



Eccentricity queries and beyond using hub labels

Guillaume Ducoffe

► To cite this version:

Guillaume Ducoffe. Eccentricity queries and beyond using hub labels. Theoretical Computer Science, 2022, 10.1016/j.tcs.2022.07.017 . hal-03738888

HAL Id: hal-03738888

<https://hal.science/hal-03738888>

Submitted on 26 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Eccentricity queries and beyond using Hub Labels

Guillaume Ducoffe^{1,2}

¹National Institute for Research and Development in Informatics, Romania

²University of Bucharest, Romania

Abstract

Hub labeling schemes are popular methods for computing distances on road networks and other large complex networks, often answering to a query within a few microseconds for graphs with millions of edges. In this work, we study their algorithmic applications beyond distance queries. Indeed, several implementations of hub labels were reported to have good practical performances, both in terms of pre-processing time and maximum label size. There are also a few relevant graph classes for applications for which we know how to compute hub labelings with sublinear labels in quasi linear time. These positive results raise the question of what can be computed efficiently given a graph and a hub labeling of small maximum label size as input. We focus on *eccentricity queries* and *distance-sum queries*, for several versions of these problems on directed weighted graphs, that is in part motivated by their importance in facility location problems. On the negative side, we show conditional lower bounds for these above problems on unweighted undirected sparse graphs, via standard constructions from “Fine-grained” complexity. Specifically, under the Strong Exponential-Time Hypothesis (SETH), answering to these above queries requires $\Omega(|V|^{2-o(1)})$ pre-processing time or $\Omega(|V|^{1-o(1)})$ query time, even if we are given as input a hub labeling of maximum label size in $\omega(\log |V|)$. However, things take a different turn when the hub labels have a *sublogarithmic* size. Indeed, for every $\varepsilon > 0$ there exists a $\delta_\varepsilon > 0$ such that the following hold, being given a hub labeling of maximum label size $\leq k$: after pre-processing the labels in total $O(2^{\delta_\varepsilon \cdot k} \cdot |V|^{1+\varepsilon})$ time, we can compute both the eccentricity and the distance-sum of any vertex in $O(2^{\delta_\varepsilon \cdot k} \cdot |V|^\varepsilon)$ time. Our data structure is a novel application of the orthogonal range query framework of (Cabello and Knauer, *Computational Geometry*, 2009). It can also be applied to the fast *global* computation of some topological indices. Finally, as a by-product of our approach, on any fixed class of unweighted graphs with *bounded expansion*, we can decide whether the diameter of an n -vertex graph in the class is at most k in $f_\varepsilon(k) \cdot n^{1+\varepsilon}$ time for any $\varepsilon > 0$, for some “explicit” function f_ε that depends on ε and the graph class considered. This result is further motivated by the empirical evidence that many classes of complex networks have bounded expansion (Demaine et al., *Journal of Computer and System Sciences*, 2019), and that some of these networks, such as the online social networks, have a relatively small diameter.

1 Introduction

We refer to [11, 25] for basic notions and terminology in Graph Theory. Real-world networks can be conveniently represented as a weighted graph $G = (V, E, w)$, with typical interpretations of the weight function $w : E \rightarrow \mathbb{R}_{>0}$ being the latency of a link, the length of a road between two locations, etc. If $\forall e \in E, w_e = 1$, then we can drop the weight function, calling the graph unweighted and

writing $G = (V, E)$. Unless stated otherwise, all graphs considered in this note are finite, weighted, directed and strongly connected¹. For every ordered pair of vertices (u, v) , the distance $\text{dist}_G(u, v)$ is defined as the smallest weight of a path from u to v . We omit the subscript if the graph G is clear from the context. We note that distance computation is an important primitive in many scenarios, ranging from Satellite Navigation devices to forwarding network packets. However, in some practical or theoretical applications, the “right” notion of distance is a twist of the one presented above. Let us give a few examples (taken from [2]).

- **Source distance:** $\text{dist}(u, v)$. This is the standard distance.
- **Max distance:** $\max\{\text{dist}(u, v), \text{dist}(v, u)\}$. It turns the distances in a directed graph into a metric (*i.e.*, it is symmetric and it satisfies the triangular inequality).
- **Min distance:** $\min\{\text{dist}(u, v), \text{dist}(v, u)\}$. As a concrete application of the latter, Abboud et al. [2] cited the problem of optimally locating a hospital (minimizing the time for either the patient coming to the hospital, or an ambulance drive being sent at the patient’s home).
- **Roundtrip distance:** $\text{dist}(u, v) + \text{dist}(v, u)$. It also turns the distances in a directed graph into a metric. Many distance computation techniques for *undirected* graphs can be generalized to roundtrip distances, such as *e.g.*, the existence of low-stretch sparse spanners [52].

A *distance oracle* is a data structure for computing distances. Next, we present a well-known family of distance oracles, of which we explore in this work the algorithmic properties.

Hub labeling. A 2-hopset or, as we call it here, a hub labeling, assigns to every vertex v an ordered pair $(L^+(v), L^-(v))$ of vertex-subsets along with the distances $\text{dist}(v, x)$, for $x \in L^+(v)$, and $\text{dist}(x, v)$, for $x \in L^-(v)$. This labeling must ensure that, for any ordered pair of vertices (u, v) , we have $\text{dist}(u, v) = \min\{\text{dist}(u, x) + \text{dist}(x, v) \mid x \in L^+(u) \cap L^-(v)\}$. In particular, we can compute $\text{dist}(u, v)$ in $O(|L^+(u)| + |L^-(v)|)$ time – and more generally, the source, max-, min- and roundtrip distances from u to v in $O(|L^+(u)| + |L^-(u)| + |L^+(v)| + |L^-(v)|)$ time. It has been observed in [44] that most distance labeling constructions are based on hub label schemes, thereby further motivating our study of the latter as a first step toward a more general analysis of the algorithmic properties of distance oracles.

The maximum label size is defined as $\max\{\frac{|L^+(v)| + |L^-(v)|}{2} \mid v \in V\}$ [23], and it is the main complexity measure for hub labelings². Every graph admits a hub labeling of maximum label size $\Theta(n)$, and this is essentially optimal [35]. Indeed, for an unweighted graph and any of its m arcs uv , we must have one of u or v that is contained in $L^+(u) \cap L^-(v)$. As a result, the cumulated size of all the labels must be at least $\Omega(m)$, and so the maximum label size must be at least $\Omega(m/n)$ – that is $\Omega(n)$ for dense graphs. However, it has been experimentally verified that many real-world graphs, and especially road networks, admit hub labelings with small maximum label size [18]. From a more theoretical point of view, every n -node planar graph admits a hub labeling with maximum label size $O(\sqrt{n} \log n)$ [35]. Bounded-treewidth graphs and graphs of bounded *highway dimension* – a parameter conjectured to be small for road networks – admit hub labelings with (poly)logarithmic

¹The strongly connected assumption could be easily dropped, but it makes the presentation simpler by avoiding the pathological case of infinite distances.

²The *bit size* is larger by at least a logarithmic factor, because we need to encode the vertices’ IDs and the distances values for all the vertices inside the labels.

maximum label size [3, 35]. Graphs of bounded tree-depth even have hub labelings with *constant* maximum label size, that can be computed in quasi linear time [41].

Eccentricity and Distance-sum computations. Since many hub labeling schemes are readily implemented [6, 4, 5, 23, 45], our approach consists in assuming a graph to be given with a hub labeling, and to study whether this additional information helps in solving faster certain distance problems. Specifically, let $D(\cdot, \cdot)$ represent any distance function on directed graphs (source, min-, max- or roundtrip). The *eccentricity* of a vertex u , denoted by $e^D(u)$, is the maximum distance from u to any other vertex, *i.e.*, $e^D(u) = \max_v D(u, v)$. The *diameter* and the radius of a graph – w.r.t. D – are, respectively, the maximum and minimum eccentricities amongst its vertices. The *distance-sum* of a vertex u , that we denote by $s^D(u)$, is the sum of all the distances from u to any other vertex, *i.e.*, $s^D(u) = \sum_v D(u, v)$. A median is a vertex minimizing its distance-sum, while the Wiener index is equal to $\sum_u s^D(u)$. We stress that both computing the radius and the median set of a graph are fundamental *facility location* problems: with applications to optimal placement of schools, hospitals and other important facilities on a network. The problems of computing the diameter and the Wiener index found applications in network design and chemistry [54], respectively. In Sec. 3, we consider other topological indices from chemical graph theory.

1.1 Related work

Computing the eccentricities in a graph is a notoriously “hard” problem in P, for which no quasi linear-time algorithm is likely to exist. For *weighted* n -vertex graphs, the problem is subcubic equivalent to All-Pairs Shortest-Paths (APSP). It is conjectured that we cannot solve APSP in $O(n^{3-\varepsilon})$ time in dense graphs for any $\varepsilon > 0$ [1]. For the special case of *unweighted* graphs, we can compute the eccentricities in truly subcubic time using fast matrix multiplication [56, 63]. However if we assume the Strong Exponential-Time Hypothesis (SETH), then it cannot be improved to $O(n^{2-\varepsilon})$ time even for sparse graphs, for any $\varepsilon > 0$ [53]. Under a related hypothesis, the so-called “Hitting Set Conjecture”, similar hardness results were proved for median computation [2]. For all that, some heuristics and exact algorithms for diameter, radius and eccentricities computations do perform well in practice [13, 22, 57, 58], which has been partially justified from a theoretical viewpoint [26, 60, 61].

These hardness results have renewed interest in characterizing the graph classes for which we can compute some of these above distance problems in truly subquadratic time, or even in quasi linear time. In this respect, several recent results were proved by using classical tools from Computational Geometry [2, 14, 17, 15, 28, 30, 29, 36]. Perhaps the most impressive such result is the truly subquadratic algorithm of Cabello for computing the diameter and the Wiener index of planar graphs [15]. It is worth of mentioning that Cabello’s techniques were the starting point of recent breakthrough results for exact *distance oracles* on planar graphs [19, 37]. In this paper, we ask whether conversely, having at hands an exact distance oracle with certain properties can be helpful in the design of fast algorithms for some distance problems on graphs. We give a partial answer for hub labelings.

Hub Labelings. Computing a hub labeling with optimal maximum label size is NP-hard [8], but it can be approximated up to a logarithmic factor in polynomial time [18]. The use of hub labels beyond distance queries was already considered in [32] and [64] for k -nearest neighbours queries and shortest-path counting, respectively. Several generalizations of hub labels were proposed. For

instance, a 3-hopset assigns a pair of labels $(L^-(v), L^+(v))$ for every vertex v (as for hub labelings), but it computes an additional *global* set of distances, denoted by E_2 . This labeling must ensure that, for every vertices u and v , $\text{dist}(u, v) = \min\{\text{dist}(u, x) + \text{dist}(x, y) + \text{dist}(y, v) \mid x \in L^+(u), y \in L^-(v), (x, y) \in E_2\}$ [38]. A *distance labeling scheme* is given by an encoding function L and a decoding function D . It assigns a label $L(v)$ to every vertex v s.t., for every vertices u and v , we have $\text{dist}(u, v) = D(L(u), L(v))$ [35]. These extensions are to be investigated in future work.

1.2 Results

We first address the following types of queries on the vertices v of a graph G .

- **Eccentricity query:** compute the eccentricity of v ;
- **Distance-sum query:** compute the distance-sum of v .

