Veronica Cipriani
203602

Human Mobility in an urban context: Public transportation and Points-Of-Interest

The study case of the city of Trento

ABSTRACT

The main aim of this paper is to understand the connection between *human mobility* and *points of interest*. In particular, the focus will be on identified hotspot in the city. In fact, as underpinned by many previous researches (Zeng et al, 2017), people movements are highly related to these so called *points of interests*.

In order to prove this theory, the case of the city of Trento is analysed. Different datasets are used and multiple analysis are implemented. First, some visualizations allow to display the location of the different points of interest in the city. Then, bus stops and trips are also plotted, in order to compare the number of stops with the number of POIs in different areas. Also, the city is divide in areas, in order to understand which category of point of interest is the most influential in shaping urban transportation.

Keywords

Human Mobility Science, Human Behaviour, Human Trajectories, Mobility, Public Transport, Urban Transport Planning, Point of Interest, Mapping, Statistics, Google Maps, Python, R, Data Visualization, Linear Regression, Correlation.

1. INTRODUCTION

Everyday, a significant portion of our time is spent trying to get from one place to another. On average, in Europe people walk around 6 km per day[1], and the distance covered by means of transport is even more significant. For this reason, the analysis of human trajectories has become more and more important throughout the years. Defined as the quantitative study of individual and collective trajectories, *human mobility* is the science that aims to understand and predict human mobility patterns and movements. It has a central role in many aspects of everyday life. In particular, many planning aspects are deeply connected with individual movements within the area of a city. Epidemic control, urban planning, traffic forecasting systems are just few examples of a field that needs to understand human mobility in an urban context. In particular, different studies have analysed the relationship between human movements and means of transport. As soon as the connection between these elements appeared clearly relevant, urban planning started to be highly influenced by human mobility habits. Human mobility science was born exactly to answer all the questions related to human movements. In fact, the study of individual and collective mobility patter is the central core of this subject.

Given the increasing number of sources of data (mobile phone with GPS, CDRs dataset, ad hoc surveys, and so

[1] *Quanto camminiamo nel corso della nostra vita?,* Focus, July 3rd, 2017.

on), the study of human mobility became more attractive. The foremost goal is to generate models that can first describe, and then predict regularities in human trajectories.

On one side, many studies proved interesting connection between human movements and POIs in cities. On the other hand, we stated that public transport planning is rooted in human mobility (Pappalardo et al., 2019; Jiang et al., 2017). Starting from these two theories, the aim of this research is to underline a hypothetic relationship between points of interest and routes of public transport. The idea is to check whether the planner keeps in mind the trajectories that each individual follows during the day when planning bus stops and routes.

2. RELATED WORK & RESEARCH QUESTION

Urban transport planning needs to carefully takes into account travel demand characterization, in order to design a service that answers as better as possible the needs of the citizens. Traditionally, two different models have been considered: on one side, gravity models; on the other, theories of intervening opportunities. The former considers as positive element the number of people in the origin and destination and the distance as negative factor in the "equation of movements". On the other side, the second theory states that distance has no great influence, as the most relevant factor is the number of intervening opportunities between the two spots (Lenormand et al., 2016; Pappalardo et al., 2019).

Therefore, in an urban context, people move according to specific rules. These movements shape the structure of the city itself. For this reason, urban mobility patterns have been analysed to understand the main characteristics of sub regions by the exploration of POIs (Xu et al., 2019), but also the structure of a city itself (Bertaud, 2001). In order to do so, many studies extracted mobility patterns from different type of trajectory data: CDRs (Jiang et al., 2015), GPS (Zhang et al., 2012), taxi trips (Liu et al. 2012, Peng et al., 2012), and even bicycle (Kaltenbrunner et al., 2010). This data is then used to develop models that are both able to descript and predict human movements in the urban context. Since people implicitly choose what they want by travelling, they *vote with their feet* (Liu et al., 2017). In particular, these movements are analysed in the dual form of individual and collective trajectories. As for the latter, many different factors influence the choice of type of transport as well as trip routes (Cohen, 2019). In this framework, one important element that has been often analysed is the presence and influence of points of influence.

The other key element of this analysis is the concept of point of interest. A point of interest is a location that someone might find useful or interesting. Basically, points of interest are able to explain – at least partly – the patterns that shapes individual's movements. In fact, different studies proved that there is some kind of relationship between people movements and POIs (Zeng et al., 2017). Since people movements are highly related to their activities, POIs plays a leading role in shaping spatial patterns. Moreover, some studies underline the possibilities of analysing the influence of different POIs according to their category. (Huang et al., 2018; Zeng et al., 2017).

