Analysis of human mobility for architecture and urban studies

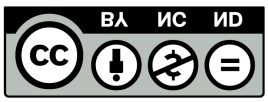
Yuji Yoshimura

TESI DOCTORAL UPF / 2016

DIRECTOR DE LA TESI

Dr. Josep Blat

DEPARTAMENT OF INFORMATION AND COMMUNICATION TECHNOLOGIES



Acknowlegements

I wish to acknowledge some of the many people who have helped make this work possible. Without their supports and helps, this work would not have been realized.

First, I am very grateful to my supervisor Josep Blat, for giving me the chance to work in his laboratory. Thank you for all the time in the supervision of my thesis and for the scientific discussion.

I would like to also thank Fabien Girardin for his continuous guidance, advice and encouragement throughout my doctoral studies and dissertation work.

I am grateful to Prof. Carlo Ratti who hosted me in his SENSEable City Laboratory at the Massachusetts Institute of Technology. I would also like to acknowledge the generous support I have received from all my SENSEable City Lab friends. I would like to extend my personal thanks to Stanislav Sobolevsky, Alexander Amini, David Lee, Michael Szell, Riccardo Campari, Sebastian Grauwin, Dietmar Offenhuber, Simone Mora, Zolzaya Dashdorj, all of whom contributed in various ways to the making of this project.

I would also like to extend my appreciation to Anne Krebs, the socio-economic Studies and Research division of the Louvre Museum, for her assistance in the experimental setups for my research.

Prof. Jaume Barceló played a fundamental role in helping me reach my goals to the completion of this thesis through constant discussions and encouragements. I express my sincere gratitude to Prof. Jaume Barceló for the time he spent debating and discussing with me.

Finally, I am indebted to my mother and sister, Harumi and Momoyo, and my grandmother and aunt. Thank you for your continuous support, love and encouragement.

Abstract

The goal of this dissertation is to provide insights into spatial analysis for architecture and urban studies through examining human mobility in the built environment based on large-scale datasets. Three studies are presented. The first one analyzed visitors' mobility in the Louvre Museum. We observed how visitors' paths superpose, independent from their length of stay, and uncovered that this is the cause for the congestion in the museum. The second study focused on pedestrians' behavioral differences between a discount day and normal days in the historical center of Barcelona. We found that pedestrians in a discount day tend to stay a shorter time than those during normal days, but that the former visit a larger number of nodes than the latter. The third study dealt with customers' mobility patterns considering their purchase activities across the city. We uncovered that the fraction between the number of stores of category A2 (Groceries) and the number of transactions conducted in this category is 2.15, revealing that the number of stores was almost two times greater than the number of transactions conducted in this category. Our findings provide important insights for understanding human mobility in architecture and urban areas, and also suggest the potential for managing crowds and reducing congestions, which is one of the most critical aspect of urban and architectural management.

Resumen

El objetivo de esta tesis es proporcionar conocimientos sobre el análisis espacial de la arquitectura y los estudios urbanos mediante el examen de la movilidad humana en el entorno construido sobre la base de conjuntos de datos a gran escala. Se presentan tres estudios. El primero analiza la movilidad de los visitantes en el museo del Louvre. Hemos observado como se superponen los caminos de los visitantes, independientemente de la duración de su estancia, y descubrimos que esta es la causa de congestión en el museo. El segundo estudio se centró en las diferencias de comportamiento de los peatones entre un día de descuento y los días normales en el centro histórico de Barcelona. Se encontró que los peatones en un día de descuento tienden a permanecer un tiempo más corto de lo normal durante el día, pero que visitan un mayor número de nodos en comparación con los días normales. El tercer estudio aborda las pautas de movilidad de los clientes teniendo en cuenta sus actividades de compra a través de la ciudad. Hemos descubierto que la fracción entre el número de tiendas de categoría A2 (productos alimenticios) y el número de transacciones realizadas en esta categoría es de 2.15, revelando que el número de tiendas fue casi dos veces mayor que el número de transacciones realizadas en esta categoría. Así pues, nuestros resultados proporcionan información importante para la comprensión de la movilidad humana en la arquitectura y las zonas urbanas y, también, sugieren el potencial para la gestión de multitudes y la reducción de la congestión, que es uno de los aspectos más críticos de la gestión urbana y arquitectónica.

Resum

L'objectiu d'aquesta tesi és proporcionar coneixements sobre l'anàlisi espacial de l'arquitectura i els estudis urbans mitjançant l'examen de la mobilitat humana en l'entorn construït sobre la base de conjunts de dades a gran escala. Es presenten tres estudis. El primer analitza la mobilitat dels visitants al museu del Louvre. Hem observat com es superposen els camins dels visitants, independentment de la durada de la seva estada, i hem descobert que aquesta és la causa de congestió al museu. El segon estudi es va centrar en les diferències de comportament dels vianants entre un dia de descompte i els dies normals al centre històric de Barcelona. Es va trobar que els vianants en un dia de descompte tendeixen a romandre un temps més curt del normal durant el dia, però que visiten un major nombre de nodes en comparació amb els dies normals. El tercer estudi aborda les pautes de mobilitat dels clients tenint en compte les seves activitats de compra a través de la ciutat. Hem descobert que la fracció entre el nombre de botigues de categoria A2 (productes alimentaris) i el nombre de transaccions realitzades en aquesta categoria és de 2.15, revelant que el nombre de botigues va ser gairebé dues vegades més gran que el nombre de transaccions realitzades en aquesta categoria . Així doncs, els nostres resultats proporcionen informació important per a la comprensió de la mobilitat humana en l'arquitectura i les zones urbanes i, també, suggereixen el potencial per a la gestió de multituds i la reducció de la congestió, que és un dels aspectes més crítics de la gestió urbana i arquitectònica.

