



ELSEVIER

Information Processing Letters 80 (2001) 213–219

Information  
Processing  
Letters

www.elsevier.com/locate/ipl

# On approximating planar metrics by tree metrics

Goran Konjevod<sup>a,1</sup>, R. Ravi<sup>b,\*</sup>, F. Sibel Salman<sup>c,2</sup>

<sup>a</sup> Department of Computer Science and Engineering, Arizona State University, Tempe, AZ 85282-5406, USA

<sup>b</sup> GSIA, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA

<sup>c</sup> Krannert School of Management, Purdue University, West Lafayette, IN 47907, USA

Received 8 September 1999; received in revised form 14 June 2000

Communicated by F.B. Schneider

---

## Abstract

We combine the results of Bartal [Proc. 37th FOCS, 1996, pp. 184–193] on probabilistic approximation of metric spaces by tree metrics, with those of Klein, Plotkin and Rao [Proc. 25th STOC, 1993, pp. 682–690] on decompositions of graphs without small  $K_{s,s}$  minors (such as planar graphs) to show that metrics induced by such graphs (with unit lengths on the edges) can be probabilistically approximated by tree metrics with an  $O(\log \text{diam } G)$  distortion. This improves upon Bartal's result that holds for general  $n$ -node metrics with a distortion of  $O(\log n \log \log n)$ . The main ingredient of our proof is that *weak* probabilistic partitions suffice for the construction of tree metrics with low distortion, in contrast to strong partitions used by Bartal. We also show that our result is the best possible by providing a lower bound of  $\Omega(\log \text{diam } G)$  for the distortion of any probabilistic approximation of the square grid by tree metrics. © 2001 Elsevier Science B.V. All rights reserved.

*Keywords:* Algorithms; Combinatorial problems; Metrics; Tree metrics; Planar graphs; Probabilistic approximations

---

## 1. Introduction

### 1.1. Metric approximations

Approximating a given metric by structurally simpler metrics has been an area of much research motivated by several different perspectives such as functional analysis [9], graph theory [7], distributed com-

putation [14], approximation algorithms [12], and computational biology [1]. A popular simple metric used to approximate a given metric is an additive or tree metric, represented by distances arising from a tree containing the given points. The distances in the additive metric must dominate the distances in the graph but not be much greater. The *distortion* is the maximum ratio of any distance in the simpler metric to the original distance. However, even for the simple metric induced by an  $n$ -cycle, if we restrict ourselves to approximating by a tree metric, the distortion must be at least  $\Omega(n)$  [15].

To overcome this obstacle, Karp [10] introduced the notion of probabilistic approximation and showed how the cycle metric can be 2-approximated by a uniform distribution on the path metrics obtained by deleting an edge of the cycle. The goal is to give a

---

\* Corresponding author. Supported by an NSF CAREER grant CCR-9625297.

E-mail addresses: goran@asu.edu (G. Konjevod), ravi@cmu.edu (R. Ravi), fs2c+@andrew.cmu.edu (F.S. Salman).

<sup>1</sup> Work done while in the Department of Math Sciences, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA. Partially supported by an NSF CAREER grant CCR-9625297.

<sup>2</sup> Work done while in GSIA, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA.

polynomially computable probability distribution over a set of simpler metrics each of which dominates the given metric, and at the same time to keep the expected distortion for any edge small. Alon et al. [2] extended this result to show that every metric on  $n$  points can be probabilistically approximated by tree metrics with distortion  $2^{O(\sqrt{\log n \log \log n})}$ .

Bartal [4] achieved a better distortion of  $O(\log^2 n)$  and improved this bound to  $O(\log n \log \log n)$  [5]. The technique underlying his results is an efficient computation of (strong) probabilistic partitions of the graph, where the induced graph on each cluster of the partition has a small diameter. The idea is to define a distribution over the set of strong partitions such that for every edge, the probability it is cut by a partition is small. This notion can be extended to define weak probabilistic partitions, where we no longer insist that the diameters of the induced graphs on the clusters be small, but only that any pair of nodes in a cluster are sufficiently close in the original graph (i.e., we impose a bound on the “weak” diameter of the clusters, hence the name).

