Internet Traffic Exploration in Milan

Andrea Cappello - cppndr96@gmail.com

Sorbonne Université, Paris (France)

Abstract—This paper is the summary of the work done for the Cellular Network 2020 course's project. It is an analysis of the data generated by *Telecom Italia*'s network during November and December 2013. In particular, in this work I created a visualization of the evolution of the traffic per each hour to help the analysis of patterns in the Internet traffic.

I. INTRODUCTION

To carry out a proficient analysis of data, it is important to visualize the data in a meaningful manner. Transform the data into forms easier to understand is fundamental to tell a story, to highlight patterns and exclude outliers [1]. In this context, the data utilized come from the initiative Big Data Challenge 2014 organized by Telecom Italia [2]. This work focuses on the analysis of internet usage in Milan.

II. DATA DESCRIPTION

The data cover two months of network utilization. It contains temporal and spatial information. Each day's usage is saved in a single file following the TSV format with 8 columns. In this columns, among other information, we find:

- the **square ID** which identify the cell in the city of Milan;
- the **time interval** which indicates the beginning of a 10 minutes interval, the time is counted in milliseconds since 01/01/1970 UTC;
- the **country code** of the user;
- the **internet** volume measured in the cell and generated by the user identified by the country code.

The map of Milan is divided into 10000 square cells by a 100x100 grid as shown in Fig. 1. Each cell is assigned to a **square ID**.
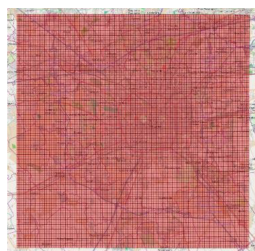

Fig. 1. 100x100 grid of Milan

III. ANALYSIS PROCESS

One of the first steps in every analysis should concentrate on the understanding of the data and its visualization. This process help to familiarize with the data and to understand what the data describe.

In this data exploration work, I focused on the Internet volume generated in the area of Milan. This can be considered the basis for further analysis since it provides a useful way to visualize the internet volume hour by hour over the two months. The final result is a video where each frame shows the network usage over one hour.

In Fig. 2, you can see an example of the outcome of this work. It represents a 3D plot of the grid that divide the area of Milan and it shows on a bar in the third dimension the volume of internet for the singular cell. The view point is placed south-west of Milan.
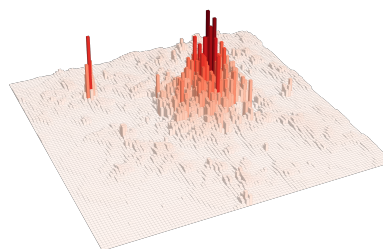

Fig. 2. 3D Internet volume - 06/12/2013

Fig. 2 represents the total traffic generated on 06/12/2013 and highlights as Milan downtown is essentially the fulcrum. Nevertheless, it is interesting to see the second spike that appears. It is nearby *Fiera Milano*, where important international exhibitions happen. This spike can be seen in other few days throughout the two months, always during or close to the weekend. From this visualization, we can infer that during those days, a big event was happening.

IV. CONCLUSIONS

As demonstrated in section III, this visualization can help to identify unusual events happening in the territory. This is only an example of how visualize data can help and further work can be done isolating the internet volume generate by cell ID and country code. By doing so, it is possible to have an intuition on places that most attracts foreigner tourists.

ACKNOWLEDGMENT

REFERENCES

[1] Tableau. Big data is here and we need to know what it says. https://tabsoft.co/2m4zfxy. 2020.
[2] Telecom Italia. The telecom italia big data challenge. https://theodi.fbk.eu/openbigdata/. 2014.
[3] Nicolas P. Rougier. Experimental 3d axis for matplotlib. https://github.com/rougier/matplotlib-3d. 2020.