

Milani, Rudy; Nistor, Marian Sorin; Moll, Maximilian; Pickl, Stefan

Article — Published Version

On the Correlation and Predictability of Topological Measures in Transportation Networks

Operations Research Forum

Provided in Cooperation with:

Springer Nature

Suggested Citation: Milani, Rudy; Nistor, Marian Sorin; Moll, Maximilian; Pickl, Stefan (2025) : On the Correlation and Predictability of Topological Measures in Transportation Networks, Operations Research Forum, ISSN 2662-2556, Springer International Publishing, Cham, Vol. 6, Iss. 2, <https://doi.org/10.1007/s43069-025-00471-8>

This Version is available at:

<https://hdl.handle.net/10419/330800>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



On the Correlation and Predictability of Topological Measures in Transportation Networks

Rudy Milani¹ · Marian Sorin Nistor¹ · Maximilian Moll¹ · Stefan Pickl¹

Received: 16 December 2024 / Accepted: 6 May 2025 / Published online: 5 June 2025

© The Author(s) 2025

Abstract

The computation of topological measures in large-scale complex networks, such as those found in transportation systems, is often a resource-intensive process. These measures, however, are critical for a comprehensive understanding of network structures and for optimizing their design. A key challenge lies in selecting the appropriate metrics that encapsulate the essential information of the network, thereby reducing the computational burden. Traditional methods involve identifying correlations between various topological measures to infer missing data. In this paper, we introduce an enhanced analytical framework comprising three stages aimed at selecting a subset of metrics to efficiently summarize network characteristics and predict measures that are costly to compute. The methodology involves: a correlation analysis of topological metrics; a principal component analysis to reduce dimensionality and highlight the essential features; and the application of SHAP and recursive feature elimination to assess the predictive significance of each metric. We demonstrate the utility of this approach using metro and road networks from 46 cities in the EU/EEA region, yielding promising results in identifying relationships between metrics and predicting missing data.

Keywords Network analysis · Metro networks · Road networks · Correlation analysis · Principal component analysis · SHAP · Recursive feature elimination

✉ Rudy Milani
rudy.milani@unibw.de

Marian Sorin Nistor
sorin.nistor@unibw.de

Maximilian Moll
maximilian.moll@unibw.de

Stefan Pickl
stefan.pickl@unibw.de

¹ Faculty of Computer Science, Universität der Bundeswehr München, Werner-Heisenberg-Weg 39, 85577 Neubiberg, Germany

1 Introduction

A wide set of real-world systems, ranging from neurobiology to statistical physics [26], has been successfully modeled using complex network theory [6]. This approach allows for an in-depth analysis of the structural properties underlying these systems, which is crucial for understanding their dynamics and optimizing their functionality. To this end, various topological measures and indices have been employed to capture critical network characteristics [12]. However, the computational complexity of these measures becomes a significant challenge, especially when dealing with large-scale networks. Consequently, selecting a subset of informative metrics that can adequately represent the network while minimizing computational costs is essential.

Several studies have explored the relationships between topological measures to address this challenge, often focusing on specific domains such as internet AS- and router-level networks [14, 18], brain networks [16, 17], and other real-world systems [8]. For example, [19] proposed a framework for selecting a subset of topological measures that can sufficiently describe network properties. Subsequently, [12] advanced this methodology by calculating correlation coefficients and principal components across various network types. More recently, further authors took this pipeline and carried out experiments on medium/large-size datasets consisting of both model-generated graphs and real networks [2, 13]. These works have laid the foundation for identifying meaningful subsets of metrics that capture the essence of a network's topology.

However, the focus of the methodologies in the literature is restricted to the determination of the statistical relationship between the topological metrics. Thus, no insights on the possibility of predicting specific high-cost measures from a subset of selected metrics are produced. Nevertheless, in order to reduce the complexity burden of computing most of the topological metrics it is necessary to determine a standardized pipeline for the forecasting procedure.

Hence, this paper presents an extended analytical pipeline designed to identify key relationships between topological measures and evaluate those that summarize the most information of the network. We leverage correlation analysis, principal component analysis (PCA), and machine learning (ML) prediction techniques combined with SHAP and recursive feature elimination (RFE) to enhance the efficiency of network characterization. Our analysis spans transportation networks from 46 EU/EEA cities, with the aim of uncovering new insights into the relationships between topological measures and optimizing the prediction of costly metrics.

The primary contributions of this study are as follows:

- **Enhancement of an existing analytical pipeline:** We refined a pre-existing methodology [12] for analyzing and selecting relevant features that effectively summarize the fundamental characteristics of complex networks. This improvement involved extending the PCA framework by incorporating factor loadings and introducing an additional phase dedicated to selecting the optimal subset of features to enhance the prediction accuracy of the un-computed metrics.
- **Comprehensive analysis of topological metrics:** We conducted an in-depth examination of the topological metrics and their correlations within metro and road

networks across various European cities, offering insights into the structural properties of these transportation networks.

- Identification of high-accuracy predictive subsets: We uncovered a subset of correlated features that are instrumental in achieving high-accuracy predictions in the context of transportation networks.

The remainder of this paper is organized as follows. In Section 2, we introduce the topological measures and networks used in the analysis. In Section 3, we present the main contribution of the paper: the extended methodology proposed. In Section 4, we first apply the new approach to metro networks, and then to road networks. In this part, we gather all the results of the corresponding steps together with a brief discussion. After reporting all the outcomes for both types of transportation networks, we focus on comparing the results obtained in Section 5. Finally, we summarize the paper and draw possible directions for future development.

2 Background

In this section, we first define the topological measures that are studied throughout this paper in Section 2.1. Then, we introduce the networks investigated, and the procedure that led us to obtain this data in Section 2.2.

2.1 Topological Measures

We mathematically define a real-world system as a graph $G = (V, E)$ representing a finite and connected network with the set of vertices $V = V(G)$, and the set of edges $E = E(G)$. Furthermore, we assume the network to be connected and without self-loops. In the sequel, we present the set of topological measures considered in this paper, along with explanations about their relevance in characterizing the network topology.

The first parameter introduced is the most basic one: the network density (ND) $ND(G) = \frac{2|E|}{|V|(|V|-1)}$, which is the result of the ratio between the actual number of edges and the maximum possible ones. A high ratio denotes a dense network, whereas a low ratio denotes a sparse network [28]. The degree $D_G(v)$, which counts the number of adjacent edges for the vertex v in G , is considered an index of the connectivity strength of a vertex in the network [28]. However, we need topological measures that can describe the whole network. For this purpose, we consider the average degree (AD) $AD(G) = \frac{\sum_{v \in V} D_G(v)}{|V|}$.

Another valuable feature for the transportation networks is the path length $d_G(v, w)$ between a pair of vertices $v, w \in G$, which is the length of any shortest path as a number of edges [27]. From this, we can derive the average path length (AVP) of a network G as $AVP(G) = \frac{\sum_{v \in V} \sum_{w \in V \setminus \{v\}} d_G(v, w)}{|V|(|V|-1)}$ [27] which is the average of the shortest paths of all possible pairs of vertices in a network. Closely related to the path length is the eccentricity of a vertex $\epsilon(v, G) = \max_{w \in V \setminus \{v\}} d_G(v, w)$, which represents the distance of the longest shortest path between that vertex and any other vertex in the network [10]. For our scopes, we use the average eccentricity (VE) of a network G defined as $VE(G) = \frac{\sum_{v \in V} \epsilon(v, G)}{|V|}$.

We also consider the assortativity parameters which provide more insight into the correlation between vertices. Commonly studied are the average nearest neighbor degree [1, 24] and the degree assortativity [20, 21]. Average nearest neighbor degree $nn(k, G)$ of vertices of degree k in graph G is defined as $nn(k, G) = \sum_{l=1}^{|V|-1} l P_G(l|k)$ where $P_G(l|k)$ is the conditional probability that a vertex of degree k is linked with a vertex of degree l in graph G . This aspect captures the dependencies between degrees of neighboring vertices in a network [24]. From this formulation, we can then compute the average nearest neighbor degree (ANN) of a graph G as $ANN(G) = \frac{\sum_{k=1}^{|V|-1} nn(k, G)}{|V|}$. The degree assortativity (DA) $DA(G)$ of a network G is Pearson's correlation coefficient of the degrees between a pair of connected vertices and is defined as $DA(G) = \frac{|E|^{-1} \sum_{e \in E} j_e k_e - [(2|E|)^{-1} \sum_{e \in E} (j_e + k_e)]^2}{(2|E|)^{-1} \sum_{e \in E} (j_e^2 + k_e^2) - [(2|E|)^{-1} \sum_{e \in E} (j_e + k_e)]^2}$, where j_e, k_e are the degrees of the vertices connected with the edge e [20]. The DA identifies if the vertices with a similar degree tend to connect more often (assortativity mixing, when $0 < DA(G) \leq 1$), tend to mostly not connect (disassortative $-1 \leq DA(G) < 0$), or tend to be neutral (non-assortativity $DA(G) = 0$).

The clustering coefficient (CC) $c(v, G)$ of a vertex captures how close its neighbors are to a complete network [29]. Given three vertices $v_i, v_j, v_k \in V$ having v_i connected with v_j and with v_k , it represents the likeliness that v_j is also connected with v_k . If v_j is also connected with v_k then the three vertices form a triangle — complete subgraph or closed triple, and the number of triangles in G that includes vertex v is denoted by $\delta(v, G)$. Otherwise, they form a triple — a path of length two from v_j to v_k with v_i as center vertex, and the number of triples in G that includes vertex v is denoted by $\tau(v, G)$. Hence the CC of a vertex v is defined as $c(v, G) = \frac{\delta(v, G)}{\tau(v, G)}$, where $\delta(v, G) = |\{(v_i, v_j, v_k) \in V^3 | \{(v_i, v_j), (v_i, v_k), (v_j, v_k)\} \subseteq E\}|$ and $\tau(v, G) = |\{(v_i, v_j, v_k) \in V^3 | \{(v_i, v_j), (v_j, v_k)\} \subseteq E, (v_i, v_k) \notin E\}|$. Therefore, we generalize this measure for a network G by considering the average over all the vertices: $CC(G) = \frac{\sum_{v \in V} c(v, G)}{|V|}$.

Remaining on the topic of clustering, we now introduce the rich-club coefficient $RC_k(G)$ which is a structural parameter of a network that captures the tendency of hubs of vertices to connect [5]. The vertices with a high degree also called rich vertices, have a higher likelihood to interconnect into subgraphs (clubs), when compared with low-degree vertices [5]. This measure is calculated using the following formula: $RC_k(G) = \frac{2|E|_{>k}}{|V|_{>k}(|V|_{>k}-1)}$, where $|E|_{>k}$ represents the number of edges between the vertices of degree greater than or equal to k , and $|V|_{>k}$ is the number of vertices with a degree greater than or equal to k .

Moving to centrality measures we focus on the vertex betweenness centrality $BV(v, G)$ [11], or simply vertex betweenness (BV). This measure is a centrality parameter that determines the occurrence of a vertex on the shortest paths between any pair of vertices. In our study, we also consider the edge betweenness centrality $BE(e, G)$ [11], or simply edge betweenness (BE), which is the equivalent of $BV(v, G)$ for edges. Mathematically, these parameters can be expressed as $BV(v, G) = \frac{\sum_{s \neq v \neq t} \frac{\sigma_{st, G}(v)}{\sigma_{st, G}}}{(|V|-1)(|V|-2)}$ and $BE(e, G) = \frac{\sum_{s \neq t} \frac{\sigma_{st, G}(e)}{\sigma_{st, G}}}{(|V|-1)(|V|-2)}$ where $\sigma_{st, G}(\cdot)$ is the number of the shortest path (in G) between the vertices s and t that includes the input vertex or edge (depending on the measure used) and $\sigma_{st, G}$ is the number of all the shortest

path (in G) between the vertices s and t . As mentioned before, we can then define the corresponding general metrics for the network G just by considering the average.

Related to the connectivity of a network, we have the algebraic connectivity (AC) $AC(G)$ which represents the second smallest eigenvalue of the Laplacian matrix of G [7]. This parameter is useful for measuring how difficult it is to cut a graph into independent components (not connected anymore to the remaining part of the graph). The last measure that we used in this work is the average vertex coreness $AVC(v, G)$, which gives an insight into the position of the vertex v and indicates its influence on spreading dynamics [3]. In fact, for larger coreness values we have vertices that are more central in the network. The average vertex coreness is defined as the maximum k such that this vertex is still in the k -core but removed in the $(k + 1)$ -core. Finally, in order to have a measure related to the full network, we compute the average and define $AVC(G)$ as the average coreness (AVC) of G .

2.2 Transportation Networks Data

This paper analyzes transportation networks from 46 European cities within the EU/EEA region, specifically focusing on metro and road networks. A comprehensive list of the cities included in the analysis is provided in [Appendix](#). An example of both network types for a single city, Munich, is shown in Fig. 1, where the metro and road networks are depicted graphically. In these graphs, vertices represent either metro stations or street junctions, while edges denote the direct connections between these points. The following paragraphs provide further details on the data collection and processing methodologies essential for the analysis.

