Embedding Bounded Bandwidth Graphs into ℓ_1

Douglas E. Carroll *[1], Ashish Goel **[2], and Adam Meyerson***[1]

[1] UCLA
[2] Stanford University

Abstract. We introduce the first embedding of graphs of low *bandwidth* into ℓ_1, with distortion depending only upon the bandwidth. We extend this result to a new graph parameter called *tree-bandwidth*, which is very similar to (but more restrictive than) treewidth. This represents the first constant distortion embedding of a non-planar class of graphs into ℓ_1. Our results make use of a new technique that we call *iterative embedding* in which we define coordinates for a small number of points at a time.

* Department of Computer Science, University of California, Los Angeles. Email: dcarroll@cs.ucla.edu.

** Departments of Management Science and Engineering and (by courtesy) Computer Science, Stanford University. Research supported in part by NSF CAREER award 0133968, and by an Alfred P. Sloan Fellowship. Email: ashishg@stanford.edu.

*** Department of Computer Science, University of California, Los Angeles. Email: awm@cs.ucla.edu.

1 Introduction

Our main result is a technique for embedding graph metrics into ℓ_1, with distortion depending only upon the *bandwidth* of the original graph. A graph has bandwidth k if there exists some ordering of the vertices such that any two vertices with an edge between them are at most k apart in the ordering. While this ordering could be viewed as an embedding into one-dimensional ℓ_1 with bounded *expansion* (any two vertices connected by an edge must be close in the ordering), the *contraction* of such an embedding is unbounded (there may be two vertices which are close in the ordering but not in the original metric). Obtaining an embedding with bounded distortion (in terms of *both* expansion and contraction) turns out to be non-trivial.

In fact, our results can be extended to a new graph parameter that we call *tree-bandwidth*. We observe that metrics based on trees are easy to embed into ℓ_1 isometrically, despite the fact that even a binary tree can have large bandwidth. The *tree-bandwidth* parameter is a natural extension of bandwidth, where vertices are placed along a tree instead of being ordered linearly. We prove that the shortest path metric of an unweighted graph can be embedded into ℓ_1 with distortion depending only upon the *tree-bandwidth* of the graph (thus independent of the number of vertices).

We achieve these results by introducing a novel technique for *iterative embedding* of graph metrics into ℓ_1. The idea is to partition the graph into small sets and embed each set separately. The coordinates of each specific point are determined when the set containing that point is embedded. Two embeddings will be computed for each set of points. One is generated via some local embedding technique, and maintains accurate distances between the members of the same set. The other embedding copies a set of "parent" points; the goal is to maintain small distances between points and their parents. These two sets of coordinates will be carefully combined to generate the final coordinates for the new set of points. We then proceed to the next set in the ordering.

For ease of exposition we use a very simple local embedding technique in this paper. However, we have also proven a more general result in which we show that with iterative embedding, any reasonable local embedding technique suffices for embedding into ℓ_1 with distortion dependent only upon the *tree-bandwidth* (proof omitted). This leaves open the possibility that the dependence on the tree-bandwidth could be improved with a different local embedding technique.

The motivation for our work is a conjecture (stated by Gupta *et al.* [9] and others) that excluded-minor graph families can be embedded into ℓ_1 with distortion dependent only upon the set of excluded minors. This is one of the major conjectures in metric embedding, and many previous results have resolved special cases of this conjecture. However, all previous ℓ_1 embedding results either yield distortion dependent upon the number of points in the metric [4, 15], or apply only to a subset of the planar graphs [13, 9, 6]. While our results do not resolve the conjecture, we are able to embed a well-studied subclass of graphs (bandwidth-k graphs) with distortion independent of the number of points in the metric. This is the first such result for a non-planar graph class. In addition, our definition of *tree-bandwidth* is similar to (although possibly weaker than) *treewidth.* While we conjecture that there exist families of graphs with low treewidth but unbounded tree-bandwidth, it is interesting to note that weighted treewidth-k graphs can be embedded with constant distortion into *weighted* tree-bandwidth-$O(k)$ graphs.

