Weizmann Institute of Science

Thesis for the degree
Master of Science

Submitted to the Scientific Council of the
Weizmann Institute of Science
Rehovot, Israel

Eliminating Steiner Vertices in Graph Metrics

Author:
Tamar Zondiner

Supervisor:
Dr. Robert Krauthgamer

February 29, 2012

Abstract

Given an edge-weighted undirected graph G and a subset of "required" vertices $R \subseteq V(G)$, called the terminals, we want to find a minor G' with possibly different edge-weights, that retains distances between all terminal-pairs exactly, and is as small as possible. We prove that every graph G with n vertices and k terminals can be reduced (in this sense) to a minor G' with $O(k^4)$ vertices and edges. We also give a lower bound of $\Omega(k^2)$ on the number of vertices required.

The $O(k^4)$ upper bound on the size of the minor is achieved using a specific construction for minors, which we call Oriented Minors. For this specific method we show that the upper bound is tight. The $\Omega(k^2)$ lower bound is proved by an even stronger claim; there are planar graphs G such that any *planar graph* that preserves distances between terminals in G has $\Omega(k^2)$ vertices. When restricting the graphs G and G' to trees, we prove that $2k-2$ vertices are sufficient and necessary.

Another version of this problem requires that $V(G') = R$ and asks for G' that approximates the distances between terminals within a constant factor. Previous results proved that this is possible in trees and in outerplanar graphs, and termed this problem *Steiner Point Removal*. We study a particular planar graph G that we suspected would give a super-constant lower bound on the approximation factor. We refute this suspicion, finding an outerplanar minor of G achieving constant approximation. An interesting generalization of this result is that for any distance metric on the terminals $\{0, 1, .., k\}$ adhering to a certain monotonicity rule, there exists an outerplanar graph that approximates the metric within a constant factor.

1 Introduction

Suppose that we are given an undirected graph G with non-negative edge-lengths, and a subset of k terminal vertices $R \subseteq V(G)$, and are only interested in the distances between terminal-pairs (henceforth: *terminal distances*). If $|V(G)| \gg k$, a lot of the information in the graph may be redundant, making it beneficial to find a smaller graph G' that contains the terminals and preserves their distances. We seek a bound $f^*(k)$ such that for every edge-weighted graph G and set of k terminals, there exists a minor G' of G of size $f^*(k)$ with new edge-lengths, that preserve terminal distances exactly. We show that it is always possible to find such a minor of G with $O(k^4)$ vertices and edges, and that some graphs G might require $|V(G')| \geq \Omega(k^2)$ for every such minor G'. We are not aware of previous publications on this problem. [1]

This problem is one instance of a general genre of problems; given an edge-weighted graph G and a set of terminals R, is there a structurally (or topologically) similar graph G' with new edge-lengths, such that G' preserves terminal distances and is small. Instances of this problem may vary in terms of the possible input graphs G (e.g. restrict them to planar graphs), the structural similarity required of G' (a minor of G, a member of the same graph family etc.), the number of non-terminal vertices allowed in G' (if any), the degree to which G' preserves the terminal distances in G (retains them exactly or approximates them) and more. Many instances of this genre are natural, but only a few of them were previously asked, and even fewer were answered.

The problem of approximating terminal distances by a graph without non-terminals was introduced by Gupta [Gup01] and termed the *Steiner Point Removal (SPR)* problem, where Steiner points refer to the non-terminals. We extend this term to include the removal of some, but not all, Steiner points, effectively reducing the number of non-terminals in the graph. Algorithmically, the input to this problem is an edge-weighted graph G and a set of terminals R. The output is an edge-weighted graph G' with $R \subseteq V(G')$

[1] We recently posted a paper [KZ12], that includes our main results from Section 3 on this problem, as well as additional results on bounded treewidth graphs.

that preserves (exactly or approximately) the terminal distances. Ideally, this graph is small (has size proportional to k) and preserves terminal distances faithfully (either exactly or within a small constant factor). The graph G' can then potentially replace the original graph G in various computations involving terminal distances.

For example, given a planar graph G and k terminals, we can use SPR to approximate the minimal-length TSP tour that visits every terminal more efficiently. If we can find a planar graph G' with $f(k)$ vertices that preserves terminal distances exactly, then we can use it to replace G in the approx. TSP computation. The runtime of the approx. TSP computation on G' is a function of k, whereas on G it is a function of $|V(G)|$, which is potentially much larger. However, since the construction of G' also takes time, a significant reduction in runtime occurs especially when TSP queries on several subsets of the terminals are computed. In this case we can find G' as a preprocessing stage, with runtime polynomial in $|V(G)|$, and run queries in time that depends only on k.

Trivially, the complete graph on the terminals can always retain distances exactly, by setting the edge-length of every edge to the distance in G between its endpoints. It is easy to see that in some cases the complete graph is the only graph without non-terminals that preserves terminal distances exactly. To see this, consider the star graph $K_{1,k}$ with unit edge-lengths, and let the k leaves in the star graph be its terminals. The distance between any two terminals is 2, making the complete graph the only graph without non-terminals that preserves these distances exactly. By allowing a single non-terminal vertex in the graph, we can avoid this dense structure. Alternatively, we can use $K_{1,k-1}$ as a tree on the terminals (without non-terminals) that approximates all terminal distances within a factor of 2. These two options exemplify the two main questions addressed in this work:

1. How many non-terminals are needed to guarantee that there exists a graph G' that preserves terminal distances **exactly**, while adhering to certain structural requirements (same graph family, minor, etc.)?

2. In what graph-families $\mathcal{F}$ does every graph $G \in \mathcal{F}$ have a graph G'

only on the terminals that **approximates** terminal distances within a constant factor, while adhering to certain structural requirements (same graph family, minor, etc.)?

1.1 Maintaining Distances Exactly

Some minors of G are what we call *Oriented Minors*. They are constructed by specific operations and are accompanied by a calculation of *induced edge-lengths* (see definitions 2.5, 2.6 and 2.7). Informally, every vertex in such a minor corresponds to a vertex in G, with distances greater or equal to the distances in G.

Any oriented minor of G is also a minor of G, and in minor-closed graph families, it also belongs to the same graph family. Therefore, we aim to find upper bounds on oriented minors G', and lower bounds on family-preserving graphs G' (graphs which belong to the same family as G), whenever possible.

Our Results: Trees form a relatively simple family of input graphs to analyze, due to the uniqueness of simple paths between any two vertices. However, removing all non-terminals cannot be done without modifying terminal distances, as exemplified above by a star graph. In Section 3.1 we prove that for any edge-weighted tree G and any subset of k terminals, there exists an oriented minor of G with at most $k-2$ non-terminals, that preserves terminal distances exactly. We further prove that this bound is tight, even for family-preserving graphs G' (i.e. when G' is only required to be a tree, not necessarily a minor of G).

In Section 3.2 we give a construction that for any graph and k terminals, creates an oriented minor with at most $O(k^4)$ non-terminals, that preserves terminal distances exactly. We prove a matching lower bound of $\Omega(k^4)$ non-terminals for oriented minors, and a lower bound of $\Omega(k^2)$ non-terminals for family-preserving reductions in the family of planar graphs. This last bound also implies that in every family of graphs containing all planar graphs, $\Omega(k^2)$ non-terminals are needed to guarantee the existence of minors that preserve terminal distances.

Related Work: Reducing the size of a graph while retaining distances between vertex-pairs is analogous to reducing it while retaining $s-t$ cuts between vertex-pairs. Gomory and Hu [GH61] proved that it is possible to construct a weighted tree preserving $s-t$ cuts between all vertex-pairs in a graph.

Another related question regarding flows is the Mimicking Network problem. Given a graph G and a set of k terminals, a Mimicking Network is a weighted graph G' that preserves flows (or cut sizes) between all possible partitions of the terminals. Hagerup et al. [HKNR95] first introduced the problem of finding the minimal number of non-terminals needed to guarantee the existence of a Mimicking Network, and proved that for any graph and k terminals, $O(2^{2^k})$ non-terminals are enough. A lower bound proving that at least one non-terminal might be required was proved by Chaudhuri et al. [CSWZ98] along with upper bounds for specific graph families.

Coppersmith and Elkin [CE06] consider the problem of preserving terminal distances exactly, while reducing the number of edges in the graph (instead of vertices). They prove that for every weighted graph $G=(V,E)$ and set of $O(|V|^{\frac{1}{4}})$ terminals R, there exists a weighted graph $G'=(V,E')$ preserving terminal distances exactly, such that $E' \subseteq E$ and $|E'| \leq O(|V|)$.

1.2 Approximating terminal distances (without non-terminals)

Previous Results: The question of approximating terminal distances within a constant factor by graphs without non-terminals was first introduced by Gupta [Gup01], where it was proved that for any weighted tree and set of terminals, there exists a weighted tree without non-terminals, that approximates all terminal distances within a factor of 8. In [CGN+06] it was stated that the approximating tree proved to exist in [Gup01] is, in fact, a minor of the original tree. Chan et al. [CXKR06] later proved a matching lower bound of $8(1-o(1))$. Basu and Gupta [BG08] proved that for any weighted outerplanar graph there exists an outerplanar graph without non-terminals, approximating terminal distances within a factor of 15. Englert et al. [EGK+10] proved a randomized version of this problem for all minor-excluded graph families, with an expected approximation factor depending only on the size of the excluded minor.

Our Results: A known open question in this area is whether or not the results of [Gup01] and [BG08] extend to planar graphs, i.e. if it is true that for every planar graph G and set of terminals R there exists a minor G' without non-terminals that approximates terminal distances within a constant factor. At the beginning of our research we conjectured that this property does not extend to planar graphs. We tried proving this conjecture using an interesting planar example, suggested to us by Anastasios Sidiropoulos. This planar graph is a discretized version of a hyperbolic plane with all terminals on the outer face, so every minor without non-terminals is in fact an outerplanar graph. We suspected that no outerplanar graph can give a constant approximation to the terminal distances in this graph, which would imply that no minor can do so. We then found that the opposite is true - there exists an outerplanar minor of this graph that achieves constant approximation to all terminal distances. We then generalized this construction, proving that outerplanar graphs can achieve constant factor approximation to every metric with a similar monotonicity property.

