X$ lie along a simple s_0-t_0 path, then we are done by Corollary 3.7. Otherwise, there is a distinct pair of vertices $u \in A_{\mathcal{G}}(x)$ and $v \in A_{\mathcal{G}}(y)$, a value $0 \le k \le m$, and an edge $(s, t) \in E(G_k)$ such that u and v are children of (s, t) (one of s or t is the least common ancestor of u and v in $T_{\mathcal{G}}$). Without loss of generality, we may assume that $F(s) \le F(t)$.

First, we can assume that that $|d(x, s) - d(y, s)| \le \frac{d(x,y)}{4D}$ and $|d(x, t) - d(y, t)| \le \frac{d(x,y)}{4D}$ since using Corollary 3.7, we have the inequalities

$$\begin{aligned}
\mathbb{E}\,|F(x) - F(y)| &\ge \Big|\mathbb{E}\,|F(x) - F(s)| - \mathbb{E}\,|F(y) - F(s)|\Big| \\
&= |d(x, s) - d(y, s)|, \\
\mathbb{E}\,|F(x) - F(y)| &\ge \Big|\mathbb{E}\,|F(x) - F(t)| - \mathbb{E}\,|F(y) - F(t)|\Big| \\
&= |d(x, t) - d(y, t)|.
\end{aligned}$$

By construction, at least one of u or v is in $R_{\mathcal{G}}$. Without loss of generality, suppose $u \in R_{\mathcal{G}}$. Since $F(x)$ and $F(y)$ are independent conditioned on $\{F(s), F(t)\}$, we need only exhibit some variance in $F(x)$ conditioned on $\{F(s), F(t)\}$ in order to prove a lower bound on $\mathbb{E}\,|F(x) - F(y)|$. Since $\mathcal{G}$ is $\frac{1}{4}$-regular, we have $\frac{1}{4} d(s, t) \le d(s, u) \le \frac{3}{4} d(s, t)$. Let $\mathcal{E}_{\mathrm{rand}}$ be the event that $F(u)$ is determined by line (2). Letting $T = 4 \cdot \exp\left(\frac{1 + \log_2 \Phi}{D^2}\right)$, and recalling that $u \in R_{\mathcal{G}}$, we see that $\mathcal{E}_{\mathrm{rand}}$ occurs precisely when the event $\mathcal{E}_{\mathrm{short}} = \{|F(s) - F(t)| \le T \cdot d(s, t)\}$ occurs. Applying Lemma 3.6, we see additionally that $|F(s) - F(t)| \sim Y(\delta_1, \ldots, \delta_h; T)$ for some sequence $\{\delta_i\}_{i=1}^h$ satisfying the conditions of the lemma.

In particular, applying Lemma 3.8, we see that $\Pr(\mathcal{E}_{\mathrm{rand}}) \ge \frac{1}{2}$, and

$$\mathbb{E}\left[|F(s) - F(t)| \,\middle|\, \mathcal{E}_{\mathrm{rand}}\right] \ge \frac{1}{2}\,\mathbb{E}\,|F(s) - F(t)| = \frac{d(s, t)}{2}.$$

Assuming that $\mathcal{E}_{\mathrm{rand}}$ occurs, we have two cases.

1. $d(u, s) \le d(u, t)$: In this case,
$$\frac{|F(u) - F(s)|}{d(u, s)} = \frac{|F(s) - F(t)|}{d(s, t)} \left(1 + \frac{\varepsilon(u)}{D}\right) \quad (4)$$
2. $d(u, t) < d(u, s)$: In this case,
$$\frac{|F(u) - F(s)|}{d(u, s)} = \frac{|F(s) - F(t)|}{d(s, t)} \left(1 - \frac{d(u, t)}{d(u, s)} \frac{\varepsilon(u)}{D}\right) \quad (5)$$

In this second case, observe that since $\mathcal{G}$ is $\frac{1}{4}$-regular, we have $\frac{d(u,t)}{d(u,s)} \ge \frac{1}{4}$. Symmetric statements hold for $|F(u) - F(t)|$.

Assume without loss of generality that $d(x, s) \le d(u, s)$. In this case, using the same reasoning as in Corollary 3.7, we have

$$\begin{aligned}
&\mathbb{E}\left[|F(s) - F(x)| \,\middle|\, F(s), F(t)\right] \\
&\qquad = \frac{d(x, s)}{d(u, s)}\,\mathbb{E}\left[|F(s) - F(u)| \,\middle|\, F(s), F(t)\right].
\end{aligned}$$