We note that at each step, our embedding technique requires the existence of a previously embedded "parent" set such that each point of the new set is close to one of the parents, but no point in the new set is close to any other previously embedded set. This property implies the existence of a hierarchy of small node separators (small sets of nodes which partition the graph), which is exactly the requirement for a graph of low *treewidth.* However, we also need each point to be close to *some* member of the parent set, which motivates our definition of the *tree-bandwidth* parameter.

1.1 Related Work

A great deal of recent work has concentrated on achieving tight distortion bounds for ℓ_1 embedding of restricted classes of metrics. For general metrics with n points, the result of Bourgain[4] showed that embedding into ℓ_1 with $O(\log n)$ distortion is possible. A matching lower bound (using expander graphs) was introduced by LLR [11]. It has been conjectured by Gupta *et al.* [9], and Indyk [10] that the shortest-path metrics of planar graphs can be embedded into ℓ_1 with constant distortion. Gupta *et al.* [9] also conjecture that excluded-minor graph families can be embedded into ℓ_1 with distortion that depends only on the excluded minors. In particular, this would mean that for any k the family of treewidth-k graphs could be embedded with distortion $f(k)$ independent of the number of nodes in the graph[3]. Such results would be the best possible for very general and natural classes of graphs.

Since Okamura and Seymour [13] showed that outerplanar graphs can be embedded isometrically into ℓ_1, there has been significant progress towards resolving several special cases of the aforementioned conjecture. Gupta *et al.* [9] showed that treewidth-2 graphs can be embedded into ℓ_1 with constant distortion. Chekuri *et al.* [6] then followed this by proving that k-outerplanar graphs can be embedded into ℓ_1 with constant distortion. Note that all these graph classes not only have low treewidth, but are planar. We give the first constant distortion embedding for a non-planar subclass of the bounded treewidth graphs.

Rao [15] proved that any minor excluded family can be embedded into ℓ_1 with distortion $O(\sqrt{\log n})$. This is the strongest general result for minor-excluded families. Rabinovich [14] introduced the idea of *average distortion* and showed that any minor excluded family can be embedded into ℓ_1 with constant *average distortion.*

Graphs of low treewidth have been the subject of a great deal of study. For a survey of definitions and results on graphs of bounded treewidth, see Bodlaender [2]. More restrictive graph parameters include domino treewidth [3] and bandwidth [7], [8].

[3] There is a lower bound of $\Omega(\log k)$ arising from expander graphs.

2 Definitions and Preliminaries

Given two metric spaces (G, ν) and (H, μ) and an embedding $\Phi : G \to H$, we say that the *distortion* of the embedding is $\|\Phi\| \cdot \|\Phi^{-1}\|$ where

$$\|\Phi\| = \max_{x,y \in G} \frac{\mu(\Phi(x), \Phi(y))}{\nu(x,y)}, \qquad \|\Phi^{-1}\| = \max_{x,y \in G} \frac{\nu(x,y)}{\mu(\Phi(x), \Phi(y))}$$

Parameter $\|\Phi\|$ will be called the *expansion* of the embedding and parameter $\|\Phi^{-1}\|$ is called the *contraction*. We will define bandwidth and then present our definition of the generalization tree-bandwidth.

Definition 1. *Given graph $G = (V, E)$ and linear ordering $f : V \to \{1, 2, ..., |V|\}$ the* bandwidth *of f is $max\{|f(v) - f(w)||(v, w) \in E\}$. The* bandwidth *of G is the minimum bandwidth over all linear orderings f.*

Definition 2. *Given a graph $G = (V, E)$, we say that it has* tree-bandwidth k *if there is a rooted tree $T = (I, F)$ and a collection of sets $\{S_i \subset V | i \in I\}$ such that:*

