Centro de Informática U·F·P·E

Pós-Graduação em Ciência da Computação

Exploring Human Activity Behavior and Mobility Data in Carpooling

Por

Vinícius Cezar Monteiro de Lira

Dissertação de Mestrado

Universidade Federal de Pernambuco
posgraduacao@cin.ufpe.br
www.cin.ufpe.br/~posgraduacao

RECIFE
2014

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

VINÍCIUS CEZAR MONTEIRO DE LIRA

Exploring Human Activity Behavior and Mobility Data in Carpooling

ESTE TRABALHO FOI APRESENTADO À PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO DO CENTRO DE INFORMÁTICA DA UNIVERSIDADE FEDERAL DE PERNAMBUCO COMO REQUISITO PARCIAL PARA OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIA DA COMPUTAÇÃO.

ORIENTADOR(A): VALÉRIA CESARIO TIMES
CO-ORIENTADOR(A): PATRÍCIA TEDESCO

RECIFE, AGOSTO/2014

Catalogação na fonte
Bibliotecária Jane Souto Maior, CRB4-571

Lira, Vinícius Cezar Monteiro de Lira
Exploring human activity behavior and mobility data in carpooling / Vinícius Cezar Monteiro de Lira. - Recife: O Autor, 2014.
104 f., fig., tab.

Orientador: Valéria Cesário Times.
Dissertação (mestrado) - Universidade Federal de Pernambuco. CIn, Ciência da Computação, 2014.

Inclui referências.

1. Banco de dados. 2. Mineração de dados. 3. Mobilidade. 4. Carpool. I. Times, Valéria Cesário (orientadora). I. Título.

025.04 CDD (23. ed.) MEI2014 – 130

Serviço Público Federal
Universidade Federal de Pernambuco
Centro de Informática
Pós-Graduação em Ciência da Computação

Ata de Defesa de Dissertação de Mestrado do Centro de Informática da Universidade Federal de Pernambuco, 28 de agosto de 2014.

Ao vigésimo oitavo dia do mês de agosto do ano de dois mil e catorze às dez horas, no Centro de Informática da Universidade Federal de Pernambuco, teve início a **milésima quadringentésima vigésima sétima** defesa de dissertação do Mestrado em Ciência da Computação, intitulada **"Exploring Human Activity Behavior and Mobility Data in Carpooling**" sob orientação da **profa. Valéria Cesário Times,** do candidato **Vinícius Cezar Monteiro de Lira** o qual já havia preenchido anteriormente as demais condições exigidas para a obtenção do grau de mestre. A Banca Examinadora, composta pelos professores Ana Carolina Brandão Salgado, pertencente ao Centro de Informática desta Universidade Renata Galante, pertencente ao Departamento de Informática Aplicada da Universidade Federal do Rio Grande do Sul e Patrícia Cabral de Azevedo Restelli Tedesco, pertencente ao Centro de Informática desta Universidade sendo a primeira presidente da banca examinadora, decidiu **Aprovar** o trabalho. E para constar lavrei a presente ata que vai por mim assinada e pela Banca Examinadora. Recife, 28 de agosto de 2014.

____
Maria Lília Pinheiro de Freitas
(secretária)

____
Profa. Ana Carolina Brandão Salgado
Centro de Informática/UFPE

____
Profa. Renata Galante
Departamento de Informática Aplicada/UFRGS

____
Profa. Patrícia Cabral de Azevedo Restelli Tedesco
Centro de Informática /UFPE

Dedico essa conquista a meus pais!
Sou ciente e eternamente grato por todo esforço e dedicação que eles tiveram por mim.

Thanks

First of all, I thank God for more this achievement in my life and for all the experiences that I had during my master's degree. I am sure that it was a period with a huge gain for myself and for my professional career.

I am especially grateful to my parents for their unconditional love and support in my life. I owe them all I am. I thank as well to all family, who supported me along the completion of the Master program.

I strongly thank my advisors, Valeria Cesário Times and Patricia Tedesco, for their support, patience and guidance. They gave me encouragement to apply for the Master program and later the confidence to take part in the SEEK Project[1], where I had the opportunity to try an enriching international experience living in Italy.

I thank my Italian supervisors in the SEEK Project, Chiara Renso and Salvo Rinzivillo, our meetings and discussions made possible the publication of two papers during my master's degree.

Without doubt, I had the luck to have a really great team of advisors.

Finally, I thank all my friends (Brazilians, Italians, Chinese, Indians, Austrian) that somehow contributed to my growth during this period.

[1] www.seek-project.eu

“Stay hungry. Stay foolish”

Steve Jobs

Abstract

The analysis of human movements has been the subject of several studies since the 70s. In recent years, the exponential growth of location aware devices must allow the study of the behavior of the individuals' mobility from their trajectories collected. However, a significant part of the available literature is focused on the development of techniques for analyzing trajectories of people from a purely geometric point of view, while a smaller part, but increasingly group is looking at the semantic aspects of mobility. This work presents a contribution to the latest trend, and is concerned with the definition of semantic regularity profiles and the applicability of these concepts to the practice of carpooling.

We propose a semantic regularity profile based on the entropy of the spatial and temporal frequency of visits to certain categories of places. We analyze the user's behavior with respect to regularity and irregularity, identifying users who are more or less loyal to certain locations, in contrast to the irregularity of visiting different places. In a different point of view, an analysis over the place perspective was also performed. A web tool was developed to show on map, for each place of a given category, the computed information about the loyal behavior of their visitors.

From the study about regularity, we have evidences that some human activities are not strictly associated to a unique POI (Point of Interest) and neither to a specific schedule of the day. Bringing to the carpooling context, in some situations it is worth for a person to change his destination or the time to perform an activity if there is a possibility of ride for him due all the benefits involved with the carpooling practice. This dissertation also presents a novel matching method for carpooling that is oriented to the passenger's intended activity, aiming to boost the possibilities of rides. Three algorithms for carpool matching are proposed, which manipulates differently the spatial and temporal dimensions. Using a real data set of trajectories, we conducted experiments and our results showed that the proposed matching algorithms improved the traditional carpooling approach in +46.84% when the spatial dimension was considered, in +50.89% when the temporal dimension was prioritized and in +82.30% when both dimensions were tackled.

Keywords: Human Mobility Behavior. Regularity. Loyalty. Carpooling.

Main Abbreviations

Abbreviation	Meaning
CR	*Carpooling Request*
MCR	*Mapped Carpooling Request*
GSM	*Groupe Spécial Mobile*
GPS	*Global Position System*
OLAP	*On-Line Analytical Processing*
OT	*Occasional Trip*
POI	*Point of Interest*
RT	*Routine Trip*
MRT	*Mapped Routine Trip*

List of Figures

Summary

Chapter 1: Introduction

This chapter discusses the context of the domain addressed, the motivation of the study and the research objectives documented here. Also, a brief overview of how this document is structured is given.

1.1. Contextualization

The constant acquisition of a moving object's positions has become technically possible thanks to the positioning sensors such as GPS (Global Positioning System) (SPACCAPIETRA et al., 2008). The increase in the use of devices that enables the tracking of moving objects generates a large amount of mobility data allocated in time and space and was motivated mainly due to the decreasing cost of acquisition of such devices. Moving objects are considered as every physical object in the real world equipped with a device that allows the tracking of its geographical position at a given time (GUTTING et al., 2000). Boats, cars, people, animals and airplanes coupled with this type of device, are typical examples of mobile objects (MORENO; ARANGO, 2011). A new field of study and application development has emerged under the name of Spatial-Temporal path or trajectory of mobile objects, or simply trajectory, which can be defined as the evolution of an object's position in space over a given interval time (SPACCAPIETRA et al. , 2008).

Furthermore, unprecedented amount of user-generated data on human movement has been collected through the introduction of location-based services in social media applications of smartphones provided by some virtual social networks (e.g. Facebook, Foursquare, Twitter etc.). It has enabled people to share their activity related choices performing "*check-ins*" during their visits to venues. This data contains detailed geo-location information, which reflects extensive knowledge about human movement behavior. In addition, the venue category information for each check-in is recorded from which user activities can be inferred (HASAN et al., 2013).

On recent years the amount of "geo-social information" increased substantially contributing as an important source of knowledge about the human mobility behavior. To have an idea about this increasing, in August 2006, Flickr introduced the geo-tagging feature; by 2007, more than 20 million geo-tagged photos were uploaded to Flickr. In August 2011, Flickr announced its 6 billionth photo, with an increase of 20% year-on-year over the last 5 years.[2] In 2010, the geo-tagging feature was also added to Twitter by generating a large amount of geo-information daily. On the first year the average number of Tweets sent per day was 50 million[3] while in March 2012 it has increased to 340

[2] Source: http://blog.flickr.net/en/2011/08/04/6000000000/
[3] Source: http://blog.twitter.com/2011/03/numbers.html

million.[4] Nowadays people are considered as sensors, producing signals on events they are directly involved in or they are present (RENSO et al., 2013).

Data analysis of human mobility data has been shown to be a highly multidisciplinary field. The capture of large amounts of these data allowed the analysis of such data for various areas of research, ranging from traffic managers to advertising and social studies of human behavior (FURLETTI et al., 2013). Many researchers have shown interest, generating works in various subfields of computer science, including: (GRUMBACH et al, 2001) database (ERWIG; SCHNEIDER, 2003) (GRUMBACH et al , 2003.) mining and knowledge discovery on spatial data (FURLETTI et al . , 2013) (Bettini et al . , 2000) (BOGORNY; WACHOWICZ , 2008) (ERWIG; SCHNEIDER, 2002) (FRANK , 2003) (HORNSBY; EGENHOFER , 2002) , geographic information systems (GIS) (HORNSBY , 2001) (MARK , 2003) (MILLER , 1991) (MILLER; WU , 2000) (MILLER , 2003) , among others.

Several works are related to the studies of human behavior using mobility data, most of them are more concerned with detection of partner from the geometric point of view. However, few works, but increasingly, consider semantic information in their analysis. The use of semantic information in analysis of mobility data allows a better understanding of the proposal of its generation. For example, if we consider a mobility register performed by a person having spatial coordinates with latitude equal to 43° 42' 45.0714 and longitude, 10° 24' 7.941, stamped at time 20:00, what does this register means? Seeing only these properties could be difficulty to have some interesting interpretation from this sample of data. However, considering semantic, we have the information that this geographic position correspond to the Pizzeria Le Scuderie in the city of Pisa. It gives a high interpretation level of the data, opening more possibilities of analysis.

In addition, studies on transport modeling bring different approaches to support management decisions and make predictions. One of the approaches for transportation modeling is the activity-based modeling. In this approach, the demand trajectories is derived from the activities that individuals need/wish to perform. This approach reflects the scheduling of activities in time and space (CHO et al., 2013), (SHIFTAN et al., 2003). Based on the concepts of the activity-based approach, where a trajectory is seen from the

[4] Source: http://blog.twitter.com/2012/03/twitter-turns-six.html

point of view of activity, we analyze the human activity using mobility data with semantic information. We explore such data to calculate regularity measures about users' visits to places to perform some activity. For a given activity, we compute for a user a measure to estimate how much regular or how much flexible he is in his choices of place to visit.

In this dissertation the carpooling topic is also addressed. The term Carpooling means the practice where a group of two or more persons "sharing" a car ride, instead of all using different cars. This practice benefits the participants, once that they can share the cost of the ride, and also the environment, avoiding pollutant emissions. Recent researches about mobility analysis applied to carpooling have been extended in order to identify the behaviors that people constantly follow, such as groups of trajectories with common routes, popular itineraries or casual trips (GIANNOTTI et. al., 2011), (TRASARTI et al, 2011), (MA; WOLFSON., 2013). From the study about users' regularity and no regularity, we have evidence that the casual trips, those that are not daily performed by a person, are more related to activities that do not require to go to a specific place (e.g. go to supermarkets, malls, restaurants, bars, entertainment center, among others). From these analysis and still keeping the focus on the analysis of the human activity, we also propose a carpool matching method that is oriented to the intended activity of the user.

1.2. Motivation

The appearance and wide distribution of position-enabled personal devices also boosted new disciplines in the direction of studying the mobility behavior of individuals from mobility data. However, a significant part of the available literature is devoted to developing techniques to analyze people's trajectories from a purely geometric point of view, while a smaller, but steadily increasing part is looking at the semantic aspects of mobility (FURLETTI et al, 2013), (HASAN et al., 2013), (MANSO et al., 2010), (PARENT et al., 2013), (QIN et al, 2012), (WU et al., 2014).

The human life can be seen as both random and regular, and its inherent complexity is manifested for instance in the spatiotemporal tracks of individuals (QIN et al., 2012). Particularly, our work presents a study of one aspect of the human mobile lifestyle related to the tendency of mobile individuals to be *regular* or *irregular* when choosing the places and the time to perform some activities. This characteristic is called the *semantic regularity* of an individual. This approach concentrates on the place and time the persons perform some activity. For specific activities, like going from home to work/school/university, is expected a very regular, and hence predictable, behavior. For other activities like going to a restaurant the behavior may depend on several factors like the individual preference, the actual distance from the current location, the type of cuisine offered, the restaurant availability in the area and so on.

The temporal interval when the activity is performed is another dimension to consider when the corresponding behavior is studied. For example, a user may be willing to go for shopping always at the same time (say always at 18:30 after work) whereas another user may go at a different time of the day. The idea is to be able to distinguish the regular persons that tend to go regularly to the same place to perform an activity, from those who try new possibilities from a spatial and/or temporal point of view. There are evidences that the regularity is not an intrinsic property of the individual but, on the contrary, it depends on the particular activity to be performed. This means that there are users who, for example, have their preferred supermarket, but visit all the available restaurants.

Regularity profiles can be useful from several points of view. They give a quantitative measure of the people's regularity habits. This can be useful in the recommendation and advertisement field: (1) where new places to visit could be suggested to the regular users trying to encourage them to discover new places; (2) where irregular users could be incentivated through loyalty campaigns.

In different points of view, regularity profiles computed from mobility data can be useful to characterize the geographical POIs (Points of Interest) of a region as well. When we talk about information of regularity of the user's visits to a POI, implicitly we are talking about information of loyalty between a POI and the user. Our work presents a web tool, called MAPMOLTY (MAPping MObility loyaLTY), which was developed to compute and show on map information about the users' loyalty behavior based on their visits to POIs to perform a given activity. To the best of our knowledge, this is the first web tool providing a loyalty map of Points of Interests (POIs).

The analysis enabled by the tool may be useful in different scenarios. For example, an urban manager may quickly discover attractions that are visited by occasional visitors like tourists or loyal visitors like dwellers. This may be useful in supporting decision making in traffic management and/or urban planning. On the other hand, the loyalty analysis results may be used for marketing purpose by the owners of the POIs to better plan advertising to targeted individuals and analysis of competitors. The loyalty indicators are also useful for developing services for the citizen like a recommendation system suggesting new destinations according to the observed visitors' behavior.

Another application of these semantic regularity concepts is the transportation management, mainly when applied to carpooling systems. In the computer science field, many works address the carpooling analysis over trajectories data by mainly focusing on spatial and temporal issues. In general, they use the raw trajectories or manually annotated trajectories by people and then, they find matching using some geometric method, a sequence algorithm, or a simple destination with a schedule matching (COLLOTTA et al., 2012), (TRASARTI et al, 2011). This dissertation also addresses the carpooling topic by proposing a novel carpool matching method. Our method is activity-oriented. Benefiting of the semantic information about the places, our carpooling approach considers alternative destinations and schedules for allowing the passenger to perform the intended activity. Differently from recent works about carpooling, our method has a different focus since we pay attention to the intended person activity, rather than the path or final destination of the ride.

The analysis of raw data, which is a sequence of geographic points ordered by time and usually traced by a GPS device, does not enable an easy interpretation about the movement of a moving object due the lack of semantic. The use of semantic in trajectories aggregates a more high level of meaning to the data, thus ceasing to be purely geometrical. Analyzing a trajectory with semantic becomes easier to infer or to predict the intention of the moving object when the trajectory was performed. For example, given one semantically enriched trajectory could mean that a person drove from his house toward a gym at a specific time of day to perform the activity: "training".

Our method approaches the carpooling problem on the point of view of the person's intended activity. For some activities a person can choose between a large number of places to perform the activity desired. Thus, these categories of destination are more flexible with respect to changes of the final destination of the ride. Our proposal is to increase the possibilities of ride in carpooling exploring the intended activity of the passenger (i.e. the hitchhiker). For example, the passenger may be able to go to different markets to perform the activity "buy foods". About the time, in some cases the person can change his behavior about temporal aspects. For example, a person may not care about the period of the day to go to a gym, when there is not a temporal appointment for that place.

1.3. Objectives

The main objective of this work is to propose a study about the human activity and using human mobility data, where we address the following issues: (*i*) for a given user, to measure the regularity level about the place's choices to perform certain activity; (*ii*) for a given a place, to calculate overall perspectives of its customer loyalty. Based on this study, we also propose an activity-oriented carpooling focus on the intended activity of the passenger, thus giving a different point of view for the carpooling practice. To contribute to the main goal of this work, the specific objectives are listed as follows.

(1) Propose a measure of regularity by using entropy in order to quantify and analyze the users in terms of the regularity of their visits to certain locations to perform a given activity, for some categories of place;

(2) Develop a web tool called MAPMOLTY to show on a map, for each POI of a given category, loyalty indicators about the frequency behavior of their visitors;

(3) Based on this analysis of the user behavior about regularity of visits to places, we propose an activity-oriented matching method for carpooling, where a person can get a ride to reach a place focusing on the intended activity. This method relaxes the hard constraint that a person has to reach a specific place, but it is based on the idea that some activities can be performed in different places (e.g., a different supermarket to go for shopping) and different schedules (i.e. postponing or anticipating the time). We also compared our proposed method with a baseline approach following the traditional carpooling.

1.4. Structure of the dissertation

This dissertation is structured as follows:

Chapter 2 – Theoretical Foundation and Related Work: introduces the main concepts related to the field approached, by including a discussion of related work, and listing definitions and relevant studies to help in the understanding of the work detailed in this dissertation;

Chapter 3 – Investigating semantic regularity of human mobility lifestyle: shows an approach to study the mobility behavior of individuals presenting a definition of user regularity profiles. Our method is based on the entropy of both spatial and temporal frequency of visits to places. This allowed us to define the concept of regular or irregular user behavior identifying users who are more or less loyal to same places in contrast to the irregularity in visiting different places.

Chapter 4 – A Matching Method for Carpooling: this chapter introduces the proposed matching method for carpooling that is oriented on the passenger's intended activity. We presents three different algorithms, where each algorithm relax differently the spatial and temporal constraints of the search for carpool matching. In the experiments, given a dataset of trajectories, we analyze the percent of trajectories that could be avoided applying our method.

Chapter 5 – Conclusions: this chapter presents the main contributions of the work discussed in this dissertation and gives suggestions for future research on the covered area.

Chapter 2: Theoretical Foundations and Related Works

This chapter shows the theoretical foundation of this dissertation, which includes basic concepts and an analysis of the related work on human mobility data and carpooling.

2.1. Introduction

To present the theoretical foundation of this dissertation, this chapter is organized as follows. In Section 2.2, the Activity-Based Approach is shortly discussed. Section 2.3 details the different types of data mobility. The concepts related to carpooling are shown in Section 2.4. Section 2.5 and Section 2.6 discuss, respectively, the state of the art about human activity analysis over mobility data and about regularity in human mobility behavior. Section 2.7 presents some methodologies for the extraction of routine trajectories and Section 2.8 summarizes the most important studies on the carpooling field. Finally, some final considerations about this chapter are given in Section 2.9.

2.2. Activity-Based Approach

The need for realistic representations of behavior in travel demand modeling is well acknowledged in the literature. As result in the field of travel demand analysis, there is an increasing realization that the traditional trip-based modeling approach, that is statistically-oriented, needs to be replaced by a more behaviorally-oriented activity-based modeling approach (BHAT; KOPPELMAN, 2003).

Thus the Activity-based modeling treats travel as being derived from the demand for personal activities (CHO et al., 2013), (MAZZULLA, 2009), (VERKEHRS, 2000). Travel decisions become part of a broader activity scheduling process based on modeling the demand for activities rather than merely trips (MAZZULLA, 2009), (SHIFTAN et al., 2003). It provides better understanding of travel behavior placing primary emphasis on activity participation. As shown in Figure 1, this approach brings the focus on sequences or patterns of activity behavior (using the whole day or longer periods of time as the unit of analysis). It enables a better analysis of responses to policies and their effect on traffic (SHIFTAN et al., 2003).

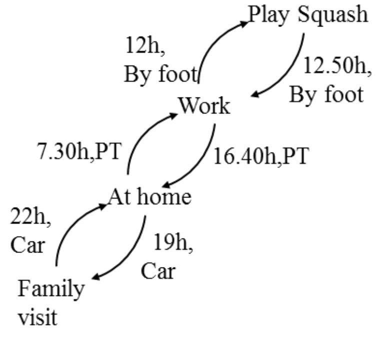


Figure 1: Activity-based approaches reflect the scheduling of activities in time and space.

2.3. Mobility Data

On this section we present the main current types of mobility data.

2.3.1. Trajectories of Moving Objects

Several works in the literature address the analysis of trajectory data. Even the definition of what a trajectory is can have several variants. The most intuitive is that a trajectory represents the spatiotemporal evolution of a moving object. Spaccapietra (SPACCAPIETRA et al., 2008) define trajectory as "the user defined record of the evolution of the position (perceived as a point) of an object that is moving in space during a given time interval in order to achieve a given goal".

Raw trajectories are well fitted for applications that aim only at locating some moving object or computing statistics on the spatiotemporal characteristics of trajectories (PARENT et al., 2013). However, most application analysis require complementing raw data with additional information from the application context. For example, interpreting trajectories of persons within a city requires some knowledge about the features of the city (e.g. map, points of interest). Thanks to city information, spatiotemporal coordinates can be replaced with street and crossing names, or with names of places of interest, such as shops, restaurants, and museums (PARENT et al., 2013).

Adding knowledge to raw trajectories is known as a semantic enrichment process (PARENT et al., 2013). A semantic trajectory is the representation of the trajectory and a set of interpretations on the moves and on the stops made by the moving object during its journey. The same trajectory can have different interpretations depending on the context of the application and the modeled domain. Figure 2 shows an example of a trajectory with implicit semantic information. In this figure, the trajectory *A* shows how the data is usually collected in its raw format. With this representation, the trajectory has no semantic aggregate and can only be seen as a set of points that form a route without associated information about the path taken. Regarding the two other trajectories (*B* and *C*) of Figure 2, the data from the same trajectory are crossed with spatial information from different domains, causing it to be semantically enriched.

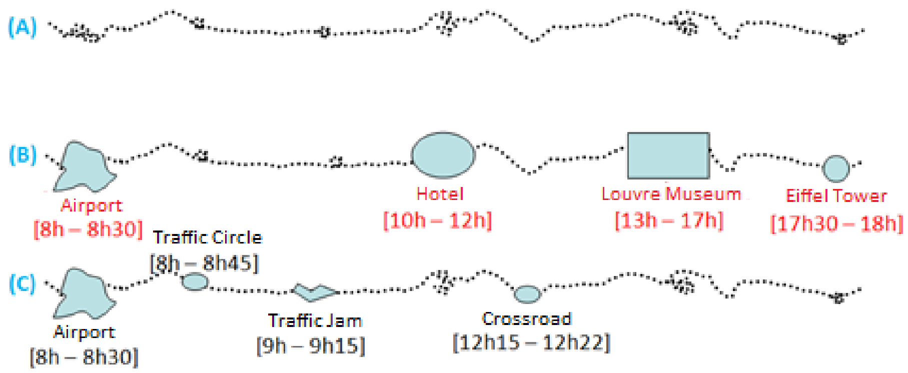


Figure 2: Semantics of a trajectory (adapted from ALVAREZ et al. (2007))

In this dissertation, we work with the representation B shown in Figure 2, where a stop is usually associated to a place where the user goes to perform some activity. Then, we are concerned with the visited place of the moving object, not about events in traffic.

2.3.2. GSM Data

GSM (*Groupe Spécial Mobile*) is the most popular standard for mobile phones in the world, nowadays used by more than 1.5 billion of people across more than 210 countries and territories. GSM networks consist of a number of base stations, where each of them is responsible for a particular spatial area (known as "cell"). Hence, for each GSM-enabled device we can collect information about the base stations that it was served at different timestamps (BELLEMANS et al. 2012).

Two main drawbacks are related to these data: the position is not precise (the granularity is the area of the cell) and the socio-demographic information of the user is usually not available due to privacy reasons.

2.3.3. Location-Based Social Network Data

With the rapid development of mobile devices, wireless networks and Web 2.0 technology, a number of location-based social networking services, e.g., Loopt and Foursquare, have emerged in recent years. These location-based social networking services allow users to connect with friends, explore places (e.g., restaurants, stores, cinema theaters, etc), share their locations, and upload photos, video, and blogs (YE; YIN, 2010). People share their activity related choices performing "*check-ins*" during their visits to places. This data contains detailed geo-location information, which reflects extensive knowledge about human movement behavior. In addition, the place category information for each check-in is recorded from which user activities can be inferred

(HASAN et al., 2013). On recent years the amount of "geo-social information" increased substantially contributing as an important source of knowledge about the human mobility behavior.

2.3.4. Bluetooth and RFID Data

The movement of a Bluetooth device within an area can be tracked by considering the distance of the device from Bluetooth receivers and using trilateration approaches. The disadvantage of this technique is that it can only cover a limited area, being mainly used for indoor tracking of objects. Therefore, they cannot really be used for outdoor object tracking (RENSO et al. 2013).

Another technology generating mobility data is the use of RFID *(Radio-Frequency Identification)* tags, commonly used in supermarkets and these tags allow tagged objects be tracked by sensors via radio signals. As in Bluetooth technology, RFID readers can locate tags within a limited area so it is hard to apply this technology to the outdoor tracking of moving objects (RENSO et al. 2013).

2.4. Carpooling

Known problems in big cities about traffic are: traffic jams and CO2 emissions. In most of the cases, to avoid these problems, changes to the traffic environment are required. However, perform changes to an existing infrastructure is expensive, have significant long-term effects, no guarantee for success and sometimes are not trivial (existing spatial zones, restricted by local and federal regulations, legislation, etc.). A simple and inexpensive solution to try to improve the traffic is the carpooling practice.

Carpooling is one of the many travel alternatives promoted by transport policies to reduce the amount of vehicles on the road. It was promoted during World War II to deal with oil and rubber shortages and during the oil crisis of the 1970s (GILBERT and PERL, 2008). Nowadays, carpooling is promoted by mobility management policies to put more emphasis on the issue of sustainable transport. The main targets here are a reduction of transport-related pollution (PM_{10},NO_x, and CO_2), noise nuisance reduction and decrease of congestion levels and minimize the necessity for parking spaces (VANOUTRIVE et al, 2011), (DIXIT et al., 2012).

