Project unit 1: networks theory

Social network analysis
Data 8°B

Student:
Karina Campos Almeida

Teacher:
Didier Omar Gamboa Angulo

EuroRoads network analysis

***Abstract*—This paper explores the structural and functional characteristics of the EuroRoads network, a critical transportation infrastructure spanning Europe, by applying network analysis techniques, we examine the connectivity, centrality measures, degree distribution, and community detection within this network. Our analysis provides insights into the network's robustness, efficiency, and potential areas for improvement, highlighting the importance of strategic planning in enhancing the resilience and functionality of Europe's road transportation system.**

I. Introduction

The international E-road network, known as the EuroRoads network, represents a critical infrastructure spanning across Europe, this network consists of nodes and edges, where the nodes symbolize cities, and the edges denote direct E-road connections between these cities. The EuroRoads network is undirected, meaning that each road connection is bidirectional, allowing travel in both directions.

The importance of the EuroRoads network lies in its multifaceted applications and the potential insights it offers, from infrastructure planning and logistics to emergency management and economic impact analysis, understanding this network can significantly enhance various operational and strategic decisions, analyzing the connectivity of cities helps optimize transportation routes, minimize travel time, and improve fuel efficiency.

The EuroRoads dataset was notably utilized in the research paper "Robust network community detection using balanced propagation" by Lovro Šubelj and Marko Bajec, this study introduced an enhanced community detection algorithm, balanced propagation, which aimed to improve the stability and performance of the traditional label propagation method, the EuroRoads network served as a practical test case, demonstrating the algorithm's ability to identify communities that correspond to different geographical regions of Europe, thus showcasing its robustness and applicability in real-world scenarios.

II. Network characteristics

To begin the analysis of the EuroRoads network, I started by plotting the network, the original file contained nodes identified only by numbers, however, I had access to information about the cities corresponding to each node. To facilitate better interpretation and analysis of the information, I included the city names in the graph, below is the code used to load the data, assign city names to the nodes, and draw the network.

```
network_file = 'out.subelj_euroroad_euroroad'
names_file = 'ent.subelj_euroroad_euroroad.city.name
    '

edges = []
with open(network_file, 'r') as f:
    for line in f:
        if line.startswith('%'):
            continue
        u, v = map(int, line.split())
        edges.append((u, v))

node_names = []
with open(names_file, 'r') as f:
    for line in f:
        node_names.append(line.strip())

G = nx.Graph()
G.add_edges_from(edges)

name_mapping = {i+1: name for i, name in enumerate(
    node_names)}
G = nx.relabel_nodes(G, name_mapping)

plt.figure(figsize=(12, 12))
pos = nx.spring_layout(G, seed=42)
nx.draw(G, pos, with_labels=True, node_size=50,
    font_size=8, font_color='black', node_color='
    lightblue')
plt.title("EuroRoads network")
plt.show()
```

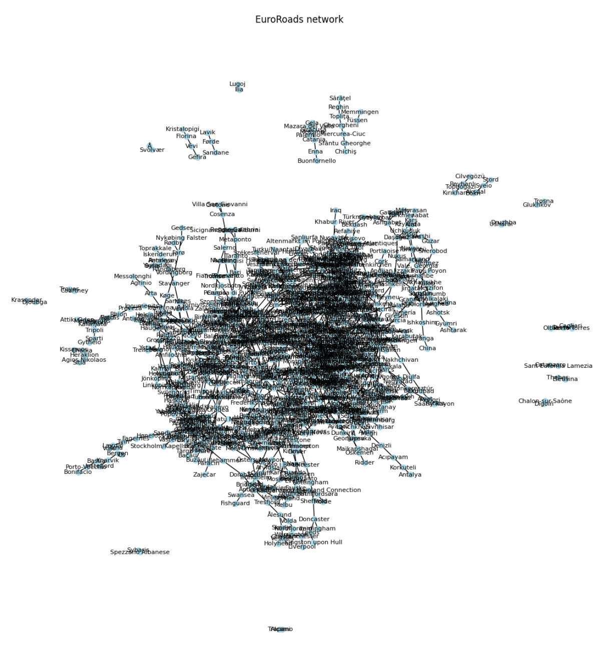

Fig. 1: Network plot

The resulting graph provides a clear visual representation of the EuroRoads network, with the city names facilitating the identification of connections between them, this visualization is crucial for understanding the structure and characteristics of the network.

