



Road networks structure analysis: A preliminary network science-based approach

Selim Reza¹ · Marta Campos Ferreira¹ · J.J.M. Machado² ·
João Manuel R.S. Tavares²

Accepted: 17 September 2022 / Published online: 29 September 2022
© Springer Nature Switzerland AG 2022

Abstract

Road network studies attracted unprecedented and overwhelming interest in recent years due to the clear relationship between human existence and city evolution. Current studies cover many aspects of a road network, for example, road feature extraction from video/image data, road map generalisation, traffic simulation, optimisation of optimal route finding problems, and traffic state prediction. However, analysing road networks as a complex graph is a field to explore. This study presents comparative studies on the Porto, in Portugal, road network sections, mainly of Matosinhos, Paranhos, and Maia municipalities, regarding degree distributions, clustering coefficients, centrality measures, connected components, k-nearest neighbours, and shortest paths. Further insights into the networks took into account the community structures, page rank, and small-world analysis. The results show that the information exchange efficiency of Matosinhos is 0.8, which is 10 and 12.8% more significant than that of the Maia and Paranhos networks, respectively. Other findings stated are: (1) the studied road networks are very accessible and densely linked; (2) they are small-world in nature, with an average length of the shortest pathways between any two roads of 29.17 units, which as found in the scenario of the Maia road network; and (3) the most critical intersections of the studied network are 'Avenida da Boavista, 4100-119 Porto (latitude: 41.157944, longitude: -8.629105)', and 'Autoestrada do Norte, Porto (latitude: 41.1687869, longitude: -8.6400656)', based on the analysis of centrality measures.

Keywords Complex network analysis · Degree centrality · Closeness centrality · Betweenness centrality · Eigenvector centrality · Power-law distribution · Community detection

1 Introduction

Watts and Strogatz [1], and Barabasi and Albert [2], are among the first to reveal the collective dynamics of small-world and scale-free networks, allowing the complex network analysis to gain importance as a new multidisciplinary research direction in network science. The analysis of road networks as a complex graph arises from this foundation.

✉ João Manuel R.S. Tavares
tavares@fe.up.pt

Extended author information available on the last page of the article

Roads are commonly spatial networks with nodes and edges embedded in space due to their geographical features. The road network model construction considers that the nodes represent intersections, and the edges represent the road sections that directly connect two intersections. Gao et al. [3], discovered strong signatures in these networks, typically real physical entities connecting points in geographic space. Current metropolitan zones as well newly proposed urban expansion areas can benefit from modelling and analysing their road networks as a complex graph. This approach focuses on understanding, identifying, visualising, and exploring the networks. The term complex refers to the fact that it is impossible to predict collective behaviour from individual components. Understanding the overall network makes it capable of predicting and possibly controlling the behaviours such as traffic congestion spreading, speed, travel times, trip fares, timings, and headway (Ahmadzai et al. [4]).

The study of a road network's structural characteristics aids in understanding its emergent behaviour. The statistical qualities of a street network, for example, can be used as proxies for a city's growth history, ranging from self-organised to highly planned metropolitan regions (Duan et al. [5]). Traffic congestion is another intriguing emergent phenomenon arising from vehicular network interaction. Interestingly, it can also be analysed using graph theory: for example, the centrality metrics of a street network are highly dependent on vehicular density. Therefore, recognising critical structural locations is vital for planning and implementing road infrastructure solutions like inserting overpasses or roundabouts to maintain a smooth traffic flow (Davidovic et al. [6]).

In this study, the road network of Porto, in Portugal, was analysed in terms of the degree distribution, power-law fitting and centrality measures: degree, closeness, betweenness, and Eigenvector centrality. It also includes a comprehensive exploration of community detection and page rank analysis. The results demonstrate beneficial characteristics of the studied networks, mainly its high reachability and fluctuation from following the power-law distribution, unlike other real-world networks, such as the protein-protein interaction networks (Khojasteh et al. [7]). In addition, this analysis allows to discover the most critical intersections of a network and distinguish the groups of highly connected components compared to others. The outcomes of this study might be essential for city planners and researchers to acquire a deep insight into their domains.

This article is organised as follows: Section 2 presents a summary of related state-of-the-art researches along with their limitations and future scopes; Section 3 introduces the formulation of the critical adopted concepts; Section 4 presents the experimental setup and results; Section 5 is devoted to discussing the overall performance of the proposed approach, and finally, Section 6 draws the conclusions.

2 Related works

This section summarises related state-of-the-art works and their limitations and future potential. The literature usually analyses road networks based on centrality measures and community detection. Although complex network analysis consists of a practical tool to extract hidden relationships between the lanes and intersections of a road network, few works were found in the literature. However, the proposed work might pave the way for attracting more importance to this topic.

Montis et al. [8], employed a weighted network model to investigate the structure of the road network that represents interurban commuting traffic in the Sardinia region, in

Italy. The authors quantitatively analysed the topological and weighted properties of the resulting network and the interplay between topological and dynamical aspects. The analysis also included the socio-demographic characteristics such as population and monthly income.

Masucci et al. [9], studied the growth of London's street network, in UK, in its dual representation. The results revealed that logistic laws could analytically describe the growth of the network. Moreover, the robust log-normal distributions govern the network's properties, characterising the network's connectivity and consistent small-world properties over time.

A similar kind of approach was adopted by Zheng et al. [10], to analyse the topological properties of the Beijing, in China, public transport network. The results demonstrated that the both the node strength and cumulative strength follow the power-law distribution, and revealed the characteristics of scale-free and small-world phenomena. Although the node strength refers to how strongly a node is connected to the others, it can only provide information about the importance of a node. And hence, the power-law fitting of the degree distribution should also be analysed. The proposed work focuses on addressing this drawback.

Sardonis et al. [11], applied centrality analysis to identify critical traffic jam areas of a road network. To assess variance in centrality values when a network's structure changes by removing or adding individual nodes, the authors developed the concepts of centrality interference and centrality resilience. However, community detection should also occur to further verify correlations between different nodes. Yang et al. [12], aimed at solving these shortcomings by proposing an improved K-means algorithm to identify the spatial correlation patterns of traffic states in a road network in Beijing. A similar approach was presented by Hong et al. [13], but with a different algorithm called hierarchical clustering models: Infomap algorithm. The results suggested that community distribution follows the urban spatial interaction within suburbs and urban centres, particularly in Guangzhou, in China.

Feng et al. [14], used network science to conduct a statistical analysis to assess the city's size's effects on its structure and quantify structural characteristics. The findings confirmed statistical laws in organising urban spatial components and looked into the relationship between topological changes in urban networks. Total et al. [15], analysed transportation networks as a complex graph revealing the relationship between several socioeconomic frameworks and their usability as economic performance indicators. However, a more in-depth analysis needs to be performed in community detection or centrality measures separately and based on their aggregated analysis.

