Distance oracles for timetable graphs - research

František Hajnovič
Faculty of Mathematics, Physics and Informatics
Comenius University
Bratislava, Slovakia
ferohajnovic@gmail.com

November 2011

Abstract

In this paper, we would like to summarize the work done on distance oracles, shortest path queries and shortest timetable connection queries. The list of papers is not comprehensive, but offers a good look into the state of research in this area.

Keywords: **distance oracles**, **timetable graphs**, **shortest path**

---

1 List of main definitions

1. **?? Highway dimension** - bla bla

2 List of main results

1. **?? Thorup and Zwick compromises** - bla bla

3 <List of main DO methods

1. **?? Voronoi duals** - bla bla

4 List of considered papers

Main papers:

1. **??** Highway Dimension, Shortest Paths, and Provably Efficient Algorithms [**?**]
2. **??** Approximate Distance Oracles [**?**]
3. **??** Approximate Shortest Path and Distance Queries in Networks [**?**]
4. **??** Distance Labeling in Graphs [**?**]

Other - distance oracles related:

1. **??** A Sketch-Based Distance Oracle for Web-Scale Graphs [**?**]
2. **??** Distance Oracles for Sparse Graphs [**?**]

3. **??** Ramsey Partitions Based Approximate Distance Oracles [?]
4. **??** A Compact Routing Scheme and Approximate Distance Oracle for Power-Law Graphs [?]
5. **??** Improved Distance Oracles for Avoiding Link-Failure [?]
6. **??** Distance Oracles for Spatial Networks [?]
7. **??** f-Sensitivity Distance Oracles and Routing Schemes [?]
8. **??** Exact Distance Oracles for Planar Graphs [?]
9. **??** Reachability and distance queries via 2-hop labels [?]

Other - timetable related

1. **??** Using Multi-level Graphs for Timetable Information in Railway Systems [?]
2. **??** Engineering Time-Expanded Graphs for Faster Timetable Information [?]
3. **??** Timetable Information: Models and Algorithms [?]
4. **??** Experimental comparison of shortest path approaches for timetable [?]

5 Paper summaries - main

5.1 Approximate Shortest Path and Distance Queries in Networks

Author(s): Christian Sommer
Year: 2010
Institution(s): The University of Tokyo

In short

Concern:
In his PhD thesis, Christian Sommer investigates the problem of *efficiently computing exact and approximate shortest path queries* in graphs. Apart from that, the thesis gives a clear understanding of the problematics, terminology and related work in the area.
Main definitions:
Main results:
Main DO methods:

In short

The thesis provides three *main results*:

1. $\mathfrak{RESULT}$There is no exact DO for sparse graphs, that would be of size $O(m)$ and have a constant query time. More specifically, given query time t and multiplicative stretch α, a required space for an exact DO is $n^{1+\Omega(1/\alpha t)}$.
2. Adapting the DO of Thorup and Zwick ([?]) and adjusting it for power-law graphs. The adjustment takes advantage of the high-degree nodes in power-law graphs and selects them for *landmarks* in the algorithm. A theoretical proof shows why such an approach yields a good heuristics, efficiently approximating shortest paths e.g. in Internet-like topologies.
3. An approximation DO for general undirected graphs with positive edge weights is presented. Based on random sampling ang graph Voronoi duals, it provides good theoretical trade-offs between stretch and preprocessing and query times. Even better is the performance in practice. Compared to DO of Thorup and Zwick, it offers a different type of trade-off: stretch of the answer is reduced when a longer query time is allowed.

5.2 Highway Dimension, Shortest Paths, and Provably Efficient Algorithms

Author(s):

- Ittai Abraham
- Amos Fiat
- Andrew V. Goldberg
- Renato F. Werneck

Year: 2010

Institution(s):

- Microsoft Research
- Tel-Aviv University

Introducing a notion of a highway dimension, this paper established a parameter which considerably influences the efficiency of several *exact* methods for shortest path computing in general graphs. These, while showing very good results in practice, lacked theoretical guarantees on their time complexities. In the results of this paper, low highway dimension was shown to bind the time complexities of the methods to practical values.

