Deterministic Online Embedding of Metric Spaces into Low Dimensional Spaces

Noam Licht
Department of Computer Science, University of Haifa, Israel

Ilan Newman[1]
Department of Computer Science, University of Haifa, Israel

Yuri Rabinovich
Department of Computer Science, University of Haifa, Israel

Abstract

We study online embeddings of metric spaces into Euclidean spaces of a constant dimension $d > 1$, against an adaptive adversary. While the case of $d = 1$ is well understood, for higher dimensions little is known. In particular, even for $d = 2$ it remains unknown whether the worst-case distortion grows exponentially with the number of exposed points, as it does in the case for the line, or whether it is polynomial, as in the case for unbounded d.

Our first result is about fixed *solid* graphs, i.e., K_5, whose edges are solid intervals, equipped with the shortest-path metric. We show that if the input points arrive from such a metric space, they can indeed be online-embedded into $\mathbb{R}^2$ with a polynomial distortion. This refutes the previously believed conjecture that the topological non-embeddability of K_5 into the plane could be exploited for establishing exponential lower bounds.

The second results is about online embeddings of tree metrics of a certain type, including, e.g., ultrametrics and HST's. Somewhat surprisingly, we show that for metrics from this class the worst-case online embedding into $\mathbb{R}^d$ is not much worse that the offline embedding, both being $n^{\Theta(1/d)}$, and this holds even when $d = \Theta(\log n)$. This is in a stark contrast to the more common situation where the online-offline gap is typically huge, and even exponential. This result allows us to transfer results about probabilistic embeddings of metrics into HST's to low-dimensional Euclidean spaces, in an almost optimal possible manner.

2012 ACM Subject Classification Theory of computation → Design and analysis of algorithms; Theory of computation → Online algorithms

Keywords and phrases online embedding, metric embedding, online algorithms, design of algorithms

Digital Object Identifier 10.4230/LIPIcs.ESA.2026.39

Related Version *Full Version*: `https://arxiv.org/abs/2607.10624`

1 Introduction

The modern theory of low-distortion embeddings of finite metric spaces began to take shape with the appearance of classical results of Johnson and Lindenstrauss [8][2] and Bourgain [6][3], in the last decades of the 20'th century. It soon became clear that this theory provides powerful tools for numerous theoretical and practical algorithmic problems. Today, it has evolved into a well-established discipline employing advanced mathematical concepts, and playing a crucial role in solving complex algorithmic problems across a variety of domains.

[1] corresponding author
[2] *Any n-point Euclidean metric can be efficiently embedded into $\ell_2^{\log n/\epsilon^2}$ with $(1+\epsilon)$-distortion.*
[3] *Any n-point metric can be efficiently embedded into Euclidean space with distortion $O(\log n)$.*

34th Annual European Symposium on Algorithms (ESA 2026).
Editors: Philip Bille, Seth Pettie, and Sabine Storandt; Article No. 39; pp. 39:1–39:18
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

The study of online metric embedding is a relatively recent development. In this setting, data points are presented one by one, and decisions must be made as they arrive. The goal, as before, is to preserve the distances or metric properties as much as possible. While the online setting is better suited for needs of modern real-time data processing, achieving good guarantees on metric distortion in this setting becomes significantly harder.

The first publication explicitly dedicated to online embeddings was [7] from 2010. One of the key observations of this paper was that a large part of Bartal's probabilistic offline embedding procedure into Hierarchically Well Separated Trees (henceforth, *HST*'s) of [2] can be implemented online. This line of research, continued in [3], and strengthened in [5], yields, e.g., that: *Any metric space (X, d) can be probabilistically online embedded into a distribution of a non-contracting ultrametrics (and hence tree-metrics) with dilation $O(\log q \cdot \log \min(q, \Delta))$. Here q is the number of exposed points, and it does not need to be known in advance. Δ stands for the* aspect ratio *of (X, d), i.e., the ratio between the largest and the smallest distances therein. The result is essentially tight.*

Importantly, in this probabilistic setting one considers only the *expected* dilation of the distances. The guarantee on distortion holds with high probability but is not assured. More importantly, the adversary is assumed to be non-adaptive, i.e., she must choose its input metric (X, d) before the exposure/embedding dialogue starts. This is a severe restriction.

In this paper we are interested in the deterministic version of online embeddings, where the embedding is produced according to a deterministic protocol, and the adversary is allowed to be adaptive. That is, the metric (X, d) does not need to be set in advance, and the choice of the next exposed point and its distances may depend on the embedding produced so far. This setting was studied explicitly in [10], where tight exponential bound for online embedding of (X, d) into the line was proved, and $(1+\epsilon)$-distortion online embedding into ℓ_∞ of a huge dimension. In addition, [10] establishes a lower bound of $\Omega(n^{0.5})$ (implicit already in [9]) for distortion of an online deterministic embedding into ℓ_2 of unbounded dimension. Interestingly, the "hard" input metric is quite simple, and it offline embeds into the line with a constant distortion. An essentially matching upper bound was recently obtained in [5] by means of an elegant novel construction.

The focus of the present paper are deterministic online embeddings into ℓ_2^k of a low, and in particular constant, dimension.

Our first result deals with embeddings into the plane. Currently, it is not even known whether the deterministic online embedding into the plane can ensure distortion[4] polynomial in the number of the exposed points, or exponential distortion is unavoidable, as it is the case of embedding into the line. Let us remark that for probabilistic online embedding a polynomial (expected) distortion is indeed feasible, e.g., by randomly projecting into the plane the embedding of [5]. To put this in context, the worse *offline* distortion of embedding a metric into the plane is linear in its size.

A standard technique for lower-bounding the distortion of embedding of (X, d) into $\mathbb{R}^2$, both offline and online, is to exploit the topological non-embeddability of (X, d) into $\mathbb{R}^2$. In the offline setting, one usually takes an ϵ-net of (X, d) and argues that it cannot be embedded too well. E.g., such an argument was used in [11] for (offline) embedding submetrics of the 2-dim sphere into $\mathbb{R}^2$. In the online setting, one starts with a small ϵ-net, allocates a small subset of X whose embedding is problematic, and locally amplifies the error. This approach was used, e.g., in [10] to establish exponential lower bounds for deterministic online embedding into tree metrics.

[4] Distortion is the product of the worst metric expansion and the worst contraction. It is defined formally in Section 2.

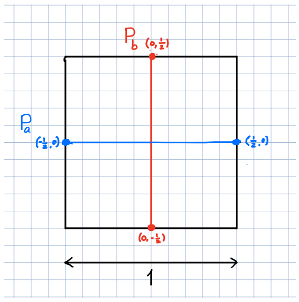

Figure 1 The horizontal path P_a and vertical path P_b should be embedded in the grid with endpoints as shown.

In view of the results in [10], it seemed that the topological non-embeddability into $\mathbb{R}^2$ of e.g., the *solid* K_5 could be exploited, in a similar manner, to establish exponential lower bounds on embedding its submetrics into the plane. Such metrics indeed embed badly offline. Our first finding is that, surprisingly, this topological obstacle can be handled at a relative small cost to the distortion. The following basic problem captures the topological obstacle, and will serve as a gadget for further constructions.

The *two-path game* is a two players game whose board is the unit square $[-\frac{1}{2},\frac{1}{2}]^2 \subset \mathbb{R}^2$. There are two unit-length paths: a blue path and a red one. They form a metric space, where the distance between two points of the same color is their distance in the corresponding unit-length interval, while the distance between any blue and any red point is 1. The endpoints of the blue path are glued to $(-\frac{1}{2},0),(\frac{1}{2},0)$ respectively, and the endpoints of the red path are glued to $(0,-\frac{1}{2}),(0,\frac{1}{2})$ respectively, see Figure 1. Both path must lie in the board. The Challenger in his turn exposes points at his will, specifying their position on the paths. The Embedder, in his turn, must place the exposed point on the board. The Challenger's goal is to maximize the metric distortion, the Embedder goal is to keep it low.

A standard analysis shows that if the q points are given in advance, the Embedder can ensure distortion of $O(q)$, and this is tight. What about the online situation? Assuming that q is known in advance, we show that Embedder can ensure $O(q^2)$ distortion. When q is not known, a slightly weaker upper bound of $O(q^2 \log^{2+\epsilon} q)$ can be ensured at every round q.

