Social Network Analysis Project

Road Network

Samuel Tsegai

Dataset

- The analysis focuses on a road network dataset in a specific region of the United States (Chesapeake bay region).
- Intersections are **nodes**, and road segments are **edges** that connect the nodes.
- The network analysis covers microscopic, mesoscopic, and macroscopic viewpoints.
- Analyzing this graph provides insights into the structural properties of the road network.
- The findings can be valuable for transportation planning and optimization.

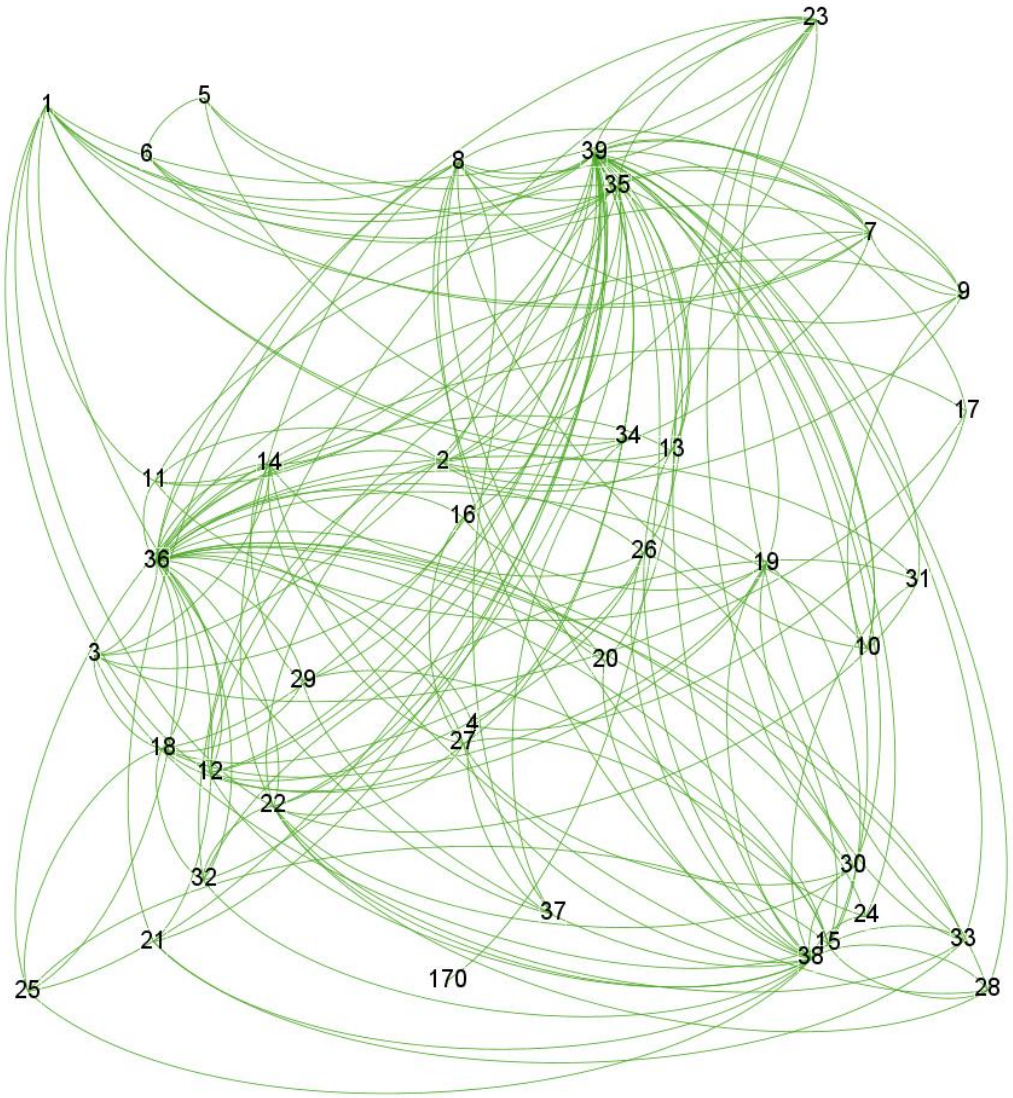
Reference: http://networkrepository.com/road-chesapeake.php

