A Report on

Navigability of Real-World Networks

GitHub Repository Link

Prepared & Submitted

By

Mr. KUNAL SANJAY SANGHVI

WSU ID.: - 011809708

Mr. ARNAV PRAMOD JAIN

WSU ID.: - 011809037

Under the Guidance of

Dr. ASSEFAW GEBREMEDHIN

WASHINGTON STATE UNIVERSITY

Department of COMPUTER SCIENCE

Pullman Campus, WA – 99164

CPT_S 591 ELEMENTS OF NETWORK SCIENCE

Contents

Table of Figures 3
I. Abstract 4
II. Introduction 4
III. Problem Definition 5
IV. Project Flow 6
1. Data Gathering 6
2. Algorithm Implementation 6
3. Simulation Framework & Network Visualization 6
4. Data Analysis 6
V. Dataset 7
1. Transportation Network 7
a. Feature set of transportation network 7
b. Network Visualization 8
2. Social Network 8
a. Feature set of social network 9
b. Network Visualization 9
VI. Algorithms 10
1. Navigation Strategies: Source to Destination Navigation using Dijkstra & A* Algorithm 10
a. Transportation Network 10
b. Social Network 11
2. Community Detection 11
a. Transportation Network 12
b. Social Network 13
3. Shortest Path Algorithm 13
a. Dijkstra 13
b. Floyd Warshall 14
c. Page Rank 15
d. Random-Walk 15
VII. Measures 16
1. Centrality Measures 16
a. Transportation Network 17
b. Social Network 18

2. Degree Distribution and Clustering Coefficient 18
a. Transportation Network 19
b. Social Network 20
3. Route Planning 21
a. Landmark 21
b. Congestion/Roadblocks 22
4. Small World Phenomenon 23
a. Transport Network 24
b. Social Network 25
5. Hubs Analysis 26
a. Visualizing Transport Network with and without top 10 hubs 26
b. Visualizing Social Network with and without top 10 hubs 26
c. Transport Network Navigability Analysis on removal of Top 10 Hubs Individually 27
d. Social Network Navigability Analysis on removal of Top 10 Hubs Individually 27
e. Hubs & Authority Scores 28
f. Relation of Navigability Analysis with Easley – Kleinberg Small World Phenomenon 29
VIII. Related Work 30
IX. Conclusion 31
X. References 32

Table of Figures

Figure 1. Project Flow 6
Figure 2. Properties of Transportation network 7
Figure 3. Transport Network Visualization 8
Figure 4. Properties of Social Network 9
Figure 5. Social Network Visualization 9
Figure 6. Results of the analysis of navigation strategies on Transportation Network 10
Figure 7. Results of the analysis of navigation strategies on Social Network 11
Figure 8. Community Detection using Louvain Algorithm on Transportation Network 12
Figure 9. Community Detection using Louvain Algorithm on Social Network 13
Figure 10. Shortest Path Length of Transportation Network using Dijkstra's Algorithm 14
Figure 11. Shortest Path Length of Social Network using Dijkstra's Algorithm 14
Figure 12. Shortest Path Length of Transportation Network using Floyd Warshall Algorithm 14
Figure 13. Shortest Path Length of Social Network using Floyd Warshall Algorithm 14
Figure 14. Shortest Path Length of Transportation Network using Page Rank 15
Figure 15. Shortest Path Length of Social Network using Page Rank 15
Figure 16. Shortest Path Length of Social Network using Random-Walk 15
Figure 17. Shortest Path Length of Transportation Network using Random-Walk 15
Figure 18. Different Centrality Measures on Transportation Network 17
Figure 19. Different Centrality Measures on Social Network 18
Figure 20. Degree Distribution of Transportation network 19
Figure 21. Degree Distribution of Social Network 20
Figure 22. Navigability using Landmark. 21
Figure 23. Change in Shortest Path Length when Congestion/Roadblocks is introduced. 22
Figure 24. Transport Network Visualization without roadblock/congestion 22
Figure 25. Transport Network Visualization with roadblocks/congestion 23
Figure 26. Transport Network Visualization along with its Small World Model 24
Figure 27. Properties of Real-World Transport Network and Small World Model 24
Figure 28. Social Network Visualization along with its Small World Model 25
Figure 29. Properties of Real-World Social Network and Small World Model 25
Figure 30. Transport Network Visualization with and without top 10 hubs 26
Figure 31. Social Network Visualization with and without top 10 hubs 26
Figure 32. Top 10 Hubs removed individually for Transport Network 27
Figure 33. Top 10 Hubs removed individually for Social Network 27
Figure 34. Hub & Authority Scores for Transport Network 28
Figure 35. Hub & Authority Scores for Social Network 29

I. Abstract

This study compares and analyzes four navigation algorithms: Dijkstra's, Floyd-Warshall, PageRank, and Random-walk algorithms to determine their applicability to different types of networks. It draws on publicly available datasets on social and transportation networks and a literature review of relevant research articles. The study evaluates each algorithm's performance on these datasets to provide insights into which algorithm is best suited for a given network structure. The study investigates how network structure impacts the emergence of navigation strategies in scenarios like landmarks, congestions, or roadblocks. The study aims to provide insights into the development of more efficient and effective navigation systems in social and transportation networks.

II. Introduction

Navigation is a fundamental aspect of our daily lives, whether we are trying to find our way through a city, navigating a complex transportation network, or navigating a social network. Over the years, researchers and practitioners have developed a range of navigation algorithms that can help individuals find their way through complex networks. These algorithms play a crucial role in guiding individuals through social and transportation networks. The importance of navigation algorithms in transportation networks cannot be overstated. In urban areas, traffic congestion has become a significant problem, leading to increased travel time, fuel consumption, and air pollution. Navigation algorithms can help drivers avoid congested areas and find the fastest route to their destination, thereby reducing travel time and fuel consumption. Moreover, navigation algorithms can also help emergency vehicles reach their destination quickly, potentially saving lives.

Similarly, these strategies are also essential in social networks. Social networks are complex webs of relationships that are constantly changing, and navigation algorithms can help individuals navigate these networks efficiently. For example, social network analysis can be used to pinpoint high priority individuals (influential), which can help businesses identify key opinion leaders and target their marketing efforts more effectively. Moreover, navigation algorithms can also help individuals find new friends or colleagues with similar interests, leading to new opportunities for collaboration or socialization. Despite the importance of navigation algorithms in social and transportation networks, it can be challenging to determine which algorithm is best suited for a given network structure. This is because different algorithms have different strengths and weaknesses, and their applicability to different types of networks can vary. For example, some algorithms may perform well in transportation networks but not in social networks, while others may perform well in both types of networks.

One popular pathfinding algorithm between two points in a network is Dijkstra's algorithm. This algorithm is extensively used in transportation networks, where the goal is to find the fastest route between two locations. However, it may not be suitable for networks with many nodes or edges, as it can be computationally expensive. To address this, researchers have developed variations of Dijkstra's algorithm, such as A* search, which uses heuristics to improve its performance. The PageRank algorithm is frequently used in social networks to identify the most influential people. Initially created to rank web pages based on their significance, it can also be used in social networks. The algorithm used in this study is a variation of the Floyd-Warshall algorithm, which is an all-pairs shortest-path algorithm that calculates the shortest routes between all vertex pairs.

While these algorithms have been used successfully in transportation and social networks, their applicability to other types of networks is less clear. Moreover, the structure of the network can have a significant impact on the emergence of navigation strategies. Landmarks and shortcuts are common strategies used in transportation networks, where the goal is to find the fastest route between two locations.

Considering these challenges, this study aims to determine the navigability and applicability of different types of networks. The study will draw on publicly available datasets on social and transportation networks, as well as a literature review of relevant research articles and publications.

III. Problem Definition

The problem addressed in this study is the challenge of determining the most suitable navigation strategy for a given network structure in social and transportation networks. With various navigation algorithms available, it can be difficult to determine which algorithm will perform best for a particular network. This problem has significant practical implications, as efficient navigation is crucial in both social and transportation networks. For transportation networks, efficient navigation can help reduce traffic congestion, travel time, and fuel consumption, which can reasonably help improve the environment and public health. For social networks, efficient navigation can help individuals find new friends, collaborators, or job opportunities, and businesses can use navigation algorithms to identify key opinion leaders and target their marketing efforts more effectively.

The study aims to address this problem by comparing and analyzing four popular navigation algorithms: Dijkstra's, Floyd-Warshall, A* Search, Random-walk, and PageRank algorithms. Understanding how these navigation strategies emerge in different network structures can help researchers and practitioners develop more efficient and effective navigation systems. All in all, the problems addressed in this study is critical for developing efficient navigation systems in social and transportation networks. The observations/results of this study will provide valuable insights into the selection of navigation strategies, metrics, and parameters, which can have significant practical implications in various domains.

IV. Project Flow

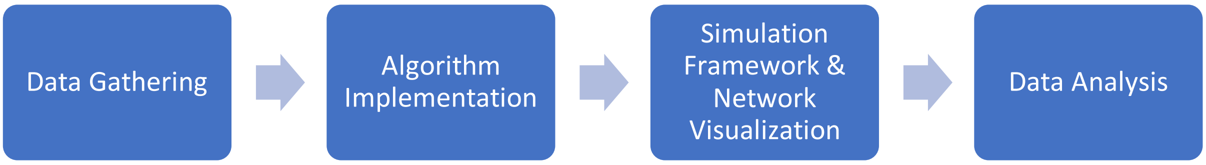

Figure 1. Project Flow

1. Data Gathering

Collect data on the social and transportation networks of interest, including information on nodes (people, places, intersections), edges (connections, roads, friendships), and any other relevant attributes. Convert the data into a network representation using a suitable software tool or library (e.g., NetworkX, Gephi, or Cytoscape).

