

# MOBILITY ATLAS BOOKLET: AN URBAN DASHBOARD DESIGN AND IMPLEMENTATION

L. Gabrielli<sup>1\*</sup>, M. Rossi<sup>2</sup>, F. Giannotti<sup>1</sup>, D. Fadda<sup>1</sup>, S. Rinzivillo<sup>1</sup>

<sup>1</sup> Istituto di Scienza e Tecnologie dell'Informazione (ISTI), Consiglio Nazionale Delle Ricerche (CNR), Pisa, Italy

<sup>2</sup> Università di Pisa, Italy

**KEY WORDS:** Data Mining, Big Data, Urban Dashboard

## ABSTRACT:

The new data sources give the possibility to answer analytically the questions that arise from mobility manager. The process of transforming raw data into knowledge is very complex, and it is necessary to provide metaphors of visualizations that are understandable to decision makers. Here, we propose an analytical platform that extracts information on the mobility of individuals from mobile phone by applying Data Mining methodologies. The main results highlighted here are both technical and methodological. First, communicating information through visual analytics techniques facilitates understanding of information to those who have no specific technical or domain knowledge. Secondly, the API system guarantees the ability to export aggregates according to the granularity required, enabling other actors to produce new services based on the extracted models. For the future, we expect to extend the platform by inserting other layers. For example, a layer for measuring the sustainability index of a territory, such as the ability of public transport to attract private mobility or the index that measures how many private vehicle trips can be converted into electrical mobility.

## 1. INTRODUCTION

### Track Urban dashboard design and implementation

Typically, official demographic data are collected systematically every ten years, during the nationwide official census. However, census data, while very rich with information and details, have two major drawbacks: the temporal lag between census, during which there is no information on mobility, and the focus on systematic mobility, i.e., the individual mobility which happens almost every day with recurrent modalities and frequency, leaving out an increasingly relevant segment of non-systematic mobility, which, by its nature, is difficult to capture with traditional methods. Thanks to Big Data, we can thus increase our analytical capability with an informative base that can be updated almost continuously, and that includes all presences and not only the systematic one. Our proposal to overcome the limitations of traditional approaches is to use mobile phones because mobile devices are today one of the principal means by which people disseminate digital tracks of their everyday activities. In particular, mobile phones and the data they produce revealed to be a high-quality proxy for studying people mobility in different domains, such as environmental monitoring (Lane et al., 2010, Quercia et al., 2010), transportation planning (Calabrese et al., 2011), smart cities and social relationship analysis (Eagle et al., 2009, Wang et al., 2011).

The broad availability of location-aware services allows the collection of a vast repository of movement data. These new sources of data give an unprecedented opportunity to have a social microscope of the individual, collective, and global behaviors. We focus on mobility data, such as mobile phone data or such as the GPS tracks from car navigation devices, which represent society-wide proxies of human activities. These big mobility data help us understand human mobility, and discover the hidden patterns and profiles that characterize the trajectories followed by individuals during daily business. Mobility Data Mining research fields have produced a wide set of analytical methods to analyze, transform, aggregate and interpret spatiotemporal data.

\*Corresponding author – lorenzo.gabrielli@isti.cnr.it

The process to transform raw mobility data into high level knowledge usable by domain experts and decision makers is usually very complicated, and it requires expertise and experience that are difficult to find and enroll. The majority of methods of Mobility Data Mining (MDM) are centered around the concept of trajectory, i.e., a sequence of spatiotemporal locations of a moving object over time. To leverage the power of MDM methods and to guarantee a broad diffusion of the analytical results, we propose a new paradigm capable of bridging the gap between the complex models derived from the analytical methods and the accessibility of content of non-expert users.

We propose a paradigm where complex analytical processes are summarized into a set of quantitative estimators of the main properties of mobility in a territory. We call such estimators mobility indicators and, for each region, we propose to compute a selection of measurements to provide a general overview of the mobility in the selected place. Here, we illustrate Mobility Atlas Booklet, a powerful analytical service for policymakers, businesses, public administrations, and individual citizens. The tool makes territorial information accessible through an API system and a set of easily navigable dashboards.

The availability of the huge amount of CDRs allowed to carry out an extensive experimentation over Tuscany, investigating how people use and live one of the biggest Italian regions. Furthermore, the period under analysis is fascinating because was the theater of many cultural and recreational events which attracted people from surroundings and more distant locations. It is worth pointing out that if we count individuals present on the territory at the city level (e.g., the whole City) we are unable to observe peaks. For this reason, we have decided to select some areas of the city, where POIs are present, in which to specialize the analysis.

## 2. THE MOBILITY ATLAS BOOKLET

### 2.1 Problem definition

While some statistics about systematic movements may also be extracted from census, this is partially true for occasional move-

ments. Due to this, for instance, it would be interesting to know who has attended an event and where they come from or how visitors are attracted in certain municipalities. This would enable to know the spread and importance of an event by measuring the attractiveness over the surrounding territory. The amount of mobility that is created for occasional reasons is impressive and certainly higher than that happening systematically and/or due to work activities. These are important destinations for tourism by Italian and foreign citizens. A difference with the movements for working activity is that not only the surrounding municipalities but the individuals of quite all the towns occasionally travel to the major cities. Individuals occasionally travel to many more places than those they visit for working reasons.

