Analysis of Road Networks Based on Graph Theory: Comparison Between Efficient and Inefficient Road Networks

Marco Raimondi (2037011), Ahmet Onur Akman (2044006), Andrea Mazzucco (2054552)

29/01/2023

Introduction

Every year TomTom releases a report covering 404 cities across 58 countries on 6 continents, based on a traffic index that ranks urban congestion worldwide.

We are interested in finding features that characterize the road networks, and so to compare these among cities (see Tab. 1) with a low congestion level and cities with a high one according to this ranking. We want to analyze the road networks exploiting graph theory. We will analyze the road networks computing 5 centrality scores for every node (degree, closeness, betweenness, eigenvector, and PageRank) and modularity of the network, and then compare these results of the six cities, in order to identify major differences that can help us to recognize if a city's road system efficiency can relate to these metrics or not. The target is to verify if the scores we will compute could be reliable metrics for ranking the efficiency of a city's road network. In addition, this can be useful for urban planners in order to recognize critical intersections, and improve the layout of the road by implementing new solutions.

Data

The networks of the six cities are created by gathering information about the nodes and edges from OpenStreetMap. The programming code we will use is Python, with some well-known libraries for data analysis. We will gather the graph of the road networks using OSMnx. This is a Python package that allows to download of geospatial data from OpenStreetMap and model, project, and analyze real-world street networks. We will exploit the package NetworkX for the network analysis.

Inefficient cities				**Efficient cities**			
City	**Rank**	**Nodes**	**Edges**	**City**	**Rank**	**Nodes**	**Edges**
Istanbul	1	8335	21510	Buenos Aires	67	2963	5791
Mumbai	5	3224	7748	Sao Paulo	68	3862	7410
Bengaluru	10	10496	28255	Dubai	244	2770	5261

Table 1 *Graph features of the studied cities.*

The districts of each city for the construction of the graphs were chosen by inspecting the most critical areas of vehicular flow. This data is made available by the *TomTom Traffic Index 2021*, which allows us to identify the most congested roads in a city. In Figures 1 and 2, the example of the city of Dubai is shown.

Figure 1 *Dubai road network graph.*

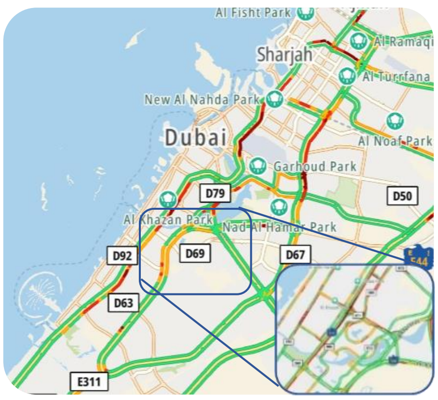

Figure 2 *Dubai roads coloured according to the level of road congestion.*

A) Degree Centrality

The first computed feature is the degree centrality, which represents the importance of a node in a graph based on the number of connections it has.