2. Algorithm Implementation

The study plans to implement navigation algorithms, including Dijkstra's algorithm, Random-walk, A* search, Floyd-Warshall algorithm, and PageRank algorithm, using Python programming language. The study aims to develop performance metrics to compare algorithms' efficiency, accuracy, and scalability. Apart from this the project also aims to implement algorithms to perform tasks like community detection and route planning.

3. Simulation Framework & Network Visualization

The study plans to develop a simulation framework to evaluate navigation algorithms' performance on networks with different properties. The simulations will consider real-world aspects like traffic congestion, road closures, and delays. The study will use a range of NetworkX libraries (matplotlib, NumPy, etc.) and built-in functions to visualize networks and its different characteristic properties. The goal is to gain a comprehensive understanding of the extent of navigability of these networks.

4. Data Analysis

The analysis phase involves using statistical analysis and visualization techniques to evaluate the chosen data. This will also investigate the relationship between network structure and navigation strategies to gain insights into the factors that influence their performance. The findings will inform the development of more effective and efficient navigation systems for social and transportation networks.

V. Dataset

1. Transportation Network

The "road-minnesota.mtx" dataset available on networkrepository.com[1] represents a road network of Minnesota state in the United States. It depicts a sparse matrix in the Matrix Market format, where each row and column correspond to a node in the graph, and the non-zero entries in the matrix represent the edges between nodes. It has 2,816 nodes and 5,574 edges. Specifically, the nodes in the graph correspond to intersections or endpoints of roads in Minnesota, and the edges represent the roads connecting these intersections. This dataset can be used to study various aspects of transportation networks, such as traffic flow, congestion, routing algorithms, and network resilience. It can also be used to develop and test models for predicting travel times and estimating the impact of infrastructure changes on traffic patterns. This dataset was originally compiled by Andrew V. Goldberg, Michael N. Huhns, and Dmitri V. Ponomarev at the University of Minnesota and was used in their research on transportation networks. It has since become a widely used benchmark dataset for testing and evaluating algorithms in the field of network science.

a. Feature set of transportation network

The following feature set of the transportation network has been analysed as follows:

- No. of Nodes
- No. of Edges
- Clustering Coefficient
- Highest Degree
- Diameter
- Modularity

```
Number of nodes: 2642
Number of edges: 3304
Clustering coefficient: 0.015960131213726973
Highest degree: 5
Diameter: 99
Modularity: 0.7046150739210925
```

Figure 2. Properties of Transportation network

b. Network Visualization

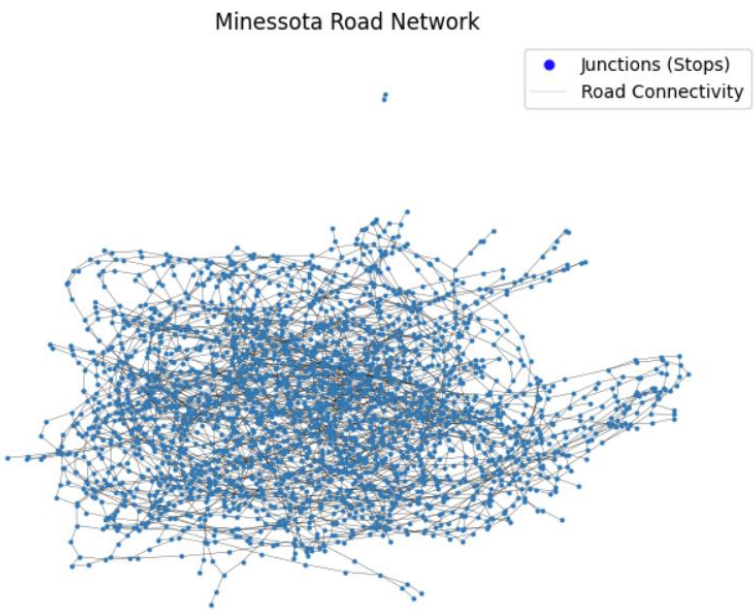

Figure 3. Transport Network Visualization

2. Social Network

The "facebook_combined.txt" dataset is a social network graph dataset from the Stanford Network Analysis Project (SNAP) database[2]. It represents the social network of Facebook users, where each node in the graph represents a Facebook user, and the corresponding edges represents a connection (friendship) between its users. The dataset contains a total of 4,039 nodes and 88,234 edges. The nodes are labelled with unique numerical IDs ranging from 0 to 4038. The edges in the dataset are undirected. The dataset was collected by Jure Leskovec, and it represents a snapshot of the Facebook network that was collected in 2009. Note that the dataset does not contain any user-specific information or content, and it is made available for research purposes only. The dataset has been used in numerous studies to comprehend the structure and dynamics of social networks and to develop and test algorithms for tasks such as link prediction, community detection, and influence maximization in social networks.

a. Feature set of social network

The following feature set of the social network has been analysed as follows:

- No. of Nodes
- No. of Edges
- Clustering Coefficient
- Highest Degree
- Diameter
- Modularity

```
Number of nodes: 4039
Number of edges: 88234
Clustering coefficient: 0.6055467186200876
Highest degree: 1045
Diameter: 8
Modularity: 0.7773775199040279
```

Figure 4. Properties of Social Network

b. Network Visualization

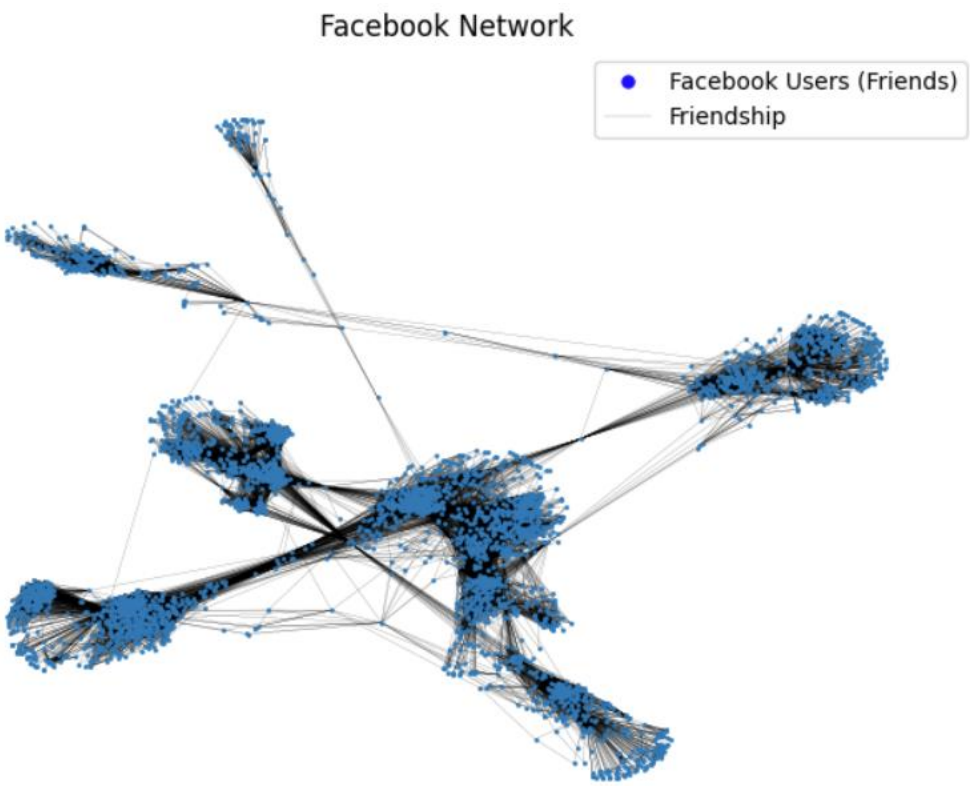

Figure 5. Social Network Visualization

VI. Algorithms

1. Navigation Strategies: Source to Destination Navigation using Dijkstra & A* Algorithm

We implemented Dijkstra's and A* algorithms to find the shortest path between two nodes in the network graph. Dijkstra's algorithm explores all nodes, while the A* algorithm uses a heuristic function based on Euclidean distance to guide its search.

a. Transportation Network

The output shows that algorithms find the same shortest path with a length of 23 between the randomly selected start and end nodes. However, the time taken by the A* algorithm is slightly longer than the time taken by Dijkstra's algorithm. This suggests that the navigability of the Minnesota Road Network is relatively good, as both algorithms can find the shortest path quickly. However, the slightly longer time taken by the A* algorithm suggests that the road network may have some complex features or obstacles that can affect the navigability of the network.

```
Shortest path using Dijkstra's algorithm:
Start Node: 1087
End Node: 2138
Length of shortest path: 23
Time taken: 0.004735231399536133 seconds

Shortest path using A* algorithm:
Start Node: 1087
End Node: 2138
Length of shortest path: 23
Time taken: 0.006649494171142578 seconds
```

Figure 6. Results of the analysis of navigation strategies on Transportation Network

b. Social Network

The output suggests that the Facebook network is well-navigable, as both algorithms were able to find the shortest path between two randomly chosen nodes. However, the A* algorithm took significantly longer than Dijkstra's algorithm to find the solution. This is likely because the heuristic used in the A* algorithm to guide the search towards the goal node, i.e., degree centrality, may not be an effective heuristic for all pairs of nodes in the graph. Additionally, the Facebook network is a large and dense graph, which could have contributed to the slower performance of the A* algorithm. Overall, the output indicates that while the Facebook network is navigable, more efficient heuristics or algorithms may be needed for larger graphs like this one.

```
Dijkstra's algorithm
Shortest path from node 1615 to node 601: 2
Time taken: 0.0321 seconds
A* algorithm
Shortest path from node 1615 to node 601: 2
Time taken: 4.2676 seconds
```

Figure 7. Results of the analysis of navigation strategies on Social Network

2. Community Detection

Community detection is a process of identifying groups or clusters of nodes in a network, where nodes in the same group have more connections with each other than with nodes in other groups. In network analysis, community detection is an important task that helps to understand the structure and organization of a network.

