

## METRIC EMBEDDING VIA SHORTEST PATH DECOMPOSITIONS\*

ITTAI ABRAHAM<sup>†</sup>, ARNOLD FILTSE<sup>‡</sup>, ANUPAM GUPTA<sup>§</sup>, AND OFER NEIMAN<sup>¶</sup>

**Abstract.** We study the problem of embedding shortest-path metrics of weighted graphs into  $\ell_p$  spaces. We introduce a new embedding technique based on low-depth decompositions of a graph via shortest paths. The notion of shortest path decomposition (SPD) depth is inductively defined: A (weighed) path graph has SPD depth 1. General graph has an SPD of depth  $k$  if it contains a shortest path whose deletion leads to a graph, each of whose components has SPD depth at most  $k - 1$ . In this paper we give an  $O(k^{\min\{1/p, 1/2\}})$ -distortion embedding for graphs of SPD depth at most  $k$ . This result is asymptotically tight for any fixed  $p > 1$ , while for  $p = 1$  it is tight up to second order terms. As a corollary of this result, we show that graphs having pathwidth  $k$  embed into  $\ell_p$  with distortion  $O(k^{\min\{1/p, 1/2\}})$ . For  $p = 1$ , this improves over the best previous bound of Lee and Sidiropoulos that was exponential in  $k$ ; moreover, for other values of  $p$  it gives the first embeddings whose distortion is independent of the graph size  $n$ . Furthermore, we use the fact that planar graphs have SPD depth  $O(\log n)$  to give a new proof that any planar graph embeds into  $\ell_1$  with distortion  $O(\sqrt{\log n})$ . Our approach also gives new results for graphs with bounded treewidth, and for graphs excluding a fixed minor.

**Key words.** metric embeddings,  $\ell_p$  spaces, sparsest cut, pathwidth

**AMS subject classifications.** 41, 60, 30L05

**DOI.** 10.1137/19M1296021

**1. Introduction.** Low-distortion metric embeddings are a crucial component in the modern algorithmist toolkit. Indeed, they have applications in approximation algorithms [LLR95], online algorithms [BBMN15], distributed algorithms [KKM<sup>+</sup>12], and for solving linear systems and computing graph sparsifiers [ST04]. Given a (finite) metric space  $(V, d)$ , a map  $\phi : V \rightarrow \mathbb{R}^D$ , and a norm  $\|\cdot\|$ , the *contraction* and *expansion* of the map  $\phi$  are the smallest  $\tau, \rho \geq 1$ , respectively, such that for every pair  $x, y \in V$ ,

$$\frac{1}{\tau} \leq \frac{\|\phi(x) - \phi(y)\|}{d(x, y)} \leq \rho .$$

The *distortion* of the map is then  $\tau \cdot \rho$ .

In this paper we will investigate embeddings into  $\ell_p$  norms; the most prominent of which are the Euclidean norm  $\ell_2$  and the cut norm  $\ell_1$ ; the former for obvious reasons, and the latter because of its close connection to graph partitioning problems, and in particular, the sparsest cut problem. Specifically, the ratio between the sparsest cut and the multicommodity flow equals the distortion of the optimal embedding

---

\*Received by the editors October 28, 2019; accepted for publication (in revised form) December 20, 2021; published electronically April 11, 2022. A preliminary version of this paper was published in the proceedings of STOC'18. A full version of this paper is also available at <https://arxiv.org/abs/1708.04073>.

<https://doi.org/10.1137/19M1296021>

**Funding:** The third author was supported in part by NSF awards CCF-1536002, CCF-1540541, and CCF-1617790. The fourth author was supported in part by ISF grant 1817/17, and by BSF grant 2015813.

<sup>†</sup>VMWare Research, Palo Alto, CA 94043 USA (iabraham@vmware.com).

<sup>‡</sup>The Department of Computer Science, Bar-Ilan University, Ramat Gan, 5290002, Israel (arnold273@gmail.com).

<sup>§</sup>Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213 USA (anupamg@cs.cmu.edu).

<sup>¶</sup>Computer Science, Ben-Gurion University of the Negev, Beer Sheva, 84105, Israel (neimano@cs.bgu.ac.il).

into  $\ell_1$ . (Note that every  $\ell_1$  metric is a linear combination of cut metrics. See [LLR95, GNRS04] for more details on the connection to multicommodity flows.)

We focus on embedding of metrics arising from certain graph families. Indeed, since general  $n$ -point metrics require  $\Omega(\log n/p)$ -distortion to embed into  $\ell_p$ -norms, much attention was given to embeddings of restricted graph families that arise in practice. (Embedding an (edge-weighted) graph is short-hand for embedding the shortest path metric of the graph generated by these edge-weights.) Since the class of graphs embeddable with some distortion into some target normed space is closed under taking minors, it is natural to focus on minor-closed graph families. A long-standing open problem in this area to decide whether all nontrivial minor-closed families of graphs embed into  $\ell_1$  with distortion depending only on the graph family, and not the size  $n$  of the graph.

While this question remains unresolved in general, there has been some progress on special classes of graphs. The class of outerplanar graphs (which exclude  $K_{2,3}$  and  $K_4$  as a minor) embeds isometrically into  $\ell_1$ ; this follows from results of Okamura and Seymour [OS81]. Following [GNRS04], Chakrabarti et al. [CJLV08] show that every graph with treewidth-2 (which excludes  $K_4$  as a minor) embeds into  $\ell_1$  with distortion 2 (which is tight, as shown by [LR10]). Lee and Sidiropoulos [LS13] showed that every graph with pathwidth  $k$  can be embedded into  $\ell_1$  with distortion  $(4k)^{k^3+1}$ . See subsection 1.3 for additional results.

We note that  $\ell_2$  is a potentially more natural and useful target space than  $\ell_1$  (in particular, finite subsets of  $\ell_2$  embed isometrically into  $\ell_1$ ). Alas, there are only few (natural) families of metrics that admit constant distortion embedding into Euclidean space, such as “snowflakes” of doubling metrics [Ass83], doubling trees [GKL03], and graphs of bounded bandwidth [BCM13]. All of these families have bounded doubling dimension. (For definitions, see section 2.)

**1.1. Our results.** In this paper we develop a new technique for embedding graphs into  $\ell_p$  spaces with small distortion. We introduce the notion of shortest path decomposition (SPD) of bounded depth. Every (weighted) path graph has an SPD of depth 1. A graph  $G$  has an SPD of depth  $k$  if there exists a *shortest path*  $P$ , such that deleting  $P$  from the graph  $G$  (that is, deleting all the vertices on  $P$  and the adjacent edges) results in a graph whose connected components all have SPD of depth at most  $k - 1$ . (An alternative definition appears in Definition 1.) Our main result is the following.

**THEOREM 1** (embeddings for SPD families). *Let  $G = (V, E)$  be a weighted graph with an SPD of depth  $k$ . Then there exists an embedding  $f : V \rightarrow \ell_p$  with distortion  $O(k^{1/p})$ .*

*Remark.* Since finite subsets of  $\ell_2$  embed isometrically into  $\ell_p$  for any  $1 \leq p \leq \infty$ , we get that the distortion of Theorem 1 is never larger than  $O(\sqrt{k})$ .

*Graph families with SPD of small depth.* We will show that graphs of pathwidth  $k$  have SPD of depth  $k + 1$ , and thus obtain the following result as a simple corollary of Theorem 1.

**THEOREM 2** (pathwidth theorem). *Any graph with pathwidth  $k$  embeds into  $\ell_p$  with distortion  $O(k^{\min\{1/p, 1/2\}})$ .*

Note that this is a superexponential improvement over the best previous distortion bound of  $O(k)^{k^3}$  by Lee and Sidiropoulos [LS13]. Their approach was based on probabilistic embedding into trees, which implies embedding only into  $\ell_1$ . Such an approach cannot yield distortion better than  $O(k)$ , due to known lower bounds for the

TABLE 1

Our and previous results for embedding certain graph families into  $\ell_p$ . (For  $H$ -minor-free graphs,  $g(H)$  is some function of  $|H|$ .)

Graph family	Our results.	Previous results
Pathwidth $k$	$O(k^{1/p})$	$(4k)^{k^3+1}$ into $\ell_1$ [LS13]
Treewidth $k$	$O((k \log n)^{1/p})$	$O(k^{1-1/p} \cdot \log^{1/p} n)$ [KLMN05] $O((\log(k \log n))^{1-1/p} (\log^{1/p} n))$ [KK16]
Planar	$O(\log^{1/p} n)$	$O(\log^{1/p} n)$ [Rao99]
$H$ -minor-free	$O((g(H) \log n)^{1/p})$	$O( H ^{1-1/p} \log^{1/p} n)$ [AGG <sup>+</sup> 19]+[KLMN05]
$H$ -minor-free	$O(1)$ into $\ell_\infty^{O(g(H) \cdot \log^2 n)}$	$O( H ^2)$ into $\ell_\infty^{O(3^{ H } \log  H  \log n)}$ [KLMN05]

diamond graph [GNRS04] that has pathwidth  $k + 1$ . Our embedding holds for any  $\ell_p$  space, and we can overcome the barrier of  $\Theta(k)$ . In particular, we obtain embeddings of pathwidth- $k$  graphs into both  $\ell_2$  and  $\ell_1$  with distortion  $O(\sqrt{k})$ . Moreover, an embedding with this distortion can be found efficiently via semidefinite-programming; see, e.g., [LLR95], even without access to the actual path decomposition (which is NP-hard even to approximate [BGHK92]). We remark that graphs of bounded pathwidth can have arbitrarily large doubling dimension (exhibited by star graphs that have pathwidth 1), and thus our result is a noteworthy example of a nontrivial Euclidean embedding with constant distortion for a family of metrics with unbounded doubling dimension.

Since graphs of treewidth  $k$  have pathwidth  $O(k \log n)$  (see, e.g., [KS93]), Theorem 2 provides an embedding of such graphs into  $\ell_p$  with distortion  $O((k \log n)^{1/p})$ . This strictly improves the best previously known bound, which follows from a theorem in [KLMN05] (who obtained distortion  $O(k^{1-1/p} \log^{1/p} n)$ ), for any  $p > 2$ , and matches it for  $1 \leq p \leq 2$ . While [KK16] obtained recently a distortion bound with improved dependence on  $k$ , their result  $O((\log(k \log n))^{1-1/p} (\log^{1/p} n))$  has suboptimal dependence on  $n$ .

Moreover, we derive several other results for planar graphs, and more generally graphs excluding a fixed minor. Even though these families have unbounded pathwidth, we show that they have SPD of depth  $O(\log n)$ . These results are summarized in Table 1, they either improve on the state-of-the-art, or provide matching bounds using a new approach.

In section 8 we show that we can slightly modify our construction of Theorem 1 so that the dimension of the host space will be  $O(k \log n)$ , while maintaining the same distortion guarantee. This implies that graphs excluding  $H$  as a minor admit an embedding into  $\ell_\infty^{O(g(H) \cdot \log^2 n)}$  with constant distortion (this constant is independent of  $H$ ). See Theorem 5 and the discussion therein.

Our result of Theorem 2 (and thus also Theorem 1) is asymptotically tight for any fixed  $p > 1$ . The family exhibiting this fact is the diamond graphs.

**THEOREM 3** (see [NR03, LN04, JS09]). *For any fixed  $p > 1$  and every  $k \geq 1$ , there exists a graph  $G = (V, E)$  with pathwidth- $k$ , such that every embedding  $f : V \rightarrow \ell_p$  has distortion  $\Omega(k^{\min\{1/p, 1/2\}})$ .*

The bound in Theorem 3 was proven first for  $p = 2$  in [NR03], generalized to  $1 < p \leq 2$  in [LN04] and for  $p \geq 2$  by [JS09] (see also [MN13, JLM11]).

For the case of  $p = 1$ , we show that Theorem 1 is tight up to second order terms.

**THEOREM 4.** *For every  $k \geq 1$ , there exists a graph  $G = (V, E)$  with SPDdepth  $O(k)$ , such that every embedding  $f : V \rightarrow \ell_1$  has distortion  $\Omega(\sqrt{\frac{k}{\log k}})$ .*

The proof appears in section 9. The family exhibiting the lower bound is a slight modification of the diamondfold graphs [LS11]. We note that the diamondfold graphs (and also our modification of them) contain large grids as a minor and therefore do not have a small pathwidth, or even treewidth. In particular, for embedding graphs with pathwidth  $k$  into  $\ell_1$ , only the trivial  $\Omega(\log k)$  lower bound is known.

**1.2. Technical ideas.** Many known embeddings [Bou85, Rao99, KLMN05, ABN11] are based on a collection of 1-dimensional embeddings, where we embed each point to its distance from a given subset of points. We follow this approach, but differ in two aspects. First, the subset of points we use is not based on random sampling [Bou85] or probabilistic clustering [Rao99]. Rather, inspired by the works of [And86] and [AGG<sup>+</sup>19], the subset used is a geodesic shortest path. The second is that our embedding is not 1-dimensional but 2-dimensional: this seemingly small change crucially allows us to use the structure of the shortest paths to our advantage.

The SPD induces a collection of shortest paths (each shortest path lies in some connected component). A natural initial attempt is to embed a vertex  $v$  relative to a geodesic path  $P$  using two dimensions:<sup>1</sup>

- The *first* coordinate  $\Delta_1$  is the distance to the path  $d(v, P)$  (see Figure 1).
- The *second* coordinate  $\Delta_2$  is the distance  $d(v, r)$  to the endpoint of the path, called its “root” (see Figure 1).

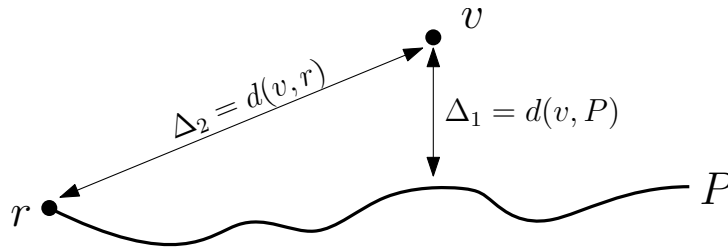


FIG. 1. An illustration of our initial attempt. The first coordinate  $\Delta_1$  is the distance to the path  $d(v, P)$ . The second coordinate  $\Delta_2$  is the distance  $d(v, r)$  to the endpoint of the path, called its “root”.