We proceed to analyze the case when the input points are drawn from a fixed solid graph G, i.e., the edges are unit length intervals glued at the vertices, and the distances are induced by the shortest-path metric of G. To avoid graph-automorphism issues, we assume that the vertices and edges are uniquely labeled, and that the input points are specified by their position on the edges. Using the ideas from the two-paths game, we construct a deterministic online embedding of the exposed points that is quadratic in q, and polynomial in $|V(G)|$. We further study the weighted solid G's, and obtain similar upper bounds.

Our second direction of study is embedding of ultrametrics and a certain generalization of HST into ℓ_2^k. It is well known that q points from an ultrametric embed isometrically into ℓ_2^{q-1}. Due to rigidity of ℓ_2, i.e., the uniqueness of the embedding up to translations, rotations and reflections, this embedding can be produced online as well. The optimal offline embeddings of ultrametrics into ℓ_2^k for k's logarithmic in q are studied in [1], establishing an essentially tight distortion of $k^{O(1)}q^{1/k}$. For k's sub-logarithmic in q, the minimal dimension required for $(1+\epsilon)$ embedding is studied in [4]. In view of the powerful results mentioned above about the (probabilistic) embedding of metric spaces into a distribution of HST's, it make sense to ask also about deterministic online embeddings of such metrics into ℓ_2^k.

For ultrametrics as an input, we show that q online exposed points can be embedded into ℓ_2^k with distortion $O(q^{(2/k)})$. In particular, ensuring a constant distortion for embedding into $\ell_2^{\theta(\log q)}$. This is surprisingly close to the optimal offline bounds. However, here we assume that q is known in advance. Removing this requirement, and possibly improving the distortion, is left for the future study. In addition, we introduce a natural class of metrics called α-trees that generalize the HST's, and obtain for them similar bounds.

To sum up, our contribution is mainly in showing that some non-trivial metric classes can be effectively embedded online, even in low dimensional Euclidean space contrary to prior expectations. It is our hope that these findings will be of help in the future study of online embedding arbitrary metric spaces in ℓ_2^k.

2 Notations and Preliminaries

Distortion of embedding: Let (X,d), (Y,μ) be two metric spaces and $\phi : X \longrightarrow Y$ an embedding of (X,d) into (Y,μ). The standard definition of (multiplicative) metric distortion of ϕ is given by

$$dist(\phi) = \max_{x,y\in X} \frac{\mu(\phi(x),\phi(y))}{d(x,y)} \cdot \max_{u,v\in X} \frac{d(u,v)}{\mu(\phi(u),\phi(v))}$$

The first term of this product is called the expansion of ϕ, and the second term is called the contraction. In the offline setting, the distortion of embedding (X,d) into (Y,μ) can also be defined as the minimal possible expansion of a non-contracting ϕ, or, equivalently, as a minimal possible contraction of a non-expanding ϕ. In the online setting, the above two definitions are both valid, but not necessarily equivalent, as an appropriate scaling might not be known in advance.

Solid graph metrics. Let $G = (V,E)$ be a connected graph with nonnegative weights (lengths) $\ell(e)$ on the edges. It may contain parallel edges and self-loops. Its "solid" one-dimensional realization is obtained by treating each edge $e \in E$ as one-dimensional intervals of length $\ell(e)$, whose endpoints are glued at the vertices as in G. The solid graph equipped with its geodesic (i.e., shortest path) metric, is an *infinite* metric space, whose points are labeled by the edge containing them, and their position on this edge. The distance between two such points can be easily computed by their labels, and the edge weights $\ell(e)$ of G.

In the online setting, embedding points from the solid (X_G,μ) induced by a given weighted graph G, when a point x lying on the edge $e=(u,v)$ is exposed, we first ensure that the corresponding endpoints u and v are already embedded, and only then address the embedding of x. We also assume that G,l are known in advance. We do so to avoid dealing with graph isomorphism problems, although most results for online embedding of X_G will have a similar upper bounds even if the "end points" of the edge in which the given point lies are not given.

Ultrametrics. A metric space (X,d) is ultrametric, abbreviated UM, on a *finite* set X if for every three points $x,y,z \in X$, $d(x,z) \le \max\{d(x,y),d(y,z)\}$. It is folklore (and easy) that each of the following conditions characterize UM.

(a) (X,d) is UM if there exists a weighted rooted tree $T=(V,E)$ with root r, and leaves $X \subset V$, such that $d_T(r,x)$ is the same for every $x \in X$, and $d = d|_X$ is the metric induced by d_T on the leaves X.

(b) X is a subset of the vertex set of a weighted undirected connected graph $G=(V,E)$, and such that $d(x,y) = \min_{P_{x,y}} \max_{e\in P_{x,y}} w(e)$, where P_{xy} ranges over all paths from x to y.

By (a), UM is a special case of a tree metric. It is quite obvious that any metric on a finite set X of size n can be embedded (offline) into UM with distortion at most $n-1$. This is so as the graph $G = (X, E)$ for which $w(x, y) = d(x, y)$ represent a UM d' on X (by (b) above). Since the weight of the heaviest edge in any path is at most the weighted length of the path and at least this length divided by $n-1$, we conclude that $d/(n-1) \le d' \le d$. Moreover, this is best possible as shown by the the unit weighted path of length n.

α-Trees. Let $T = (V, E)$ be an edge-weighted rooted tree with root $r \in V$, and $0 < \alpha < 1$. T is an α- tree (α-tree) if for any root-to-leaf path $(r = v_0, v_1, \ldots, v_\ell)$, $\quad w(v_i, v_{i+1}) \le \alpha \cdot w(v_{i-1}, v_i)$.

The metric induced by α-trees is obviously a tree metric. It is a strict generalization of Bartal's HST [2], since we do not require that all edges going out from a node have the same weight. As such, they are interesting metric spaces, and they have played an important role in probabilistic offline embedding of metric spaces into Euclidean spaces. α-trees are mostly interesting when α is bounded away from 1, e.g., $\alpha = 1/2$. In this case, it is easy to see that the metric induced by an α-tree is close to an HST. Namely, can be embedded into HST with distortion $\frac{\alpha}{1-\alpha}$. It worth also mentioning that an ultrametric is at most α^{-1} far from an α-HST. This does not imply, however, that the online embedding of an α-tree is, up to constants, as easy as that of α-HST's. We find this class of metrics interesting in its own right, while its online embedding is considerably more complicated than that of ultrametrics.

The 2-Paths game. Let $P_a = [0,1] \times \{a\}$ and $P_b = [0,1] \times \{b\}$ equipped with with the natural line metric d on each. We define the following metric space $(P_a \cup P_b, \mu)$ to be: for every x, y on the same path $\mu(x, y) = |x - y|$. For $x \in P_a, y \in P_b$ $\mu(x, y) = 1$.

The goal of the game is to *deterministically online* map $(P_a \cup P_b, \mu)$ into the unit square $S = [-\frac{1}{2}, \frac{1}{2}]^2 \subseteq \mathbb{R}^2$ with small metric distortion with respect to μ. Namely, q adversarially labeled points in $[0,1] \times \{a, b\}$ are exposed, one at a time. In turn, each point is embedded into S. The crux is that the endpoints of P_a and P_b are restricted to be mapped as follows: $(0, a)$ and $(1, a)$ to $(-\frac{1}{2}, 0)$ and $(\frac{1}{2}, 0)$, respectively, and $(b, 0)$ and $(b, 1)$ to $(0, -\frac{1}{2})$ and $(0, \frac{1}{2})$, respectively, see Figure 1.

3 Our Results

The 2-Paths game. It is easy to see that any q points in $(P_a \cup P_b, \mu)$ can be embedded *offline* into the unit planar box $S = [-1/2, 1/2]^2$ with the restrictions imposed on the endpoints and distortion $O(q)$. We show,

▶ **Theorem 1.** *There exists an* **online** *procedure that can embed any q points in the 2-path game, with distortion $O(q^2)$, when q is specified in advance. If q is not known, a distortion $O(n^2 \log^{2+\epsilon} n)$ is achieved at each step $n = 1, \ldots,$ for any $\epsilon < 1$.*

Solid graph metrics. We use the results for the 2-path game in order to online embed more complex structures into the plane with relatively low distortion. The idea is that the 2-path online algorithm allows us to "cross" lines. The metrics in this section are of the form (X_G, μ), where G is a labeled (weighted / unweighted) graph that is known in advance. The embedding algorithms have the following 2-phase form:

(a) A prepossessing phase: this is an offline phase which gets as input G.
This phase creates a high-level mapping, using empty positions for the "1-intersection" gadget. The image of every point of the metric that is not inside some "1-intersection" territory will be determined in advance in this phase.