As clearly presented in this review of previous studies, many different analysis and methodologies have been implemented. The varied framework enlightens the complexity of the topic. Starting from the findings of previous researches, the main aim of this paper is to further analyse the relationship between points of interest and urban transportation. In particular, this connection is tested on the case study - the city of Trento.

3. METHODOLOGY AND ANALYSIS

The analysis focuses on Trento, a city that counts 118714 inhabitants and a surface of 157,88 km^2[2]. The municipality is divided into 48 district, that are plotted and separately analysed thought-out the paper. As mentioned in the research question, the aim is to study urban transportation and compare it with the distribution

[2] WIKIPEDIA, *Trento*, in Wikipedia, 2019.

of points of interests. In order to do that, data of bus trips are analysed and POIs are plotted on the map of the city. For the analysis, different datasets are used.

3.1 Datasets description

First of all, a dataset of points of interests is used in the analysis[3]. This dataset contains more than 800 POI in the area of Trento and of the Comunità della Valle dei Laghi. Collected in 2013, this dataset contains different types of location (restaurants, hotels, museum, public transport spots, and so on) and also includes descriptions and geographical coordinates.

In order to plot the bus trips, data that contains stops, trips and routes of buses in the city are collected from Trentino Trasporti[4]. The GTFS data are downloaded and then grouped into a zip file, and then analysed in Python with *geopandas*.

Finally, in order to plot geo-referenced data on the actual map of Trento, a shape file is downloaded from Trento official website[5]. In particular, 48 different district are identified in the city.

3.2 Data Cleaning

As mentioned in the previous paragraph, different types of data are used for the analysis. The first step is to perform some data cleaning. Starting from the POIs dataset, only the points of interest of the city of Trento are selected and a subset is created. Then, in order to be able to plot them on a map from GoogleMaps or from a shape file, latitude and longitude of every point are needed. Since only a portion of observed location are provided with coordinates, latitude and longitude are computed from the given addresses thanks to *geopy*[6]. Still, some coordinates could not be found. The remaining missing elements are then dropped, as it is not possible to geo-locate and plot them. Moreover, since some interesting POIs are not considered in the dataset, they are manually added (Universities, canteens, Duomo di Trento and Buonconsiglio). The address is extracted from Google Maps and then the aforementioned code is used to get coordinates of each point.

Eventually, the shape file is then used to assign to each POI the corresponding district[7]. Since the polygons of each district are EPSG:25832, while latitude and longitude in the csv are in EPSG:4326, a *pyprog* transformation is used to change latitude and longitude to the same coordinates reference system as the polygon, in order to be able to plot a scatter on the map of Trento. Thanks to the library *shapefile*, is then possible to check which district a single POI belongs to, and a new column '*Quartiere'* is added to the csv dataset. The same library is used to add to each stop of the bus trips dataset the corresponding district.

Eventually, a new dataset is compiled. Each district corresponds to a row in the new dataset, and each rows contains numerical attributes. First, the number of bus stops are registered for each district. Then, both the number of total POIs and the specific number of points of interest per categories are listed for every district. The final dataset is then saved and it is used for linear regression in R.

3.3 Data Analysis

The implemented analysis is entirely focused on supervised learning methods and data visualization. In particular, the aim is to find – if there is *any* – a correlation between number of POIs in each specific area of the city and the number of bus stops in the same district. The analysis uses different tools, that are presented as follows.

First of all, some exploratory analyses are computed. In order to have a general overview, the number of points of interest per category and the number of bus stops per district are counted. In *Table 3.3.1,* for simplicity, only the district with more than 10 points of interest are

[3] https://dati.trentino.it/dataset/poi-trento

[4] https://www.trentinotrasporti.it/open-data

[5] https://www.comune.trento.it/Aree-tematiche/Politiche-sociali-e-abitative/Quartieri/Localizzazione-dei-quartieri-e-dei-poli-sociali2

[6] pulizia_dati.py

[7] zone_trento.py

listed. Although the threshold was arbitrary decided, the average number of POIs per district (9.61) is taken as starting point.

QUARTIERE	POIS	STOPS
Centro Storico	141	20
Santissimo	40	12
San Bernardino	14	7
San Pio X	10	7
San Giuseppe	27	27
Solteri-Cento Chiavi	36	28
Bolghera	20	17
Clarina	19	25
San Martino	27	27
Piedicastello	11	18
Campotrentino	11	9

Table 3.3.1: Number of points of interest and bus stops per district.