Contents

List of Figures

List of tables

1. Introduction

1.1 The research context and background

Architectural design, planning and programing

The research area of this dissertation is human mobility analysis, framed within *Architectural Programing* (Cherry, 1998). Architectural programming is "a process leading to the statement of an architectural problem and the requirements to be met in offering a solution" (Peña, 2001, p14). Architectural programing discovers the client's needs and goals, and seeks sufficient information in order to understand those requirements better.

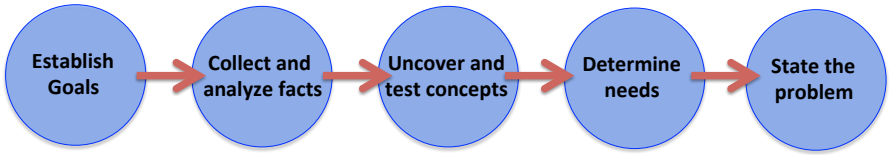


Figure 1.1 Five steps of architectural programing

Figure 1.1 presents the five steps of architectural programing, namely, (1) Establish goals, (2) Collect and analyze facts, (3) Uncover and test concepts, (4) Determine needs and (5) State the problem. Within these five steps, this dissertation focuses on the second step, Collect and analyze facts, in particular, for human mobility in the built environment. To collect data about human activities and analyze them scientifically can be an essential ingredient to establish the goals and concepts of the project, and stating potential problems to achieve them.

Architectural programing is different from architectural design, although they are strongly related to each other. Architectural programing is the analysis in the pre-design phase of a project, while architectural design is a problem solving phase, which should provide answers to the stated problem through the design process.

Architectural design can be defined as "a conceptual activity of formulating an idea intended to be expressed in a visible form or carried into action" (Terzidis, 2006, p1). *Architectural planning* is "the act of devising a scheme, program, or method worked out

before hand for the accomplishment of an objective" (Terzidis, 2006, p1). Thus, architectural planning contains architectural programing. Architectural planning is different from architectural design, resulting in programing and design being also different.

Throughout the architectural history, there have been a variety of attempts for designing architecture from different kinds of perspectives. The design process using emerging technologies is relevant to this dissertation, and we summarily discuss some of its aspects next.

Mathematics offers a common language for different fields such as art, engineering, science and architecture (Kappraff, 2001), and it has had a strong relation with architecture since the Greeks: geometry bridged the technological and aesthetical aspects of architecture, and it has frequently been used to determine the forms of architecture [e.g., see the works of Lissitzky, van Doesburg (Tzortzi, 2007)]. Topology through graphs has been applied to architecture (in particular, cities) to represent the spatial connectivity (Alexander, 1964; Alexander, 1965; Alexander, 1966; Alexander, 1968; March & Steadman, 1971). Mathematical proportion has been used to analyze architectural expression as an artwork (Rowe, 1947). A mathematical concept as "function" has been used as a metaphor to criticize the classical system of ornaments for architecture (Forty, p174). Design tools such as Shape Grammar (Stiny, 1980) enable generating plans for architecture automatically. Parametric Design and Algorithmic Design (Terzidis, 2006) use the parameters and algorithms to generate architectural forms. Algorithmic architecture "involves the design of software programs to generate space and form from the rule-based logic inherent in architectural programs, typologies, building code, and language itself" (Terzidis, 2006, p7).

On the other hand, scientific research and their methodologies have not been explored enough yet to analyze human activities in architectural spaces. There have been few attempts to scientifically and quantitatively analyze human activities in architecture, most of them falling into those classified within the qualitative method, because of using interviews, questionnaires and direct observation (Peña, 1969; Preiser, 1988; Lynch, 1960; Gehl, 2011). Few attempts

can be found in quantifying human behaviors for the analysis of architectural and urban spaces.

This dissertation employs a quantitative research method for spatial analysis for architecture through mobility analysis, instead of using qualitative techniques.

The next section reviews the human mobility analysis, which has been conducted within the framework of the architecture.

Human mobility analysis in architecture

Urban researchers have shown their interest in understanding the human activities, including their mobility in the built environment (Hillier, 1996; Barabási, 2010). The corresponding analyses are often conducted at several scales, from the regional one (González et al, 2008; Ratti et al, 2006), through the district scale (Shoval et al, 2013; Delafontain et al, 2012; Eagle & Pentland, 2005; Kostakos et al, 2010) to the building scale (Versichele et al, 2012; Kanda et al, 2007; Hui et al, 2009; Yoshimura et al, 2014). The research issues are diverse: to reveal the universal laws of human mobility (Gonzalez et al, 2008), to uncover the pattern of people's activities (Ratti et al, 2006), to find the transition probabilities between places for touristic activities (Shoval et al, 2013) and to describe customers' purchase behaviors (Hui et al, 2009).

Architects and urban researchers have been trying to analyze the human movement in order to uncover relationships between the spatial layout of the built environment (i.e., spatial forms) and human behaviors (Hillier, 1996; Hillier & Tzortzi, 2006; Choi, 1999; Porta et al, 2011). The spatial forms and their configurations largely impact on human activities, e.g., how people explore and stay longer or shorter in the built environment (Cho, 1999). That is, our activities, spatial cognitions, even feelings, may be widely determined and framed by the spatial layouts and their configurations (Ellard, 2010). *Space Syntax*, the topological representation of urban spaces, has helped to identify that it is key to the concentration of people's presence, considering the cognitive roles of human perceptions (Hillier and Hanson 1984, Hillier, 1996).