Using (4) or (5), along with $\mathbb{E}\left[|F(s) - F(t)| \,\middle|\, \mathcal{E}_{\mathrm{rand}}\right] \ge \frac{1}{2} d(s, t)$, we have

$$\mathbb{E}\left[|F(s) - F(x)| \,\middle|\, \mathcal{E}_{\mathrm{rand}}, \varepsilon(u) = \varepsilon\right] = \rho\, d(x, s) \left(1 + \lambda \frac{\varepsilon}{D}\right),$$

where $|\lambda| \geq \frac{1}{4}$, and $\rho \geq \frac{1}{2}$, where we have used the fact that $|F(s)-F(t)|$ and $\varepsilon(u)$ are independent, even conditioned on $\mathcal{E}_{\text{rand}}$. Since $\varepsilon(u) \in \{-1, 1\}$ uniformly at random and $F(x)$ and $F(y)$ are independent conditioned on $\{F(s), F(t)\}$, this variance in $|F(s) - F(x)|$ due to $\varepsilon(u)$ translates to a lower bound for $F(x)$ and $F(y)$, i.e.

$$\mathbb{E}\,|F(x)-F(y)| \geq \Pr(\mathcal{E}_{\text{rand}}) \cdot \frac{1}{2} \cdot \frac{|\lambda|}{D} \cdot \rho\, d(x,s) \geq \frac{d(x,s)}{32D}. \quad (6)$$

Now, if $d(x,y) = d(x,s) + d(s,y)$, then from our assumption $|d(x,s) - d(y,s)| \leq \frac{d(x,y)}{4D}$, we see that $d(x,s) \geq d(x,y)/4$, and (6) finishes the proof. Otherwise, $d(x,y) = d(x,t)+d(t,y)$. In this case, we know that $d(x,s) \geq \frac{1}{4} d(s,t)$, else $|d(x,s) - d(y,s)| \leq \frac{d(x,y)}{4D}$ would imply that the shortest x-y path goes through s which we have assumed is not the case. So again, $d(x,s) \geq \frac{1}{4} d(x,y)$, and again (6) finishes the proof. $\square$

Combining Lemma 3.9 and Corollary 3.7, we arrive at the main result of this section.

Theorem 3.10. *If (X,d) is a $\frac{1}{4}$-regular SP bundle with construction tree $T_{\mathcal{G}}$ and $\Phi = \Phi(T_{\mathcal{G}})$, then for every $s \geq 1$, there exists a random mapping $F : X \to \mathbb{R}$ which satisfies the following for every $x, y \in X$,*

1. $\|F(x) - F(y)\|_\infty \leq s \cdot d(x,y)$.
2. $\min\left\{1, \sqrt{\frac{\log s}{\log \Phi}}\right\} d(x,y) \lesssim \|F(x) - F(y)\|_1 \leq d(x,y)$.

Proof. Observe that for any $D \geq 2$, setting $T = 4 \cdot \exp\left(\frac{1+\log_2 \Phi}{D^2}\right)$, the random map $F : X \to \mathbb{R}$ defined in this section satisfies $\|F(x) - F(y)\|_\infty \leq 2T \cdot d(x,y)$ and $\frac{d(x,y)}{128D} \leq \|F(x) - F(y)\|_1 \leq d(x,y)$. The first inequality follows from Lemma 3.6 and the fact that $\|\mathbf{Y}_h(\delta_1, \ldots, \delta_h; T)\|_\infty \leq 2T$ as long as $\delta_i \leq \frac{1}{2}$ for every $i \in [h]$ (which holds by construction). The second set of inequalities follows from Lemma 3.9 and Corollary 3.7.

Choosing $D \approx \max\left\{2, \sqrt{\frac{\log \Phi}{\log s}}\right\}$ yields $\|F(x)-F(y)\|_\infty = O(s) \cdot d(x,y)$, in addition to (2) above. Rescaling the map F by a constant yields condition (1) exactly, finishing the proof. $\square$

3.3 Embeddings for trees

We now handle the case of trees.

Theorem 3.11. *If (X,d) is the shortest-path metric on an n-point weighted tree, then for every $s \geq 1$, we have $c_1^{\text{dom}(s)} = O\left(\sqrt{\frac{\log n}{s}}\right)$.*

Proof. Let $T = (X, E)$ be the associated weighted tree, and orient T according to an arbitrary root $r \in X$. An *edge coloring of* T is a map $\chi : E \to \mathbb{N}$. We recall that a *monotone path in* T is a contiguous subset of some root-leaf path. An edge coloring $\chi : E \to \mathbb{N}$ is called *monotone* if every color class $\chi^{-1}(c)$ is a monotone path in T. The following lemma is well-known (see, e.g. [21, 19]—our notion is a generalization of the "caterpillar dimension" of T).

