Journal Pre-proof

Urban Mobility Insights: A Dataset for Exploring Network Topology and City Dynamics

D.D. Herrera-Acevedo , D. Sierra-Porta

PII: S2352-3409(25)00798-X
DOI: https://doi.org/10.1016/j.dib.2025.112076
Reference: DIB 112076

To appear in: *Data in Brief*

Received date: 29 October 2024
Revised date: 2 September 2025
Accepted date: 11 September 2025

Please cite this article as: D.D. Herrera-Acevedo , D. Sierra-Porta , Urban Mobility Insights: A Dataset for Exploring Network Topology and City Dynamics, *Data in Brief* (2025), doi: https://doi.org/10.1016/j.dib.2025.112076

ARTICLE INFORMATION

Article title

Urban Mobility Insights: A Dataset for Exploring Network Topology and City Dynamics

Authors

D. D. Herrera-Acevedo[a,*] (ORCID: 0000-0002-2060-7040) and D. Sierra-Porta[b,**] (ORCID: 0000-0003-3461-1347)

Affiliations

[a]Universidad Tecnológica de Bolívar. Facultad de Ingeniería, Parque Industrial y Tecnológico Carlos Vélez Pombo Km 1 Vía Turbaco, Cartagena de Indias, Bolívar, 130010, Colombia.

[b]Universidad Tecnológica de Bolívar. Facultad de Ciencias Básicas, Parque Industrial y Tecnológico Carlos Vélez Pombo Km 1 Vía Turbaco, Cartagena de Indias, Bolívar, 130010, Colombia.

Corresponding author's email address and Twitter handle

acevedod@utb.edu.co (D. D. Herrera-Acevedo), dporta@utb.edu.co (D. Sierra-Porta)

Keywords

Urban Network Structures, Complex Network Analysis, Sustainable Urban Development

Abstract

This article presents a comprehensive dataset capturing the urban network structures and sociodemographic variables of 65 cities worldwide for the year 2023, based on the Urban Mobility Readiness Index (UMRi) developed by the Oliver Wyman Forum. The dataset includes key metrics such as graph entropy, node degree, clustering coefficient, graph diameter, GDP per capita, and population density, among others, which are essential for analysing the relationship between network topology and urban mobility readiness. By offering detailed insights into these urban networks, this dataset serves as a valuable resource for cities not currently included in major mobility rankings, allowing them to evaluate their mobility readiness in relation to established indices like the UMRi. Urban planners and researchers can leverage this data to explore complex urban mobility dynamics and develop strategies to enhance transportation systems, particularly in rapidly growing or underserved regions. The dataset is structured for seamless integration with various analytical tools, making it a vital asset for both urban planning and research aimed at fostering sustainable and efficient urban development.

SPECIFICATIONS TABLE

Subject	Urban Mobility, Network Topology, Complex Systems.
Specific subject area	Urban mobility readiness, sustainable urban development, complex network analysis, graph theory.
Type of data	Table, Chart & Raw and Filtered.
Data collection	The data were collected from 65 cities worldwide in 2023, using the Urban Mobility Readiness Index developed by the Oliver Wyman Forum. Key metrics

	such as graph entropy, node degree, clustering coefficient, GDP per capita, and population density were gathered from public data sources and mobility research reports. The data were processed and validated using standard complex network analysis methods, ensuring consistency and accuracy for urban mobility evaluation.
Data source location	The urban network data for this study were collected from 65 cities worldwide. The primary metrics were derived from the Urban Mobility Readiness Index and supplemented by publicly available datasets. Data processing and validation took place at the Universidad Tecnológica de Bolívar, Cartagena, Colombia (10.391049, -75.479426).
Data accessibility	Repository name: Mendeley Data. Data identification number: 10.17632/gmyt9wrgst.1 Direct URL to data: https://data.mendeley.com/datasets/gmyt9wrgst/1
Related research article	This article supports tables that are used in: Herrera-Acevedo, D. D., & Sierra-Porta, D. (2025). Network structure and urban mobility sustainability: A topological analysis of cities from the urban mobility readiness index. Sustainable Cities and Society, 119, 106076. https://doi.org/10.1016/j.scs.2024.106076.

VALUE OF THE DATA

- The dataset provides a detailed collection of urban network structures and socio-demographic variables for 65 cities worldwide, capturing critical metrics such as graph entropy, node degree, clustering coefficient, graph diameter, GDP per capita, and population density. This data is essential for understanding the complex relationship between network topology and urban mobility readiness, offering a solid foundation for analysing and improving urban transportation systems.
- This dataset enables cities that are not currently included in major mobility rankings, such as the Urban Mobility City Rankings by the Oliver Wyman Forum, to assess their urban network features and relate them to established indices. By doing so, these cities can better understand their mobility readiness and identify areas for improvement, thereby contributing to more inclusive and comprehensive mobility and sustainability rankings.
- Urban planners and researchers can leverage this dataset to explore the structural aspects of urban networks in relation to socio-demographic variables, facilitating the development of targeted strategies for enhancing transportation systems. The data's format supports seamless integration with statistical and network analysis tools, making it invaluable for a wide range of urban planning and research applications, particularly in cities undergoing rapid development or facing unique infrastructural challenges.
- This dataset offers critical insights that can be related to the Urban Mobility Readiness Index, particularly for cities that are not currently represented in such rankings. By analysing these insights, cities of all sizes can gain a better understanding of their standing in terms of urban mobility and sustainability, providing them with the tools needed to align with global standards and improve their infrastructure and policies accordingly.
- Beyond serving as a complement to the UMRi, this dataset uniquely integrates socio-demographic variables with graph-theoretical indicators of urban networks. This integration enables specific research applications such as (i) investigating the relationship between network connectivity patterns and public transit performance, (ii) exploring how graph