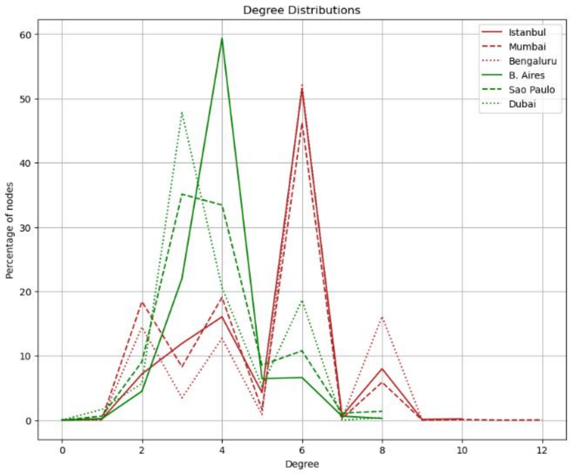

Figure 3 *Degree centralities comparison.*

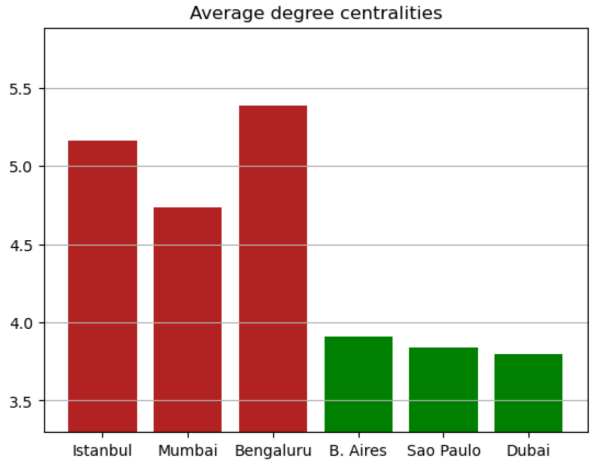

Figure 4 *Average degree centrality comparison.*

As shown in Fig. 3 and 4, cities with inefficient road networks tend to have an average degree centrality higher than those of efficient cities, indicating that probably crossroads where a high number of roads intersect can lead to traffic congestion. In fact, it can be seen that the degree distributions of the cities with an efficient road network are almost symmetrical, with the peak coinciding with the values 3 or 4, while the distributions for the inefficient cities are slightly left-skewed, with the peak coinciding with the values 6, meaning that a lot of nodes have a high degree. Even though these results should be supported by other measures to relate to the overall efficiency, this distribution implies that a more decentralized road network leads to a better traffic flow.

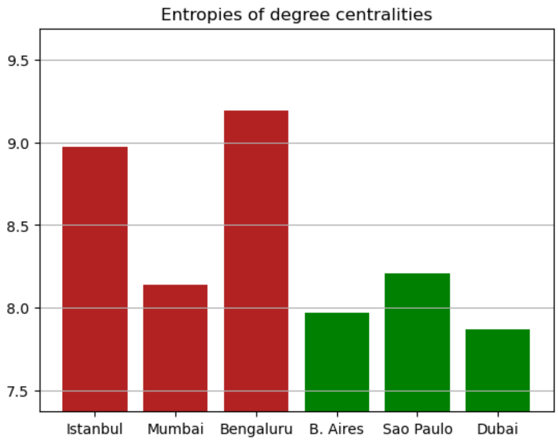

Figure 5 *Entropies of degree centrality comparison.*

In Fig. 5, the entropies of degree centrality distributions of nodes for each city are given. It is clearly seen that the entropies of efficient cities tend to be lower than the inefficient ones. This tells us that the efficient cities have more predictability in the connections of the crossroads. The unbalanced importance of intersections in inefficient cities could be causing congestion in some parts of the city, where vehicles have to visit in order to transport to other parts of the city.

B) Closeness Centrality

Closeness centrality can show a greater value of a node, the more likely it is to be closer and central to the rest of the nodes in the network, and so easily reachable. It is a useful measure to relate to a node's accessibility in its network.

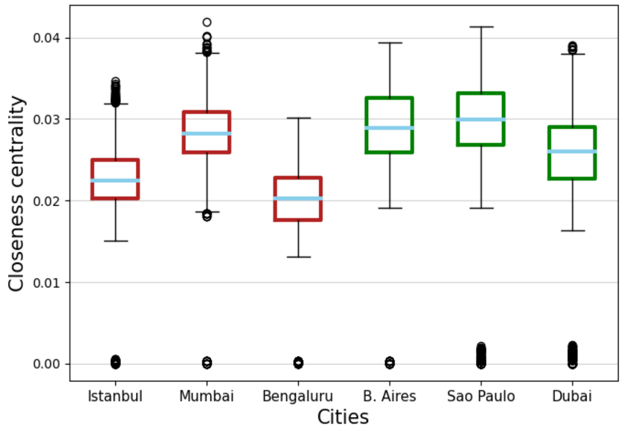

Figure 6 Closeness centralities comparison.