### 1.2. Our results

In this paper, we first show that weak probabilistic partitions suffice to guarantee probabilistic approximations by  $k$ -HSTs (see definition in Section 2) with the same distortion of  $O(\log^2 n)$  as in Bartal [4], closely following his construction.

Bartal gives a lower bound of  $\Omega(\log n)$  for approximating the metric of any graph, leaving a small gap to close, as the upper bound is  $O(\log n \log \log n)$ . We show that for any positive integer  $s$ , the metric of a graph  $G$  without a  $K_{s,s}$  minor can be probabilistically approximated by tree metrics with distortion  $O(\log \text{diam } G)$ . For example, planar graphs exclude  $K_{3,3}$ . We obtain this improvement by randomizing a weak-diameter decomposition procedure due to Klein et al. [11] to get weak probabilistic partitions, and using the previous result to build the  $k$ -HSTs.

We then show that this result is asymptotically optimal by providing a lower bound of  $\Omega(\log \text{diam } G)$  for the distortion of any probabilistic approximation of the square grid by tree metrics. Our lower bound proof uses the easy direction of Yao’s minimax theorem [17] and bounds the average distortion of any edge of

the square grid in any tree-metric approximation by extending a result of Alon et al. [2].

Finally, we turn to applications of probabilistic approximation by tree metrics to network design problems [3,6,16]. In these applications, it is convenient if the tree metric is a spanning tree of the original metric. We give a simple “lifting” procedure to convert the approximating  $k$ -HSTs into spanning trees of the original metric while only doubling the expected distortion.

## 2. Probabilistic approximation of metrics via weak partitions

We use the following definitions from Bartal [4].

**Definition 2.1.** Given a metric space  $M$  over a finite set  $V$ , denote the distance between two points  $x, y \in V$  by  $d_M(x, y)$ . A family  $\mathcal{F}$  of metric spaces over  $V$ ,  $\alpha$ -probabilistically approximates  $M$ , if

$$d_M(x, y) \leq d_N(x, y), \quad \text{for all } N \in \mathcal{F}, x, y \in V,$$

and there exists a probability distribution  $\mathcal{D}$  on  $\mathcal{F}$  such that

$$\mathbf{E}_{\mathcal{D}}[d_N(x, y)] \leq \alpha d_M(x, y),$$

for all  $x, y \in V$ , where  $N \in \mathcal{F}$  is chosen according to the distribution  $\mathcal{D}$ . We call  $\alpha$  the distortion.

**Definition 2.2.** Let  $G = (V, E, w)$  be a weighted connected graph. Consider a partition of the vertex set  $V = \bigcup_i V_i$ . The subgraph of  $G$  induced by  $V_i$ ,  $G[V_i]$  is called a *cluster* of the partition. The *strong diameter* of the cluster is the diameter of  $G[V_i]$ , whereas the *weak diameter* of the cluster is the maximum distance in  $G$  between two vertices of  $V_i$ .

**Definition 2.3.** An  $\ell$ -(weak) partition of  $G$  is a partition of the vertex set  $V = \bigcup_i V_i$  such that the strong (weak) diameter of  $G[V_i]$  is at most  $\ell$ , for all  $i$ .

For a probability distribution  $\mathcal{D}$  over all  $\ell$ -partitions of  $G$ , let  $x_e$  denote the probability that edge  $e$  crosses a boundary between some two clusters, i.e., its endpoints are in different clusters.

**Definition 2.4.** For any  $0 \leq r \leq \text{diam } G$ ,  $\lambda, \rho \geq 0$ , an  $(r, \rho, \lambda)$ -(weak) probabilistic partition of  $G$  is a

probability distribution  $\mathcal{D}$  over the set of  $(r\rho)$ -(weak) partitions of  $G$ , such that

$$x_e \leq \frac{\lambda w_e}{r}, \quad \text{for all } e \in E.$$

**Definition 2.5.** Let  $k > 1$ . A  $k$ -hierarchically well-separated tree ( $k$ -HST) is a rooted weighted tree such that

- (1) the edge lengths from any vertex to all of its children are the same, and
- (2) the length of any edge is at most  $1/k$  times the length of its parent edge.

