Distance Oracles for Sparse Graphs

Christian Sommer
Dept. of Computer Science
University of Tokyo and NII
Tokyo, Japan
sommer@nii.ac.jp

Elad Verbin*
ITCS
Tsinghua University
Beijing, China
elad.verbin@gmail.com

Wei Yu*
ITCS
Tsinghua University
Beijing, China
zig.wei@gmail.com

Abstract— Thorup and Zwick, in their seminal work, introduced the *approximate distance oracle*, which is a data structure that answers distance queries in a graph. For any integer k, they showed an efficient algorithm to construct an approximate distance oracle using space $O(kn^{1+1/k})$ that can answer queries in time $O(k)$ with a distance estimate that is at most $\alpha = 2k-1$ times larger than the actual shortest distance (α is called the *stretch*).

They proved that, under a combinatorial conjecture, their data structure is optimal in terms of space: if a stretch of at most $2k-1$ is desired, then the space complexity is at least $n^{1+1/k}$. Their proof holds even if infinite query time is allowed: it is essentially an "incompressibility" result. Also, the proof only holds for dense graphs, and the best bound it can prove only implies that the size of the data structure is lower bounded by the number of edges of the graph. Naturally, the following question arises: what happens for sparse graphs?

In this paper we give a new lower bound for approximate distance oracles in the cell-probe model. This lower bound holds even for sparse ($polylog(n)$-degree) graphs, and it is not an "incompressibility" bound: we prove a three-way tradeoff between space, stretch and query time. We show that, when the query time is t, and the stretch is α, then the space $\mathcal{S}$ must be

$$\mathcal{S} \geq n^{1+\Omega(1/t\alpha)} / \lg n \ . \tag{1}$$

This lower bound follows by a reduction from lopsided set disjointness to distance oracles, based on and motivated by recent work of Pătraşcu.

Our results in fact show that for *any* high-girth regular graph, an approximate distance oracle that supports efficient queries for all subgraphs of G must obey Eq. (1). We also prove some lemmas that count sets of paths in high-girth regular graphs and high-girth regular expanders, which might be of independent interest.

Keywords-distance oracle; data structures; lower bounds; cell-probe model; lopsided set disjointness

1. Introduction

An *approximate distance oracle* is a data structure that answers distance queries $d(u,v)$ for a (connected) graph $G = (V,E)$. If the reported distance $\tilde{d}(u,v)$ satisfies $d(u,v) \leq \tilde{d}(u,v) \leq \alpha \cdot d(u,v)$ for all $u,v \in V$, the distance oracle is said to have *multiplicative stretch* α. In all our lower bounds, the graphs are *unweighted*.[1]

Thorup and Zwick [29] coined the term *distance oracle* and gave a method to preprocess an undirected graph $G = (V,E)$ in time $\tilde{O}(kmn^{1/k})$ to create, for any integer k, a data structure of size $O(kn^{1+1/k})$ that can answer distance queries with stretch $2k-1$ in time $O(k)$. They also prove that their oracles are essentially optimal in terms of the stretch-space tradeoff, assuming a girth conjecture by Erdős and others (Section 1.1). The preprocessing and query time were improved upon later. Baswana and Sen [7] showed how to get a preprocessing time of $O(n^2)$ for unweighted graphs. Mendel and Naor [20] gave a data structure with query time $O(1)$, sacrificing a constant factor in the stretch. If the input is restricted to special classes of graphs, the girth-based lower bound may not apply. Indeed, there are better constructions for some restricted graph classes. For digraphs with bounded tree-width, Chaudhuri and Zaroliagis [9] gave an algorithm with linear preprocessing time and almost constant query time. Thorup [28] and Klein [17], [18] independently proposed efficient $(1+\epsilon)$–approximate distance oracles for planar graphs. For a brief overview of results on distance oracles, including the new lower bounds introduced in this paper, see Table I.

1.1. Thorup and Zwick's Girth-based Lower Bound

For sufficiently dense graphs, an information-theoretic space lower bound was proven using the following girth conjecture by Erdős and others [13], [14]:

Conjecture. *There exists a graph $G = (V,E)$ with $|V| = n$ nodes and $|E| = \Omega(n^{1+1/k})$ edges and girth $g(G) > 2k$.*

The girth conjecture was proven for certain values of k (1, 2, 3, and 5); for an overview see Hoory [16], and for

*This work was supported in part by the National Natural Science Foundation of China Grant 60553001, and the National Basic Research Program of China Grant 2007CB807900,2007CB807901.

[1]As usual, $n = |V|$ denotes the number of nodes, $m = |E|$ denotes the number of edges of the graph G, and $\tilde{O}$ is the big-Oh notation hiding poly-logarithmic factors. As usual, we abbreviate $[n] = \{1, 2, \ldots, n\}$. All logarithms are base-2 unless explicitly stated otherwise. $\ln$ denotes $\lg_e$. The *distance* $d_G(u,v)$ between u and v in G is the length, in edges, of the shortest path between u and v. Three other definitions we use in the introduction are: the *girth* of G, denoted by $g(G)$, is the length of the shortest cycle in G. G is called *r-regular* if each vertex has exactly r neighbors. Two (or more) paths are called *vertex-disjoint* if they do not have any vertices in common.

Graphs	Preprocessing	Space	Query	Stretch
general [13], [14], [29]		$\Omega(n^{1+1/k})$		$< 2k+1$
general (Lemma 10)		$n^{1+\Omega(1/t)}$	t	$\leq$ additive 1
general (Theorem 1)		$n^{1+\Omega(1/\alpha t)}/\lg n$	t	$\leq \alpha$
		$n^{1+\Omega(1/\alpha)}/\lg n$	$O(1)$	$\leq \alpha$
		$n^{1+\Omega(1/t)}/\lg n$	t	$O(1)$
general [29], [30]	$\tilde{O}(kmn^{1/k})$	$O(kn^{1+1/k})$	$O(k)$	$2k-1$
general [20], [21]	$O(mn^{1/k}\lg^2 n)$	$O(n^{1+1/k})$	$O(1)$	$O(k)$
unweighted [7]	$O(n^2)$	$O(kn^{1+1/k})$	$O(k)$	$2k-1$
unweighted [6]	$O(m+n^{23/12})$	$O(n^{3/2})$	$O(1)$	3 plus additive 8
geometric (sparse spanner required) [15]	$O(n\lg n)$	$O(n\lg n)$	$O(1)$	$1+\epsilon$
planar, directed [28]	$O(n\lg^3 n\lg(nW))$	$O(n\lg n\lg(nW))$	$O(\lg\lg(nW))$	$1+\epsilon$
planar, undirected [17], [28]	$O(n\lg^3 n)$	$O(n\lg n)$	$O(1)$	$1+\epsilon$
bounded tree-width, directed [9]	$O(n)$	$O(n)$	$O(\alpha(n))$	1

Table I
DISTANCE ORACLES FOR UNDIRECTED GRAPHS (EXCEPT [9], [28]). O-NOTATIONS HIDE ϵ'S. THE TOP PART OF THE TABLE LISTS LOWER BOUNDS, THE BOTTOM PART LISTS UPPER BOUNDS. W DENOTES THE LARGEST INTEGER EDGE WEIGHT. THE STRETCH IS MULTIPLICATIVE UNLESS STATED OTHERWISE. THE RESULTS IN **BOLD** ARE INTRODUCED IN THIS PAPER. THE QUERY TIME $\alpha(n)$ IN [9] DENOTES THE INVERSE ACKERMANN FUNCTION.

the detailed connection to spanners and distance oracles see Thorup and Zwick [29] and Althöfer et al. [4].