3 Methodology

This section describes the employed techniques' formulation and discusses their scopes and working principles.

3.1 Centrality measures

In a network, the nodes' centrality measures their importance. There are various categories of centrality measures throughout the literature in network science. The four most widely applied centrality measures are: degree centrality, closeness centrality, betweenness centrality, and Eigenvector centrality (Borgatti [16]; Newman [17]; Otte and Rousseau [18]).

3.1.1 Degree centrality

The most natural and computationally efficient way to convey the relevance of a specific node is to use degree centrality (Borgatti [16]), defined as the number of edges that a node connects, frequently used in a variety of networks, including biological and social network mapping (Otte and Rousseau [18]; Bodendorf and Kaiser [19]). This concept applies to urban transportation studies to evaluate road networks (Crucitti et al. [20]), and simulate traffic flow characteristics (Jayasinghe et al. [21]).

Let $G(V,E)$ be a graph, where V and E represent nodes and edges, respectively. A single/isolated node and edge can be represented by v or v_i and e or e_i , respectively, with $i = 1, 2, 3, \dots, N$, where N is the number of nodes. (Throughout the text, these notations are followed.) Hence, the degree of centrality of a v node is determined as:

$$C_D(v) = \text{deg}(v). \quad (1)$$

Now, the degree centrality of a whole graph can be formulated considering the node with highest degree centrality. Hence, if v^* represents the node with the highest value of degree centrality, then, the degree centrality of the entire graph can be calculated as (Freeman [22]):

$$C_D(G) = \frac{\sum_{i=1}^N [C_D(v^*) - C_D(v_i)]}{N^2 - 3N + 2}. \quad (2)$$

3.1.2 Closeness centrality

The concept of farness can be used to explain closeness centrality. For a given network node, the summation of the lengths of the shortest pathways from that node to all others is called the farness (Borgatti [16]). The reciprocal of farness is, in fact, closeness centrality, and the greater a node's value of closeness centrality, the more likely it is to be closer and central to the rest of the nodes in the network. As in the case of degree centrality, it is widely used in biological, economic and social network analysis (Chea and Livesay [23]; Newman and Mark [17]; Porta et al. [24]). However, an attempt to use it in order to explain urban road networks and capture their various elements using closeness centrality was made by Crucitti et al. [20], but with less success.

The base of the mathematical formulation of closeness is the average shortest path length between a specific node and all other nodes in the network. Suppose N is the total number of nodes of G graph; then, in terms of mathematics, for a v_i node, the closeness centrality is formulated as (Sabidussi [25]):

$$C_C(v_i) = \frac{1}{\sum_{j=1}^N d(v_i, v_j)}. \quad (3)$$

where $i \neq j$ and $d(v_i, v_j)$ are the shortest distances between v_i and v_j nodes, respectively.

3.1.3 Betweenness centrality

The percentage of shortest paths that cross through a particular node of interest is known as betweenness centrality. It indicates how close a node is to other node pairs in the network under analysis. A node with a high betweenness centrality value

demonstrates that it is easily reachable to other nodes by the shortest paths or lies on many shortest paths. This concept applies in a variety of applications across several domains. For example, researchers are using it to understand better land-use intensity in different cities across the United States and Europe, and also to disclose the driving factors of the evolution of Paris' urban fabric, in France, (Porta et al. [24]; Barthelemy et al. [26]). However, Gao et al. [3], have discovered that betweenness centrality is not a good predictor of urban traffic flows in transportation-related studies.

The mathematical formulation of the betweenness centrality is as follows. If $g_{v_i v_j}$ are the number of shortest paths connecting v_i and v_j nodes, and $g_{v_i v_j}(v)$ is the number that v node is on, then the betweenness centrality is formulated as (Brandes [27]):

$$C_B(v) = \sum_{v_i \neq v_j \neq v} \frac{g_{v_i v_j}(v)}{g_{v_i v_j}}. \quad (4)$$

3.1.4 Eigenvector centrality

Eigenvector centrality is a measure of a node's importance that considers the relevance of its neighbours. For example, a node with x unpopular Facebook friends would have a lower Eigenvector centrality than a node with x famous Facebook friends. The algorithm uses a power iteration strategy to calculate the Eigenvalue. The scores of its incoming neighbours within each interaction allow obtaining the centrality score. The Eigenvector is L2-normalised after each iteration in the power iteration approach, resulting in normalised findings by default (Hansen [28]).

If λ is a constant and $A_{v_i v_j}$ is the adjacency matrix, its value is 1 (one) if v_i and v_j nodes are connected, otherwise, is 0 (zero); then, the Eigenvector centrality can be formulated as (Newman [17]):

$$C_E(v_i) = \frac{1}{\lambda} \sum_{v_j \in G} A_{v_i v_j} C_E(v_j). \quad (5)$$

where $C_E(v_j)$ is the Eigenvector centrality of node v_j .

There will be a non-zero Eigenvector solution for many different λ eigenvalues. On the other hand, the prerequisite condition that all entries in the Eigenvector have to be non-negative means that only the highest Eigenvalue yields the necessary centrality score.

In essence, the four types of centrality aforementioned describe the network characteristic of a given node as follows:

- Degree centrality: the larger the value, the more edges the node connects;
- Closeness centrality: the larger the value, the more central and closer to other nodes in a network a given node is;
- Betweenness centrality: the larger the value, the shortest paths amongst different pairs of nodes have to pass through the node;
- Eigenvector centrality: A high eigenvector score means that a node connects to many nodes with high scores.

3.2 Average nearest neighbour degree

The average nearest neighbour degree (Xia et al. [29]), which can be abbreviated as knn , has a substantial effect on whether a given node likes to link to others who likewise have a large number of connections:

$$Kmn_{v_i} = \frac{1}{N(v_i)} \sum_{v_j \in N(v_i)} K(v_j). \quad (6)$$

Hence, it states that to calculate knn of v_i node, it is necessary to add the degrees of each of their v_j neighbours, $K(v_j)$, and divide it by the neighbours of v_i node, $N(v_i)$. On the other hand, in case of a weighted graph, if $K^w(v_i)$ is the weighted degree of v_i node and $w_{v_i v_j}$ represents the weight of the edge linking v_i and v_j , then, the node's average nearest neighbour degree can be calculated as (Yao et al. [30] & Barrat et al. [31]):

$$Kmn_{v_i} = \frac{1}{K^w(v_i)} \sum_{v_j} w_{v_i v_j} K(v_j). \quad (7)$$

3.3 Community

A community corresponds to a subset of nodes in a graph where connections between nodes are denser than connections with the remaining of the network; i.e., groups of nodes that are more connected to each other than the rest of the network (Yang et al. [32]). The critical feature essential to extract relevant information from networks is community detection. Some vital aspects of community detection are:

- It permits nodes' functions to be classified based on their structural placements in their communities;
- It demonstrates the hierarchical organisation generally found in real-world networks;
- It boosts the speed and efficiency of network data storage, including its processing and analysis (Reddy et al. [33]).