Lemma 3.12. *Every n-point tree T admits a monotone edge coloring such that every root-leaf path contains at most $O(\log n)$ colors.*

Let $\chi : E \to \mathbb{N}$ be the coloring guaranteed by Lemma 3.12, and let $s \geq 1$ be given. Let $\{\varepsilon_i\}_{i=1}^\infty$ be a set of i.i.d. random variables satisfying $\Pr(\varepsilon_i = 0) = 1 - \frac{1}{s}$, $\Pr(\varepsilon_i = 1) = \frac{1}{2s}$, and $\Pr(\varepsilon_i = -1) = \frac{1}{2s}$. We define a random embedding $F : X \to \mathbb{R}$ as follows. For a point $x \in X$, let $s_1, s_2, \ldots, s_k$ be the set of maximal χ-monochromatic segments on the path from the root to x. Then we set

$$F(x) = s \cdot \sum_{i=1}^{k} \mathsf{len}(s_i) \cdot \varepsilon_{\chi(s_i)},$$

where we have extended χ to the monochromatic segments $\{s_i\}_{i=1}^k$ in the natural way.

First, it is easy to see that $\|F\|_{\text{Lip}} \leq s$ with probability 1. Now, fix $x, y \in X$ with least common ancestor $u \in X$, and let $s_1, \ldots, s_k$ and $t_1, \ldots, t_h$ be the set of maximal χ-monochromatic segments on the path from u to x and y, respectively. Note that since χ is monotone, the color classes of these segments are all pairwise disjoint. First, we have

$$\begin{aligned} &\mathbb{E}\left[|F(x) - F(y)|\right] \\ &\leq \quad s \cdot \left(\sum_{i=1}^{k} \mathsf{len}(s_i)\, \mathbb{E}|\varepsilon_{\chi(s_i)}| + \sum_{i=1}^{h} \mathsf{len}(t_i)\, \mathbb{E}|\varepsilon_{\chi(t_i)}|\right) \\ &= \quad \sum_{i=1}^{k} \mathsf{len}(s_i) + \sum_{i=1}^{h} \mathsf{len}(t_i) = d(x,y). \end{aligned}$$

On the other hand, observe that $\varepsilon_i \sim |\varepsilon_i| \cdot \sigma_i$ where $\{\sigma_i\}_{i=1}^\infty$ is a family of i.i.d. ± 1 Bernoulli random variables independent from the family $\{\varepsilon_i\}_{i=1}^\infty$. We use $\mathbb{E}_\sigma$ and $\mathbb{E}_\varepsilon$, respectively, to denote expectations over these random variables. Using Fubini's theorem and Khintchine's inequality (see e.g. [24, §5.5]), we have

$$\begin{aligned} &\mathbb{E}\left[|F(x) - F(y)|\right] \\ &= \quad s \cdot \mathbb{E}\left|\sum_{i=1}^{k} \mathsf{len}(s_i)\varepsilon_{\chi(s_i)} + \sum_{i=1}^{h} \mathsf{len}(t_i)\varepsilon_{\chi(t_i)}\right| \\ &= \quad s \cdot \mathbb{E}_\varepsilon \mathbb{E}_\sigma \left|\sum_{i=1}^{k} \mathsf{len}(s_i)|\varepsilon_{\chi(s_i)}|\sigma_{\chi(s_i)} + \sum_{i=1}^{h} \mathsf{len}(t_i)|\varepsilon_{\chi(t_i)}|\sigma_{\chi(t_i)}\right| \\ &\approx \quad s \cdot \mathbb{E}_\varepsilon \sqrt{\sum_{i=1}^{k} \mathsf{len}(s_i)^2\, \varepsilon_{\chi(s_i)}^2 + \sum_{i=1}^{h} \mathsf{len}(t_i)^2\, \varepsilon_{\chi(t_i)}^2} \end{aligned}$$

Now, we let $M = |\{i \in [k] : \varepsilon_{\chi(s_i)} \neq 0\}| + |\{i \in [h] : \varepsilon_{\chi(t_i)} \neq 0\}|$. Observe that $\mathbb{E}\, M = (k+h)/s$. Let $\mathcal{E}$ be the event that $M \leq \max\left\{1, \frac{2(k+h)}{s}\right\}$, and note that $\Pr(\mathcal{E}) \geq \frac{1}{2}$. Using Cauchy-Schwartz and $k, h \leq O(\log n)$, we have

$$\begin{aligned} &\mathbb{E}\left[|F(x) - F(y)|\right] \\ &\gtrsim \quad s \cdot \Pr\left[\mathcal{E}\right] \cdot \mathbb{E}\left[\sqrt{\sum_{i=1}^{k} \mathsf{len}(s_i)^2\, \varepsilon_{\chi(s_i)}^2 + \sum_{i=1}^{h} \mathsf{len}(t_i)^2\, \varepsilon_{\chi(t_i)}^2}\;\middle|\; \mathcal{E}\right] \\ &\gtrsim \quad \frac{s}{\max\{1, \sqrt{(\log n)/s}\}} \cdot \\ &\qquad \left(\sum_{i=1}^{k} \mathsf{len}(s_i)\mathbb{E}\left[\varepsilon_{\chi(s_i)}^2 \,|\, \mathcal{E}\right] + \sum_{i=1}^{h} \mathsf{len}(t_i)\mathbb{E}\left[\varepsilon_{\chi(t_i)}^2 \,|\, \mathcal{E}\right]\right) \\ &\gtrsim \quad \min\left\{1, \sqrt{\frac{s}{\log n}}\right\} d(x,y), \end{aligned}$$

noting that $\mathbb{E}\left[\varepsilon_{\chi(s_i)}^2 \,|\, \mathcal{E}\right] = \mathbb{E}\left[\varepsilon_{\chi(t_i)}^2 \,|\, \mathcal{E}\right] \approx 1/s$. $\square$

