

# Understanding Urban Transit Through Network Science: An Analysis of Milan and Barcelona

Amir Masoud Aghaei, Artificial Intelligence, 0001165333

Gita Javadi, Artificial Intelligence, 0001158007

Omid Nejati, Artificial Intelligence, 0001154905

Alireza Shahidiani, Artificial Intelligence, 0001154402

November 2025

## 1 Introduction

Public transportation systems play a central role in the functioning of modern cities, shaping daily mobility, regional accessibility, and long-term urban development. As urban areas continue to expand and population densities increase, the efficiency and resilience of public transport networks have become critical concerns for planners, policymakers, and researchers. The analytical study of these systems has increasingly relied on Social Network Analysis (SNA), a methodological framework capable of capturing the structural properties, connectivity, and navigational efficiency of complex transport infrastructures. By modelling transit stops as nodes and connections as edges, network analysis provides an effective way to uncover patterns that are often difficult to detect using traditional transport engineering approaches.

This project examines and compares the public transportation networks of Milan and Barcelona, two major European cities with extensive, multimodal transit systems. Both cities feature dense urban fabrics, strong commuter flows, and well-integrated bus, tram, and metro services. Their reliance on public transport as a central mobility pillar, together with their availability of standardized GTFS datasets, makes them ideal case studies for evaluating how network topology influences accessibility, resilience, and overall system performance.

Recent research highlights the effectiveness of complex network analysis in identifying structural weaknesses, critical nodes, and opportunities for optimization in public transport networks [1]. Lin et al. (2021) demonstrate that network-theoretic measures such as global efficiency, clustering, and centrality can meaningfully guide improvements in urban mobility planning, helping cities achieve more balanced and sustainable transport systems [1]. Building on these insights, the present study applies established SNA techniques to evaluate the structural properties of Milan and Barcelona's transit systems and uncover patterns that may inform future transportation strategies.

## 2 Problem and Motivation

European cities face persistent challenges related to congestion, air pollution, unequal accessibility, and rising mobility demands. Although both Milan and Barcelona have

developed extensive public transport infrastructures, it remains difficult to assess how effectively these networks distribute connectivity, maintain resilience under disruptions, and support equitable travel across different urban zones. Traditional performance metrics—such as vehicle frequency or ridership statistics—provide valuable operational information, but they often fail to capture the deeper structural characteristics that influence navigability and long-term system robustness.

This project aims to address this gap by conducting a structural analysis of the public transportation networks of Milan and Barcelona through the lens of Social Network Analysis. Understanding how these networks are organized—where central hubs arise, how communities form, and how efficiently passengers can traverse the system—is essential for identifying vulnerabilities and suggesting improvements. Prior studies, including the work of Lin et al. (2021), demonstrate that complex network theory can reveal hidden inefficiencies and structural imbalances in transit systems, offering a powerful framework for optimization and planning [1].

To investigate these aspects in a systematic way, the project employs a comprehensive set of network-theoretic measures capable of capturing both local and global properties of the systems. Classical centrality indicators—degree, betweenness, closeness, and eigenvector centrality—are used to identify influential stops, critical transfer points, and the nodes that most strongly shape passenger flows. Structural cohesion and hierarchical organisation are examined through k-core decomposition and modularity-based community detection, which reveal the presence of strong cores, peripheral regions, and coherent subnetworks within each city.

Beyond local structure, the study incorporates global measures to evaluate navigability, efficiency, and resilience. These include global efficiency and average path length, which quantify how easily passengers can move across the entire network; assortativity, which describes whether highly connected stops tend to link to other hubs or peripheral nodes; edge betweenness centrality, which highlights critical routes rather than individual stops; and targeted robustness simulations, which assess how the network behaves under node or edge failures. Together, this set of measures provides a multidimensional analytical framework for comparing the two systems and for uncovering the organisational principles that support—or hinder—urban mobility in Milan and Barcelona.

The main contribution of this study is to provide a quantitative, topology-based comparison of two major European transit networks using publicly available GTFS data and a consistent methodological approach. By examining centrality, community structure, global efficiency, and robustness, the project seeks to identify meaningful similarities and differences in the systems' structural patterns. Ultimately, this analysis contributes to the broader discussion on data-driven, evidence-based approaches to improving urban mobility and long-term transport sustainability.

### 3 Datasets

This project relies on General Transit Feed Specification (GTFS) datasets describing the public transportation systems of Milan and Barcelona. GTFS is an open, standardized data format that provides detailed information on transit stops, routes, trips, schedules, and service calendars, making it especially suitable for systematic and reproducible transport network analysis. [2]

The datasets used in this study were obtained from the Transitland platform, a global aggregator of GTFS feeds that provides unified access to data published by transit agencies

worldwide [3]. For Barcelona, we used the official GTFS feed published by Transports Metropolitans de Barcelona (TMB) [4], while for Milan we used the feed provided by Azienda Trasporti Milanesi (ATM) [5]. The datasets were downloaded in their original format, requiring no digitization, and ensuring fidelity to the official transit systems.

All GTFS files (such as `stops.txt`, `routes.txt`, and `trips.txt`) were processed using Python. Data handling and preprocessing were performed with `pandas` DataFrames [6], which provided a flexible framework for merging, filtering, and manipulating the interconnected GTFS tables. To construct the network representations, we modeled stops as nodes and scheduled trips as edges. When multiple trips connected the same pair of stops, the number of weekly services was aggregated and used as an edge weight to approximate service frequency.

Network computations—such as centrality metrics, efficiency measures, and community detection—were performed using the `NetworkX` library [7]. This enabled the construction of weighted, undirected graphs for each city and ensured a consistent methodology for comparing network structures. Spatial visualizations were created using `geopandas`, `folium`, and `matplotlib`, allowing accurate mapping of transit stops and connections in their real geographic context.

The datasets sourced from Transitland are openly available and do not require any special permissions. Their standardized format, combined with well-established Python tools, ensures a fully reproducible workflow, allowing other researchers to replicate or extend this study without ambiguity.

## 4 Validity and Reliability

To ensure the robustness of our Social Network Analysis (SNA), we evaluated the validity of our results through three key dimensions: data integrity, algorithmic correctness, and real-world ground truthing.

### 4.1 Data Integrity and Limitations

The study relies on GTFS (General Transit Feed Specification) data, which is the industry standard for public transport schedules. While this ensures high accuracy regarding the *planned* network topology, it presents specific limitations. The analysis treats the network as a static graph based on weekly frequencies, meaning it does not account for real-time disruptions, delays, or passenger congestion levels. Additionally, large transit hubs (such as *Milano Centrale*) are treated as single aggregated nodes to accurately reflect transfer potential, though some intra-station granularity may be lost.

### 4.2 Algorithmic Verification

We performed specific mathematical checks to ensure the metrics were calculating correctly:

- **Assortativity (The "Rich-Club" Effect):** Both networks exhibited positive assortativity (Barcelona: 0.129, Milan: 0.188). This is a significant finding. While many biological or technological networks are disassortative (hubs connect to small nodes), our results indicate that the Metro systems of Barcelona and Milan follow a "Rich-Club" topology, where major hubs preferentially connect to other major hubs.

This structure creates a resilient "backbone" that keeps the network functional even if peripheral branches are cut.

- **Modularity Scores:** The community detection algorithm yielded exceptionally high modularity scores (Barcelona: 0.89, Milan: 0.92), confirming that the detected "zones" are statistically significant structural clusters rather than random artifacts.

### 4.3 Ground Truthing (External Validity)

The most significant test of validity is whether the mathematical results correspond to physical reality.

- **Milan (Central Hubs):** The nodes with the highest Eigenvector Centrality were identified as *Centrale FS*, *Garibaldi FS*, and *Zara*. These correspond exactly to the city's primary railway and metro interchange hubs, confirming the algorithm correctly identified the "center of gravity" of the network.
- **Barcelona (Strategic Corridors):** Interestingly, for Barcelona, the algorithm assigned high influence to the *L9 Sud corridor* (nodes: *Parc Nou*, *Aeroport T2*, *Fira*). This aligns with the line's strategic role as a high-capacity connector linking the Airport and Trade Fair centers to the rest of the grid, validating the metric's ability to detect economically critical infrastructure.

