CSEN 272 Web Search and Information Retrieval

Project 1 Report

Samuel Karumanchi (SCU ID:07700009611)

Aim: Implementing the PageRank algorithm used for ranking webpages and analyzing the results of the algorithm by varying its parameters.

Observations on PageRank with Varying Damping Factor (d)

Convergence Criterion and Runtime Analysis

When the L1 norm difference between iterations is less than 1e-6, the PageRank algorithm converges. The findings demonstrated that a slower rate of convergence was associated with larger damping factor (d) values. The algorithm took fewer iterations to converge when d was set to 0.75 than when d was set to 0.95. This is due to the fact that a smaller damping factor raises the likelihood of teleportation, which more evenly distributes ranks and expedites stability. Higher damping values, on the other hand, make the algorithm more dependent on link topologies, which slows convergence as updates spread more slowly throughout the network.

Changes in Top-Ranked Pages

Across a range of d values, the top-ranked pages stayed largely stable. Random teleportation, however, had a greater effect for lower values of d, which led to some lower-linked pages ranking somewhat higher than anticipated. On the other hand, pages with more inbound links were given preference by the algorithm when d was raised, which further enhanced their dominance in the ranks. This illustrates that while lower values add more unpredictability to the ranking process, greater damping factors emphasize the significance of robust connection structures.

Impact on Overall PageRank Distribution

A tiny group of highly connected pages received noticeably higher scores as d rose, indicating that the PageRank values were more biased. There was less variation between pages and a more balanced rank distribution when d was set to 0.75. This implies that teleportation has a greater influence on page significance for lower d values. Higher d values cause the algorithm to more properly represent the link structure, giving well-connected pages much higher weights while devaluing less-linked ones.

Running the Code

Project 1 Code.py input.txt d > output.txt