Analyzing its characteristics provides valuable insights into its structure and connectivity, which are essential for various applications such as infrastructure planning, logistics, and emergency management.

```
size_of_network = G.number_of_nodes()
number_of_links = G.number_of_edges()
clustering_coefficient = nx.average_clustering(G)

# Diameter - considering only connected components
eccentricity = {node: nx.eccentricity(G.subgraph(c))[node]
                for c in nx.connected_components(G) for node in c}
diameter = max(eccentricity.values())

radius = min(eccentricity.values())
periphery = [node for node, ecc in eccentricity.items() if ecc == diameter]
center = [node for node, ecc in eccentricity.items() if ecc == radius]

# Sample of eccentricity
eccentricity_sample = dict(list(eccentricity.items())[:5])

{
    "Size of Network": size_of_network,
    "Number of Links": number_of_links,
    "Clustering Coefficient": clustering_coefficient,
    "Diameter": diameter,
    "Eccentricity (sample)": eccentricity_sample,
    "Radius": radius,
    "Periphery": periphery,
    "Center": center
}
```

The network consists of 1174 nodes, each representing a city, this large number of nodes indicates an extensive network that connects numerous cities across Europe, the connectivity between these cities is established through 1417 direct E-road connections or edges, these connections form the backbone of the transportation network, facilitating the movement of goods and people across the continent.

One of the key metrics used to analyze the network is the average clustering coefficient, which measures the degree to which nodes in the network tend to cluster together, for the EuroRoads network, the average clustering coefficient is approximately 0.0167. This relatively low value suggests that the cities are not highly clustered, meaning there are few closed triplets or triangles in the network, this characteristic indicates that the network is more linear and spread out rather than forming tightly-knit clusters of cities, such a structure is typical for large-scale transportation networks where direct connections are prioritized over forming local clusters.

The diameter of the network, which is the longest shortest path between any two nodes, is 62, this large diameter reflects the broad geographic spread of the network, indicating that it takes up to 62 edges to travel between the two most distant cities in the network, for example, the cities of Gytheio in Greece and Rennesøy in Norway are part of the network's periphery, highlighting the extensive reach of the E-road network across Europe, the distance between these cities illustrates the challenge in connecting remote locations within a large infrastructure network.

Eccentricity values provide further insights into the network's structure by measuring the maximum distance from each node to any other node in the network, for instance, the city of Newport has an eccentricity of 12, indicating that the farthest city from Newport is 12 edges away. Similarly, cities like Warrington and Preston have eccentricity values of 14, reflecting their relatively central positions within the network, these values suggest that while some cities are centrally located and well-connected, others are more peripheral, impacting their accessibility and travel times.

The radius of the network is 1, which seems unusually low for such a large network, this value indicates the presence of highly central nodes directly connected to many other nodes, such central nodes are crucial for maintaining the network's connectivity and efficiency. Cities identified as central nodes, such as Porto-Vecchio in France and Sassari in Italy, have the lowest eccentricity values, making them pivotal points in the network, these cities likely serve as major hubs for transportation and logistics, facilitating efficient movement across the network.

The network's periphery includes cities like Gytheio and Rennesøy, which are the farthest from other nodes in terms of eccentricity, this peripheral cities are critical for understanding the network's reach and identifying potential areas for infrastructure improvement to enhance connectivity.

III. Centrality measures

Centrality measures are critical for understanding the importance of nodes within a network, they help identify key nodes that play significant roles in terms of connectivity, control over information flow, and influence within the network, for the EuroRoads network, there were selected and analyzed four centrality measures: degree centrality, betweenness centrality, closeness centrality, and eigenvector centrality, this measures provide different perspectives on node importance and contribute to a comprehensive understanding of the network's structure.

```
degree_centrality = nx.degree_centrality(G)
betweenness_centrality = nx.betweenness_centrality(G, normalized=True)
closeness_centrality = nx.closeness_centrality(G)
eigenvector_centrality = nx.eigenvector_centrality(G, max_iter=1000)

top_degree = sorted(degree_centrality.items(), key=lambda x: x[1], reverse=True)[:5]
top_betweenness = sorted(betweenness_centrality.items(), key=lambda x: x[1], reverse=True)[:5]
top_closeness = sorted(closeness_centrality.items(), key=lambda x: x[1], reverse=True)[:5]
top_eigenvector = sorted(eigenvector_centrality.items(), key=lambda x: x[1], reverse=True)[:5]


{
    "Top Degree Centrality": top_degree,
    "Top Betweenness Centrality": top_betweenness,
    "Top Closeness Centrality": top_closeness,
    "Top Eigenvector Centrality": top_eigenvector
```

```
}
```

