Department of Computer Science,
Faculty of Mathematics, Physics and Informatics
in Bratislava,
Comenius University

Distance oracles for timetable graphs

(Master thesis)

František Hajnovič

Program: Informatics 9.2.1
Advisor: doc. RNDr. Rastislav Královič, PhD.

Bratislava 2013

I hereby declare that I wrote this thesis by myself, only with the help of the referenced literature, under the careful supervision of my thesis advisor.

....

Acknowledgements

I would like to thank ...

Abstract

This thesis deals with the shortest path queries on timetable graphs - i.e. graphs that represent timetables (of e.g. Slovak bus network). Firstly we investigate, if some properties (such as low highway dimension) are propagated (and to what extent) from the underlying graph to the timetable graph. Based on these facts, we show how to compute reasonably fast an accurate distance oracle on timetable graphs, that would efficiently answer shortest path queries, thus finding quick timetable connections between any pair of nodes.

Key words: **distance oracles**, **timetable graphs**, **timetable**, **highway dimension**

Abstrakt

V tejto práci...

Kľúčové slová: **distance oracles**, **timetable graphs**, **timetable**, **highway dimension**

Contents

1 Introduction

World is getting smaller every day, as new technologies constantly make communication and travelling faster and more effective then yesterday. Road network, Internet and many other networks are becoming more evolved and denser which also brings along new problems. In order to take advantage of such huge networks, we must have efficient algorithms that operate on these networks and give us answers to many questions. Among many others, one that we take great interest in is the question: "What is the shortest path from place x to place y"?

In different networks, this question can make different sense. In the road network, we would like to obtain a sequence of intersections we have to visit in order to reach our destination, driving the shortest possible time (or the smallest possible distance) . GPS devices and the likes of Google maps have to deal with this problem. In the case of Internet network, we might be interested in the shortest path to a destination computer in terms of router hops. In a network of social acquaintances, the smallest number of persons connecting us with e.g. Mark Knopfler could be expressed as a shortest path problem. Many problems in artificial intelligence (e.g. planning of actions to obatain some goal) can be expressed (or include) as a shortest path problem.

The importance of finding a shortest path in a graph is also obvious from the amount of algorithms and approaches we have nowadays to tackle this problem. Many traditional methods (like Dijkstra's algorithm) have many variations for different settings.

1.1 Motivation

1.2 Approach

1.3 Goals

1.4 Organization

2 Preliminaries

In this section, we would like to provide the definitions and terminology used throughout the thesis.

2.1 Objects

First, we will formalize the notion of a timetable and its derived graph forms, the underlying graph and notions related to these terms.

Definition 1. ***Timetable***
A timetable is a set $T = \{(x, y, p, q)|\ p, q \in N,\ p < q\}$.

- *Elements of* T *(the 4-tuples) are called an* ***elementary connection***
- *Pairs* (x, p) *and* (y, q) *such that* $(x, y, p, q) \in T$ *form the set of* ***events*** Ev_T *(e.g. train coming to or leaving a station)*
- x *(*y*) and* p *(*q*) are the* ***departure (arrival) city*** *and* ***time****.*

Intuitively, an elementary connection corresponds to moving from one stop to the next one, e.g. with a bus. Thus we disregard the notion of *lines* in our timetables.

Place		Time	
From	To	Departure	Arrival
A	B	10:00	10:45
B	C	11:00	11:30
B	C	11:30	12:10
B	A	11:20	12:30
C	A	11:45	12:15

Table 1: An example of a timetable - the set of elementary connections (between pairs of cities). An example of an event is a pair (A, 10:00)

Definition 2. ***Connection***
A connection from a to b is a sequence of elementary connections $(e_1, e_2, ..., e_k)$*,* $e_i = (e_x^i, e_y^i, e_p^i, e_q^i)$*,* $k \geq 1$*, such that* $e_x^1 = a$*,* $e_x^k = b$ *and* $\forall i \in \{2, ..., k\} : (e_x^i = e_y^{i-1},\ e_p^i \geq e_q^{i-1})$.

- *Connection* ***starts*** *at the departure time* e_p^1.
- ***Length*** *of the connection is* $e_q^k - e_p^1$.