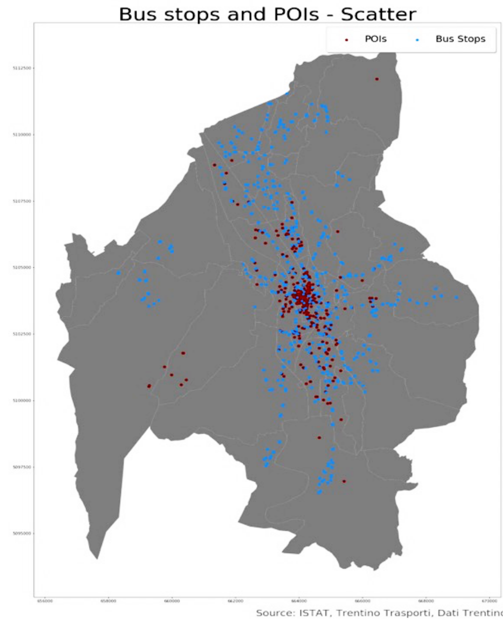

Figure 3.3.2: POIs and bus stops in the city of Trento.

Given that the increasing number of POIs seems to correspond to a high number of stops, the correlation of this subset of data is computed. However, the value of the corresponding ρ[8] in R is only 0.26. Thus, the number of point of interests itself does not seem the most relevant factor in designing public transport. In fact, some of the district with the highest density of bus stops (Martignano, Villazzano, Povo, Gardolo, Spini-Ghiaie and Mattarello) have more than 20 stops, but have less than 10 POIs. However, this preliminary result does not take into account the size of the district nor the fact that some areas are not reachable or crossable by buses.

Data visualization is another useful tool to explore the relationship between POIs and buses. *Figure 3.3.2* shows the different distribution of points of interest (*dark red*) and bus stops (*blue*).

It is clearly visible that some district present very few stops and POIs, while the highest concentration of the scatter corresponds to the city centre and surroundings.

However, the fact that the city centre is not usually crossable by bus is quite clear. In fact, the highest concentration of POIs is in the centre, while relatively few buses reach those spots. On the other side, it is also quite interesting to check whether there is any pattern in the distribution of POIs and, consequentially, if some specific categories have a particular influenced in the design of public transports. In order to confirm this hypothesis, some further analysis in R are implemented. A linear regression model is performed on the dataset containing the number of bus stops, the number of points of interest and the number of points of interest per category. The best fitting on the data is given by the model reported in *Figure 3.3.3*. There seems to be no clear pattern, therefore the conclusion should be that the type of point of interest generally do not provide any further information on the relationship *POIs-buses*. Different categories do not seem to stand for higher of lower concentration of bus stops.

[8] Pearson correlation formula: $\rho = \frac{\sum(x-\mu_x)(y-\mu_y)}{\sqrt{\sum(x-\mu_x)^2 \sum(y-\mu_y)^2}}$

```
lm(formula = n_fermate ~ n_poi + Ristorante + Tempo.libero +
    Parcheggio + Servizi + Alimentari + Museo.monumento + Istit
uzioni +
    Universita_e_altro, data = df)

Residuals:
     Min       1Q   Median       3Q      Max
-10.4132  -4.5400  -0.9888   2.6509  23.5868

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         11.4132     1.5164   7.527 9.73e-09 ***
n_poi               -1.3095     0.5928  -2.209   0.0340 *
Ristorante           3.4786     1.6863   2.063   0.0468 *
Tempo.libero         1.4585     2.4329   0.600   0.5528
Parcheggio           5.4321     3.1708   1.713   0.0958 .
Servizi              4.9127     2.7392   1.793   0.0818 .
Alimentari           2.2549     1.8454   1.222   0.2301
Museo.monumento      2.4113     2.1001   1.148   0.2589
Istituzioni          4.6695     2.3191   2.013   0.0520 .
Universita_e_altro   3.9344     1.4556   2.703   0.0107 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.393 on 34 degrees of freedom
Multiple R-squared:  0.3056,    Adjusted R-squared:  0.1218
F-statistic: 1.662 on 9 and 34 DF,  p-value: 0.1371
```

Figure 3.3.3: Linear regression model.

Given that the model only explains more or less 12% of the variability, some more complex function should probably be considered. Anyway, for the purpose of the analysis it is interesting to notice the most statistical significant variables: *n_poi, Ristorante, Universita_e_altro.* In particular, the fact that universities are central elements in this relationship may suggest that the 16000 enrolled students are an important part of public transport users.