Unfortunately, this embedding may have unbounded expansion: If two vertices  $u, v$  are separated by some shortest path, in future iterations  $v$  may have a large distance to the root of a path  $P$  in its component, while  $u$  has zero in that coordinate (because it’s not in that component), incurring a large stretch. The natural fix is to enforce a Lipschitz condition on every coordinate: for  $v$  in cluster  $X$ , we *truncate* the value  $v$  can receive at  $O(d_G(v, V \setminus X))$ . That is, a vertex close to the boundary of  $X$  cannot get a large value. Using the fact that the SPD has depth  $k$ , each vertex will have only  $O(k)$  nonzero coordinates, which implies expansion  $O(k^{1/p})$ .

To bound the contraction, for each pair  $u, v$  we consider the *first* path  $P$  in the SPD that lies “close” to  $\{u, v\}$  or separates them to different connected components. Then we show that at least one of the two coordinates should give sufficient contribution.

But what about the effect of truncation on contraction? A careful recursive argument shows that the contribution to  $u, v$  from the first coordinate (the distance from the path  $P$ ) is essentially not affected by this truncation. Hence the argument in cases (a) and (b) of Figure 2 still work. However, the argument using the distance to

<sup>1</sup>In fact, we use different dimensions for each connected component.

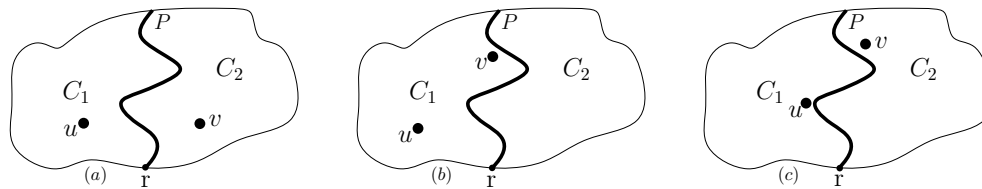


FIG. 2. A shortest path  $P$ , rooted at  $r$ , partitions cluster  $X$  into clusters  $C_1$  and  $C_2$ . In cases (a) and (b), the first coordinate (the distance to  $P$ ) provides sufficient contribution. In case (c) the second coordinate (the distance to root  $r$ ) provides the contribution.

the root of  $P$ , case (c), can be ruined. Solving this issue requires some new nontrivial ideas. Our solution is to introduce a probabilistic sawtooth function that replaces the simple truncation. The main technical part of this paper is devoted to showing that a collection of these functions for all possible distance scales, with appropriate random shifts, suffices to control the expected contraction in case (c) for *all relevant pairs simultaneously*.

**1.3. Other related work.** There has been work on embedding several other graph families into normed spaces: Chekuri et al. [CGN<sup>+</sup>06] extend the Okamura and Seymour bound for outerplanar graphs to  $k$ -outerplanar graphs, and showed that these embed into  $\ell_1$  with distortion  $2^{O(k)}$ . Rao [Rao99] (see also [KLMN05]) embed planar graphs into  $\ell_p$  with distortion  $O(\log^{1/p} n)$ . For graphs with genus  $g$ , [LS10] showed an embedding into Euclidean space with distortion  $O(\log g + \sqrt{\log n})$ . Finally, for  $H$ -minor-free graphs, combining the results of [AGG<sup>+</sup>19, KLMN05] give  $\ell_p$ -embeddings with  $O(|H|^{1-1/p} \log^{1/p} n)$  distortion.

Following [And86, Mil86], the idea of using geodesic shortest paths to decompose the graph has been used for many algorithmic tasks: MPLS routing [GKR04], directed connectivity, distance labels and compact routing [Tho04], object location [AG06], and nearest neighbor search [ACKW15].

Given a tree  $T$ , Matoušek [Mat99] recursively defined the *caterpillar dimension*  $\text{cdim}(T)$  as follows: the  $\text{cdim}$  of a singleton vertex is 0. The  $\text{cdim}$  of a tree is  $k$  if there are a set of paths  $P_1, \dots, P_s$  intersecting in a single vertex, such that removing the edges in all these paths (as opposed to vertices in our definition) results in connected components each with  $\text{cdim}$  at most  $k - 1$ . Matoušek showed that every tree  $T$  with  $\text{cdim}(T)$  embeds into every  $\ell_p$  space with distortion  $O_p(\log(\text{cdim}(T)))^{\min\{\frac{1}{2}, \frac{1}{p}\}}$ .

In a follow up paper, [Fil20a] (the second author) generalized our definition of SPD to partial-SPD (allowing the lower level in the partition hierarchy to be general subgraph rather than only a shortest path). Given a weighted planar graph  $G = (V, E, w)$  with a subset of terminals  $K$ , a face cover is a subset of faces such that every terminal lies on some face from the cover. Given a face cover of size  $\gamma$ , using our embedding result for SPD [Fil20a] shows that the terminal set  $K$  can be embedded into  $\ell_1$  with distortion  $O(\sqrt{\log \gamma})$ .

In another follow-up paper, the second author [Fil20b], created strong sparse partitions for graphs with bounded SPDdepth. These were later used to obtain constant distortion solution for the Steiner point removal problem, and also creating approximation algorithms for the universal Steiner tree problem, and universal traveling salesman problem [JLN<sup>+</sup>05].

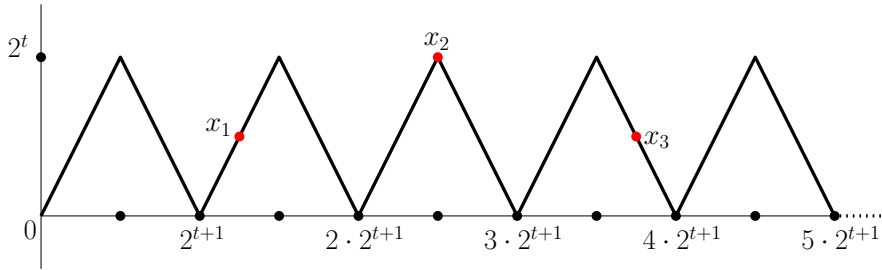


FIG. 3. The graph of the “sawtooth” function  $g_t$ . The points  $x_1 = 5 \cdot 2^{t-1}$  and  $x_3 = 15 \cdot 2^{t-1}$  are mapped to  $2^{t-1}$ , while  $x_2 = 10 \cdot 2^{t-1}$  is mapped to  $2^t$ .

**2. Preliminaries and notation.** For  $k \in \mathbb{Z}$ , let  $[k] := \{1, \dots, k\}$ . For  $p \geq 1$ , the  $\ell_p$ -norm of a vector  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  is  $\|x\|_p := (\sum_{i=1}^d |x_i|^p)^{1/p}$ , where  $\|x\|_\infty := \max_i |x_i|$ .

*Doubling dimension.* The doubling dimension of a metric is a measure of its local “growth rate.” Formally, a metric space  $(X, d)$  has *doubling dimension*  $\lambda_X$  if for every  $x \in X$  and radius  $r$  the ball  $B(x, r)$  can be covered by  $2^{\lambda_X}$  balls of radius  $\frac{r}{2}$ . A family is *doubling* if the doubling dimension of all metrics in it is bounded by some universal constant.

*Graphs.* We consider connected undirected graphs  $G = (V, E)$  with edge weights  $w : E \rightarrow \mathbb{R}_{>0}$ . Let  $d_G$  denote the shortest path metric in  $G$ ; we drop subscripts when there is no ambiguity. For a vertex  $x \in V$  and a set  $A \subseteq V$ , let  $d_G(x, A) := \min_{a \in A} d(x, a)$ , where  $d_G(x, \emptyset) := \infty$ . For a subset of vertices  $A \subseteq V$ , let  $G[A]$  denote the induced graph on  $A$ , and let  $d_A := d_{G[A]}$  be the shortest path metric in the induced graph. Let  $G \setminus A := G[V \setminus A]$  be the graph after deleting the vertex set  $A$  from  $G$ .

*Special graph families.* Given a graph  $G = (V, E)$ , a *tree decomposition* of  $G$  is a tree  $T$  with nodes  $B_1, \dots, B_s$  (called *bags*) where each  $B_i$  is a subset of  $V$  such that the following properties hold:

- For every edge  $\{u, v\} \in E$ , there is a bag  $B_i$  containing both  $u$  and  $v$ .
- For every vertex  $v \in V$ , the set of bags containing  $v$  form a connected subtree of  $T$ .

The *width* of a tree decomposition is  $\max_i \{|B_i| - 1\}$ . The *treewidth* of  $G$  is the minimal width of a tree decomposition of  $G$ .

A *path decomposition* of  $G$  is a special kind of tree decomposition where the underlying tree is a path. The *pathwidth* of  $G$  is the minimal width of a path decomposition of  $G$ .

A graph  $H$  is a *minor* of a graph  $G$  if we can obtain  $H$  from  $G$  by edge deletions/contractions and vertex deletions. A graph family  $\mathcal{G}$  is *H-minor-free* if no graph  $G \in \mathcal{G}$  has  $H$  as a minor.

**2.1. The sawtooth function.** An important component in our embeddings will be the following sawtooth function. For  $t \in \mathbb{N}$ , we define  $g_t : \mathbb{R}_+ \rightarrow \mathbb{R}$  to be the *sawtooth* function w.r.t.  $2^t$  as follows. For  $x \geq 0$ , if  $q_x := \lfloor x/2^{t+1} \rfloor$ , then

$$g_t(x) := 2^t - |x - (q_x \cdot 2^{t+1} + 2^t)|.$$

Figure 3 can help visualize this function. The following observation is straightforward.

**OBSERVATION 1.** *The sawtooth function  $g_t$  is 1-Lipschitz, bounded by  $2^t$ , and periodic with period  $2^{t+1}$ .*

To make the proofs cleaner, we define an auxiliary function given parameters  $\alpha \in [0, 1], \beta \in [0, 4]$ :

$$(1) \quad g_{t,\alpha,\beta}(x) := g_t(\beta \cdot x + \alpha \cdot 2^{t+1}).$$

Note that by Observation 1,  $g_{t,\alpha,\beta}$  is  $\beta$ -Lipschitz and bounded by  $2^t$ . The proof of the following lemma appears in section 7.

LEMMA 1 (sawtooth lemma). *Let  $x, y \in \mathbb{R}_+$ . Let  $\alpha \in [0, 1], \beta \in [0, 4]$  be drawn uniformly and independently. The following properties hold:*

1.  $\mathbb{E}_{\alpha,\beta} [ g_{t,\alpha,\beta}(x) ] = 2^{t-1}$ .
2.  $\mathbb{E}_{\alpha,\beta} [ |g_{t,\alpha,\beta}(x) - g_{t,\alpha,\beta}(y)| ] = \Omega(\min\{|x - y|, 2^t\})$ .

**3. Shortest path decompositions.** Our embeddings will crucially depend on the notion of shortest path decompositions (SPDs). In the introduction we provided a recursive definition for SPD. Here we show an equivalent definition which will be more suitable for our purposes.

DEFINITION 1 (SPD). *Given a weighted graph  $G = (V, E, w)$ , an SPD of depth  $k$  is a pair  $\{\mathcal{X}, \mathcal{P}\}$ , where  $\mathcal{X}$  is a collection  $\mathcal{X}_1, \dots, \mathcal{X}_k$  of partial partitions of  $V$ ,<sup>2</sup> and  $\mathcal{P}$  is a collection of sets of paths  $\mathcal{P}_1, \dots, \mathcal{P}_k$ , where  $\mathcal{X}_1 = \{V\}, \mathcal{X}_k = \mathcal{P}_k$ , and the following properties hold:*

1. For every  $1 \leq i \leq k$  and every subset  $X \in \mathcal{X}_i$ , there exist a unique path  $P_X \in \mathcal{P}_i$  such that  $P_X$  is a shortest path in  $G[X]$ .
2. For every  $2 \leq i \leq k$ ,  $\mathcal{X}_i$  consists of all connected components of  $G[X \setminus P_X]$  over all  $X \in \mathcal{X}_{i-1}$ .

In other words,  $\bigcup_{i=1}^k \mathcal{P}_k$  is a partition of  $V$  into paths, where each path  $P_X$  is a shortest path in the component  $X$  it belongs to at the point it is deleted.

For a given graph  $G$ , let  $\text{SPDdepth}(G)$  be the minimum  $k$  such that  $G$  admits an SPD of depth  $k$ . For a given family of graphs  $\mathcal{G}$ , let  $\text{SPDdepth}(\mathcal{G}) := \max_{G \in \mathcal{G}} \{\text{SPDdepth}(G)\}$ . In the following, we consider the  $\text{SPDdepth}$  of some graph families.

**3.1. The SPD depth for various graph families.** One advantage of defining the SPD is that several well-known graph families have bounded depth SPD.

- *Pathwidth.* Every graph  $G = (V, E, w)$  with *pathwidth*  $k$  has an  $\text{SPDdepth}$  of  $k+1$ . Indeed, let  $\mathcal{T} = (\mathcal{B}_1, \dots, \mathcal{B}_s)$  be a path decomposition of  $G$ , where  $\mathcal{B}_1, \mathcal{B}_s$  are the two bags at the end of this path. Choose arbitrary vertices  $x \in \mathcal{B}_1$  and  $y \in \mathcal{B}_s$ , and let  $P$  be a shortest path in  $G$  from  $x$  to  $y$ . By the definition of a path decomposition, the path  $P$  contains at least one vertex from every bag  $\mathcal{B}_i$ . Hence, deleting the vertices of  $P$  would reduce the size of each bag by one; consequently, each connected component of  $G \setminus P$  has pathwidth  $k - 1$ , and by induction  $\text{SPDdepth } k$ . Finally, a connected component of pathwidth 0 is necessarily a singleton, which has  $\text{SPDdepth } 1$ .
- *Treewidth.* Since every treewidth- $k$  graph has pathwidth  $O(k \log n)$ , treewidth- $k$  graphs have  $\text{SPDdepth } O(k \log n)$ .
- *Planar.* Using cycle separators [Mil86] as in [Tho04, GKR04], every planar graph has  $\text{SPDdepth } O(\log n)$ ; this follows as each cycle separator can be constructed as a union of two shortest paths.

