On the Hardness of Embeddings Between Two Finite Metrics

Matthew Cary, Atri Rudra, and Ashish Sabharwal

Computer Science and Engineering
University of Washington, Box 352350
Seattle, WA 98195-2350, U.S.A.
{cary, atri, ashish}@cs.washington.edu

Abstract. We improve hardness results for the problem of embedding one finite metric into another with minimum distortion. This problem is equivalent to optimally embedding one weighted graph into another under the shortest path metric. We show that unless $\mathsf{P} = \mathsf{NP}$, the minimum distortion of embedding one such graph into another cannot be efficiently approximated within a factor less than $9/4$ even when the two graphs are unweighted trees. For weighted trees with the ratio of maximum edge weight to the minimum edge weight of α^2 ($\alpha \geq 1$) and all but one node of constant degree, we improve this factor to $1+\alpha$. We also obtain similar hardness results for extremely simple line graphs (weighted). This improves and complements recent results of Kenyon et al. [13] and Papadimitriou and Safra [18].

1 Introduction

For two n-point metric spaces (X, ρ) and (Y, σ), the *expansion* of a bijection $\varphi : X \to Y$ is defined as $exp(\varphi) = \max_{a \neq b \in X} \frac{\sigma(\varphi(a), \varphi(b))}{\rho(a,b)}$. The *distortion* of φ, denoted $dist(\varphi)$, is the product of $exp(\varphi)$ and $exp(\varphi^{-1})$. The expansion of φ^{-1} is also referred to as the *contraction* of φ and denoted $con(\varphi)$. The *distortion between X and Y*, denoted $dist(X, Y)$, is the minimum distortion over all such bijections and may be thought of as a difference measure between these metric spaces. This paper addresses the computational hardness of the problem of embedding one finite metric space into another with minimum distortion.

The notion of distortion was originally studied for infinite metrics [12] in the analysis of Banach spaces. More recently the embedding of finite metrics into Euclidean and other L_p metrics has been very successful for applications in theoretical computer science, including approximation, learning, on-line algorithms, high-dimensional geometry, and others [6,17,16,11]. This notion has been extended in such directions as embedding a finite metric into a distribution of metrics which has again found great success in approximation algorithms [1,8]. This continues to be an active area of research [2,15].

L. Caires et al. (Eds): ICALP 2005, LNCS 3580, pp. 1412-1423, 2005.

We point out that the problems addressed in the works mentioned above are combinatorial in nature– that is, they are concerned with embedding a finite metric into another *class* of metrics and the focus is on providing bounds for the distortion itself. However, we are interested in the algorithmic problem of embedding a *specific* metric into another *specific* metric– i.e. we are interested in the worst case ratio of the distortion obtained by the algorithm under consideration and the best possible distortion. This problem was introduced by Kenyon et al. [13]. The recent work of Bădoiu et. al. [3] considers the algorithmic question of finding embeddings of a specific metric into a class of metrics.

In addition to the fact that the problem of finding low-distortion embeddings between two finite metrics is a very natural question that by itself merits investigation, the problem is also likely to have much wider use than theoretical computer science. To mention three examples, theorem proving and symbolic computation [20], database problems such as queries over heterogeneous structured databases [21], and matching gels from electrophoresis [10] can all be expressed as tree embedding problems. The problem has several other applications as well [13].

We note a basic fact that any n-point metric may be realized as the shortest path metric of a weighted undirected graph over n nodes, for example by making a complete graph whose adjacency matrix is the matrix of metric distances. Due to this correspondence, we will exclusively focus on the problem of optimally embedding one graph into another. We will implicitly identify a graph with the metric given by shortest paths on that graph. For a set of weighted graphs, their *weight ratio* is the ratio of the maximum to the minimum weights of edges in the graphs.

1.1 Previous Results

The only upper bounds on this problem known to us are by Kenyon et al. [13]. Given two point sets on the real line with the L_1 distance metric that have distortion less than $3 + 2\sqrt{2}$, there is a polynomial time algorithm to find an embedding with the minimum distortion. Their second result finds the minimum distortion between an arbitrary graph and a tree, in polynomial time if the degree of the tree and the distortion are constant. Their algorithm is exponential in the degree of the tree and doubly-exponential in the distortion. Both algorithms are based on dynamic programming; the latter is similar to those based on tree decompositions of graphs.

The situation for hardness results is a little more clear. Determining if there is an isometry—a distortion 1 embedding—between two graphs is the graph isomorphism problem, which is not known to be in P but which is probably not NP-hard either. Kenyon et al. [13] show the problem is NP-hard to approximate within a factor of 2 for general graphs and a factor of 4/3 in the case where one of the graphs is an unweighted tree and the other is a weighted graph with weights 1/2 or 1. Papadimitriou and Safra [18] show that it is NP-hard to approximate within a factor of 3 the distortion between any two finite metrics realized as point sets in $\mathbb{R}^3$ where the distance metric is the L_2 norm.

1.2 Our Results

Unweighted Trees (Section 3.3) The problem is NP-hard to approximate within a factor less than 9/4 for unweighted trees. As far as we know, this is the first hardness result for embedding an unweighted graph into another. It also improves the factor of 2 result for general graphs [13] even when the graphs are unweighted.