The last part of the analysis consists again in some data visualization tools. First, the bus trips are plotted as a network. This representation can give additional information that might help in understanding whether public transport design take POIs into consideration or not. The multigraph is generated starting from a zip file of GTFS data, containing trips, stops and routes as inputs. The output is visible in *Figure 3.3.4.* The heat map of points of interest is also plotted, giving the result displayed in *Figure 3.3.5.* While the comparison of bus stops and POIs in the scatter of *Figure 3.3.2* suggests a poor relationship between POIs and bus stops, the network representation gives some very useful, additional information. In fact, even though it has already been proved that no significant correlation between increasing number of stops and POIs exists, the city centre appears here as the centre of routes as well.

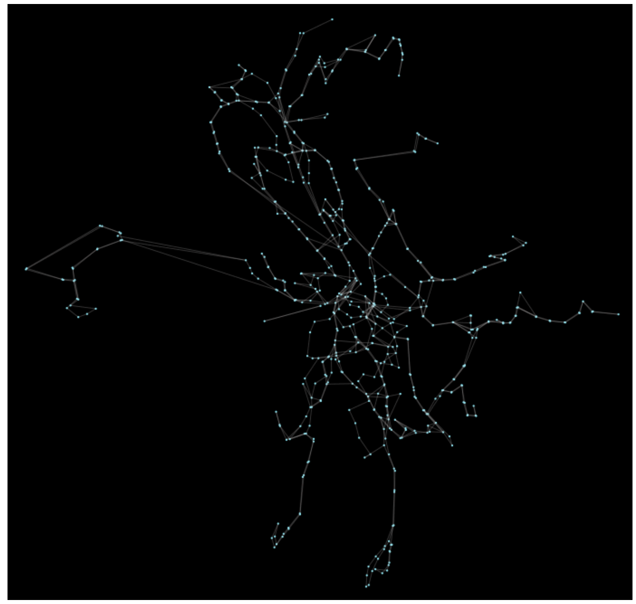

Figure 3.3.4: Bus trips represented as network.

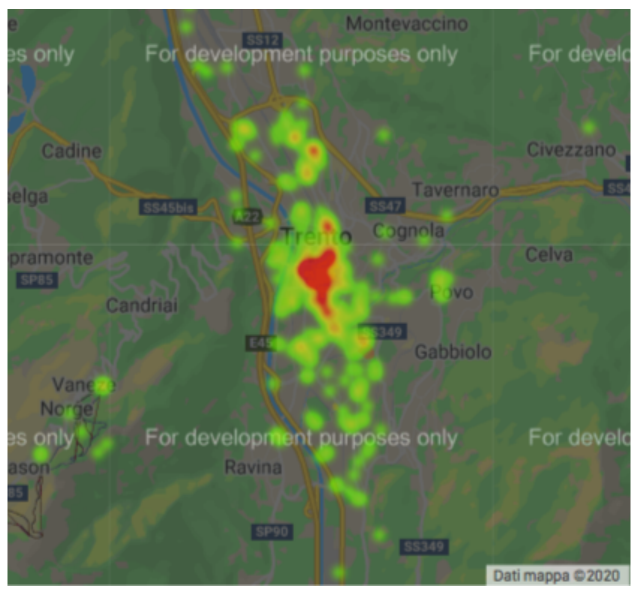

Figure 3.3.5: Heat map of the points of interest.

To conclude, even though bus stops are not always positioned near a POI, the comparison between the heat map and the network representation of bus trips shows a relatively close relationship between the most dense areas of POIs and the most travelled routes.

3.4 Analysis of the implemented methodology

The presented analysis has some advantage, but also some important limitations. For example, while the GTFS dataset contains all the bus routes, the choice of points of interest is more arbitrary – as some of them are also been added in order to make the dataset more complete[9].

Limitations due to data may also be due to other factors. Apart from arbitrariness of choices, there is lack of some information (*i.e.*, missing addresses and coordinates in POIs' dataset) and different reference systems in values encoded from different sources. Anyway, these problems are easily overcome as explained in *Paragraph 3.2.* Considering again limitation in data, GTFS files do not include the number or passenger per trips nor – as consequence – any attributes of passengers. This put some limitations when the analysis tries to define the most influent types of POIs.

Having considered the data, the analysis itself may be source of limits. The best quality of this analysis, mainly based on visualization, is the ease of interpretation. Having plotted the information in different forms, the result is reported and easily interpretable. On the other hand, the best quality can also be seen as the worst flaw of this approach. In fact, drawing conclusion only by looking at some graph may be limiting. In order to avoid this drawback, visual analysis should be completed by some deeper data analysis. In this case, linear correlation and linear regression are used in order to support the obtained results.

4. CONCLUSIONS

The aim of the analysis is to disclose the relationship between points of interest and public transport design. Starting from the already proved relationship between POIs and human movements (Zeng et al., 2017), the idea is to investigate whether – while designing urban transport – POIs are taken into consideration as proxy of human movements.

Although at first sight the relationship appears to be not so significant, further analysis show that there are some consistent similarities between the distribution of points of interest and bus routes. On the other hand, no provable relationship between the type of POIs and stops and trips has been identified. Anyway, it is interesting to notice that some single categories may play an important role.

So, a future approach may be to understand if some specific categories of POIs influences the routes of buses in relation to specific users. An example can also be found in this paper. In fact, an interesting approach could be trying to understand the most interesting POIs for the most usual bus users, in order to understand whether Trentino Trasporti is able to target its '*users'* and offer an efficient service to them. For example, as mentioned in *Paragraph 3.3* in relation to the linear regression, the category *University* itself may not be influent, but it become significant when considering that usually students take the bus to go to lesson – especially to go to Povo. This kind of application may underline that only some categories of POIs are to be taken into account in the regression. Unfortunately, as aforementioned, the available data does not include any attributes of users nor the number of passenger per buses, so it is quite difficult to really understand the hypothetical relationship between type of point of interest and closeness of bus stops.

To conclude, despite the limitations of data, the analysis points out some interesting aspect that seem to be taken into account while designing bus routes. The problem is fully disclosed and widely presented using different tools.
Moreover, some suggestions and new approach to the problem also arise from the analysis itself.

[9] pulizia_dati.py

5. DATASET

BUS TRIPS, Trento.
Retrieved:
https://www.trentinotrasporti.it/open-data

POINTS OF INTEREST, Trento, 2013.
Retrieved:
https://dati.trentino.it/dataset/poi-trento

SHAPE FILE, Poli Sociali, Trento.
Retrieved:
https://www.comune.trento.it/Aree-tematiche/Politiche-sociali-e-abitative/Quartieri/Localizzazione-dei-quartieri-e-dei-poli-sociali2

6. REFERENCES

BERTAUD A., *Metropolis: A Measure of the Spatial Organization of 7 Large Cities*, 2001.
Retrieved:
http://alainbertaud.com/wp-content/uploads/2013/06/AB_Metropolis_Spatial_Organization.pdf

COHEN K., *Human Behavior and New Mobility Trends in the United States, Europe, and China*, Working Paper – Fondazione Enrico Fermi, 2019.
Retrieved:
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3475381

FIK T., MULLIGAN G., *Spatial Flows and Competing Central Places: Towards a General Theory of Hierarchical Interaction*, Environment and Planning A,1990
Retrieved:
10.1068/a220527

GIGLIO S., BERTACCHINI F., BILOTTA E., PANTANO P., *Machine learning and points of interest: typical tourist Italian cities*, Current Issues in Tourism 2019.
Retrieved:
https://doi.org/10.1016/j.tourman.2018.12.007

HUANG L., YANG Y., ZHAO X., GAO H., YU H., *Mining the Relationship between Spatial Mobility Patterns and POIs*, Wireless Communications and Mobile Computing 2018.
Retrieved:
https://www.researchgate.net/publication/325087202_Mining_the_Relationship_between_Spatial_Mobility_Patterns_and_POIs

JIANG S., FERREIRAJ., GONZALEZ M. J., *Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore*, in Proceedings of the 4th International Workshop on Urban Computing (UrbComp'15), ACM, 2015.
Retrieved:
https://www.researchgate.net/publication/310789943_Activity-Based_Human_Mobility_Patterns_Inferred_from_Mobile_Phone_Data_A_Case_Study_of_Singapore

JIANG S., GUAN W., ZHANG W., CHEN X., YANG L., *Human mobility in space from three modes of public transportation,* Physica A: Statistical Mechanics and its Applications, Elsevier, vol. 483(C), pages 227-238, 2017.
Retrieved:
10.1016/j.physa.2017.04.182

KALTENBRUNNER A., MEZA R., GRIVOLLA J., CODINA J., BANCHS R., *Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system*, Pervasive Mob. Comput., 6, pp. 455–466, 2010.
Retrieved:
http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=64D39AE1245368924714199D874506FC?doi=10.1.1.709.1846&rep=rep1&type=pdf

LENORMAND M., BASSOLAS A., RAMASCO J.J., *Systematic comparison of trip distri- bution laws and models*. Journal of Transport Geography, Elsevier, 51, pp.158-169, 2016.

