Project Overview

- **Goal**
  - The goal of this project is to see how interconnected congress people are by analyzing their twitter interactions with a random graph.
- **Dataset**
  - https://snap.stanford.edu/data/congress-twitter.html
  - Nodes: 475, Edges: 13,289 (Accepted in the project proposal)

Data Processing

- I loaded my file in using a reader and adding each line
- No transformations or cleaning except for only using the nodes and not the weights

Code Structure

- **Struct Graph**
  - Holds an adjacency list of the source and target nodes in my graph using a HashMap
- **Impl Graph**
  - fn new()
    - Creates an empty graph and returns a new Graph using an empty HashMap
  - Fn add_edge()
    - Adds a directed edge from source to target
    - Inputs are the source and target nodes and outputs and updated adjacency list
    - Uses entry().or_insert_with to do this
  - Fn from_edgelist()
    - Constructs graph from edgelist, input is file path and output is graph; reads lines from files, parses source and target nodes, calls add_edge()
  - Fn bfs()
    - Perform Breadth-First Search to compute shortest path lengths from start,, output is HashMap of node and distance, uses a queue and a hash map.
  - Fn generate_random_graph()
    - Generates random directed graph with nodes and edges, outputs graph, randomly selects node pairs and adds edges from add_edge()
  - Fn average_path_length()
    - Calculates average length of the shortest path between node pairs, outputs a float, performs bfs from each node, sums all non-zero distances, divides by total number of paths.

Tests

- running 3 tests
- test tests::test_add_edge ... ok
- test tests::test_average_path_length ... ok
- test tests::test_bfs_simple ... ok
- Test_add_edge: tests to makes sure edges can be added
- Test_bfs_simple: simple bfs test to make sure it works properly

- Test_average_path_length: Makes sure the average function is working

Results

Loading Congress graph...
Generating random graph...
Calculating average path lengths...
Average path length (Congress): 2.3572
Average path length (Random): 2.1420

The results are marginal at best, with the results shows that the average path length of congress people are around equal if not less connected than that of random nodes.

Usage Instructions

I simply built my code by analyzing a graph of edges, using BFS to find the shortest path of each node, creating a random graph of edges, and comparing the average path lengths of both