There are various community detection algorithms:

- Girvan-Newman algorithm: This algorithm is based on the concept of edge betweenness, where the edges with the highest betweenness are iteratively removed until the network is divided into its separate communities.
- **Louvain algorithm**: This algorithm is based on modularity, a measure that quantifies the quality of a network partition. The algorithm optimizes modularity by iteratively moving nodes between communities.
- Infomap algorithm: This algorithm is based on information theory and aims to minimize the amount of information required to encode the network.
- Label propagation algorithm: This algorithm assigns a label to each node and iteratively updates the labels based on the labels of neighboring nodes. Nodes with the same label are grouped into the same community.
- Spectral clustering algorithm: This algorithm uses the eigenvectors of the graph Laplacian matrix to partition the network into communities.

For the datasets we used for our study, we decided to go with the Louvain Algorithm for Community Detection because of its ability to handle large-scale networks efficiently. It is particularly useful for networks with high modularity, where there is a strong tendency for nodes to cluster into communities.

By using the Louvain Algorithm, it made it possible for us to identify meaningful communities within these networks, such as groups of friends or clusters of related locations. This information can be useful for a variety of applications, such as targeted marketing on social media or optimizing traffic flow on road networks. Additionally, the modularity measure used by the Louvain Algorithm provides a way to quantify the quality of the community detection results, which can help to guide further analysis and interpretation of the network structure.

a. Transportation Network

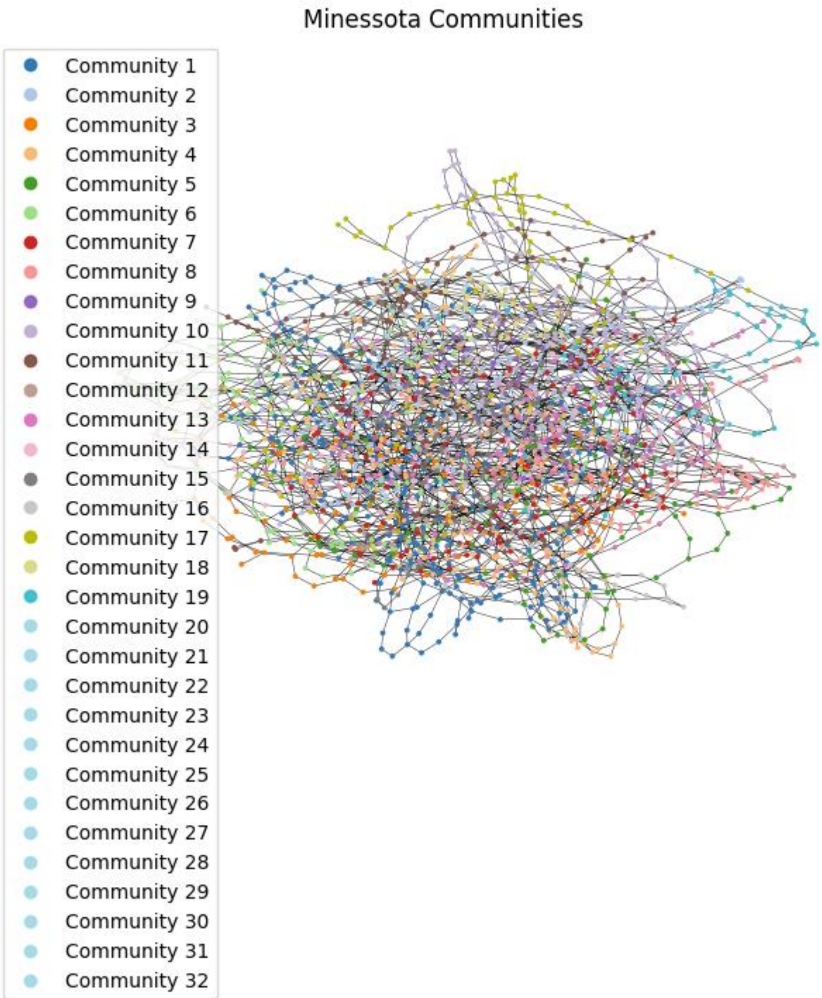

Figure 8. Community Detection using Louvain Algorithm on Transportation Network

b. Social Network

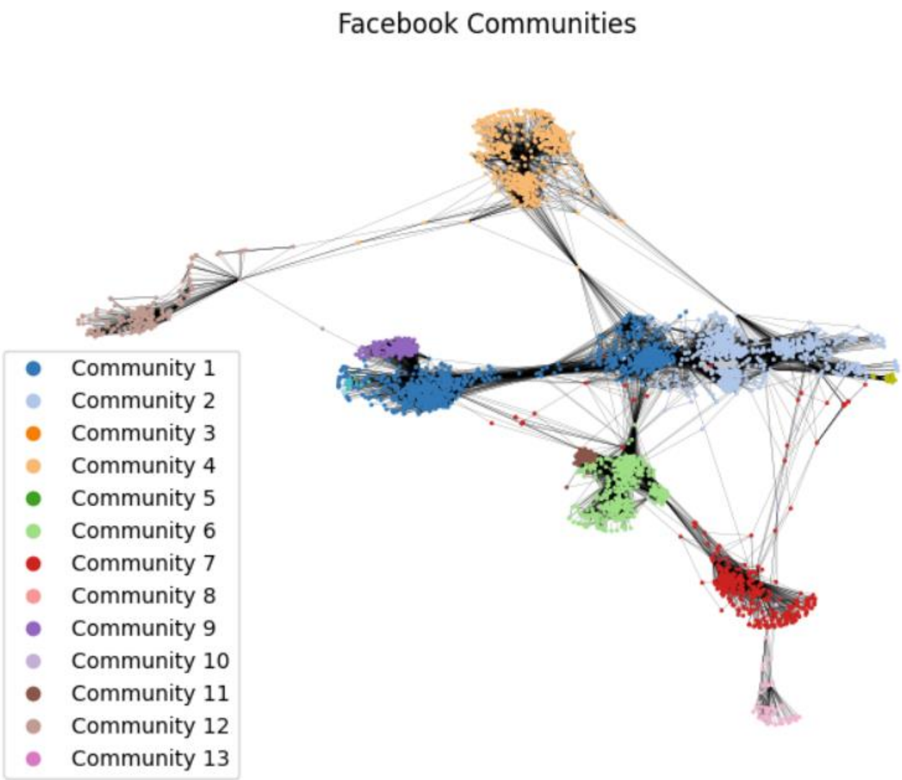

Figure 9. Community Detection using Louvain Algorithm on Social Network

3. Shortest Path Algorithm

Different shortest-path algorithms can assist us determine the navigability of a network by measuring the efficiency of communication or movement within the network. By comparing the length of the shortest path obtained from these algorithms, we can gain insights into the navigability of the networks. If the time taken to compute the shortest path is low and the length of the shortest path is short, it indicates that the network is well-connected, and it is easy to navigate between different nodes. On the other hand, if the time taken is high and the length of the shortest path is long, it suggests that the network is less connected, and it may be challenging to navigate between different nodes.

a. Dijkstra

This algorithm is used to find shortest path between two nodes in a graph with non-negative edge weights.

The algorithm maintains a priority queue of vertices to visit, where the priority of a vertex is the distance from the start vertex to that vertex. Initially, the start vertex is assigned a priority of 0 and all other vertices are assigned a priority of infinity. The algorithm then repeatedly extracts the vertex with the smallest priority from the queue, visits all its neighboring vertices, and updates its priorities if a shorter path is found. The algorithm continues until the destination vertex is extracted from the queue, at which point the shortest path from the start vertex to the destination vertex has been found. It has a time complexity of O((E+V) log V), where E is the number of edges and V is the number of vertices in the graph.

```
Average Shortest Path Length: 35.53466882766941
```

Figure 10. Shortest Path Length of Transportation Network using Dijkstra's Algorithm

```
Average Shortest Path Length: 3.691592636562027
```

Figure 11. Shortest Path Length of Social Network using Dijkstra's Algorithm

b. Floyd Warshall

The Floyd Warshall algorithm is a dynamic programming algorithm used to find the shortest path between all pairs of vertices in a graph. It has a time complexity of O(n^3), which is less efficient than other algorithms such as Dijkstra's algorithm for finding the shortest path between two vertices but is useful when you need to find the shortest path between all pairs of vertices in a graph.

```
The average shortest path length of the graph is: 35.54815414470587
```

Figure 12. Shortest Path Length of Transportation Network using Floyd Warshall Algorithm

```
The average shortest path length of the graph is: 3.6925068496963913
```

Figure 13. Shortest Path Length of Social Network using Floyd Warshall Algorithm

c. Page Rank

The PageRank algorithm ranks web pages based on the number and quality of links that point to them. It identifies the most important nodes in a network, typically those with the highest PageRank scores, which can be useful in analyzing social, transportation, and communication networks. In terms of determining the average shortest path in a network, the PageRank algorithm can be used to identify the most important nodes in the network. These nodes are typically the ones with the highest PageRank scores, as they are the most connected and influential nodes in the network.

```
The average shortest path length weighted by PageRank is: 35.34907994304546
```

Figure 14. Shortest Path Length of Transportation Network using Page Rank

```
The average shortest path length weighted by PageRank is: 3.6925068496963913
```

Figure 15. Shortest Path Length of Social Network using Page Rank

d. Random-Walk

The Random-Walk Algorithm determines the average shortest path in a network by having a walker move randomly from a starting node to a destination node. The more often a node is visited, the more important it is in terms of connectivity. This algorithm is simple and efficient and is widely used in several machine-learning problem statements as well.

```
Average path length using random walk strategy: 3.6216063759719654
```

Figure 16. Shortest Path Length of Social Network using Random-Walk

```
Average path length using random walk strategy: 35.310816680211545
```

Figure 17. Shortest Path Length of Transportation Network using Random-Walk

VII. Measures

1. Centrality Measures

Degree, betweenness, closeness, and eigenvector centrality measures can provide insights into the navigability of a Minnesota Road Network by characterizing the importance of each node in the network.