According with (VANOUTRIVE et al, 2011), carpooling is the sharing of car journeys so that there is the driver and at least one passenger traveling in a car (Dixit, et al, 2012). In this dissertation, a passenger is the hitchhiker, in other words, the person who is not the driver. Furthermore, the terms *carpooling, ridesharing and car-sharing* may or may not be used interchangeably. As common sense, “ride-sharing exists when two or more trips are executed simultaneously, in a single vehicle.” (MORENCY, 2007, p. 240). We use the term carpooling since it is widely known in the literature (A.LIYANA et al. 2012) (ARNOULD et al., 2011) (CHEN et al., 2011) (CICI et al. 2013) (PERTERER et al., 2013). Carpooling is also seen as a more environmentally friendly and sustainable way to travel as sharing journeys reduces carbon emissions, traffic congestion on the roads, and the need for parking spaces. The term *Car-Sharing* is regularly understood as a service in which a car provided by a company can be booked by persons who only occasionally need a ‘rental’ car (VANOUTRIVE et al, 2011) (KATZEV, 2003) (FEDOR et al., 2012).

A particular set of carpooling determinants are the (dis)incentives present in mobility management schemes which aim to increase the popularity of carpooling. The rationale behind the promotion of carpooling is that every carpooling employee implies one car less on the road. In addition to the benefits to the environment, the quoted benefits of carpooling are self-evident: driving costs may be shared, commuters are not dependent on schedules and/or public transport networks (VANOUTRIVE et al, 2011). Typically carpooling can also be motivated by an incentive such as a faster HOV (high-occupancy vehicle) lane or a toll reduction.

2.4.1. Casual Carpooling (Slugging)

Casual carpooling (also called “slugging”) is a system of carpooling without trip-by-trip pre-arrangement (BELLEMANS et al. 2012). In others words, casual carpooling refers to the sharing of a ride with a driver and one or more passengers, where the ridesharing between the individuals is not established in advance but coordinated on the spot (MA; WOLFSON., 2013). It is the carpooling practice in ad hoc mode, informal carpools for purposes of commuting and eventual rides.

2.5. Human Activity Analysis in Mobility Data

Jovicic (JOVICIC, 2001) defines activity as a physical engagement of an individual in something that satisfies his or/and family needs. According to him, activities are motivated by economical, physiological and sociological needs of an individual. Activities can be grouped into various categories, e.g. work, shop, recreation, mandatory, optional, etc.

Most existing human mobility studies do not focus on the human activities. One reason for this is the lack of explicit large scale activity information data (Wu et al., 2014). These studies focus on the exterior characteristics of movements like geographical heterogeneity and distance decay, population density, geographical and social distances, urban morphology and the spatial distribution of places. Basically these works neglects activities, the driving force that underlies human movements.

Nowadays, social media services have become increasingly used. They have also become an indispensable part of many people's lives to record life footprints, including both locations and travel demands (WU et al., 2014). Human mobility datasets have received increasing attention by the research community to understand human behavior. In (CHO et al., 2013), they used data from two online location-based social networks and data from cell phone location as well. They found that humans experience is a combination of periodic movement that is geographically limited and seemingly random jumps correlated with their social networks. Their results showed that social relationships can explain about 10% to 30% of all human movement, while periodic behavior explains 50% to 70%. (FURLETTI et al., 2013) propose an algorithm for automatically annotating raw trajectories with the activities performed by the users. They analyzed the stop points trying to infer the Point of Interest (POI) that the user has visited. Based on the category of the POI and a probability measure based on the gravity law, they inferred the activity performed.

In more recent studies, Hasan et al. (HASAN et al., 2013) and Wu et al. (WU et al., 2014) used the activity category as a dimension in their analysis. In (Hasan et al., 2013), they discovered a scaling law showing the relationship between the popularity of a place and the probability to select this place as a destination. Their results showed a strong influence of urban contexts on peoples' activity participation and destination choices. Wu et al. (WU et al., 2014) proposed an activity-based model composed of two parts. For the first part, they find the transition probability between activities varying over time, and then they construct a temporal transition probability matrix to represent the transition probability of travel demands during a time interval. In the second part of the work, they suggest that the travel demands can be divided into two classes, locationally mandatory activity and locationally stochastic activity, according to whether the demand is associated with a fixed location or not.

2.6. Human Regularity Mobility

Human mobility studies may be broadly based on the kind of data analyzed: standard "paper and pencil" surveys, and mobility data like GSM cell data, GPS location data, social networks. Among the works that analyses the human dynamics based on surveys there is the pioneering paper from Jiang (JIANG et al., 2012) where a Chicago area activity-based travel survey data describe people movements and activities. They investigated the inherent daily activity structure of individuals, the variation of individual daily activities, how they grow and fade over time, the clusters of individual behaviors and the revelation of their related socio-demographic information.

Another work studying the mobility lifestyle based on a survey is reported in (VANDERSMISSEN et al., 2009). Here authors base their study on a survey conducted in 2006 in Quebec to investigate the flexibility of people to change their residential area after a work change.

Many works base their findings on the analysis of GSM data. The pioneering work on reality mining by Pentland *et al.* (EAGLE; PENTLAND, 2005) posed the basis for the analysis of human mobility with GSM data. They carried on an accurate experiment involving 100 volunteers in the campus. One of the analysis they did is about regularity using entropy, mainly on the home and work locations of users.

Jiang (JIANG et al., 2012) analyze GSM data to find general human mobility laws. They find that human trajectories show a high degree of temporal and spatial regularity. In fact each individual is characterized by a time-independent characteristic travel distance and a significant probability to return to a few highly frequented locations. The regularity of users is often associated with their predictability: the assumption is that the more a user is regular, the more predictable his/her behavior is. This is the main focus of paper (SONG et al., 2010) where the predictability of human mobility has been studied based on the GSM data. Similar to our approach, they propose to use entropy to capture the randomness of human movements.

Cellular phones positioning data is also the basis of the analysis conducted by the work detailed in (QIN et al., 2012) on entropy and predictability of human mobility. They investigate individuals movements trying to predict the next location based on the current one. They combine entropy with clustering to find out which are the reasons of the randomness and thus improve the prediction accuracy.

The work (WILLIAMS et al., 2012) investigates the regularity of routines using the notion of *IVI-irregularity* (*intervisit interval irregularity*) based on a neurophysiology measure. Basically, they are interested in discovering the regularity of visiting to a certain place at similar time. The more the times of the visits are similar, the more the pattern is called regular. For example, an individual visiting a location at very similar times each week is considered to have a highly regular pattern for that location. They tested their method on three real datasets, one is the Foursquare check-ins, one a Wireless Lan and finally the London metro.

The basic differences between these works and this dissertation are: (1) the current work develops a method for regularity respect to a given activity - identified by the POI; (2) in our work, we measure this regularity in numerical values, qualifying in a range between 0 and 1, where 1 means highest regularity and 0 lowest regularity (highest irregularity); (3) another point of difference is the temporal regularity which is not considered in most of these works except the (WILLIAMNS et al., 2012) (EAGLE; PENTLAND, 2005)'s work where they study the temporal regularity using Hidden Markov model, while this current work uses entropy and (WILLIAMS et al., 2012) where they use the inter-visits intervals.

2.7. Extraction of Routine Trajectories

People perform different activities during the day. Some of these activities are daily performed and require a movement by the person from a place *A* to a place *B* as example: go to work, go to home, let children at school, go to the university, etc. To perform daily activities, people consequently generate daily similar trajectories with common spatial and temporal properties.

In this dissertation, we use the contribution of (TRASARTI et al., 2011). Their work in based on the idea that the daily mobility of each user can be essentially summarized by a set of single trips that the user performs during the day. The authors propose to identify the *user mobility profiles* from a set of trajectories, where the interest is on the trips that are part of the person habits, therefore neglecting occasional variations that divert from his typical behavior. The user's mobility profile is defined by a set of user routine trips. The extraction of such profile is made based on techniques of clustering trajectories.

The whole process is divided into three phases, as shown in Figure 3 (a) *Trajectories identification:* divide the whole history of the user into trips; and Figure 3*(b) Detection and removal of outlier group:* grouping similar trajectories and discard outliers. The clustering algorithm k-medoid is used to group similar trajectories. In grouping these trajectories two parameters are considered: Threshold$^{\text{group spatial}}$ and Threshold$^{\text{group temporal}}$. The threshold$^{\text{group temporal}}$ refers to the size of the time window of the trajectories belonging to the group (intervals of 1 hour, 2 hours) and the threshold$^{\text{group spatial}}$ refers to the distance window of the trajectories of the group. With respect to Figure 3*(c), Representatives selection of the mobility profile:* From the group of similar trajectories, the medoid trajectory is chosen to represent a group of routine trajectories. In the selection procedure only groups with a minimal amount of similar trajectories (Threshold$^{\text{group support}}$) will be considered.

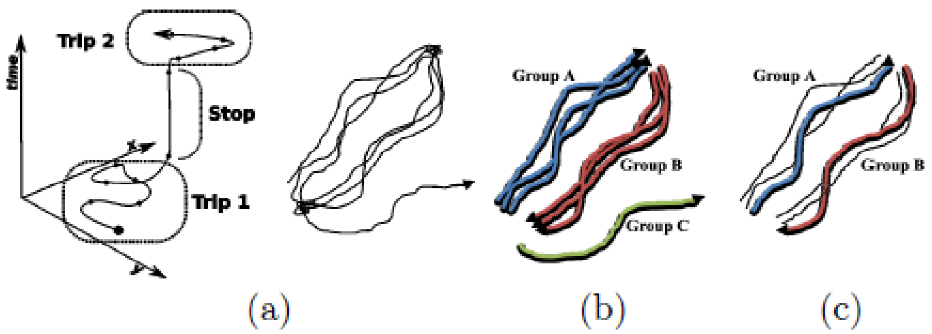


Figure 3: Routine Trajectories Extraction (Trasarti et al, 2011)

They applied the proposed analytical framework to two different scenarios: GPS vehicular data and (simulated) GSM mobile phone data. To understand how the process is affected by the parameters, using different configurations they analyzed the percentage of users with a mobility profile. When looking at the results obtained with a low spatial constraint, they must consider that the clustering method groups together fewer trips and thus pruning using the support threshold becomes more effective. They also found that the temporal threshold for the mobility profile construction does not seem to have much influence, only when assigned very high values.

2.8. Carpooling Systems

There are some current available web systems for carpooling (CARPOOLWORLD, 2014), (DJENGO, 2014), (CARPOOLING.COM, 2014), (CARPOOL.CA, 2014). These systems have similar functionalities like: search for occasional lifts and regular lifts; and match of the lift for groups of users like groups in school, events or companies. The users can have two different roles in the system: driver or passenger. They fill their profile providing some personal information like name, gender, age and information about their trips, such as address of departure and arrival. However, the matching is based on information inputted by the users and notifications services keep the contact between the driver and passengers. These systems require the user to fill information about their daily trips or causal trips.

Other common scenario of study for the application of the carpooling practice is the solution of traffic congestion in regions where large companies dominate the traffic situation. Bellemans (BELLEMANS, 2012) advise to introduce services of this kind using a two-step process: (1) an agent based simulation is used to investigate opportunities and inhibitors and (2) online matching is made available. They argue that incorporating complex negotiations between agents is required for successful carpooling, because inhibiting factors like rerouting and rescheduling have to be considered. Their model relies on comprehensive information which is obtained from big data sources as GPS, GSM and Bluetooth. These sources offer the chance to gather information for a large portion of the population. However, sophisticated extraction methods are required to derive semantic information for the agents' behavior from the raw sources. The web systems (DJENGO, 2014) and (CARPOOLWORLD, 2014) also contain functionalities focused on companies.

Some recent works on carpooling have focus on the real time constraints of their services (SGHAIER et al., 2011), (SGHAIER et al., 2010), (DIMITRIESKI et al., 2013), (DIXIT et al., 2012), (PUKHOVSKIY and LEPSHOKOV, 2011). In (DIMITRIESKI et al., 2013), they describes the design concepts, distribution and cloud computing strategies the authors feel any future global carpool and ride-sharing solution could follow, making it very scalable and ubiquitous enough to successfully reach and serve a global user base. (SGHAIER et al. 2011) and (SGHAIER et al., 2010) addressed the problem of optimizing dynamic requests processing for setting up an optimized dynamic carpool service. (SGHAIER et al., 2010) propose a Distributed dijkstra for the implementation of a Real Time Carpooling system based on the multi-agent concept. In turn, (SGHAIER et al., 2011) adopt a subdivision principle and the multi-agent concept that permit, jointly to a real time locating module, to perform a distributed process. The latter mainly concerns the optimized dynamic assignment of available cars' offers to instantaneously issued users' queries while ensuring traceability, communication and security services.

Minett and Pearce (MINETT and PEARCE, 2011) aims to find out if casual carpooling reduces energy consumption, and if so, how much. Their results estimate that casual carpooling in San Francisco is conserving in the order of 1.7 to 3.5 million liters of gasoline per year, or 200-400 liters for each participant, much of which comes from the impact on the rest of the traffic. In (KELLEY, 2007), a technology is introduced that will allow casual carpooling to function in areas without high-occupancy vehicle lane (HOV), by providing an administrative system that records actual carpooling behavior so that the access to an HOV lane can be made available. The author also addresses some of the current shortcomings associated with casual carpooling such as personal safety, the "free rider" problem, and the disincentive to maximize the number of passengers sharing a ride.

There are several works that address the carpooling issue, but few works have been done in the computer science field by focusing on the casual carpooling and other restrict group consider semantic information in their methods. In addition, most of the work related to carpooling, do not consider the activity as an attribute in their models. One recent work done by Cho (CHO et al., 2013) follow the activity-based approach, In (CHO et al., 2013), the authors use an ontology in an activity-based microsimulation by providing a carpooling case. They introduce related studies and basic knowledge about using methodologies, and provide an example of using Ontology in an agent-based carpooling simulation. While no explicit evidence is presented in this paper, at least they focus in recognizing that Ontology is a useful and appropriate method for the activity-based microsimulation research. However, only a conceptual design and framework are suggested, and this study is a clearly preliminary step.

A recent study in Computer Science about Casual Carpooling (slugging) was done by Ma et al. (MA et al., 2013). In this work, the authors formally define the slugging problem and its generalization. The authors provide proofs of their computational time complexity. For the variants of the slugging problem that are constrained by the vehicle capacity and travel time delay, they prove NP-completeness and also propose some effective heuristics. They performed the experiments in a GPS trajectory data set containing 60 thousand trips, and their results showed that their heuristics can achieve close-to-optimal performances, which means as much as 59% saving in vehicle travel distance.

2.9. Conclusions

There are different types of mobility data (e.g. GPS-Trajectories, GSM data, LSBN data, RFID data, etc.) generated from different types of sensors (e.g. GPS-Devices, Cellphone, RFID antennas, etc.). Thanks to the popularization of GPS-enable devices, increasingly content of mobility data is generated daily. Such data is generated in different types of domain, but as common structure, it has spatial and temporal properties representing a geographic position of a moving object at a specific timestamp.

The analysis of human activity using large information of mobility data is a recent topic in the literature, mainly when we consider the semantic aspects within the data. Several works address the topic over the geometric point of view, in many cases, due the lack of semantic information in the analyzed data. However, thanks to the increasingly advance of mapping service applications and Location-Based Social Network, semantic information has been associated to the mobility data.

These semantic enriched data can be useful for new analysis and this current work takes advantage of it. After a detailed investigation in the literature, we felt the need to measure the behavior of the user according to his behavior of regularity or irregularity about visits to places intending to perform a given activity. Such analysis can be applied to the carpooling topic, where depending on the activity and the user’s flexibility level (i.e. not strictly regular), the user can be reallocated or not to different destinations and at different time intervals.

Chapter 3: Investigating semantic regularity of human mobility lifestyle

This chapter is intended to study the mobility behavior of individuals presenting a definition of user regularity profiles. We measure the user's regularity behavior, identifying those who are more or less loyal to same places in contrast to the irregularity in visiting different places. The measurement is done based on the entropy of both spatial and temporal frequency of visits to places.

3.1. Introduction

This chapter is organized as follows. Section 3.2 presents some preliminary concepts to assist in understanding of the work. Section 3.3 presents the proposed semantic regularity profiles, showing also the computation of the user's regularity measures about the user's choices to visit different places. In order to give an analytical visualization of the proposed measure, the Section 3.4 shows in plotted graphs the result of the experiments in two real dataset of mobility data. Aiming a different point of view of analysis, the Section 3.5 presents a web tool called MAPMOLTY for allowing the analysis of users' loyalty in different types of POIs (Points of Interest). Section 3.6 concludes the chapter.

3.2. Concepts used in this work

Follows some concepts used in this work. We start with the definition of a point, which is defined as in Definition 1:

Definition 1 (Point): A spatiotemporal point p is a tuple <*x*,*y*,*t*>, where *x* and *y* are the spatial coordinates that represent a geographical position and t is the timestamp in which the point was collected.

A raw trajectory represents the raw representation of a trajectory considering only the spatiotemporal properties. It is formally defined in Definition 2.

Definition 2 (Raw Trajectory): represents a finite sequence of spatiotemporal points assigned to a moving object and denoted by <IdObject, <Point>$^+$>, where IdObject is the identifier of the moving object, <Point>$^+$ represents the sequence of points that captures the position of the object. Thus, formalizing: Raw Trajectory = ((id, x1, y1, t1),… , (id, xk, yk, tk)), where for all i ε [1; k], ti <ti +1.

Note that no semantic information is associated with raw trajectories defined according to Definition 2. From the simple and exclusive use of the raw representation of the trajectory, it becomes very difficult to find some interesting semantic pattern (BOGORNY; WACHOWICZ, 2009). A trajectory, in addition to being a spatiotemporal data type has some implied semantic information such as stop moments, called ***stop***, and motion, called ***moves***, where both are not easily extracted from its geometry (SPACCAPIETRA et al ., 2008). In this work, we also use the term ***trip*** to denominate the term move, proposed by (SPACCAPIETRA et al ., 2008). Thus, we formalize the term “trip” as in Definition 3:

Definition 3 (Trip): The trajectory segment between two stops is called Trip and indicates the movement.

In human mobility data usually the last point of a trip is associated to a Point of Interest (POI), representing the place that the person visited. We define a POI in this work as in Definition 4:

Definition 4 (Point of Interest (POI)): A Point of Interest (POI) is a geographical object that is interesting for a specific application, and is usually associated to a human activity. Formally, we define a POI as a tuple *POI = {(s,n,c,t)}* where *s* is the representative spatial point and the other three properties correspond to the semantic information involved. Thus *n* is the name, or label of the POI, and *c* is a category assigned to the POI and *t* represents the opening hours.

An example of POI is the Eiffel Tower: the representative spatial point *s* is the center of the tower, the category can be, for example, “tourist attraction” or “monument” or “tower”, depending on the application, the label “Eiffel Tower” denotes the name and the interval [00:00 - 00:00] is the business hour, in this case, meaning that it is always open for visits.

The time is by itself a continuous measure, in this dissertation for the temporal dimension we discretize the time in periods. Therefore, we have the two following definitions:

Definition 5 (Interval): Correspond to a period of time, where each period has a start hour and an end hour. It is represented as a tuple $<t_{start}, t_{end}>$, where t_{start} represents the start hour and t_{end} represents the end hour and $t_{start} < t_{end}$.

Examples of Intervals: [13:00, 14:00], [08:00, 14:00], [22:00, 23:59].

Definition 6 (Set of Intervals): Basically, it is a collection of Interval (Definition 5) corresponding to slices of the time representing a set of intervals of a given time cycle (for example a day). It is represented by a collection of the tuple of Interval $\{<Interval>^{+}\}$. The set of intervals are linear and ordered such that given two intervals t_1 and t_2, $t_1.t_{start}$ is smaller than $t_2.t_{start}$ and as well as $t_1.t_{endt}$ is smaller than $t_2.t_{end}$.

Examples of set of intervals are: {[00:00, 03:59], [04:00, 07:59], [08:00, 11:59], [12:00, 15:59], [16:00, 19:59], [20:00, 23:59]}, {[00:00, 12:00], [12:00, 23:59]}.

A visit dataset correspond to a repository of information about visits of users to POIs. This data format can be derived mainly from GPS trajectories, by looking for the stop points of theses trajectories, or by collecting the check-ins done by people in location-based social networks. The **Visits dataset** should provide the mobility information to associate a person p to a POI *poi_id* he visited. It is required that each visit is represented by a tuple as follow in the Definition 7:

Definition 7 (Visit): A visit is defined as tuple *<VisitID; UserID; POI; timestamp>*, where *VisitID* is the visit unique identifier, *UserID*, *POI* represent the place formally defined in the Definition 4 and the *timestamp* correspond to the time when *UserID* arrived at *POI*.

Many web services, like for example Foursquare, provide this set of attributes of a visit. Using the Visit dataset become possible associate the specific activity performed by the user during the visit, like stop at a supermarket to do shopping. We assume to have a static mapping, when associating POI categories to activities, therefore, for the remaining of this dissertation, we might refer to POI category to implicitly indicate an activity.

3.3. Semantic Regularity Profile

This section introduces the measures to estimate the regularity of a user when performing different activities (LIRA et al., 2014a). The semantic regularity behavior is measured according to two dimensions: spatial and temporal.

The spatial measure indicates how much a user tends to visit the same places to perform a given activity. Consider for example the shopping activity: in an urban area there are several shops belonging to the category supermarkets. However, despite the possibility of diversifying the choice, a user who tends to go always to the same shop is more *spatially regular* respect to the shopping activity. In other words, it is possible to say that he/she is *loyal* to that specific shop. Naturally, by their intrinsic nature, some places are visited more regularly like, for example, home, work, schools and universities. Other places are inclined to have irregular visits like restaurants, bars, shops and gas stations.

The temporal dimension measures the regularity of the user to perform an activity in a preferred temporal interval. When a user tends to change the time slots to visit some kind of place, then we say the user has a low temporal regularity. On the contrary, a user has a high temporal regularity when he usually uses the same temporal interval to perform a given activity.

The process of computing the regularity measures is structured into three phases: *(i)* data collection of users' visits to POIs; *(ii)* estimation of spatial and temporal regularity measures; and *(iii)* extraction of the semantic regularity profiles. The overall process is illustrated in Figure 4 and the three steps are detailed as follows.

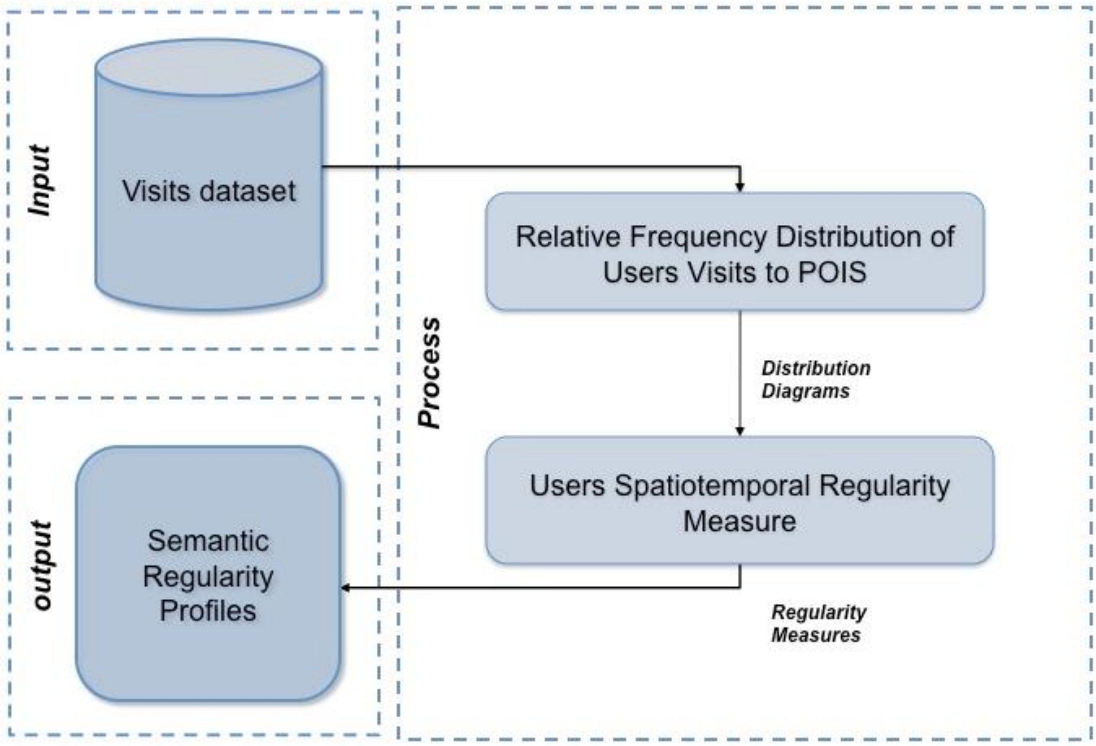


Figure 4: Semantic Regularity Profiles' Process

3.3.1. Relative Frequency Distribution Of Users Visits to POIs

Given a user *u* and a category *C*, we want to estimate the spatial and temporal regularity of *u* to perform the activity associated with *C*. For this, we first compute a spatial and a temporal distribution considering the visits of the user. The *Spatial Frequency Distribution* (SFD) of a given user *u* in a place *p* is defined as in Equation 1.

$$SFD(u,p) = \#\,visits\ of\ u\ to\ p$$

Equation 1: SFD(u,p)

The probability of observing u in a place p of category C is considered for the spatial dimension. Formally, for a POI p of category C the spatial relative frequency distribution of u is defined as in Equation 2.

$$SRFD(u,C,p) = P(u \text{ in } p/C) = {}^{SFD(u,p)}/_{\#visits\ to\ C}$$

Equation 2: SRFD(u,C,p)

For example, for the category Supermarket, each supermarket instance (Carrefour, WalMart, etc) has calculated the probability SRFD. Figure 5 illustrates the SRFD estimation for the category Supermarket and for a given user. This figure shows that the user stopped at 7-Eleven 20% of the times she/he stopped at a supermarket, 32% at Carrefour and 48% at Walmart.