Weighted Trees (Section 3.2) The problem is NP-hard to approximate within a factor less than $1 + \alpha$ for any $\alpha \geq 1$ and tree graphs with weight ratio $\Omega(\alpha^2)$. This is the first hardness result for embedding trees into trees and improves the bound of 2 for general graphs [13] at the expense of a larger weight ratio. Our result also holds when all but one node of the underlying graphs have degree ≤ 4; the problem is known to be easy in the unweighted case when all nodes have constant degree and the distortion is small [13]. This result also improves the bound of 3 by Papadimitriou and Safra [18].

Weighted Line Graphs (Section 4) The problem is NP-hard to approximate within a factor of α for any $\alpha > 1$ and line graphs with weight ratio $\Omega(\alpha^2 n^4)$, where n is the number of nodes in the two graphs. This is the only bound known for graphs with constant degrees and large weights.

2 Preliminaries

We begin with some basic properties of the distortion resulting from embedding a weighted undirected graph G into another such graph H. Let $[m]$ denote the set of integers from 1 to m. Let d_G and d_H denote the shortest path distances in G and H, respectively. Fix a bijection $\varphi : G \to H$. We state the following results for $exp(\varphi)$. Analogous results hold for $con(\varphi)$ which is nothing but $exp(\varphi^{-1})$.

Lemma 1 ([13]). *φ achieves its maximum expansion at some edge in G, i.e., $exp(\varphi) = \max_{\{a,b\} \in E(G)} \frac{d_H(\varphi(a),\varphi(b))}{d_G(a,b)}$.*

Corollary 1. *If G and H are unweighted then $exp(\varphi)$ is an integer.*

Lemma 2. *If G and H are unweighted and H has no edge-subgraph that is isomorphic to G then $exp(\varphi) \geq 2$.*

Proof. Let u and v be nodes of G such that $(u, v) \in E(G)$ but $(\varphi(u), \varphi(v)) \notin E(H)$. Such nodes must exist because H has no edge-subgraph isomorphic to G. $d_G(u, v) = 1$ and $d_H(\varphi(u), \varphi(v)) \geq 2$, implying an expansion of at least 2. □

We now state the problem we use in the reductions for our NP-hardness proofs. It is a generalization of the Hamiltonian cycle problem [9]. Let $\mathcal{G} = (V, E)$ be a directed graph over n vertices. $\mathcal{G}$ has a *disjoint cycle cover* if there is a collection of vertex-disjoint cycles in $\mathcal{G}$ that contain every node in V, i.e., there exists a permutation $\sigma : [n] \to [n]$ such that for all $i \in [n]$, $(v_i, v_{\sigma(i)}) \in E$. $\mathcal{G}$ has a *loose* disjoint cycle cover if it has a disjoint cycle cover after adding two arbitrarily chosen edges to E.

The *loose directed disjoint cycle cover testing* problem is a property testing problem defined as follows. Given a directed graph $\mathcal{G}$, output 1 if $\mathcal{G}$ has a disjoint cycle cover and 0 if $\mathcal{G}$ does not even have a loose disjoint cycle cover. Note that in the remaining scenario, one is allowed to output anything.

Lemma 3. *The loose directed disjoint cycle cover testing problem is* NP*-hard for graphs with indegree* ≤ 4 *and outdegree* $= 3$.

Proof. This can be shown by an extension of the ideas used in the NP-completeness proof of the directed disjoint cycle cover problem in an earlier paper by the authors [5] using in addition the fact that the Vertex Cover problem is hard to approximate [7]. We omit the details. □

Finally, we mention a combinatorial result about sum-free sequences that is used in one of our constructions. A sequence of n integers is *k-way sum-free* if all n^k sums of k integers (not necessarily distinct) in it are distinct. Khanna et al. [14] suggest a greedy algorithm to construct 3-way sum-free sequences. Their result can be generalized to the following.

Lemma 4. *There exists a strictly increasing sequence of size n in $[n^{2k-1}]$ that is k-way sum-free and is computable in time $O(n^{2k-1})$.*

3 Hardness of Embeddings between Tree Graphs

Consider the problem of finding a minimum distortion embedding between two given undirected tree graphs. We give reductions from the loose directed disjoint cycle cover testing problem to the decision version of this embedding problem on weighted as well as unweighted trees. The result for the weighted case holds even for graphs with all but one node of degree at most 4. We begin with a general construction that will be used in both reductions.

Given a directed graph $\mathcal{G}$ with outdegree $= 3$ and indegree ≤ 4, we will construct a source tree $\mathcal{S}$ and a destination tree $\mathcal{D}$ with the property that there exist $0 < a < b$ such that

1. if $\mathcal{G}$ has a disjoint cycle cover then $dist(\mathcal{S}, \mathcal{D}) \leq a$, and
2. if $\mathcal{G}$ has no *loose* disjoint cycle cover then $dist(\mathcal{S}, \mathcal{D}) \geq b$.

It follows from Lemma 3 that it is NP-hard to approximate $dist(\mathcal{S}, \mathcal{D})$ within a factor less than b/a.

3.1 The Construction

We describe in this section the construction of $\mathcal{S}$ and $\mathcal{D}$ from $\mathcal{G}$. Let $\mathbb{Z}^+$ denote the set of positive integers and $s : \mathbb{Z}^+ \rightarrow \mathbb{Z}^+$ be a strictly increasing monotonic function. Let $v_1, \ldots, v_n$ be the vertices of $\mathcal{G}$.