<sup>2</sup>That is, for every  $X \in \mathcal{X}_i, X \subseteq V$ , and for every different subset  $X, X' \in \mathcal{X}_i, X \cap X' = \emptyset$ .

- *Minor-free.* Finally, every  $H$ -minor-free graph admits a balanced separator consisting of  $g(H)$  shortest paths (for some function  $g$ ) [AG06], and hence has an SPDdepth  $O(g(H) \cdot \log n)$ .

Combining these observations with Theorem 1, we get the following set of results.

**COROLLARY 1.** *Consider an  $n$ -vertex weighted graph  $G$ . Theorem 1 implies the following:*

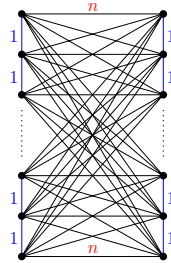
- *If  $G$  has pathwidth  $k$ , it embeds into  $\ell_p$  with distortion  $O(k^{1/p})$ .*
- *If  $G$  has treewidth  $k$ , it embeds into  $\ell_p$  with distortion  $O((k \log n)^{1/p})$ .*
- *If  $G$  is planar, it embeds into  $\ell_p$  with distortion  $O(\log^{1/p} n)$ .*
- *For every fixed  $H$ , if  $G$  excludes  $H$  as a minor, it embeds into  $\ell_p$  with distortion  $O(\log^{1/p} n)$ , where the constant in the big- $O$  depends on  $H$ .*

As mentioned in section 1, we get a substantial improvement for the pathwidth case. Our result for treewidth improves upon that from [KLMN05] for  $p > 2$ ; they got  $O(k^{1-1/p}(\log n)^{1/p})$  distortion compared to our  $O((k \log n)^{1/p})$ . Our result appears to be closer to the truth, since the distortion tends to  $O(1)$  as  $p \rightarrow \infty$ . Our results for planar graphs match the current state-of-the-art results.

Finally, our results for minor-free graphs depend on the Robertson–Seymour decomposition, and hence are currently better only for large values of  $p$ . (It remains an open question to improve the SPDdepth of  $H$ -minor-free graphs to have a  $\text{poly}(|H|) \log n$  dependence, perhaps using the ideas from [AGG<sup>+</sup>19].) In general, we hope that our results will be useful in getting other embedding results, and will spur further work on understanding shortest path separators.

We note that there exist graphs with large SPDdepth. For instance, the clique graph  $K_n$  has SPDdepth of  $\frac{n}{2}$ , as each shortest path contains at most two vertices. Moreover, there are sparse graphs with very large SPDdepth. Specifically, an  $n$ -vertex constant degree expander has SPDdepth of  $n^{\Omega(1)}$ . Indeed, denote by  $k$  the SPDdepth of some constant degree expander  $G$ . According to Theorem 5,  $G$  can be embedded into  $\ell_\infty^{O(k \log n)}$  with distortion  $O(1)$ . However, according to Rabinovich [Rab08], in order to embed a constant degree expander into  $\ell_\infty$  with distortion  $D$ ,  $n^{\Omega(1/D)}$  coordinates are required. It follows that  $k = n^{\Omega(1)}$ .

On the other hand, there are graphs with SPDdepth 2 that contain arbitrarily large cliques. For example, see the graph drawn on the right. The graph consists of two sets  $\{L, R\}$  each containing  $n$  vertices. The graph restricted to the vertices of  $L$  (resp.,  $R$ ) consist of a shortest path with unit weight edges. In addition, for every pair of vertices  $v \in L$  and  $u \in R$  we add an edge of weight  $n$ . Note that  $G$  contains the full bipartite graph  $K_{n,n}$  as a subgraph (and, in particular,  $K_n$  as a minor). It is straightforward that  $G$  has SPDdepth 2. Note also that by subdividing each edge of weight  $n$  to  $n$  unit weight edges, we will get an unweighted graph of SPDdepth 3 that contains  $K_n$  as a minor.



**4. The embedding algorithm.** Let  $G = (V, E)$  be a weighted graph, and let  $\{\mathcal{X}, \mathcal{P}\} = \{\{\mathcal{X}_1, \dots, \mathcal{X}_k\}, \{\mathcal{P}_1, \dots, \mathcal{P}_k\}\}$  be an SPD of depth  $k$  for  $G$ . By scaling, we can assume that the minimum weight of an edge is 1; let  $M \in \mathbb{N}$  be the minimal such that the diameter of  $G$  is strictly bounded by  $2^M$ . Pick  $\alpha \in [0, 1]$  and  $\beta \in [0, 4]$  uniformly and independently.

For every  $i \in [k]$  and  $X \in \mathcal{X}_i$ , we now construct an embedding  $f_X : V \rightarrow \mathbb{R}^D$  (for some number of dimensions  $D \in \mathbb{N}$ ). This map  $f_X$  consists of two parts.

*First coordinate: Distance to the path.* The first coordinate of the embedding



implements the distance to the path  $P_X$ , and is denoted by  $f_X^{\text{path}}$ . Let  $X_1, \dots, X_s \in \mathcal{X}_{i+1}$  be the connected components of  $G[X \setminus P_X]$  (note that it is also possible that  $s = 0$ ). We use a separate coordinate for each  $X_j$ , and hence  $f_X^{\text{path}} : V \rightarrow \mathbb{R}^s$ . Moreover, for  $v \in X$  we truncate at  $2d_G(v, V \setminus X)$  in order to guarantee Lipschitz-ness. In particular, the coordinate corresponding to  $X_j$  is set to

$$\left(f_X^{\text{path}}\right)_{X_j}(v) = \begin{cases} \min \{d_X(v, P_X), 2d_G(v, V \setminus X)\} & \text{if } v \in X_j, \\ 0 & \text{otherwise.} \end{cases}$$

See Figure 4 for an illustration.

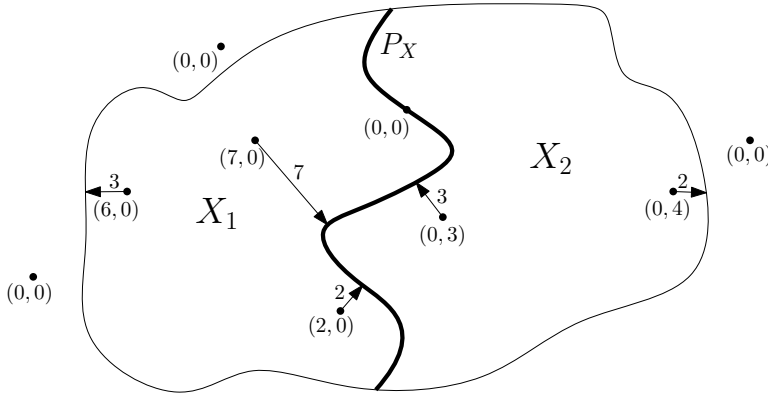


FIG. 4. The set  $X \in \mathcal{X}_i$  is surrounded by a closed curve. The path  $P_X$  partitions  $X$  into  $X_1, X_2 \in \mathcal{X}_{i+1}$ . The embedding  $f_X^{\text{path}}$  consists of two coordinates and is represented in the figure by a horizontal vector next to each vertex, where the first entry is w.r.t.  $X_1$  and the second w.r.t.  $X_2$ . Each point on  $P_X$  or not in  $X$  maps to 0 in both the coordinates. Each point in  $X_1$  maps to  $\min \{d_X(v, P_X), 2d_G(v, V \setminus X)\}$  in the first coordinate and to 0 in the second.

*Second coordinate: Distance to the root.* The second part is denoted  $f_X^{\text{root}}$ , which is intended to capture the distance from the root  $r$  of the path. Again, to get the Lipschitz-ness, we would like to truncate the value at  $2d_G(v, V \setminus X)$  as we did for  $f_X^{\text{path}}$ . However, a problem with this idea is that the root  $r$  can be arbitrarily far from some pair  $u, v$  that needs contribution from this coordinate. And hence, even if  $|d_G(u, r) - d_G(v, r)| \approx d_G(u, v)$ , there may be no contribution after the truncation. So we use the sawtooth function.

Specifically, we replace the ideal contribution  $d_G(v, r)$  by the sawtooth function  $g_t(d_G(v, r))$ , where the scale  $t$  for the function is chosen such that  $2^t \approx d_G(v, V \setminus X)$ . To avoid the case that two nearby points use two different scales (and hence to guarantee Lipschitz-ness), we take an appropriate linear combination of the two distance scales closest to  $2d_G(v, V \setminus X)$ . Recall that the sawtooth function does not guarantee contribution for  $u, v$  due to its periodicity: we may be unlucky and have  $g_t(d_G(v, r)) = g_t(d_G(u, r))$  even when  $d_G(v, r)$  and  $d_G(u, r)$  are very different. To guarantee a large enough contribution for all relevant pairs simultaneously, we add a random shift  $\alpha$ , and apply a random “stretch”  $\beta$  to  $d_G(v, r)$  before feeding it to  $g_t$ . Lemma 1 then shows that many of the choices of  $\alpha$  and  $\beta$  give substantially different values for  $u, v$ .

Formally, the mapping is as follows. The function  $f_X^{\text{root}}$  consists of  $M + 1$  coordinates, one for each distance scale  $t \in \{0, 1, \dots, M\}$ . The coordinate corresponding to

$t$  is denoted by  $f_{X,t}^{\text{root}}$ . Let  $r$  be an arbitrary endpoint of  $P_X$ ; we will call  $r$  the “root” of  $P_X$ . Let  $t_v \in \mathbb{N}$  be such that  $2 d_G(v, V \setminus X) \in [2^{t_v}, 2^{t_v+1})$ . Set  $\lambda_v = \frac{2 d_G(v, V \setminus X) - 2^{t_v}}{2^{t_v}}$ . Note that  $0 \leq \lambda_v < 1$ . For  $v \in X$ , we define

$$(2) \quad f_{X,t}^{\text{root}}(v) = \begin{cases} \lambda_v \cdot g_{t,\alpha,\beta}(d_X(v, r)) & \text{if } t = t_v + 1, \\ (1 - \lambda_v) \cdot g_{t,\alpha,\beta}(d_X(v, r)) & \text{if } t = t_v, \\ 0 & \text{otherwise.} \end{cases}$$

Recall that  $g_{t,\alpha,\beta}$  was defined in (1). For all nodes  $v \notin X$ , we set  $f_X^{\text{root}}(v) = \vec{0}$ .

Define the map  $f_X = f_X^{\text{path}} \oplus f_X^{\text{root}}$ , and the final embedding is

$$f = \bigoplus_{i=1}^k \bigoplus_{X \in \mathcal{X}_i} f_X,$$

i.e., the concatenation of all the constructed embeddings. Before we start the analysis, let us record some simple observations.

**OBSERVATION 2.** *For the map  $f$  defined above, the following hold:*

- *The number of coordinates in  $f$  does not depend on  $\alpha, \beta$ .*
- *For every  $X \in \mathcal{X}_i$  and  $v \notin X$ , the map  $f_X(v)$  is the constant vector  $\vec{0}$ .*
- *For every  $X \in \mathcal{X}_i$  and  $v \in X$ , the map  $f_X$  is nonzero in at most three coordinates.*

Hence, since  $\mathcal{X}_i$  is a partial partition of  $V$  and the depth of the SPD is  $k$ , we get that  $f(v)$  is nonzero in at most  $3k$  coordinates for each  $v \in V$ .

**5. The analysis.** The main technical lemmas now show that the per-coordinate expansion is constant, and that for every pair there exists a coordinate for which the expected contraction is constant.

**LEMMA 2 (expansion bound).** *For any vertices  $u, v$ , every coordinate  $j$ , and every choice of  $\alpha, \beta$ ,*

$$|f_j(v) - f_j(u)| = O(d_G(u, v)).$$

**LEMMA 3 (contraction bound).** *For any vertices  $u, v$ , there exists some coordinate  $j$  such that*

$$\mathbb{E}_{\alpha,\beta} [|f_j(v) - f_j(u)|] = \Omega(d_G(u, v)).$$

Given these two lemmas, we can combine them together to show that the entire embedding has small distortion. (A proof of the composition lemma can be found in section 6.)

**LEMMA 4 (composition lemma).** *Let  $(X, d)$  be a metric space. Suppose that there are parameters  $\rho, \tau$  and a function  $f : X \rightarrow \mathbb{R}^s$ , drawn from some probability space such that*

1. *For every  $u, v \in X$  and every  $j \in [s]$ ,  $|f_j(v) - f_j(u)| \leq \rho \cdot d(v, u)$ .*
2. *For every  $u, v \in X$ , there exists  $j \in [s]$  such that  $\mathbb{E}[|f_j(v) - f_j(u)|] \geq \frac{1}{\tau} \cdot d(v, u)$ .*
3. *For every  $v \in X$ ,  $\|f(v)\|_0 \leq k$ , that is,  $f(v)$  has support of size at most  $k$  (formally, there is a subset of indices  $I_v \subseteq [s]$  of size  $\leq k$  such that  $\forall j \notin I_v, f_j(v) = 0$ ).*

*Then, for every  $p \geq 1$ , there is an embedding of  $(X, d)$  into  $\ell_p$  with distortion  $O(\rho^3 \tau^2) \cdot k^{1/p}$ . In particular, if  $\rho$  and  $\tau$  are constant, then the distortion is constant.*

Moreover, if there is an efficient algorithm for sampling such an  $f$ , then there is a randomized algorithm that constructs the embedding efficiently (in expectation).

Theorem 1, our embedding for graphs with low depth SPDs, immediately follows by applying the composition lemma (Lemma 4) to Lemmas 2 and 3, and Observation 2.

**5.1. Bounding the expansion: Proof of Lemma 2.** In this section we bound the expansion of any coordinate in our embedding. Recall that the embedding of  $v$  lying in some component  $X$  consists of two sets of coordinates: its distance from the path, and its distance from the root. As mentioned in the introduction, since points outside  $X$  are mapped to zero, maintaining Lipschitz-ness requires us to truncate the contribution of  $v$  of any coordinate to its distance from the boundary. This truncation (either via taking a minimum with  $d_G(v, V \setminus X)$ , or via the sawtooth function), means that our proofs of expansion require more care. Let us now give the details.