Thorup and Zwick's lower bound proof roughly works as follows: All $2^{\Omega(n^{1+1/k})}$ subgraphs G' of the graph G in the conjecture also have large girth $g(G') \geq g(G)$. For a distance oracle with stretch smaller than $2k+1$, at least $\Omega(n^{1+1/k})$ bits of space are necessary, since it must distinguish between any two different subgraphs G' and G'', and it cannot omit any edges as there is no alternative short path [29, Prop. 5.1]. Lower bounds for multiplicative graph spanners [11] follow similar arguments. Woodruff [31] recently gave a lower bound for additive graph spanners without using Erdős' girth conjecture.

The girth conjecture itself is tight in the sense that no larger girth is possible [3], and its implications for distance oracles have also been proven to be almost tight [29] in the sense that the upper bound almost matches the lower bound.

Even though the lower bound by Thorup and Zwick is tight, it is weaker than it seems: the hard instances for their distance oracle are rather dense graphs, which contain roughly $n^{1+1/k}$ edges. For stretch 3, these graphs have $n^{3/2}$ edges. Indeed, they essentially state that we cannot compress a (specially-constructed) graph to less than its original size, or, alternatively, that the size of the data structure must be at least $\Omega(m)$ (bits). The lower bound does not refer to the query time at all, it even holds if the query algorithm is allowed to access the complete data structure. For sparse graphs, such a lower bound does not prove much, since storing the entire graph only costs quasi-linear space.

1.2. Our Contributions

In the context of computing distances, many graphs that are interesting in practice, for example road maps augmented with flight connections, are sparse. Our main result in this paper is a three-way tradeoff stating that, in the cell-probe model [32] (see definition in Section 2), when the query time and the stretch are small, then the space is large. The main result of this paper is the following, which is restated and proved as Theorem 18:

Theorem 1 (Main Theorem)**.** *A distance oracle with stretch α using query time t requires space $\mathcal{S} = n^{1+\Omega(1/t\alpha)}/\lg n$ in the cell-probe model with w-bit cells, even on graphs with maximum degree $poly(tw\alpha/\lg n)$, where we require $\alpha = o\left(\frac{\lg n}{\lg(w\lg n)}\right)$.*

Throughout the paper, $poly(x)$ means a polynomial in x of unspecified constant degree. $poly(x)$ is always allowed to be at least any constant.

Corollary 2. *A distance oracle with stretch α and query time $t = O(1)$ requires space $\mathcal{S} = n^{1+\Omega(1/\alpha)}/\lg n$ in the cell-probe model with w-bit cells, even on graphs with maximum degree $poly(w\alpha/\lg n)$, where we require $\alpha = o\left(\frac{\lg n}{\lg(w\lg n)}\right)$.*

Our lower bound in Corollary 2 is tight up to the constant hidden in the Ω with respect to the distance oracle by Mendel and Naor [20].

Corollary 3. *A distance oracle with stretch $\alpha = O(1)$ and query time t requires space $\mathcal{S} = n^{1+\Omega(1/t)}/\lg n$ in the cell-probe model with w-bit cells, even on graphs with maximum degree $poly(tw/\lg n)$, where we require $w = n^{o(1)}$.*

We can also lower-bound the query time in terms of the space and stretch:

Corollary 4. *A distance oracle with stretch less than α using space $\mathcal{S}$ requires query time $t = \Omega\left(\frac{\lg n}{\alpha\lg(\mathcal{S}\lg n/n)}\right)$ in the cell-probe model with w-bit cells, even when restricted to graphs with maximum degree $poly(tw\alpha/\lg n)$, where we require $\alpha = o\left(\frac{\lg n}{\lg(w\lg n)}\right)$.*

Unlike the lower bound by Thorup and Zwick [29], which is information theoretic, our bound uses new techniques based on Pătraşcu's [26] recent communication lower bounds for Lopsided Set Disjointness (LSD) and the lower bounds for data structures that follow from it.

Pătraşcu proves lower bounds for reachability data structures on the butterfly graph, which is a very *structured* graph, by reducing from LSD. His proof heavily uses the structure of the butterfly graph, and is not generalizable for our application. In this paper, we apply similar ideas, but in a very *unstructured* way: our arguments work on any graph, and use little knowledge about the graph; namely, we only care about the degrees of the vertices in the graph ($poly(w \lg n)$ suffices), and its girth. This allows us to prove hardness of distance approximation on a large class of (high-girth) graphs and their subgraphs.

1.3. Sketch of the Proof Technique

We now give a rough sketch of the proof of the main theorem, highlighting some interesting technical details. The proof is a reduction from the communication problem LSD, in which Alice gets a set $S_A \subseteq [N \cdot B]$ of cardinality N, Bob gets a set $S_B \subseteq [N \cdot B]$, and they need to decide whether $S_A \cap S_B = \emptyset$. Strong lower bounds are known for LSD (see Section 2). We prove that a distance oracle with good parameters implies a good protocol for LSD, and derive our lower bound from the contrapositive of this claim.

A standard way to perform such reductions is to translate one query to the data structure into a communication protocol where Alice sends Bob $t \lg \mathcal{S}$ bits and Bob replies with tw bits. However, lower bounds in communication complexity are usually loose up to constant multiplicative factors. The reduction puts this loose multiplicative factor in the exponent of $\mathcal{S}$. Therefore, this framework can only prove lower bounds on space of the form $\mathcal{S} \geq x^{\Omega(1)}$, where x is some expression that depends on the problem parameters. This is obviously useless for our purposes, since there is a trivial distance oracle that takes space n^2 (just pre-compute all answers) and we wish to prove a lower bound larger than n.

Pătraşcu, in several recent papers [25], [26] found a way to prove lower bounds on $\mathcal{S}$ that hold up to a polylogarithmic multiplicative factor. He considers performing k queries in parallel, where $k = n/polylog(n)$. In [26] he shows that good space and good query time for k range counting queries implies a good communication protocol for LSD, and he deduces a lower bound for range counting. We use the same idea here.

Our goal is thus to show that a good distance oracle somehow implies a good protocol for LSD. Fix some graph $G = (V, E)$ whose girth g is large. Let the universe size of the LSD problem be $N \cdot B = |E|$. Choose an arbitrary bijection $f : E \to [NB]$ between the edge-set E and the elements of the universe $[NB]$.

In the reduction, Bob transforms his set S_B into a subgraph of G, $G' = (V, E') = (V, E \setminus f(S_B))$. Namely, Bob constructs a subgraph G' of G where the missing edges are the ones that correspond to his input S_B. Alice constructs a set of queries based on her input S_A, in a way that we shall specify next. To get a good protocol for LSD, Alice plays the role of the querier, and Bob plays the role of the data structure. Bob builds a distance oracle for the graph G'.

Now consider one query to the data structure, which asks for $\tilde{d}_{G'}(u, v)$ where u and v are close to each other in G, namely $d_G(u, v) = \ell$ where $\ell < \frac{g}{\alpha+1}$. Let $p_{u,v}$ be the path of length ℓ between u and v in G. A query to the distance oracle gives an answer $\tilde{d}_{G'}(u, v)$, which is an approximation to $d_{G'}(u, v)$. We know that the girth of G' is large, $g(G') \geq g$, and we chose u, v to be quite close. We then get that the query will return $\leq \alpha\ell$ if and only if all edges on the path $p_{u,v}$ are in E', and will return a number larger than $\alpha\ell$ otherwise.