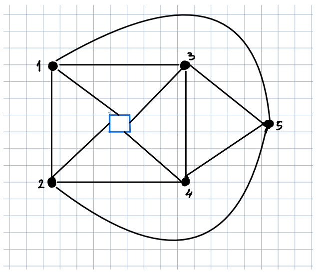

Figure 2 The embedding of the metric in solid K_5. The blue rectangular zone is a 1-intersection territory in which the relevant points on the edge $(2, 3)$ and the edge $(1, 4)$ are mapped, using the 2-paths game. Other points embed linearly along the edges that are pre-embedded in advance.

(b) The online phase: for a next exposed point we check whether it is inside some "1-intersection" territory. If so we map the point according to the "1-intersection" algorithm. Otherwise its image has already been determined in the previous phase.

As a simple example consider the metric that is induced by a solid unweighted K_5. Since K_5 is not planar, it can be easily verified that any *offline* embedding of uniformly spaced q points inside its edges has $\Omega(q)$ distortion. A straight forward implementation of the strategy above results in an *online* embedding of *any* q points with distortion $O(q^2)$. The reader is referred to Figure 2 and the explanation below it.

Using the strategy outlined above (and a generalization of the 2-paths game to k paths), we obtain the following result for *unit-weighted* graphs. Let G be *unit* weighted graph that is known in advance, and (X_G, μ) the corresponding solid metric space.

▶ **Theorem 2.** *There exists an online algorithm that can embed q points from X_G into $\mathbb{R}^2$ with distortion $O(mD \cdot q^2)$, where G is unit-weighted with diameter D, and $|E(G)| = m$.*

We note that the diameter of the input metric is bounded by $|V(G)|$, but the aspect ratio of X_G is unbounded. For a worst case of $D = \theta(n)$, a possibly cubic dependence in $n = |V(G)|$ and quadratic in q is asserted. To improve this dependence we use a more complex algorithm that decomposes G into low conductance pieces. We get,

▶ **Theorem 3.** *Let G_n be a unit-weighted graph that is known in advance. There exists an online embedder that can embed q points from X_G into $\mathbb{R}^2$ with distortion $O(mD^{1/2} \log n \cdot q^2 + nD)$.*

Theorems 2, and 3, are stated for the case that the number of points, q, is known in advance. For q not known in advance, we could still prove the same results with an additional $\log q$ factor.

For weighted graphs we prove the following.

▶ **Theorem 4.** *Let G be a* **weighted** *graph with arbitrary weights, known in advance. There exists an embedder that can online embed q points from X_G and with distortion $O(n^3 \cdot q^2)$, where $n = |V(G)|$.*

The idea in this case, following the strategy above, is to decompose G into low diameter pieces rather than low-conductance. We note that all results above are independent of the aspect ratio.

Sub Classes of Tree Metrics

Ultrametrics. We show that q points taken from an ultrametric can be embedded online in low dimensional Euclidean spaces, with relatively low distortion.

▶ **Theorem 5.** *Any q points from (X, d) that is UM can be online embedded into Euclidean $\mathbb{R}^k$ with distortion $O(k \cdot q^{2/k})$. Here we assume that q (or a bound on it) is known in advance.*

In particular, UM can be embedded with logarithmic distortion in logarithmic dimensional space, and with distortion $O(q^2)$, $O(q)$ into the line, plane, respectively. We comment in Remark 21 in Appendix C that for $k = O(\log q)$ the leading k factor in the theorem can be eliminated, resulting in a *constant* distortion. We also note that to achieve a constant distortion, logarithmic dimension is needed even in an offline embedding.

The idea is to use condition (b) in the definition of UM (Section 2), and show that a UM can be clusterable online in such a way that an online embedding is made possible. For this we prove a more general result for appropriately clusterable metric spaces in Theorem 11

α-Trees. We prove that α-trees can be offline embedded into low dimensional Euclidian spaces with relative low distortion. This holds for any $\alpha < 1$.

▶ **Theorem 6.** *Let (X, d) be the metric that is induced by an α-tree, with $\alpha < 1$. Assume that the root is exposed first. Then, (X, d) can be online embedded into Euclidean $\mathbb{R}^k$ with distortion $O(\frac{k \cdot q^{2/k} \log q}{\log(1/\alpha)})$. Here we assume that q is known in advance.*

The proof in this case is conceptually more complex. We first show that, if α that is relatively small w.r.t to $1/q$, and the tree is exposed in a "natural" order in which descendants are exposed after their predecessors, a small distortion online embedding exists. We then show how to deal with larger α's and adversarial order.

4 Proofs

We present here some of the proofs and ideas. Some of the proofs appear in the Appendix, but due to space limitations, some will appear in the final (or archive) paper.

4.1 The 2-Paths game

We describe here the online embedding procedure assuming that q is known. At the end of the proof we comment on the changes that need to be done if q is not known in advance.

Recall that we have two unit length paths P_a, P_b. Our goal is to embed any q points in $P_a \cup P_b$ into $[-\frac{1}{2}, \frac{1}{2}]^2$ with endpoints as shown in Figure 3 left part. Instead, we will embed $P_a \cup P_b$ in the shifted scaled square $[-2,\ 4q+2] \times [-2,\ 4q+2]$, where $\phi((0,a)) = [-2, 0)$, $\phi((1,a)) = (4q+2, 0)$ and $\phi((0,b)) = (0,-2)$, $\phi((1,b)) = (0, 4q+2)$ (see Figure 3 right part). The original problem can be easily reduced to this one with the same distortion. We divide each line P_a, P_b into $q+1$ closed subintervals of length $\frac{1}{q+1}$, $I_i = [\frac{i-1}{q+1}, \frac{i}{q+1}] \times \{c\}$, $i = 1, \ldots q+1$. We refer to such subintervals as "segments" and denote I_i^c the subinterval I_i of the c-"colored" path. We will pre-embed the boundary points of all segments linearly between the fixed end-points of P_a, P_b, even if they are not queried by the challenger (that is, the adversary). Furthermore, before any query is made we will commit and embed the whole first segment I_1^a as straight line between its endpoints, and I_1^b along the path starting at $(0, -2)$ going straight to $(4, -2)$, then going up to $(4, 2)$, and finally to the intended endpoint $(0, 2)$, see Figure 4 left part. With this in mind, every time a point $x \in I_i \subseteq P_c$, $c \in \{a, b\}$ is

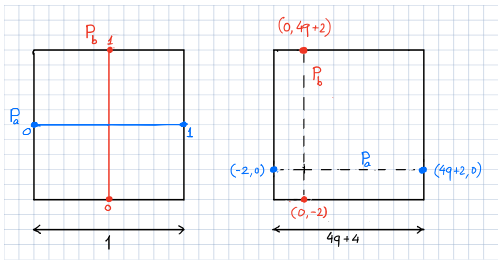

Figure 3 Left figure: the horizontal path P_a and vertical path P_b should be embedded in the grid with endpoints as shown. Right figure: the paths in the scaled shifted square.

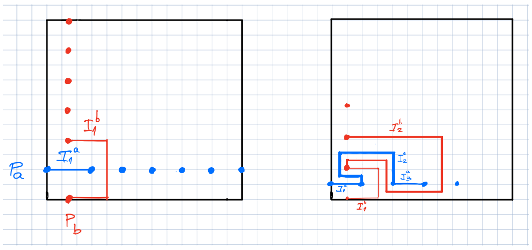

Figure 4 Left part: the initial embedding; The dots are the end points of the segments, and are embedded in advance even if not queried. The initial embedding starts with embedding of I_1^a, I_1^b as shown. Right part: An example of an initial query sequence I_3^a, I_2^a, I_2^b : At the beginning, after I_1^a, I_1^b are embedded, all intervals are open except I_1^a that is closed, and I_2^a that is active. After the embedding of I_3^a (making it closed), and the embedding of I_2^a, the segment I_2^b becomes active. Then, the embedding of I_2^b makes I_4^a active.

queried, we will embed *the entire segment* I_i^c as a continuous curve of vertical and horizontal line segments, and embed x linearly along this curve. The whole point is to determine how the image $\phi(I_i)$ is embedded. This embedding will have the following properties.

