CSCI 565: Distributed Systems (Fall 2024)

Programming Assignment: Evaluating the GHS Distributed Minimum Spanning Tree Algorithm

1 Overview

In this programming assignment, you are tasked to implement and evaluate an existing distributed minimum spanning tree algorithm and compare its experimental results against the theoretical analysis. By the time you finish this assignment, you will not only have a deeper understanding of the specific algorithm, but also you will gain experience in evaluation via simulation and large-scale graph datasets.

You are allowed to complete this assignment in Java, C++, or Python.

Before starting this assignment, you might want to review relevant textbook chapter on graph algorithms (Chapter 10) and the classic paper on the GHS distributed minimum spanning tree algorithm [1].

2 Assignment Details

This assignment is intended to be open-ended, which is the inherent nature of research.

In this programming assignment, you will implement the GHS algorithm [1]. There are open-source implementations of the algorithm on the Internet. You are allowed to use one, but make sure you cite the source and verify the code's correctness if you do so. You are encouraged to implement it on your own since that will help you really understand the algorithm.

The original paper on the GHS algorithm provided a theoretical analysis of the algorithm's communication cost and processing time. You need to conduct experiments using large scale graph datasets to validate the theoretical results. More importantly, you need to evaluate the scalability of the algorithm. Please choose one dataset from the Stanford SNAP website [2].

3 Deliverables

You need to hand in all of your source files, a README file, a Makefile if you use C/C++, and a report in PDF written in LaTeX. Zip up these files and submit to Canvas. Do not submit your object files or executables!

Your report in PDF should include (i) the theoretical results of communication cost and processing time from the paper, (ii) how you evaluated the algorithm, (iii) what dataset you used and why you chose it, (iv) results summarized in tables or figures, and (v) results analyzed in English. In

your analysis, describe whether your experimental results are in line with the theoretical analysis, whether there are any anomalies. If there are any, provide your explanations.

4 Grading

- Your own implementation: 10 bonus points.
- Report: 100 points.

  If you have any questions, feel free to post them to the forums on Ed Discussion. We prefer that you post questions to Ed Discussion rather than sending emails because others in the class might also benefit from the answers. However, please do not post any source code there.

References

[1] Robert G. Gallager, Pierre A. Humblet, and P. M. Spira, "A distributed algorithm for minimum-weight spanning trees," ACM Transactions on Programming Languages and Systems, vol. 5, no. 1, pp. 66–77, January 1983.

[2] Jure Leskovec and Rok Sosic. "SNAP: A general purpose network analysis and graph mining library in C++." http://snap.stanford.edu/snap, June 2014.