Project proposal

Analysis of bus public transport in Malaga using graphs

Jesús Moncada Ramírez

Learning from Networks 2022-23

Motivation

Public transport is currently becoming more and more important because of its necessity for **avoiding pollution** and the **increase in petrol prices**. First, they are fundamental to having good air quality in a city, the impact of the population's private vehicles always overcomes the pollution of public ones. Furthermore, due to the increase in energy prices (including petrol) in the last few months, governments are looking for a cheaper and global solution to the transport problem.

Urban Buses Map of Málaga

Once we understand the reasons why a good public transport system is essential, now we will say it should satisfy the following characteristics: multi-modal, informative, frequent, fast, and comfortable. These are fields that a council should take care of to create a convenient transit through the city and make the citizens choose public transport over their private vehicles. In this project, we are going to analyze some of these requirements (**frequent and fast**) in the bus system of the city of Málaga, Spain and we will propose improvements if some are detected. We will make the analysis of Málaga because it has been impossible to find the data for Padova.

Objectives

The motivation of the project heads to the following objectives:

- Create a **graph** representing all the bus lines and stops of the EMT (*Empresa Malagueña de Transportes*) (Transport Company of Malaga), the company in charge of the transport by bus in the city. (In the *Method* section we will explain how this graph will be built).
- Compute some **graph analytics at the node level**, for example, to identify the better-connected stops (related to centralities) or those that present a minimum waiting time…
- Compute some **graphs analytics at the graph level and network features**, for example, to identify the quality of the whole network (size related to frequency), to discover the size of the whole system (percentage of the land covered by buses), to detect neighborhoods that are more connected than others…
- Relate the results obtained in the node level and graph level analysis to some **intrinsic characteristics of the city** as the population in the different areas or the socioeconomic level. Check if that relationship is appropriate or if it could be improved.

- Analyze the presence of some **network patterns and motifs**, for example, to detect redundant routes or different ways of arriving at the same point...
- Do some experiments with **random graphs** and see how they could be useful in this topic. Compare some of the measures obtained with those computed for random graphs.
- Create an **analysis report with the conclusions** of the work, including the interpretation of all the computed measures.

Method

As has been explained we are going to build a graph. Traditionally, in most studies related to network analysis on transportation networks, **stations and stops are treated as nodes**, whereas the **edges link consecutive stations** along specific routes. Because of this, our graph will present multiple edges (different lines that connect the same two stops) but no self-loops (it doesn't make sense that a line connects a stop with itself), therefore it won't be a simple graph. We want it also to represent, somehow, the quality of relations between stops so we will need a **weighted graph**. In this way, a larger weight should represent a faster connection between two stops, while a slower one, a poorer one. The exact formula for the weights has still to be defined, but a first approximation could be: the weight of an edge is inversely proportional to the number of buses of that line in a time interval (a day, a week, the rush hour...) and inversely proportional to the distance between these two stops.

After developing the graph, those algorithms and concepts specified in the objectives will be coded and executed using **Python** programming language and **Jupyter Notebook**. A Jupyter Notebook containing the work may be a simple way of presenting the code, with explanations and results. The library used to maintain the graph data structure will be **NetworkX**. It is defined by its creators as "a package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks". It may be a good option because also includes functionalities to compute many graphs analytics and to display graphs graphically.

Dataset

The dataset used in this project will be obtained from Malaga's council open data site (*Datos abiertos, Ayuntamiento de Málaga*). On this web page, the administration of the city provides several CSV files about the bus system organization, all will be used in our python script. There is data about the following:

- Stops: information about every stop name and geographical coordinates.
- Horary: information about the schedule of every line.
- Calendar: information about the days of operation of the bus transport system.
- Routes: information about the stops that a line has in its path.
- Services: with data about every trip a bus in a line performs.

A phase of data pre-processing will be required first of applying our algorithms to create the mentioned data structures from the CSV files.

Intended experiments

The only planned experiment is the execution of the Python script created during the realization of the project. It will be executed in a laptop with processor Intel® Core™ i7, 16Gb of RAM, and Intel® Iris® Xe Graphics graphic card. The computing of this script shouldn't be a problem because there will be approximately 1000 nodes in the general graph (stops).

References

- Rahman, A. S., Magalingam, P., Kamaruddin, N. B., Samy, G. N., Maarop, N., & Perumal, S. (2020, May). Graph analysis study of a city bus transit network. In *Journal of Physics: Conference Series* (Vol. 1551, No. 1, p. 012004). IOP Publishing.
- Sara Cabodi. (2020). *Public Transport Network analysis* [Master's thesis, Dipartimento di Automatica ed Informatica Politecnico di Torino]. Webthesis. https://webthesis.biblio.polito.it/15242/1/tesi.pdf