entropy or clustering coefficients correlate with sustainability outcomes, (iii) benchmarking underrepresented cities against global leaders in mobility readiness, and (iv) using network metrics as explanatory variables in econometric or machine learning models of urban efficiency. Unlike existing resources that focus solely on infrastructure readiness scores or demographic statistics, our dataset provides the raw graph structures (GraphML files) and derived topological metrics, making it possible to replicate, extend, or combine with other spatial and environmental datasets. This dual availability of structural data and processed indicators ensures that researchers and policymakers can not only validate mobility indices but also design novel methodologies for assessing resilience, accessibility, and sustainability in diverse urban contexts.

BACKGROUND

Understanding the structural organization of cities is crucial for assessing how urban areas are progressing towards improved urban mobility and sustainable development [1, 2]. As global urbanization continues to accelerate, the need for effective transportation systems becomes increasingly vital for economic growth, social cohesion, and environmental sustainability [3, 4]. While numerous studies and indices, such as the Urban Mobility City Rankings by the Oliver Wyman Forum, provide valuable insights into the mobility readiness of certain cities, a significant gap remains in the inclusion of many other cities, particularly those that are rapidly developing or underrepresented in current analyses.

The UMRi [5, 6], developed by Olivier Wyman Forum in collaboration with the University of California, Berkeley, is a comprehensive tool for evaluating and ranking cities worldwide in terms of their preparedness to adopt and develop sustainable and innovative urban mobility systems. The UMRi assesses multiple dimensions of urban mobility, including infrastructure, innovation, efficiency, sustainability, and resilience.

Specifically, the UMRi evaluates the quality and extent of transportation infrastructure, such as roads, public transportation networks, and electric vehicle infrastructure. It also considers how cities are embracing and fostering new mobility technologies like autonomous vehicles, shared mobility services, and micromobility solutions. The index further measures the efficiency of transportation systems in managing the flow of people and goods while minimizing travel times. In terms of sustainability, the UMRi examines policies and practices aimed at reducing the environmental impact of urban transportation, such as the use of renewable energy sources and the promotion of less polluting modes of transport. Finally, the resilience of a city, or its ability to adapt to unforeseen challenges and changes, such as natural disasters or transportation crises, is also a critical factor in the UMRi.

This gap underscores the importance of expanding the scope of urban mobility studies to include a broader range of cities. By doing so, we can better understand the diverse challenges and opportunities these cities face in achieving sustainable mobility. The absence of many cities from these rankings means that their progress in terms of mobility and sustainable development remains largely unmeasured and unaddressed.

The dataset described in this article addresses this gap by enabling cities not currently included in major mobility rankings to assess their urban network features in relation to established indices. By analysing the structural aspects of their urban networks, these cities can gain insights into their mobility readiness and identify areas for improvement. This approach not only helps in benchmarking their progress against more prominent cities but also offers a pathway for these cities to contribute to the global discourse on sustainable urban development.

In essence, these data democratize access to critical urban mobility insights, ensuring that cities of all sizes and stages of development can engage in the pursuit of sustainable and efficient transportation systems.

DATA DESCRIPTION

This dataset provides a comprehensive collection of urban network structures and socio-demographic variables for 65 cities across different continents for the year 2023. The data were meticulously gathered and processed to analyse the relationship between network topology and UMRi. The dataset encompasses a wide range of metrics essential for analysing urban mobility networks. These metrics can be categorized into several key areas:

1. Mobility Readiness and Public Transit: Metrics such as the UMRi (both absolute and scored values), Sustainable Mobility (absolute and scored), and Public Transit (absolute and scored) reflect the readiness and performance of cities in terms of sustainable and efficient mobility.
2. Demographics and Economic Indicators: Population (in millions), Surface Area (km^2), Population Density (people/km^2), and GDP per capita ($) provide critical context regarding the socio-economic and spatial characteristics of the cities.
3. Graph and Network Characteristics: This category includes detailed graph-theoretical metrics such as the number of nodes, number of edges, average degree (), total and average edge length, average streets per node, intersection count, total and average street length, circuity, self-loop proportion, Average Betweenness Centrality (Normalized), Average Closeness Centrality (Normalized), Average Degree Centrality (Normalized), Graph Entropy, Graph Diameter, Graph Density, Clustering Coefficient, Average Path Length, Constraints, Assortativity Degree, Mean Degree, Reciprocity, and Diversity. These metrics are crucial for understanding the structural properties and the overall complexity of urban mobility networks.