The metro network data were collected individually from the official websites of metro operators. In cases where the available information was insufficient, external sources were used. Each of the 46 metro networks is represented by an individual table that lists metro stations per line in each city. A custom algorithm was then developed to generate adjacency matrices based on these tables. In the case of cities with more than one metro system, such as London (which operates both the London Underground

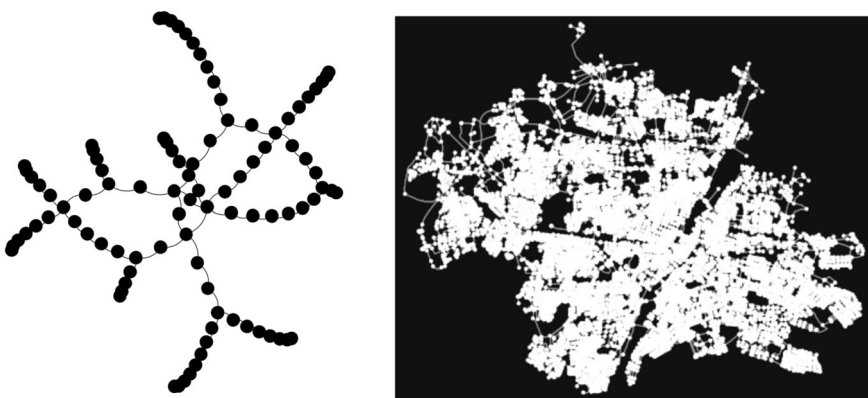


Fig. 1 Graph visualization of Munich metro on the left and of road system on the right

and the Docklands Light Railway), the networks were combined, with shared stations merged into one dataset. This method was applied to other cases, such as the Amsterdam Metro, where light rail line 51 was treated as part of the regular metro network.

Although all networks are required to be connected, exceptions were made for cities like Berlin and Naples, where some metro lines operate as isolated components. In these cases, only the largest connected component was considered for the analysis. A similar approach was taken for networks with walking connections, such as Valencia Metro's line 10, where transfers occur by walking between nearby stations. These connections were treated as regular edges, adding new links between previously unconnected components.

The process of obtaining road network data was more streamlined than for metro networks. The OpenStreetMap package for Python [4] was used to extract road data for the selected cities. However, preprocessing was required before the data could be analyzed. Specifically, we transformed the directed road networks into undirected graphs, eliminating self-loops caused by circular roads. The structure of these graphs was simplified by retaining only the vertices that represent junctions of multiple streets.

3 Methodology

In this section, we propose an extended (from [12]) analytical pipeline designed to analyze the relationships between topological measures in complex networks. This methodology is particularly well-suited for large, complex networks where the computational cost of calculating all topological measures is prohibitive. By understanding the relationships among these measures, it becomes possible to limit the number of metrics that need to be computed, thereby reducing computational demands while maintaining a comprehensive understanding of the network structure.

The examination of feature interdependencies encompasses various methodologies, ranging from elementary statistical correlation to sophisticated ML assessments. Therefore, our comprehensive approach incorporates three distinct analytical phases: correlation analysis, PCA, and training of the predictive models and feature selection.

The preliminary step in the proposed methodology consists of the computation of the topological measures discussed in Section 2.1. These parameters serve as the cornerstone upon which all the subsequent analysis are built.

Moving the focus on the actual pipeline, the first stage, represented in the yellow boxes of Fig. 2, involves a statistical analysis to examine the correlations between the topological measures. By calculating pairwise correlations between metrics, we identify dependencies and potential redundancies. This analysis enables the identification of clusters of correlated measures, facilitating dimensionality reduction while preserving the network's essential structural characteristics [22]. As observed in previous studies [12], the relationships between topological measures may vary across different contexts, making it crucial to reassess these relationships for each application.

In the second step, shown in the orange boxes of Fig. 2, PCA is employed to reduce the dimensionality of the dataset by transforming the correlated measures into a smaller set of uncorrelated variables, known as principal components [23]. This process captures the most significant patterns of variation within the data, allowing the complex

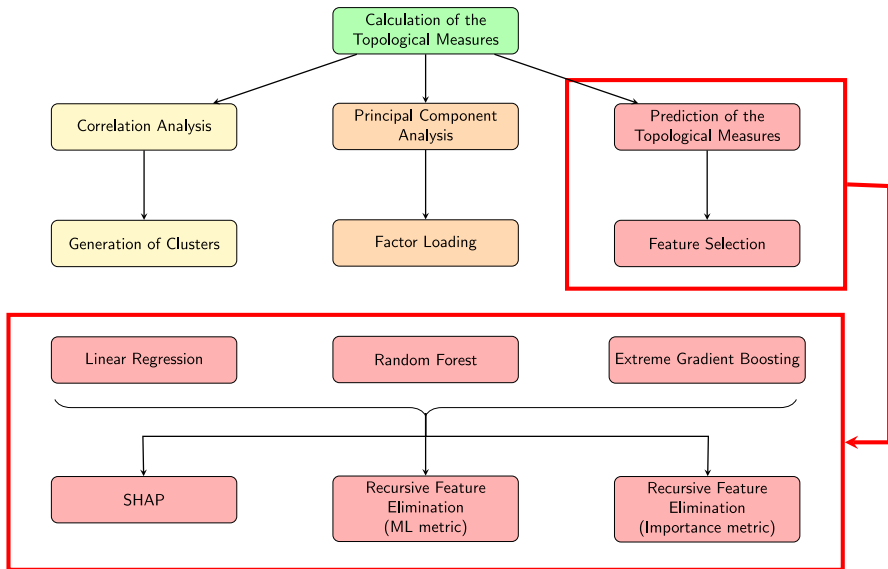


Fig. 2 Proposed pipeline for the analysis of the topological metrics in networks

network's structure to be summarized using a reduced set of orthogonal variables. To ensure accurate results, the data are first standardized to avoid skewed eigenvectors due to varying magnitudes. The factor loadings of the topological measures are also computed, providing insight into the importance of each variable in relation to the principal components. Thereby, this procedure enhances the interpretability of the PCA results and provides further discoveries on the clusters obtained in the correlation analysis.

The final stage, portrayed in the red boxes of Fig. 2, considers the application of ML algorithms to predict the topological measures based on a subset of features. Specifically, linear regression, random forests, and extreme gradient boosting (XGB) are the predictive models used in our study. The selection of these ML algorithms is justified by the limited amount of data points. Linear regression, random forests, and XGB perform efficiently when trained considering small datasets while more complex algorithms, e.g., neural networks, achieve remarkable results only when properly trained with large amount of data.

Additionally, two feature selection techniques are employed: SHAP and RFE. SHAP is a game-theoretic approach to explain the contribution of each feature to the model's predictions [15], while RFE iteratively removes the least important features based on their contribution to the prediction accuracy (evaluated using the coefficient of determination R^2) or the importance weights of the features [9]. These methods allow for the identification of the most informative subset of features, optimizing the trade-off between prediction accuracy and computational efficiency.

By integrating these three components (correlation analysis, PCA, and predictive modeling) the proposed methodology offers a comprehensive approach to efficiently analyze complex networks. It not only reduces the computational burden but also provides a systematic way to predict uncalculated topological measures, ultimately

enhancing our understanding of the network's structure. Moreover, the most appropriate subset of measures summarizing the underlying structure of the complex networks is determined from the comparison of the statistical results and ML outputs.

4 Results

In this section, we present the results of our analysis for both metro and road networks, following the methodology described previously. The computations were performed using parallel processing on a 2x AMD EPYC 7763 CPU, with 256 cores ($2 \times 64 +$ hyperthreading) and 1007GB of RAM. The analysis utilized Python libraries such as *NetworkX 3.1*, *SHAP 0.41.0*, and *Scikit-learn 1.2.2*.

4.1 Metro Networks

The first step in this analysis was to calculate the topological measures. The values obtained for each metric are listed in [Appendix](#) in Table 15, while the computational time for each metric is shown in Table 16. These computations reveal that, despite their complexity, the calculations were completed in under a second due to the relatively small size of the metro networks. The number of vertices ranged between 8 and 309, and the number of edges between 7 and 356, classifying these networks as small to medium in size.

A key observation in this analysis is that the CC for most metro networks is zero. This outcome is likely due to the linear structure of most metro systems, where stations are typically connected in a sequence, forming lines rather than triangles, which are necessary for high clustering values.

Moving to the correlation analysis, Pearson's correlation coefficients for each pair of topological measures are shown in Fig. 3. Strong correlations (absolute values greater than 0.60, as generally considered in the literature [25]) were observed in 25 out of 66 possible comparisons, with 11 positive and 14 negative correlations. To better visualize these relationships, we generated a correlation graph shown in Fig. 4, which revealed two principal clusters. The first cluster includes BE, BV, ND, and AC, while the second contains AVC, ANN, CC, DA, AVP, VE, and AD. The CC and DA are only weakly connected to this second cluster, and the RC remains largely uncorrelated.

The next step in the analysis involved calculating the principal components using PCA. The explained variance for the first 12 principal components is illustrated in Fig. 5. After standardization, we observe that the first principal component accounts for 61.48% of the total variance. However, to adequately represent the dataset (i.e., capturing at least 95% of the explained variance), it is necessary to consider five principal components.

Examining the factor loadings for the topological measures in the first five components, as shown in Table 1, we find that the first principal component is predominantly influenced by a cluster of measures, including ND, BE, BV, and AC. As previously discussed in the correlation analysis, this cluster exhibits a strong negative correlation with another major cluster centered around AD. The second principal component

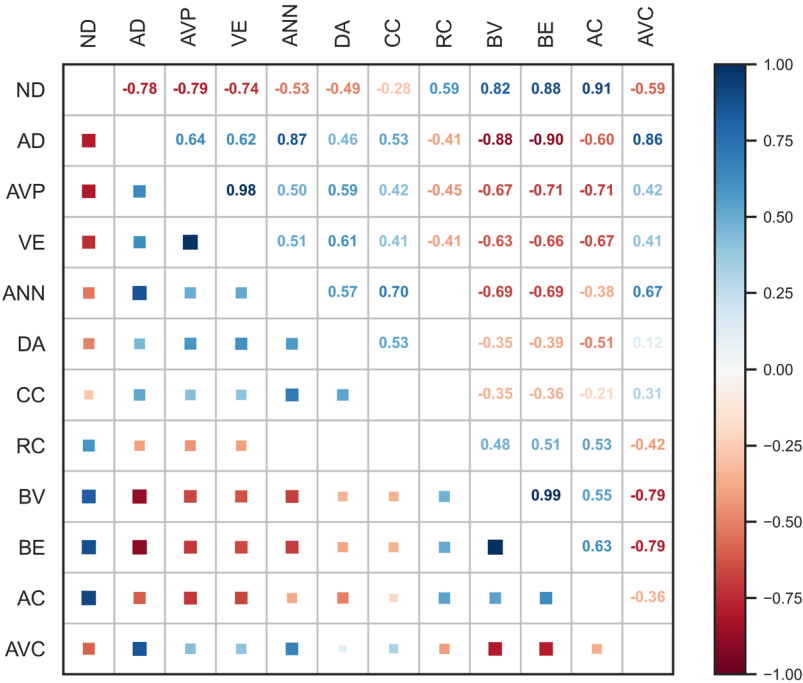


Fig. 3 Pearson’s correlation coefficients of the topological measures for the metro networks (colored in blue for the positive correlation and red for negative correlation)

highlights isolated features, showing a strong positive correlation with the CC and the RC, along with a moderate positive correlation with the DA. This suggests that the first component already encapsulates significant information from the second cluster. The remaining three components consist of more diverse combinations of features, but they do not reveal any substantial or influential correlations.

After conducting various statistical analyses of the relationships between features in metro networks, the focus has been turned to predicting individual features using the remaining ones as inputs. Specifically, we compared the performance of linear regression, random forest, and XGB for forecasting, while employing SHAP and RFE for feature selection. The RFE process was conducted using both feature importance weights (RFEI) and changes in the R^2 metric (RFEM). To validate the analysis, we divided the full dataset of 11 topological metrics (with one as the output) from the 46 cities into training and test sets, following a 3 : 1 ratio. This resulted in a training set of 34 data points and a test set of 12.