Degree centrality measures the number of direct connections a node has, it is a straightforward indicator of a node's activity and connectivity within the network, in the EuroRoads network, the cities with the highest degree centrality are:

- Moscow (0.0085)
- Paris (0.0068)
- Liège (0.0068)
- Berlin (0.0068)
- Munich (0.0068)

Moscow, with the highest degree centrality, serves as a major hub in the network, indicating its significant connectivity to other cities. Paris, Liège, Berlin, and Munich follow closely, highlighting their roles as important connection points within the European road network, these cities likely facilitate a large volume of traffic and serve as critical junctures for transportation and logistics.

```
highlight_nodes = [”Moscow”, ”Paris”, ”Li ge”, ”
    Berlin”, ”Munich”]

def highlight_and_zoom(G, pos, highlight_nodes,
    radius=2):
    for node in highlight_nodes:
        neighbors = nx.
    single_source_shortest_path_length(G, node,
    cutoff=radius).keys()

        subgraph = G.subgraph(neighbors)

        plt.figure(figsize=(8, 8))
        sub_pos = {n: pos[n] for n in subgraph.nodes
    ()}
        nx.draw(subgraph, sub_pos, with_labels=True,
     node_size=50, font_size=8, font_color='black',
    node_color='lightblue')
        nx.draw_networkx_nodes(subgraph, sub_pos,
    nodelist=[node], node_size=100, node_color='red
    ')
        plt.title(f”Zoom on {node}”)
        plt.show()

highlight_and_zoom(G, pos, highlight_nodes)
```