1. The image of $\phi(I_i^a)$, $i = 1 \dots q+1$ is a continuous polyline along the grid lines. It will be composed of at most 6 horizontal and vertical lines segments. Its *end points* are the predefined end points of I_i^a in the shifted scaled grid, $\phi(\frac{i-1}{q+1}, a) = (4(i-1)-2, 0)$ and $\phi(\frac{i}{q+1}, a) = (4i-2, 0)$. We note that this does not mean that the entire image of I_i^a will be on that horizontal line. Similarly, the *the end points* of I_i^b are at $(0, 4(i-1)-2), (0, 4i-2)$ respectively.
2. The images of pairwise disjoint segments do not intersect. The image of a segment never intersects itself.

At any step $i = 1, \dots, q$ several non crossing segments of P_a, P_b have already been embedded in the square. In what follows we define how to embed the interval I in which the next point x is in.

▶ **Definition 7.** *At step i, a segment I_j^c is denoted: **Closed** if $\phi(I_j)$ is already embedded. **Active** if the embedding $\phi(I')$ of a segment I' of different color passes through the straight line between the end-points of $\phi(I_j)$, and **Open** otherwise.*

Initially, I_1^a and I_1^b are closed, and I_2^a is active due to the fact that the image $\phi(I_1^b)$ passes through its endpoints (Figure 4 left part). All other segments are open.

Formal online algorithm and proof. We will keep the following invariant.

▶ **Invariant 4.1.** *At any step of the algorithm, there is exactly one active interval. Additionally the active interval alternate between P_a and P_b.*

Following is the formal definition of the algorithm:

Algorithm. On input query $x \in I_i \subseteq P_c$ do:

1. If I_i is closed, then $\phi(x)$ is already defined.
2. If I_i is open, then map the interval to the straight line between its endpoints.
3. Otherwise I_i is active. We assume in what follows that $c = a$. The case $c = b$ is similar. Let I_j be the last previously active interval, which by the invariant above $I_j \subset P_b$, and let I_k^b be the first open segment of P_b. Let $s_i = 4i - 2$, namely, $(s_i, 0)$ is the starting end-point of I_i. Let $s_j = 4j - 2$, $s_k = 4k - 2$, that is $(0, s_j)$ and $(0, s_k)$ are the starting point of I_j, I_k respectively.
   Then I_i^a is linearly mapped into the following sequences of line segments, $(s_i, 0) \to (s_i + 1, 0) \to (s_i + 1, s_j + 3) \to (-1, s_j + 3) \to (-1, s_k + 1) \to (s_i + 4, s_k + 1) \to (s_i + 4, 0)$. See Figure 5 ($I_i^a$ is embedded into the black doted line). The case when $c = b$ is symmetric.

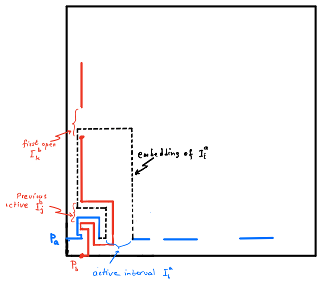

Figure 5 The embedding of an active interval I_i^a is depicted as s dashed line.

Consider, e.g., the following sequence of queried intervals: I_3^a, I_2^a, I_2^b. An embedding of it is shown in Figure 4 right part. The idea is that in every step $i \le q$ there will be a suitable open interval (not yet embedded) allowing the current active interval to "go through".

▶ **Lemma 8.** *Invariant 4.1 is maintained after every step of the algorithm as long as at most q queries are made.*

Proof. The proof is formally by induction on the number of actual queries ℓ that are made (for q fixed and known in advance). The base case is obvious.

Suppose that the lemma holds, and that $x_\ell \in I_i$ as described in the algorithm above, for $\ell \le q$.

The only part of the algorithm that can fail is the 3rd case where it needs to find an open interval I_k. Assume the active interval is a a-interval. Since at the beginning there are q open intervals of P_b, and each query can close only one interval, there is $I_k \subseteq P_b$ that is open. Similarly when the active interval is a b-interval, there are $q - 1$ open intervals of P_a, and at least one query was made to a in order to switch the active interval, so there is an open interval $I_k \subseteq P_b$.

Assume that $\ell \le q$. Then, the active interval changes only in the case when the query point x is in the current active interval, namely, when the 3rd case in the algorithm happens. Assume w.l.o.g, as in the description in the algorithm, that $I_i \in P_a$ is this active interval. In this case I_i becomes closed, $I_k \subseteq P_b$ becomes active, which proves the invariant. The proof for $I_i \subseteq P_b$ is similar. ◀

▶ **Lemma 9.** *At any step ϕ is injective. Namely, for every pair of embedded points $x \neq y$, $\phi(x) \neq \phi(y)$.* ◀

Proof of Theorem 1. By Lemma 9 and the fact that the grid is integral, the minimum distance between $x \in P_a$ and $y \in P_b$ is 1. Additionally, as each subinterval $I_i \in P_c,\ c \in \{a, b\}$ is stretched to a path of integral length, and the queried points are mapped linearly along the subinterval, there is no contraction in the resulting embedding.

Every interval $I_i \subseteq P_c$ is embedded into a union of at most six horizontal or vertical $O(q)$-long straight line segments. Thus the expansion is $max_{x,y} \frac{l(x,y)}{\mu(x,y)} \leq \frac{O(q))}{1/(q+1)} = O(q^2)$, implying a $O(q^2)$ distortion.

If q is not known in advance, instead of splitting P_a, P_b into uniform intervals, we split the paths into infinitely many intervals of decreasing lengths. We define a converging series $f(x) = \frac{c}{xlog^{1+\epsilon/2}x}$, for c such that $\Sigma_1^\infty f(x) = 1$. $f(i)$ will describe the percentage of the path length that the i'th interval contains. The algorithm will remain exactly the same, except for the definition of the 3'rd case. Because there isn't an integral grid, we will translate the addition of constant in the path coordinate, to an addition of a percentage of the interval length (for example $[s_i + 1, s_j + 3]$ is translated into $[s_i + \frac{1}{4}f(i), s_j + \frac{3}{4}f(j)]$).

Lemma 8 and 9 remains the same, but we adapt the claim of 8 to note that after q queries, the active interval is at most I_{q+1}, so $dist(\phi) \leq f(q+1)^{-2} = O(q^2 log^{2+\epsilon} q)$. ◀

4.2 Applications to "solid" graph metrics

The proofs of Theorems 2, 3 and 4 , appear in Appendix B. For this we also introduce and use a generalization of the 2-paths game to k-paths in Appendix A.

4.3 Ultrametrics

Theorem 5 will follow from the more general Theorem 11 below whose proof is in Appendix C, and Observation 23 (Appendix D). We need the following,

▶ **Definition 10** (Clusterable metric space). *A finite metric (X, d) is (γ, ϵ)- clusterable, if there is a non-trivial partition $X = C_1 \cup \ldots \cup C_\ell$, called "clusters", such that:*
1. *For every x, y in different clusters $d(x, y) \geq diam(X)/\gamma$.*
2. *For every cluster C_i, $diam(C_i) \leq \epsilon \cdot diam(X)$.*

A finite metric (X, d) is (γ, ϵ)- recursively-clusterable if it is either (γ, ϵ)-clusterable to singletons, or it is (γ, ϵ)-clusterable and each non-singleton class of it is (γ, ϵ)-recursively-clusterable.

▶ **Theorem 11.** *Let (X, d), $|X| = q$ be (γ, ϵ)-recursively-clusterable, with $\epsilon < \frac{1}{2\gamma\sqrt{k}q^{1/k}}$ then (X, d) can be online embedded into Euclidean $\mathbb{R}^k$ with distortion $O(2\gamma^2\sqrt{k}q^{1/k})$. We assume here that q is known in advance.*

The definition is meaningful when $\epsilon << 1/\gamma$. In general, a UM is not clusterable well enough to directly use the theorem. Observation 23 (Appendix D) bridges this gap.

4.4 α-Tree Metrics - Theorem 6

Let $T = (X, E)$ be an α-tree and d the corresponding metric. Assume that X is exposed in an arbitrary order $x_0 = root$, $x_1, \ldots x_{q-1}$, with the root being first. We denote by $T^{(t)}$ the subtree that is defined by $X^{(t)} = \{x_i,\ i \leq t\}$ and $d^{(t)}$ the metric induced on it. The exposure order is a **natural order** on $V(T)$ if every point v appears before any $u \in T_v \setminus \{v\}$. That is, predecessors appear before their ancestors.

If u is a predecessor of v then $diam(T_v) < 2\alpha \cdot diam(T_u)$. Hence, assuming a small enough α, we could wish to embed d_T online and so to maintain the structure of subtrees. Namely, similarly to the embedding of UM, to embed each subtree T_v rooted at v "near" the embedding of v, using v as a 'fat point'However, we can do this only if the tree is exposed in a *natural order*. In addition, even in this "simple" case, we would need α to be quite small.