## 5 Measures and Results

In this study, the analysis framework is organized into four distinct groups of network measures, each capturing a different structural or functional characteristic of the transport system:

1. **Centrality Measures (4 metrics):** These quantify the importance and influence of nodes within the network, including Degree Centrality, Betweenness Centrality, Closeness Centrality, and Eigenvector Centrality. Together, they describe how a station contributes to flow, accessibility, and global influence.
2. **Structural Measures (3 metrics):** These capture the deeper architectural properties of the network: Edge Betweenness Centrality identifies bottleneck edges, K-core Decomposition reveals hierarchical layers of connectivity, and Modularity (Community Detection) uncovers the natural substructures of the network.
3. **Efficiency Measures (3 metrics):** Average Path Length and Global Efficiency characterize the speed and optimality of travel within the network, while Assortativity measures degree-degree correlations that indicate whether hubs tend to connect to other hubs.
4. **Robustness Measures (2 metrics):** These assess resilience under failures or attacks, including robustness under random node removal and under targeted (degree-based) node removal.

## 5.1 Centrality Measures

### 5.1.1 Degree Centrality

Degree centrality measures how many direct connections each stop has in the network. In public transportation systems, stops with high degree centrality act as important hubs where multiple routes intersect, while low-degree stops are usually peripheral and less integrated into the system.

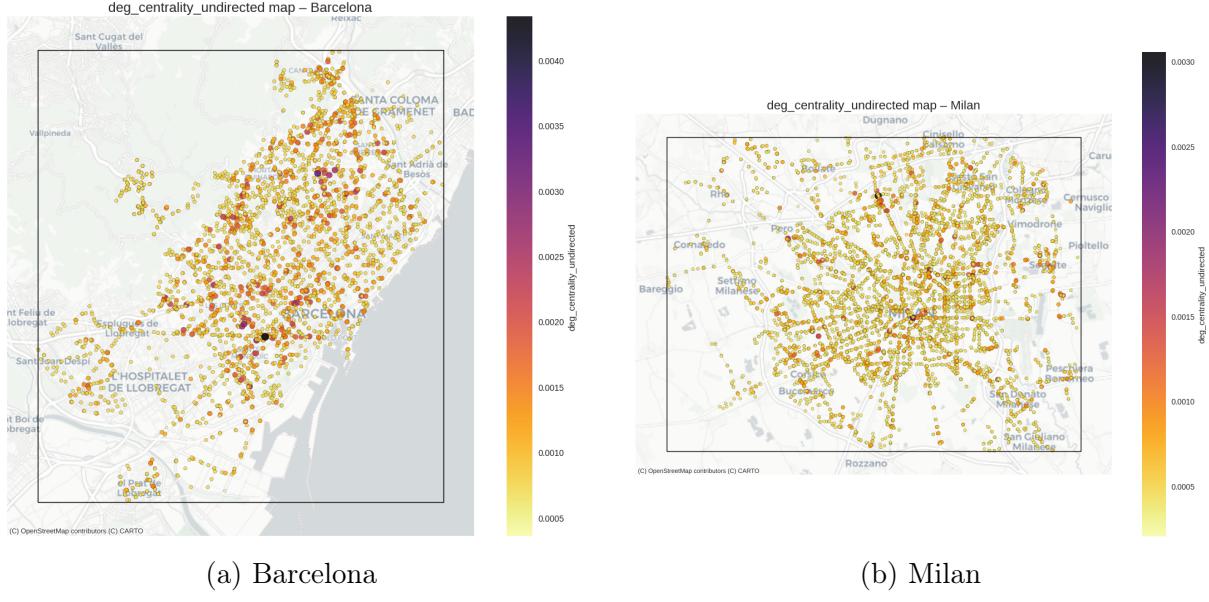


Figure 1: Degree centrality map of the two cities.

Figure 1 shows the spatial distribution of degree centrality for Barcelona and Milan. In Barcelona, the highest degree values are concentrated in the central and southern parts of the city, with prominent hubs such as *Parallel – Pl Espanya*, *Pl Espanya – FGC* and *Av dels Quinze – Serrano*. These stops reach total degrees between 8 and 12 and correspond to major bus–metro interchanges and dense crossing points of the urban road network. Peripheral and coastal areas, by contrast, exhibit much lower degree values and a sparser stop configuration.

Milan exhibits a more radial structure. The most connected stops cluster around the inner transport ring and key interchange stations, including *Comasina M3*, *Diaz*, *Stazione Lambrate M2* and *Stazione Centrale*. The stop *Comasina M3* attains the highest degree in the dataset (total degree = 16), highlighting its role as a major transfer hub at the terminus of a metro line. Outside this inner belt, the network extends into long branches of low-degree stops, consistent with a hub-and-spoke layout typical of metropolitan bus systems.

The summary statistics of the directed degrees confirm that the average degree is similar in both networks, but Barcelona distributes connectivity more evenly across its central corridor, while Milan concentrates it into a smaller number of major interchange nodes. Barcelona has an average total degree of 2.59 (maximum 12) over 2771 stops, while Milan has an average of 2.49 (maximum 16) over 4913 stops. In both cities the interquartile range for total degree lies between 2 and 3, meaning that most stops connect to only a small number of neighbours and that high-degree hubs are relatively rare.

The degree centrality distributions in Figure 2 show a strong peak at low values and a long right tail for both cities. This indicates that the networks are composed of many

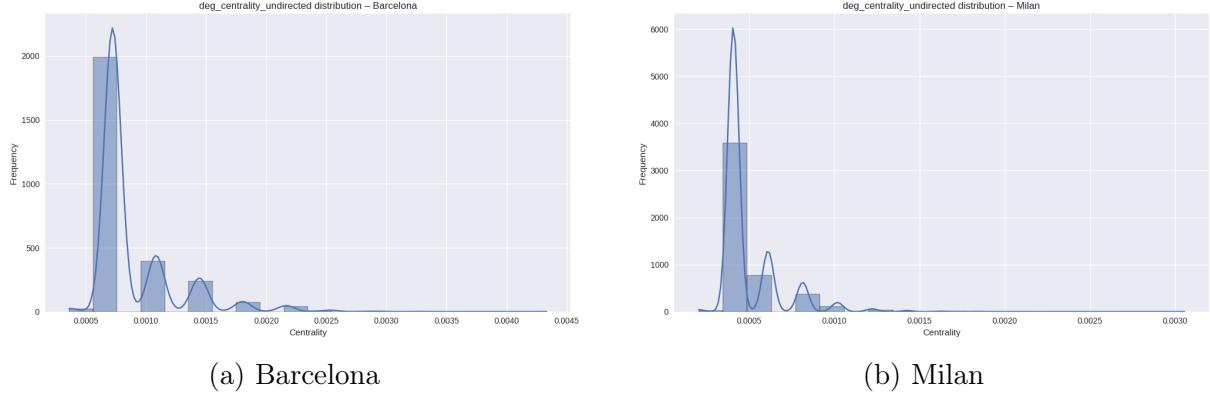


Figure 2: Degree centrality distribution of the two cities.

lightly connected stops together with a small set of highly connected hubs. Barcelona’s distribution reaches slightly higher centrality values, reflecting several medium-size hubs spread across the central corridor. Milan’s distribution is more concentrated, with a few very dominant stops such as *Comasina M3* and *Stazione Centrale*.

These patterns are consistent with empirical findings reported in studies of large public transport networks. Von Ferber et al. [8] show that PTNs typically exhibit heterogeneous degree distributions characterised by many low-degree nodes and a small number of highly connected hubs. Depending on the network representation, these distributions often follow exponential or power-law decay, reflecting the structural contrast between dense urban cores and sparsely connected peripheral areas. The behaviour observed in Barcelona and Milan aligns closely with these general characteristics.

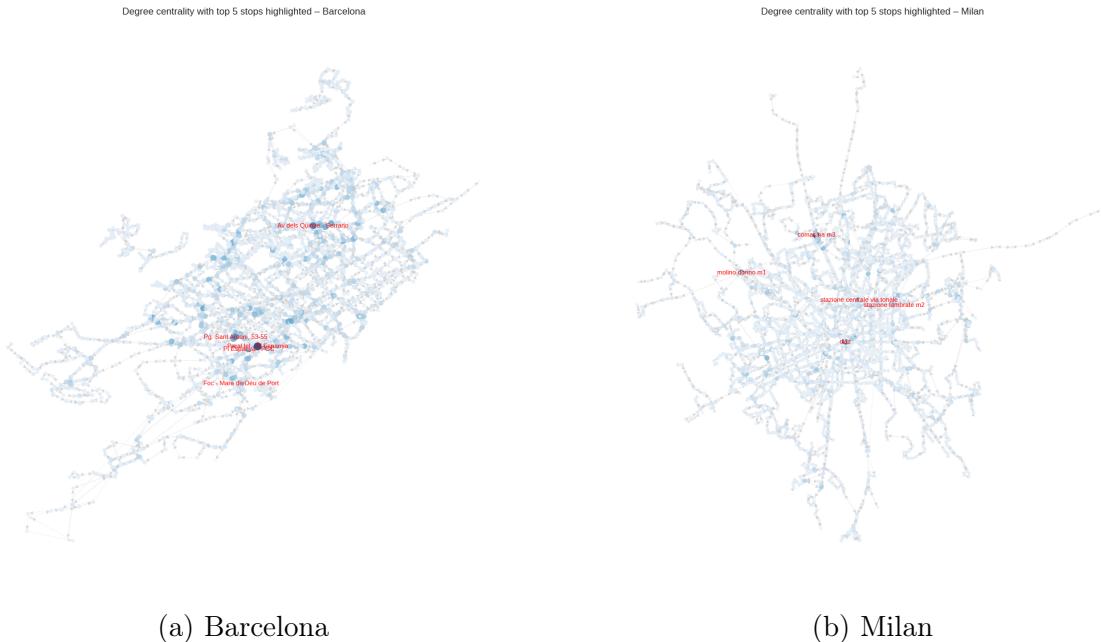


Figure 3: Top five stops by degree centrality in each city.