Consider any level  $i$ , any set  $X \in \mathcal{X}_i$ , and any pair of vertices  $u, v$ . It suffices to show that  $\|f_X(v) - f_X(u)\|_\infty = O(d_G(u, v))$ . To begin, we may assume that both  $u, v \in X$ . Indeed, if both  $u, v \notin X$ , then  $f_X(v) = f_X(u) = \vec{0}$  and we are done. If one of them, say  $v$ , belongs to  $X$  while the other  $u \notin X$ , then  $f_X(u) = \vec{0}$  while  $f_X(v)$  is bounded by  $2^{t_v+1} \leq 4d_G(v, V \setminus X) \leq 4d_G(u, v)$  in each coordinate.

Moreover, we may also assume that the shortest  $u$ - $v$  path in  $G$  contains only vertices from  $X$ . Indeed, suppose their shortest path in  $G$  uses vertices from  $V \setminus X$ , then  $d_G(u, V \setminus X) + d_G(v, V \setminus X) \leq d_G(u, v)$ . But since both  $f_X(v), f_X(u)$  are bounded in each coordinate by  $4 \cdot \max\{d_G(u, V \setminus X), d_G(v, V \setminus X)\}$ , we have constant expansion. Henceforth, we can assume that  $d_G(u, v) = d_X(u, v)$ . We now bound the expansion in each of the two parts of  $f_X$  separately.

**Expansion of  $f_X^{\text{path}}$ .** Let  $X_v, X_u$  be the connected components in  $G[X \setminus P_X]$  such that  $v \in X_v$  and  $u \in X_u$ . Consider the first case  $X_v \neq X_u$ , then  $P_X$  intersects the shortest path between  $v$  and  $u$ . In particular,

$$\begin{aligned} \|f_X^{\text{path}}(v) - f_X^{\text{path}}(u)\|_\infty &\leq \min\{d_X(v, P_X), 2d_G(v, V \setminus X)\} \\ &\quad + \min\{d_X(u, P_X), 2d_G(u, V \setminus X)\} \\ &\leq d_X(v, P_X) + d_X(u, P_X) \leq d_X(v, u) = d_G(v, u). \end{aligned}$$

Otherwise,  $X_v = X_u$  and the two vertices lie in the same component. Now  $\|f_X^{\text{path}}(v) - f_X^{\text{path}}(u)\|_\infty$  equals

$$\left| \min\{d_X(v, P_X), 2d_G(v, V \setminus X)\} - \min\{d_X(u, P_X), 2d_G(u, V \setminus X)\} \right|.$$

Assuming (without loss of generality (w.l.o.g.)) that the first term is at least the second, we can drop the absolute value signs. Now the bound on the expansion follows from a simple case analysis. Indeed, suppose  $d_X(u, P_X) \leq 2d_G(u, V \setminus X)$ . Then we get

$$\begin{aligned} \|f_X^{\text{path}}(v) - f_X^{\text{path}}(u)\|_\infty &= \min\{d_X(v, P_X), 2d_G(v, V \setminus X)\} - d_X(u, P_X) \\ &\leq d_X(v, P_X) - d_X(u, P_X) \leq d_X(u, v) = d_G(u, v). \end{aligned}$$

The other case is that  $d_X(u, P_X) > 2d_G(u, V \setminus X)$ , and then

$$\begin{aligned} \|f_X^{\text{path}}(v) - f_X^{\text{path}}(u)\|_\infty &= \min\{d_X(v, P_X), 2d_G(v, V \setminus X)\} - 2d_G(u, V \setminus X) \\ &\leq 2d_G(v, V \setminus X) - 2d_G(u, V \setminus X) \leq 2d_G(u, v). \end{aligned}$$

Hence the expansion is bounded by 2.

**Expansion of  $f_X^{\text{root}}$ .** Let  $r$  be the root of  $P_X$ . For  $t \in \{0, 1, \dots, M\}$ , let  $p_t$  (resp.,  $q_t$ ) be the “weight” of  $v$  (resp.,  $u$ ) on  $g_{t,\alpha,\beta}$ —in other words,  $p_t$  is the constant in (2) such that  $f_{X,t}^{\text{root}}(v) = p_t \cdot g_{t,\alpha,\beta}(d_X(v, r))$ . Note that  $p_t \in \{0, \lambda_v, 1 - \lambda_v\}$  is chosen deterministically, and is nonzero for at most two indices  $t$ .

First, observe that for every  $t$ ,

$$\begin{aligned}
 |f_{X,t}^{\text{root}}(v) - f_{X,t}^{\text{root}}(u)| &= |p_t \cdot g_{t,\alpha,\beta}(d_X(v, r)) - q_t \cdot g_{t,\alpha,\beta}(d_X(u, r))| \\
 &\leq \min\{p_t, q_t\} \cdot |g_{t,\alpha,\beta}(d_X(v, r)) - g_{t,\alpha,\beta}(d_X(u, r))| + |p_t - q_t| \cdot 2^t \\
 &\leq \min\{p_t, q_t\} \cdot \beta \cdot |d_X(v, r) - d_X(u, r)| + |p_t - q_t| \cdot 2^t \\
 (3) \quad &\leq O(d_G(u, v)) + |p_t - q_t| \cdot 2^t.
 \end{aligned}$$

The first inequality used that  $g_{t,\alpha,\beta}$  is bounded by  $2^t$ , and the second inequality that  $g_{t,\alpha,\beta}$  is  $\beta$ -Lipschitz; both follow from Observation 1. The last inequality follows by the triangle inequality (since we assumed that the shortest path from  $v$  to  $u$  is contained within  $X$ ).

Hence, it suffices to show that  $|p_t - q_t| = O(d_G(u, v)/2^t)$ . Indeed, for indices  $t \notin \{t_u, t_u + 1, t_v, t_v + 1\}$ ,  $p_t = q_t = 0$ , hence  $|p_t - q_t| = 0$ . Let us consider the other cases. Without loss of generality, assume that  $d_G(v, V \setminus X) \geq d_G(u, V \setminus X)$  and hence  $t_v \geq t_u$ .

- $t_u = t_v$ . In this case,  $|p_{t_v} - q_{t_v}| = |(1 - \lambda_v) - (1 - \lambda_u)| = \lambda_v - \lambda_u = |p_{t_v+1} - q_{t_v+1}|$ . Moreover, this quantity is

$$\begin{aligned}
 \lambda_v - \lambda_u &= \frac{2d_G(v, V \setminus X) - 2^t}{2^t} - \frac{2d_G(u, V \setminus X) - 2^t}{2^t} \\
 &= \frac{2(d_G(v, V \setminus X) - d_G(u, V \setminus X))}{2^t} \\
 &\leq \frac{2d_G(u, v)}{2^t}.
 \end{aligned}$$

Hence, we get that  $|p_t - q_t| = O(d_G(u, v)/2^t)$  for both  $t \in \{t_v, t_v + 1\}$ .

- $t_u = t_v - 1$ . It holds that

$$\begin{aligned}
 \lambda_v + (1 - \lambda_u) &\leq 2 \cdot \frac{2d_G(v, V \setminus X) - 2^{t_v}}{2^{t_v}} + \frac{2^{t_u+1} - 2d_G(u, V \setminus X)}{2^{t_u}} \\
 &= \frac{2d_G(v, V \setminus X) - 2d_G(u, V \setminus X)}{2^{t_u}} \leq \frac{2d_G(u, v)}{2^{t_u}}.
 \end{aligned}$$

If we define  $\chi := \lambda_v + (1 - \lambda_u)$ , we conclude that

$$\begin{aligned}
 |p_{t_v+1} - q_{t_v+1}| &= \lambda_v \leq \chi = O(d_G(u, v)/2^{t_v+1}), \\
 |p_{t_v} - q_{t_v}| &= |1 - \lambda_v - \lambda_u| \leq \chi = O(d_G(u, v)/2^{t_v}), \\
 |p_{t_u} - q_{t_u}| &= 1 - \lambda_u \leq \chi = O(d_G(u, v)/2^{t_u}).
 \end{aligned}$$

- $t_u < t_v - 1$ . By the definition of  $t_v$  and  $t_u$ ,

$$2d_G(v, u) \geq 2(d_G(v, V \setminus X) - d_G(u, V \setminus X)) \geq 2^{t_v} - 2^{t_u+1} \geq 2^{t_v-1}.$$

In particular, for every  $t \leq t_v + 1$ ,  $|p_t - q_t| \leq 1 \leq \frac{2d_G(u, v)}{2^{t_v-1}} = O\left(\frac{d_G(u, v)}{2^t}\right)$ .

**5.2. Bounding the contraction: Proof of Lemma 3.** Recall that we want to prove that for any pair  $u, v$  of vertices, the embedding has a large contribution between them. A natural proof idea is to show that vertices  $u, v$  would eventually be separated by the recursive procedure. When they are separated, either one of  $u, v$  is far from the separating path  $P$ , or they both lie close to the path. In the former case, the distance  $d(v, P)$  gives a large contribution to the embedding distance, and in the latter case the distance from one end of the path (the “root”) gives a large contribution.

However, there’s a catch: the value of  $v$ ’s embedding in any single coordinate cannot be more than  $v$ ’s distance to the boundary, which causes problems. Indeed, if  $u, v$  fall very close to the path  $P$  at some step of the algorithm, they must get most of their contribution at this level, since future levels will not give much contribution. How can we do it without assigning large values? This is where we use the sawtooth function: it gives a good contribution between points without assigning any vertex too large a value in any coordinate.

Formally, to bound the contraction and prove Lemma 3, for nodes  $u, v$  we need to show that there exists a coordinate  $j$  such that  $\mathbb{E}_{\alpha, \beta}[|f_j(v) - f_j(u)|] = \Omega(d_G(u, v))$ . For brevity, we define

$$(4) \quad \Delta_{uv} := d_G(u, v).$$

Fix  $c = 12$ . Let  $i$  be the minimal index such that there exists  $X \in \mathcal{X}_i$  with  $u, v \in X$ , and at least one of the following holds:

1.  $\min \{d_X(v, P_X), d_X(u, P_X)\} \leq \Delta_{uv}/c$  (i.e., we choose a path close to  $\{v, u\}$ ).
2.  $v$  and  $u$  are in different components of  $X \setminus P_X$ .

Note that such an index  $i$  indeed exists: if  $v$  and  $u$  are separated by the SPD, then condition 2. holds. The only other possibility that  $v$  and  $u$  are never separated is when at least one of them lies on one of the shortest paths. In such a case, surely condition 1. holds. By the minimality of  $i$ , for every  $X' \in \mathcal{X}_{i'}$  such that  $i' < i$  and  $u, v \in X'$ , necessarily  $\min\{d_{X'}(v, P_{X'}), d_{X'}(u, P_{X'})\} > \Delta_{uv}/c$ . Therefore, the ball with radius  $\Delta_{uv}/c$  around each of  $v, u$  is contained in  $X$ . In particular,  $\min \{d_G(v, V \setminus X), d_G(u, V \setminus X)\} > \Delta_{uv}/c$ .

Suppose first that condition 2. occurs but not condition 1. Let  $j$  be the coordinate in  $f_X^{\text{path}}$  created for the connected component of  $v$  in  $X \setminus P_X$ . Then

$$\begin{aligned} \left| (f_X^{\text{path}})_j(v) - (f_X^{\text{path}})_j(u) \right| &= \min \{d_X(v, P_X), 2d_G(v, V \setminus X)\} - 0 \\ &\geq \min \left\{ \frac{\Delta_{uv}}{c}, 2\frac{\Delta_{uv}}{c} \right\} = \frac{\Delta_{uv}}{c}. \end{aligned}$$

Next assume that condition 1. occurs. Without loss of generality, assume that  $d_X(v, P_X) \leq d_X(u, P_X)$ , so that  $d_X(v, P_X) \leq \Delta_{uv}/c$ . Suppose first that  $d_X(u, P_X) \geq 2\Delta_{uv}/c$ . Then in the coordinate  $j$  in  $f_X^{\text{path}}$  created for the connected component of  $u$  in  $X \setminus P_X$ , we have

$$\begin{aligned} \left| (f_X^{\text{path}})_j(v) - (f_X^{\text{path}})_j(u) \right| &\geq \left| \min \{d_X(u, P_X), 2d_G(u, V \setminus X)\} \right. \\ &\quad \left. - \min \{d_X(v, P_X), 2d_G(v, V \setminus X)\} \right| \\ &\geq \min \left\{ 2\frac{\Delta_{uv}}{c}, 2\frac{\Delta_{uv}}{c} \right\} - \frac{\Delta_{uv}}{c} = \frac{\Delta_{uv}}{c}. \end{aligned}$$

(It does not matter whether or not  $v, u$  are in the same connected component.) Thus it remains to consider the case  $d_X(u, P_X) < 2\Delta_{uv}/c$ . Let  $r$  be the root of  $P_X$ . Let  $v'$

(resp.,  $u'$ ) be the closest vertex on  $P_X$  to  $v$  (resp.,  $u$ ) in  $G[X]$ . Then by the triangle inequality

$$d_X(v', u') \geq d_X(v, u) - d_X(v, v') - d_X(u, u') \geq \frac{c-3}{c} \Delta_{uv} .$$

In particular,

$$\begin{aligned} |d_X(v, r) - d_X(u, r)| &\geq |d_X(v', r) - d_X(u', r)| - d_X(v, v') - d_X(u, u') \\ (5) \qquad \qquad \qquad &\geq \frac{c-6}{c} \Delta_{uv} = \frac{1}{2} \Delta_{uv} , \end{aligned}$$

where we used that  $P_X$  is a shortest path in  $G[X]$  (implying  $|d_X(v', r) - d_X(u', r)| = d_X(v', u')$ ). See Figure 5 for an illustration.