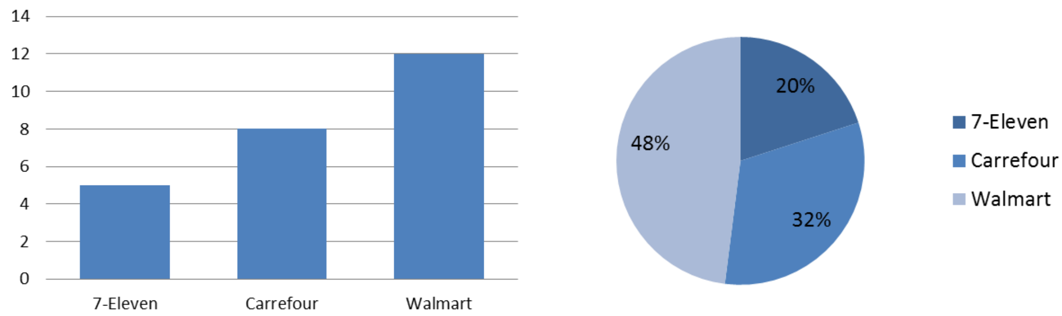


Figure 5: Number of visits to places (SFD) (left) and the respective SRFD (right) for the category supermarket of a specific user

For the temporal dimension, we compute the *Temporal Frequency Distribution* (TFD) counting the number of visits of a user u to a place of category C in an interval t (Definition 5) of the day as defined in Equation 3.

$$TFD(u,C,t) = \#\ visits\ of\ u\ to\ C\ in\ t$$

Equation 3: TFD(u,C,t)

The probability of observing u in a place p of category C at a time t is computed. Formally, for a POI of category C the temporal relative frequency distribution of u is defined as in Equation 4.

$$TRFD(u,C,t) = P(u \text{ in } t/C) = {}^{TFD(u,C,t)}/_{\#visits\ to\ C}$$

Equation 4: TRFD(u,C,t)

For efficiency issues, the time is discretized in Set of Intervals as defined in Definition 6. The estimation of the probability function *TRFD* is performed on this temporal grid. The intervals can be adjusted according to the application scenario considered. In our experiments, we consider the intervals within a day, ranging from 0h00 to 23h59. Examples of possible values are [06:00 - 07:00], [07:01 - 09:00] and [09:01 - 12:30].

Figure 6 shows an example of a *TRFD* computation for the category Supermarket and for a given user. This figure illustrates that the user stopped at supermarkets 25% of the times during the temporal slot [04:00-08:00], 15% of the times during the period [08:00-12:00], etc.

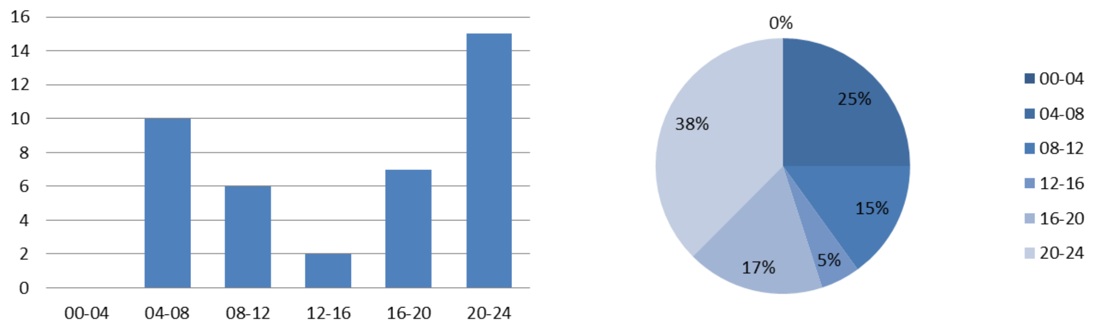


Figure 6: Number of visits by intervals (TFD) (left) and the respective TRFD (right) for the category supermarket of a specific user.

3.3.2. User Spatiotemporal Regularity Measure

In the previous section, the regularity of a user for a specific category was defined. Now our intention is to extend the concept of regularity to the whole mobility of a user, by taking into account the behavior of visiting all the available categories. The methodology proceeds by estimating the spatial and temporal variability separately, based on Shannon entropy estimation (SHANNON, 1948) of the frequency distributions.

Given a user *u* and a place category *C*, his *Spatial Entropy* (*SH*) is defined in Equation 5.

$$SH(u,C) = -\sum_{p \in C} SRFD(u,p,C) log SRFD(u,p,C)$$

Equation 5: SH(u,C)

Analogously, the *Temporal Entropy* (TH) is defined in Equation 6.

$$TH(u,C) = -\sum_{p \in T} TRFD(u,t,C) log TRFD(u,t,C)$$

Equation 6: TH(u,C)

Thus, for the spatial dimension, the entropy values are normalized for all categories in the interval [0,1]. The *Spatial Maximum Entropy* (SMH) for each category is computed using the Equation 7, *where |C| represents the number of places of category C.*

$$SMH(C) = log|C|$$

Equation 7: SMH(C)

For the temporal dimension, the entropy values are also normalized for all categories in the interval [0,1]. By means of the *Temporal Maximum Entropy* (TMH) for each category is calculated using the Equation 8, *where |I| represents the number of intervals of the day.*

$$TMH(I) = log|I|$$

Equation 8: TMH(I)

Given a user u and a category C, the *Semantic Spatial Regularity* (SSR) for C is defined in Equation 9.

$$SSR(u, C) = 1 - \frac{SH(u,C)}{SMH(C)}$$

Equation 9: SSR(u,C)

Given a user u, a set of interval I and a category C, the *Semantic Temporal Regularity* (STR) for C is defined in Equation 10.

$$STR(u, C) = 1 - \frac{TH(u,C)}{TMH(I)}$$

Equation 10: STR(u, C)

Both measures, SSR and STR, yield values in the interval [0,1]. The spatial measure SSR tends to one when the user chooses the same place *p* when performing activity *C*. The value tends to zero when the user always visit a distinct place when performing *C*. The temporal measures tend to one when the user performs activity *C* at the same time interval. The value tends to zero when the activity is performed always in a distinct interval. We can realize that for the regularity measure computation, we do not consider the order of the visits, but only the visit's choices about the places and the interval represented in their relative frequencies.

An example for the Gym category illustrating the idea behind the spatial and temporal regularity measures is shown in Figure 7. This figure shows that this user usually goes to a specific gym (black marker) where the relative frequency value is 92% while other gyms have very low and zero values. Therefore, the user is very regular in visiting gyms and we can see this as a loyalty measure of the user of that particular gym. About temporal aspects, the situation is different since the gym is visited during the temporal interval [16:00, 20:00] 10 times which is the maximum, but the other time slots frequency is not considerably different (with 8 times, 5 times, etc), thus making the temporal regularity lower.

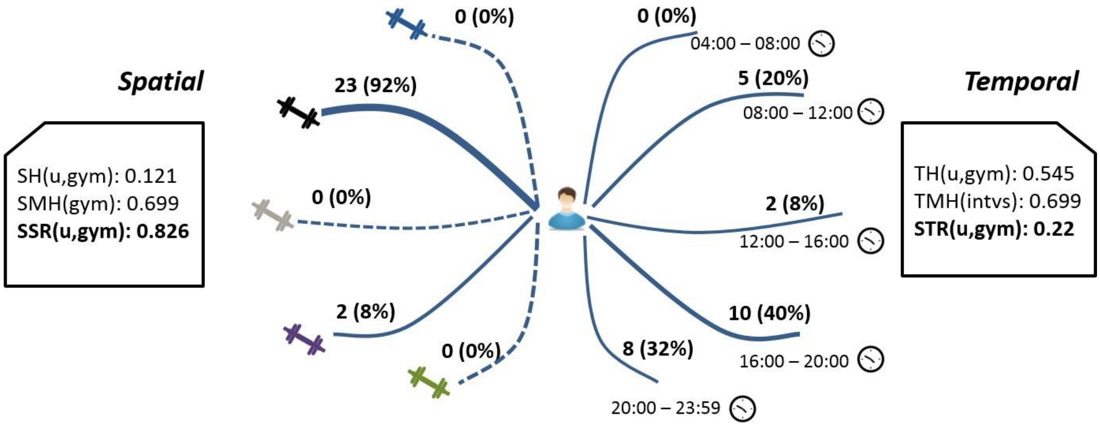


Figure 7: Example of SSR and STR for Gyms.

The corresponding semantic regularity profile is defined combining the regularities on space and time according to the Definition 8.

Definition 8 (Semantic Regularity Profile - SRP(u)): A semantic regularity profile for a user u is denoted by *SRP(u)* and consists of a set of tuples < *Ci,SSR(u,Ci),STR(u,Ci)* >, for all *Ci in C1,C2,...,Cn.*

Having the regularity profiles for each user, it allows us to analyze and compare them to identify groups of users with similar characteristics. These similarities may help in outlining some classes of users representing a typical behavior. For example, it could be possible to find a group of users very regular in visiting supermarkets and bars, but not loyal to any restaurant or cinemas and so on. Nevertheless, methods to compute profile similarity are out of the scope of this work and they are sketched here only to motivate the use of this global measure.

3.4. Experiments

Using the process shown in Figure 4 we compute using two real datasets, the semantic regularity measures (spatial and temporal) that are seen as a regularity measure with respect to each activity. The analysis is based on the concept of regularity quadrant. For each activity, the regularity measures were plotted in a chart where the X and Y-axes have values ranging from 0 to 1. The X-axis denotes the spatial measure, while the Y-axis denotes the temporal measure. This leads to an easy interpretation of the plots since a point located at position (0,0) indicates the minimum spatial and temporal regularity for that activity category. We interpret as a user having high irregularity for spatial and temporal aspects (e.g. this would be the case of a user who shops in different supermarkets at different times of the day). On the contrary, a user represented as a point at position (1,1) shows regular movements since she/he does not change the place and the time where a given activity is performed. Using this idea, the Cartesian plane is splitted into four quadrants with equal areas. The profiles of users who have higher regularity in both space and time will be quadrant top right (TR), the users who has high spatial regularity and not temporal will be in quadrant bottom right (BR), while temporal regularity and not spatial is in quadrant bottom left (BL), and neither spatial nor temporal regularity in quadrant top left (TL).

In order to give an analytical visualization of our proposed measure, we computed the semantic regularity profile SRP(u) (defined in Section 3.3, Definition 8) for identifying the semantic regularity of users' activities on two mobility datasets: (i) A GPS dataset tracing cars in Florence (Italy) and (ii) a set of check-ins generated from a Location-based Social Network (LBSN), called Brightkite[5]. The results show the POI categories with the highest amount of stops or check-ins according to the dataset used in the experiments. These experiments are aimed at validating the application of the proposed measure in these two different application domains and are described in the following sections.

Intending to limit the sparsity of the spatial and temporal frequency distributions for each category, in both datasets and for both dimensions (spatial and temporal) only the first five higher frequency distribution values are considered. This is to avoid many zero values that would affect the entropy and maximum entropy computation. The idea behind this is that, when many POIs are available for a given category, the users could not be able to visit, for example, all the restaurants around the city, neither visit restaurants in all the possible intervals of the day. To compute the semantic temporal regularity, the following set of intervals (Definition 6) were used: {[00:00:01, 04:00:00], [04:00:01, 08:00:00], [08:00:01, 12:00:00], [12:00:01, 16:00:00], [16:00:01, 20:00:00] and [20:00:01-00:00:00]}.

3.4.1. The OctoFlorence Dataset

[5] http://snap.stanford.edu/data/loc-brightkite.html

The OctoFlorence dataset consists of 33055 trajectories representing the movements of 1142 users moving by car in the city of Florence (Italy) during the period of 1st of May to 31st of May 2011(GIANNOTTI et. al., 2011). This dataset has been collected by OctoTelematics[6] Company putting a GPS box in the car of users who obtained an insurance discount. The Visits dataset, input of the method, is obtained from these trajectories by computing the stops in the trajectories. The stops detection from the trajectories used a spatiotemporal threshold of 50 meters and 20 minutes. This means that a user staying within a circular area of 50 meters for at least 20 minutes is considered as stopped.

We needed to adapt our input data composed of trajectories into a visits dataset composed by tuples of *Visits* (Section 3.2, Definition 7) as was shown in Section 3.3 to be required for the computation of the measures. Since the dataset is devoid of semantic information, it requires a semantic enrichment step to adapt the data to the computation of the spatial measure. In our case, the annotation is done associating the stop point of each trajectory to a specific POI. A *Visit* (Section 3.2, Definition 7) requires as semantic information a *POI_ID* and *CAT_ID*. To provide such information, we used the Foursquare API.

We proceed annotating only the stops to the POIs downloaded from Foursquare using a spatial buffer of 300 meters around each stop and considering the 5 closest places. From the result obtained from Foursquare, the most frequent place category found in the buffer of the nearby resulting data is associated to the given stop. Figure 8 shows an example where among the 5 closest places from Foursquare, 3 POIs are restaurants and 2 are stores. Even if the closest POI is a store, it is not the majority category since the higher number of places is of the category restaurant with 3 POIs. The closest POI inside this category is selected. However, it is worth to point out that the methodology is parametric to the way POIs are assigned to stops and it can assume that such an assignment has been done with any available method, like in (FURLETTI et al., 2013).

[6] http://www.octotelematics.it

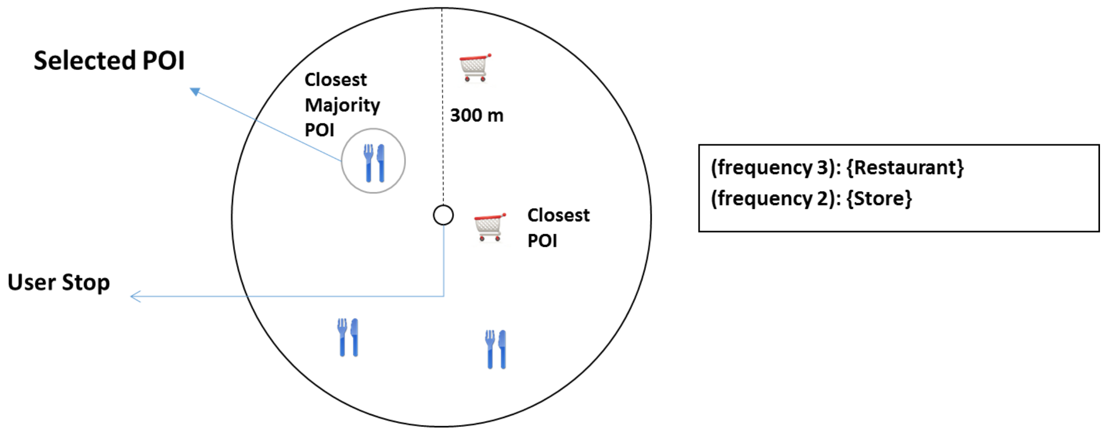


Figure 8: The majority category for the association of POIs

The entire dataset comprised 1637 different POIs visited by the users. Since the categories of POIs resulted from the Foursquare were very sparse and often too detailed, we manually grouped the similar categories into one single category. For example, Indian restaurant, Chinese restaurants, Fast food, etc, were grouped into the larger category labelled Restaurants. We used this larger category as the category to be mapped to activities for computing the semantic regularity measures.

Figure 9 starts our analysis showing how users are located in the different quadrants for the categories: Gym and Store. For the Gym category, the regularity profiles tend to stay in the TR and mainly in the BR quadrants thus showing high spatial regularity. About the temporal aspects, there are different behaviors: some are regulars and others irregulars. As expected, the result for the Gym category shows that many users tend to change the time to go to the gym, but not the location: the explanation is that usually people have a monthly subscription to a specific gym. Analyzing the Store category we can observe a high irregular temporal behavior where most of the users are concentrated in the BL and BR quadrants. In fact, a person decides according to his preferences when to go to a store during its business hour. Looking at the spatial results, we find regular users concentrated in the quadrant TR and BR, as well as irregular users concentrated in the TL and BL quadrants.

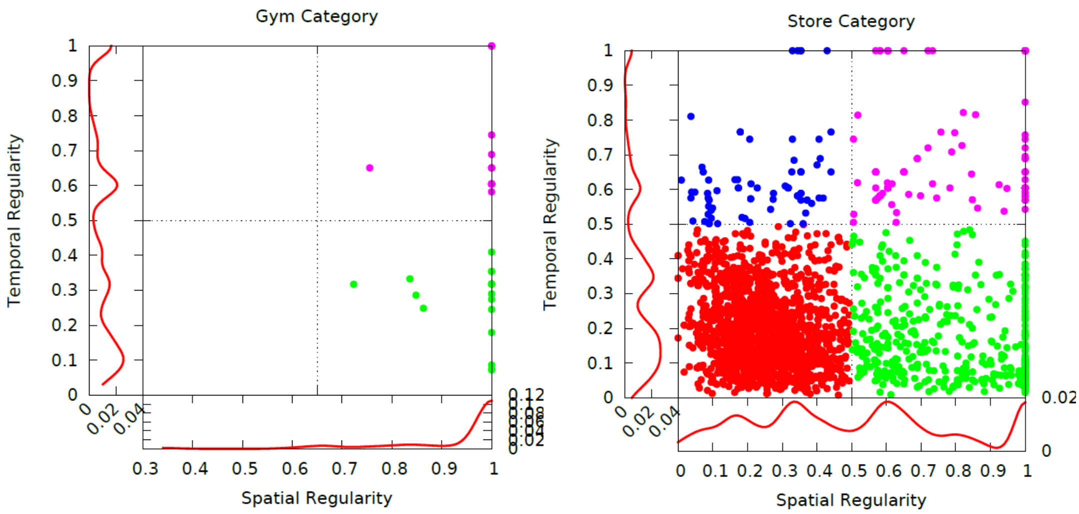


Figure 9: Plots indicating the collocation of users visiting Gyms(left) and Stores (right)

Figure 10 shows our second analysis: the percentage of users belonging to each quadrant for each activity category. The first group of bars represents the percentage of users belonging to the quadrant TR (regular both in space and in time) for each category, the second group of bars represents the percentage of users belonging to the quadrant BR (regular in space and not in time) and so on. Seeing for the store activity category, a high percentage of users is concentrated in the quadrant BL, which is quite intuitive since there is a high tendency for users to change the shops where to go shopping and visiting them at different temporal slots. On the contrary, universities are mainly present in the quadrants TR and BR, where the spatial regularity is high. This is also intuitive, since students or professors attending a university usually go to the same university, even at different temporal slots. The same applies to the gym and offices since people have tendency to visit the same places for studying and working.

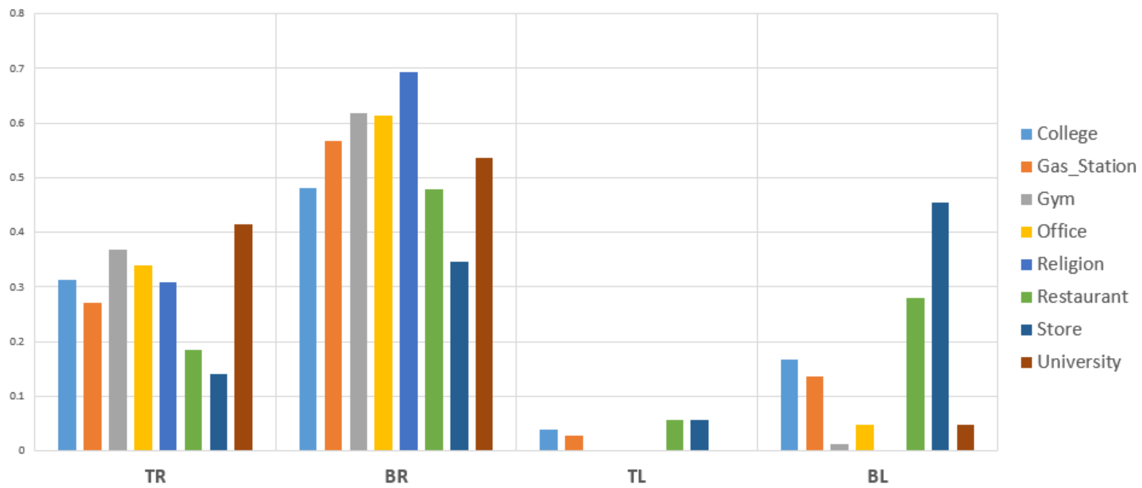


Figure 10: Percentage of users belonging to each quadrant for each activity category

After having discussed the overall view for the regularity of users with respect to the activities, the analysis of the results goes more in depth to try to understand if (semantic) regularity is a characteristic of the user or is more related to the specific activity. Examples of this kind of analysis are the plots shown in Figure 11 and Figure 12. Figure 11 depicts the spatial regularity trend of the (anonymized) user 427158. The blue horizontal bar represents the median of spatial value, while the X-axis shows the categories visited and the Y-axis the respective spatial measure. This user has a different behavior depending on the activities since he/she shows high spatial regularity for offices (work places) and residential, while for restaurant his/her behavior has low regularity level. About the temporal profile, in Figure 12, the user behavior is more regular for the POI categories religion and restaurant. The low temporal regularity for residential and office may be explained because the user can leave and come back to home and work several times during the day like lunch time, extra commitments and delays.

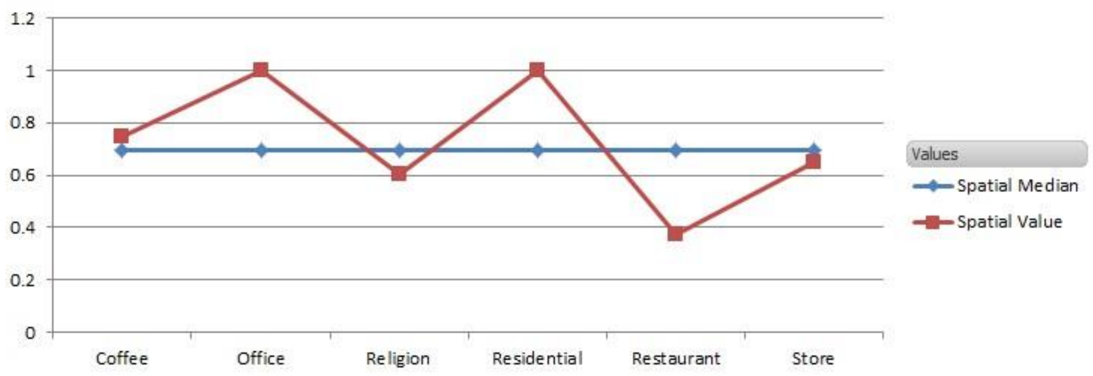


Figure 11: Spatial Regularity Trend for user 427158

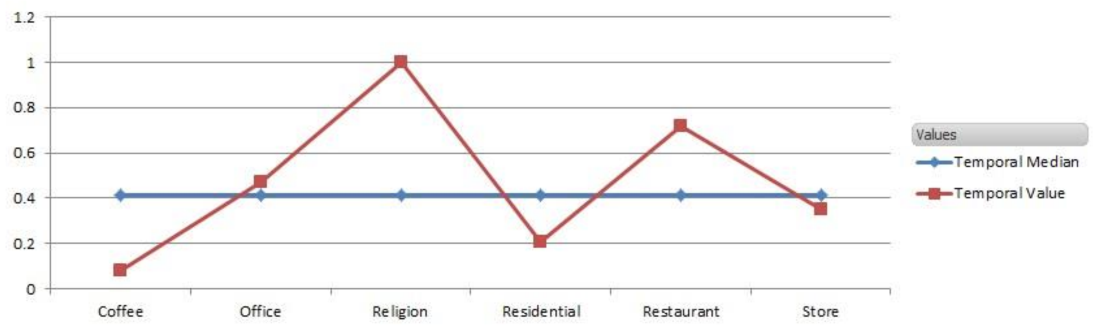


Figure 12: Temporal Regularity Trend for user 427158

3.4.2. The Brightkite Dataset

The Brightkite dataset has a total of 968.784 check-ins performed by 2806 users around the world between March 22nd, 2008 and October 18th, 2010. In this dataset, the check-ins are composed by the user identification, the geographic coordinates of the check-in and the time instant when the user performed the check-in. Since the dataset was obtained without the semantic information required for using our method, we also used the Foursquare API to annotate semantically the places where users performed the check-ins. A check-in at a specific place means that a user is trying to notify his friends that he is at a given place at a given time. Thus, since a check-in corresponds to a user's intentional action to inform about his/her location, the check-ins used in this experiment were annotated with the closest place indicated by the Foursquare API. Figure 13 shows an example where a supermarket is associated to the Check-In to be the closest place.

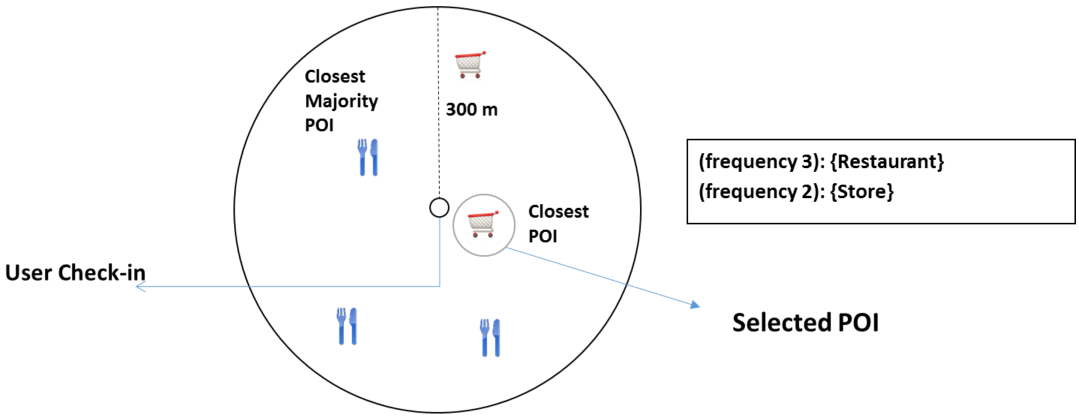


Figure 13: The closest place for the association of POIs with checkins.

In Figure 14, we have our first analysis for the Brightkite dataset. This Figure shows the quadrants of users' profiles for the POI Categories University and Restaurant. About the University category clearly it is possible to notice a very regular spatial behavior (quadrants TR and BR), where most of the users are distributed close to the value 1 on the spatial dimension. This indicates a loyal attitude for the users in visiting the same place for an activity. For the temporal aspects, there is a more equal distribution of the values showing different patterns of behavior, some more irregular and others more regular. On the other hand, the Restaurant category shows that most people tend to change the place when they go eating and also the time when they go to this place category, i.e. indicating that most of them are irregular in space and time. In addition, there are few loyal users (quadrant TR) meaning that they tend to go to the same restaurant and mostly at the same time. These users could be, for example, the target for an advertisement recommendation trying to encourage them to visit different restaurants or for sending customized promotions. The temporal part is also important since it could be used for some special service based on public transportation or carpooling to favor those users who go to a specific place always at the same time.