We will need two types of gadgets, a *center gadget* and for each $i \in [n]$, a *size gadget* T_i. The center gadget is a rooted tree consisting of n leaves, all at depth 1. All edges in this gadget have weight $y \in \mathbb{Z}^+$. Its root is denoted by c_r and leaves by c_ℓ. The size gadget T_i is a rooted tree consisting of $s(i)$ leaves, all at depth 1. All edges in T_i have weight 1. The root of T_i is denoted by g_r and the leaves by g_ℓ.

The source tree $\mathcal{S}$ is constructed as follows (see Fig. 1). Start with a copy of the center gadget and associate with each c_ℓ node of it a distinct vertex v_i of $\mathcal{G}$. For any $i \in [n]$, let the successors of v_i in $\mathcal{G}$ be the vertices v_{i_1}, v_{i_2}, and v_{i_3}. Attach to the c_ℓ node corresponding to v_i copies of the three size gadgets T_{i_1}, T_{i_2}, and T_{i_3} by adding edges with weight $x \in \mathbb{Z}^+$ to the g_r nodes of these gadgets. Copies of any size gadget T_i in $\mathcal{S}$ will henceforth be denoted by S_i.

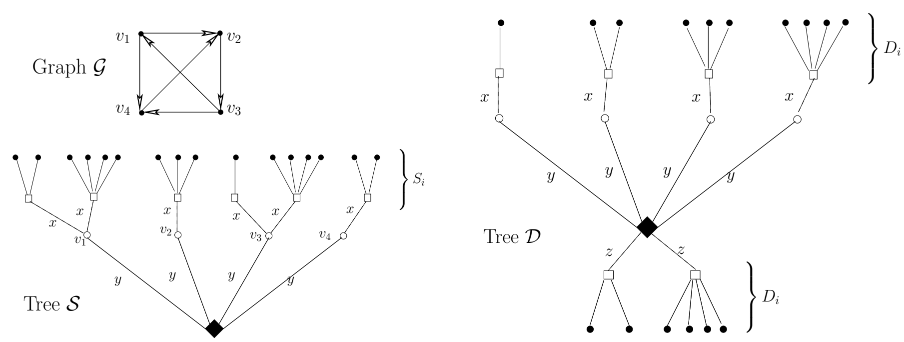

Fig. 1. A directed graph and the source and destination trees corresponding to it. For simplicity of depiction, $s(i) = i$. Unmarked edges have a weight of 1.

The destination tree $\mathcal{D}$ is constructed similarly. As before, start with a copy of the center gadget. Fix an arbitrary ordering of its c_ℓ nodes. For all $i \in [n]$, attach to the i^{th} c_ℓ node a copy of the size gadget T_i by adding an edge of weight x to its g_r node. These n size gadgets are called *non-spare* size gadgets. Now let $\mathcal{P}$ be the multi-set $\{i \mid \text{gadget } T_i \text{ is used in } \mathcal{S}\}$. We may assume that $\mathcal{P} \supseteq [n]$, otherwise a disjoint cycle cover cannot exist. For each $i \in \mathcal{P} \setminus [n]$, attach a copy of the size gadget T_i directly to the c_r node by adding edges of weight $z \in \mathbb{Z}^+$ to their g_r node. These are called *spare* size gadgets. Copies of any size gadget T_i in $\mathcal{D}$ will henceforth be denoted by D_i.

Note that both $\mathcal{S}$ and $\mathcal{D}$ have the same number of nodes and for every $i \in [n]$, the same number of copies of the size gadget T_i. Further, $\mathcal{S}$ and $\mathcal{D}$ each have exactly one c_r node, n c_ℓ nodes, and $3n$ g_ℓ nodes (recall the outdegree of every vertex of $\mathcal{G}$ is 3). Consider a mapping φ from $\mathcal{S}$ to $\mathcal{D}$. Let A and B be sets of nodes in $\mathcal{S}$ and $\mathcal{D}$, respectively. φ *fully* maps A to B if $\{\varphi(u) \mid u \in A\} = B$. φ maps A *exactly* to B if A and B are size gadgets with g_r nodes a and b, respectively, φ fully maps A to B, and $\varphi(a) = b$.

The basic idea of the construction is that $\mathcal{S}$ encodes the input graph $\mathcal{G}$ while $\mathcal{D}$ is setup so that the relationships between the c_ℓ nodes and the non-spare size gadgets induce (via a low distortion embedding) a permutation on the vertices of $\mathcal{G}$. This construction balances two conflicting desires. On one hand, it must be possible to match unused size gadgets to the spare gadgets with small distortion when a disjoint cycle cover exists. Thus, the spare gadgets cannot be too far from the successor-selection part $\mathcal{D}$. On the other hand, a node corresponding to a vertex in $\mathcal{G}$ must be far enough from size gadgets not corresponding to its own successors so that choosing a predecessor incorrectly gives large distortion.