To prove the theorem we use three independent steps. Theorem 12 asserts that, in the "simple" case where the exposure is in the *natural* order *and* α small enough, a relatively easy embedding exists. Then we show, in Theorem 13, how to *online* embed an α-tree into another α'-tree where the possibly adversarial exposure order w.r.t the input becomes natural w.r.t the host tree. Lastly, in Theorem 14, we show that for natural order exposure and β arbitrary small, an α-tree can be online embedded into a β-tree with small distortion. Finally we can combine the above to prove Theorem 6. The proofs of Theorems 12, 13 and 14 will appear in the full paper.

▶ **Theorem 12.** *Let (X, d) be the metric that is induced by an α-tree, with $\alpha < \frac{1}{32\sqrt{k}q^{1/k}}$. Assume that $X = \{x_0, \ldots, x_{q-1}\}$ is exposed in the natural order and that q is known in advance. Then, (X, d) can be online embedded into Euclidean $\mathbb{R}^k$ with distortion $O(\sqrt{k}q^{1/k})$.*

▶ **Theorem 13.** *Let (X, d) be the metric that is induced by an α-tree and suppose that root appears first. Then (X, d) can be online embedded into an $\frac{\alpha}{1-\alpha}$-tree $\tilde{T} = (X', E')$ inducing a metric $\tilde{d}$ that dominates d and expand it by at most $\frac{1+\alpha}{1-\alpha}$. Further, the order of exposure w.r.t $\tilde{T}$ is natural.*

▶ **Theorem 14.** *Let T be an α-tree and $\beta < \alpha$. Assume that T is exposed in the natural order, then $d = d_T$ can be online embedded in the natural order into a β-tree $\tilde{T}$ with distortion $\frac{\log 1/\beta}{\beta \cdot \log(\frac{1-\alpha}{\alpha})}$.*

Combining the above theorems, the proof of Theorem 6 follows:

Proof. Let $T = (X, E)$ be an α-tree with root o. Let d_T the corresponding metric. We first use Theorem 13 to online embed X into a tree T_1 that is a $\alpha_1 = \frac{\alpha}{1-\alpha}$-tree rooted at o where the exposure order on T induces a natural order on T_1. Theorem 13 implies that the resulting metric d_1 distorts d_T by $\frac{1+\alpha}{1-\alpha}$.

Next we set $\beta = \frac{1}{32\sqrt{k}q^{1/k}}$ which is the parameter needed for Theorem 12. We use Theorem 14 to online embed d_1 into a β-tree T_2 where the original exposure order is the natural order for T_2. The resulting metric d_2 distorts d_1 by $O(\frac{\sqrt{k}q^{1/k}(\frac{\log q}{k}+\log k)}{\log(\frac{1}{\alpha})}) = O(\frac{q^{1/k}}{\log(1/\alpha)} \cdot (\frac{\log q}{\sqrt{k}} + \sqrt{k}\log k))$.

Finally, we can now online embed d_2 using Theorem 12 into $\mathbb{R}^k$, resulting in a ℓ_2 metric that distorts d_2 by $O(\sqrt{k}q^{1/k})$. Multiplying the 3 terms implies that the total distortion is $O(\frac{q^{2/k}}{\log(1/\alpha)} \cdot (\log q + k \log k))$. ◀

We comment that for simplicity, in the proof of Theorem 12 the construction uses grids with points far apart. Instead, we could use spherical codes. This will omit the k factor from the distortion estimate, for $k = \log q$, but this is less significant in this case, as for $\alpha = \Omega(1)$ (e.g., $\alpha = 1/2$) we would still get a $\log q$ distortion even for logarithmic dimension.

References

1 Mihai Badoiu, Julia Chuzhoy, Piotr Indyk, and Anastasios Sidiropoulos. Embedding ultrametrics into low-dimensional spaces. In Nina Amenta and Otfried Cheong, editors, *Proceedings of the 22nd ACM Symposium on Computational Geometry, Sedona, Arizona, USA, June 5-7, 2006*, pages 187–196. ACM, 2006. doi:10.1145/1137856.1137886.

2 Yair Bartal. On approximating arbitrary metrices by tree metrics. In Jeffrey Scott Vitter, editor, *Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing, Dallas, Texas, USA, May 23-26, 1998*, pages 161–168. ACM, 1998. doi:10.1145/276698.276725.

3 Yair Bartal, Nova Fandina, and Seeun William Umboh. Online probabilistic metric embedding: A general framework for bypassing inherent bounds. In Shuchi Chawla, editor, *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA 2020, Salt Lake City, UT, USA, January 5-8, 2020*, pages 1538–1557. SIAM, 2020. doi:10.1137/1.9781611975994.95.

4 Yair Bartal, Nathan Linial, Manor Mendel, and Assaf Naor. Low dimensional embeddings of ultrametrics. *Eur. J. Comb.*, 25(1):87–92, 2004. doi:10.1016/J.EJC.2003.08.003.

5 Sujoy Bhore, Arnold Filtser, and Csaba D. Tóth. Online duet between metric embeddings and minimum-weight perfect matchings. In David P. Woodruff, editor, *Proceedings of the 2024 ACM-SIAM Symposium on Discrete Algorithms, SODA 2024, Alexandria, VA, USA, January 7-10, 2024*, pages 4564–4579. SIAM, 2024. doi:10.1137/1.9781611977912.162.

6 Jean Bourgain. On lipschitz embedding of finite metric spaces in hilbert space. *Israel Journal of Math*, 52(1-2):46–45, 1984.

7 Piotr Indyk, Avner Magen, Anastasios Sidiropoulos, and Anastasios Zouzias. Online embeddings. In Maria J. Serna, Ronen Shaltiel, Klaus Jansen, and José D. P. Rolim, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, 13th International Workshop, APPROX 2010, and 14th International Workshop, RANDOM 2010, Barcelona, Spain, September 1-3, 2010. Proceedings*, volume 6302 of *Lecture Notes in Computer Science*, pages 246–259. Springer, 2010. doi:10.1007/978-3-642-15369-3_19.

8 William B. Johnson and Joram Lindenstrauss. Extensions of 1-lipschitz mappings into a hilbert space. In *Conference in modern analysis and probability*, volume 26 of *Contemporary Mathematics*, pages 189–206, 1984.

9 Ilan Newman and Yuri Rabinovich. A lower bound on the distortion of embedding planar metrics into euclidean space. In Ferran Hurtado, Vera Sacristán, Chandrajit Bajaj, and Subhash Suri, editors, *Proceedings of the 18th Annual Symposium on Computational Geometry, Barcelona, Spain, June 5-7, 2002*, pages 94–96. ACM, 2002. doi:10.1145/513400.513412.

10 Ilan Newman and Yuri Rabinovich. Online embedding of metrics. In Susanne Albers, editor, *17th Scandinavian Symposium and Workshops on Algorithm Theory, SWAT 2020, June 22-24, 2020, Tórshavn, Faroe Islands*, volume 162 of *LIPIcs*, pages 32:1–32:13. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2020. doi:10.4230/LIPIcs.SWAT.2020.32.

11 Anastasios Sidiropoulos, Mihai Badoiu, Kedar Dhamdhere, Anupam Gupta, Piotr Indyk, Yuri Rabinovich, Harald Räcke, and R. Ravi. Approximation algorithms for low-distortion embeddings into low-dimensional spaces. *SIAM J. Discret. Math.*, 33(1):454–473, 2019. doi:10.1137/17M1113527.

A k-paths intersection

We define below a generalization of the 2-intersection that we call "k-paths-intersection-gadget". This will allow us to use the algorithm for that problem as a black box in online embedding of solid graph metrics.

To do the above we note that in the 2-paths game, if we embed two lines of length l (rather than 1), of pairwise distance d (instead of 1), into a $s \times s$ grid (rather then $\theta(q) \times \theta(q)$), Then, by scaling, we get $O(\frac{qs}{l})$ expansion and $O(\frac{dq}{s})$ contraction.

Let $X_k = [0, l] \times [k]$ be a disjoint union of k paths of length l, each equipped with the line metric ℓ. The metric (X_k, μ) is defined by: for each $x, y \in [0, l] \times [i],\ \mu(x, y) = \ell(x, y)$ and for $x \in [0, l] \times [i],\ y \in [0, l] \times [j],\ i \neq j,\ \ \mu(x, y) = d$. Namely, the distance between points on different paths is d. The goal is to online embed q points from X_k into the square $[0, s]^2$ where the ends points are fixed in advance in an arbitrary way that is presented to the embedding algorithm. For our use we assume that $q >> k, l, d$.