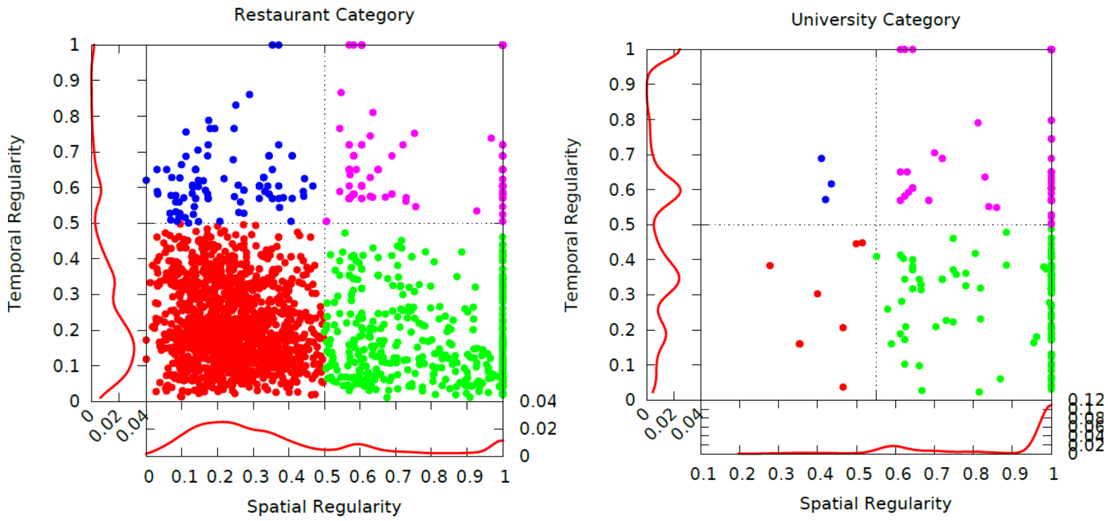


Figure 14: Plots indicating the collocation of users visiting Restaurants(left) and University(right)

Our second analysis for this dataset is presented in the Figure 15. The first group of bars representing the quadrant TR has a high percentage of users with high values for the spatial and temporal regularity measures, mainly for the categories: Mall, Bakery, and Hospital. Activities having a high percentage of users with a high spatial regularity are found in the second group (BR): Religion, Residential, University and Gym. The third group (TL) shows the POI categories with high temporal regularity and low spatial regularity. There are no categories with high percentage of users in this third group. The last group (BL) with low spatial and temporal regularity contains mainly the categories Restaurant, Stores and Bar.

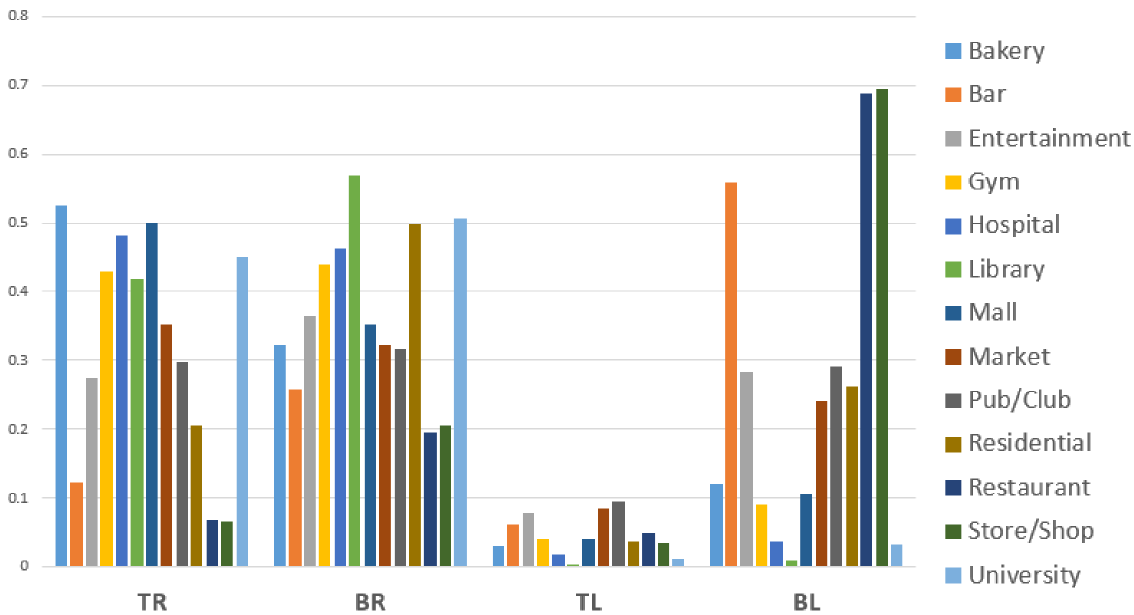


Figure 15: Percentage of users belonging to each quadrant for each activity category

As a further step comparing the two datasets used in the experiments is possible identify that some categories of places like Restaurant, Market, Stores and Bar have a certain spatial and temporal irregularity about the users' visits, while the categories Gym, College and University maintained their spatial and temporal regularity. From the results it is also observed that the percent of users for the 3th quadrant is low in both datasets. This demonstrates that it is uncommon that a person keeps a regular behavior about temporal aspects and an irregular behavior about the spatial aspects to perform a specific activity.

3.5. MAPMOLTY

When we talk about the user regularity in his choices to visit a specific POI to perform an activity, indirectly we are talking about the loyalty level between the user and the visited POI. A person to be loyal respect to a place means that for a particular place category (e.g. Restaurant, Bar), the person often goes to the same place even if there are others. Thus, a person that usually chance their choices about his visits to place of a given category does not have a loyal relationship with any place of this category.

With the semantic regularity profiles (Section 3.3, Definition 8), we profile the user considering the regularity of his/her visits. For profiling the user, the computation for the spatial measure takes into account the frequency distribution of his/her visits. Such information can also be applied for the analysis of loyalty between a place and a user. Given a user and given an activity, a high relative frequency distribution value related to a place represent a high loyalty level between them (i.e. the user and the place), so contrary, low relative frequency distribution value represents no loyalty behavior for this place.

From the user profile analysis, emerged the idea of profile also the POIs with respect to the loyalty behavior of their visitors. Changing the focus of the analysis from the persons and by bringing our attention to the POIs, we developed a web tool called MAPMOLTY (MAPping MObility loyaLTY)[7] (LIRA et al., 2014b). MAPMOLTY profile the POI of a given category respect to loyalty indicators, where each indicator is a measure about the POI related to the behavior of visits of their visitors. Given a dataset of mobility data, like tracks or check-ins of individuals, and a set of Point Of Interest,

[7] https://mapmolty.isti.cnr.it

MAPMOLTY computes a number of measures, the loyalty indicators, to summarize the loyalty level of the visitors of each POI.

The analysis enabled by the tool may be useful in different scenarios. Some examples of questions that could be supported by the tool to help in the answers are:

a) Where should I place a restaurant if I want to favor the attendance of loyal customers instead of more occasional customers? What if I want to favor the overall number of visits?

b) Which are the Bars where the customers are less loyal?

c) Which are the most loyal places of a given category? Which are the less?

d) Which are the categories of places where people tend to be more loyal?

e) How is the loyalty frequency of visitors to a given POI comparing to its competitors?

f) How is the behavior about loyalty of the visitors of a specific area for a given city?

From the loyalty measure of the user we can derive a loyalty map of a territory. We can discover that some areas have the tendency to be visited by loyal users, while other areas are more characterized by occasional visitors.

3.5.1. The Tool

An overview of the tool is illustrated in Figure 16. As we can see, it is organized into three components as described below. Each component is described in the next sections. The *Visits Dataset* is detailed in the Section 3.5.2, the *CORE* and the *User Interface* are addressed the Sections 3.5.3 and 3.5.4 respectively.

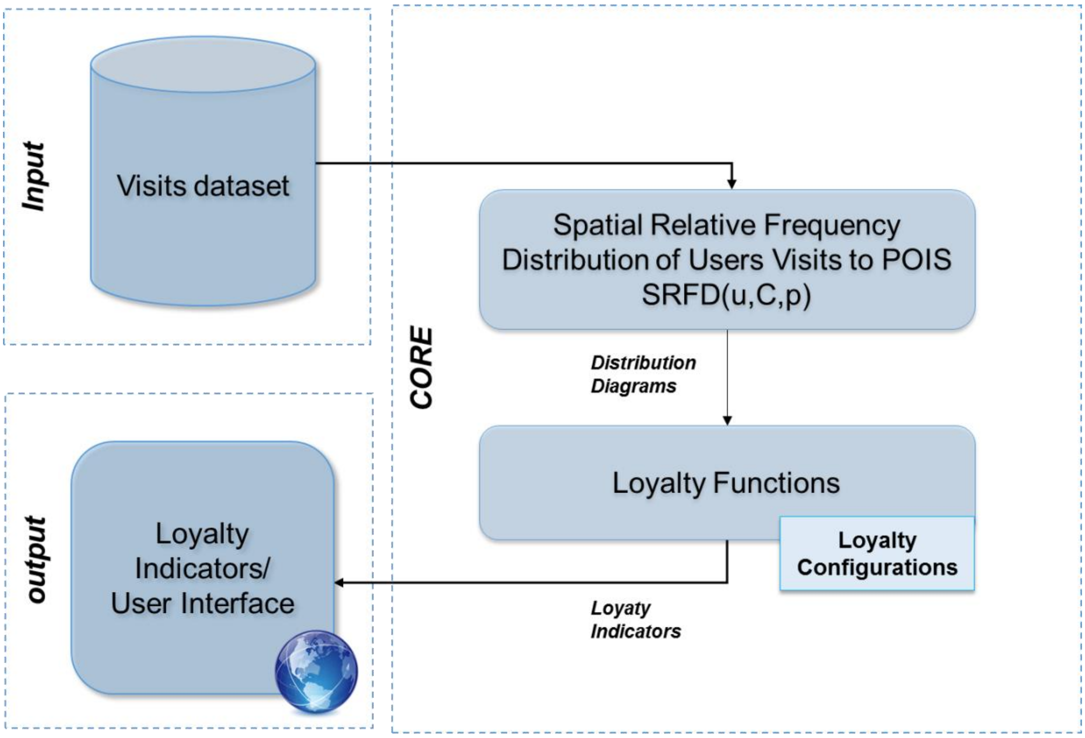


Figure 16: The overview of MAPMOLTY with the three components

3.5.2. Visits Dataset

We require that the **visits dataset** provides the mobility information to associate a person p to a POI *poi_id* he visited. The format of visit dataset was presented in the Section 3.2, through the Definition 7.

MAPMOLTY works with many *different* data sources like: Location Based Mobile Social Networks (e.g. Foursquare, Jiepang, Brightkite) and GPS traces. However, a transformation may be needed to convert the mobility data into the visits dataset format. For example, let us consider a sequence of time stamped GPS points for an individual (called trajectory). A trajectory can be transformed into a sequence of visits in a two-step process: (1) we detect stops, i.e. subsequences of points where the user stands still for a minimum amount of time; (2) we associate each stop to the closest POIs provided by the POIs dataset[8]. For the stop detection, we use two parameters: δ, a spatial tolerance threshold and τ, a temporal tolerance threshold. In our experiments, we used $\delta = 50m, \tau = 20min$, meaning that we detect a stop if the user stays in an area of radius 50 m for at least 20 minutes.

[8] More sophisticated Stop-POI association techniques can be used

MAPMOLTY uses PostGres SQL 9.3 [9] with PostGIS 2.1.1 [10] as Data Base Management System (DBMS) to store the data. PostGIS provides the spatial extension for the PostgreSQL database, allowing storage and query of geographical data.

3.5.3. Core

This module analyzes the visits dataset to derive loyalty indicators about the POIs. MAPMOLTY works with three main different visualization properties: the marker, the radius of the circle and the opacity of the circle. The CORE module computes for each POI, all the loyalty indicator values and also, all the property visualization that will be showed in the *User Interface* module.

So far, MAPMOLTY contains these following loyalty indicators: (1) Number of Visits; (2) Number of Visitors; (3) Number of Loyal Visitors; (4) Number of Non Loyal Visitors; (5) Average Relative Frequency of All Visitors; (6) Average Relative Frequency of Loyal Visitors; (7) Average Relative Frequency of Non Loyal Visitors; (8) Average Visits by Visitors. The computation of these indicators will be detailed in the Section 3.5.3.2.

A visitor is loyal to a place for performing an activity when his visits to this place are frequent and regular compared to other available places. MAPMOLTY measures the regularity of a user computing his spatial distribution over the frequency of visits to places of a given specific activity. Starting from the spatial distributions of visits showed in the Section 3.3.2, MAPMOLTY computes the loyalty indicators and the property visualization for each POI, by considering the *Loyalty Configurations* that are detailed as follows.

3.5.3.1. Loyalty Configurations

There are 3 groups of configurations in MAPMOLTY: *Fidelity Configuration, Marker Visualization Configuration, Circle's Size and Circle Opacity Visualization Configuration*. An important point of the tool is that these configurations can be personalized according to the user's preference by changing the visualization perspective on the *User Interface* component. The interface to set the tool configuration is a dialog

[9] http://www.postgresql.org/download/

[10] http://postgis.net/install

page divided into four tabs. Each tab is related to a group of configurations. These configurations are listed below:

a) Fidelity Configuration.

i. **Relative Freq. for a Loyal User (Min_{RFLU}).** Defines how the MAPMOLTY considers a person loyal to a place. This configuration represents the minimum value of the SRFD(u,C,p) for a user *u* be considered loyal to the POI *p* of the category C. Then if a given user *u*, in respect a place *p* of the category *C* has a SRFD(u,C,p) equal or higher than Min_{RFLU}, it means that the user *u* is loyal about visits to the place *p*. Figure 17 shows the tab to configure the Relative Freq. for a Loyal User (MinRFLU).

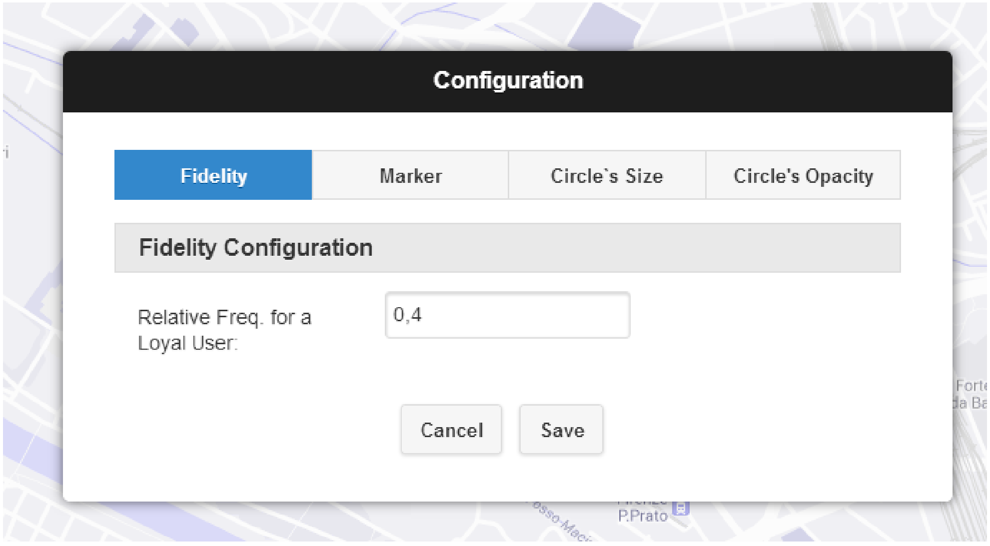


Figure 17: Fidelity Configuration

b) Marker Visualization Configuration:

i. **Loyal Indicator.** Given the list of loyalty indicators, one item from the list is configured for the Marker Visualization. Then, for a POI p, the Marker Visualization Property will be showed according to his value $V_{MarkerInd}(p)$ for the loyalty indicator configured.

ii. **Split Configuration.** The values of the indicators are discretized considering from two until four groups of value intervals defined by 3 limit values (S_1, S_2, S_3). The first divisor splits the visualization marker property in *Red Marker*, where $V_{MarkerInd} < S_1$, and *Orange Marker*, where $S_1 \leq V_{MarkerInd}$, if the second split S_2 is not configured, or $S_1 \leq V_{MarkerInd} < S_2$, if S_2 is configured. Including a value for S_2, introduce the light blue, where $S_2 \leq V_{MarkerInd}$, if the third split

S_3 is not configured, or $S_2 \le V_{MarkerInd} < S_3$, if the S_3 value is configured. And including the S_3 introduce the dark blue, where $S_3 \le V_{MarkerInd}$. Figure 18 shows the tab to set the marker visualization configuration and it is possible to see the marker configuration splits.

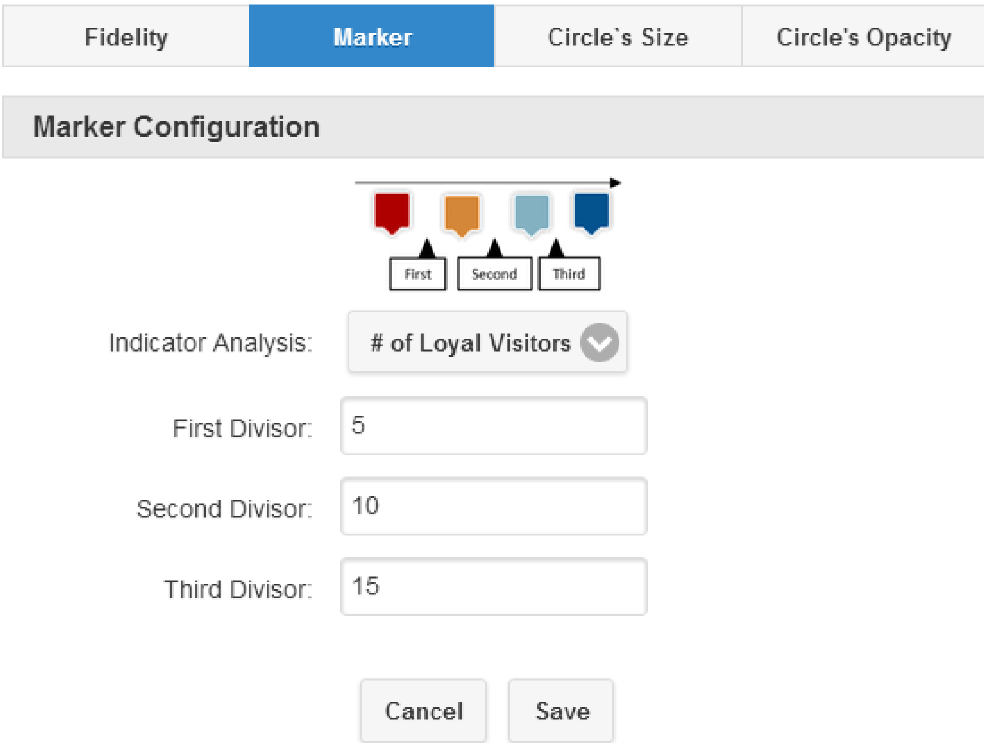


Figure 18: Marker Configuration

c) Circle Size and Circle Opacity Configuration:

i. **Loyal Indicator.** Given the list of loyalty indicators, one item from the list is configured by the user for the Circle Size Visualization and another for the Circle Opacity Visualization. Then, for a POI *p*, the Circle Size Visualization and the Circle Opacity Visualization Properties will be showed according to their values $V_{SizeInd}(p)$ and $V_{OpacityInd}(p)$ respect to the loyalty indicator configured for each.

ii. **Normalize Using Limit.** It is a Boolean property. Each property: circle size and circle opacity, contains their own limits and they are computed in the same way in the core module. When it is true, two limits values (L_{min} and L_{max}) are configured by the user to normalize the data according to the range (L_{min} and L_{max}). When it is false, the normalization is done using the minimum (V_{min}) and the maximum (V_{max}) values for the considered loyalty indicator using all

POIs of the given Category *C*. Then the values for all the POIs are normalized according to the range ($V_{min,}$ V_{max}). Figure 19 shows the increasing of the radius, how higher is the radius higher is the value of the loyalty indicator configured. Figure 20 shows the changing of opacity that follows the same idea as the radius, i.e. how higher the opacity, higher is the value of the loyalty indicator configured for the opacity property.

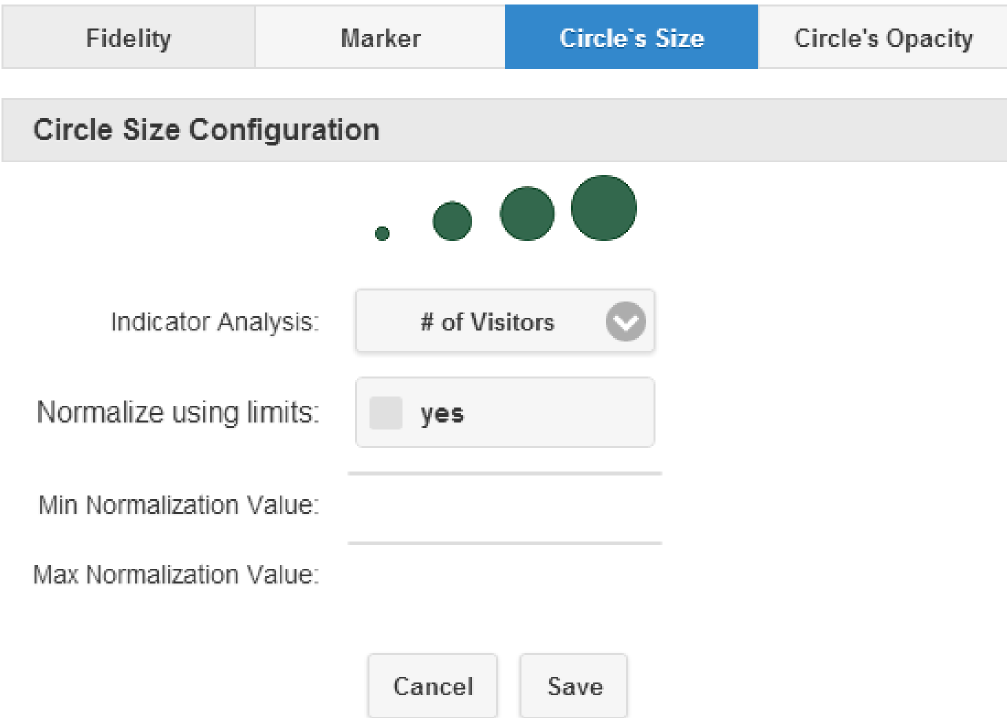


Figure 19: Circle Size Configuration

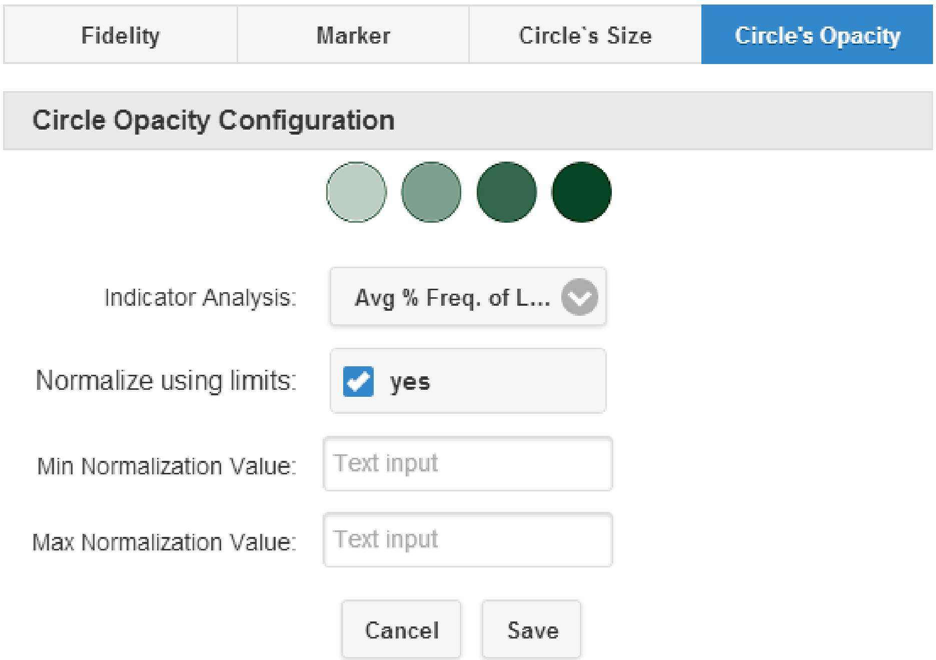


Figure 20: Circle Opacity Configuration

3.5.3.2. Loyalty Indicator

So far, MAPMOLTY contains a total of eight loyalty indicators. These indicators are implemented as SQL procedures developed in PostGres SQL. One peculiarity of this web tool is that it provides a first kernel of features and it can be easily extended with new functionalities. For example, we can incorporate new analysis functions in the database like statistics functions (median, mode and so on) over the data and consequently update the web application to display this new information. Then the tool can easily be extended including new indicators.