The main observed limit of traditional data sources is the difficulty in finding data continuously over time and space. Many dimensions of our social life have big data proxies, as the mobile calls data for mobility (Liu et al., n.d.). To overcome this limit, we propose the use of mobile data to measure the presence of people because almost the entire population has at least one mobile device and uses it at every stage of the day.

## 2.2 The Sociometer

The Mobile phone data can be processed using various data mining techniques. In our case, we propose to pre-process the mobile phone data with the methodology Sociometer (Gabrielli et al., 2015) in order to extract information on the type of users present in the area under study. The Sociometer is a Data Mining process able to classify the call behavior of people to quantify the different type of city users within a territory (Figure 1). The basic statistical unit of our analytical process is the Individual Call Profile (ICP). ICPs are the set of aggregated spatio-temporal profiles of a user computed by applying spatial and temporal rules on the raw CDRs. The structure is a matrix where the temporal aggregation is by week, where each day of a given week is grouped in weekdays and weekend. Given for example a temporal window of 28 days (4 weeks), the resulting matrix has eight columns (2 columns for each week, one for the weekdays and one for the weekend). A further temporal partitioning is applied to the daily hours. A day is divided into several timeslots, representing interesting times of the day. This partitioning adds to the matrix new rows. Numbers in the matrix represent the number of events (in this case the presence of the user) performed by the user in a particular period within a specific timeslot. The analytical process consists of several phases. The first phase involves the *ICP Building*, then we group similar ICPs (*Prototypes Extraction*), the third part of the process involves labeling the centroid of each cluster computed in the previous phase w.r.t. the definitions introduced above (*Prototype Labeling*). The last phase is the *Label Propagation*, each point for each cluster is labeled propagating the value of its stereotype. At the end of the process, we can quantify the stock of individuals present in the area and flows of individuals among the different regions.

Next, we post-process the Sociometer's results to calculate the daily distributions of presence stratified by population category (resident, worker, visitor). We also applied a post-processing step on the class *Visitors* to distinguish the short-term visitors and people in-transit. We called them *Passing by*, i.e., users who made a single call in all the period, and thus we registered their presence only for a single day.

The main advantage of our proposal is to provide high level knowledge extracted from the raw data using sophisticated analytical

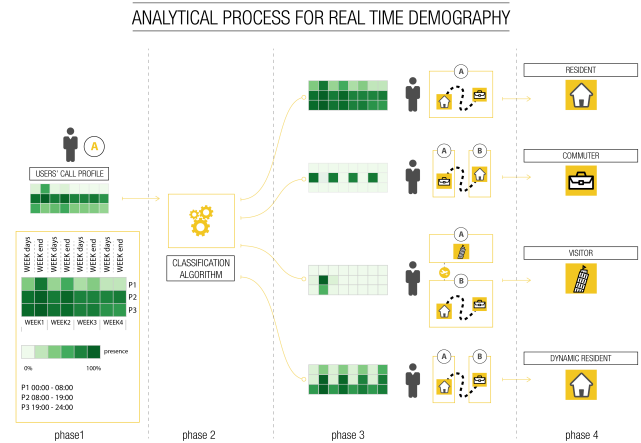


Figure 1. Sociometer: Starting from raw call data record, in phase 1, we first build for each user, for each zone an Individual Call Profile(ICP). Then we apply a clustering algorithm to group users with similar behavior (phase 2). From each cluster, we extract a centroid (phase 3), and we label it w.r.t. the closest representative archetypes (phase 4).

methods (Giannotti et al., 2011, Guidotti et al., 2016), without requiring to the domain expert and decision maker the capabilities of managing the data herself, and to enable her to evaluate how many persons were attracted by a particular event. Specifically, we can find that the increase in presence is mainly due to residents or to people who occasionally visit the city. This information helps to verify the attractiveness and diffusion of the event itself and therefore to implement security plans. The proposed approach brings together assessed methods of Mobility Data Mining, (Giannotti et al., 2011) and modern visual analytics techniques for data exploration (Rinzivillo et al., 2008, Andrienko et al., 2011).

The calculation of the time series raises another question: what is the spatial granularity with which to calculate them? One of the problems faced during the development of the work was the choice of the minimum level of spatial granularity. The studies carried out so far have focused on the municipal level, but it is evident that the chosen level is not adequate. To carry out an appropriate analysis of the territory it would be useful to be able to analyze it at the district level. In this sense, we propose to use a new data source calculated by the Land Registry Agency for statistical purposes.