Thus, using one query to the distance oracle we can tell apart the case that all edges of $p_{u,v}$ are in E' and the case where at least one edge is missing. Now, if we perform k queries of this form, we can check k paths in this way. Therefore, using k queries we can check whether $k\ell$ edges are all in the graph, or if at least one of these $k\ell$ edges is missing. Setting $NB = |E|$ and $N = k\ell$, this is almost the LSD problem! By doing a standard transformation from a data structure to a communication protocol, we connect the parameters of the data structure to those of a protocol for LSD: Alice sends roughly $tk \lg(\mathcal{S}/k)$ bits to Bob, and Bob sends roughly tkw bits to Alice.

There is one problem with this: Alice's input may not necessarily map to a collection of k vertex-disjoint paths of length ℓ each.

There is a trick of Pătraşcu [26] that allows to get a lower bound for LSD even when only a non-negligible fraction of Alice's inputs map to a set of vertex-disjoint paths. In short, the trick is to allow a preliminary round of communication, where Alice and Bob choose the bijection f from some large set of bijections (the existence of such a set is proved using the probabilistic method, see Claim 13). For this trick to work, we need to guarantee that there is a large ensemble of sets of vertex-disjoint paths in G (we refer to this as the *path-count*). If the path-count is large enough, then we get a strong lower bound. The details of this argument are in the proof of Theorem 11.

We then consider the construction by Lubotzky et al. (LPS) [19] of high-girth regular expanders (see Theorem 9). To prove the main theorem, we just need to prove that the path-count of the LPS graph is large. We prove this using Lemma 14. Surprisingly, in proving this, we only need to use the fact that the graph is regular and that the girth is large (see Lemma 19). To prove Lemma 19, we count and see that an r-regular graph has $\geq |V|\,(r-1)^{\ell}/2$ paths of length ℓ, as long as $\ell < g(G)$. We then use the union bound to argue

that if we have to avoid a $(1/4\ell)$-fraction of the vertices of the graph, we still have $\geq |V|\,(r-1)^{\ell}/4$ paths. We thus can count sets of paths by choosing them one at a time and arguing that we have many options for each choice. In the expander-based version of this lemma (Lemma 14) we use a lemma by Alon et al. [2] (Lemma 16) to improve some of the parameters.

Our proofs are quite modular: the main theorem follows as a combination of Theorem 11, Lemma 14, and the construction of LPS graphs (Theorem 9). All of these results are relatively independent from each other, so each part of the argument can be changed in order to prove different things.

2. Preliminaries

A class of graphs $G = (V, E)$ is considered *sparse* if $|E| = \tilde{O}(|V|)$. In this paper, we will sometimes consider a graph to be *sparse* if $|E| \leq n \cdot poly(w, \lg n)$.

The following is the definition of the ensemble of sets of vertex-disjoint paths. We use this definition extensively throughout the paper.

Definition 5. *For a graph $G = (V, E)$ and two positive integers k, ℓ, let $\mathcal{P}(G, \ell, k)$ be the set whose elements are all possible sets $P \subseteq E$ where P can be written as a union of k vertex-disjoint paths in G, each of length exactly ℓ. When the context is clear, we will denote this simply by $\mathcal{P}$.*

In the cell-probe model [22], [32], a *cell* has w bits and the *space* of a data structure is measured as the number of cells it occupies, say $\mathcal{S}$ cells. The query time is measured by the worst-case number of cells that a query accesses. The most typical values are $w = \lg n$ or $w = polylog(n)$, but larger (or smaller) values may be interesting as well.

Note that it is impossible to prove a lower bound on distance oracles in the cell-probe model that holds for one particular graph G. This is due to the fact that the query algorithm can hard-code G and can then answer queries in constant time using a constant number of probes. Thus, we make the following definition to capture the notion of a graph (or rather a family of graphs) being "hard". For a graph $G = (V, E)$, we say that there is a distance oracle with space $\mathcal{S}$, query time t, and stretch α for a *base-graph* G if for any subgraph $G' = (V, E')$ of G, where $E' \subseteq E$, a data structure can be constructed for G' that uses space $\mathcal{S}$, and such that for any $u, v \in V$ the data structure returns in time t an estimate $\tilde{d}(u, v)$ such that $d_{G'}(u, v) \leq \tilde{d}(u, v) \leq \alpha \cdot d_{G'}(u, v)$. When we prove a tradeoff on the values of $\mathcal{S}, t, \alpha$, we consider this tradeoff evidence that the base-graph G is "hard". Most of our lower bounds in the rest of the paper will be of this form.

In the lopsided (asymmetric) set disjointness problem (LSD), Alice and Bob receive sets S_A and S_B, respectively. Their goal is to determine whether $S_A \cap S_B = \emptyset$ using some communication protocol. Lopsided set disjointness is parameterized by the size of Alice's set $|S_A| = N$ and by B, the fraction between N and the size of the universe NB.

The communication complexity of LSD was bounded by Miltersen et al. [23] as follows:

Lemma 6 (Miltersen et al. [23])**.** *There exists some constant $\mathsf{C} > 0$ such that in a one-sided error protocol for LSD, either Alice sends $\mathsf{C}N \lg B$ bits or Bob sends NB^{C} bits.*

Andoni et al. [5] extended the bound to include protocols with two-sided error as well, and Pătraşcu [26], [27] improved it to the following:

Lemma 7 (Pătraşcu [27, Thm. 1.4])**.** *There exists some constant $\mathsf{C} > 0$ such that in a bounded error protocol for LSD, either Alice sends $\mathsf{C}N \lg B$ bits or Bob sends NB^{C} bits.*

Our reduction from distance oracles to LSD implies a lower bound for deterministic and randomized distance oracles.

2.1. High-Girth Regular Expanders

All of our explicit lower bounds use high-girth graphs. Some of them use high-girth expanders. We use the same construction of Lubotzky et al. [19] both for high-girth expanders and for "just" high-girth graphs, since it allows us much freedom in the choice of parameters.

Definition 8 ([2], [10])**.** *Let $G = (V, E)$ be an r-regular graph with n vertices. Let A be the adjacency matrix of G, i.e. we enumerate the vertices in an arbitrary order $\{v_1, \ldots, v_n\}$, and set $A_{i,j}$ to be 1 if $(v_i, v_j) \in E$ and 0 otherwise. Let $B = A/r$, that is $B_{i,j} = 1/r$ if $(v_i, v_j) \in E$ and $B_{i,j} = 0$ otherwise. Let $\lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_{n-1}$ be the eigenvalues of B. They do not depend on the ordering $\{v_1, \ldots, v_n\}$. We have that $\lambda_0 = 1$. Let $\lambda(G) = \max(\lambda_1, |\lambda_{n-1}|)$. $\lambda(G)$ is called the* (normalized) second-largest eigenvalue *of G. It is a real number in the range $[0, 1]$. G is called* Ramanujan *if $\lambda(G) \leq \frac{2\sqrt{r-1}}{r}$.*

We will not heavily use expansion. The interested reader is referred to [2], [10].

Theorem 9 (Corollary of Lubotzky et al. [19])**.** *For every large enough n_0, r_0 with $n_0 > 8r_0^3$, there exists a graph $G = (V, E)$ with the following properties:*

1) *$|V| = n$ with $n_0 \leq n \leq 8n_0$*
2) *G is r-regular, where $r_0 \leq r \leq 2r_0$*
3) *The girth of G is at least $g(G) \geq \frac{1}{2} \lg_r n$*
4) *$\lambda(G) \leq \frac{2\sqrt{r-1}}{r}$*

Proof: The graph claimed to exist is a Ramanujan graph $X^{p,q}$ as in Lubotzky et al. [19, pp. 262–263]. The graph has $n = q(q^2 - 1)$ nodes and it is $(p+1)$-regular. The girth is at least $2 \lg_p q$ (the girth for the other case of the Jacobi symbol [19, Case i, p. 263] is even larger for p, q large enough).

