Urban Street Networks, a comparative study of Portuguese cities

Bruno Casteleiro
(up201505347@fc.up.pt)
Pedro Ribeiro - Teacher advisor[1]
[1]*DCC, Faculdade de Ciências da Universidade do Porto*

In this work we use centrality measures in order to compare a set of Portuguese cities. These measures are very important in the study of complex networks since relevant structural aspects of these networks can be studied and understood. By using the 3 most simple centrality measures, namely degree, betweenness and closeness, $4km^2$ samples of 5 Portuguese cities are compared in order to find relevant aspects and patterns of each city and, by analysing statistical graphs, find relations between cities.

I. INTRODUCTION

Centrality measures are used to quantify how important are the nodes of a network, and this was always the main focus for using them. However, nowadays, attentions are more focused to the distribution of centrality across all nodes of a network, just like a shared resource, and thus allowing to relate different networks by analyzing these distributions. In this study, we make conclusions based on the 2 types of analysis described, meaning that we will both observe the importance of nodes and relate that to the actual geographical place it represents and also observe the different centrality distributions in order to find city relations.

This analysis take primal representations of each city as its networks. In this type of representation, street intersections represent nodes and streets represent links between nodes. Favoring primal over other representations makes it possible to consider real street distances. Other commonly used representations, like the dual representation, convert streets to nodes, losing the notion of continuity and thus losing the notion of real distance.

II. DATA COLLECTION AND PROCESSMENT

As in any network study, having data, and a source to get it, is essential. In an optimal scenario, the choosed data source provides not only complete information to work with, but also correct information. Although heavily relying on crowd-sourcing - thus having some inconsistency here and that -, we found OpenStreetMap (OSM) to be a sufficiently robust and well formatted data source to use for this study.

To make the process of downloading data from OSM easier, the Overpass API was used. Overpass "acts as a database over the web"[1] for OSM, and give us the possibility to query for data in a variant of the so desired JSON format, the Overpass JSON. Using this API, it is still necessary to manually discover the boundary box coordinates of the desired region one wants to download and report it to our script by an external file, however, its now possible to download various regions at the same time and in JSON format, which is more parsing friendly than many other formats available by default.

Having downloaded every city's OSM data, its then only necessary to create the corresponding Network. To OSM, streets are considered "ways", collections of ordered nodes and some tags detailing the street. Moreover, nodes also contains important information such as its latitude and longitude. Since we wanted to study the cities in their primal form, thus preserving its geographical properties, our parse algorithm simply convert each OSM node into a graph node and each valid OSM way in edges that connect sequentially from node to node. Here, a valid way is every street of some allowed type, namely primary, secondary, tertiary or residential, so a filter on the document ways is first applied. Figure 1 shows Faro's raw network. Nodes and edges are then finally saved in different csv files.

Since Gephi, our choosed tool for network analysis and visualization, doesn't fully support weighted Networks, a final additional processment over the Network is needed in order to compute both Closeness and Betweenness centrality values for each node. This is accomplished with R scripting, using the TNET[2] library. Also, R was the selected tool to compute the graphics of Closeness and Betweenness cumulative distributions.

Each network was finally imported to Gephi for visualization and remaining analysis.

III. APPLICATION AND ANALYSIS

Starting from the very top of Portugal and coming all the way down, the 5 choosed cities were Viana do Castelo, Porto, Coimbra, Lisboa and Faro. The main reason for choosing this cities were their geographical position in the country, since we've tried to cover all the territory from North to South. Other city filter that doesn't fall under geographical reasons is difficult to apply in Portugal since every city belongs to the self-organized category of cities, meaning that there's not a city that has been fully organized under the control of some superior entity. Basic properties of each sample are listed in I, namely the number of nodes N, links K and average link length (street real distance) L. Its interesting to observe how, despite being a small city, Faro has a larger num-

FIG. 1. Raw network representation of Faro city

Sample	N	K	L
Viana	457	572	85.03
Porto	454	598	95.21
Coimbra	429	518	97.19
Lisboa	1089	1393	62.23
Faro	1123	1561	51.16

TABLE I. Basic properties of the primal network representations of each sample. N is the number of nodes, K the number of links and L the average link length.

ber of nodes and edges than, for example, Porto. This only means that Faro's sample ended up containing a lot more valid streets relating to the filter described at II, something that can be confirmed by the good amount of detail in Figure 1. On the other hand, there's also a substantial difference in terms of average link length, this time favoring Porto, from where we can imply that Porto's streets are way bigger than Faro's.