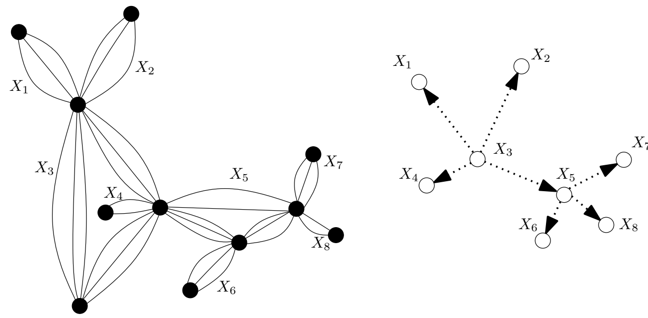

Figure 1: An SP bundle tree and the corresponding oriented tree

We end this section with the follow question.

QUESTION 6. *What is the right bound for* $c_1^{\mathrm{dom}(s)}(X)$ *when* X *is an* n*-point tree metric?*

If the answer is, say, $O\left(\sqrt{\frac{\log\log n}{\log s}}\right)$, this would show (via Theorem 1.3) that every n-point tree metric admits an $O(1)$-distortion embedding into $\ell_1^{O(\log n)^{1+\delta}}$ for every $\delta > 0$, improving over the $O(\log n)^2$ bound of [7].

3.4 Reduction to regular SP bundles

In this section, we complete the proof of Theorem 1.5. First, we define an *oriented SP bundle tree* as a metric space (X, d) with an associated graph G satisfying the following property. There exists a set of SP bundle metrics $V_\Lambda = \{X_1, X_2, \ldots, X_k\}$ with $|X_i \cap X_j| \le 1$ for every $i, j \in [k]$ and a tree $\Lambda_G = (V_\Lambda, D_\Lambda)$ oriented with respect to a fixed root $r_\Lambda \in V_\Lambda$ (i.e. with all arcs pointing away from r_Λ). Furthermore, $X = \bigcup_{i=1}^k X_i$ and the bundles $X_1, \ldots, X_k$ satisfy the following property: $|X_i \cap X_j| = 1$ if and only if $(X_i, X_j) \in D_\Lambda$, and in this case $X_i \cap X_j$ contains an endpoint of X_j. See Figure 1. We say that the oriented SP bundle tree is ε-regular if every X_i has an ε-regular construction sequence.

By applying the Structure Theorem (stated and proved in Section 4), it suffices to prove Theorem 1.5 for a $\frac{1}{4}$-regular oriented SP bundle tree (X, d) where each bundle component X_i has a construction sequence $\mathcal{G}_i$ satisfying $\Phi(\mathcal{G}_i) = O(n)$ and an oriented tree Λ_G with $|\Lambda_G| = O(n)$.

We first define a tree metric $T = (X, E_T)$. We flatten every bundle X_i to a single path metric P_i with endpoints s_i and t_i and such that $d_{P_i}(x, s_i) = d(x, s_i)$ for every $x \in X_i$. Then using the node identifications between the X_i's that are present in G, we arrive at a (weighted) tree T. Let $s \ge 1$ be given and let $F_T : X \to \mathbb{R}$ be the random map guaranteed by Theorem 3.11 applied with parameter s.

Now let $s_1 \in X_1, s_2 \in X_2, \ldots, s_k \in X_k$ be chosen so that for every $(X_i, X_j) \in D_\Lambda$, we have $s_j \in X_i \cap X_j$. For each $i \in [k]$, let $F_i : X_i \to \mathbb{R}$ be the random map guaranteed by Theorem 3.10 applied with parameter s. We define a random map $F_G : X \to \mathbb{R}$ inductively on the structure of Λ_G. Assume that X_1 is the root of Λ_G, and then set $F_G(x) = F_1(x)$ for $x \in X_1$. Now if $(X_i, X_j) \in D_\Lambda$ and F_G has already been defined on X_i, then we extend F_G to X_j by setting $F_G(x) = \varepsilon_j \cdot [F_j(x) - F_j(s_j)] + F_i(s_j)$ for all $x \in X_j$, where $\{\varepsilon_j\}_{j=1}^k$ is a family of i.i.d. uniform ± 1 random variables.

