Navigating the Networks: A Study of Public Transportation in Madrid and Brussels

Tara Sabooni, Artificial Intelligence, 0001133567
Bahar Hamzehei, Artificial Intelligence, 0001111961
Yasaman Samadzadeh, Artificial Intelligence, 0001113331
Mohammed Oubia, Artificial Intelligence, 0001135872

GitHub Repository: `https://github.com/oubia/Social-Analysis`

1 Introduction

This project focuses on the analysis of Public Transport Networks of Madrid and Brussels. This is a vital area in urban planning that examines public transportation systems' design, efficiency, and functionality. By studying data on service frequency, travel times, connectivity, and coverage, Public Transport Network Analysis aims to identify patterns, improve accessibility, and enhance the sustainability and equity of urban mobility. Research in this area has a long history, and a wealth of reliable data is available, including real-time tracking, smart card usage, and passenger journey records from major cities. These data sources provide valuable insights for analyzing public transport networks.

2 Problem and Motivation

Cities around the world are grappling with growing challenges tied to increased car usage, such as environmental damage and urban congestion. Public transport has become a key solution to address these issues, offering a more sustainable alternative to private vehicles. Recognizing its importance, the United Nations Human Settlements Programme (UN-Habitat) [9] issued guidelines in 2015 to promote sustainable urban development through improved public transport [2]. However, assessing how well these strategies and systems work remains a challenge, making detailed performance evaluations essential for understanding and enhancing public transportation networks.

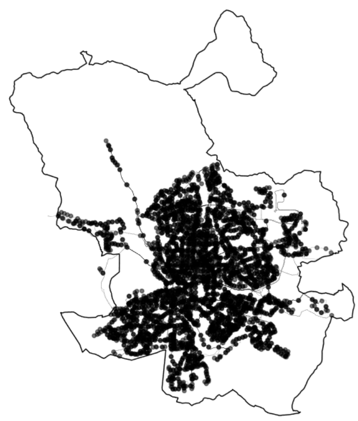
To contribute to this ongoing discourse, this project focuses on analyzing the public transport networks of Brussels and Madrid. These two cities were chosen because they are not excessively large, making their transport systems more manageable for in-depth analysis. Moreover, both cities are popular tourist destinations where public transport is crucial for connecting visitors and residents to key attractions and services. Their reliance on efficient public transport systems underscores the need for a thorough understanding of how these networks operate.

The study uses weekly service frequency as a key metric to investigate service clusters and high-activity zones within the networks. By applying centrality measures and community detection algorithms, the project will evaluate connectivity and identify strengths and weaknesses in each city's system. Through this comparative analysis, we aim to uncover patterns and insights that can guide transportation planning, enhance accessibility, and support the sustainable development of public transport networks in similar urban contexts.

3 Datasets

The present project relies on General Transit Feed Specification (GTFS) [1] data for the public transportation networks of Madrid and Brussels. GTFS is an open data format, widely adopted and created to standardize public transit system information and make it more accessible to developers, researchers, and policymakers. It consists of a set of text files (mostly in .txt format) that, when taken together, describe the operation aspects of a transit system. These files contain detailed information about stops, routes, trips, stop times, schedules, and service calendars.
Both datasets were obtained from [5]. All these datasets are publicly available and are provided in their original GTFS format, thus avoiding the need for digitization. This provides an inherent digital nature to the data, which in turn guarantees high precision and homogeneity apt for spatial and temporal analysis of transit systems. The data collection process was straightforward due to the easily accessible GTFS data, which transit agencies publish freely to support both public transit apps and academic research. Both datasets were licensed for free use without restrictions, allowing us to include them in our study without any limitations. The standardized format of GTFS greatly simplified data handling and processing, thus enabling us to focus on analysis rather than data preparation.
The processing, storage, and analysis of the data were performed using Python libraries tailored to handle the particular demands of data manipulation. The data was structured using pandas DataFrames [4] to offer a flexible and efficient base for handling the interrelated GTFS files. In the network analysis, transportation systems were modeled as graphs: stops were represented as nodes while trips were represented as edges and number of weekly trips had been considered as weight (Figure 1) and NetworkX [3] library is used for computing metrics such as connectivity, centrality, and efficiency. For plotting and visualizing results, libraries such as folium, geopandas, and matplotlib were employed, enabling the creation of detailed spatial and graphical representations of the data.

(a) Brussels



(b) Madrid

Figure 1: Public transport network of the considered cities.

4 Validity and Reliability

4.1 Validity

The validity of the data used in this project is ensured by sourcing it from the Mobility Database, a trusted platform that hosts curated datasets, guaranteeing the authenticity and accuracy of the data [6]. For this study, we selected datasets for Madrid and Brussels from their official transit operators. For Brussels, the dataset from Société des Transports Intercommunaux de Bruxelles/Maatschappij voor het Intercommunaal Vervoer te Brussel (STIB/MIVB) [7] was used, and for Madrid, data was sourced from Empresa Municipal de Transportes de Madrid (EMT Madrid)[8]. Both datasets are the latest available versions for November, reflecting the current state of the transit systems.

