Approximate Distance Oracles with Improved Stretch for Sparse Graphs

Liam Roditty[1] and Roei Tov[1]

Bar Ilam University, Israel{liam.roditty,roei81}@gmail.com

Abstract. Thorup and Zwick [30] introduced the notion of approximate distance oracles, a data structure that stores for an n-vertices, m-edges weighted undirected graph $G = (V, E)$ distance estimations. Its size is $O(kn^{1+1/k})$ and given a pair of vertices $u, v \in V$ at distance $d(u, v)$ it returns in $O(k)$ time an estimation that is bounded by $(2k-1)d(u, v)$, i.e, a $(2k-1)$-multiplicative approximation (stretch). Thorup and Zwick [30] presented also a lower bound based on the girth conjecture of Erdős.
For sparse unweighted graphs (i.e, $m = \tilde{O}(n)$) the lower bound does not apply. Pǎtraşcu and Roditty [21] used the sparsity of the graph and obtained a distance oracle that uses $\tilde{O}(n^{5/3})$ space, has $O(1)$ query time and a stretch of 2. Pǎtraşcu et al. [22] presented infinity many distance oracles with fractional stretch factors that for graphs with $m = \tilde{O}(n)$ converge exactly to the integral stretch factors and the corresponding space bound of Thorup and Zwick.
It is not known, however, whether graph sparsity can help to get a stretch which is better than $(2k - 1)$ using only $\tilde{O}(kn^{1+1/k})$ space. In this paper we answer this open question and prove a separation between sparse and dense graphs by showing that using sparsity it is possible to obtain better stretch/space tradeoffs than those of Thorup and Zwick. We show that for every $k \geq 2$ there is a distance oracle that uses $O(kmn^{1/k} \log n)$ space and provides in $O(k)$ time an estimation $d^*(u, v)$ that satisfies $d(u, v) \leq d^*(u, v) \leq (2k-1)d(u, v)-4$, for $k > 2$, and $d(u, v) \leq d^*(u, v) \leq 3d(u, v) - 2$, for $k = 2$.
Another contribution of this paper is a refined stretch analysis of Thorup and Zwick distance oracles that allows us to obtain a better understanding of this important data structure. We present a simple quantitative measure for every vertex that allows to characterize the exact scenarios in which every query that involves this vertex produces an estimation of stretch strictly better than $2k - 1$.

Keywords: Graph algorithms · Approximate shortest paths · Approximate distance oracles.

1 Introduction

An approximate distance oracle is a succinct data structure that supports efficient approximate distance queries. Thorup and Zwick [30] showed that given an undirected graph $G = (V, E)$ with m edges and n vertices and an integer

$k \geq 1$, there is a data structure of size $O(kn^{1+1/k})$ that for every pair of vertices $u, v \in V$ returns in $O(k)$ time an estimation $\hat{d}(u,v)$ which is a $(2k-1)$ multiplicative approximation (stretch) of $d(u,v)$, that is, $d(u,v) \leq \hat{d}(u,v) \leq (2k-1)d(u,v)$, where $d(u,v)$ is the length of the shortest path between u and v in G.

Thorup and Zwick [30] presented also a conditional lower bound based on the girth conjecture of Erdős[1]. More specifically, they proved that if there is a graph of $\Omega(n^{1+1/k})$ edges whose girth is $2k+2$ then any distance oracle with stretch $t \leq 2k$ requires $\Omega(n^{1+1/k})$ bits on some input. A careful examination of the conditional lower bound proof reveals that it relies on the stretch of the estimation for vertex pairs that share an edge. Therefore, it still might be possible to obtain a stretch better than $2k-1$ using the same space, for non-adjacent vertex pairs or for sparser graphs.

In this paper we show that for every $k \geq 2$ there is a distance oracle that uses $O(kmn^{1/k}\log n)$ space and provides in $O(k)$ time an estimation $d^*(u,v)$ that satisfies $d(u,v) \leq d^*(u,v) \leq (2k-1)d(u,v) - 4$, for $k > 2$, and $d(u,v) \leq d^*(u,v) \leq 3d(u,v) - 2$, for $k = 2$. Notice that for sparse graphs with $m = \widetilde{O}(n)$ the space is the same as the space of Thorup and Zwick distance oracles (up to poly-logarithmic factors).

Pătraşcu, Roditty and Thorup [22] proved, based on a set intersection hardness conjecture, that any distance oracle with stretch strictly less than $3 - \frac{2}{\ell+1}$, for any fixed integer $\ell > 1$, and constant query time, requires $\tilde{\Omega}(n^{1+\frac{1}{2-1/\ell}})$ space. Similarly to the conditional lower bound of Thorup and Zwick, the conditional lower bound of Pătraşcu et al. is obtained by considering vertex pairs at a specific distance. More specifically, they show that for any $\ell > 1$, a distance oracle that for every pair of vertices at distance $\ell + 1$, provides in constant query time an estimation with stretch strictly below $3 - \frac{2}{\ell+1}$ requires $\tilde{\Omega}(n^{1+\frac{1}{2-1/\ell}})$ space. Notice that for $k = 2$ our distance oracle has stretch of $3 - \frac{2}{d(u,v)}$, for every $u, v \in V$ and uses $\tilde{O}(n^{1.5})$ space for sparse graphs with $m = \tilde{O}(n)$. It follows from [22] that obtaining a stretch which is strictly better than $3 - \frac{2}{d(u,v)}$ requires $\tilde{\Omega}(n^{1.5+\varepsilon})$ space, where $\varepsilon > 0$.

Pătraşcu et al. [22] showed also that there are infinitely many distance oracles for sparse graphs with fractional stretch factors. Their distance oracles converge exactly to the integral stretch factors and the corresponding space bound of Thorup-Zwick distance oracles. Our new construction implies that for space $\tilde{O}(km^{1+1/k})$ a stretch that is strictly better than the corresponding integral stretch of $2k-1$ is possible.

Moreover, in light of the girth conjecture lower bound, our new distance oracles with $O(k)$ query time, reveal a true separation between sparse and dense graphs, as the space/stretch tradeoffs for sparse graphs (e.g, $m = \tilde{O}(n)$) are strictly better than those that are possible for dense graphs (e.g, $\widetilde{\Omega}(n^{1+1/k})$).

The distance oracles of Thorup and Zwick, beside being an important data structure on their own, are also extremely useful as a tool in many applications.

[1] The girth is the length of the shortest cycle in an unweighted graph.

They were a crucial building block in several important dynamic graph algorithms along the last decade (e.g, [27, 8, 17, 18]). They also play a pivotal role in designing distance labeling and compact routing schemes as was already shown by Thorup and Zwick [29] and in subsequent works (e.g, [9, 2, 26, 25]). Distance oracles were also implemented and tested (e.g, [24, 13]) and found useful on real world graphs. Therefore, any further understanding that we gain on the basic properties of distance oracles is of great interest.

Along these lines, another contribution that we make in this paper, is a refined analysis of the stretch of Thorup and Zwick distance oracles. At the base of the distance oracles there is an hierarchy of vertex sets $A_0, A_1, \ldots, A_k$, where $A_0 = V$, $A_k = \emptyset$ and A_i is formed by picking each vertex of A_{i-1}, independently, with some probability p. For every $u \in V$ the distance $d(u, A_i)$ between u and A_i is computed and saved. We introduce a simple parameter the *average distance* which is roughly defined[2] for every $i \in [1, k-1]$ as the distance between u and A_i divided by i, that is $d(u, A_i)/i$. Our refined analysis characterizes several cases in which the stretch is strictly better than $2k-1$ using only the average distance, which can be easily computed using the current information saved with the distance oracle. Roughly speaking, if there exist $i, j \in [1, k-1]$ such that $i \neq j$ and $d(u, A_i)/i \neq d(u, A_j)/j$, then the stretch is strictly better than $2k-1$ for every distance query that includes the vertex u.