Formally, we assume that each side (edge) of the square contains $k+2$ uniformly spaced points (including the corners). We refer to the set T of the $4k$ *internal* points on the sides (that is, excluding the corners) as "Terminals". The *input* will be an injective mapping $t : \cup_{i=1}^{k}\{(0,i),(l,i)\} \mapsto T$ that specifies for each $i \in [k]$, the terminals that are the images of the endpoints of the ith path. This assignment is known to the algorithm in advance. The embedder must map the endpoints of each path, $[0,i]$ and $[l,i]$ to the corresponding terminals $t(*)$.

We refer to this embedding problem as the "k-paths intersection" with parameters l,d, that is, the length of the paths and pairwise distance between paths, and s which is the target box size in $\mathbb{R}^2$. We assume (for our applications) that $d,l \geq 1$, although the whole procedure is invariant under scaling.

▶ **Theorem 15.** *There is an online procedure that* **online** *embeds q points of X_k into $\mathbb{R}^2$ with expansion $O(\frac{sq}{l})$ and contraction $O(\frac{dkq}{s})$, implying a total distortion of $O(\frac{dkq^2}{l})$. The number of points, q, is known in advance.*

Proof. In order to define the online embedding we first define a scaled grid $[0,3(3k+1)]^2$ and define the embedding therein. Later, in order to embed in a $[s]^2$ box we just scale the grid by $a_1 = \frac{s}{3(3k+1)}$ which will scale the expansion by a_1 and scale down the contraction by a_1 resulting in the same overall distortion. We do not optimize the result here as this gadget is mostly needed for our applications, and is less interesting on its own.

The terminals T will be the points $\cup_{i=1}^{k}\{(3i,0),(3i,9k+3),(0,3i),(9k+3,3i)\}$. The indices $i \in [k]$ in this definition do not necessarily correspond to the endpoints of the i-path, as the mapping of endpoints is arbitrary.

Given the mapping of the endpoints to terminals, we define now a set of paths $\{\tilde{P}_i\}_1^k$ in the grid corresponding to the intended line metrics. First, for paths for which one terminal is on the vertical edge and the other is on an horizontal edge, e.g., w.l.o.g. $x = (3i,0)$ an y is $(0,3j)$. The path will be $(3i,0) \to (3i,3j) \to (0,3j)$.

Once this is done we declare all rows and all columns that contain a segment that is already assigned as "used". We also declare all rows and all columns that contain an assigned terminal as used. This define at most $2k$ rows and $2k$ columns as used.

Next, we define the intended paths for these path with endpoints both on a vertical edge or a horizontal edge (not necessarily the same). This is done as follows: for each remaining path in an arbitrary order do:

Let x,y be the assigned terminals of some path P_ℓ to be currently defined.

- If x and y are on the same bottom edge of the square, i.e., x is $(3i,0)$ and y is $(3j,0)$, for some i,j. Let $r \in [k]$ be such that the $3r$'th row is yet unused (there will be such r). The path will be $(3i,0) \to (3i,3r) \to (3j,3r) \to (3j,0)$, and the $3r$'th row is now marked as used. A analogous definition is done for x,y on the top edge or on the same vertical edge, using a previously unused column instead of row.
- If x and y are on opposite edges e.g., for $x = (3i,0)$ and $y = (3j,9k+3)$ the path will start as before, going to the first $3r$ unused row, and finally going up from $(3j,3r)$ to $(3j,9k+3)$, see Figure 6.

▷ Claim 16. Let $\tilde{P}_i,\ i = 1,\ldots,k$ be the paths that corresponds to the terminal mapping above. Then, any two path meet in at most two integral points. Further, any point in the grid belongs to at most two paths, and if two path share a point, one passes through it vertically, and the other horizontally.

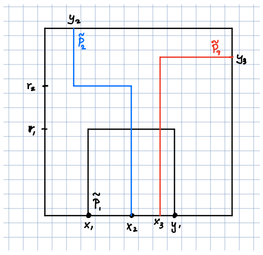

Figure 6 The three types of paths in the grid. The red path with terminal in horizontal and vertical edges is allocated first.

Proof. By definition, each path is composed of at most three line segments. More over, there are $3k$ rows and $3k$ columns of type $3i$, $i \in [k]$. After the first stage, at most $2k$ of these rows, and at most $2k$ of these columns are declared used. Thus leaving at least k unused rows, as needed for the at most k paths whose endpoints lie on the horizontal edges, and with the same argument for columns.

In addition at most one path travel at any given row or at any given column of the grid. Hence paths can meet only if one is horizontal and the other is vertical. Further, this can happen only at integral points. ◁

Now the online embedder attempts to use $\tilde{P}_i$ for P_i where points are mapped linearly along$\tilde{P}_i$.

At each intersection (these are known in advanced, as the paths $\tilde{P}_{i=1}^k$ are defined beforehand), we use "1-intersection" gadget to resolve the conflict, with a square of size 2 centered at the intersection integral point. Note that no two 1-intersection gadgets intersect in the grid.

Distortion analysis. *Expansion:* Since every $\tilde{P}$ is of length at most $18k + 6$ and P that is mapped to $\tilde{P}$ embeds linearly in $\tilde{P}$ then the expansion between points on $\tilde{P}$ is bounded by $O(k/l)$. Moreover, the embedding of points in a path P may deviate from its mapped line $\tilde{P}$ only in 1-intersection gadgets. Each such expansion adds additively to the total expansion. Further, for two points on the same path that are in an intersection gadget, their distance $l' = \Omega(l/k)$ could increase by a factor $O(\frac{q}{l'}) = O(\frac{qk}{l})$ (see last par. in Section 4.1). This results in a total $O(\frac{qk}{l})$ expansion inside a path. For two points not on the same path the original length is d while the resulting length is $O(k)$ implying an $O(k/d)$ expansion for such points. Hence the overall expansion is $O(\frac{qk}{l})$. Scaling this embedding to a $[s]^2$ box (by multiplying by a_1, see first paragraph in the proof) results in $O(\frac{qs}{l})$ expansion.

Contraction: For two points on the the same line P the contraction is determined by the 1-intersection regions. For such two points the contraction is $O(dq)$ (again see last par. in Section 4.1). Otherwise, for two points of which at least one is not in a 1-intersection gadget, the distance may become $O(1)$ while the initial distance is d. Thus the total contraction is bounded by $O(dq)$, and once scaled down by a_1 it becomes $O(\frac{dqk}{s})$, as claimed in the Theorem. ◀

B Proof of Theorems 2, 3 and 4

Here we consider the metric space X_G for connected *unit* weighted graph G. The graph G is known in advance.

Proof of Theorem 2. In order to online map q labeled points from X_G into the plane, we first embed *offline, part* of G as follows. We subdivide each edge $(u,v) \in E$ into a 3-path, (u,u',v',v) where $distance(u,u') = distance(v,v') = 1/4$ and $distance(u',v') = 1/2$. We will do a preprocessing where we commit to an embedding of all edges (u,u') that are embedded as straight lines, as follows.

Let $\{v_i\}_{i=1}^n$ be a fixed enumeration of the vertices, deg_i the degree of v_i, $x_i = \Sigma_{j=1}^{i-1} deg_j,\ i = 1, \ldots$ "n, and $x_1 = 0$. We first embed $V(G)$ by $\phi(v_i) = (x_i, 0)$. Further, for each $v_i \in V$ we use an arbitrary enumeration of the edges $\{(v,u) \in E\} = (e_1, \ldots, e_{deg_i})$), and embed the first 1/4-length part of $e_j = (v_i, v'_j)$ as a straight vertical line with end point $\phi(v'_j) = (x_i + j - 1, 1), j = 1, \ldots deg_i$.

Now, in the *online* part, any point in (u,u'') is already determined, and for the $(u'v')$ edges, we use a "m-intersection" gadget of size m to connect the endpoints of all edges, where all $\phi(v')$ are below the box of the m-intersection. The parameters of the "m-intersection" are $s = m,\ d = O(D),\ l = 0.5$, because the distance between two point is at most $2D$, and the part of each edge $(u'v')$ in the "m-intersection" gadget is 0.5. The result in Section A implies that the distortion is $O(mD \cdot q^2)$. ◀

We note that this embedding is oblivious to the structure of the graph which is assumed to be worst case. Namely, if we knew e.g., that G_n can be embedded in the planar grid, so that no two non-adjacent edges intersect, then we could do much better.