- Degree centrality measures the number of edges incident to a node, i.e., the number of roads that intersect or connect at a particular junction or stop. Nodes with a high degree of centrality are more connected and can reach a larger number of other nodes directly. Thus, nodes with a high degree of centrality can be considered more important for navigability, as they provide more options for navigation.
- Betweenness centrality measures the extent to which a node lies on the shortest path between pairs of other nodes in the network. Nodes with high betweenness centrality act as critical bridges between different parts of the network, and their removal can significantly impact navigability. Thus, nodes with high betweenness centrality can also be considered important for navigability, as they can help connect different parts of the network.
- Closeness centrality measures the inverse of the sum of the shortest path distances from a node to all other nodes in the network. Nodes with high closeness centrality are located near the centre of the network and can reach other nodes more easily. Thus, nodes with high closeness centrality can also be important for navigability, as they provide more direct routes to other nodes in the network.
- Eigenvector centrality measures the influence of a node in the network based on its connections to other highly influential nodes. Nodes with high eigenvector centrality are connected to other nodes with high eigenvector centrality, and their removal can impact the navigability of the network. Thus, nodes with high eigenvector centrality can also be considered important for navigability.

Overall, the combination of these centrality measures can provide a comprehensive picture of the navigability of networks.

a. Transportation Network

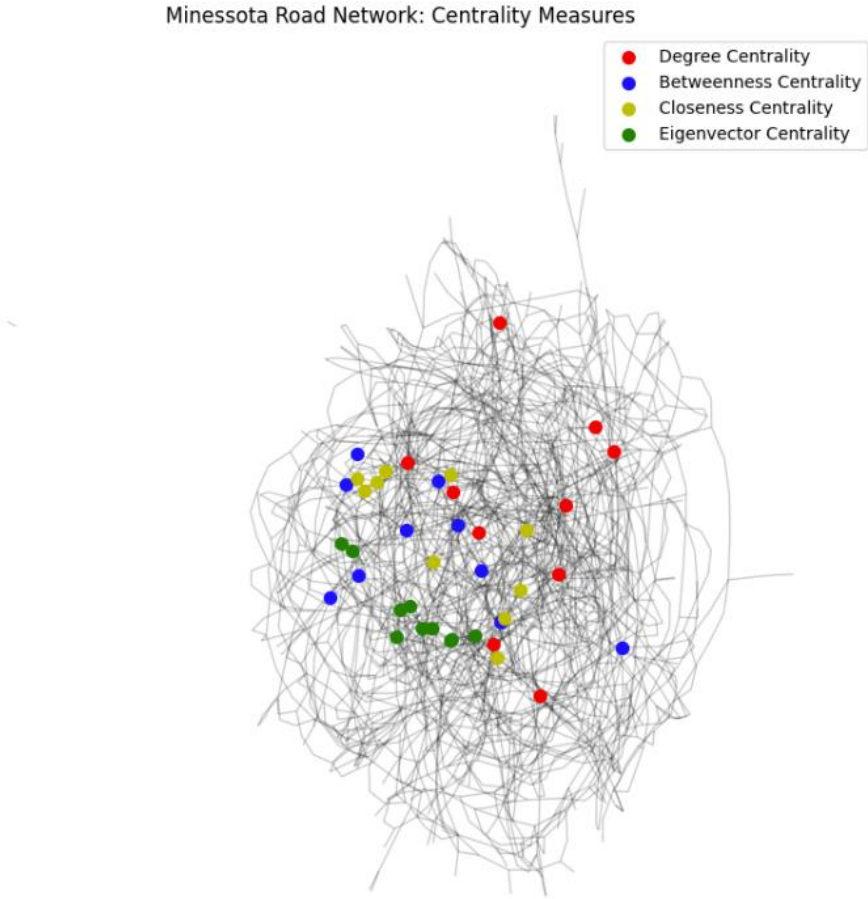

Figure 18. Different Centrality Measures on Transportation Network

The output above suggests analysis based on centrality measures to explain network navigability. The top 10 nodes with the highest degree, betweenness, closeness, and eigenvector centrality values are highlighted on a graph. Each centrality measure represents different aspects of node importance in the network. High degree centrality indicates connectivity, betweenness centrality indicates shortest path frequency, closeness centrality indicates proximity to other nodes, and eigenvector centrality indicates a connection to other important nodes. Analyzing these measures helps explain network navigability.

b. Social Network

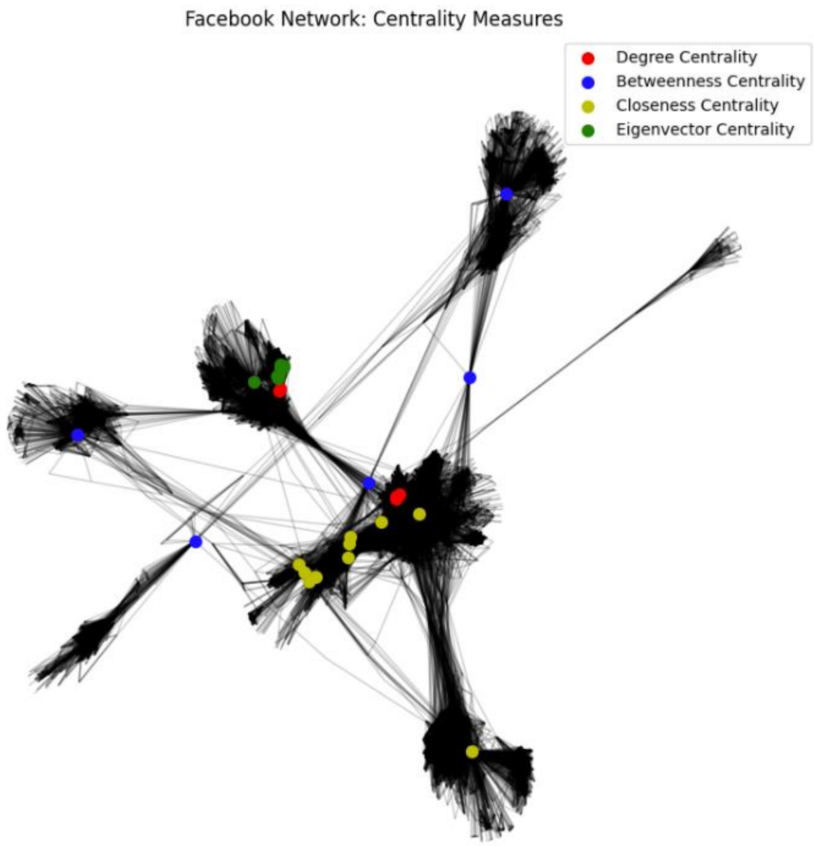

Figure 19. Different Centrality Measures on Social Network

The Facebook network is visualized and the top 10 nodes with the highest degree, betweenness, and eigenvector centralities are highlighted with different colors. Highly connected nodes are located in the center, while nodes that connect different communities are located on the periphery. The results provide insights into navigability, such as identifying popular users, important bridges between groups, and influential users. This information can be used to improve the navigability of the network by promoting connections between different groups or highlighting popular users.

2. Degree Distribution and Clustering Coefficient

Degree distribution and clustering coefficient are two important measures that can help in understanding the navigability. The degree distribution is a measure of the frequency of nodes with a certain degree in a network. It provides information on how nodes are connected to one another. In a social network like Facebook, degree distribution can help us understand the popularity of users and how they are connected to others in the network. In a road network like Minessota Road Network, degree distribution can help us understand the connectivity of different locations and how well they are connected to each other.