1. $\forall i, |S_i| \leq k$
2. $V = \bigcup S_i$
3. *the S_i are disjoint*
4. $\forall (u, v) \in E$, *u and v lie in the same set S_i or $u \in S_i$ and $v \in S_j$ and $(i, j) \in F$.*
5. *if c has parent p in T, then $\forall v \in S_c, \exists u \in S_p$ such that $d(u, v) \leq k$.*

We claimed that *tree-bandwidth* was a generalization of *bandwidth*. Intuitively, we can divide a graph of low bandwidth into sets of size k (the first k points in the ordering, the next k points in the ordering, and so forth). We then connect these sets into a path. This gives us all the properties required for tree-bandwidth except for the fifth property – there may be some node which is not close to any node which appeared prior to it in the linear ordering. We can fix this problem by defining a new linear ordering of comparable bandwidth. The proof of this fact has been deferred until the full version of the paper.

Lemma 1. *Graph $G = (V, E)$ with bandwidth b has tree-bandwidth at most $2b$.*

We will now define *treewidth* and show the close relationship between the definitions of treewidth and tree-bandwidth.

Definition 3. *(i) Given a connected graph $G = (V, E)$, a* DFS-tree *is a rooted spanning subtree $T = (V, F \subset E)$ such that for each edge $(u, v) \in E$, v is an ancestor of u or u is an ancestor of v in T.*

(ii) The value *of DFS-tree T is the maximum over all $v \in V$ of the number of ancestors that are adjacent to v or a descendent of v.*

(iii) The edge stretch *of DFS-tree T is the the maximum over all $v, w \in V$ of the distance $d(v, w)$ where w is an ancestor of v and w is adjacent to v or a descendent of v.*

We use the following definition of treewidth due to T. Kloks and related in a paper of Bodlaender [2]:

Definition 4. *Given a connected graph $G = (V, E)$, the* treewidth *of G is the minimum value of a DFS-tree of a supergraph $G' = (V, E')$ of G where $E \subset E'$.*

The following proposition follows immediately from the definition of tree-bandwidth:

Proposition 1. *Given a connected graph $G = (V, E)$, the* tree-bandwidth *of G is the minimum* edge stretch *of a DFS-tree of G.*

Thus, treewidth and tree-bandwidth appear to be related in much the same way that cutwidth and bandwidth are related (see [2] for instance). The close relationship between treewidth and tree-bandwidth is cemented by the following observation (the proof is deferred until the full version of the paper):

Lemma 2. *Any metric supported on a weighted graph $G = (V, E)$ of treewidth-k can be embedded with distortion 4 into a weighted graph with tree-bandwidth-$O(k)$*

Thus, a technique for embedding *weighted* tree-bandwidth-k graphs into ℓ_1 with $O(f(k))$ distortion would immediately result in constant distortion ℓ_1-embeddings of weighted treewidth-k graphs.

2.1 Bounded Bandwidth Example

To see that previous constant distortion embedding techniques do not handle bounded bandwidth graphs consider the following example. Construct a graph G by connecting k points in an arbitrary way, then adding k new points connected to each other and the previous k points in an arbitrary way, and repeat many times.

Clearly the graph G generated in this way has bandwidth $\leq 2k-1$. However, note that if $k \geq 3$ and some set of $2k$ consecutively added points contains $K_{3,3}$ or K_5 then G is not planar and thus previous constant distortion ℓ_1-embedding techniques cannot be applied [13, 9, 6]. G does have bounded treewidth, so Rao's algorithm [15] can be applied but it only guarantees $O(\sqrt{\log n})$ distortion.

2.2 Bounded Tree-Bandwidth Example

To show that bounded tree-bandwidth graphs form a broader class than the bounded bandwidth graphs consider the following example. Let $G = (V', E')$ consist of k copies of an arbitrary tree $T = (V, E)$. Construct G' from G as follows:

1. For $x \in V$, let $\{x_1, ..., x_k\}$ be the k copies of x in V'.
2. For each $x \in V$, connect $\{x_1, ..., x_k\}$ in an arbitrary way.

