Analysis of Road Network Topologies through Centrality Measures and Graphlet Analysis

Tommaso Bianco ID : 2037864
Matteo Villani ID : 2090299
Federico Violin ID : 2061746

Project Repository:
https://github.com/Mattolo4/Network-graphs-analysis?tab=readme-ov-file

1 ABSTRACT

This project explores the structural characteristics of two distinct Road Networks using networks analysis techniques. Focusing on the quantification of connectivity and influence within these networks, we calculated closeness and betweenness centralities as well as graphlet analysis for graphlet of dimensions 3 (triangular graphlet), 4 (square graphlets) and 5 (start graphlets). Closeness centrality is computed to identify intersections that, on average are closest to all other intersections in the road network. Betweenness centrality is calculated to pinpoint crucial intersections that act as bridges for traffic flow. Graphlet analysis provides the understanding of local structural motifs within the networks. Trough the application of these methodologies, we aim to view key similarities and differences between the two road networks. The analysis will highlight nodes with significant centrality values in order to indicate critical locations for traffic, to make some strategic intervention. Graphlet analysis will highlight topological patterns, in order to understand better the difference between two cities, one from USA and one from Europe. The project leverages mathematical and computational tools to process and analyse large-scale road networks, facilitating the extraction of meaningful information.

2 INTRODUCTION

In our time, the study and optimization of road networks play a central node for efficiency and sustainability of modern cities. The evolving of cities must be related to understanding the underlying dynamics of transportation systems. The aim of this project is to create a general view for understanding the meaningful information of two different cities: one from USA, that is Philadelphia, and one from Europe, Berlin but it can be expanded in order to compare other two different cities from all over the world. The two selected road networks under examination serve as representative cases, each containing unique characteristic influenced by geographical and urban factors. By employing mathematical and computational tools, we want to reveal the complexities embedded in these networks and extract meaningful insights that contribute to the broader discourse on urban mobility. We choose this project because we found interesting the argument of traffic flow and urban topologies of the cities, we define some important points that describe the relevance of our case of study and are the motivation for the creation of the project:

- Relevance to Urban Development: The project focus on the optimization of road networks, a critical component of modern cities.
- Impact on Transportation Efficiency: Understanding the meaningful information of road networks, such as central nodes and topologies allows us to highlight the areas where actions can be taken to update transport efficiency and limit traffic with queues.
- Strategic Points: Identifying nodes with significant centrality values informs strategic intervention points.
- Comparation of different cities: In order to find the difference of cities topologies between different country or different continents.
- Practical Applications: in urban planning, traffic management, infrastructure development, enhancing the functionality of urban spaces.

Those points summarize the motivations of why we choose this project.

3 DATASET

In order to create this project, the first part was to individuate the datasets for working, we focused on road networks so we tried to find some relevant Networks. In the end we find two important datasets that are of the following cities:

- Philadelphia:
  - Nodes: 12,981
  - Edges: 28,376
- Berlin:
  - Nodes: 13,389
  - Edges: 40,003

Datasets are chosen from this repository: DATASETSREPOSITORY. The data are updated to May 1, 2016. Each Road Network is a TNTP Data file: TNTP is tab delimited text files, with each row terminated by a semicolon. The files have the following format:

- First lines are metadata; each item has a description.
- Comment lines start with '$\sim$ '.
- Network files (must be named "network"_net.tntp) and it's one per link; links are directional, going from 'init node' to 'term node'. The standard order of the fields in the network files is {Init node , Term node , Capacity , Length , Free Flow Time , B , Power , Speed limit , Toll , Link Type}.

Datasets, however, have a problem, namely the fact that, although there are TNTP files that contain the geographical coordinates of each node, the latter are wrong and lead to some completely different geographical areas. It is a problem for our project because we also wanted to analyse the coordinates to better visualize the point, to see where targeted interventions could be made to limit traffic or to create new infrastructures. For our project, in order to analyse correctly the dataset for our purpose, we select the 'length' factor for weighting the graph obtained by the TNTP file. This factor has been standardized to km in order to have most comparable results (length indicates the distance from one node to another in miles for American datasets and in km for European datasets).

