

Near-Optimal Distance Oracles for Vertex-Labeled Planar Graphs

Jacob Evald

Department of Computer Science, University of Copenhagen, Denmark

Viktor Fredslund-Hansen  

Department of Computer Science, University of Copenhagen, Denmark

Christian Wulff-Nilsen  

Department of Computer Science, University of Copenhagen, Denmark

Abstract

Given an undirected n -vertex planar graph $G = (V, E, \omega)$ with non-negative edge weight function $\omega : E \rightarrow \mathbb{R}$ and given an assigned label to each vertex, a vertex-labeled distance oracle is a data structure which for any query consisting of a vertex u and a label λ reports the shortest path distance from u to the nearest vertex with label λ . We show that if there is a distance oracle for undirected n -vertex planar graphs with non-negative edge weights using $s(n)$ space and with query time $q(n)$, then there is a vertex-labeled distance oracle with $\tilde{O}(s(n))^1$ space and $\tilde{O}(q(n))$ query time. Using the state-of-the-art distance oracle of Long and Pettie [12], our construction produces a vertex-labeled distance oracle using $n^{1+o(1)}$ space and query time $\tilde{O}(1)$ at one extreme, $\tilde{O}(n)$ space and $n^{o(1)}$ query time at the other extreme, as well as such oracles for the full tradeoff between space and query time obtained in their paper. This is the first non-trivial exact vertex-labeled distance oracle for planar graphs and, to our knowledge, for any interesting graph class other than trees.

2012 ACM Subject Classification Theory of computation \rightarrow Data structures design and analysis; Theory of computation \rightarrow Shortest paths; Mathematics of computing \rightarrow Graph algorithms

Keywords and phrases distance oracle, vertex labels, color distance oracle, planar graph

Digital Object Identifier 10.4230/LIPIcs.ISAAC.2021.23

Related Version *arXiv Version*: <https://arxiv.org/abs/2110.00074>

Funding *Jacob Evald*: This research is supported by the Starting Grant 7027-00050B from the Independent Research Fund Denmark under the Sapere Aude research career programme.

Viktor Fredslund-Hansen: This research is supported by the Starting Grant 7027-00050B from the Independent Research Fund Denmark under the Sapere Aude research career programme.

Christian Wulff-Nilsen: This research is supported by the Starting Grant 7027-00050B from the Independent Research Fund Denmark under the Sapere Aude research career programme and by Basic Algorithms Research Copenhagen (BARC).

Acknowledgements We are grateful for the thorough and useful feedback provided by the reviewers.

1 Introduction

Efficiently answering shortest path distance queries between pairs of vertices in a graph is a fundamental algorithmic problem with a wide range of applications. An algorithm like Dijkstra's can answer such a query in near-linear time in the size of the graph. If we allow for precomputations, we can break this bound, for instance by simply storing the answers to all possible queries in a look-up table. However, a fast query time should preferably not come at the cost of a large space requirement. A *distance oracle* is a compact data structure that can answer a shortest path distance query in constant or close to constant time.

¹ We use \tilde{O} -notation to suppress $\text{poly}(\log n)$ -factors.



A lot of research has focused on approximate distance oracles which allow for some approximation in the distances output. This is reasonable since there are graphs for which the trivial look-up table approach is the best possible for exact distances. However, for restricted classes of graphs, it may be possible to obtain exact oracles with a much better tradeoff between space and query time. Indeed, for any planar n -vertex digraph, there is an exact distance oracle with space close to linear in n and query time close to constant [7, 2, 12].

A related problem is that of obtaining a vertex-labeled distance oracle. Here, we are given a graph with each vertex assigned a label. A query consists of a pair (u, λ) of a vertex u and a label λ and the output should be the distance from u to the nearest vertex with label λ . Each vertex is given only one label but the same label may be assigned to multiple vertices. To give some practical motivation, if the graph represents a road network, a label λ could represent supermarkets and the output of query (u, λ) gives the distance to the nearest supermarket from the location represented by u .

Note that this is a generalization of the distance oracle problem since vertex-to-vertex distance queries can be answered by a vertex-labeled distance oracle if each vertex is given its own unique label. If L is the set of labels, a trivial vertex-labeled distance oracle with constant query time is a look-up table that simply stores the answers to all possible queries, requiring space $O(n|L|)$. This bound can be as high as quadratic in n .

Our main result, which we shall state formally later in this section, is that for undirected edge-weighted planar graphs, the vertex-labeled distance oracle problem can be reduced to the more restricted distance oracle problem in the sense that up to $\log n$ -factors, any space/query time tradeoff for distance oracles also holds for vertex-labeled distance oracles. Hence, the tradeoff from [12] translates to vertex-labeled distance oracles, assuming that the planar graph is undirected. To the best of our knowledge, this is the first non-trivial upper bound for vertex-labeled distance oracles in any interesting graph class other than trees [8, 15]. A strength of our result is that any future progress on distance oracles in undirected planar graphs immediately translates to vertex-labeled distance oracles.

1.1 Related work on vertex-labeled distance oracles

Vertex-labeled distance oracles have received considerably more attention in the approximate setting. With $(1 + \epsilon)$ multiplicative approximation, it is known how to get $\tilde{O}(n)$ space and $\tilde{O}(1)$ query time both for undirected [11] and directed planar graphs [13] and it has been shown how oracles with such guarantees can be maintained dynamically under label changes to vertices using $\tilde{O}(1)$ time per vertex relabel.

For general graphs, vertex-labeled distance oracles with constant approximation have been presented [9, 3, 14] with state of the art being an oracle with $O(kn|L|^{1/k})$ space, $4k - 5$ multiplicative approximation, and $O(\log k)$ query time, for any $k \in \mathbb{N}$.

1.2 Our contributions

We now state our reduction and its corollary:

► **Theorem 1.** *If there is an exact distance oracle for n -vertex undirected edge-weighted planar graphs with $s(n)$ space, $q(n)$ query time, and $t(n)$ preprocessing time, then there exists an exact vertex-labeled distance oracle for such graphs with $s(n) + O(n \log^2 n)$ space, and with $O(q(n) \log n + \log^3 n)$ query time, and $t(n) + \text{poly}(n)$ preprocessing time.*

Plugging in the distance oracle of Long and Pettie et al. [12] gives the following corollary which can be seen as a generalization of their result:

► **Corollary 2.** *For n -vertex undirected edge-weighted planar graphs, there exist exact vertex-labeled distance oracles with the following tradeoffs between space and query time:*

1. $n^{1+o(1)}$ space and $\tilde{O}(1)$ query time,
2. $\tilde{O}(n)$ space and $n^{o(1)}$ query time.

All oracles have preprocessing time polynomial in n .