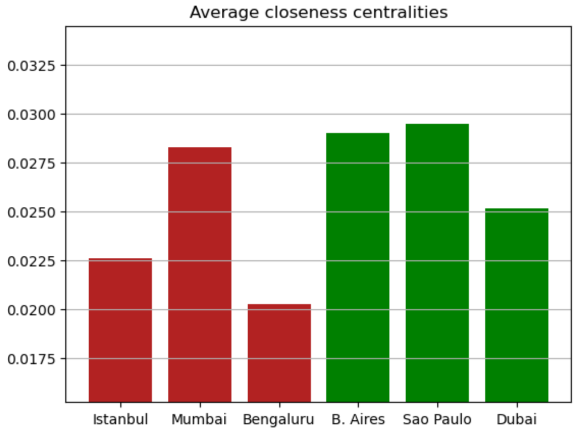

Figure 7 Average closeness centrality comparison.

The results, shown in Fig. 6 and 7, indicate that efficient cities have a higher average closeness centrality. This suggests that in efficient road networks, every place can be reached easier than in inefficient road networks. This also causes fewer central nodes in the network, which gives drivers fewer options in their trips to other parts of the city. These smaller number of central nodes might be causing the digestion of the traffic.

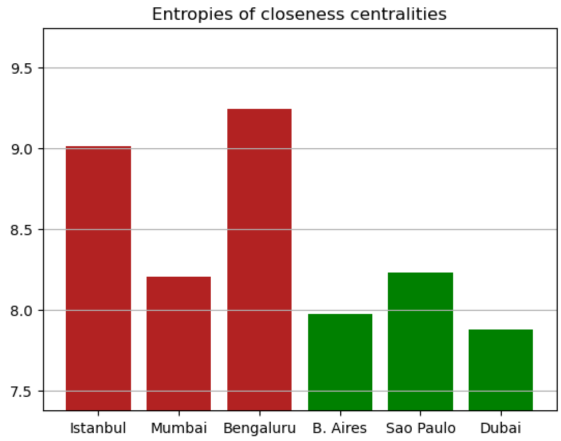

Figure 8 Entropies of closeness centrality comparison.

In Fig. 8, the entropies of closeness centrality distributions of nodes for each city are given. It is observable that the entropies of efficient cities are lower than the inefficient ones, implying a more balanced centrality amongst their nodes. This once again tells us that the efficient cities have less variation in the importance of their crossroads, which could be the reason for a smoother traffic flow. This observation also implies that the drivers in inefficient cities will need to take longer routes to reach their destinations if they correspond to the inaccessible nodes.

C) Betweenness Centrality

Betweenness centrality is useful to observe the control of a node on the flow in the network. In fact, a node with a high value of betweenness centrality is considered a key intermediary in the graph, lying on a high number of shortest paths.

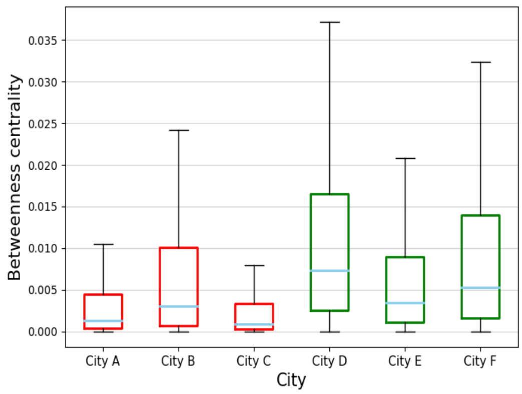

Figure 9 Betweenness centralities comparison.