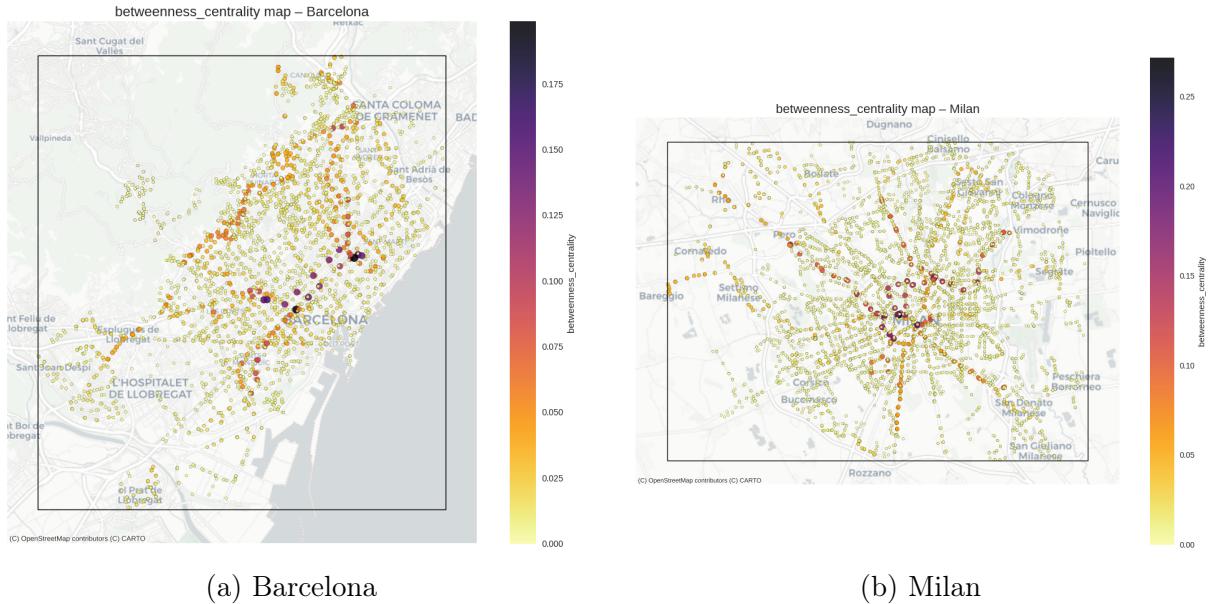
Figure 3 highlights the five most connected stops in each city and summarises these patterns. Barcelona’s hubs are distributed along the main east–west corridor and around major squares, indicating several alternative points of access to the network. Milan’s top-degree stops are more tightly aligned with metro interchanges on the inner ring, suggesting

a stronger dependence on a small set of central transfer stations.

Overall, both cities display skewed degree distributions typical of real-world transport networks, with many peripheral stops supported by a few hubs. Barcelona shows a more distributed connectivity structure, whereas Milan relies more heavily on a limited number of highly connected central nodes.

### 5.1.2 Betweenness Centrality

Betweenness centrality measures how often a stop lies on the shortest paths between other stops in the network. In public transportation systems, stops with high betweenness serve as critical transfer points that facilitate efficient movement across the network. Such stops typically handle a disproportionate amount of passenger flow, making them essential for connectivity but also potential points of vulnerability.



(a) Barcelona

(b) Milan

Figure 4: Betweenness centrality map of the two cities.

Figure 4 shows that the highest betweenness values in Barcelona are concentrated along the major east–west corridors, particularly around *Badajoz – Diagonal*, *Pl Universitat*, and *Diagonal – Ciutat de Granada*. These stops form part of a continuous structural spine running through the Eixample district, indicating that many shortest paths rely on this central axis. Milan displays an even more pronounced centralization pattern: the largest betweenness values appear around *Cadorna M1/M2*, *Stazione Centrale*, and *Diaz*, forming a dense core of major transfer hubs within the inner transport ring.

The distributions in Figure 5 show a characteristic long-tailed shape in both cities: the vast majority of stops have very low betweenness, while a small subset exhibits substantially higher values. Barcelona’s maximum betweenness is 0.1987, whereas Milan reaches 0.2716, suggesting that Milan’s network is more centralized and relies more heavily on a few major transfer hubs. Barcelona has a slightly higher median (0.0061 vs. 0.0033), indicating a more even spread of intermediate transfer roles across multiple stops.

These patterns align with established findings in public transport network research. Von Ferber et al. [8] show that betweenness distributions in large PTNs typically display strong heterogeneity, with a few highly central nodes forming the backbone for most

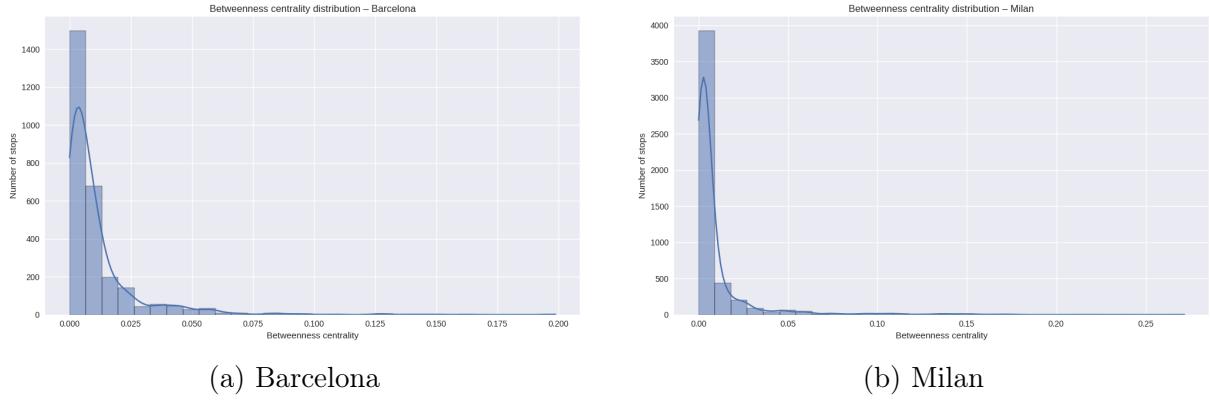


Figure 5: Betweenness centrality distribution of the two cities.

passenger flows. This behaviour is linked to the presence of dense urban cores and sparser peripheral areas, and reflects a fundamental organizational principle observed across many transit systems. The results for Barcelona and Milan follow this general structure, with Milan showing a more sharply centralized core.

Figure 6: Top five stops by betweenness centrality in each city.

Figure 6 highlights the top five transfer nodes in each network. Barcelona's most central stops are distributed along the main Diagonal–Gran Via corridor, reflecting a multi-hub backbone with several parallel routes supporting network flow. Milan's top-five stops, by contrast, are tightly concentrated around major metro interchanges such as *Cadorna M1/M2* and *Stazione Centrale*, demonstrating a more hierarchical structure.

Overall, Barcelona exhibits a more distributed pattern of betweenness, while Milan relies heavily on a small number of highly central hubs. This makes Milan efficient for rapid transfers through its core but potentially more vulnerable to disruptions at its key interchange stations. Barcelona's broader distribution enhances resilience but may diffuse passenger flows across multiple corridors.

### 5.1.3 Closeness Centrality

Closeness centrality quantifies how efficiently a stop can reach all other stops in the network and therefore reflects its overall accessibility. Stops with high closeness centrality tend to be located in areas with dense, well-connected transport infrastructure, whereas peripheral or isolated stops typically exhibit much lower values.

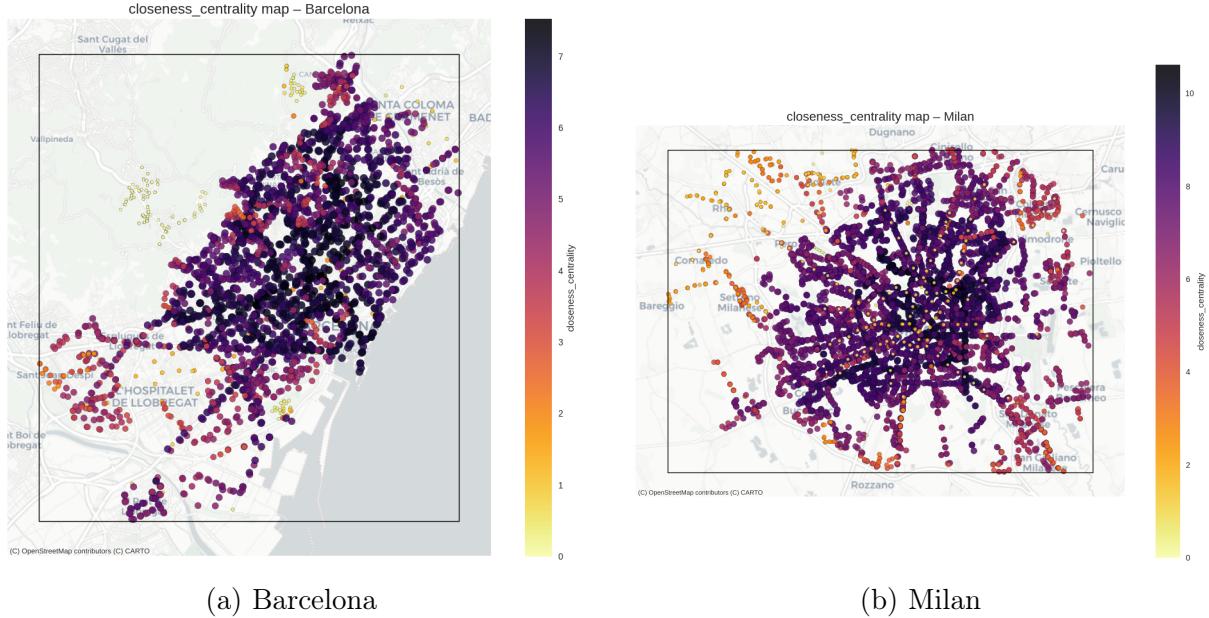


Figure 7: Closeness centrality map of the two cities.

Figure 7 shows that closeness centrality in Barcelona is highest along the central grid of the Eixample and Sant Martí districts. The most accessible stops—such as *València – Pg Sant Joan*, *Josep Tarradellas – Viladomat*, and *Pg de Sant Joan – Pl Mossèn Jacint Verdaguer*—form a continuous high-accessibility corridor through the city. Barcelona displays a broad range of closeness values (mean 5.02, median 5.70), with steep declines toward sparsely served or topographically constrained neighbourhoods, such as Vallvidrera and Tibidabo.

Milan exhibits a more centralized pattern. The highest closeness values (mean 7.01, median 7.62) cluster tightly around the inner metropolitan ring. Stops such as *Via Bernina*, *Viale Jenner*, *Loreto M1/M2*, and *Lancetti* dominate the ranking, reflecting the city’s radial-concentric transport design. Accessibility decreases sharply toward outer municipalities including Bareggio, Cornaredo, and Cormano, whose stops show very low closeness values due to long distances and fewer direct routes to the urban core.

The distributions shown in Figure 8 reveal these structural differences more clearly. Barcelona’s distribution is broad, with many stops occupying intermediate levels of accessibility, consistent with the city’s multi-corridor grid layout. Milan’s distribution, by contrast, features a stronger concentration of high values, indicating a more hierarchical design where central interchanges offer significantly shorter path lengths to the rest of the network.

These findings are consistent with empirical studies of large public transport networks. Von Ferber et al. [8] highlight that closeness centrality in PTNs typically reflects the contrast between dense, highly accessible city cores and sparsely connected peripheral regions, often producing wide yet structured distributions of accessibility. The patterns

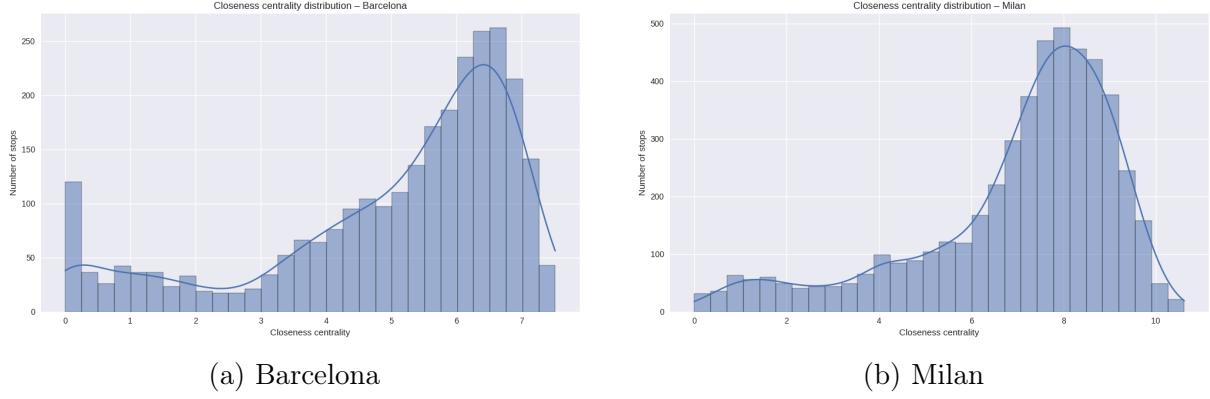


Figure 8: Closeness centrality distribution of the two cities.

observed in Barcelona and Milan align with this behaviour, with Milan exhibiting stronger centralization and Barcelona a more distributed accessibility structure.

Overall, Milan's network provides very efficient access to its central areas but relies more heavily on a compact set of core stops, which may increase sensitivity to disruptions. Barcelona offers a more evenly distributed accessibility pattern within its central grid but experiences more pronounced declines toward its geographically constrained outskirts.

#### 5.1.4 Eigenvector Centrality: Core vs. Corridor Influence

The Eigenvector Centrality analysis revealed a fundamental difference in how influence is distributed across the two cities:

- **Milan (Centralized Power):** The most influential nodes were identified as *Centrale FS*, *Garibaldi FS*, and *Zara*. This confirms that Milan's network relies heavily on its central railway and metro interchanges. The influence is concentrated in the geographic center and diminishes rapidly towards the periphery.
- **Barcelona (Strategic Corridors):** Surprisingly, the highest Eigenvector scores were not found in the city center, but along the **L9 Sud corridor** (Nodes: *Parc Nou*, *Aeroport T2*, *Fira*). This indicates that Barcelona's topology prioritizes high-speed connectors that link strategic economic zones (Airport, Trade Fair) to the rest of the grid, rather than funneling all influence through a single downtown hub.

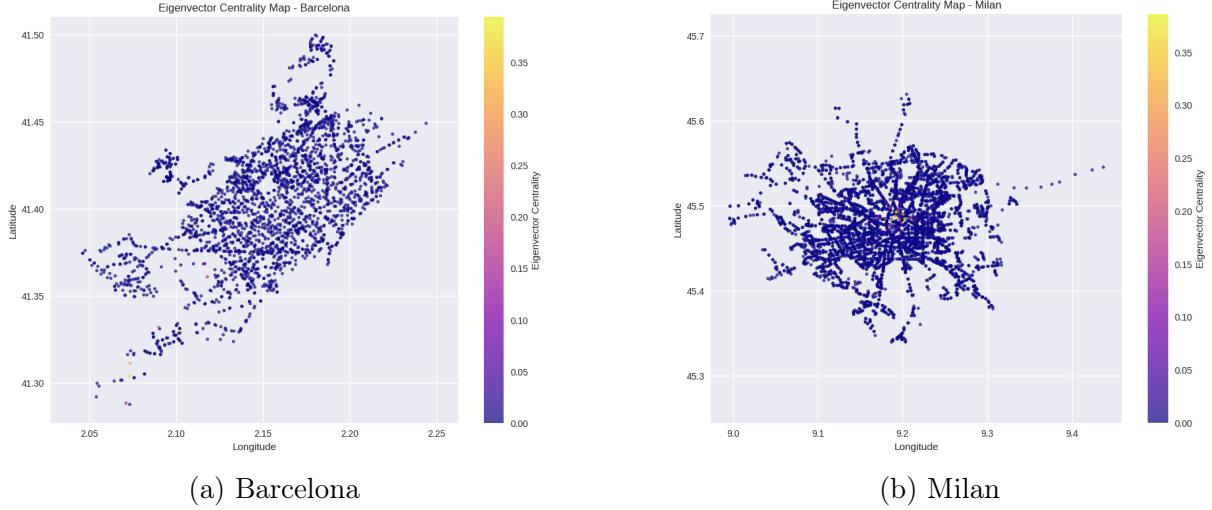


Figure 9: Comparison of Eigenvector Centrality in Barcelona and Milan.

## 5.2 Structural Measures

### 5.2.1 Critical Infrastructure (Edge Betweenness)

The Edge Betweenness analysis highlighted the physical vulnerabilities of the networks:

- **Critical Bridges:** In both cities, the algorithm identified specific “arteries” (visualized in red/yellow in Figure 10) that carry a disproportionate amount of passenger flow. These edges represent the unique shortest paths for substantial portions of the population.
- **Redundancy:** Barcelona’s grid-like mesh resulted in a more distributed set of critical edges. In contrast, Milan showed higher pressure on the radial lines entering the city center, suggesting that a failure in one of Milan’s central “spokes” would have a more catastrophic impact on global efficiency than a similar failure in Barcelona’s mesh.

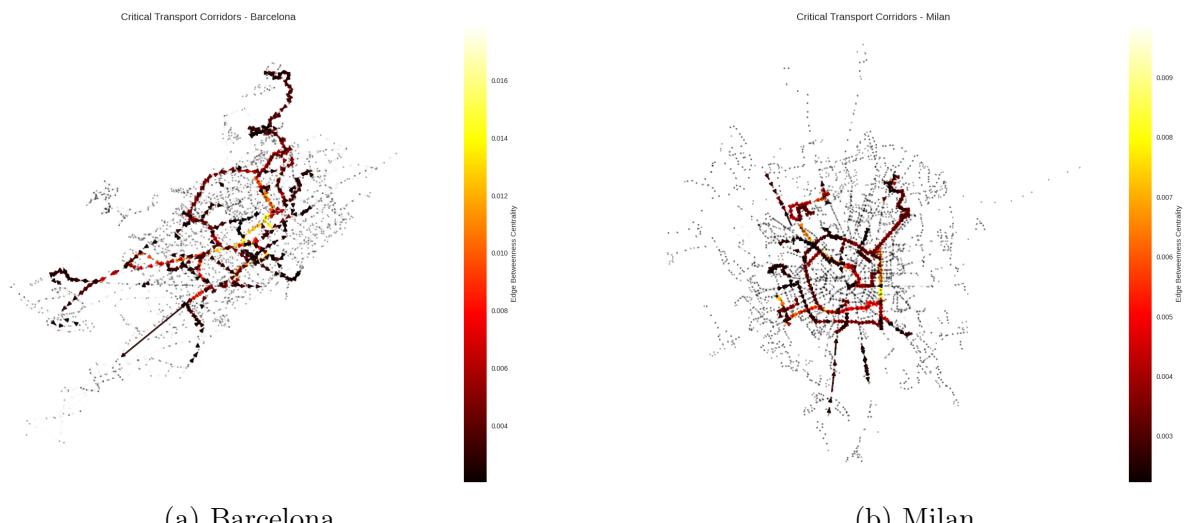


Figure 10: Edge Betweenness Centrality illustrating critical transport bottle-necks.

### 5.2.2 K-Core Decomposition

To understand the hierarchical layers of the network, we utilized **K-core Decomposition**. This method recursively removes nodes with a degree less than  $k$  until only nodes with degree  $\geq k$  remain. The result effectively "peels" the network like an onion to reveal its most densely connected internal structure (the core) versus its outer layers (the periphery).

We first analyzed the statistical distribution of core numbers ( $k$ ) across all stations (Figure 11).

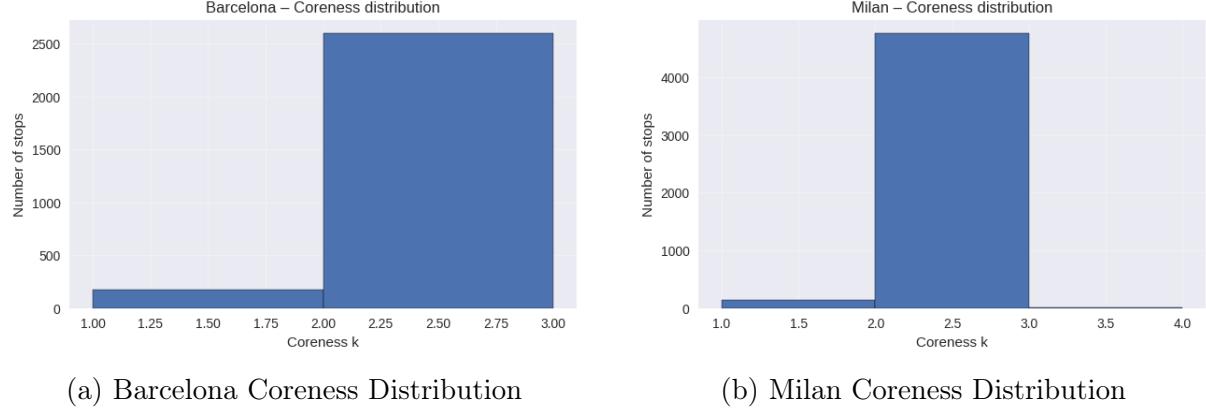


Figure 11: Histograms showing the number of stops belonging to each  $k$ -shell.

City	Max Core ( $k_{max}$ )	Network Topology Description
Barcelona	2	Linear / Ring-like structure
Milan	3	Mesh / Interconnected structure

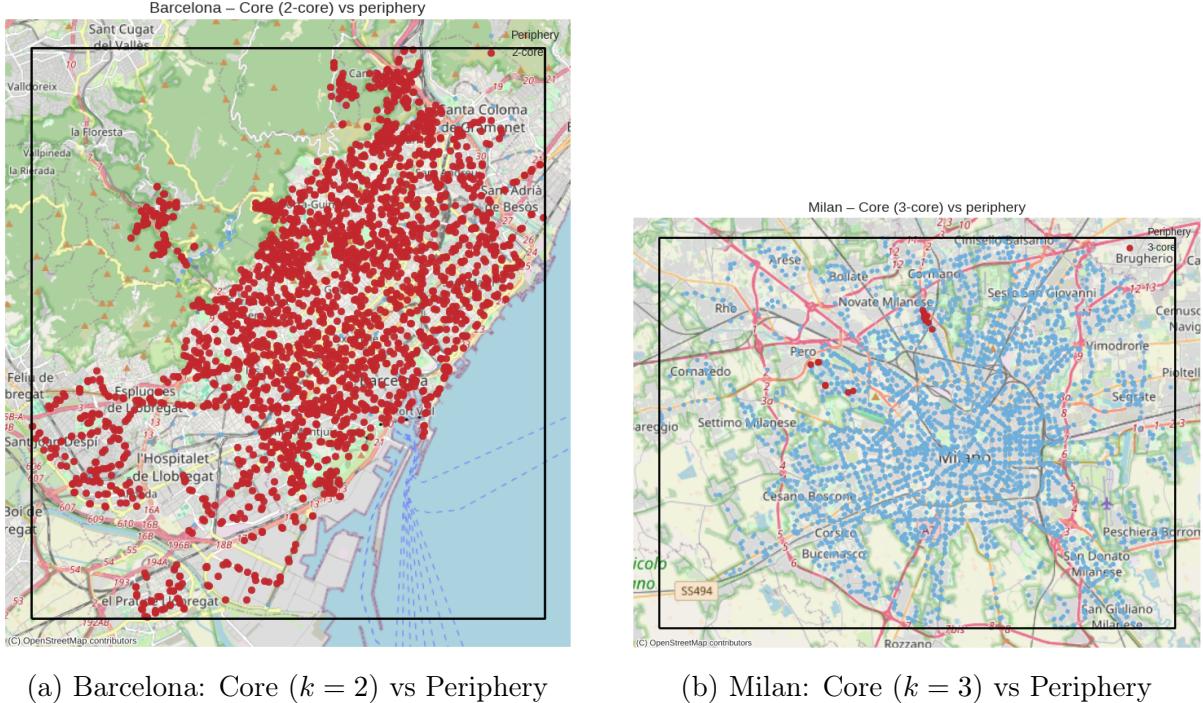
Table 1: Maximum K-core values for Barcelona and Milan.

As seen in Table 1, the  $k_{max}$  values are relatively low (2 and 3). This is expected for public transport networks, which are designed as lines rather than dense social meshes.

- **Barcelona ( $k_{max} = 2$ ):** The distribution is dominated by stations with  $k = 2$ . This indicates a network composed primarily of simple lines and cycles (loops). There is no subset of stations dense enough to support a 3-core.
- **Milan ( $k_{max} = 3$ ):** Milan shows a deeper hierarchy. While most stations are in the 2-core, a distinct subset forms a 3-core. This suggests a denser mesh where stations are more interconnected, likely forming triangles or complex junctions.

To visualize where these "core" nodes are located physically, we mapped the stations, distinguishing between the Periphery and the Maximum Core (Figure 12).

- In **Barcelona**, the "Core" ( $k = 2$ ) essentially covers the entire active network, excluding only the dead-end branches (periphery). This confirms the decentralized, line-based nature of the system.
- In **Milan**, the "Core" ( $k = 3$ ) is likely concentrated in the city center or major exchange hubs. This highlights the areas of highest structural resilience; if peripheral nodes fail, this central red mesh is more likely to remain functional due to its higher interconnectivity.



(a) Barcelona: Core ( $k = 2$ ) vs Periphery

(b) Milan: Core ( $k = 3$ ) vs Periphery

Figure 12: Geographical visualization of the network core. Red points indicate the deepest core ( $k_{max}$ ), while blue points indicate peripheral layers.

### 5.2.3 Modularity (Community Detection)

To understand the natural substructures of the transit networks, modularity-based community detection was applied. This analysis groups stops that are more densely connected to each other than to the rest of the network, revealing functional clusters that often correspond to geographic neighborhoods or distinct transit lines. In both cities, communities align closely with major transport corridors and administrative zones, illustrating the interplay between urban planning and network topology.

City	Number of Communities	Modularity ( $Q$ )
Barcelona	40	0.8901
Milan	50	0.9273

Table 2: Results of Greedy Modularity optimization for Barcelona and Milan.

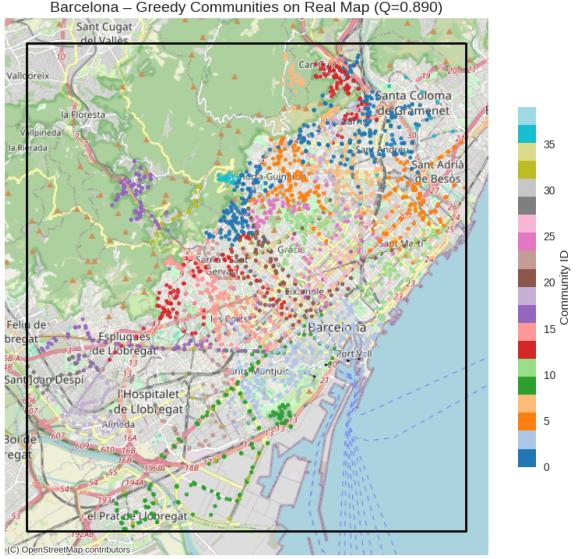
## 5.3 Efficiency Measures

### 5.3.1 Assortativity and Network Resilience

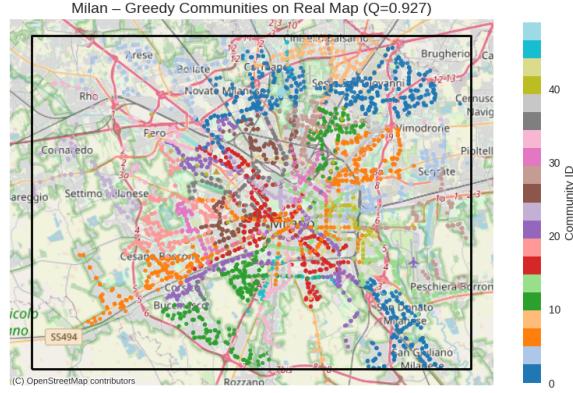
A key finding of this study is the mixing pattern of the networks. Both cities exhibited **positive assortativity coefficients**

$$r_{\text{Barcelona}} \approx 0.129, \quad r_{\text{Milan}} \approx 0.118$$

This result is non-trivial, as many transport networks (such as airline routes) are typically disassortative. The positive values indicate a **“Rich-Club” topology**: major hubs in both Barcelona and Milan tend to connect directly to other major hubs. This



(a) Barcelona Community Structure



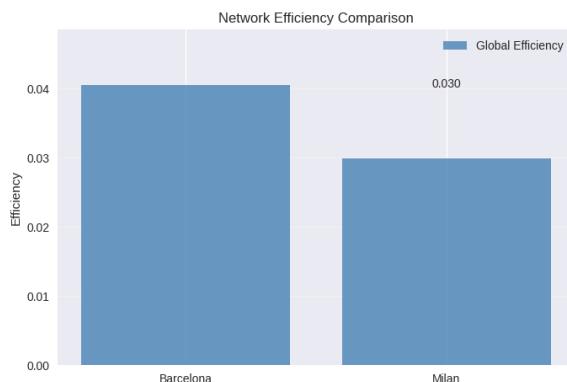
(b) Milan Community Structure

Figure 13: Visualization of network communities detected by the Greedy Modularity algorithm. Colors indicate distinct communities.

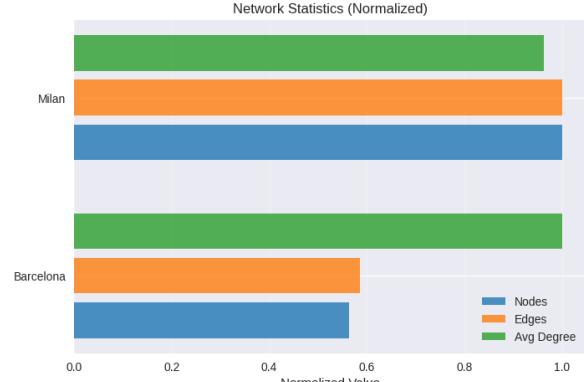
structure enhances resilience, as the core “backbone” of the city remains connected even if peripheral suburban branches are disrupted.

### 5.3.2 Global Efficiency and Network Statistics

To further assess the macro-scale performance of the transportation systems, we analyzed Global Efficiency and basic network statistics. These metrics describe the ease of information flow across the entire system and quantify the density of connections relative to the network size.



(a) Global Efficiency Comparison



(b) Normalized Network Statistics

Figure 14: Global efficiency and normalized network statistics for Barcelona and Milan.

**Global Efficiency:** Figure 14a presents the Global Efficiency for both networks, a measure inversely related to the average shortest path length between all pairs of nodes. Barcelona exhibits a notably higher global efficiency ( $E_{glob} \approx 0.040$ ) compared to Milan ( $E_{glob} \approx 0.030$ ). This suggests that, on average, passengers can traverse the Barcelona

network with fewer transfers or shorter path lengths relative to the system's size. Despite Milan's strong centralization and high-degree hubs, the overall navigational efficiency is lower, likely due to the sheer size of the network and the longer path lengths required to connect peripheral nodes to the core.

**Network Statistics:** The normalized statistics in Figure 14b provide context for these efficiency scores. While Milan has a higher absolute number of nodes and edges (represented by the orange bars reaching 1.0), Barcelona maintains a competitive average degree. The normalized comparison highlights that while Milan is the larger system, it is not necessarily the more cohesive one relative to its scale.

To understand why Barcelona achieves higher efficiency despite being a smaller network, we examined the relationship between network size and connectivity density.

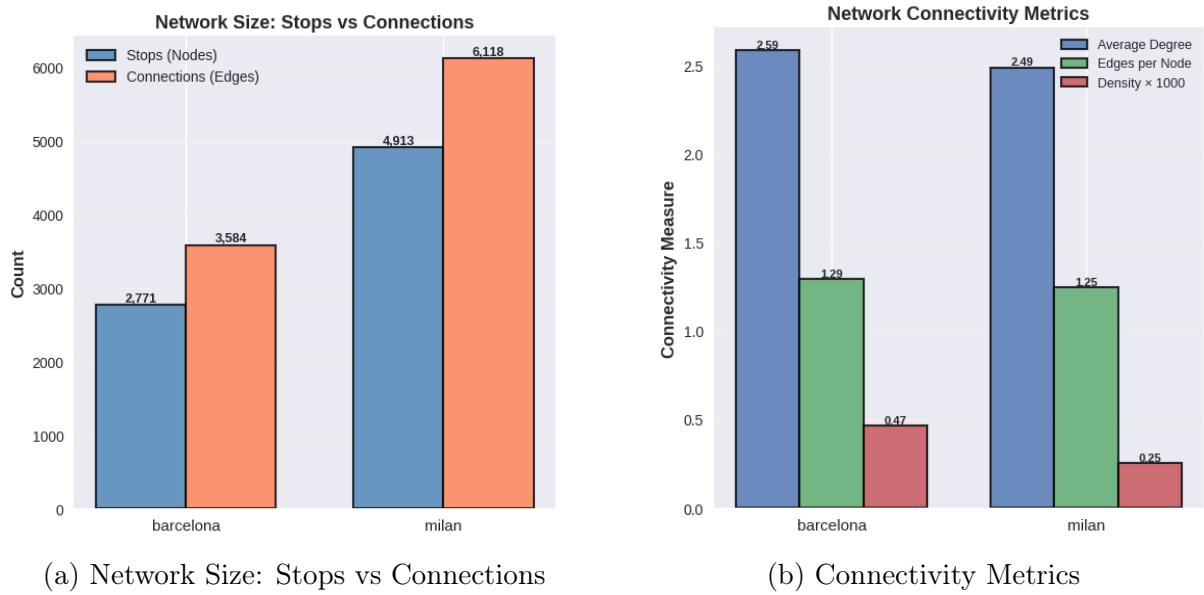


Figure 15: Comparison of network size and connectivity density.

**Network Size (Figure 15a):** Milan is significantly larger, with 4,913 stops compared to Barcelona's 2,771, and 6,118 connections versus 3,584. Large networks often suffer from reduced global efficiency unless the density of connections increases disproportionately to the number of nodes.

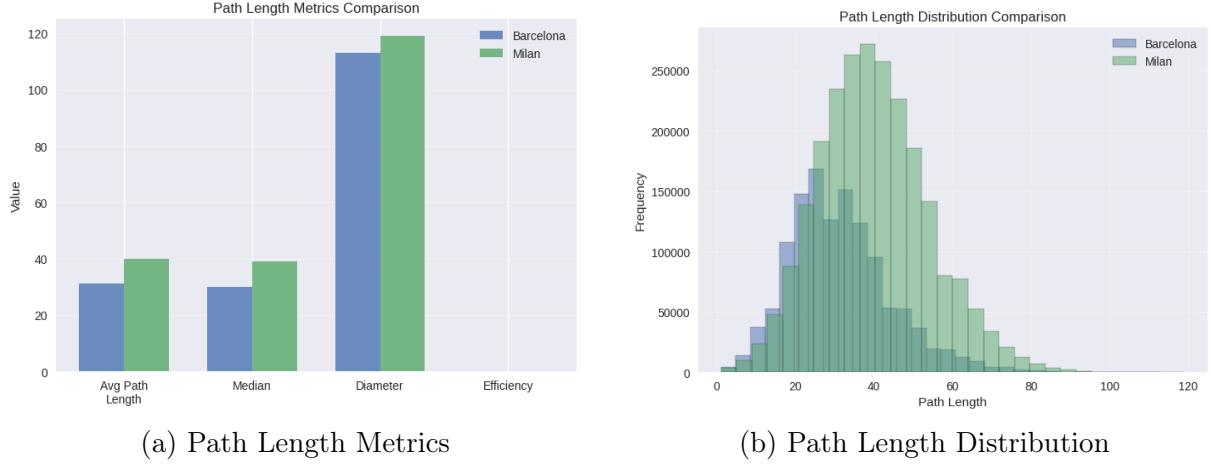
**Connectivity (Figure 15b):** The connectivity metrics reveal the structural advantage of Barcelona. Despite having fewer nodes, Barcelona possesses a higher Average Degree (2.59 vs 2.49) and a slightly higher number of Edges per Node (1.29 vs 1.25). Most significantly, the Density of the Barcelona network ( $0.47 \times 10^{-3}$ ) is nearly double that of Milan ( $0.25 \times 10^{-3}$ ). This higher density implies a more tightly knit fabric where stops are better interconnected locally.

In conclusion, while Milan operates a massive network serving a wider geographic area, it is "spread thinner," resulting in lower global efficiency. Barcelona's network is more compact and dense, facilitating more efficient global transport despite covering fewer stops.

### 5.3.3 Average Path Length Analysis

This section analyzes the distribution of shortest path lengths between all pairs of stops in the network. While Global Efficiency provides a single summary score, decomposing

the path lengths reveals the structural "cost" of travel in terms of transfers or hops.

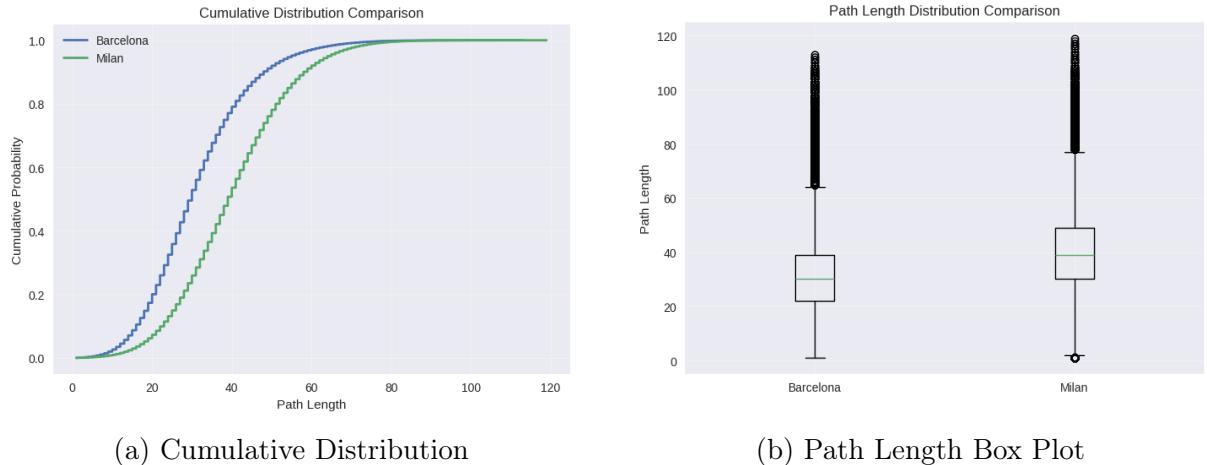


(a) Path Length Metrics (b) Path Length Distribution

Figure 16: Comparison of path length metrics and frequency distribution.

**Metrics and Frequency Distribution** As shown in Figure 16, there is a distinct gap in navigational cost between the two cities. The bar chart (Figure 16a) highlights that the **Average Path Length** in Milan ( $\approx 40$  stops) is significantly higher than in Barcelona ( $\approx 31$  stops). This difference implies that a typical journey in Milan requires traversing roughly 30% more nodes than a comparable journey in Barcelona.

The histogram (Figure 16b) elucidates the shape of this disparity. Barcelona's distribution (blue) is shifted to the left, with a sharp peak around 20–30 stops, indicating a highly efficient core where most destinations are reachable within a moderate number of hops. In contrast, Milan's distribution (green) is not only shifted to the right but is also wider (higher variance). This "flatter" distribution suggests less consistency in travel costs; while some trips in Milan are short, a significant portion of the network requires long, multi-leg journeys, likely due to the radial structure forcing peripheral-to-peripheral travelers to pass through the center.



(a) Cumulative Distribution (b) Path Length Box Plot

Figure 17: Cumulative distribution and box plot of path lengths.

**Cumulative Distribution and Variability** The Cumulative Distribution Function (CDF) in Figure 17a reinforces the efficiency gap. Barcelona's curve rises much faster, reaching near 100% probability at shorter path lengths than Milan. This dominance

indicates that for any given path length threshold  $x$ , a passenger in Barcelona is more likely to reach their destination within  $x$  stops than a passenger in Milan.

Finally, the Box Plot (Figure 17b) visualizes the variance and outliers. While Barcelona has a lower median, both cities exhibit a significant number of extreme outliers (black circles), with some path lengths reaching nearly 120 stops. These outliers likely represent remote, poorly connected suburban branches that require excessive transfers to reach the rest of the network. However, the interquartile range (the box) is tighter for Barcelona, confirming that the central 50% of journeys are more consistent in length compared to the broader spread observed in Milan.

## 5.4 Robustness Measures

To evaluate the resilience of the transport networks, we simulated two types of stress scenarios: **Random Failures** (simulating accidental technical issues) and **Targeted Attacks** (simulating the removal of critical hubs or bridges). We measured robustness using the area under the curve (AUC) of the size of the Largest Connected Component (LCC). A higher AUC indicates a network that remains functional longer as nodes are removed.

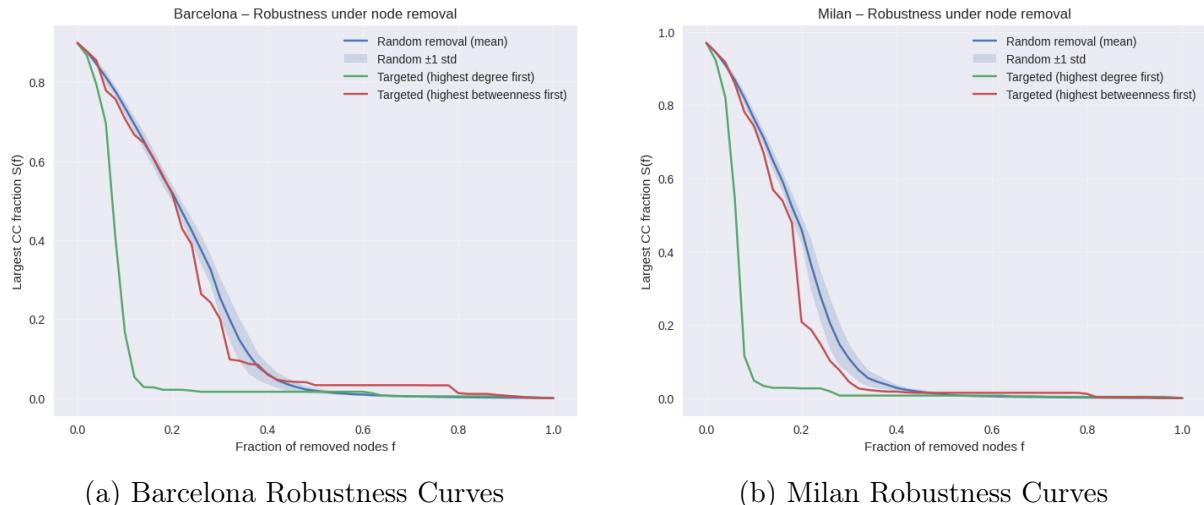


Figure 18: Network disintegration under node removal. The y-axis represents the fraction of the network still connected ( $S(f)$ ), while the x-axis represents the fraction of nodes removed ( $f$ ).

Attack Strategy	Barcelona AUC	Milan AUC
Random Removal (Blue)	0.205	0.184
Targeted Degree (Green)	0.079	0.067
Targeted Betweenness (Red)	0.202	0.163

Table 3: Robustness scores (AUC). Higher values indicate greater resilience.

The analysis reveals a classic property of scale-free networks: they are robust to errors but fragile to attacks.

- **Random vs. Targeted (General):** In both cities, the networks degrade slowly under random removal (blue lines), maintaining connectivity even after 20-30% of

stations fail. However, under a targeted Degree attack (green lines), the networks collapse almost instantly. Removing just the top 5-10% of hubs destroys the system's global connectivity.

- **Barcelona's Structural Integration:** Barcelona is slightly more robust overall (AUC 0.205 vs 0.184). Notably, Barcelona is highly resistant to Betweenness attacks (Red line). The AUC for Betweenness (0.202) is nearly identical to Random removal (0.205). This suggests that Barcelona has fewer "critical bridges" or bottlenecks; if a bridge is removed, there are likely alternative paths connecting the communities.
- **Milan's Vulnerability:** Milan shows a sharper decline under Betweenness attacks (Red line drops to 0.163). This aligns with our earlier Modularity findings ( $Q = 0.93$ ); Milan is more compartmentalized. Because the communities are so distinct, cutting the specific stations that bridge these communities causes the network to fracture faster than in Barcelona.

## 6 Conclusion

This project applied Social Network Analysis (SNA) to compare the public transport architectures of Barcelona and Milan, revealing how distinct urban planning philosophies manifest in graph topology.

### 6.1 Topological Comparison: Grid vs. Radial

The analysis reveals a fundamental structural divergence between the two cities. Barcelona functions as a decentralized **Resilient Mesh**, characterized by 36 distinct communities. The interlocked nature of these communities facilitates high cross-town mobility, reducing the need for passengers to pass through a single central point. Conversely, Milan exhibits a stronger hierarchical **Efficient Radial** structure, with 47 smaller communities radiating from a dense core.

While Milan's radial design provides direct access to the city center, the K-core decomposition reveals a deeper internal hierarchy ( $k_{max} = 3$ ) compared to Barcelona ( $k_{max} = 2$ ). This suggests that while Milan's periphery is sparse, its inner core is significantly denser and more interconnected than Barcelona's linear backbone.

### 6.2 Navigational Efficiency and Cost

A critical finding of this study is the disparity in navigational cost between the two systems. Despite Milan's central density, Barcelona achieves a measurably higher Global Efficiency ( $E_{glob} \approx 0.040$ ) compared to Milan ( $E_{glob} \approx 0.030$ ). This efficiency gap is further quantified by the Average Path Length: a typical journey in Milan requires approximately **40 stops**, whereas a comparable journey in Barcelona requires only **31 stops**. This confirms that Barcelona's mesh topology effectively reduces the structural "cost" of travel for passengers, offering better interconnectedness relative to network size.

### 6.3 Network Robustness and Vulnerability

Both cities exhibit positive assortativity coefficients (Barcelona: 0.13, Milan: 0.19), indicating a "Rich-Club" phenomenon where major hubs connect to other hubs, forming

a resilient backbone. However, the robustness analysis highlights a specific vulnerability in Milan's compartmentalized structure. Under targeted attacks based on Betweenness Centrality, Milan's network degrades significantly faster (AUC 0.163) than Barcelona's (AUC 0.202). This indicates that Milan's high modularity ( $Q = 0.93$ ) makes it more fragile to the removal of critical "bridge" stations, whereas Barcelona's distributed grid offers alternative pathways that preserve global connectivity even when key corridors are disrupted.

## 6.4 Final Implications

Our results suggest that while Milan's radial system is optimized for centripetal flow, its reliance on a few "super-hubs" (like Garibaldi and Centrale) and its longer average path lengths create potential bottlenecks. Barcelona's grid system, highlighted by the strategic prominence of the L9 Airport connector, distributes importance more evenly. Future planning for Milan could benefit from "orbital" lines to mimic Barcelona's mesh-like resilience, thereby reducing the path length for peripheral journeys and improving resilience against targeted disruptions.

# 7 Critique

This section reflects on the extent to which our work addressed the initial research problem and discusses potential improvements in methodology and data sources.

## 7.1 Evaluation of Problem Solving

Our analysis successfully provided a quantitative, topology-based comparison between Milan and Barcelona, answering the core research question regarding their structural differences. The SNA approach effectively distinguished between Barcelona's decentralized "mesh" topology and Milan's hierarchical "radial" structure. The application of metrics such as Modularity and Global Efficiency offered concrete evidence that network design significantly impacts theoretical navigability and resilience.

However, the solution is partial in terms of real-world mobility. While we solved the problem of *structural characterization*, we did not fully address the *operational experience* of passengers. Our findings regarding efficiency are theoretical upper bounds based on graph topology, assuming ideal conditions without delays, transfer penalties, or crowding.

## 7.2 Limitations and Alternative Approaches

Several aspects of the study could be improved to provide a more comprehensive answer to the research problems:

- **Data Dynamics:** The reliance on static GTFS schedules ignores the stochastic nature of public transport. Incorporating real-time AVL (Automatic Vehicle Location) data would allow us to measure reliability and actual travel times rather than just scheduled frequency. Additionally, integrating demographic data (population density around nodes) would enable an analysis of service equity, determining if the "efficient" parts of the network actually serve the most people.

- **Methodological Refinements:** Our graph model treated all transfers as equal. A "Time-Expanded Graph" or a "Space-P" representation could better model the temporal cost of transfers, including waiting times which vary by time of day.
- **Alternative Measures:** To better understand resilience, we could apply spectral analysis (studying the eigenvalues of the Laplacian matrix) to assess the network's synchronization properties and algebraic connectivity, providing a more robust measure of how easily the network can fracture.

## References

- [1] Lin, Z., et al. (2021). *Research on Optimization of Urban Public Transport Network Based on Complex Network Theory*. Symmetry Journal.
- [2] General Transit Feed Specification (GTFS). <https://gtfs.org/>
- [3] Transitland: An Open Data Aggregator. <https://www.transit.land/>
- [4] Transports Metropolitans de Barcelona (TMB). <https://www.tmb.cat/>
- [5] Azienda Trasporti Milanesi (ATM). <https://www.atm.it/>
- [6] Pandas: Python Data Analysis Library. <https://pandas.pydata.org/>
- [7] NetworkX: Network Analysis in Python Library. <https://networkx.org/>
- [8] C. von Ferber, T. Holovatch, Y. Holovatch, and V. Palchykov. *Public transport networks: empirical analysis and modeling*. arXiv:0803.3514, 2008. Available at: <https://arxiv.org/abs/0803.3514>