More recently, “The ability to collect time-space data in such high resolutions in time (seconds) and space (metres) for long periods of time opens up the possibility of drawing new lines of inquiry and creates opportunities to formulate new research questions that could not be previously asked” (Shoval, 2008). The spatial analysis through human movement has been based on datasets, and it can be classified into three groups, depending on the datasets to be obtained and the way of collecting them. We present a categorization based on data collection instead of the results of spatial analysis, because the data collection process is a critical step that deeply impacts on the final outcome of a study: the data has to be rich enough to describe human behavior yet it has to avoid the inclusion of errors that might hamper the reliability of the subsequent data analysis.

The first group covers low-tech or manual data collection techniques. Observations, questionnaires and interview-based surveys have been the most common methods to acquire data of humans’ movements and behaviors. These methods provide insights into people’s actual behavior and enable obtaining socio-demographic attributes (i.e., gender, age) and inner thoughts of the subjects (e.g., motivations). They enhance the understanding of the use of the built environment and provide basic knowledge for spatial management.

The second group relates to active data collection techniques, which makes use of emerging technologies such as GPS, WiFi access points and wearable sensors (Sparacino, 2002; Tschacher et al., 2012). These techniques involve the subjects to actively take part in the process of data collection. These techniques provide more objective and precise time, space and routing data than the traditional / manual based methods, because they do not depend on observers’ or interviewees’ subjectivity. For example, mobile devices with GPS (Asakura & Iryo, 2007; Shoval et al., 2013; Rhee, Shin, Hong, Lee, Kim & Chong, 2011) and actively used RFID tags (Kanda et al., 2007) provide sample data of human movement at fine granularity both spatial and temporal.

The third group is based on passive data collection techniques, including computer-vision aided image analysis (Antonini et al., 2006), Laser Range Finders (Nakamura et al., 2005), passive mobile

positioning data (Ratti et al. 2006, Ahas et al. 2008), Bluetooth detection (Yoshimura et al., 2014), social media analysis (Hawelka et al, 2014) and user generated content (Girardin et al. 2008). These techniques collect human movement and behavior without users' active participation, because they are based on the systematic observation in the framework of "unobtrusive measures" (Webb et al. 2000), making use of unconsciously left people's digital footprints.

These three groups are complementary in terms of way of detection, detection range and sample size.

The first group requires a large human effort, both from respondents and observers, particularly in the transcription and digitalization of the data. Manual data entry is a time-consuming task and data entry errors might lie unseen. Moreover, the observation and tracking of pedestrians may involve errors caused by the observer being overwhelmed, while interviews and questionnaires could be biased through self-reporting errors typical of these qualitative research techniques. In addition, large efforts to take detailed notes in a sequence of days increase the risks of non-response and decrease their quality and accuracy, which depends on the ambiguous memories to determine the exact location of the people's visits. Furthermore, due to the above mentioned features of the way these data are collected, they are likely to result in a relatively small sample size.

The active data collection techniques largely improve on the shortcomings of the first class, with examples such as travel diary methods combined with GPS (Fraijer et al, 2000; Ohmori et al, 2005). However, these techniques require that the users carry specific mobile devices (e.g., active RFID), which added to the high cost of tags, result in a small sample size as well. In addition, they can potentially deliver biased results, because participants adapt their behaviors when they know they are being observed, or because they are likely to provide socially desirable responses to the survey. Furthermore, other reasons leading to small scale datasets are that the techniques require a pervasive sensor infrastructure, and a participatory process to provide mobile devices or tags in advance of the experiment. This led previous researchers to fail to involve mass participation.

Conversely, the passive data collection techniques can provide a large scale dataset of human movement, because they are based on unobtrusive methods. However, the users' socio-demographic attributes and their inner thoughts are neither available nor obtained in most cases.

This dissertation analyzes sequential human movements in the topologically represented architectural and urban space through a passive data collection approach. We deal with passive data collection techniques, because they have advantages over both manual based ones and active ones in some specific contexts. The specific context we consider contains large buildings, malls, undergrounds and urban districts for shopping purposes, where pedestrians can walk without restrictions and their length of stay can be no more than few hours. Tracking a pedestrian from his/her entry into one of these till his/her exit by an observer requires a large effort, but results in just a single sample. Inside buildings, the most common state-of-the-art-technologies (i.e., GPS, mobile phone tracking) do not work. RFID needs to prepare large-scale infrastructures and registration of visitors in advance.

Indeed, for shopping behaviors in an urban district, many previous researches use small scale datasets collected through interviews, observations or GPS tracking. Marketing researchers tend to analyze them in qualitative terms to uncover psychological aspects (see Teller et al, 2008 for a review), while urban researchers are likely to apply modeling and simulation approaches to discuss shopping behaviors (see Timmermans, 2009 for a review). Although within the modeling approach researchers use the small sample datasets to estimate variables for their model or to validate their constructed models, this approach is based on the strong assumption that the consumer behaves as a rational agent (Simon, 1978). That is, a consumer always seeks to maximize her/his benefit (e.g., by selecting the most relevant items from all options) and minimize the cost (e.g., by choosing the shortest path from the origin to destination). There are alternative which have been proposed such as the bounded rationality approach (Zhu & Timmermans, 2008) or the heuristic approach (Hagen, Borgers & Timmermans, 1991), but these assumptions and hypotheses do not match well the reality, especially, the shoppers' behaviors for the sales periods when they become impulsive consumers (Laroche et

al. 2003; Liao, Shen, Chu, 2009; Tinne, 2011; Virvilaite, Saladiene & Bagdonaite, 2009).