Before stating our results, let us discuss about naive resolutions for these two types of queries. If we compute APSP in $O(|V|^3)$ time, then we can compute all eccentricities and distance-sums in additional $O(|V|^2)$ time. In doing so, we get a trivial data structure in $O(|V|)$ -space and $O(1)$ query time. Therefore, the main challenge here is to decrease the pre-processing time. A hub labeling (or more generally, a distance-labeling) of maximum label size k is an implicit representation of the distance-matrix in $O(k|V|)$ space. It allows us to answer these above two types of queries in $O(k|V|)$ time. However, this query time is in $\Omega(|V|)$, even in the very favourable case of $k = O(1)$. If for instance, the graph considered is sparse, then this is no better than computing a shortest-path tree from scratch. Our first observation is that, *in general*, hub labelings do *not* help. Namely:

Theorem 1. *Under SETH, for any $\varepsilon > 0$, any data structure for answering eccentricity queries or distance-sum queries requires $\Omega(|V|^{2-\varepsilon})$ pre-processing time or $\Omega(|V|^{1-\varepsilon})$ query time. This holds even if a hub labeling of maximum label size $\omega(\log |V|)$ is given as part of the input.*

Proof. We consider undirected graphs, for which all the aforementioned distances are equivalent (up to a factor 2 for the roundtrip distance). In particular, a split graph is a graph whose vertex-set can be bipartitioned into a clique K and a stable set S . We observe that any unweighted split graph $G = (K \cup S, E)$ admits a hub labeling with maximum label size $O(|K|)$, that can be computed in $O(|K||S|)$ time: simply store, for each vertex, its distance from and to any vertex of K , that is either 1 or 2. If we further assume $|K| = n^{o(1)}$, then this trivial hub labeling can be computed in $n^{1+o(1)}$ time.

Under SETH, for any $\varepsilon > 0$, we cannot compute the diameter of n -vertex split graphs in $O(n^{2-\varepsilon})$ time, and this result holds even if $|K| = \omega(\log n)$ [12]. Assume the existence of a data structure for eccentricity queries, with $P(n)$ pre-processing time and $Q(n)$ query time. In particular, we can compute the diameter in $P(n) + n \cdot Q(n)$ time. Therefore, $P(n) = \Omega(n^{2-\varepsilon})$, or $Q(n) = \Omega(n^{1-\varepsilon})$.

In the same way, assume the existence of a data structure for distance-sum queries, with $P'(n)$ pre-processing time and $Q'(n)$ query time. Observe that the diameter of a non-complete split graph is either 2 or 3. It is folklore (see, *e.g.*, [14]), that an undirected unweighted graph has diameter at most 2 if and only if the distance-sum of any vertex of degree d is equal to $2(n-1) - d$. In particular, we can compute the diameter of an unweighted undirected split graph in $P'(n) + n \cdot Q'(n)$ time. As a result, under SETH we must also have $P'(n) = \Omega(n^{2-\varepsilon})$, or $Q'(n) = \Omega(n^{1-\varepsilon})$. \square

A framework for fast queries computation. Positive results can be derived for hub labelings with sublogarithmic labels. Our algorithm is a novel application of a popular framework for fast diameter computation, using orthogonal range queries [2, 14, 17, 27, 29]. In what follows, the $O_\varepsilon(\cdot)$ notation hides a constant factor that depends on ε .

Theorem 2. *Let $\varepsilon > 0$ be arbitrary. For source, min-, max- and roundtrip distances, for every graph $G = (V, E, w)$, if we are given a hub labeling with maximum label size $\leq k$, then we can compute a data structure for answering eccentricity queries and distance-sum queries with $2^{O_\varepsilon(k)} \cdot |V|^\varepsilon$ query time. This takes $2^{O_\varepsilon(k)} \cdot |V|^{1+\varepsilon}$ pre-processing time.*

We stress that in practice, the bottleneck of Theorem 2 is the computation of a hub labeling. For this task, we are bound to use heuristics [23]. In a few restricted classes, such as graphs of bounded tree-depth, *constant-size* hub labels can be computed in quasi linear time [41]. – Below, we give an application of this result to graph classes of bounded expansion. – Let us cite, as another example of graphs with small hub labels, the graphs of bounded vertex-cover. Indeed, we can easily derive a hub labeling of maximum label size k from a vertex-cover of cardinality k . However, for the *unweighted* graphs of bounded vertex-cover, there exist slightly faster methods for eccentricities and distance-sum computations than our Theorem 2 [14, 20]. We hope that our results will encourage the quest for other graph classes that admit constant-size hub labels. For instance, it was conjectured in [3] that all graphs of highway dimension at most h admit a hub labeling with maximum label size $O(h)$. Our results in this paper show that solving this conjecture in the affirmative would be a first step toward fast diameter computation within these graphs, with applications to road networks.

Our techniques are inspired from those in [1] for bounded treewidth graphs, of which they are to some extent a generalization. Indeed, the algorithms in [1] for graphs of treewidth at most k parse a hub labeling of maximum label size in $O(k \log n)$, but where every label is composed of $O(\log n)$ levels, each containing $O(k)$ vertices. Roughly, it allows the authors from [1] to build an algorithm of logarithmic recursion depth, where at each call we may assume to be given hub labels of maximum size $O(k)$. However, in [1], all the nodes considered have the same $O(k)$ vertices in their labels (the latter forming a balanced separator). This is no more true for general hub labels of maximum size $O(k)$, a case which requires more complex range queries than in [1]. Incidentally, for *roundtrip distances*, we improved on our way the $2^{O_\varepsilon(k^2)} n^{1+\varepsilon}$ -time algorithm from [1] for directed graphs of treewidth at most k :

Corollary 1. *For the roundtrip distance, for any n -vertex directed graph of treewidth at most k , we can compute all the eccentricities and distance-sums in $2^{O_\varepsilon(k)} n^{1+\varepsilon}$ time for any $\varepsilon > 0$.*

Application to graph classes of bounded expansion. Our initial motivation for this work was to study constant diameter computation within graph classes of *bounded expansion*. Without entering too much into technical details, the graph classes of bounded expansion are exactly those whose so-called shallow minors are all sparse (see Sec. 4 for a formal definition). In particular, it generalizes bounded-degree graphs, bounded-treewidth graphs, and more generally properly minor-closed graph classes [47]. In [24], evidence was given that many classes of complex networks exhibit a bounded-expansion structure. Among those networks, social networks are well-known to obey the so-called “small-world” property [62], that implies a relatively small diameter. In this context, our theoretical results for constant-diameter computation within graph classes of bounded expansion might be a first step toward more practical algorithms for computing the diameter of social networks and other complex networks with similar properties.

Theorem 3. *Let $\varepsilon > 0$ be arbitrary. For every class of graphs \mathcal{G} of bounded expansion, for every n -vertex unweighted graph $G \in \mathcal{G}$ and positive integer k , we can decide whether the diameter of G is at most k in $O(f_\varepsilon(k) \cdot n^{1+\varepsilon})$ time, for some function f_ε that only depends on \mathcal{G} and ε .*

Diameter computation within unweighted graph classes of bounded expansion has already received some attention in the literature. In particular, for the special case of undirected graphs of maximum degree 3, Dahlgaard and Evald proved that under SETH, we cannot compute the diameter in truly subquadratic time [33]. But their hardness results hold for bounded-degree graphs of super-logarithmic diameter. In contrast to this negative result, testing whether a graph has diameter at most some *constant* k can be written as a first-order formula of size $O(k)$. Therefore, in any class of unweighted graphs of bounded expansion, we can derive from a prior work of Dvořák et al. a quasi linear-time *parameterized* algorithm for constant diameter computation [31]. Unfortunately, the hidden dependency in the parameter k is rather huge due to the use of Courcelle’s theorem [21]. Recently, we proposed a different approach for constant diameter computation within nowhere dense graph classes – a broad generalization of the graph classes of bounded expansion –, based on a VC-dimension argument [30]. However, the running time of our algorithm for deciding whether the diameter of an n -vertex graph is at most k was of order $\tilde{O}(n^{2-f(k)})$ for some super-exponential function f .

Our Theorem 3 improves on these previous works by using *low tree-depth decompositions*: a covering of n -vertex graphs of bounded expansion by relatively few subgraphs of bounded *tree-depth*, so that for a fixed k , each k -vertex subgraph is contained in at least one subgraph of this covering [47]. Combined with previous results on hub labelings within bounded tree-depth graph classes [41], it allows us to prove the existence, for every fixed k , of $O(f(k))$ -size hub labels for the pairs of vertices at distance $\leq k$, for some “explicit” function f – about a tower of exponentials of height four. Then, Theorem 3 follows from our Theorem 2. We left open whether our approach could be generalized to *nowhere dense* graph classes.

1.3 Organization of the paper

Our main technical contribution (Theorem 2) is proved in Sec. 2. In Sec. 3.1 and 3.2, we discuss applications of Theorem 2 to, respectively, bounded-treewidth graphs (Corollary 1) and the computation of topological indices. In particular, in Sec. 3.2, we exploit recent results of Cabello from [16], and we present a new type of distance information which can be computed from the orthogonal range query framework. Finally, we conclude this paper in Sec. 4 with our results for graph classes of bounded expansion (Theorem 3).

2 Properties of Hub Labels

2.1 Range Queries

We first review some important terminology, and prior results. Let V be a set of k -dimensional points, where each point $\vec{p} = (p_1, p_2, \dots, p_k) \in V$ is assigned some value $f(\vec{p})$. A box is the cartesian product of k intervals $I_i = (l_i, u_i)$, for $1 \leq i \leq k$, denoted by $\mathcal{R} = I_1 \times I_2 \times \dots \times I_k$. Note that we allow each of the k intervals to exclude either of its ends, and that we allow these ends to be infinite. Furthermore, $\vec{p} \in \mathcal{R} \cap V$ if and only if $\forall 1 \leq i \leq k, p_i \in I_i$. A *range query* asks for some information about the points within a given box. We use the following types of range queries:

- **Max-Query:** Compute a point $\vec{p} \in \mathcal{R}$ maximizing $f(\vec{p})$;
- **Sum-Query:** Compute the sum $\sum_{\vec{p} \in \mathcal{R}} f(\vec{p})$;
- **Count-Query:** Count $|V \cap \mathcal{R}|$ (the latter can be obtained from the above sum-query by setting $\forall \vec{p} \in V, f(\vec{p}) = 1$).

We refer to [14] for a thorough treatment of range queries and their applications to distance problems on graphs. The k -dimensional range tree is a classic data structure in order to answer range queries efficiently for static point sets. Evidence of its practicality for graph problems was given in [46].

Lemma 1 ([14]). *Let $B(n, k) = \binom{k + \lceil \log n \rceil}{k}$. For all the aforementioned types of range queries, for every k -dimensional point set V of size n , we can construct a k -dimensional range tree in $O(k^2 B(n, k)n)$ time, that allows to answer a query in $O(2^k B(n, k))$ time. Furthermore, $B(n, k) = 2^{O_\varepsilon(k)} \cdot n^\varepsilon$ for every $\varepsilon > 0$.*

2.2 Proof of Theorem 2

We are now ready to prove our main algorithmic tool for this paper. Our approach is essentially independent from the type of query and the distance considered. We first present our results for eccentricity queries, postponing the slight changes to be made for distance-sum queries until the end of this section. Similarly, we postpone the specific parts for each distance function to the end of the section. In what follows, let $D(\cdot, \cdot)$ be an arbitrary distance function (*i.e.*, source, min-, max- or roundtrip).