Up to logarithmic factors, the full tradeoff between space and query time in their paper similarly extends to vertex-labeled distance oracles in undirected edge-weighted planar graphs.

The rest of the paper is organized as follows. In Section 2, we introduce basic definitions and notation and present tools from the literature that we will need for our oracle. In Section 3 we state the key lemmas but defer their proofs until later sections, and thus immediately present our reduction by describing how to obtain a vertex-labeled distance oracle given a distance oracle as a black box. In Section 4, we present a point location structure similar to [7] but with some important modifications to improve space in our setting.

2 Preliminaries

Let $G = (V, E, \omega)$ be a graph with edge weight function $\omega : E \rightarrow \mathbb{R} \cup \{\infty\}$. We denote by $V(G) = V$ and $E(G) = E$ the vertex and edge-set of G , respectively, and by $n = |V(G)|$ the number of vertices of G . A graph G' is said to be a subgraph of G if $V(G') \subseteq V(G)$ and $E(G') \subseteq E(G)$. We denote by $u \rightsquigarrow_G v$ a *shortest path* from u to v in G , by $\mathbf{d}_G(u, v)$ the weight of $u \rightsquigarrow_G v$, and write $u \rightsquigarrow v = u \rightsquigarrow_G v$ and $\mathbf{d}(u, v) = \mathbf{d}_G(u, v)$ when G is clear from context. For a shortest path $p = u \rightsquigarrow v = (u = p_1), p_2, \dots, (p_k = v)$ we define vertex p_i to occur *before* p_j on p if $i < j$ and similarly for edges $p_i p_{i+1}$ and $p_j p_{j+1}$. Thus statements such as “the first/last vertex/edge on p satisfying some property P ” will always be made w.r.t. this ordering. We also write $p \circ p'$ to denote the concatenation of paths (or edges) p and p' , assuming the last vertex of p equals the first vertex of p' . Given $u, v, v' \in V$; we say that v is *closer* than v' to u in G if $\mathbf{d}_G(u, v) < \mathbf{d}_G(u, v')$ or $\mathbf{d}_G(u, v) = \mathbf{d}_G(u, v')$ and $v < v'$, assuming some lexicographic ordering on vertices. We denote by $V(p)$, respectively $E(p)$, the set of vertices, respectively edges, on a path p .

Assume in the following that G is undirected. G is said to be connected, respectively biconnected, if any pair of vertices are connected by at least one, respectively two, vertex-disjoint paths. For a rooted spanning tree T in G and for any edge $e = uv$ not in T , we define *the fundamental cycle* of uv w.r.t. T as the cycle obtained as the concatenation of uv and the two paths of T from the root to u and v , respectively.

2.1 Planar graphs and embeddings

An *embedding* of a planar graph G assigns to each vertex a point in the plane and to each edge a simple arc such that its endpoints coincide with those of the points assigned to its vertices. A *planar embedding* of G is an embedding such that no two vertices are assigned the same point and such that no pair of arcs coincide in points other than those corresponding to vertices they share. A graph is said to be *planar* if it admits a planar embedding. When we talk about a planar graph we assume that it is *planar embedded* and hence some implicit, underlying planar embedding of the graph. When it is clear from the context we shall refer interchangeably to a planar graph and its embedding, its edges and arcs and its vertices and points. Thus the term graph can refer to its embedding, an edge to its corresponding arc and a vertex to its corresponding point in the embedding.

Assumptions about the input

Unless stated otherwise, we shall always assume that G refers to a graph which is weighted, undirected and planar with some underlying embedding. Furthermore, we shall make the structural assumption that G is triangulated. Triangulation can be achieved by standard techniques, i.e. adding to each face f an artificial vertex and artificial edges from the artificial vertex to each vertex of $V(f)$ with infinite weight. This transformation preserves planarity, shortest paths and ensures that the input graph consists only of simple faces. We also assume that shortest paths in the input graph are unique; this can be ensured for any input graph by either randomly perturbing edge weights or with e.g. the deterministic approach in [6] which gives only an $O(1)$ -factor overhead in running time. Finally, it will be useful to state the following lemma when talking about separators in a graph with unique shortest paths:

► **Lemma 3.** *Let $u, v, x, y \in V(G)$. Then $u \rightsquigarrow v$ and $x \rightsquigarrow y$ share at most one edge-maximal subpath.*

Proof. Assume that $x \rightsquigarrow y$ intersects $u \rightsquigarrow v$ and let a resp. b be the first resp. last intersection along $u \rightsquigarrow v$. Since G is undirected, uniqueness of shortest paths implies that $a \rightsquigarrow b$ is a subpath of $u \rightsquigarrow v$ shared by $x \rightsquigarrow y$. ◀

Edge orderings, path turns and path intersections

For an edge $e = uv$ of a planar embedded graph H , we let \langle_e^H be the clockwise ordering of edges of H incident to v starting at e (ignoring edge orientations). Hence \langle_e^H is a strict total order of these edges and e is the first edge in this order.

For vertices $u, v \in V(H)$, $x \in V(u \rightsquigarrow v) \setminus \{u, v\}$ and $y \in V(H) \setminus V(u \rightsquigarrow v)$, let pq be the last edge shared by $u \rightsquigarrow v$ and $x \rightsquigarrow y$. Furthermore let qz resp. qz' be the edge following pq in the traversal of $u \rightsquigarrow v$ and $x \rightsquigarrow y$, respectively. We say that $x \rightsquigarrow y$ *emanates from the left* of $u \rightsquigarrow v$ if $qz' \prec_{pq}^H qz$, and otherwise it *emanates from the right*. We dually say that $y \rightsquigarrow x$ *intersects $u \rightsquigarrow v$ from the left (right)* if $x \rightsquigarrow y$ *emanates from the left (right)*.

Given a face f of H , vertices $u \in V(f)$, and $v, v' \in V$, let H'_f be a copy of H with an artificial vertex f^* embedded in the interior of f along with an additional edge f^*u . Define the paths $p_v = f^*u \circ u \rightsquigarrow_{H'_f} v$ and $p_{v'} = f^*u \circ u \rightsquigarrow_{H'_f} v'$ and assume that neither path is a prefix of the other. By assumption and Lemma 3, p_v and $p_{v'}$ share exactly one edge-maximal subpath $f^* \rightsquigarrow x$. We say that $u \rightsquigarrow_H v$ makes a *left turn* w.r.t $u \rightsquigarrow_H v'$ from f if $x \rightsquigarrow v$ emanates from the left of p_v , and otherwise it makes a *right turn*; we will omit mention of f when the context is clear. Note that the notion of a turn is symmetric in the sense that $u \rightsquigarrow_H v$ makes a left turn w.r.t $u \rightsquigarrow_H v'$ iff $u \rightsquigarrow_H v'$ makes a right turn w.r.t $u \rightsquigarrow_H v$.