Their construction requires unequal primes p, q congruent to $1 \mod 4$. In the following, we prove their existence in the range we need them for r and n.

The Bertrand-Chebyshev theorem states that for every $m > 1$ there is always at least one prime p such that $m < p < 2m$. This generalizes to some arithmetic progressions [8], [12], [24]. Breusch [8, p. 505] proves

> [...] daß für $x \geq 7$ zwischen x und $2x$ stets Primzahlen einer jeden der vier Progressionen $3n+1$, $3n+2$, $4n+1$, $4n+3$ liegen.

that for every $x \geq 7$ there is a prime of the form $4n+1$ in the interval between x and $2x$.

Since the intervals for p and q do not overlap we can choose p and q as desired. ∎

In short, there exist regular high-girth expanders. In fact, we could use much weaker constructions, since for the rest of the results, we only need $g(G) \geq \Omega(\lg_r |V|)$ and $\lambda(G) \leq 0.1$. Any reasonable construction of regular high-girth expanders would suffice.

3. Warm-up

One key part in Pătraşcu's result is the reduction from set disjointness to reachability oracles. The reachability problem is, given a sparse *directed* graph $G = (V, E)$, can we construct a data structure using less than n^2 space such that reachability queries (deciding whether there is a directed path from u to v) can be answered efficiently?

Theorem (Pătraşcu [26, Thm. 2])**.** *A reachability oracle using space $\mathcal{S}$ in the cell-probe model with w-bit cells, requires query time $t = \Omega(\lg n / \lg \frac{\mathcal{S}w}{n})$.*

Pătraşcu reduces a variant of LSD to the problem of reachability queries in a butterfly graph. The following direct reduction from reachability oracles to distance oracles yields the same lower bound for the latter.

Lemma 10. *A distance oracle with* **additive** *stretch less than 2 using space $\mathcal{S}$ in the cell-probe model with w-bit cells, requires query time $t = \Omega(\lg n / \lg \frac{\mathcal{S}w}{n})$.*

Proof: We give a reduction from reachability oracles for the directed butterfly graph G in Pătraşcu's reduction [26, Reduction 13] to a distance oracle for the same graph, interpreted as undirected, say G'. If node v is reachable from u in G, the distance in G' is equal to d; if v is not reachable from u in G, the distance in G' must be at least $d+2$ since the butterfly graph G' and its subgraphs are bipartite. Therefore, if there is an algorithm that approximates distances with additive stretch less than 2 using t probes in a data structure of size $\mathcal{S}$, then there is an algorithm that, using the same t probes in the same data structure of size $\mathcal{S}$, that answers reachability queries by distinguishing distances d and distances $\geq d+2$. ∎

4. Lower Bounds for Approximate Distance Oracles as a Function of Path-Counts

In this section we show a lower bound on the space complexity of any approximate distance oracle on *any* base-graph G, based on essentially two parameters of G: the girth of G, and the *path-count* of G, which is the cardinality of the set $\mathcal{P}(G, \ell, k)$.

In a later section (Section 5.2), we substitute a particular graph and get the explicit lower bound stated in Theorem 1. The theorem we prove in the current section is quite general, and we believe it might be of future use.

We now prove that, if the path-count is large, then a strong lower bound holds. In other words, a graph G with a large path-count is a hard base-graph for approximate distance oracles.

Theorem 11. *Let $G = (V, E)$ be a graph, such that an α-approximate distance oracle exists for the base-graph G, using query time t and space $\mathcal{S}$. Let k, ℓ be two positive integers, such that $\ell < \frac{g(G)}{\alpha+1}$. Assume $|E| \geq k\ell(2tw/\ell)^{1/\mathsf{C}}$. Then,*

$$\mathcal{S} \geq k \cdot \left(\frac{|\mathcal{P}(G, \ell, k)|^{1/k\ell}}{e(|E|/k\ell)^{1-\mathsf{C}}} \right)^{\ell/t} / ((e\,|E|)^{1/tk} e) \ .$$

The proof proceeds as follows: In Section 4.2 we show how to reduce the data structure problem to a communication problem, that is, we prove that if there exists a good data structure then there is a good protocol for LSD. Then, in Section 4.3, we use the LSD lower bound to show that there cannot be a good data structure.

For the rest of Section 4, we fix the graph $G = (V, E)$, and the parameters w, α, t, $\mathcal{S}$, k, ℓ. And we define $N = k\ell$ and $B = |E|/N$. Notice that now the condition $|E| \geq k\ell(2tw/\ell)^{1/\mathsf{C}}$ implies that $B \geq (2tw/\ell)^{1/\mathsf{C}}$.

4.1. Sketch of the Proof

As the computations get rather tedious in some parts, we first sketch the calculations to provide a brief but non-rigorous outline (various terms and approximations are omitted).

In Section 4.2, we prove that the existence of the data structure implies a communication protocol for LSD where Alice sends roughly $\lg\left((\mathcal{S}/k)^{tk} B^N / |\mathcal{P}|\right)$ bits and Bob sends ktw bits. Then, in Section 4.3, we see that $ktw < NB^{\mathsf{C}}$, so by the LSD lower bound, Alice's communication must be at least $\mathsf{C}N \lg B$. However, Alice sends $\lg\left((\mathcal{S}/k)^{tk} B^N / |\mathcal{P}|\right)$ bits, thus $\lg\left((\mathcal{S}/k)^{tk} B^N / |\mathcal{P}|\right) \geq \mathsf{C}N \lg B$, and by simplifying we get $\mathcal{S} \geq k\left(|\mathcal{P}| / B^{N(1-\mathsf{C})}\right)^{1/tk}$. Substituting $N = k\ell$ and $B = |E|/k\ell$ and simplifying, we get $\mathcal{S} \geq k\left(|\mathcal{P}|^{1/k\ell} / (|E|/k\ell)^{1-\mathsf{C}}\right)^{\ell/t}$, which is the lower bound stated in Theorem 11.

4.2. Step 1: A data structure implies a protocol for LSD

We now prove that the data structure can be translated to a protocol for LSD.

Lemma 12. *If $\ell < \frac{g(G)}{\alpha+1}$, then there exists a protocol for LSD with parameters N and B, where Alice sends $tk\lg(e\mathcal{S}/k)+N\lg(eB)+\lg(eBN)-\lg|\mathcal{P}(G,\ell,k)|$ bits and Bob sends ktw bits.*

Proof: We begin by defining a bijection from the universe $[NB]$ to E. For now, any bijection will do (we will change this later). Denote the bijection by $f:[NB]\to E$. Both Alice and Bob know G, so we can assume they both know f.

Now, in the LSD problem, Alice receives a set $S_A \subseteq [NB]$ of cardinality $|S_A| = N$, and Bob may receive any set $S_B \subseteq [NB]$. We now show a protocol for LSD based on the existence of the data structure. Bob will use his set S_B to construct a set of edges, $E' = E \setminus f(S_B)$. Namely, an edge $e \in E$ is in E' if and only if its corresponding element is *not* in Bob's set S_B. Bob will preprocess the graph $G' = (V, E')$ creating the data structure, and from now on, Bob will "play" the role of the data structure. Note that G' is a subgraph of G, so by the assumption of Theorem 11 this data structure has space $\mathcal{S}$, query time t, and it is an α-approximate distance oracle for G'.