Decomposition-based techniques. Let us fix a hub labeling for $G = (V, E, w)$ of maximum label size $\leq k$. It implies that for every vertex v , we must have $1 \leq \min\{|L^-(v)|, |L^+(v)|\} \leq \max\{|L^-(v)|, |L^+(v)|\} \leq 2k - 1$. For every $1 \leq i, j \leq 2k - 1$ such that $i + j \leq 2k$, let $V_{i,j} = \{v \in V \mid |L^-(v)| = i, |L^+(v)| = j\}$. In particular, $(V_{i,j})_{1 \leq i, j \leq 2k-1}$ is a partition of V . For a fixed choice of i, j and any vertex u , we want to compute: $e^D(u, V_{i,j}) = \max_{v \in V_{i,j}} D(u, v)$. Indeed, $e^D(u) = \max_{i,j} e^D(u, V_{i,j})$. For that, let us fix a total ordering over V . For any vertex $u \in V$ and $X \subseteq L^+(u)$, $Y \subseteq L^-(u)$, we now define:

$$e^D(u, V_{i,j}, X, Y) = \max\{D(u, v) \mid v \in V_{i,j}, L^+(u) \cap L^-(v) = X, \\ L^-(u) \cap L^+(v) = Y\}.$$

We observe that $e^D(u, V_{i,j}) = \max_{X \subseteq L^+(u), Y \subseteq L^-(u)} e^D(u, V_{i,j}, X, Y)$, that follows from the definition of hub labelings. Finally, let $(x, y) \in X \times Y$ be arbitrary. Recall that we totally ordered the vertices of the graph. We further reduce the computation of all eccentricities to computing:

$$e^D(u, V_{i,j}, X, Y, x, y) = \max\{D(u, v) \mid v \in V_{i,j}, \\ L^+(u) \cap L^-(v) = X, L^-(u) \cap L^+(v) = Y, \\ x \text{ is the least vertex of } X \text{ s.t. } \text{dist}(u, v) = \text{dist}(u, x) + \text{dist}(x, v), \\ y \text{ is the least vertex of } Y \text{ s.t. } \text{dist}(v, u) = \text{dist}(v, y) + \text{dist}(y, u)\},$$

for all $u \in V$ and $\varphi = (V_{i,j}, X, Y, x, y)$. As before we can observe that $e^D(u, V_{i,j}, X, Y) = \max_{x \in X, y \in Y} e^D(u, V_{i,j}, X, Y, x, y)$. Furthermore, because the maximum label size is $\leq k$, for any fixed vertex u there are at most $O(k^4 4^k) = 2^{O(k)}$ tuples $(V_{i,j}, X, Y, x, y)$ to consider. We will show how to reduce the computation of $e^D(u, V_{i,j}, X, Y, x, y)$ to $2^{O(k)}$ range queries. The following technical lemma, inspired from [14, 17], is the gist of our approach in the paper.

Lemma 2. Let a hub labeling with maximum label size $\leq k$ be given for a graph $G = (V, E, w)$. In $O(2^{O(k)}|V|)$ time, we can map every vertex $v \in V$ to a $(2(|L^-(v)| + |L^+(v)|) - 1)$ -dimensional point-set $\{\overrightarrow{p(v, s, t)} \mid 1 \leq s \leq |L^-(v)|, 1 \leq t \leq |L^+(v)|\}$.

Moreover, for every $u \in V$ and $\varphi = (V_{i,j}, X, Y, x, y)$, let $V_{u,\varphi}^{\leq}$ (resp., $V_{u,\varphi}^{>}$) be the set of all vertices $v \in V_{i,j}$ s.t.: $L^+(u) \cap L^-(v) = X$, $L^-(u) \cap L^+(v) = Y$, x is the least vertex of X such that $\text{dist}(u, v) = \text{dist}(u, x) + \text{dist}(x, v)$, y is the least vertex of Y such that $\text{dist}(v, u) = \text{dist}(v, y) + \text{dist}(y, u)$, and $\text{dist}(u, v) \leq \text{dist}(v, u)$ (resp., $\text{dist}(u, v) > \text{dist}(v, u)$). In $2^{O(k)}$ time, we can compute a family of $2^{O(k)}$ boxes $\mathcal{R}\langle \overrightarrow{\ell_X}, \overrightarrow{\ell_Y} \rangle$ where $\overrightarrow{\ell_X} = (\ell_{x^*})_{x^* \in X} \subseteq [i+1]^X$, $\overrightarrow{\ell_Y} = (\ell_{y^*})_{y^* \in Y} \subseteq [j+1]^Y$, and we have:

$$v \in V_{u,\varphi}^{\leq} \text{ (resp., } V_{u,\varphi}^{>} \text{)} \iff \exists (\overrightarrow{\ell_X}, \overrightarrow{\ell_Y}) \text{ s.t. } \overrightarrow{p(v, \ell_x, \ell_y)} \in \mathcal{R}\langle \overrightarrow{\ell_X}, \overrightarrow{\ell_Y} \rangle.$$

If $v \in V_{u,\varphi}^{\leq}$ (resp., $V_{u,\varphi}^{>}$), then there exists a unique $(\overrightarrow{\ell_X}, \overrightarrow{\ell_Y})$ s.t. $\overrightarrow{p(v, \ell_x, \ell_y)} \in \mathcal{R}\langle \overrightarrow{\ell_X}, \overrightarrow{\ell_Y} \rangle$.

Proof. For every $v \in V_{i,j}$ and $1 \leq s \leq i$, $1 \leq t \leq j$, the point $\overrightarrow{p(v, s, t)} = (p_\ell(v, s, t))_{1 \leq \ell \leq 2(i+j)-1}$ is defined as follows:

- The elements $p_1(v, s, t), p_2(v, s, t), \dots, p_i(v, s, t)$ are the vertices in $L^-(v)$ totally ordered. In particular, these i first coordinates form an i -dimensional point $\overrightarrow{q^-(v)} = (q_\ell^-(v))_{1 \leq \ell \leq i}$ that is a common prefix to all the points $\overrightarrow{p(v, s', t')}$.
- The elements $p_{i+1}(v, s, t), p_{i+2}(v, s, t), \dots, p_{i+j}(v, s, t)$ are the vertices in $L^+(v)$ totally ordered. These j consecutive coordinates form a j -dimensional point $\overrightarrow{q^+(v)} = (q_\ell^+(v))_{1 \leq \ell \leq j}$ that is also common to all the points $\overrightarrow{p(v, s', t')}$.
- The elements $p_{i+j+1}(v, s, t), p_{i+j+2}(v, s, t), \dots, p_{2i+j-1}(v, s, t)$ are equal to the values: $\text{dist}_G(q_r^-(v), v) - \text{dist}_G(q_s^-(v), v)$, for $1 \leq r \leq i, r \neq s$. In particular, these coordinates may be different between $\overrightarrow{p(v, s, t)}$ and $\overrightarrow{p(v, s', t')}$, for $s \neq s'$.
- The elements $p_{2i+j}(v, s, t), p_{2i+j+1}(v, s, t), \dots, p_{2(i+j)-1}(v, s, t)$ are equal to: $\text{dist}_G(v, q_t^+(v)) - \text{dist}_G(v, q_t^+(v))$, for $1 \leq r \leq j, r \neq t$. In particular, these coordinates may be different between $\overrightarrow{p(v, s, t)}$ and $\overrightarrow{p(v, s', t')}$, for $t \neq t'$.
- Finally, we set the last coordinate to $p_{2(i+j)-1}(v, s, t) = \text{dist}_G(q_s^-(v), v) - \text{dist}_G(v, q_t^+(v))$.

Then, let $u \in V$ and $\varphi = (V_{i,j}, X, Y, x, y)$ be fixed. In order to restrict ourselves to the vertices $v \in V_{i,j}$ s.t. $L^+(u) \cap L^-(v) = X$, we define a family of $2^{O(k)}$ range queries over the points $\overrightarrow{q^-(v)}$, $v \in V_{i,j}$.

1. We encode the set of $|X|$ indices in which the nodes of X must be found. That is, let $X = (x_1, x_2, \dots, x_{|X|})$ be totally ordered. For a fixed choice of $1 \leq \ell_1 < \ell_2 < \dots < \ell_{|X|} \leq i$, we restrict ourselves the vertices v such that $\forall 1 \leq r \leq |X|$, $q_{\ell_r}^-(v) = x_r$. We stress that there are only $\binom{i}{|X|} = 2^{O(k)}$ possibilities.
2. In order to exclude $L^+(u) \setminus X$ from $L^-(v)$, we encode, for each vertex z in this subset, the least index ℓ s.t. $q_\ell^-(v)$, and so, all subsequent vertices of $L^-(v)$, is greater than z ; if no such index ℓ exists, then by convention we associate to z the value $i+1$. Specifically, for every

$z \in L^+(u) \setminus X$, we pick $1 \leq \ell_z \leq i+1$ and we add the following range constraints for our query:

$$\begin{cases} \forall 1 \leq \ell < \ell_z, q_\ell^-(v) < z \\ \forall \ell_z \leq \ell \leq i, q_\ell^-(v) > z \end{cases}$$

Since we totally ordered $L^+(u)$, each possibility is fully characterized by: (i) the set of indices $\{\ell_z \mid z \in L^+(u) \setminus X\}$; (ii) and an ordered partition of $L^+(u) \setminus X$ such that two vertices z, z' are in the same group if and only if $\ell_z = \ell_{z'}$. Hence, the number of possibilities here is at most $\sum_{\ell=1}^{|L^+(u)|-|X|} \binom{i+1}{\ell} \cdot 2^{|L^+(u)|-|X|} \leq 2^{k+1} \cdot 2^{|L^+(u)|-|X|} = 2^{O(k)}$.

Furthermore, in the exact same way as above, in order to restrict ourselves to the points $v \in V_{i,j}$ s.t. $L^-(u) \cap L^+(v) = Y$, we can define a family of $2^{O(k)}$ range queries over the points $\overrightarrow{q^+(v)}$, $v \in V_{i,j}$. For the remainder of the proof, let us fix one query over the points $\overrightarrow{q^-(v)}$, $v \in V_{i,j}$, and one query over the points $\overrightarrow{q^+(v)}$, $v \in V_{i,j}$. Note that we can represent the latter as two sequences of indices $\overrightarrow{\ell_X} = (\ell_{x^*})_{x^* \in X}$ and $\overrightarrow{\ell_Y} = (\ell_{y^*})_{y^* \in Y}$. We want to further restrict ourselves to the vertices v s.t. x is the least vertex of X satisfying $\text{dist}(u, v) = \text{dist}(u, x) + \text{dist}(x, v)$, and in the same way y is the least vertex of Y satisfying $\text{dist}(v, u) = \text{dist}(v, y) + \text{dist}(y, u)$. The above condition on x is equivalent to have $\forall x' \in X$:

$$\begin{cases} (1) x' < x \implies \text{dist}_G(u, x') + \text{dist}_G(x', v) > \text{dist}_G(u, x) + \text{dist}_G(x, v) \\ (2) x' > x \implies \text{dist}_G(u, x') + \text{dist}_G(x', v) \geq \text{dist}_G(u, x) + \text{dist}_G(x, v) \end{cases}$$

which can be rewritten as:

$$\begin{cases} (1) \text{dist}_G(q_{\ell_{x'}}^-(v), v) - \text{dist}_G(q_{\ell_x}^-(v), v) > (\text{dist}_G(u, x) - \text{dist}_G(u, x')) \\ (2) \text{dist}_G(q_{\ell_{x'}}^-(v), v) - \text{dist}_G(q_{\ell_x}^-(v), v) \geq (\text{dist}_G(u, x) - \text{dist}_G(u, x')) \end{cases}$$

and under this form, can be encoded as additional range constraints over the points $\overrightarrow{p(v, \ell_x, \ell_y)}$, for the indices between $i+j+1$ and $2i+j-1$. We proceed similarly for the desired condition on the vertex y . Finally, in order to complete these inequalities into a range query for $V_{u,\varphi}^{\leq}$, we further impose: $p_{2(i+j)-1}(v, \ell_x, \ell_y) = \text{dist}_G(q_{\ell_x}^-(v), v) - \text{dist}_G(v, q_{\ell_y}^+(v)) \leq (\text{dist}_G(y, u) - \text{dist}_G(u, x))$. Indeed, in this situation, $\text{dist}_G(u, v) = \text{dist}_G(u, x) + \text{dist}_G(x, v) \leq \text{dist}_G(v, y) + \text{dist}_G(y, u) = \text{dist}_G(v, u)$. We proceed similarly for $V_{u,\varphi}^>$ by changing direction of the last inequality. \square

We stress that the point-sets constructed in the proof of Lemma 5 are different than those presented in prior works [14, 17]. Indeed, this is because we need to account for different vertices not having exactly the same vertices in their respective hub labels. For that, we need to increase the dimension of our point-sets by a constant factor compared to [14, 17].

We split the proof for eccentricity queries in two lemmas so as to take into account additional technicalities for the min- and max-distance. Specifically:

Lemma 3. *Let a hub labeling with maximum label size $\leq k$ be given for a graph $G = (V, E, w)$. If $D(\cdot, \cdot)$ is the source distance or the roundtrip distance, then for any $\varepsilon > 0$, after a pre-processing in $2^{O_\varepsilon(k)} |V|^{1+\varepsilon}$ time we can compute $e^D(u, V_{i,j}, X, Y, x, y)$, for any $u \in V$ and $\varphi = (V_{i,j}, X, Y, x, y)$, in $2^{O_\varepsilon(k)} |V|^\varepsilon$ time.*

Proof. We create $O(k^4)$ range trees, for point sets of various dimensions. More specifically, let $i, j \in \{1, \dots, 2k-1\}$ such that $i+j \leq 2k$ be fixed. We create $ij = O(k^2)$ different $(2(i+j)-1)$ -dimensional range trees, that are indexed by all possible pairs $(s, t) \in [i] \times [j]$. For every vertex $v \in V_{i,j}$, we insert the point $\overrightarrow{p(v, s, t)}$ (defined in Lemma 2) in the range tree with same index (s, t) . The corresponding value $f_D(\overrightarrow{p(v, s, t)})$ depends on the distance $D(\cdot, \cdot)$ considered. This will be discussed at the end of the proof. By Lemmas 1 (applied $O(k^4)$ times) and 2, this overall pre-processing phase can be executed in total $O(k^4) \cdot 2^{O_\varepsilon(k)} \cdot |V|^{1+\varepsilon} = 2^{O_\varepsilon(k)} \cdot |V|^{1+\varepsilon}$ time.

Answering a query. In what follows, let $u \in V$ and $\varphi = (V_{i,j}, X, Y, x, y)$ be fixed. Applying Lemma 2, we compute $2^{O(k)}$ boxes for the points $\overrightarrow{p(v, s, t)}$ to which the vertices $v \in V_{u,\varphi} = V_{u,\varphi}^{\leq} \cup V_{u,\varphi}^{>}$ were mapped. For every such box $\mathcal{R}\langle\overrightarrow{\ell_X}, \overrightarrow{\ell_Y}\rangle$, let $v_{x,y} \in V_{i,j}$ be such that $\overrightarrow{p(v_{x,y}, \ell_x, \ell_y)} \in \mathcal{R}\langle\overrightarrow{\ell_X}, \overrightarrow{\ell_Y}\rangle$ and $f_D(\overrightarrow{p(v_{x,y}, \ell_x, \ell_y)})$ is maximized.

- If $D(\cdot, \cdot)$ is the source distance, then we set $\forall(v, s, t) \ f_D(\overrightarrow{p(v, s, t)}) = \text{dist}_G(q_s^-(v), v)$. In particular, we have: $e^D(u, V_{i,j}, X, Y, x, y) = \text{dist}_G(u, x) + \max_{\overrightarrow{\ell_X}, \overrightarrow{\ell_Y}} \max\{f_D(\overrightarrow{p(v, \ell_x, \ell_y)}) \mid \overrightarrow{p(v, \ell_x, \ell_y)} \in \mathcal{R}\langle\overrightarrow{\ell_X}, \overrightarrow{\ell_Y}\rangle\}$. Note that in this special case, Y, y are irrelevant.
- If $D(\cdot, \cdot)$ is the roundtrip distance, then we set $\forall(v, s, t) \ f_D(\overrightarrow{p(v, s, t)}) = \text{dist}_G(q_s^-(v), v) + \text{dist}_G(v, q_t^+(v))$. In particular, we have: $e^D(u, V_{i,j}, X, Y, x, y) = \text{dist}_G(u, x) + \text{dist}_G(y, u) + \max_{\overrightarrow{\ell_X}, \overrightarrow{\ell_Y}} \max\{f_D(\overrightarrow{p(v, \ell_x, \ell_y)}) \mid \overrightarrow{p(v, \ell_x, \ell_y)} \in \mathcal{R}\langle\overrightarrow{\ell_X}, \overrightarrow{\ell_Y}\rangle\}$.

We are done by applying Lemma 1 for max-queries. \square

Lemma 4. *Let a hub labeling with maximum label size $\leq k$ be given for a graph $G = (V, E, w)$. If $D(\cdot, \cdot)$ is either the min-distance or the max-distance, then for any $\varepsilon > 0$, after a pre-processing in $2^{O_\varepsilon(k)}|V|^{1+\varepsilon}$ time we can compute $e^D(u, V_{i,j}, X, Y, x, y)$, for any $u \in V$ and $\varphi = (V_{i,j}, X, Y, x, y)$, in $2^{O_\varepsilon(k)}|V|^\varepsilon$ time.*

Proof. We only detail the necessary modifications for the proof of Lemma 3. Let $i, j \in \{1, \dots, 2k-1\}$ such that $i+j \leq 2k$ be fixed. We create $2ij = O(k^2)$ different $(2(i+j)-1)$ -dimensional range trees, that are indexed by $(s, t, 0)$ and $(s, t, 1)$ for all possible pairs $(s, t) \in [i] \times [j]$. – In particular, we need twice more range trees than for Lemma 3. – For every $v \in V_{i,j}$ and $1 \leq s \leq i, 1 \leq t \leq j$, we insert two identical copies of $\overrightarrow{p(v, s, t)}$ in the range trees that are indexed by $(s, t, 0)$ and $(s, t, 1)$, with different values associated:

- For the index $(s, t, 0)$, $f_0(\overrightarrow{p(v, s, t)}) = \text{dist}_G(q_s^-(v), v)$;
- For the index $(s, t, 1)$, $f_1(\overrightarrow{p(v, s, t)}) = \text{dist}_G(v, q_t^+(v))$.

Let $u \in V$ and $\varphi = (V_{i,j}, X, Y, x, y)$ be fixed. Applying Lemma 2, we compute a family of $2^{O(k)}$ boxes $\mathcal{R}_0\langle\overrightarrow{\ell_X}, \overrightarrow{\ell_Y}\rangle$ for the points $\overrightarrow{p(v, s, t)}$ to which the vertices $v \in V_{u,\varphi}^{\leq}$ were mapped. In the same way, we compute a family of $2^{O(k)}$ boxes $\mathcal{R}_1\langle\overrightarrow{\ell_X}, \overrightarrow{\ell_Y}\rangle$ for the points $\overrightarrow{p(v, s, t)}$ to which the vertices $v \in V_{u,\varphi}^{>}$ were mapped. In doing so, if $D(\cdot, \cdot)$ is the min-distance then we have: $e^D(u, V_{i,j}, X, Y, x, y) = \max_{\overrightarrow{\ell_X}, \overrightarrow{\ell_Y}} \max\{\text{dist}_G(u, x) + f_0(\overrightarrow{p(v, \ell_x, \ell_y)}) \mid \overrightarrow{p(v, \ell_x, \ell_y)} \in \mathcal{R}_0\langle\overrightarrow{\ell_X}, \overrightarrow{\ell_Y}\rangle\} \cup \{\text{dist}_G(y, u) + f_1(\overrightarrow{p(v, \ell_x, \ell_y)}) \mid \overrightarrow{p(v, \ell_x, \ell_y)} \in \mathcal{R}_1\langle\overrightarrow{\ell_X}, \overrightarrow{\ell_Y}\rangle\}$. It is straightforward to adapt the above to the max-distance, *i.e.*, by reversing the respective roles of \mathcal{R}_0 and \mathcal{R}_1 . \square

Lemmas 3 and 4 complete the proof of Theorem 2 for eccentricity queries. For adapting our approach to distance-sum queries, the key observation is that, for any fixed $u \in V$, the sets $V_{u,\varphi}^{\leq}$ and $V_{u,\varphi}^{>}$, over all possible tuples $\varphi = (V_{i,j}, X, Y, x, y)$, form a partition of V . In particular, for any distance $D(\cdot, \cdot)$ we have: $s^D(u) = \sum_{\varphi} \left[\sum_{v \in V_{u,\varphi}^{\leq}} D(u, v) + \sum_{v \in V_{u,\varphi}^{>}} D(u, v) \right]$. Then, for a fixed φ , applying Lemma 2 we compute a family of $2^{O(k)}$ boxes $\mathcal{R}(\vec{\ell}_X, \vec{\ell}_Y)$ for the points $\overrightarrow{p(v, s, t)}$ to which the vertices $v \in V_{u,\varphi}^{\leq}$ were mapped. We have that $\{v \in V \mid \overrightarrow{p(v, \ell_x, \ell_y)} \in \mathcal{R}(\vec{\ell}_X, \vec{\ell}_Y)\}$, for all above boxes $\mathcal{R}(\vec{\ell}_X, \vec{\ell}_Y)$, is a partition of $V_{u,\varphi}^{\leq}$. Furthermore, we defined in Lemmas 3 and 4 some value D_u and function $f_D(\cdot)$ so that $\forall v \in V_{u,\varphi}^{\leq}, \overrightarrow{p(v, \ell_x, \ell_y)} \in \mathcal{R}(\vec{\ell}_X, \vec{\ell}_Y) \implies D(u, v) = D_u + f_D(\overrightarrow{p(v, \ell_x, \ell_y)})$. As a result: $\sum_{v \in V_{u,\varphi}^{\leq}} D(u, v) = D_u \cdot |V_{u,\varphi}^{\leq}| + \sum_{(\vec{\ell}_X, \vec{\ell}_Y)} \sum \{f_D(\overrightarrow{p(v, \ell_x, \ell_y)}) \mid \overrightarrow{p(v, \ell_x, \ell_y)} \in \mathcal{R}(\vec{\ell}_X, \vec{\ell}_Y)\}$. The latter computation reduces to a count-query and a sum-query for each box considered. We proceed similarly for the vertices $v \in V_{u,\varphi}^{>}$. \square

3 Applications

3.1 Bounded-treewidth graphs

A *tree-decomposition* for an undirected graph G is a pair (T, \mathcal{X}) , where T is a tree and $\mathcal{X} = (X_t)_{t \in V(T)}$ is a collection of subsets of $V(G)$ satisfying the following two properties:

- for every vertex v of G , $\{t \in V(T) \mid v \in X_t\}$ induces a nonempty subtree of T ;
- for every edge uv of G , there exists a $t \in V(T)$ s.t. $u, v \in X_t$.

The *width* of a tree-decomposition is equal to $\max_{t \in V(T)} |X_t| - 1$. The *treewidth* of an undirected graph G is the minimum width over its tree-decompositions. Finally, the treewidth of a directed graph G is the treewidth of its underlying graph (obtained by removing the arcs orientation)³. We refer to [9] for a compendium of many algorithmic applications of bounded-treewidth graphs, and to [34] for more recent such applications in the field of Fine-Grained complexity in P. Some real-world graphs have bounded-treewidth, *e.g.*, the control-flow graphs of well-structured programs in C and various other programming languages [59]. Eccentricity and distance-sum computations for bounded-treewidth graphs have been considered in [1, 14, 17]. In particular, with the notable exception of roundtrip distance, for all other distance functions considered in this article, on any directed graph of order n and treewidth at most k we can compute the eccentricity and the distance-sum of all the vertices in total $2^{O_\varepsilon(k)} n^{1+\varepsilon}$ time for any $\varepsilon > 0$. However, if one is interested in the roundtrip distance, then the algorithms proposed in [1] – which, to the best of our knowledge, were the best ones known for these two problems before our work – run in $2^{O_\varepsilon(k^2)} n^{1+\varepsilon}$ time. We replace this quadratic dependency in the treewidth by a linear one, that is optimal under SETH.

Proof of Corollary 1. We start by reminding the reader of the overall strategy in [1]. Let G be a directed graph of order n and treewidth at most k .

1. We compute a tree-decomposition of width $O(k)$. It can be done in $2^{O(k)} n$ time [10].

³There also exist directed variants of treewidth [42]. We do *not* address these variants in the paper.

2. We find a node $t \in V(T)$ such that $C = X_t$ is a balanced separator. *I.e.*, we can partition V into A, B, C s.t. $\max\{|A|, |B|\} \leq 2n/3$ and all the paths between A and B intersect C . It can be computed in linear time if a tree-decomposition is given.
3. Then, for every $a \in A$, we compute $e_A(a) = \max\{\text{dist}_G(a, b) + \text{dist}_G(b, a) \mid b \in B\}$ and $s_A(a) = \sum\{\text{dist}_G(a, b) + \text{dist}_G(b, a) \mid b \in B\}$. We proceed similarly for every vertex $b \in B$ (*i.e.*, by reversing the roles of A and B).
4. Add all possible arcs between the vertices in C , so that for every $c, c' \in C$ the arc cc' has weight equal to $\text{dist}_G(c, c')$. In doing so, we get a supergraph G' . We recurse on $G'[A \cup C]$ and $G'[B \cup C]$ separately.

If Step 3 can be executed in $T(n, k)$ time, then the total running time of the algorithm is in $O(\max\{2^{O(k)}n, T(n, k)\} \log n)$ time [1]. We prove in what follows that we can solve Step 3 in $2^{O_\varepsilon(k)}n^{1+\varepsilon}$ time for any $\varepsilon > 0$. For that, for every vertex v , let us define $L^-(v) = (\text{dist}_G(c, v))_{c \in C}$ and $L^+(v) = (\text{dist}_G(v, c))_{c \in C}$. These labels can be computed in total $k^2n^{1+o(1)}$ time by executing Dijkstra algorithm backward and forward from every vertex of C . We may not get a hub labeling, however we have the following weaker property: $\forall a \in A, b \in B, \text{dist}_G(a, b) = \min\{\text{dist}_G(a, c) + \text{dist}_G(c, b) \mid c \in C\}$, and in the same way $\text{dist}_G(b, a) = \min\{\text{dist}_G(b, c) + \text{dist}_G(c, a) \mid c \in C\}$. We are left solving a bi-chromatic version of the eccentricity and distance-sum queries studied in Sec. 2. Up to doubling the number of range trees we use – in order to store the points corresponding to vertices in A and B separately –, and modifying the range queries of our framework accordingly, our techniques for Theorem 2 can still be applied to the bi-chromatic versions of these queries. \square

3.2 Topological indices

The Wiener index has attracted attention in recent “fine-grained” analysis of polynomial-time solvable problems. Indeed, the best-known algorithms for diameter computation on bounded-treewidth graphs and planar graphs can also be applied to the computation of the Wiener index on these graph classes [14, 15]. In this section, we extend this line of work to a broader family of topological indices – with practical applications to chemistry. Our list excludes important such indices, *e.g.*, the so-called Randić index [49]. This is because the latter is linear-time computable on any graph. By contrast, the naive algorithms for computing all the indices below run in cubic time. We chose to study these specific indices in part because they illustrate some interesting applications of the orthogonal range query framework to distance problems on graphs.

In what follows, *we only consider the source distance, for undirected unweighted graphs*. This is because most chemical graphs are undirected. Given a graph $G = (V, E)$ and a hub labeling of maximum label size $\leq k$, we may assume $L^-(v) = L^+(v) = L(v)$ for all vertices v . For every $1 \leq i \leq k$, let $V_i = \{v \in V \mid |L(v)| = i\}$. The following result simplifies Lemma 2 for undirected graphs:

Lemma 5. *Let a hub labeling with maximum label size $\leq k$ be given for a graph $G = (V, E)$. In $O(2^{O(k)}|V|)$ time, we can map every vertex $v \in V$ to an $(2|L(v)| - 1)$ -dimensional point-set $\{\overrightarrow{p(v, s)} \mid 1 \leq s \leq |L(v)|\}$. Moreover, for every $u \in V$ and $\varphi = (V_i, X, x)$, let $V_{u, \varphi}$ be the set of all vertices $v \in V_i$ s.t.: $L(u) \cap L(v) = X$, and x is the least vertex of X such that $\text{dist}(u, v) = \text{dist}(u, x) + \text{dist}(x, v)$. In $2^{O(k)}$ time, we can compute a family of $2^{O(k)}$ boxes $\mathcal{R}(\overrightarrow{\ell_X})$ where $\overrightarrow{\ell_X} = (\ell_{x^*})_{x^* \in X} \subseteq [i+1]^X$ and:*

$$v \in V_{u, \varphi} \iff \exists \overrightarrow{\ell_X} \text{ s.t. } \overrightarrow{p(v, \ell_x)} \in \mathcal{R}(\overrightarrow{\ell_X}).$$

If $v \in V_{u,\varphi}$, then there exists a unique $\overrightarrow{\ell_X}$ s.t. $\overrightarrow{p(v, \ell_x)} \in \mathcal{R}(\overrightarrow{\ell_X})$.

Note in particular that for every $1 \leq i \leq k$ and $v \in V_i$, all points $\overrightarrow{p(v, s)}$ (as defined above) start with a common i -dimensional prefix $\overrightarrow{q(v)}$, a.k.a., the ordered list of all the vertices in $L(v)$.

Finally, recall that in any hub labeling of an n -vertex m -edge graph, there must be a label of size at least $\Omega(m/n)$. It implies that if a graph $G = (V, E)$ admits a hub labeling with maximum label size $\leq k$, then we have $|E| = O(k \cdot |V|)$. We often use this observation in order to remove the dependency on $|E|$ in our running times.

Wiener index and its relatives. Let us recall that the Wiener index of a graph G is equal to $\sum_{u \neq v} \text{dist}_G(u, v)$. The following result is a direct consequence of Theorem 2 (and in particular, of the distance-sum variant of Lemma 3):

Corollary 2. *For every graph $G = (V, E)$, if we are given a hub labeling with maximum label size $\leq k$, then we can compute its Wiener index in $2^{O_\varepsilon(k)} \cdot |V|^{1+\varepsilon}$ time for any $\varepsilon > 0$.*

Several other indices can be computed using similar methods as for the Wiener index. For instance, the *Hyper-Wiener* index of G is equal to $\frac{1}{2} \sum_{u \neq v} [\text{dist}_G(u, v) + \text{dist}_G(u, v)^2]$ [50]. By Corollary 2, we can reduce the computation of the latter to $\sum_{u \neq v} \text{dist}_G(u, v)^2$.

Lemma 6. *For every graph $G = (V, E)$ and integer $\alpha \geq 1$, if we are given a hub labeling with maximum label size $\leq k$, then we can compute $\sum_{u \neq v} \text{dist}_G(u, v)^\alpha$ in $O(\alpha 2^{O_\varepsilon(k)} \cdot |V|^{1+\varepsilon})$ time for any $\varepsilon > 0$.*

Proof. We store the points $\overrightarrow{p(v, s)}$, for every $1 \leq i \leq k$, $v \in V_i$ and $1 \leq s \leq i$ (as defined in Lemma 5), in a family of $O(k^2)$ range trees for point-sets of various dimensions in $O(k)$. Each such tree is indexed by a pair (i, s) . Then, let $u \in V$ be fixed. We have $\sum_v \text{dist}(u, v)^\alpha = \sum_{\varphi=(V_i, X, x)} \sum_{v \in V_{u, \varphi}} \text{dist}(u, v)^\alpha$. Therefore, in what follows, let $\varphi = (V_i, X, x)$ be fixed. Applying Lemma 5, we compute a family of $2^{O(k)}$ rectangles $\mathcal{R}(\overrightarrow{\ell_X})$ s.t. $\{v \in V \mid \overrightarrow{p(v, \ell_x)} \in \mathcal{R}(\overrightarrow{\ell_X})\}$ partitions $V_{u, \varphi}$. Furthermore, for every $v \in V_{u, \varphi}$, we have:

$$\text{dist}_G(u, v)^\alpha = \sum_{t=0}^{\alpha} \binom{\alpha}{t} \text{dist}_G(u, x)^t \text{dist}_G(x, v)^{\alpha-t}.$$

We define $\alpha+1$ functions f_t over the point-sets, $0 \leq t \leq \alpha$, so that $f_t(\overrightarrow{p(v, s)}) = \text{dist}_G(q_s(v), v)^t$. Doing so, we obtain: $\sum_{v \in V_{u, \varphi}} \text{dist}(u, v)^\alpha = \sum_{\overrightarrow{\ell_X}} \sum_{t=0}^{\alpha} \left[\binom{\alpha}{t} \text{dist}_G(u, x)^t \right] \cdot \sum \{f_{\alpha-t}(\overrightarrow{p(v, \ell_x)}) \mid \overrightarrow{p(v, \ell_x)} \in \mathcal{R}(\overrightarrow{\ell_X})\}$. As a result, $\sum_{v \in V_{u, \varphi}} \text{dist}(u, v)^\alpha$ can be derived from $2^{O(k)}$ sum-queries, applied $\alpha+1$ times to different families of range trees. We stress that all these families of range trees contain the same point-sets, but with different values, chosen according to f_t . \square

Corollary 3. *For every graph $G = (V, E)$, if we are given a hub labeling with maximum label size $\leq k$, then we can compute its Hyper-Wiener index in $2^{O_\varepsilon(k)} \cdot |V|^{1+\varepsilon}$ time for any $\varepsilon > 0$.*

The Schultz molecular topological index of G (MTI) is equal to $\sum_{u \neq v} (d(u) + d(v)) \cdot \text{dist}_G(u, v)$, where $d(u)$ and $d(v)$ are the degrees of vertices u and v , respectively [55].