4 METHODS AND ALGORITHMS

Our project concentrates in one ipynb files that contains all the algorithms and the results. Before entering on the details of the algorithm we want to specify some important decisions that we've done for the project. First of all, we decided to create two graphs: one directed and one undirected, for each road networks of the two cities Philadelphia and Berlin. This decision is done because we calculated closeness and betweenness centralities in directed graph, but we use graphlet matcher in undirected graph in order to count correctly the graphlets in directed graph (eliminating the redundancy). We divided our project in some major phases:

- Read the TNTP file and create the directed/undirected graph: in order to do this, we create a Pandas dataframe that contains all the data, we used the function pd.read_csv from Pandas library to do this. Then, we create the direct/indirect graph weighted graph.
- Calculate the closeness Centrality: we use the function nx.closeness_centrality() that computes the closeness centrality for each node in the graph.
- Calculate the betweenness Centrality: we use the function nx.betweenness_centrality() that computes the betweenness centrality for each node in the graph.
- Graphlet Isomorphism Problem: we create a function for solve this problem on undirected graph. The function is called find_motif_occurrences, it receives in input a graph and a motif and returns in output the motif count and a list of motifs occurrences in the graph. The purpose of this function is to analyze a given graph and identify occurrences of a specified motif within in. The motif is treated as a subgraph, and the function uses the GraphMatcher class from NetworkX to find subgraph ismorphisms, effectively locating instances of the motif within the larger graph.
- Graphlet Counting: we divide this in two different parts: graphlet counting for undirected graph, that is done by the function find_motif_occurrences; and graphlet counting for directed graph. For this task we create a function called directed_motifs that takes in input: the directed graph, the number of motifs in the undirected graph and the list of the motifs found in indirected graph. For each motifs found in indirect graph we generate the subgraph of the graph G that is isomorphic to and we count the possible instances by normalizing the permutations.

5 RESULTS

5.1 CENTRALITY

Since we are studying a road network, the intersection with the highest closeness centrality is the one that, on average, is closest to all other intersections. This means that, from this intersection, a vehicle can reach any other intersection in the network with the least travel distance, so the node serves as an efficient travel hub that is well-located to quickly access different parts of the road network.