The fact that no standard description of community structure exists (Fortunato and Hric [34]) poses the most significant obstacle in community detection. As a result, detecting communities in large-scale networks is computationally infeasible.

3.3.1 Girvan-Newman algorithm

By gradually eliminating edges from the original network, the Girvan–Newman algorithm finds communities. The communities are the remaining network's connected components. The Girvan–Newman algorithm, rather than attempting to develop a metric that identifies which edges are most vital to communities, concentrates on edges that are most probable "between" communities (Girvan and Newman [35]).

The Girvan-Newman algorithm follows four general steps:

- Computing the edge betweenness centrality for every edge in the graph;

- Deleting the edge with the highest value of betweenness centrality;
- Again, calculating the edge betweenness centrality for every leftover edge;
- Repeating the first three steps until there are no more edges.

3.3.2 Modularity

Modularity refers to how densely connected components can separate into unique communities or clusters that interact more than others. By limiting the transmission of perturbations through the web, the modularity found in empirical data bounds to enhance the network stability and robustness (Massol et al. [36]).

By considering a network with V nodes and E edges that is divided into n_c communities, each with V_c nodes connected by E_c edges and $c = 1, \dots, n_c$ communities, the partition's modularity over all n_c communities is calculated as (Clauset et al. [37]):

$$Modularity = \sum_{c=1}^{n_c} \left[\frac{E_c}{E} - \left(\frac{D_c}{2E} \right)^2 \right], \quad (8)$$

where D_c represents the sum of degrees of V nodes in c community.

3.3.3 Louvain method

The Louvain method is a hierarchical clustering methodology that iteratively integrates communities into a single node and performs modularity clustering on condensed graphs to detect communities in large networks. It maximises each community's modularity score, which measures the assigned nodes to communities, entailing to calculate how much more densely connected a community is than a random network (Lu et al. [38]).

It is a simple, efficient, and easy way to locate communities in large networks, for example, a few hundred million nodes to billions of edges. Hence, it aims to identify community hierarchies and allow for zooming into communities to find sub-communities and even the sub-sub-communities (Blonde et al. [39]).

The algorithm consists of two general steps. The first one allocates each node to its community and then tries to discover its highest positive modularity gain by transferring each node to its adjacent communities in the second step. The node returns to its original community if no positive gain is acquired. The gain in modularity, ΔQ provided by transferring an isolated v node into a c community can be estimated as (Blondel et al. [39]):

$$\Delta Q = \frac{W_{v,in}}{2s} - \gamma \frac{\sum_{tot} \cdot W_v}{2s^2}. \quad (9)$$

where the graph's size is s , $W_{v,in}$ is the total weights of the connections from v node to others in c , and the sum of the weights of the links incident to v node is denoted by W_v , and γ and \sum_{tot} represent the weights' links and the sum of the weights of the links incident to nodes in c , respectively.

For the direct network case, the modularity gain is:

$$\Delta Q = \frac{W_{v,in}}{s} - \gamma \frac{W_v^{out} \cdot \sum_{tot}^{in} + W_v^{in} \cdot \sum_{tot}^{out}}{s^2}, \quad (10)$$

where W_v^{out} and W_v^{in} are the outer and inner weighted degrees of v node and Σ_{tot}^{in} , Σ_{tot}^{out} are the sum of in-going and out-going links incident to nodes in c , respectively, (Dugue and Perez [40]).

The first step will continue until no single step improves the modularity. The second step entails creating a new network, the communities from the first step. By adding the weights of the links between nodes in the two communities, it is possible to determine the weights of the links connecting the new nodes. After completion, the replication of the first step can create large communities with more modularity. Repetition of the preceding two rounds will occur to reach no modularity gain. The output is a list of sets of the partition of $G(V,E)$ graph, where each set suggests a community and its constituent nodes.

4 Experiments

The experiments were conducted on a local device with an Intel® Core™ i7-10750H processor at 2.6 GHz and 16GB of RAM and without GPU support. The developed code uses the Python NetworkX package (Hagberg et al. [41]). The extraction of the Porto road network real-world data was performed using the Python OSMnx module (Boeing [42]). The studied network has a total of 17851 nodes and 40091 edges. The comprehensive Comparison used three municipalities of the entire network: Matosinhos (11961 nodes and 28459 edges), Paranhos (1926 nodes and 4508 edges), and Maia (1776 nodes and 4145 edges). Figure 1 illustrates the networks of the three studied regions, where the intersections are represented in white, while the lanes are in red.

4.1 Results

This section provides this research's findings and outcomes. The first part consists of presenting the network's degree distribution study findings. Then, results are obtained based on page rank, and average path distance is displayed, followed by centrality measurements and community detection findings.

4.1.1 Degree distribution

The degree of a node within a network corresponds to the number of connections it has to other nodes, and the distribution degree is the probability distribution of such degrees



Fig. 1 Visualization of the three studied regions of Porto: (on the left) Matosinhos, (on the middle) Paranhos and (on the right) Maia (intersections and lanes are represented in white and red, respectively)

over the entire network. The network has a binomial distribution of k degrees since each of n nodes is individually linked with p probability:

$$P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}. \quad (11)$$

Figure 2 depicts the degree distribution of the Porto road network as a histogram and in regular and log scales. The maximum and minimum degrees of the network are 9 and 1 (one), respectively, with a mean value of 2.83. Nonetheless, for Matosinhos, the mean value was 4.76 with maximum and minimum values of 12 and 1, respectively.

