

# Data-Driven Planning for Sustainable Tourism in Tuscany

Io Flament, Cristina Lozano, Momin M. Malik

Data Science For Social Good Europe Summer Fellowship 2017

7 August, 2017

## Introduction

The region Tuscany and the city of Florence are top tourist destinations globally. The metropolitan city of Florence (1 million inhabitants) is the choice of over 5M tourists per (year 2016), with the number of tourists growing year-on-year (2016 has a 10% growth compared to 2012). The city center alone receives ~3.5 Million tourists every year.

While tourism is a fundamental pillar of the economy of the city, there is a lively debate on the balance between the economic benefits of tourism, the impacts on the city fabric and resources, the degradation of communities and the loss of identity of places.

The project partner (Toscana Promozione Turistica) and data partner (Vodafone Italy) are providing anonymized data that will be used to identify patterns of tourism in time and space. The project is designed to:

- Support local authorities to understand and measure tourism through data, above and beyond traditional surveys and official aggregated statistics
- Support local authorities in exploring and designing solutions for sustainable tourism in the city

## Project Description

The city and the metropolitan area of Florence are home to a vast array of historical and cultural sites of exceptional value. Tourism tends to be concentrated at certain times and on a small number of iconic places, putting pressure on the infrastructure and quality of life. While the aggregate behavior of tourism is known, as well as the key data for specific sites (e.g. museums ticket sold) it is usually infeasible to reconstruct, measure and segment the end-to-end experience of visitors, such as:

- Where are they at specific times?
- Where do they come from?
- Where do they go next?
- How do they move within the city?
- How do they spend their time in the city? What do they do?
- Which patterns emerge from their visit experience?

The project is:

- Measuring the tourism flows and presence in space and time by using telecom traffic (CDR data), tickets of museums or monuments, and arrivals and departures at Florence

- airport.
- Identifying behaviors and patterns of tourism in space and time and categories of tourist movements
- Giving suggestions for how to link data and policy in terms of data collection and data sharing, and give possibilities for analysis and basing decisions on certain data

## Data and Method

The primary sources of data for this project are telecommunications data provided by Vodafone Italy and hourly museum logs per user provided by the Firenzecard. Some secondary sources of data are the hourly plane arrivals to the Florence airport, the hourly arrivals of cruise ships to the port of Livorno, daily tourist info point check-ins, as well as demographic information for the city of Florence. Each data source has its strengths and weaknesses: telecommunications data give relative numbers everywhere whereas museum logs, airports, ports, buses, info points, etc., give absolute numbers, but only in a few times and places. The goal is to link these data sources to get a better picture of some of the key tourist interactions in the city.

Telecommunications data has a unique ID per user, a timestamp for the call, and a lat/lon location. The lat/lon values are for the tower that picked up the call, not the actual location of the user. We can generate the estimated coverage zones for each tower by making a voronoi plot with these points. Once we do that, we can then pull out aggregated paths between tower regions for telecommunications for users with frequent calls per day (>15) to get a sense of the movement around the region. A similar study can be done for Firenzecard users using their museum entry logs. We can then compare the generated paths for the card and the telecommunications data to see if any correlations or patterns emerge.