2 Preliminaries

In this paper, we discuss properties of edge-weighted, connected, undirected graphs. An SPR (Steiner Point Removal) instance is a triple $\langle G, \ell_G, R\rangle$ of such a graph G with edge-lengths $\ell_G: E(G) \to \mathbb{R}^+$, and a set of terminals $R \subseteq V(G)$. An SPR "output" instance is a triple $\langle G', \ell_{G'}, R\rangle$ of a graph G' with edge-lengths $\ell_{G'}: E(G') \to \mathbb{R}^+$ such that $R \subseteq V(G')$. Unless otherwise specified, we use the following notations: $n := |V(G)|$, $m := |E(G)|$, and $k := |R|$. The shortest-path distances of a graph G according to its edge-lengths define a metric, denoted d_G on the vertices $V(G)$, so that $d_G(u, v)$ is the length of the shortest-path between u and v in G for every $u, v \in V(G)$. The distance metric of the SPR instance $\langle G, \ell_G, R\rangle$ induced by the terminals is denoted $d_G|_{R\times R}$. We will use the notions of weighted graphs and graph distance interchangeably throughout this work.

Given an SPR instance $\langle G, \ell_G, R\rangle$, we want to find an SPR instance $\langle G', \ell_{G'}, R\rangle$ that induces a distance metric $d_{G'}|_{R\times R}$ on the terminals that is *faithful* to $d_G|_{R\times R}$.

Definition 2.1 *The SPR instance* $\langle G', \ell_{G'}, R\rangle$ *is* faithful *to the SPR instance* $\langle G, \ell_G, R\rangle$ *if terminal distances are preserved* exactly*, i.e.*

$$d_{G'}(t_1, t_2) = d_G(t_1, t_2) \qquad \forall t_1, t_2 \in R.$$

Definition 2.2 *The SPR instance* $\langle G', \ell_{G'}, R\rangle$ *is* α-approximately faithful *to the SPR instance* $\langle G, \ell_G, R\rangle$ *if terminal distances are preserved up to a factor* $\alpha \geq 1$*, i.e.*

$$d_G(t_1, t_2) \leq d_{G'}(t_1, t_2) \leq \alpha \cdot d_G(t_1, t_2) \qquad \forall t_1, t_2 \in R.$$

We require that the graph G' is structurally (or topologically) similar to the graph G. Three requirements are made of G' in this work: *family-preserving*, *minor*, and *Oriented-Minor*.

Definition 2.3 *Given a graph family* $\mathcal{F}$ *and a graph* $G \in \mathcal{F}$*, a graph* G' *is* family-preserving *of* G *if it also belongs to* $\mathcal{F}$*.*

Definition 2.4 *A* minor *of a graph G is any graph G' that is the result of a series of edge-removals, vertex-removals, and edge-contractions applied to G.*

We are free to assign any edge-length function to G', since the claim that G' is a minor of G is only structural. We hereby define an *Oriented Minor* G' of G, a minor with specific *induced edge lengths*, assigned according to the series of operations performed to create G'.

Definition 2.5 *An* Oriented Minor *of a graph G is the result of a series of edge-removals, vertex-removals and* oriented edge-contractions *performed on G.*

Definition 2.6 *An* oriented edge-contraction *of an edge (u, v) is the contraction of one of its vertices, wlog u, into the other, so that the new vertex created by the edge-contraction is named v.*

Definition 2.7 *The* Induced Edge-Lengths *of an oriented minor are created along with the oriented minor itself as follows:*

1. *Let G' and $\ell_{G'}$ be the oriented minor and induced edge-lengths created thus far (at first G' is set to G and $\ell_{G'}$ is set to ℓ_G).*
2. *Edge-removal removes the edge from the range of $\ell_{G'}$.*
3. *Vertex-removal removes all the edges adjacent to the vertex from $\ell_{G'}$.*
4. *The oriented contraction of (u, v) into v does the following for every $(u, w) \in E(G')$:*
   - *If $(v, w) \in E(G')$ set $\ell_{G'}(v, w)$ to $\min\{\ell_{G'}(v, w), \ell_{G'}(v, u)+\ell_{G'}(u, w)\}$*
   - *Otherwise, add (v, w) to the range of $\ell_{G'}$, and set $\ell_{G'}(v, w)$ to $\ell_{G'}(v, u) + \ell_{G'}(u, w)$.*

   This action also removes all the edges adjacent to u from the range of $\ell_{G'}$.

Note that if G' is an oriented minor of G, it holds that $V(G') \subseteq V(G)$. It is easy to verify that $\forall u, v \in V(G')$ the inequality $d_{G'}(u, v) \geq d_G(u, v)$ holds.

Defining the Goals: The Steiner Point Removal (SPR) problem is a general genre of problems, and has many instances depending on the promised graph family $\mathcal{F}$ to which G belongs, the required property of the returned graph G' (family-preserving, minor, or oriented minor), and a bound on either the number of vertices allowed in $|V(G')|$ or the maximal factor to which terminal shortest paths may be increased.

The following notations are used to denote the values we wish to bound in the *Exact Faithfulness* section:

- $f^*(k, \mathcal{F})$ denotes the minimal size of $V(G')$ such that for every SPR instance $\langle G, \ell_G, R\rangle$ with $G \in \mathcal{F}$ there exists an SPR instance with $G' \in \mathcal{F}$ such that G' is faithful to G.
- $f^{minor}(k, \mathcal{F})$ denotes the minimal size of $V(G')$ such that for every SPR instance $\langle G, \ell_G, R\rangle$ with $G \in \mathcal{F}$ there exists an SPR instance such that G' is a faithful minor of G.
- $f^{orient}(k, \mathcal{F})$ denotes the minimal size of $V(G')$ such that for every SPR instance $\langle G, \ell_G, R\rangle$ with $G \in \mathcal{F}$ there exists an SPR instance such that G' is a faithful oriented minor of G and $\ell_{G'}$ is the induced edge-lengths for G'.

The following notations are used to denote the values we wish to bound in the α-*Approximate Faithfulness* section:

- $\alpha^*(k, \mathcal{F})$ denotes the minimal approximation $\alpha \geq 1$ such that for every SPR instance $\langle G, \ell_G, R\rangle$ with $G \in \mathcal{F}$ there exists an α-approximately faithful SPR instance $\langle G', \ell_{G'}, R\rangle$ with $G' \in \mathcal{F}$ and $V(G') = R$.
- $\alpha^{minor}(k, \mathcal{F})$ denotes the minimal approximation $\alpha \geq 1$ such that for every SPR instance $\langle G, \ell_G, R\rangle$ with $G \in \mathcal{F}$ there exists an α-approximately faithful SPR instance $\langle G', \ell_{G'}, R\rangle$ where G' is a minor of G and $V(G') = R$.

If $\langle G, \ell_G, R\rangle$ is an oriented minor with induced edge-lengths, it is also a minor, hence $f^{minor}(k, \mathcal{F}) \leq f^{orient}(k, \mathcal{F})$. In a minor-closed graph family $\mathcal{F}$, any minor is also a member of the family. Therefore $f^*(k,\mathcal{F}) \leq f^{minor}(k,\mathcal{F})$ and $\alpha^*(k,\mathcal{F}) \leq \alpha^{minor}(k,\mathcal{F})$ in any minor-closed family $\mathcal{F}$.

3 Exact Faithfulness

The goal of this section is to bound the minimal number of non-terminals required to guarantee that for every SPR instance $\langle G, \ell_G, R\rangle$ there exists a faithful SPR instance $\langle G', \ell_{G'}, R\rangle$ such that G' is either a minor of G, an oriented-minor of G, or belongs to the same family as G. As exemplified in a star graph it is impossible, even on simple graphs, to find such an instance without allowing that $V(G')$ contains some non-terminals.

We begin with the simple case of tree graphs, and prove that $f^*(k, \textsc{Trees}) \leq f^{orient}(k, \textsc{Trees}) \leq 2k-2$. We show a matching lower bound for infinitely many k; for $k = 2^i$, we prove that $2k - 2 \geq f^*(k, \textsc{Trees})$. We later discuss minors of general graphs, and bound $\Omega(k^2) \leq f^{minor}(k, \textsc{AllGraphs}) \leq O(k^4)$ and $f^{orient}(k, \textsc{AllGraphs}) = \Theta(k^4)$.

3.1 Trees

In trees, it is simple to calculate the shortest-path between any two terminals, as it is the only simple path between them. This simplifies the analysis, which we consider as more of a baseline. In this section we prove that

$$f^*(k, \textsc{Trees}) = f^{minor}(k, \textsc{Trees}) = f^{orient}(k, \textsc{Trees}) = 2k - 2.$$

We start by proving the upper bound $f^{orient}(k, \textsc{Trees}) \leq 2k - 2$. This bound would imply that $f^*(k, \textsc{Trees}) \leq f^{minor}(k, \textsc{Trees}) \leq 2k - 2$. We later show that these bounds are tight. First we give the simpler proof that $f^{orient}(k, \textsc{Trees}) \geq 2k - 2$, and then generalize this result, proving that $f^*(k, \textsc{Trees}) \geq 2k - 2$.

3.1.1 Upper Bound for trees

Theorem 3.1 *Let $\langle G, \ell_G, R\rangle$ be an SPR instance where G is a tree. Then there exists a faithful SPR instance $\langle G', \ell_{G'}, R\rangle$ such that G' is an oriented minor of G, $\ell_{G'}$ is its induced edge-lengths, and $|V(G')| \leq 2k - 2$.*

In other words, $f^{orient}(k, \textsc{Trees}) \leq 2k - 2$.

The instance $\langle G', \ell_{G'}, R\rangle$ *can be computed in time polynomial in* $|V(G)|$.

We construct the oriented minor G' and its induced length function $\ell_{G'}$ using Algorithm 1.