A. Centrality as a shared resource

To graphically show the distribution of node centralities across the network we've used colored maps in which every node is color-ranked by its centrality value. Figure 2 displays the color spectrum being used as the rank. The network links get to be colored too, but with a special color that is represented by the average centrality value of the link-connected nodes.

FIG. 2. Centrality color-rank, in which colors from left to right represent centrality values from the smallest to the biggest.

Figure 3, 4 and 5 show, respectively, the degree, closeness and betweenness distribution maps of Porto's network. The degree centrality is the simplest measure of the considered 3 and its based on the idea that a node is more important if it has a larger number of connections to other nodes. Since we're using primal representations, nodes are limited by geographic constraints, this means that other than recognizing the most important intersections in the network, there's no real relevance in the information provided by the degree centrality. Viana do Castelo is the only city that displayed an interesting degree map, because almost every node as high degree ranking.

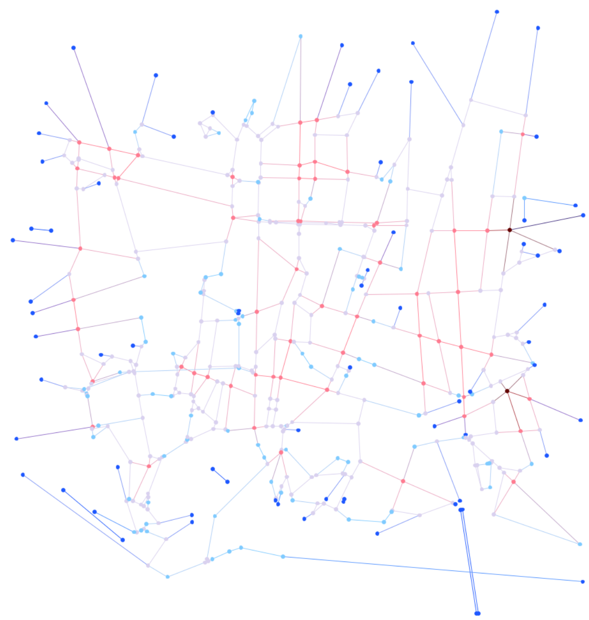

FIG. 3. Porto's degree centrality color map.

The closeness centrality is a measure that favors nodes that are closer to all other nodes in the network, or, by other words, it is a measure that highly ranks nodes that can reach others quickly. Surprisingly, and as seen in others studies[3] where central network nodes were favored across many samples, this centrality measure hasn't displayed any curious pattern in the Portuguese cities. Anyway, for Porto, the Avenida dos Aliados (presented at the center) is showing high ranking, which is a good sign. Moreover, Faro's closeness map has displayed the higher concentration of high rank nodes, which is another good sign of correct information since its average link length is the smallest of the 5 cities, which means that streets are closer to each other.

The betweenness centrality ranks nodes based on their presence in the shortest paths of all pairs of nodes in the network. This means that if a node is part of the majority of transactions between the other nodes, it will have a high rank. This last centrality measure presented consistent results across samples, allowing us to spot the continuity of important routes across various intersections and changes in direction.

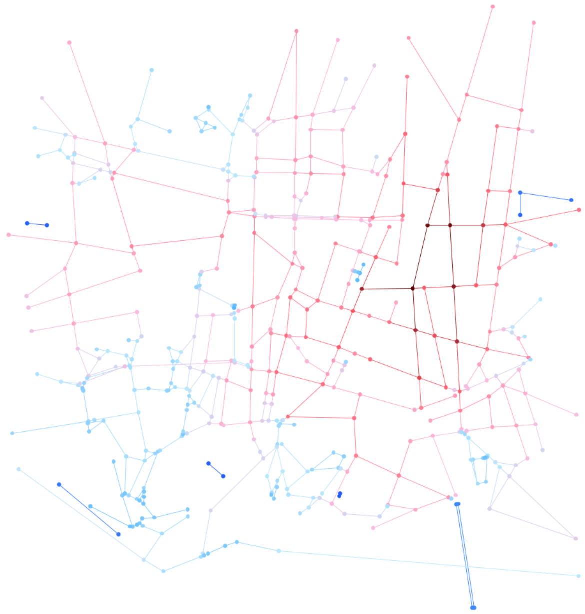

FIG. 4. Porto's closeness centrality color map.