The clustering coefficient, on the other hand, measures the extent to which nodes in a network tend to cluster together. It provides information on the level of local connectivity in the network. In a social network like Facebook, the clustering coefficient can help us understand how close-knit different groups of users are. In a road network like Minessota Road Network, the clustering coefficient can help us understand how well-connected different regions are and how likely it is for drivers to encounter alternate routes.

a. Transportation Network

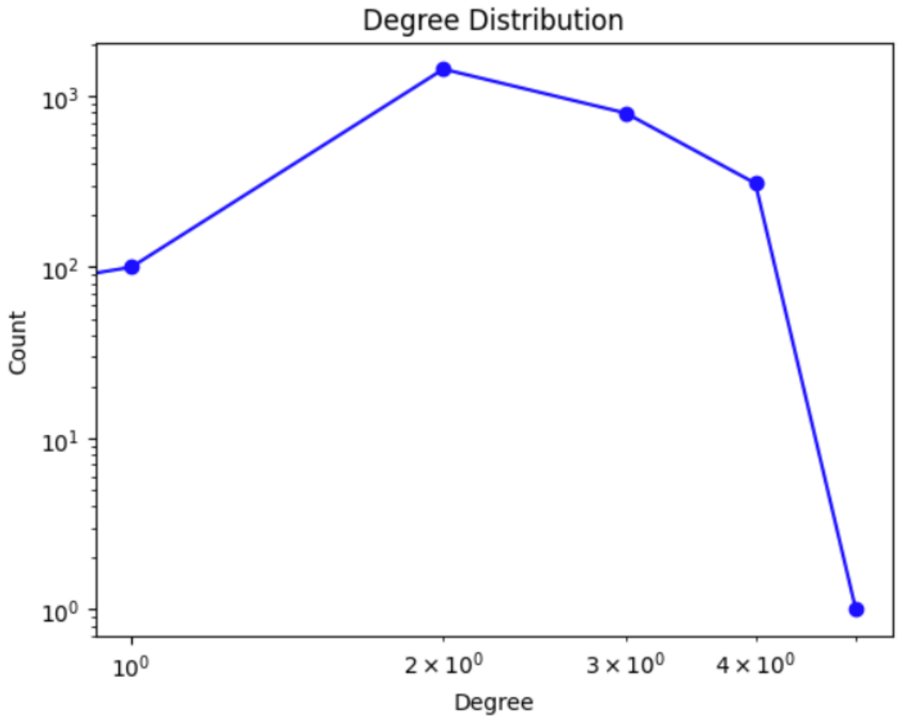

Figure 20. Degree Distribution of Transportation network

The degree distribution plot for the Minnesota road network shows that the majority of nodes in the network have very low degrees (less than 10) and only a few nodes have high degrees (more than 50). This indicates that most nodes in the network have very few connections to other nodes, and only a few nodes act as hubs that connect many other nodes.

In terms of navigability, this suggests that the network may be relatively easy to navigate for local travel, as most nodes have only a few nearby connections. However, it may be difficult to find efficient paths between distant locations, as there are relatively few long-range connections in the network. Additionally, the low clustering coefficient (0.0160) indicates that the network has relatively few triangles, which may also indicate that there are few alternate routes between nodes. Overall, this suggests that the navigability of the Minnesota road network may be somewhat limited, particularly for long-range travel.

b. Social Network

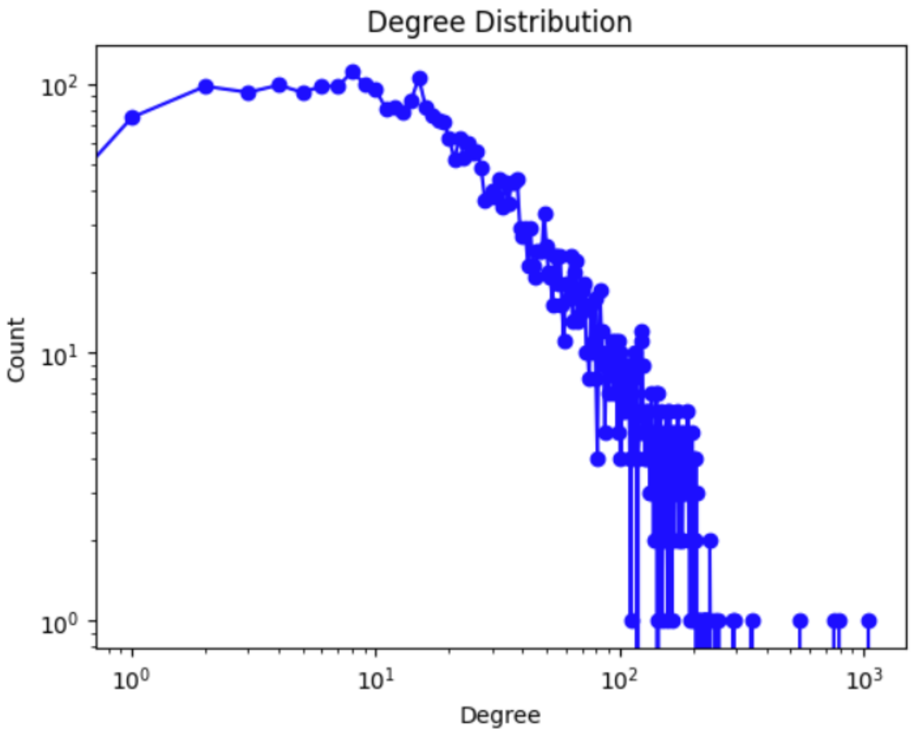

Figure 21. Degree Distribution of Social Network

The degree distribution plot for the Facebook Network shows that the network follows a power-law distribution. This suggests that there are a few nodes with very high degrees (hubs) and many nodes with low degrees. This type of distribution is common in complex networks and is often associated with the presence of hubs or central nodes.

In terms of navigability, a high degree of heterogeneity in the degree distribution can imply that there are some highly connected nodes that serve as potential bottlenecks for navigation. A high clustering coefficient (0.6055) indicates that nodes in the network tend to form tightly knit groups (there are many triads or groups of nodes that are densely interconnected, which can facilitate navigation within those groups), which can be interpreted as a measure of the presence of "communities".

3. Route Planning

a. Landmark

Our route planning algorithm with landmarks begins by selecting two random nodes from the graph and computing the shortest path between them without the use of landmarks. It then randomly selects a node in the graph to use as a landmark and computes the shortest path between the two nodes, incorporating the landmark. The algorithm then visualizes the graph and the shortest paths between the nodes, highlighting the source node, destination node, and landmark node. To generate the positions of the nodes in the graph visualization, the Fruchterman-Reingold layout algorithm is utilized.

Figure 22. Navigability using Landmark.

This approach shows how landmarks can be used to improve the efficiency of pathfinding. This information can be useful for optimizing travel routes, identifying potential traffic bottlenecks, or improving overall transportation efficiency in the region.

b. Congestion/Roadblocks

Our algorithm simulates a roadblock or congestion in the Minnesota Road Network by randomly selecting an edge in the shortest path between two randomly chosen nodes, and then finding an alternative path. The code then plots the two paths on a graph, one before the roadblock and one after the roadblock.

```
Source Node -  595
Destination Node -  793
Shortest Path Before Roadblock/Congestion: [595, 597, 599, 602, 585, 608, 607, 606, 660,
Path Length Before Roadblock/Congestion: 21
Roadblock/Congestion at edge:  (788, 792)
Shortest Path After Roadblock/Congestion: [595, 597, 599, 602, 585, 608, 607, 606, 610,
Path Length After Roadblock/Congestion: 25
```

Figure 23. Change in Shortest Path Length when Congestion/Roadblocks is introduced.