2.2 Voronoi Diagrams

The definitions in this subsection will largely be made in a manner identical to those of [7], but are included as they are essential to a point location structure which will be presented in Section 4. Given a planar graph $G = (V, E, \omega)$, $S \subseteq V$, the *Voronoi diagram of S in G* , denoted by $VD(S)$ in G is a partition of V into disjoint sets, $Vor(u)$, referred to as *Voronoi cells*, with one such set for each $u \in S$. The set $Vor(u)$ is defined to be $\{v \in V \mid d(u, v) < d(u', v) \text{ for all } u' \in S \setminus \{u\}\}$, that is the set of vertices that are closer to u than any other site in terms of $d(\cdot, \cdot)$. We shall simply write VD when the context is clear.

It will also be useful to work with a dual representation of Voronoi diagrams. Let VD_0^* be the subgraph of G^* s.t. $E(VD_0^*)$ is the subset of edges of G^* where $uv^* \in VD_0^*$ iff u and v belong to different Voronoi cells in VD . Let VD_1^* be the graph obtained by repeatedly

contracting edges of VD_0^* incident to degree 2 vertices until no such vertex remains². We refer to the vertices of VD_1^* as *Voronoi vertices*, and each face of the resulting graph VD_1^* can be thought of as corresponding to some Voronoi cell in the sense that its edges enclose exactly the vertices of some Voronoi cell in the embedding of the primal. We shall restrict ourself to the case in which all vertices of S lie on a single face h . In particular, h^* is a Voronoi vertex, since each site is a vertex on the boundary of h in the primal. Finally, let VD^* be the graph obtained by replacing h^* with multiple copies, one for each edge. We note that since there are $|S|$ Voronoi sites (and thus faces in VD^*), the number of Voronoi vertices in VD^* is $O(|S|)$ due to Euler's formula. Furthermore, [7] show that when assuming unique shortest paths and a triangulated input graph, VD^* is a ternary tree. It follows that the primal face corresponding to a Voronoi vertex f^* consists of exactly three vertices, each belonging to different Voronoi cells. We refer to the number of sites in a Voronoi diagram as its *complexity*.

Finally, they also note that a *centroid decomposition*, T^* , can be computed from VD^* s.t. each node of T^* corresponds to a Voronoi vertex f^* and the children of f^* in T^* correspond to the subtrees resulting from splitting the tree at f^* , and s.t. the number of vertices of each child is at most a constant fraction of that of the parent. We remark that $VD^*(S)$ can be computed by connecting all sites to a super-source and running a single-source shortest paths algorithm, and its centroid decomposition in time proportional to $|V(VD^*(S))|$.

2.3 Separators and decompositions

In the following, we will outline the graph decomposition framework used by our construction. As part of the preprocessing step, we will recursively partition the input graph using balanced fundamental cycle separators until the resulting graphs are of constant size. We shall associate with the recursive decomposition of G a binary decomposition tree, \mathcal{T} , which is a rooted tree whose nodes correspond to the regions of the recursive decomposition of G . We will refer to nodes and their corresponding regions interchangeably. The root node of \mathcal{T} corresponds to all of G . The following lemma states the invariants of the decomposition that will be used in our construction:

► **Lemma 4.** *Let $G = (V, E, \omega)$ be an undirected, planar embedded, edge-weighted, triangulated graph and let T be a spanning tree³ of G . Then there is an $\tilde{O}(n)$ time algorithm that returns a binary decomposition tree \mathcal{T} of G s.t.*

1. *for any non-leaf node $G' \in \mathcal{T}$, its children G'_l , respectively G'_r , corresponds to the non-strict interior, respectively non-strict exterior of some fundamental cycle in G' w.r.t. T ,*
2. *for any child node, it contains at most a constant fraction of the faces of its parent,*
3. *for any leaf node it contains a constant number of faces of G ,*
4. *for all nodes at depth i , \mathcal{T}_i , $\sum_{G' \in \mathcal{T}_i} |V(G')| = O(n)$*

Properties 1-3 follow from recursively applying a classic linear time algorithm for finding fundamental cycles. Property 4 follows from employing standard techniques that involve contracting degree-two vertices of the separators found at each level of recursion and weighting the resulting edges accordingly. This transformation results in a decomposition where the sum of faces of all regions at any level is preserved. We stress that our construction does not rely on the usual sparse simple cycle separators (of size $O(\sqrt{n})$) but rather fundamental cycle separators of size $O(n)$.

² Formally, given a degree 2 vertex v with incident edges vw, vv' , we replace these edges by wv' , concatenate their arcs and embed wv' using this arc in the embedding.

³ For our purposes, the spanning tree will be a shortest path tree.

3 The vertex labeled distance oracle

In this section we describe our reduction which shows our main result. The reduction can be described assuming Lemma 4 and the existence of the point location structure which we will state in the following lemma, the proof of which is deferred to Section 4:

► **Lemma 5.** *Let $G = (V, E, \omega)$ be an undirected, planar embedded, edge-weighted graph with labeling $l : V \rightarrow L$ and let p be a shortest path in G . There is a data structure $O_{G,p}$ with $O(|V| \log |V|)$ space which given $u \in V$ and $\lambda \in L$ returns a subset $C \subset V$ of constant size, s.t. if v is the vertex with label λ closest to u and $v \rightsquigarrow u$ intersects p , then $v \in C$. Each such query takes time at most $O(\log^2 |V|)$.*

3.1 Preprocessing

Given the input graph $G = (V, E, \omega)$, the preprocessing phase initially computes the decomposition tree, \mathcal{T} , of Lemma 4. Associated with each non-leaf node $G' \in \mathcal{T}$ is a fundamental cycle separator of $ab \in E(G')$ w.r.t. the shortest path tree T rooted at some $c \in V(G')$. For such a G' we shall refer to $S_1(G') = c \rightsquigarrow_{G'} a$ and $S_2(G') = c \rightsquigarrow_{G'} b$. Thus the fundamental cycle separator is given by $S_1(G') \circ ab \circ S_2(G')$. The preprocessing phase proceeds as follows: For all non-leaf nodes $G' \in \mathcal{T}$, compute and store data structures $O_{G',S_1(G')}$ and $O_{G',S_2(G')}$ of Lemma 5. Finally, a distance oracle D with $O(s(|V|))$ space capable of reporting vertex-to-vertex shortest path distances in time $O(t(|V|))$ is computed for G and stored alongside the decomposition tree and the point location structures.

Space complexity

The decomposition tree \mathcal{T} can be represented with $O(|V| \log |V|)$ space and D with $O(s(n))$ space. For each node $G' \in \mathcal{T}$, we store data structures $S_1(G')$ and $S_2(G')$, so by Lemma 4 and 5, we get

$$\sum_{i=0}^{\infty} \sum_{G' \in \mathcal{T}_i} |V(G')| \log |V(G')| = \sum_{i=0}^{c \log n} O(|V| \log |V|) = O(|V| \log^2 |V|)$$

for a total space complexity of $O(s(|V|) + |V| \log^2 |V|)$.