5.3 Approximate Distance Oracles

Author(s):
- Mikkel Thorup
- Uri Zwick

Year: 2004

Institution(s):
- AT&T Labs Research
- Tel-Aviv University

This paper by Thorup and Zwick carries *one of the most important results* in the sphere of distance oracles, further used in other works in this area. In particular they have shown that given an undirected weighted graph of n vertices and m edges and a chosen integer $k \geq 1$, we can build a distance oracle answering shortest path queries in the following manner:
- preprocessing takes $O(kmn^{1/k})$ expected time
- resulting distance oracle is of size $O(kn^{1+1/k})$
- answering queries takes $O(k)$ time
- stretch of the anwser(i.e. the worst ratio of returned path against the optimal value) is $2k - 1$ at most

First, they show a simple randomized algorithm (later derandomized) for metric spaces, which is then analyzed to show that it satisfy the mentioned bounds. An adjustnemnt is introduced later on, which no longer relies on a metric being provided explicitly.

5.4 Distance Labeling in Graphs

Author(s):

- Cyril Gavoille
- David Peleg
- Stéphane Pérennes
- Ran Raz

Year: 2004

Institution(s):

- Université Bordeaux
- The Weizmann Institute of Science, Israel
- Université de Nice-Sophia Antipolis

The paper considers the problem of *distance labeling in graphs*. Distance labeling itself is an approach that assigns each node of the graph some label in such a way, that queries for distance between any pair of vertices can be answered using only the information provided by the labels.

This paper talks about various lower and upper bounds on the length of such labels for individual classes of graphs, as well as *time complexity of the extraction of the distance* from the labels.

Distance labeling $< L, f >$ for a graph G consists of **node labeling** L and **distance decoder** f. The first is a function that assigns each node a label. Second is a function that computes the distance between the nodes, given their labels. $< L, f >$ is a **Distance labeling scheme** for family (class) of graphs $\mathcal{G}$ if it is a correct distance labeling for every $G \in \mathcal{G}$.

Now we are interested in such a distance labeling scheme, that minimizes the maximal label length in all graphs of the given graph family. Following results of that kind were obtained:

Table 1: Results of [?] on the necessary length of labels. In each entry we consider a n-vertex subclass of the given class

Class of graphs	**Lower bound**	**Upper bound**
general	$n/2 - O(1) = \omega(n)$	$11n + O(\log n \log\log n) = O(n)$
unweighted binary trees	$\log^2 n/8 - O(\log n) = O(\log^2 n)$	
binary trees with integral weights up to M	$\theta((\log M + \log n) \log n)$	$\theta((\log M + \log n) \log n)$
planar	$\omega(n^{1/3})$	
bounded degree	$\omega(\sqrt{n})$	
having $r(n) - separator$	$O(R(n) \log n + \log^2 n)$	

In the last row, $R(n) = \sum_{i=0}^{\log_{3/2} n} r(n(2/3)^i)$. We say a graph class $\mathcal{G}$ has a $\boldsymbol{r(n) - separator}$ if for every connected graph $G \in \mathcal{G}$ of n vertices there is a separator of size at most $r(n)$ such that upon its deletion we obtain graphs from $\mathcal{G}$. Moreover, all of these created components have size at most $r(2n/3)$.

approximations

Other results are concerning the time complexity of the distance decoder. The main result is rather technical, so we'll demonstrate it on a simpler corollary:

Let $\mathcal{G}$ be the family of all graphs. For infinitely many n, we have a $G \in \mathcal{G}$, for which:

1. There is an exact (with stretch 1) distance labeling scheme with maximal label $\leq 3 \log n + o(\log n)$.
2. However, for any distance labeling scheme with stretch ł2 that satisfies $\sum_{u \in V(G)} |L(u, G)| \leq n^2/2 - O(n \log n)$ (that is also the one from point 1.)), the time and space complexities of distance decoder function f are larger than any constant size stack of exponentials ($2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}$ - constant number of times).