It is easily checked that $\Pr(\|F_G\|_{\mathrm{Lip}} \le s) = 1$ since the same property holds for each F_i. We define a final random map $F : X \to \mathbb{R}$ by $F = \frac{1}{2} F_T \oplus \frac{1}{2} F_G$ (in other words, we choose F to be one of F_T or F_G each with probability $\frac{1}{2}$). Obviously $\Pr(\|F\|_{\mathrm{Lip}} \le s) = 1$.

Now consider any pair $u, v \in X$ with $u \in X_i$ and $v \in X_j$. We assume that a shortest path P from u to v in G intersects at least three distinct bundles (otherwise an easier variant of the following argument suffices). In this case, let X_k be the least common ancestor of X_i and X_j in Λ_G, and let $a_u, a_v \in X_k$ be the two nodes on which any u-v shortest path P enters and exits X_k (so the nodes are visited by P in order u, a_u, a_v, v). It is not difficult to see that $d(u, v) = d_T(u, a_u) + d_{X_k}(a_u, a_v) + d_T(a_v, v)$. As a consequence, we see that either $d_T(u, v) \ge \frac{1}{2} d(u, v)$ or $d_{X_k}(a_u, a_v) \ge \frac{1}{2} d(u, v)$. In the former case, we use $\mathbb{E}\,|F(u) - F(v)| \ge \mathbb{E}\,|F_T(u) - F_T(v)|$, and thus get the necessary contribution from F_T. In the latter case, we get the desired contribution using

$$\begin{aligned} \mathbb{E}\,|F(u) - F(v)| &\ge \frac{1}{2}\mathbb{E}\,|F_G(u) - F_G(v)| \\ &\ge \frac{1}{2}\mathbb{E}\,|F_k(a_u) - F_k(a_v)|, \end{aligned}$$

where the last inequality follows from the fact that $\mathbb{E}\,|F_G(a_u) - F_G(u)| = \mathbb{E}\,|F_G(a_v) - F_G(v)| = 0$ (by our use of the random variables $\{\varepsilon_j\}$), and the fact that $\{|F_G(a_u) - F_G(u)|, |F_G(a_v) - F_G(v)|, |F_G(a_u) - F_G(a_v)|\}$ are mutually independent random variables.

4. THE STRUCTURE THEOREM

For any rooted tree T, we recall that $\Phi(T)$ is the number of nodes in T with more than one child. The main theorem of this section follows.

THEOREM 4.1 (**Structure theorem for SP metrics**). *Let* (X, d) *be an arbitrary* n*-point SP metric, and* $\varepsilon > 0$. *Then there exist a pair of metric spaces* $(X_1, d_1), (X_2, d_2)$ *which satisfy the following conditions.*

1. *For* $i = 1, 2$, (X_i, d_i) *is an* ε*-regular oriented SP bundle tree with associated graph* G_i *and an associated oriented tree* Λ_i

2. *For* $i = 1, 2$, $|\Lambda_i| = O(n)$.

3. *Let* $\mathcal{F}$ *be the collection of all SP bundle metrics occuring as vertices of* Λ_i *for* $i = 1, 2$. *Then we can associate to every* $Y \in \mathcal{F}$ *a construction tree* $T_{\mathcal{G}_Y}$ *such that*
$$\sum_{Y \in \mathcal{F}} \Phi(T_{\mathcal{G}_Y}) \le O(n).$$

4. X *admits an* $O(1)$*-embedding into the product metric* $X_1 \times X_2$.

The preceding theorem follows from the next two results.

LEMMA 4.2 (**Regularization**). *Let* (X, d) *be an SP bundle metric. Then for every* $\varepsilon > 0$, *there exists an SP bundle metric* (X', d) *which contains* (X, d) *as a sub-metric, such that* (X', d) *has an associated construction sequence which is* ε*-regular. Furthermore, we can associate a construction tree* $T'_{\mathcal{G}}$ *to* (X', d) *such that* $\Phi(T'_{\mathcal{G}}) \le |X|$.

PROOF. If, at some point in the construction sequence, we have an edge (s,t) with child x for which, say, $d(s,x) < \varepsilon d(s,t)$, then we simply modify the construction sequence so that (s,t) has a child m with $d(s,m) = d(m,t) = \frac{1}{2}d(s,t)$, and make x the child of the new edge (m,s). For any fixed $\varepsilon > 0$, this goes on for only finitely many steps. Furthermore, it is easy to see that $\Phi(T_{\mathcal{G}})$ does not increase under this operation (we are only adding children of degree 1). $\square$

THEOREM 4.3 (**Product embedding**). *Let (X,d) be an arbitrary n-point SP metric. Then there exist a pair of oriented SP bundle trees $(X_1,d_1),(X_2,d_2)$ such that $|X_1| + |X_2| \le O(n)$ and X admits an $O(1)$-embedding into the product metric $X_1 \times X_2$.*