While the resulting graph G' clearly has tree-bandwidth k, a complete binary tree of depth d has bandwidth $\Omega(d)$ [7], thus G' may have bandwidth $\Omega(\log n)$.

Note again that if $k \geq 5$ and G' contains $K_{3,3}$ or K_5 then G' is not planar and thus previous constant distortion ℓ_1-embedding techniques cannot be applied.

Also note that there are trees T with $|V| = n$ such that any ℓ_2-embedding of T has distortion $\Omega(\sqrt{\log\log n})$ [5]. Since Rao's technique embeds first into ℓ_2 this gives a lower bound of $\Omega(\sqrt{\log\log n})$ on the distortion achievable using Rao's technique to embed G' into ℓ_1. The technique presented in this paper embeds these examples into ℓ_1 with distortion depending only on k.

Apart from being interesting from a technical viewpoint, bounded tree-bandwidth graphs may also be a good model for phylogenentic networks with limited introgression/reticulation [12]. This is a fruitful connection to explore, though it is outside the scope of this paper.

3 Algorithm

Given a graph G of tree-bandwidth k, it must have a tree-bandwidth-k decomposition $(T, \{X_i\})$. We will embed the sets X_i one set at a time according to a DFS ordering of T. When set X_i is embedded, all members of that set will be assigned values for each coordinate. Note that once a point is embedded, *its coordinates will never change - all subsequently defined coordinates will be assigned value zero for these points.* Note that when new coordinates are introduced, these are considered to be coordinates that were never used at any previous point in the algorithm.

For each set we will obtain two embeddings: one derived by extending the embedding of the parent of X_i in T and one local embedding using a simple deterministic embedding technique. We prove the existence of a method for combining these two embeddings to provide an acceptable embedding of the set X_i.

At stage i, our algorithm will compute a weight for each partition S of X_i. We would like these weights to look like $w_M(S)$ - the distance between the closest pair of points separated by S. The embedded distance between two points x, y in X_i will be the sum of weights over partitions separating x from y. The weights suggested above will guarantee no contraction and bounded expansion within X_i. We can transform weighted partitions into coordinates by introducing $w_M(S)$ coordinates for each partition S, such that the coordinate has value 1 for each $x \in S$ and value -1 for each $x \in X_i - S$.

This approach will create entirely new coordinates for each point. Since points in X_i are supposed to be close to points in $X_{p(i)}$, this can create large distortion between sets. Instead of introducing all new coordinates, we would like to "reuse" existing coordinates by forcing points in X_i to take on values similar to those taken on by points in $X_{p(i)}$.

To reuse existing coordinates we will choose a "parent" in $X_{p(i)}$ for each point $x \in X_i$ and identify x with its parent $p(x)$. The critical observation here is that each point in X_i is within distance k of some point in $X_{p(i)}$. Therefore, the partition weights (and hence distances) established by these coordinates are good approximations of the target values we would like to assign.

More precisely, for each point $x \in X_i$ there is at least one closest point in $X_{p(i)}$. Choose an arbitrary such point to be the parent of x. After identifying points in this way, each parent coordinate induces a partition S on X_i between points whose parents have values 1 and -1 in that coordinate. We can define $w_P(S)$ to be the number of parent coordinates inducing partition S. If $|w_P(S) - w_M(S)|$ is always small then the independent local weightings agree and we get a good global embedding.

Unfortunately, there are cases in which $w_P(S) - w_M(S)$ can be large. However, we can successfully combine the two metrics by using the following weighting: $w_F(S) = \max(w_M(S), w_P(S) - \mu)$. The key property of this weighting is that we do not activate too many new coordinates (since $w_P(S)$ not much less than $w_F(S)$) nor do we deactivate too many existing coordinates ($w_P(S)$ not much more than $w_F(S)$). In addition, we can show that $w_F(S)$ does not contract nor greatly expand distances between points of X_i.

