Prajusha Reddy
Link to dataset: https://snap.stanford.edu/data/roadNet-CA.html

Project Introduction

For my project, I decided to conduct graphical analysis on California's road network. I used the dataset from the Stanford Large Network Dataset Collection, and the link to that is provided above. With the dataset, I aimed to find the degree centrality of each node (intersection/endpoint) in the road network, and wanted to find the top 10 intersections/endpoints with the most connections (roads) leading to them.

Modules and Functions

main.rs

- **main()**
- **top_nodes_centrality()**
- **top_nodes_degree()**

graph.rs

- **read_lines()**
- **read_graph_from_file()**
- **calculate_degree()**
- **calculate_degree_centrality()**
- **calculate_average_distance()**
- **bfs_distances()**

Tests

- **test_calculate_degree_centrality()**
- **test_calculate_degree()**

Analysis

Output:

For the top ten intersections and endpoints in the California road network, the code outputs the number of connections (roads), as well as the degree centrality, which I calculated by dividing the degree of each node by the total number of nodes minus one.

```
Top 10 Nodes by Degrees:
1. Node 590397: 12.00
2. Node 1774834: 10.00
3. Node 3943295: 10.00
4. Node 3660686: 9.00
5. Node 881638: 8.00
6. Node 2792954: 8.00
7. Node 6992304: 8.00
8. Node 314328: 8.00
9. Node 5324043: 8.00
10. Node 2351315: 8.00

Top 10 Nodes by Degree Centrality:
1. Node 590397: 0.000001600337
2. Node 1774834: 0.000001333614
3. Node 3943295: 0.000001333614
4. Node 3660686: 0.000001200253
5. Node 4490654: 0.000001066892
6. Node 7388485: 0.000001066892
7. Node 7007560: 0.000001066892
8. Node 2376823: 0.000001066892
9. Node 5119635: 0.000001066892
10. Node 1381904: 0.000001066892
```

As you can see, the top node, Node 590397, has the highest number of road connections (12). Respectively, its degree centrality is 0.0000016, which is the number of roads upon that intersection or endpoint. The nodes with the highest degree centralities are the hubs of the network. In this case, the top 10 nodes would be the most central intersections/endpoints in the California road network. I wanted to conduct this analysis because these could represent roads leading to important places, such as tourist destinations. Additionally, they could be hubs for peak traffic which could be helpful in the future to make places more efficient.

I also calculated the average distance of a subset of 1000 nodes (to make it more efficient), because I was curious. I employed the BFS Algorithm (Breadth-First Search Algorithm) in order to do this.

```
Average Distance (Subset): 0.43
```

This is the average distance that I calculated between a subset of 1000 nodes.

Sources:

https://docs.rs/graphrs/latest/graphrs/
https://en.wikipedia.org/wiki/Centrality
https://towardsdatascience.com/notes-on-graph-theory-centrality-measurements-e37d2e49550a
https://omicsforum.ca/t/what-are-the-differences-between-degree-and-betweenness-centrality/163
https://doc.rust-lang.org/std/collections/struct.HashMap.html
https://docs.rs/petgraph/latest/petgraph/
https://doc.rust-lang.org/std/collections/struct.VecDeque.html