Algorithm 1 ReduceTree($\langle G, \ell_G, R\rangle$)

$G' \leftarrow G$
$\ell_{G'} \leftarrow \ell_G$
while there exists a vertex $v \in V(G') \setminus R$ such that $deg_{G'}(v) = 1$ **do**
  Remove v from G' along with its edges
end while
while there exists a vertex $v \in V(G') \setminus R$ such that $deg_{G'}(v) = 2$ **do**
  Let u, w be the neighbors of v in G'.
  Use *Oriented Edge-Contraction* to contract v into u
  ($\ell_{G'}(u, w)$ is set to the induced length $\ell_{G'}(u, v) + \ell_{G'}(v, w)$).
end while
return $\langle G', \ell_{G'}, R\rangle$.

It is clear from the construction of G' that it is an oriented minor of G, and that $\ell_{G'}$ is its induced edge-lengths.

Claim 3.2 *The returned SPR instance* $\langle G', \ell_{G'}, R\rangle$ *is faithful to* $\langle G, \ell_G, R\rangle$.

Proof No terminals are removed during the algorithm, so $R \subseteq V(G') \subseteq V(G)$.

Every non-terminal vertex with degree 1 is never on a shortest-path between two terminals, and can thus be removed without changing the metric $d_{G'}|_{R\times R}$.

When contracting a non-terminal vertex with degree 2, the 2-edge path is replaced by an edge with the same length, thus retaining the original distance.

As a result, any two vertices $u, v \in V(G')$ have $d_{G'}(u, v) = d_G(u, v)$, and the returned SPR instance is faithful. ■

We denote the non-terminals in G' at the end of the algorithm by S.

Lemma 3.3 $|S| \leq k-2$.

Proof After the first loop terminates, every non-terminal $v \in V(G')$ has $\deg_{G'}(v) \geq 2$. Since the degrees of remaining vertices in G' don't change during the second loop, once it terminates there are no non-terminals with degrees 1 or 2, i.e. $\forall v \in S$ it holds that $\deg_{G'}(v) \geq 3$. The graph G' is a connected tree, so it holds that $\sum_{v \in V(G')} \deg_{G'}(v) = 2|E(G')| = 2|V(G')| - 2$. It now holds that:

$$
\begin{aligned}
2|V(G')| - 2 = 2(k + |S|) - 2 &= \sum_{v \in V(G')} \deg_{G'}(v) \\
&= \sum_{v \in R} \deg_{G'}(v) + \sum_{v \in S} \deg_{G'}(v) \geq k + 3|S|
\end{aligned}
$$

$\Rightarrow |S| \leq k-2$. ■

It is easy to see that the algorithm described above has polynomial time complexity in $|V(G)|$. Lemma 3.3 implies that $|V(G')| \leq 2k-2$, making $\langle G', \ell_{G'}, R \rangle$ a faithful oriented minor with induced edge lengths and $|V(G')| \leq 2k-2$, thus proving Theorem 3.1.

3.1.2 Lower Bound for trees

Both the proof for $f^{orient}(k, \text{TREES})$ and the proof for $f^*(k, \text{TREES})$ use the complete binary tree as an example. Let G_k be a complete binary tree with $k = 2^i$ leaves, denoted $L(G_k)$. Let the terminals R_k be these leaves. Let $\tilde{G}_k$ be the result of the oriented contraction of the root of G_k into one of its children, and the edge-lengths $\ell_{\tilde{G}_k}$ be the induced edge-lengths of $\tilde{G}_k$. The triple $\langle \tilde{G}_k, \ell_{\tilde{G}_k}, L(G_k) \rangle$ is an SPR instance, with $2k-2$ vertices and k terminals.

Claim 3.4 *For every $k = 2^i$, the only faithful oriented minor of $\langle \tilde{G}_k, \ell_{\tilde{G}_k}, L(G_k) \rangle$ with induced edge lengths, is $\langle \tilde{G}_k, \ell_{\tilde{G}_k}, L(G_k) \rangle$ itself. In other words,*
$f^{orient}(k, \text{TREES}) \geq |V(\tilde{G}_k)| \geq 2k-2$.

Proof It is easy to see that any operation on $\tilde{G}_k$ increases the length of the induced shortest path. Induced distances only increase when performing a series of operations, so if no single operation can be performed without increasing distances, no series of natural operations can be performed without doing so.

Thus proving that all $2k-2$ vertices in $\langle \tilde{G}_k, \ell_{\tilde{G}_k}, L(G_k)\rangle$ are needed in any faithful oriented minor with induced distances, i.e. $f^{orient}(k, \textsc{Trees}) \geq |V(\tilde{G}_k)| \geq 2k-2$. ■

We now prove the more general claim, stating that for infinitely many values of k, $f^*(k, \textsc{Trees}) \geq 2k-2$. We do so by proving that any SPR instance $\langle G', \ell_{G'}, L(\tilde{G}_k)\rangle$ that is faithful to $\langle \tilde{G}_k, \ell_{\tilde{G}_k}, L(G_k)\rangle$, and with a tree G', has at least $2k-2$ vertices.

Theorem 3.5 *For any $k = 2^i$, any triple $\langle G', \ell_{G'}, R\rangle$ such that G' is a tree, and is faithful to $\langle \tilde{G}_k, \ell_{\tilde{G}_k}, L(G_k)\rangle$ has at least $2k-2$ vertices. Furthermore, in any such tree there doesn't exist a vertex $v \in V(G')$ at equal distances from each of the terminals of G'.*

Thus, $f^{orient}(k, \textsc{Trees}) \geq 2k-2$.

Proof We proceed by induction on i.

When $i = 1$ this is clear. The tree must be a single edge between two vertices, with length 2. Clearly, there is no vertex within equal distance to the two vertices.

Consider the tree $\tilde{G}_{2k}$. Let T_1 and T_2 be the two complete binary trees in $\tilde{G}_{2k}$.

Lemma 3.6 *There doesn't exist a vertex $x \in V(G')$ used both by a path between $v_1, v_2 \in L(T_1)$ and between $u_1, u_2 \in L(T_2)$*

Proof of Lemma 3.6 Suppose towards contradiction that this is not the case, and denote this vertex by x. Suppose wlog that $d_{G'}(v_1, x) \leq d_{G'}(v_2, x)$, and $d_{G'}(u_1, x) \leq d_{G'}(u_2, x)$. Then $d_{G'}(u_1, v_1) \leq \frac{1}{2} d_{G'}(v_1, v_2) + \frac{1}{2} d_{G'}(u_1, u_2)$. Since T_1 and T_2 are the two subtrees of a complete binary tree, the distance between leaves of these subtrees must be larger than any internal path, thus causing a contradiction. ■

No vertices are used by internal shortest paths between terminals in T_1 and terminals in T_2. This means that there exist two distinct subtrees in G' as well, which we denote T'_1 and T'_2. These subtrees are connected by only one edge, since the graph G' is a tree. This edge must retain distances, and the distances between any terminal in T'_1 to any terminal in T'_2 are equal. Therefore, the connection must be between a point equally close to all terminals in T'_1 and a point equally close to all terminals in T'_2.

From the induction hypothesis, the trees T'_1 and T'_2 have a central point only if they each have at least $2k-1$ vertices. Therefore, $|V(G')| \geq 2(2k-1) = 4k-2$ vertices, which proves Theorem 3.5. ■

3.2 General Graphs

We start by proving an intuitive upper bound of $f^{orient}(k, \textsc{AllGraphs}) \leq O(k^4)$. An immediate corollary of this result is the following:

Corollary 3.7 *For any minor-closed family of graphs $\mathcal{F}$, any SPR instance $\langle G, \ell_G, R\rangle$ with $G \in \mathcal{F}$ has a faithful graph $G' \in \mathcal{F}$ with $O(k^4)$ vertices. In other words, $f^*(k, \textsc{AllGraphs}) \leq O(k^4)$.*

This corollary is useful when discussing graph-family specific problems on terminals, as it effectively reduces complexities depending on n to ones depending solely on k, assuming that the graph-family is minor-closed.

We later prove that $f^{orient}(k, \textsc{AllGraphs}) = \Theta(k^4)$, as well as a lower bound of $f^{minor}(k, \textsc{AllGraphs}) \geq \Omega(k^2)$.

3.2.1 Upper Bound for minors of General Graphs

Theorem 3.8 *For every SPR instance $\langle G, \ell_G, R\rangle$ there exists an instance $\langle G', \ell_{G'}, R\rangle$ s.t. G' is a faithful oriented minor of G and $\ell_{G'}$ is its induced edge lengths, such that $|V(G')| \leq k^4$, and $|E(G')| = O(k^4)$. Furthermore, G' can be found in time polynomial in $|V(G)|$.*

We construct the oriented minor (and induced edge lengths) as follows. Consider all the shortest paths in a graph G, and suppose that the lengths of shortest paths are unique. This assumption is reasonable since otherwise we can either enforce a lexicographic order on the paths, and choose the shortest paths accordingly, or perturbate the path lengths by ϵ small as we like.

Observation 3.9 *Let u, v, x, y be terminals in R. Then assuming the shortest path uniqueness, their shortest paths in G, denoted $p_G(u, v)$ and $p_G(x, y)$, converge and diverge at most once.*

This is easy to see since otherwise the shortest path between the first convergence and last divergence is not unique, in contradiction to the unique path lengths.

We construct the graph G' as follows. First we mark all the vertices in which there is convergence or divergence of two paths between terminals. There are $\binom{k}{2}$ paths, and thus $\leq k^4$ such vertices.

Clearly, two paths starting at the same terminal mark that terminal as a convergence vertex. So, a vertex v is not marked if it is both not a terminal and also one of the following holds:

1. The vertex v is never used in any shortest-path between terminals in G.
   In this case, v and its edges can be safely removed from G.

2. The vertex v is only traversed using the same two edges $(u, v), (v, w)$. In this case, v can be orientally contracted into w, and all shortest paths using v (i.e. using the path between u and w) retain their lengths in G' with its induced edge-lengths.