3.2 Query

Let $G' \in \mathcal{T}$ and consider the query $\mathbf{d}_{G'}(u, \lambda)$. If G' is a leaf node, the query is resolved in time $O(t(n))$ by querying D once for each vertex of G' . If G' is a non-leaf node, the query is handled as follows: First, data structures $O_{G',S_1(G')}$ and $O_{G',S_2(G')}$ are queried with u and λ , resulting in two “candidate sets”, C_1 and C_2 , one for each query. By Lemma 5, $C_1 \cup C_2$ contains the nearest vertex with label λ for which $u \rightsquigarrow_{G'} v'$ intersects either S_1 or S_2 if such a vertex exists. Compute $d_{G'} = \min \{\mathbf{d}_G(u, c) \mid c \in C_1 \cup C_2\} \cup \{\infty\}$ by querying D once for each vertex of $C_1 \cup C_2$. The query then recursively resolves $d_{G''} = \mathbf{d}_{G''}(u, \lambda)$ where G'' is a child of G' in \mathcal{T} containing u . Finally, the query returns $\min \{d_{G'}, d_{G''}\}$.

Correctness

Denote by v the vertex of G with label λ nearest to u in G' , and consider the case in which $u \rightsquigarrow_{G'} v$ intersects $S_1(G')$ or $S_2(G')$. In this case, $v \in C$ by Lemma 5 and $C \neq \emptyset$, so

$$d_{G'} = \min \{\mathbf{d}_G(u, c) \mid c \in C\} = \mathbf{d}_G(u, v) = \mathbf{d}_{G'}(u, v) = \mathbf{d}_{G'}(u, \lambda) \leq d_{G''}$$

■ **Algorithm 1** Query procedure for the distance oracle.

```

1: procedure QUERY( $u, \lambda, G'$ )
2:   if  $G'$  is a leaf node in  $\mathcal{T}$  then
3:     return  $\min \{d_{G'}(u, v) \mid v \in V(G') \text{ and } l(v) = \lambda\}$ 
4:   else
5:      $C_1 \leftarrow O_{G', S_1(G')}(u, \lambda)$ ;  $C_2 \leftarrow O_{G', S_2(G')}(u, \lambda)$ 
6:      $G'' \leftarrow$  A child of  $G'$  in  $\mathcal{T}$  containing  $u$ 
7:      $d_{G'} \leftarrow \min \{d_G(u, c) \mid c \in C_1 \cup C_2\} \cup \{\infty\}$ 
8:      $d_{G''} \leftarrow$  QUERY( $u, \lambda, G''$ )
9:     return  $\min \{d_{G'}, d_{G''}\}$ 

```

with the inequality following from definition of v . Note that in case u is a vertex of either S_1 or S_2 , the correct estimate is returned at the current level, but for a simpler description, the recursion proceeds anyways. Otherwise $u \rightsquigarrow_G v$ intersects neither S_1 and S_2 in which case, the path must be fully contained in the (unique) child node, G'' , of G' containing u . In this case, the query reports $d_{G''} = \mathbf{d}_{G''}(u, \lambda) = \mathbf{d}_{G''}(u, v) \leq d_{G'}$, showing the correctness.

Time complexity

At each level of the recursion, $O_{G', S_1(G')}$ and $O_{G', S_2(G')}$ are queried in time $O(\log^2 |V|)$. Furthermore, D is queried $|C| = O(1)$ times in total time $O(1) \cdot O(t(|V|)) = O(t(|V|))$. By Lemma 4, the query is recursively resolved on a problem instance which is a constant fraction smaller at each level of recursion, giving rise to the recurrence relation $T(n) = T(n/a) + O(t(n) + \log^2 n)$. When G' is a leaf node, then by Lemma 4, G' consists of a constant number of faces, described by the base case $T(n) = O(t(n))$ when $n \leq b$ for sufficiently small b . It is easily verified that a solution to the recurrence is bounded by $O(\log^3 n + t(n) \log n)$. This shows the main theorem, and the rest of this paper is devoted to proving Lemma 5.

4 The point location data structure

Our point-location data-structure uses techniques similar to those of [7] for point location in additively weighted Voronoi diagrams, but with some crucial differences in order to save space.

Both structures rely on being able to determine left/right turns of shortest paths in shortest path trees rooted at sites in G , but to facilitate this, the data structure of [7] explicitly stores an (augmented) shortest path tree rooted at each site as well as a data structure for answering least common ancestor (LCA) queries. The point location structure thus requires $\Theta(|S|n)$ space where S is the number of sites, and since S may be large as $\Theta(\sqrt{n})$ (corresponding to the size of a sparse balanced separator in a planar graph), this translates to $\Theta(n^{3/2})$ space for their problem. This will not work in our case since the number of sites can in fact be as high as $\Theta(n)$, leading to a quadratic space bound.

The second issue with applying the techniques from [7] directly to our setting, is that it requires us to store a Voronoi diagram for each label, for each shortest path. Each vertex of the path separator would then be a site in the stored Voronoi diagram but as each separator may be large, i.e. $\Theta(n)$, we may use as much as $\Theta(n|L|)$ space over *all* labels of L for a single separator. What we need is for the number of sites involved for a label λ to be proportional to the number of vertices with label λ ; this would give a near-linear bound on the number of sites when summing over all $\lambda \in L$ across all levels of the recursive decomposition. We address these issues in the following sections.

4.1 MSSP structure

To compactly represent shortest path trees, our point location structure uses an augmented version of the multiple-source shortest-path (MSSP) data structure of Klein [10]. It cleverly uses the persistence techniques of [5] in conjunction with top trees [1] to obtain an implicit representation of shortest path trees rooted at each site. Top-trees allow for shortest path distance queries and least-common ancestor (LCA) queries in time $O(\log n)$ per query while using $O(n \log n)$ space, and can easily be augmented to support turn queries, as we shall see shortly. To be used as a black box, the MSSP structure relies on being initialized from a face of G . In our construction, we wish to use it for querying left/right turns of paths and distances from vertices residing on shortest paths of fundamental cycle separators, and thus some further preprocessing is required. The guarantees of the augmented MSSP structure used for the point location structure are summarized in the following lemma:

► **Lemma 6.** *Let $G = (V, E, \omega)$ be an edge-weighted planar graph, f be a face of G and let T_u denote the shortest path tree rooted at u . Then there exists a data structure $\text{MSSP}(G, f)$ with $O(n \log n)$ space which given $u \in V(f)$ and $c, v \in V$ supports queries*

1. $\text{DIST}(u, v)$: report $\mathbf{d}_G(u, v)$,
2. $\text{LCA}(u, c, v)$: report the least common ancestor of v, c in T_u ,
3. $\text{TURN}(u, c, v)$: report whether $u \rightsquigarrow_{T_u} c$ makes a left or right turn w.r.t. $u \rightsquigarrow_{T_u} v$ or if one is a prefix of the other.

in time $O(\log n)$ per query. The data structure can be preprocessed in $O(n \log n)$ time.