Put informaly, for almost any size of graph, there is such graph for which we can find a labeling producing only short labels. But for that graph, there will not be any distance labeling scheme that would answer distance queries in a small time/space complexity.

6 Paper summaries - distance oracles related

6.1 A Sketch-Based Distance Oracle for Web-Scale Graphs

Author(s):
- Atish Das Sarma
- Sreenivas Gollapudi
- Marc Najork
- Rina Panigrahy

Year: 2010
Institution(s):
- Microsoft Research
- Georgia Institute of Technology

This paper considers real-world large-scale graphs and tries to provide algorithms to answer shortest path queries that would be efficient and providing accurate enough results at the same time.

The provided algorithm looks the following in the preprocessing phase:

$$SAMPLING \xrightarrow{\log|V|\ times} \forall u \in V\ SAMPLE[u] \xrightarrow{k\ times} \forall u \in V\ SKETCH[u]$$

$\boldsymbol{SAMPLING}$ is just picking random vertices (seeds) from V. $\boldsymbol{SAMPLE[u]}$ for some $u \in V$ is a set of couples (w_i, δ_i) where w_i is the closest seed to u from a given sampling, and δ_i is its distance. Sampling is done $\log|V|$ times - starting with 1 sampled seed and doubling each subsequent time. Finally, $\boldsymbol{SKETCH[u]}$ is a union of $SAMPLE[u]$ created when we iterated the whole process k times.

The **preprocessing** thus takes $O(k \cdot n^2)$, using one breath-first search from each seed set.
The algorithm $\boldsymbol{ONLINE - COMMON - SEED(x, y)}$ for answering queries compares the two $SKETCH$es of the two vertices and finds a commond seed s, which minimizes the distance $d(x,s) + d(s,y)$. In case of no common seed, we output inf. This algorithm always gives upper bound on actual shortest distance.

Further, there is a proof that for $k = \Theta(n^{1/c} polylog(n))$, we get a $2c - 1$ approximation of the actual distance with high probability.

Another algorithm called $\boldsymbol{ONLINE - BOURGAIN(x, y)}$ provides a lower bound on the actual distance. Unlike $ONLINE - COMMON - SEED$, it takes a maximum of $|d(x,S) - d(y,S)|\ \forall S$ where S are seed sets from $SAMPLING$. Again, a similiar theorem applies: for $k = \Theta(n^{1/c})$, we get a $O(2c - 1)$ approximation of the actual distance with high probability.

Both algorithms answer queries with time complexity $O(\log n)$.

Algorithms are also modified to directed graphs.

Experiments were run on a large crawl of the web graph, containing about 65000000 web pages and 420000000 distinct URLs. For $k = 1$, thus only $\log n$ guaranteed approximatio, they still obtained 1.2 approximation ratio with $ONLINE - COMMON - SEED$ and 2.14 ratio with $ONLINE - BOURGAIN$, thus showing that the theoretical guarantee could be possibly improved.

6.2 Distance Oracles for Sparse Graphs

Author(s):
- Christian Sommer
- Elad Verbin
- Wei Yu

Year: 2009

Institution(s):
- The University of Tokyo and NII
- Tsinghua University

6.3 Ramsey Partitions Based Approximate Distance Oracles

Author(s): Chaya Fredman
Year: 2008
Institution(s): The Open University Of Israel

6.4 A Compact Routing Scheme and Approximate Distance Oracle for Power-Law Graphs

Author(s):

- Wei Chen
- Christian Sommer
- Shang-Hua Teng
- Yajun Wang

Year:

Institution(s):

- Microsoft Research
- MIT

6.5 Improved Distance Oracles for Avoiding Link-Failure

Author(s):

- Rezaul Alam Chowdhury
- Vijaya Ramachandran

Year: 2002

Institution(s): The University of Texas at Austin

Paper attempts to solve the problem of preprocessing an edge-wieghted **directed** graph to answer queries for shortest path avoiding specific link.