So we understand connection as a (valid) sequence of elementary connections.

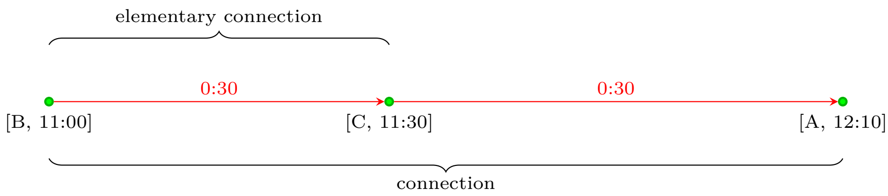

Figure 1: Connection and elementary connection

Next, we continue with the underlying graph - a graph representing basically the map on top of which the timetable operates.

Definition 3. ***Underlying graph (UG graph)***
The underlying graph of a timetable T UG_T *is an oriented graph* $G = (V, E)$*, where* V *is the set of all timetable cities and* $E = \{(x, y)|\ \exists (x, y, p, q) \in T\}$

Note, that we do not specify the weights of the edges in the underlying graph - they will be specified based on the current usage of the UG. Most of the time, however, we will work with UG where the

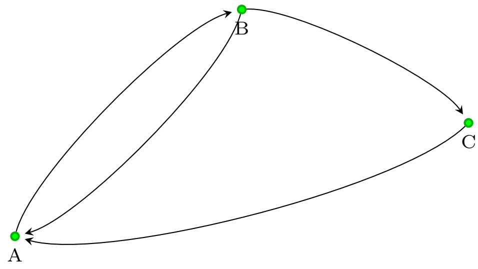

Figure 2: Underlying graph of the timetable in picture 1

weight of each arc is the length of the shortest elementary connection on that arc. More specifically, $w(x, y) = \min_{(x,y,p,q)\in T}(q - p)$ $(x, y) \in E(UG_T)$.
If we want to represent the timetable itself by a graph, there are two most common options [MH-SWZ07].

Definition 4. ***Time-expanded graph (TE graph)***
Let T be a timetable. Time-expanded graph from T TE_T is an oriented graph $G = (V, E)$ whose vertices are events of T ($V = Ev_T$). The edges of G are of two types

1. $([x, p], [y, q]), ;\ \forall (x, y, p, q) \in T$ *- the so called* ***connection edges***
2. $([x, p], [x, q]), [x, p], [x, q] \in V, p < q$ *and* $\nexists [x, r] \in V : p < r < q$. *- the so called* ***waiting edges***

Weight of the edge $([x, p], [y, q])$ is $w([x, p], [y, q]) = q - p$.

Informally, an edge in TE graph represent either the travelling with an elementary connection or waiting for the next event in the same city.