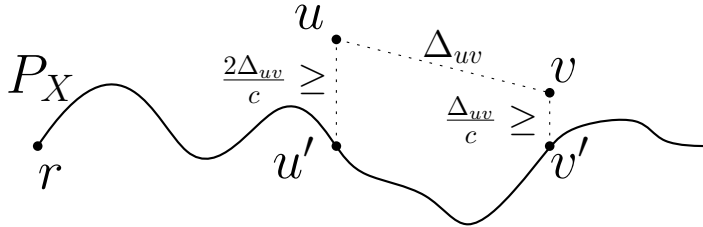


FIG. 5.  $P_X$  is a shortest path with root  $r$ .  $v$  (resp.,  $u$ ) is at distance at most  $\frac{\Delta_{uv}}{c}$  (resp.,  $\frac{2\Delta_{uv}}{c}$ ) from  $v'$  (resp.,  $u'$ ), it's closest vertex on  $P_X$ . By triangle inequality  $d_X(v', u') \geq (1 - \frac{3}{c})\Delta_{uv}$ . As  $u', v'$  lay on the same shortest path starting at  $r$ ,  $|d_X(v', r) - d_X(u', r)| = d_X(v', u')$ . Using the triangle inequality again we conclude  $|d_X(v, r) - d_X(u, r)| \geq |d_X(v', r) - d_X(u', r)| - \frac{3}{c} \Delta_{uv} \geq (1 - \frac{6}{c})\Delta_{uv}$ .

Set  $x = d_X(v, r)$  and  $y = d_X(u, r)$ . Assume first that  $d_G(v, V \setminus X) \geq d_G(u, V \setminus X)$ . In particular,  $t_v \geq t_u$ . By the definition of  $t_v$ ,  $2d_G(v, V \setminus X) \leq 2^{t_v+1}$ . Thus

$$(6) \qquad \qquad \qquad 2^{t_v} \geq \frac{\Delta_{uv}}{c} = \Omega(\Delta_{uv}).$$

CLAIM 1. Let  $t \geq t_v$ , then there is a constant  $\phi$  such that

$$\mathbb{E}_{\alpha, \beta} [|g_{t, \alpha, \beta}(x) - g_{t, \alpha, \beta}(y)|] \geq \Delta_{uv} / \phi.$$

Proof. According to property 2 of Lemma 1,

$$\mathbb{E}_{\alpha, \beta} [|g_{t, \alpha, \beta}(x) - g_{t, \alpha, \beta}(y)|] = \Omega(\min\{|x - y|, 2^t\}) \stackrel{(5) \& (6)}{=} \Omega(\Delta_{uv}) .$$

Set  $S = \max\{8\phi, \frac{8c}{2}\}$ . Note that  $p_{t_v} + p_{t_v+1} = (1 - \lambda_v) + \lambda_v = 1$ . Let  $t \in \{t_v, t_v + 1\}$  be such that  $p_t \geq \frac{1}{2}$ . We consider two cases:

- If  $|p_t - q_t| \cdot 2^t > \frac{\Delta_{uv}}{S}$ , then

$$\begin{aligned} \mathbb{E}_{\alpha, \beta} [|f_{X, t}^{\text{root}}(v) - f_{X, t}^{\text{root}}(u)|] &= \mathbb{E}_{\alpha, \beta} [|p_t \cdot g_{t, \alpha, \beta}(x) - q_t \cdot g_{t, \alpha, \beta}(y)|] \\ &\geq |p_t \cdot \mathbb{E}_{\alpha, \beta} [g_{t, \alpha, \beta}(x)] - q_t \cdot \mathbb{E}_{\alpha, \beta} [g_{t, \alpha, \beta}(y)]| \\ (7) \qquad \qquad \qquad &= |p_t - q_t| \cdot 2^t = \Omega(\Delta_{uv}), \end{aligned}$$

where the equality follows by property 1 of Lemma 1.

- Otherwise, using inequality (6),  $q_t \geq p_t - \frac{\Delta_{uv}}{2^{t^v} \cdot S} \geq \frac{1}{2} - \frac{2c}{2\Delta_{uv}} \cdot \frac{\Delta_{uv}}{S} \geq \frac{1}{4}$ . In particular,

$$\begin{aligned}
\mathbb{E}_{\alpha,\beta} [|f_{X,t}^{\text{root}}(v) - f_{X,t}^{\text{root}}(u)|] &= \mathbb{E}_{\alpha,\beta} [|p_t \cdot g_{t,\alpha,\beta}(x) - q_t \cdot g_{t,\alpha,\beta}(y)|] \\
&\geq \min\{p_t, q_t\} \cdot \mathbb{E}_{\alpha,\beta} [|g_{t,\alpha,\beta}(x) - g_{t,\alpha,\beta}(y)|] \\
&\quad - |p_t - q_t| \cdot 2^t \\
(8) \qquad \qquad \qquad &\geq \frac{1}{4} \cdot \frac{\Delta_{uv}}{\phi} - \frac{\Delta_{uv}}{S} = \Omega(\Delta_{uv}),
\end{aligned}$$

where in the first inequality we used property 1 of Lemma 1, and in the second inequality we used Claim 1.

Finally, recall that we assumed  $d_G(v, V \setminus X) \geq d_G(u, V \setminus X)$  for the proof above. The other case ( $d_G(v, V \setminus X) < d_G(u, V \setminus X)$ ) is completely symmetric.

**6. The composition lemma: Proof of Lemma 4.** We restate the lemma for convenience.

LEMMA 4 (composition lemma). *Let  $(X, d)$  be a metric space. Suppose that there are parameters  $\rho, \tau$  and a function  $f : X \rightarrow \mathbb{R}^s$ , drawn from some probability space such that*

1. *For every  $u, v \in X$  and every  $j \in [s]$ ,  $|f_j(v) - f_j(u)| \leq \rho \cdot d(v, u)$ .*
2. *For every  $u, v \in X$ , there exists  $j \in [s]$  such that  $\mathbb{E}[|f_j(v) - f_j(u)|] \geq \frac{1}{\tau} \cdot d(v, u)$ .*
3. *For every  $v \in X$ ,  $\|f(v)\|_0 \leq k$ , that is,  $f(v)$  has support of size at most  $k$  (formally, there is a subset of indices  $I_v \subseteq [s]$  of size  $\leq k$  such that  $\forall j \notin I_v, f_j(v) = 0$ ).*

*Then, for every  $p \geq 1$ , there is an embedding of  $(X, d)$  into  $\ell_p$  with distortion  $O(\rho^3 \tau^2) \cdot k^{1/p}$ . In particular, if  $\rho$  and  $\tau$  are constant, then the distortion is constant. Moreover, if there is an efficient algorithm for sampling such an  $f$ , then there is a randomized algorithm that constructs the embedding efficiently (in expectation).*

*Proof.* Fix  $n = |X|$  and set  $m = 48\rho\tau \cdot \ln n$ . Let  $f^{(1)}, f^{(2)}, \dots, f^{(m)} : X \rightarrow \mathbb{R}^s$  be functions chosen i.i.d. according to the given distribution. Set  $g = m^{-1/p} \bigoplus_{i=1}^m f^{(i)}$ . We argue that, with high probability,  $g$  has distortion  $16\rho^3\tau^2 \cdot k^{1/p}$  in  $\ell_p$ .

Fix some pair of vertices  $v, u \in V$ . Set  $d(v, u) = \Delta$ . The upper bound follows from property 1 and property 3 of the above lemma:

$$\begin{aligned}
\|g(v) - g(u)\|_p^p &= \sum_{i=1}^m \sum_{j \in I_v \cup I_u} \left( m^{-1/p} \cdot |f_j^{(i)}(v) - f_j^{(i)}(u)| \right)^p \\
&\leq \sum_{i=1}^m \sum_{j \in I_v \cup I_u} \frac{1}{m} \cdot (\rho \cdot \Delta)^p \leq 2k \cdot (\rho \cdot \Delta)^p,
\end{aligned}$$

thus  $\|g(v) - g(u)\|_p \leq 2^{\frac{1}{p}} \cdot \rho \cdot k^{1/p} \cdot \Delta$ .

Next, for the contraction bound, let  $j$  be the index of property 2 w.r.t.  $v, u$ . Set  $\mathcal{F} = \{f : |f_j(v) - f_j(u)| \geq \Delta/2\tau\}$  to be the event that we draw a function with significant contribution to  $v, u$ . Then using properties 1 and 2,

$$\begin{aligned}
\frac{\Delta}{\tau} &\leq \mathbb{E}[|f_j(v) - f_j(u)|] \\
&\leq \Pr[\overline{\mathcal{F}}] \cdot \frac{\Delta}{2\tau} + \Pr[\mathcal{F}] \cdot \rho\Delta \leq \frac{\Delta}{2\tau} + \Pr[\mathcal{F}] \cdot \rho\Delta,
\end{aligned}$$

which implies that  $\Pr[\mathcal{F}] \geq \frac{1}{2\rho\tau}$ . Let  $Q_{u,v}^{(i)}$  be an indicator random variable for the event  $f^{(i)} \in \mathcal{F}$ , and set  $Q_{u,v} = \sum_{i=1}^m Q_{u,v}^{(i)}$ . By linearity of expectation,  $\mathbb{E}[Q_{u,v}] \geq \frac{m}{2\rho\tau} = 24 \cdot \ln n$ . By a Chernoff bound

$$\begin{aligned} \Pr[Q_{u,v} \leq 12 \cdot \ln n] &\leq \Pr\left[Q_{u,v} \leq \frac{1}{2} \cdot \mathbb{E}[Q_{u,v}]\right] \\ &\leq \exp\left(-\frac{1}{8} \mathbb{E}[Q_{u,v}]\right) \\ &\leq \exp(-3 \ln n) = n^{-3}. \end{aligned}$$

By taking a union bound over the  $\binom{n}{2}$  pairs, with probability at least  $1 - \frac{1}{n}$  for every  $u, v \in V$ ,  $Q_{u,v} > 12 \ln n = \frac{m}{4\rho\tau}$ . If this event indeed occurs, then the contraction is indeed bounded:

$$\begin{aligned} \|g(v) - g(u)\|_p^p &\geq \sum_{i=1}^m \left(m^{-1/p} \cdot |f_j^{(i)}(v) - f_j^{(i)}(u)|\right)^p \\ &\geq \frac{1}{m} \sum_{i:Q_{u,v}=1} |f_j^{(i)}(v) - f_j^{(i)}(u)|^p \\ &\geq \frac{Q_{u,v}}{m} \cdot \left(\frac{\Delta}{2\rho\tau}\right)^p = \frac{1}{4\rho\tau} \cdot \left(\frac{\Delta}{2\rho\tau}\right)^p. \quad \square \end{aligned}$$

In particular, for every  $u, v$ ,  $\|g(v) - g(u)\|_p \geq \left(\frac{1}{4\rho\tau}\right)^{\frac{1}{p}} \cdot \frac{\Delta}{2\rho\tau}$ . Combining the upper and lower bounds, we conclude that  $g$  has distortion  $2^{\frac{1}{p}} \cdot \rho \cdot k^{1/p} \cdot 2\rho\tau \cdot (4\rho\tau)^{\frac{1}{p}} = 2^{1+\frac{3}{p}} \rho^{2+\frac{1}{p}} \tau^{1+\frac{1}{p}} \cdot k^{1/p}$ .

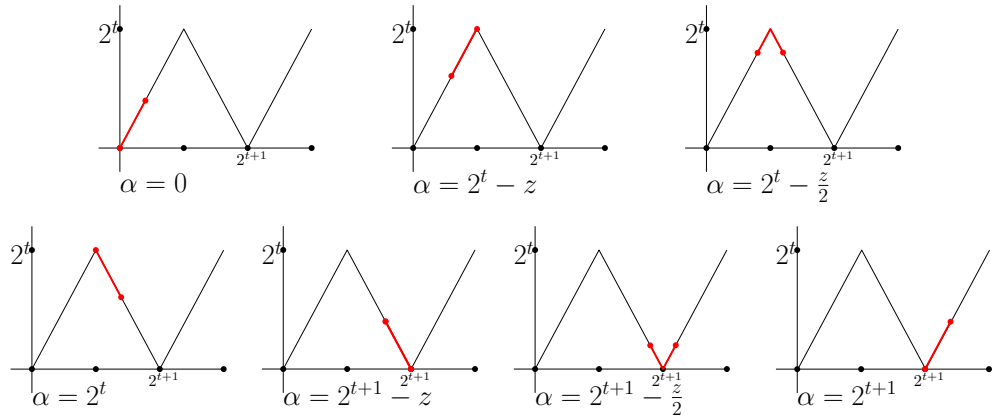


FIG. 6.  $\alpha$  is going from 0 to  $2^{t+1}$ .  $z \leq 2^t$ . In each of the figures the leftmost red point represents  $\alpha$  while the rightmost red point represents  $z + \alpha$ . Each of the middle figures represent a moment when  $g_t(z + \alpha) - g_t(z)$  changes its derivative. (Figure in color online.)

**7. The sawtooth lemma: Proof of Lemma 1.** We restate Lemma 1 for convenience.

LEMMA 1 (sawtooth lemma). Let  $x, y \in \mathbb{R}_+$ . Let  $\alpha \in [0, 1]$ ,  $\beta \in [0, 4]$  be drawn uniformly and independently. The following properties hold:



1.  $\mathbb{E}_{\alpha,\beta} [g_{t,\alpha,\beta}(x)] = 2^{t-1}$ .
2.  $\mathbb{E}_{\alpha,\beta} [|g_{t,\alpha,\beta}(x) - g_{t,\alpha,\beta}(y)|] = \Omega(\min\{|x - y|, 2^t\})$ .

Property 1 is straightforward, as by Observation 1  $g_t$  is periodic with period length  $2^{t+1}$ . Indeed, for every fixed  $\beta$ ,  $\mathbb{E}_\alpha [g_{t,\alpha,\beta}(x)] = \mathbb{E}_\alpha [g_t(\beta x + \alpha \cdot 2^{t+1})] = \mathbb{E}_\alpha [g_t(\alpha \cdot 2^{t+1})] = 2^{t-1}$ . The following claim will be useful in the proof of property 2.

CLAIM 2. For  $z \in [0, 2^{t+1}]$ ,  $\mathbb{E}_{\alpha \in [0,1]} [|g_t(z + \alpha \cdot 2^{t+1}) - g_t(\alpha \cdot 2^{t+1})|] = \frac{(2^{t+1}-z)z}{2^{t+1}}$ .

*Proof.* Set  $(*) = \mathbb{E}_{\alpha \in [0,1]} [|g_t(z + \alpha \cdot 2^{t+1}) - g_t(\alpha \cdot 2^{t+1})|]$ . By substituting the variable of integration,  $(*) = \frac{1}{2^{t+1}} \cdot \int_0^{2^{t+1}} |g_t(z + \alpha) - g_t(\alpha)| d\alpha$ . First, assume that  $z \leq 2^t$ , then there are five “phase changes” in  $|g_t(z + \alpha) - g_t(\alpha)|$  from 0 to  $2^{t+1}$  at  $2^t - z$ ,  $2^t - \frac{z}{2}$ ,  $2^t$ ,  $2^{t+1} - z$ ,  $2^{t+1} - \frac{z}{2}$ . (See Figure 6 for an illustration.)