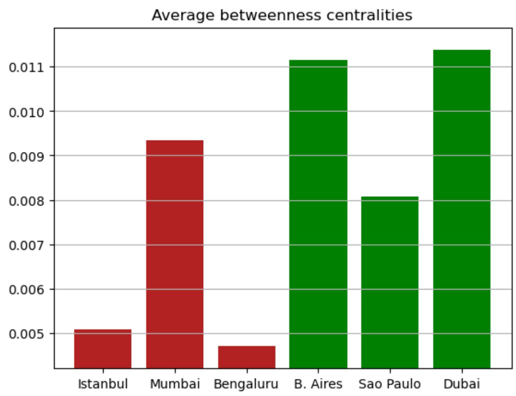

Figure 10 Average betweenness centrality.

As shown in Fig. 9 and 10, cities with inefficient road networks tend to have an average betweenness centrality lower than the ones with efficient road networks. This means that the nodes of the inefficient cities are relatively "unimportant" for their network, compared to the efficient city nodes. This again causes an uneven distribution for the traffic flow, hence greater digestion on the important crossroads, while other nodes are not efficiently used.

One interesting point to observe is that the betweenness of Mumbai's nodes seems to be way higher than expected. We choose to reason it with that the number of nodes in this network is lower than other cities in its class, therefore the shortest paths capture nodes across the graph better than other inefficient cities.

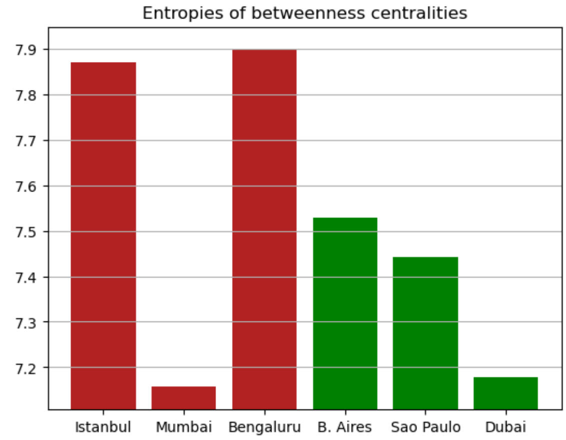

***Figure 11** Entropies of betweenness centralities.*

Except for the unexpected result mentioned above, it is seen that the entropies of the distributions of betweenness centralities in the inefficient cities are higher than the efficient ones, in Fig. 11. This again shows how unbalanced the control of nodes in inefficient cities, therefore the drivers are forced to use same certain intersections to reach where they are going, hence the congestion in those parts of the cities.

D) Eigenvector Centrality

Eigenvector centrality is a measure of the centrality of a node in a graph based on the idea that connections to high-centrality nodes contribute more to the centrality of a node than connections to low-centrality nodes. It is used to identify the most influential nodes in a network. A node with a high eigenvector centrality value has many connections to other high-centrality nodes and is therefore considered influential, while a node with a low eigenvector centrality value is less influential. Results are given in Fig. 12 and 13.