## Key Deliverables

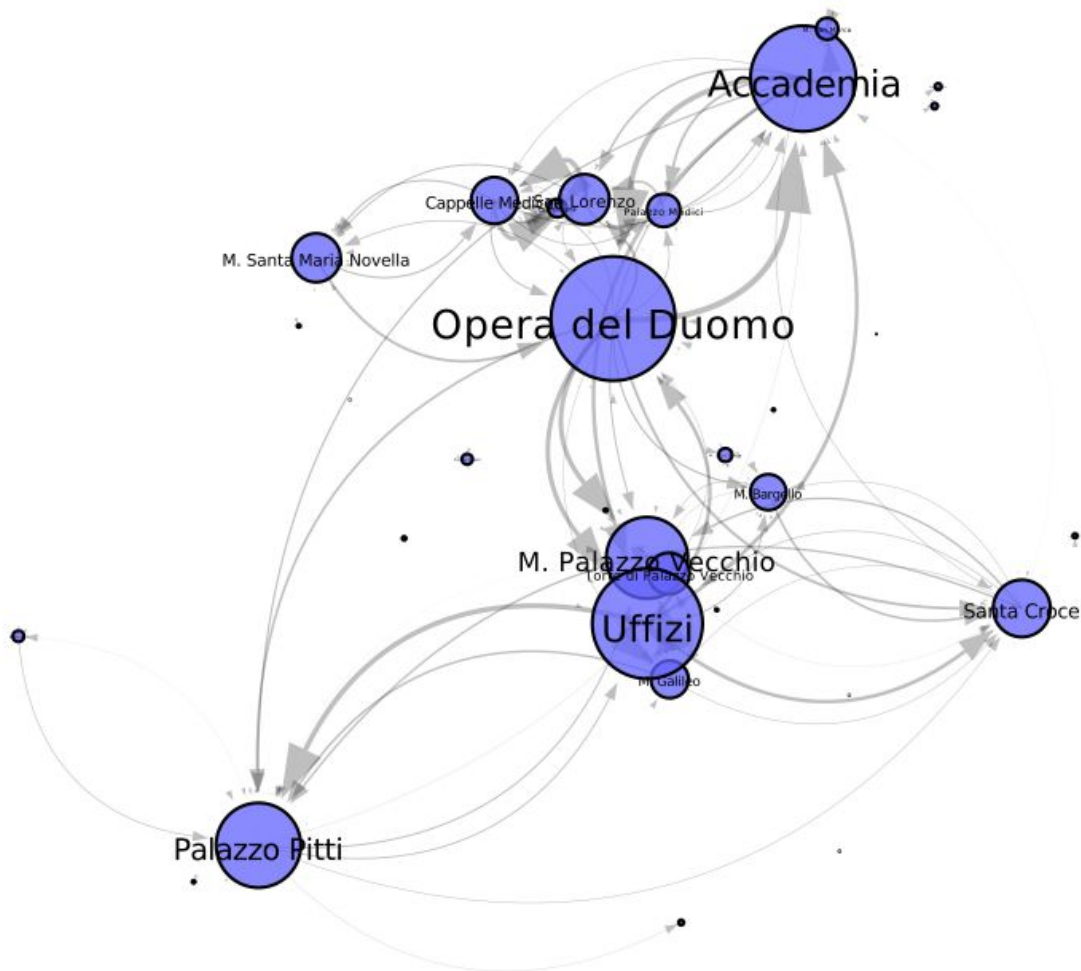
Since this project is intended to help Toscana Promozione Turistica further their understanding of the patterns of tourists, our deliverables primarily focus on key descriptive statistics and analyses that paint a picture of what the situation on the ground has been in recent history. Our goal is to give them better information about the current state of tourism to help them inform future tourism policy changes. These are the key deliverables for our project:

- Descriptive analyses are able to identify and measure, with reasonable accuracy, patterns that correspond to types of tourist behaviours
- A written report describing quantitative measures of tourist behavior in Florence and recommendations for future data practice and usage, with accompanying dynamic visualizations
- Recommendations formulated in ways that are actionable for the project partners
- Code used to generate the analyses is documented and can be reproduced by Toscana Promozione at a later stage

## Preliminary Results

Our first results have been summary statistics about busiest regions of the city per day, hour, and month for the telecommunications data as well as the museum logs. We have also calculated distributions of trip length based on the telecommunications presence. The goal is to summarize and visualize all of this information in a concise and easily understandable fashion.

More recently we have begun network analysis on the Firenzecard logs and locations. The example below shows the museum network where each museum is plotted in its geographic location and the size of the point corresponds to the relative number of people who visit each attraction. The size of the links in the network have width corresponding to how common it is to transition from one museum to the next. From this analysis, we can clearly see which attractions are most visited as well as some strong ordering preferences for how museums are visited within the day.



We plan to continue with analysis of the Firenzecard network as well as extend this type of visualization to the telecommunications towers.