

Efficient Vertex-Label Distance Oracles for Planar Graphs ^{*†}

Shay Mozes Eyal E. Skop

June 28, 2021

Abstract

We consider distance queries in vertex-labeled planar graphs. For any fixed $0 < \epsilon \leq 1/2$ we show how to preprocess a directed planar graph with vertex labels and arc lengths into a data structure that answers queries of the following form. Given a vertex u and a label λ return a $(1 + \epsilon)$ -approximation of the distance from u to its closest vertex with label λ . For a directed planar graph with n vertices, such that the ratio of the largest to smallest arc length is bounded by N , the preprocessing time is $O(\epsilon^{-2} n \lg^3 n \lg(nN))$, the data structure size is $O(\epsilon^{-1} n \lg n \lg(nN))$, and the query time is $O(\lg \lg n \lg \lg(nN) + \epsilon^{-1})$. We also point out that a vertex label distance oracle for undirected planar graphs suggested in an earlier version of this paper is incorrect.

1 Introduction

Imagine you are driving your car and suddenly see you are about to run out of gas. What should you do? Obviously, you should find the closest gas station. This is the *vertex-label distance query problem*. Various software applications like Waze and Google Maps attempt to provide such a functionality. The idea is to preprocess the locations of service providers, such as gas stations, hospitals, pubs and metro stations in advance, so that when a user, whose location is not known a priori, asks for the distance to the closest service provider, the information can be retrieved as quickly as possible. A dual situation is, for example, when a taxi company wants to dispatch a taxi from the station closest to the location where the taxi is required. Clearly, this problem can be solved using a *vertex-label distance oracle* by transposing the original graph.

We study this problem from a theoretical point of view. We model the network as a planar graph with labeled vertices (e.g., a vertex labeled as a gas

^{*}This work was partially supported by Israel Science Foundation grants 794/13 and 592/17, and by the Israeli ministry of absorption.

[†]An extended abstract of this work was presented in the 13th Workshop on Approximation and Online Algorithms (WAOA 2015), held in Patras, Greece, 17-18 September 2015.

station). We study distance oracles for such graphs. A *vertex-label distance oracle* is a data structure that represents the input graph and can be queried for the distance between any vertex and the closest vertex with a desired label. We consider approximate distance oracles, which, for any given fixed parameter $\epsilon > 0$, return a distance estimate that is at least the true distance queried, and at most $(1+\epsilon)$ times the true distance (this is known as a $(1+\epsilon)$ -stretch). One would like an oracle with the following properties: queries should be answered quickly, the oracle should consume little space, and the construction of the oracle should take as little time as possible. We use the notation $\langle O(S(n))_{space}, O(T(n))_{time} \rangle$ to express the dependency of the space requirement and query time of a distance oracle on n , the number of vertices in the graph.

Our results and approach For directed planar graphs we give a $(1 + \epsilon)$ -stretch $\langle O(\epsilon^{-1}n \lg n \lg(nN))_{space}, O(\lg \lg n \lg \lg(nN) + \epsilon^{-1})_{time} \rangle$ vertex-label distance oracle whose construction time is $O(\epsilon^{-2}n \lg^3 n \lg(nN))$. Throughout the paper, N is the ratio of the largest to smallest arc length. To the best of our knowledge, no non-trivial directed vertex-label distance oracles were proposed prior to the current work.

Consider a vertex-vertex distance oracle for a graph with label set L . If the oracle works for general directed graphs then the vertex-label problem can be solved easily; add a distinct apex v_λ for each label $\lambda \in L$, and connect every λ -labeled vertex to v_λ with a zero length arc. Finding the distance from a vertex u to label λ is now equivalent to finding the distance between u and v_λ . The main difficulty in applying this approach to oracle for directed planar graphs is that adding apices breaks planarity. In particular, it affects the separability of the graph. Thus, the reduction does not work with oracles that depend on planarity or on the existence of separators.

Our contribution is in realizing and showing that the internal workings of vertex-vertex distances oracles for planar graphs due to Thorup [23] can be extended to support vertex labels. Achieving this modification is non-trivial since introducing the apices needs to be done in a manner that guarantees correctness without compromising efficiency. Thorup's oracles rely on the existence of fundamental cycle separators in planar graphs, a property that breaks when apices are added to the graph. We observe, however, that once the graph is separated, Thorup's oracle does not depend on planarity. We therefore postpone the addition of the apices till a later stage in the construction of the distance oracle, when the graph has already been separated. We show that, nonetheless, approximate distances from any vertex to any label in the entire graph can be efficiently approximated.

An earlier version of this paper claimed a simplified and more efficient vertex label distance oracle for undirected planar graphs. This claim turns out to be incorrect, as we briefly explain in Section 7.

2 Related Work

We summarize related work on approximate distance oracles. For general graphs, no (2)-stretch approximate vertex-vertex distance oracles with nearly-linear space consumption are likely to exist [20]. For any integer $k \geq 2$, Thorup and Zwick [24] presented a $(2k-1)$ -stretch $\langle O(kn^{1+1/k})_{space}, O(k)_{time} \rangle$ distance oracle for undirected graphs whose construction time is $O(kmn^{1/k})$. Wulff-Nilsen [26] achieved the same result with preprocessing of $O(kn^{1+\frac{c}{k}})$ for a universal constant c . Several more improvements of [24] have been obtained for unweighted or sparse graphs ([2], [3], [4]). The current state of the art is due to Chechik [6]. She presented a $(2k-1)$ -stretch $\langle O(n^{1+1/k})_{space}, O(1)_{time} \rangle$ distance oracle construction algorithm for undirected graphs with $O(n^2 + m\sqrt{n})$ construction time.

In contrast, vertex-vertex oracles for planar graphs with stretch less than 2 have been constructed. Thorup [23] gave a $\langle O(\epsilon^{-1}n \lg n \lg(nN))_{space}, O(\lg \lg(nN) + \epsilon^{-1})_{time} \rangle$ $(1+\epsilon)$ -stretch distance oracle for directed planar graphs, and a $(1+\epsilon)$ -stretch $\langle O(\epsilon^{-1}n \lg n)_{space}, O(\epsilon^{-1})_{time} \rangle$ (simplified) distance oracle for undirected planar graphs.

Our result is based on Thorup's directed oracle, which is described in section 4. Klein [14] independently gave a distance oracle for undirected planar graphs with the same bounds. Kawarabayashi, Klein and Sommer [12] have shown a $\langle O(n)_{space}, O(\epsilon^{-2} \lg^2(n))_{time} \rangle$ undirected $(1+\epsilon)$ -stretch distance oracle constructed in $O(n \lg^2 n)$ time, inspired by [23]. They give a trade-off of $\langle O(\frac{\epsilon^{-1}n \lg n}{\sqrt{r}})_{space}, O(r + \sqrt{r}\epsilon^{-1} \lg n)_{time} \rangle_{\forall r \leq n}$ oracle algorithms. Sommer et al. [13] have shown better tradeoffs for oracles for undirected planar graphs. For the case where $N \in poly(n)$, they achieve $(1+\epsilon)$ -stretch $\langle O^*(n \lg n)_{space}, O^*(\epsilon^{-1})_{time} \rangle$ oracle, where O^* hides $\lg(\epsilon^{-1})$ and $\lg^* n$ factors.

The vertex-label distance query problem was introduced by Hermelin, Levy, Weimann and Yuster [11]. For any integer $k \geq 2$, they gave a $(4k-5)$ -stretch $\langle O(kn^{1+1/k})_{space}, O(k)_{time} \rangle$ vertex-label distance oracle (expected space) for undirected general (i.e., non-planar) graphs. This is not efficient when the number l of distinct labels is $o(n^{1/k})$ (since the trivial solution of storing all pairwise vertex-label distances is better in that case). They also presented a (2^k-1) -stretch $\langle O(knl^{1/k})_{space}, O(k)_{time} \rangle$ oracle for undirected graphs. Chechik [5] improved the latter result. In the same paper she also presented a $(4k-5)$ -stretch $\langle O(knl^{1/k})_{space}, O(k)_{time} \rangle$ (expected space) oracle for undirected graph.

For undirected planar graphs, Li, Ma and Ning [16], building on [14], constructed a $(1+\epsilon)$ -stretch vertex-labeled oracle with $\langle O(\epsilon^{-1}n \lg n)_{space}, O(\epsilon^{-1} \lg n \lg \Delta)_{time} \rangle$ bounds. Here, Δ is the (hop) diameter of the graph, which can be $\theta(n)$. It is also shown in [16] how to avoid the $\lg \Delta$ factor when $\Delta = O(\lg n)$. The construction time of their oracle is $O(\epsilon^{-1}n \lg^2 n)$.

Łącki, Oćwieja, Pilipczuk, Sankowski, and Zych [18] developed dynamic vertex-labeled distance oracles for undirected general and planar graphs, and used them to maintain approximate solutions for dynamic Steiner and subgraph TSP problems. They describe a generic scheme for converting certain distance

oracles for undirected planar graphs into dynamic vertex-label distance oracle. Applying their scheme to one of the slower variants of Thorup’s distance oracles, they obtained a $\langle O(\epsilon^{-1}n \lg n \lg(nN))_{space}, O(\epsilon^{-1} \lg n \lg(nN))_{time} \rangle$ $(1 + \epsilon)$ -stretch undirected vertex-labeled distance oracle that also supports merging labels.