Based on similar ideas we also show that if $D(u) = \{\Delta_1, \ldots, \Delta_\ell\}$ is the set of all possible distances of $u \in V$ with other vertices in the graph then there is at most one value $\Delta \in D(u)$ for which the stretch of the distance estimation is exactly $2k-1$, that is, only for vertices v that satisfy $d(u, v) = \Delta$ it might be that $\hat{d}(u, v) = (2k-1)d(u, v)$.

1.1 Related Work.

Since their introduction by Thorup and Zwick [30] distance oracles were studied by many researchers. Chechik [12, 11], presented a $(2k-1)$-stretch distance oracle with $O(1)$ query time and $O(n^{1+1/k})$ space. (See also [32, 19].)

Pătraşcu and Roditty [21] showed a distance oracle for weighted undirected graphs with stretch 2 and size $O(n^{4/3}m^{1/3})$. For $m = o(n^2)$, this distance oracle has $o(n^2)$ size and stretch 2. Pătraşcu, Roditty and Thorup [22] showed for every integer $k \geq 0$ and $\ell > 0$ distance oracles, that use $\tilde{O}(m^{1+1/(k \pm 1/\ell)})$ space and answer distance query in $O(k+\ell)$ time with stretch $2k+1 \pm 2/\ell$. Sommer, Verbin, and Yu [28] provided a unconditional lower bound. They showed that c-approximate distance oracles with constant query time require space $n^{1+\Omega(1/c)}$.

Agarwal, Godfrey and Har-Peled [3], Agarwal and Godfrey [4] and Porat and Roditty [23], presented distance oracles for sparse graphs with low space/stretch tradeoffs, at the expense of a larger query time. For the most recent advances in this direction see [4]. Elkin and Pettie [16] presented for every $k \geq 1$, a distance oracle that uses $k^{O(\log\log k)}n^{1+1/k}$ space and in $O(n^{\varepsilon})$ time produces an estimation with a multiplicative error of $O(1)$ and additive error of $k^{O(\log\log k)}$.

[2] In the formal definition we take the ceiling of the average distance.

Spanners are closely related to distance oracles. Given a graph $G = (V, E)$ a subgraph H of G is an (α, β)-spanner of G, if for every $u, v \in V$, $d_H(u, v) \leq \alpha \cdot d_G(u, v) + \beta$, where $d_{G'}(u, v)$ is the distance between u and v in G'. Any weighted undirected graph has $(2k-1, 0)$-spanner with $O(n^{1+1/k})$ edges [5]. As in the distance oracle case, the space/stretch tradeoffs are optimal assuming the girth conjecture. Elkin and Peleg [15] were the first to bypass the girth conjecture in the context of spanners and initiated the study of (α, β)-spanners. Roughly speaking, in such spanners distant vertices have a good multiplicative approximation and close vertices have a good additive approximation. They showed that for every integer $k \geq 1$ and $\varepsilon > 0$ there is a $(1+\varepsilon, \beta)$-spanner with $O(\beta n^{1+1/k})$ edges, where β depends on k and ε but independent on n. Baswana et al. [6] presented a $(k, k-1)$-spanner with $O(kn^{1+1/k})$ edges. Parter [20] bypassed the girth conjecture for non-adjacent vertices by introducing the notion of hybrid multiplicative spanners in which the stretch for non-adjacent vertices is k and for adjacent vertices is $2k-1$. For more results in this direction see [20, 7] and references therein.

Additive spanners were studied as well. Dor, Halperin and Zwick [14] showed an additive 2-spanner with $O(n^{3/2})$ edges. Chechik [10] presented an additive 4-spanner with $\tilde{O}(n^{7/5})$ edges [10], and Baswana et al. [6] presented an additive 6-spanner with $O(n^{4/3})$ edges. Woodruff [31] showed that an additive 6-spanner with $\tilde{O}(n^{4/3})$ edges can be constructed in $\tilde{O}(n^2)$ time. Abboud and Bodwin [1] proved that any spanner with $O(n^{4/3-\varepsilon})$ edges has at least $n^{o(1)}$ additive stretch.

1.2 Paper Organization.

In the next section we present ingredients from the distance oracle of Thorup-Zwick that are required in order to present our results. In Section 3 we present our new distance oracles for sparse graphs. In Section 4 we present our refined stretch analysis for Thorup-Zwick distance oracle.

2 Recap of the Distance Oracle of Thorup and Zwick

Let $G = (V, E)$ be an n-vertices m-edges undirected unweighted graph. For every $u, v \in V$, let $d(u, v)$ be the length of the shortest path between u and v. Let $N(u)$ be the vertices that are neighbours of u and let $deg(u) = |N(u)|$. For every set $A \subseteq V$, let $p_A(u)$ be the closest vertex to u from A, where ties are broken in favor of the vertex with a smaller identifier, and let $d(u, A) = d(u, p_A(u))$. Notice that by this definition, if v is on a shortest path between u and $p_A(u)$, then $p_A(u) = p_A(v)$.

Let $B(u, \ell) = \{v \in V \mid d(u, v) < \ell\}$, and let $B(u, \ell, X) = \{v \in X \mid d(u, v) < \ell\}$, where $X \subseteq V$. In the degenerated case of $B(u, 0)$ we assume that $B(u, 0)$ contains only u. Let $E' \subseteq E$, the set $V(E')$ contains the endpoints of the edges in E'. Finally, let $L(u, \ell) = \{v \in V \mid d(u, v) = \ell\}$.

In their seminal paper Thorup and Zwick [30] showed that there is a data structure of size $O(kn^{1+1/k})$ that returns a $(2k-1)$ multiplicative approximation

Algorithm 1: $\text{dist}(u, v)$

```
i ← 0;
while p_i(u) ∉ B_i(v) do
    i ← i + 1;
    swap u and v;
return d(u, p_i(u)) + d(v, p_i(u));
```

(stretch) of the distances of an undirected weighted graph in $O(k)$ time. Their data structure is constructed as follows. Let $k \geq 1$ and let $A_0, A_1, \ldots, A_k$ be sets of vertices, such that $A_0 = V$, $A_k = \emptyset$ and A_i is formed by picking each vertex of A_{i-1} independently with probability[3] $n^{-1/k}$, for every $1 \leq i \leq k-1$. For every $u \in V$, let $p_i(u) = p_{A_i}(u)$ and $\ell_i(u) = d(u, A_i) = d(u, p_i(u))$. We set $p_0(u)$ to u, $p_k(u)$ to be null and set $\ell_k(u)$ to ∞.

For every $0 \leq i \leq k-1$, let $B_i(u) = B(u, \ell_{i+1}(u), A_i)$. The *bunch* of $u \in V$ is $B(u) = \cup_{i=0}^{k-1} B_i(u)$. Thorup and Zwick proved the following bound:

Lemma 1. *[Lemma 3.2 and Theorem 3.7 from [30]] For every $u \in V$ and $i \in [0, k-1]$ the size of $B_i(u)$ is $O(n^{1/k})$ in expectation or $O(n^{1/k} \log n)$ in the worst case if a deterministic sampling procedure is used.*

In the distance oracle we store for every vertex $u \in V$ the set $B(u)$ and the value of $d(u, v)$, for every $v \in B(u)$, in a 2-level hash table. We store also $p_i(u)$ for every $0 \leq i \leq k-1$. Thus, the total size of the distance oracle is $O(kn^{1+1/k})$ in expectation or $O(kn^{1+1/k} \log n)$ in the worst case[4].

We will now present the query algorithm of the distance oracle. Algorithm $dist(u, v)$, presented in Algorithm 1, looks for the smallest even i such that $p_i(u) \in B_i(v)$ or $p_{i+1}(v) \in B_{i+1}(u)$. Since both $p_{k-1}(u) \in B_{k-1}(v)$ and $p_{k-1}(v) \in B_{k-1}(u)$ the algorithm always stops. Let $f(u, v)$ be the largest value that i reached to during the run of $dist(u, v)$. In other words, $f(u, v)$ is the largest value such that for every even $j < f(u, v)$, it holds that $p_j(u) \notin B_j(v)$ and for every odd $j < f(u, v)$ it holds that $p_j(v) \notin B_j(u)$. Since $dist(u, v)$ always stops it follows that $f(u, v) \leq k-1$.