We report the results in terms of the selected features and R^2 values, starting with the linear regression model. For brevity, we focus on the most significant findings, while the full results for the selected feature sets are provided in Table 2, and the corresponding R^2 values are shown in Table 3. Notably, the models struggled to predict features that were disconnected or isolated from the main clusters, such as the DA, CC, and RC. In these cases, the R^2 values were negative, indicating the difficulty of accurately predicting these features. Indeed, these topological measures

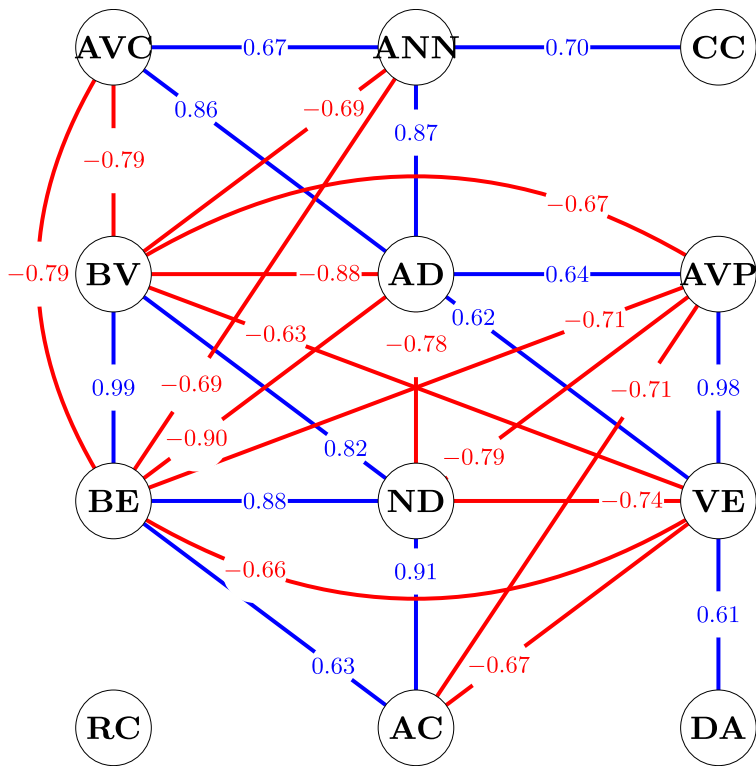
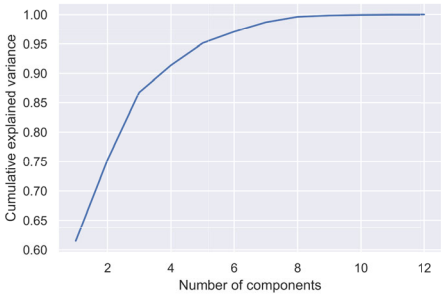


Fig. 4 Representation of the correlation cluster graph for the topological measures of metro networks

Components	Explained variance
1	0.614798
2	0.137898
3	0.115102
4	0.046124
5	0.037489
6	0.019499
7	0.016221
8	0.009424
9	0.002221
10	0.000851
11	0.000356
12	0.000016

(a) Explained Variance for each Principal Component



(b) Cumulative Explained Variance

Fig. 5 Explained variance for the first 12 principal components for the metro networks

Table 1 Factor loading of the topological measures in the metro network

	<i>ND</i>	<i>AD</i>	<i>AVP</i>	<i>VE</i>	<i>ANN</i>	<i>DA</i>	<i>CC</i>	<i>RC</i>	<i>BV</i>	<i>BE</i>	<i>AC</i>	<i>AVC</i>
Component 1	0.916753	-0.927415	-0.851200	-0.824763	-0.769048	-0.581155	-0.522792	0.508777	0.901286	0.933128	0.762366	-0.739380
Component 2	0.248772	0.069502	-0.011106	0.035675	0.485519	0.561569	0.651170	0.705988	0.135966	0.145240	0.234948	-0.135463
Component 3	-0.175218	-0.309114	0.401758	0.412831	-0.318305	0.437010	-0.071592	-0.054325	0.277288	0.210422	-0.413772	-0.574792
Component 4	0.169109	-0.046640	0.161675	0.166005	-0.044749	-0.226866	0.460966	-0.380784	0.098675	0.118839	0.184870	-0.005339
Component 5	0.138479	-0.068125	0.277132	0.318558	-0.059133	-0.069858	-0.228573	0.197411	-0.143254	-0.090488	0.340573	0.058970

Table 2 Selected features for linear regression

Features	Output											
	ND	AD	AVP	VE	ANN	DA	CC	RC	BV	BE	AC	AVC
ND	■	■	✓	✓	■	●	■	●	✓	●	●	■
AD	■	■	■	■	■	●	■	✓	✓	✓	●	■
AVP	✓	■	■	✓	✓	✓	●	■	■	■	●	✓
VE	■	■	■	■	■	✓	●	■	■	■	●	■
ANN	●	✓	✓	■	■	✓	●	✓	●	●	●	■
DA	■	●	✓	✓	■	■	■	✓	■	■	■	✓
CC	■	■	■	■	■	■	■	■	■	■	■	■
RC	■	■	■	■	■	■	■	■	■	■	■	■
BV	✓	■	■	■	■	■	■	■	■	■	■	■
BE	✓	✓	■	✓	■	■	■	■	■	■	■	■
AC	✓	■	■	■	■	■	■	■	■	■	■	■
AVC	■	■	■	■	■	■	■	■	■	■	■	■
SHAP ✓, RFEM ●, RFEI ■												

Table 3 Linear regression R^2

	ND	AD	AVP	VE	ANN	DA	CC	RC	BV	BE	AC	AVC
R^2_{SHAP}	0.950745	0.948226	0.976000	0.972994	0.791868	-0.467867	-24.254429	-0.926574	0.999126	0.999338	0.288190	0.630156
R^2_{RFEM}	0.950686	0.831016	0.965388	0.964198	0.516257	-2.170150	-22.421979	-2.628122	0.998289	0.998659	0.812253	0.743163
R^2_{RFEI}	0.953685	0.937340	0.982685	0.980552	0.526916	-2.699641	-21.206942	-1.713796	0.998530	0.998692	0.814438	0.762572
R^2	0.947773	0.906828	0.969001	0.960048	0.683127	-1.401943	-18.993547	-1.036491	0.997830	0.998360	0.787994	0.580374

are inherently unpredictable through this set of information as attested by the poor correlations shown with the other variables. For this reason, it is necessary to extend the data included in the analysis to obtain acceptable results in their prediction.

A further observation arises from the feature selection process: only for the CC prediction did we identify a clear rule for selecting the subset of features. For the other metrics, the choice of input features varied depending on the method used, largely due to the absence of a definitive optimal subset. This issue was also encountered in the experiments with random forest and XGB.

For the remaining topological measures, prediction errors were generally low. It is worth noting that in some cases, the full model did not provide optimal forecasts, particularly for the ANN and AVC. However, when we compared these metrics with those obtained by selecting specific features, we found that the coefficient of determination could be improved to 0.79 using SHAP for ANN, and to 0.76 using RFEI for AVC. Overall, prediction errors did not vary significantly between using the full dataset or a reduced set of features.

Similar conclusions can be drawn from the random forest results. The selected features and corresponding coefficients of determination for the test set are listed in Tables 4 and 5. In this case, the recursive feature elimination methods tended to include a larger number of features. However, this did not result in improved prediction accuracy. Interestingly, SHAP achieved comparable performance using only four features at a time, indicating that the output can be well approximated with a minimal set of relevant inputs.

Another noteworthy observation is the comparison between the R^2 values for linear regression and random forest. In most cases, linear regression outperformed random forest. Exceptions occurred with the DA, CC, and AC, where random forest provided better approximations. However, as observed in the linear regression analysis, DA and CC belong to the class of isolated features, making it challenging to select appropriate input features for their prediction. For AC, despite the random forest's better performance, the coefficient of determination for linear regression was still scientifically acceptable at 0.79.

Finally, the selected features varied depending on the regression model used. Only two instances shared the exact same subset of features for both linear regression and random forest: BV with SHAP, and BE with SHAP and RFEI. In other cases, the models selected different combinations of topological measures, although some overlap was observed when predicting the same output.

The considerations mentioned earlier were also validated through experiments using the XGB method. The variables selected for this context are provided in Table 6, with the corresponding R^2 values shown in Table 7. Notably, there was an exceptional case involving the isolated feature DA. For this feature, the RFEI model produced a positive coefficient of determination, a result that had not been achieved with any previous method. Additionally, for the AVC, the model was able to approximate the output reasonably well using only the features selected by the RFEM method.

Moreover, when comparing the results with those obtained using SHAP, it becomes clear that the ANN is a key feature. ANN is included in the RFEI model but not in the SHAP model, highlighting its importance. A similar observation can be made

Table 4 Selected features for random forest

Features	Output											
	ND	AD	AVP	VE	ANN	DA	CC	RC	BV	BE	AC	AVC
ND	■	✓	✓	✓	■	■	✓	■	✓	■	■	●
AD	●	■	●	●	●	●	✓	●	✓	●	●	✓
AVP	■	■	■	■	■	■	●	■	■	■	■	●
VE	●	■	✓	■	■	■	●	●	■	●	●	●
ANN	●	✓	✓	■	■	■	✓	●	●	●	●	●
DA	●	■	✓	✓	■	■	■	■	●	●	●	●
CC	●	■	●	●	■	■	■	■	●	●	●	●
RC	●	■	●	●	■	■	■	■	■	■	■	■
BV	✓	●	●	●	■	■	■	■	■	■	■	■
BE	✓	✓	●	●	■	■	■	■	■	■	■	■
AC	■	■	✓	✓	■	■	●	■	■	■	■	■
AVC	■	✓	✓	✓	■	■	■	■	■	■	■	■
SHAP ✓, RFEM ●, RFEI ■												

Table 5 Random forest R^2

	ND	AD	AVP	VE	ANN	DA	CC	RC	BV	BE	AC	AVC
R^2_{SHAP}	0.938164	0.889520	0.950552	0.921131	0.190528	-0.059183	-8.206427	-3.199260	0.894742	0.936816	0.964030	0.270231
R^2_{RFEM}	0.926842	0.891924	0.940929	0.927943	0.206184	-0.287366	-13.365895	-1.879339	0.918395	0.952274	0.957341	0.064595
R^2_{RFEI}	0.949491	0.890548	0.951762	0.920138	0.243719	-0.096437	-7.798809	-3.212647	0.900252	0.941660	0.960443	0.239398
R^2	0.927336	0.894174	0.942309	0.923821	-0.060730	0.108959	-9.004830	-3.366023	0.906618	0.933599	0.942903	0.155532

Table 6 Selected features for XGB

Features	Output											
	ND	AD	AVP	VE	ANN	DA	CC	RC	BV	BE	AC	AVC
ND	■	✓	■	✓	■	■	✓	■	✓	■	✓	■
AD	✓	■	■	■	■	●	■	■	✓	■	■	■
AVP	✓	■	■	✓	■	●	✓	■	✓	✓	■	■
VE	■	■	■	■	■	■	■	■	■	■	■	■
ANN	■	✓	■	■	■	✓	✓	■	✓	■	■	■
DA	●	■	■	■	■	■	■	■	■	✓	■	■
CC	■	■	■	■	■	■	■	■	■	■	■	■
RC	■	■	■	■	■	■	■	■	■	■	■	■
BV	✓	✓	■	■	■	■	■	■	■	■	■	■
BE	✓	■	■	■	■	■	■	■	■	■	■	■
AC	■	■	✓	■	■	✓	■	■	■	■	■	■
AVC	■	■	■	■	■	✓	■	■	■	■	■	■
SHAP ✓, RFEM ●, RFEL ■												

Table 7 XGB R^2

	ND	AD	AVP	VE	ANN	DA	CC	RC	BV	BE	AC	AVC
R^2_{SHAP}	0.956054	0.905015	0.922705	0.929900	0.125844	-0.073378	-5.858772	-5.879550	0.946106	0.925570	0.946009	0.121962
R^2_{RFEM}	0.898683	0.742185	0.898391	0.917288	-0.067203	-0.147108	-6.941562	-4.420400	0.902773	0.975978	0.918739	0.570213
R^2_{RFEI}	0.902764	0.854853	0.918221	0.891585	0.342405	0.294456	-5.429628	-4.563038	0.877861	0.915890	0.928731	0.427139
R^2	0.915969	0.793333	0.921423	0.891544	0.291111	-0.360931	-1.615500	-4.533205	0.885512	0.910154	0.929637	0.367852

when comparing RFEI selections, where AVP was chosen instead of BV, resulting in a reduction of approximately 0.14 in the R^2 value.

In conclusion, selecting a specific subset of topological measures can improve accuracy while reducing the time required to compute excluded metrics. However, it is important to note that this selection is influenced by the predictive model used. For metro networks, we found that linear regression provided the best performance across both the restricted and full datasets.

4.2 Road Networks

In this subsection, we present the results from the analysis of road networks for the same 46 European cities previously analyzed in the metro study. However, the road networks differ significantly in scale compared to the metro networks. The size of these graphs ranges from Lausanne, the smallest network with 1,358 vertices and 1,859 edges, to London, the largest, with 127,540 vertices and 160,010 edges. As a result, this section focuses on the analysis of large-scale graphs. In [Appendix](#), detailed information on the fundamental measures calculated for each network is provided in [Table 17](#), while [Table 18](#) outlines the computation times for each topological measure across cities.

One notable observation is the dramatic increase in computational time compared to metro networks. For instance, calculating BE and BV requires several hours, while VE and AVP can be computed in a few minutes. Consequently, identifying and selecting the most representative features of a graph for computation becomes a critical task to optimize efficiency.