This dissertation deals with spatial analysis through human mobility in the specific context mentioned above. First, it is set in the mesoscopic scale in a built environment. The mesoscopic scale indicates that people move between points in a building or city, and their movements are tracked over the building or city scale (see Yoshimura et al, 2015 for more detail). Indeed, we study a large museum, the historical center of Barcelona and the city of Barcelona. Second, we deal with visitors' behaviors in the large art museum and shoppers' behaviors in the district and the city. Third, the analyses of those human movements derive from large-scale datasets, unlike most previously conducted human mobility researches, based on small scale datasets. This scale change is significant, because quantitative change in the data scale causes qualitative changes in the analysis:

"When we increase the scale of the data that we work with, we can do new things that weren't possible when we just worked with smaller amounts" (Mayer-Schönberger & Cukier, 2013, p10)

Thus, we reveal what the large scale datasets of human mobility can provide, uncovering hidden aspects of spatial impacts, showing its usefulness for spatial analysis which is much larger than the traditionally used small-scale datasets.

1.2 Main Research Questions

The main research question of this dissertation is:

How can the scientific analysis through datasets improve architecture and urban planning?

This question is not typical one for architects, because architecture is not considered a science (Forty, 2000, p 100). It is rather conceived of as an art, i.e., a highly subjective activity, based on intuition, imagination, creativity, rather than on objectively conducted experiments or observable evidences. Therefore, it is significant to ask how the scientific research method, in this case

based on objective analysis of datasets, can contribute to architecture. We will answer this too general question by formulating three specific sub-questions, which apply a scientific approach to architecture and urban studies. Namely, we introduce quantitative data and the analysis of human activities in architecture in terms of mobility, which are framed in the context summarized before. The research sub-questions in this dissertation are:

(1) *Which are the factors that affect people's behaviors in architecture? How can we use those factors to improve people's experiences in architecture?* Although a person is the main actor who uses the space, the quantitative analysis of human activities is rarely conducted in the field of architecture, largely due to the lack of adequate datasets. There have been few means to capture and collect human behaviors in a quantitative way such as large scale datasets. As a consequence, the scientific method is rarely applied for the design process of architecture. The first sub-question of this dissertation focuses on analyzing visitors' sequential movements and their length of stay in a large art museum. Within this research question we shed light on visitors' behaviors and their differences in terms of mobility patterns in terms of the different length of stay, which is a critical aspect in designing architecture.

(2) *Which are the factors that affect people's behaviors in the urban district? How can we use those factors in urban planning to enhance their experiences?* The pedestrians' behavior in a shopping district can be considered as key not only for retailers, but also for urban planning and its management. However, it has been extremely difficult to obtain behaviors through conventional data collection and analytical methodologies are lacking. The interview-based research tends to remain just at small scale samples, which are meant to represent "typical" pedestrian behavior. The modeling approach cannot correctly reproduce their behaviors, because the approach is based on the strong assumption that pedestrians are rational agents, which is contrary to the shopping behaviors, characterized by the emotional, unplanned and "on the spot decision-making". Against this framework, architects tend to make a urban

planning in the commercial districts based on their previous experiences, intuitions, imaginations and artistic drawings, rather than on the scientific analysis of human behaviors. Within this research question, we shed some light on pedestrians' behaviors and their differences during discount sale and normal days, in terms of their sequential movement and length of stay in the district.

(3) *Which are the factors that generate human mobility over the city? How can we use those factors to improve the attractiveness of the district?* The retail shops can be classified into two types: the primary and secondary stores (Jacobs, 1969). The primary store strongly attracts customers from other districts as well as nearby, and distributes them to the secondary stores. Although they are one of the most important triggers for people's mobility patterns, the analysis of the actual stores' attracting power and distribution power has received little attentions. For this purpose, we propose *Urban Association Rules* to analyze the customers' sequential movements in the city. We focus on large department stores, which are classified as primary and hub shops, and located in three different urban contexts. Within this research question, we shed some light on how customers' mobility patterns are generated around the primary store, considering the store's attracting power and distribution power.

We answered these questions by analyzing three different kinds of large scale datasets of human movements and their behaviors: (1) visitors' sequential movements between places and their length of stay at each place in a large art museum, (2) pedestrians' sequential movements between places and their length of stay in the historical center of Barcelona, and (3) customers' sequential movements between retail shops and the volume of their expenditure in each shop.