## 2.3 Architectural design

After the analysis of the requirements we identify the dimensions, the measurements and the metrics to be calculated to address the analysis questions (Tab. 1). We have modeled three dimensions: *User, Date e Region*. After describing the characteristics of the analysis dimensions, we report the design of the Data Warehouse *Presences* (Fig. 2).

*User* contains the user category obtained by applying the Sociometer algorithm to the mobile phone data (Gabrielli et al., 2015). The categories obtained are as follows: *Resident, Commute, Dynamic resident, Visitors* and *Passing by*. The Sociometer is a Data Mining process able to classify the call behavior of people to quantify the different type of city users within a territory. The analytical process consists of several phases. The first phase

involves the *ICP Building*<sup>1</sup>, then we group similar ICPs (*Prototypes Extraction*), the third part of the process involves labeling the centroid of each cluster computed in the previous phase w.r.t. the definitions introduced above (*Prototype Labeling*). The last phase is the *Label Propagation*, each point for each cluster is labeled propagating the value of its stereotype. At the end of the process, we can quantify the stock of individuals present in the area and flows of individuals among the different regions.

*Date* allows to navigate the analysis with respect to a time hierarchy. *Date* has two hierarchies: the first allows to analyze the presence up to the daily granularity (year, month and day); while the second allows you to measure the average weekly presence (day of the week, our).

*Region* allows us to navigate the analysis with respect to a spatial dimension. The granularities chosen range from the regional level up to the minimum granularity, very similar to the city districts. The choice of the minimum granularity is linked to the need to divide the municipal territory into areas larger than the census areas and as similar as possible to the administrative districts. The territorial subdivision has been obtained through the intersection of different official data sources. The administrative subdivision is derived from the ISTAT data sources, while the subdivision by districts is derived from fiscal and cadastral data<sup>2</sup>.

Requirements	Dimensions	Measures	Metrics
How many people are present within the territory?	User, Region	Number of users	cardinality of IDs, aggregate w.r.t. the area of analysis.
which is the intended use of the territory?	User, Region	Number of users	cardinality of IDs in a region, w.r.t the users categories.
What are the periods in which the territory is most visited?	User, Region, Date	Number of users	cardinality of IDs w.r.t. category and time.

Table 1. Dimensions, measurements and metrics necessary to meet our requirements.

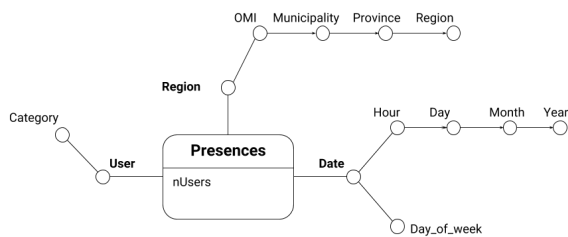


Figure 2. Data Warehouse Design.

**Fact Table: Presences** The Fact of analysis is defined as the number of *Presences* detected with minimal spatial granularity the neighborhood and as minimum time granularity the hour. The analysis measure *nUsers* is derived as the distinct number of users who have made at least one call with respect to the minimum spatial and temporal granularity. Please note that the calculation of the number of users detected is stratified according to the categories provided by the Sociometer.

<sup>1</sup>Individual Call Profiles (ICPs) are set of profiles of a user computed by applying spatial and temporal aggregation on the raw CDRs. The temporal aggregation is by week, where each day of a given week is grouped in weekdays and weekend.

<sup>2</sup>[http://www.agenziaentrate.gov.it/geopoi\\_omi](http://www.agenziaentrate.gov.it/geopoi_omi)

## 2.4 Population of the DataWarehouse

The CDR dataset used for the analysis has a considerable size, includes about 70 million CDR, from which we extract 3.5 million ICP for 1.5 million people. The period of analysis covered the month of November 2015. The population of the DataWarehouse requires the junction between the various data sources including geographic information (antennas, spatial subdivision) and the category of users. Table 2 shows the final result.

timestamp	OMI	caller	duration	label
2015/11/10 23	B1 - Pisa	4F80460	120	visitor
2015/11/10 01	B1 - Pisa	2B01359	10	resident
2015/11/10 01	B1 - Pisa	2B19935	301	visitor
⋮	⋮	⋮	⋮	⋮

Table 2. Final dataset, containing the starting data together with the spatial information, the user profile obtained from the Sociometer and the time hierarchy.

## 2.5 Implementing the Data Warehouse

We have made different versions of the platform as our system is constantly evolving. The first version stores the data on a standard DBMS. Another version of the system, currently in the alpha test phase, implements a distributed data storage system using the Hadoop Framework. Thanks to the use of the primitives provided by the PySpark module we can carry out the operations of large amounts of data efficiently.

In the second case, the output of the Data Engineering process has the structure shown in the Table 3. In Table 3 we decide to materialize in a single record all the information concerning the analysis and the dimensions, including the hierarchies. The *Date* dimension contains information about the time, day, month, year, and day of the week. The dimension *Regions* it shows the whole space hierarchy, starting from the district up to the region. Finally, the calculated measure will be taken at the run time by calculating the distinct users contained in the field *ids*.

year	month	day	day of week	hour	region	province	municipality	neighborhood	label	ids
2015	11	22	5	7	Toscana	Firenze	Signa	R1	Resident	15...