Descriptions of DIST and LCA are available in [10] and [1], and a description of how to implement TURN is provided in Appendix A in terms of the vocabulary and interface specified in [1] for completeness. A top-tree representing any shortest path tree rooted at a vertex on f can be accessed in time $O(\log n)$ by using persistence in the MSSP structure. Lemma 6 then readily follows from applying the bound of Lemma 13 in Appendix A.

4.2 Label sequences

To address the second issue, we first need to make the following definition and state some of its properties when applied in the context of planar graphs:

► **Definition 7.** *Let $G = (V, E)$ be a graph, $p = p_1, \dots, p_k$ a sequence of vertices and $S \subseteq V$. The label-sequence of p w.r.t. S is a sequence $M_{G,S,p} \in S^k$ satisfying $M_{G,S,p}(i) = \arg \min_{s \in S} \text{dist}_G(s, p_i)$. The alternation number on p w.r.t. S in G is defined as $|M_{G,S,p}| = \sum_{i=1}^{k-1} [M_{G,S,p}(i) \neq M_{G,S,p}(i+1)]$.*

When G , S and p are clear from the context, we shall simply write M , and also note that the sequence is well-defined due the tie-breaking scheme chosen in the preliminaries. The alternation number can be thought of as the number of times consecutive vertices on p change which vertex they are closest to among S when “moving along” p .

When G is an undirected planar graph and p is a shortest path in G , it can be observed⁴ that M is essentially a Davenport-Schinzel sequence of order 2, and it immediately follows that the alternation number is “small” in the sense it is proportional to S while being agnostic towards the length of p altogether.

⁴ We thank the anonymous reviewer for this observation which saved a tedious proof.

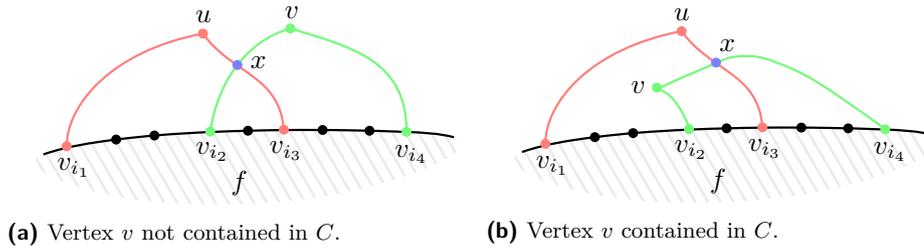


Figure 1 Illustration of the proof of Lemma 10. The concatenation of $u \rightsquigarrow v_{i_1}$, $v_{i_1} \rightsquigarrow v_{i_3}$, and the reverse of $u \rightsquigarrow v_{i_3}$ forms a cycle C and $v \rightsquigarrow v_{i_2}$ intersects $u \rightsquigarrow v_{i_3}$ in x . Note that w.l.o.g. C is not necessarily simple, that $v \rightsquigarrow v_{i_2}$ may intersect the cycle more than once and that v may be a vertex of C .

► **Definition 8** (Davenport-Schinzel [4]). A sequence u_1, u_2, \dots, u_k over an alphabet Σ on n symbols is said to be a (n, s) -Davenport-Schinzel sequence if

1. $u_i \neq u_{i+1}$ for all $1 \leq i < k$ and
2. There do not exist $s + 2$ indices $1 \leq i_1 < i_2 < \dots < i_{s+2} \leq k$ for which $u_{i_1} = u_{i_3} = \dots = u_{i_{s+1}} = u$ and $u_{i_2} = u_{i_4} = \dots = u_{i_{s+2}} = v$ for $u \neq v \in \Sigma$.

► **Lemma 9** (Davenport-Schinzel [4]). Let U be a $(n, 2)$ -Davenport-Schinzel sequence of length m . Then $|U| \leq 2n - 1$.

For a sequence of $\mathcal{S} = u_1, u_2, \dots, u_k$ over an alphabet Σ , the *contraction* of \mathcal{S} is the subsequence obtained from \mathcal{S} by replacing every maximal substring s, s, \dots, s of \mathcal{S} consisting of identical symbols s by a single occurrence of s . As an example with $\Sigma = \{0, 1, 2\}$, the contraction of $0, 0, 1, 2, 2, 1, 1, 0, 1, 0, 0, 2$ is $0, 1, 2, 1, 0, 1, 0, 2$.

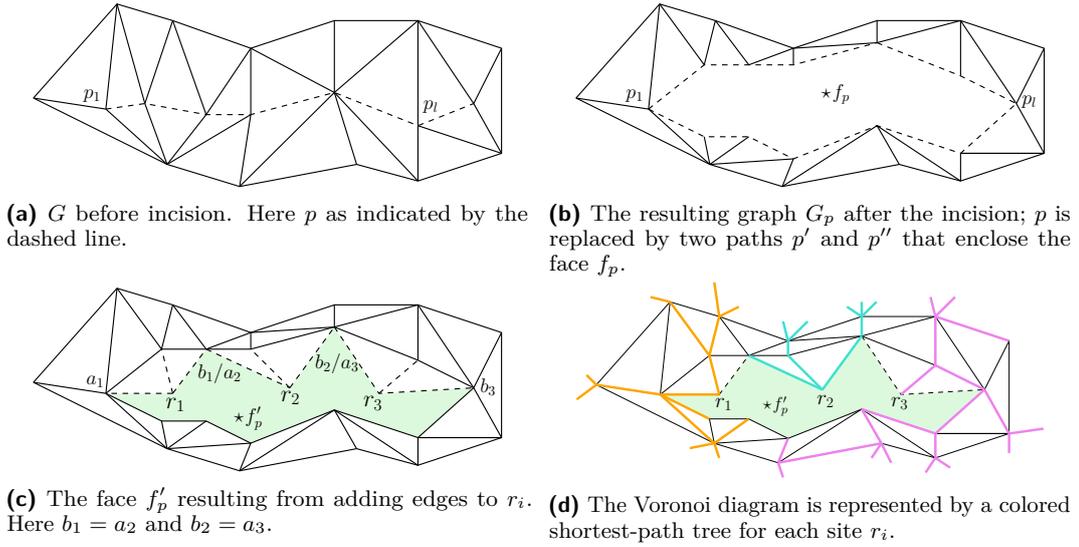
► **Lemma 10.** Let G be an undirected, weighted planar graph, let $S \subseteq V$, and let p be a shortest path of G contained in (the boundary of) a face of G . Then the contraction of M is a $(|S|, 2)$ -Davenport-Schinzel sequence.