Two functions are considered:

- $distance(x, y, u, v)$ - shortest distance from x to y avoiding edge (u, v).
- $path(x, y, u, v)$ - shortest path from x to y avoiding edge (u, v).

Assumption is that the time between two successive link failures are long enough to compute a new data structure in the background.

They build upon algorithms $DT - 1$ and $DT - 2$ presented in paper by Demetrescu and Thorup. Those are not very complicated - $DT - 1$ divides each shortest path from x to y to $O(\log n)$ segments, storing shortest distance from x to y avoiding each of these segments in a table. $DT - 2$ divides each shortest-path tree $T(x)$ into $O(\sqrt{n})$ bands of independent paths. Again, shortest distance avoiding these bands are precomputed.

The algorithms introduced in this paper are called $CR - 1$ and $CR - 2$. Their respective properties are shown in the table below.

Table 2: Results of [**?**]

Algorithm	Preprocessing time	Space	Query time
$CR - 1$	$O(mn^2 \log n + n^3 \log^2 n)$	$O(n^2 \log n)$	$O(1)$
$CR - 2$	$O(mn \log^2 n + n^2 \log^3 n)$	$O(n^2 \log^2 n)$	$O(\log n)$

6.6 Distance Oracles for Spatial Networks

Author(s):

- Jagan Sankaranarayanan
- Hanan Samet

Year: 2009

Institution(s): University of Maryland

6.7 f-Sensitivity Distance Oracles and Routing Schemes

Author(s):
- Shiri Chechik
- Michael Langberg
- David Peleg
- Liam Roditty

Year: 2011

Institution(s):
- The Weizmann Institute of Science
- Open University of Israe
- Bar-Ilan University

6.8 Exact Distance Oracles for Planar Graphs

Author(s):

- Shay Mozes
- Christian Sommer

Year: 2011

Institution(s):

- Brown University
- MIT

6.9 Reachability and distance queries via 2-hop labels

Author(s):

- Edith Cohen
- Eran Halperin
- Haim Kaplan
- Uri Zwick

Year: 2002

Institution(s):

- AT&T Labs Research
- Tel-Aviv University

7 Paper summaries - timetable graphs related

7.1 Using Multi-level Graphs for Timetable Information in Railway Systems

Author(s):
- Frank Schulz
- Dorothea Wagner
- Christos Zaroliagis

Year: 2002
Institution(s):
- University of Konstanz
- University of Patras

7.2 Engineering Time-Expanded Graphs for Faster Timetable Information

Author(s):

- Daniel Delling
- Thomas Pajor
- Dorothea Wagner

Year: 2008

Institution(s): University of Karlsruhe

7.3 Timetable Information: Models and Algorithms

Author(s):

- Matthias Muller-Hannemann
- Frank Schulz
- Dorothea Wagner
- Christos Zaroliagis

Year: 2006

Institution(s):

- Darmstadt University of Technology
- University of Karlsruhe
- University of Patras

7.4 Experimental comparison of shortest path approaches for timetable

Author(s):

- Evangelia Pyrga
- Frank Schulz
- Dorothea Wagner
- Christos Zaroliagis

Year: 2004

Institution(s):

- University of Karlsruhe
- University of Patras

References

[AFGW10] Ittai Abraham, Amos Fiat, Andrew V. Goldberg, and Renato Fonseca F. Werneck. Highway dimension, shortest paths, and provably efficient algorithms. In Moses Charikar, editor, *SODA*, pages 782–793. SIAM, 2010.

[CHKZ03] Edith Cohen, Eran Halperin, Haim Kaplan, and Uri Zwick. Reachability and distance queries via 2-hop labels. *SIAM J. Comput.*, 32(5):1338–1355, 2003.

[CLPR10] Shiri Chechik, Michael Langberg, David Peleg, and Liam Roditty. f-sensitivity distance oracles and routing schemes. In Mark de Berg and Ulrich Meyer, editors, *Algorithms – ESA 2010*, volume 6346 of *Lecture Notes in Computer Science*, pages 84–96. Springer Berlin / Heidelberg, 2010. ISBN 978-3-642-15774-5.