The first step of the proposed analytical pipeline involves calculating the correlation coefficients, illustrated in [Fig. 6](#). In this case, we observed that the interactions between features were relatively weak. Out of 66 possible pairs, only 10 coefficients had an absolute value greater than 0.60. This aspect is justified by the structural difference between the two transportation networks: metro networks generally have simpler, linear structures, where stations are sequentially connected along defined lines, while, road networks exhibit a more complex topology, characterized by a dense interconnection of junctions and streets. Additionally, the latter networks span much larger areas, containing significantly more vertices and edges (up to hundreds of thousands). Such a large scale introduces greater variability and structural heterogeneity. This complexity results in weaker correlations among the topological metrics, since there are many possible routes and structural variations leading to less uniformity across the network. Moreover, all of the strong correlations were positive, while the negative correlations were of moderate magnitude [\[25\]](#), falling within the range of 0.40 to 0.60. Despite this, we were still able to clearly define multiple clusters of topological measures, shown in [Fig. 7](#).

The first prominent cluster comprises the AD, DA, AVC, and ANN. Within this cluster, the relationships are strong, with all four variables being closely interconnected. A second visible group includes ND, BE, and BV. This set of features can be expanded by incorporating AC and the RC through positive medium correlations.

Additionally, this extended cluster has a negative influence on the final pair of metrics: AVP and VE. All of the medium negative correlations occur between these

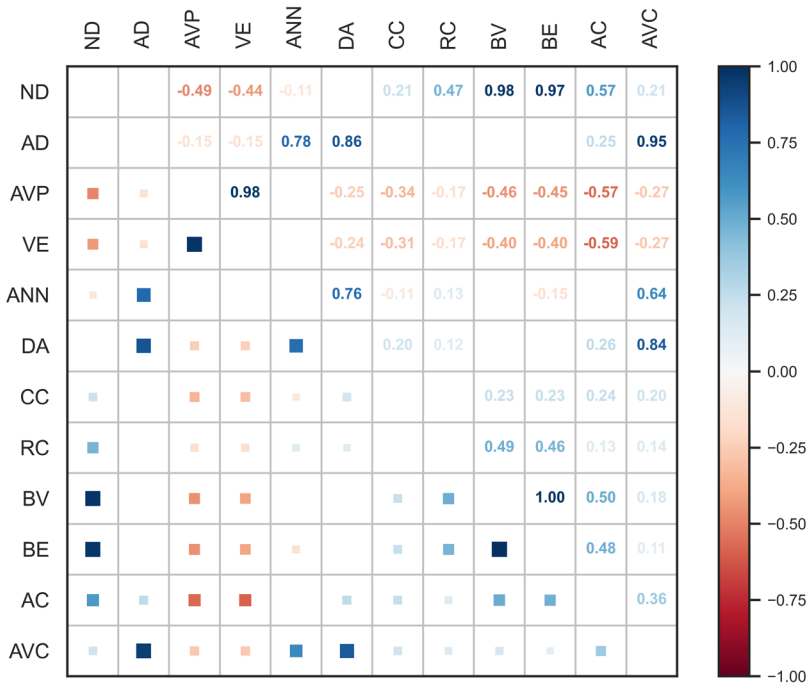


Fig. 6 Pearson's correlation coefficients of the topological measures for the road networks (colored in blue for the positive correlation and red for negative correlation)

two groups. Furthermore, AVP and VE exhibit a very high correlation with each other, forming the last identifiable group of features in road networks. The CC, in contrast, remains completely isolated from all other clusters in this analysis.

To assess the relative importance of each cluster, we examined the results of the PCA. First, it is noteworthy to observe the explained variance of each principal component, as shown in Fig. 8. Unlike the metro network analysis, the first two principal components account for substantial variance, with values of 0.39 and 0.28, respectively. Despite the high explained variance of these components, at least five orthogonal directions are needed to reach an acceptable total variance of approximately 0.93.

Subsequently, we computed the factor loadings of the topological measures, as presented in Table 8. From this analysis, we observe that the first principal component is positively correlated with the cluster comprising VE and AVP. It also shows a negative correlation with the extended group of features. These correlations, particularly in absolute terms, are the strongest, confirming that these two sets of topological measures are closely interconnected.

The second principal component provides insights into the remaining highly correlated cluster. The subsequent components primarily involve combinations of different measures, resulting in medium-level interactions with the other features. Notably, the fourth principal component exhibits a significant factor loading for the CC, an

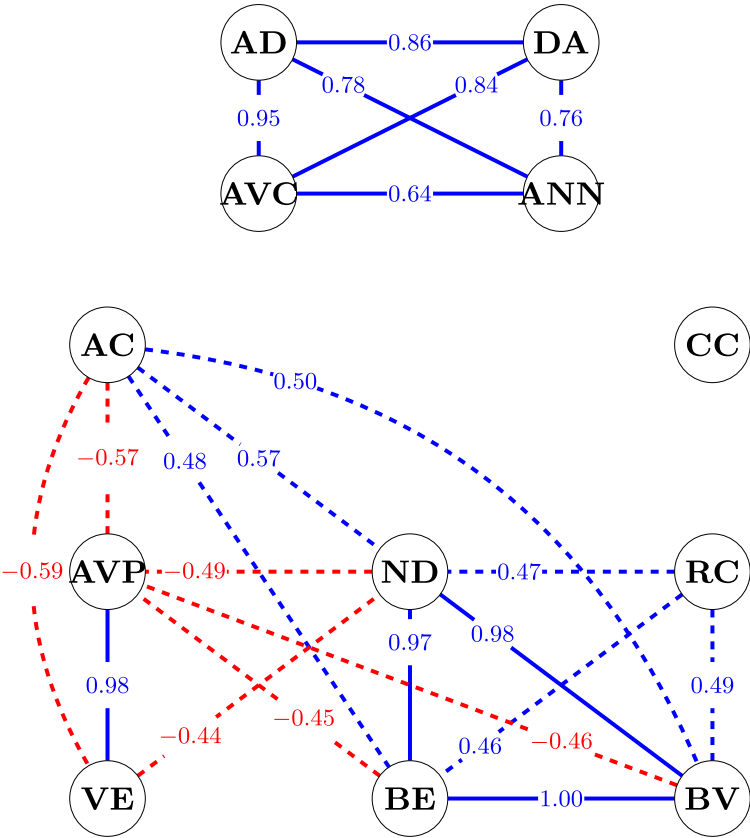
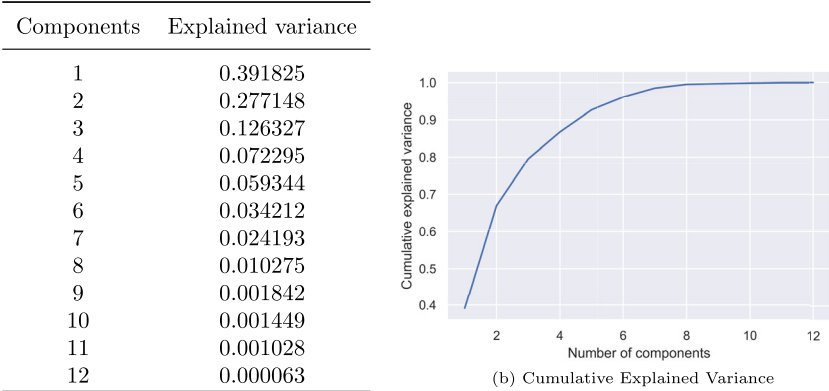


Fig. 7 Representation of the correlation cluster graph for the topological measures of road networks. The dotted lines represent a medium correlation (absolute value between 0.40 and 0.60) while normal lines are the strong correlations (absolute value > 0.60)



(a) Explained Variance for each Principal Component

(b) Cumulative Explained Variance

Fig. 8 Explained variance for the first 12 principal components for the road networks

Table 8 Factor loading of the topological measures in the road network

	<i>ND</i>	<i>AD</i>	<i>AVP</i>	<i>VE</i>	<i>ANN</i>	<i>DA</i>	<i>CC</i>	<i>RC</i>	<i>BV</i>	<i>BE</i>	<i>AC</i>	<i>AVC</i>
Component 1	-0.830784	-0.433530	0.754811	0.725210	-0.219527	-0.453105	-0.374711	-0.452145	-0.808894	-0.777577	-0.733414	-0.576986
Component 2	0.422919	-0.858633	-0.119435	-0.094536	-0.832054	-0.829109	0.055400	0.136075	0.434472	0.498413	0.035254	-0.739405
Component 3	0.289035	0.119248	0.534376	0.560549	0.260029	-0.017133	-0.433386	0.547254	0.335170	0.311539	-0.223995	0.059965
Component 4	0.074395	0.014545	0.275185	0.330661	-0.099629	0.073856	0.757944	-0.188115	0.110463	0.110640	-0.107074	0.124782
Component 5	-0.156346	-0.108907	-0.155908	-0.157396	0.021430	0.103035	0.275993	0.608491	-0.100855	-0.107787	-0.376684	-0.074519

otherwise isolated feature. Additionally, this component has a strong positive correlation with the DA, suggesting a potential relationship between these two metrics.

Following the same training approach used for the metro networks, we validated the earlier results through the prediction phase of the analytical pipeline. In this case, we focus on the most significant findings, while the remaining outcomes are presented in the accompanying tables.

The first model considered was linear regression and the corresponding results are reported in Tables 9 and 10. As shown in Table 10, the predictions did not consistently yield optimal performance. A notable example is the prediction of the ANN. Despite being a highly correlated feature, it remains challenging to predict accurately, regardless of the input features selected. Table 9 demonstrates that all measures correlated with ANN were chosen as key inputs, yet the R^2 value remained negative. Similar results were observed for the CC, though in this case, the poor prediction accuracy can be attributed to the isolation of CC from other feature clusters.

Overall, the trend indicates that prediction accuracy improves with careful feature selection. This was confirmed across multiple tests, where the coefficient of determination increased by approximately 0.25 when a specific set of features was selected.

However, this strong result is not reflected in the other two models (see Tables 11, 12, 13 and 14). In particular, with random forest, there were no significant changes (less than 0.10) when using different subsets of features, as shown in Table 12. The only exceptions, where performance improved beyond this threshold, were for AVP and CC, both using SHAP.

An interesting observation arises when examining the selected features listed in Table 11. The choice of topological measures is fairly consistent across the different methodologies applied. This consistency is also mirrored in the R^2 results: using the same subset of inputs tends to produce similar forecasting outcomes. Furthermore, since the prediction errors are comparable to those obtained with the full model, the most informative features have been successfully selected.

Before transitioning to the next model analysis, it is worth noting the results related to the ANN. While the coefficient of determination was always positive (around 0.64), this value is still not satisfactory.

Better results were achieved with the XGB model. The topological measures selected for each prediction using XGB are presented in Table 13, while the corresponding coefficients of determination from the testing phase are listed in Table 14. As mentioned earlier, the RFEM method yielded good performance in predicting the ANN. Specifically for the XGB case, we observed a notable difference between RFEM and SHAP. However, when comparing SHAP with RFEI (which performs similarly to the full model), the only variation in feature selection was the inclusion of AVP. Despite the lack of a significant correlation between ANN and AVP, it is clear that AVP has a considerable influence on the model's ability to predict ANN.

Overall, the accuracies achieved with XGB and random forests were comparable. However, as in previous cases, poor results were obtained for isolated features, reinforcing the importance of directly calculating these metrics.