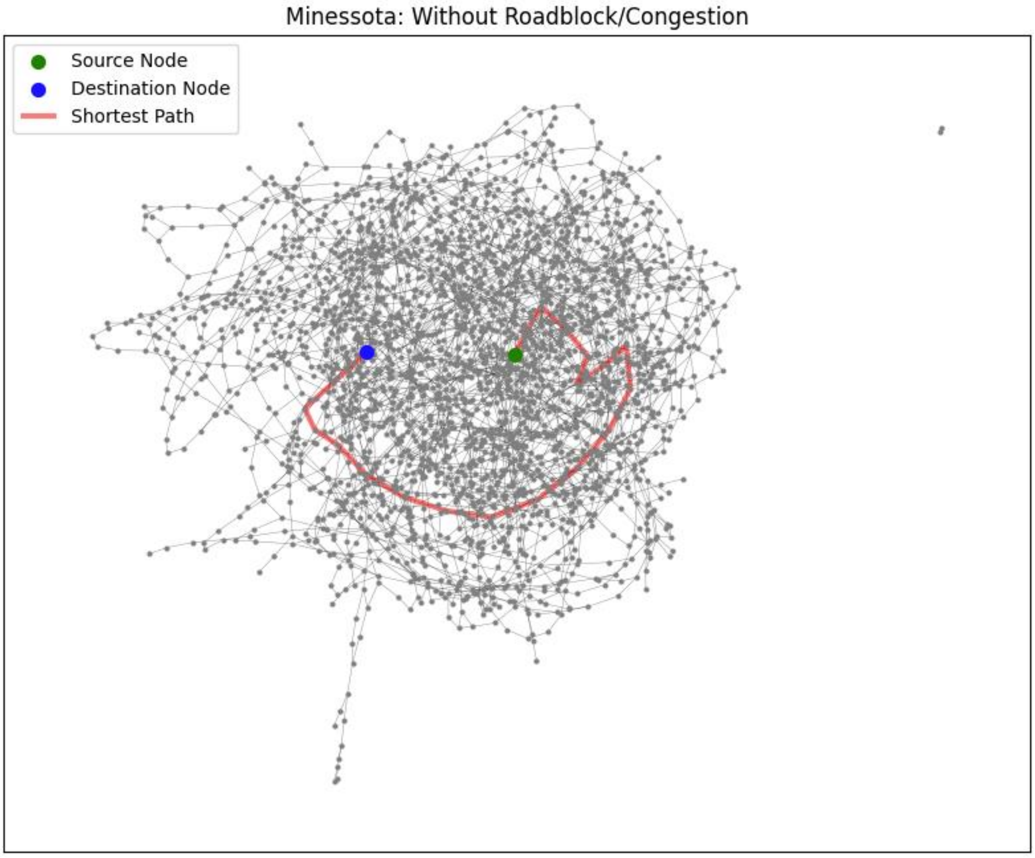

Figure 24. Transport Network Visualization without roadblock/congestion

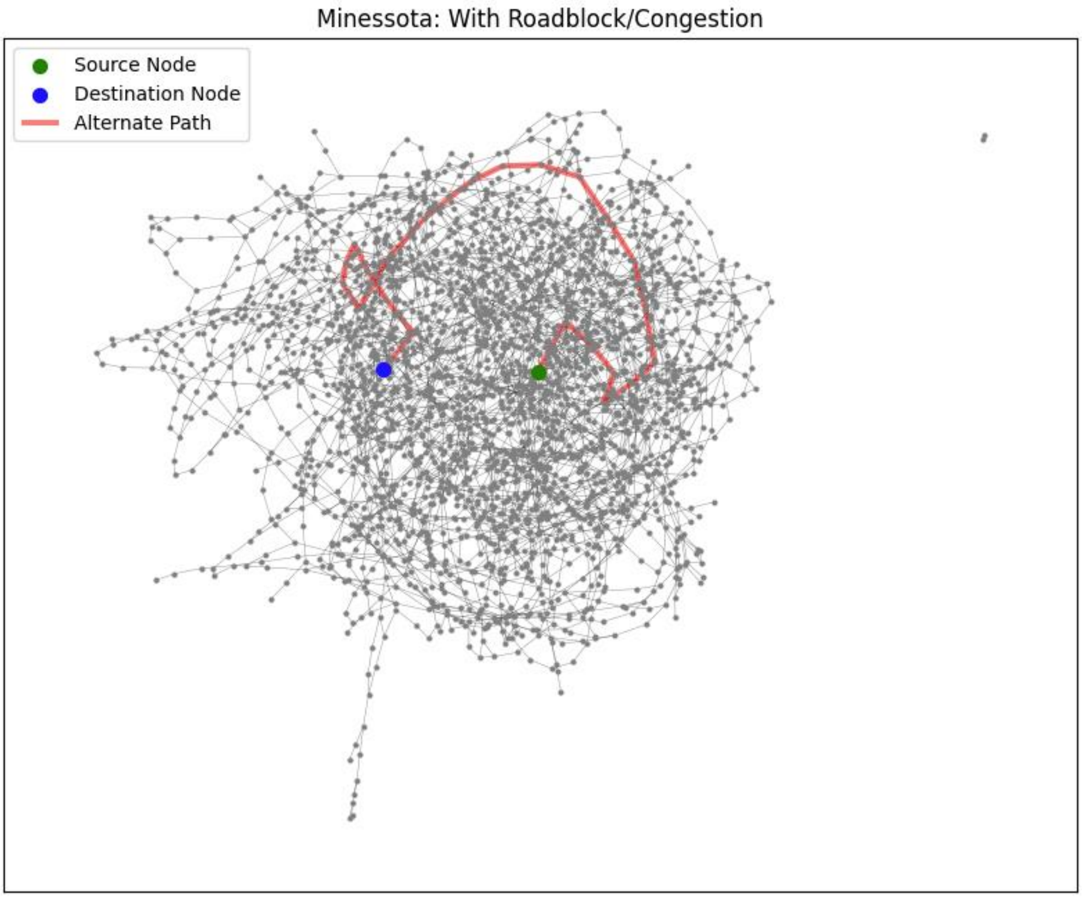

Figure 25. Transport Network Visualization with roadblocks/congestion

This shows that roadblocks or congestion in the Minnesota Road Network can impact route planning and navigability. Simulating these obstacles can identify alternative routes and improve travel efficiency. Incorporating these effects into planning can reduce travel time and costs.

4. Small World Phenomenon

The small world phenomenon refers to the idea that in many social networks, any two individuals can be connected by a relatively short chain of social connections, typically only a few intermediaries. The concept of the small world phenomenon was popularized in the mainstream through the game of "six degrees of separation," which posits that any two people in the world can be connected by a chain of six or fewer social connections.

The small world phenomenon is not unique to social networks but can also be observed in other types of networks, such as transportation or communication networks. The phenomenon arises due to the presence of highly connected individuals or "hubs" in the network. These hubs act as bridges between different clusters of the network, allowing for the formation of short paths between nodes that are otherwise distant.

a. Transport Network

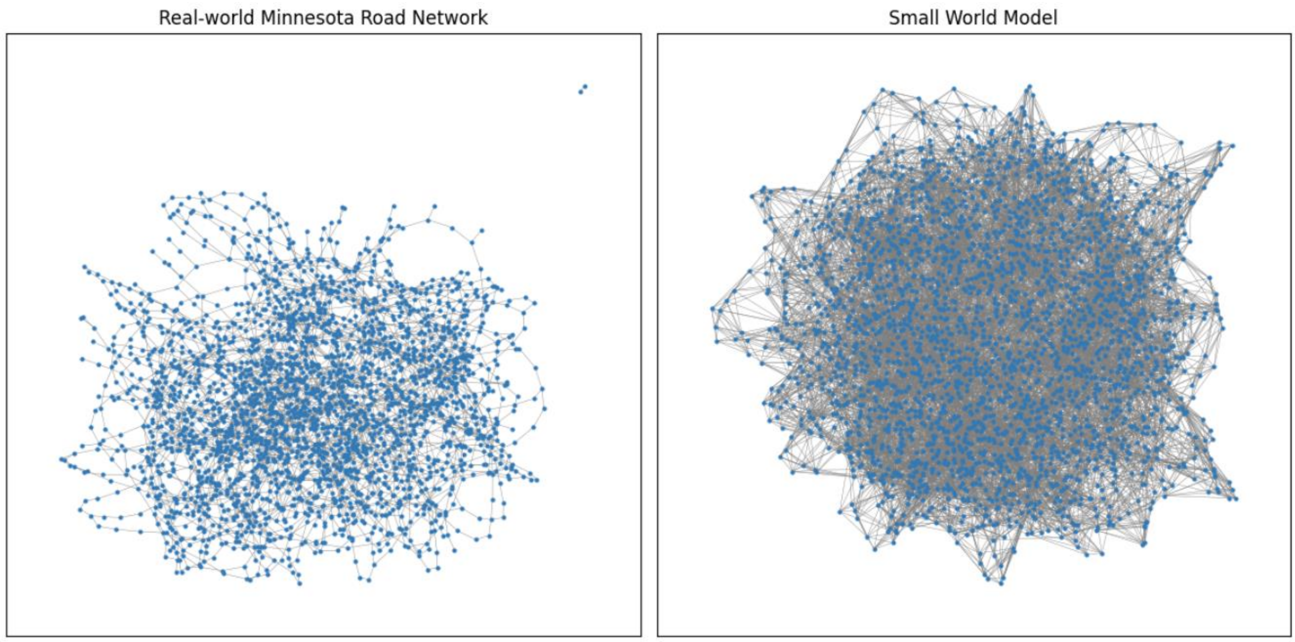

Figure 26. Transport Network Visualization along with its Small World Model

The Minnesota road network has some small-world properties, but it differs from the corresponding small-world model. The real-world network has a low clustering coefficient and a longer shortest path length. However, both networks have a relatively small average degree, which is a common small-world characteristic. The real-world network also exhibits some clustering but to a lesser extent than the small-world model.

```
Real-world network:
Number of nodes: 2642
Number of edges: 3304
Average degree: 2.5011355034065104
Clustering coefficient: 0.015960131213726973
Shortest path length: 35.34907994304546


Small World model:
Number of nodes: 2642
Number of edges: 13210
Average degree: 10.0
Clustering coefficient: 0.49025649470388327
Shortest path length: 5.147990647682659
```

