



ul. Piotrowo 3, 60-965 POZNAŃ  
tel. : 61 665 34 20  
fax.: 61 665 34 21  
e-mail: office\_dcf@put.poznan.pl

Poznań, Poland, 18.07.2019

## PhD Thesis Review

Submitted by: **Guralnik Roman Igorevich**  
Title: **Incremental Algorithms for Solving Optimization Problems on Big Graphs**  
Supervisors: **Prof. Boris Novikov** (Saint-Petersbourg State University, Russia)  
Reviewer: **Prof. Robert Wrembel**

Poznan University of Technology  
Faculty of Computing  
Institute of Computing Science  
Piotrowo 2  
60-965 Poznań, Poland  
tel: +48 61 665 2991, +48 665 2999, mob: +48 601 392346  
e-mail: Robert.Wrembel@cs.put.poznan.pl

### 1. Overview

For over 10 years graph processing has been one of the most popular topics in data processing research worldwide and for the last 3 years it became number one topic at the top data processing conferences, i.e., SIGMOD, VLDB, and ICDE<sup>1</sup>. Moreover, the latest issue of the VLDB Journal<sup>2</sup> is devoted entirely to graph technologies.

Graphs are data structures applicable to representing data from multiple domains, including among others: social networks, process (workflow) management, bioinformatics, good transportation and traffic control. Typically, linked open data are represented as graphs by means of RDFs. Thus, solutions to efficient storage and processing of graphs are very important issues of great practical impact.

The most common algorithms applied to graphs include: finding the shortest path (in multiple variations), finding minimum spanning tree, finding clique, solving the traveling salesman problem, computing a transitive closure, finding connected components, partitioning, clustering, similarity search.

<sup>1</sup> R. Wrembel, A. Abelló, I.-Y. Song: DOLAP data warehouse research over two decades: Trends and challenges. Information Systems, vol. 85, 2019, <https://doi.org/10.1016/j.is.2019.06.004>

<sup>2</sup> The VLDB Journal, vol. 28, issue 3, 2019

Bx. N 09/2 - 352 07 04.09.2019



ul. Piotrowo 3, 60-965 POZNAŃ  
tel. : 61 665 34 20  
fax.: 61 665 34 21  
e-mail: office\_dcf@put.poznan.pl

Problems (also the aforementioned ones) related to graph processing are computationally very complex. For this reason, providing efficient (even if sub-optimal) algorithms for processing data is challenging and of great importance. The problem of computational complexity goes into the extreme for big data, i.e., graphs of millions of nodes (which is typical to social networks and transportation).

The last domain is strongly related to economy, since reducing costs of goods delivery may increase profitability of a business. Typical problems dealt with in this area include: (1) finding the shortest or (2) finding the fastest route for delivering goods from a source to a destination.

The PhD thesis by Roman Igorevich Guralnik addresses these computationally complex problems on: (1) vehicle routing with precedence and capacity constraints with time windows and (2) the problem of resource-constrained shortest path with increasing edge weights and incremental shortest path trees updates.

In this context, the PhD thesis by R. Guralnik addresses a challenging and important problem that is being addressed for years by multiple international research communities, cf. to the latest PVLDB paper<sup>3</sup>. The difficulty of the problem is suitable for a doctoral dissertation.

## 2. Contribution

In my opinion, the **research contributions** of the PhD dissertation by Roman Guralnik are as follows. First - the development of an algorithm solving the resource constrained shortest path problem (RCSP). Second - an algorithm for real-time rerouting of vehicles.

In the shortest path problem applied to vehicle routing, graph edges represent path with associated costs. In the resource constrained shortest path problem, graph edges have additionally associated resources (e.g., fuel consumption) that are consumed while travelling through a given edge. In this context, R. Guralnik developed an algorithm that uses previously calculated and stored data as a basis for modifying a vehicle route (path) as a results of some changes to a road network (graph topology). The solution is described in Chapter 2.

An incremental algorithm was proposed for efficient update of the already selected route. The algorithm is based on the solutions developed by other researchers, i.e.,

- R. Muhandiramge, N. Boland: Simultaneous solution of Lagrangean dual problems interleaved with preprocessing for the weight constrained shortest path problem. *Networks*, vol. 53, no. 4, 2009
- S. Pallottino, M.G. Scutella: A new algorithm for reoptimizing shortest paths when the arc costs change. *Operations Research Letters*, vol. 31, no. 2, 2003

R. Guralnik modified the algorithm by Pallottino and Scutella algorithm to re-optimize the shortest path trees. The solution is described in Chapter 3.