The official transit providers maintain rigorous quality control processes, ensuring that the datasets include precise geographic coordinates for transit stops and accurate route information. This enables the construction of spatial networks where nodes represent stops and edges denote trips between them. By utilizing the Mobility Database and officially maintained data, this study ensures the spatial networks faithfully represent the actual public transit systems of Madrid and Brussels, achieving a high level of validity.

4.2 Reliability

The reliability of this study is ensured through a systematic and reproducible methodology for processing and analyzing the spatial datasets. Data from STIB/MIVB and EMT Madrid were handled using Python libraries, such as pandas for data manipulation and NetworkX for network analysis, which provide robust and deterministic tools for ensuring consistent results. Spatial networks were constructed by transforming geographic coordinates into a common projection system, with data cleansing steps applied to address missing or inconsistent entries, ensuring uniformity and reliability across the datasets.

Standard tools and well-documented methodologies were employed, contributing to the study's transparency and reproducibility. Analyses, such as centrality measures and community detection, were conducted using deterministic or controlled stochastic algorithms, with random seeds and hyperparameters explicitly documented to guarantee repeatability. This transparent approach ensures that the methods and results can be replicated under identical conditions, affirming the reliability of the analysis.

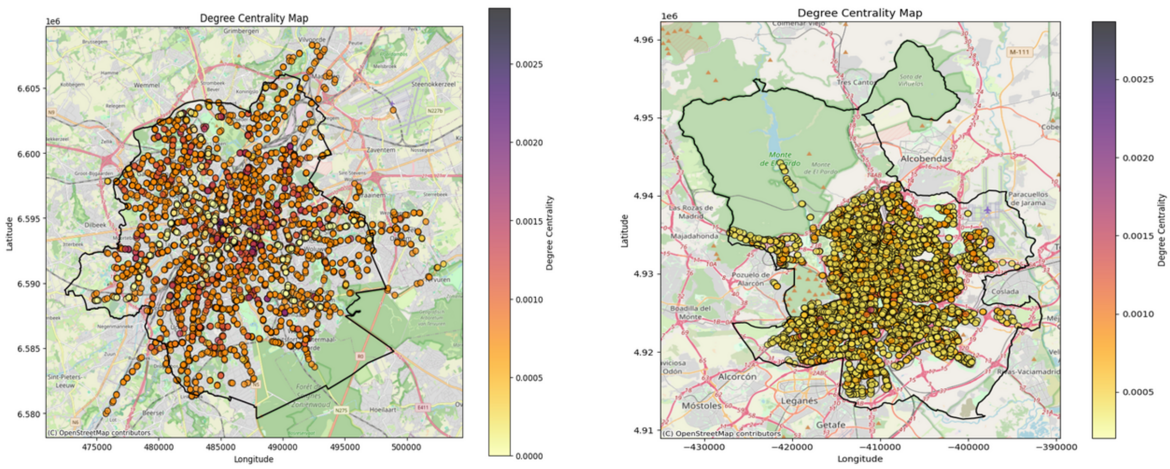
5 Measures and Results

Network analysis through visualization gives a first look at some of its properties, although it poorly gives a deep analysis. Therefore, one should use mathematical metrics when describing the outstanding features quantitatively in the network structure. First, a range of centrality metrics in the nodes was calculated: Degree Centrality, Eigenvector Centrality, Betweenness Centrality, and Closeness Centrality. Next, groups of nodes were analyzed with such techniques as the identification of k-Cores and k-means clustering. Lastly, the overall cohesiveness of the network was calculated by computing Density and further analyzed its structure by the distribution of centrality measures.

5.1 Centrality Measures

5.1.1 Degree Centrality

Degree Centrality is a measure that quantifies the importance of a node by the number of connections it has within the network. The maps above illustrate the degree centrality of stops in the Brussels (Figure 2a) and Madrid (Figure 2b) transit networks. Darker colors indicate higher centrality, highlighting the most connected nodes that serve as critical hubs in the transportation system. The central areas of both cities exhibit the highest degree centrality, reflecting the dense interconnectivity of stops in the urban core. These hubs play a pivotal role in the efficiency and resilience of the transit network, facilitating the movement of passengers and linking peripheral areas to the city center.



(a) Brussels

(b) Madrid

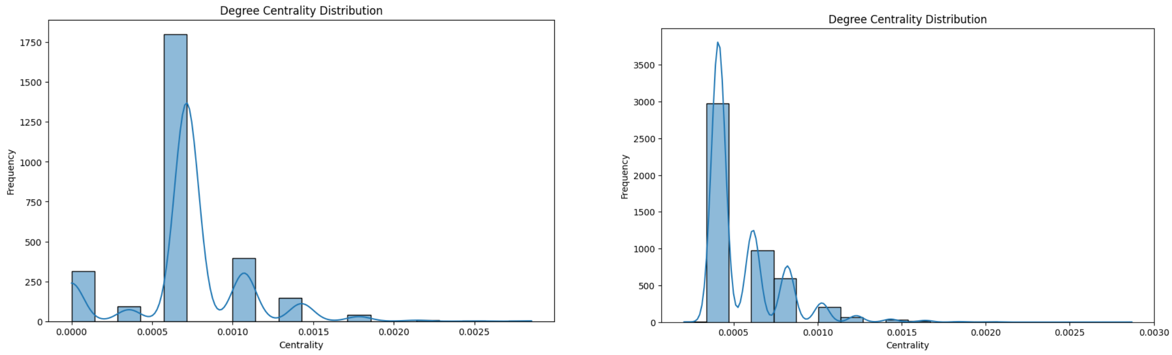
Figure 2: The degree centrality map of the two cities.