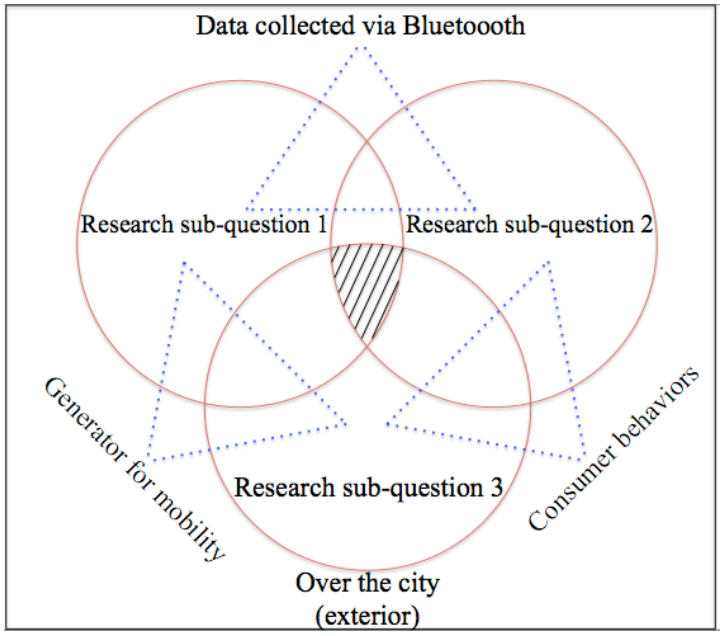


Figure 1.2. Three research sub-questions and their relationship

The three research sub-questions are strongly interrelated to each other, resulting in the consistency of this dissertation. Figure 1.2 presents a diagram of the three research sub-questions and the relationships between them. The first and second sub-questions enable us to uncover the factors which affect people's behaviors, but in different built environments: the former is implemented in an indoor space (a museum) and the latter is in exterior spaces (an urban district). In addition, the analytical conditions are similar, as we applied the same technologies for data collection in both cases, i.e, Bluetooth proximity detection techniques (Delafontain et al, 2012; Kostakos et al, 2010). By analyzing datasets collected via the same technology, we discover the similarity and dissimilarity of human movement and behaviors, in a significant better way than previous research, based on data collected via different data methods. Furthermore, in the first and third cases, we deal with the generator of the human mobility in architecture (inside building) and over the city (exterior space), respectively. The human movement is generated when people are attracted by the magnets (i.e., department stores, Mona Lisa), and when people are distributed into the surroundings (i.e., small stores, other artworks). In order to clarify this, we employed large scale credit card information provided by a Spanish bank (BBVA) for the third case.

We analyze how a store attracts people and distribute them into the retail shops in the district. This is a novel viewpoint with respect to conventionally conducted human mobility analyses: most researches deal with the human mobility itself, but our research uncovers it from the generator of such mobility. Finally, the second and third cases treat the consumers' mobility patterns in the exterior space, but at different scales. The former is analyzed at the district scale, while the latter does it over the city scale. In addition, the third research sub-question analyzes their mobility considering the purchase evidences through the transaction activities, while the second one does it without considering any spending activities, but during the discount period.

Thus, this dissertation analyzes human mobility from two sides: the mobility itself and the generator of such mobility, in the different built environments: interior and exterior spaces, and three different scales for the research target: building, district and city. By analyzing those three cases, we discuss how we can enhance knowledge of human behaviors through them, and how those scientific analyses can improve architecture and urban planning. In this way, these scientific analyses of human behaviors should contribute to improve the quality of architecture and urban planning adding completely different perspectives than those, which the architect has traditionally been using to design architecture. This is the science based software design process, rather than the intuition or art based hardware design.

Next, we will discuss further how these questions are framed and targeted within research areas, and which methodological approach is taken.

1.3 Research areas and some methodological aspects

This dissertation applies quantitative methods to collect human mobility data, which are analyzed with the purpose of spatial analysis in architecture. This dissertation is framed within the recent emergence of Big Data, and datification, i.e., many things around us can be measured and showed as quantitative data (Mayer-Schönberger & Cukier, 2013). New kinds of data, which could not

be quantified before, appear as quantifiable and can be analyzed now. The field of architecture and urban planning is one of the fields, which is potentially most influenced by this change.

Research methodology

The research methodology of this dissertation is based on a systematic observation in the framework of the "unobtrusive measures" (Webb et al. 2000), making use of digital footprints unconsciously left by people. This approach is an alternative to qualitative research conducted through traditional methods, but it can also be considered as complementary rather than exclusive of it. Unobtrusive measures through emerging technologies do not require manual effort, counting or human tracking. The sample size via those measures can be much larger than that based on manual effort. The techniques we use can provide a large scale dataset of human movement, on which quantitative research can be carried out.

Analytical framework

The analytical framework for this dissertation is largely based on association analysis. We analyze the correlations among the independent factors. This is different from uncovering the causality by relying on randomly chosen small-scale samples, which can be most likely best obtained from qualitative survey data. Although causality must be explored, the quantity from making a larger number of proposals is more useful than the quality of each one in the beginning of the planning process.

Datasets for this dissertation and privacy issues

This dissertation is based on datasets obtained by different technologies. Some of the datasets are based on a Bluetooth detection technique, other datasets are based on credit card information provided by the Spanish bank BBVA.

In order to avoid privacy issues for the former case, Bluetooth sensor features a preudonymization layer based on SHA (Secure Hash Algorithms) (Stallings W, 2011). This software layer has been

implemented in order to convert MAC addresses into pseudonyms, which enables us to generate a univocal link between a MAC address and its hash code. Because SHA encryption makes it unfeasible to retrieve the original MAC address of the device being tracked, it permits to generate anonymous data trajectories without invading privacy. For the latter case, the data is aggregated and hashed for anoymization by BBVA prior to sharing. This is done in accordance to all local privacy protection laws and regulations. As a result, "customers are identified by randomly generated IDs, connected with certain demographic characteristics and an indication of a residence location at the level of zip code for direct customers of BBVA and country of residence for all others" (Sobolevsky et al., 2014).