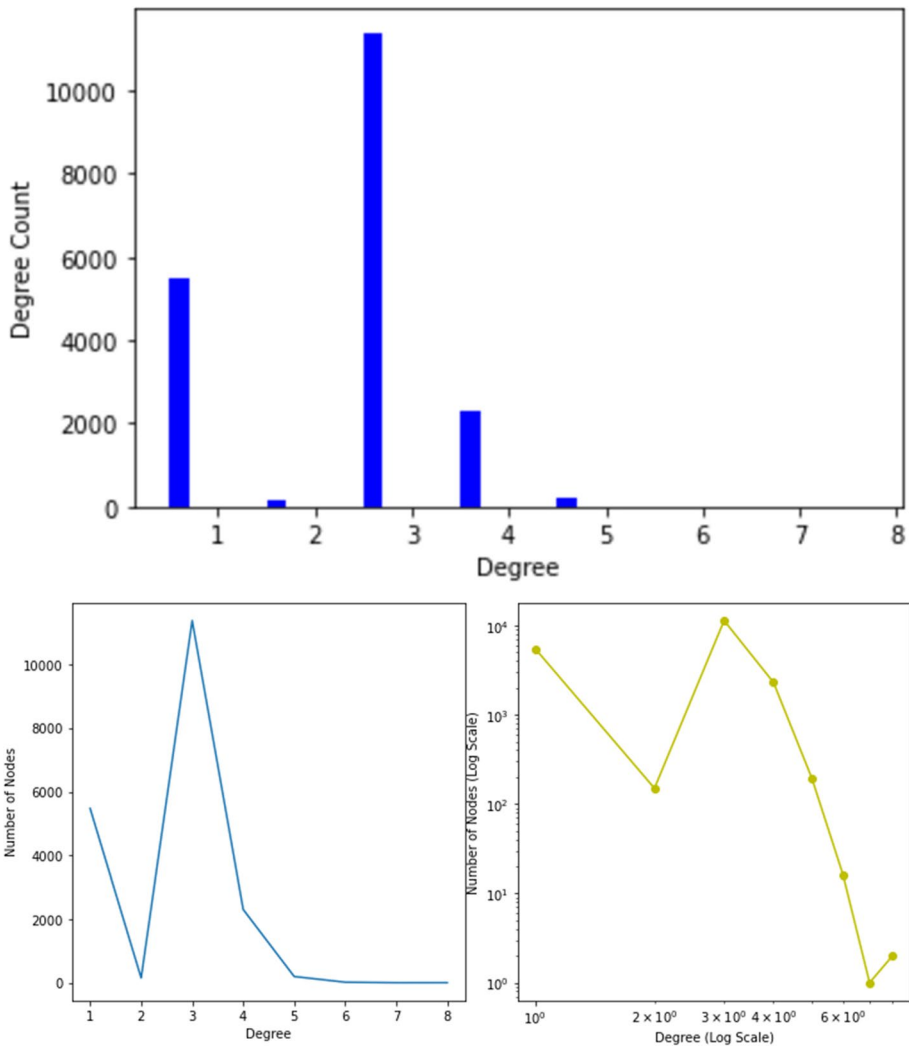


Fig. 2 Degree distribution of Porto road network as (on the top) a histogram and also according to (on the bottom-left) both normal and (on the bottom-right) log scales

Contradictorily, for Paranhos and Maia, the maximum and minimum values were 1, 10, and 2, 8 with a mean of 4.68 and 4.73, respectively.

The power-law fitting of the network was also analysed. Figure 3 illustrates the power-law fit of the degree distribution compared with Log-normal and Stretched-exponential distribution. The calculated best minimal value for power-law fit was -55.32 , which further indicates the rarity of full power low fit of the real-world transportation network, as can be seen by the dotted red line in the figure. The degree distribution analysis demonstrates that only a few intersections possess a high number of connections compared to the others. Also, the power-law analysis depicts that the road networks are not scale-free and should not be expected to follow such distribution.

4.1.2 Average neighbour degree

Calculating the average nearest neighbour degree allows for evaluating whether a given node prefers to connect to others with many connections. The results showed the value of the average nearest neighbour degree of the road network of Maia to be 2.6 and 7.5% higher than that of Matosinhos and Paranhos (Figs. 4 and 5), respectively.

It means that the intersections of Maia are more dependent on their neighbouring counterparts than in the other studied networks.

The box-plot analysis of the average nearest neighbour degree of the Paranhos road network follows a normal distribution, and the other two possess either positive or negative skews. In addition, many outliers for the Matosinhos road network depict a large numerical fluctuation compared to the values within the inter-quartile range.

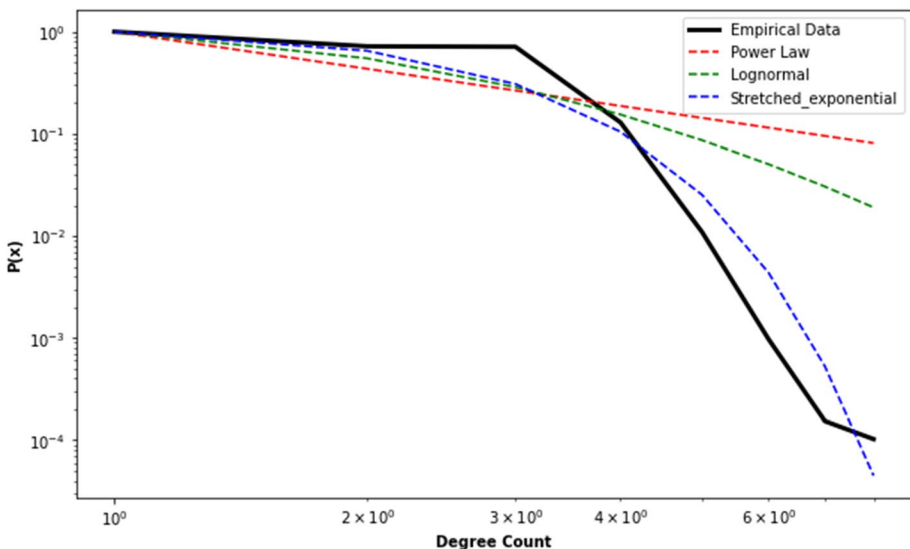


Fig. 3 Degree distribution of Porto road network as a power-law fit according to Log-normal and Stretched-exponential distributions

Fig. 4 Box plot of the average neighbor degrees of the road network of Matosinhos, Paranhos and Maia