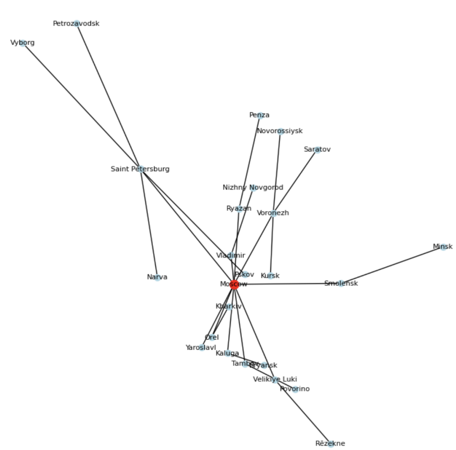

Fig. 2: Zoom in Moscow

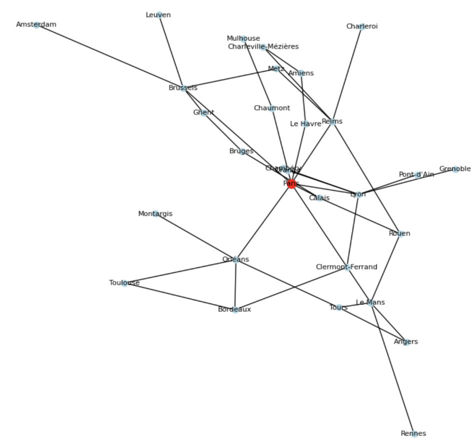

Fig. 3: Zoom in Paris

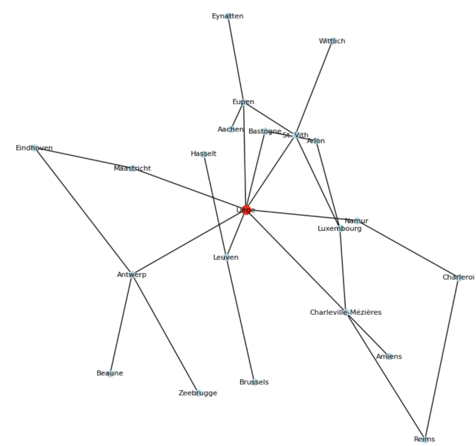

Fig. 4: Zoom in Liège

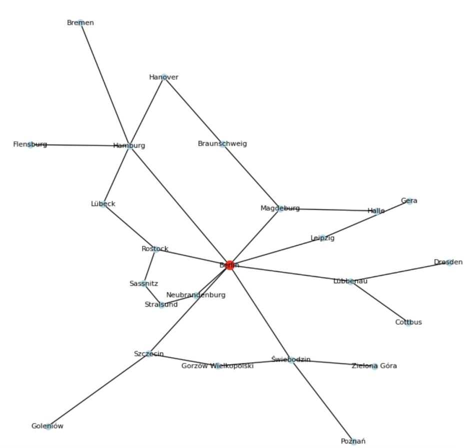

Fig. 5: Zoom in Berlin

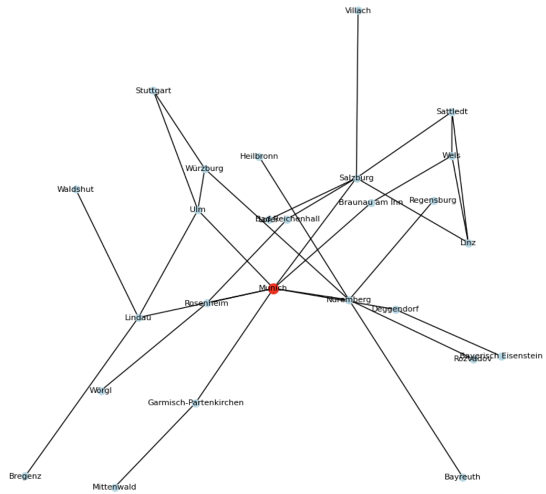

Fig. 6: Zoom in Munich

Betweenness centrality measures the extent to which a node lies on the shortest paths between other nodes, it reflects the node's role in controlling information flow and its potential to act as a bottleneck in the network. The top cities by betweenness centrality in the EuroRoads network are:

- Brest (0.2148)
- Moscow (0.2125)
- Saint Petersburg (0.1955)
- Le Mans (0.1751)
- Rennes (0.1744)