The graphs depict the degree centrality distribution for the transportation networks of Brussels and Madrid. Degree centrality reflects the connectivity of a stop within the network.

Brussels: (Figure 3a) The distribution peaks around a low centrality value (0.0005), indicating that most stops have limited connections. A few key hubs with higher centrality values (0.001 to 0.0025) play a significant role in the network.

Madrid: (Figure 3b) The distribution shows a broader range of centrality values, with a pronounced peak and a longer tail (up to 0.003). This suggests a greater number of highly connected hubs and denser sub-regions compared to Brussels.

Insights: Both networks exhibit skewed distributions, with most nodes being peripheral stops. Madrid's broader distribution highlights a more robust network with multiple well-connected stops, while Brussels relies heavily on a few central stops, potentially increasing vulnerability to disruptions.



(a) Brussels (b) Madrid

Figure 3: The degree centrality distribution of the two cities.

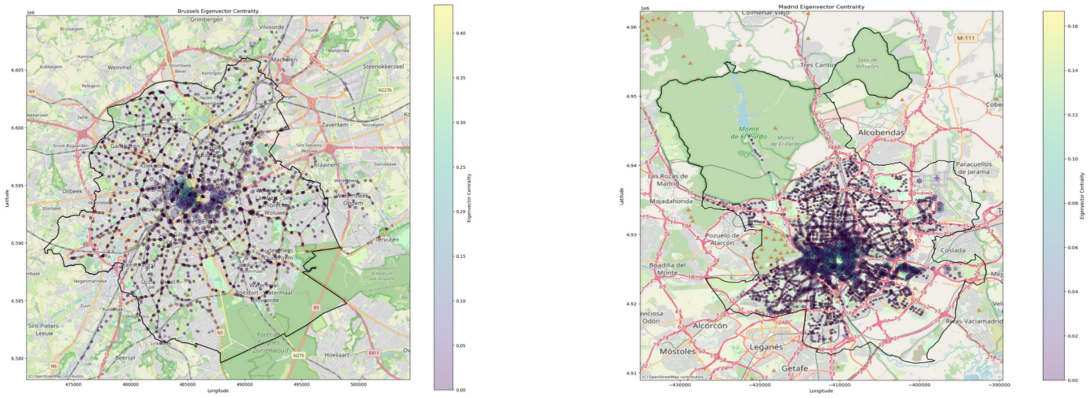
5.1.2 Eigenvector Centrality

Eigenvector centrality measures how important a transit stop is based on its connections and the influence of the stops it connects to. In public transport networks, it helps pinpoint key hubs that keep the system running smoothly.

A comparison of Madrid and Brussels reveals a key difference in the layout of their transit networks. Madrid's network has a higher share of stops with higher values of eigenvector centrality compared to Brussels (Figure 5). Additionally, their high-centrality nodes—the most important stops—are distributed differently. In Madrid, some key stops are spread throughout the city, not just concentrated in the center. This design connects suburban and peripheral areas more effectively, reducing pressure on the city center and supporting better accessibility across the entire network. Brussels, on the other hand, clusters most of its important stops in the city center. This setup works well for a compact city structure, prioritizing efficient movement in and out of the central areas. However, it may leave outlying regions less well-connected, potentially increasing travel times for suburban commuters (Figure 4).

These differences highlight how each city's transport network reflects its urban layout and priorities—Madrid's system is geared for citywide connectivity, while Brussels focuses on central efficiency.

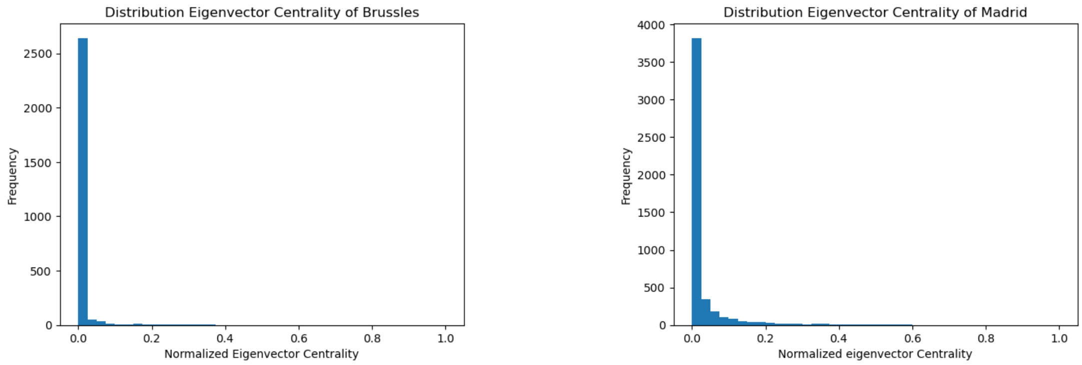
Madrid's approach of spreading high-eigenvector centrality nodes supports balanced citywide connectivity but increases the risk of localized disruptions at critical hubs. Brussels' focus on a centralized network works well for its compact urban design but leaves the system vulnerable to central bottlenecks and marginalizes outer areas. Each city's vulnerabilities reflect the trade-offs inherent in their network designs, emphasizing the importance of targeted interventions to enhance resilience.