3.1 MIN-SEPARATOR Embedding

We can prove that any reasonable local embedding technique suffices to obtain $O(f(k))$ distortion. However, that proof is quite involved and is omitted from this abstract. Instead, for ease of exposition, we will employ a simple local embedding technique which we call a MIN-SEPARATOR embedding and which is described below. The MIN-SEPARATOR embedding returns similar embeddings for independently embedded metrics with similar distances. This is a very useful property and greatly simplifies our overall algorithm and analysis[4].

MIN-SEPARATOR embedding: Given metric (G, d), we assign a weight for each of the distinct partitions of G. To each partition S we assign weight $w_{M(G)}(S) = d(S, G - S) = \min\{d(x, y) | x \in S, y \in G - S\}$. Note that when the source metric is clear we will denote these weights as $w_M(S)$. We then transform these weighted partitions into coordinates by introducing $w_M(S)$ coordinates for each partition S such that the coordinate has value 1 for each $x \in S$ and value -1 for each $x \in G - S$. The distances in this embedding become $d_{M(G)}(x, y) = \sum_{S \in 2^G : x \in S, y \in G - S} w_{M(G)}(S) = \sum_{S \in 2^G : x \in S, y \in G - S} d(S, G - S)$

Lemma 3. *The MIN-SEPARATOR embedding does not contract distances and does not expand distances by more than* 2^k.

Proof. First we show that MIN-SEPARATOR does not contract the distance between x and y. The proof is by induction on the number of points in the metric (G, d).

[4] It is conceivable that a different local embedding technique might result in a better dependence on k.

If $|G| = 2$, then there is only one non-trivial partition and it has weight $d(x, y)$. For larger graphs, there must be some point z other than x, y. Let $B = G - \{z\}$; by the inductive hypothesis the claim holds on set B. However, we observe that the embedded distance $d_{M(B)}(x, y)$ is at most the embedded distance $d_{M(G)}(x, y)$. For any partition of B, we can consider two new partitions of G (one with z on each side) and observe that the total weight MIN-SEPARATOR places on these partitions must be at least the weight MIN-SEPARATOR placed on the original partition of B (this because of triangle inequality).

We now show that MIN-SEPARATOR does not expand distances by more than 2^k. For each partition S which separates x, y, $w_{M(G)}(S) \leq d(x, y)$ and since there are $< 2^k$ partitions which separate x, y, $d_{M(G)}(x, y) \leq 2^k d(x, y)$.

3.2 Combining the Local Embeddings

The algorithm EMBED-BAND relies on three critical properties of the tree-bandwidth decomposition:

1. Each node in X_i is within distance k of a node in the parent of X_i.
2. The nodes of X_i are not adjacent to any previously embedded nodes except those in the parent of X_i.
3. The number of points in X_i is at most k.

The first property enables us to prove that

$$w_P(S) - \mu \leq w_F(S) \leq w_P(S) + 2k \tag{1}$$

This is key in bounding the distortion between sets, since it indicates that we never introduce or "zero-out" too many coordinates for any partition S of X_i.

The second property means that we don't need to bound expansion between too many pairs of points. As long as we can prove that distances between points in X_i and $X_{p(i)}$ don't expand too much, the triangle inequality will allow us to bound expansion between all pairs of points.

The third property allows us to bound the distortion of the local embedding (MIN-SEPARATOR) as well as to bound the total number of coordinates introduced or zeroed out, since there are only 2^k partitions of set X_i with k points.

3.3 Example: Embedding a Cycle

It is instructive to observe what happens when embedding a cycle (see figure 1). It is clear that the first two points in the cycle (X_1) can be embedded acceptably. As we embed subsequent sets we embed the descendents of these two points. Because the pairs of points in consecutive sets diverge, each new point inherits the values of all of the coordinates of its parent. Additionally, new coordinates are added to separate the pairs of points. The union of these coordinates is enough to establish the distances between these pairs of points as they diverge.