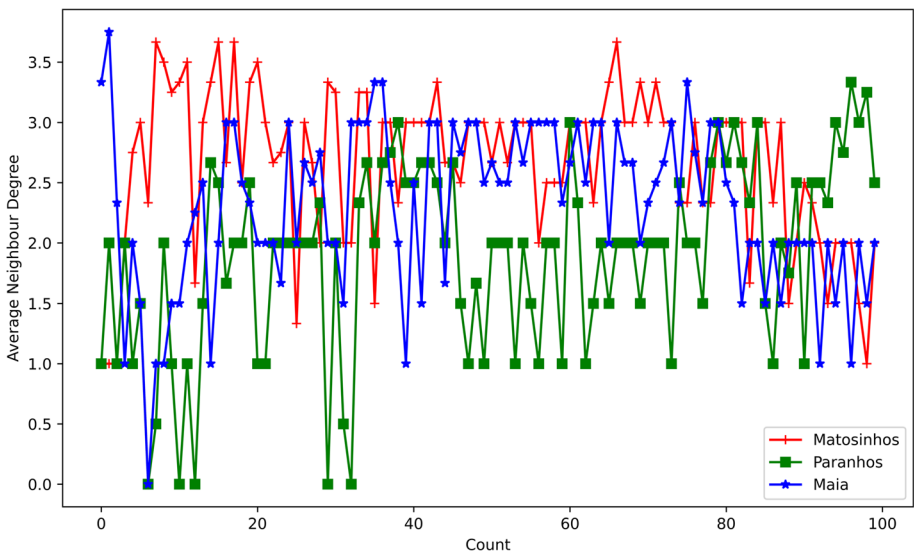
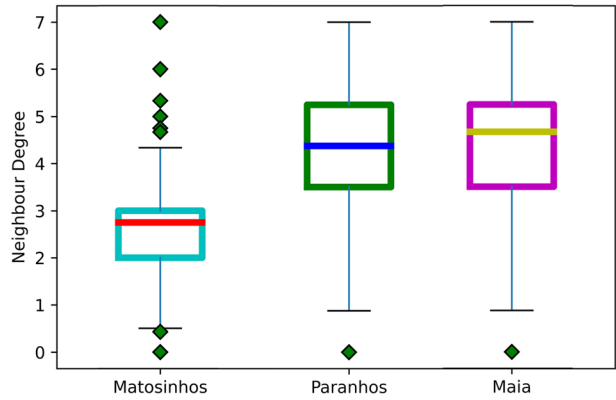


Fig. 5 Comparative analysis of the average neighbor degrees of the road networks of Matosinhos, Paranhos and Maia

4.1.3 Centrality measures

The Maia road networks degree centrality’s maximum and minimum values were 0.0294 and 0.0074, respectively, with an average value of 0.017, which is 97.7 and 86.5% higher relatively to the networks of Matosinhos and Paranhos, respectively. In the case of closeness centrality, the Paranhos road network showed an average value of 0.0344, with a maximum and minimum of 0.051 and 0.0, respectively. The average value of closeness centrality of Paranhos is 42.4% higher and 128.78% lower than that of Matosinhos and Maia, respectively.

Figure 6 illustrates the visualization of the three studied networks in terms of closeness centrality. Following the analysis of the closeness centrality, it can be observed that the intersections of the Matosinhos road network are closer to each other than of the other two networks.



Fig. 6 Visualization of the three studied regions in terms of closeness centrality: (on the left) Matosinhos, (on the middle) Paranhos and (on the right) Maia

The average Eigenvector centrality of the Maia road network is 0.0256, which is 96.9 and 88.04% higher than that of Matosinhos and Paranhos, respectively, as is illustrated in Fig. 7. Table 1 presents the summary of Eigenvector centrality of the three studied regions. In summary, the Eigenvector of the Matosinhos road network indicates that it contains the most intersections that are better connected to other vital intersections within the network.

It statistically depicts the Eigenvector centrality values for the three studied sub-regions. For the road network of Maia, most of its values are close to zero. While the other two networks do not follow a normal distribution, they illustrate positively skewed distributions.

In the case of betweenness centrality, the Maia road network possessed a value of 0.02065, which is 91.3 and 70.9% greater than the road networks of Matosinhos and Paranhos, respectively, as is illustrated in Fig. 8. Table 2 presents more insights into the betweenness centrality of the road networks. The analysis of betweenness centrality demonstrates that the intersections of the Matosinhos road network have less influence over the flow of information within the network.

Fig. 7 Box plot of the Eigenvector centrality of the studied three regions

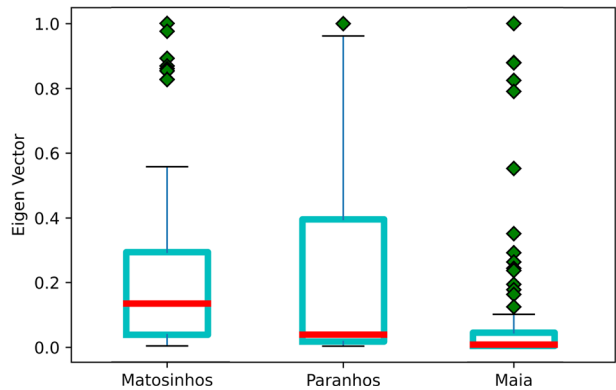


Table 1 Eigenvector Centrality of studied network

Eigenvector Centrality	Max	Min	Average
Matosinhos	0.27	$5.7e^{-272}$	0.000789
Paranhos	0.263	$5.4e^{-97}$	0.00306
Maia	0.272	$9.4e^{-06}$	0.0256

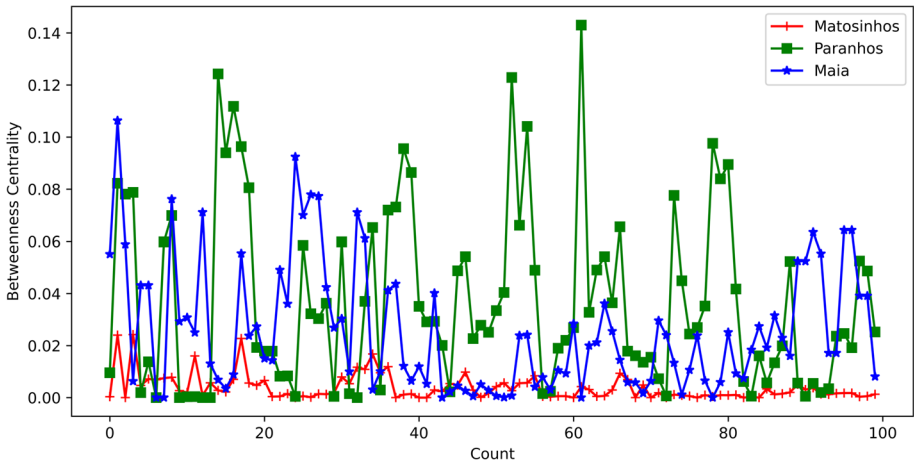


Fig. 8 Comparison of the betweenness centralities of Matosinhos, Paranhos, and Maia

Table 2 Betweenness Centrality of studied road networks

Betweenness Centrality	Max	Min	Average
Matosinhos	0.198	0.0	0.00179
Paranhos	0.123	0.0	0.0060
Maia	0.130	$1.34e^{-05}$	0.02065