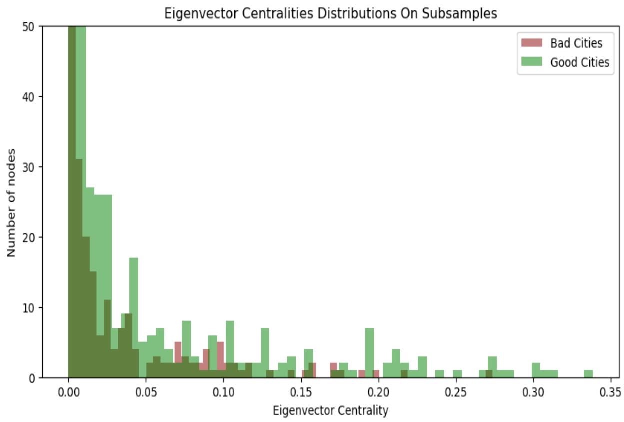

***Figure 12** Eigenvector centrality distributions of efficient (green) and inefficient (red) cities.*

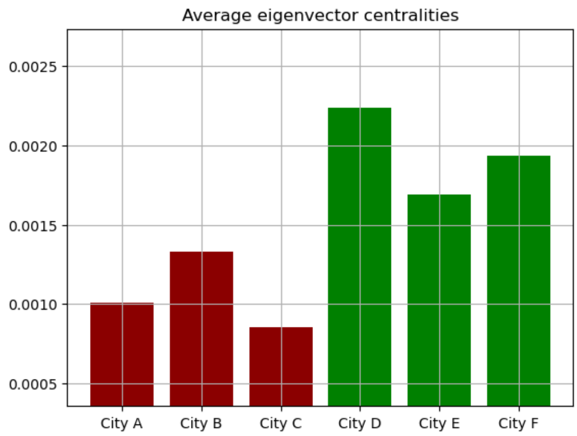

***Figure 13** Average eigenvector centrality.*

After the calculation of eigenvector centralities, we merged efficient cities' nodes as one and then did the same for the inefficient cities. Later we sampled 10000 nodes from each at random. You will notice that the y-axis in Fig. 12 goes up to 50, not capturing the actual value of the first few bins. We chose this way because it was hard to interpret the meaning when we used the same styles as in the other sections, the distributions were not varying, and most of the nodes had scores very close to zero. The first bins, therefore, have up to 9000 nodes for each class. By this method, we tried to visualize the distribution of the outliers.

In these two figures, we see that the efficient cities have relatively higher scores. This implies that the inefficient cities contain groups of intersections placed further from well-connected nodes, making them rather isolated. This means leaving some parts of the city relatively out of use and disconnected.

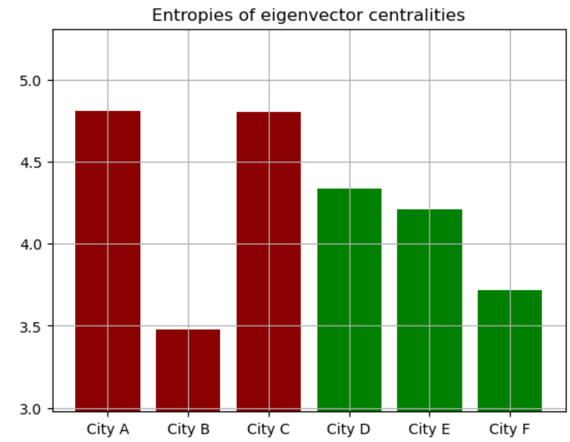

Figure 14 *Entropies of eigenvector centralities.*

Fig. 14 shows that efficient cities tend to have more balanced eigenvector centrality scores amongst their nodes. The difference here is, the results are not as distinguishing as the others in this study, and again, the result in Mumbai is unexpectedly low.
Again, we reason this with the small size of the graph, hence more balancly connected nodes. Considering this, and keeping in mind that the average eigenvector centrality in Mumbai is not very high (see Fig. 13), one can suggest that this city does not contain as “influential nodes” as others, causing this balanced distribution.

E) PageRank

PageRank represents a ranking of the nodes in the graph based on the importance of the neighboring nodes with incoming links. As the definition implies similarity to the eigenvector centrality, which is rightful, this measure focuses on the incoming connections in contrast. Results are given in Fig. 15 and 16.