(a) Brussels

(b) Madrid

Figure 4: The eigenvector centrality of two cities.



(a) Brussels

(b) Madrid

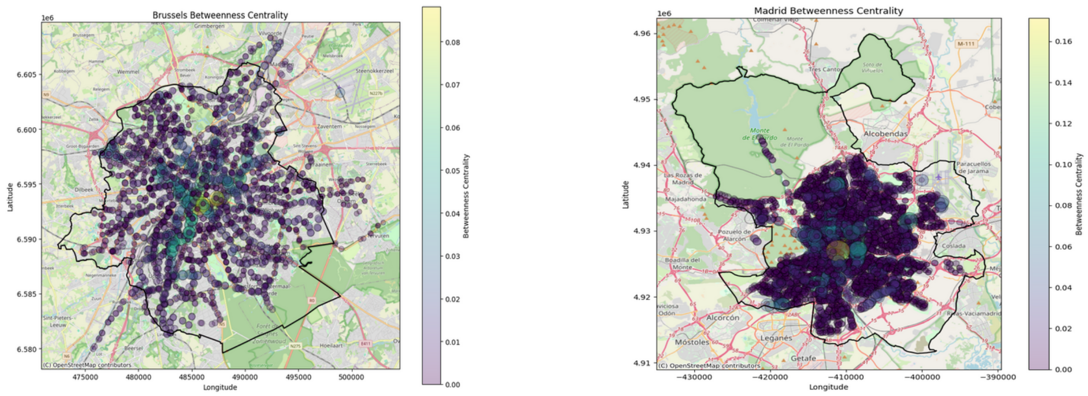
Figure 5: The eigenvector centrality distribution of two cities.

5.1.3 Betweenness Centrality

Betweenness centrality, a network theory metric, measures how often a node (in this case, a transit stop) is part of the shortest path between other nodes. In public transportation, stops with high betweenness centrality play a crucial role in connecting different parts of the network, making them indispensable for efficient passenger movement. The betweenness centrality distribution shown in Figure 6, suggests that in both cities the central areas stops had shown higher betweenness centrality.

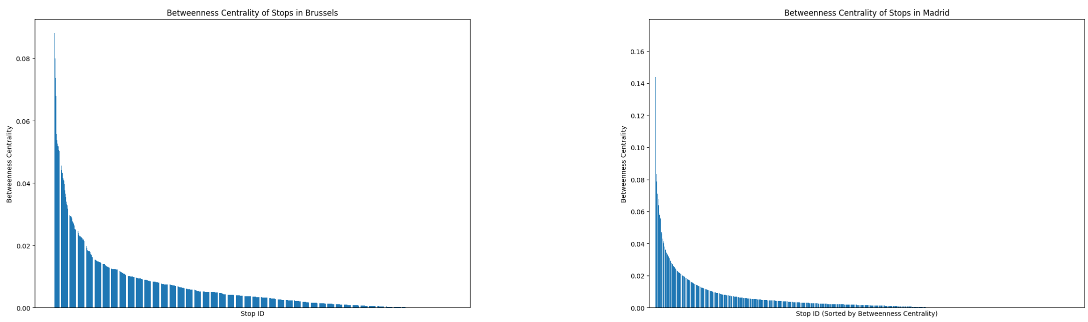
The analysis reveals significant differences between the transit networks of Brussels and Madrid. Brussels has a lower maximum betweenness centrality score 0.088 (Figure 7a) with a relatively even distribution among key stops, reflecting a decentralized network design. Connectivity is spread across several pivotal stops, which enhances network resilience but may require more complex navigation for passengers. In contrast, Madrid's maximum betweenness centrality is much higher 0.171 (Figure 7b), indicating a centralized network where a few key stops serve as critical transfer points, facilitating efficient commutes but increasing vulnerability to disruptions at these hubs.

Overall, Madrid's network relies heavily on a few high-centrality hubs, making it efficient but susceptible to failures at these points. Meanwhile, Brussels' more decentralized structure enhances reliability and resilience but might pose challenges for passenger navigation due to the lack of dominant transfer points. These differences highlight distinct design philosophies in the two cities' transit systems.



(a) Brussels (b) Madrid

Figure 6: The betweenness centrality of two cities.



(a) Brussels (b) Madrid

Figure 7: Bar plot of betweenness centrality of stops of two cities.

