

Analisis of the STCP transport network

Diogo Bogas

DCC/FCUP

December 12, 2016

Abstract

This theme aims to do an initial analysis of the complex network that represents all the bus lines belonging to STCP. The student will start by creating the network, with automatic data extraction from sources like the website stcp.pt. Then, with all the gathered data, there should be a preliminary data analysis with focus on node centrality, degree distribution and characteristic patterns.

0.1 Introduction

A network is a catalog of a systems components often called nodes or vertices and the direct interactions between them, called links or edges. Some good examples of networks, are social networks like Facebook and Twitter. This article uses networks to represent the services offered by a bus company operating in Oporto, with the intent of answering some questions like : how many bus stops are there, which one has the most lines going through it and how many times do i have to change bus so i can go from one side of the network to the other.

0.2 Data Gathering and Processing

0.2.1 Data source

The first decision made in this study wasn't what to study about the network, was to know were to get the data. The STCP website was a great place to start, as it has all the information on bus lines and bus stops. So, how do we get the information programatically? The best way to do it, is to read the answer the server gives the client when a request is made. Luckily, this answer is in form of JSON objects, which are easily parsable. The information the site gives about bus stops and bus lines is enough for the common user, but we need more. There are some aspects we need to derive if we want to do a complete study on the proposed network. Therefore, every piece of information was locally stored.

0.2.2 Data storing

While reading the information from the server answers, i noticed that each time i wanted to read the data i had to , programatically,connect to the site ,and that took too long to be viable.

A simple answer was to create two .txt files with the information gathered.

One of the files (AllLines.txt) had all the information about each busline. Each bus line is represented in a single line by a code, a direction, an integer that tells us the total number of stops in that line, and finally, all the stops that compose said line.

The other file (AllStops.txt) has all the information about all the stops. Each stop is represented in a single line by a stop code, an address, a zone,

a name and a pair of floating point numbers that represent it's geographical location.

All of this information can be refreshed, as there are methods that do so.

There is the need for some structures in which to represent all this data we have, so it can be easily used to answer some of the questions raised above.

Using the JAVA programming language, and after analysing the data, two structures came up as obvious:

- Stop, composed by:
 - a stopCode
 - an address
 - a zone
 - a name
 - a longitude
 - a latitude

The stopCode is the ID of a stop, as it is unique. Zones in this network are parts of the city, this will be clear on the visual part of this study.

- BusLine, composed by:
 - direction
 - code
 - description
 - pubcode
 - LineStops

The direction tells us which way a bus goes through the line. The code and pubcode are different things. The pubCode is, as the name suggests, public, and it is how we humans distinguish lines from one another. The code variable however, is the server way to distinguish bus lines. The description is a sentence that indicates where the line begins and where it ends. LineStops is the group of stops a certain line serves.

0.2.3 Data processing

0.3 Results