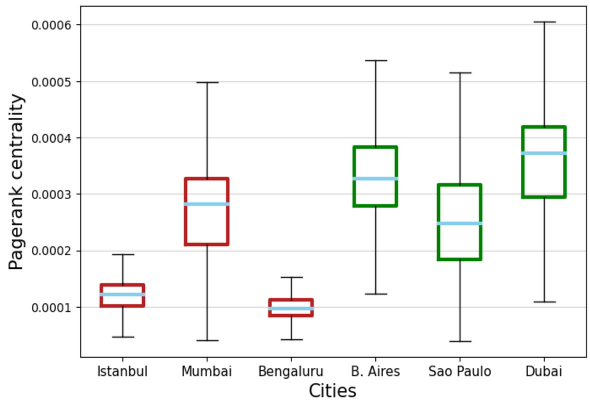

Figure 15 *PageRank centrality comparison.*

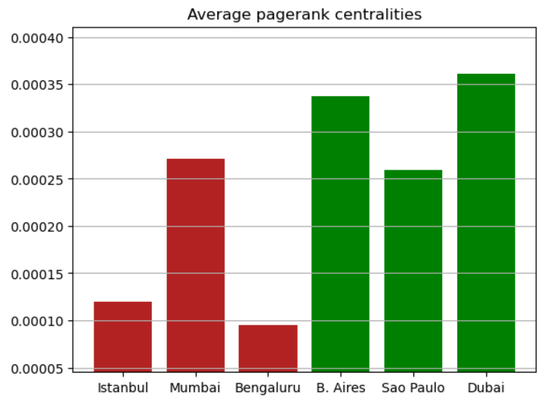

Figure 16 *Average PageRank centrality comparison.*

It is seen that the intersections of the inefficient cities tend to have lower values. The reason behind this could be either the lack of central nodes in these cities or isolated nodes far from the central intersections.

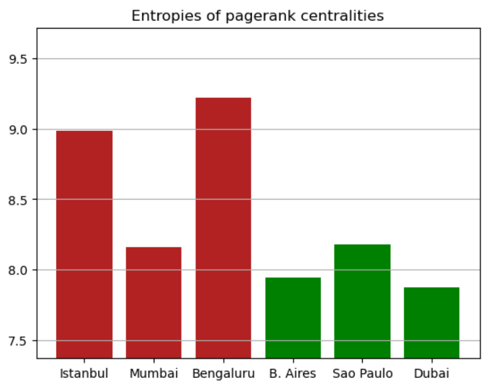

Figure 17 *Entropies of PageRank comparison.*

Fig. 17 shows the imbalance in the distribution in the inefficient cities. As some nodes in these networks are very well connected to the influential nodes, some nodes are left afar and disconnected. This causes the travels from and to these parts of the cities to depend on the same parts of the city, causing digestion and delays.

F) Modularity

Modularity is a measure of the density of connections within individual clusters (communities) in a graph relative to connections between clusters. It is used to identify the presence of distinct communities or modules in a graph. A high modularity value indicates that there are many connections within the communities and few connections between communities, while a low modularity value indicates that there are few connections within the communities and many connections between communities.
We have first obtained the partition of communities with the highest modularity in each city using the Louvain algorithm, built in Community API. Then we calculated modularities using these communities.

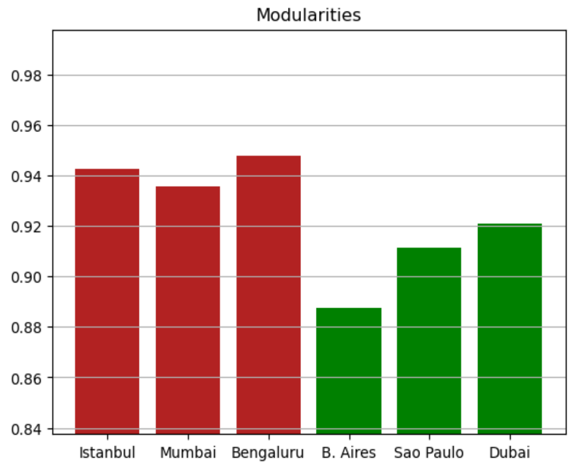

Figure 18 *Modularities comparison.*

From Fig. 18, which shows the comparison among the values of modularities of the cities, it can be realized that coming up with a partition that leads to better-separated groups in efficient cities is more difficult than the inefficient ones. Overall, we see a better-connected structure in these cities, hence less digestion between different parts of the city. The high fragmentation in inefficient cities give more importance and digestion on the (small number of) roads connecting the different parts of the cities.