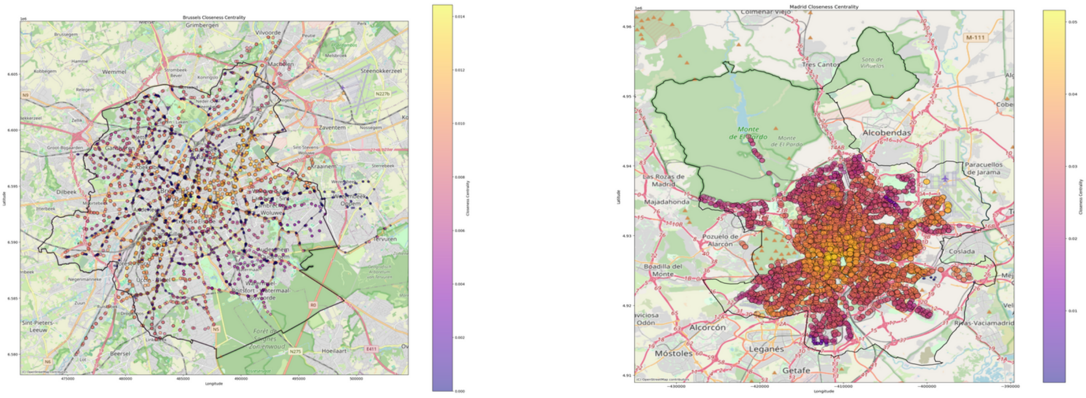
5.1.4 Closeness Centrality

Closeness centrality measures how efficiently a node connects to all others in a network, reflecting accessibility. When comparing the transit networks of Madrid, Spain, and Brussels, Belgium, distinct structural differences become apparent due to their unique geographic and demographic contexts.

In Madrid, the closeness centrality is significantly concentrated in the city center, as shown in Figure 8b, indicating a centralized network design with a strong reliance on key hubs. This hub-and-spoke pattern, typical of metropolitan transit systems, ensures efficient connectivity by prioritizing access to central areas. Moving away from the center, the network becomes more diffused, though still maintaining reasonable accessibility, showcasing a deliberate effort to integrate peripheral areas effectively.

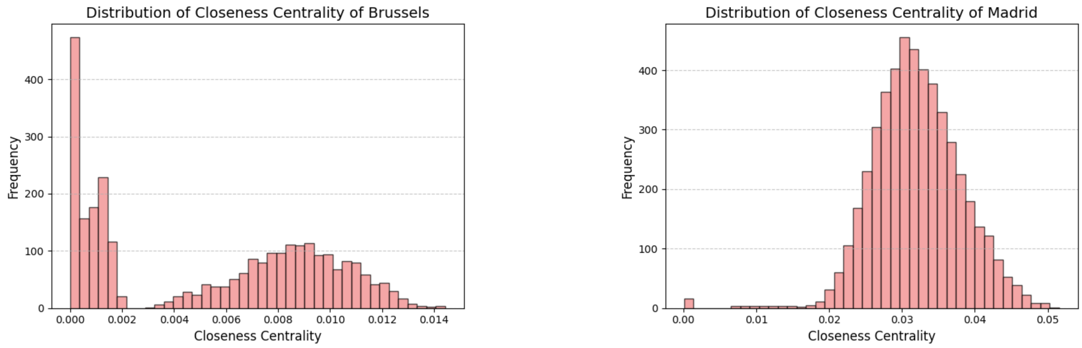
In contrast, Brussels exhibits a more evenly distributed network, with slight increases in closeness centrality observed in the city center as illustrated in Figure 8a. However, the centrality values are generally lower and more scattered compared to Madrid, reflecting a less centralized design. This pattern aligns with the city's diverse and less densely populated regions, emphasizing distributed access over reliance on specific hubs.

These differences highlight the strategic priorities of each city's transit system. Madrid focuses on centralized connectivity to enhance access to critical urban hubs, while Brussels adopts a more distributed approach, catering to the rural and suburban characteristics of the region.



(a) Brussels (b) Madrid

Figure 8: Closeness centrality map of two cities.



(a) Brussels (b) Madrid

Figure 9: Closeness centrality distribution of two cities.

5.2 Community Measures

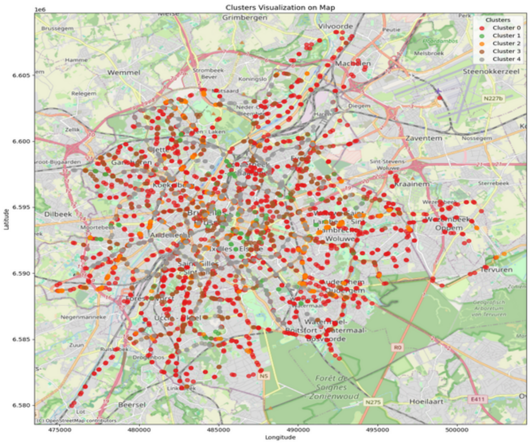
5.2.1 K-means Clustering

The K-means clustering results for Brussels and Madrid reveal distinct spatial patterns in their public transportation networks. Clustering was based on centrality measures derived from the stops and stop times datasets, leveraging geographical coordinates to identify connectivity patterns.