Bartal uses strong probabilistic partitions of a graph recursively to construct a  $k$ -HST that approximates the metric of the graph.

**Theorem 2.6** [4]. For all  $r \geq 0$ , if there is a polynomially computable  $(r, \rho, \lambda)$ -probabilistic partition of every subgraph of  $G$ , then  $G$  can be  $\alpha$ -probabilistically approximated by a polynomially computable distribution on a set of  $k$ -HSTs each of diameter  $O(\text{diam } G)$ , for  $\alpha = O(\rho\lambda k \log_k(\text{diam } G))$ .

We extend this theorem to weak probabilistic partitions.

**Theorem 2.7.** For all  $r \geq 0$ , if there is a polynomially computable  $(r, \rho, \lambda)$ -weak probabilistic partition of every subgraph of  $G$ , then  $G$  can be  $\alpha$ -probabilistically approximated by a polynomially computable distribution on a set of  $k$ -HSTs each of diameter  $O(\text{diam } G)$ , for  $\alpha = O(\rho\lambda k \log_k(\text{diam } G))$ .

**Proof.** We follow Bartal’s construction but deviate from it in the recursive definition of the radius bounds on clusters to accommodate weak partitions. We first construct the family of  $k$ -HSTs that approximate distances in  $G$  and then bound the expected distortion.

For  $i = 1, \dots, \log_k \text{diam } G$ , let  $r_i = \text{diam } G / (\rho k^i)$ . First, find an  $(r_1, \rho, \lambda)$ -weak probabilistic partition of  $G$ . Then, find an  $(r_2, \rho, \lambda)$ -weak probabilistic partition of each of the resulting subgraphs of  $G$ . At a generic iteration  $i$ , find an  $(r_i, \rho, \lambda)$ -weak probabilistic partition of the current graph  $G_i$ . Let  $C_{i+1}^1, \dots, C_{i+1}^s$  be the resulting clusters. For each  $j \in \{1, \dots, s\}$ , recursively construct the  $k$ -HST of  $C_{i+1}^j$ . Denote these trees

$T_{i+1}^j$  and their roots  $q_{i+1}^j$ , for  $j = 1, \dots, s$ . We construct  $T_i$  by adding a root node  $q_i$  and connecting it to  $q_{i+1}^j$ , for all  $j = 1, \dots, s$ , by an edge of length  $r_i \rho k / 2 = \text{diam } G / (2k^{i-1})$ . We stop when  $G_i$  consists of a single vertex, thus the leaves of the tree are the vertices of  $G$ . Let  $T = T_1$ . Note that  $T$  is a  $k$ -HST with the additional property that all the edge lengths at a level are equal, and it has depth at most  $\log_k(\text{diam } G)$ .

First we show by induction that the distance in  $T$  between any pair of nodes in  $V(G)$  is at least the distance in  $G$ . The claim holds trivially when  $G_i$  has a single node. Assume  $d_{T_{i+1}}(u, v) \geq d_G(u, v)$ . If  $u$  and  $v$  are both in cluster  $C_{i+1}^j$ , then

$$d_{T_i}(u, v) = d_{T_{i+1}^j}(u, v) \geq d_G(u, v).$$

Otherwise,

$$d_{T_i}(u, v) \geq \text{diam } G / k^{i-1} = \rho r_i k,$$

and the weak diameter of the cluster at the  $i$ th level containing both  $u$  and  $v$  is at most  $\rho r_{i-1} = \rho r_i k$ , hence

$$d_{T_i}(u, v) \geq d_G(u, v).$$

We will now show that the expected distortion of distances in  $T$  is  $O(\lambda \rho k \log_k(\text{diam } G))$ . The main deviation from Bartal’s proof is that we always relate distances in  $T$  to those in the original graph, as our probabilistic partitions are weak.