1.4 Results and their impact

Next we summarize the results and contributions of this dissertation to the areas mentioned before. The results in the context of inside architecture are:

- We discovered that the visiting style of short (less than 1:30 min) and long (more than 6 hours) are not as significantly different as one could expect. Both types of visitors tend to visit a similar number of key locations in the museum while the longer stay type visitors just tend to do so more extensively. We speculate that they could cause uneven distribution of the quantity of visitors, resulting in the congestion/vacancies in the museum spaces.

Our results help to uncover the spatial problems deriving from people's activities. This can be basic information for re-designing, renovating or rehabilitating architecture. As a concrete example of the Louvre Museum, the obtained results help planning to re-arrange the spatial layout together with the key artworks to be dispersedly, which make visitors' path not to be intersected during a whole day (Yoshimura et al., 2014). More generally, we can say that architect needs to consider not only the spatial factors such as the distribution of the number of visitors among spaces, but also the temporal factors of human behaviors for designing architecture. Visualization of visitors' flow helps to understand how they move

in a dynamic way considering the temporal factors. This could result in improving the quality of architecture's spatial design. The typical and traditional architects tend to consider the hardware (i.e., the building itself), while our vision focuses on the software (i.e., human activities).

Second, the results of the context of pedestrian behaviors are:

- We discovered that in a discount day pedestrians tend to stay a shorter period than in normal days, but in the former they visit a larger number of nodes than in the latter. In a discount day, pedestrians actively explore the district by visiting all nodes including those which are rarely visited by pedestrians in other normal days.

- We found, however, that in a discount day pedestrians tend to spend longer time in some streets than they do in other normal days, depending on the street they visit, and the direction of their walking.

- We also found that pedestrians' sequential movements between nodes have patterns in terms of the number of visited nodes and their order. Most pedestrians use a few path types, and most of path types are used only by a few pedestrians. This tendency is stronger in a discount day than in other normal days.

These findings show that collecting relevant datasets about pedestrians' behaviors, could be a basic information for use when regulating pedestrians' incoming and outgoing flow in a district. This would be helpful for architects in renovating and rehabilitating processes for a district. The rehabilitation of a district is a significant architect task for an architect, but this planning is normally conducted without the information we show that can be obtained. Perhaps rearranging the stores location in a district, pedestrians' behaviors could be altered in some ways (i.e., length of stay in the district), resulting in enhancing the quality of the district.

Third, the results of the context of the analysis of customers' behaviors are:

- We discovered that five of six groups from three selected focal shops present similar behaviors with regard to the category B1 (hotel, hostel B&B and restaurant, pub and café), but the corresponding LIFT values vary greatly. This means that there is a large difference in purchasing activities in this category between the focal shop's customers and people in the district where the focal shop is located.

- We also found that, in the district of the focal shop PA, 34.8% of existing shops are classified into the category A2 (groceries). However, people, who visited this district did not make many transactions in this category. In contrast, customers of shop PA were likely to make transactions in this category within the area. The fraction between the number of stores of category A2 and the number of transactions conducted in this category is 2.15. This reveals that the number of stores was almost two times greater than the number of transactions conducted in this category.

These findings are useful for urban planners and neighborhood associations in their efforts to economically develop regular areas, revitalize deteriorated districts or re-habilitate neighborhoods. For example, because the customers' origin and destination of the shop AD is known in terms of the shop category, coupons or royalty cards can be introduced between the relevant stores to increase the number of transactions in this district. Conversely, shop AD can provide similar coupons for their customers to use in stores of category B1 in other districts. Thus, city planners together with the commercial entities can attract or distribute more customers depending on the strategy of the district. This analysis is also useful for city authorities to determine whether they should allow or limit the opening of new stores of the same or different categories within the area. This makes the city more efficient and well balanced.

In a wider context, our contributions provide important insights for understanding human mobility in architecture and urban settings. Human mobility can be considered as a result of the interaction between pedestrians' intention to move, and the generator of such mobility, for instance the attractors, with parameters such as the attracting power and distribution power of stores dispersed over the city or the artworks inside architecture. This dissertation analyzes

human mobility from both those sides, which is not typical in the previous research. In addition, this dissertation contributes to enhance the knowledge about the pedestrian' mobility patterns in different scales: architecture, district and over the city. Although the built environments for the research are largely different from inside to outside, we can discover some similarities of the human behaviors, which can be the underlying patterns behind human mobility. Finally, the presented methodologies make it possible to generate the basic information for the pre-design process for architecture, and for renovation and rehabilitation for urban planning. They are based on the scientific analysis and its results, indicating that this dissertation shows that the scientific research could improve the quality of architecture and urban planning through integrating those methods in the process of designing architecture.

1.5 Organization of this dissertation

Five papers contribute to the body of this dissertation. The second chapter of this dissertation focuses on the analysis of visitors' behaviors in the Louvre Museum. We employed a Bluetooth detection technique to collect a large scale dataset about visitors' behaviors, and we analyzed them in order to uncover the behavioral differences between the shorter stay type and longer stay type. The chapter consists of the following two papers:

Yoshimura, Y. Girardin, F., Carrascal, J.P., Ratti, C, Blat J (2012), "New tools studing visitor behaviors in museums: a case study at the Louvre" in Information and Communication Technologies in Tourism 2012. Proceedings of the international conference in Helsingborg (ENTER 2012) Eds Fucks M, Ricci F, Cantoni L (Springer Wien New York, Morlenback) 391-402.