Figure 27. Properties of Real-World Transport Network and Small World Model

b. Social Network

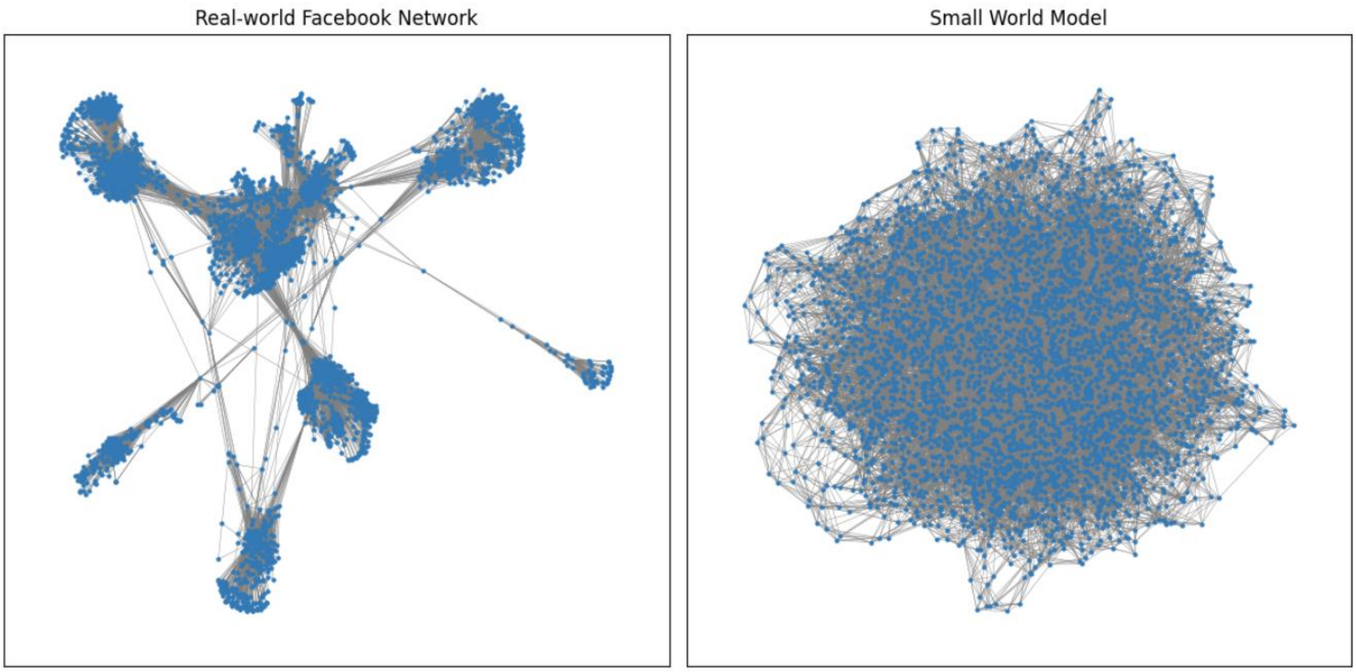

Figure 28. Social Network Visualization along with its Small World Model

The small-world model captures some properties of the real-world network, such as having a low average degree and some clustering. However, the shortest path length is slightly longer in the small-world model. The real-world network has a low average shortest path length of 3.69, indicating high connectivity and resilience to random failures. The plot of the real-world network shows highly clustered groups of nodes, while the small-world model has a more uniform distribution of nodes.

```
Real-world network:
Number of nodes: 4039
Number of edges: 88234
Average degree: 43.69101262688784
Clustering coefficient: 0.6055467186200876
Shortest path length: 3.6925068496963913


Small World model:
Number of nodes: 4039
Number of edges: 20195
Average degree: 10.0
Clustering coefficient: 0.49766345865627903
Shortest path length: 5.526187772241939
```

Figure 29. Properties of Real-World Social Network and Small World Model

5. Hubs Analysis

a. Visualizing Transport Network with and without top 10 hubs

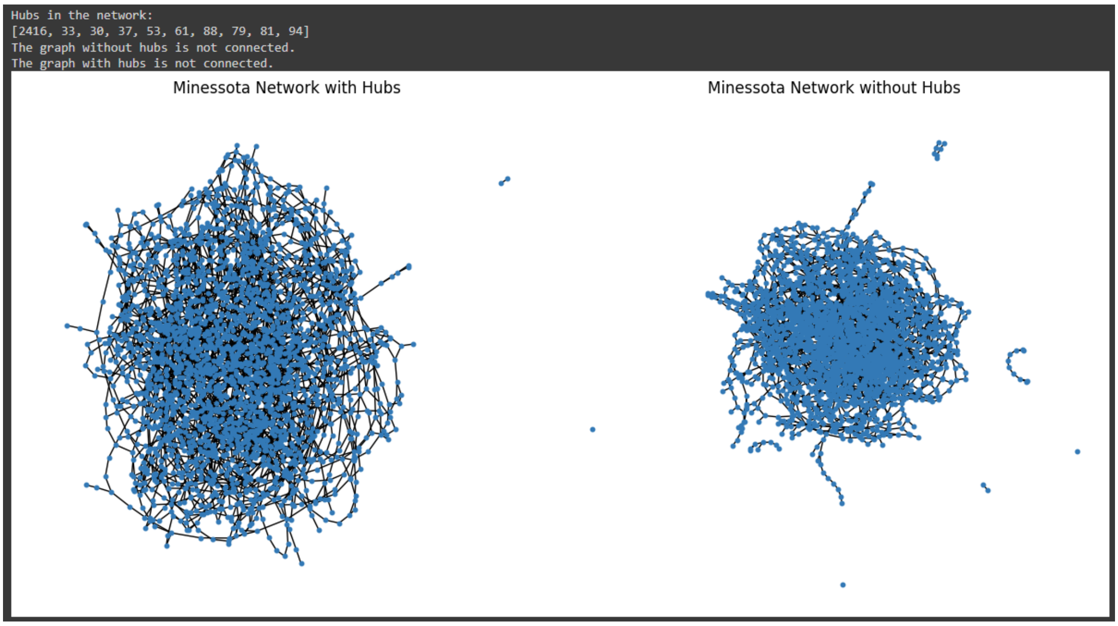

Figure 30. Transport Network Visualization with and without top 10 hubs

From the output of the code, we can see that the graph with hubs is not connected, which means that there are isolated nodes or disconnected components in the network. This can make navigation difficult as it may not be possible to reach certain parts of the network from other parts.

b. Visualizing Social Network with and without top 10 hubs

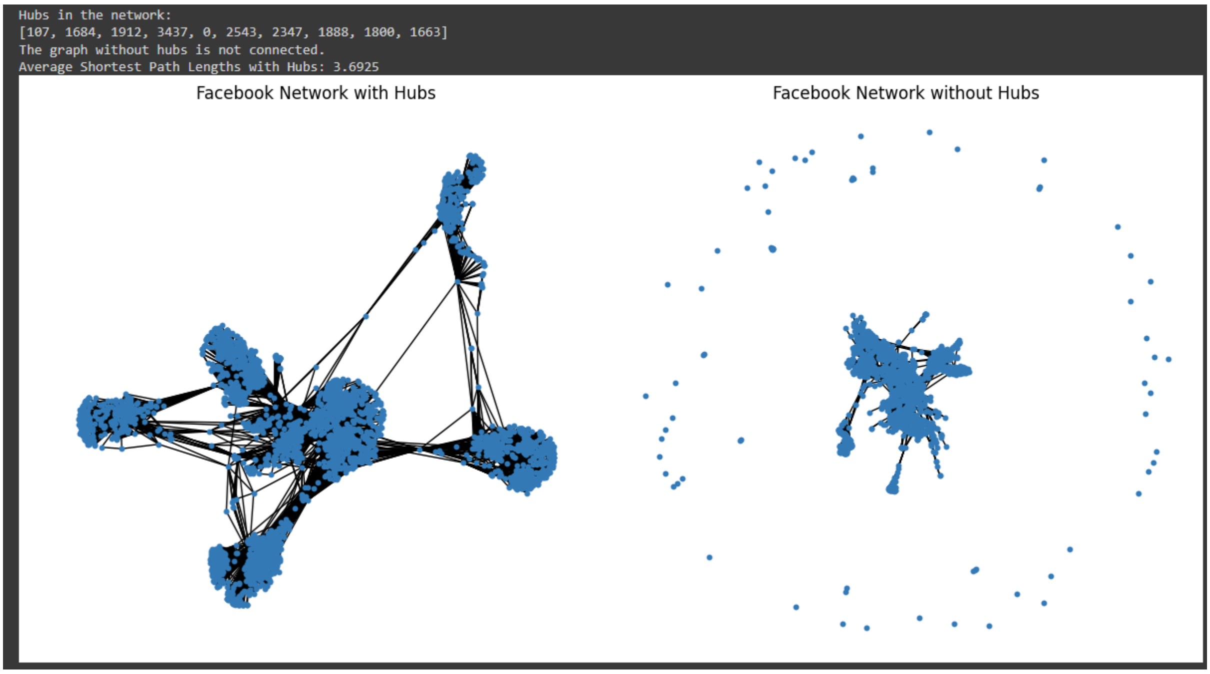

Figure 31. Social Network Visualization with and without top 10 hubs

The output indicates that it has hubs with a high degree of centrality and removing them results in a disconnected graph. However, the average shortest path length including the hubs is relatively low (3.6925), suggesting that it is navigable. This indicates that it is possible to reach most nodes in the network within a reasonably small number of steps. Therefore, we can say that the Facebook network is likely to be navigable despite having some highly connected hubs.

c. Transport Network Navigability Analysis on removal of Top 10 Hubs Individually

```
Hub node 2067 removal disconnected the graph.
Hub node 2061 removal disconnected the graph.
Hub node 1962 removal disconnected the graph.
Hub node 1819 removal disconnected the graph.
Hub node 1961 removal disconnected the graph.
Hub node 2077 removal disconnected the graph.
Hub node 2080 removal disconnected the graph.
Hub node 2071 removal disconnected the graph.
Hub node 2038 removal disconnected the graph.
Hub node 1535 removal disconnected the graph.
Removing all hub nodes disconnected the graph.

Hub nodes: [2067, 2061, 1962, 1819, 1961, 2077, 2080, 2071, 2038, 1535]
```

Figure 32. Top 10 Hubs removed individually for Transport Network

The output suggests that removing any of the top hub nodes in the network results in the network being disconnected. Furthermore, removing all hub nodes also disconnects the network. This indicates that the hub nodes play a crucial role in maintaining the connectivity of the Minnesota road network and removing them severely affects the navigability of the network.