Table 9 Selected features for linear regression

Features	Output											
	ND	AD	AVP	VE	ANN	DA	CC	RC	BV	BE	AC	AVC
ND	■	■	■	■	■	✓	■	✓	■	✓	■	■
AD	■	■	■	■	■	✓	■	■	■	■	■	■
AVP	■	■	■	■	■	■	■	■	■	■	■	■
VE	■	■	■	■	■	■	■	■	■	■	■	■
ANN	■	■	■	■	■	■	■	■	■	■	■	■
DA	■	■	■	■	■	■	■	■	■	■	■	■
CC	■	■	■	■	■	■	■	■	■	■	■	■
RC	■	■	■	■	■	■	■	■	■	■	■	■
BV	■	■	■	■	■	■	■	■	■	■	■	■
BE	■	■	■	■	■	■	■	■	■	■	■	■
AC	■	■	■	■	■	■	■	■	■	■	■	■
AVC	■	■	■	■	■	■	■	■	■	■	■	■
SHAP ✓, RFEM ●, RFEI ■	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 10 Linear regression R^2

	ND	AD	AVP	VE	ANN	DA	CC	RC	BV	BE	AC	AVC
R^2_{SHAP}	0.982941	0.943263	0.925412	0.921262	-1.170507	0.643634	-0.134785	-0.053726	0.996018	0.981424	0.422221	0.894990
R^2_{RFEM}	0.763239	0.947141	0.829198	0.902390	-1.614260	0.346339	-0.169483	0.044717	0.996673	0.996369	0.325781	0.890164
R^2_{RFEI}	0.763239	0.880538	0.876592	0.882708	-1.763658	0.621421	-0.196800	0.050838	0.996083	0.995319	0.463975	0.920215
R^2	0.748714	0.730139	0.843971	0.852106	-1.738818	0.418725	-1.356248	0.131662	0.998584	0.997837	0.102107	0.621594

Table 11 Selected features for random forest

Features	Output											
	ND	AD	AVP	VE	ANN	DA	CC	RC	BV	BE	AC	AVC
ND	■		✓	●	■	✓	✓	✓	✓	✓	✓	●
AD	●	■	●	■	■	✓	■	●	●	✓	●	●
AVP	✓	■	■	■	■	✓	■	●	✓	✓	■	●
VE		■	✓	■	■	✓	■	●			■	
ANN		✓	●	■	■	✓	■	●		●	●	■
DA		✓	●	■	■	✓	■	●			●	■
CC		✓	●	●	■	✓	■	●	●		●	■
RC	■				■	✓	■	✓	✓	✓	●	■
BV	✓	■	●	■	■	✓	■	●	■	✓	■	●
BE	✓	■	✓	■	■	✓	■	●	■	✓	●	●
AC	✓	■	✓	✓	■	✓	■	●	✓	✓	●	●
AVC	✓	■	✓	✓	✓	✓	■	●	●	✓	●	■

SHAP ✓, RFEM ●, RFEL ■

Table 12 Random forest R^2

	ND	AD	AVP	VE	ANN	DA	CC	RC	BV	BE	AC	AVC
R^2_{SHAP}	0.826513	0.786998	0.835218	0.908193	0.647517	0.693857	-0.114572	0.095183	0.933388	0.904024	0.490160	0.859182
R^2_{RFEM}	0.724716	0.773535	0.901575	0.903498	0.663890	0.495067	-0.340280	0.080388	0.913590	0.905001	0.409934	0.807671
R^2_{RFEI}	0.816161	0.811773	0.849848	0.909967	0.615401	0.593065	-0.242279	0.090905	0.950692	0.908551	0.492554	0.825163
R^2	0.808640	0.772625	0.786533	0.887647	0.646603	0.593095	-0.268207	0.044489	0.916335	0.915817	0.457599	0.782199

Table 13 Selected features for XGB

Features	Output												
	ND	AD	AVP	VE	ANN	DA	CC	RC	BV	BE	AC	AVC	
ND	■	●	✓	■		●	✓	✓	✓	✓	✓	✓	■
AD	■	■	●	■	✓	●	✓	■	●	■	●	●	■
AVP	✓	■	■	■	■	●	■	●	●	✓	■	✓	■
VE	●	■	✓	■	■	●	✓	●	●	●	●	✓	■
ANN	●	✓	●	■	■	●	✓	●	●	■	●	●	■
DA	●	✓	✓	●	■	■	■	■	●	✓	●	●	■
CC	●	■	■	●	■	■	■	■	●	■	●	●	■
RC	●	■	■	●	■	■	■	■	●	■	●	●	■
BV	✓	●	■	●	■	●	■	■	■	■	●	✓	■
BE	■	■	■	●	■	●	■	■	■	■	●	■	■
AC	✓	■	✓	✓	✓	●	■	✓	✓	■	●	■	■
AVC	■	■	■	■	■	■	■	■	■	■	■	■	■
SHAP ✓, RFEM ●, RFEI ■													

Table 14 XGB R^2

	ND	AD	AVP	VE	ANN	DA	CC	RC	BV	BE	AC	AVC
R^2_{SHAP}	0.909292	0.724856	0.903402	0.882579	0.291743	0.630523	-0.530387	-0.024946	0.961840	0.947753	0.463920	0.816234
R^2_{RFEM}	0.805525	0.588517	0.902173	0.888480	0.846979	0.573294	-1.113806	0.020257	0.943276	0.824694	0.447121	0.772353
R^2_{RFEI}	0.845489	0.743405	0.853026	0.819157	0.794181	0.693664	-1.275593	-0.015827	0.933938	0.947408	0.408047	0.795994
R^2	0.830658	0.719017	0.853922	0.800807	0.786731	0.526808	-1.251441	-0.000107	0.938814	0.949034	0.457871	0.794754

5 Discussion

This section provides a comparative analysis between the results obtained from the transportation networks analyzed here and those found in the existing literature. Beginning with the correlation analysis, certain recurring patterns are evident in both metro and road networks. Specifically, two primary clusters of features exhibit strong correlations in both network types. The first cluster, referred to as the Distance-Connectivity group, includes BE, BV, ND, and AC. In road networks, the correlations involving AC are of moderate strength. Nevertheless, this metric is included due to its multiple negative correlations with VE and AVP, similar to other measures within this group.

The second identifiable cluster consists of Degree-associated features, including AD, AVC, and ANN. For road networks, this cluster also includes a strong correlation with DA, a relationship that is not observed in metro networks. This discrepancy may stem from the structural differences between the two types of networks: metro systems typically consist of long, linear vertex sequences, while road networks exhibit more complex interconnections. As a result, road networks demonstrate more intricate dependency patterns between vertex degrees, offering deeper insights into their connectivity.

A third notable set of correlated features includes VE and AVP, which together form a Distance cluster. An important observation is the lack of significant correlations involving the CC and the RC. These two metrics do not exhibit notable correlations with other features, making them difficult to predict or approximate, as demonstrated in the forecasting analyses. Nonetheless, these measures have been shown to correlate with others in different types of networks [12], suggesting that their lack of significance in transportation networks may be context-specific.

The clusters of highly correlated features identified in this study can be compared with those reported in the literature. For instance, [12] identified four clusters based on distance, degree, intra-connectivity, and inter-connectivity. When comparing these findings with our results, the Distance-Connectivity group observed in our study appears to be a combination of the Distance and Intra-connectivity clusters from their work. Furthermore, the Degree cluster in our analysis includes the ANN rather than the CC. It is important to note that [12] study covered a broad range of networks, including infrastructure and social systems, which may explain the absence of some correlation links observed in our transportation network analysis. This highlights the fact that the relationships between topological measures are highly context-dependent, a conclusion supported by multiple authors [2, 16].

From the PCA, we gathered insights regarding the variance explained by a smaller number of principal components. Factor loadings enabled us to evaluate the relative importance of each feature. In metro networks, the first principal component was strongly correlated with the Distance-Connectivity group, while in road networks, it was more strongly linked to the Distance cluster. Notably, the first principal component explained 61% of the variance in metro networks, compared to only 39% in road networks. This difference is likely due to the varying scale and complexity of the networks analyzed; metro networks tend to be smaller or medium-sized, whereas road networks often encompass a significantly larger number of vertices and edges.

Consequently, it is necessary to consider multiple features in larger graphs, such as those representing road networks.

Finally, when comparing the features selected in the prediction phase with the identified correlation groups, some overlap is evident. However, the most correlated measures are not always selected for improving prediction accuracy. This can be attributed to the fact that having features outside of a specific cluster is often necessary to avoid redundancy. Over-reliance on highly correlated features may result in redundant information, which can limit the effectiveness of the predictive models. Conversely, metrics that exhibit multiple correlations are easier to predict, while those with fewer or no correlations remain challenging. The selection of features and prediction accuracy are thus highly dependent on the specific model and experiment. A generalized approach applicable to all scenarios is difficult to achieve. For instance, linear regression outperformed other models in metro networks, while XGB proved more effective for road networks. Overall, selecting an appropriate subset of features for training machine learning models enhances the predictive performance and reduces computational complexity.

6 Conclusion and Future Work

In this paper, we present an enhanced analytical pipeline for network analysis. Our approach is structured into three distinct phases. First, we compute the correlation coefficients of the topological measures and visualize the resulting clusters to identify patterns and relationships. Next, we apply PCA to assess how much of the data can be explained by a reduced set of components, allowing us to extract relevant information from the previously identified clusters. Finally, a predictive modeling phase is carried out to guide the selection of the most informative features. In particular, we employ a game-theoretic approach known as SHAP and the RFE algorithm (using feature importance and the coefficient of determination as metrics) to select inputs for the training phase. This methodology was tested on 46 EU/EEA metro and road networks, yielding promising results in terms of discovering correlations between features and predicting missing information. Furthermore, our findings are compared with existing literature, offering new perspectives and insights for future research.

In comparison with the work of [12], our study reveals that the relationships between features can vary significantly depending on the network context. This suggests the need for further investigation across different types of networks to better understand the correlation patterns that emerge in various domains. A natural extension of this study would involve exploring railway and tram networks to assess whether similar patterns hold. Additionally, comparing the selected features for prediction with the most highly correlated measures could help validate the reliability of the feature selection algorithms used. Lastly, optimizing the selection of features for ML models by taking into account the computational time required for each metric would be valuable. This would enable a more balanced approach to gaining insights into the network while minimizing the computational burden.

Appendix. Metro and road network data

In this appendix, are reported additional information concerning the metro and road network considered in our study. Specifically, we first list in alphabetical order all the cities included in our analysis:

- | | | | |
|------------------------------|------------------------------|------------------------------------|--------------------------|
| 1. Amsterdam,
Netherlands | 13. Frankfurt,
Germany | 26. Milan, Italy | 38. Sevilla, Spain |
| 2. Athens,
Greece | 14. Genoa, Italy | 27. Munich,
Germany | 39. Sofia,
Bulgaria |
| 3. Barcelona,
Spain | 15. Glasgow, UK | 28. Naples, Italy | 40. Stockholm,
Sweden |
| 4. Berlin,
Germany | 16. Hamburg,
Germany | 29. Nuremberg,
Germany | 41. Toulouse,
France |
| 5. Bilbao, Spain | 17. Helsinki,
Finland | 30. Oslo, Norway | 42. Turin, Italy |
| 6. Brescia, Italy | 18. Lausanne,
Switzerland | 31. Palma de
Mallorca,
Spain | 43. Tyne and
Wear, UK |
| 7. Brussels,
Belgium | 19. Lille, France | 32. Paris, France | 44. Valencia,
Spain |
| 8. Bucharest,
Romania | 20. Lisbon,
Portugal | 33. Porto,
Portugal | 45. Vienna,
Austria |
| 9. Budapest,
Hungary | 21. London, UK | 34. Prague,
Czech
Republic | 46. Warsaw,
Poland |
| 10. Catania, Italy | 22. Lyon, France | 35. Rennes,
France | |
| 11. Charleroi,
Belgium | 23. Madrid,
Spain | 36. Rome, Italy | |
| 12. Copenhagen,
Denmark | 24. Malaga,
Spain | 37. Rotterdam,
Netherlands | |
| | 25. Marseille,
France | | |

For the sake of completeness, we include in the following pages the tables reporting the different topological measures computed (in Table 15 for the metro networks and Table 17 for the road networks) and the corresponding time required for the calculation of the specific values (in Table 16 for the metro networks and Table 18 for the road networks).

Table 15 Topological measures of metro networks for European cities

City, Country	$ V $	$ E $	N/D	AD	AVP	VE	ANN	DA	CC	RC	BV	BE	AC	AVC
Amsterdam Metro, Netherlands	53	53	0.038462	2.000000	11.299710	23.264151	2.309524	0.129108	0.020000	0.127313	0.201955	0.213202	0.007858	1.056604
Athens Metro, Greece	61	62	0.033880	2.032787	10.917486	22.590164	2.225490	0.163418	0.000000	0.184564	0.168093	0.176088	0.009957	1.131148
Barcelona Metro, Spain	149	168	0.015237	2.255034	11.153183	22.751678	2.547139	0.062184	0.004000	0.061021	0.069069	0.066388	0.010276	1.711409
Berlin U-Bahn, Germany	170	182	0.012670	2.141176	14.478455	31.817647	2.215540	0.079077	0.004000	0.037601	0.080229	0.079552	0.005488	1.400000
Bilbao Metro, Spain	48	47	0.041667	1.958333	11.775709	23.229167	2.018293	-0.016216	0.000000	0.029393	0.234255	0.250547	0.008369	1.000000
Brescia Metro, Italy	17	16	0.117647	1.882353	6.000000	12.235294	1.966667	-0.066667	0.000000	0.125490	0.333333	0.375000	0.034054	1.000000
Brussels Metro, Belgium	60	60	0.033898	2.000000	10.951412	21.783333	2.080065	0.027778	0.000000	0.079201	0.171576	0.182524	0.010227	1.283333
Bucharest Metro, Romania	47	48	0.044403	2.042553	7.278446	13.468085	2.113095	-0.067655	0.000000	0.040572	0.139521	0.151634	0.032201	1.382979
Budapest Metro, Hungary	48	48	0.042553	2.000000	7.662234	14.062500	2.146742	-0.142020	0.000000	0.023138	0.144831	0.159630	0.023686	1.166667
Catania Metro, Italy	11	10	0.181818	1.818182	3.745455	7.090909	2.000000	-0.333333	0.000000	0.215909	0.305051	0.374545	0.088744	1.000000
Charleroi, Belgium	46	49	0.047343	2.130435	10.282126	21.391304	2.825351	0.380088	0.023000	0.282904	0.210957	0.209839	0.010261	1.217391
Copenhagen Metro, Denmark	22	21	0.090909	1.909091	6.212121	11.500000	2.000000	0.000000	0.000000	0.098086	0.260606	0.295815	0.037742	1.000000
Frankfurt, Germany	98	108	0.022722	2.204082	12.813802	27.602041	2.511784	-0.029219	0.000000	0.108352	0.123060	0.118646	0.006217	1.530612
Genoa Metro, Italy	8	7	0.250000	1.750000	3.000000	5.500000	1.916667	-0.166667	0.000000	0.291667	0.333333	0.428571	0.152241	1.000000
Glasgow Subway, UK	15	14	0.133333	1.866667	5.333333	10.733333	1.961538	-0.076923	0.000000	0.143590	0.333333	0.380952	0.043705	1.000000