Together, these variables provide a comprehensive view of urban mobility, blending socio-economic factors with advanced network analysis to offer deep insights into the functioning and efficiency of city transportation systems.

The primary objective of this dataset is to serve as a foundational resource for examining how urban network structures impact mobility efficiency and sustainability across various global cities.

Urban planners and researchers can leverage this dataset to explore complex urban mobility dynamics and devise strategies for enhancing transportation systems, not only in cities that rank high on mobility indexes but also in those that may not typically be included in such rankings. By offering a detailed analysis of network structures and sociodemographic variables, this dataset provides a robust foundation for classifying and improving urban areas, particularly in rapidly growing or underserved regions. The dataset is formatted to support seamless integration with statistical and network analysis tools, making it an invaluable resource for a wide range of urban planning and research applications.

The dataset is organized into a comprehensive directory structure that facilitates easy access to various types of data. Within the main directory, there is a subdirectory named “GRAPHS” which contains files for each city. For each city, there is a “.GRAPHML” file that describes the graph network structure of the urban network, alongside a “.NPY” file storing basic statistical metrics derived from the graph. Additionally, two ".CSV" files accompany each city's data: one capturing centrality measures of the graph (e.g., degree centrality, betweenness centrality), and another containing advanced structural metrics, such as constraints and assortativity degree, which are vital for understanding the network's resilience and connectivity patterns.

Beyond the "GRAPHS" subdirectory, the dataset also includes the original data extracted from the Oliver Wyman Forum, transcribed in the "CITIES-MOBILITY_DATA.ODS" file. Furthermore, there are two files in the "RESULTS_CITIES" directory: one with a ".CSV" extension and the other with a ".XLSX" extension. These files essentially contain the same data, providing a comprehensive overview of the final results of the graph analysis.

EXPERIMENTAL DESIGN, MATERIALS AND METHODS

4.1. Obtaining Networks

The urban network graphs for the cities featured in the fifth edition of the UMRi were obtained using a systematic approach leveraging Python libraries such as OSMNX and NUMPY. These tools facilitated the extraction and storage of the graphs and their basic statistics, which are vital for analysing urban mobility.

The process began by selecting the cities from the dataset that needed to be processed. For each selected city, the street network was extracted from OPENSTREETMAP, focusing on the drive network type—essentially, the streets accessible to vehicles. This network was then transformed into a graph structure that captures the city's connectivity in an easily analysed format.

In addition to generating the graph, basic statistics were calculated. These statistics include the number of intersections (nodes), streets (edges), total street length, and other fundamental metrics that provide insight into each city's transportation infrastructure.

To ensure transparency, reproducibility, and ease of reuse, all scripts, notebooks, and supplementary documentation used to obtain, process, and analyse the urban networks are openly provided in the accompanying Mendeley Data repository. The repository is organized into clearly structured directories containing (i) an executable Python script (*"Network.py"*) that automates the retrieval of street networks from OpenStreetMap via OSMnx and stores each city's network in *".graphml"* and *".npy"* formats along with the corresponding basic statistics, and (ii) a Jupyter notebook (*"Preparing_data_v2.ipynb"*) that applies NetworkX and igraph functions to compute the advanced graph-theoretical metrics described in this article, including centrality measures, clustering coefficients, graph entropy, and other topological indicators. The workflow is fully documented in a README file that provides detailed usage instructions, input/output descriptions, and environment requirements. The Mendeley Data repository includes an executable script (*"Network.py"*) for network extraction and a Jupyter notebook (*"Preparing_data_v2.ipynb"*) that computes all topological metrics reported in this article. Both resources contain commented code cells and examples of execution that illustrate the step-by-step process, from downloading networks to generating the final CSV and XLSX files. This ensures that users can reproduce the workflow directly, adapt it to other cities, or extend it with additional network metrics. The inclusion of runnable scripts goes beyond pseudocode, providing a transparent and verifiable foundation for replication.

All analyses were conducted in Python 3.11.4, employing OSMnx 1.3.0 for network extraction, NumPy 1.25.2 and Pandas 2.0.3 for data handling, NetworkX 3.2.1 and igraph 0.10.8 for graph analysis, and joblib 1.3.2 to parallelize computations on complex or large-scale networks. Default parameter settings were used unless explicitly stated; for example, Louvain community detection was executed with resolution parameter γ = 1.0. In addition, the repository provides sample input data (*"cities-mobility_data.ods"*) and example outputs for selected cities, enabling users to verify the pipeline step by step. This level of documentation guarantees that researchers, policymakers, and practitioners can replicate the dataset, adapt the workflow to additional cities, or extend the analysis to other urban contexts with minimal effort. By making the entire computational process openly accessible and version-controlled, this dataset not only supports reproducibility but also facilitates transparency and long-term reusability in urban mobility research.