Table 3. Structure of the records stored in the distributed data storage system.

## 3. WEB ANALYTICS

The Mobility Atlas Booklet (MAB) has been implemented as a *web application* that allowed to visualise interesting indicators related to the presence of different categories of users within a selected geographic area. This information will be available at different spatio-temporal granularity.

To increase information about an area, we extracted the set of Central Italy Points of Interest (POIs) from the Geofabrik DB<sup>3</sup>. These POIs can be used to understand the expected type of users. For instance, if most of the people flows into the analysed area in the evening we can hypothesize that this area is residential. Given the aim of our study, we chose to consider just four categories: *Hotels, Restaurants, Sport, Tourism*.

<sup>3</sup><http://download.geofabrik.de/index.html>

Figure 3 shows the elements involved in the process that brings from memorised data to their graphical visualisation on the web platform. The API layer described in 3.1 is at the centre of the figure. It provides the access to the data through a request system based on some parameters provided in input. The designed layer returns aggregated data in a format suitable to represent views of interest. We developed two specific functions (*getStat*, *getOd*) for the API system, in order to provide aggregated data related to the presence time series and origin destination matrixes.

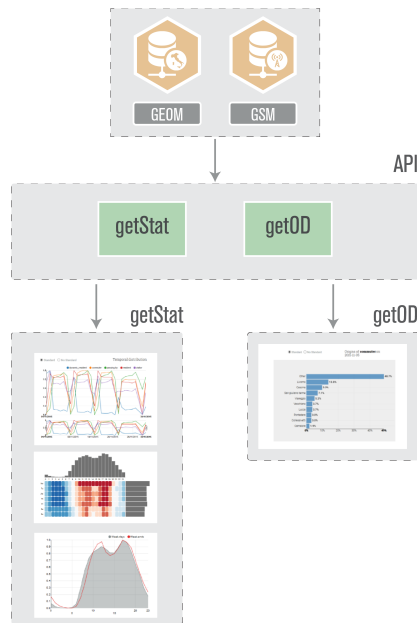


Figure 3. Graphical representation of the interactions between APIs and the data archiving system, and between APIs and the analysis showed on the web platform.

### 3.1 Layer API

Restful APIs provide access to data while computations are carried out through map/reduce libraries provided by the PySpark layer. The results of the computations performed by the function are saved in a cache memory for a faster check of the recurrent computations.

### 3.2 *getStat* function

The API function *getStat* returns the presence time series in a given area with respect of the day of the month, the day of the week and the time.

To improve peaks detection, we rescaled the distribution by normalizing it w.r.t. the typical daily distributions as proposed in (Furletti et al., 2017). The normalization procedure foresees the computation of the typical distribution of a week for each time series obtaining two values on a weekly basis for each day:  $avg^n$ ,  $std^n$ . The  $avg$  is the average number of distinct users for the  $n$ -th day of the week (0 = monday, 6 = sunday) and  $std$  is the standard deviation of the same day. Using those values we rescaled the time series as follows:

$$v_{normalized}^d = \frac{v^d - avg^n}{std^n}$$

where:  $n$  is the relative day of the week of the absolute day  $d$ .

The method we use in your time series analysis is introduce only presenting our model to the final user. The scaling is used to represent the time series on the same reference scale, allowing the end user to observe the variation of the signal, for each category, over time. In addition, the user can study the distributions, separately for each POI covered by at least an antenna, and observing several peaks confirming if events were present.

**Input parameters** The input parameter for the *getStat* function is the spatial hierarchy of the area of which we want to know the users inflow. Based on the spatial level analysed, we used 4 different formats for the input parameters (Table 4). The function checks the accuracy of the provided parameters.

Spatial level	Format input parameter
Regional	/region
District	/region/district
Municipality	/region/district/municipality
OMI	/region/district/municipality/omi

Table 4. description of the different input formats that the *getStat* function can take on. The format changes based on the level of spatial granularity required.

*GetStat* function returns a JSON object with two aggregates. Each piece of information is calculated with respect to the different population categories measured with the Sociometer technique.

The first time series represents the average weekly presence with respect to the days of the week, the hours of the day and a combination of the two (for instance, the combination 23 and 7 will include the count of users at 11 pm on the seventh day of the week Sunday). The second time series returns, for each user category, the count of presences with respect to the tuple year, month and day.

### 3.3 *getOD* function

For each time series it is possible to visualise the number of persons present. An interesting feature that we offer is the chance to go back to the residence of a given group of users who are present in an area in a given time moment.

**Input parameters** The *getOD* function has two input parameters: the level of spatial granularity for which statistics is required and the time moment with respect to such statistics is calculated.

The *getOD* function returns a JSON object which contains information on the residence of users who were present in the analysed days. Knowing their residences, it is possible to build an origin destination matrix that highlights the origins of the presences.

**API Rest** The access to the dataset are managed by a RESTful interface, exposed as a set of URLs. The selection and projections on the database can be controlled by means of parameters passed within the URL. Each call returns the resulting data as a JSON document. The RESTful functions are available at the following links:

[http://kdd.isti.cnr.it/mab\\_api/getStat/users/<region>/<district>/<municipality>/<zone>](http://kdd.isti.cnr.it/mab_api/getStat/users/<region>/<district>/<municipality>/<zone>)  
[http://kdd.isti.cnr.it/mab\\_api/getOD/users/<region>/<district>/<municipality>/<zone>?timefilter=YYYY-MM-DD](http://kdd.isti.cnr.it/mab_api/getOD/users/<region>/<district>/<municipality>/<zone>?timefilter=YYYY-MM-DD)

where <region>, <district>, <municipality>, and <zone> are placeholders to specify parameter for each level of the geographic hierarchy (i.e. respectively Toscana, Pisa, Pisa, B1). The function `getOD` has an additional parameter to specify the time constraints for the selection (i.e. 2015-11-02).

### 3.4 Layer Analytics

The *web application* has been conceived in order to allow users to navigate areas at different levels of spatial granularity. For such task, the application provides an interactive map that allows to select areas for which we want to visualise information about presences. This map is loaded through Leaflet, a mobile-friendly library written in javascript used to build interactive maps. Its usability and flexibility allowed us to manage spatial layers through the selection on the map or through the drop-down menu in the website header. The interaction between the application and the user as well as that between different views is managed via Javascript/JQuery functions. These functions enable navigation of thematic views through the selection of the area of interest. Moreover, they manage the coordination of the different views: filters applied to the main view will also be applied to the related views.

The analytics layer interacts with the APIs mentioned above to represent results returned by `getStat` and `getOD` functions through the use of views. The final product shows up as a single-page *web application*, developed using web technologies that allow us to visualise information of interest at different levels of spatio-temporal granularity.

### 3.5 Mobility Atlas Booklet at work

In this section we show how we implemented the interaction between the final user interface and the underlying analytic engine. Specifically, we describe how each object in the view is calculated and can be used by the final user. Based on the selected area, the application sends a call to the `getStat` function provided by the API, with the hierarchy related to the chosen area as the input parameter.

**Temporal distribution** A line chart shows the results returned by the `getStat` function, related to the presences with respect to the day of the month. This kind of representation can properly highlight possible presence peaks in the *time series*.

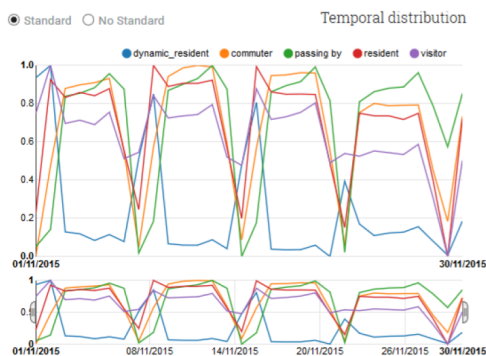


Figure 4. Temporal distribution. Example of a distribution of the presences in the analysed area based on the day of the month, stratified for user categories.

Figure 4 shows an example of a normalised daily distribution of the presences for the different categories of users, for each of

them we used a line of a different colour. The radio button within the chart allows to visualise non-normalised results of this analysis, providing a reference to the real figures recorded. The views legend enables the selection/deselection of the user categories in the main chart and in the related ones as well. This feature allows us to analyse the distribution of presences for different user categories directly on the *web application*. Below this chart there is a dashboard for the selection of a different time window, allowing to focus the analysis on a specific time moment.

Figure 3 indicates the presence based on day of the week, which indicates that the group called dynamic residents appear to be active only on the weekend, instead all other groups showed remarkably similar patterns. This pattern is often repeated on points of interest that do not have particular events. In other cases it is possible to notice significant variations.

**Temporal Matrix** The `getStat` function returns information used to implement this view and to generate a dashboard showing the presences in the territory compared to a standard week.

Figure 5 shows the density of presences with respect to the combination of the day of the week and time. The colouring shows with cold hues a low number of presences, while a higher presence is marked by warm hues. The presences distribution is shown separately on the two axes that are related to the two dimensions mentioned above. The matrix linked to the main view shows the values compared to the user categories selected with the *Temporal Distribution* dashboard.

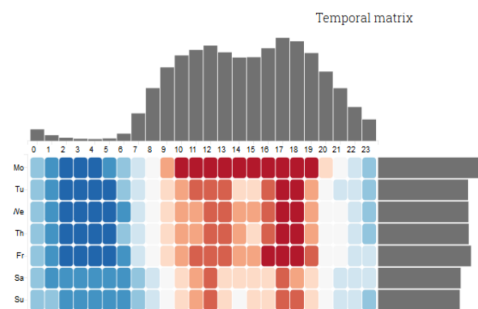


Figure 5. *Temporal matrix*. Time grid showing the representation of a typical week in the selected area.

**Daily distribution** The view called *Daily distribution* aims to show whether presences in the analyzed area display a different temporal distribution between weekdays and weekend. This view allows to classify the analyzed area as a residential area or a working area. The design of this view is inspired by the work of (Combes et al., 2017) and was realized reshaping the information returned by the `getStat` API.

Line chart in Figure 6 represents hourly distribution during weekday (grey area) compared to hourly distribution during weekend (red line). Also, this chart is linked to the main chart and the result can thus be shown with respect to the active user categories in the *Temporal Distribution* dashboard.

**Origin distribution** The main goal of this view is to show the residences i.e. the travel origin of users *visitors*, *commuters*, and *passers-by* observed in the daily distribution graph in Figure 4. In this case, the data represented are those returned by `getOD` function. The bar chart in Figure 7 shows an overview of the users origins in the selected category in the area on a given day.

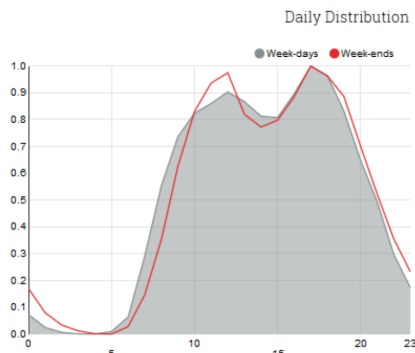


Figure 6. Daily distribution. Distribution of presences by time of the day, stratified for weekdays and weekends.

In this graph it is possible to decide whether to display the results normalized or not. The graph shows aggregated origins at the provincial or municipal levels. This feature is useful to understand the origins of admittances during a busy day, for example to measure the attractiveness of a given territory compared to the surrounding areas, for both tourist and business scopes.

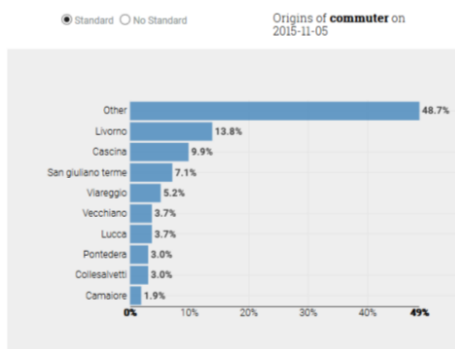


Figure 7. Origin distribution. Distribution of origins for different users categories on the selected day. The horizontal bar chart shows, on the y-axis, provinces/municipalities of residence.

**POI Tool** We provide an additional tool to improve understanding of the territory. Each territory is characterized by the Points Of Interest it contains. If a territory is mainly industrial, it will not show tourist attractions, whereas a historic center will be full of restaurants or monuments (Furletti et al., 2013). By turning this layer on, it is possible to see the density and types of Points Of Interest, thus allowing for a better explanation to distributions of presences, especially in presence of peaks. With this feature enabled, it is possible to see on the map the POI categories that belong to classes *Hotels*, *Restaurants*, *Sports* and *Tourism*. The POIs are shown on the map as red dots, as in Figure 8.

#### 4. USE CASES

In this section we present two relevant use cases of our approach using the *web application* available at the [http://kdd.isti.cnr.it/mab\\_omi/](http://kdd.isti.cnr.it/mab_omi/). The web application is designed to provide an overview of several spatio-temporal aggregates over an area selected by the user, by navigating and selecting administrative boundaries. We demonstrate the capabilities of the web application by analyzing the mobile data collected in Tuscany in November 2015, highlighting the decision support the application can



Figure 8. Map enriched with *layer point of interest*. In the figure, the layer containing hotel category is highlighted.

provide to a decision maker to better understand how presences on her territory are distributed.

There are a large set of possible analytics we may explore. For this paper we are focusing on two main objectives. First, the analysis of the impact of very large events on the urban territory. Here we are focusing on the *Lucca Comics and Games* event, a yearly event that attract people interested in comics and games. In the second use case we compare the normal behavior of two districts in a city, each with a different functional destination, the sea district and an industrial district in Pisa. These two cases are chosen to highlight the features of the web application. The reader may refer to the web site to explore other geographic context in Tuscany.

#### 4.1 Lucca Comics&Games 2015

*Lucca Comics and Games* is one of the most relevant events for comics and games. During the period of collection of our mobile data, the event was held in Lucca from October 29th and November 1st 2015. This event is interesting for the large amount of people that is able to attract in Lucca during that week. Our application is able to provide a precise point of view to estimate the impact of visitors in the city and the potential impact on the neighboring places.

Since we want to analyze the number of presences along time, we may take advantage of the **Temporal Distribution**. The graphical widget (Figure 9) shows a pick in the presence of *visitors* during the period of study. Outside the temporal window of the event, the series of the presence of visitors falls back to standard values in the rest of the month. Other categories of users, are not affected by the event.

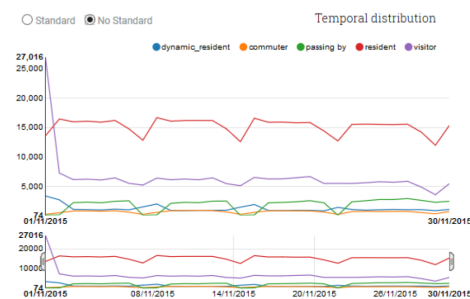


Figure 9. *Temporal Distribution* of presences in the municipality of Lucca.

From the **Temporal Matrix** we may observe the evolution of presences during days of week and hours of day. The impact of the large event is evident and affects both weekdays and weekend (Figure 10). *Visitors* are mainly concentrated on Sunday,

from 9am to 7pm, the time interval where the festival is open. *Residents*, on the contrary, do not show any perturbation in their usually profile of presences. This proves how the tool is able to capture the attraction versus external visitors.

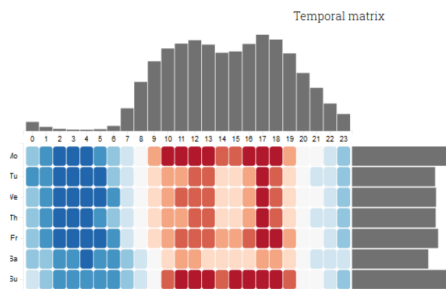


Figure 10. *Temporal Matrix* for Lucca municipality.

By considering the city as a whole, we may notice how the festival produces an increase in the number of the visitors mainly in the centre of the city. In the periphery of the city, the number of *visitors* is below the number of *residents*, that maintain a coherent series as the rest of the month. In the city centre, the presence of *visitors* is more than twice the number of *residents*. The combination of spatial and temporal indicators allows us to conclude that the event has an impact mainly concentrated in the historical center of the city.

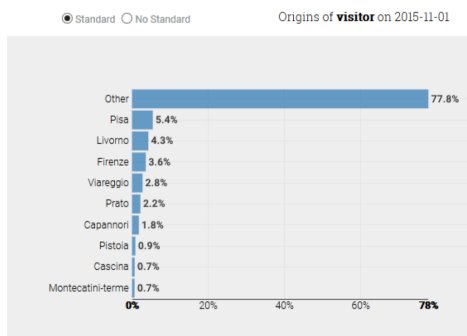


Figure 11. Municipalities of origins of presences.

To verify the range of attractiveness of the event, we use the **Origin distribution** to evaluate the origin of the *visitors* arriving in the city during the event. We can see that the event attracts a large part of visitors coming from "Other", i.e. outside the Tuscany Region (Figure 11). "Other" in this case means that the residence is not calculated from the available data sets then it follows that they live outside Tuscany region.

#### 4.2 Land use: Industrial district and maritime district

In this section we focus how analysis to a higher granularity, by analyzing zones within a city to investigate their land use. In particular we focus how discussion on the comparison of two districts in Pisa, namely E1 and E2 zones. E1 is an industrial district, whereas E2 is the district of the city on the sea coastline.

Zone E1 does not contains point of interests for tourists or visitors. This is confirmed by the **Temporal Matrix** which shows a general increment in the presence during working hours and weekdays. During the weekends, there is a general decrement of presences in the area and it is interesting to note from the *Temporal Distribution* (Figure 13) how the different user profiles are

distributed in the area. Even if the area concentrates companies and services, it does not have a large number of commuters. This suggest an high frequency of worker that reside within the city. This conclusion is highly valuable for the mobility manager of the city, to better plan and deploy the public transportation service.

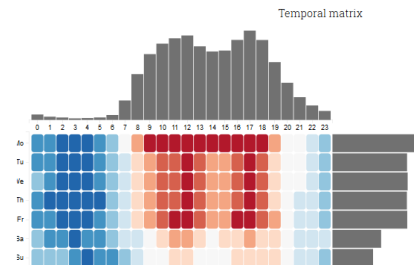


Figure 12. *Temporal Matrix* of Zone OMI E1.

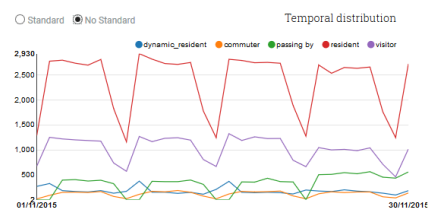


Figure 13. *Temporal Distribution* in Zone OMI E1.

Zone E2 has the peculiarity of been located on the sea side. Despite the period of the year non particularly warm in Italy, the **Temporal Distribution** shows peaks of presence of *visitors* during the weekends (Figure 14). The difference of behaviors in the two zones is more evident by comparign the **Temporal Matrix** of Zone E2. In E2, there are peaks of presence mainly during weekend and during lunch time in the weekdays (Figure 15).