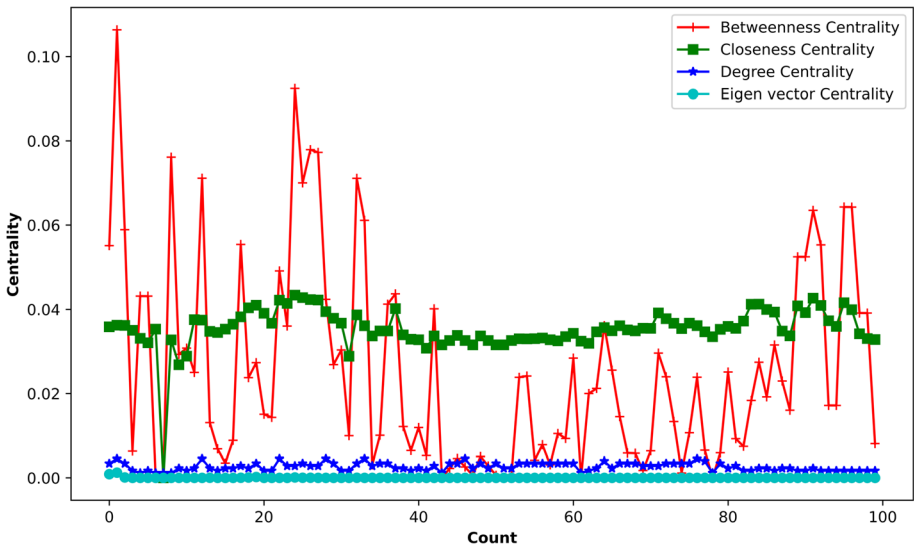


Fig. 9 Comparison of four centrality measures calculated for Maia road network (in each case, the first 50 readings were taken into account)

Finally, to have a broader look at how different centrality measures evolves, a graphical comparison is shown in Fig. 9, which allows to perceive the four types of centrality measures of the road network of Maia.

4.1.4 Connected components

An induced subgraph is a component of an undirected graph where pathways connect any two vertices, but not to any other vertices in the graph. Connected components form a partition of the graph vertices, meaning that the connected components are non-empty, pairwise disjoint, and the union of connected components forms the set of all vertices. Porto road network is a connected graph, and the number of the connected component is 1 (one), which indicates the reachability of the network, implying that one can reach all other nodes by traversing edges from any network node.

4.1.5 Community analysis

Using the Louvain algorithm, the modularity value of the network was found to be 0.9469, which indicates that it has dense connections between the nodes within modules, but sparse connections between nodes in different modules. The number of communities of the Paranhos road network found by the Girvan–Newman algorithm was 3. However, it should not be so small; thus, it indicates the algorithm's shortcomings. On the other hand, the most effective community detection algorithm was the Louvain method, which was used to found the total of the communities of Porto, Matosinhos, Paranhos and Maia of 147, 19, 24, and 16, respectively, Figs. 10 and 11.

From the analysis of the community detection concepts, it can be realized that the road network of Paranhos contains more structural similarities and is very tightly bound within those detected regions than the others.

4.1.6 PageRank

PageRank computes a ranking of the nodes in the graph based on the structure of the incoming links. The maximum and minimum page rank values of the Matosinhos road network were calculated as 0.0002 and $1.68e^{-05}$, respectively, with an average of $8.36e^{-05}$.

Fig. 10 Visualization of found communities of Porto road network (found communities are highlighted using different colours)





Fig. 11 Additional visualization of found communities of studied networks: (on the left) Matosinhos, (on the middle) Paranhos and (on the right) Maia (found communities are highlighted using different colours)

Compared with Paranhos and Maia road networks, it was 18472.7 and 18473.2% lower than of Paranhos and Maia networks, respectively. Table 3 presents more insights into the networks under analysis.

From the analysis of the obtained values, it can be concluded that the intersections of the Maia road network possess a higher rank than of the other regions, which indicates their importance within the network. It also suggests that one intersection will likely receive more lanes from other intersections.

4.1.7 Average path distance

The average path distance between the nodes of each studied network was also measured. The average shortest path of the Matosinhos, Paranhos and Maia road networks were 50.09, 26.74, and 29.17 units, respectively. Figure 12 illustrates the analysis of path distances of the road network of Maia with the help of a histogram.

The low values of average path distance of the three studied networks demonstrate that they facilitate quick transfer of traffic and, hence, reduce transportation costs.

5 Discussion

A comprehensive analysis of Porto's road network consisted of the degree distribution, including power-law fitting, centrality measures, degree centrality, closeness centrality, betweenness centrality, eigenvector centrality, community detection, and page rank.

The degree distribution showed that most nodes possess a relatively small degree, but a few nodes have a considerable degree connected to many other nodes. The best minimal value of power-law fit was -55.32 , which indicates that the networks do not fully follow the power-law distribution.

Table 3 PageRank values of studied networks

Page Rank	Max	Min	Average
Matosinhos	0.0002	$1.68e^{-05}$	$8.36e^{-05}$
Paranhos	0.0013	0.000106	0.000519
Maia	0.0061	0.00182	0.00366

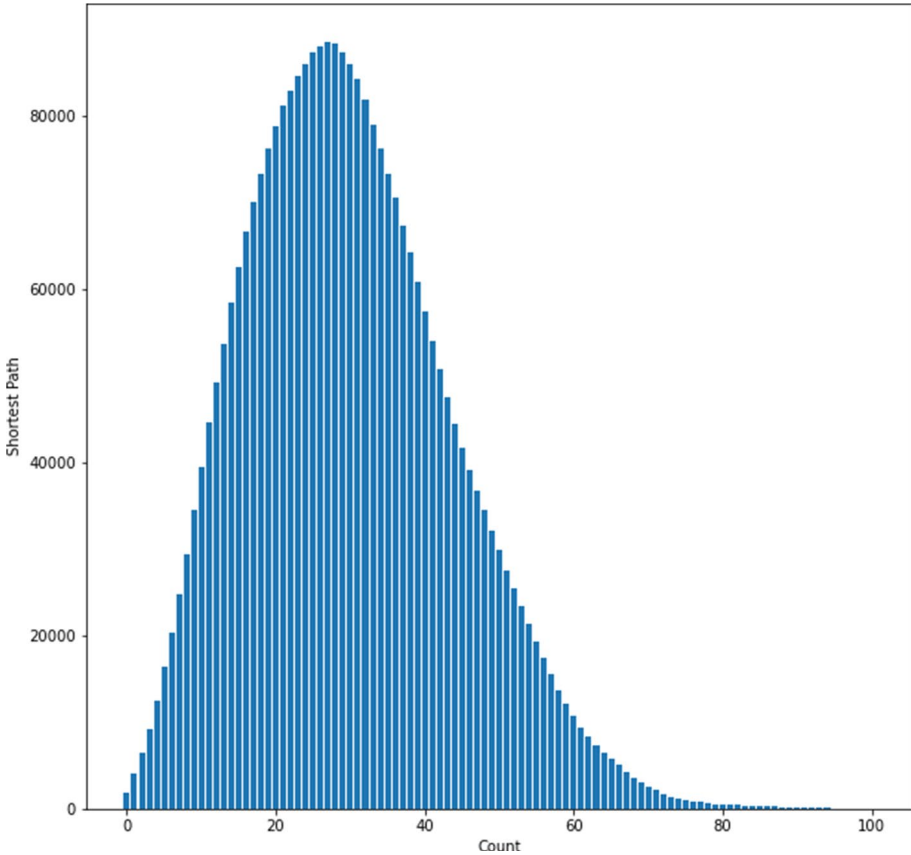


Fig. 12 All shortest path distances of the Maia road network

The nodes of the Maia road network seemed to have the strongest tendencies to connect to others who are also popular, while the nodes of the Matosinhos road network displayed the weakest such tendencies, as indicated by the box plot diagram of Fig. 4.