Theorem 3 is an improvement of Theorem 2, for unit weight graphs, w.r.t the dependence on the diamemeter D (that is always at most n in this case).

The idea of the proof of Theorem 3 is to use a decomposition into small diameter components by deleting a sublinear amount of edges. As a result, the embedding of points inside each components, using Theorem 2, will improve, and at the same time the k-intersection gadget for interconnecting the components will be smaller.

We stress that we do not assume any special structure of the given graph.

We leave further details to the full paper publication.

Proof of Theorem 4. In the case of a weighted graph, we first normalize the weights such that the *minimum* weight is 1. Define D as the diameter of the graph after normalization. Then the exact same algorithm as in the proof of Theorem 2 will yield a distortion of $O(mD \cdot q^2)$. Here D becomes the aspect ratio of (X_G, d) and can be very large. We design below an algorithm which eliminates the dependence on aspect ratio.

We assume that a weighted graph $G = (V,E)$ with weights $d : E \mapsto \mathbb{R}^+$ forms a metric space, namely where $d(u,v) = distance(u,v)$ is given in advance. Recall that a labeled point-query will be any $x \in (u,v)$ and $\alpha \in [0,1]$, namely such that $d(x,u) = \alpha d(u,v)$ and $d(x,v) = (1-\alpha)d(u,v)$.

Our online embedding is based on an *offline* preprocessing phase of $G = G_n$ that is known in advance. The grand goal of this preprocessing is to cluster G into small diameter (w.r.t that of G) clusters. Then we allocate space in the plane for the online embedding for each cluster, and a disjoint space for the interconnection between clusters. This clustering is going to be recursive. More specifically, on each recursion level we decopose the current cluster H of diameter D' (initially G itself) into a collection of connected subgraphs (subclusters) $\{C_i\}$, each of diameter smaller than $D'/2$. Then we proceed recursively on each subcluster. While doing this clustering, an additional output is an allocation of a box in the plane $B(C_i)$ for each C_i inside the box $B(H)$ in which C_i is going to be embedded (if a point in it is to be asked), and in addition, the allocation of a box in $B(H)$ in which a A-intersection gadget will be implemented in order to embed the (solid) edges connecting these clusters.

We leave further details to the final paper. ◀

C Proof of Theorem 5 and 11

▶ **Definition 17** (Shifted grid). *Let $G_{k,q} = (\{0, 1, \ldots, q^{1/k} - 1\})^k$, with a fixed arbitrary injection*
$G : \{0, \ldots, q-1\} \to G_{k,q}$*, for which* $G(0) = (0, \ldots 0)$.

For a scale s and a vector $v \in \mathbb{R}^k$ the shifted image $\phi_{v,s} : \{0, \ldots, n-1\} \to \mathbb{R}^k$ is defined $i \mapsto \ v + s \cdot G(i)$.

The idea behind the proof of Theorem 11 is simple. The parameters are chosen so that the metric induced on points between distinct subclusters of any class (cluster) C is closed to the uniform metric. Hence we will embed each cluster C, namely the at most q points in it, in a "fat" point using $\phi_{v,s}$ as the embedding, where v is the place we embed the first point in C and each point x_t that is in a different subcluster in C, at $\phi_{v,s}(t) = v + s \cdot G(t)$ for s large enough and proportional to $diam(C)$.

This will expand the distances between points in C that are of distance large enough by approximately $s\sqrt{k}q^{1/k}$ and with no contraction. Points in subclusters of C will be embedded "inside" the corresponding "fat" points, as by the clustering properties, the diameter of a subcluster is very small w.r.t $diam(C)$. Further, by our choice of parameters, there will be enough place for each point of every subcluster in the "fat" point representing the cluster, with minimal "interaction" between subclusters.

To follow the strategy above, we need to show that the cluster-structure is well defined online, during the partial exposure of points. This is the essence of Observation 19 below.

We assume that the points of X are adversarially exposed in order $x_0, x_1, \ldots, x_{q-1}$ that is not known in advance. We assume however, that d is (γ, ϵ)-recursively-clusterable and that n, γ and ϵ are known in advance. We do not assume that we know any other fact about the actual structure of the recursive (γ, ϵ)-partition of X. In particular, we do not assume a prior knowledge on the number of classes at any recursion level in the recursive decomposition of X.

▶ **Notations 18.** *Let (X, d) be (γ, ϵ)-recursively clusterable, and $X = \{x_0, \ldots x_{t-1}\}$ enumerated according to the exposure order. The superscript (t) is used to define properties / quantities of the exposed metric and of the embedding at time t. Thus e.g., for the exposed sequence $x_0, x_1, \ldots \ldots$ we denote $X^{(t)} = \{x_0, \ldots, x_t\}$.*

For $x, y \in X$ we denote $C(x, y)$ the smallest cluster that contains both x and y. We denote $C^{(t)}(x, y) = C(x, y) \cap X^{(t)}$, namely the cluster containing x, y at time t. We do not necessarily know $C(x, y)$ at time $t < q$, but we will show that for $x = x_j, y = x_\ell \ , j, \ell \le t$, namely, that are exposed at t or before, we do know $C^{(t)}(x, y)$.

▶ **Observation 19.** *For (X, d) as above, let $x = x_t$ and $y \in X^{(t-1)}$ for which $d(x, y)$ is the smallest. Let $C^{(t)}(x, y)$ be the smallest cluster that contains both x, y at time t, and let $C_\gamma(x, y) = \{z \in X^{(t)} |\ d(z, x) \le \gamma d(x, y)\}$. Then $C^{(t)}(x, y) = C_\gamma(x, y)$.*

The proof of Observation 19 is omitted and will appear in the full paper.

Proof of Theorem 11 . We define next the online embedding $\phi : X \longrightarrow \mathbb{R}^k$, and denote the induced metric by $\tilde{d}$.

Let $N = 2\gamma$.

1. $\phi(x_0) = \bar{0} = (0, \ldots, 0)$.
2. At time t embed $x = x_t$ as follows.
   Let $y \in X^{(t-1)}$ be the closest to x_t, and $C^{(t)}(x, y) = C^{(t)}_\gamma(x, y) = \{z \in X^{(t)},\ d(z, x) \le \gamma d(y, x)\}$.

Let $y_1, y_2 \in C^{(t)}(x,y)$ be the first and second exposed points in $C^{(t)}(y,x)$, for which $d(y_1,y_2) \geq d(y,x)$. Note that $C^{(t)}(x,y)$ contains x, y and hence $y_1 \neq y_2$ are well defined. We set $L = L(C(x,y)) = d(y_1,y_2)$ and set $\phi(x) = \phi(y_1) + NL \cdot G(t)$.

The theorem now follows from Lemma 20, items 3, and item 2, applied for $C^{(q)} = X$. ◀

▶ **Lemma 20.**

1. *For any cluster C, $\tilde{diam}(C) \leq N\sqrt{k}q^{1/k} diam(C)$.*
2. *$\tilde{d} \leq 2\gamma^2\sqrt{k}q^{1/k} \cdot d$.*
3. *$\tilde{d}$ is not contracting.*

Proof. For the first item consider a class $C \subseteq X$ with its first two exposed points y_1, y_2, and $L = d(y_1,y_2)$. Since all points in C are embedded by a shifted embedding at $\phi(y_1) + v + N \cdot L \cdot G(t)$ for positive vectors v and corresponding t, it immediately follows that $\tilde{diam}(C) \leq N \cdot L \cdot \sqrt{k}(q^{1/k} - 1) + \tilde{d}(C')$ where C' is the subcluster in C achieving the maximum embedded diameter. The first item above corresponds to the furthest distance between two points in the shifted $G_{k,q}$ with scale $N \cdot L$.

The same argument can be made for C' and corresponding scale $L' \leq diam(C') \leq \epsilon L$. Thus plugging this for C' in the above, and using induction we get.

$$\tilde{diam}(C) \leq NL\sqrt{k}(q^{1/k} - 1) + N \cdot \epsilon L \cdot diam(C'') \leq \ldots \leq$$

$$NL\sqrt{k}(q^{1/k} - 1)\sum_{i \geq 0} \epsilon^i \leq NL\sqrt{k}(q^{1/k} - 1)\frac{1}{1-\epsilon} \leq NL\sqrt{k}q^{1/k} \leq N\sqrt{k}q^{1/k} diam(C)$$

where the last inequality is since $d(y_1,y_2) \leq diam(C)$.