Alice constructs the set $P = f(S_A)$. For now, assume that $P \in \mathcal{P}(G,\ell,k)$ and call this assumption the *perfect bijection scenario*; we shall remove this assumption later. Under this assumption, P can be written as the union of k vertex-disjoint paths, each of length ℓ. Let $(u_1,v_1),\ldots,(u_k,v_k)$ be the endpoints of these paths.

Now we continue describing the protocol for LSD, in the perfect bijection scenario. For every pair (u_i,v_i), we know that $d_G(u_i,v_i) = \ell$ (it cannot be smaller since then we would find a cycle of length $\leq 2\ell$). Denote the path of length ℓ between u_i and v_i by p. We see that, on one hand, an α-approximate distance query on the pair (u_i,v_i) will return $\tilde{d}(u_i,v_i) \leq \alpha\ell$ if all of the edges of the path p are in E'. On the other hand, the approximate distance query will return a number $\geq g(G)-\ell$ if at least one of the edges of p is not in E', since there are no cycles shorter than $g(G)$, and since an approximate distance oracle never returns an underestimate of the distance, but always an overestimate or the correct value. After querying all k distances $(u_1,v_1),\ldots,(u_k,v_k)$, if all of the k queries return distances at most $\leq \alpha\ell$, then we conclude that $S_A \cap S_B = \emptyset$, otherwise we conclude that $S_A \cap S_B \neq \emptyset$. Thus, all Alice and Bob need to do in order to get an answer to LSD is to simulate k queries to the data structure.

Now, still assuming the perfect bijection scenario, Alice and Bob simulate the data structure by communication. This is standard [23], [26], but we include the details for completeness: Bob computes the data structure itself, based on E'. Alice computes the set P and the pairs $(u_1,v_1),\ldots,(u_k,v_k)$. Alice then considers which cells should be probed in the first round of each of the k queries, and sends the set of probed cells to Bob. Note that we send it as a set, not one by one. This set can be sent using $\lg\binom{\mathcal{S}}{k}$ bits. Bob replies with the contents of these cells, using wk bits. Now Alice sends the set of cells that should be probed in the second round, using another $\lg\binom{\mathcal{S}}{k}$ bits, and then Bob replies, using another wk bits, and so on. Overall, Alice sends $t\lg\binom{\mathcal{S}}{k} \leq t\cdot\lg\left(\frac{e\mathcal{S}}{k}\right)^k = tk\lg(e\mathcal{S}/k)$ bits, and Bob sends wtk bits. We remark that it is important to send the set of cells to be probed as a set, taking $\lg\binom{\mathcal{S}}{k}$ bits, rather than one by one, taking $k\lg\mathcal{S}$ bits. The latter would cause the lower bound to become so weak as to be meaningless (the same is necessary in Pătraşcu's lower bound [27]).

Now we want to get rid of the perfect bijection assumption. To do this, we include a round of communication that is performed before starting the protocol. This round of communication will choose the bijection f such that the perfect bijection scenario holds for this particular f. Alice encodes the bijection and sends $\lg(\ln\left((eB)^N\right)\cdot\frac{(eB)^N}{|\mathcal{P}|})$ bits to Bob, who sends 0 bits to Alice. To do this, instead of having only one bijection $f:[NB]\to E$ at the start, Alice and Bob share knowledge of m bijections, $f_1,\ldots,f_m$, all from $[NB]$ to E. These bijections must have the property that for any set $S_A \subseteq [NB]$ of cardinality N, there exists an i such that choosing $f = f_i$ puts us in the perfect bijection scenario, that is, $\exists i \in [m] : f_i(S_A) = \mathcal{P}(G,\ell,k)$. If we find such a family of bijections, then Alice and Bob can get to the perfect bijection scenario by sending $\lg m$ bits – the index of the bijection they use – and then continue as before. Recall that Alice and Bob share all $f_1,\ldots,f_m$ beforehand, without exchanging any communication.

The existence of such a family of bijections is proved in the following claim using the probabilistic method. It is a tailored restatement of Pătraşcu [26, Lemma 11].

Claim 13. *There exists a set of bijections, $f_1,\ldots,f_m : [NB]\to E$, where $m = \ln\left((eB)^N\right)\cdot\frac{(eB)^N}{|\mathcal{P}|}$, such that for each $S_A \subseteq [NB]$ with $|S_A| = N$, there exists a bijection f_i such that $f_i(S_A) \in \mathcal{P}(G,\ell,k)$.*

Proof: Fix some S_A of cardinality N. Let h be a randomly-chosen bijection from $[NB]$ to E. The probability that $h(S_A) \in \mathcal{P}$ is $q = |\mathcal{P}|/\binom{|E|}{N} \geq |\mathcal{P}|/(eB)^N$.

Let $m' = \ln\left((eB)^N\right)$. Note that $e^{m'}$ is an upper bound on the number of sets $S_A \subseteq [NB]$ of cardinality N. Thus, consider $m = m'/q$ bijections selected uniformly at random. The probability that all of them fail to map S_A to an element of $\mathcal{P}$ is $(1-q)^{m'/q} < e^{-m'} = (eB)^{-N}$. By the union bound, we get that a set of m'/q randomly-chosen permutations has probability > 0 to have the property in the lemma. Thus there is such a set. ∎

Thus, we got an LSD protocol where Alice sends $\lg N + \lg\ln(eB) + N\lg(eB) - \lg|\mathcal{P}(G,\ell,k)|$ bits (in order to get to the perfect bijection scenario) and then she sends another $tk\lg(e\mathcal{S}/k)$ in the perfect bijection scenario, and Bob sends tkw bits. By using the inequality $\lg N + \lg\ln(eB) \leq \lg(eBN)$ the lemma follows. ∎

4.3. Step 2: The lower bound for LSD implies a lower bound for distance oracles

We have seen that a good data structure implies a good communication protocol. However, there is a lower bound on the communication problem LSD (Lemmas 6 and 7). We use it to derive a lower bound on the space usage of distance oracles.

Recall that if a protocol computes LSD with parameters N and B, then either Alice sends at least $\mathsf{C}N\lg B$ bits or Bob must send at least NB^{C} bits (Lemma 7).

We saw that Bob's communication is tkw. However, by the condition of Theorem 11, $B \geq (2tw/\ell)^{1/\mathsf{C}}$, so $NB^{\mathsf{C}} \geq k\ell\cdot 2tw/\ell = 2ktw$. Thus, Bob's communication is strictly less than NB^{C} bits. From the lower bound on LSD, it follows that Alice must communicate at least $\mathsf{C}N\lg B$ bits. Substituting all we know so far, this means that

$$tk\lg(e\mathcal{S}/k) + N\lg(eB) + \lg(eBN) - \lg(|\mathcal{P}|) \geq \geq \mathsf{C}N\lg B\ .$$

Use $N = k\ell$ and move terms around to get

$$\lg(e\mathcal{S}/k) \geq \lg\left((|\mathcal{P}|)^{1/tk}\right) + \frac{\mathsf{C}\ell}{t}\lg B - \frac{\ell}{t}\lg(eB) - \lg(eBN)/tk\ .$$

Exponentiating and moving terms around, we get

$$e\mathcal{S}/k \geq (|\mathcal{P}|)^{1/tk}(eB^{1-\mathsf{C}})^{-\ell/t}\cdot(eBN)^{-1/tk}\ ,$$

or equivalently,

$$\mathcal{S} \geq k\cdot\left(\frac{|\mathcal{P}|^{1/k\ell}}{eB^{1-\mathsf{C}}}\right)^{\ell/t}\cdot((eBN)^{-1/tk}/e)\ ,$$

which is what we wanted to prove.