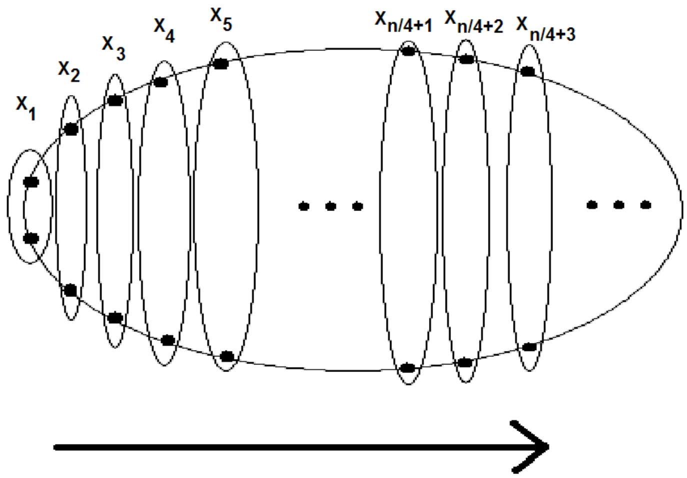

Fig. 1. Embedding a Cycle

After embedding half the points in the cycle, the pairs of points in subsequent sets begin to converge. Whenever the distance induced by the parent points exceeds the target distance of the current points (represented by the MIN-SEPARATOR distance), we set the values of μ coordinates establishing that distance to zero *for the new points*. Because points in consecutive sets are within distance k of their parents, the distances between consecutive pairs of points cannot decrease by more than $2k$ per step. Thus, zeroing μ coordinates at each step is more than sufficient to compensate for the decreasing distances.

It might appear that zeroing μ coordinates at each step would contract distances between points and their ancestors, but recall that we also define β new coordinates at each step to separate the current points from all previously embedded points and prevent such contractions.

4 Analysis

The central result of this paper follows directly from the lemmas below:

Theorem 1. *Algorithm* EMBED-BAND *embeds tree-bandwidth-k graphs into* ℓ_1 *with distortion* $\leq 2\beta = 4 \cdot 2^k \mu = 16k \cdot 2^{2k}$.

Lemma 4. *The distances between points embedded simultaneously are not contracted.*

Proof. If x, y are in the same tree node X_i, then the distance $d_E(x, y)$ is at least as large as the distance $d_{M(X_i)}(x, y)$ returned by MIN-SEPARATOR. This is because for every partition we use $w_F(S) = \max\{w_M(S), w_P(S) - \mu\} \geq w_M(S)$.

Input: Assume $G = (V, H)$ has tree-bandwidth decomposition $(T = (I, F), \{X_i | i \in I\})$. Let $p(i)$ be the parent of $i \in T$. Assume that $p(i)$ appears before i in the ordering of the nodes of I. X_1 is the root of T.

1. $\mu \leftarrow 4k2^k$
2. for each of the $2^{k-1} - 1$ non-trivial partitions S of X_1:
   (a) $w_M(S) \leftarrow \min\{d(x, y) | x \in S, y \in X_1 - S\}$
   (b) define $w_M(S)$ new coordinates
   (c) for each new coordinate c set:
   $x_c \leftarrow 1$ if $x \in S$,
   $x_c \leftarrow -1$ if $x \in X_1 - S$
3. FOR $i \leftarrow 2$ TO $|I|$
   (a) for each $x \in X_i$, let $p(x)$ be the parent of x (closest node to x) in $X_{p(i)}$.
   (By identifying nodes x with their parents $p(x)$, each existing coordinate induces a partition on the points of X_i.)
   (b) for each of the $2^{k-1} - 1$ non-trivial partitions S of X_i:
      i. $w_M(S) \leftarrow \min\{d(x, y) | x \in S, y \in X_i - S\}$
      ii. $w_P(S) \leftarrow$ # of existing coordinates which induce S via $X_{p(i)}$
      iii. $w_F(S) \leftarrow \max(w_M(S), w_P(S) - \mu)$
      iv. if $w_F(S) > w_P(S)$ then:
         A. for all the $w_P(S)$ coordinates that induce partition S set $x_c \leftarrow p(x)_c$ for all $x \in X_i$
         B. define $w_F(S) - w_P(S)$ new coordinates
         C. for each new coordinate c set:
         $x_c \leftarrow 1$ if $x \in S$,
         $x_c \leftarrow -1$ if $x \in X_i - S$
         ($x_c \leftarrow 0$ for all previously embedded points)
      v. If $w_F(S) \leq w_P(S)$ then:
         A. for $w_P(S) - w_F(S)$ of the coordinates that induce partition S set $x_c \leftarrow 0$ for all $x \in X_i$
         B. for the $w_F(S)$ remaining coordinates that induce partition S set $x_c \leftarrow p(x)_c$ for all $x \in X_i$
   (c) $x_c \leftarrow p(x)_c$ for all coordinates c which do not partition X_i
   (d) define an additional $\beta = 2 \cdot 2^k \mu$ coordinates and set $x_c \leftarrow 1$ for all $x \in X_i$
4. NEXT i