To prove Theorem 4.3, we need to introduce an additional type of composition procedure for SP metrics and their associated graphs. Consider any weighted, 2-connected series-parallel graph $G = (V,E)$. As discussed in [13], every such graph has a composition sequence $\mathcal{G} = \langle G_0, G_1, \ldots, G_m\rangle$ where $G_m = G$ and G_0 is a single weighted edge (s_0,t_0). One forms G_{i+1} from G_i as follows: For some edge $(u,v) \in E(G_i)$, we put $V(G_{i+1}) = V(G_i) \cup \{x\}$ and $E(G_{i+1}) = E(G_i) \cup \{(u,x),(x,v)\}$, where x is a new node not present in $V(G_i)$. The new edges (u,x) and (x,v) are allowed to have arbitrary non-negative weights. We call this a *standard composition.*

If the edge weights are such that $d(u,v) = d(u,x)+d(x,v)$, we call the refer to the new path u-x-v as *taut*; if $d(u,x) + d(x,v) \ge \alpha \cdot d(u,v)$, we refer to the new path as *α-slack*. A standard composition $\mathcal{G}$ is called *α-slack-taut* if every newly created path u-x-v in the composition sequence is either taut or α-slack. The following is a simple generalization of a lemma of [13].

LEMMA 4.4. *Given any 2-connected series-parallel graph $G = (V,E)$, one may construct a series-parallel graph $H = (V,E')$ which satisfies:*

- *H has an α-slack-taut construction sequence.*
- *For every $x,y \in V$, $\frac{1}{\alpha}d_G(x,y) \le d_H(x,y) \le d_G(x,y)$,*

where d_G and d_H are the shortest-path metrics on G and H, respectively. Furthermore, this construction is polynomial time in the size of G.

Thus, by incurring a distortion of at most 41 we can assume that G has a 41-slack-taut composition sequence $\mathcal{G} = \langle G_0,\ldots,G_m\rangle$. Next we reduce the case of α-slack-taut series-parallel graphs to the case of series-parallel bundle trees.

LEMMA 4.5. *Given any 2-connected series-parallel graph $G = (V,E)$ that is 41-slack-taut, one may construct two oriented SP bundle trees (L,d_L) and (R,d_R) such that*

- *$V(G) \subseteq V(L)$ and $V(G) \subseteq V(R)$*
- *$|V(L)|,|V(R)| = O(|V(G)|)$*
- *For every $x,y \in V$, $\frac{1}{11}d_G(x,y) \le \frac{1}{2}(d_L(x,y)+d_R(x,y)) \le d_G(x,y)$,*
- *For every $x,y \in V$, $d_L(x,y) \le d_G(x,y)$ and $d_R(x,y) \le d_G(x,y)$*

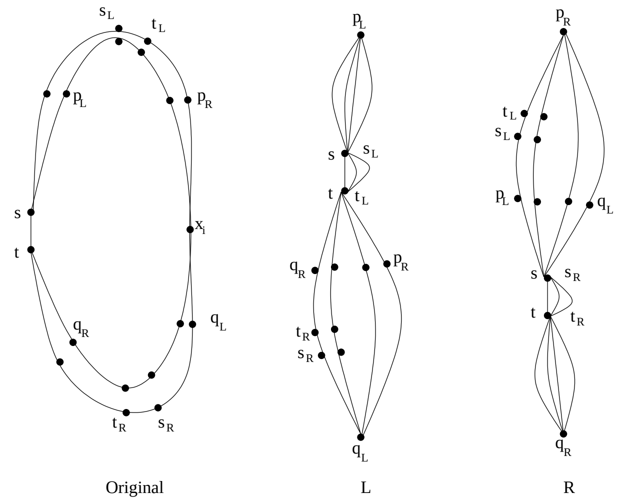

Figure 2: The bundle trees L and R with key points labeled. Note that x_i may be positioned anywhere on the slack loop, and may not necessarily fall where it is pictured.

Note that in [13] the step of converting a slack-taught graph to (a distribution over) trees of bundles is accomplished by cutting slack bundles. This technique cannot be used here because we need each of our component maps to be Lipschitz. Instead, we will proceed by "folding" slack bundles across a diameter, in a procedure similar to Charikar and Sahai [7]. We describe the folding procedure.

Folding algorithm. We will describe the construction of L; the construction of R is similar. A basic folding step concerns an edge (s,t) and a node x_i for which the edges (s,x_i) and (x_i,t) were added in some step of G's composition sequence. We assume inductively that we have oriented SP bundle trees (G_s,d_s) and (G_t,d_t) with corresponding oriented trees Λ_s and Λ_t. Furthermore, we suppose that s,x_i are the endpoints of the root bundle in Λ_s and x_i,t are the endpoints of the root bundle in Λ_t, and that the equalities $d_s(x_i,s) = d(x_i,s)$ and $d_t(x_i,t) = d(x_i,t)$ are satisfied.