Yoshimura, Y. Sobolevsky, S, Ratti, C, Girardin, F, Carrascal J.P., Blat, J, Sinatra, R (2014), "An analysis of visitors' behavior in the Louvre Museum: A study using Bluetooth data" in Environment and Planning B: Planning and Design, 41 (6) 1113-1131.

The third chapter presents the analysis of the difference of pedestrians' behaviors between discount and normal days in the

historical center of Barcelona. We reveal the difference of pedestrian behaviors in terms of the number of visited nodes, its order and their length of stay in the district. This chapter consists of the paper:

Yoshimura, Y. Amini, A. Sobolevsky, S, Ratti, C (2015b). "An analysis of pedestrians behaviors through non-invasive Bluetooth monitoring" in Applied Geography (submitted).

The fourth chapter presents the analysis of customers' mobility patterns between retail shops, and the strength of attracting power and distribution power of retail shops in the city. We uncover the similarity/dissimilarity of spatial structures and their features of the district in Barcelona. We disclose the relationship between both types of activities (daily-use/non-daily use shop) through customers' mobility patterns. This chapter consists of two papers:

Yoshimura, Y. Amini, A. Sobolevsky, S, Blat, J, Ratti, C (2015c). "Analysis of customer' spatial distribution through transaction datasets" in the special issue of Springer's "Transactions on Large-Scale Data-and Knowledge Centered Systems" (printing).

Yoshimura, Y., Sobolevsky, S., Bautista Hobin, J N., Ratti, C., Blat, J (2015a). "Urban Association Rules: uncovering consumer behaviors in urban settings through Transaction data", Environment and Planning B (submitted).

References

Ahas R, Aasa A, Roose A, Mark U, Silm S, 2008, "Evaluating passive mobile positioning data for tourism surveys: An Estonian case study", Tourism Management 29(3): 469-486.

Alexander C, 1964, Notes on the Synthesis of Form, Harvard University Press.

Alexander C, 1977, A Pattern Language: Towns, Buildings, Construction, Oxford University Press.

Antonini G, Bierlaire M, Weber M, 2006, "Discrete choice model of pedestrian walking behavior", Transportation Research Part B: Methodological, 40 (8), 667-687.

Appleyard D, Gerson M S, Lintell M, 1981, Livable Streets, University of California Press.

Asakura Y, Iryo T, 2007, ”Analysis of tourist behaviour base don the tracking data collected using a mobile communication instrument”, Transportation Research Part A: Policy and Practice 41(7): 684-690.

Barabási A L, 2010, Bursts:The Hidden Patterns Behind Everything We Do, from Your E-mail to Bloody Crusades (Dutton Books, USA)

Bierlaire M, Robin T, 2009, “Pedestrians Choices” In: Timmermans H (ed), Pedestrian Behavior. Models, Data Collection and Applications, 1-26.

Borgers A W J, Timmermans H J P, 1986, “City center entry points, store location patterns and pedestrian route choice behaviour: A microlevel simulation model”, Socio-Economic Planning Sciences, 20 25-31.

Borgers A, Kemperman A, Timmermans H, 2009, “Modeling pedestrian movement in shopping street segments”. In: Timmermans H (ed.) Pedestrian Behavior: Models, Data Collection and Applications, Bingley, Yorks: Emerald 87-111.

Borgers A, Timmermans H J P, 1986, “A model of pedestrian route choice and demand for retail facilities within inner-city shopping areas”, Geographical Analysis, 18 (2) 115-128.

Borgers A, Timmermans H J P, 2005, “Modelling pedestrian behaviour in downtown shopping areas”, in Proceedings of CUPUM conference London CD-ROM paper 83, http://128.40.111.250/cupum/searchpapers/papers/paper83.pdf

Borgers A, Timmermans H, 2010, “A model of pedestrian route choice and demand for retail facilities within inner-city shopping areas”, Geograph. Anal. 18 (2) 115-128

Chalmers A F, 2003, What is this thing called science? Queensland University Press, Open University Press and Hackett Publishing Company.

Cherry E, 1998, Programming for Design: From Theory to Practice, Wiley.

Choi Y K, 1999, “The morphology of exploration and encounter in museum layouts” Environment and Planning B: Planning and Design 26(2) 241-250

Delafontaine M, Versichele M, Neutens T, Van de Weghe N, 2012, “Analysing spatiotemporal sequences in Bluetooth tracking data” Applied Geography 34 659-668

Demoulin N, Zidda P, 2008, “On the Impact of Loyalty Cards on Store Loyalty: Does the Customers’ Satisfaction with the Reward Scheme Matter?” Journal of Retailing and Consumer Services 15 (5) 386-98

Dijkstra J, Timmermans H, de Vries B, 2009, “Modeling Impulse and Non-Impulse Store Choice Processes in a Multi-Agent Simulation of Pedestrian Activity in Shopping Environment” segments”. In: Timmermans H (ed.) Pedestrian Behavior: Models, Data Collection and Applications, Bingley, Yorks: Emerald 63-85.

Dijkstra J, Timmermans H, Jessurun J, 2014, “Modeling planned and unplanned store visits within a framework for pedestrian movement simulation”, Transporation Research Procedia 2 559-566.

Dijkstra J, Timmermans, H J P, 2005, “Modelling behavioural aspects of agents in simulating pedestrian movement”, in Proceedings of CUMPUM conference London, CD-ROM paper 63, http://128.40.111.250/cupum/searchpapers/papers/paper63.pdf

Eagle N, Pentland A, 2005, “Reality mining: sensing complex social systems” Personal and Ubiquitous Computing 10(4) 255-268.