The contributions presented in this dissertation were also evaluated by experiments. The obtained results are promising and show the applicability of the contributions.

---

<sup>3</sup> Y. Wang, G. Li, N. Tang: Querying Shortest Paths on Time Dependent Road Networks. *PVLDB* (11):12, 2019



ul. Piotrowo 3, 60-965 POZNAŃ  
tel. : 61 665 34 20  
fax.: 61 665 34 21  
e-mail: office\_dcf@put.poznan.pl

The content of the dissertation has been published as: two papers published in international conference proceedings (ACM and Springer CCIS) and one book chapter (IOS Press) - all listed at DBLP, as well as one Russian journal:

- Guralnik, R.: *Incremental Rerouting Algorithm for single-vehicle VRPPD. Proc. of Int. Conference on Computer Systems and Technologies (CompSysTech), ACM, 2017*
- Novikov, B., Guralnik R.: *The Algorithm for Constrained Shortest Path Problem Based on Incremental Lagrangian Dual Solution. Proc. of Int. Baltic Conference on Databases and Information Systems, Springer CCIS, vol. 838, 2018*
- Novikov, B., Guralnik R.: *Efficiency thresholds for the incremental constrained shortest path algorithm. In Databases and Information Systems X Selected Papers from the 13th Int. Baltic Conference. IOS Press, 2019*
- Guralnik, R.: *Some problems on graph databases. Institute for System Programming of Russian Academy of Science, vol. 28, no. 4, 2016 (in Russian)*

For this reason, the quality of the contributions have already been verified and approved by an international research community. His publication record fulfills requirements of a PhD.

### 3. Detailed evaluation

#### 3.1. Major comments, observations, and questions

On page 104 the author states: "we have analyzed different papers and have described basic algorithms which were offered 10-15 years ago". Why he considers only so old papers on the topic? At DBLP there are 75 papers published on the shortest path problem only in 2019!

Section 2.2: it is unclear how in the second phase alternative paths are selected. Are all possible alternative paths traversed? It seems that only the costs associated to edges (paths, routes) are modified and considered - see step I "Take into account changes in price matrices ...". What about changes in a topology, e.g., removing a given path (road closing)?

Conditions stated as inequalities 2.1, 2.2, and 2.3: what would happen if some costs  $m_i' > m_i$  and some  $m_i' < m_i$ , i.e., some paths in  $I$  increased their costs and some decreased?

Page 132, step II: what do you mean by "Extensive experiments show that the best results can be achieved ..." - the best results in what sense? What is the criterion (measure) for comparing the results?

Section 2.3: I failed to see a clear statement whether the experimental evaluation confirms or not the computational complexity obtained theoretically.

Section 3.7: are the obtained solutions really optimal and not sub-optimal?

Figure 3.2 present the performance of the algorithm contributed in this dissertation for two different graphs (road connections). Do the performance depends on the shape of a graph, i.e., the average number of in-going and out-going edges, the average length of a path?



ul. Piotrowo 3, 60-965 POZNAŃ  
tel. : 61 665 34 20  
fax.: 61 665 34 21  
e-mail: office\_dcf@put.poznan.pl

### 3.2. Minor issues

Abbreviation RCSP is used from the very beginning of the dissertation, but it is explained in Section 1.5, which is too late. It should be explained when used for the 1st time.

Page 103, the 2nd paragraph: should be "the Internet".

Page 127, the 3rd paragraph: term "standard data" was used but it is totally unclear what the author meant by this term. Further in the paragraph: "with great probability" - this statement is vague. What do you mean by "great", 0.9999? What is the threshold for you to be able to say "great"?

Page 129, the 1st paragraph: "that 4" should be replaced by "that the first four".

Page 130, 139 - don't use "." at the end of titles of chapters and sections.

The tables on page 138 are missing numbers and captions.

Page 139: the 1st paragraph includes repetitions.

Page 139: the 2nd paragraph: "the less" should be replaced with "the lower".

Page 134, 4th paragraph: "several iteration" should be replaced with "several iterations", "First" should be replaced with "The first".

Table 3.1: a clear display of the units of the presented values is missing. I can only guess that they are measured in seconds.

### 4. Final recommendation

The above comments, observations, and questions show the area of potential improvement and should be treated as starting points for discussion. In my opinion, the reviewed dissertation by Roman Igorevich Guralnik contributed solutions to research and technological problems in the area of graph processing with practical application. Since the proposed solutions are adequate for a PhD, the dissertation by R. I. Guralnik **should be accepted for an oral defense**.