Table 15 continued

City, Country	V	E	ND	AD	AVP	VE	ANN	DA	CC	RC	BV	BE	AC	AVC
Hamburg U-Bahn, Germany	91	96	0.023443	2.109890	11.654212	23.725275	2.145789	-0.012417	0.000000	0.059386	0.119710	0.121398	0.009112	1.450549
Helsinki Metro, Finland	25	24	0.080000	1.920000	8.126667	16.480000	2.000000	0.000000	0.000000	0.085455	0.309855	0.338611	0.016836	1.000000
Lausanne, Switzerland	29	28	0.068966	1.931034	8.344828	16.896552	2.000000	-0.000000	0.000000	0.072944	0.272031	0.298030	0.016909	1.000000
Lille Metro, France	60	60	0.033898	2.000000	14.337853	31.500000	2.037037	-0.034483	0.000000	0.017565	0.229963	0.238964	0.004913	1.183333
Lisbon Metro, Portugal	49	51	0.043367	2.081633	7.802721	15.346939	2.496230	-0.164384	0.000000	0.081951	0.144739	0.152995	0.020441	1.408163
London Underground + DLR, UK	309	356	0.007481	2.304207	14.369794	28.142395	2.830357	0.126384	0.047000	0.035133	0.043550	0.040365	0.005547	1.543689
Lyon Metro, France	40	40	0.051282	2.000000	6.917949	13.350000	2.066667	-0.078431	0.000000	0.027972	0.155735	0.172949	0.029569	1.275000
Madrid Metro, Spain	276	308	0.008116	2.231884	16.902899	43.221014	2.746507	0.240821	0.007000	0.033384	0.058040	0.054880	0.002174	1.452899
Malaga, Spain	22	21	0.090909	1.909091	6.147186	12.272727	2.138889	-0.009615	0.000000	0.400673	0.257359	0.292723	0.031151	1.000000
Marseille Metro, France	28	28	0.074074	2.000000	6.026455	11.857143	2.090909	-0.076923	0.000000	0.040258	0.193325	0.215231	0.037839	1.250000
Milan Metro, Italy	106	109	0.019587	2.056604	12.175741	25.349057	2.074361	-0.129621	0.000000	0.010384	0.107459	0.111704	0.006967	1.301887
Munich U-Bahn, Germany	96	99	0.021711	2.062500	10.307675	19.093750	2.179342	-0.013765	0.000000	0.042071	0.099018	0.104118	0.012684	1.354167
Naples Metro, Italy	18	17	0.111111	1.888889	6.333333	13.000000	1.968750	-0.062500	0.000000	0.118056	0.333333	0.372549	0.030384	1.000000
Nuremberg U-Bahn, Germany	48	48	0.042553	2.000000	9.118794	18.229167	2.157346	0.032258	0.000000	0.064500	0.176496	0.189975	0.014905	1.104167
Oslo Metro, Norway	101	102	0.020198	2.019802	14.494059	28.069307	2.089286	0.070312	0.000000	0.041835	0.136304	0.142099	0.006010	1.158416

Table 15 continued

City, Country	V	E	ND	AD	AVP	VE	ANN	DA	CC	RC	BV	BE	AC	AVC
Palma de Mallorca, Spain	16	15	0.125000	1.875000	4.966667	9.812500	2.000000	0.000000	0.000000	0.139423	0.283333	0.331111	0.049940	1.000000
Paris Metro, France	303	358	0.007825	2.363036	12.020480	24.184818	2.688760	-0.011342	0.016000	0.010350	0.036613	0.033577	0.011286	1.617162
Porto, Portugal	80	80	0.025316	2.000000	15.050316	31.425000	2.148259	0.175824	0.000000	0.106587	0.180132	0.188129	0.004280	1.050000
Prague Metro, Czech Republic	58	58	0.035088	2.000000	9.505747	17.396552	2.158730	0.101684	0.000000	0.185243	0.151888	0.163892	0.016093	1.086207
Rennes Metro, France	15	14	0.133333	1.866667	5.333333	10.733333	1.961538	-0.076923	0.000000	0.143590	0.333333	0.380952	0.043705	1.000000
Rome Metro, Italy	73	72	0.027397	1.972603	13.755708	27.767123	2.015625	-0.016129	0.000000	0.019083	0.179658	0.191051	0.005582	1.000000
Rotterdam Metro, Netherlands	62	62	0.032787	2.000000	12.422528	23.758065	2.028302	-0.040268	0.000000	0.022834	0.190375	0.200363	0.007809	1.322581
Sevilla, Spain	22	21	0.090909	1.909091	7.666667	16.000000	1.975000	-0.050000	0.000000	0.095455	0.333333	0.365079	0.020357	1.000000
Sofia Metro, Bulgaria	34	34	0.060606	2.000000	7.490196	14.852941	2.034483	-0.044693	0.000000	0.042424	0.202819	0.220300	0.018117	1.441176
Stockholm Metro, Sweden	100	100	0.020202	2.000000	13.320808	26.600000	2.137262	-0.029240	0.000000	0.028589	0.125723	0.133208	0.005934	1.070000
Toulouse Metro, France	37	36	0.054054	1.945946	8.708709	16.945946	2.020833	-0.011236	0.000000	0.057330	0.220249	0.241909	0.016985	1.000000
Turin Metro, Italy	21	20	0.095238	1.904762	7.333333	15.238095	1.973684	-0.052632	0.000000	0.100251	0.333333	0.366667	0.022338	1.000000
Tyne and Wear Metro, UK	60	60	0.033898	2.000000	12.196045	24.100000	2.083726	-0.192053	0.000000	0.023421	0.193035	0.203267	0.007169	1.450000
Valencia, Spain	145	150	0.014368	2.068966	21.081609	47.103448	2.387719	0.250911	0.026000	0.074128	0.140431	0.140544	0.001858	1.193103
Vienna U-Bahn, Austria	98	104	0.021881	2.122449	10.670524	20.551020	2.347714	0.016296	0.000000	0.055980	0.100735	0.102601	0.012449	1.336735
Warsaw Metro, Poland	27	26	0.074074	1.925926	7.236467	15.000000	2.030303	-0.015625	0.000000	0.080515	0.249459	0.278326	0.021198	1.000000

Table 16 Time (seconds) for the evaluation of the topological measures of metro networks for European cities

City, Country	<i>ND</i>	<i>AD</i>	<i>AVP</i>	<i>VE</i>	<i>ANN</i>	<i>DA</i>	<i>CC</i>	<i>RC</i>	<i>BV</i>	<i>BE</i>	<i>AC</i>	<i>AVC</i>
Amsterdam Metro, Netherlands	0.000000	0.000994	0.006017	0.005998	0.001003	0.002000	0.013522	0.001003	0.017000	0.016991	0.031001	0.001000
Athens Metro, Greece	0.000000	0.001001	0.005169	0.004783	0.003015	0.001065	0.010000	0.000000	0.016046	0.017516	0.070009	0.000991
Barcelona Metro, Spain	0.000000	0.000000	0.021999	0.023073	0.003028	0.004057	0.008543	0.001425	0.084610	0.092601	0.081525	0.002004
Berlin U-Bahn, Germany	0.000000	0.000000	0.034006	0.040528	0.006002	0.005998	0.012998	0.002008	0.139882	0.175579	0.167473	0.002881
Bilbao Metro, Spain	0.000000	0.000000	0.003000	0.002000	0.000998	0.001004	0.008961	0.000000	0.009036	0.009000	0.028000	0.001002
Brescia Metro, Italy	0.000000	0.000000	0.000000	0.001000	0.000000	0.000000	0.007996	0.000000	0.001002	0.001999	0.009002	0.000000
Brussels Metro, Belgium	0.000000	0.000000	0.003000	0.002001	0.001002	0.000998	0.005997	0.000000	0.011990	0.011508	0.021037	0.000000
Bucharest Metro, Romania	0.000000	0.000000	0.004003	0.001997	0.001000	0.002004	0.012995	0.001001	0.008998	0.011004	0.123509	0.000000
Budapest Metro, Hungary	0.000000	0.000000	0.002999	0.002010	0.001035	0.001967	0.007031	0.000989	0.009011	0.011000	0.110496	0.001001
Catania Metro, Italy	0.000000	0.000000	0.000000	0.000000	0.001000	0.000000	0.005969	0.000998	0.000000	0.001032	0.005967	0.001033
Charleroi, Belgium	0.000000	0.000000	0.003013	0.001997	0.001000	0.001002	0.010500	0.000000	0.009050	0.011003	0.034958	0.001002
Copenhagen Metro, Denmark	0.000000	0.000000	0.000998	0.001000	0.000000	0.001001	0.012004	0.000000	0.002999	0.002001	0.028296	0.000996
Frankfurt, Germany	0.000000	0.000000	0.015997	0.012003	0.002002	0.001995	0.009002	0.000997	0.071554	0.077785	0.048991	0.001004
Genoa Metro, Italy	0.000000	0.000000	0.000000	0.000000	0.000000	0.001009	0.006995	0.000000	0.000000	0.000000	0.006030	0.000000
Glasgow Subway, UK	0.000000	0.000000	0.000000	0.000000	0.000000	0.000999	0.005999	0.000000	0.001006	0.000999	0.008001	0.000000
Hamburg U-Bahn, Germany	0.000000	0.000000	0.010002	0.009002	0.002004	0.001998	0.009001	0.000994	0.039006	0.053559	0.142538	0.000998
Helsinki Metro, Finland	0.000000	0.000000	0.001003	0.000000	0.001019	0.000000	0.008012	0.000000	0.001961	0.003007	0.008999	0.000000
Lausanne, Switzerland	0.000000	0.000000	0.003209	0.001992	0.000000	0.001000	0.013992	0.001003	0.005006	0.005990	0.019003	0.001001
Lille Metro, France	0.000000	0.000000	0.004004	0.003999	0.000970	0.000998	0.008040	0.000000	0.014000	0.014000	0.018624	0.001000
Lisbon Metro, Portugal	0.000000	0.001000	0.003999	0.001997	0.001000	0.002002	0.014012	0.000993	0.010003	0.011558	0.052962	0.001000

Table 16 continued

City, Country	<i>N</i> <i>D</i>	<i>A</i> <i>D</i>	<i>A</i> <i>V</i> <i>P</i>	<i>V</i> <i>E</i>	<i>A</i> <i>N</i> <i>N</i>	<i>D</i> <i>A</i>	<i>C</i> <i>C</i>	<i>R</i> <i>C</i>	<i>B</i> <i>V</i>	<i>B</i> <i>E</i>	<i>A</i> <i>C</i>	<i>A</i> <i>V</i> <i>C</i>
London Underground + DLR, UK	0.000000	0.000000	0.103987	0.101185	0.006008	0.005000	0.009036	0.001999	0.399284	0.704214	0.281174	0.003999
Lyon Metro, France	0.000961	0.000000	0.001000	0.002008	0.001038	0.000992	0.007000	0.000000	0.005998	0.007004	0.032980	0.001000
Madrid Metro, Spain	0.000000	0.000000	0.068048	0.064554	0.004174	0.003998	0.008008	0.001000	0.333636	0.308774	0.072404	0.002962
Malaga, Spain	0.000000	0.000000	0.000999	0.000000	0.000999	0.000000	0.007003	0.000000	0.001997	0.001964	0.013087	0.000000
Marseille Metro, France	0.000000	0.000000	0.001001	0.000000	0.001002	0.000965	0.005996	0.001002	0.001999	0.003001	0.022995	0.000000
Milan Metro, Italy	0.000000	0.000000	0.011006	0.010007	0.002564	0.001964	0.008999	0.001001	0.054532	0.066550	0.076280	0.001034
Munich U-Bahn, Germany	0.000000	0.000000	0.011999	0.012007	0.001995	0.001999	0.010005	0.001001	0.055312	0.054998	0.144193	0.002001
Naples Metro, Italy	0.000000	0.000000	0.001533	0.000461	0.000000	0.001017	0.011086	0.000000	0.002210	0.001015	0.014378	0.001021
Nuremberg U-Bahn, Germany	0.000000	0.000000	0.003411	0.003534	0.001516	0.001545	0.015476	0.000540	0.010677	0.012285	0.044012	0.001051
Oslo Metro, Norway	0.000000	0.000000	0.013012	0.020025	0.002997	0.002995	0.010535	0.000998	0.082476	0.066527	0.157169	0.001001
Palma de Mallorca, Spain	0.000566	0.000000	0.000000	0.000000	0.001009	0.001011	0.009362	0.000000	0.001089	0.001067	0.009296	0.000000
Paris Metro, France	0.000000	0.000000	0.084479	0.080675	0.004004	0.004591	0.006999	0.002005	0.671602	0.707328	0.484080	0.004000
Porto, Portugal	0.000000	0.001003	0.006960	0.007039	0.000955	0.001045	0.008996	0.000000	0.025002	0.028208	0.028603	0.001040
Prague Metro, Czech Republic	0.000000	0.000000	0.003002	0.003967	0.001034	0.000997	0.006001	0.000000	0.013049	0.015550	0.066404	0.000999
Rennes Metro, France	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.008998	0.000000	0.002002	0.000995	0.013005	0.000000
Rome Metro, Italy	0.000000	0.000000	0.008973	0.007998	0.003000	0.002024	0.008983	0.001000	0.027668	0.033414	0.055536	0.002006
Rotterdam Metro, Netherlands	0.000000	0.000000	0.005001	0.005997	0.001001	0.001000	0.010001	0.000000	0.023999	0.019534	0.082877	0.000000
Sevilla, Spain	0.000000	0.000000	0.000000	0.001003	0.000000	0.001003	0.007144	0.000000	0.002001	0.002001	0.007960	0.000000
Sofia Metro, Bulgaria	0.000000	0.000000	0.000998	0.000000	0.000999	0.000000	0.006000	0.000000	0.003000	0.004000	0.008002	0.001000