Fig. 2. Algorithm EMBED-BAND

Lemma 5. *The distances between points embedded simultaneously are expanded by at most a factor of* 2^k

Proof. Recall that for each partition S of X_i, we compute three weights: a local weight, a "parent" weight, and the final weight which we use to embed the current tree node.

$$\begin{aligned} w_M(S) &= min\{d(x,y)|x \in S, y \in X_i - S\} \\ w_P(S) &= \#\text{ of existing coordinates that induce } S \text{ via } X_{p(i)} \\ w_F(S) &= max(w_M(S), w_P(S) - \mu) \end{aligned}$$

If for all partitions S separating x and y, we have $w_F(S) = w_M(S)$, then the embedded distance will be the same as that from MIN-SEPARATOR, which is at most $2^k d(x,y)$.

Otherwise, at least one partition separating x and y has $w_F(S) = w_P(S) - \mu$. Note that by the triangle equality, $d(x,y) \leq d(p(x),p(y)) + 2k$ for all x, y. Thus every such partition has $w_F(S) \leq w_P(S) + 2k$, so by summing and observing that there are only 2^k possible partitions of k points, we have $d_E(x,y) \leq d_E(p(x),p(y)) - \mu + 2k2^k$. Applying our inductive hypothesis to points in the parent set and using $\mu = 4k2^k$ gives the desired bound.

Lemma 6. *The distances between points in different sets are expanded by at most* $2\beta = 4 \cdot 2^k \mu$ *where* $\mu = 4k2^k$*.*

Proof. Consider $x \in X_i$ and $y \in X_j$. X_i and X_j are connected by a unique path Q in T. Assume WLOG that $X_{p(i)}$ is in Q. Our proof will be by induction on the length of Q.

If $length(Q) = 1$, this means $X_j = X_{p(i)}$ and by triangle inequality we have $d_E(y,x) \leq d_E(y,p(x)) + d_E(p(x),x)$. The distortion of the first quantity is bounded because these points are in the same tree node. The second quantity is bounded by β plus the sum of differences in partition weights since we re-use coordinates when possible. Combining these, and observing that $p(x)$ is closer to x than y is, we obtain $d_E(y,x) \leq 2\beta d(y,x)$. If $length(Q) > 1$, there must be a point $z \in X_{p(i)}$ such that z lies on a shortest path between x and y in G. By the induction hypothesis, $d_E(x,z) \leq 2\beta d(x,z)$ and $d_E(z,y) \leq 2\beta d(z,y)$. Thus, $d_E(x,y) \leq d_E(x,z) + d_E(z,y) \leq 2\beta d(x,z) + 2\beta d(z,y) = 2\beta d(x,y)$ since z is on the shortest path between x and y.