For a single POI *p* in a given category C, MAPMOLTY computes the loyalty indicator using the set of users U that visited *p*. This computation is done using the fidelity configuration $\mathrm{Min}_{\mathrm{RFLU}}$ and considering the values of *SFD* (*u*, *p*) and the *SRFD* (*u*,*C*,*p*) for each u in U. It is noteworthy that *SFD* is the Spatial Frequency Distribution and *SRFD* is the Spatial Relative Frequency Distribution, both defined in Section 3.3.1. The loyalty indicators and their computations are listed below:

(i) **Number of visits:** Represents the number of visits performed by U in the POI p. Its computation is done according to the Equation 11.

$$\# \, \boldsymbol{of} \, Visits = \sum_{\boldsymbol{n=1}}^{|\boldsymbol{U}|} \boldsymbol{SFD}$$

Equation 11: # of Visitors Computation

(ii) **Number of Visitors:** represents the number of users U that visited p. Its computation is done according to the Equation 12.

$$\# \, of \, Visitors = |U|$$

Equation 12: # of Visitors Computation

(iii) **Number of Loyal Visitors:** represents the number of loyal user in U related to p given the activity category C. To characterize whether a User is loyal, it depends on the configuration MinRFLU, defined in the section 3.5.3.1 of this chapter. Its computation is done according to the Equation 13.

$$\# \, of \, Loyal \, Visitors = \sum_{i=0}^{|U|} f(u_n, C, p)$$

$$Where,$$

$$f(u, C, p) = \begin{cases} 1, & SRFD(u, C, p) \geq Min_{RFLU} \\ 0, & SRFD(u, C, p) < Min_{RFLU} \end{cases}$$

Equation 13: # of Loyal Visitors Computation

(iv) **Number of Non Loyal Visitors:** represents the number of non-loyal user in U related to p. It is basically the complement of the previous indicator, *Number of Loyal Visitors*. Its computation is done according to the Equation 14.

$$\# \, of \, Non \, Loyal \, Visitors = |U| - \# \, of \, Loyal \, Visitor$$

Equation 14: # of Non Loyal Visitors Computation

(v) **Average Relative Frequency of All Visitors:** represents the average of the relative frequency distribution SRFD of all visitors *U* related to *p*. Its computation is done according to the Equation 15.

$$Avg \, \% \, Freq. of \, All \, Visitors = \frac{\sum_{n=1}^{|U|} SRFD\,(u_n, p)}{|U|}$$

Equation 15: Avg % Freq. of All Visitors Computation

(vi) **Average Relative Frequency of Loyal Visitors:** represents the average of the relative frequency distribution (SRFD) of all loyal visitors of *U* related to *p* given the activity category *C*. Its computation is done according to the Equation 16.

$$Avg\ \%\ Freq.of\ Loyal\ Visitors = \frac{\sum_{i=0}^{|U|} h(u_n, C, p)}{\sum_{i=0}^{|U|} f(u_n, C, p)}$$

Where,

$$h(u, C, p) = \begin{cases} SRFD(u, C, p), & SRFD(u, C, p) \geq Min_{RFLU} \\ 0, & SRFD(u, C, p) < Min_{RFLU} \end{cases}$$

$$\boldsymbol{f(u, C, p)} = \begin{cases} \mathbf{1}, & \boldsymbol{SRFD(u, C, p) \geq Min_{RFLU}} \\ \mathbf{0}, & \boldsymbol{SRFD(u, C, p) < Min_{RFLU}} \end{cases}$$

Equation 16: Avg % Freq. of Loyal Visitors Computation

(vii) **Average Relative Frequency of Non Loyal Visitors:** represents the average of the relative frequency distribution (SRFD) of all non-loyal visitors of *U* related to *p* given the activity category *C*. Its computation is done according to the Equation 17.

$$Avg\ \%\ Freq.of\ non\ Loyal\ Visitors = \frac{\sum_{i=0}^{|U|} w(u_n, C, p)}{\sum_{i=0}^{|U|} g(u_n, C, p)}$$

Where,

$$w(u, C, p) = \begin{cases} 0, & SRFD(u, C, p) \geq Min_{RFLU} \\ SRFD(u, C, p), & SRFD(u, C, p) < Min_{RFLU} \end{cases}$$

$$g(u, C, p) = \begin{cases} 0, & SRFD(u, C, p) \geq Min_{RFLU} \\ 1, & SRFD(u, C, p) < Min_{RFLU} \end{cases}$$

Equation 17: Avg % Freq. of non Loyal Visitors Computation

(viii) **Average Visits by Visitors:** represents the average of all visits performed by *U* to *p*. Its computation is done according to the Equation 18.

$$Avg\ Visits\ by\ Visitors = \frac{\sum_{n=1}^{|U|} SFD\ (u_n, p)}{|U|}$$

Equation 18: Avg Visitis by Visitors Computation

3.5.4. The User Interface

The user interface is implemented as a web application, where the user can interact with the map and visualize the information computed from the core component. This tool has been developed using the ASP MVC 4 framework[11]. This technology has a Model-View-Controller architecture that allows an easy separation between the data manipulation (server side) and the interaction of the user with the web application (client side). MAPMOLTY uses JQuery Mobile 1.3.2 with JQuery 1.9.1 [12] to implement the visual widgets used for the visualization for different types of web-browser devices. The web map widget also uses the JavaScript Google Maps API V3 [13] for the visualization and interaction with the map. The presentation web site of the tool is available at http://mapmolty.isti.cnr.it. At this website, it is possible to find screenshots and demo videos of the tool.

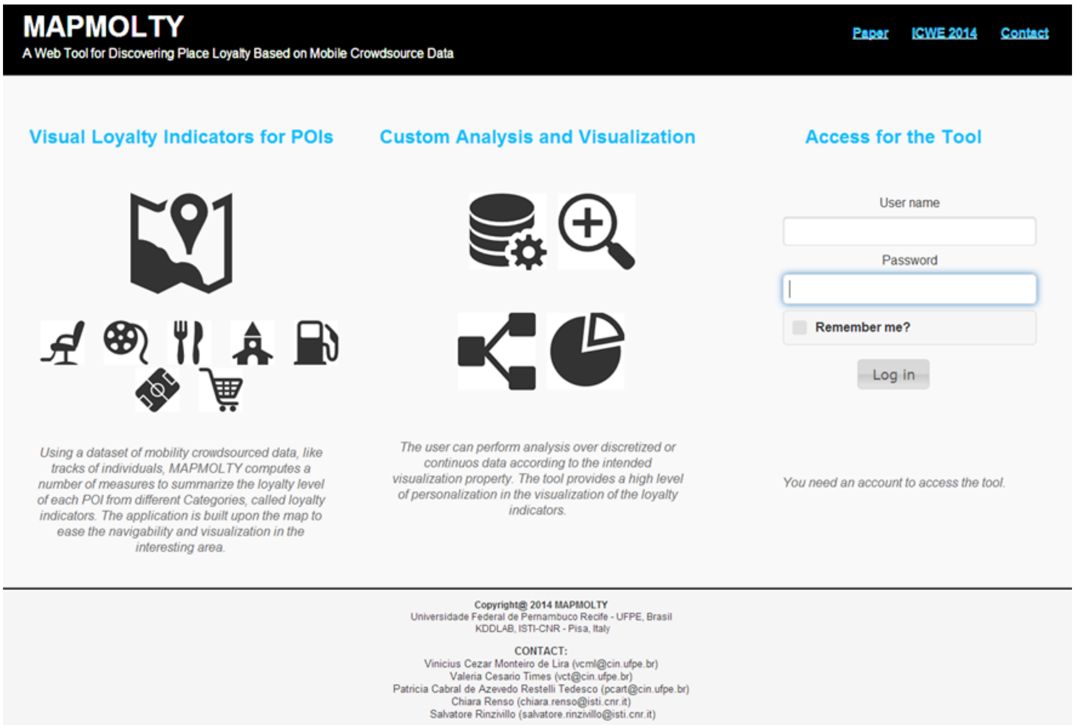


Figure 21: MAPMOLTY Home Page

3.5.4.1. Access to the Tool

To access MAPMOLTY it is necessary a user account. Each user keeps his configuration and his available datasets for analysis. The login is done providing a user name and a password as showed in Figure 22. After the login, the system proposes a list of Datasets from which to choose the area of interest as showed in the Figure 23.

[11] http://www.asp.net/mvc/mvc4

[12] http://jquerymobile.com/download/

[13] https://developers.google.com/maps/documentation/javascript/

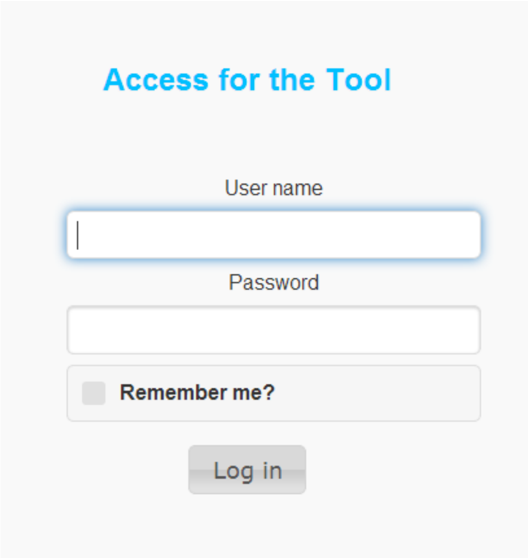


Figure 22: Login Form

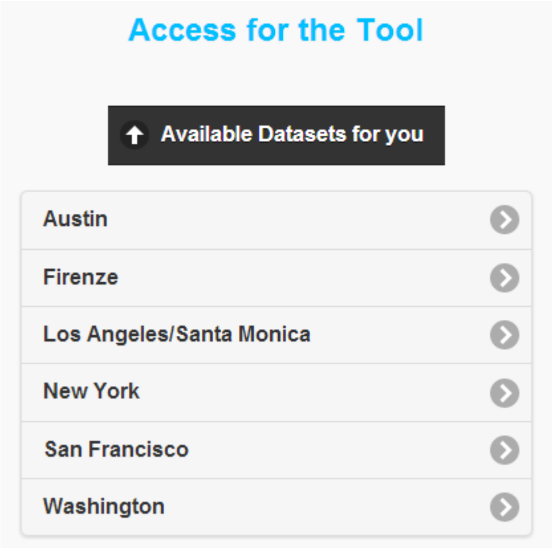


Figure 23: Available Dataset

3.5.4.2. Loyalty Analysis

Once the area of analysis has been selected, the user can select the place category. So far, the categories used in MAPMOLTY are: Bar, College, Entertainment, Gym, Mall, Medical Center, Multiplex, Religion, Restaurant, Square, Store, Theater and University. After selecting the super-category, the system shows a summary of the available indicators on a map. Each POI is indicated in the map by three visualization properties: marker color, circle size and circle opacity. Based on the loyalty indicators and loyalty configuration informed by the user, MAPMOLTY computes and plots the values of these visualization properties. Figure 24 shows the visualization of loyalty indicators for Restaurants in Florence.

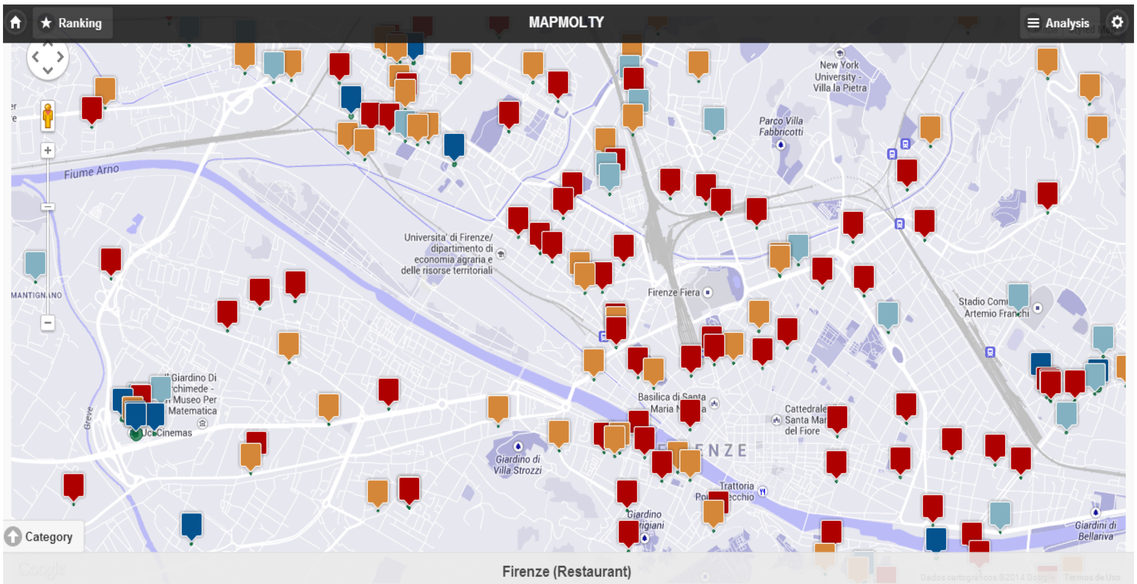
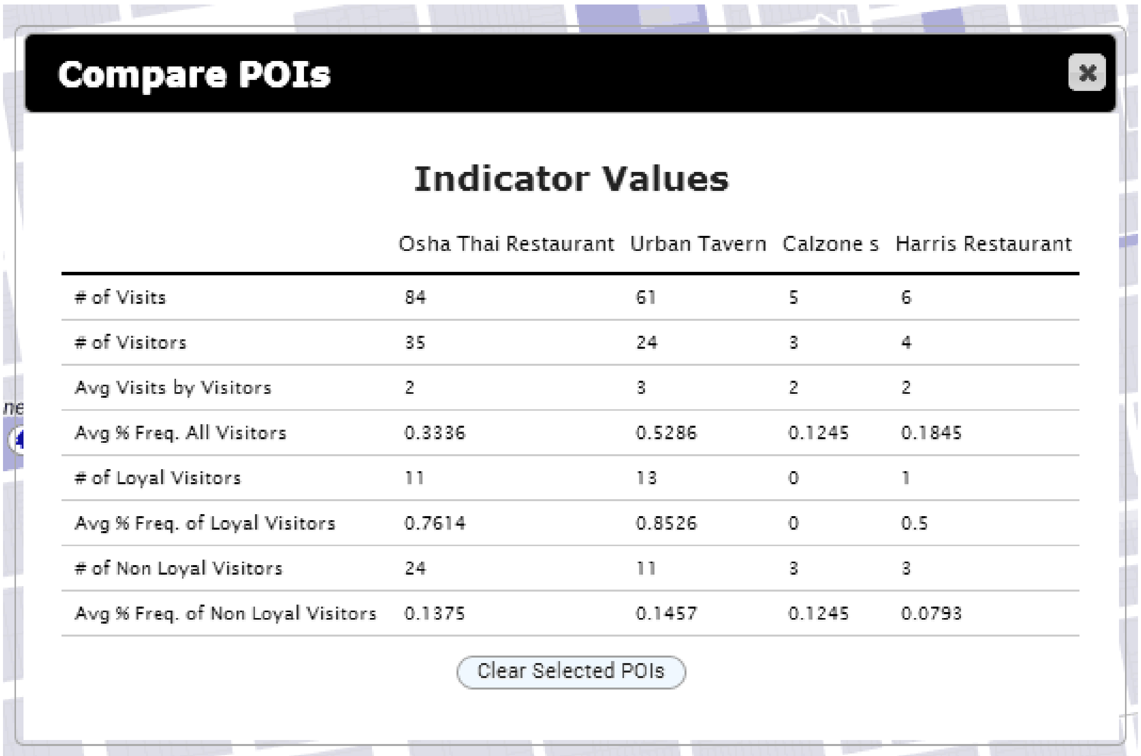


Figure 24: Loyalty Visualization for the Restaurant Category in Florence

Clicking on the marker the user can visualize the detailed information about the POI like the name, the category, sub-category, photo and all the indicators values as shown in Figure 25. Other interesting features provided by the web interface is the comparison between two places or more POIs according to their loyalty indicator values. This comparison can be visualized in a grid, as shown in Figure 26, or alternatively, the user can also select one of the indicator to plot the values in bar graph facilitating the analysis, as shown in Figure 27.

Figure 25: Detailed information about the POI



Compare POIs

Indicator Values

	Osha Thai Restaurant	Urban Tavern	Calzone s	Harris Restaurant
# of Visits	84	61	5	6
# of Visitors	35	24	3	4
Avg Visits by Visitors	2	3	2	2
Avg % Freq. All Visitors	0.3336	0.5286	0.1245	0.1845
# of Loyal Visitors	11	13	0	1
Avg % Freq. of Loyal Visitors	0.7614	0.8526	0	0.5
# of Non Loyal Visitors	24	11	3	3
Avg % Freq. of Non Loyal Visitors	0.1375	0.1457	0.1245	0.0793

Clear Selected POIs

Figure 26: Grid Compare POIs

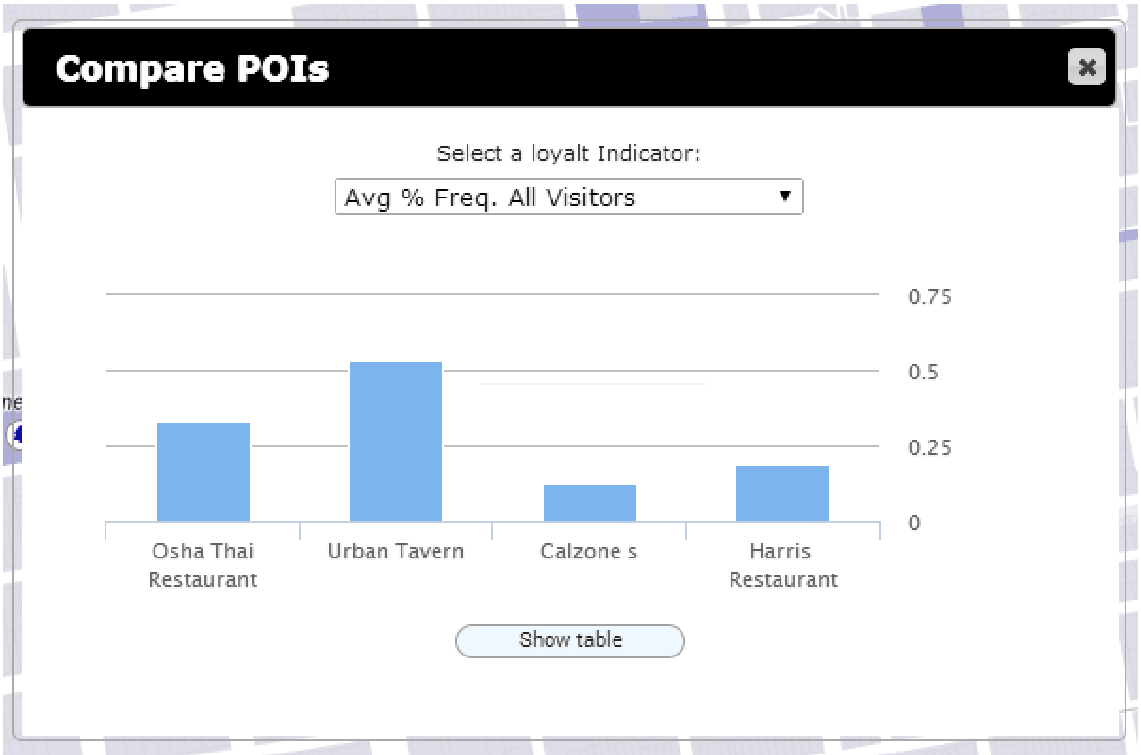


Figure 27: Bar Graph Compare POIs

MAPMOLTY also enables the visualization of the Top 15 POIs respect to the selected loyalty indicator. From the result list the user can click in one item to view details about the POI as illustrated in Figure 28.

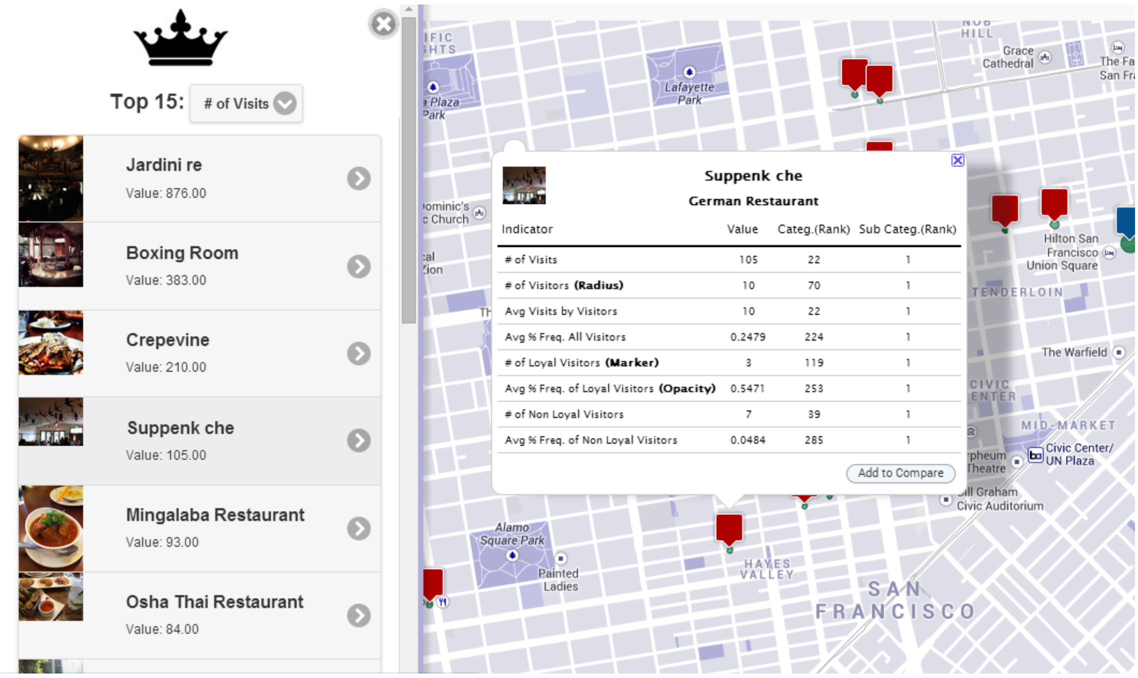


Figure 28: TOP 15 Visualization

MAPMOLTY also provides other functionalities, including:

Filter by Subcategory. The user can filter the POIs by selecting which POI's subcategories should be plotted on the map.

Clear the map. Using this functionality MAPMOLTY removes all the circles and markers on the map and closes all the opened windows.

Search for POI. Once the category for analysis is selected and the POIs are plotted on the map, it can be useful for the user to find a specific POI. MAPMOLTY facilitates the search of POI in the map providing an autocomplete text input.

3.6. Conclusion

We proposed a definition of semantic regularity profile, representing the spatial and temporal regularity measure of a user to perform some activity. We formally introduced a definition of semantic regularity measures and we corroborate our idea experimenting on two human mobility datasets. We discussed the results and noticed that the majority of tracked people tend to follow an expected trend.

The regularity profiles can be useful from several points of view. On the one hand they characterize one specific aspect of the user lifestyle, thus giving a quantitative measure of the people's regularity habits under observation. This can be useful in the recommendation and advertisement field, where new places to visit could be suggested to the regular users trying to encourage them to discover new places. Another application is transportation management: once known the semantic regularity profile of a user, it is possible to organize a carpooling system based on the flexibility of users to change their actual destination to perform a given activity, like changing the supermarket where to go shopping. On the other hand, regularity profiles computed from mobility data can be useful to characterize the geographical places as well. Using these information it is possible discover, for example, that a given restaurant is visited mostly by regular users while other restaurants are visited mostly by irregular users as it happens for the restaurants for tourists.

About visits to a place, regularity and loyalty are two terms that share the same interpretation. Based on it, a web tool called MAPMOLTY was developed giving a different interpretation of the computed data about the visiting behavior of users. This new interpretation has focus on the POIs. MAPMOLTY profiles the POIs of a given category providing loyalty information about their visitors. Three visualization properties related to the loyalty indicator values represent the POI on a map and detailed information is available for the user. The tool offers many analysis functionalities and a high level of customization features for the user. Two interesting features are the possibility of comparison of the loyalty indicator values between two or more different POIs of the same category and the Top 15 visualization according to the selected loyalty indicator.

With these information provided by the tool, it is possible to discover, for example, that a given restaurant is visited mostly by regular users while other restaurants are visited mostly by irregular users as it happens for the restaurants for tourists. A competitor's visualization is also possible from this kind of analysis about the POIs and their visitors' behavior. This analysis about the POIs and their users' behavior is also addressed in the present work.

Chapter 4: A Matching Method for Carpooling

This chapter presents a method to improve carpooling with the use of semantic information. Our objective is to boost the possibilities of carpooling proposing a carpool matching method that is oriented to the passenger's intended activity.

4.1. Introduction

The previous chapter showed that some people are more regular or more flexible about their visits to some types of places. The regularity analysis was conducted over two different perspectives: spatial and temporal. Therefore, each person has his/her own preferences about which POI (spatial perspective) and which time (temporal perspective) should be better to perform some intended activities. The regularity for a specific POI depends on many factors, such as the existence of contract with the POI, working relationship, proximity from home, convenience, pleasantness, effort to reach and so on.

However, we noticed that to perform some types of activities a person does not need to go always to a specific POI. In other words, it means that some human activities are not strictly associated with a unique POI. For example, if a person wants to have dinner at an Italian restaurant, the user has a wide amount of POIs to choose to perform this activity. Therefore, for some categories of POI, the user can change his destination POI and he can still perform the same activity desired. Another observation is that, during the day, some activities can be performed by a person in different intervals of the day. For example, in a category of POIs like gym, a user may not have a fixed interval of the day to train. In some situations, the user can freely change the time to go to a gym according to his obligations of the day.

Bringing to the Carpooling context, in some situations it is worth for a person to change his destination or the time at which an activity is to be performed if there is a ride possibility for him due all the benefits involved with the carpooling. From the knowledge about the user's intended activity, the carpooling approach can take advantage of it to boost the possibilities of rides. We present a novel matching method for carpooling that is oriented to the intended activity of the passenger.

This chapter is structured as follows. Section 4.2 shows the proposed method for carpool matching which is based on the passenger's intended activity. The method presents three matching algorithms that manipulate in different ways the spatial and temporal dimension. Section 4.3 presents an evaluation process that is used to investigate the efficiency of our method. In Section 4.4, an experiment based on real data of trajectories is presented and the results collected from this experiment are discussed. Section 4.5 displays the final considerations about this chapter.

4.2. A Novel Matching Method for Carpooling

In a Carpooling Context, the term *Matching* is commonly used to express that there is a possibility of carpooling between two or more people. In general, it means that the paths performed by these people have stretches of common interest, and therefore they can share a vehicle.

We have evidences that some trips are more related to activities executed with lower obligation about the destination place and the time (e.g. go to restaurants, go to bars, etc.). It means that to perform some kinds of activities (e.g. drink, eat, shop) a person does not need to be strict in his/her choices of place and time. For example, activities like shopping, training, watching movies in a cinema, eating out of home, and so on, are occasionally performed by a person during the week and there are a wide possibility of places and time to perform these activities.

Our carpool matching method is activity-oriented. This implies that our search method for carpool matching considers alternative destinations and schedules for the passenger to perform the intended activity. This reallocation of the destination is performed respecting the intended activity, therefore even when the passenger is reallocated he still can perform the wished activity at the new place suggested. For the schedule reallocation, it is proposed a time change for the person to perform his intended activity. This temporal reallocation is controlled by a parameter called time window that is defined according to the Definition 9.

Definition 9 (*Time Window (tw)*) *A time window corresponds to a time offset onwards or backwards of given time or time interval. It is represented by an integer number.*

Figure 29 shows a set of intervals sliced by time intervals of 1 hour. There is a time window of 2 hours of interval onwards or backwards of the current interval (22:00 – 23:00).

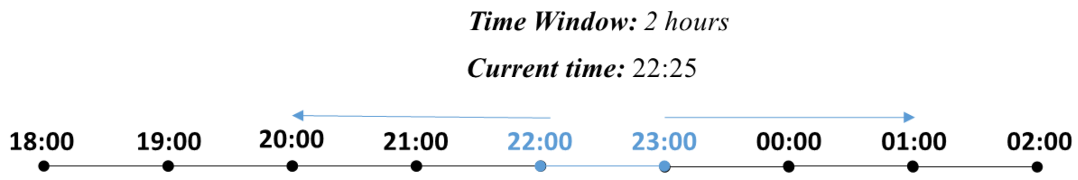


Figure 29: Time window example

In our method a **driver** is the person who has a car and can offer a ride. In turn, a **passenger** is the hitchhiker, in simple words, the person who is not the driver. We also recall that in this dissertation we associate the POI category to the activity performed in that kind of POI, thus we refer to the POI category or activity as synonyms.

To do so, we propose the method shown in Figure 30. In the next sections, the components of this method are detailed.