Now we create L' by gluing G_s and G_t together at x_i, and adding the edge (s,t). If we are in the taut case $d(s,t) = d(s,x_i)+d(x_i,t)$, then L' is itself an oriented SP bundle tree (where s,t simply join the new root bundle). Otherwise, we are in the 41-slack case, and L' needs to be modified by folding. We refer to Figure 2 in the following description.

1. Let $u = d(s,t)$ and define $z_i \ge u$ s.t. $d(s,t)+d(s,x_i)+d(x_i,t) = 2u + 40z_i$. Such a z_i is guaranteed to exist by the 41-slack-taut structure.

2. Consider every simple (s,t) path through x_i. Insert a new point called p_L on each path so that it is at distance $5z_i$ from s and $5z_i + u$ from t. Coalesce all these new points into a single new point p_L.

3. Create another point q_L which is diametrically opposite p_L. In other words, it is made up of all the points that are $15z_i + u$ away from s and $15z_i$ away from t.

4. Find all points at distance $10z_i$ from s and $10z_i + u$ from t, and call them s_L. Coalesce all such points with the point s to form a single vertex.

5. Find all points at distance $10z_i+u$ from s and $10z_i+2u$ away from t and call them t_L. Coalesce all such points with the point t to form a single vertex.

This produces a new oriented SP bundle tree L. The analysis is deferred to the full version.

5. MARKOV TYPE

In this section, we will show that every series-parallel metric has Markov type 2. The following lemma is basic, and we omit its proof.

LEMMA 5.1. *Let $X_1, \ldots, X_k$ be metric spaces with Markov type 2. If $X = X_1 \times \cdots \times X_k$ is equipped with the product metric d_X, then X has Markov type 2 as well, and*

$$M_2(X) \leq \sqrt{k} \cdot \sqrt{M_2(X_1)^2 + \cdots + M_2(X_k)^2}.$$

In light of the preceding lemma and Theorem 4.3, it suffices to prove that every oriented SP bundle tree has Markov type 2. Using Proposition 1.4, we need to demonstrate the existence of nice potential functions on such metric spaces.

LEMMA 5.2. *If (X,d) is an oriented SP bundle tree with associated weighted graph $G = (X,E)$, then there exists a 9-Lipschitz map $\psi : X \to \mathbb{R}$ such that, for every path $P = \langle v_0, v_1, \ldots, v_m \rangle$ in G, one has*

$$d(v_0, v_m) \leq \max\{|\psi(v_0) - \psi(v_t)| : 0 \leq t \leq m\}$$

PROOF. Due to lack of space, we defer some details. Let $\Lambda_G = (V_\Lambda, D_\Lambda)$ be the associated oriented tree (whose vertices are SP bundles in G). Let $V_\Lambda = \{X_1, X_2, \ldots, X_k\}$, let $X_r \in V_\Lambda$ be the root of Λ_G, and let s be one of the endpoints of the SP bundle X_r. We consider the map $\psi : X \to \mathbb{R}$ given by $\psi(x) = d(x,s)$.

Let $u, v \in X$ be arbitrary points with $u \in X_i, v \in X_j$. Let us assume that some u-v path P in G intersects at least three distinct bundles from V_Λ (otherwise, the proof is only simpler). Let X_k be the least common ancestor of X_i and X_j in Λ_G, and let $a_u, a_v \in X_k$ be the points at which P enters and exits X_k. Then every u-v path $\gamma = \langle v_0, v_1, \ldots, v_m \rangle$ can be decomposed into three segments $\gamma_1, \gamma_2, \gamma_3$ which connect u and a_u, a_u and a_v, and a_v and v, respectively. Furthermore, we can pass to a sub-path $\hat{\gamma}_2$ of γ_2 such that $\hat{\gamma}_2$ connects a_u and a_v and is completely contained in X_k.

Now, it is not difficult to see that $d(u, a_u) = |\psi(u) - \psi(a_u)|$ and $d(v, a_v) = |\psi(v) - \psi(a_v)|$, and using the triangle inequality, we have $|\psi(v) - \psi(a_v)| + |\psi(u) - \psi(a_u)| \leq 3 \cdot \max\{|\psi(u) - \psi(v_t)| : 0 \leq t \leq m\}$. On the other hand, if s_k is the endpoint of X_k closest to s, then for every $x \in X_k$, we have $d(x, s_k) = d(x,s)$, hence inside X_k, ψ behaves like the function $\psi'(y) = d(y, s_k)$. Since $\hat{\gamma}_2$ is contained completely inside X_k, this reduces to an analysis just for bundles, which is very similar to the analysis of [25] for the Laakso graphs.