5. Counting Paths in Regular High-Girth Expanders

To prove Theorem 1, we need to count paths in regular high-girth expanders. We use such expanders with degree $poly(wt\alpha/\lg n)$. The expression $wt\alpha/\lg n$ could be $o(1)$, in which case the degree will be $O(1)$.

For a reader who is not interested in expander graphs, or who is dealing with regular high-girth graphs that might not be expanders, Section 6 contains analogues of the results of this section, which are almost as strong, and do not require good expansion.

We are interested in lower-bounding the cardinality of $\mathcal{P}(G,\ell,k)$ when G is a high-girth expander.

Lemma 14. *Let $G = (V,E)$ be an r-regular graph and k,ℓ be two positive integers, such that the following three conditions hold: (i) $\lambda(G) \leq 0.1$; (ii) $|V| \geq 20k\ell$; (iii) $\ell < g(G)$, then*

$$|\mathcal{P}(G,\ell,k)| \geq \binom{|V|}{k}\cdot(r/8)^{k\ell}\ .$$

Proof: Denote $N = k\cdot\ell$. As a warm-up, we first count the number of ways to choose just one path. We know that ℓ is smaller than the girth of G, so there are exactly $|V|\cdot(r-1)^{\ell}/2$ ways to choose a path of length ℓ, since we can start at any vertex and we are allowed to progress to any of the neighbors, but not to backtrack. We divide by two since we counted each path twice: each path can be started from both of its endpoints.

To count collections of vertex-disjoint paths, we choose the k paths one by one and prove that, no matter which paths we already chose, there are at least $|V|\cdot(r/8)^{\ell}$ ways to choose the next one. The difficulty in proving this is that after choosing some of the paths, there are some vertices that we are not allowed to use anymore.

Suppose that we are constructing a member of $\mathcal{P}$ and we have already chosen some of the paths, and used the vertices $A \subseteq V$ (where $|A| < N + k \leq 2N$). We are not allowed to use these vertices again, since the paths need to be vertex-disjoint. We claim that:

Claim 15. *Let $G = (V,E)$ be an r-regular graph and k,ℓ be two positive integers, such that the following three conditions hold: (i) $\lambda(G) \leq 0.1$; (ii) $|V| \geq 20k\ell$; (iii) $\ell < g(G)$. Let $A \subseteq V$ be a set of vertices of cardinality $\leq 2N$. Then the number of different paths of length ℓ in G that do not use any vertices of A is at least $|V|\cdot(r/8)^{\ell}$.*

Before proving this claim, we show how it yields the lemma. By repeatedly using Claim 15, we get that the total number of ways to choose k paths of length ℓ each is at least

$$\left(|V|\cdot(r/8)^{\ell}\right)^{k}/k! = |V|^{k}\cdot(r/8)^{N}/k!\ .$$

The $k!$ factor in the denominator arises since we potentially counted each collection $k!$ times: we have chosen the paths in different orders, while the order of choosing the paths should not matter. So, the cardinality of $\mathcal{P}$ is at least

$$\frac{|V|^{k}\cdot(r/8)^{N}}{k!} \geq \binom{|V|}{k}\cdot(r/8)^{N}\ .$$

. Thus, if we prove Claim 15 then Lemma 14 follows. ∎

We now prove the claim.

Proof of Claim 15: The set A of disallowed vertices is of cardinality $\leq 2N$. Assume w.l.o.g. that A is of cardinality exactly $2N$ (this only makes our situation worse than it really is).

Now, choose a path of length ℓ in G by a random walk: start at a uniformly random vertex, and at each step, pick a random outgoing edge except the one we came through. Since the girth is larger than ℓ, then the path that we choose this way is a simple path. There are exactly $|V|\cdot(r-1)^\ell/2 \geq |V|\cdot(r/4)^\ell$ paths that can be produced this way and each one is produced exactly twice (once from each direction), so we just need to estimate the probability that this path manages to avoid A. It is enough to prove that this probability is $\geq 1/2^\ell$.

If the $\ell+1$ vertices in the path were chosen uniformly and independently from V, then this would be trivial. Fortunately, since G is an expander, then a random walk on it "behaves like" a set of independently-chosen vertices in various aspects, including the aspect we need, as shown by Alon et al. [2]:

Lemma 16 (Alon et al. [2])**.** *Let $G=(V,E)$ be an r-regular graph. Let $W_1,\ldots,W_{\ell+1}\subset V$ be some sets of vertices (that may be equal to each other), each of cardinality at least $\mu|V|$, and suppose that $\mu \geq 6\lambda(G)$. The probability that a random walk of length ℓ stays inside $W_1,W_2,\ldots,W_{\ell+1}$ is at least $\mu(\mu-2\lambda(G))^{\ell-1}$.*

In our case, set all W_i's equal to $V\setminus A$. Set $\mu=0.9$. We know that $\lambda(G)\leq 0.1$ so the conditions of Lemma 16 hold, and thus the probability that the walk stays inside $V\setminus A$ for all of the steps is at least $0.9\cdot 0.7^{\ell-1}\geq 1/2^\ell$. ■

5.1. Lower Bound for Base-Graphs that are High-Girth Regular Expanders

We now combine Lemma 14 with Theorem 11 to get a lower bound for any regular expander base-graph, based only on its degree and girth. The following theorem may be of independent interest.

Theorem 17. *Let $G=(V,E)$ be an r-regular expander graph with $|V|=n$ vertices. Suppose there is a distance oracle for a base-graph G that uses space $\mathcal{S}$, query time t, and achieves stretch α. Assume $w=n^{o(1)}$, $g(G)\geq 2\alpha$, $\lambda(G)\leq 0.1$, and $r\geq (4tw\alpha/g(G))^{1/\mathsf{C}}$. Then, $\mathcal{S}\geq nr^{\Omega(g(G)/\alpha t)}/\lg n$.*

Proof: The proof requires some tedious calculations, and not much else. We assume $w=n^{o(1)}$, but in fact the calculations below can be generalized to even larger word size: up to $w=n^c$ where c is some small constant.

Let $\ell=\lfloor g(G)/2\alpha\rfloor \leq \lg n$. Let $k=|V|/20\ell$.

By Lemma 14, it holds that $|\mathcal{P}(G,\ell,k)|\geq (r/8)^{k\ell}$. To apply Theorem 11, we need to check that $|E|\geq k\ell(2tw/\ell)^{1/\mathsf{C}}$. This can be seen by $|E|=|V|\cdot r/2 = 10k\ell\cdot r \geq 10\ell k\cdot(2tw/\ell)^{1/\mathsf{C}}$. So, we get that the conditions for Theorem 11 hold.