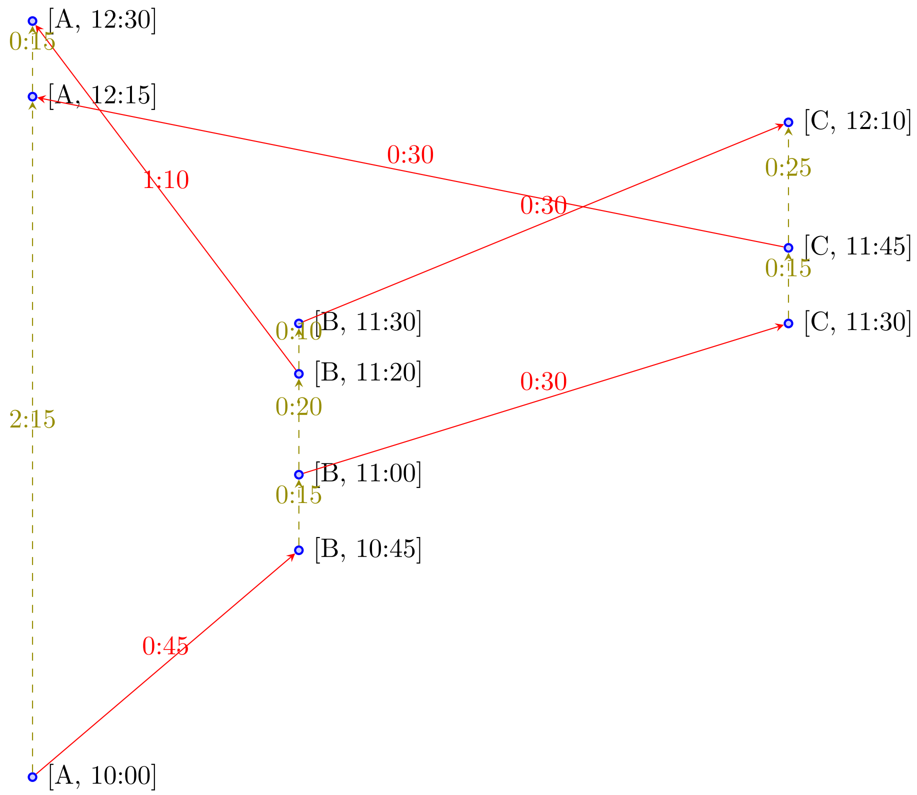

Figure 3: Time-expanded graph of the timetable in picture 1. Nodes are the events. There are connection and waiting edges.

Definition 5. ***Time-dependent graph (TD graph)***
Let T be a timetable. Time-dependent graph from T TD_T is an oriented graph $G = (V, E)$ whose vertices are the timetable cities and $E = \{(x, y)|\ \exists(x, y, p, q) \in T\}$. Furthermore, the weight of an edge $(x, y) \in E$ is a piece-wise linear function $w(x, y) = f_{x,y}(t) = q - t$ where q is:

- $\min\{q|(x, y, p, q) \in T,\ p \geq t\}$
- ∞ *if* $p < t\ \forall(x, y, p, q) \in T$

Intuitively, the TD graph is simply the UG graph where each arc carries a function specifying the traversal time of that arc at any time.

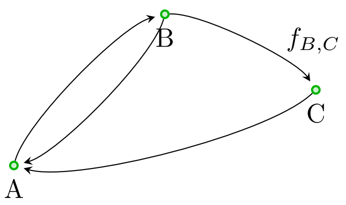

Figure 4: Time-dependent graph of the timetable in picture 1

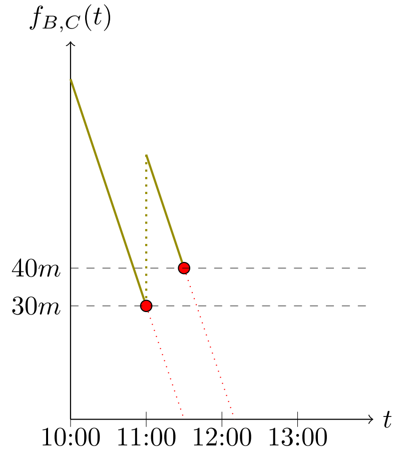

Figure 5: Piece-wise linear function - for the arc (B, C)

3 Related work

Name	Description	El. conns.	Cities	UG arcs	Time range	Height (h)
air01	domestic flights (US)	601489	287	4668	1 month	24374
cpru	regional bus (SVK)	37148	871	2415	1 day	239
cpza	regional bus (SVK)	60769	1108	2778	1 day	370
montr	public transport (Montreal)	7153	217	349	1 day	363
sncf	country-wide rails (FRA)	90676	2646	7994	1 day	488
zsr	country-wide rails (SVK)	932052	233	588	1 year	60308

Table 2: Timetables datasets

4 Data

In this section we would like to introduce the timetable datasets we were working with and provide the results of the analysis which we carried out on the data. The main reason for this analysis is that it gives some insight into the properties of the timetables, and thus may contribute to the make an oracle based method with better qualities.

5 Underlying shortest paths

6 Neural network approach

7 Application TTBlazer

8 Conclusion

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.

[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`.

[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`.

[Haj12] František Hajnovič. Distance oracles for timetable graphs - research, 2012.

[Mar] Mária Markošová. Základy umelej inteligencie 1 - prednášky.

[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.

[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.

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

[Ďur97] Pavol Ďuriš. Tvorba efektívnych algoritmov - prednášky, 1997.

[Wik] Wikipedia. London underground - wikipedia.