First, we relate the diameter of  $T_i$  to that of  $G$ . Let  $T_i$  denote a subtree of  $T$  whose root is a vertex at the  $i$ th level of  $T$  and let  $\gamma(T_i)$  denote the maximal length of a root-leaf path in  $T_i$ . We prove by induction that  $\gamma(T_i) \leq 1/2(1 + 1/(k-1)) \text{diam } G / k^{i-1}$ . Assume the claim holds for  $i + 1$ . Then,

$$\begin{aligned} \gamma(T_i) &= \max_j \gamma(T_{i+1}^j) + \frac{1}{2} \frac{\text{diam } G}{k^{i-1}} \\ &\leq \frac{1}{2} \left( 1 + \frac{1}{k-1} \right) \frac{\text{diam } G}{k^i} \\ &\quad + \frac{1}{2} \frac{\text{diam } G}{k^{i-1}} \frac{1}{2} \frac{\text{diam } G}{k^{i-1}} \left( 1 + \frac{1}{k-1} \right). \end{aligned}$$

Hence,

$$\text{diam } T_i \leq \left( 1 + \frac{1}{k-1} \right) \frac{\text{diam } G}{k^{i-1}}. \tag{1}$$

Bartal’s proof of the upper bound on the expected distance in  $T$  carries over almost directly. The

proof shows by induction on  $i$  that for some  $h \leq \log_k(\text{diam } G)$ ,

$$\mathbf{E}[d_{T_i}(u, v)] \leq \lambda \rho k \left(1 + \frac{1}{k-1}\right) (h-i) d_{G_i}(u, v).$$

Assume the claim holds for  $i+1$ . Let  $A_i$  be the event that a shortest  $u-v$  path is within some cluster  $C_{i+1}^j$  produced at iteration  $i$ . Then,

$$\begin{aligned} \mathbf{E}[d_{T_i}(u, v) | A_i] &= \mathbf{E}[d_{T_{i+1}^j}(u, v)] \\ &\leq \lambda \rho k \left(1 + \frac{1}{k-1}\right) (h-i-1) d_{C_{i+1}^j}(u, v) \quad (2) \end{aligned}$$

$$= \lambda \rho k \left(1 + \frac{1}{k-1}\right) (h-i-1) d_{G_i}(u, v). \quad (3)$$

Let  $x_i(u, v)$  be the probability that some shortest  $u-v$  path crosses between some clusters produced at iteration  $i$ . Then,

$$\begin{aligned} \mathbf{E}[d_{T_i}(u, v)] &\leq x_i(u, v) \text{diam } T_i + (1 - x_i(u, v)) \mathbf{E}[d_{T_i}(u, v) | A_i] \\ &\leq \lambda \frac{d_{G_i}(u, v)}{r_i} \text{diam } T_i + \mathbf{E}[d_{T_i}(u, v) | A_i] \\ &\leq \lambda \frac{d_{G_i}(u, v)}{\text{diam } G} \rho k^i \left(1 + \frac{1}{k-1}\right) \frac{\text{diam } G}{k^{i-1}} \\ &\quad + \lambda \rho k \left(1 + \frac{1}{k-1}\right) (h-i-1) d_{G_i}(u, v) \\ &= \lambda \rho k \left(1 + \frac{1}{k-1}\right) d_{G_i}(u, v) \\ &\quad + \lambda \rho k \left(1 + \frac{1}{k-1}\right) (h-i-1) d_{G_i}(u, v) \\ &= \lambda \rho k \left(1 + \frac{1}{k-1}\right) (h-i) d_{G_i}(u, v). \end{aligned}$$

The first inequality is derived by examining whether the shortest  $u-v$  path crosses cluster boundaries. The second inequality follows since in an  $(r_i, \rho, \lambda)$ -weak probabilistic partition,  $x_i(u, v) \leq \sum_{e \in P} x_e$ , and thus  $x_e \leq \lambda w_e / r_i$  implies  $x_i(u, v) \leq \lambda \cdot d_{G_i}(u, v) / r_i$ . The third inequality uses (1), the definition of  $r_i$  and (2).

Finally, since  $h \leq \log_k(\text{diam } G)$ ,  $\mathbf{E}[d_T(u, v)] \leq \lambda \rho k \left(1 + \frac{1}{k-1}\right) \log_k(\text{diam } G) d_G(u, v)$ .  $\square$

**Remark.** Bartal (personal communication, 1997) has pointed out that the  $\text{diam } G$  bound in the above proof

can be replaced by  $n$  by using the idea of  $\frac{1}{n}$ -forcing weak partitions as in [4].

### 3. Probabilistic partition of graphs without $K_{s,s}$ minors

In this section we randomize the decomposition of Klein et al. [11] to show that given a graph  $G$  and parameters  $\delta$  and  $s$ , either  $G$  has a  $K_{s,s}$  minor, or there is a  $(\delta, O(s^2), s)$ -weak probabilistic partition of any subgraph of  $G$ . This result together with Theorem 2.7 yields a family of  $k$ -HSTs that  $O(s^3 k \log_k \text{diam } G)$ -probabilistically approximates the metric of  $G$ .

We first describe their decomposition procedure. Let  $\delta > 0$  be given. Find a breadth-first-search tree of  $G$  starting from any vertex as the root. If the depth of the tree is less than  $\delta$ , stop. Otherwise choose  $\eta \in \{1, 2, \dots, \delta\}$ , and divide  $V$  into clusters of vertices whose distance from the root is between  $\eta + i\delta$  and  $\eta + (i+1)\delta - 1$ , for  $i = 0, 1, \dots$ , by deleting edges that cross the clusters. Recursively apply the same procedure to each cluster. The depth of the recursion is  $s$ , if  $G$  has no  $K_{s,s}$  minor (thus,  $s = 3$  for a planar graph). In [11],  $\eta$  is chosen to minimize the number of edges removed. Using this decomposition, Klein et al. show the following:

**Theorem 3.1.** *A graph  $G$  without a  $K_{s,s}$  minor can be decomposed into subgraphs of weak diameter  $4\delta s^2$  by removing a constant  $(s/\delta)$  fraction of edges, where  $s < \delta < \text{diam } G$ .*

Our key observation is that the bound on the weak diameter is independent of the choice of the root or the choice of  $\eta$  at any level of the recursion. We simply pick the root and  $\eta$  uniformly at random from the set of all available vertices and  $\{1, \dots, \delta\}$ , respectively.

**Theorem 3.2.** *If a graph  $G$  does not contain a  $K_{s,s}$  minor, then the decomposition procedure outputs a  $(\delta, 4s^2, s)$ -weak probabilistic partition of  $G$ .*

**Proof.** By Theorem 3.1, the weak diameter of the clusters is at most  $4\delta s^2$ . Therefore, each of the partitions that may result as a consequence of particular choices of the cutting parameter  $\eta$  is a  $4\delta s^2$ -weak partition of  $G$ . Since  $\eta$  and the root are chosen uniformly

at random, the probability that a given edge will be deleted is at most  $1/\delta$  at every level of the recursion. Since the recursion terminates after at most  $s$  levels, the union bound gives  $x_e \leq w_e s/\delta$ .  $\square$

The following result easily follows from Theorems 2.7 and 3.2.

**Corollary 3.3.** *Let  $G$  be a weighted graph without a  $K_{s,s}$  minor. Then  $G$  can be  $\alpha$ -probabilistically approximated by a polynomially computable distribution on a set of  $k$ -HSTs, for  $\alpha = O(s^3 k \log k(\text{diam } G))$ .*

#### 4. Lower bounds

In this section we show that the metric of a 2-dimensional  $n \times n$  grid, which we denote by  $G(n, 2)$ , cannot be  $\alpha$ -probabilistically approximated by a tree for any  $\alpha = o(\log n)$ . The result is an extension of a theorem of Alon et al. [2].

An often used method of proving lower bounds for (the running time or the performance guarantee of) randomized algorithms is that of Yao [17,13], based on the minimax principle of linear programming. Consider a randomized algorithm as a distribution  $p$  on a set of deterministic algorithms  $A \in \mathcal{A}$ . If  $q$  is any distribution on the set of inputs  $I \in \mathcal{I}$ , then

$$\min_{A \in \mathcal{A}} \mathbf{E}_q[\text{cost}(I, A)] \leq \mathbf{E}_p[\max_{I \in \mathcal{I}} \text{cost}(I, A)]. \quad (4)$$

**Definition 4.1.** Let  $T$  be a spanning tree of the graph  $G$ . The distortion of an edge  $e$  with respect to  $T$  is the length of the unique path in  $T$  between the endpoints of  $e$ . We denote this by  $\text{dist}_T(e)$ .

Alon et al. [2] prove the following lower bound on the expected distortion.

**Theorem 4.2.** *Let  $T$  be a spanning tree of  $G(n, 2)$ . If  $q$  is the uniform distribution on the edges of  $G(n, 2)$ , then*

$$\mathbf{E}_q[\text{dist}_T(e)] \geq \frac{1}{2048} \ln n - O(1).$$

Note that from (4) it follows that a lower bound for expected distortion of an edge (like the one in Theorem 4.2) translates to the same lower bound for probabilistic approximation by spanning trees. Thus we only

need to show that the proof from [2] can be extended to the more general notion of probabilistic approximation by trees that are not necessarily subgraphs of the original graph.

First of all, we may assume that all the grid vertices are the leaves of the tree we consider (otherwise, add edges of length zero and make the grid vertices leaves). Further, we can reduce the maximum degree of the tree to 4 by splitting any vertex of degree greater than 4 into two vertices (dividing the neighbors between the two vertices) and connecting them by edges of length 0.

**Definition 4.3.** We call a tree  $T$  a *super-spanning tree* of the grid  $G = G(n, 2)$ , if

- (1)  $d_T(x, y) \geq d_G(x, y) \forall x, y \in V(G)$ ,
- (2) every vertex of  $G$  is a leaf of  $T$ , and
- (3) degree of every vertex of  $T$  is at most 4.

Note that a super-spanning tree of a grid need not be a spanning tree. However, the only obstacle is the requirement (2) which can be satisfied by adding edges of cost 0. Thus it is sufficient to generalize Theorem 4.2 to super-spanning trees.

**Lemma 4.4** [2]. *If  $A$  is a set of at least  $\alpha^2$  vertices of  $G(n, 2)$ , and  $|A| \leq n^2/2$ , then there are at least  $\alpha$  rows or at least  $\alpha$  columns that  $A$  intersects but does not fill.*

**Lemma 4.5** (variant of Lemma 6.4 of [2]). *Let  $T'$  be a subtree of a super-spanning tree of  $G(n, 2)$ , and  $A$  be the set of leaves of  $T'$ . If  $A$  contains at least  $\alpha^2$  vertices (of  $G(n, 2)$ ), where  $|A| \leq n^2/2$ , and  $B$  is a set of at most four vertices in  $T'$ , then there are at least  $\alpha/2$  vertices in  $A$  that have neighbors outside  $A$  and have distance at least  $\alpha/16$  from each vertex of  $B$ .*

**Proof.** By the previous lemma there is a set  $S$  of at least  $\alpha$  vertices in  $A$  that are in distinct rows (or columns), and have neighbors outside  $A$  (hence  $T'$ , as well). Because distances in  $T$  overestimate distances in  $G$ , any vertex of  $B$  can be within  $\alpha/16$  of at most  $\alpha/8$  vertices of  $S$  in  $T'$ . So, at least  $\alpha/2$  vertices of  $S$  are more than  $\alpha/16$  away from any vertex of  $B$ .  $\square$

**Lemma 4.6** (variant of Lemma 6.5 of [2]). *If  $T$  is a super-spanning tree of  $G(n, 2)$  and  $1 \leq \alpha \leq n/16$ ,*

then there are at least  $n^2/(64\alpha)$  edges  $e$  such that  $\text{dist}_T(e) \geq \alpha/16$ .

**Proof.** Any tree on  $m > 1$  vertices with maximum degree 4 can be split in two smaller subtrees, each of which has at least  $(m^2 - 1)/4 \geq m^2/8$  leaves by deleting a single edge. We use this observation starting with  $T$ , and repeatedly delete an edge from the biggest remaining component so that this component is split as evenly as possible. In this way, we delete  $\lceil n^2/(8\alpha^2) \rceil - 1$  edges. The average number of leaves in a component is  $8\alpha^2$ , and each of the  $\lceil n^2/(8\alpha^2) \rceil$  components has at least  $\alpha^2$  leaves.

The average number of deleted edges incident with a component is less than 2, so at least half the components are incident with at most 4 deleted edges. Let  $D$  be the set of such components and let  $T'$  be one of the components in  $D$ . Let  $A$  be the set of leaves of  $T'$  and  $B$  be the set of vertices in  $T'$  incident on the deleted edges. Then, by Lemma 4.5 there are at least  $\alpha/2$  leaves in  $T'$  that are at distance at least  $\alpha/16$  from each vertex of  $B$ , and that have neighbors outside  $A$ . The paths in  $T$  from any of these  $\alpha/2$  vertices in  $A$  to their neighbors outside  $A$  must go through a vertex of  $B$ . Applying the same reasoning to each of the  $n^2/(16\alpha^2)$  components in  $D$ , and taking into account double-counting, we get a total of  $n^2/(64\alpha)$  edges whose distortion in  $T$  is more than  $\alpha/16$ .  $\square$

**Theorem 4.7** (Theorem 6.6 in [2]). *If  $T$  is a super-spanning tree of  $G = G(n, 2)$ , and an edge is chosen uniformly at random from  $E(G)$ , then*

$$\mathbf{E}[\text{dist}_T(e)] \geq \frac{1}{2048} \ln n - O(1).$$

The proof uses Lemma 4.6 and is otherwise identical to that in [2].

## 5. Obtaining a spanning tree of the original metric

Many recent network design approximation algorithms use probabilistic approximations by tree metrics, e.g., the buy-at-bulk network design problems [3], the group Steiner tree problem [6], and the communication cost spanning tree problem [16]. The key idea is that the expected cost of an optimal solution in the tree metrics is at most  $O(\log^2 n)$  times the optimum and

furthermore, it is easy to find the minimum or near-minimum cost network in a tree metric. To transform the solution in the tree metric to one in the original metric, it is convenient if the tree is a spanning tree of the original metric, so the solution edges form a subforest of this spanning tree. However, this is not the case in the  $k$ -HST construction due to the addition of new internal nodes.

We show that HSTs can be made spanning trees of the original metric (not necessarily of the original graph) while doubling the expected distortion. This is similar to a lifting procedure of Jiang, Lawler and Wang [8]. The resulting spanning trees are not guaranteed to be hierarchically well-separated, but this is not a concern in the mentioned network design applications.

**Theorem 5.1.** *Any  $k$ -HST  $T'$  resulting from weak or strong probabilistic partitions as described in the proof of Theorem 2.7 can be replaced by a tree  $T$  whose vertex set is  $V(G)$ , such that*

$$d_G(u, v) \leq d_T(u, v) \leq 2d_{T'}(u, v)k/(k-1)$$

for any  $u, v \in V(G)$ .

**Proof.** By construction of  $T'$  described in the proof of Theorem 2.7, the set of leaves of  $T'$  is exactly the set of vertices of  $G$ . Also,  $d_G(u, v) \leq d_{T'}(u, v)$ .

Let  $v$  be an arbitrary vertex of  $G$  in  $T'$  such that the parent  $w$  of  $v$  is not a vertex of  $G$ . Shrink the edge  $wv$  and identify  $v$  with  $w$ . Repeat this until all the vertices of the tree are vertices of the original graph  $G$ . Finally, multiply the length of every edge by  $2k/(k-1)$ . Let  $T$  denote the resulting tree.

Clearly,  $d_T(u, v) \leq 2d_{T'}(u, v)k/(k-1)$ , as the distance could only have decreased in the shrinking process. To prove the lower bound, let  $u$  and  $v$  be leaves of  $T'$  and  $w$  be their least common ancestor. Let  $l_w$  be the length of the edges from  $w$  to its children. From the definition of a  $k$ -HST, it follows that  $d_{T'}(u, v) \leq 2l_wk/(k-1)$ . Thus,  $d_T(u, v) \geq 2l_wk/(k-1) \geq d_{T'}(u, v) \geq d_G(u, v)$ .  $\square$

Note that at the end of the shrinking algorithm, a spanning tree of the metric is obtained and we may replace the updated edge weights by the original weights in the metric without increasing the distortion.

## References

- [1] R. Agarwala, V. Bafna, M. Farach, B. Narayanan, M. Paterson, M. Thorup, On the approximability of numerical taxonomy, in: Proc. 7th Annual ACM-SIAM Symposium on Discrete Algorithms, 1996, pp. 365–372.
- [2] N. Alon, R.M. Karp, D. Peleg, D. West, A graph-theoretic game and its application to the  $k$ -server problem, *SIAM J. Comput.* 24 (1) (1995) 78–100.
- [3] B. Awerbuch, Y. Azar, Buy-at-bulk network design, in: Proc. 38th Annual IEEE Symposium on Foundations of Computer Science, 1997, pp. 542–547.
- [4] Y. Bartal, Probabilistic approximation of metric spaces and its algorithmic applications, in: Proc. 37th Annual IEEE Symposium on Foundations of Computer Science, October 1996, pp. 184–193.
- [5] Y. Bartal, On approximating arbitrary metrics by tree metrics, in: Proc. 30th Annual ACM Symposium on Theory of Computing, 1998, pp. 161–168.
- [6] N. Garg, G. Konjevod, R. Ravi, A polylogarithmic approximation algorithm for the group Steiner tree problem, *J. Algorithms* 37 (2000) 66–84.
- [7] R.L. Graham, P.M. Winkler, On isometric embeddings of graphs, *Trans. Amer. Math. Soc.* 288 (1985) 527–536.
- [8] T. Jiang, L. Lawler, L. Wang, Aligning sequences via an evolution: complexity and approximation, in: Proc. 26th Annual ACM Symposium on Theory of Computing, May 1994, pp. 760–769.
- [9] W.B. Johnson, J. Lindenstrauss, Extensions of Lipschitz mappings into a Hilbert space, in: *Contemp. Math.*, Vol. 26, Amer. Math. Soc., Providence, RI, 1984, pp. 189–206.
- [10] R.M. Karp, A  $2k$ -competitive algorithm for the circle, Manuscript, August 1989.
- [11] P. Klein, S.A. Plotkin, S. Rao, Excluded minors, network decomposition and multicommodity flow, in: Proc. 25th Annual ACM Symposium on Theory of Computing, May 1993, pp. 682–690.
- [12] N. Linial, E. London, Y. Rabinovich, The geometry of graphs and some of its algorithmic applications, *Combinatorica* 15 (2) (1995) 215–245.
- [13] R. Motwani, P. Raghavan, *Randomized Algorithms*, Cambridge University Press, Cambridge, 1995.
- [14] D. Peleg, A.A. Schaeffer, Graph spanners, *J. Graph Theory* 13 (1) (1989) 99–106.
- [15] Y. Rabinovich, R. Raz, Lower bounds on the distortion of embedding finite metric spaces in graphs, *Discrete Comput. Geometry* 19 (1995) 79–94.
- [16] B.Y. Wu, G. Lancia, V. Bafna, K.-M. Chao, R. Ravi, C.Y. Tang, A polynomial-time approximation scheme for minimum routing cost spanning trees, in: Proc. 9th Annual ACM-SIAM Symposium on Discrete Algorithms, 1997, pp. 21–32.
- [17] A.C.-C. Yao, Probabilistic computations: Towards a unified measure of complexity, in: Proc. 18th Annual IEEE Symposium on Foundations of Computer Science, 1977, pp. 222–227.