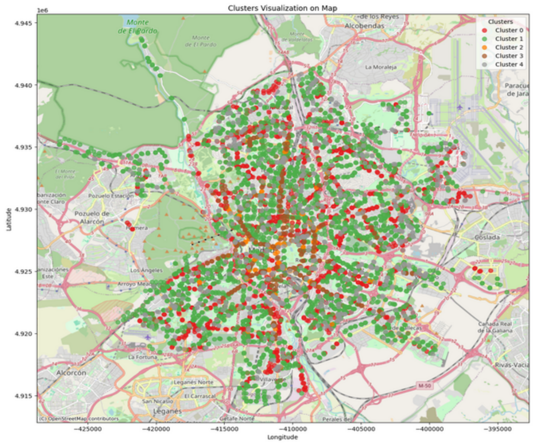
Brussels: The clusters in Brussels (Figure 10a) highlight a centralized structure, with dense clusters forming around the city center and major transit corridors. The segmentation reflects the reliance of suburban stops on central hubs, underscoring the importance of central stops for accessibility. Smaller cluster sizes indicate a focused distribution of connectivity.

Madrid: Madrid's clusters (Figure 10b) exhibit a decentralized pattern with larger, well-distributed clusters across urban and suburban areas. This design reduces dependence on the central core, enhancing resilience and equitable transit coverage. Peripheral clusters highlight Madrid's focus on balanced network design.

Insights: Compared to Madrid's decentralized layout, Brussels relies heavily on central stops, which could create vulnerabilities during disruptions. Madrid's design ensures redundancy and improved accessibility across the city. These differences reflect contrasting transit planning strategies: Brussels focuses on centralization for efficiency, while Madrid emphasizes decentralization for resilience and inclusivity.



(a) Brussels



(b) Madrid

Figure 10: K-means clustering map of the two cities.

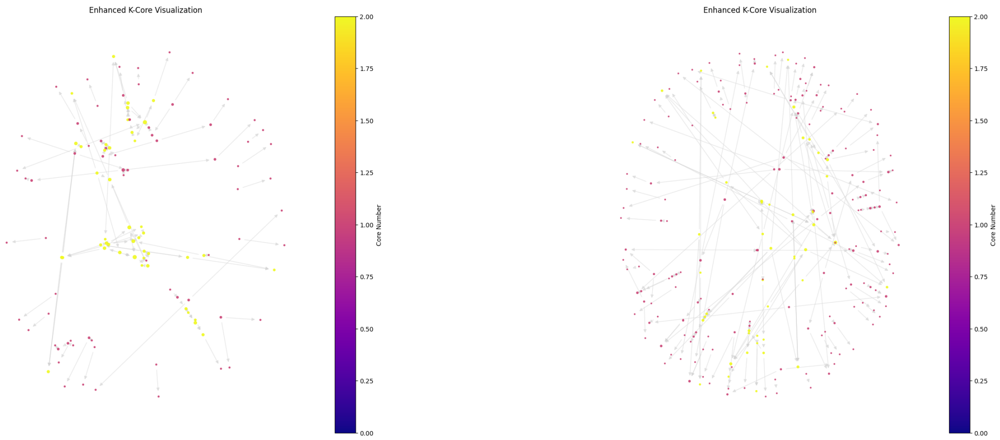
5.2.2 K-Cores and Core Periphery Structure

The K-Core decomposition of Brussels and Madrid's public transportation networks provides a clear depiction of their hierarchical connectivity. The K-Core analysis groups stops based on their connectivity strength, where higher-core numbers represent stops with robust connections, forming the backbone of the network.

Brussels: The enhanced K-Core visualization for Brussels (Figure 11a) highlights a compact core with a small number of high-core stops surrounded by lower-core stops. The radial K-Core layout (Figure 12a) further demonstrates the centralized structure, where the network heavily relies on a limited number of high-core stops at the center. This centralization makes the network efficient but could create vulnerabilities in case of disruptions at core stops.

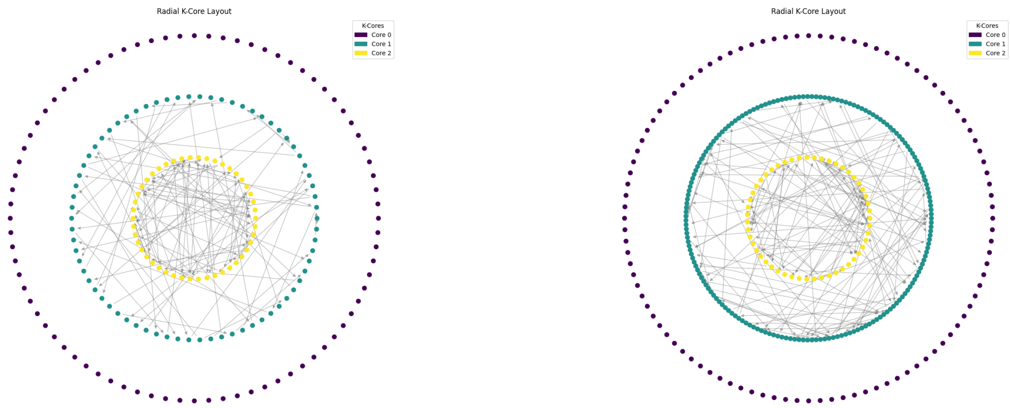
Madrid: In contrast, Madrid's enhanced K-Core visualization (Figure 11b) displays a larger and more evenly distributed core structure. The radial layout (Figure 12b) illustrates a broader distribution of higher-core stops, reflecting a decentralized network design. This enhances the network's resilience and ensures better coverage across both urban and suburban areas.