[CR02] Rezaul Alam Chowdhury and Vijaya Ramachandran. Improved distance oracles for avoiding link-failure. In Prosenjit Bose and Pat Morin, editors, *ISAAC*, volume 2518 of *Lecture Notes in Computer Science*, pages 523–534. Springer, 2002. ISBN 3-540-00142-5.

[CSTW09] Wei Chen, Christian Sommer, Shang-Hua Teng, and Yajun Wang. A compact routing scheme and approximate distance oracle for power-law graphs. *Distributed Computing*, ?(Disc 2009):379–391, 2009. URL `http://research.microsoft.com/pubs/81716/msr-tr-2009-84.pdf`.

[DPW09] Daniel Delling, Thomas Pajor, and Dorothea Wagner. Engineering time-expanded graphs for faster timetable information. In Ravindra Ahuja, Rolf Möhring, and Christos Zaroliagis, editors, *Robust and Online Large-Scale Optimization*, volume 5868 of *Lecture Notes in Computer Science*, pages 182–206. Springer Berlin / Heidelberg, 2009. ISBN 978-3-642-05464-8.

[Fre08] Chaya Fredman. *Ramsey Partitions Based Approximate Distance Oracles.* PhD thesis, The Open University Of Israel, 2008.

[GPPR04] Cyril Gavoille, David Peleg, Stéphane Pérennes, and Ran Raz. Distance labeling in graphs. *Journal of Algorithms*, 53(1):85 – 112, 2004. ISSN 0196-6774. URL `http://www.sciencedirect.com/science/article/pii/S0196677404000884`.

[MHSWZ07] Matthias Müller-Hannemann, Frank Schulz, Dorothea Wagner, and Christos Zaroliagis. *Algorithmic Methods for Railway Optimization*, volume 4359 of *Lecture Notes in Computer Science*, chapter Timetable Information: Models and Algorithms, pages 67 – 90. Springer, 2007.

[MS10] Shay Mozes and Christian Sommer. Exact shortest path queries for planar graphs using linear space. *CoRR*, abs/1011.5549, 2010.

[PSWZ04] Evangelia Pyrga, Frank Schulz, Dorothea Wagner, and Christos D. Zaroliagis. Experimental comparison of shortest path approaches for timetable information. In Lars Arge, Giuseppe F. Italiano, and Robert Sedgewick, editors, *ALENEX/ANALC*, pages 88–99. SIAM, 2004. ISBN 0-89871-564-4.

[SGNP10] Atish Das Sarma, Sreenivas Gollapudi, Marc Najork, and Rina Panigrahy. A sketch-based distance oracle for web-scale graphs. In Brian D. Davison, Torsten Suel, Nick Craswell, and Bing Liu, editors, *WSDM*, pages 401–410. ACM, 2010. ISBN 978-1-60558-889-6.

[Som10] Christian Sommer. *Approximate Shortest Path and Distance Queries in Networks.* PhD thesis, Graduate School of Information Science and Technology, The University of Tokyo, 2010.

[SS09] Jagan Sankaranarayanan and Hanan Samet. Distance oracles for spatial networks. In *ICDE*, pages 652–663. IEEE, 2009. ISBN 978-0-7695-3545-6.

[SVY09] Christian Sommer, Elad Verbin, and Wei Yu. Distance oracles for sparse graphs. In *FOCS*, pages 703–712. IEEE Computer Society, 2009. ISBN 978-0-7695-3850-1.

[SWZ02] Frank Schulz, Dorothea Wagner, and Christos Zaroliagis. Using multi-level graphs for timetable information in railway systems. In David Mount and Clifford Stein, editors, *Algorithm Engineering and Experiments*, volume 2409 of *Lecture Notes in Computer Science*, pages 43–59. Springer Berlin / Heidelberg, 2002. ISBN 978-3-540-43977-6.

[TZ05] Mikkel Thorup and Uri Zwick. Approximate distance oracles. *J. ACM*, 52(1):1–24, 2005.