Proof. Define $v_1, \dots, v_k = p$ and assume for the sake of contradiction that for some $1 \leq i_1 < i_2 < i_3 < i_4 \leq k$ and $u, v \in S$ with $u \neq v$, it holds that $u = M(i_1) = M(i_3)$ and $v = M(i_2) = M(i_4)$. Then the concatenation of $u \rightsquigarrow v_{i_1}$, $v_{i_1} \rightsquigarrow v_{i_3}$, and the reverse of $u \rightsquigarrow v_{i_3}$ forms a cycle. Thus, either $v \rightsquigarrow v_{i_2}$ intersects $u \rightsquigarrow v_{i_1} \cup u \rightsquigarrow v_{i_3}$ or $v \rightsquigarrow v_{i_4}$ intersects $u \rightsquigarrow v_{i_1} \cup u \rightsquigarrow v_{i_3}$; see Figure 1a and 1b. By symmetry, we only need to consider the former case. If $v \rightsquigarrow v_{i_2}$ intersects $u \rightsquigarrow v_{i_3}$ in some vertex x then $v \rightsquigarrow x$ has the same weight as $u \rightsquigarrow x$. By the “closer than”-relation, $u = M(i_3) = M(i_2) = v$, contradicting our assumption that $u \neq v$. A similar contradiction is obtained if $v \rightsquigarrow v_{i_2}$ intersects $u \rightsquigarrow v_{i_1}$. ◀

► **Corollary 11.** Let G, S and p be as in Lemma 10. Then $|M_{G,S,p}| = O(|S|)$.

We remark that M can be readily computed in polynomial time by adding a super-source connected to each vertex of S and running an SSSP algorithm. Furthermore M can be represented with $O(S)$ space, by storing only the indices for which $M(i) \neq M(i + 1)$ and $M(i)$ for each such index.

We will now describe how to achieve $O(n)$ space for storing Voronoi diagrams for all labels $\lambda \in L$ at any level of the recursive decomposition. We do so by modifying the preprocessing steps and query scheme of [7] in a manner suitable for application of Lemma 10 and Corollary 11.



■ **Figure 2** Preprocessing steps for the point-location structure.

4.3 Preprocessing

Let us briefly recall the statement of Lemma 5; that is, we let G be an undirected, edge-weighted, planar embedded graph with associated labeling $l : V \rightarrow L$ and let $p = p_1 \rightsquigarrow p_k$ be a shortest path in G . Given a query $(u, \lambda) \in V \times L$, we want to identify a small “candidate” set of vertices $C \subseteq V$ such that if v is the vertex with label λ closest to u and $u \rightsquigarrow v$ intersects p , then $v \in C$.

Here, we first describe how to compute a data structure which provides the guarantees of Lemma 5, but restricts itself to the case only where $u \rightsquigarrow v$ intersects p from the left. The description of the data structure for handling paths that intersect p from the right is symmetric (e.g. by swapping the endpoints of p). Lemma 5 thus readily follows from the existence of such structures.

First, a copy, G_p , of G is stored and an incision is added along p in G_p . This results in a planar embedded graph G_p , which has exactly one more face than G . Define by $p' = p'_1, \dots, p'_l$ and $p'' = p''_1, \dots, p''_l$ the resulting paths along the incision, where $p'_1 = p'_l$ and $p''_1 = p''_l$. We denote by f_p the face whose boundary vertices are $V(p') \cup V(p'')$. An illustration of this is provided in Figure 2a and 2b.

Next, the MSSP data structure of Lemma 6, $MSSP(G_p, f_p)$, is computed and stored as part of the point-location data structure. This structure will be used for the point location query.

Centroid decompositions of Voronoi diagrams

The following preprocessing is now done for *each* label $\lambda \in L$: First a copy, G_p^λ , of G_p is made. Next, $M_{G_p, S_\lambda, p'}$ is computed for $S_\lambda = \{v \in V \mid l(v) = \lambda\}$. For convenience we assume that $M(0) = \text{NIL}$. The preprocessing phase now consists of modifying G_p^λ before computing the Voronoi diagram and the associated centroid decomposition associated with λ as follows: For $i \leftarrow 1, \dots, l$, whenever $M(i) \neq M(i-1)$, a new vertex is added to G_p^λ and embedded in f_p along the curve formed by the deleted arc of the embedding of p . Denoting by r_i the most recently added vertex after iteration i , edge $p'_i r_i$ with $\omega(p'_i r_i) = d_G(M(i), p'_i)$ is added to G_p^λ

and embedded for all i . Once again, it is fairly easy to see that G_p^λ is planar embedded. See Figure 2c for an illustration of this. Denote by a_i and b_i the endpoints of the first and last edges added incident to r_i (w.r.t. the order in which they were added). Denote by f'_p that has $\{r_i, a_i, b_i \mid 1 \leq i \leq l\} \cup V(p'')$ as its boundary vertices. Now the Voronoi diagram, its dual and subsequently its corresponding centroid decomposition, $T_{p,\lambda}^*$, is computed in (the now modified) G_p^λ using $R = \{r_i \mid 1 \leq i \leq l\}$ as Voronoi sites, see Figure 2d. The intuition is that each site in R corresponds to a contiguous subsequence $M(k), \dots, M(l)$ of M for which $M(j) = M(j+1) = v$ where v is the vertex with label λ closest to p'_j for $k \leq j < l$. This implies that the number of sites is proportional to the number of vertices with label λ instead of the length of the original separator p : By Lemma 11, $|M| = O(|S_\lambda|)$ and since $|R| = |M|$ it follows that $|R| = O(|S_\lambda|)$ bounds the complexity of $T_{p,\lambda}^*$, which is stored as part of the data structure. As aforementioned, each centroid $c \in T_{p,\lambda}^*$ corresponds to some degree three Voronoi vertex, f_c^* , with vertices, $\{x_1, x_2, x_3\}$ in the corresponding primal face f_c s.t. each x_j belongs to a different Voronoi site r_{i_j} for $j \in \{1, 2, 3\}$. For each such j , the centroid c stores a pointer to its corresponding face f_c , the first vertex p_{k_j} of p' on $r_{i_j} \rightsquigarrow_{G_p'} x_j$ and the weight $\omega(p_{k_j} r_{i_j})$.

Space complexity

The space used for storing the MSSP structure is $O(|V| \log |V|)$ by Lemma 6. For each centroid, we store a constant amount of data, so the space required for storing the centroid decompositions is

$$\sum_{\lambda \in L} O(1) \cdot |T_{p',\lambda}^*| = \sum_{\lambda \in L} O(|S_\lambda|) = O(V)$$

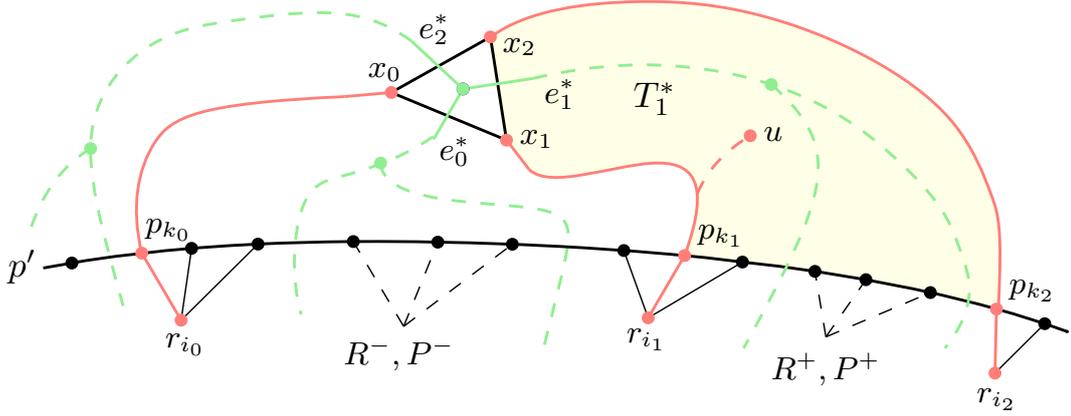
since $S_\lambda \cap S_{\lambda'} = \emptyset$ for $\lambda \neq \lambda' \in L$ as each vertex has exactly one label. The total space used is thus $O(|V| \log |V|)$.

4.4 Handling a point location query

We now show how to handle a point location query. Note that in the following, we can assume that the vertices of p' appear in increasing order of their indices when traversing the boundary of f_p in a clockwise direction. Recall that given $u \in V$ and $\lambda \in L$, we wish to find a subset $C \subset V$ of constant size, s.t. if v is the closest vertex with label λ where $u \rightsquigarrow v$ intersects p from the left, then $v \in C$. The query works by identifying a subset $P \subseteq V(p')$ s.t. for some $p'_k \in P$ it holds that $M_{G_p, S_\lambda, p'}(k) = v$. We first show how to identify the subset by recursively querying the centroid decomposition $T_{p,\lambda}^*$ according to the following lemma, which we note is modified version of the query in [7]:

► **Lemma 12.** *Given a query $(u, \lambda) \in V \times L$, consider the centroid decomposition tree $T_{p,\lambda}^*$ computed from G_p^λ in the preprocessing. Let c be a centroid $c \in T_{p,\lambda}^*$ corresponding to some Voronoi vertex, f_c^* , with associated primal triangle containing vertices $\{x_0, x_1, x_2\} = V(f_c)$ where x_j belongs to the Voronoi cell of r_{i_j} for $j \in \{0, 1, 2\}$ and $i_0 < i_1 < i_2$. Furthermore let e_j^* be the dual edge incident to f_c^* , s.t. $e_j = x_j x_{(j+1) \bmod 3}$, let p_{k_j} be the successor of r_{i_j} on $r_{i_j} \rightsquigarrow_{G_p^\lambda} x_j$, let $P_j = p_{k_j} \rightsquigarrow_{G_p} u$, and let T_j^* be the subtree of $T_{p,\lambda}^*$ attached to c by e_j^* for $j \in \{0, 1, 2\}$. Finally, let $j^* = \arg \min_{j \in \{0, 1, 2\}} \{\mathbf{d}_{G_p}(p_{k_j}, u) + \omega(r_{i_j} p_{k_j})\} = \arg \min_{j \in \{0, 1, 2\}} \{\mathbf{d}_{G_p^\lambda}(r_{i_j}, u)\}$. Then*

1. *If $p_{k_{j^*}} \rightsquigarrow_{G_p} u$ emanates from the left of P_{j^*} or $u \in P_{j^*}$, then the site closest to u in G_p^λ belongs to $R^- = \{r_{(i_{j^*-1}) \bmod 3}, \dots, r_{i_{j^*}}\}$ and the second vertex on the shortest path from that site to u in G_p^λ belongs to $P^- = \{p_{(i_{j^*-1}) \bmod 3}, \dots, p_{i_{j^*}}\}$; furthermore, $T_{j^*}^*$ is the centroid decomposition tree for G_p^λ when restricted to shortest paths from sites in R^- through successors in P^- .*



■ **Figure 3** Illustration of Lemma 12. The dashed green lines represent Voronoi edges in the centroid decomposition and the red lines the primal shortest paths to the primal vertices of the centroid. In this case, $j^* = 1$, so u is contained in primal faces spanned by the subtree T_1^* contained in the region highlighted in yellow.

2. otherwise, the site closest to u in G_p^λ belongs to $R^+ = \{r_{i_{j^*}}, \dots, r_{i_{(j^*+1) \bmod 3}}\}$ and the second vertex on the shortest path from that site to u belongs to $P^+ = \{p_{i_{j^*}}, \dots, p_{i_{(j^*+1) \bmod 3}}\}$; furthermore, $T_{(j^*+1) \bmod 3}^*$ is the centroid decomposition tree for G_p^λ when restricted to shortest paths from sites in R^+ through successors in P^+ .

Proof. By symmetry, we only consider the first case since the second case occurs when $p_{k_{j^*}} \rightsquigarrow_{G_p} u$ emanates from the right of P_{j^*} and the first case also includes the case where $u \in P_{j^*}$.

By the choice of j^* , $p_{k_{j^*}} \rightsquigarrow_{G_p} u$ cannot intersect any of the paths $P_{j'}$ with $j' \in \{(j^* - 1) \bmod 3, (j^* + 1) \bmod 3\}$ since these two paths are subpaths of shortest paths from sites $r_{i_{j'}} \neq r_{i_{j^*}}$ in G_p^λ and we assume unique shortest paths. Let x be the first vertex of $p_{k_{j^*}} \rightsquigarrow_{G_p} u$ such that either $x = u$ or the path emanates from the left of P_{j^*} at x . The rest of $p_{k_{j^*}} \rightsquigarrow_{G_p} u$ following x will not intersect P_{j^*} again due to uniqueness of shortest paths. Thus u belongs to the region of the plane enclosed by paths $P_{(j^*-1) \bmod 3}$, P_{j^*} , edge $e_{(j^*-1) \bmod 3}$, and path $p_{k_{(j^*-1) \bmod 3}}, \dots, p_{k_{j^*}}$. Note that $T_{j^*}^*$ is the subtree of $T_{p', \lambda}^*$ spanning the primal faces contained in this region. Hence, $T_{j^*}^*$ is the centroid decomposition tree for G_p^λ when restricted to shortest paths from sites in R^- through successors in P^- . ◀

For an illustration of Lemma 12, see Figure 3. The Lemma implies a fast recursive point location scheme. On query (u, λ) , obtain centroid c from $T_{p', \lambda}^*$. Since weights of edges from sites have been precomputed, $MSSP(G_p, f_p)$ is applied to find j^* . $MSSP(G_p, f_p)$ is also used to determine if $p_{k_{j^*}} \rightsquigarrow_{G_p} u$ emanates from the left of P_{j^*} and hence whether the first or second case of the lemma applies. The point location structure now recurses on a subset of sites and vertices of p' and on a subtree of $T_{p', \lambda}^*$, depending on which case applies.