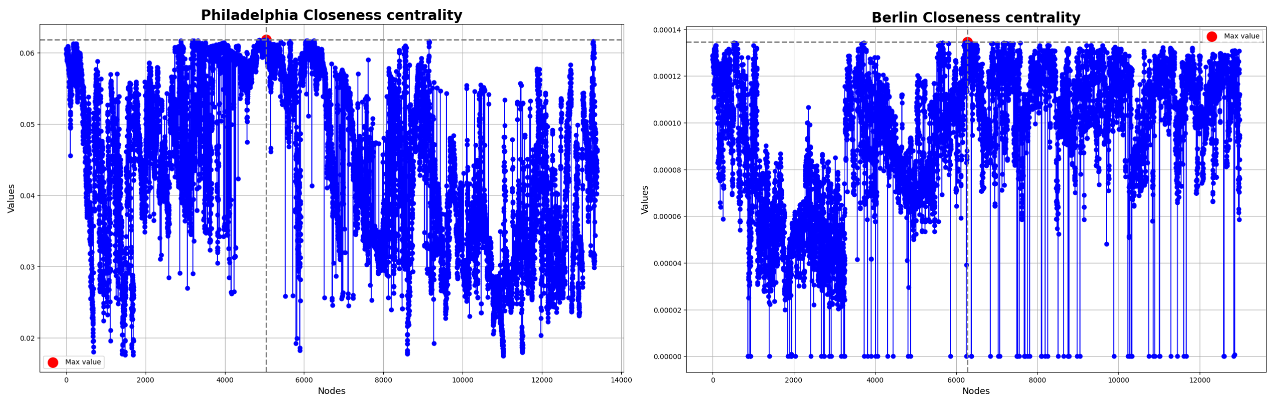


Figure 1: Closeness Centrality Plot

- In Berlin's closeness centrality plot , each node has low results, we can see that the maximum value is near to 0.00014. This can be related to the length weight factor. As a future improvement to have a better interpretation of the data, we can also identify the geographic coordinates of the nodes with the highest values to evaluate the traffic flow of a specific intersection.
- In Philadelphia's closeness centrality plot , each node has a higher result than the Berlin dataset, we can see that the maximum value is near to 0.06. As a future improvement, also here, we can identify the geographic coordinates of the nodes with the highest value to evaluate the traffic flow of a specific intersection.

In our case of study, the intersection with the highest betweenness centrality is crucial for maintaining the overall connectivity of the road network. It acts as a key bridge or bottleneck in facilitating traffic flow between different intersections; it is likely to be a major connector between routes and regions in the network. Congestion at this intersection could have a significant impact on the connectivity of the entire road network just because it is a key point for monitoring and managing traffic.

Comparing the betweenness centrality plots of the two cities, we can observe that the mean value of nodes in Philadelphia is higher than in Berlin. This suggests that the traffic flow in Berlin is concentrated in only a small number of nodes, whereas in Philadelphia, the traffic flow is distributed across all nodes.

Analyzing both metrics (Betweenness and Closeness Centralities) together provides a comprehensive view of the network's structure and potential points of interest for various planning and management purposes.

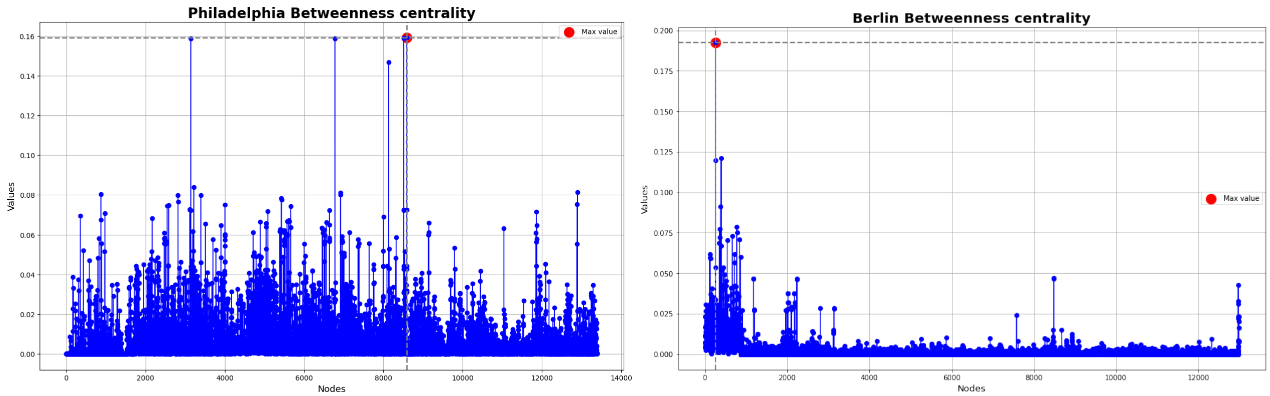
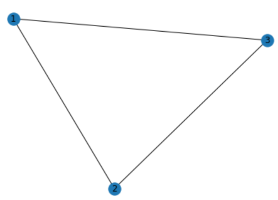


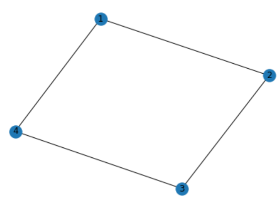
Figure 2: Betweenness Centrality Plot

5.2 GRAPHLETS

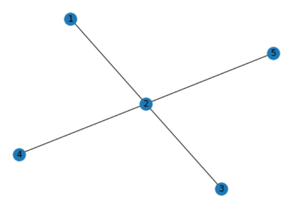
For our case of study, we want to concentrate the project on visualizing three different graphlets that are the following, for each of them we want to indicate differences between these two networks of Berlin and Philadelphia in order to make a topology comparison. We used undirected graph in order to extrapolate the number of graphlet and used it for finding the most popular directed graphlet in directed graph.



(a) Triangular Motifs



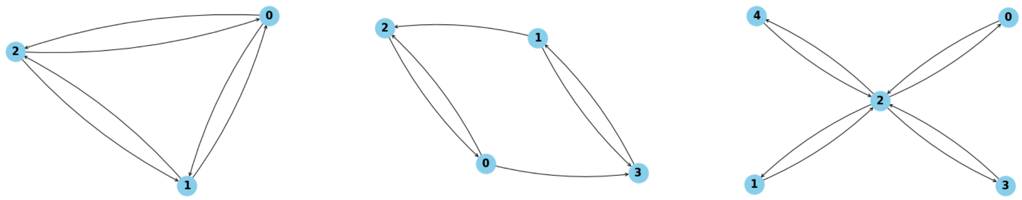
(b) Square Motifs



(c) Star Graphlet

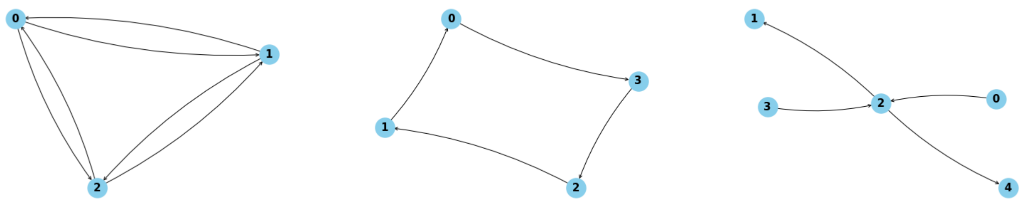
Figure 3: Tested Graphlets

The most popular directed triangular graphlets of Berlin and Philadelphia are equal, but we can say that by looking at the most popular directed square and star graphlets in Berlin are all in one direction of travel, in Philadelphia them are bidirectional. Through this result we can assume that Philadelphia has wider streets than Berlin that allow two-way traffic; this could also be due to the fact that the Berlin dataset only concerns the city centre, while Philadelphia also contains suburban neighbourhoods. By looking at the count of these different graphlets we can say that Berlin has much more triangular and star graphlets than Philadelphia, the number of square graphlets are very similar, this comparison is done for topological reason.



(a) Triangular Motifs (b) Square Motifs (c) Star Graphlet

Figure 4: Graphlet Philadelphia



(a) Triangular Motifs (b) Square Motifs (c) Star Graphlet

Figure 5: Graphlet Berlin

6 DETAILED CONTRIBUTION OF EACH MEMBER

The results we obtained were completely successful for us. Even if this project was quite challenging, we achieved good results and managed to finish before the schedule time. In order to finish the project, we made approximately 15 video calls of 1:30 each. In these video calls we brought our part of the project and brought everything together, combining ideas and discussing possible changes. All members of the group developed the available code. After all we can say that we divided the work to be done very well. Here we write the principal points that were computed by each member:

- TOMMASO BIANCO (2037864): he worked on Directed Graph's algorithms, on the report and on the documentation.
- MATTEO VILLANI (2090299): He worked on the plot of the results for better interpretation of the results obtained, he put together all the code in one ipynb file, he worked on the documentation.
- FEDERICO VIOLIN (2061746): He worked on Undirect Graph's algorithms, he searched the datasets and worked on the documentation.