The most critical nodes from the centrality measures are indicated in Table 4. The most critical nodes of the networks based on betweenness centrality measures were the 'Avenida da Boavista, 4100-119 Porto (latitude: 41.157944, longitude: -8.629105)', and 'Autoestrada do Norte, Porto (latitude: 41.1687869, longitude: -8.6400656)', respectively.

Table 4 Top three most important nodes (latitude and longitude) of each studied network measured using the betweenness centrality

Network	Best (Lat. and Lon.)	2nd Best (Lat. and Lon.)	3rd Best (Lat. and Lon.)
Porto	41.162321,-8.6554258	41.1616354,-8.6513866	41.1618465,-8.652652
Matosinhos	41.1889938,-8.6714335	41.1896475,-8.6758741	41.2013094,-8.6566653
Paranhos	41.1736115,-8.6058176	41.1831689,-8.6134596	41.1742979,-8.6103207
Maia	41.2449145,-8.1803343	41.2536122,-8.1781785	41.2533825,-8.178357

The Porto road network possesses 147 communities from the Louvain community detection algorithm, which means that there are 147 groups of nodes with dense connections with similar structural properties.

Only one connected network component exists, indicating that the network is highly accessible; i.e., from any node in the network, all other nodes may be reached by traversing edges from any node in the network.

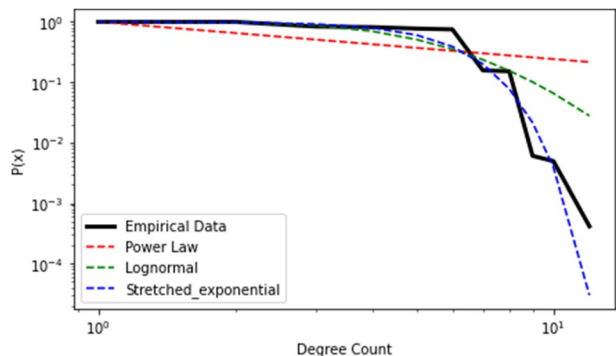
The achieved modularity values was equal to 0.9469, which is relatively high and indicates dense connections between the network nodes. The page rank analysis found that the Maia road network nodes have higher values than of the Matosinhos and Paranhos networks. Moreover, the average path distance revealed low values, which indicates the small world characteristic of the road network.

The same experiments using identical parameters were performed on the Ingolstadt road network, a city of Bavaria, in Germany, and the results were consistent with those obtained as to the Porto road network. For example, the degree distribution did not follow the power-law, as shown in Fig. 13. Interestingly, the minimal value for power-law fit was -248.71 , which is much less than the one of the Porto road network. However, different outcomes are expected for road networks of sub-continent regions like India and Bangladesh, because of their different structural properties, which is part of the future scope of this study.

6 Conclusions

From a network science perspective, analysing road networks is fundamental for investigating valuable insights of a current and future urban settlement. It is a compelling approach to identify, visualise and explore different network behaviours that can even serve to their prediction and controlling purposes. The beautiful and historic city of Porto, in Portugal, and its three sub-parts: Matosinhos, Paranhos, and Maia, were selected to conduct this study. The three sub-parts were selected to find valuable insights within the each sub-urban area as they might possess different structures and characteristics compared to the whole city network. Python OSMnx module aimed to capture the network data, and the NetworkX package of Python allowed for analysis of the results. Some beneficial insights found into the studied network are:

Fig. 13 Power-law fitting of the degree distribution of the Ingolstadt road network



- Most nodes of the studied road networks have a relatively small degree while a few have a considerable degree;
- The studied networks do not fully follow the power-law distribution;
- The bigger the network, the lower the tendency to connect to other popular nodes;
- The higher the value of modularity, the higher the reachability of the network;
- The smaller the average path distance, the higher the representability of small-world phenomena.

This information can be beneficial in modelling real-world road networks. Although the current analysis only considered the Porto road network, it can efficiently analyse other road networks in any part of the world. For example, the Ingolstadt road network, in Germany, was also analyzed using a similar approach. Interestingly, the outcomes were found to be identical to the ones of the presented results. However, road networks of Africa and other less developed parts of the world should also be studied in the near future.

Particular concentration on finding the road network's subgraph patterns will be necessary for future works. Also, more analysis should be performed on small-world phenomena; for example, by studying small-world coefficients (σ , ω) and the congestion spreading analysis.

Acknowledgements This article results from the project Safe Cities - “Inovação para Construir Cidades Seguras”, with reference POCI-01-0247-FEDER-041435, co-funded by the European Regional Development Fund (ERDF), through the Operational Programme for Competitiveness and Internationalization (COMPETE 2020), under the PORTUGAL 2020 Partnership Agreement.

Data availability The used data is public available at: <https://www.openstreetmap.org/> (accessed in July 2022).

Declarations

Conflict of interests The authors declare no conflict of interest.

References

1. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. *Nature* **393**(6684), 440–442 (1998)
2. Barabási, A.-L., Albert, R.: Emergence of scaling in random networks. *Science* **286**(5439), 509–512 (1999)
3. Gao, S., Wang, Y., Gao, Y., Liu, Y.: Understanding urban traffic-flow characteristics: a rethinking of betweenness centrality. *Environment and Planning B: Planning and Design* **40**(1), 135–153 (2013)
4. Ahmadzai, F., Rao, K.L., Ulfat, S.: Assessment and modelling of urban road networks using integrated graph of natural road network (a gis-based approach). *Journal of Urban Management* **8**(1), 109–125 (2019)
5. Duan, Y., Lu, F.: Robustness of city road networks at different granularities. *Physica A Stat. Mechan. Appl.* **411**, 21–34 (2014)
6. Davidović, S., Bogdanović, V., Garunović, N., Papić, Z., Pamučar, D.: Research on speeds at roundabouts for the needs of sustainable traffic management. *Sustainability* **13**(1), 399 (2021)
7. Khojasteh, H., Khanteymoori, A., Olyaei, M.H.: Comparing protein–protein interaction networks of sars-cov-2 and (h1n1) influenza using topological features. *Sci. Rep.* **12**(1), 1–11 (2022)
8. De Montis, A., Barthélemy, M., Chessa, A., Vespignani, A.: The structure of interurban traffic: a weighted network analysis. *Environ. Plann. B Plann. Design* **34**(5), 905–924 (2007)
9. Masucci, A.P., Stanilov, K., Batty, M.: Exploring the evolution of london's street network in the information space: A dual approach. *Phys. Rev. E* **89**(1), 012805 (2014)