Table 16 continued

City, Country	<i>ND</i>	<i>AD</i>	<i>AVP</i>	<i>VE</i>	<i>ANN</i>	<i>DA</i>	<i>CC</i>	<i>RC</i>	<i>BV</i>	<i>BE</i>	<i>AC</i>	<i>AVC</i>
Stockholm Metro, Sweden	0.000000	0.000000	0.008965	0.008029	0.000969	0.002038	0.007197	0.000000	0.034051	0.038559	0.079265	0.001002
Toulouse Metro, France	0.000000	0.000000	0.002008	0.001992	0.000000	0.001000	0.008998	0.000000	0.007004	0.006003	0.039002	0.001001
Turin Metro, Italy	0.000000	0.000000	0.000999	0.001001	0.001016	0.000993	0.010008	0.000000	0.002997	0.003003	0.018119	0.001001
Tyne and Wear Metro, UK	0.000000	0.000000	0.007998	0.007005	0.001009	0.001000	0.008999	0.001000	0.017143	0.022534	0.021515	0.001005
Valencia, Spain	0.000000	0.000000	0.019082	0.022483	0.002996	0.002000	0.008000	0.002001	0.091573	0.108531	0.104833	0.003006
Vienna U-Bahn, Austria	0.000000	0.000997	0.009000	0.014997	0.001000	0.002006	0.006992	0.000000	0.042529	0.037029	0.071569	0.001003
Warsaw Metro, Poland	0.000000	0.000000	0.000998	0.000999	0.001002	0.001000	0.009099	0.000000	0.003001	0.003959	0.014195	0.001037

Table 17 Topological measures of road networks for European cities

City, Country	V	E	ND	AD	AVP	VE	ANN	DA	CC	RC	BV	BE	AC	AVC
Amsterdam, Netherlands	12,577	18,497	0.000234	2.941401	39.378503	67.615489	3.351498	0.198036	0.065000	0.001276	0.003052	0.002129	0.000304	1.857677
Barcelona, Spain	8833	13,818	0.000354	3.128722	38.536394	73.968867	3.314761	0.228239	0.049000	0.000782	0.004251	0.002789	0.000673	1.913846
Berlin, Germany	27,991	41,820	0.000107	2.988103	67.946565	134.206173	3.240147	0.133620	0.055000	0.000415	0.002392	0.001625	0.000122	1.863170
Big Athens Ring, Greece	27,504	45,098	0.000119	3.279378	84.909130	167.246800	3.443555	0.160967	0.034000	0.001795	0.003051	0.001883	0.000088	1.945608
Bilbao, Spain	2550	3611	0.001111	2.832157	27.530846	54.598824	3.175358	0.144757	0.078000	0.001399	0.010412	0.007624	0.001258	1.852549
Brescia, Italy	6706	8963	0.000399	2.673129	41.448647	78.724277	2.986898	0.045753	0.058000	0.002328	0.006034	0.004624	0.000329	1.745750
Brussels, Belgium	2286	3407	0.001304	2.980752	31.750002	64.994751	3.189089	0.165696	0.065000	0.030885	0.013463	0.009319	0.000414	1.926509
Bucharest, Romania	16,648	24,194	0.000175	2.906535	57.244066	113.576225	3.190907	0.143179	0.057000	0.000624	0.003379	0.002366	0.000190	1.869294
Budapest, Hungary	23,999	36,155	0.000126	3.013042	69.755331	140.691237	3.213403	0.168701	0.053000	0.000287	0.002865	0.001929	0.000110	1.889995
Catania, Italy	7673	10,768	0.000366	2.806725	48.972235	107.796299	3.109059	0.091306	0.080000	0.000435	0.006254	0.004548	0.000213	1.814675
Charleroi, Belgium	5257	7322	0.000530	2.785619	32.492524	62.764885	3.110573	0.074149	0.059000	0.002627	0.005993	0.004438	0.000621	1.838501
Copenhagen Municipality, Denmark	7005	10,183	0.000415	2.907352	51.593797	97.681799	3.378410	0.138623	0.061000	0.010568	0.007225	0.005067	0.000215	1.846966
Frankfurt, Germany	9469	13,669	0.000305	2.887105	40.207894	71.291794	3.129995	0.115645	0.073000	0.000629	0.004142	0.002942	0.000361	1.845285
Genoa, Italy	8963	10,813	0.000269	2.412808	45.980310	83.581502	3.002511	-0.031012	0.034000	0.002590	0.005020	0.004252	0.000165	1.572911
Glasgow City, UK	16,246	20,861	0.000158	2.568140	47.371723	94.032254	3.113136	0.062334	0.048000	0.000241	0.002855	0.002271	0.000142	1.665333

Table 17 continued

City, Country	V	E	ND	AD	AVP	VE	ANN	DA	CC	RC	BV	BE	AC	AVC
Hamburg, Germany	21,867	30,372	0.000127	2.777884	71.006658	144.648557	3.102478	0.119489	0.083000	0.000289	0.003202	0.002338	0.000116	1.803128
Helsinki, Finland	9617	13,114	0.000284	2.727254	49.888059	99.658001	3.115504	0.128433	0.102000	0.000939	0.005085	0.003804	0.000191	1.733389
Lausanne, Switzerland	1358	1859	0.002018	2.737850	28.240144	64.044919	3.013090	0.024674	0.078000	0.002492	0.020089	0.015191	0.000533	1.825479
Lille, France	3419	5090	0.000871	2.977479	38.698261	88.446037	3.146971	0.148565	0.071000	0.001581	0.011033	0.007603	0.000188	1.903481
Lisbon, Portugal	10,090	14,555	0.000286	2.885035	38.564306	73.061249	3.083847	0.151133	0.084000	0.000407	0.003724	0.002650	0.000369	1.865510
London, UK	127,540	160,010	0.000020	2.509174	134.143548	267.821640	3.036988	-0.014970	0.038000	0.000590	0.001044	0.000838	0.000018	1.659448
Lyon, France	4244	6576	0.000730	3.098963	35.511524	75.198398	3.208356	0.171200	0.063000	0.001162	0.008136	0.005400	0.000658	1.910226
Madrid, Spain	30,768	47,080	0.000099	3.060322	40.778483	75.050020	3.228753	0.191071	0.062000	0.001849	0.001293	0.000866	0.000213	1.926872
Malaga, Spain	10,952	15,850	0.000264	2.894449	35.367326	61.423850	3.091381	0.118723	0.055000	0.000929	0.003139	0.002231	0.000301	1.865778
Marseille, France	13,000	17,587	0.000208	2.705692	43.356149	104.849615	3.106465	0.059054	0.098000	0.001002	0.003259	0.002465	0.000231	1.765615
Milan, Italy	12,884	19,810	0.000239	3.075132	54.547232	105.689149	3.223614	0.177049	0.110000	0.000499	0.004157	0.002754	0.000332	1.933716
Munich, Germany	14,080	20,801	0.000210	2.954687	45.769784	85.598224	3.162331	0.120671	0.055000	0.000293	0.003180	0.002200	0.000220	1.876065
Naples, Italy	7826	10,897	0.000356	2.784820	35.668602	66.429083	3.082851	0.072177	0.068000	0.000722	0.004431	0.003273	0.000361	1.813953
Nuremberg, Germany	8067	10,825	0.000333	2.683773	43.600060	90.276187	3.104386	0.083331	0.070000	0.000680	0.005282	0.004028	0.000233	1.737201
Oslo, Norway	8247	11,215	0.000330	2.719777	38.112188	76.433612	3.135082	0.091428	0.090000	0.000534	0.004501	0.003398	0.000355	1.773493
Palma, Spain	7105	10,487	0.000416	2.952006	36.142755	74.922449	3.308668	0.177013	0.065000	0.001386	0.004948	0.003446	0.000285	1.854469
Paris, France	9605	14,966	0.000324	3.116294	39.397240	69.692972	3.257076	0.120109	0.062000	0.000502	0.003998	0.002632	0.000696	1.935867
Porto, Portugal	5204	7307	0.000540	2.808224	31.299655	57.268447	3.052890	0.067935	0.080000	0.001689	0.005825	0.004284	0.000542	1.843966

Table 17 continued

City, Country	V	E	ND	AD	AVP	VE	ANN	DA	CC	RC	BV	BE	AC	AVC
Prague, Czech Republic	21,503	29,804	0.000129	2.772078	50.227571	102.124494	3.152889	0.096263	0.061000	0.000429	0.002290	0.001685	0.000148	1.778915
Rennes, France	4225	5799	0.000650	2.745089	38.685475	74.668166	3.198164	0.138099	0.072000	0.009593	0.008924	0.006671	0.000565	1.781775
Rome, Italy	43,164	60,026	0.000064	2.781299	62.317524	132.395098	3.119074	0.094109	0.068000	0.000175	0.001421	0.001038	0.000066	1.801409
Rotterdam, Netherlands	11,302	16,228	0.000254	2.871704	42.217502	101.803398	3.200585	0.162599	0.060000	0.000861	0.003648	0.002602	0.000115	1.830649
Seville, Spain	10,219	15,296	0.000293	2.993639	38.755375	68.069283	3.109249	0.169285	0.063000	0.000927	0.003695	0.002534	0.000421	1.907134
Sofia, Bulgaria	11,544	16,980	0.000255	2.941788	45.898110	90.123008	3.171503	0.094564	0.036000	0.000460	0.003890	0.002703	0.000314	1.856462
Stockholm, Sweden	13,649	18,007	0.000193	2.638582	47.924038	91.940215	3.062187	0.066786	0.066000	0.000778	0.003438	0.002661	0.000057	1.710675
Toulouse, France	8867	12,442	0.000317	2.806361	37.352932	66.439495	3.073783	0.120726	0.094000	0.001212	0.004101	0.003002	0.000446	1.855193
Turin, Italy	11,307	17,784	0.000278	3.145662	47.372613	95.295215	3.357630	0.241318	0.058000	0.001575	0.004102	0.002664	0.000395	1.904838
Tyne and Wear, UK	46,098	54,690	0.000051	2.372771	62.502480	123.414226	2.953304	-0.077966	0.029000	0.001065	0.001334	0.001143	0.000041	1.520196
Valencia, Spain	8005	12,186	0.000380	3.044597	41.045532	96.708432	3.191113	0.179617	0.052000	0.001019	0.005004	0.003368	0.000255	1.906683
Vienna, Austria	16,083	23,891	0.000185	2.970963	50.355838	95.532488	3.315368	0.123420	0.067000	0.000549	0.003069	0.002108	0.000200	1.861655
Warsaw, Poland	19,202	26,552	0.000144	2.765545	49.564319	93.362306	3.116886	0.099272	0.076000	0.000422	0.002529	0.001867	0.000164	1.782210

Table 18 Time (seconds) for the evaluation of the topological measures of road networks for European cities