Lemma 7. *For every graph $G = (V, E)$, if we are given a hub labeling with maximum label size $\leq k$, then we can compute its MTI in $2^{O_\varepsilon(k)} \cdot |V|^{1+\varepsilon}$ time for any $\varepsilon > 0$.*

Proof. Let $u \in V$ and $\varphi = (V_i, X, x)$ be fixed. We have:

$$\begin{aligned}
& \sum_{v \in V_{u,\varphi}} (d(u) + d(v)) \cdot \text{dist}(u, v) \\
&= \sum_{v \in V_{u,\varphi}} (d(u) + d(v)) \cdot [\text{dist}(u, x) + \text{dist}(x, v)] \\
&= d(u) \text{dist}(u, x) \cdot |V_{u,\varphi}| + d(u) \cdot \sum_{v \in V_{u,\varphi}} \text{dist}(x, v) \\
&\quad + d(u, x) \cdot \sum_{v \in V_{u,\varphi}} d(v) + \sum_{v \in V_{u,\varphi}} d(v) \text{dist}(x, v).
\end{aligned}$$

Let us define four functions over the point-sets so that: $f_0(\overrightarrow{p(v, s)}) = 1$ (counting), $f_1(\overrightarrow{p(v, s)}) = \text{dist}_G(q_s(v), v)$, $f_2(\overrightarrow{p(v, s)}) = d(v)$, $f_3(\overrightarrow{p(v, s)}) = d(v) \text{dist}_G(q_s(v), v)$. These functions require the knowledge of the degree sequence of the graph, that can be computed in $O(|V| + |E|) = O(k \cdot |V|)$ time. In order to compute the MTI of G , we can apply the same multi-valued range query framework as for Lemma 6 w.r.t. the functions f_0, f_1, f_2, f_3 . \square

Harary index and beyond. Another important topological index is the Harary index, defined for any graph G as $\sum_{u \neq v} \frac{1}{\text{dist}(u, v)}$ [48]. The Reciprocal complementary Wiener index of a graph G is equal to $\sum_{u \neq v} \frac{1}{\text{diam}(G) + 1 - \text{dist}(u, v)}$, where $\text{diam}(G)$ is the diameter [40]. In a recent paper [16], Cabello exploited an efficient batched evaluation of rational functions from Aronov and Katz [7], that he applied to the computation of the Harary index on bounded-treewidth graphs. The following Lemma 8 and Lemma 9 hold in the RealRAM model (see [16] for a more detailed discussion about the minimal assumptions needed for the computational model). Namely, as in Sec. 2.1, let V be a set of k -dimensional points, where each point $\vec{p} = (p_1, p_2, \dots, p_k) \in V$ is assigned some value $f(\vec{p})$. An *inverse shifted-weight query* (ISW) is given by a pair (\mathcal{R}, δ) , for some box \mathcal{R} and non-negative value δ . It outputs $\sum_{\vec{p} \in V \cap \mathcal{R}} \frac{1}{\delta + f(\vec{p})}$.

Lemma 8 ([16]). *Let V be a set of n points in k dimensions, and assume that each point \vec{p} has a positive value $f(\vec{p})$. Consider ISW queries specified by $(\mathcal{R}_1, \delta_1), (\mathcal{R}_2, \delta_2), \dots, (\mathcal{R}_m, \delta_m)$. These values can be computed in total $O(2^{O_\varepsilon(k)}(n + m)^{1+\varepsilon})$ time for any $\varepsilon > 0$.*

Lemma 9. *For every graph $G = (V, E)$, if we are given a hub labeling with maximum label size $\leq k$, then we can compute both its Harary index and its Reciprocal complementary Wiener index in $2^{O_\varepsilon(k)} \cdot |V|^{1+\varepsilon}$ time for any $\varepsilon > 0$.*

Proof. Let us first discuss the case of the Harary index. For every $v \in V$ and $1 \leq s \leq |L(v)|$, we augment the point $\overrightarrow{p(v, s)}$ (as defined in Lemma 5) with one more coordinate, set equal to v . In particular, we map v to an $2|L(v)|$ -dimensional point-set. We create $O(k^2)$ range trees of various dimensions in $O(k)$, that are indexed by (i, s) for every $1 \leq i \leq k$ and $1 \leq s \leq i$. For every such pair (i, s) , the corresponding range tree is $2i$ -dimensional; we insert in it all points $\overrightarrow{p(v, s)}$ s.t. $v \in V_i$ and $q_s(v) \neq v$ (the s^{th} vertex of $L(v)$ is different from v itself). This is because we want to assign

to this point the value $f(\overrightarrow{p(v, s)}) = \text{dist}(q_s(v), v)$, that needs to be positive so that we can apply Lemma 8 later in the proof.

Then, let $u \in V$ be fixed. We consider all possible tuples $\varphi = (V_i, X, x)$. In particular:

$$\begin{aligned} \sum_{u \neq v} \frac{1}{\text{dist}(u, v)} &= \sum_{\varphi} \sum_{v \in V_{u, \varphi} \setminus \{u\}} \frac{1}{\text{dist}(u, v)} = \sum_{x \in L(u) \setminus \{u\}} \frac{1}{\text{dist}(u, x)} \\ &+ \sum_{\varphi} \sum_{v \in V_{u, \varphi} \setminus (L(u) \cup \{u\})} \frac{1}{\text{dist}(u, v)}. \end{aligned}$$

Obviously, $\sum_{x \in L(u) \setminus \{u\}} \frac{1}{\text{dist}(u, x)}$ can be computed in $O(k)$ time from the hub label. Furthermore, for any fixed $\varphi = (V_i, X, x)$, applying Lemma 5, we compute a family of $2^{O(k)}$ boxes $\mathcal{R}(\overrightarrow{\ell_X})$ s.t. $\{v \in V \mid \overrightarrow{p(v, \ell_x)} \in \mathcal{R}(\overrightarrow{\ell_X})\}$ partitions $V_{u, \varphi}$. However, in order to avoid over-counting, we need to exclude all vertices of $L(u) \cup \{u\}$ from the latter set. This was already done implicitly for vertex x , since having $\text{dist}(x, x) = 0$ prevented the corresponding points to be inserted in the range trees. Nevertheless, we may also have $\text{dist}(u, x') = \text{dist}(u, x) + \text{dist}(x, x')$ for some $x' \in L(u) \setminus \{x\}$. Let $\tilde{L}(u) = L(u) \cup \{u\}$. We replace each box $\mathcal{R}(\overrightarrow{\ell_X})$ by $|\tilde{L}(u)| + 1$ boxes, that are obtained by imposing further constraints on the last coordinate. Specifically, let $q_0(u) = -\infty$, $q_{|\tilde{L}(u)|+1}(u) = +\infty$, and for every $1 \leq s \leq |\tilde{L}(u)|$, let $q_s(u)$ be the s^{th} vertex of $\tilde{L}(u)$. For every $0 \leq t \leq |\tilde{L}(u)|$, the box $\mathcal{R}_t(\overrightarrow{\ell_X})$ further imposes $p_{2i}(v, \ell_x) \in (q_t(v), q_{t+1}(v))$. In doing so, we can now partition $V_{u, \varphi} \setminus (L(u) \cup \{u\})$. Let $\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_m$ be the resulting family of boxes, for some $m = O(k2^{O(k)}) = 2^{O(k)}$. In order to compute $\sum_{v \in V_{u, \varphi} \setminus (L(u) \cup \{u\})} \frac{1}{\text{dist}(u, v)} = \sum_{v \in V_{u, \varphi} \setminus (L(u) \cup \{u\})} \frac{1}{\text{dist}(u, x) + \text{dist}(x, v)}$, it suffices to solve the ISW queries $(\mathcal{R}_1, \text{dist}(u, x)), (\mathcal{R}_2, \text{dist}(u, x)), \dots, (\mathcal{R}_m, \text{dist}(u, x))$. Overall, we define $2^{O(k)}$ ISW queries per vertex. By Lemma 8, we can solve all queries in total $2^{O_\varepsilon(k)}|V|^{1+\varepsilon}$ time.

Finally, we discuss the changes to bring to the above strategy so that we can also compute the Reciprocal complementary Wiener index. First, we need to compute the diameter, which by Theorem 2 can be done in $2^{O_\varepsilon(k)}|V|^{1+\varepsilon}$ time. The value of a point $\overrightarrow{p(v, s)}$ is now defined as $f(\overrightarrow{p(v, s)}) = -\text{dist}(q_s(v), v)$. Furthermore, given a vertex $u \in V$ and a tuple $\varphi = (V_i, X, x)$, we specify as ISW queries $(\mathcal{R}_1, \text{diam}(G) + 1 - \text{dist}(u, x)), (\mathcal{R}_2, \text{diam}(G) + 1 - \text{dist}(u, x)), \dots, (\mathcal{R}_m, \text{diam}(G) + 1 - \text{dist}(u, x))$, where $\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_m$ is the same family of $m = 2^{O(k)}$ hyper-rectangles as the one described earlier in the proof for computing the Harary index. We *cannot* apply Lemma 8 directly because we assigned negative values to the points $\overrightarrow{p(v, s)}$. However, a careful analysis of the proof of Lemma 8 in [16] shows that we only need to ensure that we never evaluate a rational function at a singularity (which is always the case under the sufficient conditions stated in the lemma). In the case of the Reciprocal complementary Wiener index, the rational functions to be evaluated are of the form $\rho \rightarrow \frac{1}{\rho - \text{dist}(x, v)}$. They are evaluated at $\rho = \text{diam}(G) + 1 - \text{dist}(u, x)$ for some vertices u s.t. $\text{dist}(u, v) = \text{dist}(u, x) + \text{dist}(x, v)$. In particular, since we always have $\text{diam}(G) + 1 - \text{dist}(u, v) > 0$, we never need to evaluate these rational functions at a singularity. \square

Padmakar-Ivan index and Szeged index. For the last two indices which we study in this section, we need a few more terminology. For every two vertices u and v in G , we denote n_{uv} the number of vertices that are closer to u than to v . The Szeged index of G is defined as $\sum_{uv \in E} n_{uv}n_{vu}$ [39]. Similarly, let n_{uv}^e be the number of edges that are closer to u than to v (where the distance between

an edge $e = xy$ and a vertex u is defined as $\min\{\text{dist}(x, u), \text{dist}(y, u)\}$. The Padmakar-Ivan index of G is defined as $\sum_{uv \in E} [n_{uv}^e + n_{vu}^e]$ [43].

Lemma 10. *Let $\varepsilon > 0$ be arbitrary. For every graph $G = (V, E)$, if we are given a hub labeling with maximum label size $\leq k$, then after a pre-processing in $2^{O_\varepsilon(k)}|V|^{1+\varepsilon}$ time, for any vertices u and v we can compute the value n_{uv} in $2^{O_\varepsilon(k)}|V|^\varepsilon$ time.*

Proof. Consider the point-sets defined in Lemma 5. For every $v \in V$ and $1 \leq s, t \leq |L(v)|$, we define a point $\overrightarrow{p(v, s, t)}$ which is obtained by concatenating $\overrightarrow{p(v, s)}$ with $\overrightarrow{p(v, t)}$. Observe that all these points have a dimension in $O(|L(v)|) = O(k)$. If $v \in V_i$, then we store the point $\overrightarrow{p(v, s, t)}$ in some range tree indexed by (i, s, t) . Overall, there are $O(k^3)$ range trees.