10. Xiao, Z., Jian-Ping, C., Jia-Li, S., Li-Dong, B.: Analysis on topological properties of beijing urban public transit based on complex network theory. *Acta Physica Sinica* **61**(19) (2012)
11. Scardoni, G., Laudanna, C.: Identifying critical traffic jam areas with node centralities interference and robustness. *Networks & Heterogeneous Media* **7**(3), 463 (2012)
12. Yang, Y., Cao, J., Qin, Y., Jia, L., Dong, H., Zhang, A.: Spatial correlation analysis of urban traffic state under a perspective of community detection. *Int. J. Modern Phys. B* **32**(12), 1850150 (2018)
13. Hong, Y., Yao, Y.: Hierarchical community detection and functional area identification with osm roads and complex graph theory. *Int. J. Geogr. Inf. Sci.* **33**(8), 1569–1587 (2019)
14. Feng, W., Li, B., Chen, Z., Liu, P.: City size based scaling of the urban internal nodes layout. *PLoS ONE* **16**(4), 0250348 (2021)
15. Tsiotas, D.: Drawing indicators of economic performance from network topology: The case of the interregional road transportation in greece. *Res. Transp. Econ.* **90**, 101004 (2021)
16. Borgatti, S.P.: Centrality and network flow. *Soc. Netw.* **27**(1), 55–71 (2005)
17. Newman, M.: *Networks*, 2nd edn. Oxford University Press, Oxford (2018)
18. Otte, E., Rousseau, R.: Social network analysis: a powerful strategy, also for the information sciences. *J. inform. Sci.* **28**(6), 441–453 (2002)
19. Bodendorf, F., Kaiser, C.: Detecting opinion leaders and trends in online social networks. In: *Proceedings of the 2nd ACM Workshop on Social Web Search and Mining*, pp. 65–68 (2009)
20. Crucitti, P., Latora, V., Porta, S.: Centrality measures in spatial networks of urban streets. *Phys. Rev. E* **73**(3), 036125 (2006)
21. Jayasinghe, A., Sano, K., Nishiuchi, H.: Explaining traffic flow patterns using centrality measures. *Int. J. Traff. Transport Eng.* **5**(2), 134–149 (2015)
22. Freeman, L.C.: Centrality in social networks conceptual clarification. *Soc. Netw.* **1**(3), 215–239 (1978)
23. Chea, E., Livesay, D.R.: How accurate and statistically robust are catalytic site predictions based on closeness centrality? *BMC Bioinformatics* **8**(1), 1–14 (2007)
24. Porta, S., Strano, E., Iacoviello, V., Messori, R., Latora, V., Cardillo, A., Wang, F., Scellato, S.: Street centrality and densities of retail and services in bologna, italy. *Environment and Planning B: Planning and Design* **36**(3), 450–465 (2009)
25. Sabidussi, G.: The centrality index of a graph. *Psychometrika* **31**(4), 581–603 (1966)
26. Barthelemy, M., Bordin, P., Berestycki, H., Gribaudo, M.: Self-organization versus top-down planning in the evolution of a city. *Scientif. Rep.* **3**(1), 1–8 (2013)
27. Brandes, U.: On variants of shortest-path betweenness centrality and their generic computation. *Soc. Netw.* **30**(2), 136–145 (2008)
28. Hansen, D., Shneiderman, B., Smith, M.A.: *Analyzing Social Media Networks with nodeXL: Insights from a Connected World*. Morgan Kaufmann (2010)
29. Xia, S., Xiong, Z., Luo, Y., Dong, L., Zhang, G.: Location difference of multiple distances based k-nearest neighbors algorithm. *Knowl.-Based Syst.* **90**, 99–110 (2015)
30. Yao, D., van der Hoorn, P., Litvak, N.: Average nearest neighbor degrees in scale-free networks. [arXiv:1704.05707](https://arxiv.org/abs/1704.05707) (2017)
31. Barrat, A., Barthelemy, M., Pastor-Satorras, R., Vespignani, A.: The architecture of complex weighted networks. *Proc. Nat. Acad. Sci.* **101**(11), 3747–3752 (2004)
32. Yang, B., Liu, D., Liu, J.: *Discovering Communities from Social Networks: Methodologies and Applications*. In: *Handbook of Social Network Technologies and Applications*, pp 331–346. Springer (2010)
33. Reddy, P.K., Kitsuregawa, M., Sreekanth, P., Rao, S.S.: A graph based approach to extract a neighborhood customer community for collaborative filtering. In: *International Workshop on Databases in Networked Information Systems*, pp 188–200. Springer (2002)
34. Fortunato, S., Hric, D.: Community detection in networks: a user guide. *Phys. Rep.* **659**, 1–44 (2016)
35. Girvan, M., Newman, M.E.: Community structure in social and biological networks. *Proc. Nat. Acad. Sci.* **99**(12), 7821–7826 (2002)
36. Chapter four-Island biogeography of food webs. In: Bohan, D.A., Dumbrell, A.J., Massol, F. (eds.) *Networks of Invasion: A Synthesis of Concepts*. *Advances in Ecological Research*, vol. 56, pp. 183–262 Academic Press (2017)
37. Clauset, A., Newman, M.E., Moore, C.: Finding community structure in very large networks. *Phys. Rev. E* **70**(6), 066111 (2004)
38. Lu, H., Halappanavar, M., Kalyanaraman, A.: Parallel heuristics for scalable community detection. *Parallel Comput.* **47**, 19–37 (2015)
39. Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. *J. Stat. Mechan. Theor. Exp.* **2008**(10), 10008 (2008)
40. Dugué, N., Perez, A.: *Directed Louvain: Maximizing Modularity in Directed Networks*. PhD thesis, Université d'Orléans (2015)

41. Hagberg, A., Swart, P., S Chult, D.: Exploring Network Structure, Dynamics, and Function Using Networkx. Technical Report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States) (2008)
42. Boeing, G.: Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Comput. Environ. Urban. Syst.* **65**, 126–139 (2017)

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

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Selim Reza¹  · Marta Campos Ferreira¹  · J.J.M. Machado²  ·
João Manuel R.S. Tavares² 

Selim Reza
up202003355@fe.up.pt

Marta Campos Ferreira
mferreira@fe.up.pt

J.J.M. Machado
jjmm@fe.up.pt

¹ Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal

² Departamento de Engenharia Mecânica, Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal