B534 Project 1

Group 6

Qiuwei Shou and Chao-Hong Chen

There are 3 main steps of our SequentialPageRank program:

1. Parse input file and create two data structures:
   1. degs: a HashMap which contains the outdegree of each node.
   2. adjMatrix: a HashMap which contains a list of node for every vertex , where every node in has a link point to .
2. Calculate page rank:
   1. Initialize all rank values to where is the number of nodes.
   2. For each vertex , update it according to the formula:  
      , where is the outdegree of and is the damping factor.   
      However, if the graph contains dangling nodes (no outbound links) just use this formula will cause . There are two ways to solve this:
      1. Before running the algorithm, find all dangling nodes and add outbound links to every other node, no need to change the formula.
      2. When updating dangling node use the formula:

Both method works but the second one is more efficient and more accurate so we choose the second approach.

The program will iterate this update process for give number of times.

1. Sort according to rank values and output.

The output file in the report is produced using damping factor 0.85 and 100 iteration.

The command to generate the output file in the report:

java SequentialPageRank pagerank.input.1000.urls.6 chen464\_SequentialPageRank\_output.txt 100 0.85