Answering a query. In what follows, let $u, v \in V$ be fixed. For every $i \in \{1, \dots, k\}$, $X \subseteq L(u), Y \subseteq L(v)$ and $(x, y) \in X \times Y$, we get a tuple $\varphi = (V_i, X, Y, x, y)$ and we define:

$$\begin{aligned} V_{u,v,\varphi} = \{z \in V_i \mid & L(u) \cap L(z) = X, L(v) \cap L(z) = Y, \\ & x \text{ is the least vertex of } X \text{ such that } \text{dist}(u, z) = \text{dist}(u, x) + \text{dist}(x, z), \\ & y \text{ is the least vertex of } Y \text{ such that } \text{dist}(v, z) = \text{dist}(v, y) + \text{dist}(y, z)\}. \end{aligned}$$

Let $n_{uv}(\varphi)$ be the number of vertices in $V_{u,v,\varphi}$ that are closer to u than to v . In particular, $n_{uv} = \sum_{\varphi} n_{uv}(\varphi)$. Therefore, we only need to explain how to compute $n_{uv}(\varphi)$ for any fixed tuple $\varphi = (V_i, X, Y, x, y)$. For that, we first adapt the proof of Lemma 2 in order to compute a family of $2^{O(k)}$ boxes $\mathcal{R}(\overrightarrow{\ell_X}, \overrightarrow{\ell_Y})$ with the following properties: $z \in V_{u,v,\varphi} \iff \exists(\overrightarrow{\ell_X}, \overrightarrow{\ell_Y})$ s.t. $\overrightarrow{p(z, \ell_x, \ell_y)} \in \mathcal{R}(\overrightarrow{\ell_X}, \overrightarrow{\ell_Y})$; furthermore, if $z \in V_{u,v,\varphi}$ then there is a *unique* $(\overrightarrow{\ell_X}, \overrightarrow{\ell_Y})$ s.t. $\overrightarrow{p(z, \ell_x, \ell_y)} \in \mathcal{R}(\overrightarrow{\ell_X}, \overrightarrow{\ell_Y})$.

Let $r_\varphi(u, v) = \text{dist}(v, y) - \text{dist}(u, x)$. For every $z \in V_{u,\varphi}$ we have:

$$\begin{aligned} \text{dist}(u, z) < \text{dist}(v, z) & \iff \text{dist}(u, x) + \text{dist}(x, z) < \text{dist}(v, y) + \text{dist}(y, z) \\ & \iff \text{dist}(x, z) - \text{dist}(y, z) < [\text{dist}(v, y) - \text{dist}(u, x)] \\ & \iff \text{dist}(q_{\ell_x}(z), z) - \text{dist}(q_{\ell_y}(z), z) < r_\varphi(u, v), \end{aligned}$$

which can be verified by modifying correspondingly the range constraints defining $\mathcal{R}(\overrightarrow{\ell_X}, \overrightarrow{\ell_Y})$. We are left performing sum-queries over the family of all modified boxes. \square

Finally, we show that our above strategy for computing the values n_{uv} can be generalized to the computation of the values n_{uv}^e .

Lemma 11. *Let $\varepsilon > 0$ be arbitrary. For every graph $G = (V, E)$, if we are given a hub labeling with maximum label size $\leq k$, then after a pre-processing in $2^{O_\varepsilon(k)}|V|^{1+\varepsilon}$ time, for any vertices u and v we can compute the value n_{uv}^e in $2^{O_\varepsilon(k)}|V|^\varepsilon$ time.*

Proof. We modify the framework of Lemma 10 so that it applies to edges rather than to vertices. Since $|E| = O(k \cdot |V|)$, this increases the runtimes of Lemma 10 by a factor at most k , that is absorbed by the $2^{O_\varepsilon(k)}$ notation. For that, recall that we totally ordered the vertex-set V . In particular, every edge e has a canonical representation uv , where $u < v$. For every $1 \leq i, j \leq k$, let $E_{i,j} = \{e = uv \in E \mid u \in V_i, v \in V_j\}$. If $e = uv \in E_{i,j}$ then, let $(\overrightarrow{p(u, s, t)})_{1 \leq s, t \leq i}$ and $(\overrightarrow{p(v, s', t')})_{1 \leq s', t' \leq j}$ be the points corresponding to u and v in the proof of Lemma 10; we define a family of $(ij)^2 = O(k^4)$ points $\overrightarrow{p(e, s, t, s', t')}$ that are obtained from the concatenation of $\overrightarrow{p(u, s, t)}$

with $\overrightarrow{p(v, s', t')}$ for every $1 \leq s, t \leq i$ and $1 \leq s', t' \leq j$. In doing so, we can adapt the proof of Lemma 10 in order to compute, for every $u, v \in V$ and $\varphi = (V_i, V_j, X, X', Y, Y', x, x', y, y')$ the value:

$$n_{uv}^e(\varphi) = \sum \{e = zz' \in E_{i,j} \mid \begin{aligned} &L(u) \cap L(z) = X, \quad L(v) \cap L(z) = Y, \\ &L(u) \cap L(z') = X', \quad L(v) \cap L(z') = Y', \\ &x \text{ is the least vertex of } X \text{ such that } \text{dist}(u, z) = \text{dist}(u, x) + \text{dist}(x, z) \\ &y \text{ is the least vertex of } Y \text{ such that } \text{dist}(v, z) = \text{dist}(v, y) + \text{dist}(y, z) \\ &x' \text{ is the least vertex of } X' \text{ such that } \text{dist}(u, z') = \text{dist}(u, x') + \text{dist}(x', z') \\ &y' \text{ is the least vertex of } Y' \text{ such that } \text{dist}(v, z') = \text{dist}(v, y') + \text{dist}(y', z') \\ &\text{dist}(u, e) < \text{dist}(v, e) \end{aligned} \}$$

by reducing to $2^{O(k)}$ sum-queries. We are done, as for any u and v we have $n_{uv}^e = \sum_{\varphi} n_{uv}^e(\varphi)$, and there are $2^{O(k)}$ tuples φ to be considered. \square

The following result now follows from Lemmas 10 and 11.

Corollary 4. *For every graph $G = (V, E)$, if we are given a hub labeling with maximum label size $\leq k$, then we can compute its Padmakar-Ivan and Szeged indices in $2^{O_{\varepsilon}(k)} \cdot |V|^{1+\varepsilon}$ time for any $\varepsilon > 0$.*

4 Graph classes of bounded expansion

A graph H is an r -shallow minor of a graph G if it can be obtained from some subgraph of G by the contraction of pairwise disjoint subgraphs of radius at most r ; a graph family \mathcal{G} has *bounded expansion* if, for any r , every r -shallow minor for any graph in \mathcal{G} has edge density at most some constant $d_r(\mathcal{G})$ [47]. A remarkable property of graph classes of bounded expansion, which we use in this section, is the existence of low tree-depth decompositions.

Specifically, an undirected graph $G = (V, E)$ has tree-depth at most k if there exists a rooted tree T of height $\leq k$ such that $V(T) = V$, and if $u, v \in V$ are adjacent in G then they have an ancestor-descendent relationship in T [47]. A directed graph has tree-depth at most k if its underlying graph (obtained by removing the orientation of its arcs) has tree-depth at most k .

Lemma 12 ([51]). *For every fixed k and n -vertex graph G , we can decide whether the tree-depth of G is at most k , and if so compute a corresponding rooted tree T , in $2^{O(k^2)}n$ time.*

Lemma 13 ([41]). *For every fixed k and n -vertex graph G , if G has tree-depth at most k and a corresponding rooted tree T is given, then we can compute hub labels of size at most k in $O(k^2n)$ time.*

A *low tree-depth decomposition* with parameter k for a graph G is a vertex-coloring such that the union of any $i \leq k$ color classes induces a subgraph of tree-depth at most i [47]. We denote by $\chi_k(G)$ the minimum number of colours in a low tree-depth decomposition with parameter k for G (for $k = 1$, this is exactly the chromatic number of G).

Lemma 14 ([47]). *A graph class \mathcal{G} has bounded expansion if and only if $\forall k, \limsup\{\chi_k(G) \mid G \in \mathcal{G}\} < \infty$.*

Before presenting our proof of Theorem 3, what now remains to discuss is the computational cost for computing a low tree-depth decomposition for graph classes of bounded expansion. For that, for any graph G , let $\nabla_r(G)$ be the maximum density of an r -shallow minor of G . The following result is a straightforward consequence of the definition of graph classes of bounded expansion.

Lemma 15 ([47]). *A graph class \mathcal{G} has bounded expansion if and only if $\forall k, \limsup\{\nabla_k(G) \mid G \in \mathcal{G}\} < \infty$.*

We will use the following constructive result of Nešetřil and Ossona de Mendez for low tree-depth decompositions:

Lemma 16 ([47]). *For every graph G and integer k , there exists a polynomial P_k of degree $\approx 2^{2^k}$ such that we can compute a low tree-depth decomposition with parameter k for G with $P_k(\nabla_{2^{k-2}+1}(G))$ colours, in total $O(P_k(\nabla_{2^{k-2}+1}(G)) \cdot |G|)$ time.*

We are now ready to present the main result in this section:

Proof of Theorem 3. Let $c = P_k(\nabla_{2^{k-2}+1}(G))$. By Lemma 16, a low tree-depth decomposition with parameter k for G and c colours can be computed in total $O(c \cdot n)$ time. Consider all possible unions of k color classes sequentially (there are $\binom{c}{k}$ possibilities). Every one of them induces a subgraph of tree-depth at most k . In particular, by Lemma 12, we can compute a corresponding rooted tree for this subgraph in $2^{O(k^2)}$ time. Then, by Lemma 13, we can compute a hub labeling for this subgraph, with maximum label size $\leq k$, in $O(k^2 n)$ time. By concatenating all these hub labelings, each vertex of G gets assigned a label of cardinality $\leq k \binom{c}{k}$. This resulting labeling is not necessarily a hub labeling for G , as we may not have $\text{dist}_G(u, v) = \min\{\text{dist}_G(u, x) + \text{dist}_G(x, v) \mid x \in L^+(u) \cap L^-(v)\}$ for all ordered pairs of vertices (u, v) . However, since every path of length at most k is contained in the union of at most k color classes (and so, in one of the $\binom{c}{k}$ subgraphs of bounded tree-depth on which our above algorithm iterates), if $\text{dist}_G(u, v) \leq k$, then given u and v this above labeling correctly computes $\text{dist}_G(u, v)$. In particular, if the diameter is at most k then we indeed computed a hub labeling for G . Finally, we use this labeling as input for the algorithm of Theorem 2 in order to decide in $2^{O_\varepsilon(k \binom{c}{k})} n^{1+\varepsilon}$ time whether the diameter of G is at most k .

The total running time is in $2^{O_\varepsilon(k \binom{c}{k})} n^{1+\varepsilon}$. Note that by Lemma 15, $c \approx [d_{2^{k-2}+1}(\mathcal{G})]^{2^{2^k}}$ for some constant $d_{2^{k-2}+1}(\mathcal{G})$ depending only on \mathcal{G} and k . \square

References

- [1] A. Abboud, F. Grandoni, and V. Vassilevska Williams. Subcubic equivalences between graph centrality problems, APSP and diameter. In *Proceedings of the twenty-sixth annual ACM-SIAM symposium on Discrete algorithms*, pages 1681–1697. SIAM, 2014.
- [2] A. Abboud, V. Vassilevska Williams, and J. Wang. Approximation and fixed parameter subquadratic algorithms for radius and diameter in sparse graphs. In *Proceedings of the twenty-seventh annual ACM-SIAM symposium on Discrete Algorithms (SODA)*, pages 377–391. SIAM, 2016.
- [3] I. Abraham, D. Delling, A. Fiat, A. V. Goldberg, and R. F. Werneck. Highway dimension and provably efficient shortest path algorithms. *Journal of the ACM (JACM)*, 63(5):1–26, 2016.