To complement the dataset and facilitate its reuse, we provide a comprehensive data dictionary that defines each variable, its unit of measurement, and the method of calculation or source. This dictionary is included both in the manuscript (Table 1) and in the accompanying Mendeley Data repository as a separate file. The dictionary ensures that users can clearly understand and replicate each indicator, whether it is a sociodemographic measure (e.g., population, GDP per capita) or a graph-theoretical metric (e.g., clustering coefficient, graph entropy).

Table 1. Data Dictionary for the Urban Mobility Dataset

Variable	**Description**	**Units**	**Calculation / Source**
Pop	Population	millions of inhabitants	Source: UMRi dataset (2023)
SA	Surface area	km²	Source: UMRi dataset (2023)
PD	Population density	people/km²	Pop / SA
GDP	Gross Domestic Product per capita	USD	Source: UMRi dataset (2023)
UMRi	Urban Mobility Readiness Index (score)	dimensionless	From Oliver Wyman Forum, UMRi dataset
SMi	Sustainable Mobility Index (score)	dimensionless	From Oliver Wyman Forum, UMRi dataset
PTi	Public Transit Index (score)	dimensionless	From Oliver Wyman Forum, UMRi dataset
K Avg	Average node degree	dimensionless	Mean degree of nodes (NetworkX/igraph)
m	Number of edges	count ($\times 10^4$)	Extracted from OSMnx graph
IC	Intersection count (nodes)	count ($\times 10^4$)	OSMnx basic_stats
SLT	Total street length	meters ($\times 10^7$)	Sum of edge lengths (OSMnx)
SSC	Street segment count	count ($\times 10^5$)	OSMnx basic_stats
ELT	Total edge length	meters ($\times 10^7$)	OSMnx basic_stats
ELa	Average edge length	meters	Mean of all edge lengths
SNa	Average streets per node	dimensionless	OSMnx basic_stats
SLa	Average street length	meters	Mean length of street segments
Ca	Circuity average	dimensionless	Ratio of edge length to straight-line distance
SLP	Self-loop proportion	dimensionless ($\times 10^{-3}$)	Proportion of edges forming loops

BCa	Average betweenness centrality (normalized)	dimensionless ($\times 10^{-3}$)	igraph betweenness centrality
CCa	Average closeness centrality (normalized)	dimensionless	igraph closeness centrality
DCa	Average degree centrality (normalized)	dimensionless ($\times 10^{-4}$)	igraph degree centrality
GE	Graph entropy	dimensionless	Shannon entropy of degree distribution
GDi	Graph diameter	edges (steps)	Longest shortest-path length (igraph)
GDe	Graph density	Dimensionless ($\times 10^{-4}$)	Ratio of edges to possible edges
CC	Clustering coefficient	dimensionless	Average clustering coefficient (igraph)
PLa	Average path length	edges (steps)	Mean shortest path length (igraph)
Const	Constraints	dimensionless	igraph structural constraint
AD	Assortativity degree	dimensionless	igraph assortativity measure
MD	Mean degree	dimensionless	igraph mean degree
Rec	Reciprocity	dimensionless	Ratio of reciprocated edges (igraph)
Div	Diversity	dimensionless	igraph diversity index

4.2. Advanced Metrics Calculation

To obtain more advanced metrics, which are crucial for deeper network analysis [7] and Topological Data Analysis [8, 9], additional computations were performed using specialized tools. These metrics include measures of centrality (e.g., betweenness, closeness, degree, and Katz centrality), graph entropy, graph diameter, graph density, clustering coefficient, and average path length. These were computed using iGraph in Python, given its robust capabilities for handling complex network analysis efficiently.

Due to the complexity of some urban networks, particularly in cities with highly dense or large-scale street networks, the calculation of these advanced metrics required parallelization. This approach significantly reduced processing time, which could otherwise extend to several hours for more intricate graphs.

Figure 1 provides a comparative visual analysis of three cities, showcasing variations in their urban mobility network structures. Each plot displays the community structure detected through the Louvain algorithm, which is effective in identifying significant clusters in large-scale and complex networks [10]. Key clusters are highlighted based on the betweenness centrality of their nodes, providing insight into the importance of different communities within the urban mobility networks. The visualizations reveal distinct patterns in connectivity, community importance, and clustering coefficient, reflecting how variations in urban planning and infrastructure impact each city's mobility readiness. This analysis emphasizes the role of influential communities in shaping mobility efficiency, as represented by their UMRi scores, and highlights the structural adaptability of different urban environments.

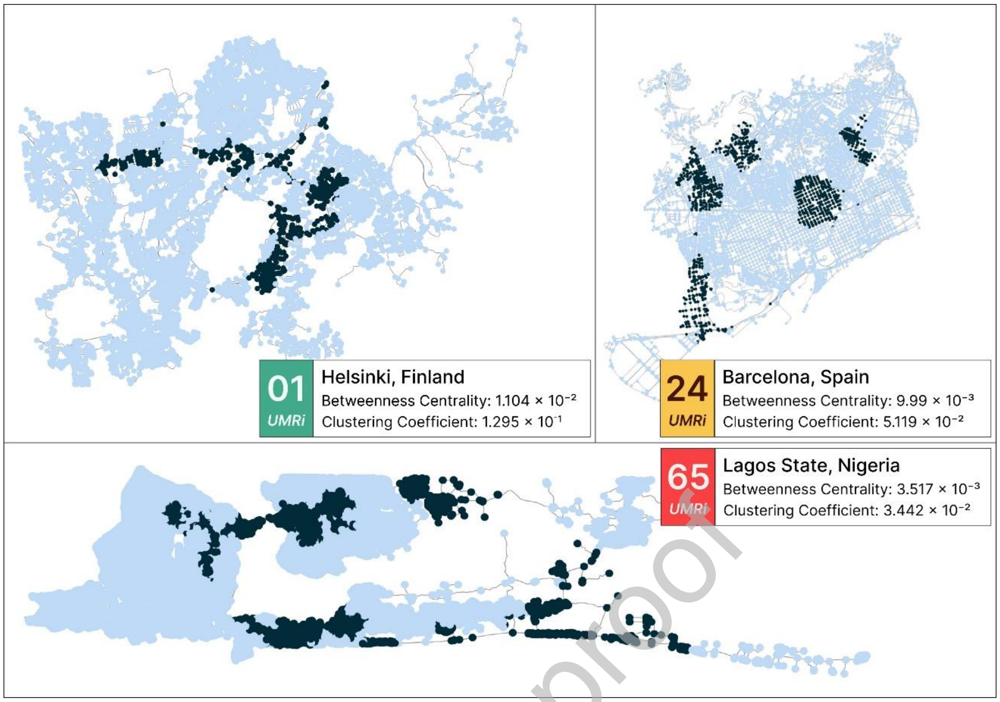


Figure 1: Comparative analysis of urban mobility networks in three cities based on their UMRi scores. The community structures were identified using the Louvain algorithm, and key nodes were highlighted based on the betweenness centrality of the most important communities, determined by their normalized betweenness centrality scores. Helsinki and Barcelona, with higher UMRi scores, display community clusters with average betweenness centrality values of 0.011 and 0.009, respectively, indicating a higher level of node importance. In contrast, Lagos, with a lower UMRi score, has a significantly lower average betweenness centrality of 0.003, suggesting a less central role of key nodes within the network. Additionally, the clustering coefficient shows greater cohesion in Helsinki (0.01259) and Lagos (0.0344) compared to Barcelona (0.0511), further reflecting structural differences in urban mobility readiness across regions.

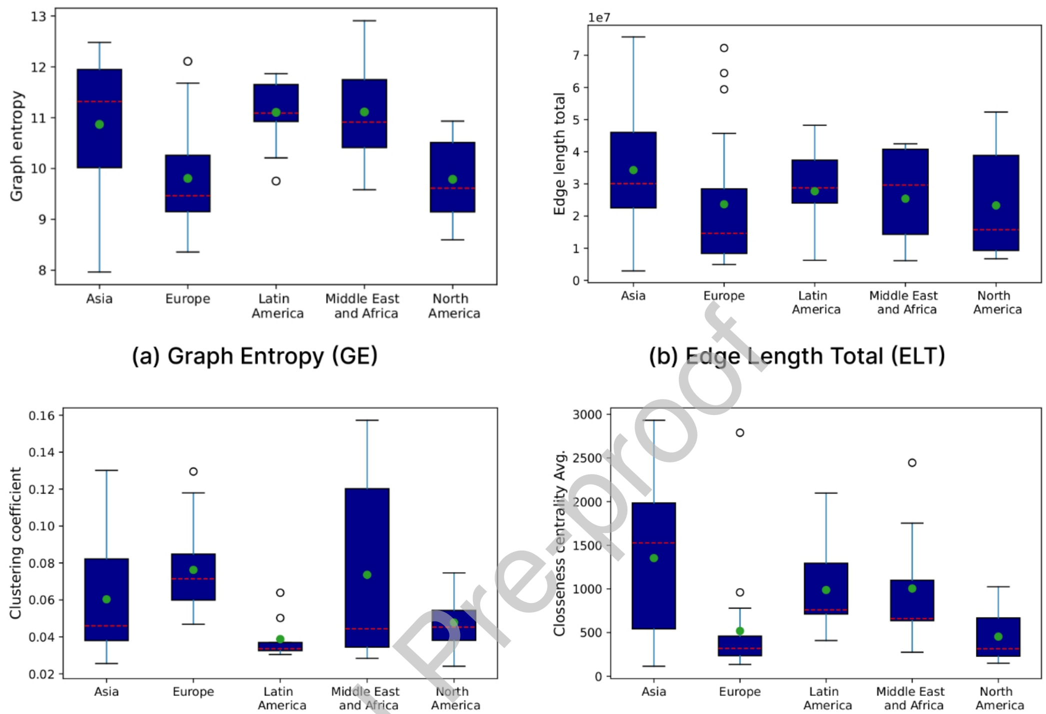
Figure 2 presents a comparative analysis of various graph metrics across different regions or continents. These metrics include Graph Entropy (GE), Edge Length Total (ELT), Clustering Coefficient (CC), and Closeness Centrality Average (CCa), each of which provides valuable insights into the structural characteristics of urban mobility networks.

Subfigure 2(a) illustrates the distribution of Graph Entropy (GE) across regions, showing significant variation between continents. Europe and Asia exhibit higher entropy values, indicative of more complex network structures. In contrast, regions like Latin America and the Middle East and Africa tend to have lower entropy values, suggesting less complex urban networks.

Subfigure 2(b) compares the total edge length (ELT) of urban networks by region. North America shows the highest total edge length, which could be attributed to the extensive infrastructure in larger cities. Conversely, regions such as Latin America and Asia demonstrate more modest edge lengths, reflecting differences in urban planning and network development.

Subfigure 2(c) presents the Clustering Coefficient (CC) distribution, which varies considerably across regions. Europe and North America display higher clustering coefficients, indicating a higher tendency of nodes to cluster together, suggesting a more interconnected and potentially resilient urban network. Lower clustering coefficients are observed in Latin America and Asia.

Finally, subfigure 2(d) shows the average Closeness Centrality (CCa) by region, where Europe and Asia exhibit higher averages. This suggests that nodes in these networks are more centrally located, potentially leading to greater efficiency in mobility within these cities. In contrast, regions like Latin America show lower values, reflecting more dispersed network structures with less direct connectivity between nodes.



(a) Graph Entropy (GE)

(b) Edge Length Total (ELT)

(c) Clustering Coefficient (CC)

(d) Closeness Centrality Avg (CCa)

Figure 2: Comparative boxplots for graph metrics across different regions: (a) Graph Entropy (GE), (b) Edge Length Total (ELT), (c) Clustering Coefficient (CC), and (d) Closeness Centrality Average (CCa). Horizontal red lines represent medians while green dots represent means.

Additionally, Table 2 presents a summary of the key network metrics calculated across the studied urban mobility networks. These statistics include measures such as the mean, standard deviation, minimum, and maximum values for various metrics, providing a comprehensive overview of the structural properties of the networks. Notably, the metrics range from basic properties like the average node degree (K Avg) to more complex characteristics such as graph entropy (GE) and closeness centrality average (CCa). This summary is crucial for understanding the variation and central tendencies in the dataset, which reflect the diverse nature of urban mobility networks in different regions.

Table 2: Summary of Network Metrics. The following abbreviations are used in the table: UMR Index (Scr) as UMRi, Sustainable Mobility (Scr) as SMi, Public Transit (Scr) as PTi, Population (millions) as Pop, Surface Area (km2) as SA, Population Density (people/km2) as PD, GDP per capita ($) as GDP, average node degree as K Avg, total edge length as ELT, average edge length as ELa, average streets per node as SNa, intersection count as IC, total street length as SLT, street segment count as SSC, average street length as SLa, circuity average as Ca, self-loop proportion as SLP, Average Betweenness Centrality (Normalized) as BCa, Average Closeness Centrality (Normalized) as CCa, Average Degree Centrality (Normalized) as DCa, Graph Entropy as GE, Graph Diameter as

GDi, Graph Density as GDe, Clustering Coefficient as CC, Average Path Length as PLa, Constraints as Const, Assortativity Degree as AD, Mean Degree as MD, Reciprocity as Rec, and Diversity as Div.

Metric	mean	std	min	25%	50%	75%	max
$m \times 10^4$	45.31	31.76	5.33	18.73	40.73	64.38	135.73
$ELT \times 10^7$	2.68	1.90	0.29	1.01	2.43	4.01	7.57
$IC \times 10^4$	14.34	9.87	1.16	5.88	12.84	20.68	41.24
$STL \times 10^7$	1.49	1.05	0.16	0.56	1.37	2.22	4.43
$SSC \times 10^5$	2.48	1.70	0.30	1.00	2.21	3.48	7.11
$GDe \times 10^{-4}$	1.19	1.32	0.05	0.25	0.69	1.77	6.12
$SLP \times 10^{-3}$	2.64	1.47	0.57	1.76	2.18	3.30	7.76
$BCa \times 10^{-3}$	5.47	4.44	0.79	2.16	3.96	7.60	21.86
$DCa \times 10^{-4}$	2.38	2.63	0.09	0.50	1.38	3.55	12.24
K Avg	5.29	0.42	4.66	4.95	5.27	5.55	6.70
ELa	63.68	26.30	29.39	45.46	59.67	73.13	173.58
SNa	2.93	0.19	2.46	2.81	2.92	3.04	3.46
SLa	63.84	25.48	29.82	46.22	61.38	72.91	180.24
Ca	1.07	0.03	1.02	1.05	1.06	1.08	1.16
CCa	832.85	713.61	113.84	291.99	618.23	1118.48	2931.48
GE	10.42	1.18	7.97	9.42	10.38	11.38	12.91
GDi	216.19	94.72	76.00	142.50	192.00	273.25	428.00
CC	0.06	0.03	0.02	0.04	0.05	0.08	0.16
PLa	74.95	28.29	34.68	52.22	68.09	87.77	151.34
Const	0.45	0.05	0.33	0.41	0.44	0.48	0.60
AD	0.26	0.13	0.03	0.16	0.24	0.36	0.59
MD	4.76	0.52	3.72	4.42	4.84	5.16	5.74
Rec	0.76	0.16	0.31	0.71	0.81	0.87	0.95
Div	2.90	0.69	1.27	2.54	3.01	3.39	4.29

The analysis of the Spearman correlations between the Sustainable Mobility Index (SMi) and various sociodemographic and topological variables has revealed several significant relationships that offer valuable insights into the factors driving sustainable urban mobility. One of the strongest positive correlations was observed between SMi and GDP per capita (□□ = 0.77, □□ < 0.01), indicating that higher economic prosperity is strongly associated with better sustainable mobility practices. This finding aligns with the understanding that wealthier cities are more capable of investing in advanced infrastructure and sustainable mobility solutions.

On the other hand, SMi exhibited a strong negative correlation with Population Density (□□ = −0.39, □□ < 0.01) and the average street length (SLa) (□□ = −0.61, □□ < 0.01). These results suggest that densely populated areas and longer street segments are potentially less conducive to sustainable mobility, possibly due to congestion and the inefficiency of urban sprawl. Additionally, moderate positive correlations were found between SMi and Average Degree Centrality (DCa) (□□ = 0.44, □□ < 0.01), highlighting the importance of network connectivity in enhancing urban mobility. These correlations are critical as they underscore the need to focus on economic growth, efficient urban planning, and robust network structures to improve sustainability in urban mobility.

The main objective of this result is to identify the key socio-demographic and topological factors that influence sustainable mobility across cities. By understanding these relationships, urban planners and policymakers can better design strategies that target the most impactful variables, thereby promoting more sustainable and efficient urban mobility systems. Figure 3 visually illustrates these correlations, providing a clearer understanding of the strength and direction of the relationships between SMi and the variables studied. This plot allows one to quickly identify which factors have the most significant impact on sustainable mobility.

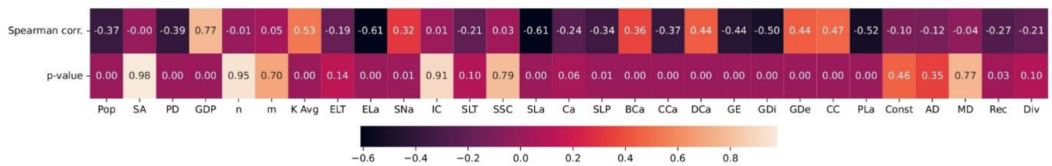


Figure 3: Spearman correlation matrix showing the relationships between the Sustainable Mobility Index and various socio-demographic and topological variables. Significant correlations are highlighted.

4.3. Quality Assurance

To guarantee the reliability of the dataset, several quality assurance procedures were implemented during data acquisition and processing. First, all street networks were retrieved from OpenStreetMap using OSMnx and automatically checked for completeness, ensuring that the extracted graphs contained valid nodes, edges, and attributes. The executable script (*"Network.py"*) was designed to record processing time, detect errors, and flag incomplete downloads, allowing us to repeat the extraction when necessary. Second, the computed statistics (e.g., number of nodes, total edge length) were cross-validated against OSMnx's basic_stats outputs and compared with descriptive indicators reported in the Urban Mobility Readiness Index to detect inconsistencies. Third, advanced metrics computed with igraph were validated by performing internal consistency checks (e.g., verifying that the sum of normalized centrality values was within expected ranges, ensuring graph diameter was finite, and confirming clustering coefficients remained between 0 and 1). Additionally, correlation analyses were performed among topological and sociodemographic variables to identify outliers or anomalous values, which were manually inspected. All validation steps and intermediate results are documented in the Mendeley Data repository, where users can find reproducible scripts and logs. Together, these measures ensure the accuracy, consistency, and robustness of the dataset, making it reliable for further research and urban mobility analysis.

LIMITATIONS

During the process of obtaining the urban network graphs for various cities, it is important to be mindful of potential challenges related to the extraction process using OSMnx from OpenStreetMap. Specifically, the format "City, Country" might not always be sufficient to accurately obtain the desired graph. In some cases, a more detailed and precise reference may be required to ensure that the correct urban area is captured. This is particularly relevant for cities with multiple regions or complex administrative boundaries, where a broader or more specific query might be necessary to avoid incomplete or inaccurate data.

Furthermore, cities with a higher number of nodes or more intricate network structures present additional challenges. The complexity of these graphs demands considerable computational power, both for their extraction and subsequent analysis. As the complexity of the graph increases, so too does the dependency on robust computational resources to handle the processing and analysis efficiently. This limitation highlights the importance of ensuring access to adequate computational infrastructure when working with large-scale or complex urban network data.

