

UNIVERSITY OF TARTU
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE
Institute of Computer Science

Volodymyr Floreskul

Memory-Efficient Fast Shortest Path Distance Estimation in Large Graphs

Master's thesis (30 EAP)

Supervisor: Konstantin Tretyakov, MSc

Author: “.....” May 2013

Supervisor: “.....” May 2013

Approved for defence

Professor: “.....” May 2013

TARTU 2013

Abstract

Shortest path computation is one of the most critical primitives in graph algorithms. In large graphs there is often a need to use approximate methods as simple exact algorithms require unacceptably large amounts of time. One family of techniques that is simple and scalable at the same time is based on upper bound distance approximation using a fixed set of selected nodes called *landmarks*. If we know the distances from all nodes in the graph to a landmark u , the distance between a pair of nodes s and t can be approximated using the triangle inequality: $d(s, t) \leq d(s, u) + d(u, t)$.

In a similar way it is also possible to calculate approximated shortest paths. The obtained accuracy can be increased by using a set of k landmarks, but this leads to linear increase of memory usage and preprocessing time with sublinear approximation error reduction.

In this work we introduce a simple but powerful modification to this approach that we call *pruned landmark trees*. The idea of this improvement is based on the fact that in majority of cases it is sufficient to keep the distances only to the r closest landmarks rather than the whole set of landmarks, where r is much smaller than k . We describe three shortest path approximation algorithms that use the proposed modification.

The performance of the presented methods was tested on big real-world social network graphs including DBLP, Orkut, Twitter and Skype. The obtained results have been compared against the results of regular landmark-based techniques. The comparison showed that pruned landmark tree-based algorithms can be used to significantly reduce the used memory consumption while achieving both comparable accuracy and query execution time.

Non-exclusive licence to reproduce thesis

I, Volodymyr Floreskul, (date of birth: 29.12.1989),

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright

Memory-Efficient Fast Shortest Path Distance Estimation in Large Graphs,
supervised by Konstantin Tretyakov,

2. I am aware of the fact that the author retains the right referred to in point 1.

3. This is to certify that granting the non-exclusive licence does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Tartu, 20.05.2013