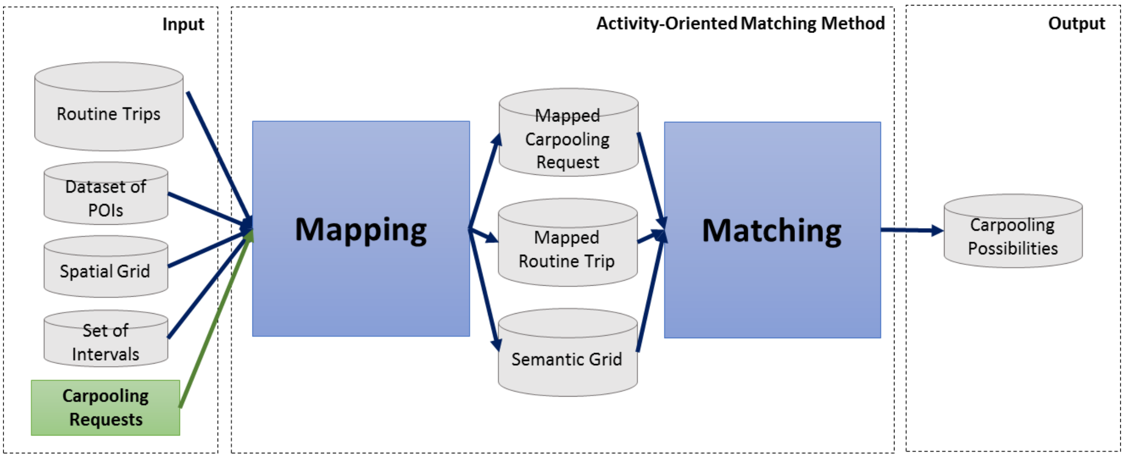


Figure 30: Activity-Oriented Matching Method for Carpooling

4.2.1. Data Input

The activity-oriented matching method for Carpooling that is proposed in this dissertation receives as input a dataset of POIs, a set of Intervals, a spatial grid, a set of carpooling requests and a dataset of routine trajectories. The **dataset of POIs** provides geographic information (latitude and longitude) and semantic information (POI Category and opening time). The semantic information provided are used for identifying which activity can be performed in that POI and when the POI is available for the users to perform the referred activity. The **set of Interval**, as defined in Definition 6 of Section 3.2, corresponds to slices of the day. The **Spatial Grid** overlaps and slices the geographic region of interest in cells. Each cell divides the grid into small pieces and is defined by a rectangle with 4 different geographic points. The Grid corresponds to the region of interest for the practice of carpooling and intends to facilitate the semantic analysis of the user’s interest for being in a specific geographic area. For carpooling purposes, the diagonal of the cell represents the maximum straight distance that a person should walk to start the ride (in case of the pickup cell) and/or walk to reach the desired place (in case of the arrival cell), after leaving the car. Figure 31 exemplifies a grid 24x30 for the city of Florence.

Figure 31: Example of Spatial Grid

For the two last remains input, **Carpooling Request** and **Routine Trip**, we have respectively the Definition 10 and Definition 11, which are given as follows.

Definition 10 (Carpooling Request (CR)) A carpooling request represents an intention of ride from a passenger. It expresses the user's necessity to move from a start point, to a destination place at a given time to perform some activity. We formally represent a Carpooling Request by the tuple: <*RequestID*, *DeparturePoint*, *DepartureTime*, *DestinationPOI*, *ArrivalTime*>, where *RequestID* represents the ID of the Request; the *DeparturePoint* and the *DepartureTime* represent respectively, the intended pick-up point for the passenger and the intended schedule to start the ride; the *ArrivalPOI* and the *ArrivalTime* represents the drop off point of the user and the intended schedule for it.

An example of a carpooling request is a request for a student that needs to move from his house (*DeparturePoint*) at 08:00 A.M. (*DepartureTime*), to arrive at his University (*DestinationPOI*) at 09:00 A.M (*ArrivalTime*). Thus a **set of carpooling request**s is nothing more than the collection of carpooling requests. The last input, the **Routine Trips**, we defined it according to the Definition 11, as mentioned before.

Definition 11 (Routine Trip (RT)) A Routine Trip (RT) corresponds to a trip that is usually made by the user. A RT represents a regular behavior about temporal and spatial aspects of a user to perform daily or almost daily a determined trip. We formally represent a Routine Trip by the nested tuple <IdRoutineTrip, <Points>$^+$>, where IdRoutineTrip represent the Id of the Trip and the <Points>$^+$ represent the sequence of points of the trip.

A common example of a routine trip is the daily path from home to a specific school performed by a person who brings his child to the school early morning. In our method, the routine trips are considered as fixed routes during the day that a passenger could take a ride. The user who performs a routine trip assumes a role of a potential driver. Therefore, the idea behind this identification is to use these routine trips as routes that are performed constantly and may serve as transportation to other people. In this work, we use the term ***route*** to represent also the routine trip of a user. The idea is similar to the route of a bus in which there is an itinerary defined for a certain interval of the day

4.2.2. Mapping Phase

The mapping phase of our method adapts the requests (CRs) and the routine trips (RTs) to fit in the algorithms of the next phase of the method, the matching. The RT and the CR are mapped into the spatial grid. This mapping aims to facilitate the matching between the CR and the RT. The mapping process performed is different for each type of trip (RT or CR) since they have different purposes.

For the mapping of a RT in a grid, we represent the whole path (the sequence of points) as a sequence of visited cells and their respective visited intervals. All the cells intersecting the trip are mapped and assigned with a sequential number. The cell containing the start point is the first cell of the sequence labeled with the number 1 as the first sequence number. The next covered cells by the trip have the sequence number defined by the previous cell's sequence number plus one unit. The sequence ends in the cell with the final point of the trip (the stop). Furthermore, each mapped cell is associated to one or more intervals of the day according to the input set of intervals. The association is done by looking for the timestamps of the trip's points located inside of a determined cell. Thus, we formally define a Mapped Routine Trip (MRT) according to Definition 12.

Definition 12 (Mapped Routine Trip (MRT)) A Mapped Routine Trip represents a composed tuple $< RTID, LstSeq* >$, where RTID is the RT unique identifier and *LstSeq* is the nested tuple $< Seqn, Cell, Intv >$ representing a sequence (*Seqn*) of cells (*Cell*) visited by the route in their respective time interval (*Intv*).

Figure 32 illustrates the mapping of a RT in a Grid3x3. This figure shows the composed tuple built from a RT mapped according to the set of intervals and the spatial grid given as an example.

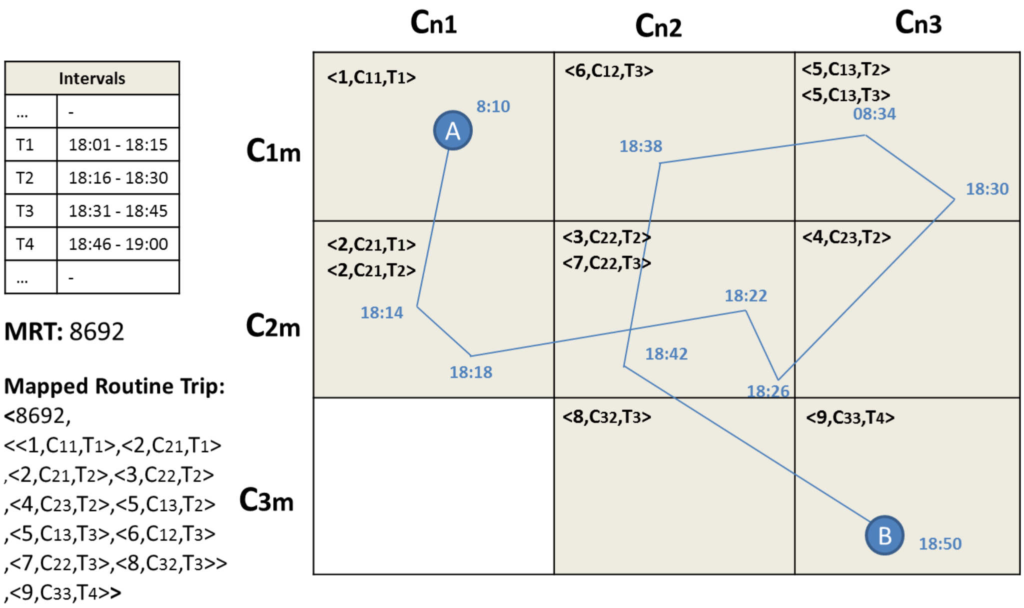


Figure 32: An Example of a Routine Trip Mapping

For the mapping of a carpooling request we basically map the departure point and the destination POI according to their positions on the spatial Grid. Thus, each point is related to one of the cells of the grid. Regarding the temporal properties, the departure schedule and the arrival schedule are each of them associated to its corresponding interval from the set of intervals.

Definition 13 (Mapped Carpooling Request (MCR)) A mapped carpooling request represents a simple tuple <*RequestId, CellDeparture, IntvDeparture, CellArrival, IntvArrival, CatgID*>, where RequestId is the CR unique identifier, *CellDeparture* and *IntvDeparture* represent respectively, the departure cell and departure interval of the CR, *CellArrival* and *IntvArrival* represent the arrival cell and arrival interval of the CR, *CatgID* denotes the category of the destination POI intended to be visited by the passenger.

Figure 33 shows an example of a mapping for a carpooling request to a Japanese Restaurant. In this example, the departure point and the arrival point of the request are mapped into the cell C_{22} and C_{33} respectively. As explained with the discussion about the MCR definition (Definition 13), the same association is done for the temporal aspects. Therefore, in this example, this mapped carpooling request represents the passenger's request for ride at the time interval T_2 in the cell C_{22} to go to the cell C_{33} arriving at the time interval T_4.

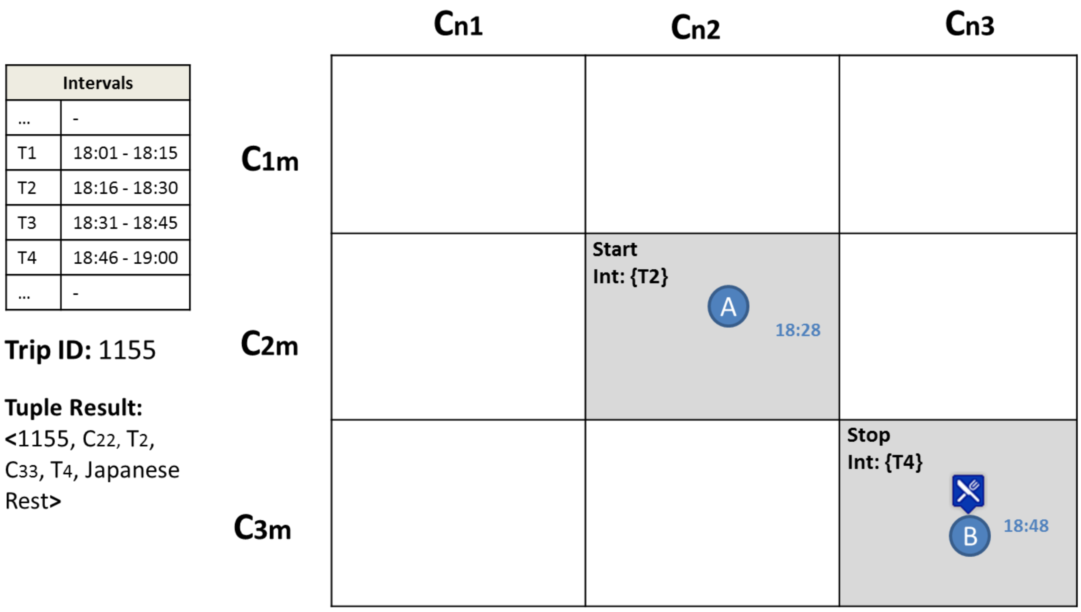


Figure 33: An Example of a Carpooling Request Mapping

The dataset of POIs corresponds to a collection of POIs and according to the Definition 4, each POI contains spatial (i.e. the geographic position) and temporal (i.e. the business hours) properties. Using the spatial properties, we map the POIs on the Grid to the corresponding cell. In turn, using the temporal properties and the set of intervals, we also map the available cell according to the business time of the place. Mapping the dataset of POIs to the Spatial Grid using these two properties gives us a semantic map about the availability of the POIs for a given category. We call this semantic visualization as Semantic Grid that is defined according to Definition 14.

Definition 14 (Semantic Grid (SG)) A Semantic Grid represents a tuple *<CategoryID,LstCellIntv∗>*, where *CategoryID* is the Category unique identifier and *LstCellIntv* is the nested tuple *<Cell,Intvopen,Intvclosed>* representing a cell (Cell) where is possible to find a POI of the given category (*CategoryID*) that is available during the interval of the day between *Intvopen* and *Intvclosed*.

An example of a Semantic Grid illustrating the category Japanese Restaurant is shown in Figure 34. We can observe that the cells containing at least one instance of the POI of the given activity category (i.e. Japanese Restaurant) and that is geographically positioned within them are marked (i.e. blue color) in the Grid as available cells (Figure

34(a)). Then, according to the temporal properties of the POI about the opening hours, the POI availability in a cell is also interpreted in terms of temporal aspects (Figure 34(b)).

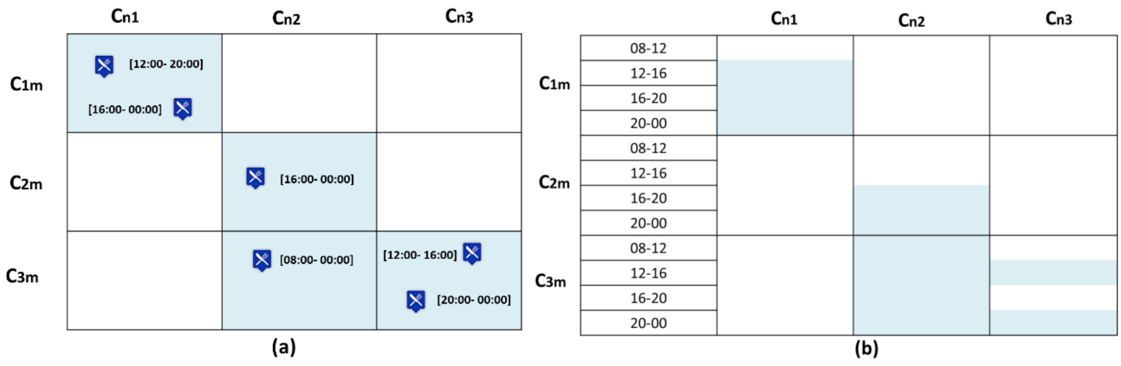


Figure 34: Semantic Grid Example illustrating the Category Japanese Restaurant

4.2.3. Matching Phase

We recall that our matching method focuses on the requested trips intending to supply them by means of a ride. To do so, the proposed method intends to match the CTs using the RTs as possibilities of ride.

To identify the possible trips that can be avoided, our matching method scans the set of those CRs (requests of rides) that can possibly be supplied using the RTs. We use three different algorithms to compute the matching between a RT and a CR: *Spatial Reallocation Matching*, *Temporal Reallocation Matching and Spatiotemporal Reallocation Matching*. These algorithms manipulate the spatial and temporal dimensions in a different way. The algorithms are basically composed by two steps.

The first step of the algorithms corresponds to checking the possibility of picking-up of the passenger through the function *GetPickupCell (params)*. The second step is responsible to find the possible destinations for the request using the function *GetDestinationCell (params)*. The functions of the first step (*GetPickupCell*) and of the second step (*GetDestinationCell*) receive input parameters that are used as filters. Depending on the algorithm chosen (*Spatial Reallocation Matching*, *Temporal Reallocation Matching or Spatiotemporal Reallocation Matching*), different input parameters are required.

The term *Reallocation* represents the possibility of making spatially and/or temporally changes in the original trip. The idea behind the use of these different searching algorithms is the possibility of analyzing separately the impact of reallocating trips according to different dimensions (spatial and/or temporal). These algorithms have some assumptions aiming to avoid small ride and noises: (1) the CR and the RT are performed by different users; (2) the *Cell Departure* and the *Cell Arrival* of the same CR are always different.

4.2.3.1 Spatial Reallocation Matching (SpatialRM)

The Spatial Reallocation Matching (*SpatialRM*) considers new possibilities of destination from the same category allowing the user to perform the activity desired. For example, given a CR (carpooling request) to a specific Italian restaurant and given a MRT (route), the *SpatialRM* tries to find destinations considering other available Italian restaurants in the route path.

Algorithm 1 presents the *Spatial Reallocation Matching*. This algorithm receives as input a MCR (θ), a MRT (γ) and a Semantic Grid (*SG*) and returns all the possibilities of carpool matching between the MCR and the MRT. The first step corresponds to the passenger's pickup. For the *SpatialRM*, the function *GetPickupCell* checks if the MRT contains the departure cell of the MCR representing a possibility of pickup for the passenger in his departure cell (*θ. CellDeparture*) within an interval between *θ. IntvDeparture* and *θ. IntvArrival* (lines 2-4). It means that the passenger would be picked-up in an interval between the departure and arrival intervals of the MCR without changing his schedule, and thus with no delays at his arrival.

Next (lines 6-13), for each pickup found in the first step, the function *GetDestinationCell*, using the *SG*, searches for destination cells from the MRT where the following conditions must be satisfied: (*i*) if the destination cell's sequence number is higher than the current pickup cell's sequence number (line 8); (*ii*) with an interval constraint between *θ. IntvDeparture and θ. IntvArrival* (line 9); (*iii*) if the MRT(γ) contains at least one opened POI of the requested category (*CatgID*) in the drop time interval considering the arrival cell of MCR(θ) (lines 10-13). The cells returned by the function *GetDestinationCell* for each possible pickup cell are seen as possibilities of carpooling and are returned as output of the algorithm (line 15 and line 18).

Algorithm: Spatial Reallocation Matching (*SpatialRM*)

Input: MCR (θ); MRT (γ); Semantic Grid (*SG*).

Output: *lstCarpooling <pickup,destination> (*List of carpooling possibilities*)*

```
lstCarpooling<pickup,destination> = Ø:
pickups = γ.LstSeq. GetPickupCell (
      Cell == θ.CellDeparture
      AND Intv BETWEEN θ.IntvDeparture AND θ.IntvArrival);
FOR EACH (p IN pickups) {
      cellsDest = γ.LstSeq. GetDestinationCell (
            Grid == Sg
            γ.Seqn >= p. Seqn
            AND γ.Intv BETWEEN θ.IntvDeparture AND θ.IntvArrival
            AND EXISTS (
            Cell IN (Sg.getCell(Sg.CatgID == θ.CatgID
              AND γ.Intv BETWEEN Sg.cell.Intvopen AND Sg.cell.Intvclosed))
            );
      FOR EACH (d in cellsDest) {
            lstCarpooling.add(p,d);
      }}
}
RETURN lstCarpooling;
```

Algorithm 1: Spatial Reallocation Matching

The Spatial Reallocation Matching algorithm guarantees that the passenger (user who made the request) can satisfy his request by accepting a ride to some cell that contain a POI of the same category (CatgID) to perform the same activity desired in an interval between the original departure interval (θ.IntvDeparture) and the original arrival interval (θ.IntvsArrival).

Figure 35 shows an example of a Spatial Reallocation Matching. The MCR 99155 could be avoided using the RTM 8692 and arriving at the cell C_{13}, at the interval T_3, or arriving at the cell C_{32} at the interval T_3. The Semantic Grid contains the information about the Japanese Restaurant like the cells and the intervals that is possible to find opened places.

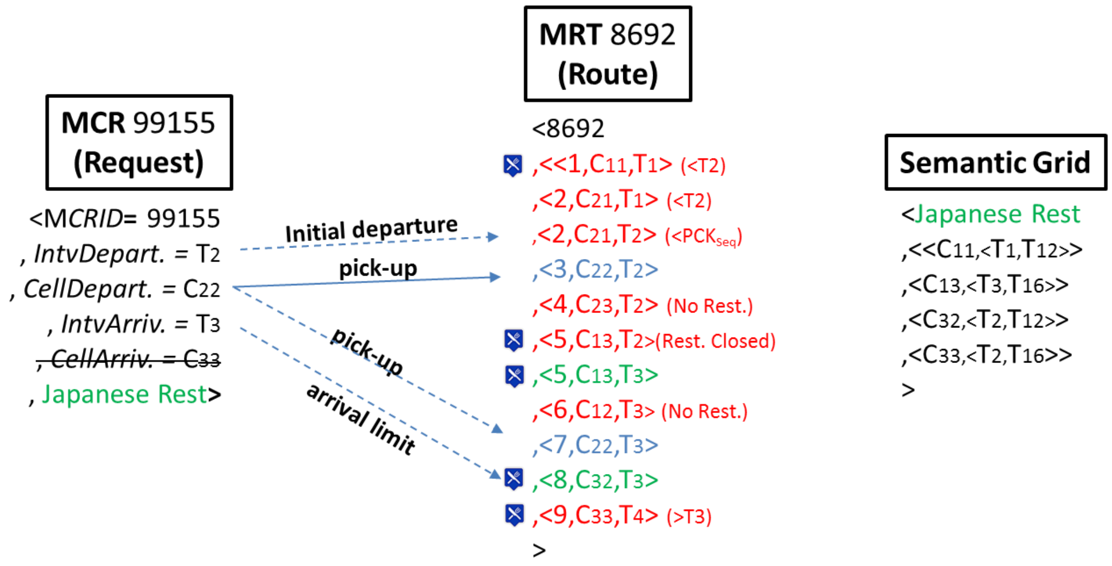


Figure 35: An Example of Spatial Reallocation

4.2.3.2 Temporal Reallocation Matching (TemporalRM)

Whereas the Temporal Reallocation Matching consists in an algorithm that finds routes (MRT) to satisfy the requests (MCRs) using a Time Window. This algorithm proposes a change in the passenger's schedule, but keeps the exactly intended POI destination. We propose that the user may be able to postpone or anticipate his trip to reach a specific POI. For example, a user could go to the gym in another interval of the day. The algorithm Temporal Reallocation Matching is presented as the Algorithm 2.

The Algorithm 2 receives the same input parameters of the Spatial Reallocation Matching plus the Time Window (*tw*) and returns the possibilities of carpool matching. The algorithm starts with the function *GetPickupCell* by checking if there is a pickup possibility according to the cell *θ.CellDeparture* and during the interval between (*θ.IntvDeparture*) minus the time window (*tw*) and *θ.IntvArrival* plus the time window (*tw*) (lines 2-4).

Then, the function *GetDestinationCell* searches for destinations, where the following conditions must be satisfied: (*i*) if the destination cell's sequence number is higher than the current pickup cell's sequence number (line 7); (*ii*) if the destination cell is the same as the intended cell (*θ.CellArrival*) (line 8); (*iii*) if the interval constraint between *θ.IntvDeparture* - *tw* and *θ.IntvArrival* + *tw* is satisfied (line 9); (*iv*) if contains at least one open POI of the requested category (CatgID) within the drop time interval and at the destination cell (*θ.CellArrival*) of MRT(γ) (lines 10-14). The cells returned by the function *GetDestinationCell* for each possible pickup cell are seen as possibilities of carpooling and are returned as output of the algorithm (line 14 and line 17).

Algorithm: Temporal Reallocation Matching (*TemporalRM*)

Input: MCR (θ); MRT (γ); Semantic Grid (*SG*); Time Window (t*w*).

Output: *lstCarpooling <pickup,destination> (*List of carpooling possibilities*)*

```
1   lstCarpooling<pickup,destination> = Ø:
2   pickups = γ.LstSeq. GetPickupCell (
3       Cell == θ.CellDeparture
4       AND Intv BETWEEN θ.IntvDeparture - wi AND θ.IntvArrival + wi);
5   FOR EACH (p IN pickups) {
6       cellsDest = γ.LstSeq. GetDestinationCell (
7           γ .Seqn >= p. Seqn
8           AND γ.Cell == θ.CellArrival
9           AND γ.Intv BETWEEN θ.IntvDepart. – wi AND θ.IntvArriv. + wi
10          AND EXISTS (
11          Cell IN (Sg.getCell(Sg.CatgID == θ.CatgID
12            AND γ.Intv BETWEEN Sg.Cell.Intvopen AND Sg.Cell.Intvclosed)
13            AND Sg.Cell == θ.CellArrival)
14          );
15      FOR EACH (d in cellsDest) {
16          lstCarpooling.add(p,d);
17      }}
18  }
19  RETURN lstCarpooling;
```

Algorithm 2: Temporal Reallocation Matching

The Temporal Reallocation Matching algorithm only guarantees that, accepting a ride, the passenger arrives at the same cell destination that was the original and desired destination (*θ.CellArrival*), but the arrival and departure schedule can be postponed or made in advance according to the time window (*tw*).

Figure 7 shows an example of use of the Temporal Reallocation Matching. As in the previous figure (Figure 35), the semantic Grid contains the information about the cells and intervals that is possible to find opened Japanese Restaurant. In this example, the algorithm returns two possibilities to avoid the CR 99155 from the RT 8692 using the cell C_{22} as pick-up in two different intervals (T_2 or T_3) and arriving at the intended destination cell C_{33}, but delaying the arrival interval from T_3 to T_4.

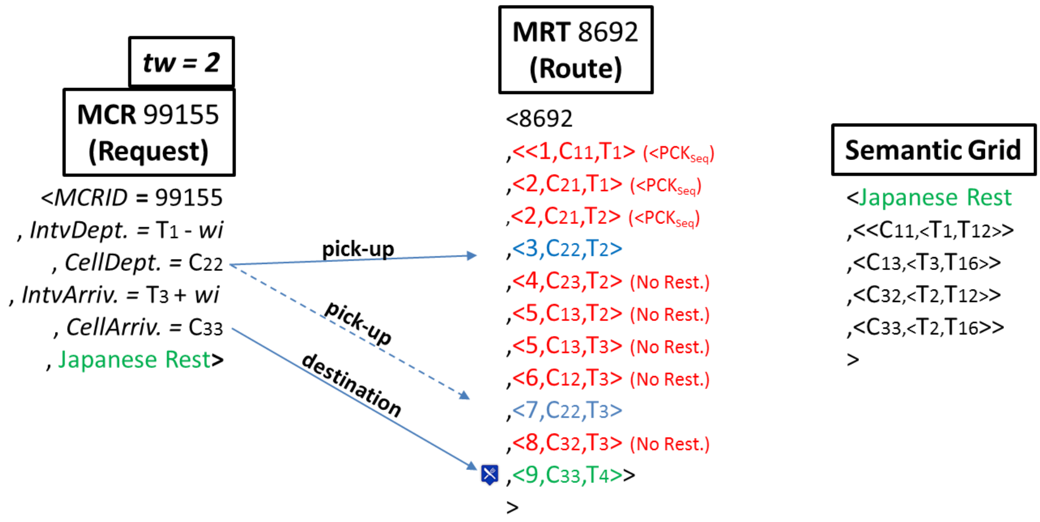


Figure 36: An Example of Temporal Reallocation

4.2.3.3 Spatiotemporal Reallocation Matching (SpotempRM)

Basically, it is the combination of the Spatial Reallocation and Temporal Reallocation algorithms. The Spatiotemporal Reallocation Matching searches for carpooling possibilities considering the possibility of the user to change the place to perform his desired activity and to change the interval of the day to perform this activity respecting a time window (*tw*). For example, given a CR (request) for a specific market, the proposed matching tries to find possible routes to go to other reachable markets in different intervals of the day. Thus, we suppose that the user can still perform the activity of *shopping*, in many places and at a different time interval. The algorithm is less restrictive and is described as the *Algorithm 3* as follows.