Insights: The comparative analysis shows that Brussels emphasizes centralization for efficiency, making it dependent on a few core stops. On the other hand, Madrid adopts a decentralized approach, reducing network vulnerability and enhancing connectivity. This analysis aids in understanding the fundamental differences in transit network planning between the two cities.



(a) Brussels (b) Madrid

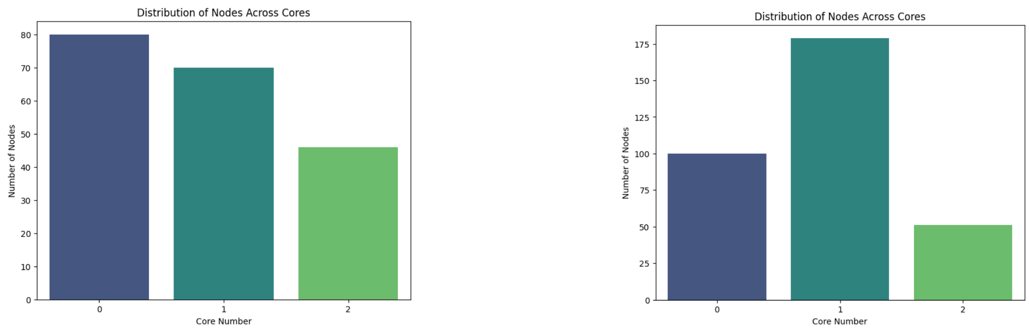
Figure 11: K-Core visualization of the two cities.



(a) Brussels (b) Madrid

Figure 12: Radial K-Core Layout of two cities.

Brussels: The distribution in Brussels (Figure 13a) shows a relatively balanced core structure with a significant number of nodes in the periphery (core 0), followed by core 1 and a smaller set in the central core (core 2). This distribution highlights a centralized network design where fewer stops form the backbone of the network, contributing to its structural integrity.
Madrid: In contrast, Madrid's distribution (Figure 13b) indicates a larger and more evenly distributed network, with a noticeable increase in the number of nodes in core 1. The central core (core 2) is less prominent compared to the intermediate core, suggesting a decentralized approach that reduces reliance on a few central stops.



(a) Brussels (b) Madrid

Figure 13: Distribution of nodes across cores in two cities.

5.2.3 Modularity

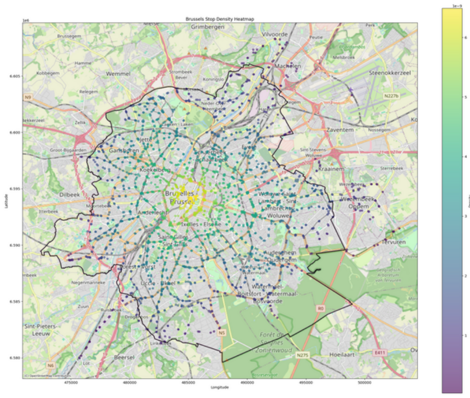
Modularity measures how well a network is divided into communities. Brussels (0.922) has slightly higher modularity than Madrid (0.897), indicating its public transportation network forms more distinct, localized communities, potentially enhancing reliability. Madrid's lower modularity suggests fewer but more interconnected communities (Figure 14), improving accessibility but increasing interdependence and vulnerability to disruptions.

(a) Brussels (b) Madrid

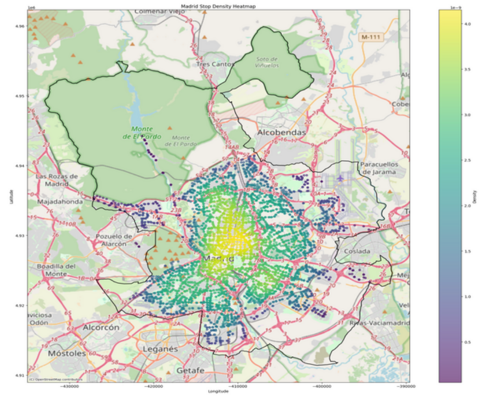
Figure 14: Communities distribution of the considered cities.

5.2.4 Density and Connectivity

Density indicates how interconnected the network is, comparing actual connections to the maximum possible. In terms of density and connectivity, Madrid's network is denser (Figure 15b), supporting high passenger volumes and efficient travel while minimizing isolated nodes. High-centrality hubs ensure smooth connectivity across the city, aligning with the needs of a bustling metropolitan area. In contrast, Brussels sparser network includes a significant number of underutilized or disconnected stops (Figure 15a, which aligns with the region's less urbanized and more regional transit needs.



(a) Brussels



(b) Madrid

Figure 15: Density heatmap of two cities.