VISUALISATION AND BASIC NETWORK STATISTICS

- Number of nodes: 39
- Number of edges: 170

- **It is Undirected and Unweighted**

- No Isolated components

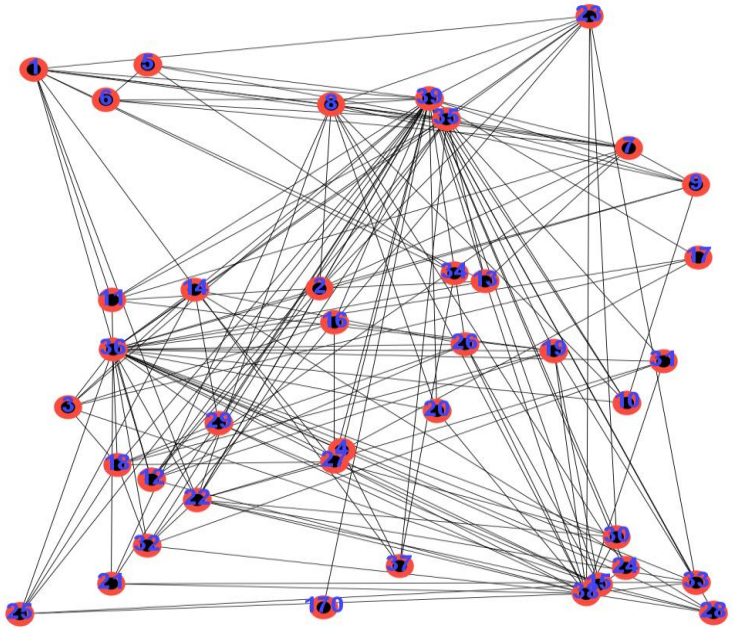


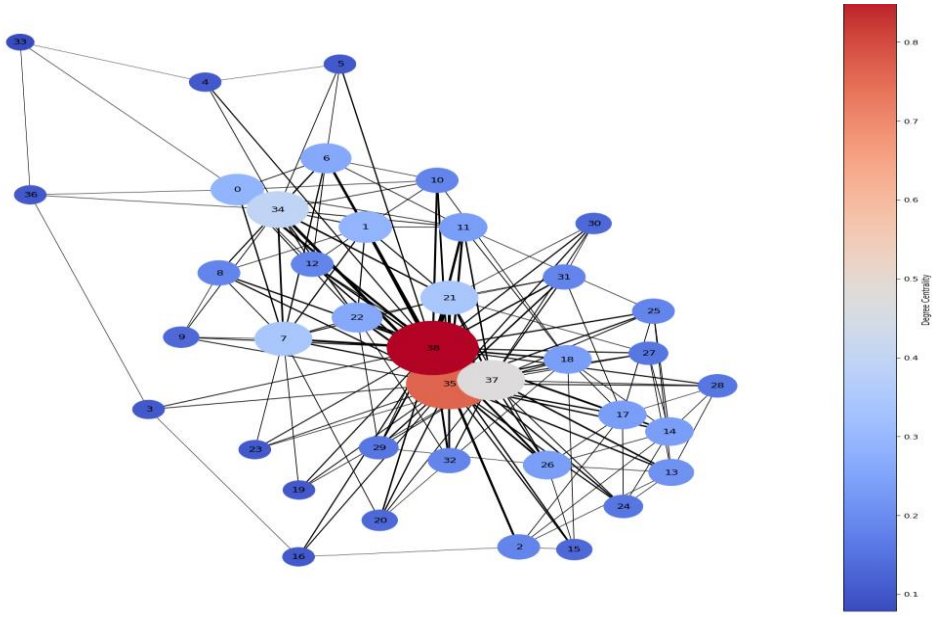
Exploration of the network

- **Average degree**: 8.717
- **Diameter**: 3
- **Average clustering coefficient**: 0.450
- **Density**: 0.229

- **Assortativity coefficient**: -0.3757

(this is **disassortative network** ,meaning high degree nodes tend to connect with low degree nodes)



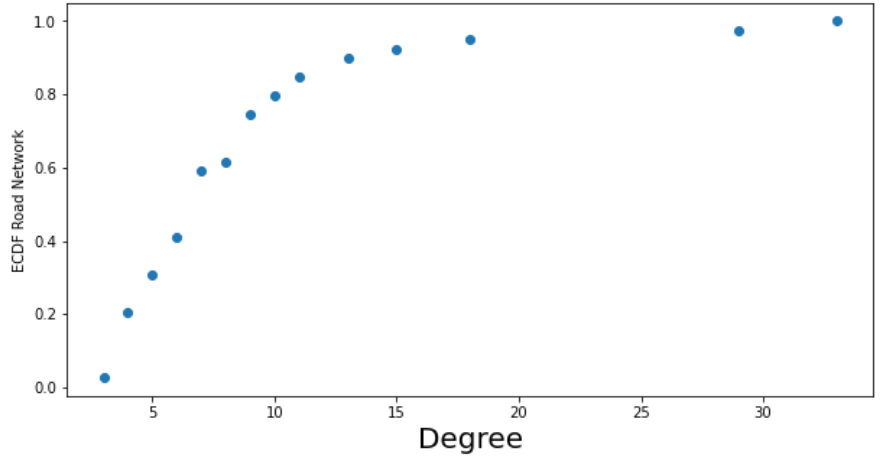


Assortativity is a preference for a network's nodes to attach to others that are similar in some way

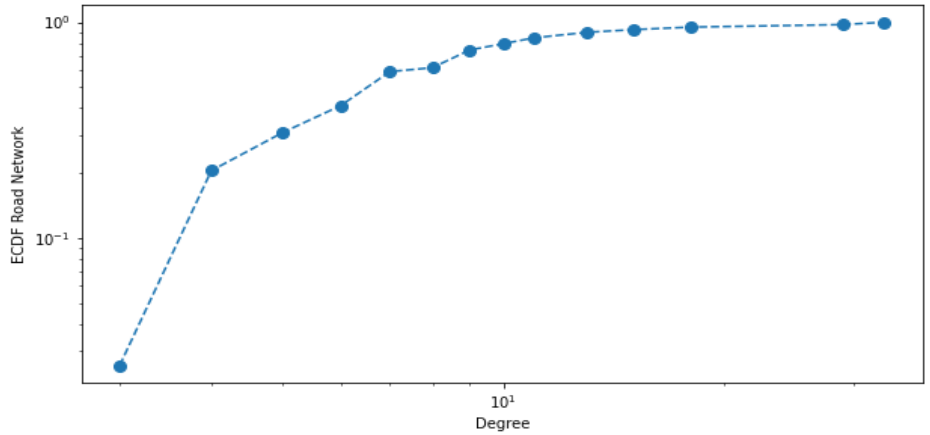
Connectivity

- No Isolated components
- Number of connected components is 1
- The number of total triangle is 582

ECDF in linear scale



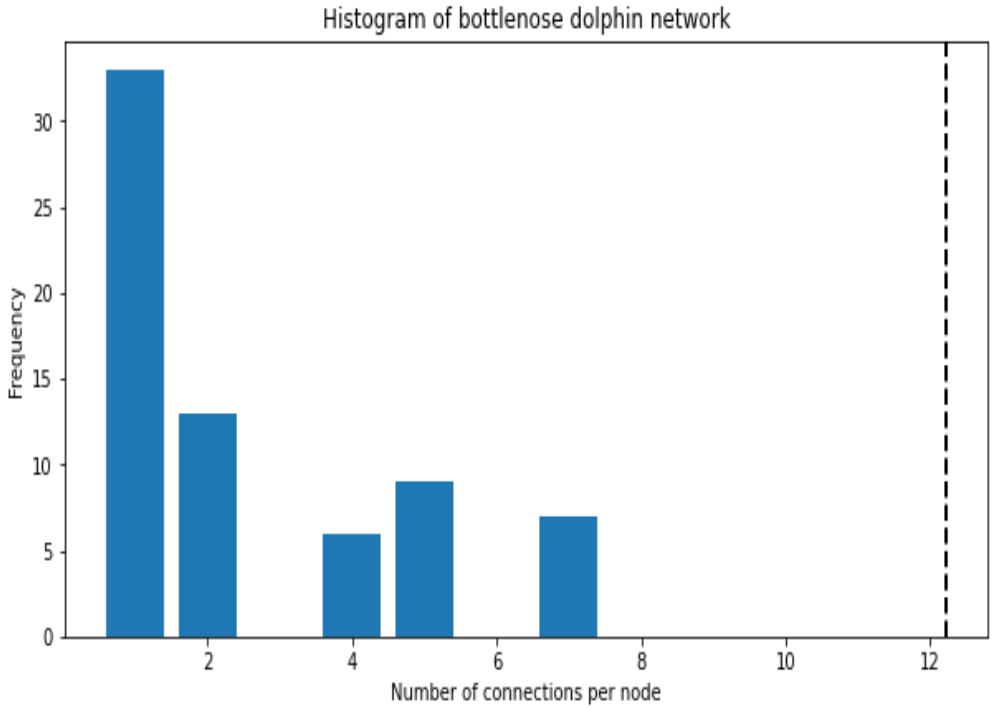
ECDF in log scale



ECDF linear and log plots to visualize the degree distribution of the network. We found that the degree distribution is right-skewed, meaning that most nodes have low degree values and a few nodes have very high degree values. The ECDF log plot helped us visualize this more clearly.

Bottlenose dolphin network

- The black dashed line represents the mean number of connections per node. We can see from the plot that the majority of nodes have a relatively small number of connections. However, there are also a small number of nodes with a large number of connections, up to a maximum of around 12 connections per node.
- Overall, this plot gives us a good visual understanding of the distribution of connections in the bottlenose dolphin network.
- For example, in a road network, we could use similar measures to analyze traffic flow, identify important intersections or highways, and optimize routes for efficient travel. Additionally, the same concepts of centrality, clustering, and community detection can be applied to identify key nodes, clusters of connected roads, and communities of roads that are heavily connected to each other.

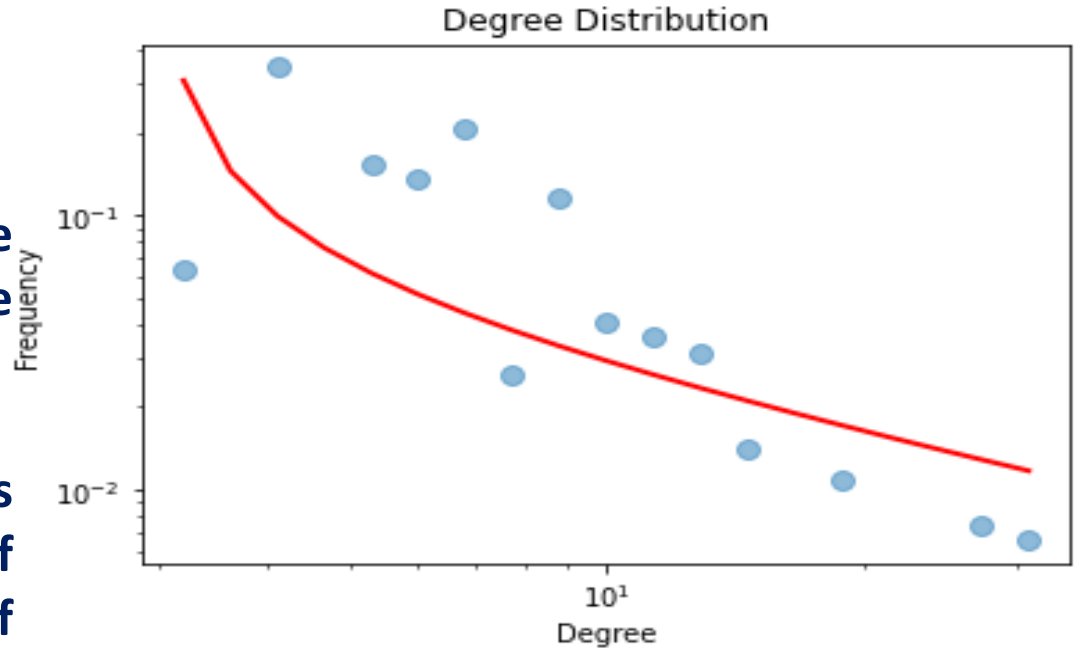


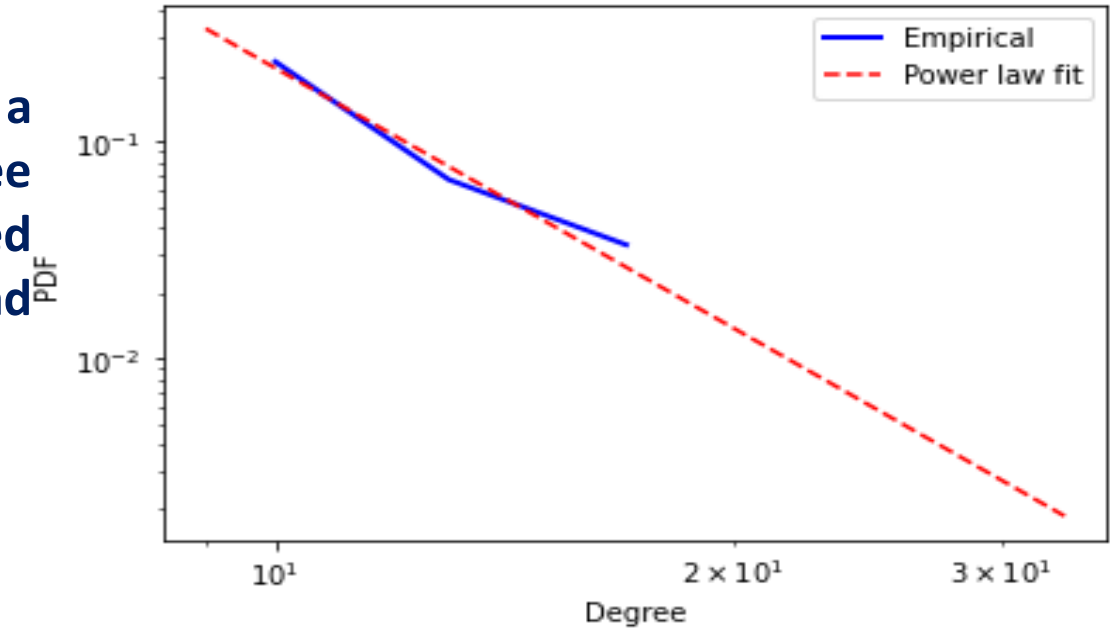
Network structure

- The plot helps to understand the distribution of node degrees in the network and can provide insights into the network's structure and dynamics.

- The log-log plot shows a heavy-tailed distribution, follows power law distribution ,where there are a small number of nodes with very high degrees (hubs) and a large number of nodes with low degrees.

- The power law fit shows that the distribution follows a power law, suggesting that the network exhibits a scale-free structure that is organized around a few highly connected nodes (hubs) critical for the network's connectivity and function.





Comparison with Random Network

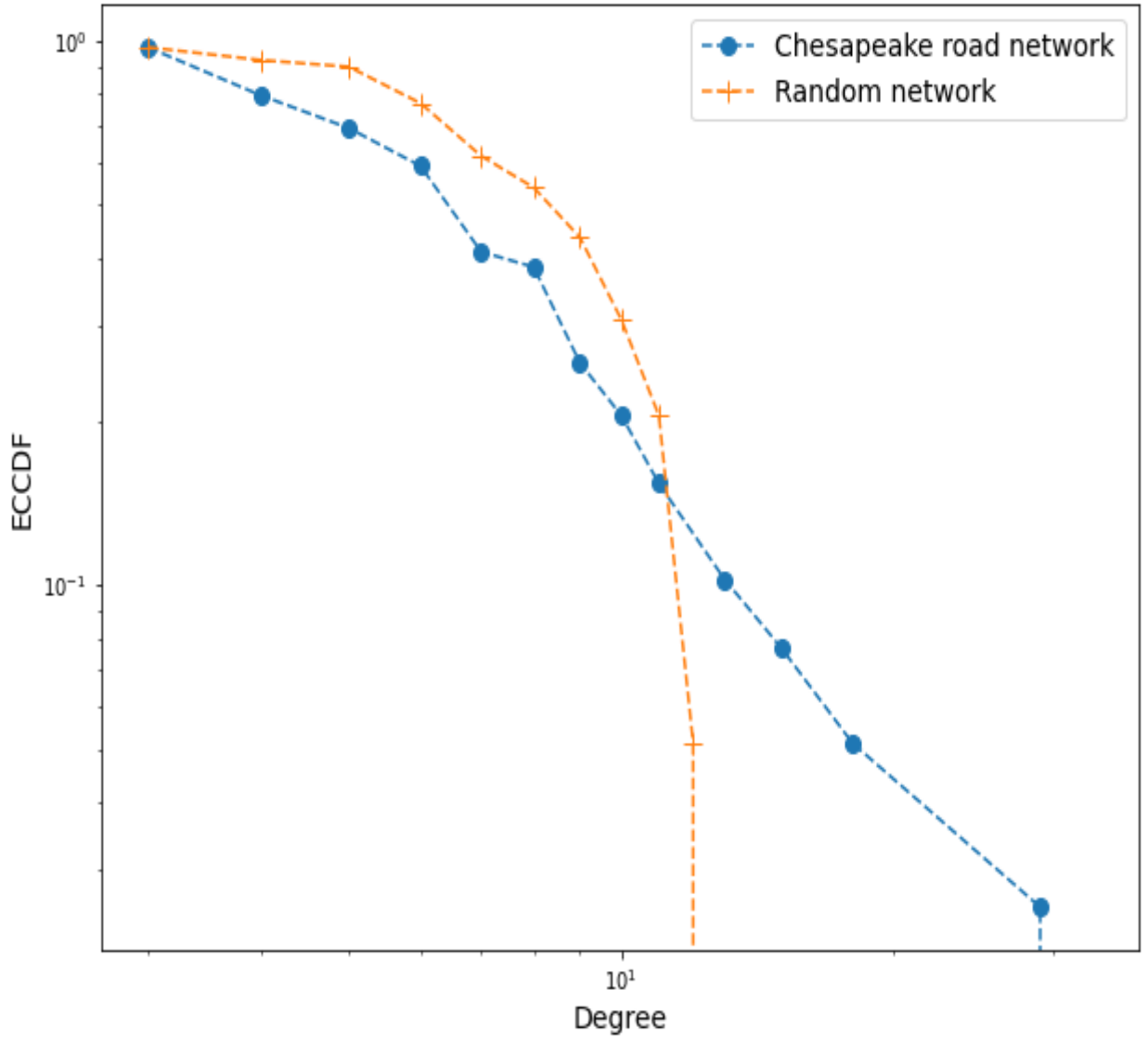
With the same number of nodes and edges It then computes the empirical complementary cumulative distribution function (ECCDF) for both networks and plots them on a log-log scale.

- **Using a Random Network with:**
  - **Number of nodes: 39**
  - **Number of links: 170**
  - **P = Density = Real-network density**

- **we can see that the degree distribution for the Chesapeake road network is highly right-skewed, with a long tail of high-degree nodes. This indicates the presence of highly connected nodes or "hubs" in the network.**

- **The ECCDF for the random network, on the other hand, shows a much more even distribution of degrees, indicating a lack of highly connected nodes.**

Real-network vs Random Network



Network Centrality

- This is based on the assumption that important nodes have many connections

Degree Centrality

```
10 most important nodes for Degree
Centrality:
(38, 0.8684210526315789)
(35, 0.763157894736842)
(37, 0.47368421052631576)
(34, 0.39473684210526316)
(7, 0.3421052631578947)
(21, 0.3421052631578947)
(0, 0.2894736842105263)
(1, 0.2894736842105263)
(6, 0.2631578947368421)
(22, 0.2631578947368421)
```

Eigenvector Centrality

```
10 most important nodes for Eigenvector
Centrality:
(38, 0.384635842084886)
(35, 0.345721957360852)
(37, 0.2250000830021762)
(21, 0.220949591126525 07)
(34, 0.20480242534727228)
(7, 0.20076771090174994)
(1, 0.20002557867157084)
(22, 0.18613518494319473)
(11, 0.17165715269597087)
(0, 0.1633142941719281)
```

Betweennes Centrality

```
10 most important nodes for Betweennes
Centrality:
(38, 0.3408661232459859)
(35, 0.20490490490490493)
(37, 0.05425726478358057)
(0, 0.050475200475200474)
(34, 0.04850100559482711)
(7, 0.026075556281505936)
(21, 0.024242132002261222)
(4, 0.014187408924251025)
(3, 0.01339045812730023)
(22, 0.010918281836305797)
```

Page Rank Centrality

```
10 most important nodes for Page Rank:
(38, 0.08971906295006934)
(35, 0.07852977389496887)
(37, 0.04944674407581577)
(34, 0.04272461711958419)
(7, 0.0363866761385929)
(21, 0.03566798375058441)
(0, 0.03248362605748524)
(1, 0.030333582406409 97)
(6, 0.02876087264229895)
(22, 0.027999842326237766)
```

Closeness Centrality

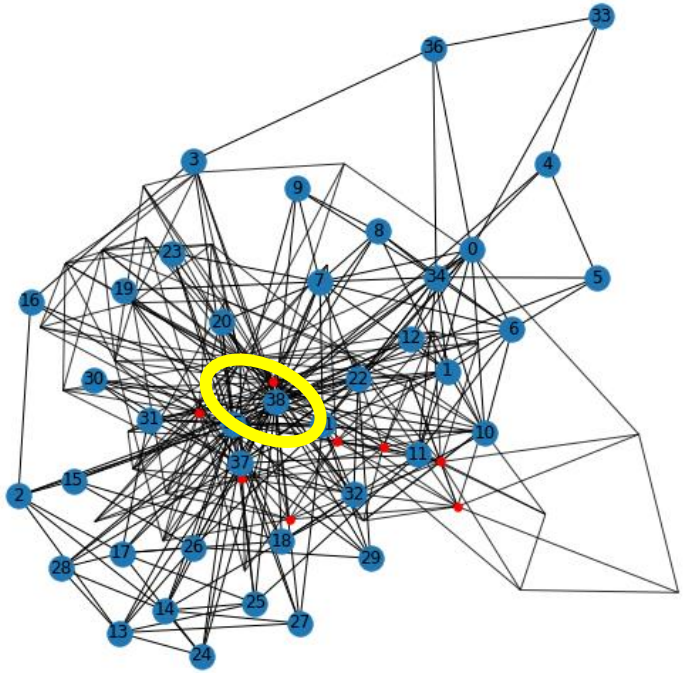
10 most important nodes for closeness Centrality:
(38, 0.8837209302325582)
(35, 0.7916666666666666)
(34, 0.6229508196721312)
(7, 0.6031746031746031)
(21, 0.6031746031746031)
(0, 0.5846153846153846)
(37, 0.5846153846153846)
(1, 0.5757575757575758)
(6, 0.5757575757575758)
(22, 0.5757575757575758)

Combined normalized centrality measures of top six

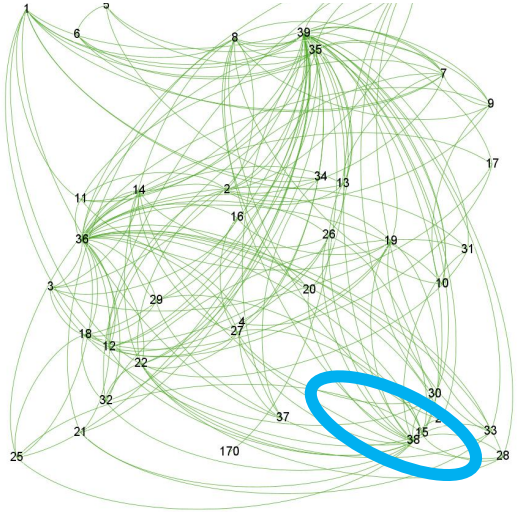
Node 38: 1.6861876871020793
Node 35: 1.3945819147755054
Node 37: 0.8034908071616644
Node 34: 0.6913362883332166
Node 21: 0.6237691237086819
Node 7: 0.605933166892 8774

HUBS

- 99-Percentile degree is about 31
- 1 nodes has a degree <= 31:
  - Node [38]



- **Hubs** are important in **scale free network** because they have a large number of connections to other nodes in the network, and therefore play a significant role in the flow of information or traffic in the network.



- **Hubs** can act as bottlenecks, as their high degree makes them more vulnerable to failure or disruption, which can significantly impact the overall network structure and function.

Transitivity

- **Transitivity** value is : 0.223

Transitivity is a measure of how interconnected the nodes of a graph are, and it reflects the probability that if node A is connected to node B and node B is connected to node C, then node A is also connected to node C. In our case, the transitivity value of 0.223 indicates that there is some clustering in the network, but it is not extremely dense.

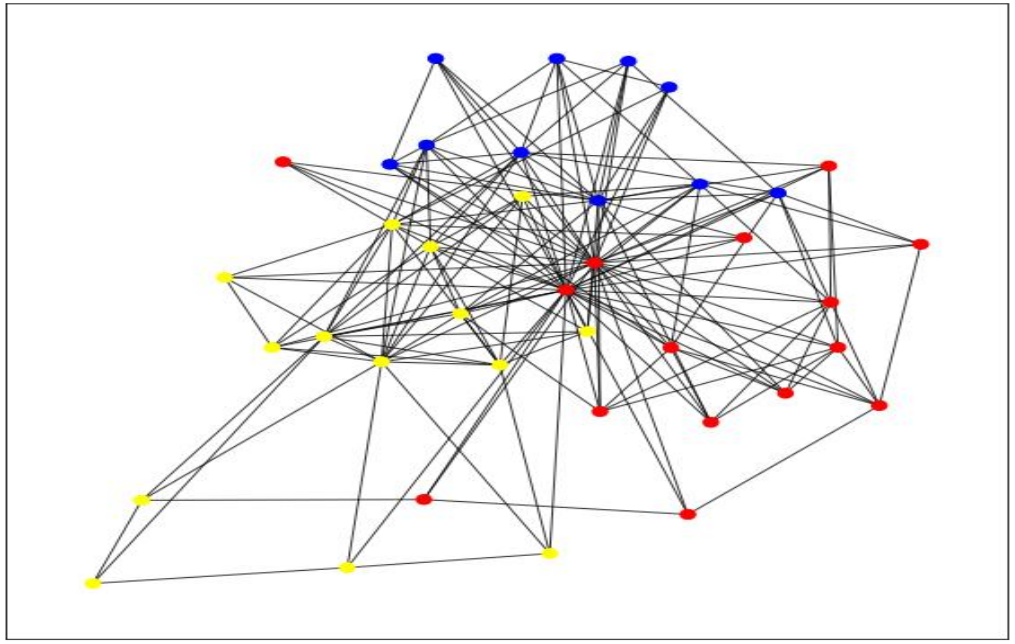
Communities

- Using set of **greedy partition** we obtain 3 communities

Community: 0 Number of elements 15

Community: 1 Number of elements 14

Community: 2 Number of elements 10



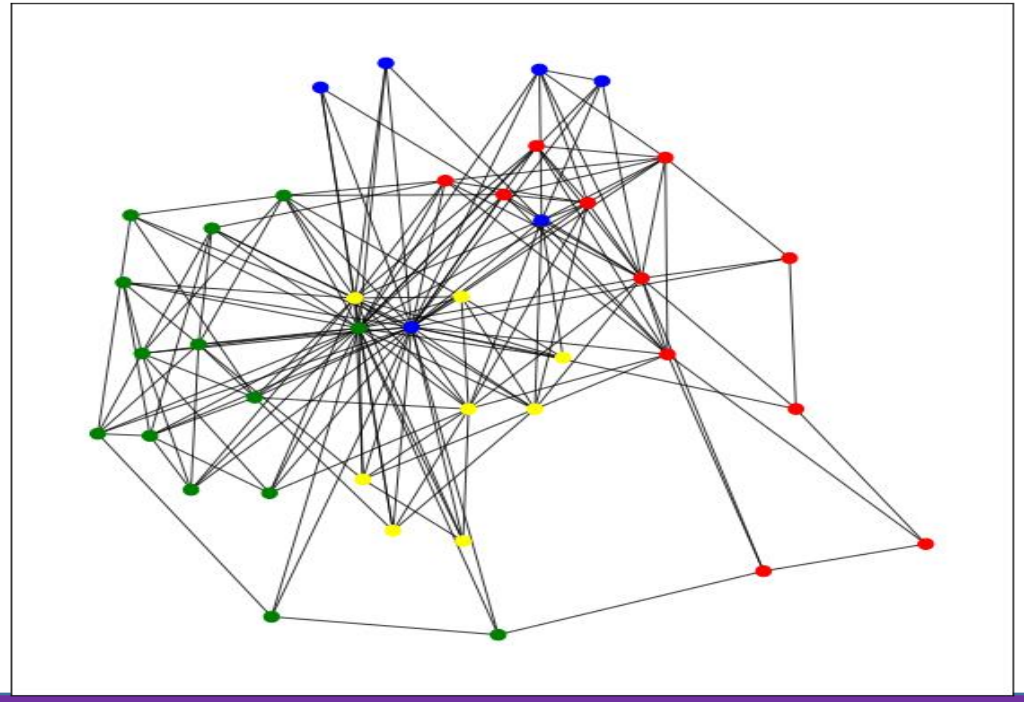
- Using set of **Louvain partition** we obtain 4 communities

Community: 0 Number of elements 11

Community: 1 Number of elements 12

Community: 2 Number of elements 8

Community: 3 Number of elements 8



Coverage, modularity and performance

Greedy

```
---
Coverage: 0.5941176470588235
Modularity: 0.2489446366782006 6
Performance: 0.717948717948718---
```

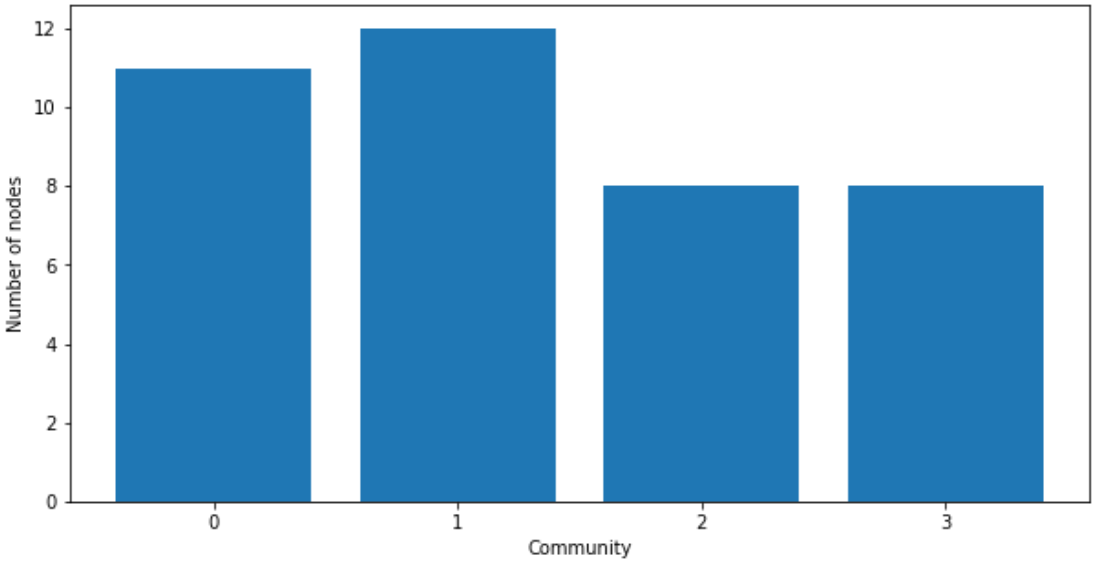
Louvain

```
---
Coverage: 0.5176470588235295
Modularity: 0.2594982698961937 3
Performance: 0.7692307692307693---
```

- **The coverage measures** the fraction of edges in the network that are within the community. Higher coverage indicates that more edges are captured within communities.
- **Modularity measures** how well the detected communities align with the network's internal structure. It ranges between -1 and 1, where values closer to 1 indicate better community structure.
- **Performance measures** how well the detected communities separate nodes in the network. It ranges between 0 and 1, where values closer to 1 indicate better separation of nodes into communities.

The results show that the Louvain method has higher Modularity and performance than the greedy method

Louvain Distribution



Conclusions

- The network has **39 nodes** and **170 edges** ,We found that the network has an **average degree of 8.717**, The **average clustering** coefficient of the network is **0.45**, indicating that intersections tend to be moderately densely interconnected and a **density of 0.229**. The diameter of the network is 3, indicating that it would take at most 3 road segments to travel between any two intersections in the network.
- The **assortativity** coefficient was -**0.3757**, indicating a **disassortative network**.
- . we computed the combined normalized measures of centrality to obtain a more **comprehensive ranking of nodes**, and we found that nodes **38, 37, 35, 34, 21,and 7** are the most important nodes in the network.
- **ECDF** plot for the degree distribution of the network, which showed that the distribution is heavy-tailed. It is a real network (scale free network)
- We also found **Hubs** that the 99th percentile of the degree distribution was **31.47999**
- We also computed the **coverage, modularity, and performance measures** for the two community sets obtained using the **greedy** and **Louvain methods**. We found that the Louvain method identified more communities than the greedy method and had a higher modularity score
- The analysis provides valuable insights into the structure and properties of the road network, which could be useful for transportation planners and policymakers in making decisions about road infrastructure improvements and traffic management.

THANK YOU !