First, the function *GetPickupCell* finds the possible pickup cells from the *RTM* within an interval between θ*.IntvDeparture - tw* and θ*.IntvArrival + tw*. Next, the function *GetDestinationCell* searches for all available destination cells in the *Semantic Grid (SG)* of the category (θ*.CatgID*) visited by the RTM during the interval between θ*.IntvDeparture - tw* and θ*.IntvArrival + tw*. The Spatiotemporal Reallocation Matching only guarantees that the passenger can accept a ride to reach a POI of the same category to perform the desired activity satisfying the *MCR*.

Algorithm: Spatiotemporal Reallocation Matching

Input: MCR (θ), MRT (γ), Time Window (t*w*), Semantic Grid (*SG*)

Output: True, if there is matching; False, if there is not.

```
lstCarpooling<pickup,destination> = Ø;
pickups = γ.LstSeq.GetPickupCell (
Cell == θ.CellDeparture
AND Intv BETWEEN θ.IntvDeparture - tw AND θ.IntvArrival + tw);
FOR EACH (p in pickups) {
      cellsDest = γ.LstSeq. GetDestinationCell (
      Grid == Sg
      AND CatgID == θ.CatgID
      AND Intv BETWEEN θ.IntvDeparture - wi and θ.IntvArrival + wi
      AND EXISTS (
            Cell IN (Sg.getCell(Sg.CatgID = θ.CatgID
               AND γ.Intv BETWEEN Sg.cell.Intvopen ANDSg.cell. Intvclosed))
      );
      FOR EACH (d IN cellsDest) {
            lstCarpooling.add(p,d);
      }}
}
RETURN lstCarpooling;
```

Algorithm 3: Spatiotemporal Reallocation Matching

Figure 37 shows an example of applying the Spatiotemporal Reallocation for the MRT 8692 and the MCR 99155. As a result, the algorithm returns true because there are four possibilities for the passenger to take a ride using the route 8692: departing from the cell C_{22} in the interval T_2 and: arriving at the cell C_{13} in the interval T_3 (Possibility 1); arriving at the cell C_{32} in the interval T_3 (Possibility 2); arriving at the cell C_{33} in the interval T_4 (Possibility 3). There are also the possibilities of departing from the cell C_{22}

in the interval T_3 and arriving at the cell C_{32} in T_3 (Possibility 4); and arriving at the cell C_{33} in T_4 (Possibility 5) as well.

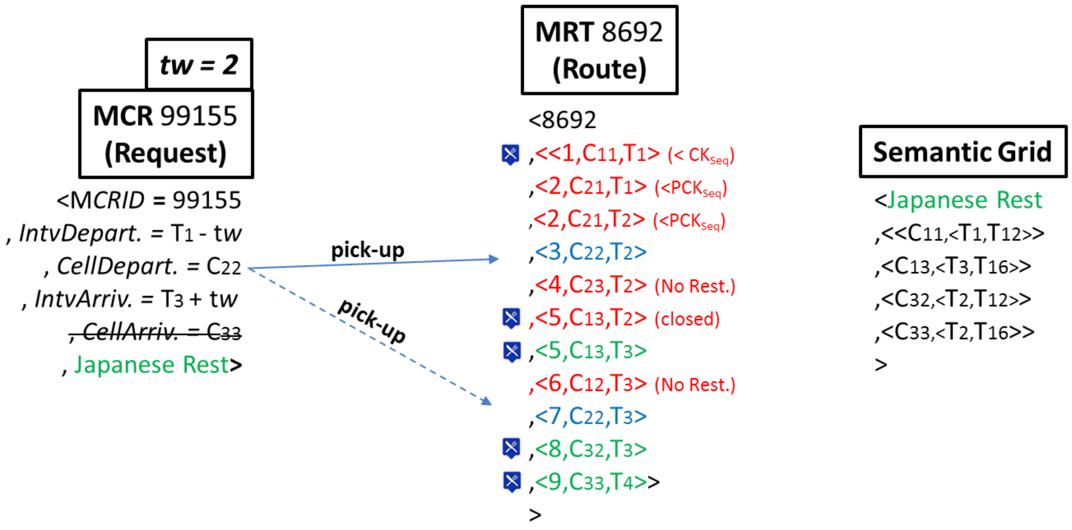


Figure 37: An Example of Spatiotemporal Reallocation

4.3. Evaluation Process

Intending to corroborate our idea, we propose an evaluation process to assess the efficacy of our method. The intention is to evaluate the amount of trips that could be *"potentially avoided"* from a given dataset of trajectories using our proposed method for carpool matching. Since the acceptance of the ride by the passenger is subjective (the ride can be accepted or not according to the passenger's convenience), the term *"potentially avoided"* is used to indicate that for one given trip there is at least one possibility of carpooling that could prevent its execution.

We focus on the avoidance of trips not commonly performed, which have a characteristic of being more flexible about the place and the time to perform the intended activity (e.g. Shopping, eating, watching cinema, play soccer, etc.). We are not aimed at evaluating our method by avoiding routine trips, since these trips represent a strict relation between the person and the destination POI (e.g. trips to home, trips to work). Therefore, there is no sense in using our matching method to return results with reallocation of destinations to places like home and work.

Figure 38 shows the whole evaluation process. The evaluation process is composed by four modules: Input, Pre-Processing, Matching and Output. Each one of these modules is described in the following sections.

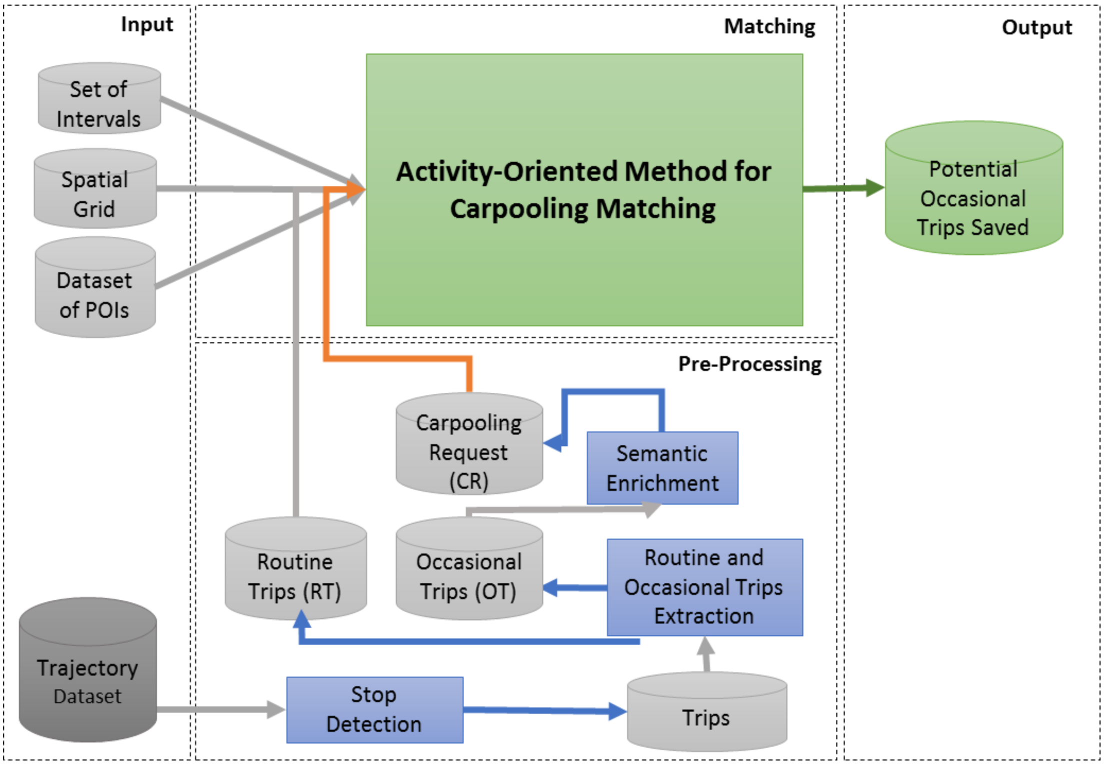


Figure 38: Evaluation Process

4.3.1. Data Input of the Evaluation Process

The data acquisition of the evaluation process is composed basically by the following data inputs that are required by our matching method: a Set of Intervals, a Spatial Grid and a Dataset of POIs. Furthermore, it receives a **dataset of Trajectories** containing a historical set of raw trajectories represented as a sequence of points that indicate the spatiotemporal path of a person.

4.3.2. Pre-Processing Phase

The first step in the Pre-Processing is the stop detections aiming to segment the data set of trajectories into trips. Using two parameters: δ, a spatial tolerance threshold and τ, a temporal tolerance threshold, we determine the stops from the trajectories. Once we have the stops, by the Definition 3, we also have the trips.

The second step aims to separate the routine and occasional trips. We use the method described in Section 2.7 for the extraction of a set of the user's routine trips from a set of trajectories. The extraction is made clustering similar trajectories and discarding outliers. From the group of similar trajectories, the medoid trajectory is chosen to represent a group of routine trajectories. Once we detect the clusters of routine trips and their respective representative, we classify all other trips that do not belong to any similarity cluster as occasional trips. The routine trip was already defined according to Definition 11, while the occasional trips are defined according to Definition 15.

Definition 15 (Occasional Trip (OT)) A Occasional Trip (OT) is a not Routine Trip (RT). Therefore, an OT is a trip not often performed by the user that has an eventual characteristic.

We recall that the intention of this evaluation process is to apply our carpool matching method to compute the amount of occasional trips that may be potentially avoided. For this, we adapt each occasional trip to a carpooling request (Definition 10). To adapt the occasional trip to our carpooling request, it is required a *semantic enrichment* aiming to understand the intended activity of the trip. This enrichment is done using the dataset of POIs. In simple way, for each end point of the trip, we can associate it to a specific POI. At the end of this step, each occasional trip has a semantic destination. Thus, the occasional trips are associated, for example, to electronic stores, Japanese restaurants, cinemas and so on.

We convert a semantically enriched OT into a CR keeping the information about the departure point, the time departure, and also the information about the POI destination and the arrival time. Figure 39 shows the carpooling request 99155 (CR) built from the occasional trip (OT) 99155.

It is important to notice that the methods: "*Stop Detection*", "*Routine and Occasional Trip Extraction*" and *"Semantic Enrichment"* work like black boxes in our evaluation process where, depending of the application, different implementations of these methods can be applied.

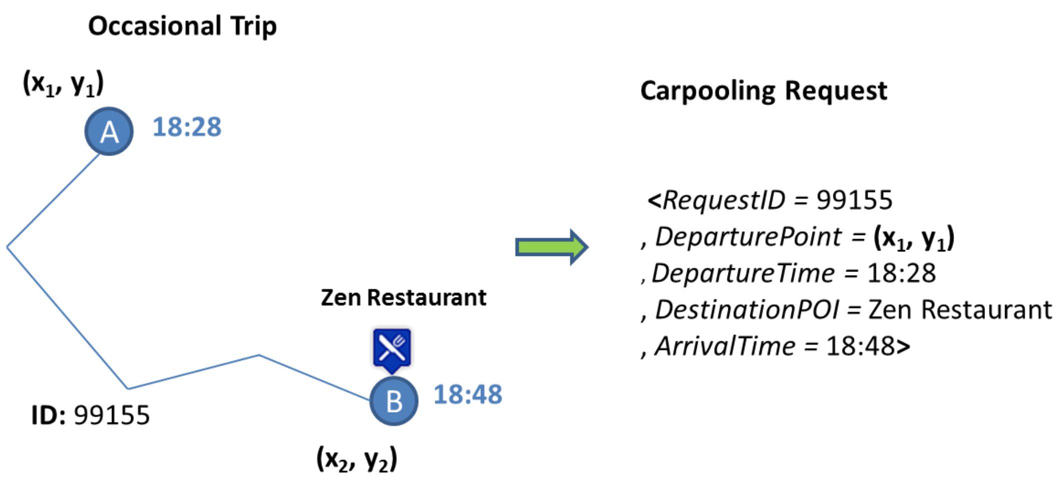


Figure 39: Example of an Occasional Trip converted into a Mapped Carpooling Request

4.3.3. *Application of our Matching Method*

During the pre-processing the given dataset of trajectories was divided into two different groups: routine trips and occasional trips. We recall that, in our method, the routine trips are considered as fixed routes during the day that a passenger could take a ride. On the other hand, the occasional trips correspond to trips not daily performed, i.e. trips performed not often by a person. Thus, the potential passengers of the occasional trips could benefit from the routine trips to hitchhike and then go to their intended destinations. In our evaluation process, the occasional trips are seen as carpooling requests.

Therefore, given the group of routines trips (RT), the set of carpooling requests (CR, originated from the occasional trips) and the others required inputs (Spatial Grid, dataset of POIs and Set of Interval), the evaluation process uses the activity-oriented matching method for carpooling to identify carpooling requests (CR) that could be satisfied using our search method.

4.3.4. *Evaluation Process Output*

In the evaluation process, an OT is converted into a CR to be adapted in our proposed method. Thus, from the relation between the occasional trips (OTs) and the carpooling requests (CRs), if an OT can be avoided by some RT, it also means that the corresponding CR can also be supplied and vice-versa.

As evaluation output of the evaluation process, we relate the carpooling request (CR) that were satisfied by our method, with the original occasional trip (OT). Thus, we have the total amount of occasional trips that could be potentially avoided.

We used as evaluation metrics the percentage of occasional trips (OT) that could be potentially avoided if our carpool matching was applied. In other words, it means the percentage of carpooling requests that could be satisfied applying our method.

For a given function ϕ representing one of the carpool matching algorithm (Traditional Matching (Baseline), Spatial Reallocation, Temporal Reallocation, Spatiotemporal Reallocation), we computed from a Set of OTs (Θ) the percentage of OTs that could be potentially avoided. We performed this evaluation considering different values for the parameters: cell size α and the time window size *i*. Therefore, we defined the Equation 19 to compute the metric used in our tests for evaluating the aforementioned algorithms:

$$Pot^{OTAvd}(\phi, \alpha, i) = \sum_{t=1}^{|\Theta|} \phi(\Theta n, \alpha, i) \Big/ |\Theta|$$

where,

$$\phi(\Theta n, \alpha, i) \begin{cases} \mathbf{1}, if\ there\ is\ matching \\ \mathbf{0}, if\ there\ is\ not\ matching \end{cases}$$

Equation 19: Percentage of occasional trips (OT) that could be potentially avoided

4.4. Experiments

4.4.1. Baseline: Traditional Matching (BaselineM)

We conducted tests to compare our activity-oriented matching method with a traditional matching for carpooling that follows the traditional carpooling purpose, which is pick up and drop off the user in the arranged places and times. This traditional matching is seen as our baseline and is detailed as follows.

Without the use of semantics the baseline algorithm just keeps the focus on the spatial and temporal aspects. The baseline strictly checks the match between a MRT (γ) and a MCR (θ). In this scenario, the MCR is also interpreted as a request of ride, but in this case, without the activity property being assigned to it. For the spatial case, the idea is that, respectively, the passenger is picked-up and left by a MRT in the same departure and arrival cells of the carpooling request (MCR). Similarly for the temporal case, the passenger is picked-up and left in an interval between the departure interval and the arrival interval of the request.

In the temporal case, we suppose that the user is able to start the ride in a schedule that is equal to or later than the departure interval, but before the arrival interval of the request (MCR). Regarding the arrival interval, we consider that the user is able to arrive to his destination cell in the arrival interval of the request (MCR) or before this time interval. The baseline algorithm is also composed by three validations as follows: pick-up validation, destination validation and direction validation. The baseline algorithm receives as input parameters: a MRT and a MCR; and returns a Boolean value as result, which is true if there is a match, and false otherwise. The traditional matching algorithm is shown as Algorithm 4:

Algorithm: Traditional Matching (BaselineM)

Input: MCR (θ), MRT (γ)

Return: True, if there is matching; False, if there is not.

```
BEGIN:
pickups = γ.LstSeq.Get(
    Cell == θ.CellDeparture
    AND Intv BETWEEN θ.IntvDeparture AND θ.IntvArrival
)
if exists (pickups) then
     cellsDest = γ.LstSeq.Get(
     Cell == θ.CellArrival
     AND Intv BETWEEN θ.IntvDeparture AND θ.IntvArrival
     )
     for each d in cellsDest
       if (d.Seqn > Min(pickups.Seqn)) then
           RETURN true;
       end if;
     end for;
end if;
RETURN false;
END:
```

Algorithm 4: Traditional Matching (BaselineM)/Baseline

4.4.2. Inputs of the Experiment

Trajectory dataset. The experiments were performed on a dataset of real trajectories collected in the city of Florence. The database corresponds to a set of data with 44.278 trips, made by 5.048 users moving in the city of Florence during the weekdays for the period from 01/05/2011 to 31/05/2011. The stops trajectories were detected using a spatiotemporal threshold of 50 meters and 20 minutes. After the pre-processing phase using the methodology of Trasarti (TRASARTI et al, 2011), a total of 36.593 non routine trips were classified as occasional trips (OT). We recall that such selected trips did not belong to any cluster of similar trajectories that were extracted for composing the routine trips (RT). The remaining amount of trips, 7.685 trips, composes the similarity clusters for the extraction of the RT. From these 7.685 similar trips, a total of 1.122 routes (RT) were extracted to summarize these similar trips.

POI's dataset. To provide the information about the place inside the Spatial Grid and also to enrich semantically the trips, we used the API from Foursquare. Regarding the unknown business hours we set some default values according to the category. For example, when not provided the business hour for a restaurant, we considered the open interval as [10:00-00:00] and we considered some of them as always open for public places like square, hospital and airport.

Set of intervals and **Time Window**. The set of Intervals is represented by the time slices. In our experiments, we used a set of intervals with fixed size of 30 minutes. For the time window, we performed the experiment with four different sizes of time window (i): 30 minutes, 1 hour, 2 hours and 4 hours.

Spatial Grid. The Spatial Grid covers an area of 173.49 km of Florence with of 11,790 km height and 14,720 km of width. The four vertices of the grid were selected considering that they are the most extreme stops of the dataset of trajectories. Thus, all the stops could be mapped into the cells. In our experiments we used 3 different combinations of size for the cells (α): 250 meters, 500 meters and 1 kilometer. We recall that the diagonal of the cell represents the maximum straight distance that a person should walk to start the ride, in the case of the pickup cell, and/or to reach the desired place, in the case of the arrival cell. Thus, how smaller the cell, smaller is the distance that the user should walk.

4.4.3. Evaluation

As detailed in the Table 1, we conducted the experiments for the follow configurations of i, ϕ and α.

Configuration	**Values**
Time window size (i)	0.5 hour, 1 hour, 2 hours and 4 hours.
Cell size (α)	0.25 kilometer, 0.5 kilometer and 1 kilometer.
Matching Algorithm (ϕ)	BaselineM, TemporalRM, SpatialRM and SpotempRM.

Table 1: Experiment Configuration

Table 2 reports the main results collected from our experiments. When compared to the baseline *BaselineM*, our results indicated that the proposed algorithms using reallocation *(TemporalRM, SpatialRM, SpotempRM)* substantially increase the possibilities of carpooling for the tested configurations of α and *i*.

For all the algorithms tested, the potential of avoided trips increases when the values of α and *i* are increased accordingly, reaching the maximum percentage of potentially avoided trips when these values are the largest. This is understandable since how larger the values of the parameters α and *i* are, less restricted about spatial and temporal aspects the algorithms that find a matching and supply the OTs are. Increasing the parameter *i*, conceptually, increases the tolerance of the passenger to change his schedule to perform his activity. Also, the more the value of the parameter α is increased, the higher the user's tolerance is to walk. Also, the smaller the number of grid cells is, the empty cells tend to be grouped with populated cells of the intended place category. It is important to notice that both *BaselineM* and *SpotempRM* do not depend on the size of the interval *i*. This is because their algorithms consider as starting time, the original time and the stop time of OT, without using the window interval.

As expected, for all investigated test configurations, the algorithm *SpotempRM* produced the best performance gains, indicating that in almost all the cases, if the passengers have the flexibility to change the place to go and the time to go, then *SpotempRM* provides the higher probability of finding an available ride to reach their destinations and perform their intended activities.

	ϕ = **BaselineM**	ϕ = **TemporalRM**	ϕ = **SpatialRM**	ϕ = **SpotempRM**
(α = 0.25km, i = 0.5$hour$)	4.10%	8.95%	28.16%	40.68%
(α = 0.25km, i = 1$hour$)	4.10%	12.96%	28.16%	51.18%
(α = 0.25km, i = 2$hour$)	4.10%	19.36%	28.16%	64.31%
(α = 0.25km, i = 4$hour$)	4.10%	27.75%	28.16%	77.28%
(α = 0.5km, i = 0.5$hour$)	9.39%	18.91%	47.14%	62.34%
(α = 0.5km, i = 1$hour$)	9.39%	25.40%	47.14%	72.40%
(α = 0.5km, i = 2$hour$)	9.39%	35.25%	47.14%	83.03%
(α = 0.5km, i = 4$hour$)	9.39%	46.73%	47.14%	91.70%
(α = 1.0km, i = 0.5$hour$)	18.85%	34.26%	65.69%	78.87%
(α = 1.0km, i = 1$hour$)	18.85%	43.94%	65.69%	86.43%
(α = 1.0km, i = 2$hour$)	18.85%	56.53%	65.69%	92.89%
(α = 1.0km, i = 4$hour$)	18.85%	69.74%	65.69%	97.50%

Table 2: Percentage of OTs Potentially Avoided PotOTAvd for each algorithm ϕ size of the cell α and time window i

Figure 40 shows the line graph of the percentage of OTs potentially avoided (Pot^{OTAvd}) for each test configuration. This graph also shows the growth of the measure Pot^{OTAvd} according to the increase in the values of the time window i.

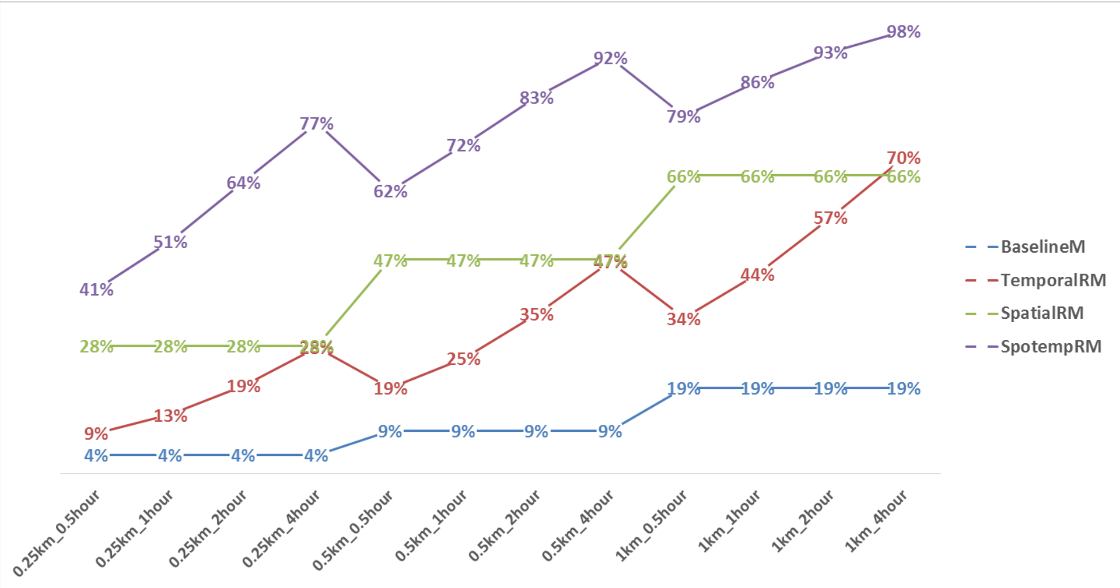


Figure 40: Percentage of OTs Potentially Avoided (Pot^{OTAvd})

The gains produced by each matching algorithms (ϕ) when compared to the baseline for the tested configurations of α and i is calculated according to Equation 20.

$$Gain(\phi) = PotOTAvd(\phi, \alpha, i) - PotOTAvd(BaselineM, \alpha, i)$$

Equation 20: Computation of the gain compared to the baseline

Table 3 shows the gains of each algorithm (ϕ) when compared to the baseline. As expected the *SpotempRM* has the best gains when compared to the other matching algorithms (*TemporalRM and SpatialRM*) for all test configurations.

	ϕ = **TemporalRM**	ϕ = **SpatialRM**	ϕ = **SpotempRM**
(α = 0.25*km*, *i* = 0.5*hour*)	+4.85%	+24.05%	+36.57%
(α = 0.25*km*, *i* = 1*hour*)	+8.86%	+24.05%	+47.08%
(α = 0.25*km*, *i* = 2*hour*)	+15.26%	+24.05%	+60.20%
(α = 0.25*km*, *i* = 4*hour*)	+23.65%	+24.05%	+73.18%
(α = 0.5*km*, *i* = 0.5*hour*)	+9.52%	+37.75%	+52.95%
(α = 0.5*km*, *i* = 1*hour*)	+16.01%	+37.75%	+63.01%
(α = 0.5*km*, *i* = 2*hour*)	+25.86%	+37.75%	+73.64%
(α = 0.5*km*, *i* = 4*hour*)	+37.34%	+37.75%	+82.30%
(α = 1.0*km*, *i* = 0.5*hour*)	+15.41%	+46.84%	+60.02%
(α = 1.0*km*, *i* = 1*hour*)	+25.09%	+46.84%	+67.58%
(α = 1.0*km*, *i* = 2*hour*)	+37.68%	+46.84%	+74.04%
(α = 1.0*km*, *i* = 4*hour*)	+50.89%	+46.84%	+78.65%

Table 3: Gains of the matching algorithms when compared to the baseline

For a more applied practical visualization of the results of our experiments, Table 4 shows the amount in kilometers of occasional trips potentially avoided. As expected the values for our proposed algorithms (*TemporalRM, SpatialRM and SpotempRM*) reached values much higher than the baseline approach (*BaselineM*). These values represent a positive impact for the traffic and the environment. If we consider that using 1 liter of gasoline, a normal car performs 10 km, thus for the *SpotempRM*, the lowest result (50.203 km) could save 5.020 liters of gasoline, while the highest (106.786 km), 10.678 liters. These savings of gas represents less pollutants for the atmosphere and also less cost for the drivers, since they can share the cost.

