

Azalea Rohr

Homework 10

April 17st, 2024

Collaboration and Sources: None

My code computes an approximation to PageRank, a method used to rank web pages. It is similar to the memoryless process of walking over vertices of a directed graph, and the next vertex depends only on the current vertex and does not depend on the previous vertices. I split my code into different modules consisting of

- Graph: implementation of a graph data structure with methods to add vertices, edges, count vertices, and get neighbors.
- Pagerank: implements functions to simulate random walks over the graph vertices and count the number of terminations of these walks
- Data generator: function to generate a data file with directed edges for testing purposes.

Within the code, it consists of the following

1. Main execution
 - a. Reads a directed graph from file "data.txt"
 - b. If the file doesn't exist, generate the file using the data generator module.
 - c. Finds the count of terminations for each vertex using random walks
 - d. Determines the top 5 vertices based on the number of terminations
 - e. Find the Page Rank for the top five vertices (approximately)
 - f. Prints and top 5 vertices and their Page Rank values
2. Reading the Graph from the file
 - a. Reads the directed graph from the file into the graph data structure
 - b. Every line of the file represents a directed edge between 2 vertices
3. Number of Terminations
 - a. Runs random walks from each vertex and counts the number of terminations for each vertex
 - b. Using the Pagerank module to simulate the random walks
4. Finding top 5 vertices
 - a. Top 5 vertices is based on termination counts
 - b. Sorts the vertices by their termination counts and outputs the top 5
5. Approximation to Pagerank
 - a. Computes the approximation to PageRank for each of the top five vertices.

- b. Divide the termination counts by $100n$, where n is the number of vertices.
- 6. Tests
 - a. For my test, I created one that verifies the correctness of the output of the program by comparing the generated output with the expected output.

And below is my output for the top 5 vertices.

vertex 530: approximate PageRank 0.0010

vertex 418: approximate PageRank 0.0010

vertex 91: approximate PageRank 0.0009

vertex 561: approximate PageRank 0.0009

vertex 895: approximate PageRank 0.0009