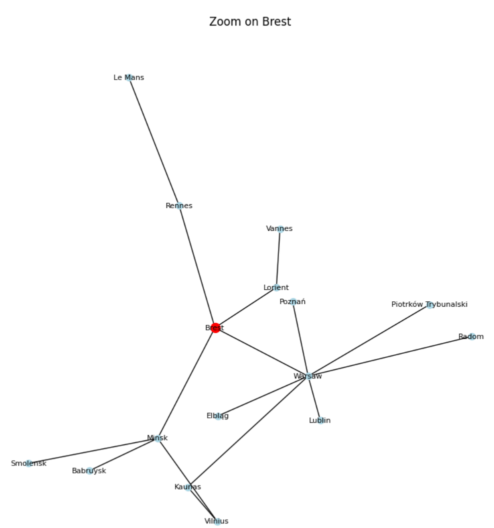

Fig. 7: Zoom in Brest

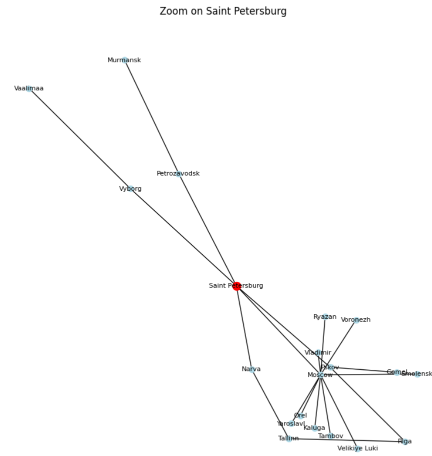

Fig. 8: Zoom in Saint Petersburg

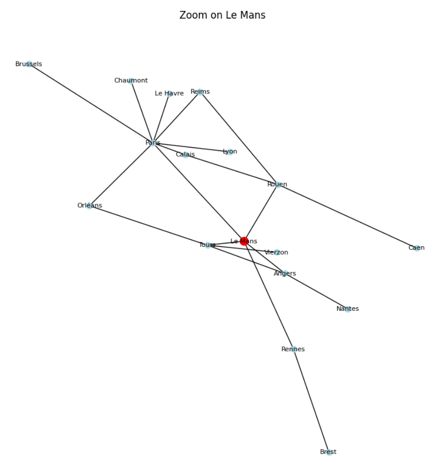

Fig. 9: Zoom in Le Mans

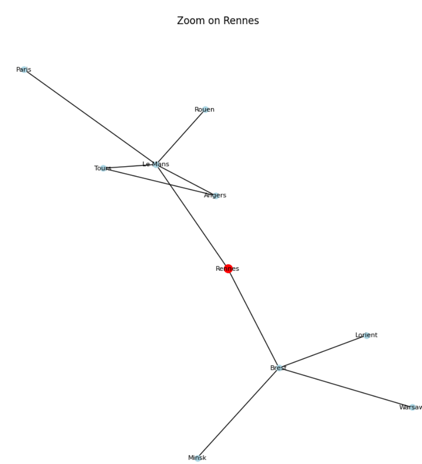

Fig. 10: Zoom in Rennes

Brest and Moscow, with the highest betweenness centrality, are crucial for efficient travel and communication across the network, they act as key transit points that connect various parts of Europe, indicating their strategic importance. Saint Petersburg, Le Mans, and Rennes also play significant roles in facilitating the flow of traffic, underscoring their importance in maintaining the network's efficiency.

Closeness centrality measures how quickly a node can reach all other nodes in the network, it reflects the efficiency of a node in disseminating information throughout the network, the cities with the highest closeness centrality are:

- Warsaw (0.0769)
- Brest (0.0768)
- Minsk (0.0754)
- Lviv (0.0741)
- Lublin (0.0740)

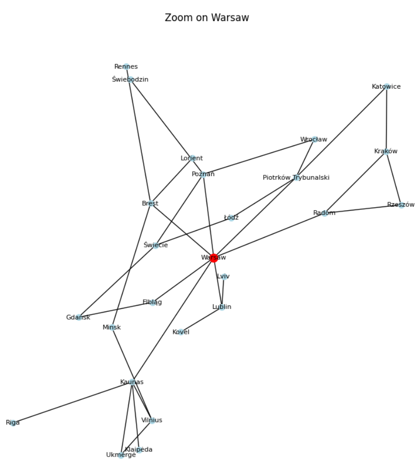

Fig. 11: Zoom in Warsaw

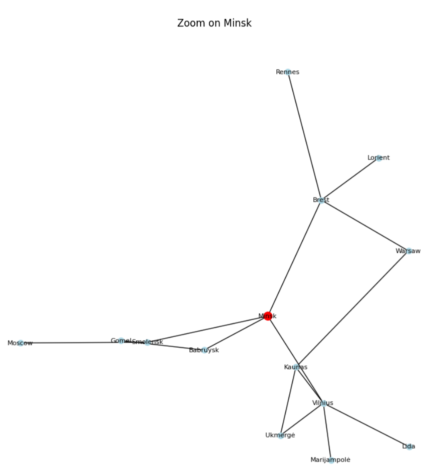

Fig. 12: Zoom in Minsk

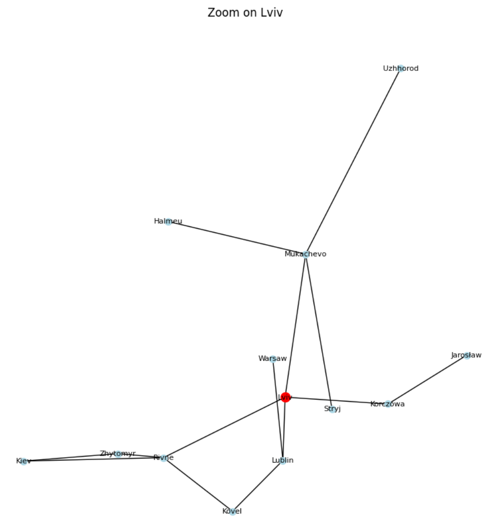

Fig. 13: Zoom in Lviv

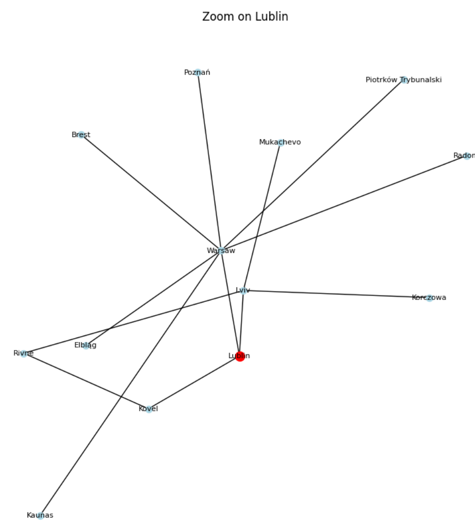

Fig. 14: Zoom in Lublin

Warsaw, Brest, and Minsk are among the most central cities, meaning they can reach other cities more quickly compared to other nodes, this high centrality suggests that these cities are well-positioned to serve as efficient hubs for transportation and communication, facilitating rapid dissemination of information and resources across Europe.

Eigenvector centrality measures a node's influence based on the connectivity of its neighbors, a high eigenvector centrality indicates that a node is connected to other well-connected nodes, reflecting its overall influence within the network, the top cities by eigenvector Centrality are:

- Paris (0.3651)
- Metz (0.3002)
- Reims (0.2808)

- Brussels (0.2363)
- Liège (0.2060)

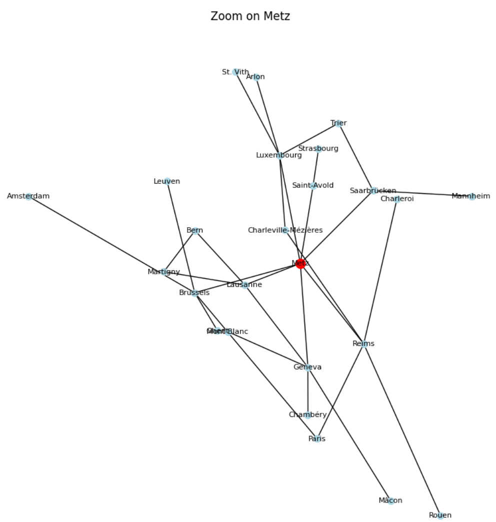

Fig. 15: Zoom in Metz

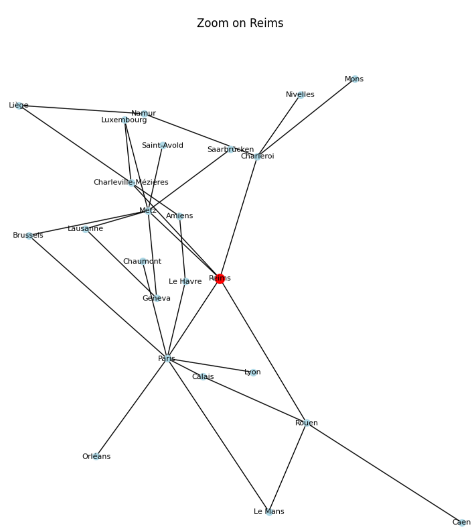

Fig. 16: Zoom in Reims

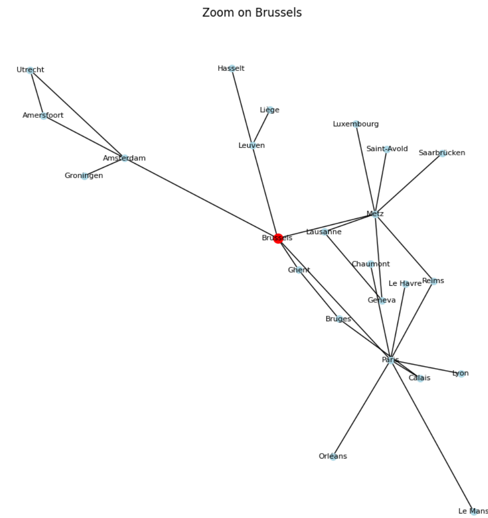

Fig. 17: Zoom in Brussels

Paris, with the highest eigenvector centrality, stands out as the most influential city in the network, its connections to other well-connected cities, such as Metz, Reims, Brussels, and Liège, enhance its overall influence. These cities likely play pivotal roles in the network, influencing not only their immediate neighbors but also the broader connectivity of the network.

IV. Degree distribution

The degree distribution of a network is a fundamental metric that describes how connections (or edges) are distributed among the nodes, it provides insights into the network's topology, identifying whether the network is homogeneous, with most nodes having similar degrees, or heterogeneous, with a few highly connected nodes.

```
degrees = [d for n, d in G.degree()]

degree_count = Counter(degrees)
deg, freq = zip(*degree_count.items())

deg = np.array(deg)
freq = np.array(freq)

plt.figure(figsize=(8, 6))
plt.loglog(deg, freq, 'bo', markersize=5)
plt.title("Degree Distribution")
plt.xlabel("Degree (d)")
plt.ylabel("Frequency")
plt.grid(True, which="both", ls="--")
plt.show()
```

The degree distribution plot for the EuroRoads network is shown below. The plot is a log-log scale graph with the degree d on the x-axis and the frequency of nodes with that degree on the y-axis.

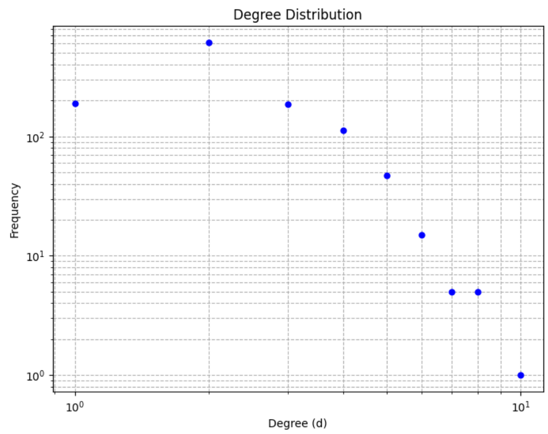

Fig. 18: Degree distribution plot

Firstly, the degree distribution shows a heavy-tailed distribution, where most nodes have a low degree, while a few nodes have a very high degree, this feature suggests that the network follows a scale-free distribution, which is typical of many real-world networks, including infrastructure and social networks. The presence of these high-degree nodes, such as Moscow and Paris, indicates the existence of major hubs within the network.

Moreover, the majority of nodes in the network have a low degree, reflecting the large number of cities connected to only a few other cities, this structure is typical for transportation networks, where many smaller cities or towns are connected to nearby larger hubs but not directly to other small cities. This configuration implies that while small cities depend on these hubs for their connectivity, direct connections between them are scarce.

The implication of this heavy-tailed distribution for the network's robustness is significant, it suggests that the network is robust against random failures but vulnerable to targeted attacks on high-degree nodes. Removing a few key hubs could significantly disrupt the network's connectivity, emphasizing the need for strategic planning to enhance network resilience.

V. Community Detection

Community detection is a crucial aspect of network analysis, allowing us to identify groups of nodes (or communities) that are more densely connected internally than with the rest of the network., for the EuroRoads network, was applied the Label Propagation Algorithm (LPA), which is particularly suited for large networks due to its computational efficiency and simplicity, LPA works by assigning labels to nodes and iteratively updating these labels based on the most frequent label among a node's neighbors until a consensus is reached within each community.

The visualization of the EuroRoads network with detected communities, using a spring layout, shows the network divided into various color-coded groups, each color represents a distinct community, providing a clear visual representation of the network's modular structure.

```python
communities = list(label_propagation_communities(G))
community_map = {node: community_id for community_id
    , nodes in enumerate(communities) for node in
    nodes}

colors = [community_map[node] for node in G.nodes()]
plt.figure(figsize=(12, 12))
pos = nx.spring_layout(G)

nx.draw_networkx_nodes(G, pos, node_size=50, cmap=
    plt.get_cmap('jet'), node_color=colors)
nx.draw_networkx_edges(G, pos, alpha=0.5)
plt.title("Community Analysis of the Network")
plt.show()
```

The application of LPA resulted in the detection of 406 communities, reflecting a high degree of modularity within the network.

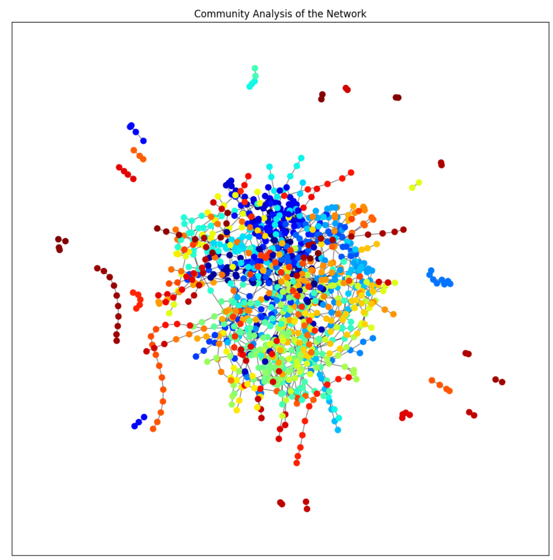

Fig. 19: Community detection plot

The visualization of the EuroRoads network with detected communities, using a spring layout, shows the network divided into various color-coded groups. Each color represents a distinct community, providing a clear visual representation of the network's modular structure, the application of LPA resulted in the detection of 406 communities, reflecting a high degree of modularity within the network.