City, Country	ND	AD	AVP	VE	ANN	DA	CC	RC	BV	BE	AC	AVC
Amsterdam, Netherlands	0.008708	0.016168	208.997041	190.898394	0.193531	0.189099	0.007209	0.110595	899.224999	654.944430	8.321645	0.126936
Barcelona, Spain	0.005650	0.010539	96.894135	74.981631	0.134353	0.128374	0.007443	0.077048	506.827851	474.537499	5.449561	0.050524
Berlin, Germany	0.017715	0.030867	1096.320881	608.371375	0.204242	0.185225	0.004771	0.737006	3219.699749	3709.580312	45.075825	0.494536
Big Athens Ring, Greece	0.017349	0.031803	1054.899072	494.492131	0.246797	0.235258	0.004622	0.344825	2970.116192	3498.475964	28.184772	0.214928
Bilbao, Spain	0.001520	0.002972	7.274398	6.907903	0.045215	0.044393	0.007948	0.021677	31.101992	35.721923	2.228243	0.022635
Brescia, Italy	0.003879	0.007611	53.071468	43.988352	0.116731	0.114155	0.009479	0.062187	237.123640	275.940767	1.908161	0.210609
Brussels, Belgium	0.001266	0.002478	5.482430	4.765464	0.041526	0.039535	0.009186	0.022058	27.265302	30.341219	0.285780	0.027225
Bucharest, Romania	0.009729	0.018252	328.025847	345.830140	0.303786	0.288712	0.009050	0.605261	1243.700011	1092.888802	17.310471	0.140947
Budapest, Hungary	0.015785	0.029051	860.402676	454.353126	0.175331	0.156758	0.003479	0.093229	2272.008394	2493.980476	51.443081	0.175052
Catania, Italy	0.004563	0.008612	70.960357	67.603036	0.138810	0.137325	0.008250	0.073884	361.302687	369.906102	5.707481	0.092450
Charleroi, Belgium	0.002825	0.005302	30.991562	31.067167	0.046805	0.046012	0.005201	0.030689	143.257823	172.367118	0.836883	0.027593
Copenhagen Municipality, Denmark	0.003636	0.007411	55.263829	53.752108	0.122891	0.117856	0.006903	0.061547	265.570585	337.819414	1.722214	0.082325
Frankfurt, Germany	0.004880	0.009904	111.068365	77.924856	0.141696	0.130319	0.007261	0.073405	559.224195	480.117207	4.175030	0.055574
Genoa, Italy	0.004736	0.009516	87.444033	67.041272	0.162609	0.154603	0.007341	0.077603	493.426712	453.886482	2.104708	0.132991
Glasgow City, UK	0.008712	0.016193	321.862709	261.460036	0.115836	0.106714	0.003693	0.176726	1217.543967	973.596741	18.938349	0.101694

Table 18 continued

City, Country	<i>N D</i>	<i>A D</i>	<i>A V P</i>	<i>V E</i>	<i>A N N</i>	<i>D A</i>	<i>C C</i>	<i>R C</i>	<i>B V</i>	<i>B E</i>	<i>A C</i>	<i>A V C</i>
Hamburg, Germany	0.013363	0.025769	725.580783	444.859971	0.172730	0.159510	0.003934	0.087607	1882.973543	1949.351104	26.902015	0.216871
Helsinki, Finland	0.005319	0.010345	100.191947	87.017784	0.172706	0.165426	0.007910	0.087693	588.295027	461.689851	9.210443	0.159403
Lausanne, Switzerland	0.000827	0.001640	1.928913	1.283558	0.011511	0.010954	0.003842	0.005701	8.073826	9.988138	0.120324	0.008422
Lille, France	0.001370	0.003393	12.627719	11.068537	0.054377	0.052160	0.007748	0.030086	65.666574	65.335806	0.156618	0.019134
Lisbon, Portugal	0.005763	0.011266	132.845638	99.033283	0.182869	0.175744	0.008399	0.090106	665.705965	432.860622	5.774817	0.101403
London, UK	0.081546	0.159827	14426.115892	16045.692152	2.225395	2.197306	0.015759	2.351946	89389.737012	99169.296649	2127.071927	1.113487
Lyon, France	0.002034	0.004406	19.524251	17.363027	0.075222	0.085095	0.008744	0.039281	97.298160	98.570587	0.813995	0.049833
Madrid, Spain	0.012499	0.020972	1230.997426	629.426697	0.216481	0.215247	0.004756	0.968889	3925.684133	4429.026140	38.963468	0.663908
Malaga, Spain	0.006003	0.012249	141.492853	130.506136	0.098660	0.087803	0.004313	0.050275	747.080316	495.056826	42.324206	0.108422
Marseille, France	0.008642	0.016155	197.571037	176.909276	0.228706	0.217401	0.008120	0.112155	922.147568	674.932282	12.705823	0.118751
Milan, Italy	0.007582	0.014704	209.467841	189.445585	0.248847	0.239370	0.009419	0.308553	919.593745	764.411776	9.637183	0.293980
Munich, Germany	0.007735	0.015168	229.079582	232.401819	0.263145	0.258919	0.008992	0.140416	1044.598972	771.404656	9.846286	0.126827
Naples, Italy	0.005169	0.009813	69.397460	64.105090	0.136178	0.140355	0.010303	0.076749	386.399691	431.106463	3.950655	0.076470
Nuremberg, Germany	0.004626	0.009131	68.303878	70.242816	0.136035	0.128789	0.007595	0.069122	381.560755	475.594117	2.984282	0.113547
Oslo, Norway	0.005630	0.010577	77.186905	57.111298	0.069818	0.058566	0.003382	0.030088	428.585345	455.853397	1.658191	0.047389
Palma, Spain	0.004053	0.008090	51.359697	53.444024	0.135368	0.132307	0.008980	0.070615	300.788654	358.222173	1.286385	0.087270

Table 18 continued

City, Country	<i>N</i> <i>D</i>	<i>A</i> <i>D</i>	<i>A</i> <i>V</i> <i>P</i>	<i>V</i> <i>E</i>	<i>A</i> <i>N</i> <i>N</i>	<i>D</i> <i>A</i>	<i>C</i> <i>C</i>	<i>R</i> <i>C</i>	<i>B</i> <i>V</i>	<i>B</i> <i>E</i>	<i>A</i> <i>C</i>	<i>A</i> <i>V</i> <i>C</i>
Paris, France	0.005464	0.010834	100.602912	74.981781	0.176433	0.169974	0.008737	0.336858	635.516402	472.253089	1.833786	0.053708
Porto, Portugal	0.002905	0.005504	30.198572	29.665719	0.092149	0.088463	0.007974	0.043253	143.279550	173.624362	0.333803	0.028542
Prague, Czech Republic	0.013257	0.025728	630.092800	512.002287	0.181224	0.159014	0.003815	0.096364	1849.581723	1927.207649	27.475664	0.414062
Rennes, France	0.001943	0.003255	18.398222	18.360547	0.074728	0.069203	0.007860	0.038912	99.698286	98.033360	1.123189	0.146416
Rome, Italy	0.030118	0.055262	1837.553791	1143.290766	0.298524	0.281754	0.003947	0.152936	8115.247589	10064.383087	1388.824745	0.399801
Rotterdam, Netherlands	0.007380	0.013766	139.228126	126.406795	0.199255	0.197546	0.008420	0.104872	807.154365	502.890889	6.160848	0.119808
Seville, Spain	0.005354	0.009730	125.556233	105.806228	0.180711	0.169387	0.008293	0.087249	638.344608	482.315488	12.070543	0.219362
Sofia, Bulgaria	0.016997	0.012414	149.960999	136.126625	0.200455	0.191835	0.008368	0.100143	821.974195	544.606920	9.977724	0.118126
Stockholm, Sweden	0.008485	0.016507	241.833753	250.934736	0.246878	0.197184	0.008822	0.455692	929.589701	728.274954	3.828852	0.119949
Toulouse, France	0.004864	0.009904	99.419466	70.555064	0.159793	0.153936	0.008477	0.084128	475.553876	465.424704	2.071956	0.051537
Turin, Italy	0.007217	0.013630	141.821731	151.648244	0.205652	0.202150	0.009111	0.116478	808.815130	476.703016	12.039782	0.116632
Tyne and Wear, UK	0.029904	0.058193	2139.861492	1237.870544	0.317945	0.285757	0.004616	1.314275	9240.494702	11716.502484	79.690288	0.725430
Valencia, Spain	0.003953	0.008310	66.079262	65.093321	0.142270	0.134132	0.008582	0.070899	373.529783	435.697695	1.726928	0.111545
Vienna, Austria	0.006057	0.011969	333.969349	309.329290	0.251905	0.237454	0.006827	0.532720	1249.865738	1030.323555	6.166231	0.139557
Warsaw, Poland	0.011515	0.022370	431.051147	481.924582	0.351275	0.334576	0.008214	0.177704	1493.024513	1489.674855	13.039317	0.160354

Author Contributions R.M.: conceptualization, formal analysis, methodology, software, writing—original draft preparation, writing—review and editing. M.S.N.: conceptualization, supervision, writing—original draft preparation, writing—review and editing. M.M.: supervision, writing—review and editing. S.P.: supervision, writing—review and editing.

Funding Open Access funding enabled and organized by Projekt DEAL. We acknowledge financial support by Universität der Bundeswehr München. Rudy Milani is funded by dtec.bw—Digitalization and Technology Research Center of the Bundeswehr project RISK.twin. dtec.bw is funded by the European Union—NextGenerationEU.

Data Availability Data available on request from the authors.

Declarations

Competing Interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A (2004) The architecture of complex weighted networks. *Proc Natl Acad Sci United States Am* 101(11):3747–3752. <https://doi.org/10.1073/pnas.0400087101>
2. Bounova G, De Weck O (2012) Overview of metrics and their correlation patterns for multiple-metric topology analysis on heterogeneous graph ensembles. *Phys Rev E* 85(1):016117
3. Borgatti SP, Everett MG (2000) Models of core/periphery structures. *Soc Netw* 21(4):375–395
4. Boeing G (2017) OSMnx: new methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Comput Environ Urban Syst* 65:126–139
5. Colizza V, Flammini A, Serrano MA, Vespignani A (2006) Detecting rich-club ordering in complex networks. *Nat Phys* 2(2):110
6. Dehmer M, Emmert-Streib F, Pickl S, Holzinger A (2016) Big data of complex networks. Chapman and Hall/CRC. <https://doi.org/10.1201/9781315370736>
7. Fiedler M (1973) Algebraic connectivity of graphs. *Czechoslovak Math J* 23(2):298–305
8. García-Robledo A, Díaz-Pérez A, Morales-Luna G (2013) Correlation analysis of complex network metrics on the topology of the internet. In: 2013 10th International conference and expo on emerging technologies for a smarter world (cewit). IEEE, pp 1–6
9. Guyon I, Weston J, Barnhill S, Vapnik V (2002) Gene selection for cancer classification using support vector machines. *Mach Learn* 46:389–422
10. Harary F (1969) Graph theory. <https://doi.org/10.21236/ad0705364>
11. Holme P, Kim BJ, Yoon CN, Han SK (2002) Attack vulnerability of complex networks. *Phys Rev E* 65(5):056109
12. Jamakovic A, Uhlig S (2008) On the relationships between topological measures in real-world networks. *Netw Heterogen Media* 3(2):345–359. <https://doi.org/10.3934/nhm.2008.3.345>
13. Kogotkova S, Oehlers M, Ermakova T, Fabian B (2018) Correlation analysis of local graph metrics. Available at SSRN 3192028
14. Li L, Alderson D, Willinger W, Doyle J (2004) A first-principles approach to understanding the internet's router-level topology. *ACM SIGCOMM Comput Commun Rev* 34(4):3–14

15. Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst* 30
16. Li C, Wang H, De Haan W, Stam C, Van Mieghem P (2011) The correlation of metrics in complex networks with applications in functional brain networks. *J Stat Mech Theory Exper* 2011(11):11018
17. Madole JW, Buchanan CR, Rhemtulla M, Ritchie SJ, Bastin ME, Deary IJ, Cox SR, Tucker-Drob EM (2023) Strong intercorrelations among global graph-theoretic indices of structural connectivity in the human brain. *NeuroImage* 275:120160
18. Mahadevan P, Krioukov D, Fomenkov M, Dimitropoulos X, Claffy K, Vahdat A (2006) The internet as-level topology: three data sources and one definitive metric. *ACM SIGCOMM Comput Commun Rev* 36(1):17–26
19. Mahadevan P, Krioukov D, Fall K, Vahdat A (2006) Systematic topology analysis and generation using degree correlations. *ACM SIGCOMM Comput Commun Rev* 36(4):135–146
20. Newman MEJ (2002) Assortative mixing in networks. *Phys Rev Lett* 89(20):208701
21. Newman MEJ (2003) Mixing patterns in networks. *Phys Rev E* 67(2):026126
22. Nistor MS, Pickl S, Raap M, Zsifkovits M (2019) Network efficiency and vulnerability analysis using the flow-weighted efficiency measure. *Int Trans Oper Res* 26(2):577–588
23. Pearson K (1901) LIII. On lines and planes of closest fit to systems of points in space. *London Edinburgh Dublin Philos Mag J Sci* 2(11):559–572
24. Pastor-Satorras R, Vázquez A, Vespignani A (2001) Dynamical and correlation properties of the internet. *Phys Rev Lett* 87(25):258701
25. Ratner B (2009) The correlation coefficient: its values range between $+1/-1$, or do they? *J Target Meas Anal Market* 17(2):139–142
26. Strogatz SH (2001) Exploring complex networks. *Nature* 410(6825):268–276
27. West DB (1996) *Introduction to graph theory*. Prentiss Hall, Upper Saddle River, NJ
28. Wasserman S, Faust K (1994) *Social network analysis: methods and applications*. Cambridge University Press, Cambridge; New York
29. Watts DJ, Strogatz SH (1998) Collective dynamics of ‘small-world’ networks. *Nature* 393(6684):440–442. <https://doi.org/10.1038/30918>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.