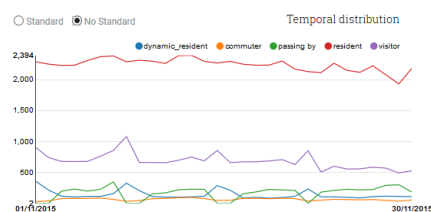


Figure 14. *Temporal Distribution* in Zone OMI E2.

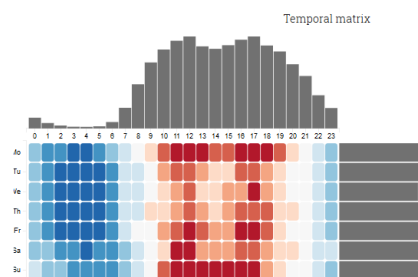


Figure 15. *Temporal Matrix* of Zone OMI E2.

## 5. CONCLUSION

The large availability of location-aware services allows the collection of a vast repository of movement data. Hence, in this work we have proposed an analytical platform that extracts information on the mobility of individuals from digital traces left by individuals by applying Data Mining methodologies. For each territory, we computed several measures to provide a general overview of the mobility through a web application. The resulting Mobility Atlas Booklet is a powerful analytical service for policymakers, businesses, public administrations, and individual citizens. The tool makes territorial information accessible through an API system, and a set of easily navigable dashboards. The result is twofold. First, we communicate information through visual analytic techniques that facilitate understanding of information to users who have no specific technical or domain knowledge. Secondly, the API system guarantees the ability to export aggregates according to the granularity required, enabling other actors to produce new services based on the extracted models. We will extend the platform by inserting layers for measuring the sustainability index of a territory, such as the ability of public transport to attract private mobility or the index that measures how many private vehicle trips can be converted into electrical mobility.

## REFERENCES

- Andrienko, G. L., Andrienko, N. V., Hurter, C., Rinzivillo, S. and Wrobel, S., 2011. From movement tracks through events to places: Extracting and characterizing significant places from mobility data. In: *2011 IEEE Conference on Visual Analytics Science and Technology, VAST 2011, Providence, Rhode Island, USA, October 23-28, 2011*, pp. 161–170.
- Calabrese, F., Colonna, M., Lovisolo, P., Parata, D. and Ratti, C., 2011. Real-time urban monitoring using cell phones: A case study in rome. *IEEE Transactions on Intelligent Transportation Systems* 12(1), pp. 141–151.
- Combes, S., de Bellefon, M.-P. and Vanhoof, M., 2017. Mining mobile phone data to detect urban areas.
- Eagle, N., Pentland, A. and Lazer, D., 2009. Inferring friendship network structure by using mobile phone data. In: *Proc. of the National Academy of Sciences*, Vol. 106number 36.
- Furletti, B., Cintia, P., Renso, C. and Spinsanti, L., 2013. Inferring human activities from gps tracks. In: *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*, ACM, p. 5.
- Furletti, B., Trasarti, R., Cintia, P. and Gabrielli, L., 2017. Discovering and understanding city events with big data: The case of rome. *Information* 8(3), pp. 74.
- Gabrielli, L., Furletti, B., Trasarti, R., Giannotti, F. and Pedreschi, D., 2015. City users' classification with mobile phone data. In: *Big Data (Big Data), 2015 IEEE International Conference on*, IEEE, pp. 1007–1012.
- Giannotti, F., Nanni, M., Pedreschi, D., Pinelli, F., Renso, C., Rinzivillo, S. and Trasarti, R., 2011. Unveiling the complexity of human mobility by querying and mining massive trajectory data. *VLDB J.* 20(5), pp. 695–719.
- Guidotti, R., Monreale, A., Rinzivillo, S., Pedreschi, D. and Giannotti, F., 2016. Unveiling mobility complexity through complex network analysis. *Social Netw. Analys. Mining* 6(1), pp. 59:1–59:21.
- Lane, N. D., Miluzzo, E., Hong, L., Peebles, D., Choudhury, T. and Campbell, A. T., 2010. A survey of mobile phone sensing. *IEEE Communications Magazine*.
- Liu, Z., Ma, T., Du, Y., Pei, T., Yi, J. and Peng, H., n.d. Mapping hourly dynamics of urban population using trajectories reconstructed from mobile phone records. *Transactions in GIS*.
- Quercia, D., Lathia, N., Calabrese, F., Lorenzo, G. D. and Crowcroft, J., 2010. Recommending social events from mobile phone location data. In: *Proc. of IEEE 10th International Conference on Data Mining (ICDM)*, pp. 971–976.
- Rinzivillo, S., Pedreschi, D., Nanni, M., Giannotti, F., Andrienko, N. V. and Andrienko, G. L., 2008. Visually driven analysis of movement data by progressive clustering. *Information Visualization* 7(3-4), pp. 225–239.
- Wang, D., Pedreschi, D., Song, C., Giannotti, F. and Barabási, A.-L., 2011. Human mobility, social ties, and link prediction. In: *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '11*, ACM, New York, NY, USA, pp. 1100–1108.