Another related work is by Abraham, Chechik, Krauthgamer and Wieder [1], who considered approximate nearest neighbor search in planar graph metrics. This is the special case of vertex-labeled distance oracle with only one label. For this easier problem they obtained a data structure whose size is nearly linear in the number of labeled vertices. However, they assume an *exact* vertex-vertex distance oracle is provided.

To the best of our knowledge, no non-trivial directed vertex-label distance oracles were proposed prior to the current work.

3 Preliminaries

In this paper we only deal with connected graphs. This is without loss of generality since each weakly connected component can be handled separately. For a graph G , we denote by $V(G)$ and $E(G)$ the set of vertices and arcs of G , respectively. Throughout the paper, all graphs are *directed* unless stated otherwise. We write $e = uv$ to denote an arc from vertex u to vertex v . We use the term edges when dealing with undirected graphs, or when we wish to ignore the orientation of arcs in directed graphs.

Let $\delta : E(G) \rightarrow \mathbb{R}^+$ be a function assigning lengths to the arcs. We assume that the lengths are normalized so that the smallest arc length is 1. The *length* of a path is the sum of lengths of its arcs. For $u, v \in V(G)$, the *distance* between u and v in G , denoted $\delta_G(u, v)$, is the length of a shortest u -to- v path in G . We denote by N the maximum length of an arc in G . Thus $\delta_G(\cdot, \cdot) = O(nN)$. We assume there are no parallel arcs since it suffices to keep just the arc with minimum length within each set of parallel arcs.

A path or a cycle P is *simple* if each vertex is the endpoint of at most two arcs of P . The concatenation of two paths P_1 and P_2 , where the last vertex of P_1 is the first vertex of P_2 , is denoted $P_1 \circ P_2$. Two paths P and Q *intersect* if $V(P) \cap V(Q) \neq \emptyset$. For a simple path Q and a vertex set $U \subseteq V(Q)$ with $|U| \geq 2$, we define \bar{Q} , the *reduction* of Q to U as a path whose vertices are U . Consider the vertices of U in the order in which they are visited by Q . For every two consecutive vertices u_1, u_2 of U in this order, there is an arc u_1u_2 in \bar{Q} whose length is the length of the u_1 -to- u_2 subpath of Q .

Let $L = \{\lambda_i\}_{i=1}^l$ be a set of l labels. A vertex-labeled graph is a graph G equipped with a function $f : V(G) \rightarrow L$. We define $V_\lambda = \{v \in V(G) | f(v) = \lambda\}$ to be the set of vertices with label λ . For a vertex-labeled G and $\lambda \in L$, we define $\delta_G(u, \lambda) = \min_{w \in V_\lambda} \delta_G(u, w)$ to be the distance in G from u to the closest λ -labeled vertex.

A *vertex-label distance oracle* is a data structure that, for a specific vertex-labeled graph G , given a vertex $v \in V(G)$ and a label $\lambda \in L$, outputs (an

approximation of) $\delta_G(v, \lambda)$. We note that this problem is a generalization of the basic distance oracle problem in which each vertex is given a unique label. Constructing an $O(nl)$ -space vertex-label distance oracle is trivial, by precomputing and storing the distance between each vertex and each possible label. The goal is, therefore, to devise an oracle which requires $o(nl)$ space, while allowing fast queries.

We assume the reader is familiar with basic definitions and properties of planar graphs, such as the definitions of planar embeddings, and planar duality. Let G be a planar embedded graph. Each vertex of G is embedded to a point in the plane, and each arc is embedded as a curve in the plane between the images of its endpoints, such that the images of distinct arcs are internally disjoint. The faces of G are the maximally connected regions of the plane after removing the image of the vertices and arcs of G . Each face is identified with the set of arcs and vertices on its boundary. We denote by G^* the planar dual of G . The vertices of G^* are the faces of G , and the arcs of G^* are in one-to-one correspondence with the arcs of G . We can therefore refer to the same arc e in both the primal G and the dual G^* . For a graph G , we define the size of the graph as $|G| = |V(G)| + |E(G)|$. An immediate consequence of Euler's formula is that in a planar graph G where each face of G is of size at least 3 (i.e. there are no parallel edges and no self loops), $|E(G)| = O(|V(G)|)$. The number of faces is also $O(|E(G)|)$. Therefore, for planar graph G , $|G| = O(|V(G)|) = O(n)$.

An undirected cycle C in a directed graph G is a set of arcs in G that, when regarded as undirected edges, form a cycle. Let G be a directed planar embedded graph, and let f_∞ be the infinite face of G . Let C be a simple undirected cycle in G . The cycle C partitions the plane into two regions. A face f is *enclosed* by C if it belongs to the part of the partition that does not contain the infinite face. A vertex $u \in V(G)$ is enclosed by C if it is incident to a face enclosed by C . It is *strictly enclosed* if, in addition, u is not a vertex of C . An arc $e \in E(G)$ is strictly enclosed by C if both faces incident to e are enclosed by C . An arc $e \in E(G)$ is enclosed by C if $e \in E(C)$ or if e is strictly enclosed by C . The *interior* of C is the subgraph induced on G by the edges enclosed by C . The *exterior* of C is the subgraph induced on G by the edges not strictly enclosed by C . Note that C belongs to both the interior and exterior of G . The strict interior and strict exterior of G are, respectively, the interior and exterior of G without the vertices and edges of C .

The following is a specialized statement of the Jordan curve theorem for planar embedded graphs.

Proposition 1 *Let G be a planar graph. Let C be an undirected simple cycle in G . Any path in G between a vertex in the interior of C and a vertex in the exterior of C contains a vertex of C .*

Let T be a rooted spanning tree of an undirected graph G . For $u \in V(G)$, let $T[u]$ denote the unique root-to- u path in T . We call $T[u]$ a branch of T . Let $u_1, u_2 \in V(G)$ be two vertices such that $e = u_1u_2 \in E(G) \setminus E(T)$. The *fundamental cycle* of e (with respect to T) is the undirected cycle composed of $E(T[u_1])$, e , and $E(T[u_2])$. Note that a fundamental cycle might not be simple

since $T[u_1]$ and $T[u_2]$ might have a common prefix. In general, a fundamental cycle can be decomposed into a simple prefix path and a simple cycle. We extend the notions of enclosure defined above for simple cycles to fundamental cycles. A face f is enclosed by C if f is enclosed by the simple cycle portion of C . The remaining definitions are as in the simple cycle case.

Proposition 2 ([22]) *For any spanning tree T of G , the set of edges of G not in T form a spanning tree of G^* .*

For a spanning tree T of G , we use T^* to denote the spanning tree of G^* consisting of the edges not in T .

An undirected fundamental cycle C is a *fundamental cycle separator* if each of the interior and exterior of C , consists of at most $3/4$ of the faces of G .¹ Lipton and Tarjan [17] show that, if G is triangulated (every face is adjacent to at most 3 vertices) then for any spanning tree T of G , there is an edge not in T whose fundamental cycle with respect to T is a fundamental cycle separator. Goodrich [7] observed that such an edge can be found by looking for an edge-separator in the dual tree T^* . Thorup [23] used a different construction that separates a graph into three balanced subgraphs. Our discussion of fundamental cycle separators differs from Thorup's, and follows that of [15]. This leads to a simpler structure of the oracle, as we discuss in Section 4.

Let G be a planar graph. For the discussion of separators we ignore the directions of arcs in G and treat it as an undirected graph (the direction of arcs does not affect separation properties of the graph). The following basic and simple lemma was proved in [15].

Lemma 1 [15, Lemma 1] *Let G be a planar graph with face weights such that no face has weight more than $1/4$ the total weight. Let T be a spanning tree of G . Let T^* be the spanning tree of the planar dual G^* of G consisting of the edges not in T . Assume that T^* has maximum degree 3. One can find in $O(|V(G)|)$ time an edge \hat{e} of T^* such that the total face weight enclosed by the fundamental cycle of \hat{e} with respect to T is at least $1/4$ the total weight and at most $3/4$ the total weight.*

Let G be a triangulated planar graph with spanning tree T . Let T^* be the spanning tree of G^* consisting of the edges not in T . Since G is triangulated, the maximum degree of T^* is 3. Let \hat{e} be the edge in Lemma 1, and let C be its fundamental cycle with respect to T . Separating G with C yields two subgraphs, the interior G_{int} , and exterior G_{ext} of C . These subgraphs inherit the embedding from G . Thus, the cycle C is the infinite face of G_{int} , and the simple cycle portion of C is a face of G_{ext} . See Figure 1 for an illustration.

One can obtain from T subtrees that span G_{ext} and G_{int} . The subgraph of T enclosed by C is a spanning tree T_{int} of G_{int} , and the subgraph of T not strictly enclosed by C is a spanning tree T_{ext} of G_{ext} . Similarly, Consider the

¹The definitions in the literature differ in the choice of constant ($3/4$ in our case) as well as in the quantity according to which the balance of the separation is defined (number of faces in our case).