Lemma 7. *The distances between points in different sets are not contracted.*

Proof. Consider $x \in X_i$ and $y \in X_j$. X_i and X_j are connected by a unique path Q in T. Assume WLOG that $X_p(i)$ is in path Q. x has a closest ancestor z in X_j which is at distance $d_E(z,y)$ from y. Consider the path from z to x that lies in Q. Intuitively, we activate at least β coordinates at each step and deactivate at most $2^k\mu$, so distances increase as $\approx (\beta - 2^k\mu)|Q|$. So

$$\begin{aligned} d_E(x,y) &\geq max((d_E(z,y) - 2^k\mu|Q|), 0) + \beta|Q| \geq d_E(z,y) - 2^k\mu|Q| + \beta|Q| \\ &\geq d(z,y) - 2^k\mu|Q| + \beta|Q| \geq d(x,y) - 2k|Q| - 2^k\mu|Q| + \beta|Q| \\ &= d(x,y) + (\beta - 2k - 2^k\mu)|Q| \geq d(x,y) \end{aligned}$$

References

1. M. Badoiu, K. Dhamdhere, A. Gupta, Y. Rabinovich, H. Raecke, R. Ravi, and A. Sidiropoulos, “Approximation Algorithms for Low-Distortion Embeddings Into Low-Dimensional Spaces”, In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*, 2005, pp. 119-128.
2. H. Bodlaender, “A partial k-arboretum of graphs with bounded treewidth”, *Theoretical Computer Science*, 209 (1998), pp. 1-45.
3. H. Bodlaender, “A Note On Domino Treewidth”, *Discrete Mathematics and Theoretical Computer Science*, 3 (1999), pp. 144-150.
4. J. Bourgain, “On Lipschitz Embeddings of Finite Metric Spaces in Hilbert Space.”, *Israel Journal of Mathematics*, 52 (1985), pp. 46-52.
5. J. Bourgain, “The Metrical Interpretation of Superreflexivity in Banach Spaces.”, *Israel Journal of Mathematics*, 56 (1986), pp. 222-230.
6. C. Chekuri, A. Gupta, I. Newman, Y. Rabinovich, and A. Sinclair, “Embedding k-Outerplanar Graphs into ℓ_1”, In *Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms*, 2003, pp. 527-536.
7. F. Chung, P. Seymour, “Graphs with Small Bandwidth and Cutwidth”, *Discrete Mathematics*, 75 (1989), pp. 113-119.
8. U. Feige, “Approximating the bandwidth via volume respecting embeddings”, In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, 1998, pp. 90-99.
9. A. Gupta, I. Newman, Y. Rabinovich, and A. Sinclair, “Cuts, trees and ℓ_1-embeddings.”, In *Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science*, 1999, pp. 399-408.
10. P. Indyk, “Algorithmic Aspects of Geometric Embeddings.”, In *Proceedings of the 42th Annual IEEE Symposium on Foundations of Computer Science*, 2001, pp. 10-33.
11. N. Linial, E. London, and Y. Rabinovich, “The geometry of graphs and some of its algorithmic applications”, *Combinatorica*, 15 (1995), pp. 215-245.
12. L. Nakhleh, T. Warnow, C. R. Linder, and K. St. John, “Reconstructing reticulate evolution in species - theory and practice.”, *Journal of Computational Biology*, (To Appear).
13. H. Okamura, P. Seymour, “Multicommodity Flows in Planar Graphs”, *Journal of Combinatorial Theory Series B*, 31 (1981), pp. 75-81.
14. Y. Rabinovich, “On Average Distortion of Embedding Metrics into the Line and into ℓ_1”, In *Proceedings of the 35th Annual ACM Symposium on Theory of Computing*, 2003, pp. 456-462.
15. S. Rao, “Small distortion and volume preserving embeddings for Planar and Euclidean metrics”, In *Proceedings of the 15th Annual Symposium on Computational Geometry*, 1999, pp. 300-306.
16. N. Robertson, P. Seymour, “Graph Minors II. Algorithmic Aspects of Tree-Width”, *Journal of Algorithms*, 7 (1986), pp. 309-322.