The SPR instance $\langle G', \ell_{G'}, R \rangle$ we create by constructing this oriented minor, G', with its induced edge-lengths, $\ell_{G'}$, is the result of performing these vertex-removals and oriented contractions. This SPR instance is clearly faithful to $\langle G, \ell_G, R \rangle$.

It is therefore possible to eliminate all vertices but the convergence and divergence vertices of shortest paths between terminals in G, thus proving that $f^{orient}(k, \textsc{AllGraphs}) \leq O(k^4)$.

It is interesting to note that the new graph is rather sparse, with $|E(G')| \leq O(k^4)$. This is an immediate result of the following claim:

Claim 3.10 *A path $p_{G'}(t_1, t_2)$ between $t_1, t_2 \in R$ uses $O(k^2)$ edges.*

Proof Suppose wlog we begin by comparing the original paths $p_G(a_1, a_2)$ with $p_G(t_1, t_2)$ for all $a_1, a_2 \in R$. This marks at most two vertices as ones not to be contracted. Overall, these comparisons mark at most k^2 vertices on the path $p_G(t_1, t_2)$.

Then we compare the rest of the terminal-pairs' paths, and mark their convergence and divergence points. These points are either not on $p_G(t_1, t_2)$, or they are also the convergence of one of the paths with $p_G(t_1, t_2)$ and thus have already been marked.

Since only marked vertices exist in G', there are at most $O(k^2)$ vertices along the path between t_1 and t_2 in G', hence also $O(k^2)$ vertices along that path. There are $\binom{k}{2}$ paths, meaning $O(k^4)$ edges in G' overall.

■

The graph G' can be easily constructed by finding the shortest paths between terminals and comparing them, and can therefore be constructed in time polynomial in the number of vertices $|V(G)|$. This completes the proof of Theorem 3.8

3.2.2 Lower Bounds for Minors of General Graphs

We proved above that $f^{orient}(k, \textsc{AllGraphs}) \leq O(k^4)$. We now show that this bound is tight, i.e. $f^{orient}(k, \textsc{AllGraphs}) = \Theta(k^4)$. Later, we prove that $f^*(k, \textsc{Planar}) \geq \Omega(k^2)$, and hence that $f^{minor}(k, \textsc{AllGraphs}) \geq \Omega(k^2)$. How to settle the gap for $f^{minor}(k, \textsc{AllGraphs})$ between $\Omega(k^2)$ and $O(k^4)$ is an interesting open problem resulting from this work.

Theorem 3.11 *For every k there exists an SPR instance $\langle G, \ell_G, R\rangle$ such that every faithful SPR instance $\langle G', \ell_{G'}, R\rangle$ of an oriented minor G' and induced edge-lengths $\ell_{G'}$ has $|V(G')| \geq \Omega(k^4)$. This holds also for planar graphs G. In other words, $f^{orient}(k, \text{AllGraphs}) \geq f^{orient}(k, \text{Planar}) \geq \Omega(k^4)$.*

Proof We construct $\langle G, \ell_G, R\rangle$ probabilistically as follows. Consider the $[0,1] \times [0,1]$ square in the 2-dimensional Euclidean plane. On each of the edges of this square, we randomly choose $\lfloor \frac{k}{4} \rfloor$ points, and place terminals in them. We connect by a straight line the vertices on the top edge with those on the bottom edge, and those on the right with those on the left. There are $\Theta(k^2)$ "horizontal" edges each meeting $\Theta(k^2)$ "vertical" edges, creating $\Theta(k^4)$ intersections. An intersection point is the intersection of only two lines; the probability of an intersection of three lines is the probability that the last point is placed at a specific number, which is 0. The probability is 0 even after applying the union bound on all potential sets of three lines.

Let the graph G include the terminals and intersection points as its vertices, and the segmented lines as its edges. The edge-lengths ℓ_G are set to be the Euclidean distances between the endpoints of the edges.

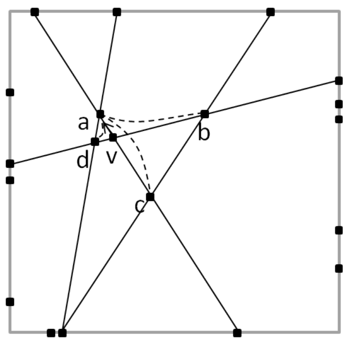

Figure 1: Illustration for contraction of v.

Let v be any of the $\Theta(k^4)$ internal intersection points. Then $\deg_G(v) = 4$, and each of its edges is used in some shortest path between terminals. Let its neighbors be a, b, c, d according to clockwise order. Suppose wlog that we contracted v into a. Then the direct path between b and c that went through v now detours through a, which means an increase of $2d_G(v,a)$ to that path. No alternative shortest path exists, since the graph distances equal the Euclidean distances, and the only shortest path between two points is the straight line between them. See Figure 1 for illustration. Hence, any single oriented contraction is not possible. Since all of the edges are used in some shortest paths, it is clear that no edge or vertex removals can be done either.

Since any further oriented edge-contractions, vertex- and edge-removals can only further increase the lengths of the induced shortest paths, no oriented minor with induced edge-lengths can be faithful to G but G itself, with $\Omega(k^4)$ vertices, as required. ■

This proof hinges on the fact that when we contract a vertex with degree> 2, the edges incident to it get the sum of two edges, clearly larger than the original Euclidean distance. This is not the case when edge-contractions don't dictate the new edge-lengths.

The freedom to select edge lengths changes the problem dramatically. We prove a weaker lower bound of $\Omega(k^2)$ for minors of general graphs, and for family-preserving planar graphs.

Theorem 3.12 *For every k there exists a planar SPR instance $\langle G, \ell_G, R\rangle$ such that every faithful SPR instance $\langle G', \ell_{G'}, R\rangle$ with a planar G' has $|V(G')| \geq \Omega(k^2)$. In other words, $f^*(k, \text{PLANAR}) \geq \Omega(k^2)$.*

Proof Our proof uses a $k \times k$ grid graph with k terminals, whose edge-lengths are chosen so that terminal distances are essentially "linearly independent" of one another. We use this independence to prove that no faithful SPR instance $\langle G', \ell_{G'}, R\rangle$ with a planar G' can have a small vertex-separator. Since G' is planar, we can apply the planar separator theorem [LT79], and obtain the desired lower bound.

For simplicity we shall assume that k is even. Consider the SPR instance $\langle G, \ell_G, R\rangle$ with a grid graph G of size $k \times k$ with vertices (x, y) for $x, y \in [0, k-1]$, the length function ℓ_G such that the length of all horizontal edges $((x, y),(x+1, y))$ is 1, and the length of each vertical edge $((x, y),(x, y+1))$ is

$1+\frac{1}{2^{x^2}\cdot k}$. Let $R_1 = \{(0,y) : y \in [0, \frac{k}{2}-1]\}$, and $R_2 = \{(x,x) : x \in [\frac{k}{2}, k-1]\}$. Let the terminals in the graph be $R = R_1 \cup R_2$, so $|R| = k$. See Figure 2 for illustration.

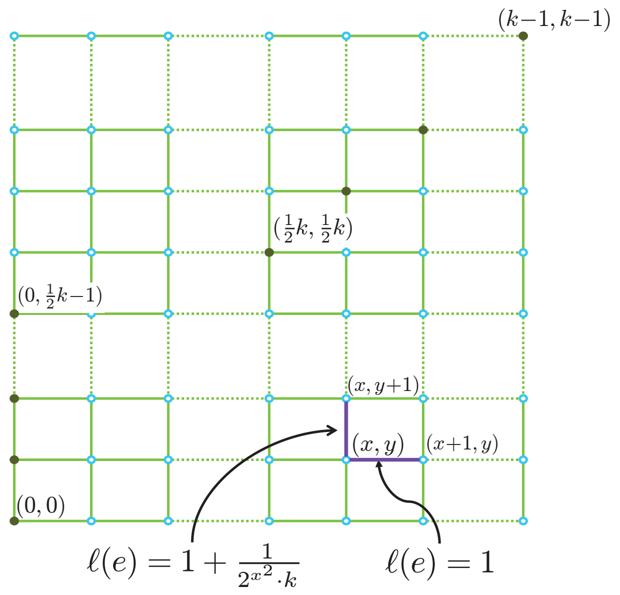

Figure 2: A grid graph G and terminals R.

It is easy to see that the shortest-path between a vertex $(0,y) \in R_1$ and a vertex $(x,x) \in R_2$ includes exactly x horizontal edges and $x-y$ vertical edges. Indeed, such paths have length smaller than $x + (x-y)(1+\frac{1}{k}) \leq 2x - y + 1$. Any other path between these vertices will have length greater than $2x-y+2$. Furthermore, the shortest path with x horizontal edges and $x - y$ vertical edges starting at vertex $(0,y)$ makes horizontal steps before vertical steps, since the vertical edge-lengths decrease as x increases, hence

$$d_G((0,y),(x,x)) = 2x - y + \frac{x-y}{2^{x^2}\cdot k}. \tag{1}$$

Assume towards contradiction that there exists a faithful SPR instance $\langle G', \ell_{G'}, R\rangle$ with a planar graph G' and $|V(G') \leq \frac{k^2}{1600}$ vertices. Since G' is planar, by the weighted version of the planar separator theorem by Lipton and Tarjan [LT79] with vertex-weight 1 on terminals and 0 on non-terminals, there exists a partitioning of V' into three sets A_1, S, and A_2 such that

$w(S) \le |S| \le 2.5 \cdot \sqrt{\frac{k^2}{1600}} < \frac{3k}{40}$, each of A_1 and A_2 has at most $w(A_i) \le \frac{2k}{3}$ terminals, and there are no edges going between A_1 and A_2. Hence, for $i \in \{1, 2\}$ it holds that $w(A_i \cup S) \ge k/3$ and $w(A_i) \ge \frac{k}{3} - \frac{3k}{40} > \frac{k}{4}$.

Without loss of generality, we claim that $A_1 \cap R_1$ and $A_2 \cap R_2$ each have $\Theta(k)$ terminals. To see this, suppose without loss of generality that A_1 is the heavier of the two sets (i.e. $w(A_1) \ge \frac{k}{2} - \frac{3k}{40}$ and $\frac{k}{4} \le w(A_2) \le \frac{k}{2}$). Suppose also that $w(A_2 \cap R_2) \ge w(A_2 \cap R_1)$. Then $w(A_2 \cap R_2) \ge \frac{k}{8}$, and $w(A_2 \cap R_1) \le \frac{1}{2} \cdot w(A_2) \le \frac{k}{4}$, implying that $w(A_1 \cap R_1) \ge w(R_1) - (w(R_1 \cap A_2) + w(R_1 \cap S)) \ge \frac{k}{2} - (\frac{k}{4} + \frac{3k}{40}) = \frac{k}{5}$. In conclusion, without loss of generality it holds that $w(A_1 \cap R_1) \ge \frac{k}{5}$ and $w(A_2 \cap R_2) \ge \frac{k}{8}$. Let $Q_1 \subseteq A_1 \cap R_1$ and $Q_2 \subseteq A_2 \cap R_2$ be two sets with the exact sizes $\frac{k}{5}$ and $\frac{k}{8}$.

Every path between a terminal in Q_1 and a terminal in Q_2 goes through at least one vertex of the separator S. Overall, the vertices in the separator participate in $\frac{k}{8} \times \frac{k}{5}$ paths between Q_1 and Q_2. See Figure 3 for illustration.

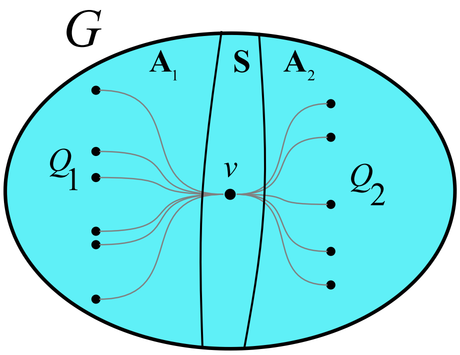

Figure 3: Terminals on different sides connected by paths going through $v \in S$.

We will need the following lemma, which is proved below.

Lemma 3.13 *Let* $\langle G', \ell_{G'}, R \rangle$, S*,* Q_1 *and* Q_2 *be as described above. Then every vertex* $v \in S$ *participates in at most* $|Q_1| + |Q_2| = \frac{k}{5} + \frac{k}{8}$ *shortest paths between* Q_1 *and* Q_2*.*

Applying Lemma 3.13 to every vertex in S, at most $\frac{3k}{40} \cdot \frac{13k}{40} = \frac{39k^2}{1600} < \frac{k^2}{40}$ shortest paths between Q_1 and Q_2 go through S, which contradicts the fact that all $\frac{k}{8} \cdot \frac{k}{5} = \frac{k^2}{40}$ shortest-paths between Q_1 and Q_2 in G' go through the separator, and proves Theorem 3.12. ■

Proof of Lemma 3.13 Define a bipartite graph H on the sets Q_1 and Q_2, with an edge between $(0, y) \in Q_1$ and $(x, x) \in Q_2$ whenever a shortest path in G' between $(0, y)$ and (x, x) uses the vertex v. We shall show that H does not contain an even-length cycle. Since H is bipartite, it contains no odd-length cycles either, making H a forest with $|E(H)| < |Q_1| + |Q_2| = \frac{k}{5} + \frac{k}{8}$, thereby proving the lemma.

Let us consider a potential $2s$-length (simple) cycle in H on the vertices $(0, y_1)$, (x_1, x_1), $(0, y_2)$, (x_2, x_2), ..., $(0, y_s)$, (x_s, x_s) (in that order), for particular $(0, y_i) \in Q_1$ and $(x_i, x_i) \in Q_2$. Every edge $((0, y), (x, x)) \in E(H)$ represents a shortest path in G' that uses v, thus

$$d_G((0, y), (x, x)) = d_{G'}((0, y), v) + d_{G'}(v, (x, x)). \tag{2}$$

If the above cycle exists in H, then the following equalities hold (by convention, let $y_{s+1} = y_1$). Essentially, we get that the sum of distances corresponding to "odd-numbered" edges in the cycle equals the one corresponding to "even-numbered" edges in the cycle.

$$\begin{aligned}
\sum_{i=1}^{s} d_G((0, y_i), (x_i, x_i)) &\overset{(2)}{=} \sum_{i=1}^{s} d_{G'}((0, y_i), v) + \sum_{i=1}^{s} d_{G'}(v, (x_i, x_i)) \\
&= \sum_{i=1}^{s} d_{G'}(v, (0, y_{i+1})) + \sum_{i=1}^{s} d_{G'}((x_i, x_i), v) \\
&\overset{(2)}{=} d_G((x_i, x_i), (0, y_{i+1})).
\end{aligned}$$

Plugging in the distances as described in (1) and simplifying, we obtain

$$\sum_{i=1}^{s} (2x_i - y_i + (x_i - y_i) \cdot \frac{1}{2^{x_i^2} \cdot k}) = \sum_{i=1}^{s} (2x_i - y_{i+1} + (x_i - y_{i+1}) \cdot \frac{1}{2^{x_i^2} \cdot k}),$$

or equivalently,

$$\sum_{i=1}^{s} \frac{y_i}{2^{x_i^2}} = \sum_{i=1}^{s} \frac{y_{i+1}}{2^{x_i^2}}$$

Suppose without loss of generality that $x_1 = \min\{x_i : i \in [1, s]\}$ (otherwise we can rotate the notations along the cycle), and that $y_1 > y_2$ (otherwise we can change the orientation of the cycle). Then we obtain

$$\frac{y_1 - y_2}{2^{x_1^2}} = \sum_{i=2}^{s} \frac{y_{i+1} - y_i}{2^{x_i^2}}.$$

However, since $y_1 > y_2$, the lefthand side is at least $\frac{1}{2^{x_1^2}}$, whereas the righthand side is $\sum_{i=2}^{s} \frac{y_{i+1}-y_i}{2^{x_i^2}} \leq s - 1 \cdot \frac{k}{2^{(x_1+1)^2}} \leq \frac{k^2}{2^{(x_1+1)^2}}$. Therefore it must hold that $2^{2x_1+1} \leq k^2$. Since $x_1 > \frac{k}{2}$ this inequality does not hold for any $k > 4$. Hence, for any s, no cycle of size $2s$ exists in H, completing the proof of Lemma 3.13. ■

This completes the proof of Theorem 3.12. It is easy to see the following corollary of the above theorem.

Corollary 3.14

$$f^{minor}(k, \textsc{AllGraphs}) \geq f^{minor}(k, \textsc{Planar}) \geq f^*(k, \textsc{Planar}) \geq \Omega(k^2)$$

4 α-Approximate Faithfulness

So far we tried to reduce the size of graphs, while retaining their structure as well as the exact distances between given terminals. We showed that this reduction is possible to an extent, but that some SPR instances exist that cannot be reduced to instances where the graphs have less than $\Omega(k^2)$ vertices. For these SPR instances, any graph on the terminals alone has either a different structure or different terminal distances. In this section we discuss the problem of constructing graphs without non-terminals, which retain the structure (or topology) of given graphs (minors or graphs in the same graph family), while allowing the distances to increase by a constant factor. This problem was first introduced in [Gup01], where the following theorem was proved.

Theorem 4.1 (Gupta [Gup01]) *Given an SPR instance $\langle T, \ell_T, R\rangle$ where T is a tree, there exists an SPR instance $\langle T', \ell_{T'}, R\rangle$ such that T' is a tree, $V(T') = R$ and for all $x, y \in R$,*

$$d_T(x, y) \leq d_{T'}(x, y) \leq 8 \cdot d_T(x, y).$$

In other words, this means that $\alpha^*(k, \textsc{Trees}) \leq 8$. It was later stated, in [CGN+06], that the graph T' is in fact a minor of T, implying that $\alpha^{minor}(k, \textsc{Trees}) \leq 8$ as well. In [BG08] it was proved that $\alpha^*(k, \textsc{Outerplnanar}) \leq 15$.

We first conjectured that for some graphs G, there exists no minor (induced oriented minor or otherwise) that retains the terminal distances within any constant factor. In other words, that there exists a graph family $\mathcal{F}$ such that $\alpha^*(k, \mathcal{F}) \to \infty$ as $k \to \infty$. More specifically, we conjectured the following:

Conjecture 4.2 $\alpha^{minor}(k, \textsc{Planar}) \to \infty$ *as* $k \to \infty$.

Our attempt to prove this conjecture failed, and in the process we found a family of metrics $\mathcal{M}$ such that every graph such that its terminal distances induce a metric in $\mathcal{M}$ has an SPR instance $\langle G', \ell_{G'}, R\rangle$ that gives a constant factor approximation, where G' is outerplanar[2].

4.1 The attempt

Let T_k be a complete binary tree with $k = 2^{2^a}$ leaves. Draw it in the plane and order the vertices of each level from left to right. The graph G_k is constructed by taking T_k, and for every level i, adding edges connecting every two consecutive vertices, and also connecting the first and last vertices. It is easy to see that G_k is planar. See figure 4 for illustration. Let ℓ_{G_k} be the uniform edge-length function, assigning the length 1 to all edges.

Number the vertices on the last level by the integers $[0, k-1]$ according to their order. These are the terminals. The minimal number of hops along edges in the last level between terminals number i and j (wlog $j > i$) is denoted by $\|i - j\|$ and equals $\min\{j - i, k - j + i\}$. Notice this is not the shortest path in G_k; in fact, $d_{G_k}(i, j) = \Theta(\log \|i - j\|) + \Theta(1)$. For simplicity we shall assume that the constants are 1, i.e. $d_{G_k}(i, j) = \log(\|i - j\|) + 1$. The actual constants do not change the proof, as the edge-lengths need only be multiplied by them to achieve a construction that approximates the true graph distances.

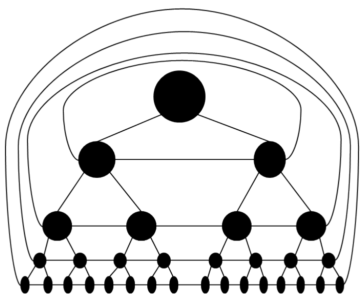

Figure 4: The graph G_2

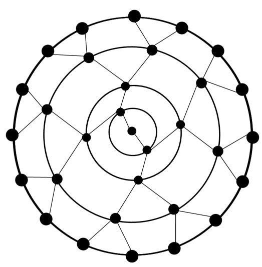

Figure 5: Another embedding of G_2 in the plane

As seen in figures 4 and 5, the graph G_k is planar, and the vertices in the last

[2]An outerplanar graph is a graph that can be drawn in the plane so that all its vertices touch the outer face.

level are all incident to the outer face. Edge-contractions, edge-removals and vertex-removals can all be done in a way that guarantees that the locations of the vertices in the embedding remain fixed, while retaining a legal embedding in the plane. Therefore, any oriented minor of this graph which includes only the terminals (incident to the outer face), is an outerplanar graph.

Theorem 4.3 *For every $k = 2^{2^a}$, the SPR instance $\langle G_k, \ell_{G_k}, R\rangle$ where R is the set of vertices in the last level, numbered by $[0, k-1]$, has an SPR instance $\langle G'_k, \ell_{G'_k}, R\rangle$ such that $V(G'_k) = R$, the graph G'_k is an oriented minor of G_k, $\ell_{G'_k}$ is some edge-length function, and for every $x_1, x_2 \in [0, k-1]$ it holds that $d_G(x_1, x_2) \leq d_{G'}(x_1, x_2) \leq O(1) \cdot d_G(x_1, x_2)$.*

4.1.1 Constructing G'_k:

An oriented minor G' of some graph G, with $V(G') = R$ and created only by oriented edge contractions, can be described as an assignment $h : V(G) \to R$ assigning a terminal to each of the vertices in $V(G)$, where $h(t) = t$ for every $t \in R$, and vertices assigned to the same terminal ($h^{-1}(t)$) induce a connected component in G. As a result of the contraction process, an edge (t_1, t_2) exists in the minor iff there exists an edge between some vertex v assigned t_1 ($v \in h^{-1}(t_1)$) and some vertex u assigned t_2 ($u \in h^{-1}(t_2)$).

The structure of the graph G_k we discuss is that of a complete binary tree of depth 2^a, with additional edges. This gives an intuitive definition to the concepts of levels and subtrees in the graph. We show the contraction process from the bottom of the tree and up, in stages. At the end of stage i, all vertices in the lowest 2^i levels of G_k are contracted into (in other words, have been assigned) the terminals.

Let $G_k|_i$ denote the subgraph of G_k induced on the vertices in the lowest 2^i levels of the graph. This graph can be described as a series of subtrees of depth 2^i with additional edges connecting the vertices of each level. We prove that we can create the assignment h (hence, the minor) while maintaining the following.

Claim 4.4 *The assignment* $h : V(G_k|_i) \to [0, k-1]$ *and the corresponding graph* $G'_k|_i$ *can be created such that:*

1. *Let* $\min(v)$ *be the smallest number of a terminal in* v*'s subtree. Then every vertex* r *at the highest level of* $G_k|_i$ *(closest to the root of* G_k*), gets* $h(r) = \min(r)$*.*

2. $G'_k|_i$ *is a legal oriented minor of* $G_k|_i$*, i.e. vertices assigned the same terminal induce a connected component in* $G_k|_i$*.*

3. *Let* $l \leq i$ *be some integer, and* $p \cdot 2^{2^l} \leq k$ *and* $((p+1) \cdot 2^{2^l} \mod k)$ *be two consecutive multiples of* 2^{2^l}*. Then there in an edge* $(p \cdot 2^{2^l}, (p+1) \cdot 2^{2^l} \mod k)$ *in* $E(G'_k|_i)$ *connecting these terminals. In other words, there exist vertices* $v \in h^{-1}(p \cdot 2^{2^l})$ *and* $u \in h^{-1}((p+1) \cdot 2^{2^l} \mod k)$ *such that* $(v, u) \in E(G_k|_i)$*.*

4. *Let* $1 \leq l \leq i$ *be some integer,* $p \cdot 2^{2^{l-1}} \leq k$ *be some multiple of* $2^{2^{l-1}}$*, and* $p' \cdot 2^{2^l}$ *be the multiple of* 2^{2^l} *that minimizes* $\|p' \cdot 2^{2^l} - p \cdot 2^{2^{l-1}}\|$*. Then there is an edge* $(p \cdot 2^{2^{l-1}}, p' \cdot 2^{2^l})$ *in* $E(G'_k|_i)$ *connecting these terminals. In other words, there exist vertices* $v \in h^{-1}(p \cdot 2^{2^{l-1}})$ *and* $u \in h^{-1}(p' \cdot 2^{2^l})$ *such that* $(v, u) \in E(G_k|_i)$*.*

Proof by induction on i.

i=0: To begin with, the terminals (leaves) are assigned to themselves, and vertices in level $2^0 = 1$ are assigned the their smaller child. Conditions (1) and (2) and satisfied. Condition (3) is satisfied by the edges at level 1, and condition (4) is satisfied vacuously. Note that the cycle edges connect every terminal with a terminal numbered by some multiple of 2, satisfying a slight variation of condition (4) (an endcase resulting from the fact that 1 is the result of $2^{2^{-\infty}}$ and not $2^{2^{-1}}$).

Suppose that the invariant holds for $i-1$. Consider the graph $G_k|_i$ and assign terminals to vertices in its lower half (the vertices of $G_k|_{i-1}$) accordingly. The construction now assigns terminals to the upper half of $G_k|_i$.

Let r be a vertex in the highest level of $G_k|_i$ and let $\mathsf{left\text{-}child}(r)$ and $\mathsf{right\text{-}child}(r)$ be r's children, $\mathsf{right}(r)$ be the next vertex in level 2^i to the right of r (connected to it by an edge) and left(r) be the previous vertex in level 2^i, to the left of r. Assign $\min(r)$ to r, left-child(r) and all of the vertices in left-child(r)'s subtree not yet assigned. Assign $\min(\text{right}(r))$ to right-child(r) and all of the vertices in its subtree not yet assigned. Do this to all the subtrees in $G_k|_i$. See Figure 6 for illustration.

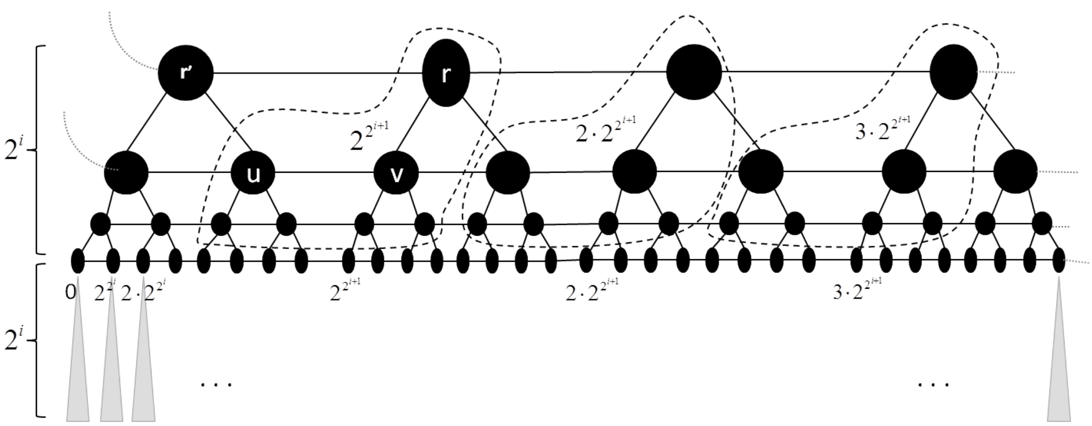

Figure 6: Illustration for assignments h.
Numbers in figure are values of h function.
Vertices inside dashed lines are assigned the same value.
Vertex r is on level 2^{i+1}, $r' = \text{left}(r)$, $v = \text{left-child}(r)$ and $u = \text{right-child}(\text{left}(r))$.

Clearly this assignment satisfies condition (1).

From the induction hypothesis, before adding vertices to $h^{-1}(\min(r))$ it was a connected component. The vertices added to $h^{-1}(\min(r))$ at this stage are r, the vertices in the top half of left-child(r)'s subtree, and the vertices in the top half of right-child(left(r))'s subtree. These vertices induce a connected component due to the subtrees' edges and the edges connecting consecutive trees. This component is connected to the previous connected component, since from the induction hypothesis on (1), the value $\min(r)$ is assigned to a vertex in level 2^{i-1}. Hence, all modified sets, $h^{-1}(\min(r))$ for some such r, induce connected components, proving (2).

Condition (3) holds for $l < i$ according to the induction hypothesis. Since the roots are all assigned to their subtree's minimal terminal, they are assigned consecutive multiples of 2^{2^i}. The edges connecting consecutive roots satisfy condition (3) for $l = i$.

Condition (4) also holds for $l < i$ according to the induction hypothesis. Every multiple of $2^{2^{i-1}}$ is the smallest value in some tree of depth 2^{i-1}, and according to the induction hypothesis some root at level 2^{i-1} is assigned that value. This root is either in the left or right subtree of some root r' at level 2^i, and its parent is assigned accordingly - either to the largest multiple of 2^{2^i} smaller than it, or to the smallest multiple of 2^{2^i} larger than it, which proves (4) and the claim. ■

4.1.2 Proving the approximation factor

Let G'_k be the minor induced by the assignment h from the previous section. We set the length $\ell_{G'_k}(i,j) = d_G(i,j)$ to every $(i,j) \in E(G'_k|_i)$. We hereby prove the following claim:

Claim 4.5 *The SPR instance $\langle G'_k, \ell_{G'_k}, [0, k-1]\rangle$ is 12-approximately faithful to the SPR instance $\langle G_k, \ell_{G_k}, [0, k-1]\rangle$.*

Proof of Claim Since $\ell_{G'_k}(v,u) = d_G(v,u)$ to every $(v,u) \in E(G'_k|_i)$, distances in G'_k dominate those in G_k. To prove that the increase is by at most a factor of 12, we first prove the following lemma about the distance to a multiple of 2^{2^i} for any i.

Lemma 4.6 *Suppose terminal x is in $[p \cdot 2^{2^i}, (p+1) \cdot 2^{2^i}]$. Then either $d_{G'_k}(x, p \cdot 2^{2^i}) \leq 2 \cdot 2^i + 2$, or $d_{G'_k}(x, (p+1) \cdot 2^{2^i}) \leq 2 \cdot 2^i + 2$.*

Proof of Lemma We use induction on i.

i=0: Since the terminals are assigned to themselves, all edges from the last level exist also in G'_k, and connect odd-numbered terminals to even-numbered terminals. The lemma holds.

Induction step: Consider a terminal $x \in [p \cdot 2^{2^i}, (p+1) \cdot 2^{2^i}]$. There exists $p' \in [0, 2^{2^{i-1}} - 1]$ such that $x \in [p \cdot 2^{2^i} + p' 2^{2^{i-1}}, p \cdot 2^{2^i} + (p'+1) 2^{2^{i-1}}]$. From the induction hypothesis, to one of these endpoints, $y \in \{p \cdot 2^{2^i} + p' 2^{2^{i-1}}, p \cdot 2^{2^i} + (p'+1) 2^{2^{i-1}}\}$, it holds that $d_{G'_k}(x,y) \leq 2 \cdot 2^{i-1} + 2$.

Let z be the terminal numbered by a multiple of 2^{2^i} that minimizes $\|y-z\|$. Then $z \in \{p2^{2^i}, (p+1)2^{2^i}\}$. From (4) in Claim 4.4, the edge (y,z) is in $E(G'_k)$, with the length $\ell_{G'_k}(y,z) = d_{G_k}(y,z)$. Hence,

$$\begin{aligned} d_{G'_k}(x,z) &\leq d_{G'_k}(x,y) + d_{G'_k}(y,z) \\ &\leq 2 \cdot 2^{i-1} + 2 + \log(2^{2^i}) + 1 \\ &\leq 2 \cdot 2^i + 2. \end{aligned}$$

■

Let $x_1, x_2 \in [0, k-1]$ be two terminals. Suppose wlog that $x_2 > x_1$ and that $\|x_2 - x_1\| = x_2 - x_1$ (since the construction is symmetric, this holds for all other cases as well). Let s denote $x_2 - x_1$. Suppose $2^{2^t} \leq s \leq 2^{2^{t+1}}$. Then in $G_k|_{t+1}$, x_1 and x_2 were either in the same subtree, or in consecutive subtrees. In both cases we can denote by q the largest multiple of $2^{2^{t+1}}$ that is smaller than x_2, and have $x_2 - q$ and $|x_1 - q|$ (x_1 may be greater or smaller than q) smaller than $2^{2^{t+1}}$. Using Lemma 4.6, condition (3) from Claim 4.4, and the edge-lengths, we have $d_{G'_k}(x_1, q) \leq 3 \cdot 2^{t+1} + 2$ and the same for x_2. Using the triangle inequality, $d_{G'_k}(x_1, x_2) \leq 6 \cdot 2^{t+1} + 4 = 12 \cdot 2^t + 4$.

The original distance is $d_G(x_1, x_2) \geq \log s \geq \log 2^{2^t} = 2^t$.
Therefore $d_{G'_k}(x_1, x_2) \leq 12 \cdot 2^t + 4 \leq 12 d_G(i,j)$, proving Claim 4.5. ■

Condition (2) in Claim 4.4 implies that G'_k is a minor of G_k, and along with Claim 4.5 proves Theorem 4.3.

4.2 A Generalization

In the previous section, we constructed an outerplanar graph (specifically, a minor of a given SPR instance) that approximates a specific distance metric on the terminals. Assuming a numbering of the terminals, the metric assigned two terminals x_1 and x_2 the logarithm of the (cyclic) difference between their indices, $\|x_1 - x_2\|$. The construction we proved used the structure of the original SPR instance, but the outcome is an outerplanar graph that mimics the metric $d_G|_{R \times R}$, independently from the structure. In this section we prove for a broader family of metrics, that they can be embedded into outerplanar graphs, using similar constructions to that of the previous section.

When viewing the construction of the outerplanar graph from the top-down instead of bottom-up, we realize that the same graph can also be constructed as follows: Split the vertices into two segments: $[0, k/2]$ and $[k/2, k]$ (where $k \equiv 0$ in the cyclic structure), by adding the edge $(0, k/2)$. Further split these segments into multiple segments of equal size, by choosing evenly spaced vertices in each segment, and adding edges between two consecutive chosen vertices, and between each chosen vertex and the closest endpoint of the containing segment. Continue this process until the segments can no longer be split. This construction gives the same graph we proved for the logarithm function, but now the construction applies also to any metric M on $[0, k-1]$ for which $d_M(i,j) = g(\|i-j\|)$ for some monotonic function g.

In fact, we can generalize this result even further, and prove the following theorem.

Theorem 4.7 *Let M be a metric on $[0, k-1]$ such that $\forall i < x < j$ it holds that $d(i,x) \leq d(i,j)$ and that $d(x,j) \leq d(i,j)$. Then there exists an SPR instance $\langle G', \ell_{G'}, [0, k-1] \rangle$ such that $V(G') = [0, k-1]$, $d_{G'}$ is 34-approximately faithful to M, and the graph G' can be drawn in the plane so that its terminals lie on a line, and all edges are drawn above the line.*

4.2.1 Constructing the graph G':

We will need the following definition. For $r > 0$, an r-net of the metric X is a subset $N \subseteq X$ such that:

1. If $p_1, p_2 \in N$ then $d_X(p_1, p_2) > r$.
2. For every $v \in X$ there is $p \in N$ that **covers** v, i.e. $d_X(v, p) \leq r$.

Such a net always exists. It can be constructed by picking vertices (in any order) that are not yet covered by any vertex in N and adding them to it, until all vertices are covered.

Given a metric d_M as in Theorem 4.7, we construct the graph G' according to Algorithm 2.

Algorithm 2 Construct $\langle G', \ell_{G'}, [0, k-1] \rangle$ (M)

1. Normalize distances in the graph so that the smallest one equals 2.
2. Construct 2^i-nets, denoted N^i, for i from $\log(d_M(0, k-1))$ to 0, so that $N^{i-1} \supseteq N^i$. Let $v_1^i, v_2^i, ..., v_{l_i}^i$ denote the vertices of the 2^i-net in increasing order.
3. Set $V(G') = [0, k-1]$ and $E(G') = \emptyset$.
4. For $i = \log(d_M(0, k-1)) - 1$ to 0 do
   (a) For every $j \in [1, l_i - 1]$ add to $E(G')$ the edge (v_j^i, v_{j+1}^i).
   (b) For every $j \in [1, l_i]$ add to $E(G')$ an edge connecting v_j^i with the vertex v_a^{i+1} that minimizes $d_M(v_j^i, v_a^{i+1})$. Break ties by minimizing $|v_j^i - v_a^{i+1}|$. Further ties can be broken arbitrarily.
5. For every edge $(v, u) \in E(G')$ set $\ell_{G'}(v, u) = d_M(v, u)$.

Claim 4.8 *The graph G' constructed above can be drawn in the plane so that its terminals lie on a line, and all edges are drawn above it.*

Proof of Claim Embed the vertices $[0, k-1]$ in a line, in order. An edge (x_1, x_2) induces a segment of the line, denoted $[x_1, x_2]$, which includes all the vertices $t \in [x_1, x_2]$. It is easy to see that if the intersection of two segments $[x_1, x_2]$ and $[y_1, y_2]$ is either $[x_1, x_2]$, $[y_1, y_2]$ a single vertex or $\emptyset$, then we can draw the edges of G' above the line so that no two edges intersect.

We show by induction on i from $\log(d_M(0, k-1))$ to 0, that the segments induced by edges added in round i are contained in or lie completely outside of segments induced by the edges from round $i+1$, and also don't intersect each other. At the beginning, the net $N^{\log(d_M(0,k-1))}$ includes a single vertex, and no edges are added. Suppose that there are no intersections by round $i+1$. Since $N^{i+1} \subseteq N^i$, the segments induced by consecutive vertices in N^i, and induced by the edges added in step 4a, are contained in larger segments in N^{i+1}.

Because of monotonicity, edges added in step 4b connect a vertex $v_a^i \in N^i$ to a vertex $x \in N^{i+1}$ which minimizes $|v_a^i - x|$. There is some b such that $v_a^i \in [v_b^{i+1}, v_{b+1}^{i+1}]$. Then $x = v_b^{i+1}$ or v_{b+1}^{i+1}. Hence, the segment induced by the edge (v_a^i, x) is contained inside the segment $[v_b^{i+1}, v_{b+1}^{i+1}]$, and from the induction hypothesis is either contained in or lies completely outside of all the other segments from previous rounds.

It is easy to see that edges added in step 4a induce segments that are contained or equal to those induced by edges added in step 4b, and that two edges added in step 4a induce segments that intersect in at most one vertex. Consider the segments induced by edges added in step 4b. If these segments are contained in two distinct segments from round $i+1$, they are disjoint. Suppose that they belong to the same segment $[v_b^{i+1}, v_{b+1}^{i+1}]$ from round $i+1$. If both edges connect to the same endpoint of that segment, then one of the segments is contained in the other. Otherwise, denote the edges by (v_b^{i+1}, v_x^i) and (v_y^i, v_{b+1}^{i+1}). In this case, it must be that $v_y^i > v_x^i$. Otherwise, from the construction, $d_M(v_b^{i+1}, v_x^i) \leq d_M(v_x^i, v_{b+1}^{i+1})$ and $d_M(v_y^i, v_{b+1}^{i+1}) \leq d_M(v_b^{i+1}, v_y^i)$. From monotonicity of M, $d_M(v_x^i, v_{b+1}^{i+1}) \leq d_M(v_y^i, v_{b+1}^{i+1})$. Put together, we get that $d_M(v_b^{i+1}, v_x^i) \leq d_M(v_x^i, v_{b+1}^{i+1}) \leq d_M(v_y^i, v_{b+1}^{i+1}) \leq d_M(v_b^{i+1}, v_y^i)$, and due to the monotonicity of M, all these terms are equal, and the edges contradict the tie-breaking rule. Since $v_y^i > v_x^i$, the segments $[v_b^{i+1}, v_x^i]$ and $[v_y^i, v_{b+1}^{i+1}]$ are disjoint. Two edges added in round i don't intersect, and induce segments contained in segments from round $i+1$, proving the induction and the claim.
■

4.2.2 Proving the approximation factor

The proof is very similar to that in Section 4.1.2, and uses a similar lemma.

Lemma 4.9 *For every $v \in V(G')$ every vertex $v_j^i \in N^i$ at distance $d_M(v, v_j^i) \leq 2^i$ is at distance $d_{G'}(v, v_j^i) \leq 7 \cdot 2^i$.*

Proof of Lemma The proof is by induction on i. When $i = 0$ the only vertex at distance smaller than 1 from v is v itself, at distance 0.

Suppose that the lemma holds for $i-1$. Because N^{i-1} is a net, there exists a vertex $v_{j'}^{i-1} \in N^{i-1}$ such that $d_M(v, v_{j'}^{i-1}) \leq 2^{i-1}$. From the triangle inequality, the distance $d_M(v_j^i, v_{j'}^{i-1}) \leq 2^i + 2^{i-1}$. The vertex $v_{j'}^{i-1}$ is connected in G' by an edge to the closest vertex to it in N^i: either v_j^i, v_{j-1}^i, or v_{j+1}^i. Since $v_{j'}^{i-1}$ is covered by a vertex in N^i, the closest vertex to it in N^i is within distance 2^i.

If $v_{j'}^{i-1}$ is connected to v_j^i then using the induction hypothesis,

$$\begin{aligned} d_{G'}(v, v_j^i) &\leq d_{G'}(v, v_{j'}^{i-1}) + d_M(v_{j'}^{i-1}, v_j^i) \\ &\leq 7 \cdot 2^{i-1} + 2^i \\ &\leq 7 \cdot 2^i \end{aligned}$$

Otherwise, suppose wlog that $v_{j'}^{i-1}$ is connected to v_{j-1}^i. We get:

$$\begin{aligned} d_{G'}(v, v_j^i) &\leq d_{G'}(v, v_{j'}^{i-1}) + d_M(v_{j'}^{i-1}, v_{j-1}^i) + d_M(v_{j-1}^i, v_j^i) && (3) \\ &\leq 7 \cdot 2^{i-1} + 2^i + d_M(v_{j-1}^i, v_{j'}^{i-1}) + d_M(v_{j'}^{i-1}, v) + d_M(v, v_j^i) && (4) \\ &\leq 7 \cdot 2^{i-1} + 2^i + 2 \cdot 2^i + 2^{i-1} && (5) \\ &= 7 \cdot 2^i. && (6) \end{aligned}$$

Using (1) the triangle inequality, (2) the induction hypothesis, the edge $(v_{j'}^{i-1}, v_{j-1}^i)$ and the triangle inequality, and (3) the lemma's statement, and the facts that $v_{j'}^{i-1}$ covers v and v_{j-1}^i covers $v_{j'}^{i-1}$. ■

We can now prove the approximation factor.

Claim 4.10 *The SPR instance $\langle G', \ell_{G'}, [0, k-1] \rangle$ constructed above induces distances such that for every two vertices $x_1, x_2 \in [0, k-1]$,*

$$d_M(x_1, x_2) \leq d_{G'}(x_1, x_2) \leq 34 d_M(x_1, x_2).$$

Proof of Claim Denote by a the power for which $2^a \leq d_M(x_1, x_2) < 2^{a+1}$ and suppose that $x_1 < x_2$. Since the distance between any two vertices in $[x_1, x_2]$ is smaller than 2^{a+1}, there can be at most one vertex in $[x_1, x_2]$ that is in I_{a+1}.

- If there is a vertex $v_p^{a+1} \in I_{a+1} \cap [x_1, x_2]$, then it is at distance smaller than 2^{a+1} from both x_1 and x_2. Using Lemma 4.9, $d_{G'}(x_1, v_p^{a+1}) \leq 7 \cdot 2^{a+1}$ and $d_{G'}(x_2, v_p^{a+1}) \leq 7 \cdot 2^{a+1}$, hence $d_{G'}(x_1, x_2) \leq 14 \cdot 2^{a+1} \leq 28 d_M(x_1, x_2)$.

- If there isn't such a vertex, then one of the following holds:

  – Both x_1 and x_2 are covered by the same vertex in I_{a+1}, using similar inequalities we get $d_{G'}(x_1, x_2) \leq 14 \cdot 2^{a+1} \leq 28 d_M(x_1, x_2)$.

  – x_1 is covered by a vertex $v_p^{a+1} < x_1$, and x_2 is covered by the vertex $v_{p+1}^{a+1} > x_2$. From the triangle inequality, $d_M(v_p^{a+1}, v_{p+1}^{a+1}) \leq 3 \cdot 2^{a+1}$. Using Lemma 4.9 and the edge $(v_p^{a+1}, v_{p+1}^{a+1})$ from the construction, we get

$$\begin{aligned} d_{G'}(i,j) &\leq 2 \cdot 7 \cdot 2^{a+1} + 3 \cdot 2^{a+1} \\ &= 17 \cdot 2^{a+1} \\ &\leq 34 d_M(i,j). \end{aligned}$$

■

Let M be a metric on $[0, k-1]$. A rotation i of the vertices is a re-ordering $[0_i, k-1_i]$ such that $0_i \equiv i$, $k - 1_i - i \equiv k - 1$ and $k - 1_i - (i-1) + j \equiv j$. We say that the segment $[v, u]$ is *contained* in the segment $[x, y]$ if for some rotation i of the the vertices it holds that $x_i \leq v_i < u_i \leq y_i$, $\|u_i - v_i\| = u_i - v_i$ and $\|y_i - x_i\| = y_i - x_i$.

We say that M is circularly monotonic in regards to containment, if for every segment $[v, u]$ contained in the segment $[x, y]$ it holds that $d_M(v, u) \leq d_M(x, y)$.

Then using a similar construction to that for Theorem 4.7, we get that every metric M that is circularly monotonic in regards to containment, can be embedded in an outerplanar graph with a constant distortion to its values.

5 Acknowledgements

It is a pleasure to thank the many people who made this thesis possible. I am heartily thankful to my supervisor Dr. Robert Krauthgamer, for his patient and insightful guidance, great advice, and friendly encouragement. Working with him and learning from him were a valuable and enjoyable experience.

I would like to thank the faculty and staff of the Department of Computer Science and Applied Mathematics for creating an inspiring atmosphere for academic pursuits, and to my fellow students and friends for making this atmosphere a warm and inviting one. This department is filled with friendly and welcoming faces, and it has been a pleasure to belong to it.

Finally, my love and thanks go to my family, whose interest and encouragement have been invaluable to me. Saying I wouldn't be here today if it weren't for them is more than just a tautology, and I can't thank them enough for all their support.

References

[BG08] A. Basu and A. Gupta. Steiner point removal in graph metrics. Unpublished Manuscript, available from `http://www.math.ucdavis.edu/~abasu/papers/SPR.pdf`, 2008.

[CE06] D. Coppersmith and M. Elkin. Sparse sourcewise and pairwise distance preservers. *SIAM J. Discrete Math.*, 20:463–501, 2006.

[CGN$^+$06] C. Chekuri, A. Gupta, I. Newman, Y. Rabinovich, and A. Sinclair. Embedding k-outerplanar graphs into ℓ_1. *SIAM J. Discret. Math.*, 20(1):119–136, 2006.

[CSWZ98] S. Chaudhuri, K. V. Subrahmanyam, F. Wagner, and C. D. Zaroliagis. Computing mimicking networks. In *ICALP*, volume 1443 of *Lecture Notes in Computer Science*, pages 556–567. Springer, 1998.

[CXKR06] T. Chan, D. Xia, G. Konjevod, and A. Richa. A tight lower bound for the Steiner point removal problem on trees. In *9th International Workshop on Approximation, Randomization, and Combinatorial Optimization*, volume 4110 of *Lecture Notes in Computer Science*, pages 70–81. Springer, 2006.

[EGK$^+$10] M. Englert, A. Gupta, R. Krauthgamer, H. Räcke, I. Talgam-Cohen, and K. Talwar. Vertex sparsifiers: New results from old techniques. In *13th International Workshop on Approximation, Randomization, and Combinatorial Optimization*, volume 6302 of *Lecture Notes in Computer Science*, pages 152–165. Springer, 2010.

[GH61] R. E. Gomory and T. C. Hu. Multi-terminal network flows. *Journal of the Society for Industrial and Applied Mathematics*, 9:551–570, 1961.

[Gup01] A. Gupta. Steiner points in tree metrics don't (really) help. In *12th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 220–227. SIAM, 2001.

[HKNR95] T. Hagerup, J. Katajainen, N. Nishimura, and P. Ragde. Characterizations of k-terminal flow networks and computing network flows in partial k-trees. In *SODA'95*, pages 641–649, 1995.

[KZ12] R. Krauthgamer and T. Zondiner. Preserving terminal distances using minors. Technical Report arXiv:1202.5675, Feb 2012.

[LT79] R. J. Lipton and R. E. Tarjan. A separator theorem for planar graphs. *SIAM J. Appl. Math.*, 36(2):177–189, 1979.