Retrieved: https://hal.archives-ouvertes.fr/hal-01175812/document

LIU Y., LIU C., LU X., TENG M., ZHU H., XIONG H., *Point-of-Interest Demand Modeling with Human Mobility Patterns*, KDD 2017 Research Paper, 2017.
Retrieved: https://www.kdd.org/kdd2017/papers/view/point-of-interest-demand-modeling-with-human-mobility-patterns

LIU Y., WANG F., XIAO Y., GAO S., *Urban land uses and traffic "source-sink areas": Evidence from GPS-enabled taxi data in Shanghai*, Landscape and Urban Plan, vol. 106, pp. 73–87, 2012.
Retrieved: https://doi.org/10.1016/j.landurbplan.2012.02.012

LIU S., PU J., LUO Q., QU H., NI L. M., KRISHNAN R., *VAIT: A visual analytics system for metropolitan transportation*. IEEE Transactions on Intelligent Transportation Systems, 14(4):1586–1596, 2013.
Retrieved: https://www.researchgate.net/publication/221589678_VAIT_A_Visual_Analytics_System_for_Metropolitan_Transportation

MORIOKA M., KURAMOCHI K., MISHINA Y., AKIYAMA T., TANIGUCHI N., *City management platform using big data from people and traffic flows*, Hitachi Review, 64(1): pp. 52 – 57, 2015.
Retrieved: https://pdfs.semanticscholar.org/8008/22a5c758e53beec462a6afe243e3d9c4b659.pdf

PAPPALARDO L., BARLACCHI G., PELLUNGRINI R., SIMINI F., *Human Mobility from theory to practice: Data, Models and Applications*, Conference Paper – The web conf 2019, 2019.
Retrieved: 10.1145/3308560.3320099

PENG C., JIN X., WONG K.C., SHI M., LIÒ P., *Collective human mobility pattern from taxi trips in urban area*, PLoS ONE 2012, 7, e34487, 2012.
Retrieved: https://doi.org/10.1371/journal.pone.0034487

Quanto camminiamo nel corso della nostra vita?, Focus, July 3rd, 2017.
Retrieved: https://www.focus.it/cultura/curiosita/quanto-camminiamo-nel-corso-della-nostra-vita

VELOSO M., PHITHAKKITNUKOON S., BENTO C., *Urban mobility study using taxi traces*, in Proceedings of the 2011 International Workshop on Trajectory Data Mining and Analysis, pp. 23–30. ACM, 2011.
Retrieved: http://wiki.amilab.dei.uc.pt/images/8/83/Veloso_2011_TDMA.pdf

WANG J., KONG X., FENG X., SUN L., *Urban Human Mobility: Data-Driven Modeling and Prediction*, ACM SIGKDD Exploration Newsletter, 2019.
Retrieved: https://www.researchgate.net/publication/333136406_Urban_Human_Mobility_Data-Driven_Modeling_and_Prediction

WIKIPEDIA, *Trento*, in Wikipedia, 2019.
Retrieved: https://it.wikipedia.org/wiki/Trento

XU Z., CUI G., ZONG M., WANG X., *Anomalous Urban Mobility Pattern Detection Based on GPS Trajectories and POI Data*, International Journal of Geo-information, 2019.
Retrieved: https://doi.org/10.3390/ijgi8070308

ZHANG W., LI S. PAN G., *Mining the semantics of origin-destination flows using taxi traces,* ACM Conference on Ubiquitous Computing-UbiComp '12, Pittsburgh, PA, USA, 5–8 September 2012.
Retrieved:

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.459.171&rep=rep1&type=pdf

ZHANG F., PFOSER D., *Using OpenStreetMap point-of-interest data to model urban change—A feasibility study*, PLoS ONE 14, 2019. Retrieved: https://doi.org/10.1371/journal.pone.0212606

ZENG W., FU C., ARIZONA S. M., SCHUBIGER S., BURKHARD R. A., MA K., *Visualizing the Relationship Between Human Mobility and Points of Interest*, IEEE Transactions on Intelligent Transportation Systems, 2017. Retrieved: https://doi.org/10.1109/tits.2016.2639320