Now we apply Theorem 11 and get that

$$\mathcal{S}\geq k\cdot\left(\frac{|\mathcal{P}(G,\ell,k)|^{1/k\ell}}{e(|E|/k\ell)^{1-\mathsf{C}}}\right)^{\ell/t}/((e\,|E|)^{1/tk}e)\ .$$

Now, $|E|\leq |V|^2\leq (\ell k)^4\leq k^8$. Thus, the term $(e\,|E|)^{1/tk}e$ in the bound is $\Theta(1)$, and we can ignore it. Furthermore $k\geq n/\lg n$, and we know that $|\mathcal{P}(G,\ell,k)|^{1/k\ell}\geq r/8$. Dropping constants where they are irrelevant, we get that

$$\mathcal{S}\geq \frac{n}{\lg n}\cdot\left(\frac{r}{(|E|/k\ell)^{1-\mathsf{C}}}\right)^{\Omega(\ell/t)}\ .$$

Now, $(|E|/k\ell)^{1-\mathsf{C}}=(10r)^{1-\mathsf{C}}$. Thus we get

$$\mathcal{S}\geq \frac{n}{\lg n}\cdot\left(\frac{r}{r^{1-\mathsf{C}}}\right)^{\Omega(\ell/t)}\ .$$

Substituting $\ell=\Theta(g(G)/\alpha)$, we get the theorem. ■

5.2. Proof of the Main Theorem

We now combine the LPS construction with Theorem 17 to get the main result of the paper.

Theorem 18. *Let $\mathcal{S}(n)$, $\alpha(n)$, $t(n)$, $w(n)$, be such that there exists a distance oracle for any graph with n vertices, which has stretch α, uses query time t, and space $\mathcal{S}$ with word-size w. Assume $\alpha=o\left(\frac{\lg n}{\lg(w\lg n)}\right)$. Then $\mathcal{S}\geq n^{1+\Omega(1/t\alpha)}/\lg n$. This even holds when restricted to graphs with maximum degree $poly(tw\alpha/\lg n)$.*

Proof: Let n_0 be some asymptotically-large number. Let $r_0=\max\{(8tw\alpha/\lg n_0)^{2/\mathsf{C}},400,(4/\mathsf{C})^{4/\mathsf{C}}\}$. Apply Theorem 9 to these two numbers. This gives a graph $G=(V,E)$ with:

- $n_0\leq |V|\leq 8n_0$
- G is r-regular where $r_0\leq r\leq 2r_0$
- $g(G)\geq \frac{1}{2}\lg_{r_0} n_0$
- $\lambda(G)\leq 0.1$ (since $r_0\geq 400$).

Apply Theorem 17 to this graph. Then

$$\mathcal{S}\geq n_0 r_0^{\Omega(g(G)/\alpha t)}/\lg(2n_0)\ .$$

Using the fact that $r_0^{g(G)}\geq\sqrt{n_0}$ and that $n=\Theta(n_0), r=\Theta(r_0)$ we get the theorem.

We need to check that Theorem 17 can indeed be applied. This is somewhat tedious. $\lambda(G)\leq 0.1$ is known from the construction of G. $w=n^{o(1)}$ holds because if $w=n^{\Omega(1)}$ then the condition $\alpha=o\left(\frac{\lg n}{\lg(w\lg n)}\right)$ implies $\alpha=o(1)$.

To see that $g(G)\geq 2\alpha$, begin by observing that $\alpha\leq\lg n$ because of the condition $\alpha=o\left(\frac{\lg n}{\lg(w\lg n)}\right)$. Furthermore, we can assume that $t\leq\lg n$, otherwise the term $n^{1/t\alpha}$ is equal to $\Theta(1)$ and we could then derive the bound $\mathcal{S}\geq\Omega(n)$ by setting $\ell=1,k=n/20,r=w$ and proceeding as before (or just using the Thorup-Zwick proof technique).

From $\alpha \leq \lg n$ and $t \leq \lg n$ we get $r_0 \leq poly(w \lg n)$. We then see that $g(G) \geq \Omega(\lg_{r_0} n_0) \geq \Omega\left(\frac{\lg n}{\lg(w \lg n)}\right) > \alpha$.

Finally, to see that $r \geq (4tw\alpha/g(G))^{1/\mathsf{C}}$, observe that $r \geq (4/\mathsf{C})^{4/\mathsf{C}}$ and for any such r it holds that $(\lg r)^{1/\mathsf{C}} \leq \sqrt{r}$. Now,

$$\begin{aligned}(4tw\alpha/g(G))^{1/\mathsf{C}} &\leq (8tw\alpha \lg r/\lg n_0)^{1/\mathsf{C}} \\ &\leq (8tw\alpha/\lg n_0)^{1/\mathsf{C}} \cdot (\lg r)^{1/\mathsf{C}} \\ &\leq (8tw\alpha/\lg n_0)^{1/\mathsf{C}} \cdot \sqrt{r} \\ &\leq \sqrt{r} \cdot \sqrt{r} = r.\end{aligned}$$

We thus proved that all conditions of Theorem 17 hold. ■

6. Counting Paths in Regular High-Girth Graphs

In the proof of Theorem 18 we used expander graphs with degree $poly(wt\alpha/\lg n)$. Instead of this, we can prove a similar theorem with a slightly worse bound on the degree (namely, $poly(w, \lg n)$), but without using expansion properties. This section follows the same outline as Section 5.

Lemma 19. *Let $G = (V, E)$ be an r-regular graph and k, ℓ be two positive integers, such that the following two conditions hold: (i) $|V| \geq 8k\ell^2$; (ii) $\ell < g(G)$, then*

$$|\mathcal{P}(G, \ell, k)| \geq \binom{|V|}{k} \cdot (r-1)^{k\ell}/4^k \ .$$

The proof of this lemma is very similar to the proof of Lemma 14, except that Claim 15 is replaced by Claim 20. Thus, we skip the proof, and just state and prove Claim 20.

Claim 20. *Let $G = (V, E)$ be an r-regular graph and k, ℓ be two positive integers, such that the following two conditions hold: (i) $|V| \geq 8k\ell^2$; (ii) $\ell < g(G)$. Let $A \subseteq V$ be a set of vertices of cardinality $\leq 2N$. Then the number of different paths of length ℓ in G that do not use any vertices of A is at least $|V| \cdot (r-1)^\ell/4$.*

Proof: The set A of disallowed vertices is of cardinality $\leq 2N$. Assume w.l.o.g. that A is of cardinality exactly $2N$ (this only makes our situation worse than it really is).

Now, choose a path of length ℓ in G by a random walk: start at a uniformly random vertex, and at each step, pick a random outgoing edge except the one we came through. Since the girth is larger than ℓ, the path chosen this way is a simple path. There are exactly $|V| \cdot (r-1)^\ell/2$ paths that can be produced this way and each one is produced exactly twice (once from each direction), so we just need to estimate the probability that this path manages to avoid A. It is enough to prove that this probability is $\geq 1/2$.

Think of the path as a sequence of $\ell+1$ random variables. The graph is an undirected regular graph, thus each of these random variables is uniformly distributed over V, but they depend on each other of course.[2] We can use the union-bound to see that the probability that none of these vertices are in A is at least $1 - (\ell+1) \cdot \frac{|A|}{|V|} \geq 1 - 2\ell \cdot \frac{2k\ell}{8k\ell^2} = \frac{1}{2}$, and this concludes the proof. ■

6.1. Lower Bound for High-Girth Regular Base-Graphs

We can now combine Lemma 19 with Theorem 11 to get a lower bound for any regular base-graph based only on its degree and girth. The following theorem may be of independent interest.

Theorem 21. *Let $G = (V, E)$ be an r-regular graph with $|V| = n$ vertices. Suppose there is a distance oracle for base-graph G that uses space $\mathcal{S}$, query time t, and achieves stretch α. Assume $w = n^{o(1)}$, $g(G) \geq 2\alpha$, and $r \geq (4twg(G)/\alpha)^{2/\mathsf{C}}$. Then, $\mathcal{S} \geq nr^{\Omega(g(G)/\alpha t)}/\lg^2 n$.*

We skip the proof of this theorem due to space restrictions. Its proof is very similar to the proof of Theorem 17.