G) Relationship between roads and betweenness centrality

It is interesting to notice how the distribution of the betweenness centrality can show interesting routes in the road network of a city. For this purpose, we have taken the graph of the entire street network of the city of Milan. The spatial distribution of the betweenness centrality highlights the continuity of relevant urban routes across a number of intersections, changes in direction, and focal urban spots. In the case of Milan, the most popular driving paths are the main ring roads, that emerge along the brighter nodes' routes (Figure 19). In most cities, betweenness centrality is also able to identify the primary structure of movement channels as different from that of secondary, local routes.

Figure 19 *Milan crossroads coloured by betweenness centrality.*

Figure 20 *Map of Milan's main roads coloured by average speed.*

Conclusions

The performed analysis demonstrates that an effective correlation exists between measures of centrality and efficiency of the road network, examining 6 different cities across the world. The analysis is highly approximate, since many variables in the areas under examination are only partially or not considered, such as population density, number and type of residential and/or commercial buildings and many more. However, many interesting aspects have emerged.

The results show that, from the computation of the degree centrality, cities with good viability possess intersections where fewer roads intersect. On the contrary, cities with an inefficient road network tend to have crossroads with many more intersecting roads. From the measure of closeness centrality, it is possible to notice that cities with a good viability system have a higher score with respect to the inefficient cities, meaning that moving from one place to another is easier. Furthermore, from the analysis of the betweenness centrality, it emerges that in good road networks the average score is higher, indicating that intersections in efficient cities are closer to other node pairs, with respect to crossroads of poor road networks. It is also important to note that, in efficient cities, the links between the various intersections have a greater importance, given by the PageRank centrality, and within the same city area (or cluster) we have fewer nodes and connections, but we have many more roads that connect different areas (or clusters). Conversely, for cities with poor viability, we have many nodes with many connections between them, but they are poorly placed as they are not central to the graph and do not play an important role within the urban infrastructure.

Finally, a correlation between betweenness centrality and the characteristics of the roads has been demonstrated, indicating that information on the road network can be obtained from the scores of the centrality measures. In conclusion, it is possible to affirm that the infrastructure of a city greatly influences its viability, and that these results can be taken into account for the construction of a city with good viability.

Code

Other details, coloured maps and more can be found in the Python Jupyter Notebook file given in the following link: https://github.com/aonurakman/LFN_Project

Links

- www.tomtom.com/traffic-index/ranking/
- www.openstreetmap.org/

References

- Boeing, G. 2017. OSMnx: New Methods for Acquiring, Constructing, Analyzing, and Visualizing Complex Street Networks. *Computers, Environment and Urban Systems* 65, 126-139. doi:10.1016/j.compenvurbsys.2017.05.004
- Cardillo, Alessio & Scellato, Salvatore & Latora, Vito & Porta, Sergio. (2006). Structural Properties of Planar Graphs of Urban Street Patterns. Physical review. E, Statistical, nonlinear, and soft matter physics. 73. 066107. 10.1103/PhysRevE.73.066107.
- Reza, S., Ferreira, M.C., Machado, J. *et al.* Road networks structure analysis: A preliminary network science-based approach. *Ann Math Artif Intell* (2022). https://doi.org/10.1007/s10472-022-09818-x

Member	Contribution
Marco Raimondi **ID:** 2037011 marco.raimondi@studenti.unipd.it	• Implementation of degree centrality, closeness centrality, and betweenness centrality; • Implementation of the relationship between roads and betweenness centrality; • Improvements in the implementation of entropy measures; • Overall structure of the code; • Improvements on the plotting of the graphs; • Documentation and research.
Ahmet Onur Akman **ID:** 2044006 ahmetonur.akman@studenti.unipd.it	• Implementation of modularity; • Implementation of eigenvector centrality, PageRank centrality; • Implementation of entropy measures; • Improvements in the implementation of degree centrality, closeness centrality, betweenness centrality; • Overall structure of the code; • Improvements on the data visualization; • Documentation and research.
Andrea Mazzucco **ID:** 2054552 andrea.mazzucco.2@studenti.unipd.it	• Documentation and research; • Improvements in the implementation of graph measures; • Improvements on the data visualization; • Review the code and fix it; • Re-organize the structure of the code.