For example, Community 0 includes the cities of Greenock, Glasgow, and Edinburgh, all located in Scotland. This regional clustering makes sense as these cities are geographically close and likely have a high degree of interaction and connectivity, similarly, Community 1 groups together Preston, Southampton, Birmingham, and Cambridge in England, the presence of Birmingham, a major city and transport hub, suggests that this community is an important cluster within the UK, facilitating significant travel and commerce.

Additionally, some communities span across several countries, illustrating the transnational nature of the E-road network, for instance, Community 2 includes Charleville-Mézières, Liège, Amiens, Maastricht, and Le Havre, indicating a cluster that spans France and Belgium, this transnational clustering demonstrates the importance of cross-border connectivity in Europe, facilitating seamless travel and trade between neighboring countries.

The results from the community detection reveal a network with many localized clusters. The high number of communities (406) indicates a highly modular structure, with various regions forming tight-knit groups, this modularity is expected in transportation networks where regional connectivity is crucial for efficient travel and logistics.

The choice of the label propagation algorithm is justified by its efficiency and effectiveness in handling large networks. LPA is non-parametric and does not require prior knowledge of the number of communities, making it ideal for exploratory analysis, its ability to quickly converge to a solution makes it practical for real-time applications and large-scale networks like EuroRoads.

VI. Conclusion

The analysis of the EuroRoads network has revealed several key insights into its structure and functionality, with 1174 nodes and 1417 edges, the network demonstrates a robust and expansive reach across Europe. The average clustering coefficient of 0.0167 indicates a predominantly linear and spread-out network, which is typical for large-scale transportation systems.

The examination of centrality measures identified critical hubs such as Moscow, Paris, and Brest, which play significant roles in maintaining network connectivity and efficiency, the degree distribution analysis revealed a heavy-tailed distribution, suggesting that the network follows a scale-free pattern, with a few highly connected nodes acting as major hubs, this characteristic highlights the network's vulnerability to targeted attacks on these key nodes but also underscores its robustness against random failures.

Community detection using the label propagation algorithm (LPA) identified 406 distinct communities, reflecting a high degree of modularity within the network, these communities often align with geographical regions, illustrating the importance of regional connectivity for efficient travel and logistics.

Overall, the EuroRoads network's structure facilitates efficient transportation across Europe but also presents challenges in terms of vulnerability and regional disparities in connectivity, strategic enhancements in network resilience and targeted infrastructure improvements could further optimize the network's performance, ensuring seamless and efficient transportation for goods and people across the continent, this analysis underscores the importance of network analysis techniques in understanding and improving large-scale transportation systems.

VII. References

[1] Jérôme Kunegis. KONECT – The Koblenz Network Collection. In Proc. Int. Conf. on World Wide Web Companion, pages 1343–1350, 2013. [http]

[2] Lovro Šubelj and Marko Bajec. Robust network community detection using balanced propagation. Eur. Phys. J. B, 81(3):353–362, 2011.

[3] Hage, P., & Harary, F. (1995). Eccentricity and centrality in networks. Social Networks, 17(1), 57–63. doi: 10.1016/0378-8733(94)00248-9

[4] Sharma, V. (2023). Comparative Analysis of Degree Centrality and Betweenness Centrality in Large Graphs. Medium. Retrieved from: https://medium.com/nerd-for-tech/comparative-analysis-of-degree-centrality-and-betweenness-centrality-in-large-graphs-e63576e052b8

[5] Zhang, J., & Luo, Y. (2017). Degree Centrality, Betweenness Centrality, and Closeness Centrality in Social Network. Atlantis Press. doi: 10.2991/msam-17.2017.68

[6] Eigenvector Centrality. (2014, October 08). Retrieved from: https://www.sci.unich.it/francesc/teaching/network/eigenvector.html

[7] Degree Distribution. (2014, October 08). Retrieved from https://www.sci.unich.it/francesc/teaching/network/distribution.html

[8] Garza, S. E., & Schaeffer, S. E. (2019). Community detection with the Label Propagation Algorithm: A survey. Physica A, 534, 122058. doi: 10.1016/j.physa.2019.122058