Using Theorem 21 together with the LPS construction from Theorem 9, we can get the following theorem, which is similar to the main result of this paper. This theorem has somewhat weaker parameters than the main results and the proof is similar to that of Theorem 18.

Theorem 22. *Let $\mathcal{S}(n)$, $\alpha(n)$, $t(n)$, $w(n)$, be such that there exists a distance oracle for any graph with n vertices, which has stretch α, uses query time t, space $\mathcal{S}$, and word-size w. Assume $\alpha = o\left(\frac{\lg n}{\lg(w \lg n)}\right)$. Then $\mathcal{S} \geq n^{1+\Omega(1/t\alpha)}/\lg^2 n$. This even holds when restricted to graphs with maximum degree $poly(tw)$.*

Acknowledgments

We thank the anonymous referees for their helpful comments and suggestions and, in particular, we thank one anonymous FOCS referee for preventing us from a major mistake. The first author thanks the Institute for Theoretical Computer Science at Tsinghua University for their invitation and their very kind hospitality. The second author would like to thank Kristoffer Arnsfelt Hansen for referring him to Lemma 16. The third author would like to thank Pinyan Lu for helpful discussions.

[2] Why are they uniformly distributed? Imagine placing r tokens at each vertex, and then move each token in each vertex to a different neighbor of the vertex; then there are still r tokens at each vertex. This shows that if the distribution is uniform at step i then it would be uniform at step $i+1$, and we proceed by induction.

REFERENCES

[1] Noga Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986.

[2] Noga Alon, Uriel Feige, Avi Wigderson, and David Zuckerman. Derandomized graph products. *Computational Complexity*, 5(1):60–75, 1995.

[3] Noga Alon, Shlomo Hoory, and Nathan Linial. The Moore bound for irregular graphs. *Graphs and Combinatorics*, 18(1):53–57, 2002.

[4] Ingo Althöfer, Gautam Das, David P. Dobkin, Deborah Joseph, and José Soares. On sparse spanners of weighted graphs. *Discrete & Computational Geometry*, 9:81–100, 1993.

[5] Alexandr Andoni, Piotr Indyk, and Mihai Patrascu. On the optimality of the dimensionality reduction method. In *47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006), 21-24 October 2006, Berkeley, California, USA, Proceedings*, pages 449–458, 2006.

[6] Surender Baswana, Akshay Gaur, Sandeep Sen, and Jayant Upadhyay. Distance oracles for unweighted graphs: Breaking the quadratic barrier with constant additive error. In *Automata, Languages and Programming, 35th International Colloquium, ICALP 2008, Reykjavik, Iceland, July 7-11, 2008, Proceedings, Part I: Tack A: Algorithms, Automata, Complexity, and Games*, pages 609–621, 2008.

[7] Surender Baswana and Sandeep Sen. Approximate distance oracles for unweighted graphs in expected $O(n^2)$ time. *ACM Transactions on Algorithms*, 2(4):557–577, 2006.

[8] Robert Breusch. Zur Verallgemeinerung des Bertrandschen Postulates, daß zwischen x und $2x$ stets Primzahlen liegen. *Mathematische Zeitschrift*, 34(1):505–526, 1932.

[9] Shiva Chaudhuri and Christos D. Zaroliagis. Shortest paths in digraphs of small treewidth. part I: Sequential algorithms. *Algorithmica*, 27(3):212–226, 2000. Announced at ICALP 1995.

[10] Giuliana Davidoff, Peter Sarnak, and Alian Valette. Elementary Number Theory, Group Theory, and Ramanujan Graphs. Cambridge University Press, 2003.

[11] Michael Elkin. Sparse graph spanners. In *Encyclopedia of Algorithms*. 2008.

[12] Paul Erdős. Über die Primzahlen gewisser arithmetischer Reihen. *Mathematische Zeitschrift*, 39(1):473–491, 1935.

[13] Paul Erdős. Extremal problems in graph theory. *Theory of graphs and its applications, Proceedings of the Symposium in Smolenice (Prague)*, pages 29–36, 1964.

[14] Paul Erdős and Horst Sachs. Reguläre Graphen gegebener Taillenweite mit minimaler Knotenzahl. *Wissenschaftliche Zeitschrift der Martin-Luther-Universität Halle-Wittenberg. Mathematisch-Naturwissenschaftliche Reihe*, pages 251–258, 1963.

[15] Joachim Gudmundsson, Christos Levcopoulos, Giri Narasimhan, and Michiel H. M. Smid. Approximate distance oracles for geometric spanners. *ACM Transactions on Algorithms*, 4(1), 2008.

[16] Shlomo Hoory. On graphs of high girth. *PhD thesis, Hebrew University*, 2002.

[17] Philip N. Klein. Preprocessing an undirected planar network to enable fast approximate distance queries. In *Proceedings of the Thirteenth Annual ACM-SIAM Symposium on Discrete Algorithms, January 6-8, 2002, San Francisco, CA, USA. ACM/SIAM, 2002*, pages 820–827, 2002.

[18] Philip N. Klein. Multiple-source shortest paths in planar graphs. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2005, Vancouver, British Columbia, Canada, January 23-25, 2005*, pages 146–155, 2005.

[19] Alexander Lubotzky, Ralph Phillips, and Peter Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[20] 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.

[21] Manor Mendel and Chaya Schwob. C-K-R partitions of sparse graphs. *CoRR*, abs/0809.1902, 2008.

[22] Peter Bro Miltersen. Cell probe complexity - a survey. *Invited talk/paper at Advances in Data Structures (Pre-conference workshop of FSTTCS)*, 1999.

[23] Peter Bro Miltersen, Noam Nisan, Shmuel Safra, and Avi Wigderson. On data structures and asymmetric communication complexity. *Journal of Computer and System Sciences*, 57(1):37–49, 1998.

[24] Pieter Moree. Bertrand's Postulate for primes in arithmetical progressions. *Computers & Mathematics with Applications*, 26(5):35–43, 1993.

[25] Mihai Patrascu. Lower bounds for 2-dimensional range counting. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing, San Diego, California, USA, June 11-13, 2007*, pages 40–46, 2007.

[26] Mihai Patrascu. (Data) STRUCTURES. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 434–443, 2008.

[27] Mihai Patrascu. Unifying the Landscape of Cell-Probe Lower Bounds. submitted, 2009.

[28] Mikkel Thorup. Compact oracles for reachability and approximate distances in planar digraphs. *Journal of the ACM*, 51(6):993–1024, 2004. Announced at FOCS 2001.

[29] Mikkel Thorup and Uri Zwick. Approximate distance oracles. *Journal of the ACM*, 52(1):1–24, 2005. Announced at STOC 2001.

[30] Liam Roditty, Mikkel Thorup, and Uri Zwick. Deterministic constructions of approximate distance oracles and spanners. In *Automata, Languages and Programming, 32nd International Colloquium, ICALP 2005, Lisbon, Portugal, July 11-15, 2005, Proceedings*, pages 261–272, 2005.

[31] David P. Woodruff. Lower bounds for additive spanners, emulators, and more. In *47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006), 21-24 October 2006, Berkeley, California, USA, Proceedings*, pages 389–398, 2006.

[32] Andrew Chi-Chih Yao. Should tables be sorted? *Journal of the ACM*, 28(3):615–628, 1981.