While the dataset offers valuable insights into the structural properties of urban mobility networks, some limitations should be acknowledged. First, as the networks were extracted from OpenStreetMap (OSM), the accuracy of the dataset is inherently dependent on the quality and completeness of OSM data [11]. In certain regions, particularly in rapidly developing or underrepresented areas, street networks may be incomplete, outdated, or unevenly detailed compared to cities in Europe or North America, introducing potential biases in the results. Second, OSM contributions are crowd-sourced, which may lead to inconsistencies in the tagging of road types, intersections, or boundaries. These inconsistencies can affect the derived graph metrics (e.g., node degree, clustering coefficient) and may vary across regions. Third, temporal considerations must be taken into account: the networks reflect a snapshot of OSM as of 2023 and therefore do not capture subsequent changes, expansions, or policy-driven transformations in mobility infrastructure. Researchers aiming to replicate or extend the analysis in the future should be aware that OSM data are continuously updated and that longitudinal comparisons would require careful version control of OSM extracts. Finally, although computational checks and validation steps were applied to ensure internal consistency, these limitations highlight the importance of interpreting the dataset as an approximation of urban mobility structures rather than a definitive or exhaustive representation.

ETHICS STATEMENT

The authors confirm that they have read and adhere to the ethical requirements for publication in Data in Brief. This work does not involve human subjects, animal experiments, or any data collected from social media platforms.

CRediT AUTHOR STATEMENT

D. D. Herrera-Acevedo: Conceptualization of this study, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Visualization, Data Curation, Writing - Original Draft. **D. Sierra-Porta**: Conceptualization of this study, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision, Project administration.

ACKNOWLEDGEMENTS

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

DECLARATION OF COMPETING INTERESTS

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

[1] Simon Elias Bibri, John Krogstie, and Mattias Kärrholm. Compact city planning and development: Emerging practices and strategies for achieving the goals of sustainability. Developments in the built environment, 4:100021, 2020. https://doi.org/10.1016/j.dibe.2020.100021.

[2] Abdulaziz I Almulhim and Patrick Brandful Cobbinah. Can rapid urbanization be sustainable? the case of saudi arabian cities. Habitat International, 139:102884, 2023. https://doi.org/10.1016/j.habitatint.2023.102884.

[3] Kim Carlotta von Schönfeld and António Ferreira. Urban planning and european innovation policy: Achieving sustainability, social inclusion, and economic growth? Sustainability, 13(3):1137, 2021. https://doi.org/10.3390/su13031137.

[4] Keshav Bhattarai, Dennis Conway, Keshav Bhattarai, and Dennis Conway. Urban growth. Contemporary Environmental Problems in Nepal: Geographic Perspectives, pages 201–334, ISBN: 978-3-030-50168-6. 2021. https://doi.org/10.1007/978-3-030-50168-6.

[5] Cayetano Medina-Molina, Noemí Pérez-Macías, and María Coronado-Vaca. Searching for complexity. application of the set-theory to the analysis of urban mobility readiness index. Discover Sustainability, 5(1):10, 2024. https://doi.org/10.1007/s43621-024-00187-5.

[6] Cayetano Medina-Molina and Maria de la Sierra Rey-Tienda. The transition towards the implementation of sustainable mobility. Looking for generalization of sustainable mobility in different territories by the application of qca. Sustainable Technology and Entrepreneurship, 1(2):100015, 2022. https://doi.org/10.1016/j.stae.2022.100015.

[7] Johannes F Knabe and Johannes F Knabe. Topological network analysis. Computational Genetic Regulatory Networks: Evolvable, Selforganizing Systems, pages 71–81, ISBN: 978-3-642-30295-4, 2013. https://doi.org/10.1007/978-3-642-30296-1_5.

[8] Elizabeth Munch. A user's guide to topological data analysis. Journal of Learning Analytics, 4(2):47–61, 2017. https://doi.org/10.18608/jla.2017.42.6.

[9] Larry Wasserman. Topological data analysis. Annual Review of Statistics and Its Application, 5(1):501–532, 2018. https://doi.org/10.1146/annurev-statistics-031017-100045.

[10] Sayan Ghosh, Mahantesh Halappanavar, Antonino Tumeo, Ananth Kalyanaraman, Hao Lu, Daniel Chavarrià-Miranda, Arif Khan, and Assefaw Gebremedhin. Distributed louvain algorithm for graph community detection. In 2018 IEEE International Parallel and Distributed Processing Symposium (IPDPS), pages 885–895, 2018. https://doi.org/10.1109/IPDPS.2018.00098.

[11] Herrera-Acevedo, D. D., & Sierra-Porta, D. (2025). Network structure and urban mobility sustainability: A topological analysis of cities from the urban mobility readiness index. Sustainable Cities and Society, 119, 106076. https://doi.org/10.1016/j.scs.2024.106076.