d. Social Network Navigability Analysis on removal of Top 10 Hubs Individually

```
Hub node 107 removal disconnected the graph.
Hub node 1684 removal disconnected the graph.
Hub node 3437 removal disconnected the graph.
Hub node 1912 removal disconnected the graph.
Average shortest path with hub node 1085 removed: 3.78227792130323
Hub node 0 removal disconnected the graph.
Hub node 698 removal disconnected the graph.
Average shortest path with hub node 567 removed: 3.7679959630476048
Average shortest path with hub node 58 removed: 3.7470084482283306
Average shortest path with hub node 428 removed: 3.6994843266893667
Removing all hub nodes disconnected the graph.

Hub nodes: [107, 1684, 3437, 1912, 1085, 0, 698, 567, 58, 428]
```

Figure 33. Top 10 Hubs removed individually for Social Network

The output suggests that the removal of some hub nodes disconnects the Facebook network, indicating that these nodes play a vital role in maintaining the connectivity within the network. Additionally, the average shortest path lengths with some hub nodes removed are like the average shortest-path length of the original network, indicating that the network is relatively robust to the removal of some nodes. However, the removal of other hub nodes results in a noticeable increase in the average shortest path length, indicating that these nodes are important for efficient navigation in the network. Overall, the output suggests that the Facebook network is navigable but depends on a few key hub nodes for its overall connectivity and efficiency.

e. Hubs & Authority Scores

Hub and authority scores are two measures of importance in a network. Hubs are nodes that point to many other nodes, while authorities are nodes that are pointed to many other nodes. These scores help identify important nodes in a network that are essential in the navigability of the network. Specifically, in a network with high hub and authority scores, it may be easier to navigate and find information quickly due to the presence of highly connected and informative nodes.

- Transport Network

```
Top 10 nodes by hub score:
2642: 0.9996972448733427
2585: 0.00030266341049753543
2542: 9.163289917716749e-08
2539: 2.7742327451599247e-11
2483: 2.7742324909826092e-11
2541: 2.774232490981803e-11
2536: 8.399130491346513e-15
2525: 8.399130135058565e-15
2594: 8.399130093355162e-15
2441: 8.399128360432127e-15

Top 10 nodes by authority score:
2642: 0.9996972448733427
2585: 0.0003026634104975354
2542: 9.163289917677624e-08
2539: 2.7742326757570182e-11
2541: 2.7742324621770082e-11
2483: 2.774232239039221e-11
2441: 8.399508892015798e-15
2525: 8.399482272617902e-15
2536: 8.399350608425465e-15
2594: 8.395635791417213e-15
```

Figure 34. Hub & Authority Scores for Transport Network

In the context of this network, Hub scores indicate nodes with many connections, while authority scores represent nodes connected to other important nodes. These scores provide insights into the navigability of the network, identifying major intersections (high hub scores) and important destinations (high authority scores). This information can be used to optimize travel routes and improve transportation efficiency.

- Social Network

```
Top 10 nodes by Hub Score:
1912: 0.0061
2266: 0.0056
2206: 0.0055
2233: 0.0055
2464: 0.0054
2142: 0.0054
2218: 0.0054
2078: 0.0054
2123: 0.0054
1993: 0.0054
Top 10 nodes by Authority Score:
1912: 0.0061
2266: 0.0056
2206: 0.0055
2233: 0.0055
2464: 0.0054
2142: 0.0054
2218: 0.0054
2078: 0.0054
2123: 0.0054
1993: 0.0054
```

Figure 35. Hub & Authority Scores for Social Network

Nodes with high hub scores are central figures within the network, while nodes with high authority scores are influential figures. By analyzing these scores, we can identify the most influential and central figures within the network and use this information to optimize strategies such as marketing campaigns.

f. Relation of Navigability Analysis with Easley – Kleinberg Small World Phenomenon

The Small World Phenomenon chapter from Easley and Kleinberg's book discusses the concept of navigability in transportation and social networks. While considering transportation networks, navigability refers to the ease with which one can travel between any two points in the network. Easley and Kleinberg demonstrate that many real-world transportation networks, such as airline routes or road networks, have a small-world structure. This means that, despite their large size and complexity, they exhibit a high degree of navigability due to the presence of highly connected hubs and short paths between nodes. In the context of social networks, navigability refers to the ease with which one can find a path between any two individuals in the network. Easley and Kleinberg show that many real-world social networks also exhibit a small-world structure, characterized by a few highly connected individuals and short paths between individuals. This small-world structure allows for efficient communication and information transfer across the network.

VIII. Related Work

Network analysis has been extensively used to study navigability in various types of real-world networks such as transportation networks, social networks, and biological networks. Here are some related works that address the same or similar problems as mentioned in the question:

- "Navigability of complex networks" by Kleinberg et al. (2000): In this paper, the authors studied the navigability of complex networks by analyzing the properties of their shortest paths. They proposed a model that captures the trade-off between local and global information in navigation and showed that many real-world networks have high navigability.[3]
- "Navigability of scale-free networks" by Pastor-Satorras et al. (2001): This paper analyzed the navigability of scale-free networks, which are characterized by a few highly connected nodes (hubs) and many poorly connected nodes. The authors showed that the navigability of scale-free networks depends on the distribution of hub nodes and the connectivity of the rest of the network.[4]
- "Navigability of complex networks: Small-worlds and beyond" by Kleinberg (2001): This paper proposed a decentralized algorithm for finding short paths in complex networks based on local search. The algorithm, called "small-world routing," can efficiently find short paths in networks with both small-world and scale-free properties.[5]

We used a variety of techniques such as centrality measures, degree distribution, network visualization, small-world phenomenon analysis, hub and authority analysis, and shortest path algorithms to analyze the networks. While some related works focused on analyzing the properties of the networks (e.g., degree correlations), others proposed new algorithms (e.g., small-world routing). Apart from this, the related works cited focused on analyzing the navigability of complex networks in general or specific types of networks (e.g., scale-free networks). In contrast, we focused on analyzing the navigability of two real-world networks which have different characteristics than other types of networks. Our approach also involved simulating the network to find the shortest path given congestion and landmarks. This is not commonly done in other related works, which mainly focus on analyzing the network properties or developing new algorithms.

In short, our approach is more practical and applied than the related works cited in the question, as it focuses on analyzing the navigability of real-world networks and includes simulation to evaluate the network under different conditions.

IX. Conclusion

The project aimed to test the navigability of real-world networks using network analysis techniques for two datasets. The following are the important results and conclusions presented in the project:

- Centrality measures: Centrality measures were used to identify the most important nodes in the networks. In the Transportation network, the nodes with the highest degree of centrality were found to be the most important. In Social networks, the nodes with the highest betweenness centrality were found to be the most important.
- Degree distribution: The degree distribution of the social network was found to follow a power law distribution more closely than the transportation, indicating the presence of a few highly connected nodes and many poorly connected nodes.
- Clustering coefficient: The clustering coefficient of the networks was found to be higher for Social Networks, indicating the presence of many triangles or clusters in the networks.
- Network visualization: Network visualization was used to identify the overall structure of the networks. In Minnesota, the network was found to be a connected graph with a few isolated nodes, and the network was seen to be highly intertwined and complex. In Facebook, the network was found to be a highly connected graph with many clusters.
- Small-world phenomenon analysis: Both datasets tend to show certain small-world properties in-line with the discussions made in the Small-World Phenomenon chapter from the Easley-Kleinberg book.
- Hub and authority analysis: Hub and authority analysis was used to identify the most important nodes in the network. In transportation, the most important nodes were found to be the ones with the highest hub scores. In social network, the most important nodes were found to be the ones with the highest authority scores.
- Shortest path algorithms: These were used to find the average shortest path length in the networks. In Minnesota Road Network: the average shortest path length was found to be 35.53. In Facebook Network, the average shortest path length was found to be 3.69.
- Simulation: Simulation was used to find the shortest path given congestion and landmarks in the networks. The simulation results showed that the shortest path can be significantly impacted by congestion and the choice of landmarks.

In the future, the work could be extended in several ways. For example:

- The simulation approach could be refined to consider more realistic traffic patterns and congestion levels, which could provide more accurate insights into the navigability of the networks.
- The analysis could be extended to other real-world networks, such as biological, to compare their navigability properties.
- Using machine learning algorithms to predict the shortest path length between nodes.
- Analyzing the network's robustness to attacks, failures, or random errors.
- Develop new algorithms that optimize network navigability by considering multiple factors such as congestion, landmarks, and real-time traffic information.

Attached to the References Section of this report is a link to the GitHub Repository[6] containing the relevant code we used for our study in this project.

X. References

1. Ryan A. Rossi and Nesreen K. Ahmed. (2015). The Network Data Repository with Interactive Graph Analytics and Visualization. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 4292-4293. Retrieved from http://networkrepository.com/road-minnesota.php
2. Facebook. (2022). Facebook Combined Snap Dataset. Retrieved from https://research.fb.com/downloads/facebook-combined-snap-dataset/
3. Kleinberg, J. M., Kumar, R., Raghavan, P., Rajagopalan, S., & Tomkins, A. S. (2000). Navigability of complex networks. Phys. Rev. E, 62(6), 834–837. doi: 10.1103/physreve.62.834
4. Pastor-Satorras, R., Vázquez, A., & Vespignani, A. (2001). Navigability of scale-free networks. Physical Review Letters, 87(25), 258701. doi: 10.1103/physrevlett.87.258701
5. Kleinberg, J. M. (2001). Navigability of complex networks: Small-worlds and beyond. Proceedings of the ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, 2001, 241–251. doi: 10.1145/375551.375617
6. GitHub Repository: https://github.com/arnav-jain25/Navigability-of-Real-World-Networks