We calculate

$$\begin{aligned} 2^{t+1} \cdot (*) &= \int_0^{2^t-z} z d\alpha + \int_0^{\frac{z}{2}} (z - 2\alpha) d\alpha + \int_0^{\frac{z}{2}} 2\alpha d\alpha + \int_0^{2^t-z} z d\alpha \\ &\quad + \int_0^{\frac{z}{2}} (z - 2\alpha) d\alpha + \int_0^{\frac{z}{2}} 2\alpha d\alpha \\ &= 2 \cdot \int_0^{2^t-z} z d\alpha + 2 \cdot \int_0^{\frac{z}{2}} z d\alpha = (2^{t+1} - z) z. \end{aligned}$$

For  $z > 2^t$ , set  $w = 2^{t+1} - z$ . Then using that  $g^t$  is periodic,

$$\begin{aligned} \mathbb{E}_{\alpha \in [0,1]} [|g_t(w + \alpha \cdot 2^{t+1}) - g_t(\alpha \cdot 2^{t+1})|] &= \mathbb{E}_{\alpha \in [0,1]} [|g_t(w + z + \alpha \cdot 2^{t+1}) \\ &\quad - g_t(z + \alpha \cdot 2^{t+1})|] \\ &= \mathbb{E}_{\alpha \in [0,1]} [|g_t(2^{t+1} + \alpha \cdot 2^{t+1}) \\ &\quad - g_t(z + \alpha \cdot 2^{t+1})|] \\ &= \mathbb{E}_{\alpha \in [0,1]} [|g_t(z + \alpha \cdot 2^{t+1}) - g_t(\alpha \cdot 2^{t+1})|]. \end{aligned}$$

Hence by the first case,  $(*) = \frac{(2^{t+1}-w)w}{2^{t+1}} = \frac{(2^{t+1}-z)z}{2^{t+1}}$ . □

For the proof of property 2 assume w.l.o.g. that  $x > y$ . Set  $z = x - y$  and  $(*) = \mathbb{E}_{\alpha,\beta} [|g_{t,\alpha,\beta}(x) - g_{t,\alpha,\beta}(y)|]$ . As  $g_t$  is a periodic function, we have that  $(*) = \mathbb{E}_\beta [\mathbb{E}_\alpha [|g_{t,\alpha,\beta}(z) - g_{t,\alpha,\beta}(0)|]]$ . The rest of the proof is by case analysis.

- If  $|x - y| \leq 2^{t-1}$ . Using Claim 2, we have

$$\begin{aligned} (*) &= \mathbb{E}_\beta \left[ \frac{(2^{t+1} - \beta z) \beta z}{2^{t+1}} \right] = \frac{1}{2^{t+1}} \cdot \frac{1}{4} \cdot \left( \frac{2^{t+1}z}{2} \beta^2 - \frac{z^2}{3} \cdot \beta^3 \Big|_0^4 \right) \\ &= \frac{1}{4} \cdot \left( \frac{16}{2} \cdot z - \frac{z^2}{2^{t+1}} \cdot \frac{2^6}{3} \right) \geq \frac{1}{4} \cdot \left( 8 - \frac{64}{3 \cdot 4} \right) \cdot z = \frac{2}{3} \cdot |x - y|, \end{aligned}$$

where in the inequality we used that  $z \leq 2^{t-1}$ .

- If  $|x - y| > 2^{t-1}$ .

As  $g_t$  is a periodic function, Claim 2 implies that for every  $w \geq 0$  it holds that  $\mathbb{E}_\alpha [|g_{t,\alpha,\beta}(w) - g_{t,\alpha,\beta}(0)|] = \frac{(2^{t+1} - (w \bmod 2^{t+1}))(w \bmod 2^{t+1})}{2^{t+1}}$ . Let  $a \in [0, 4]$  such that

$a \cdot z = 2^{t+1}$  (such  $a$  exists as  $|x - y| > 2^{t-1}$ ). The claim follows as

$$\begin{aligned}
 & (*) \cdot 2^{t+1} \\
 &= \mathbb{E}_{\beta \in [0,4]} \left[ \left( 2^{t+1} - (\beta z \bmod 2^{t+1}) \right) (\beta z \bmod 2^{t+1}) \right] \\
 &\geq \sum_{i=0}^{\lfloor \frac{4}{a} \rfloor - 1} \frac{1}{4} \cdot \int_{ia}^{(i+1)a} \left( 2^{t+1} - \left( \beta \cdot \frac{2^{t+1}}{a} \bmod 2^{t+1} \right) \right) \cdot \left( \beta \cdot \frac{2^{t+1}}{a} \bmod 2^{t+1} \right) d\beta \\
 &= \sum_{i=0}^{\lfloor \frac{4}{a} \rfloor - 1} \frac{1}{4} \cdot \int_0^a \left( 2^{t+1} - \beta \cdot \frac{2^{t+1}}{a} \right) \cdot \beta \cdot \frac{2^{t+1}}{a} d\beta \\
 &= \left\lfloor \frac{4}{a} \right\rfloor \cdot \frac{1}{4} \cdot \frac{a}{2^{t+1}} \cdot \int_0^{2^{t+1}} (2^{t+1} - \gamma) \cdot \gamma d\gamma \\
 &\geq \frac{4}{2a} \cdot \frac{1}{4} \cdot \frac{a}{2^{t+1}} \cdot \left( 2^{t+1} \frac{\gamma^2}{2} - \frac{\gamma^3}{3} \Big|_0^{2^{t+1}} \right) \\
 &= \frac{1}{2} \cdot \frac{1}{2^{t+1}} \cdot \frac{(2^{t+1})^3}{6} = \frac{(2^{t+1})^2}{12}.
 \end{aligned}$$

Property 2 now follows.

**8. Reducing the dimension.** In the previous sections we did not attempt to bound the dimension of our embedding (Theorem 1). As each point is nonzero in at most  $O(k \log n)$  coordinates (taking into account the repetitions done by the composition lemma (Lemma 4)), naively we can bound the number of coordinates by  $O(nk \log n)$ . However, we can improve further. By introducing some modifications to the embedding algorithm, we are able to bound the number of coordinates by  $O(k \log n)$ . Notice that this fact is interesting only for  $p > 2$ . For embeddings into  $\ell_2$ , one can easily reduce the dimension to  $O(\log n)$  using the Johnson–Lindenstrauss lemma [JL84]. Furthermore, for embeddings into  $\ell_p$  for  $p \in [1, 2)$ , we first embed into  $\ell_2$  (using dimension  $O(\log n)$ ). Then we embed from  $\ell_2$  into  $\ell_p$ . It is well known that  $\ell_2^d$  embeds into  $\ell_p^{O(d)}$  (for  $p \in [1, 2]$ ) with constant distortion (see [Mat13]), thus we conclude that our embedding can use only  $O(\log n)$  coordinates.

**THEOREM 5** (embeddings with bounded dimension). *Let  $G = (V, E)$  be an  $n$ -vertex weighted graph with an SPD of depth  $k$ . Then there exists an embedding  $f : V \rightarrow \ell_p^{O(k \log n)}$  with distortion  $O(k^{1/p})$ .*

*Proof.* Recall the embedding algorithm: we assumed that the minimal distance in  $G$  is 1, while the diameter is bounded by  $2^M$ . Let  $\{\mathcal{X}, \mathcal{P}\} = \{\{\mathcal{X}_1, \dots, \mathcal{X}_k\}, \{\mathcal{P}_1, \dots, \mathcal{P}_k\}\}$  be an SPD of depth  $k$  for  $G$ . For every index  $i \in [k]$  and cluster  $X \in \mathcal{X}_i$  we had two different embeddings  $f_X^{\text{path}}$  and  $f_X^{\text{root}}$ . The function  $f_X^{\text{path}}$  is a deterministic embedding that maps each point  $x \in X$  to its (truncated) distance from  $P_X$ , while using a different coordinate for each connected component in  $X \setminus P_X$ . The function  $f_X^{\text{root}}$  is an embedding that depends on random variables  $\alpha, \beta$ . It uses  $M + 1$  different coordinates that captures a randomly truncated distance to the root  $r$  of  $P_X$ .

In Lemma 2 we proved that our embedding is Lipschitz in each coordinate. In Lemma 3 we showed that for every pair of vertices  $v, u$  there is some coordinate  $j$  such that  $\mathbb{E}[|f_j(v) - f_j(u)|] = \Omega(d_G(u, v))$ . The coordinate  $j$  might come from either  $f^{\text{path}}$  or  $f^{\text{root}}$ . Denote by  $\mathcal{R}_{\text{path}} \subseteq \binom{V}{2}$  (resp.,  $\mathcal{R}_{\text{root}}$ ) the set of pairs for which the coordinates above come from  $f^{\text{path}}$  (resp.,  $f^{\text{root}}$ ). In order to replace expectation

with high probability, we invoke  $O(\log n)$  independent repetitions of our embedding (Lemma 4). We will modify each type of coordinates separately, arguing that a total of  $O(k \log n)$  coordinates suffices.

$f^{\text{path}}$ . We start with modifying the  $f^{\text{path}}$  type coordinates. First, note that as the value of this coordinates chosen deterministically, there is no reason to invoke the independent repetitions (Lemma 4). Next, consider a specific level  $i \in [k]$ . For every cluster  $X \in \mathcal{X}_{i+1}$ , let  $\Pi(X) \in \mathcal{X}_i$  be the cluster such that  $X \subseteq \Pi(X)$ . Denote by  $f_i^{\text{path}}$  the concatenation of all  $(f_{\Pi(X)}^{\text{path}})_X$  for  $X \in \mathcal{X}_{i+1}$ , and by  $f^{\text{path}}$  the concatenation of all  $f_i^{\text{path}}$  for  $i \in [k]$ . Set  $D = |\mathcal{X}_{i+1}|$ , note that  $f_i^{\text{path}}$  has exactly  $D$  coordinates, where each  $v \in V$  is nonzero in at most one coordinate. For every  $X \in \mathcal{X}_{i+1}$  pick a sequence  $\alpha^X \in \{\pm 1\}^m$ , where  $m = O(\log D)$ , such that for every different  $X, X' \in \mathcal{X}_{i+1}$  the number of coordinates where  $\alpha^X$  and  $\alpha^{X'}$  differ is at least  $\frac{m}{4}$ .<sup>3</sup> We define a new embedding  $h_i^{\text{path}} : V \rightarrow \mathbb{R}^m$ , such that for every  $v \in X \in \mathcal{X}_{i+1}$ ,  $h_i^{\text{path}}(v) = \frac{f_{\Pi(X)}^{\text{path}}(v)}{m^{1/p}} (\alpha_1^X, \dots, \alpha_m^X)$ . For  $v \in V$  that belongs to no cluster in  $\mathcal{X}_{i+1}$ , set  $h_i^{\text{path}}(v) = \vec{0}$ . Consider  $v, u \in V$ . If  $u, v$  are both belong to the same cluster  $X$ , then

$$\begin{aligned} \left\| h_i^{\text{path}}(v) - h_i^{\text{path}}(u) \right\|_p^p &= \sum_{i=1}^m \left| \alpha_i^X \cdot \left( \frac{f_{\Pi(X)}^{\text{path}}(v)}{m^{1/p}} - \frac{f_{\Pi(X)}^{\text{path}}(u)}{m^{1/p}} \right) \right|^p \\ &= \left| f_{\Pi(X)}^{\text{path}}(v) - f_{\Pi(X)}^{\text{path}}(u) \right|^p = \left\| f_i^{\text{path}}(v) - f_i^{\text{path}}(u) \right\|_p^p. \end{aligned}$$

On the other hand, if  $v \in X_v$  and  $u \in X_u$  belong to different clusters, it holds that

$$\begin{aligned} \left\| h_i^{\text{path}}(v) - h_i^{\text{path}}(u) \right\|_p^p &= \sum_{i=1}^m \frac{1}{m} \left| \alpha_i^{X_v} \cdot f_{\Pi(X_v)}^{\text{path}}(v) - \alpha_i^{X_u} \cdot f_{\Pi(X_u)}^{\text{path}}(u) \right|^p \\ \left\| h_i^{\text{path}}(v) - h_i^{\text{path}}(u) \right\|_p^p &\leq \sum_{i=1}^m \frac{1}{m} \left| f_{\Pi(X_v)}^{\text{path}}(v) + f_{\Pi(X_u)}^{\text{path}}(u) \right|^p \\ &\leq 2^p \cdot \left( \left( f_{\Pi(X_v)}^{\text{path}}(v) \right)^p + \left( f_{\Pi(X_u)}^{\text{path}}(u) \right)^p \right) \\ &= 2^p \cdot \left\| f_i^{\text{path}}(v) - f_i^{\text{path}}(u) \right\|_p^p \\ \left\| h_i^{\text{path}}(v) - h_i^{\text{path}}(u) \right\|_p^p &\geq \frac{m}{4} \cdot \frac{1}{m} \left| f_{\Pi(X_v)}^{\text{path}}(v) + f_{\Pi(X_u)}^{\text{path}}(u) \right|^p \\ &\geq \frac{1}{4} \cdot \frac{1}{2} \left( \left( f_{\Pi(X_v)}^{\text{path}}(v) \right)^p + \left( f_{\Pi(X_u)}^{\text{path}}(u) \right)^p \right) \\ &= \frac{1}{8} \cdot \left\| f_i^{\text{path}}(v) - f_i^{\text{path}}(u) \right\|_p^p. \end{aligned}$$

Note that  $h_i^{\text{path}}$  has  $m = O(\log D) \leq O(\log n)$  coordinates. Denote by  $h^{\text{path}}$  the concatenation of all  $h_i^{\text{path}}$  for  $i \in [k]$ . Then  $h^{\text{path}}$  has at most  $O(k \log n)$  coordinates, as desired. Moreover, for all  $u, v \in V$  it holds that

$$8^{-\frac{1}{p}} \cdot \left\| f^{\text{path}}(v) - f^{\text{path}}(u) \right\|_p \leq \left\| h^{\text{path}}(v) - h^{\text{path}}(u) \right\|_p \leq 2 \cdot \left\| f^{\text{path}}(v) - f^{\text{path}}(u) \right\|_p.$$

