Hessisches Kompetenzzentrum
für Hochleistungsrechnen

Comparing Importance of Webgraph Nodes via Random Walks



Project Manager
Lukas Roth

Principal Investigator
Prof. Dr. Volker Betz

Project Term
2018 - 2020

Clusters
Lichtenberg Cluster Darmstadt

Additional Software
C, C++

Institute
Department of Mathematics

University
Technische Universität Darmstadt

Introduction

We planned to test a method for ranking nodes in a network (for example web pages in the internet) that is an alternative to the page rank algorithm. This involves simulating random walks on the network and contracting the network in the run. As we intend to analyse different kinds of the method, we also use some methods to characterize the shape of the network, and run them before and after we contract the networks. Because we need to do many simulations and plan to look also at large networks we need much computation time so that we need to use the HPC.

Methods

We are interested in the probability that starting at some node in the graph it is more likely to visit a certain destination node before coming back to the starting node. Therefore, in one run we perform a random walk on the network and contract the network during the run. By contracting the network we intend to speed the algorithm up. With the random walk we check whether we reach the destination node or not. To calculate the probability, we make use of Monte Carlo Methods, that means that we perform a lot of independent runs and then average the results. The code of the algorithm is written in C++.

Results

Natalia Londono wrote her Master Thesis about this project and

it was a great success. She tested several methods of contracting networks and used them on some example graphs. To analyse and compare these methods, some methods were used to characterize the structure of large graphs. These methods are based on the theory of simplicial complexes. In addition the runtime was taken into account. However, the results don't tell much about which methods are better.

Discussion

The results suggested that it would be difficult to bring the method to a point that it achieves a good benefit, because the computation amount seems too high to be practicable. Because of that we stopped investigating further in this direction for now. But it might be worth looking at further in the future.

Last Update: 2021-07-27 19:08