The recursion stops when reaching a subtree corresponding to a bisector for two sites. The vertices of V with label λ corresponding to these two sites are reported as the set C , yielding the desired bound.

Time complexity

The $O(\log^2 |V|)$ query time bound of Lemma 5 follows since there are $O(\log |V|)$ recursion levels and in each step, a constant number of queries to $MSSP(G_p, f_p)$ are executed, each taking $O(\log |V|)$ time.

References

- 1 Stephen Alstrup, Jacob Holm, Kristian De Lichtenberg, and Mikkel Thorup. Maintaining information in fully dynamic trees with top trees. *ACM Trans. Algorithms*, 1(2):243–264, October 2005. doi:10.1145/1103963.1103966.
- 2 Panagiotis Charalampopoulos, Paweł Gawrychowski, Shay Mozes, and Oren Weimann. Almost optimal distance oracles for planar graphs. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019*, pages 138–151, New York, NY, USA, 2019. Association for Computing Machinery. doi:10.1145/3313276.3316316.
- 3 Shiri Chechik. Improved distance oracles and spanners for vertex-labeled graphs. In Leah Epstein and Paolo Ferragina, editors, *Algorithms – ESA 2012*, pages 325–336, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.
- 4 H. Davenport and A. Schinzel. A Combinatorial Problem Connected with Differential Equations. *American Journal of Mathematics*, 87(3):684, July 1965. doi:10.2307/2373068.
- 5 James R. Driscoll, Neil Sarnak, Daniel D. Sleator, and Robert E. Tarjan. Making data structures persistent. *Journal of Computer and System Sciences*, 38(1):86–124, 1989. doi:10.1016/0022-0000(89)90034-2.
- 6 Jeff Erickson, Kyle Fox, and Luvsandongov Lkhamsuren. Holiest minimum-cost paths and flows in surface graphs. In Ilias Diakonikolas, David Kempe, and Monika Henzinger, editors, *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 1319–1332. ACM, 2018. doi:10.1145/3188745.3188904.
- 7 Paweł Gawrychowski, Shay Mozes, Oren Weimann, and Christian Wulff-Nilsen. Better tradeoffs for exact distance oracles in planar graphs. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018*, pages 515–529, USA, 2018. Society for Industrial and Applied Mathematics.
- 8 Paweł Gawrychowski, Gad M. Landau, Shay Mozes, and Oren Weimann. The nearest colored node in a tree. *Theoretical Computer Science*, 710:66–73, 2018. doi:10.1016/j.tcs.2017.08.021.
- 9 Danny Hermelin, Avivit Levy, Oren Weimann, and Raphael Yuster. Distance oracles for vertex-labeled graphs. In Luca Aceto, Monika Henzinger, and Jiří Sgall, editors, *Automata, Languages and Programming*, pages 490–501, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.
- 10 Philip Klein. Multiple-source shortest paths in planar graphs. *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 146–155, January 2005.
- 11 Mingfei Li, Chu Chung Christopher Ma, and Li Ning. $(1 + \epsilon)$ -distance oracles for vertex-labeled planar graphs. In T.-H. Hubert Chan, Lap Chi Lau, and Luca Trevisan, editors, *Theory and Applications of Models of Computation*, pages 42–51, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.
- 12 Yaowei Long and Seth Pettie. Planar Distance Oracles with Better Time-Space Tradeoffs. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2517–2537. Society for Industrial and Applied Mathematics, Philadelphia, PA, January 2021. doi:10.1137/1.9781611976465.149.
- 13 Shay Mozes and Eyal E. Skop. Efficient Vertex-Label Distance Oracles for Planar Graphs. *Theory of Computing Systems*, 62(2):419–440, February 2018. doi:10.1007/s00224-017-9827-0.
- 14 Maximilian Probst. On the Complexity of the (Approximate) Nearest Colored Node Problem. In Yossi Azar, Hannah Bast, and Grzegorz Herman, editors, *26th Annual European Symposium on Algorithms (ESA 2018)*, volume 112 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 68:1–68:14, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.ESA.2018.68.
- 15 Dekel Tsur. Succinct data structures for nearest colored node in a tree. *Information Processing Letters*, 132:6–10, 2018. doi:10.1016/j.ipl.2017.10.001.

A Appendix

A description of how to implement TURN in Lemma 6 is provided here in terms of the terminology and the interface specified in [1]. Readers that are not familiar with the terminology and interface pertaining to top-trees are referred to [1].

► **Lemma 13.** *Let $G = (V, E, \omega)$ be a weighted planar graph, and let \mathcal{T} be the root-cluster of a top-tree corresponding to some tree T in G . Then in addition to DIST and LCA, we can support TURN queries: For $u, c, v \in V$, report whether $u \rightsquigarrow c$ makes a left or right turn w.r.t. $u \rightsquigarrow_T v$, or if one is a prefix of the other in time $O(\log n)$ per query.*

Proof. A description of how to perform LCA queries is found in [1]. Assume that $u, c, v \in T$ and w.l.o.g. that u is strictly more rootward in T than v . First use \mathcal{T} to determine the LCA c' of (v, c) . If c' is on both $u \rightsquigarrow c$ and $u \rightsquigarrow v$ one path is a prefix of the other. Otherwise invoke EXPOSE(u, c') and traverse \mathcal{T} until a leaf of \mathcal{T} corresponding to the edge, $e_v \in E(T)$ is reached, which connects c' to the subtree containing v in T . This can be done in time $O(\log n)$. The same is done for (u, c') and (c, c') , exposing edges $e_u, e_c \in T$. Now, if $e_c = e_v$ or $e_c = e_u$, c must be on $u \rightsquigarrow_T v$. Otherwise it is easily checked, by maintaining a cyclic order of edges in the adjacency list of c' , in constant time, whether e_c emanates to the left or right of the subpath $e_u e_v$ and hence $u \rightsquigarrow_T v$. The total time spent is $O(\log n)$. ◀