	ϕ = **BaselineM**	ϕ = **TemporalRM**	ϕ = **SpatialRM**	ϕ = **SpotempRM**
(α = 0.25km, i = 0.5$hour$)	5.333 km	8.956 km	39.482 km	50.203 km
(α = 0.25km, i = 1$hour$)	5.333 km	11.657 km	39.482 km	61.101 km
(α = 0.25km, i = 2$hour$)	5.333 km	16.986 km	39.482 km	75.998 km
(α = 0.25km, i = 4$hour$)	5.333 km	24.014 km	39.482 km	91.137 km
(α = 0.5km, i = 0.5$hour$)	10.260 km	16.532 km	59.411 km	72.281 km
(α = 0.5km, i = 1$hour$)	10.260 km	21.436 km	59.411 km	82.716 km
(α = 0.5km, i = 2$hour$)	10.260 km	29.340 km	59.411 km	94.634 km
(α = 0.5km, i = 4$hour$)	10.260 km	40.193 km	59.411 km	105.061 km
(α = 1.0km, i = 0.5$hour$)	18.030 km	29.360 km	75.800 km	86.727 km
(α = 1.0km, i = 1$hour$)	18.030 km	37.834 km	75.800 km	94.353 km
(α = 1.0km, i = 2$hour$)	18.030 km	50.191 km	75.800 km	101.248 km
(α = 1.0km, i = 4$hour$)	18.030 km	64.537 km	75.800 km	106.786 km

Table 4: Kilometers of the OTs Potentially Avoided

Figure 41 and Figure 42 show two different categories of OTs that could be potentially avoided considering different values of α and *i*. Figure 41 displays the OTs whose destination is a Golf Course and passenger's intended activity is "*Play Golf*". We observed that the *SpatialRM* and *SpotempRM* presented low values of potentially avoided OTs with *α* equals to *0.25km, 0.5km and 1km*. The main reasons for these low values are the availability of these kinds of places. Since in our dataset of Florence's POIs there are only 4 places where the user can reach a Golf Course, the algorithms *SpatialRM* and *SpotempRM* tried to find alternative destinations considering only those 4 possibilities of places.

On the other hand, Figure 42 shows the values of potentially avoided OTs having as destination Italian restaurants. In our dataset of POIs, there are 690 Italian restaurants in Florence, giving for the *SpatialRM* and *SpotempRM* test configurations, a huge number of places to be seen as possible destinations to avoid an OT. Therefore, *SpatialRM* and *SpotempRM* produced substantial and high values of potentially avoided OTs for all the tested values of α and *i*. When compared the performance gains of *SpatialRM* and *TemporalRM*, for all the test configuration of α and *i*, our results indicated that *SpatialRM* generated better performance than *TemporalRM*, where with test configuration *α=1km* and *i=0.5 hour* the *SpatialRM* avoided 71.2% OTs, while the only *TemporalRM* 28.3%. It shows the strength of the spatial dimension given the destination's availability for the user to perform the intended activity. Therefore, comparing both Figure 41 e Figure 42 we can easily see how more available and spread on the map are the places of a given intended category more possibilities to avoid the OT can be found by the *SpatialRM* and *SpotempRM's* algorithms.

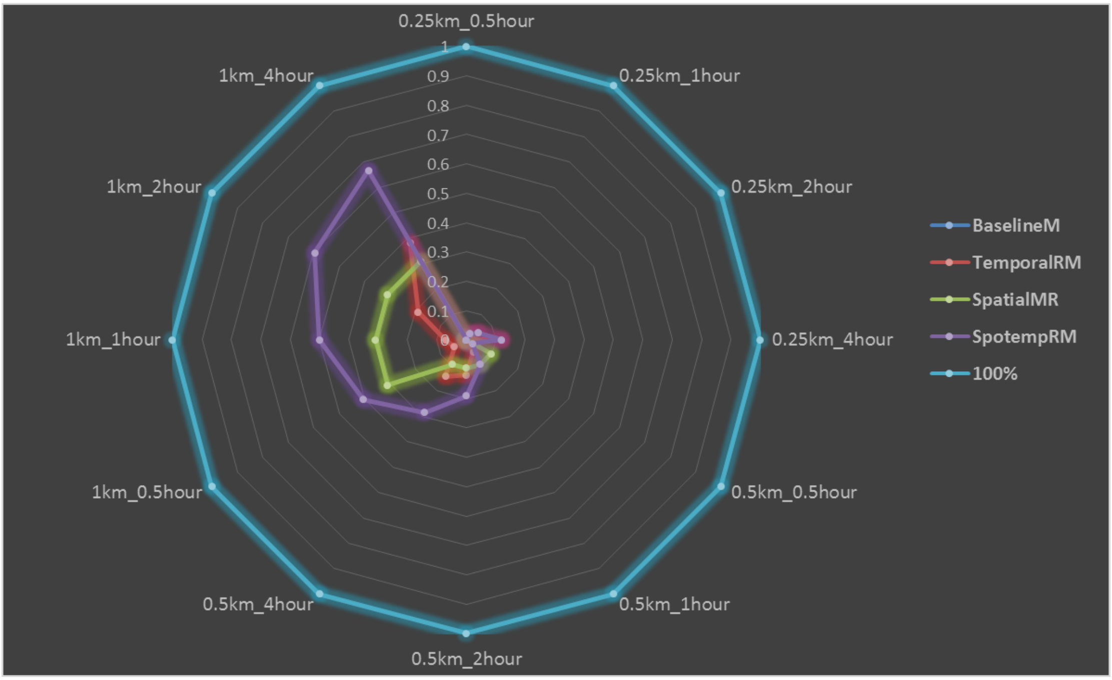


Figure 41: OTs potentially avoided for the activity Golf Course.

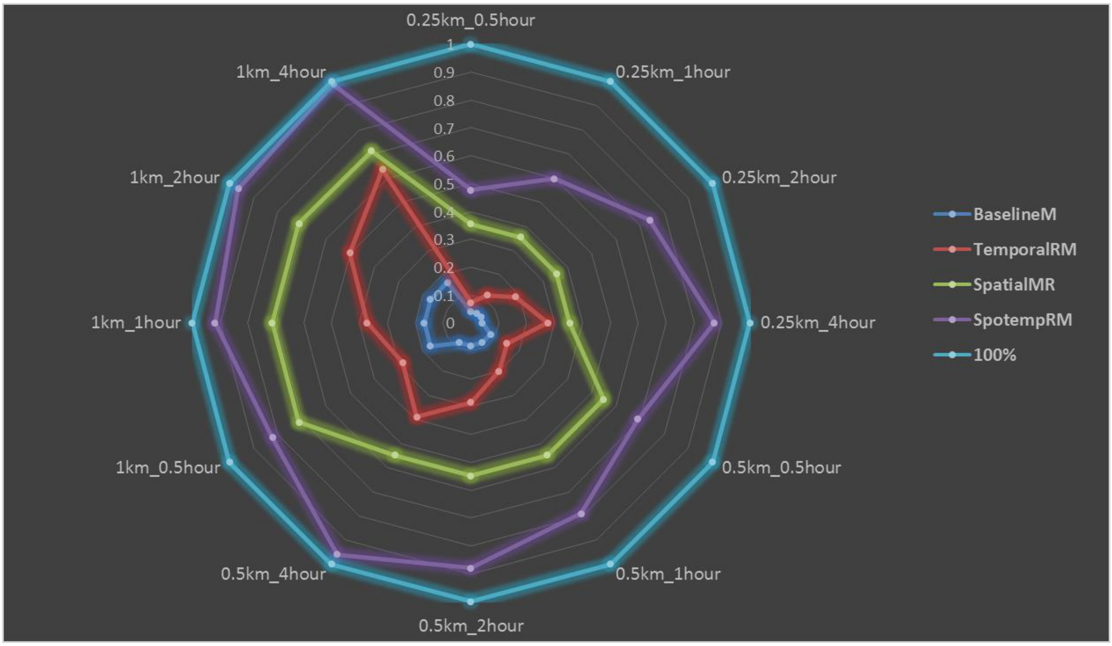


Figure 42: OTs potentially avoided for the activity Italian Restaurant.

4.5. Conclusion

From the study of people's mobility behavior presented in the Chapter 3, we noticed different patterns of behavior about visiting some place categories. These patterns were studied with respect to the regularity and irregularity of users to change the schedule and the place to perform some activities. Relying on this analysis, an activity-oriented matching method for carpooling was proposed.

Carpooling allows people to share the cost of gas and parking, cutting the expenses of the people involved and also helping to reduce the costs with construction of new roads, road maintenance and air pollution related health costs. Our carpool matching method is based on the idea that some activities can be performed by a person at a huge number of places, as well as can be postponed or anticipated according to the person's convenience to perform it.

Therefore, our matching algorithms work with the possibilities of spatial and temporal reallocation. We presented three different matching algorithms for our method, each one relaxing the temporal and spatial constraints in a different way. Our method uses a spatial grid composed by virtual cells to facilitate the search for carpool matching. Thus, supported by a Semantic Grid (SG) containing a semantic mapping of the POIs for activity, our method searches for routine trips (RTs) that could supply the carpooling requests (CRs) received as input.

We defined an Evaluation Process intending to apply our carpool matching method to discovery in a dataset of trajectories the amount of trips that could be avoided by using our method. The Evaluation Process split the dataset of trajectories into routine trips (RT) and occasional trips (OT). The group of routines trips consists of trips daily performed by the user, while the group of occasional trips differently correspond to trips not often performed by the users. The occasional trips are often related to destinations that are not home or work places, as they are related to activities that are not strictly to specific destination places and times.

A baseline algorithm was also executed to be compared with the three matching algorithms proposed. As metric evaluation, we compute the percentage of occasional trips that could be potentially avoided by our method. An experiment in a real dataset was performed. The results derived from our experiments showed how the carpooling practice can be improved using the proposed carpooling matching methods to avoid larger amount of trips. The experiment's results, showed that the algorithm *SpotempRM* reached a gain of +82.30% when compare with the baseline method. In another analysis, a total of 64.537 km could be avoided when using the algorithm *TemporalRM*, 75.800 km could be avoided when using the *SpatialRM*, 106.786 km when using the algorithm *SpotempRM*, while the baseline reached only 18.030 km.

Our proposed method can be useful for on-line carpooling recommendation systems based on the user's intended activity. It can also be useful for performing analytical queries about the adoption of the carpooling practice, providing general results or, in more specific, for determined activity of the population.

Chapter 5: Conclusions

This chapter aims to present the final considerations about the main topics discussed in this dissertation, including the main contributions obtained and some directions for future work.

5.1. Final Considerations

Huge datasets of human mobility like GPS traces from GPS devices and check-ins from Location Based Social Network are a rich source of knowledge about human mobility behavior. Many studies are concerned with human mobility and the current work also went on this direction.

Our studies on human mobility data have focused on the users' activity. First, we defined a method to measure the regularity of the users about their choices to visit some place to perform a given activity. This regularity measure is computed using two dimensions: temporal and spatial. The spatial regularity measure is related to the regularity to visit or not the same place to perform a given activity. In turn, the temporal regularity measure corresponds to user's regularity to visit a type of place to perform a given activity in certain period of time.

We also developed a web tool to show for each POI, the loyalty behavior of his visitors. This loyalty behavior is based on the users' regularity discussed previously. The web tool is called MAPMOLTY, which computes for a given user's activity, a set of loyalty indicators about the POIs displayed on a map. MAPMOLTY is a customized tool and allows different types of analysis to be done through an interaction with a map.

The study about spatial and temporal regularity contributed for the proposal of a carpool matching method based on the intended activity of the user. The matching method is oriented to the user's intended activity and uses semantic information about the POIs of a given region to boost the possibilities of carpooling. Experiments were performed on a real dataset of trajectories using our proposed algorithms for carpool matching. Our results, compared to a baseline algorithm, showed a substantial increase for carpool matching.

To conclude, the remainder of this chapter is organized as follows. Section 5.2, addresses the principal contributions obtained with the development of this work and Section 5.3 describes the future work that can be done to continue the research project described in this dissertation.

5.2. Main Contributions

The main contributions of this work are discussed below. The order in which they are listed is the same order as they were presented in this document:

(1) Semantic Regularity Profile

The definition of spatial and temporal entropy as a measure of the semantic regularity of users computed from mobility data. This measure quantifies the users' regularity with respect to their visits to POIs of a given category (activity). The regularity is computed in terms spatial and temporal data. For the spatial regularity, the measure computes how regular the user is about visiting the same place to perform a given activity. On the other hand, for the temporal regularity, the measure computes the user's regularity for visiting a category of place in a specific interval of the day.

(2) A web tool for discovering place loyalty based on mobility data

We developed a web tool called MAPMOLTY. For a given dataset of mobility data, like GPS tracks or check-ins of individuals, and a set of Points Of Interest, MAPMOLTY computes a number of measures, called loyalty indicators, to summarize the loyalty level of each POI. MAPMOLTY presents the loyalty measures of each POI to be viewed on a MAP using three visualization properties: Marker, Circle Size and Circle Opacity. The analysis is made according to the frequency distribution of the users' visits. The tool also contains a high level of users' customization and provides a first kernel of features that can be easily extended with new functionalities.

(3) An Activity-Oriented Matching Method for Carpooling

A novel matching method for Carpooling was proposed. The method is based on the following idea: to perform some activities a person has several POIs in which these activities can be performed and, in some cases, the activities can be performed by postponing or anticipating the time originally planned to execute them. The carpool matching is done considering three different algorithms according to the manipulation of the spatial and temporal dimensions. Experiments on a real dataset was performed using different parameter configurations. We also proposed an evaluation process to compute from a dataset of trajectories the amount of trips that can be avoided using our proposed matching method. Results showed that our proposed matching algorithms for carpooling improved the traditional carpooling approach in +50.89% when the temporal dimension was considered, in +46.84% when the spatial dimension was prioritized and in +82.30% when both dimensions were tackled.

5.3. Future Works

To continue the research initiated in this dissertation, this section lists proposals for future work to be performed:

(1) Extend the Semantic Regularity Profile

It would be interesting to extend the regularity profile measure including the concept of a hierarchy into the spatial and temporal dimensions. The spatial measure could be computed using a hierarchy structure considering the following aggregation level about the POI: category (i.e. highest level), subcategory, possible affiliates and the own POIs (i.e. lowest level). For example, regarding the restaurant category, there are subcategories like Italians, Chinese, Fast Food, Japanese, among others and, furthermore, some famous restaurants have branches scattered in the cities, for example MC Donald's and Pizza Hut.

Thereby, if a person usually visits different POIs to perform a given activity and those POIs belong to the same subcategory or they are branches of the same company, then it shows a certain regularity level since some properties among the visited places are equals. With respect to the temporal dimension the method could manipulate a time hierarchy by considering the phases of the day: morning, afternoon and evening. It would also be interesting to perform experiments using dynamic interval windows instead of using fixed intervals. Thus, given an interval window size, a user and an activity, the classes to compute the temporal distribution could be created by grouping near time visits respecting to the interval window size.

Through our regularity measure, we can classify the user among regular and not regular for several activities. Thus, we could apply an association algorithm to find interesting patterns about the user's behavior with respect to the execution of some activities. Given a dataset of semantic profiles, we could find as hypothetical example, users that are spatially regular about visiting gyms and are not temporally regular when visiting universities. Another work proposal is to apply cluster algorithms to group similar user's behavior and also to summarize the population behavior of a given dataset.

Another possibility of work is to adapt this regularity measure for another context like sale of products. One approach is to identify for some type of products the customers with regular or irregular behavior with regards to the acquisition of same brands of products.

(2) New features for MAPMOLTY

It would be important to improve the tool using OLAP Operators to provide faster computation of the user's distribution with respect to the time. This implies in improving the data manipulation using OLAP operators like drill down/up, dice and slices. The results of the loyalty indicator could be showed, for example, filtering the data by weeks, months or years.

Another important feature is an aggregated analysis of the POIs of a specific area. Using a spatial grid we would be able to cluster nearby POIs and compute values of loyalty to a given cell of the grid grouping the data. It enables an analysis by area and not only by POI as it is currently done.

(3) Activity-Oriented System Recommendation for Carpooling

Our matching method for carpooling returns the possibilities of rides, but without ranking them according to the best options of rides for the passenger. An open challenge is: how to sort a list of possible rides considering an activity request of the user? Then, based on the drivers' routes, possible routes could be ranked according to the intended activity requested by the passenger. A web system can be developed with friendly interface and use this method as core for the matching computation.

References

A.LIYANA, K.INA, L.IOANNA, C. B. **Carpooling (The Carpooling Box)**. **Vasa**. [S.l: s.n.]. Disponível em: <http://medcontent.metapress.com/index/A65RM03P4874243N.pdf>, 2012.

AL, C. Casual Carpool Survey Summary 2010 - Final. **511 RIDESHARE** v. 94607, n. July, 2010.

ALVARES, L. O. et al. A Model for Enriching Trajectories with Semantic Geographical Information Categories and Subject Descriptors. **Energies**, 2007.

ARNOULD, G. et al. A transport based clearing system for dynamic carpooling business services. **2011 11th International Conference on ITS Telecommunications**, p. 527–533, 2011.

BELLEMANS, T. et al. An Agent-Based Model to Evaluate Carpooling at Large Manufacturing Plants. **Procedia Computer Science**, v. 10, n. 270833, p. 1221–1227, doi:10.1016/j.procs.2012.08.001, 2012.

BHAT, C. R. e KOPPELMAN, F. S. 3 ACTIVITY-BASED MODELING OF TRAVEL DEMAND. **Handbook of Transportation Science.** doi:10.1007/0-306-48058-1_3, [S.d.].

CARPOOLING.COM. **Carpooling.com**. Disponível em: <http://www.carpooling.com> Acesso em: 24 julho 2014.

CARPOOLWORLD. **Carpoolworld**, free worldwide carpool matching. Disponível em: <http://www.carpoolworld.com/>. Acesso em: 24 julho 2014.

CHEN, C. et al. Smart Ride Share with Flexible Route Matching. **Advanced Communication Technology (ICACT), 2011 13th International Conference on.** p. 1506–1510, 2011.

CHO, E. Friendship and Mobility : User Movement In Location-Based Social Networks. **Proceeding. KDD '11 Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining**, 2011.

CHO, S. et al. An Activity-based Carpooling Microsimulation Using Ontology. **Procedia Computer Science**, v. 19, n. Ant, p. 48–55, doi:10.1016/j.procs.2013.06.012, 2013.

CICI, B. et al. Quantifying the potential of ride-sharing using call description records. **Proceedings of the 14th Workshop on Mobile Computing Systems and Applications - HotMobile ’13**, p. 1, doi:10.1145/2444776.2444799, 2013.

COLLOTTA, M et al. A NOVEL TRUST BASED ALGORITHM FOR CARPOOLING TRANSPORTATION SYSTEMS. **Energy Conference and Exhibition (ENERGYCON), 2012 IEEE International.** p. 1077–1082, 2012.

DIMITRIESKI, V. Real-Time Carpooling and Ride-Sharing: Position Paper on Design Concepts, Distribution and Cloud Computing Strategies. **Proceedings of the 2013 Federated Conference on Computer Science and Information Systems pp. 781–786 Real-Time**, 2013.

DIXIT, A. et al. Real-Time Carpooling System for Android Platform. **International Journal of Engineering and Innovative Technology (IJEIT)** v. 2, n. 6, p. 436–437, 2012.

DJENGO, **DJENGO,** carpooling solution for your company, Disponível em: <http://www.djengo.be/ (2014)>. Acesso em: 24 julho 2014.

TRANS CANADA CARPOOL.CA. **Trans Canada Carpool.ca** Do your bit… Share it. Dispoível em: <http://www.carpool.ca/About.aspx> Acesso em: 24 julho 2014.

EAGLE, N. e (SANDY) PENTLAND, A. Reality mining: sensing complex social systems. **Personal and Ubiquitous Computing**, v. 10, n. 4, p. 255–268, doi:10.1007/s00779-005-0046-3, 2005.

FEDOR, M.;; ŠEBO, J. e PETRIKOVÁ, A. Innovative application of inventory theory for determining optimal fleet size for a car-sharing system. **SAMI 2012 10th IEEE Jubilee International Symposium on Applied Machine Intelligence and Informatics**. p. 157–160, 2012.

FURLETTI, B. et al. Inferring human activities from GPS tracks. **UrbComp 2013.** 2013.

GIANNOTTI, F. et al. Unveiling the complexity of human mobility by querying and mining massive trajectory data. **The VLDB Journal**, v. 20, n. 5, p. 695–719, doi:10.1007/s00778-011-0244-8, 2011.

GUC, B. et al. Semantic Annotation of GPS Trajectories. **11th AGILE International Conference on Geographic Information Science.** p. 1–9, 2008.

GUIDOTTI, R. Human Mobility Analysis using Ranking Measures. **Master Thesis in Computer in Computer Science. Università di Pisa.** 2013.

HASAN, S.;; ZHAN, X. e UKKUSURI, S. V. Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. In: **PROCEEDINGS OF THE 2ND ACM SIGKDD INTERNATIONAL WORKSHOP ON URBAN COMPUTING - URBCOMP '13**. **Anais**... New York, New York, USA: ACM Press. Disponível em: <http://dl.acm.org/citation.cfm?doid=2505821.2505823>. Acesso em: 7 jul. 2014, 2013.

JIANG, S.;; FERREIRA, J. e GONZÁLEZ, M. C. Clustering daily patterns of human activities in the city. **Data Mining and Knowledge Discovery**, v. 25, n. 3, p. 478–510, doi:10.1007/s10618-012-0264-z, 2012.

JOVICIC, G. **Activity based travel demand modelling Activity based travel demand modelling**. [S.l: s.n.], 2001.

KATZEV, R. Car Sharing: A New Approach to Urban Transportation Problems. **Analyses of Social Issues and Public Policy**, v. 3, n. 1, p. 65–86, doi:10.1111/j.1530-2415.2003.00015.x, 2003.

KELLEY, K. L. Casual Carpooling — Enhanced. **Journal of Public Transportation**, Vol. 10, No. 4. p. 119–130, [S.d.].

LAUBE, P.;; IMFELD, S. e WEIBEL, R. Discovering Relative Motion Patterns in Groups of Moving Point Objects. p. 1–34, 2009.

LI, J. et al. Who Chooses to Carpool and Why?: Examination of Texas Carpoolers. **Transportation Research Record**, v. 2021, n. -1, p. 110–117, doi:10.3141/2021-13, 2007.

LIRA, V. M. DE et al (a). Investigating semantic regularity of human mobility lifestyle. **Proceedings of the 18th International Database Engineering & Applications Symposium on - IDEAS '14**, p. 314–317, doi:10.1145/2628194.2628226, 2014.

LIRA, V. M. DE et al (b). Mapmolty: a web tool for discovering place loyalty based on mobile crowdsource data. **14th International Conference on Web Engineering ICWE**, 2014.

MA, S. e WOLFSON, O. Analysis and evaluation of the slugging form of ridesharing. **Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - SIGSPATIAL'13**, p. 64–73, doi:10.1145/2525314.2525365, 2013.

MAZZULLA, G. An activity-based system of models for student mobility simulation. **European Transport Research Review**, v. 1, n. 4, p. 163–174, doi:10.1007/s12544-009-0017-2, 2009.

MINETT, P. e PEARCE, J. Estimating the Energy Consumption Impact of Casual Carpooling. **Energies**, 2011.

MORENO, B. N. et al. Looking inside the Stops of Trajectories of Moving Objects. **In GeoInfo**, 2010.

PALMA, A. T. et al. A clustering-based approach for discovering interesting places in trajectories. **Proceedings of the 2008 ACM symposium on Applied computing - SAC '08**, p. 863, doi:10.1145/1363686.1363886, 2008.

PARENT, C. et al. Semantic trajectories modeling and analysis. **ACM Computing Surveys**, v. 45, n. 4, p. 1–32, doi:10.1145/2501654.2501656, 2013.

PERTERER, N. et al. Come Drive with Me : An Ethnographic Study of Driver - Passenger Pairs to Inform Future In - Car Assistance. **Proceeding CSCW '13 Proceedings of the 2013 conference on Computer supported cooperative work**. p. 1539–1548, 2013.

PUKHOVSKIY, N. V. e LEPSHOKOV, R. E. Real-time carpooling system. **2011 International Conference on Collaboration Technologies and Systems (CTS)**, p. 648–649, doi:10.1109/CTS.2011.5928758, 2011.

QIN, S. M. et al. Patterns, Entropy, and Predictability of Human Mobility and Life. **PLoS ONE**, v. 7, n. 12, 2012.

RENSO et al. Mobility Data: Modeling, Management, and Understanding. Book. **Cambridge Press**, 2013.

ROCHA, J. A. M. R. et al. DB-SMoT: A direction-based spatio-temporal clustering method. **2010 5th IEEE International Conference Intelligent Systems**, p. 114–119, doi:10.1109/IS.2010.5548396, 2010.

SGHAIER, M. et al. A distributed dijkstra's algorithm for the implementation of a Real Time Carpooling Service with an optimized aspect on siblings. **13th International IEEE Conference on Intelligent Transportation Systems**, p. 795–800, doi:10.1109/ITSC.2010.5625288, 2010.

SGHAIER, M. et al. A novel approach based on a distributed dynamic graph modeling set up over a subdivision process to deal with distributed optimized real time carpooling requests. **2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)**, p. 1311–1316, doi:10.1109/ITSC.2011.6082983, 2011.

SHANNON, C. E. A mathematical theory of communication. **Bell System Technical Journal**, 27:379-423, 623-656, 1948.

SHIFTAN, Y. et al. Activity-Based Modeling as a Tool for Better Understanding. **10th International Conference on Travel Behaviour Research.** August, p. 10–15, 2003.

SONG, C. et al. Limits of predictability in human mobility. **Science (New York, N.Y.)**, v. 327, n. 5968, p. 1018–21, doi:10.1126/science.1177170, 2010.

TRASARTI, R et al.. Mining mobility user profiles for car pooling. **Proceeding KDD '11 Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining.** n. iii, p. 1190–1198, 2011.

VANDERSMISSEN, M.-H. et al. Modeling propensity to move after job change using event history analysis and temporal GIS. **Journal of Geographical Systems**, v. 11, n. 1, p. 37–65, doi:10.1007/s10109-009-0076-x, 2009.

VANOUTRIVE, T. et al. What determines carpooling to workplaces in Belgium: Location, organisation, or promotion? **Journal of Transport Geography**, v. 22, p. 77–86, 2012.

VERKEHRS-, A. Activity-based possibilities modelling : **KW Axhausen**. October, 2000.

WILLIAMS, M. J.;; WHITAKER, R. M. e ALLEN, S. M. Measuring Individual Regularity in Human Visiting Patterns. **2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing**, p. 117–122, doi:10.1109/SocialCom-PASSAT.2012.93, 2012.

WU, L. et al. Intra-urban human mobility and activity transition: evidence from social media check-in data. **PloS one**, v. 9, n. 5, p. e97010, doi:10.1371/journal.pone.0097010, 2014.

YE, M. e YIN, P. Location Recommendation for Location-based Social Networks. **Analysis, Proceeding GIS '10 Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems**. Pages 458-461. n. c, p. 458–461, doi:10.1145/1869790.1869861, 2010.