Lemma 5. *If $\mathcal{G}$ has a disjoint cycle cover then $dist(\mathcal{S},\mathcal{D}) \leq (y+z)(x+y)/(xz)$.*

Proof. As $\mathcal{G}$ has a disjoint cycle cover, there is a permutation $\sigma : [n] \to [n]$ such that for all $i \in [n]$, $(i, \sigma(i))$ is an edge in $\mathcal{G}$. We construct a small distortion embedding φ of $\mathcal{S}$ into $\mathcal{D}$. Consider any $i \in [n]$. By the definition of σ, an S_i gadget A is attached to the c_ℓ node u corresponding to $v_{\sigma(i)}$ in $\mathcal{S}$. Let φ map A exactly to the non-spare D_i gadget B of $\mathcal{D}$ and u to the c_ℓ node attached to B. This leaves $2n$ size gadgets of $\mathcal{S}$ not yet mapped. Map each of these exactly to spare size gadgets of $\mathcal{D}$. Finally, let φ map the c_r node of $\mathcal{S}$ to the c_r node of $\mathcal{D}$.

We claim that $exp(\varphi) = (y+z)/x$. By Lemma 1, we only need to consider the expansion of the edges of $\mathcal{S}$. The (g_r, g_ℓ) and (c_r, c_ℓ) edges in $\mathcal{S}$ have an expansion of 1. A (g_r, c_ℓ) edge in $\mathcal{S}$ has an expansion of 1 if the corresponding S_i gadget is mapped to a non-spare D_i gadget and $(y+z)/x$ otherwise. This proves the claim. We further claim that $exp(\varphi^{-1}) = (x+y)/x$. Again using Lemma 1, the only edges in $\mathcal{D}$ that have expansion different from 1 are the (c_r, g_r) edges in $\mathcal{D}$ that give an expansion of $(x+y)/z$. This completes the proof. □

Let φ be any embedding of $\mathcal{S}$ into $\mathcal{G}$. Since both $\mathcal{S}$ and $\mathcal{D}$ contain edges of weight 1 and all edge weights are in $\mathbb{Z}^+$, we have the following.

Proposition 1. *$exp(\varphi) \geq 1$ and $con(\varphi) \geq 1$.*

Lemma 6. *If $\mathcal{G}$ has no disjoint cycle cover and φ fully maps every non-spare D_i gadget from an S_i gadget and c_ℓ nodes from c_ℓ nodes, then both $exp(\varphi)$ and $con(\varphi)$ are at least $1 + 2y/x$.*

Proof. For $i \in [n]$, consider the S_i gadget A_i that maps to the non-spare D_i gadget B_i of $\mathcal{D}$. Let A_i be attached to the c_ℓ node u_j of $\mathcal{S}$ corresponding to vertex v_j of $\mathcal{G}$. Let B_i be attached to the c_ℓ node w_i of $\mathcal{D}$. If u_j maps to w_i and $(v_j, v_i) \in E(\mathcal{G})$, think of vertex v_i being chosen as the successor of vertex v_j in $\mathcal{G}$. Since $\mathcal{G}$ does not have a disjoint cycle cover, there must exist $i \in [n]$ such that u_j, as defined above, does not map to w_i. Fix such i and j. Let $\varphi(u_j) = w_{k_1}$ and $\varphi(u_{k_2}) = w_i$, where $k_1 \neq i$ and $k_2 \neq j$. Let r be the g_r node of A_i and r' be that of B_i. The edge (u_j, r) in $\mathcal{S}$ gives an expansion of at least $(x+2y)/x = 1+2y/x$ because φ maps u_j to w_{k_1} and r to a node within B_i. Similarly, the edge (w_i, r') in $\mathcal{D}$ gives a contraction of at least $1+2y/x$ because φ^{-1} maps w_i to u_{k_2} and r' to a node within A_i. □

Lemma 7. *If $\mathcal{G}$ has no loose disjoint cycle cover and φ fully maps every S_i gadgets to a D_i gadget, then both $exp(\varphi)$ and $con(\varphi)$ are at least $1 + 2y/x$.*

Proof. Since every S_i gadget fully maps to a D_i gadget, the center gadget of $\mathcal{S}$ fully maps to the center gadget of $\mathcal{D}$. We first consider the case when φ maps the c_r node of $\mathcal{S}$ to the c_r node in $\mathcal{D}$. Every c_ℓ node of $\mathcal{S}$ must then map to a c_ℓ node of $\mathcal{D}$ and Lemma 6 completes the proof.

Now suppose that φ maps the c_r node of $\mathcal{S}$ to a c_ℓ node w_i of $\mathcal{D}$. As all gadgets are fully mapped, there is a c_ℓ node u_j of $\mathcal{S}$ corresponding to vertex v_j of $\mathcal{G}$ be mapped to the c_r node of $\mathcal{D}$. Let B_i be the D_i gadget attached to w_i. From the arguments we made above, it follows that if we want at least one of $exp(\varphi)$ and $con(\varphi)$ to be strictly less than $1 + 2y/x$, then only one of two things can happen. First, a size gadget A_i in $\mathcal{S}$ that does not correspond to a successor of v_j is mapped to B_i and every other size gadget maps correctly w.r.t. the successor relationship in $\mathcal{G}$. In this case, $exp(\varphi) \geq 1+2y/x$ while $con(\varphi)$ may be at most $1 + y/x$. However, if A_i corresponds to vertex v_i, by adding the edge (v_j, v_i), we have a disjoint cycle cover, contradicting the absence of a *loose* cycle cover. Second, B_i and at most two other non-spare size gadgets B_k and B_ℓ in $\mathcal{D}$ are mapped from size gadgets in $\mathcal{S}$ that correspond to successors v_i, v_k and v_ℓ of v_j, and every other size gadget maps correctly w.r.t. the successor relationship in $\mathcal{G}$. In this case, $con(\varphi) \geq 1 + 2y/x$ while $exp(\varphi)$ may be at most $1 + y/x$. The successor of v_j is well-defined in this case, but v_k and v_ℓ may not be successors of the c_ℓ nodes in $\mathcal{S}$ mapped to the c_ℓ nodes of B_k and B_ℓ. If those nodes are v_s and v_t, by adding edges (v_s, v_k) and (v_t, v_ℓ), we have a disjoint cycle cover, again contradicting the absence of a loose cover. □

3.2 Hardness for Weighted Trees

We first consider general weighted trees with unbounded degree and then modify the reduction so that exactly one node in both $\mathcal{S}$ and $\mathcal{D}$ has non-constant degree. Let φ be an embedding of $\mathcal{S}$ into $\mathcal{D}$. We begin by showing that for suitably weighted $\mathcal{S}$ and $\mathcal{D}$, the distortion is large if φ does not map size gadgets correctly.

Lemma 8. *If $s(1) > n$ and φ does not fully map every S_i gadget to a D_i gadget, then $dist(\varphi) \geq x \cdot \min\{x, z\}$.*

Proof. Suppose $exp(\varphi) < \min\{x, z\}$. For $i \in [n]$, $s(i) \geq s(1) > n$. Since the center gadgets have only $n + 1$ nodes, every size gadget in $\mathcal{S}$ must have at least one node that φ maps to a size gadget in $\mathcal{D}$. Recall that all edges within size gadgets in $\mathcal{S}$ have weight 1 while every edge going out of size gadgets in $\mathcal{D}$ has weight $\min\{x, z\}$. To keep $exp(\varphi) < \min\{x, z\}$, every node of any size gadget in $\mathcal{S}$ must map within a single size gadget in $\mathcal{D}$. Since for all $i \in [n]$, $\mathcal{S}$ and $\mathcal{D}$ have the same number of S_i and D_i gadgets, respectively, this can happen only if every S_i gadget fully maps to a D_i gadget. A similar argument shows that $exp(\varphi^{-1}) < x$ only if every D_i gadget fully maps to an S_i gadget. □

Theorem 1. *For $\alpha \geq 1$, it is* NP*-hard to approximate the distortion between two trees with weight ratio $O(\alpha^2)$ within a factor less than $1 + \alpha$.*

Proof. Let $\mathcal{G}$, $\mathcal{S}$, and $\mathcal{D}$ be as in Section 3.1 with $x = \alpha + 1$, $y = \alpha(\alpha + 1)/2$, $z = x + y = (\alpha + 1)(\alpha + 2)/2$, and $s(i) = i + n$ for $i \in [n]$. The weight ratio of $\{\mathcal{S}, \mathcal{D}\}$ is $(\alpha + 1)(\alpha + 2)/2$. If $\mathcal{G}$ has a disjoint cycle cover then by Lemma 5 $dist(\mathcal{S}, \mathcal{D}) \leq 1 + 2y/x = 1 + \alpha$. If $\mathcal{G}$ does not have a loose disjoint cycle cover then by Lemmas 7 and 8, $dist(\mathcal{S}, \mathcal{D}) \geq \min\{x \cdot \min\{x, z\}, (1 + 2y/x)^2\}$, which is $(1 + \alpha)^2$. The result follows from Lemma 3. □

Let N be the number of nodes in $\mathcal{S}$ (and $\mathcal{D}$). In the above construction, $N = \Theta(n^2)$. The c_r nodes of $\mathcal{S}$ and $\mathcal{D}$ have degrees n and $3n$, respectively, which is $\Theta(\sqrt{N})$. The c_ℓ nodes have degrees 4 and 2, respectively. The g_r nodes have degrees between $n = \Theta(\sqrt{N})$ and $2n$, while the g_ℓ nodes have degree 1. By replacing each S_i and D_i gadget with a line graph, we can show the following theorem. The complete proof may be found in the full version of the paper [4].

Theorem 2. *For $0 < \epsilon \leq 1/2$ and $\alpha \geq 1$, it is* NP*-Hard to approximate the distortion between two trees with N nodes, weight ratio $\Omega(\alpha^2)$, exactly one node of degree $\Theta(N^\epsilon)$, and all other nodes of degree ≤ 4 within a factor less than $1 + \alpha$.*

3.3 Hardness for Unweighted Trees

The construction from Section 3.1 needs slight modification in order to obtain hardness results for the unweighted case. Let $\mathcal{G}$, $\mathcal{S}$, and $\mathcal{D}$ be as in Section 3.1 with $x = y = z = 1$ and $s(i) = 2c \cdot \big(f(i) + 2n^5\big)$, where $c = 4n + 2$ and f is a strictly increasing 3-way sum-free sequence of size n in $[n^5]$ guaranteed by Lemma 4. These parameters imply six useful properties of s, namely, $s(\cdot)$ is even, $s(\cdot)$ is a multiple of c, $2s(1) \gg s(n)$, $2s(n) < 3s(1)$, $|s(i) - s(j)|$ is large for $i \neq j$, and $s(1), s(2), \ldots, s(n)$ is a strictly increasing 3-way sum-free sequence. Furthermore, we have that $c > |\text{Edges}(\mathcal{G})| = 3n$. We will repeatedly use the fact that $\mathcal{S}$ and $\mathcal{D}$ each have $n + 1$ center gadget nodes and $3n$ g_r nodes.

The only change to the construction is to modify the non-spare size gadgets in $\mathcal{D}$. Instead of being depth one trees with $s(i)$ leaves, they are now depth two trees with $s(i)/2$ nodes at depth one, each of which has a single depth two leaf. The root and depth one nodes are denoted by g_r and g_ℓ as before, the depth two leaves are denoted by g'_ℓ, and the depth one and two nodes are together denoted by $\overline{g_\ell}$. All other notation is unchanged. Like the original construction, both $\mathcal{S}$ and $\mathcal{D}$ have the same number of nodes and for each S_i gadget there is a corresponding D_i gadget with the same number of nodes.

Let φ be any embedding of $\mathcal{S}$ into $\mathcal{D}$. We will prove the following lemmas in the rest of this section using Propositions 2 and 3, respectively.

Lemma 9. *If $\mathcal{G}$ has no disjoint cycle cover and φ does not fully map every S_i gadget to a D_i gadget, then $exp(\varphi) \geq 3$.*

Lemma 10. *If $\mathcal{G}$ has no disjoint cycle cover and φ does not fully map every S_i gadget to a D_i gadget, then either $con(\varphi) \geq 3$ or $exp(\varphi) \geq 5$.*

Theorem 3. *It is* NP*-Hard to approximate the distortion between two unweighted trees within a factor less than* $9/4$.

Proof. If $\mathcal{G}$ has a disjoint cycle cover then by an argument similar to Lemma 5, $dist(\mathcal{S}, \mathcal{D}) \leq 4$. Assume that $\mathcal{G}$ does not have a loose disjoint cycle cover (and hence no disjoint cycle cover either). If φ fully maps every S_i gadget to a D_i gadget then by an argument similar to Lemma 7, $dist(\varphi) \geq 9$. If it does not then Lemmas 2, 9, and 10 imply $dist(\varphi) \geq 9$. The result follows from Lemma 3. □

Proposition 2. *If any of the following fail,* $exp(\varphi) \geq 3$.

1. *No size gadget in* $\mathcal{S}$ *maps to the* $\overline{g_\ell}$ *nodes of two distinct size gadgets in* $\mathcal{D}$.
2. *Nodes of no two size gadgets in* $\mathcal{S}$ *are both mapped to the* $\overline{g_\ell}$ *nodes of a single size gadget in* $\mathcal{D}$.
3. *No node of an* S_i *gadget maps to a* $\overline{g_\ell}$ *node of a* D_j *gadget for* $j \neq i$.
4. *The* g_r *node of any* S_i *gadget* A *maps within the unique* D_i *gadget* B *whose* $\overline{g_\ell}$ *nodes* A *maps to, or possibly to the* c_r *node of* $\mathcal{D}$ *if* B *is a spare gadget.*
5. *The* c_r *node of* $\mathcal{S}$ *is not mapped to a non-spare gadget or the* g_ℓ *nodes of a spare gadget in* $\mathcal{D}$.
6. *No* c_ℓ *node in* $\mathcal{S}$ *is mapped to a non-spare gadget or the* g_ℓ *nodes of a spare gadget in* $\mathcal{D}$.

Proof. Suppose (1) fails and a size gadget A in $\mathcal{S}$ maps to the $\overline{g_\ell}$ nodes of two distinct size gadgets B and C in $\mathcal{D}$. Any $\overline{g_\ell}$ node of B is at least distance 5 away from any $\overline{g_\ell}$ node of C, while all nodes in A are within distance 2 of each other. Hence, $exp(\varphi) \geq 5/2$. By Corollary 1, $exp(\varphi) \geq 3$.

Suppose (2) fails with an S_i gadget A and an S_k gadget C mapping to the $\overline{g_\ell}$ nodes of a single D_j gadget B. A and C together have at least $s(i)+s(k)-s(j) \geq 2s(1) - s(n) \geq 2c(2+n^5)$ nodes mapped outside B. Since there are only $4n+1$ non-$\overline{g_\ell}$ nodes in $\mathcal{D}$ ($n+1$ in the center gadget and $3n$ g_r nodes), a node of A or C must map to a $\overline{g_\ell}$ node of a size gadget in $\mathcal{D}$ other than C. This violates (1).

The proofs of the remaining cases are similar and may be found in [4]. □

Proof of Lemma 9. By Proposition 2 (1), (4), (5), and (6), no node other than that of a unique S_i gadget can be mapped to any non-spare D_i gadget or the g_ℓ nodes of a spare D_i gadget. It follows that all non-spare gadgets are fully mapped. We further claim that all c_ℓ nodes of $\mathcal{S}$ are mapped to c_ℓ nodes of $\mathcal{D}$, in which case the proof is complete by Lemma 6. The claim holds because of the following. Observe that since all non-spare gadgets are fully mapped, all c_ℓ nodes of $\mathcal{S}$ must map within the center gadget of $\mathcal{D}$ to ensure $exp(\varphi) \leq 2$. Further, by the assumption in the lemma, at least one spare gadget B is partially mapped from a gadget A in $\mathcal{S}$. By (4), the g_r node of A must map to the c_r node of $\mathcal{D}$, making the latter unavailable for the c_ℓ nodes of $\mathcal{S}$. □

We begin the contraction argument by stating a straightforward but crucial property of the g'_ℓ nodes of the size gadgets in $\mathcal{D}$.

Observation 1 *If a g'_ℓ node of a size gadget B in $\mathcal{D}$ does not have as its image under φ^{-1} in $\mathcal{S}$ a node with neighbors only those nodes that are images of nodes of B, the c_ℓ node attached to B, or the c_r node of $\mathcal{D}$, then $exp(\varphi) \geq 5$.*

Define the *successor cluster* X corresponding to a vertex v of $\mathcal{G}$ to be the c_ℓ node u of $\mathcal{S}$ corresponding to v and the three size gadgets A_{i_X}, A_{j_X}, and A_{k_X} attached to it. Let $Q_X^\varphi \subseteq \{1, \ldots, n\}$ be the *multi*-set defined by $Q_X^\varphi = \{r \mid$ some non-g'_ℓ node of a D_r gadget maps under φ^{-1} to a non-c_ℓ node of $X\}$. The multiplicity of r in Q_X^φ is the number of D_r's that map in this way to X. Since the number of center gadget nodes in $\mathcal{D}$ is only $n+1$, $s_X^\varphi = \sum_{r \in Q_X^\varphi} s(r)$ can be less than $s_X = s(i_X) + s(j_X) + s(k_X)$ by at most $n+1$. However, since s_X^φ and s_X are both multiples of $c > n+1$, $s_X^\varphi \geq s_X$.

Proposition 3. *If any of the following fail, $con(\varphi) \geq 3$ or $exp(\varphi) \geq 5$.*

1. $Q_X^\varphi = \{i_X, j_X, k_X\}$.
2. *The g_r node of any D_i gadget B is mapped within the unique successor cluster X to which B's non-g'_ℓ nodes map.*
3. *The c_r node of $\mathcal{D}$ maps to the c_r node of $\mathcal{S}$.*
4. *The g_ℓ nodes of $\mathcal{S}$ are occupied only by the size gadget nodes of $\mathcal{D}$.*
5. *If a c_ℓ node of $\mathcal{D}$ is mapped to a node of a successor cluster X, then nodes from exactly three size gadgets of $\mathcal{D}$ map into X. (X may have other c_ℓ nodes of $\mathcal{D}$ mapped into it as well.)*
6. *If a c_ℓ node of $\mathcal{D}$ is mapped to the c_ℓ node of a successor cluster X, then three size gadgets of $\mathcal{D}$ fully map to the non-c_ℓ nodes of X.*
7. *If no c_ℓ node of $\mathcal{D}$ is mapped to a node of a successor cluster X, then nodes from exactly three size gadgets of $\mathcal{D}$ map into X and the c_ℓ node of X is occupied by a node from a fourth size gadget of $\mathcal{D}$.*
8. *Every successor cluster in $\mathcal{S}$ is fully mapped from exactly one c_ℓ node and three size gadgets of $\mathcal{D}$.*
9. *If a c_ℓ node v in $\mathcal{D}$ is mapped to a successor cluster X, then the root r of the size gadget B attached to v is mapped to X.*

Proof. Unless mentioned otherwise the mapping under consideration in this proof is φ^{-1}. We refer the reader to the full version [4] for most cases of the proof.

If (5) fails, let v be a c_ℓ node of $\mathcal{D}$ that is mapped to a node u of X. By (4), u is either a c_ℓ node or a g_r node. Suppose first that it is a c_ℓ node. Since (5) fails, there is a size gadget B in $\mathcal{D}$ that has a node mapping to X and another adjacent node mapping outside X. Then B contains two nodes that are mapped at least distance 3 apart because they cannot map to the c_ℓ node of X or to the c_r node of S.

Suppose on the other hand that u is the g_r node of size gadget A in X. Consider the set $\mathcal{Z}$ of size gadgets in $\mathcal{D}$ that have a node mapping to a g_ℓ node of A. Since the size of each gadget in $\mathcal{Z}$ is 1 mod c, the number of g_ℓ nodes of A is 0 mod c, and $|\mathcal{Z}| \leq |\text{Edges}(\mathcal{G})| = 3n < c$, there exists a size gadget $B \in \mathcal{Z}$ that also maps outside A. In particular, B must have a node mapped to the c_ℓ node of X which is the only node of $\mathcal{S}$ outside A within distance 2 of the g_ℓ nodes of A. Since by (4) the c_r node of $\mathcal{S}$ is already occupied by the c_r node of $\mathcal{D}$, no size gadget (other than possibly B) mapping to a node outside X can also map within X without causing $exp(\varphi^{-1}) \geq 3$. (5) now follows from (1). □

Proof of Lemma 10. From Proposition 3 (8) and (9), any non-spare size gadget in $\mathcal{D}$ and the c_ℓ node of $\mathcal{D}$ it is attached to must map within the same successor cluster in $\mathcal{S}$ under φ^{-1}. Consequently, Proposition 3 (8) can be strengthened to say that every successor cluster in $\mathcal{S}$ corresponding to a node v is fully mapped from exactly one c_ℓ node in $\mathcal{D}$, the size gadget B attached to it, and two spare size gadgets. As $s(\cdot)$ is sum-free, B must correspond to a successor of v. Since there are exactly n c_ℓ nodes in $\mathcal{D}$, this assigns a unique successor to each node v, establishing a disjoint cycle and the contradiction which proves the lemma. □

4 Hardness of Embeddings Between Line Graphs with Large Weights

A *line graph* is an acyclic connected graph of maximum degree two, that is, a line of vertices. We have the following result whose proof appears in the full version of the paper [4].

Theorem 4. *Given two line graphs with n nodes and weight ratio $\Omega(b^2)$, for any $k > 1$ and b with $b = \Omega(kn^2)$, it is NP-hard to determine if the distortion between them is less than b/k or at least b.*

Corollary 2. *For $\alpha > 0$, it is NP-hard to approximate the distortion between two line graphs with n nodes and weight ratio $\Omega(\alpha^2 n^4)$ within a factor of α.*

5 Conclusion

We have shown that the problem of finding a minimum distortion embedding between two metrics is hard to approximate within constant factors on even extremely simple graphs, such as weighted lines or unweighted trees. While our constants improve previous results, we believe they are still far from the true story: it seems likely that even approximating distortion in unweighted graphs is much harder than what we know.

One natural relaxation to the graph embedding problem is to find the distortion of embedding a constant fraction of one graph to another. While this quantity will in general be far from the true distortion, it may provide a good enough measure of graph difference for certain applications. Other notions of distortion may also be useful. Rabinovich [19] has used average distortion to study the MinCut-MaxFlow gap in uniform-demand mulitcommodity flow. Other possibly interesting measures are max-distortion, which is the maximum of expansion and contraction rather than the product, and Gromov-Hausdorff distance, which has applications in analysis. The problem remains open in all these scenarios.

References

1. Y. Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *FOCS*, pages 184–193, Burlington, VT, 1996.
2. M. Bădoiu, E. Demaine, M. Farach-Colton, M. Hajiaghayi, and A. Sidiropoulos. Ordinal embeddings of minimum relaxation: General properties, trees, and ultrametrics. In *SODA*, pages 650–659, Vancouver, BC, Jan. 2005.
3. M. Bădoiu, K. Dhamdhere, A. Gupta, Y. Rabinovich, H. Raecke, R. Ravi, and A. Sidiropoulos. Approximation algorithms for low-distortion embeddings into low-dimensional spaces. In *SODA*, pages 119–128, Vancouver, BC, Jan. 2005.
4. M. Cary, A. Rudra, and A. Sabharwal. On the hardness of embeddings between two finite metrics. Technical Report UW-CSE-TR-2005-04-02, Univ. of Washington, 2005.
5. M. Cary, A. Rudra, A. Sabharwal, and E. Vee. Floodlight illumination of inifinite wedges. In *14th Annual Fall Workshop on Comp. Geom.*, Boston, MA, Nov. 2004. Technical Report UW-CSE-2004-10-4, University of Washington.
6. M. Deza and M. Laurent. *Geometry of Cuts and Metrics*, volume 15 of *Algorithms and Combinatorics*. Springer, 1997.
7. I. Dinur and S. Safra. The importance of being biased. In *STOC*, pages 33–42, Montreal, Canada, May 2002.
8. J. Fakcharoenphol, S. Rao, and K. Talwar. A tight bound on approximating arbitrary metrics by tree metrics. In *STOC*, pages 448–455, San Diego, CA, June 2003.
9. M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company, NY, 1979.
10. F. Hoffmann, K. Kriegel, and C. Wenk. Matching 2d patterns of protein spots. In *Proc. 14th Ann. Symp. on Comp. Geom.*, pages 231–239, 1998.
11. P. Indyk. Algorithmic aspects of geometric embeddings, Oct. 2001. FOCS '01 tutorial. Available at `http://thoery.lcs.mit.edu/~indyk/tut.html`.
12. W. B. Johnson and H. Lindenstrauss, editors. *Handbook of the Geometry of Banach Spaces*. North-Holland, 2003.
13. C. Kenyon, Y. Rabani, and A. Sinclair. Low distortion maps between point sets. In *STOC*, pages 272–280, Chicago, IL, June 2004.
14. S. Khanna, N. Linial, and S. Safra. On the hardness of approximating the chromatic number. *Combinatorica*, 20(3):393–415, 2000.
15. J. Lee. On distance scales, embeddings, and efficient relaxations of the cut cone. In *SODA*, pages 92–101, Vancouver, BC, Jan. 2005.
16. N. Linial. Finite metric spaces – combinatorics, geometry and algorithms. In *Proc. Int. Congress of Math. III*, pages 573–586, Beijing, China, 2002.
17. J. Matsoušek. *Lectures on Discrete Geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer, 2002.
18. C. Papadimitriou and S. Safra. The copmlexity of low-distortion embeddings between point sets. In *SODA*, pages 112–118, Vancouver, BC, Jan. 2005.
19. Y. Rabinovich. On average distortion of embedding metrics into the line and into l1. In *STOC*, pages 456–462, San Diego, CA, June 2003.
20. R. Ramesh and I. V. Ramakrishnan. Nonlinear pattern matching in trees. *Journal of the ACM*, 39(2):295–316, 1992.
21. T. Schlieder and F. Naumann. Approximate tree embedding for querying XML data. In *ACM-SIGIR Work., XML and Info. Retrieval*, Athens, Greece, July 2000.