$f^{\text{root}}$ . Next, we modify the  $f^{\text{root}}$  type coordinates. Consider level  $i \in [k]$ , and a cluster  $X \in \mathcal{X}_i$ .  $f_X^{\text{root}} : V \rightarrow \mathbb{R}^{M+1}$  is a function that sends each vertex  $v \notin X$  to  $\vec{0}$ ,

<sup>3</sup>Such a set of sequences can be chosen greedily.

while each vertex  $v \in X$  has a specific scale  $t_v \in [0, M - 1]$ , such that  $f_X^{\text{root}}(v)$  can be nonzero only in coordinates  $t_v, t_{v+1}$ . Set  $h_X^{\text{root}} : V \rightarrow \mathbb{R}^2$  as a concatenation of  $h_{X,\text{odd}}^{\text{root}}, h_{X,\text{even}}^{\text{root}}$ , where  $h_{X,\text{odd}}^{\text{root}}$  (resp.,  $h_{X,\text{even}}^{\text{root}}$ ) is the sum of all the odd (resp., even) coordinates of  $f_X^{\text{root}}$ . That is,  $h_{X,\text{odd}}^{\text{root}} = \sum_{t=0}^{\lfloor M-1/2 \rfloor} f_{X,2t+1}^{\text{root}}$  and  $h_{X,\text{even}}^{\text{root}} = \sum_{t=0}^{\lfloor M/2 \rfloor} f_{X,2t}^{\text{root}}$ . Next, define  $h_i^{\text{root}} = \sum_{X \in \mathcal{X}_i} h_X^{\text{root}}$  as the sum of all  $h_X^{\text{root}}$  for  $X \in \mathcal{X}_i$ . Denote by  $f_i^{\text{root}}$  the sum of all  $f_X^{\text{root}}$  for  $X \in \mathcal{X}_i$ , and by  $f^{\text{root}}$  the concatenation of all  $f_i^{\text{root}}$  for  $i \in [k]$ . It is clear that the expansion is not increased in  $h_i^{\text{root}}$ , as for every  $v, u \in V$ , using the triangle inequality

$$\begin{aligned} \|h_i^{\text{root}}(v) - h_i^{\text{root}}(u)\|_p &\leq \sum_{X \in \mathcal{X}_i} \|h_{i,X}^{\text{root}}(v) - h_{i,X}^{\text{root}}(u)\|_p \\ &\leq \sum_{X \in \mathcal{X}_i} \sum_{t=0}^M \|f_{X,t}^{\text{root}}(v) - f_{X,t}^{\text{root}}(u)\|_p = \|f_i^{\text{root}}(v) - f_i^{\text{root}}(u)\|_p . \end{aligned}$$

Arguing that the expected contraction property is maintained is more involved. Consider a pair of vertices  $v, u \in V$ . Following the arguments in Lemma 3,  $i$  is the minimal index such that there exists  $X \in \mathcal{X}_i$  with  $u, v \in X$  such that either condition (1) or condition (2) hold. We can assume that condition (1) holds, and moreover, that  $d_X(v, P_X), d_X(u, P_X) \leq 2\Delta_{uv}/c$  (as otherwise the coordinate that contributes to the contraction comes from  $f^{\text{path}}$  and we have nothing to prove here). In particular, inequalities (5) and (6) and Claim 1 hold. Recall that we assumed  $t_v \geq t_u$ , and let  $t \in \{t_v, t_v + 1\}$  such that  $p_t \geq \frac{1}{2}$ . Without loss of generality, assume that  $t$  is odd. We proceed to the case analysis:

- If  $|p_t - q_t| \cdot 2^t > \frac{\Delta_{uv}}{S}$  and  $q_t \neq 0$ , note that for every odd  $t' \neq t$ ,  $f_{X,t'}^{\text{root}}(v) = f_{X,t'}^{\text{root}}(u) = 0$ . Therefore, following inequality (7)

$$\mathbb{E}_{\alpha,\beta} [|h_{i,\text{odd}}^{\text{root}}(v) - h_{i,\text{odd}}^{\text{root}}(u)|] = \mathbb{E}_{\alpha,\beta} [|f_{X,t}^{\text{root}}(v) - f_{X,t}^{\text{root}}(u)|] = \Omega(\Delta_{uv}) .$$

- Otherwise, if  $q_t = 0$ , there might be a single odd scale  $t' \leq t - 2$  such that  $q_{t'} \neq 0$  (if  $q_{t'} = 0$  for all odd scales, then the analysis above holds). We have

$$\begin{aligned} \mathbb{E}_{\alpha,\beta} [|h_{i,\text{odd}}^{\text{root}}(v) - h_{i,\text{odd}}^{\text{root}}(u)|] &= \mathbb{E}_{\alpha,\beta} [|f_{X,t}^{\text{root}}(v) - f_{X,t'}^{\text{root}}(u)|] \\ &= \mathbb{E}_{\alpha,\beta} [|p_t \cdot g_{t,\alpha,\beta}(x) - q_{t'} \cdot g_{t',\alpha,\beta}(y)|] \\ &\geq |p_t \cdot \mathbb{E}_{\alpha,\beta} [g_{t,\alpha,\beta}(x)] - q_{t'} \cdot \mathbb{E}_{\alpha,\beta} [g_{t',\alpha,\beta}(y)]| \\ &\geq p_t \cdot 2^{t-1} - q_{t'} \cdot 2^{t'-1} \geq \frac{1}{2} \cdot 2^{t-1} - 2^{t-3} = 2^{t-3} \\ &= \Omega(\Delta_{uv}) , \end{aligned}$$

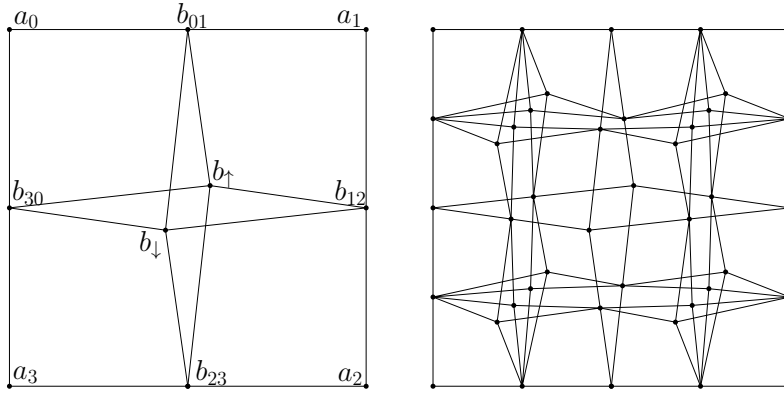
where the last equality follows by inequality (6).

- Otherwise,  $|p_t - q_t| \cdot 2^t \leq \frac{\Delta_{uv}}{S}$ . Using inequality (6),  $q_t \geq \frac{1}{4}$  (and therefore  $f_{X,t'}^{\text{root}}(v) = f_{X,t'}^{\text{root}}(u) = 0$  for every odd  $t' \neq t$ ). Following inequality (8),

$$\mathbb{E}_{\alpha,\beta} [|h_{i,\text{odd}}^{\text{root}}(v) - h_{i,\text{odd}}^{\text{root}}(u)|] = \mathbb{E}_{\alpha,\beta} [|f_{X,t}^{\text{root}}(v) - f_{X,t}^{\text{root}}(u)|] = \Omega(\Delta_{uv}) .$$

Define  $h^{\text{root}}$  the concatenation of all  $h_i^{\text{root}}$  for  $i \in [k]$ .  $h^{\text{root}}$  has exactly  $2k$  coordinates. We saw that  $h$  is Lipschitz in every coordinate. Moreover, for every  $\{u, v\} \in \mathcal{R}_{\text{root}}$ ,  $\mathbb{E}_{\alpha,\beta} [|h_{i,\text{odd}}^{\text{root}}(v) - h_{i,\text{odd}}^{\text{root}}(u)|] = \Omega(\Delta_{uv})$ .

Set  $h$  to be the concatenation of  $h^{\text{path}}$  and  $h^{\text{root}}$ . We now invoke the composition lemma (Lemma 4) to construct an embedding with distortion  $O(k^{\frac{1}{p}})$ . Recall that

FIG. 7. The two first diamondfold graphs  $\mathcal{X}_1$  and  $\mathcal{X}_2$ .

during the construction of Lemma 4 we sample and concatenate  $O(\log n)$  independent copies of  $h$  (normalized accordingly). As  $h^{\text{path}}$  is deterministic, it is enough to take only a single (nonnormalized) copy of  $h^{\text{path}}$ , and  $O(\log n)$  (normalized) copies of  $h^{\text{root}}$ . In particular, the total number of coordinates is  $O(k \log n) + O(\log n) \cdot 2k = O(k \log n)$ , as required.  $\square$

**9. Lower bounds for  $p = 1$ : Proof of Theorem 4.** Our lower bound will be proven using a slight modification of the *diamondfold* graphs presented by Lee and Sidiropoulos [LS11]. We begin by presenting the discrete version of the diamondfold graph.

**DEFINITION 2** (diamondfold graph). *Let  $\mathcal{X}_0, \mathcal{X}_1, \mathcal{X}_2, \dots$ , be a sequence of unweighted graphs defined as follows:  $\mathcal{X}_0$  is a simple cycle with four vertices:  $A = \{a_0, a_1, a_2, a_3\}$ .  $\mathcal{X}_1$  is obtained from  $\mathcal{X}_0$  by first subdividing each edge, getting new vertices  $B = \{b_{01}, b_{12}, b_{23}, b_{30}\}$ . In addition, we add two new vertices  $b_{\uparrow}, b_{\downarrow}$  and all the edges between them to  $B$ . As a result we get eight cycles of length 4 ( $\{a_0, b_{01}, b_{\uparrow}, b_{30}\}, \dots, \{a_2, b_{23}, b_{\downarrow}, b_{12}\}$ ). These cycles are called basic cycles. See Figure 7 for an illustration.*

*In general, in  $\mathcal{X}_k$  we have  $8^k$  basic cycles.  $\mathcal{X}_{k+1}$  is obtained from  $\mathcal{X}_k$  by replacing each basic cycle with a copy of  $\mathcal{X}_1$  (if an edge belong to several basic cycles, we subdivide it only once). An alternative way to construct  $\mathcal{X}_{k+1}$  will be to start with  $\mathcal{X}_1$ , and replace each of the eight basic cycles with a copy of  $\mathcal{X}_k$ .*

**THEOREM 6** (see [LS11]). *Every embedding of  $\mathcal{X}_k$  into  $\ell_1$  has distortion  $\Omega(\sqrt{\frac{k}{\log k}})$ .*

Intuitively, the diamondfold graphs are fractals. In particular, by deleting a constant number of shortest paths from  $\mathcal{X}_k$ , we obtain a graph where each connected component is a subgraph of  $\mathcal{X}_{k-1}$ . Therefore, by inductive arguments it should follow that  $\mathcal{X}_k$  has SPDdepth  $O(k)$ . Unfortunately, SPD possesses the counterintuitive property that deleting edges may increase the SPDdepth of a graph. In order to overcome this problem, and provide an upper bound on the SPDdepth, we slightly modify the construction of the diamondfold graph.

**DEFINITION 3** (buffered diamondfold graph). *Let  $\hat{\mathcal{X}}_1, \hat{\mathcal{X}}_2, \dots$ , be a sequence of weighted graphs. The first graph  $\hat{\mathcal{X}}_1$  is obtained from  $\mathcal{X}_1$  as follows: First, replace each of the vertices  $\{b_{01}, b_{12}, b_{23}, b_{30}, b_{\uparrow}, b_{\downarrow}\}$  with a copy of a star with four leaves,*

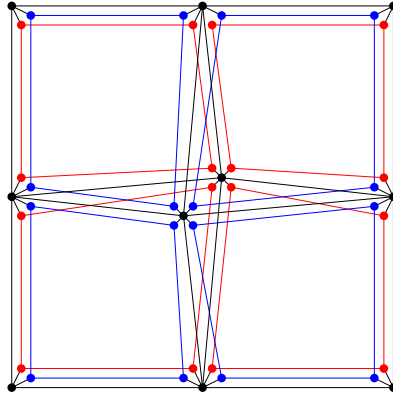


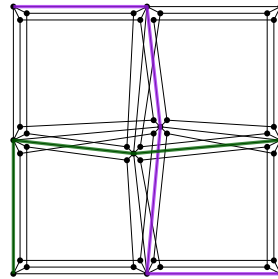
FIG. 8. The first buffered diamondfold graph  $\hat{\mathcal{X}}_1$ . Each basic cycle containing a vertex from the star of  $b_\uparrow$  (resp.,  $b_\downarrow$ ) is colored in red (resp., blue). (Figure in color online.)

such that each leaf participates in exactly one basic cycle. Similarly, replace each of the vertices  $\{a_0, a_1, a_2, a_3\}$  with a copy of a star with two leaves. See Figure 8 for an illustration. The weight of the star edges are defined to be  $\epsilon > 0$ , for  $\epsilon$  to be determined later. The weight of all the other edges remains 1. There are eight basic cycles in  $\hat{\mathcal{X}}_1$  corresponding to the eight basic cycles in  $\mathcal{X}_1$ .

In general, in  $\hat{\mathcal{X}}_k$  we have  $8^k$  basic cycles.  $\hat{\mathcal{X}}_{k+1}$  is obtained from  $\hat{\mathcal{X}}_k$  by replacing each basic cycle with a copy of  $\hat{\mathcal{X}}_1$ . An alternative way to construct  $\hat{\mathcal{X}}_{k+1}$  will be to start with  $\hat{\mathcal{X}}_1$ , and replace each of the eight basic cycles with a copy of  $\hat{\mathcal{X}}_k$ .

Each vertex in  $\mathcal{X}_k$  corresponds to some induced tree in  $\hat{\mathcal{X}}_k$  where all the edges in the tree have  $\epsilon$  weight. In particular, we can define an embedding  $f : \mathcal{X}_k \rightarrow \hat{\mathcal{X}}_k$  by sending each vertex  $v \in \mathcal{X}_k$  to one of its copies (i.e., to some vertex in its corresponding tree). It is clear that the distortion of  $f$  tends to 1 as  $\epsilon$  tends to 0. Therefore, Theorem 6 implies that for small enough  $\epsilon$ , every embedding of  $\hat{\mathcal{X}}_k$  into  $\ell_1$  has distortion  $\Omega(\sqrt{\frac{k}{\log k}})$ .