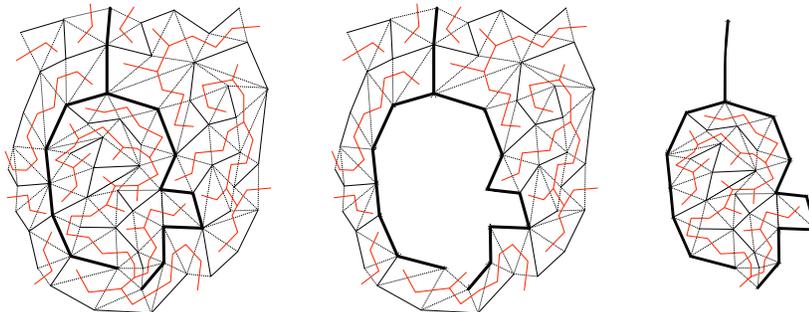


Figure 1: Separating a graph using a fundamental cycle separator. Left: part of a graph G is shown. Edges of T are black. Two branches of T on a fundamental cycle are solid thick. Edges of the dual tree T^* are shown in red. Since only a part of G is shown, this illustration does not show T^* as a tree. Center: part of the subgraph G_{ext} of G is shown. Right: the subgraph G_{int} is shown. Note that the spanning trees T_{ext} and T_{int} are subtrees of T , and that the dual spanning trees T_{ext}^* and T_{int}^* are subtrees of T^* . Observe that, e.g., the infinite face of G_{int} is not triangulated, yet the maximum degree of T_{int}^* remains 3.

two subtrees, T_{ext}^* and T_{int}^* , obtained from T^* by deleting the edge \hat{e} , and then adding \hat{e} back to both subtrees. The subtree T_{ext}^* that contains the dual vertex corresponding to the infinite face f_∞ of G is a spanning tree of G_{ext}^* . The other subtree, T_{int}^* , is a spanning tree of G_{int}^* .

We note that even though G is triangulated, G_{int} and G_{ext} are not. For example, the face of G_{int} whose boundary is C might not be a triangle. However, since the spanning trees T_{ext}^* and T_{int}^* of G^* are subtrees of T^* , their maximum degree is 3. Hence, one can continue to apply Lemma 1 recursively even though G_{int} and G_{ext} are not triangulated. We will describe such a decomposition in Section 4.

4 Thorup's Approximate Distance Oracle [23]

In this section we outline the distance oracle of Thorup [23]. This material is not new, but is necessary for understanding our results. Our description is somewhat different from that of Thorup. It does not go into all the details of Thorup's oracle. Rather, we focus on the aspects that our algorithm does not use as black boxes.

Thorup shows that the problem of constructing a distance oracle for a directed graph can be reduced to constructing a distance oracle for a restricted kind of graphs, defined in the following.

Definition 1 *A set T of arcs in a directed graph H is an α -layered spanning tree if it satisfies the following properties:*

- Regarding the arcs of T as edges, T forms a rooted spanning tree of $V(H)$.
- Each branch of T can be decomposed into no more than 3 directed shortest paths in H , each of length at most α . These paths may be of opposing directions. I.e., they need not all be directed away from the root of T .

A graph H is called α -layered if it has an α -layered spanning tree. Thorup shows that any graph G can be decomposed into α -layered graphs of total linear size for any $\alpha \in \mathbb{R}^+$. The decomposition is such that any shortest path of length at most α in G is represented in at least one of a constant number of α -layered graphs in the decomposition (See [subsection 5.1](#) for more details). Thus, a natural scaling technique can be used to answer distance queries in G by answering distance queries in a few α -layered graphs. The concept of α -layered is important because fundamental cycles of an α -layered spanning tree can be decomposed into a constant number of directed shortest paths. This property is crucial in the design of Thorup’s oracle.

Definition 2 A scale- (α, ϵ') distance oracle for a graph H is a data structure that, when queried for $\delta_H(v, w)$, returns

$$d(v, w) \in \begin{cases} [\delta_H(v, w), \delta_H(v, w) + \epsilon' \alpha] & \text{if } \delta_H(v, w) \leq \alpha \\ [\delta_H(v, w), \infty] & \text{otherwise} \end{cases}$$

Thorup shows how to construct a distance oracle for any graph G using scale- (α, ϵ') distance oracles for minors of G at several scales. This is summarized in the following lemma (See [subsection 5.1](#) for more details).

Lemma 2 ([\[23, Sections 3.1, 3.2, 3.3\]](#)) Let G be a graph. Suppose that, for any $\alpha, \epsilon' \in \mathbb{R}^+$ and any α -layered minor H of G , one can construct, in $O(p(|H|, \epsilon'))$ time, a scale- (α, ϵ') distance oracle with space bound $O(s(|H|, \epsilon'))$ and query time $O(t(\epsilon'))$ (here, p, s and t are arbitrary functions that only depend on $|H|$ and ϵ' , not on α). Then, one can construct, for any $\epsilon \in \mathbb{R}^+$, a $(1 + \epsilon)$ -stretch $\langle O(s(|G|, \frac{\epsilon}{4}) \lg(|G|N))_{space}, O(t(\frac{1}{4}) \lg \lg(|G|N) + t(\frac{\epsilon}{4}))_{time} \rangle$ distance oracle for G in $O(p(|G|, \frac{\epsilon}{4}) \lg(|G|N))$ time.

Planar graphs are closed under taking minors. Thus, by [Lemma 2](#), to show a distance oracle for planar graphs, one only needs to show how to construct scale- (α, ϵ) distance oracles for α -layered planar graphs. We next describe Thorup’s construction of such oracles. In [Section 4.1](#) we explain the recursive structure. In [Section 4.2](#) we describe a non-efficient construction, and in [Section 4.3](#) we describe how to make the construction efficient. We note again that all of these constructions are essentially due to Thorup.

4.1 The recursive decomposition

Let G be a directed α -layered planar graph with an α -layered spanning tree T . Thus, each branch in T can be decomposed into at most 3 directed shortest

paths. We assume that G is triangulated. This is without loss of generality since one can triangulate G with infinite length bidirected arcs. Clearly, this does not affect the shortest paths or the distances in G . For the description of the recursive decomposition we ignore the directions of arcs of G and treat it as an undirected graph. We stress that this is done only to define the decomposition. When describing the oracle we will, of course, take the directions of arcs into consideration.

The set of edges not in T forms a spanning tree T^* of G^* , and, since G is triangulated, the maximum degree of G^* is at most 3. We decompose G recursively using the fundamental cycle separator in [Lemma 1](#) until each subgraph contains a constant number of faces of G . The decomposition can be represented by a binary tree \mathcal{T} in the following manner. We refer to the vertices of \mathcal{T} as *nodes* to distinguish them from the vertices of G . See [Figure 2](#) for an illustration.

- Each node r of \mathcal{T} is associated with a subgraph G_r of G . The subgraph associated with the root of \mathcal{T} is G itself. The spanning tree of G is T and the spanning tree of G^* is T^*
- Each non-leaf node r of \mathcal{T} is associated with the fundamental cycle separator Sep_r found by invoking [Lemma 1](#) on G_r . The weight assignment to the faces of G_r used in the invocation of [Lemma 1](#) assigns weight 1 to each face of G_r that is also a face of the original graph G (we call these faces *original faces*), and weight 0 to all other faces of G_r (these faces are called *holes*).
- Each non-leaf node r has two children r_0, r_1 . The subgraph G_{r_0} associated with the node r_0 is the exterior of Sep_r in G_r . The subgraph G_{r_1} associated with the node r_1 is the interior of Sep_r in G_r .

In both G_{r_0} and G_{r_1} we replace each of the two branches comprising Sep_r by their reduction to their vertices incident to at least one original face.

As explained in the text following [Lemma 1](#), the spanning trees of G_{r_0} and G_{r_1} are subtrees of T_r (with the branches of Sep_r reduced as described above), and the spanning trees of $G_{r_0}^*$ and $G_{r_1}^*$ are subtrees of T_r^* .

Since the number of original faces decreases by a constant factor at each level, the depth of \mathcal{T} is $O(\lg n)$. We will bound the space required to store the decomposition tree \mathcal{T} , and all the subgraphs G_r generated by the decomposition by $O(|G| \lg |G|)$. This is where reducing the branches of Sep_r is crucial, because the sizes of the separators do not decrease along the recursion. We explain this issue in detail. Recall that the subgraphs created during the recursive decomposition have two type of faces; faces that are also original faces of G , which we call *original faces*, and faces that are not faces on G , which are called *holes*. The original faces are all triangles, and their number decreases by a constant factor at each recursive step. The holes are faces that may consist of more than 3 vertices, and their number increases by at most one at each recursive step. The problem with bounding the size of the subgraphs G_r arises

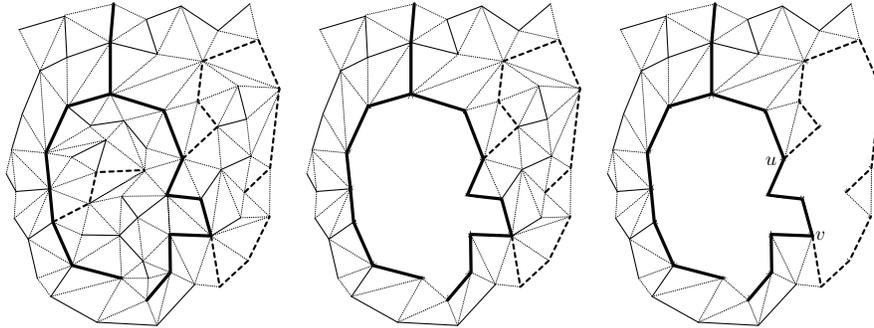


Figure 2: Illustration of the recursive decomposition. Edges of T are black. Edges of the cycle separator at the first level are solid thick, edges of the cycle separators at the second level are dashed. Left: part of the graph G is shown. Center: part of the subgraph G_0 of G . G_0 has a single hole. Right: part of the subgraph G_{00} of G_0 . G_{00} has two holes. The subpath of T between vertices u and v is reduced into a single edge because internal vertices on this subpath are not incident to original faces of G .

from vertices that are adjacent only to holes, because the number of such vertices does not necessarily decrease along the recursion. Note that such vertices are always vertices of the separator of some ancestor of r in \mathcal{T} (possibly r itself). For this reason we replace the branches of Sep_r with their reductions to vertices incident to at least one original face. This has the effect of replacing every (possibly long) maximal path of edges that are only incident to holes with a single reduced edge (that is assigned the length of the path). This change reduces the size of the subgraph without changing the distances in the subgraph. After reducing these paths each vertex in the subgraph is incident to at least one original face. Therefore, the number of vertices in a subgraph is at most 3 times the number of original faces in the subgraph. Namely, for any $r \in \mathcal{T}$, $|G_r|$ is within a constant multiplicative factor from the number of original faces in G_r . Since each original face belongs to exactly one G_r at each level of \mathcal{T} , the total size of all subgraphs in a single level of the recursion is $O(|G|)$. Thus the total size is $O(|G| \lg |G|)$.

