CS 442 Project Proposal

Group 11
Alp Ege Baştürk 21501267
Aldo Tali 21500097
Oğuz Kaan Ağaç 21503895

Project Description

Introduction

In this project, we are planning to implement Google's Pagerank algorithm using MPI in C. Pagerank is an algorithm used to rank node importances based on links of the graph. We will be implementing power iteration based Pagerank using matrix-vector multiplication. Dead ends and spider traps will be considered using teleportation.

Details

Pagerank operation depends on $r = \beta M \cdot r + \left[\frac{1-\beta}{N}\right]_N$ calculation for rank vector, where β is the transition probability. This matrix operation is iterated until convergence. Matrix-vector multiplication is a well known problem and it can be distributed to different nodes for calculation. In addition, convergence check $\sum_j |r_j^{new} - r_j^{old}| > \varepsilon$ might be implemented by reduce operation on the master. Such optimizations will be considered for performance issues. Different datasets will be used to test the algorithm performance on different graphs.

Algorithm:

$r^{old} = [1/N]_N$

$Do:$

$\qquad r^{new} = \beta M r^{old} + [\frac{1-\beta}{N}]_{NxN}$

$\qquad r^{old} = r^{new}$

$Until\ Convergence$

Possible Challenges

There are several problems that we will consider in this distributed implementation being: The distribution of the data/graph over the network in a way that each node can deal with its own local cluster of computation, minimizing the number of messages exchanged in the network both in number and in size and reaching a representation that is both scalable with the size of the input as well as with the improved running time.