Next, we argue that  $\hat{\mathcal{X}}_k$  has SPDdepth  $O(k)$ . It is straightforward to verify that  $\hat{\mathcal{X}}_1$  has SPDdepth 4. Assume by induction that  $\hat{\mathcal{X}}_k$  has SPDdepth at most  $2(k+1)$ . Consider  $\hat{\mathcal{X}}_{k+1}$  which obtained from  $\hat{\mathcal{X}}_1$  by replacing each of the eight basic cycles with a copy of  $\hat{\mathcal{X}}_k$ . By deleting two shortest paths (the purple and green paths in the figure on the right), we obtain a graph consisting of eight connected components, each component isomorphic to  $\hat{\mathcal{X}}_k$ . By the induction hypothesis,  $\hat{\mathcal{X}}_{k+1}$  has SPDdepth at most  $2 + 2(k+1) = 2(k+2)$ . Theorem 5 now follows.



**10. Conclusions.** In this paper we introduced the notion of shortest path decompositions with low depth. We showed how these can be used to give embeddings into  $\ell_p$  spaces. Our techniques give optimal embeddings of bounded pathwidth graphs into  $\ell_2$ , and also new embeddings for graphs with bounded treewidth, planar, and excluded-minor families of graphs. Our embedding for the family of graphs with SPD depth  $k$  into  $\ell_p$  has an asymptotically matching lower bound for every fixed  $p > 1$ . The lower bound for  $p = 1$  differs from the upper bound of  $O(\sqrt{k})$  by an  $O(\sqrt{\log k})$

factor. Our techniques already have been useful for other embedding results, e.g., for embedding planar graphs with small face covers into  $\ell_1$  [Fil20a]. We hope that our techniques will find further applications.

Our work raises several open questions. While our embeddings are tight for fixed  $p > 1$ , can we improve the bounds for  $\ell_1$  embedding of graphs with bounded pathwidth? Can we give better results for the SPDdepth of  $H$ -minor-free graphs? Our approach gives a  $O(\sqrt{\log n})$ -distortion embedding of planar graphs into  $\ell_1$ , which is quite different from the previous known results using padded decompositions: can a combination of these ideas be used to make progress towards the planar graph embedding conjecture?

## REFERENCES

- [ABN11] I. ABRAHAM, Y. BARTAL, AND O. NEIMAN, *Advances in metric embedding theory*, Adv. Math., 228 (2011), pp. 3026–3126, <https://doi.org/10.1016/j.aim.2011.08.003>.
- [ACKW15] I. ABRAHAM, S. CHECHIK, R. KRAUTHGAMER, AND U. WIEDER, *Approximate nearest neighbor search in metrics of planar graphs*, in Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2015, Princeton, NJ, 2015, pp. 20–42, <https://doi.org/10.4230/LIPIcs.APPROX-RANDOM.2015.20>.
- [AG06] I. ABRAHAM AND C. GAVOILLE, *Object location using path separators*, in Proceedings of the Twenty-Fifth Annual ACM Symposium on Principles of Distributed Computing, PODC 2006, Denver, CO, ACM, pp. 188–197, <https://doi.org/10.1145/1146381.1146411>.
- [AGG<sup>+</sup>19] I. ABRAHAM, C. GAVOILLE, A. GUPTA, O. NEIMAN, AND K. TALWAR, *Cops, robbers, and threatening skeletons: Padded decomposition for minor-free graphs*, SIAM J. Comput., 48 (2019), pp. 1120–1145, <https://doi.org/10.1137/17M1112406>.
- [And86] T. ANDREAE, *On a pursuit game played on graphs for which a minor is excluded*, J. Combin. Theory, Series B, 41 (1986), pp. 37–47, [https://doi.org/10.1016/0095-8956\(86\)90026-2](https://doi.org/10.1016/0095-8956(86)90026-2).
- [Ass83] P. ASSOUD, *Plongements lipschitziens dans  $\mathbb{R}^n$* , Bull. Soc. Math. France, 111 (1983), pp. 429–448, <https://eudml.org/doc/87452>.
- [BBMN15] N. BANSAL, N. BUCHBINDER, A. MADRY, AND J. NAOR, *A polylogarithmic-competitive algorithm for the k-server problem*, J. ACM, 62 (2015), 40, <https://doi.org/10.1145/2783434>.
- [BCMN13] Y. BARTAL, D. E. CARROLL, A. MEYERSON, AND O. NEIMAN, *Bandwidth and low dimensional embedding*, Theoret. Comput. Sci., 500 (2013), pp. 44–56, <https://doi.org/10.1016/j.tcs.2013.05.038>.
- [BGHK92] H. L. BODLAENDER, J. R. GILBERT, H. HAFSTEINSSON, AND T. KLOKS, *Approximating treewidth, pathwidth, and minimum elimination tree height*, Lecture Notes in Comput. Sci. 570, Springer, Berlin, Heidelberg, 1992, pp. 1–12, [https://doi.org/10.1007/3-540-55121-2\\_1](https://doi.org/10.1007/3-540-55121-2_1).
- [Bou85] J. BOURGAIN, *On lipschitz embedding of finite metric spaces in Hilbert space*, Israel J. Math., 52 (1985), pp. 46–52, <https://doi.org/10.1007/BF02776078>.
- [CGN<sup>+</sup>06] C. CHEKURI, A. GUPTA, I. NEWMAN, Y. RABINOVICH, AND A. SINCLAIR, *Embedding k-outerplanar graphs into  $\ell_1$* , SIAM J. Discrete Math., 20 (2006), pp. 119–136, <https://doi.org/http://dx.doi.org/10.1137/S0895480102417379>.
- [CJLV08] A. CHAKRABARTI, A. JAFFE, J. R. LEE, AND J. VINCENT, *Embeddings of topological graphs: Lossy invariants, linearization, and 2-sums*, in Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, Philadelphia, PA, 2008, pp. 761–770, <https://doi.org/10.1109/FOCS.2008.79>.
- [Fil20a] A. FILTSEER, *A face cover perspective to  $\ell_1$  embeddings of planar graphs*, in Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA 2020, S. Chawla, ed., Salt Lake City, SIAM, Philadelphia, 2020, pp. 1945–1954, <https://doi.org/10.1137/1.9781611975994.120>.
- [Fil20b] A. FILTSEER, *Scattering and sparse partitions, and their applications*, in Proceedings of the 47th International Colloquium on Automata, Languages, and Programming, ICALP 2020, Saarbrücken, Germany (Virtual Conference), A. Czumaj, A. Dawar, and E. Merelli, eds., Leibniz International Proceedings in Informatics (LIPIcs) 168,

- Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020, 47, <https://doi.org/10.4230/LIPIcs.ICALP.2020.47>.
- [GKL03] A. GUPTA, R. KRAUTHGAMER, AND J. R. LEE, *Bounded geometries, fractals, and low-distortion embeddings*, in Proceedings of the 44th Symposium on Foundations of Computer Science (FOCS 2003), Cambridge, MA, 2003, pp. 534–543, <https://doi.org/10.1109/SFCS.2003.1238226>.
- [GKR04] A. GUPTA, A. KUMAR, AND R. RASTOGI, *Traveling with a Pez dispenser (or, routing issues in MPLS)*, SIAM J. Comput., 34 (2004), pp. 453–474, <https://doi.org/10.1137/S0097539702409927>.
- [GNRS04] A. GUPTA, I. NEWMAN, Y. RABINOVICH, AND A. SINCLAIR, *Cuts, trees and  $\ell_1$ -embeddings of graphs*, Combinatorica, 24 (2004), pp. 233–269, <https://doi.org/10.1007/s00493-004-0015-x>.
- [JL84] W. JOHNSON AND J. LINDENSTRAUSS, *Extensions of lipschitz maps into a hilbert space*, Contemp. Math., 26 (1984), pp. 189–206, <https://doi.org/10.1090/conm/026/737400>.
- [JLM11] A. JAFFE, J. R. LEE, AND M. MOHARRAMI, *On the optimality of gluing over scales*, Discrete Comput. Geom., 46 (2011), pp. 270–282, <https://doi.org/10.1007/s00454-011-9359-3>.
- [JLN<sup>+</sup>05] L. JIA, G. LIN, G. NOUBIR, R. RAJARAMAN, AND R. SUNDARAM, *Universal approximations for tsp, steiner tree, and set cover*, in Proceedings of the 37th Annual ACM Symposium on Theory of Computing, Baltimore, MD, 2005, H. N. Gabow and R. Fagin, eds., ACM, 2005, pp. 386–395, <https://doi.org/10.1145/1060590.1060649>.
- [JS09] W. B. JOHNSON AND G. SCHECHTMAN, *Diamond graphs and super-reflexivity*, J. Topology Anal., 01 (2009), pp. 177–189, <https://doi.org/10.1142/S1793525309000114>.
- [KK16] L. KAMMA AND R. KRAUTHGAMER, *Metric decompositions of path-separable graphs*, Algorithmica, 2016, pp. 645–653, <https://doi.org/10.1007/s00453-016-0213-0>.
- [KKM<sup>+</sup>12] M. KHAN, F. KUHN, D. MALKHI, G. PANDURANGAN, AND K. TALWAR, *Efficient distributed approximation algorithms via probabilistic tree embeddings*, Distrib. Comput., 25 (2012), pp. 189–205, <https://doi.org/10.1007/s00446-012-0157-9>.
- [KLMN05] R. KRAUTHGAMER, J. R. LEE, M. MENDEL, AND A. NAOR, *Measured descent: a new embedding method for finite metrics*, Geom. Funct. Anal., 15 (2005), pp. 839–858, <https://doi.org/10.1007/s00039-005-0527-6>.
- [KS93] E. KORACH AND N. SOLEL, *Tree-width, path-width, and cutwidth*, Discrete Appl. Math., 43 (1993), pp. 97–101, [https://doi.org/https://doi.org/10.1016/0166-218X\(93\)90171-J](https://doi.org/https://doi.org/10.1016/0166-218X(93)90171-J).
- [LLR95] N. LINIAL, E. LONDON, AND Y. RABINOVICH, *The geometry of graphs and some of its algorithmic applications*, Combinatorica, 15 (1995), pp. 215–245, <https://doi.org/10.1007/BF01200757>.
- [LN04] J. R. LEE AND A. NAOR, *Embedding the diamond graph in  $\ell_p$  and dimension reduction in  $\ell_1$* , Geom. Funct. Anal., 14 (2004), pp. 745–747, <https://doi.org/10.1007/s00039-004-0473-8>.
- [LR10] J. R. LEE AND P. RAGHAVENDRA, *Coarse differentiation and multi-flows in planar graphs*, Discrete Comput. Geom., 43 (2010), pp. 346–362, <https://doi.org/10.1007/s00454-009-9172-4>.
- [LS10] J. R. LEE AND A. SIDIROPOULOS, *Genus and the geometry of the cut graph*, in Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2010, Austin, TX, SIAM, Philadelphia, 2010, pp. 193–201, <https://doi.org/10.1137/1.9781611973075.18>.
- [LS11] J. R. LEE AND A. SIDIROPOULOS, *Near-optimal distortion bounds for embedding doubling spaces into  $l_1$* , in Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, ACM, 2011, pp. 765–772, <https://doi.org/10.1145/1993636.1993737>.
- [LS13] J. R. LEE AND A. SIDIROPOULOS, *Pathwidth, trees, and random embeddings*, Combinatorica, 33 (2013), pp. 349–374, <https://doi.org/10.1007/s00493-013-2685-8>.
- [Mat99] J. MATOUŠEK, *On embedding trees into uniformly convex Banach spaces*, Israel J. Math., 114 (1999), pp. 221–237, <https://doi.org/10.1007/BF02785579>.
- [Mat13] J. MATOUŠEK, *Lecture Notes on Metric Embeddings*, Technical report, 2013, <https://kam.mff.cuni.cz/~matousek/ba-a4.pdf>.
- [Mil86] G. L. MILLER, *Finding small simple cycle separators for 2-connected planar graphs*, J. Comput. System Sci., 32 (1986), pp. 265–279, [https://doi.org/10.1016/0022-0000\(86\)90030-9](https://doi.org/10.1016/0022-0000(86)90030-9).



- [MN13] M. MENDEL AND A. NAOR, *Markov convexity and local rigidity of distorted metrics*, J. European Math. Soc., 15 (2013), pp. 287–337, <https://doi.org/10.4171/JEMS/362>.
- [NR03] I. NEWMAN AND Y. RABINOVICH, *A lower bound on the distortion of embedding planar metrics into Euclidean space*, Discrete Comput. Geom., 29 (2003), pp. 77–81, <https://doi.org/10.1007/s00454-002-2813-5>.
- [OS81] H. OKAMURA AND P. SEYMOUR, *Multicommodity flows in planar graphs*, J. Combin. Theory, Series B, 31 (1981), pp. 75–81, [https://doi.org/10.1016/S0095-8956\(81\)80012-3](https://doi.org/10.1016/S0095-8956(81)80012-3).
- [Rab08] Y. RABINOVICH, *On average distortion of embedding metrics into the line*, Discrete Comput. Geom., 39 (2008), pp. 720–733, <https://doi.org/10.1007/s00454-007-9047-5>.
- [Rao99] S. RAO, *Small distortion and volume preserving embeddings for planar and Euclidean metrics*, in Proceedings of the Fifteenth Annual Symposium on Computational Geometry, Miami Beach, FL, 1999, pp. 300–306, <https://doi.org/10.1145/304893.304983>.
- [ST04] D. A. SPIELMAN AND S. TENG, *Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems*, in Proceedings of the 36th Annual ACM Symposium on Theory of Computing, Chicago, IL, 2004, L. Babai, ed., 2004, pp. 81–90, <https://doi.org/10.1145/1007352.1007372>.
- [Tho04] M. THORUP, *Compact oracles for reachability and approximate distances in planar digraphs*, J. ACM, 51 (2004), pp. 993–1024, <https://doi.org/10.1145/1039488.1039493>.