- [4] I. Abraham, D. Delling, A. Goldberg, and R. Werneck. A hub-based labeling algorithm for shortest paths in road networks. In *International Symposium on Experimental Algorithms*, pages 230–241. Springer, 2011.
- [5] I. Abraham, D. Delling, A. V. Goldberg, and R. F. Werneck. Hierarchical hub labelings for shortest paths. In *European Symposium on Algorithms*, pages 24–35. Springer, 2012.
- [6] T. Akiba, Y. Iwata, and Y. Yoshida. Fast exact shortest-path distance queries on large networks by pruned landmark labeling. In *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*, pages 349–360, 2013.
- [7] B. Aronov and M. Katz. Batched point location in SINR diagrams via algebraic tools. *ACM Transactions on Algorithms (TALG)*, 14(4):1–29, 2018.
- [8] M. Babenko, A. Goldberg, H. Kaplan, R. Savchenko, and M. Weller. On the complexity of hub labeling. In *International Symposium on Mathematical Foundations of Computer Science*, pages 62–74. Springer, 2015.
- [9] H. Bodlaender. Treewidth: Algorithmic techniques and results. In *International Symposium on Mathematical Foundations of Computer Science*, pages 19–36. Springer, 1997.
- [10] H. Bodlaender, P. Drange, M. Dregi, F. Fomin, D. Lokshtanov, and M. Pilipczuk. A $\tilde{O}(n^5)$ -approximation algorithm for treewidth. *SIAM Journal on Computing*, 45(2):317–378, 2016.
- [11] J. A. Bondy and U. S. R. Murty. *Graph theory*. 2008.
- [12] M. Borassi, P. Crescenzi, and M. Habib. Into the square: On the complexity of some quadratic-time solvable problems. *Electronic Notes in Theoretical Computer Science*, 322:51–67, 2016.
- [13] M. Borassi, P. Crescenzi, M. Habib, W. Kusters, A. Marino, and F. Takes. Fast diameter and radius BFS-based computation in (weakly connected) real-world graphs: With an application to the six degrees of separation games. *Theoretical Computer Science*, 586:59–80, 2015.
- [14] K. Bringmann, T. Husfeldt, and M. Magnusson. Multivariate Analysis of Orthogonal Range Searching and Graph Distances. *Algorithmica*, pages 1–24, 2020.
- [15] S. Cabello. Subquadratic algorithms for the diameter and the sum of pairwise distances in planar graphs. *ACM Transactions on Algorithms (TALG)*, 15(2):1–38, 2018.
- [16] S. Cabello. Computing the inverse geodesic length in planar graphs and graphs of bounded treewidth. *ACM Transactions on Algorithms (TALG)*, 2022. To appear.
- [17] S. Cabello and C. Knauer. Algorithms for graphs of bounded treewidth via orthogonal range searching. *Computational Geometry*, 42(9):815–824, 2009.
- [18] E. Cohen, E. Halperin, H. Kaplan, and U. Zwick. Reachability and distance queries via 2-hop labels. *SIAM Journal on Computing*, 32(5):1338–1355, 2003.
- [19] V. Cohen-Addad, S. Dahlgaard, and C. Wulff-Nilsen. Fast and compact exact distance oracle for planar graphs. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 962–973. IEEE, 2017.

- [20] D. Coudert, G. Ducoffe, and A. Popa. Fully polynomial fpt algorithms for some classes of bounded clique-width graphs. *ACM Transactions on Algorithms (TALG)*, 15(3):1–57, 2019.
- [21] B. Courcelle. The monadic second-order logic of graphs. I. Recognizable sets of finite graphs. 85(1):12–75, 1990.
- [22] P. Crescenzi, R. Grossi, M. Habib, L. Lanzi, and A. Marino. On computing the diameter of real-world undirected graphs. *Theoretical Computer Science*, 514:84–95, 2013.
- [23] D. Delling, A. Goldberg, R. Savchenko, and R. Werneck. Hub labels: Theory and practice. In *International Symposium on Experimental Algorithms*, pages 259–270. Springer, 2014.
- [24] E. Demaine, F. Reidl, P. Rossmanith, F. Villaamil, S. Sikdar, and B. Sullivan. Structural sparsity of complex networks: Bounded expansion in random models and real-world graphs. *Journal of Computer and System Sciences*, 105:199–241, 2019.
- [25] R. Diestel. *Graph Theory, 4th Edition*, volume 173 of *Graduate texts in mathematics*. Springer, 2012.
- [26] F. Dragan, M. Habib, and L. Viennot. Revisiting radius, diameter, and all eccentricity computation in graphs through certificates. Technical Report 1803.04660, arXiv, 2018.
- [27] G. Ducoffe. A New Application of Orthogonal Range Searching for Computing Giant Graph Diameters. In *SOSA ’19*, 2019.
- [28] G. Ducoffe and F. Dragan. A story of diameter, radius and Helly property. *Networks*, 77(3):435–453, 2021.
- [29] G. Ducoffe, M. Habib, and L. Viennot. Fast diameter computation within split graphs. In *International Conference on Combinatorial Optimization and Applications*, pages 155–167. Springer, 2019.
- [30] G. Ducoffe, M. Habib, and L. Viennot. Diameter computation on H -minor free graphs and graphs of bounded (distance) VC-dimension. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1905–1922. SIAM, 2020.
- [31] Z. Dvorak, D. Král, and R. Thomas. Deciding first-order properties for sparse graphs. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 133–142. IEEE, 2010.
- [32] A. Efentakis, C. Efstathiades, and D. Pfoer. COLD. Revisiting hub labels on the database for large-scale graphs. In *International Symposium on Spatial and Temporal Databases*, pages 22–39. Springer, 2015.
- [33] J. Evald and S. Dahlgaard. Tight Hardness Results for Distance and Centrality Problems in Constant Degree Graphs. Technical Report arXiv:1609.08403, ArXiv, 2016.
- [34] F. Fomin, D. Lokshantov, S. Saurabh, M. Pilipczuk, and M. Wrochna. Fully polynomial-time parameterized computations for graphs and matrices of low treewidth. *ACM Transactions on Algorithms (TALG)*, 14(3):1–45, 2018.

- [35] C. Gavaille, D. Peleg, S. Pérennes, and R. Raz. Distance labeling in graphs. *Journal of Algorithms*, 53(1):85–112, 2004.
- [36] P. Gawrychowski, H. Kaplan, S. Mozes, M. Sharir, and O. Weimann. Voronoi diagrams on planar graphs, and computing the diameter in deterministic $\tilde{O}(n^{5/3})$ time. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2018)*, pages 495–514. SIAM, 2018.
- [37] P. Gawrychowski, S. Mozes, O. Weimann, and C. 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)*, pages 515–529. SIAM, 2018.
- [38] S. Gupta, A. Kosowski, and L. Viennot. Exploiting Hopsets: Improved Distance Oracles for Graphs of Constant Highway Dimension and Beyond. In *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*, volume 132 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 143:1–143:15, 2019.
- [39] I. Gutman. A formula for the Wiener number of trees and its extension to graphs containing cycles. *Graph Theory Notes NY*, 27(9):9–15, 1994.
- [40] O. Ivanciuc, T. Ivanciuc, and A. Balaban. Quantitative structure-property relationship evaluation of structural descriptors derived from the distance and reverse Wiener matrices. *J. Mol. Des.*, 1:467–487, 2002.
- [41] Y. Iwata, T. Ogasawara, and N. Ohsaka. On the Power of Tree-Depth for Fully Polynomial FPT Algorithms. In *35th Symposium on Theoretical Aspects of Computer Science (STACS 2018)*, volume 96 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 41:1–41:14. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018.
- [42] T. Johnson, N. Robertson, P. Seymour, and R. Thomas. Directed tree-width. *Journal of Combinatorial Theory, Series B*, 82(1):138–154, 2001.
- [43] P. Khadikar, S. Karmarkar, and V. Agrawal. A novel PI index and its applications to QSPR/QSAR studies. *Journal of chemical information and computer sciences*, 41(4):934–949, 2001.
- [44] A. Kosowski, P. Uznanski, and L. Viennot. Hardness of Exact Distance Queries in Sparse Graphs Through Hub Labeling. In *Proceedings of the 2019 ACM Symposium on Principles of Distributed Computing (PODC '19)*, page 272–279. ACM, 2019.
- [45] Y. Li, L. U, M. Yiu, and N. Kou. An experimental study on hub labeling based shortest path algorithms. *Proceedings of the VLDB Endowment*, 11(4):445–457, 2017.
- [46] M. Magnusson. Orthogonal Range Searching and Graph Distances Parameterized by Treewidth. Master’s thesis, Lund University, Sweden, 2018.
- [47] J. Nešetřil and P. Ossona de Mendez. *Sparsity: graphs, structures, and algorithms*, volume 28. Springer Science & Business Media, 2012.
- [48] D. Plavšić, S. Nikolić, N. Trinajstić, and Z. Mihalić. On the Harary index for the characterization of chemical graphs. *Journal of Mathematical Chemistry*, 12(1):235–250, 1993.

- [49] M. Randić. Characterization of molecular branching. *Journal of the American Chemical Society*, 97(23):6609–6615, 1975.
- [50] M. Randić. Novel molecular descriptor for structure—property studies. *Chemical Physics Letters*, 211(4-5):478–483, 1993.
- [51] F. Reidl, P. Rossmanith, F. Villaamil, and S. Sikdar. A faster parameterized algorithm for treedepth. In *International Colloquium on Automata, Languages, and Programming*, pages 931–942. Springer, 2014.
- [52] I. Roditty, M. Thorup, and U. Zwick. Roundtrip spanners and roundtrip routing in directed graphs. *ACM Transactions on Algorithms (TALG)*, 4(3):1–17, 2008.
- [53] L. Roditty and V. Vassilevska Williams. Fast approximation algorithms for the diameter and radius of sparse graphs. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing (STOC)*, pages 515–524, 2013.
- [54] D. Rouvray and R. King. *Topology in chemistry: Discrete mathematics of molecules*. Elsevier, 2002.
- [55] H. Schultz. Topological organic chemistry. 1. Graph theory and topological indices of alkanes. *Journal of Chemical Information and Computer Sciences*, 29(3):227–228, 1989.
- [56] R. Seidel. On the all-pairs-shortest-path problem in unweighted undirected graphs. *J. Comput. Syst. Sci.*, 51(3):400–403, 1995.
- [57] J. Shun. An evaluation of parallel eccentricity estimation algorithms on undirected real-world graphs. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1095–1104, 2015.
- [58] F. Takes and W. Kusters. Computing the eccentricity distribution of large graphs. *Algorithms*, 6(1):100–118, 2013.
- [59] M. Thorup. All structured programs have small tree width and good register allocation. *Information and Computation*, 142(2):159–181, 1998.
- [60] T. Wagner. Eccentricities via Parallel Set Cover (Brief Announcement). In *The 31st ACM Symposium on Parallelism in Algorithms and Architectures*, pages 43–45, 2019.
- [61] T. Wagner. Eccentricity Heuristics through Sublinear Analysis Lenses. In *APOCS*, pages 75–89, 2020.
- [62] D. Watts and S. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684):440–442, 1998.
- [63] R. Yuster. Computing the diameter polynomially faster than APSP. Technical Report arXiv:1011.6181, arXiv, 2010.
- [64] Y. Zhang and J. Yu. Hub Labeling for Shortest Path Counting. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data, SIGMOD ’20*, page 1813–1828. ACM, 2020.