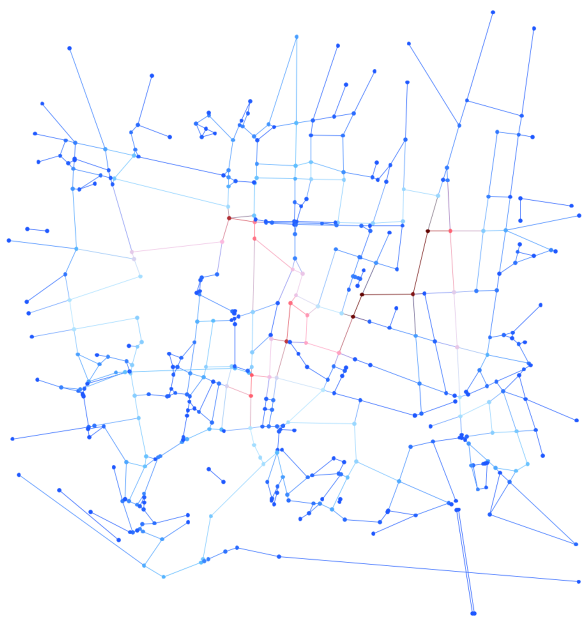

FIG. 5. Porto's betweenness centrality color map.

B. Statistical distribution of centralities

Attending to what as been said in III A about the degree centrality, only the cumulative distributions of both closeness and betweenness were taken in account. As expected, being all the considered cities categorized under the self-organized title, both centrality distributions are similar across all samples. Take for example Figure 6 and 7, representing Porto and Lisboa cumulative distributions.

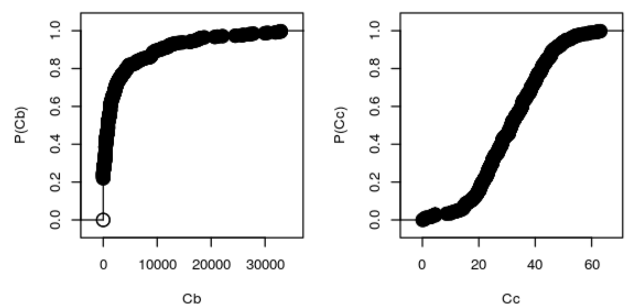

FIG. 6. Porto's betweenness and closeness cumulative distributions

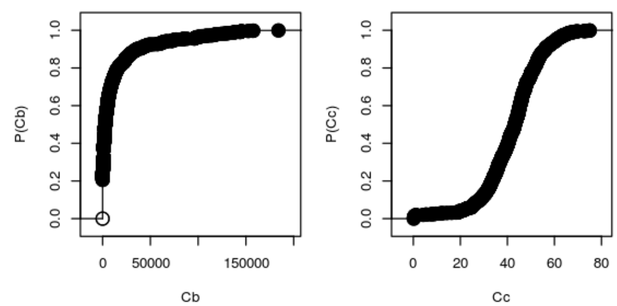

FIG. 7. Lisboa's betweenness and closeness cumulative distributions

IV. CONCLUSIONS

In terms of centrality distribution, every measure correctly revealed different notions of node importance when it comes to primal network representations. Also, as expected, the selected cities related to each other when comparing the statistical distribution of centralities, which means that there's no real difference when using different located cities in Portugal and comparing them using the same metrics used in this study.

[1] Overpass acts as a database over the web, `https://wiki.openstreetmap.org/wiki/Overpass_API`, 01 02 2018.

[2] TNET library, `https://toreopsahl.com/tnet/weighted-networks/node-centrality/`, 01 02 2018.

[3] Paolo Crucitti, Vito Latora *Centrality in networks of urban streets* 2006.