4.2 A scale- (α, ϵ) distance oracle

The main idea in obtaining an approximate distance oracle is to store just a subset of the pairwise distances in the graph, from which all approximate distances can be computed efficiently. Fix some $\epsilon > 0$.

Definition 3 (ϵ -covering connections set) *Let H be a graph. Let Q be a shortest path in H of length at most α . Let v be a vertex of H . A set $C(v, Q) \subseteq V(Q)$ is called an ϵ -covering connections set from v to Q with connection lengths*

$\ell : V(H) \times V(Q) \rightarrow \mathbb{R}^+$ if, for every $q^* \in V(Q)$, if $\delta(v, q^*) \leq \alpha$, then there exists some $q \in C(v, Q)$ s.t. $\delta_H(v, q^*) \leq \ell(v, q) + \delta_H(q, q^*) \leq \delta_H(v, q^*) + \epsilon\alpha$.

Intuitively one should think of a connection length $\ell(v, q)$ as the true distance $\delta_H(v, q)$. However, as we explain later on, to achieve efficient construction, $\ell(v, q)$ is sometimes an approximation of $\delta_H(v, q)$. An ϵ -covering connections set $C(Q, v)$ from Q to v with connections lengths $\ell : V(Q) \times V(H) \rightarrow \mathbb{R}^+$ is defined symmetrically. If $\delta(q^*, v) \leq \alpha$, then there exists some $q \in C(Q, v)$ s.t. $\delta(q^*, v) \leq \delta(q^*, q) + \ell(q, v) \leq \delta(q^*, v) + \epsilon\alpha$. We use the term *ϵ -covering connections set* (or just connections set) to refer to both the v -to- Q (or Q -to- v) connections as well as their corresponding connection lengths.

Thorup showed that, given any graph H , shortest path Q in H of length at most α , and for every vertex $v \in V(H)$ there exists an ϵ -covering connections set of size $O(\epsilon^{-1})$. The utility of ϵ -covering connections sets is summarized in the following lemma ([23, Lemma 3.5]):

Lemma 3 *Let Q be a shortest path of length at most α in graph H . Let P be a shortest u -to- w path in H of length at most α which intersects Q . For any $\epsilon > 0$, let $C(u, Q)$ and $C(Q, w)$ be ϵ -covering connections sets from u to Q and from Q to w , respectively. Let H_Q^{uw} be a graph with vertices u, w , the vertices and arcs of the reduction of Q to the connections of u and of w , and with u -to- Q and Q -to- w arcs whose lengths are the corresponding connection lengths of $C(u, Q)$ and $C(Q, w)$. Then*

$$\delta_{H_Q^{uw}}(u, w) \leq \delta_H(u, w) + 2\epsilon\alpha \quad (1)$$

A *Lowest Common Ancestor* (LCA) data structure for a tree T is a data structure that, given any two nodes x, y of T , returns the node furthest from the root that is an ancestor of both x and y . Harel and Tarjan [9] (and many other subsequent simpler and practical results) show how to construct, in linear time, an LCA data structure of linear size and constant query time.

Let u, w be vertices of G . Let r_u, r_w be leaves of \mathcal{T} such that $u \in G_{r_u}$ and $w \in G_{r_w}$. Let r be the LCA of r_u and r_w in \mathcal{T} . We denote by S_r the set of six directed shortest paths in the decomposition of Sep_r (recall that each branch in an α -layered tree can be decomposed into at most 3 directed paths, and that Sep_r is composed of two such branches). If $r_u \neq r_w$, u and w are separated by Sep_r , and so, by Proposition 1, every u -to- w path in G must intersect some path of S_r . Therefore, by Lemma 3, to be able to approximate $\delta_G(u, w)$, it suffices to keep, for every $Q \in S_r$, connections $C(u, Q)$ and $C(Q, w)$, where the connection lengths reflect distances in G (not in G_r). To stress that the connection lengths are in G , we call such connections *global* connections. If $r = r_u = r_w$, then r is a leaf of \mathcal{T} , so the size of G_r is constant. All such distances (in G) between pairs of vertices in G_r are stored explicitly by the oracle.

The distance oracle keeps the following items for every internal node $r \in \mathcal{T}$ and for every vertex $u \in G_r$:

1. global $\epsilon/2$ connections $C(u, Q)$ for all $Q \in S_r$.
2. global $\epsilon/2$ connections $C(Q, u)$ for all $Q \in S_r$.

These connections, over all $u \in G_r$ and all paths in S_r are called the connections of r . In addition, the data structure stores:

3. A mapping of each vertex $v \in V(G)$ to some leaf node $r_v \in \mathcal{T}$ s.t. $v \in G_{r_v}$.
4. A lowest common ancestor data structure over \mathcal{T} .
5. $\epsilon\alpha$ -additive approximation of $\delta_G(u, v)$ for all $u, v \in V(G)$ such that $r_u = r_v$.

We next describe how a query is performed. Given a u -to- w distance query, let r be the lowest common ancestor of r_u and r_w in \mathcal{T} . If $r = r_u = r_w$, the query algorithm returns their length approximation (item 5 above) in $O(1)$ time. Otherwise, the query algorithm computes, for each path $Q \in S_r$ the length of a shortest u -to- w path that intersects Q using the $\epsilon/2$ -covering sets $C(u, Q)$ and $C(Q, w)$. By the construction of \mathcal{T} , the number of such paths Q is constant. If $\delta(u, v) \leq \alpha$, one can compute the distance estimate within $\epsilon\alpha$ additive error (see [Lemma 3](#)) for each Q in $O(\epsilon^{-1})$ time (see also [[23](#), Lemma 3.6]). Thus, if $\delta(u, v) \leq \alpha$, an additive $\epsilon\alpha$ distance approximation is produced in $O(\epsilon^{-1})$ time.

We now bound the total space required for the oracle. Since the height of \mathcal{T} is $O(\lg n)$, and since each original face belongs to exactly one subgraph at each level of \mathcal{T} , each original face belongs to G_r for $O(\lg n)$ nodes r of \mathcal{T} . For each of the $O(1)$ shortest paths in S_r of each such node r , and each original face f (of G) in G_r , each of the 3 vertices of f has a set of $O(\epsilon^{-1})$ connections. This gives a total of $O(\epsilon^{-1}n \lg n)$ connections in items 1,2. Items 3,4 require $O(n)$ space. Since there are $O(n)$ leaves, each with a subgraph of constant size, storing the additional length approximations in item 5 also requires $O(n)$ space. Hence the total space required by the oracle is $O(\epsilon^{-1}n \lg n)$.

The simplifications in our presentation of Thorup's oracle compared to the original description in [[23](#)] stem from the fact that all fundamental cycle separators used throughout our recursive decomposition are obtained via subtrees of a single spanning tree T^* of G^* . Since G is assumed to be triangulated, T^* and all its subtrees have maximum degree 3, so there is no need to retriangulate the subgraphs G_r along the recursive decomposition. This leads to a significant simplification because it implies that all cycle separators used in our constructions consist of original arcs of G . Therefore, by [Proposition 1](#), it suffices to check connections only on the separator of the LCA of r_u and r_v in order to approximate the distance from u to v . In Thorup's construction a cycle separator Sep_r computed for some G_r may consist of artificial edges introduced to triangulate G_r . In this situation, even though Sep_r separates u from v in G_r , there may exist in G u -to- v paths that do not intersect Sep_r . Therefore, in order to approximate the distance from u to v one needs to check connections on all the separators of all the ancestors of the LCA of r_u and r_v . Thorup deals with this by defining the concept of the *frame* of G_r . The use of frames leads to additional complications since one needs to make sure that the number of paths in each frame is constant. Our construction avoids the need for frames and its associated complications.

4.3 Efficient construction

We now mention some, but not all the details of Thorup's $O(\epsilon^{-2}n \lg^3 n)$ -time construction algorithm. The computation of the connections and connection lengths is done top-down the decomposition tree \mathcal{T} . Thorup [23, Lemma 3.11, 3.12] describes a divide-and-conquer procedure that constructs the connections $C(u, Q)$ for all vertices u in a graph H , and a single shortest path Q . A symmetric procedure computes $C(Q, u)$. We summarize the procedure in the following lemma.

Let H be a graph. Let Q be a shortest path in H . Let $sssp(Q, H)$ be a function s.t. for any subgraph H_0 of H , and any vertex $q \in Q_0$, where Q_0 is the reduction of Q to $V(H_0)$, we can compute single source shortest paths from q in the graph $Q_0 \cup H_0$ in $O(sssp(Q, H)|E(H_0)|)$ time. It is easy to see that a standard implementation of Dijkstra's algorithm with priority queue implies $sssp(Q, H) = O(\lg |E(H)|)$. If H is planar, then $sssp(Q, H) = O(1)$ by the algorithm of Henzinger et al. [10].

Lemma 4 [23, Lemmas 3.11, 3.12] *Let Q be a shortest path of length at most α in a directed graph H . Connections $C(u, Q)$ and $C(Q, u)$ for all vertices $u \in V(H)$, where the connection lengths correspond to distances in H , can be computed in $O(\epsilon^{-1}sssp(Q, H)|E(H)| \lg |V(Q)|)$ total time.*

Using Lemma 4 on (planar) G_r for all internal $r \in \mathcal{T}$ and all $Q \in S_r$ takes total $O(\epsilon^{-1}|G| \lg^3 |G|)$ time (by a similar analysis to the one used to bound the space, see also Lemma 6), but does not generate global connections. Because the lemma is applied only to G_r , the connection lengths are with respect to G_r , not to G . We call such connections *local* connections (of r). For the sake of later computations of global $\epsilon/2$ connections, the algorithm first computes local $\epsilon/6$ connections for all internal $r \in \mathcal{T}$ and all $Q \in S_r$.

We next describe how to efficiently compute global connections. We discuss the connections of the form $C(Q, u)$. Computing $C(u, Q)$ is similar. Recall that global connection lengths reflect distances in the entire graph, not just in G_r . Applying Lemma 4 on G for every r would take quadratic time, which is too much. Instead, the computation is done top-down \mathcal{T} , augmenting G_r with the local connection lengths of ancestors of r in \mathcal{T} , which have already been computed, and represent distances outside G_r . This is done as follows.

Lemma 5 *Let $r \in \mathcal{T}$ and $\epsilon > 0$. Global $\epsilon/2$ connections for r can be computed in $O(\epsilon^{-2}|V(G_r)| \lg^2 |V(G)|)$ time using just G_r and local $\epsilon/6$ connections of all ancestors of r .*

Proof 1 *Let Q be a path in S_r . Let X_r^Q be the graph composed of:*

- *The vertices of G_r .*
- *The arcs of Q .*
- *For each ancestor r' of r (including r itself), for each path $Q' \in S_{r'}$, let $V(Q')^\circ$ be the vertices of Q' that have (local) connections to $V(G_r)$ or (local) connections from $V(Q)$:*

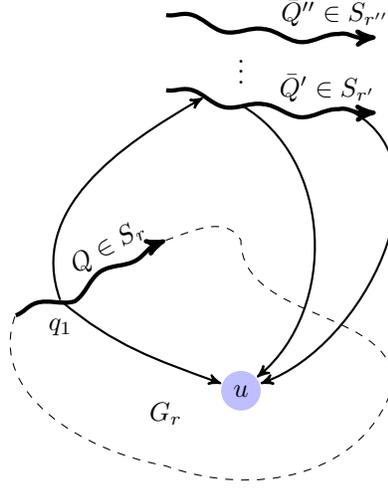


Figure 3: The figure illustrates a part of X_r^Q . The dashed circle represents Sep_r . Only solid arcs are part of X_r^Q . A shortest path from $q_1 \in Q$ to $u \in V(G_r)$ might be enclosed in G_r (local connections $C(Q, u)$) or approximated through a path which intersects a separator of an ancestor of r (in the figure, it is composed of arcs from $C(q_1, \bar{Q}')$, \bar{Q}' and $C(\bar{Q}', u)$).

- The vertices and arcs of \bar{Q}' , the reduction of Q' to vertices of $V(Q')^\circ$.
- Arcs representing the (local) connection lengths of \bar{Q}' for connections from $V(Q')^\circ$ to $V(G_r)$ and from $V(Q)$ to $V(Q')^\circ$.

Since each vertex of G_r has $O(\epsilon^{-1})$ connections to each path in the separator of each of r 's $O(\lg |V(G)|)$ ancestors, the size of X_r^Q is $O(\epsilon^{-1} |V(G_r)| \lg |V(G)|)$.

The algorithm applies [Lemma 4](#) on X_r^Q to generate $\epsilon/6$ -covering connection sets from Q in X_r^Q . We prove below that for any path $Q \in S_r$, and any $q \in V(Q)$ and $u \in V(G_r)$, $\delta_{X_r^Q}(q, u) \leq \delta_G(q, u) + \frac{\epsilon}{3}\alpha$. (The fact that $\delta_G(q, u) \leq \delta_{X_r^Q}(q, u)$ is obvious.) Hence, the $\epsilon/6$ connections computed in X_r^Q are in fact global $(\epsilon/3 + \epsilon/6)$ connections in G , i.e., global $\epsilon/2$ connections.

Let $q \in V(Q)$, and $u \in V(G_r)$. Consider a shortest q -to- u path P . Let r' be the rootmost ancestor of r such that P intersects $S_{r'}$. I.e., P is confined to $G_{r'}$. If $r' = r$ then P is confined to G_r , and because there are arcs representing local $\epsilon/6$ connections lengths for $C(Q, u)$, $\delta_{X_r^Q}(q, u) \leq \delta_G(q, u) + \frac{\epsilon}{6}\alpha$. Otherwise, r' is a strict ancestor of r . Let Q' be a path of $S_{r'}$ intersected by P . Consider the local $\epsilon/6$ connection lengths from q to \bar{Q}' in r' and the local $\epsilon/6$ connection lengths from \bar{Q}' to u . These connection lengths were calculated with respect to exact distances in $G_{r'}$. Since \bar{Q}' , $C(q, \bar{Q}')$ and $C(\bar{Q}', u)$ are all present in $X_{r'}^Q$, it follows from [Lemma 3](#) that $\delta_{X_{r'}^Q}(q, u) \leq \delta_{G_{r'}}(q, u) + 2\frac{\epsilon}{6}\alpha$. Since P is confined to $G_{r'}$, $\delta_{G_{r'}}(q, u) = \delta_G(q, u)$, and so $\delta_{X_r^Q}(q, u) \leq \delta_G(q, u) + \frac{\epsilon}{3}\alpha$, as claimed.

To analyze the running time, first note that X_r^Q can be easily constructed

in $O(|X_r^Q|)$ time by storing, when constructing the oracle, for each vertex of G its connections on each of the separators in each of the subgraphs it belongs to. The lengths of the arcs of X_r^Q can be obtained within the same time bound since they are either stored as connection lengths, or correspond to distances between vertices on a shortest path comprising part of a separator. In the latter case a distance can be retrieved in constant time by storing, when constructing the oracle, for each vertex on a shortest path of a separator, its distance from the beginning of the path.

We next show that $\text{sssp}(Q, X_r^Q) = O(1)$. That is, we show how to compute a shortest path tree in X_r^Q rooted at any vertex of Q in $O(|E(X_r^Q)|)$ time. We first go through the vertices of Q in order. For each vertex $u \in V(Q)$ we relax all the arcs $uv \in E(X_r^Q)$. Let S be the set of all reduced paths of separators of strict ancestors of r . For each path $\tilde{Q}' \in S$ (in any order), we go through the vertices of \tilde{Q}' in order. For each vertex $u \in V(\tilde{Q}')$ we relax all the arcs $uv \in E(X_r^Q)$. The computation is correct by the construction of X_r^Q ; for any $q \in V(Q)$ and any $v \in V(X_r^Q)$, every q -to- v path starts with a subpath of Q , followed by a subpath of at most one path $\tilde{Q}' \in S$, and then reaching v . Hence, the relaxation order is correct.

X_r^Q has $O(\epsilon^{-1}|V(G_r)| \lg |V(G)|)$ arcs and vertices because all ancestral separator paths are in reduced form. Hence, applying [Lemma 4](#) to X_r^Q for each path Q of S_r requires $O(\epsilon^{-2}|V(G_r)| \lg^2 |V(G)|)$ time. Since there are a constant number of shortest paths in S_r , this is also the total runtime.

Lemma 6 *Given an α -layered graph (along with an α -layered spanning tree), constructing a scale- (α, ϵ) distance oracle takes $O(\epsilon^{-2}|V(G)| \lg^3 |V(G)|)$ time.*

Proof 2 *Since a fundamental cycle separator can be found in linear time, constructing the decomposition tree \mathcal{T} takes $O(|V(G)| \log |V(G)|)$ time. Local connections are computed, for each $r \in \mathcal{T}$ and each path Q of S_r , by applying [Lemma 4](#) to G_r using Dijkstra's algorithm.² This takes $O(\epsilon^{-1}|E(G_r)| \lg^2 |V(G_r)|)$ time. Global connections are computed by applying [Lemma 5](#) to X_r^Q , which takes $O(\epsilon^{-2}|V(G_r)| \lg^2 |V(G)|)$ time by [Lemma 5](#).*

Hence, computing all connections for a single $r \in \mathcal{T}$ takes $O(\epsilon^{-2}|V(G_r)| \lg^2 |V(G)|)$. Summing over all $r \in \mathcal{T}$ in the same depth results in $O(\epsilon^{-2}|V(G)| \lg^2 |V(G)|)$ time. Summing over all $O(\lg |V(G)|)$ depths gives a total preprocessing time of $O(\epsilon^{-2}|V(G)| \lg^3 |V(G)|)$.

Finally, computing length approximations for all pair of vertices $u, v \in V(G)$ where $r_u = r_v$ (item 5 in the description of the oracle) can also be done within the same time bound using a degenerate application of [Lemma 5](#). For each leaf node r and for each $u, v \in V(G_r)$, let r' be the subgraph of r consisting of just u and v . Define $\text{Sep}_{r'}$ to be the singleton vertex u , and consider u to be the a trivial local connection from u to v with local connection length $\ell(u, v) = \delta_{G_r}(u, v)$. Considering r' as a child of r and applying [Lemma 5](#) to r' yields a global connection from u to v in $O(\epsilon^{-2} \lg^2 |V(G)|)$ time. That is,

²Since G_r is planar, one may use [\[10\]](#) instead to achieve $O(\epsilon^{-1}|E(G_r)| \lg |V(G)|)$ complexity. However this is not the bottleneck.

the length of this connection is an $\epsilon/2$ approximation of $\delta_G(u, v)$. As there are $O(|V(G)|)$ such pairs (u, v) over all leaves of \mathcal{T} , the total preprocessing time remains $O(\epsilon^{-2}|V(G)| \lg^3 |V(G)|)$.

By Lemma 2 and Lemma 6, we obtain the following theorem:

Theorem 1 *One can construct a $(1 + \epsilon)$ -stretch $(O(\epsilon^{-1}n \lg n \lg(nN))_{space}$, $O(\lg \lg(nN) + \epsilon^{-1})_{time})$ distance oracle in $O(\epsilon^{-2}n \lg^3 n \lg(nN))$ time.*

5 Scale- (α, ϵ) Vertex-Label Distance Oracle

In this section we show how to adapt Thorup's oracle (section 4) to the vertex-label case. Most of our description details how a scale- (α, ϵ) vertex-label distance oracle can be constructed efficiently. At the end of the section we explain why Thorup's argument showing that scale- (α, ϵ) distance oracles can be used to construct a general distance oracle (Lemma 2) applies in the vertex-label case as well.

Thorup's oracle supports one-to-one (vertex-vertex) distance queries, whereas here we need one-to-many distance queries. Given two vertices u, v , Thorup's oracle finds the LCA of r_u and r_v in \mathcal{T} , and uses its connections to produce the desired distance approximation. In a one-to-many query, we are given the query vertex u , but there is no analogue for v . We do not know which λ -labeled vertex in G is closest to u . The minimal distance approximation from u to a λ -labeled vertex in G_{r_u} can be computed in $O(1)$, using the vertex-vertex distance oracle data structure, as $|G_{r_u}| = O(1)$. Therefore, we assume, without loss of generality, that a shortest u -to- λ path intersects some path of $S_{parent(r_u)}$. More precisely, we assume that a shortest u -to- λ path intersects the separator of the leafmost (i.e., furthest from the root) strict ancestor r of r_u in \mathcal{T} such that G_r contains some λ -labeled vertex. The node r takes the role of the LCA of r_u and r_v in the vertex-vertex query algorithm. In order to be able to use r 's connections in a distance query, one must make sure that r 's connections represent approximate distances to λ -labeled vertices in the entire graph, not just in G_r .

We define a set \mathcal{L} of new (artificial) vertices, one per label. For every $r \in \mathcal{T}$, let $\mathcal{L}_r = \{\lambda \in \mathcal{L} | V(G_r) \cap V_\lambda \neq \emptyset\}$ be the restriction of \mathcal{L} to labels present in G_r . Let \hat{G}_r be the graph with vertex set $V(\hat{G}_r) = V(G_r) \cup \mathcal{L}_r$ whose arcs are the arcs of G_r along with a zero-length arc from each λ -labeled vertex of G_r to the corresponding vertex in \mathcal{L}_r . Note that the number of vertices and arcs in \hat{G}_r is within a constant factor of those of G_r .

In addition to the information stored in the vertex-vertex case (Lemma 6), for every $r \in \mathcal{T}$ and $\lambda \in \mathcal{L}_r$, the oracle stores connections w.r.t. λ . For every shortest path $Q \in S_r$ the oracle stores global $\epsilon/2$ connections $C(Q, \lambda)$. Before explaining how to compute these connections we discuss how a distance query is performed.

Obtaining the distance from u to λ is done by finding the leafmost strict ancestor r of r_u with $\lambda \in \mathcal{L}_r$. The algorithm estimates, for each $Q \in S_r$, the

length of a shortest u -to- λ path that intersects Q , using the connections $C(u, Q)$ and $C(Q, \lambda)$ stored for r (Since $\lambda \in \mathcal{L}_r$, r does store Q -to- λ connections).

Finding r can be done by binary search on the path from r_u to the root of \mathcal{T} . The number of steps of the binary search is $O(\lg \lg n)$. Finding whether a node r' has a vertex with label λ can be done, e.g., by storing all unique labels in $G_{r'}$ in a deterministic dictionary [8, 21]. This dictionary works in the word-RAM model. It stores k elements in $O(k)$ space, requires $O(k \log k)$ construction time, and answers whether an element x is in the dictionary in constant time. Over all nodes of \mathcal{T} this takes $O(n \log n)$ space, and $O(n \log n \log |L|) = O(n \log^2 n)$ preprocessing time, which are dominated by the overall space and construction time of the oracle. The constant query time of the deterministic dictionary allows us to perform the binary search for r' in $O(\lg \lg n)$ time. Thus, the query time of our oracle is $O(\lg \lg n + \epsilon^{-1})$.

The Construction Algorithm It remains to show how the connections are computed. We begin with the local connections. For every $r \in \mathcal{T}$, for every $Q \in S_r$, the algorithm computes $\epsilon/6$ connections sets on Q w.r.t. each vertex of \hat{G}_r by invoking Lemma 4 on \hat{G}_r .

As for the quality of approximation, we must show that the connection lengths to the artificial vertices are useful for approximate distance queries. For the local connections, the approximation is immediate because the desired distances are in \hat{G}_r .

We now show how to compute the global connections without invoking Lemma 4 on the entire input graph G at every call. The crucial point is that the connections to the artificial vertex λ from separators of ancestors of r represent distances to vertices with label λ that are not necessarily in G_r .

Lemma 7 *Let $r \in \mathcal{T}$. Global $\epsilon/2$ connections of r to label $\lambda \in \mathcal{L}_r$ can be computed using just the local $\epsilon/6$ connections of ancestors of r . Computing all global connections to \mathcal{L}_r for all $r \in \mathcal{T}$ can be done in $O(\epsilon^{-2} |V(G)| \lg^3 |V(G)|)$ time.*

Proof 3 *Let Q be a shortest path of S_r . Let \vec{X}_r^Q be the graph composed of the following: (see Figure 4 for an illustration)*

- The vertices \mathcal{L}_r .
- The arcs of Q .
- For each ancestor r' of r (including r), for each path $Q' \in S_{r'}$, let $V(Q')^\circ$ be the vertices of Q' that have local connections to \mathcal{L}_r or local connections from $V(Q)$:
 - The vertices and arcs of \bar{Q}' , the reduction of Q' to vertices of $V(Q')^\circ$.
 - Arcs representing the local connection lengths of \bar{Q}' from $V(Q')^\circ$ to \mathcal{L}_r and from $V(Q)$ to $V(Q')^\circ$.

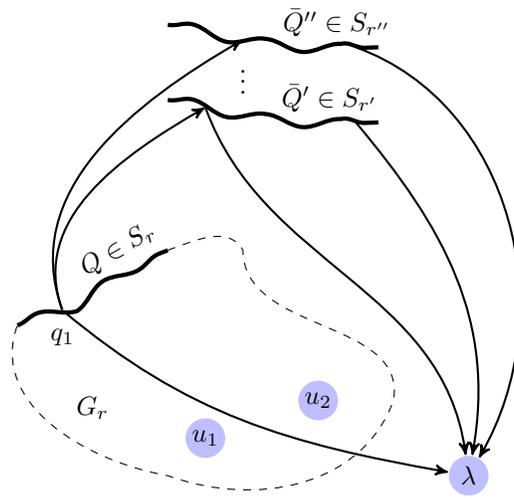


Figure 4: The figure illustrates a part of X_r^Q for the labels case, similarly to Figure 3. The vertices u_1 and u_2 are λ -labeled vertices of G_r , and are not part of X_r^Q . Paths from Q to λ -labeled vertices such as u_1 and u_2 confined to G_r are represented in X_r^Q by arcs between Q and λ . These arcs correspond to the local $\epsilon/6$ connections of λ on Q in \hat{G}_r . All solid arcs are part of X_r^Q . A shortest path from $q_1 \in V(Q)$ to $\lambda \in \mathcal{L}_r$ is approximated by connections from q_1 to a separator of an ancestor of r and from there to λ . Note that $C(Q', \lambda)$ represent distances from Q' to λ -labeled vertices that are not necessarily in G_r .

Let $q \in V(Q)$ and $\lambda \in \mathcal{L}_r$. We show that $\overrightarrow{\delta_{X_r^Q}}$ approximates the distance from q to its closest λ -labeled vertex in G with an additive error of $\frac{\epsilon}{3}\alpha$, namely:

$$\delta_{\overrightarrow{X_r^Q}}(q, \lambda) \leq \delta_G(q, \lambda) + \frac{\epsilon}{3}\alpha \quad (2)$$

Therefore, applying [Lemma 4](#) to $\overrightarrow{X_r^Q}$ with $\epsilon/6$ results in global $\epsilon/2$ connections, just as in [Lemma 5](#).

Let u_λ be a closest λ -labeled vertex to q in G . Consider a shortest q -to- u_λ path P in G . Let r' be the rootmost ancestor of r such that P intersects $S_{r'}$. As in the vertex-vertex case, if $r' = r$ then we are done, as the local connection sets of Q represent approximations of distances from vertices of Q to vertices \mathcal{L}_r in G . Otherwise, $r' \neq r$ and let \tilde{Q} be a shortest path in $S_{r'}$ that is intersected by P . Consider the local $\epsilon/6$ connection lengths from q to \tilde{Q} and from \tilde{Q} to λ in r' . These lengths were calculated with respect to exact distances in $\hat{G}_{r'}$. Since \tilde{Q} , $C(q, \tilde{Q})$ and $C(\tilde{Q}, \lambda)$ are all present in $\overrightarrow{X_r^Q}$, it follows that $\delta_{\overrightarrow{X_r^Q}}(q, \lambda) \stackrel{\text{Lemma 3}}{\leq} \delta_{\hat{G}_{r'}}(q, \lambda) + 2\frac{\epsilon}{6}\alpha$. Since P is confined to $G_{r'}$, $\delta_{\hat{G}_{r'}}(q, \lambda) = \delta_{G_{r'}}(q, u_\lambda) = \delta_G(q, \lambda)$, and the lemma follows.

The construction algorithm differs from the one in [Section 4.2](#) in the existence of the artificial vertices. Since $|\mathcal{L}_r| = O(|V(G_r)|)$ for any $r \in \mathcal{T}$, the total running time and space requirements remain as in [Section 4.2](#).

5.1 Vertex-Label Distance Oracle

We now explain why [Lemma 2](#) applies also to the vertex-label case, using our construction algorithm for scale- (α, ϵ) vertex-label distance oracles. For this we need to elaborate a bit more on the proof of [Lemma 2](#).

The proof of [Lemma 2](#) relies on two reductions [[23](#), [Lemma 3.2](#), [3.9](#)]. The first shows that for any graph G , and for any $\alpha > 0$, one can construct in linear time a family of α -layered graphs $\{G_i^\alpha\}_i$ such that

1. $\sum |G_i^\alpha| = O(|G|)$.
2. Each $v \in V(G)$ has an index $\iota(v)$ s.t. any $w \in V(G)$ has $d = \delta_G(v, w) \leq \alpha$ iff $d = \min\{\delta_{G_{\iota(v)-2}^\alpha}(v, w), \delta_{G_{\iota(v)-1}^\alpha}(v, w), \delta_{G_{\iota(v)}^\alpha}(v, w)\}$
3. Each G_i^α is a minor of G . I.e., it can be obtained from G by contraction and deletion of arcs and vertices. In particular, if G is planar, so is G_i^α .

Item (2.) means that any shortest path of length at most α in G is represented in at least one of three fixed graphs G_i^α . Thus, one can use scale- (α, ϵ) distance oracles for the α -layered graphs $\{G_i^\alpha\}$ to implement a scale- (α, ϵ) oracle of G .

The second reduction [[23](#), [Lemma 3.8](#)] is a scaling argument that shows how to construct a $(1 + \epsilon)$ -stretch distance oracle for G using scale- (α, ϵ') distance oracles for $\alpha = 2^i$ for all integers $i \in [\lceil \lg(nN) \rceil]$. The reduction does not rely on planarity. Roughly, the idea is that an additive $\frac{\epsilon}{2}\alpha$ -approximation to $\delta(u, v)$ is a

multiplicative $(1+\epsilon)$ -approximation if $\frac{\alpha}{2} \leq \delta(u, v)$. Therefore one can use binary search to identify the appropriate scale from which to get a $(1+\epsilon)$ approximate distance. See [23, Lemma 3.9] for the details.

Now consider the vertex-labeled case. Let G' be the graph obtained from G by adding apices representing the labels. All vertices with a specific label are connected in G' to the apex corresponding to this label with zero length arcs. A vertex-vertex distance oracle for G' is a vertex-label distance oracle for G , so it suffices to show the former. Applying Thorup's second reduction to G' , it suffices to show how to construct a scale- (α, ϵ) distance oracle for G' for any α, ϵ . Answering distance queries between vertices of G' that are not apices can be done using a scale- (α, ϵ) vertex-vertex distance oracle for the original planar graph G whose existence was shown by Thorup. Answering distance queries to an apex in G' can be done using a vertex-label scale- (α, ϵ) distance oracle for G . It therefore suffices to show a vertex-label scale- (α, ϵ) distance oracle for G . Let $\alpha \in \mathbb{R}^+$. We use Thorup's first reduction, and construct the vertex-label distance oracle described in the beginning of Section 5 for each minor G_i^α of G . Given $u \in V(G)$ and $\lambda \in \mathcal{L}$ with $\delta_G(u, \lambda) \leq \alpha$, let $w \in V_\lambda$ be a closest λ -labeled vertex to u . By the properties of Thorup's first reduction, there is a graph G_i^α in which the u -to- w distance is $\delta_G(u, w)$. Thus, the vertex-label distance oracle for G_i^α will report a distance of at most $\delta_G(u, \lambda) + \epsilon\alpha$.

Therefore, the main theorem follows:

Theorem 2 *A $(1+\epsilon)$ -stretch $\langle O(\epsilon^{-1}n \lg n \lg(nN))_{space}, O(\lg \lg n \lg \lg(nN) + \epsilon^{-1})_{time} \rangle$ vertex-label distance oracle can be constructed in $O(\epsilon^{-2}n \lg^3 n \lg(nN))$ time for a directed planar graph with n vertices and maximum arc length N .*

6 Reporting Approximate Shortest Path

Thorup describes how to augment his oracle to report a u -to- v path of length $(1+\epsilon)\delta(u, v)$ in time linear in the number of arcs reported. We provide here a brief description (cf. [23, Sections 2.7, 2.8, 3.7]).

The algorithm now stores additional information. The ϵ -cover construction algorithm (Lemma 4) computes shortest path trees rooted at each connection. In the original description these trees are discarded once the connection lengths have been recorded. To report shortest paths, the algorithm stores these trees for all local connections. Let r be a node of \mathcal{T} . For each $v \in G_r$, $Q \in S_r$ and global connection $q \in C(Q, v)$, (i.e., connections computed by invoking the ϵ -cover construction algorithm on X_r^Q), the algorithm records the rootmost node $r' \in \mathcal{T}$ whose separator is intersected by the q -to- v shortest path in X_r^Q .

The query algorithm for the distance between u and v uses some global connection. Let r' be the node of \mathcal{T} recorded for that connection. By choice of r' there exists a $(1+\epsilon)$ -approximate shortest path P between u and v in $G_{r'}$. The algorithm now uses the local connections of r' to find P and uses the shortest path trees stored for those connections to report the edges of P .

Storing all shortest path trees does not change the preprocessing time but increases the required space to $O(\epsilon^{-1}n \lg^2 n)$. This is because each vertex par-

anticipates in $O(\epsilon^{-1} \lg n)$ connections shortest path trees (see [23, Lemma 3.12]) in $O(\lg n)$ nodes of \mathcal{T} . Let \bar{d} be the number of arcs of the reported path. The resulting query time is $O(\epsilon^{-1} + \lg \lg n + \bar{d})$.

We extend this technique to the vertex label case. As in the vertex-vertex case, the query algorithm finds a node r' such that there exists in $G_{r'}$ a path P that $(1 + \epsilon)$ -approximates the shortest path between u and a λ -labeled vertex in G . It then finds P using the local connections of r' . The main difference is that at this point the algorithm knows the distance to the artificial vertex λ , but not the identity of the λ -labeled vertex realizing this distance. However, this information can be stored along with the shortest path trees for local connections. Consider any r . Recall that each local connection corresponds to a shortest path in \hat{G}_r . We record, with every local connection to each $\lambda \in \mathcal{L}_r$, the identity of the vertex preceding λ in the corresponding shortest path.

7 Undirected Vertex-Label Distance Oracle

The extended abstract of the current paper [19] contained a description of a simplified and more efficient version of a vertex-label distance oracle for undirected planar graphs. Unfortunately, that oracle is flawed. Specifically, Lemma 6 in [19] is false. Let Q be a shortest path in an undirected graph G . If one introduces an artificial vertex u (apex) to G and connects u to vertices of G with zero-length edges, then Q might no longer be a shortest path. The proof of Lemma 6 in [19] assumes Q remains a shortest path, and is therefore incorrect. This seems to be a fundamental problem with this approach, which we were not able to correct. This situation is not problematic in the directed case, because all the arcs that are added to the graph enter the artificially added vertices. Therefore, adding these arcs does not change distances between vertices of the graph in the directed case.

We do mention an improvement to the vertex-label distance oracle of [16] for undirected planar graphs. Given a label λ and vertex u , for any path Q , when queried for the shortest u -to- λ path which intersects Q , the algorithm in [16] uses a predecessor search in order to identify a range of $\bigcup_{v \in V_\lambda} \{C(Q, v)\}$ which is relevant for the query. In [16] searching for the predecessor is done using binary search on the vertices of Q , which takes $O(\lg \Delta)$, where Δ is the (hop) diameter of the graph. In general Δ may be as large as n . Instead, one can record, for each vertex of Q , its ordinal number along Q , and store the vertices of $\bigcup_{v \in V_\lambda} \{C(Q, v)\}$ in a fast integer predecessor data structure (e.g. [25]), which supports predecessor search in $O(\lg \lg \Delta)$ instead of $O(\lg \Delta)$. Therefore, the query time for the vertex-label distance oracle of [16] can be made $O(\epsilon^{-1} \lg n \lg \lg n)$ instead of $O(\epsilon^{-1} \lg^2 n)$.

8 Concluding Remarks

This work presents an extension of Thorup’s vertex-vertex oracle to enable vertex-label queries in planar graphs. Although our focus is on planar graphs, the algorithm works for any class of graphs which are both minor-closed and tree-path separable. By tree-path separable we mean that, given any spanning tree, there exists a constant number of root branches whose deletion separates the graph into subgraphs, each of size at most half the original size.

Related problems In this work we dealt with labeled-graphs where each vertex has exactly one label. However, there is no obstacle to deal with labeled-graphs in which each vertex has several labels. Let κ be the bound on the maximal number of labels a vertex is labeled by. Assuming $\kappa = \text{poly}(n)$, the preprocessing time and space of the oracles in this work are multiplied by κ .

We note that the oracle in this paper can be used to obtain vertex-vertex distance oracles in an apex graph G (i.e., G is a planar graph with additional κ apices). Split each apex a in G into $\text{degree}(a)$ copies, one copy for each arc incident to a . Label the copies of a with a distinct label λ_a . Note that the resulting planar graph G' has the same number of edges as G , and at most κ times the number of vertices of G . Construct a vertex-vertex distance oracle and a vertex-label distance oracle for G' . Also store explicitly the distances between any two apices in κ^2 space. Consider a u -to- v query. Let P be a shortest u -to- v path in G . If P does not visit any apex, its length can be reported by the vertex-vertex distance oracle for G' . If P does visit some apex, then let a_1, a_2 be the first and last apices on P , respectively. The path P can be decomposed into (i) a u -to- a_1 prefix, (ii) a a_1 -to- a_2 infix, and (iii) a a_2 -to- v suffix. The lengths of the 2κ possible prefixes (i) and suffixes (iii) can be found by making 2κ queries to the vertex-label distance oracle of G' . The apex-to-apex distances (ii) have been precomputed. Thus the query time is $O(\kappa^2 + \kappa(\lg \lg n \lg \lg (nN) + \epsilon^{-1}))$.

A possible direction for future work is to devise efficient label-to-label distance queries. This seems significantly more difficult. In the vertex-label query, when queried for a u -to- λ distance the algorithm used the leafmost node r in \mathcal{T} that contains u and some λ -labeled vertex to quickly answer the query. In a λ_1 -to- λ_2 query, we do not know which node $r \in \mathcal{T}$ necessarily has its separator intersected by the desired shortest path. This is because λ_1 and λ_2 labeled vertices might be scattered in any of the leaves of \mathcal{T} . Constructing a label-to-label distance oracle whose query time does not depend on the number of vertices with the queried labels remains as an interesting open problem.

References

- [1] Abraham, I., Chechik, S., Krauthgamer, R., Wieder, U.: Approximate nearest neighbor search in metrics of planar graphs. In: Approximation, Randomization, and Combinatorial Optimization. Algorithms and Tech-

- niques, APPROX/RANDOM, pp. 20–42 (2015). DOI 10.4230/LIPIcs.APPROX-RANDOM.2015.20
- [2] Baswana, S., Gaur, A., Sen, S., Upadhyay, J.: Distance oracles for unweighted graphs: Breaking the quadratic barrier with constant additive error. In: Proceedings of the 35th International Colloquium on Automata, Languages and Programming (ICALP), pp. 609–621 (2008). DOI 10.1007/978-3-540-70575-8_50
 - [3] Baswana, S., Kavitha, T.: Faster algorithms for approximate distance oracles and all-pairs small stretch paths. In: Proceedings of the 47th Annual Symposium on Foundations of Computer Science (FOCS), pp. 591–602 (2006). DOI 10.1109/FOCS.2006.29
 - [4] Baswana, S., Sen, S.: Approximate distance oracles for unweighted graphs in expected $O(n^2)$ time. ACM Transactions on Algorithms **2**(4), 557–577 (2006). DOI 10.1145/1198513.1198518
 - [5] Chechik, S.: Improved distance oracles and spanners for vertex-labeled graphs. In: Proceedings of the 20th European Symposium on Algorithms (ESA), pp. 325–336 (2012). DOI 10.1007/978-3-642-33090-2_29
 - [6] Chechik, S.: Approximate distance oracles with improved bounds. In: Proceedings of the 47th Annual ACM Symposium on Theory of Computing (STOC), pp. 1–10 (2015)
 - [7] Goodrich, M.T.: Planar separators and parallel polygon triangulation. Journal of Computer and System Sciences **51**(3), 374–389 (1995)
 - [8] Hagerup, T., Miltersen, P.B., Pagh, R.: Deterministic dictionaries. Journal of Algorithms **41**(1), 69–85 (2001). DOI 10.1006/jagm.2001.1171
 - [9] Harel, D., Tarjan, R.E.: Fast algorithms for finding nearest common ancestors. SIAM Journal on Computing **13**(2), 338–355 (1984)
 - [10] Henzinger, M.R., Klein, P.N., Rao, S., Subramanian, S.: Faster shortest-path algorithms for planar graphs. Journal of Computer and System Sciences **55**(1), 3–23 (1997)
 - [11] Hermelin, D., Levy, A., Weimann, O., Yuster, R.: Distance oracles for vertex-labeled graphs. In: Proceedings of the 38th International Colloquium on Automata, Languages and Programming (ICALP), pp. 490–501 (2011)
 - [12] Kawarabayashi, K., Klein, P.N., Sommer, C.: Linear-space approximate distance oracles for planar, bounded-genus and minor-free graphs. In: Proceedings of the 38th International Colloquium on Automata, Languages and Programming (ICALP), pp. 135–146 (2011)

- [13] Kawarabayashi, K., Sommer, C., Thorup, M.: More compact oracles for approximate distances in undirected planar graphs. In: Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 550–563 (2013)
- [14] Klein, P.N.: Preprocessing an undirected planar network to enable fast approximate distance queries. In: Proceedings of the 13th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 820–827 (2002)
- [15] Klein, P.N., Mozes, S., Sommer, C.: Structured recursive separator decompositions for planar graphs in linear time. In: Proceedings of the 45th Annual ACM Symposium on Theory of Computing (STOC), pp. 505–514 (2013). DOI 10.1145/2488608.2488672
- [16] Li, M., Ma, C.C.C., Ning, L.: $(1 + \epsilon)$ -distance oracles for vertex-labeled planar graphs. In: Proceedings of the 10th International Conference on Theory and Applications of Models of Computation, 10th International Conference (TAMC), pp. 42–51 (2013)
- [17] Lipton, R., Tarjan, R.: A separator theorem for planar graphs. *SIAM Journal of Applied Mathematics* 36 pp. 177–189 (1979)
- [18] Łącki, J., Oćwieja, J., Pilipczuk, M., Sankowski, P., Zych, A.: The power of dynamic distance oracles: Efficient dynamic algorithms for the Steiner tree. In: Proceedings of the 47th Annual ACM Symposium on Theory of Computing (STOC), pp. 11–20 (2015)
- [19] Mozes, S., Skop, E.E.: Efficient vertex-label distance oracles for planar graphs. In: Proceedings of the 13th International Workshop on Approximation and Online Algorithms (WAOA), pp. 97–109 (2015). DOI 10.1007/978-3-319-28684-6_9
- [20] Patrascu, M., Roditty, L.: Distance oracles beyond the Thorup-Zwick bound. *SIAM Journal on Computing* **43**(1), 300–311 (2014). DOI 10.1137/11084128X
- [21] Ružić, M.: Constructing efficient dictionaries in close to sorting time. In: Proceedings of the 35th International Colloquium on Automata, Languages and Programming (ICALP), pp. 84–95 (2008). DOI 10.1007/978-3-540-70575-8_8
- [22] von Staudt, K.G.C.: *Geometrie der Lage*. Bauer und Raspe, Nürnberg (1847)
- [23] Thorup, M.: Compact oracles for reachability and approximate distances in planar digraphs. *Journal of the ACM* **51**(6), 993–1024 (2004)
- [24] Thorup, M., Zwick, U.: Approximate distance oracles. *Journal of the ACM* **52**(1), 1–24 (2005)

- [25] Willard, D.E.: Log-logarithmic worst-case range queries are possible in space $\theta(n)$. *Information Processing Letters* **17**(2), 81–84 (1983). DOI 10.1016/0020-0190(83)90075-3
- [26] Wulff-Nilsen, C.: Approximate distance oracles with improved preprocessing time. In: *Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 202–208 (2012)