6 Conclusion

In summary, the analysis of centrality and community detection metrics reveals key differences in the design and structure of the transit networks of Madrid and Brussels. These differences reflect the distinct urban characteristics, population densities, and transit priorities of each city. Madrid's transit network reflects the demands of a densely populated metropolis, featuring a hub-and-spoke structure centered on efficiency. High closeness centrality at key nodes underscores its focus on connecting urban hubs and ensuring citywide accessibility, even in peripheral areas. This centralized design is further supported by eigenvector centrality, which highlights the strategic placement of important nodes linking the city center to suburban and outlying regions. However, the system's high betweenness centrality at critical hubs reveals a vulnerability—disruptions at these points could have significant ripple effects throughout the network.

To enhance resilience, Madrid could invest in fortifying major hubs and developing contingency plans for potential disruptions. Expanding feeder routes and introducing more intermodal connections would help distribute traffic away from critical points, reducing the system's reliance on these hubs and increasing overall robustness.

Brussels, in contrast, serves a smaller and less densely populated region with a more distributed transit network. Lower closeness centrality values and their wider spread reflect a design tailored to regional needs rather than centralized efficiency. Its eigenvector centrality emphasizes key nodes within the compact urban core, enabling smooth movement in and out of the city center but leaving suburban areas less integrated. On the other hand, the network's relatively even betweenness centrality enhances resilience by reducing dependence on any single hub.

To strengthen its network, Brussels could focus on bridging the gaps between its city center and suburban areas. Building stronger connections between existing clusters would cut travel times and improve accessibility, creating a more cohesive transit system that better meets the needs of its diverse population.

The community detection analysis of Brussels and Madrid's public transportation networks reveals their distinct approaches to connectivity and resilience.

In Brussels, the K-Core decomposition highlights a compact core of highly connected stops surrounded by less-connected outer layers. This centralized design emphasizes efficiency and creates well-defined local communities, as reflected in its slightly higher modularity score (0.922). However, this structure relies heavily on a small number of critical stops, leaving the network

vulnerable to disruptions at these key points. The network's relatively sparse density further underscores its alignment with the region's less urbanized and more distributed transit needs.
Madrid, on the other hand, demonstrates a decentralized and robust core structure, with a broader distribution of high-core stops. This design enhances resilience, ensuring accessibility across urban and suburban areas while accommodating high passenger volumes. Despite a slightly lower modularity score (0.897), the network's denser and more interconnected communities reflect its focus on maintaining smooth connectivity throughout a bustling metropolitan area.
These insights emphasize the strategic trade-offs in transit network planning. While Brussels prioritizes localized efficiency and reliability, Madrid focuses on widespread accessibility and resilience. Together, these findings highlight the importance of tailoring transit designs to the unique demands of a city's geography and demographics, striking a balance between centralization and adaptability.

7 Critique

The study effectively addresses the research problem by providing a comprehensive analysis of the public transportation networks in Madrid and Brussels. It captures key aspects of these systems, such as connectivity, centrality, and community structure, offering valuable insights into their operational dynamics. The use of centrality measures, k-means clustering, and modularity analysis highlights the unique strengths of each city's network—Brussels' centralized design prioritizes urban efficiency, while Madrid's decentralized approach supports resilience and broader accessibility. These findings align closely with the research goal of understanding and comparing network functionality, contributing to improved planning and sustainable development.
While the research makes significant strides, it partially solves the problem as it primarily focuses on structural aspects of the networks. Including additional metrics, such as passenger demand or real-time operational data, could enhance the understanding of network performance. Similarly, exploring temporal dynamics, such as peak versus off-peak variations, would provide a more holistic perspective. Despite these limitations, the methodology and findings lay a strong foundation for future studies and offer actionable insights for urban planners.
If the suggestions from the conclusion—such as incorporating real-time data, addressing accessibility gaps, and optimizing network design—are implemented, they would not only improve the efficiency of public transport systems but also encourage greater public transport usage. This shift could significantly reduce car dependency, leading to lower carbon emissions and promoting environmental sustainability.

References

[1] General Transit Feed Specification (GTFS). Available at: `https://gtfs.org`.

[2] Lin, D., et al. "UN-Habitat Guidelines for Sustainable Urban Development through Improved Public Transport." Available at: `https://unhabitat.org/a-guide-for-municipalities-inclusive-and-sustainable-urban-development-planning-volume-1`.

[3] Python Network Analysis Library, NetworkX. Available at: `https://networkx.org`.

[4] Pandas Library for Data Manipulation. Available at: https://pandas.pydata.org.

[5] GTFS Documentation and Resources. Available at: https://transitfeeds.com.

[6] Mobility Database. Available at: https://mobilitydatabase.org.

[7] STIB/MIVB (Société des Transports Intercommunaux de Bruxelles). Available at: https://www.stib-mivb.be.

[8] EMT Madrid (Empresa Municipal de Transportes de Madrid). Available at: https://www.emtmadrid.es.

[9] Evaluation of Public Transportation System through Social Network Analysis Approach. Available at: https://www.researchgate.net/publication/383439488_Evaluation_of_Public_Transportation_System_through_Social_Network_Analysis_Approach.

[10] Louail, T., et al. "Uncovering the spatial structure of mobility networks." *Scientific Reports*, 2020. Available at: https://www.nature.com/articles/s41598-020-65175-x.