This yields the inequality

$$\begin{aligned} d(a_u, a_v) &\leq 3 \cdot \max\{|\psi(a_u) - \psi(v_t)| : 0 \leq t \leq m, v_t \in \hat{\gamma}_2\} \\ &\leq 6 \cdot \max\{|\psi(u) - \psi(v_t)| : 0 \leq t \leq m\} \end{aligned}$$

So overall, we have

$$\begin{aligned} d(u,v) &= d(u, a_u) + d(a_u, a_v) + d(a_v, v) \\ &\leq 9 \cdot \max\{|\psi(u) - \psi(v_t)| : 0 \leq t \leq m\} \end{aligned}$$

To finish, we simply scale ψ by a factor of 9, implying that $\|\psi\|_{\mathrm{Lip}} \leq 9$, and completing the proof. □

6. REFERENCES

[1] A. Agarwal, M. Charikar, K. Makarychev, and Y. Makarychev. $O(\sqrt{\log n})$ approximation algorithms for Min UnCut, Min 2CNF Deletion, and directed cut problems. In *37th Annual ACM Symposium on Theory of Computing*. ACM, 2005.

[2] S. Arora, J. R. Lee, and A. Naor. Euclidean distortion and the Sparsest Cut. In *37th Annual Symposium on the Theory of Computing*, pages 553–562, 2005. To appear, *Journal of the AMS*.

[3] S. Arora, S. Rao, and U. Vazirani. Expander flows, geometric embeddings, and graph partitionings. In *36th Annual Symposium on the Theory of Computing*, pages 222–231, 2004.

[4] K. Ball. Markov chains, Riesz transforms and Lipschitz maps. *Geom. Funct. Anal.*, 2(2):137–172, 1992.

[5] J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.*, 52(1-2):46–52, 1985.

[6] B. Brinkman and M. Charikar. On the impossibility of dimension reduction in ℓ_1. *J. ACM*, 52(5):766–788, 2005.

[7] M. Charikar and A. Sahai. Dimension reduction in the ℓ_1 norm. In *43rd Annual Symposium on Foundations of Computer Science*, 2002.

[8] E. Chlamtac, K. Makayrchev, and Y. Makarychev. How to play unique game using embeddings. In *47th Annual Syposium on Foundations of Computer Science*, 2006.

[9] R. Durrett. *Probability: theory and examples*. Duxbury Press, Belmont, CA, second edition, 1996.

[10] U. Feige, M. T. Hajiaghayi, and J. R. Lee. Improved approximation algorithms for minimum-weight vertex separators. In *37th Annual ACM Symposium on Theory of Computing*. ACM, 2005. To appear, *SIAM J. Comput.*

[11] A. Gupta. Improved bandwidth approximation for trees and chordal graphs. *J. Algorithms*, 40(1):24–36, 2001.

[12] 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.

[13] A. Gupta, I. Newman, Y. Rabinovich, and A. Sinclair. Cuts, trees and l_1-embeddings of graphs. *Combinatorica*, 24(2):233–269, 2004.

[14] J. Heinonen. *Lectures on analysis on metric spaces*. Universitext. Springer-Verlag, New York, 2001.

[15] P. Indyk. Algorithmic applications of low-distortion geometric embeddings. In *42nd Annual Symposium on Foundations of Computer Science*, pages 10–33. IEEE Computer Society, 2001.

[16] T. J. Laakso. Plane with A_∞-weighted metric not bi-Lipschitz embeddable to $\mathbb{R}^N$. *Bull. London Math. Soc.*, 34(6):667–676, 2002.

[17] U. Lang and C. Plaut. Bilipschitz embeddings of metric spaces into space forms. *Geom. Dedicata*, 87(1-3):285–307, 2001.

[18] J. R. Lee and A. Naor. Embedding the diamond graph in L_p and dimension reduction in L_1. *Geom. Funct. Anal.*, 14(4):745–747, 2004.

[19] J. R. Lee, A. Naor, and Y. Peres. Trees and Markov convexity. In *Proceedings of the 17th annual ACM-SIAM Symposium on Discrete Algorithms*, 2006.

[20] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

[21] J. Matoušek. On embedding trees into uniformly convex Banach spaces. *Israel J. Math.*, 114:221–237, 1999.

[22] J. Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002.

[23] J. Matoušek and Y. Rabinovich. On dominated l_1 metrics. *Israel J. Math.*, 123:285–301, 2001.

[24] V. D. Milman and G. Schechtman. *Asymptotic theory of finite-dimensional normed spaces*, volume 1200 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1986. With an appendix by M. Gromov.

[25] 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.

[26] I. Newman and Y. Rabinovich. A lower bound on the distortion of embedding planar metrics into Euclidean space. *Discrete Comput. Geom.*, 29(1):77–81, 2003.

[27] S. Rao. Small distortion and volume preserving embeddings for planar and Euclidean metrics. In *Proceedings of the 15th Annual Symposium on Computational Geometry*, pages 300–306, New York, 1999. ACM.