For the second item let $x, y \in X$ be any two points and $C(x,y)$ the smallest class containing both. Then by definition of clusterability $diam(C(x,y)) \leq \gamma d(x,y)$. Plugging the lower bound on $\tilde{diam}(C(x,y))$ from 1. we get $\tilde{d}(x,y) \leq N\gamma d(x,y)\sqrt{k}q^{1/k} \leq 2\gamma^2\sqrt{k}q^{1/k} d(x,y)$, where the last inequality is by our choice of N.

For the 3rd item let $x = x_t$, and let $y, C^{(t)}(x,y) = C_\gamma(x,y), y_1, y_2 \in C^{(t)}(x,y)$ and $L = d(y_1,y_2)$ as defined in the embedding above.

Consider first any $z = z_j \in C^{(t)}(x,y)$. Let $C(z) \subset C^{(t)}(x,y)$ the cluster of z in the (γ,ϵ)-clustering of $C^{(t)}(x,y)$. Since, by Observation 19, x form its own cluster in that clustering, $x \notin C(z)$. Let $z_0 = x_r$ be the first exposed in $C(z)$. Then, by the definition of the embedding, z_0 is embedded in the shifted grid $G_{\phi(y_1),L}$ namely $\phi(z_0) = \phi(y_1) + NL \cdot G(r)$. Hence it follows that $\tilde{d}(z_0,x) \geq NL$. Further, using item 1. in the lemma, which we already proved, $\tilde{diam}(C(z)) \leq N\sqrt{k}q^{1/k} diam(C(z))$. Recalling that $diam(C(z)) \leq \epsilon \cdot diam(C^{(t)}(x,y))$ and that $d(z,x) \geq diam(C^{(t)}(x,y))/\gamma$, we conclude,

$$\tilde{d}(x,z) \geq \tilde{d}(z_0,x) - \tilde{d}(C(z)) \geq N \cdot L - \epsilon N\sqrt{k}q^{1/k} \cdot diam(C^{(t)}(x,y)) \geq$$

$$N(\ diam(C^{(t)}(x,y))/\gamma - \epsilon\sqrt{k}q^{1/k} \cdot diam(C^{(t)}(x,y))\ ) \geq Nd(x,z))(\frac{1}{\gamma} - \epsilon\sqrt{k}q^{1/k})$$

where the second to last inequality is by the (γ,ϵ)-clusterability of $C^{(t)}(x,y) > d(x,z)$ and the last is by our choice of parameters. This proves that $\tilde{d}$ is not contracting on $C^{(t)}(x,y)$.

For $z \notin C^{t)}(x,y)$, z got separated from y before or at time t. Let $C^{(t)}(z)$ be the largest cluster that contains z and not y, and let $C^{(t-1)}(y,z)$ the smallest cluster that contains both. Let $y_1(z), y_2(z)$ be the first exposed points in $C^{(t-1)}(y,z)$. Then

$$\tilde{d}(z,x) \geq \tilde{d}(y_1(z),y_2(z)) - \tilde{d}(C^{(t)}(z)) - \tilde{d}(C^{(t)}(x,y)) \geq$$

$$Nd(y_1(z),y_2(z)) - N\sqrt{k}q^{1/k} diam(C(z)) - N\sqrt{k}q^{1/k} diam(C^{(t)}(x,y)).$$

where the last inequality is by the definition of embedding and the first item in the Lemma.

Recalling that $d(y_1(z), y_2(z)) \geq \frac{1}{\gamma} diam(C(y,z)) \geq \frac{1}{\gamma} d(z,x)$ and that $diam(C(z))$, $diam(C^{(t)}(x,y)) \leq \epsilon \cdot diam(C(y,z))$, and plugging into the above we get,

$$\tilde{d}(z,x) \geq d(z,x) \cdot N(\frac{1}{\gamma} - 2\epsilon\sqrt{k}q^{1/k}) \geq d(z,x)$$

by our choice of parameters. ◀

▶ Remark 21. In the above, we embedded the near uniform metric in $\mathbb{R}^k$ by using an arbitrary non-contracting mapping into the (shifted) grid $[q^{1/k}]^k$. This has the property that it can be done online, while expanding the uniform metric by at most $\sqrt{k} \cdot q^{1/k}$. This is asymptotically optimal for $k = o(\log q)$. But for $k = \theta(\log q)$ a better embedding can be used: One can use a fixed spherical code, e.g., as guaranteed by Johnson-Lindenstrauss [8] for embedding K_n in $O(\log q)$-dim space with distortion $1 + \epsilon$ (for arbitrary small fixed ϵ). Using such a (shifted) embedding instead of shifted grid will result in a distortion of $\gamma^2 q^{1/k}$ for $\epsilon < \frac{1}{2\gamma q^{1/k}}$ in Theorem 11. However, this may be applied only for $k = \Omega(\log q)$ for which spherical codes with constant distortion exist. For such construction, an online embedding of any clusterable metric as above into $\mathbb{R}^{O(\log q)}$, with $O(1)$-distortion, is achieved. See also Remark 24.

D Online embedding of Ultrametrics - proofs

We use Theorem 11 to show that ultrametric can be embedded into $\mathbb{R}^k$ with distortion $O(\sqrt{k}q^{2/k})$, and in particular, with distortion $O(q^2)$ into the line and distortion $O(q)$ in $\mathbb{R}^2$. We note that a UM d is always $(1, \epsilon)$-clusterable for some ϵ, but ϵ depends on the metric and it is not necessarily made known in the online process. We will first embed a UM d into another UM d' to remedy the above obstacle.

▶ **Definition 22.** *Let (X, d) be a UM, and $\alpha > 1$. Let $(X, d_{(\alpha)})$ be defined by $d_{(\alpha)}(x, y) = \alpha^i$ for $i \in \mathbb{Z}$ the smallest such that $d(x, y) \leq \alpha^i$.*

▶ **Observation 23.** *Let (X, d) be a UM and $(X, d_{(\alpha)})$ as defined above. Then,*
1. *$d_{(\alpha)}$ is a UM.*
2. *$d \leq d_{(\alpha)} \leq \alpha \cdot d$.*
3. *$d_{(\alpha)}$ can be computed online while d is exposed.*
4. *$d_{(\alpha)}$ is $(1, \frac{1}{\alpha})$-recursively clusterable.*

Proof. Consider the representation of (X, d) by a weighted graph G, as asserted by (b) in the definition of UM (Section 2), and change (increase) the weight of each $e \in E(G)$ to the smallest α-power, α^i, such that $w(e) \leq \alpha^i$. This results in a weighted graph $G_{(\alpha)}$ that by definition defines a UM that is exactly $d_{(\alpha)}$. Hence items (1), (2) in the claim directly follow. For (3) note that exposing x with $d(x, y)$ for every previously exposed y defines $d_{(\alpha)}(x, y)$.

For the last item, it is enough to show that $d_{(\alpha)}$ is $(1, \frac{1}{\alpha})$-clusterable, as the metric induced on subclasses is also UM, and thus recursive clustering will automatically follow. Indeed, let $X = C_1 \cup \ldots C_\ell$ be the partition that is defined by the connected components (classes) of G_α after deleting all edges of maximum weight α^a. Then, then by definition, $d_{(\alpha)}(x, y) = \alpha^a$ for every x, y belonging to different components proving that $\gamma = 1$, and $d_{(\alpha)}(x, y) \leq \alpha^{a-1}$ for every component C_k and $x, y \in C_k$, showing that $\epsilon \leq \frac{1}{\alpha}$. ◀

Proof of Theorem 5. For a UM (X, d) with $|X| = q$, and fixed k, set $\alpha = 2\sqrt{k}q^{1/k}$. Then $d \leq d_{(\alpha)} \leq \alpha \cdot d$. Thus $d_{(\alpha)}$ distorts d by at most α. Further $d_{(\alpha)}$ is UM and by Observation 23, it is $(1, 1/\alpha)$-recursively-clusterable, meeting the condition of Theorem 11. Hence Theorem 11 implies that $d_{(\alpha)}$ is embeddable into the $\mathbb{R}^k$ with distortion at most $O(\sqrt{k}q^{1/k})$, and thus d with distortion at most $O(kq^{2/k})$. ◀

▶ Remark 24. For $k > \log q$, if a spherical code is used as discussed in Remark 21, we set $\alpha = O(q^{1/k}) = O(1)$, achieving a total of $O(1)$ distortion.