The following Lemma is implicit in [30]. We prove it explicitly since we use it in our proofs.

Lemma 2. *For every even $i \leq f(u, v)$ it holds that $\ell_i(u) \leq i \cdot d(u, v)$ and for every odd $i \leq f(u, v)$ it holds that $\ell_i(v) \leq i \cdot d(u, v)$.*

Proof. Let $\Delta = d(u, v)$. We prove that $\ell_i(u) \leq i \cdot \Delta$ for every even $i \leq f(u, v)$ by induction on i. For the base case let $i = 0$. Since $u = p_0(u)$ we have $\ell_0(u) = 0$ and the claim trivially holds. Assume now that the claim holds for every even integer smaller than i. We prove the claim for i.

[3] A deterministic implementation is also possible, see [30] for more details.

[4] Throughout the paper we will use the $\tilde{O}(\cdot)$ notation to hide small poly-logarithmic factors.

From the definition of $f(u, v)$ it follows that $p_{i-2}(u) \notin B_{i-2}(v)$ and this implies that $\ell_{i-1}(v) \leq d(v, p_{i-2}(u)) \leq \Delta + \ell_{i-2}(u)$, where the last inequality follows from the triangle inequality.

From the definition of $f(u, v)$ it follows also that $p_{i-1}(v) \notin B_{i-1}(u)$ and similarly to before this implies that $\ell_i(u) \leq d(u, p_{i-1}(v)) \leq \Delta + \ell_{i-1}(v)$. Therefore, we get that $\ell_i(u) \leq 2\Delta + \ell_{i-2}(u)$. From the induction hypothesis we have $\ell_{i-2}(u) \leq (i-2)\Delta$ and we get that $\ell_i(u) \leq i\Delta$.

For odd $i \leq f(u, v)$ we have from the first part that $\ell_{i-1}(u) \leq (i-1) \cdot \Delta$. Since $i \leq f(u, v)$ we have $p_{i-1}(u) \notin B_{i-1}(v)$, thus, $\ell_i(v) \leq d(v, p_{i-1}(u)) \leq \Delta + \ell_{i-1}(u) \leq i\Delta$, as required.

We proceed with the following useful observation on Thorup-Zwick distance oracle. Consider the set A_{i-j}, where i and j are even and $0 \leq j < i \leq f(u, v)$. From Lemma 2 it follows that $\ell_{i-j}(u) \leq (i-j) \cdot d(u, v)$ and $\ell_i(u) \leq i \cdot d(u, v)$. But what if we have a bound for $\ell_{i-j}(u)$ that is better than $(i-j) \cdot d(u, v)$, can we use it to obtain a better bound for $\ell_i(u)$?

In the next Lemma we generalize Lemma 2 and show that this is indeed possible.

Lemma 3. *For every even $i \leq f(u, v)$:*

(i) $\ell_i(u) \leq \ell_{i-j}(u) + j \cdot d(u, v)$, *for every even* $j \leq i$.
(ii) $\ell_i(u) \leq \ell_{i-j}(v) + j \cdot d(u, v)$, *for every odd* $j \leq i$.

For every odd $i \leq f(u, v)$:

(i) $\ell_i(v) \leq \ell_{i-j}(u) + j \cdot d(u, v)$, *for every even* $j \leq i$.
(ii) $\ell_i(v) \leq \ell_{i-j}(v) + j \cdot d(u, v)$, *for every odd* $j \leq i$.

Proof. We prove the claim for an even $i \leq f(u, v)$. The proof for an odd $i \leq f(u, v)$ is symmetric.

We prove (ii) by induction on j. The base case is when $j = 1$. We need to show that $\ell_i(u) \leq \ell_{i-1}(v) + d(u, v)$. Since $i \leq f(u, v)$ we have $p_{i-1}(v) \notin B_{i-1}(u)$ and thus $\ell_i(u) \leq d(u, p_{i-1}(v)) \leq \ell_{i-1}(v) + d(u, v)$ as required.

Assume now that the claim holds for every odd number smaller than j. We prove the claim for $j > 1$. From the induction hypothesis it follows that $\ell_i(u) \leq \ell_{i-(j-2)}(v) + (j-2) \cdot d(u, v)$. Since $i \leq f(u, v)$ we have $p_{i-(j-1)}(u) \notin B_{i-(j-1)}(v)$ and thus $\ell_{i-(j-2)}(v) \leq d(v, p_{i-(j-1)}(u)) \leq \ell_{i-(j-1)}(u) + d(u, v)$. Since $i \leq f(u, v)$ we have $p_{i-j}(v) \notin B_{i-j}(u)$ and thus $\ell_{i-(j-1)}(u) \leq d(u, p_{i-j}(v)) \leq \ell_{i-j}(v) + d(u, v)$. Adding these together we get $\ell_{i-(j-2)}(v) \leq \ell_{i-j}(v) + 2d(u, v)$. Substituting this in the induction hypothesis we get that $\ell_i(u) \leq \ell_{i-j}(v) + j \cdot d(u, v)$, as required.

We now turn to prove (i). We need to show that $\ell_i(u) \leq \ell_{i-j}(u) + j \cdot d(u, v)$, for every even $j \leq i$. This trivially holds for $j = 0$, so let $j \geq 2$. From (ii) we have $\ell_i(u) \leq \ell_{i-(j-1)}(v) + (j-1) \cdot d(u, v)$. Since $i \leq f(u, v)$ we have $p_{i-j}(u) \notin B_{i-j}(v)$ and thus $\ell_{i-(j-1)}(v) \leq d(v, p_{i-j}(u)) \leq \ell_{i-j}(u) + d(u, v)$. Therefore, $\ell_i(u) \leq \ell_{i-(j-1)}(v) + (j-1) \cdot d(u, v) \leq \ell_{i-j}(u) + j \cdot d(u, v)$ are required.

We finish the description of Thorup-Zwick distance oracle with a bound on $dist(u,v)$.

Lemma 4. *$dist(u,v)$ outputs an estimation that is bounded by $2\ell_{f(u,v)}(u) + d(u,v) \leq (2f(u,v){+}1)d(u,v) \leq (2k{-}1)d(u,v)$, for even $f(u,v)$ and by $2\ell_{f(u,v)}(v)$ $+d(u,v) \leq (2f(u,v)+1)d(u,v) \leq (2k-1)d(u,v)$, for odd $f(u,v)$.*

Proof. Let $i = f(u,v)$ be even. The algorithm returns $d(u,p_i(u)) + d(v,p_i(u))$. Using the triangle inequality we get $d(u,p_i(u)) + d(v,p_i(u)) \leq 2\ell_i(u) + d(u,v)$. From Lemma 2 we have $\ell_i(u) \leq i \cdot d(u,v)$ and since $i \leq k-1$ we get $d(u,p_i(u))+d(v,p_i(u)) \leq (2i+1)d(u,v) \leq (2k-1)d(u,v)$. For the case that $f(u,v)$ is odd the proof is the same with u and v switching their roles.

In order to obtain our improvements we are using a slightly different but yet a standard variant of Thorup-Zwick distance oracles (e.g [12]), which we present below. In this variant we also store the exact distance for every pair $\langle u,v\rangle \in A_{k/2} \times A_{k/2-1}$, when k is even, and every pair $\langle u,v\rangle \in A_{(k-1)/2} \times A_{(k-1)/2}$ when k is odd. In both cases the space remains $O(kn^{1+1/k}\log n)$.

The query will work as follows. Let $u,v \in V$. Let $f = \min(f(u,v), f(v,u))$. If $f \leq \lfloor k/2 \rfloor$ then we output $\min(dist(u,v), dist(v,u))$. If $f > \lfloor k/2 \rfloor$ then we output $\min\big(\ell_{k/2}(u) + d(p_{k/2}(u), p_{k/2-1}(v)) + \ell_{k/2-1}(v), \ell_{k/2}(v) + d(p_{k/2}(v), p_{k/2-1}(u)) + \ell_{k/2-1}(u)\big)$, for an even k, and $\ell_{(k-1)/2}(u) + d(p_{(k-1)/2}(u), p_{(k-1)/2}(v)) + \ell_{(k-1)/2}(v)$, for an odd k.

In the next Lemma we establish an upper bound on the query output when $f > \lfloor k/2 \rfloor$.

Lemma 5. *When $f > \lfloor k/2 \rfloor$ the query algorithm described above returns an estimation that is at most $\min(2\ell_{k/2}(u){+}2\ell_{k/2-1}(v){+}d(u,v), 2\ell_{k/2}(v){+}2\ell_{k/2-1}(u){+}d(u,v))$, when k is even and at most $2\ell_{(k-1)/2}(u) + 2\ell_{(k-1)/2}(v) + d(u,v)$, when k is odd.*

Proof. Let $a = \ell_{k/2}(u) + d(p_{k/2}(u), p_{k/2-1}(v)) + \ell_{k/2-1}(v)$. Let $b = \ell_{k/2}(v) + d(p_{k/2}(v), p_{k/2-1}(u)) + \ell_{k/2-1}(u)$. Let $A = 2\ell_{k/2}(u) + 2\ell_{k/2-1}(v) + d(u,v)$ and let $B = 2\ell_{k/2}(v) + 2\ell_{k/2-1}(u) + d(u,v)$. For even k, the query returns $\min\big(a,b\big)$. We show that this value is at most $\min(A,B)$.

Using the triangle inequality we get that $d(p_{k/2}(u), p_{k/2-1}(v)) \leq \ell_{k/2}(u) + d(u,v) + \ell_{k/2-1}(v)$. Therefore, $a \leq A$. Similarly, we get that $d(p_{k/2}(v), p_{k/2-1}(u)) \leq \ell_{k/2}(v) + d(u,v) + \ell_{k/2-1}(u)$. Therefore, $b \leq B$. Adding it all together we get that $\min(a,b) \leq \min(A,B)$, as required.

When k is odd, the query returns $\ell_{(k-1)/2}(u) + d(p_{(k-1)/2}(u), p_{(k-1)/2}(v)) + \ell_{(k-1)/2}(v) \leq \ell_{(k-1)/2}(u) + (\ell_{(k-1)/2}(u) + d(u,v) + \ell_{(k-1)/2}(v)) + \ell_{(k-1)/2}(v) = 2\ell_{(k-1)/2}(u) + 2\ell_{(k-1)/2}(v) + d(u,v)$.

It is relatively straightforward to prove that the estimation produced by the updated query algorithm has $2k-1$ stretch by combining Lemma 5 with Lemma 2.

Throughout the paper we will refer to this variant of Thorup-Zwick distance oracle as the standard variant of Thorup-Zwick distance oracle.

3 Distance Oracles with Improved Stretch

In this section prove the following Theorem:

Theorem 1. *Let $G = (V, E)$ be an n-vertices m-edges undirected unweighted graph. For every $k > 2$ there is a distance oracle that uses $O(kmn^{1/k} \log n)$ space and for every pair of vertices $u, v \in V$ returns in $O(k)$ time an estimation $d^*(u, v)$ such that:*

$$d(u, v) \leq d^*(u, v) \leq (2k - 1)d(u, v) - 4.$$

For $k = 2$, the estimation $d^*(u, v)$ satisfies: $d(u, v) \leq d^*(u, v) \leq 3d(u, v) - 2$.

Before we describe our new construction we present an additional ingredient that we need. Thorup and Zwick [30] in their distance oracle construction defined for every $w \in A_i \setminus A_{i+1}$ a cluster $C(w) = \{u \in V \mid d(u, w) < \ell_{i+1}(u)\}$. Notice that $C(w) = \{u \in V \mid w \in B_i(u)\}$. While the size of the bunch is bounded (see Lemma 1) the size of the cluster is not bounded and can be as large as $n - 1$. Thorup and Zwick [29] showed, in the context of compact routing schemes, that it is possible to bound the size of certain clusters by carefully choosing the set A_1. Their algorithm *center*(G, s) that computes A_1 is given in the Appendix for completeness.

Lemma 6. *Let $A_0 = V$. For $i \in [2, k-1]$ the set A_i is constructed by picking each vertex of A_{i-1} independently with probability p. It is possible to compute a set A_1 of expected size $O((n/p) \cdot \log n)$ by running Algorithm center$(G, n/p)$ such that for every vertex $w \in A_0 \setminus A_1$ it holds that $|C(w)| = O((1/p) \cdot \log n)$ and for every $v \in V$ and $i \in [0, k-1]$ it holds that $|B_i(v)| = O((1/p) \cdot \log n)$.*

3.1 Proof of Theorem 1

Our new distance oracle is constructed as follows. We compute $A_0, \ldots, A_{k-1}$ using Lemma 6 with $p = n^{-1/k}$, and construct the standard variant of Thorup-Zwick distance oracle. We save the graph. We save the set $L(u, \ell_1(u))$, for every $u \in V$, in a 2-level hash table. In the next Lemma we show that the space required for this is within the limits of our space bound.

Lemma 7. $\sum_{u \in V} |L(u, \ell_1(u))| = \tilde{O}(mn^{1/k})$.

Proof. Let $y \in L(u, \ell_1(u))$ and let x be a vertex that precedes y on a shortest path from u to y. Since $d(u, x) = \ell_1(u) - 1$ it follows that $x \in B_0(u)$. We will charge the edge (x, y) for saving the information that $y \in L(u, \ell_1(u))$. We will now count how many times in total we will charge the edge (x, y) for saving such information for y. Each time that (x, y) is being charged for y being in $L(w, \ell_1(w))$, for some $w \in V$, means that $x \in B_0(w)$. Hence, the number of charges to (x, y) because of y is bounded by the number of different vertices w that contain x in $B_0(w)$ which is exactly the size of $C(x)$. Since an edge is charged only for its endpoints, these are all the charges to (x, y) because of y.

The only additional charges to the edge (x, y) come from its other endpoint x in the symmetric case of (y, x). We get:

$$\sum_{u \in V} |L(u, \ell_1(u))| \leq \sum_{y \in V} \sum_{x \in (A_0 \setminus A_1) \cap N(y)} |C(x)| \leq n^{1/k} \log n \sum_{y \in V} deg(y) = \tilde{O}(mn^{1/k})$$

as required.

Since the size of the graph is subsumed by $\tilde{O}(mn^{1/k})$ and the size of the standard variant of Thorup-Zwick distance oracle is $\tilde{O}(n^{1+1/k})$, it follows that the total space used is $\tilde{O}(mn^{1/k})$.

Given a pair $u, v \in V$ the query works as follows. First, we check if $(u, v) \in E$ and if so return 1 and stop. Otherwise, we check if either $v \in L(u, \ell_1(u))$ or $u \in L(v, \ell_1(v))$ and if so return the exact distance and stop. If this is not the case we use the query of the standard variant of Thorup-Zwick distance oracle on u, v and on v, u and report the minimum of these two estimations.

Next, we analyze the stretch of the distance oracle. Let $u, v \in V$ and let $\Delta = d(u, v)$. If $(u, v) \in E$ or $u \in B_0(v)$ or $v \in B_0(u)$ then the exact distance is returned. Therefore, we can assume that $(u, v) \notin E$, $u \notin B_0(v)$ and $v \notin B_0(u)$. Let $d(u', v) = d(u, v') = \Delta - 1$, where $u' \in N(u)$ and $v' \in N(v)$. If $u' \in B_0(v)$ (respectively, $v' \in B_0(u)$) then $u \in L(v, \ell_1(v))$ (respectively, $v \in L(u, \ell_1(u))$) and the exact distance is returned. Therefore, we can assume also that $u' \notin B_0(v)$ and $v' \notin B_0(u)$. This implies that $\ell_1(v) \leq \Delta - 1$ and $\ell_1(u) \leq \Delta - 1$.

For $k = 2$ the standard variant of Thorup-Zwick distance oracle degenerates to the regular one since the additional distances stored are for pairs from $A_1 \times A_0$. The query returns $\ell_1(u) + d(v, p_1(u))$ which is bounded by $2\ell_1(u) + \Delta$. Using the bound $\ell_1(u) \leq \Delta - 1$ we get that the estimation is bounded by $3\Delta - 2$, as required.

Consider now the case that $k \geq 3$. As we have checked whether $(u, v) \in E$, we can assume that $\Delta \geq 2$. Let $f = \min\big(f(u, v), f(v, u)\big)$. In the case that $f \leq \lfloor k/2 \rfloor$ the query returns $\min(dist(u, v), dist(v, u))$. From Lemma 4 it follows that this estimation is bounded by $(2(k/2)+1)d(u, v) = (k+1)\Delta \leq (2k-1)\Delta - 4$ for even $k \geq 4$ and $\Delta \geq 2$, and bounded by $(2((k-1)/2)+1)d(u, v) = k\Delta \leq (2k-1)\Delta - 4$ for odd $k \geq 3$ and $\Delta \geq 2$.

For $f > \lfloor k/2 \rfloor$ the query returns $\min\big(\ell_{k/2}(u) + d(p_{k/2}(u), p_{k/2-1}(v)) + \ell_{k/2-1}(v), \ell_{k/2}(v) + d(p_{k/2}(v), p_{k/2-1}(u)) + \ell_{k/2-1}(u)\big)$, for an even k, and $\ell_{(k-1)/2}(u) + d(p_{(k-1)/2}(u), p_{(k-1)/2}(v)) + \ell_{(k-1)/2}(v)$, for an odd k.

Consider the case of an even k. Let $i = k/2$ and assume that i is even. It follows from Lemma 5 that $2\ell_i(u) + 2\ell_{i-1}(v) + d(u, v)$ is an upper bound for the estimation. From Lemma 3 we have $\ell_i(u) \leq \ell_1(v) + (i-1)\Delta$ and $\ell_{i-1}(v) \leq \ell_1(u) + (i-2)\Delta$. Thus, we get:

Assume now that i is odd. It follows from Lemma 5 that $2\ell_i(v) + 2\ell_{i-1}(u) + d(u, v)$ is an upper bound for the estimation. From Lemma 3 we have $\ell_i(v) \leq \ell_1(v) + (i-1)\Delta$ and $\ell_{i-1}(u) \leq \ell_1(v) + (i-2)\Delta$. Thus, we get:

Consider now the case that k is odd. Let $i = (k-1)/2$. It follows from Lemma 5 that $2\ell_i(u) + 2\ell_i(v) + d(u, v)$ is an upper bound for the estimation.

From Lemma 3 we have $\ell_i(v) \leq \ell_1(u) + (i-1)\Delta$ and $\ell_i(u) \leq \ell_1(v) + (i-1)\Delta$ if i is even or odd. Thus, we get:

4 A Refined Stretch Analysis of Thorup-Zwick Distance Oracle

In this section we present several different conditions that can be easily checked and once fulfilled by the distance oracle of Thorup-Zwick guarantee that the estimation has a stretch which is strictly better than $2k-1$.

The main parameter that we use is the *average distance* between a vertex and the sets $A_1, \ldots, A_{k-1}$. We define the average distance between $u \in V$ and A_i to be $\bar{\ell}_i(u) = \lceil \ell_i(u)/i \rceil$, where $i \in [1, k-1]$.

Let $\hat{d}(u,v) = \min(dist(u,v), dist(v,u))$. We prove the following properties:

Property 1. Let $u \in V$. If $\bar{\ell}_i(u) \neq \bar{\ell}_j(u)$ for some $i, j \in [1, k-1]$ then for every $v \in V$ the stretch of $\hat{d}(u,v)$ is strictly better than $(2k-1)$.

Property 2. Let $u, v \in V$. If $\bar{\ell}_i(u) \neq \bar{\ell}_i(v)$ for some $i \in [1, k-1]$ then the stretch of $\hat{d}(u,v)$ is strictly better than $(2k-1)$.

Property 3. Let $u, v \in V$. If $\bar{\ell}_i(u) = \bar{\ell}_i(v) = q$, for every $i \in [1, k-1]$ and $d(u,v) \neq q$ then the stretch of $\hat{d}(u,v)$ is strictly better than $(2k-1)$.

Before we turn into the technical part of this section we discuss these properties. First notice to the nice relation between these properties. If the conditions of Property 1 do not hold then the conditions of Property 2 can still hold, and if the conditions of both Properties 1 and 2 do not hold then the conditions of Property 3 can still hold.

From the implementation perspective we can verify whether Property 1 and Property 2 hold using a simple computation that does not require the actual computation of the distance oracle itself. Moreover, if Property 1 does not hold then we have $\bar{\ell}_i(u) = \ell_1(u)$, for every $i \in [1, k-1]$, since $\bar{\ell}_1(u) = \ell_1(u)$. Thus, $\ell_1(u) - 1 \leq {}^{\ell_i(u)}/_i \leq \ell_1(u)$ and we get that $\ell_i(u) \in [i\ell_1(u) - i, i\ell_1(u)]$. In such a scenario the shortest paths tree of u has a relatively well defined structure in which $|B(u, \ell_1(u))| \leq n^{1/k}$ and for every $i \in [2, k-1]$ it holds that $|B(u, i\ell_1(u) - i)| \leq n^{i/k}$ and $n^{i/k} \leq |B(u, i\ell_1(u))|$. This situation is depicted in Fig 1 in the Appendix. It is a plausible conjecture that such a well defined structure is not common. For the sake of completeness we do a small experiment on several different datasets of real world graphs to test how frequent these properties are. We elaborate more on this experiment in Section 5.2 of the Appendix.

We now turn to the technical part of this section and start by presenting a simple but yet important observation, that stems from Lemma 2 and is needed for our refined stretch analysis.

Observation 2 *Let $v \in V$ and let $i \leq k-1$ be even (resp., odd). If $\ell_i(u) > i \cdot d(u,v)$ (resp., $\ell_i(v) > i \cdot d(u,v)$) then one of the following must hold:*

(i) There is an even $j < i$ such that $p_j(u) \in B_j(v)$
(ii) There is an odd $j < i$ such that $p_j(v) \in B_j(u)$

Proof. Let i be even. Assume that neither (i) nor (ii) holds. In such a case $i \leq f(u,v)$ and from Lemma 2 we have $\ell_i(u) \leq i \cdot d(u,v)$, a contradiction. The odd case is symmetric.

Let $h \in [1, k-1]$ and let $\Delta = d(u,v)$. The following cases cover all the possible relations between the values $h\Delta$, $\ell_h(u)$ and $\ell_h(v)$.

(i) $h\Delta < \ell_h(u)$ or $h\Delta < \ell_h(v)$
(ii) $h\Delta \geq \ell_h(u) + h$ or $h\Delta \geq \ell_h(v) + h$
(iii) $\ell_h(u) \leq h\Delta < \ell_h(u) + h$ and $\ell_h(v) \leq h\Delta < \ell_h(v) + h$

Next, we bound $\hat{d}(u,v)$ in case (i).

Lemma 8. *If $u, v \in V$ satisfy case (i) then $\hat{d}(u,v) \leq (2k-3)d(u,v)$.*

Proof. In case (i) either $h\Delta < \ell_h(u)$ or $h\Delta < \ell_h(v)$. Assume, wlog, that $h\Delta < \ell_h(u)$. Thus, the condition of Observation 2 holds for u. This implies that either $p_j(u) \in B_j(v)$ for an even $j < h$ or $p_j(v) \in B_j(u)$ for an odd $j < h$. Therefore, $\min\big(f(u,v), f(v,u)\big) < h \leq k-1$ and from Lemma 4 it follows that $\hat{d}(u,v) \leq (2k-3)d(u,v)$, as required.

We proceed to case (ii) and prove:

Lemma 9. *If $u, v \in V$ satisfy case (ii) then $\hat{d}(u,v) \leq \max\big((2k-1)d(u,v) - 2h, (2k-3)d(u,v)\big)$.*

Proof. In case (ii) either $h\Delta \geq \ell_h(u)+h$ or $h\Delta \geq \ell_h(v)+h$. If $\min\big(f(u,v), f(v,u)\big) < h$ then since $h \leq k-1$ it follows from Lemma 4 that $\hat{d}(u,v) \leq (2k-3)d(u,v)$ as required. So we can assume that both $f(u,v) \geq h$ and $f(v,u) \geq h$.

Assume, wlog, that $h\Delta \geq \ell_h(u)+h$. Consider first the case that h is even and assume also that $f(u,v)$ is even. From Lemma 4 we have $dist(u,v) \leq 2\ell_{f(u,v)}(u) + \Delta$. Since $h \leq f(u,v)$ it follows from Lemma 3 that $\ell_{f(u,v)}(u) \leq (f(u,v)-h)\Delta + \ell_h(u)$. Since $\ell_h(u) \leq h\Delta - h$ we get that $\ell_{f(u,v)}(u) \leq f(u,v)\Delta - h$. Therefore, $dist(u,v) \leq 2\ell_{f(u,v)}(u) + \Delta \leq 2f(u,v)\Delta - 2h + \Delta$ and since $f(u,v) \leq k-1$ we get that $dist(u,v) \leq (2k-1)d(u,v) - 2h$, as required.

Assume now that $f(u,v)$ is odd. Since h is even it must be that $h < f(u,v)$, thus, $h \leq f(u,v)-1$. From Lemma 4 we have $dist(u,v) \leq 2\ell_{f(u,v)}(v) + \Delta$. From the definition of $f(u,v)$ it follows that $p_{f(u,v)-1}(u) \notin B(v)$ thus $\ell_{f(u,v)}(v) \leq d(v, p_{f(u,v)-1}(u)) \leq \ell_{f(u,v)-1}(u) + \Delta$. So $dist(u,v) \leq 2\ell_{f(u,v)-1}(u) + 3\Delta$. Now, since $f(u,v)-1 \geq h$ and both h and $f(u,v)-1$ are even it follows from Lemma 3 that $\ell_{f(u,v)-1}(u) \leq (f(u,v)-1-h)\Delta + \ell_h(u)$.

Since by our assumption $\ell_h(u) \leq h\Delta - h$ we get that $\ell_{f(u,v)-1}(u) \leq (f(u,v)-1)\Delta - h$. Therefore, $dist(u,v) \leq 2((f(u,v)-1)\Delta - h) + 3\Delta = (2f(u,v)+1)\Delta - 2h$. Since $f(u,v) \leq k-1$ we get $dist(u,v) \leq (2k-1)d(u,v) - 2h$, as required.

Consider now the case that h is odd and assume also that $f(v,u)$ is odd. From Lemma 4 we have $dist(v,u) \leq 2\ell_{f(v,u)}(u) + \Delta$. Since $h \leq f(v,u)$ it follows from Lemma 3 that $\ell_{f(v,u)}(u) \leq (f(v,u) - h)\Delta + \ell_h(u)$. Since $\ell_h(u) \leq h\Delta - h$ we get that $\ell_{f(v,u)}(u) \leq f(v,u)\Delta - h$. Therefore, $dist(u,v) \leq 2\ell_{f(v,u)}(u) + \Delta \leq 2f(v,u)\Delta - 2h + \Delta$ and since $f(v,u) \leq k-1$ we get that $dist(v,u) \leq (2k-1)d(u,v) - 2h$, as required.

Assume now that $f(v,u)$ is even. Since h is odd it must be that $h < f(v,u)$, thus, $h \leq f(v,u)-1$. From Lemma 4 we have $dist(v,u) \leq 2\ell_{f(v,u)}(v) + \Delta$. From the definition of $f(v,u)$ it follows that $p_{f(v,u)-1}(u) \notin B(v)$ thus $\ell_{f(v,u)}(v) \leq d(v, p_{f(v,u)-1}(u)) \leq \ell_{f(v,u)-1}(u) + \Delta$. So $dist(v,u) \leq 2\ell_{f(v,u)-1}(u) + 3\Delta$. Now, since $f(v,u)-1 \geq h$ and both h and $f(v,u)-1$ are odd it follows from Lemma 3 that $\ell_{f(v,u)-1}(u) \leq (f(v,u) - 1 - h)\Delta + \ell_h(u)$.

Since by our assumption $\ell_h(u) \leq h\Delta - h$ we get that $\ell_{f(v,u)-1}(u) \leq (f(v,u) - 1)\Delta - h$. Therefore, $dist(v,u) \leq 2((f(v,u)-1)\Delta - h) + 3\Delta = (2f(v,u)+1)\Delta - 2h$. Since $f(v,u) \leq k-1$ we get $dist(v,u) \leq (2k-1)d(u,v) - 2h$, as required.

From Lemma 8 and Lemma 9 it follows that the stretch of $\hat{d}(u,v)$ is strictly better than $2k-1$ in cases (i) and (ii). Next, we use these Lemmas to prove the main Lemma of this section.

Lemma 10. *Let $x,y \in V$. If $d(x,y) \neq \bar{\ell}_h(x)$, for some $h \in [1,k-1]$, then $dist(x,y) \leq \max\big((2k-1)d(x,y) - 2h, (2k-3)d(x,y)\big)$.*

Proof. Consider first the case that $\ell_h(x)$ is a multiplicative of h, that is, $\ell_h(x) = \Gamma h$. Since $d(x,y) \neq \bar{\ell}_h(x)$ then either $d(x,y) < \Gamma$ or $d(x,y) > \Gamma$.

If $d(x,y) < \Gamma$ then $h \cdot d(x,y) < h\Gamma = \ell_h(x)$ and we are in case (i). If $d(x,y) > \Gamma$ then $d(x,y) \geq \Gamma + 1$ and we get $h \cdot d(x,y) \geq h\Gamma + h = \ell_h(x) + h$ which is case (ii). We now turn to the case that $\ell_h(x)$ is not a multiplicative of h, that is, $\ell_h(x) = h\Gamma + r$, where $r \in [1, h-1]$. In this case we have $d(x,y) \neq \Gamma + 1$. If $d(x,y) < \Gamma + 1$ then we have $h \cdot d(x,y) \leq h\Gamma < h\Gamma + r = \ell_h(x)$ and we are in case (i). If $d(x,y) > \Gamma + 1$ then we have $h \cdot d(x,y) \geq h(\Gamma + 2) \geq \ell_h(x) + h$ and we are in case (ii). From Lemma 8 and Lemma 9 it follows that $dist(x,y) \leq \max((2k-1)d(x,y) - 2h, (2k-3)d(x,y))$ as required.

We are now ready to prove the properties.

Proof of Property 1. We show that if $\bar{\ell}_i(u) \neq \bar{\ell}_j(u)$ for some $i,j \in [1,k-1]$ then for every $v \in V$ the stretch of $\hat{d}(u,v)$ is strictly smaller than $2k-1$. Let $W \subseteq V$ be a set of vertices that satisfy $d(u,w) = \bar{\ell}_i(u)$, for every $w \in W$. For every $v \in V \setminus W$ we have $d(u,v) \neq \bar{\ell}_i(u)$ and the proof follows by applying Lemma 10 with $h = i$. For every $w \in W$ we have $d(u,w) = \bar{\ell}_i(u) \neq \bar{\ell}_j(u)$ and the proof follows by applying Lemma 10 with $h = j$.

Proof of Property 2. We show that if there is an $i \in [1,k-1]$ such that $\bar{\ell}_i(u) \neq \bar{\ell}_i(v)$ then the stretch of $\hat{d}(u,v)$ is better than $2k-1$. First we assume that both u and v do not satisfy the condition of Property 1 as if one of them satisfies the condition then the stretch of $\hat{d}(u,v)$ is strictly smaller than $2k-1$.

Therefore, for every $i \in [1, k-1]$ we have $\bar{\ell}_i(u) = q_u$, $\bar{\ell}_i(v) = q_v$ and $q_u \neq q_v$. Now if $d(u,v) = q_u$ then $d(u,v) \neq \bar{\ell}_i(v)$ and and if $d(u,v) = q_v$ then $d(u,v) \neq \bar{\ell}_i(u)$. In both cases the proof follows from Lemma 10.

Proof of Property 3. In this case $\bar{\ell}_i(u) = \bar{\ell}_i(v) = q$, for every $i \in [1, k-1]$. From Lemma 10 it follows that if $\hat{d}(u,v) \neq q$ then $\hat{d}(u,v)$ has stretch strictly better than $2k-1$.

References

1. Amir Abboud and Greg Bodwin. The 4/3 additive spanner exponent is tight. In *STOC*, 2016.
2. Ittai Abraham and Cyril Gavoille. On approximate distance labels and routing schemes with affine stretch. In *Distributed Computing - 25th International Symposium, DISC 2011, Rome, Italy, September 20-22, 2011. Proceedings*, pages 404–415, 2011.
3. Rachit Agarwal, Brighten Godfrey, and Sariel Har-Peled. Faster approximate distance queries and compact routing in sparse graphs. *CoRR*, 2012.
4. Rachit Agarwal and Philip Brighten Godfrey. Distance oracles for stretch less than 2. In *SODA*, 2013.
5. Ingo Althöfer, Gautam Das, David P. Dobkin, Deborah Joseph, and José Soares. On sparse spanners of weighted graphs. *Discrete & Computational Geometry*, 1993.
6. Surender Baswana, Telikepalli Kavitha, Kurt Mehlhorn, and Seth Pettie. Additive spanners and (alpha, beta)-spanners. *ACM Trans. Algorithms*, 2010.
7. Uri Ben-Levy and Merav Parter. New (α, β) spanners and hopsets. In Shuchi Chawla, editor, *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA 2020, Salt Lake City, UT, USA, January 5-8, 2020*, pages 1695–1714. SIAM, 2020.
8. Aaron Bernstein. Fully dynamic (2 + epsilon) approximate all-pairs shortest paths with fast query and close to linear update time. In *50th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2009, October 25-27, 2009, Atlanta, Georgia, USA*, pages 693–702, 2009.
9. Shiri Chechik. Compact routing schemes with improved stretch. In *ACM Symposium on Principles of Distributed Computing, PODC '13, Montreal, QC, Canada, July 22-24, 2013*, pages 33–41, 2013.
10. Shiri Chechik. New additive spanners. In *SODA*, 2013.
11. Shiri Chechik. Approximate distance oracles with constant query time. In *STOC*, 2014.
12. Shiri Chechik. Approximate distance oracles with improved bounds. In *STOC*, 2015.
13. Wei Chen, Christian Sommer, Shang-Hua Teng, and Yajun Wang. Compact routing in power-law graphs. In *Distributed Computing, 23rd International Symposium, DISC 2009, Elche, Spain, September 23-25, 2009. Proceedings*, pages 379–391, 2009.
14. D. Dor, S. Halperin, and U. Zwick. All-pairs almost shortest paths. *SIAM J. Comput.*, 29(5):1740–1759, 2000.
15. Michael Elkin and David Peleg. (1+epsilon, beta)-spanner constructions for general graphs. *SIAM J. Comput.*, 2004.

16. Michael Elkin and Seth Pettie. A linear-size logarithmic stretch path-reporting distance oracle for general graphs. *ACM Trans. Algorithms*, 2016.
17. Monika Henzinger, Sebastian Krinninger, and Danupon Nanongkai. Dynamic approximate all-pairs shortest paths: Breaking the o(mn) barrier and derandomization. *SIAM J. Comput.*, 45(3):947–1006, 2016.
18. Jakub Lacki, Jakub Ocwieja, Marcin Pilipczuk, Piotr Sankowski, and Anna Zych. The power of dynamic distance oracles: Efficient dynamic algorithms for the steiner tree. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 11–20, 2015.
19. Manor Mendel and Assaf Naor. Ramsey partitions and proximity data structures. In *47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006), 21-24 October 2006, Berkeley, California, USA, Proceedings*, pages 109–118, 2006.
20. Merav Parter. Bypassing erdős' girth conjecture: Hybrid stretch and sourcewise spanners. In *ICALP*, pages 608–619, 2014.
21. Mihai Patrascu and Liam Roditty. Distance oracles beyond the thorup-zwick bound. *SIAM J. Comput.*, 2014.
22. Mihai Patrascu, Liam Roditty, and Mikkel Thorup. A new infinity of distance oracles for sparse graphs. In *FOCS*, 2012.
23. Ely Porat and Liam Roditty. Preprocess, set, query! *Algorithmica*, 67(4):516–528, 2013.
24. Zichao Qi, Yanghua Xiao, Bin Shao, and Haixun Wang. Toward a distance oracle for billion-node graphs. *PVLDB*, 7(1):61–72, 2013.
25. Liam Roditty and Roei Tov. New routing techniques and their applications. In *Proceedings of the 2015 ACM Symposium on Principles of Distributed Computing, PODC 2015, Donostia-San Sebastián, Spain, July 21 - 23, 2015*, pages 23–32, 2015.
26. Liam Roditty and Roei Tov. Close to linear space routing schemes. *Distributed Computing*, 29(1):65–74, 2016.
27. Liam Roditty and Uri Zwick. Dynamic approximate all-pairs shortest paths in undirected graphs. *SIAM J. Comput.*, 41(3):670–683, 2012.
28. Christian Sommer, Elad Verbin, and Wei Yu. Distance oracles for sparse graphs. In *FOCS*, 2009.
29. Mikkel Thorup and Uri Zwick. Compact routing schemes. In *SPAA*, pages 1–10, 2001.
30. Mikkel Thorup and Uri Zwick. Approximate distance oracles. *J. ACM*, 2005.
31. David P. Woodruff. Additive spanners in nearly quadratic time. In Samson Abramsky, Cyril Gavoille, Claude Kirchner, Friedhelm Meyer auf der Heide, and Paul G. Spirakis, editors, *Automata, Languages and Programming, 37th International Colloquium, ICALP 2010, Bordeaux, France, July 6-10, 2010, Proceedings, Part I.*
32. Christian Wulff-Nilsen. Approximate distance oracles with improved query time. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2013, New Orleans, Louisiana, USA, January 6-8, 2013*, pages 539–549, 2013.

5 Appendix

5.1 Thorup and Zwick's center Algorithm

Algorithm 2: center(G, s)

$A \leftarrow \emptyset$, $W \leftarrow V$;
while $W \neq \emptyset$ **do**
  $A \leftarrow A \cup \mathit{sample}(W, s)$;
  $W \leftarrow \{w \in V \mid |C(w)| > 4n/s\}$;
return A;

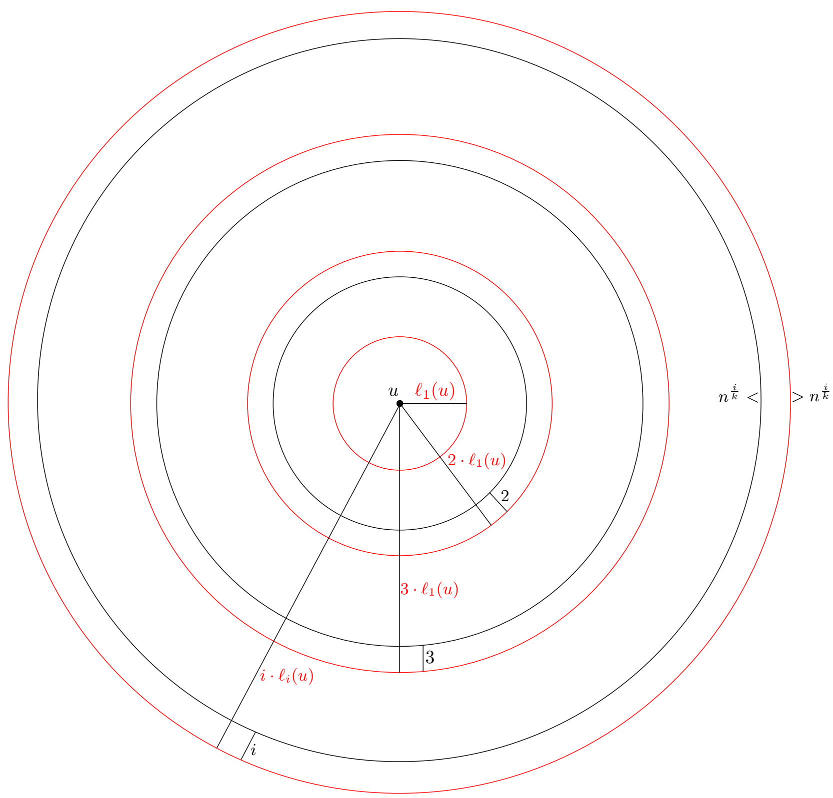

Fig. 1: A shortest paths tree of vertex u that does not satisfy Property 1. The black circle delimits $B(u, i\ell_1(u) - i)$ and the red circle delimits $B(u, i\ell_1(u))$, for $i \in [1, k-1]$.

5.2 A small experiment

Description of the experiment. In our experiment we use datasets from the Stanford Network Analysis Project (SNAP) which contains real-world graphs from various domains (see https://snap.stanford.edu/index.html). We consider in our experiment undirected graphs, where for each graph we apply the following steps:

1. Parse and load the graph edges from the dataset into an undirected graph data structure.
2. Compute using Breadth First Search (BFS) or Dijkstra's algorithm the distance matrix of every graph.
3. Construct Thorup-Zwick's Distance Oracle for $k = 2, 3, 5, 10$.
4. Calculate statistics for Properties 1, 2 and 3 and for the stretch.

For each undirected graph $G = (V, E)$ and k, we considers the following measurements:

– Measurement 1: Percentage of pairs in $V \times V$ that fulfil Property 1.
– Measurement 2: Percentage of pairs in $V \times V$ that fulfil Property 2.
– Measurement 3: Percentage of pairs in $V \times V$ that fulfil Property 3.
– Measurement 4: Percentage of pairs in $V \times V$ that fulfil at least one of the three properties.
– Measurement 5: The average stretch returned by Thorup-Zwick's Distance Oracle when one of the properties hold.
– Measurement 6: The average stretch returned by Thorup-Zwick's Distance Oracle when none of the properties hold.
– Measurement 7: The average stretch returned by Thorup-Zwick's Distance Oracle.

The code of this experiment, datasets and results are available at:
https://github.com/roei-tov/distance-oracle-experiment.

Results. As mentioned above we use datasets taken from the SNAP web site. For each graph in the dataset we consider the largest connected component of the graph.

The characteristics of the datasets are summarized in Table 1 (all graphs are unweighted) and the results for $k = 2, 3, 5, 10$ are presented in Tables 2, 3, 4, 5, respectively.

Notice that it follows from the definition of Property 1 that for $k = 2$ Property 1 can never hold, thus we excluded Measurement 1 from Tables 2. Also, since it turns out that for $k = 3, 5, 10$ all pairs of vertices in the dataset we used for our experiment fulfill at least one of the properties, we excluded Measurement 4, 5, 6 from Tables 3, 4, 5, accordingly.

ID	Dataset Name	Category	Nodes	Edges
1	Ego Facebook	Social	$4,039$	$88,234$
2	Deezer (Croatia)	Social	$54,573$	$498,202$
3	Facebook Page to Page Mutual Like (New Sites Category)	Social	$27,917$	$206,259$
4	Twitch Users Friendship (Germany)	Social	$9,498$	$153,138$
5	Github Developers Mutual Follower Relationship	Social	$37,700$	$289,003$
6	General Relativity and Quantum Cosmology collaboration	Collaboration	$4,158$	$13,428$
7	High Energy Physics - Theory collaboration	Collaboration	$8,638$	$24,827$
8	Email Enron	Communication	$33,696$	$180,811$
9	Oregon-2 Mar 31 2001	Autonomous Systems	$10,900$	$31,180$

Table 1: Characteristic of the datasets

Discussion. We can see at the table for $k = 2$ that when none of the properties hold then the average stretch is significantly higher than the average stretch in the case that Property 2 or Property 3 hold. For higher values of k we can see that the fact that Property 1 can hold changes the situation and there is no pair of vertices that does not satisfy at least one of the properties. Moreover, as we can see from the average stretch, even for a large value of k, such as $k = 10$ the average stretch is below 2. This reveals an interesting phenomena, in which, for some pairs the stretch for $k = 2$ might be worse than the stretch for $k = 10$.

Another thing to notice is that the percentage of pairs for which Property 1 hold increases when k increases. This is perhaps not so surprising since for larger values of k there are more room for indices $i, j \in [1, k-1]$ for which $\bar{\ell}_i(u) \neq \bar{\ell}_i(v)$.

Dataset ID	Prop. 2 Fulfilled (%)	Prop. 3 Fulfilled (%)	At Least One Fulfilled (%)	Avg. Stretch - Any Fulfilled	Avg. Stretch - None Fulfilled	Avg. Stretch Overall
1	48.68%	44.95%	93.64%	1.45	2.08	1.49
2	57.8%	41.62%	99.43%	1.46	2.46	1.46
3	60.35%	38.6%	98.95%	1.41	2.26	1.41
4	41.79%	40.43%	82.22%	1.55	2.11	1.65
5	39.28%	49.75%	89.03%	1.54	2.21	1.62
6	76.18%	23.04%	99.22%	1.43	2.28	1.43
7	73%	26.52%	99.52%	1.37	2.31	1.38
8	50.83%	47.57%	98.4%	1.34	2.09	1.35
9	48.43%	43.44%	91.86%	1.53	2.12	1.58

Table 2: Results for $k = 2$

Dataset ID	Prop. 1 Fulfilled (%)	Prop. 2 Fulfilled (%)	Prop. 3 Fulfilled (%)	Avg. Stretch Overall
1	55.46%	63.58%	56.33%	1.49
2	19.95%	80.35%	42.37%	1.61
3	19.65%	82.63%	39.16%	1.58
4	44.82%	67.75%	44.22%	1.75
5	20.99%	71.67%	49.26%	1.6
6	18.23%	91.41%	28.32%	1.47
7	24.08%	90.66%	28.75%	1.57
8	14.82%	73.88%	53.38%	1.55
9	11.72%	72.87%	56.61%	1.63

Table 3: Results for $k = 3$

Dataset ID	Prop. 1 Fulfilled (%)	Prop. 2 Fulfilled (%)	Prop. 3 Fulfilled (%)	Avg. Stretch Overall
1	62.59%	93.75%	56.93%	1.6
2	79.17%	94.45%	46.57%	1.67
3	72.16%	95.07%	41.9%	1.53
4	73.52%	87.02%	50.79%	1.82
5	58.38%	90.68%	38.53%	1.76
6	55.72%	99.02%	32.53%	1.7
7	58.62%	98.5%	32.19%	1.64
8	62.96%	92.16%	38.54%	1.6
9	41.61%	86.78%	42.31%	1.63

Table 4: Results for $k = 5$

Dataset ID	Prop. 1 Fulfilled (%)	Prop. 2 Fulfilled (%)	Prop. 3 Fulfilled (%)	Avg. Stretch Overall
1	90.77%	98.28%	49.55%	1.61
2	97.05%	99.65%	48.21%	1.83
3	96.44%	99.47%	44.86%	1.77
4	94.76%	96.65%	47.76%	1.71
5	94.19%	97.92%	39.63%	1.88
6	90.12%	99.88%	40.01%	1.61
7	91.58%	99.91%	39.33%	1.66
8	91.93%	99.3%	37.06%	1.77
9	91.67%	98.03%	36.41%	1.65

Table 5: Results for $k = 10$