Ellard C, 2010, You Are Her: Why We Can Find Our Way to the Moon, but Get Lost in the Mall, Anchor.

Euler L, 1736, “Solutio problematis ad geometriam situs pertinentis”, COmmentarii Academiae Scientiarum Imperialis Petropolitanae 8 128-140 = Opera Omnia (1) 7 (1911-56), 1-10.

Flick U, 2009, An Introduction to Qualitative Research, SAGE Publications Ltd.

Forty A, 2001, Words and Buildings: A Vocabulary of Modern Architecture, Thames & Hudson.

Fruin J J, 1971, Pedestrian planning and design, New York: Metropolitan Association of Urban Designers and Environmental Planners.

Gehl J, 2011, Life Between Buildings: Using Public Space (Island Press)

González M C, Hidalgo C A, Barabási A L, 2008, “Understanding individual human mobility patterns” Nature 453 779-782

Hawelka B, Sitko I, Beinat E, Sobolevsky S, Kazakopoulos P, Ratti C, 2014, “Geo-located Twitter as proxy for global mobility patterns” Cartography and Geographic Information Science 41 (3) 260-271

Hein G, 1998, Learning in the Museum (Routledge, London)

Helbing D, Molnár P, Farkas I J, Bolay K, 2001, “Self-organizing pedestrian movement”, Environment and Planning B, 28, 327-341.

Helbing D, Molnár P, Farkas I J, Bolay K, 2001, “Self-organizing pedestrian movement”, Environment and Planning B: Planning and Design, 28 (3) 361-383.

Hillier B, 1996 Space is the Machine: a configurational theory of architecture (Cambridge University Press, Cambridge)

Hillier B, Hanson J, 1984 The Social logic of space (Cambridge University Press, Cambridge)

Hillier B, Penn A, Hanson J, Grajewski T, Xu J, 1993, “Natural movement: or, configuration and attraction in urban pedestrian movement”, Environment and Planning B: Planning and Design 20 29-66.

Hillier B, Tzortzi K, 2006, “Space Syntax: The Language of Museum Space”, in A Companion to Museum Studies Ed MacDonald S (Blackwell Publishing, London) 282-301

Hoteit S, Secci S, Sobolevsky S, Ratti C, Pujolle G, 2014, “Estimating human trajectories and hotspots through mobile phone data”, Computer Networks, 64 296-307

Hui S K, Bradlow E T, Fader P S, 2009, “Testing Behavioral Hypotheses Using an Integrated Model of Grocery Store Shopping Path and Purchase Behavior” Journal of Consumer Research 36 478-493

Jacobs J, 1961, The Death and Life of Great American Cities (Random House, New York)

Kanda T, Shiomi M, Perrin L, Nomura T, Ishiguro H, Hagita N, 2007, “Analysis of people trajectories with ubiquitous sensors in a science museum” Proceedings 2007 IEEE International Conference on Robotics and Automation (ICRA’07) 4846-4853

Kappraff J, 2001, Connections: The Geometric Bridge Between Art and Science, World Scientific Pub Co Inc.

Kostakos V, O'Neill E, Penn A, Roussos G, Papadongonas D, 2010, “Brief encounters: sensing, modelling and visualizing urban mobility and copresence networks” ACM Transactions on Computer Human Interaction 17(1) 1-38

Krumme C, Llorente A,Cebrian M, Pentland A (S), Moro E, 2013, “The predictability of consumer visitation patterns”, Sci. Rep. 3, 1645; DOI:10.1038/srep01645

Laroche,M.,Pons,F.,Zgolli,N.,Cervellon,M.-C.,Kim,C.,2003. A model of consumer response to two retail sales promotion techniques.J.Bus.Res.56,513–522.

Larson J, Bradlow E, Fader P, 2005, “An exploratory look at supermarket shopping paths” International Journal of Research in Marketing 22 (4) 395-414

Leenheer J, Tammo H A Bijmolt, 2008, “Which retailers adopt a loyalty program? An empirical study” Journal of Retailing and Consumer Services 15 429-442

Lynch K, 1960, The image of the city, Cambridge, MA: The MIT Press.

March L, Steadman P, 1971, The geometory of environment: An introduction to spatial organization in design, RIBA Publications Ltd.

Mayer-Schönberger V, Cukier K, 2013, Big Data: A Revolution That Will Transform How We Live, Work and Think (John Murray, London)

Mckercher B, Shoval N, Ng E, Birenboim A, 2012, “First and Repeat Visitor Behaviour: GPS Tracking and GIS Analysis in Hong Kong” Tourism Geographies 14 (1) 147-161

Nakamura, K., Zhao, H., Shibasaki, R., Sakamoto, K., Ohga, T., & Suzukawa, N. (2006). Tracking pedestrians using multiple single-row laser range scanners and its reliability evaluation. Systems and Computers in Japan, 37(7), 1–11.

Paldino S, Bojic I, Sobolevsky S, Ratti C, González M C, 2015, ¨Urban Magnetism Through The Lens of Geo-tagged Photography”, EPJ Data Science, 4(1), 1-17

Pan Y, Zinkhan G M, 2006, “Determinants of retail patronage: A meta-analytical perspective” Journal of Retailing 82 (3) 229-243

Peña A M, Parshall S A, 2001, Problem Seeking: An Architectural Programming Primer, Wiley.

Pine J B, Gilmore J B, 1999, “The Experience Economy”, Boston: Harvard Business School Press.

Porta S, Latora V, Wang F, Rueda S, Strano E, Scellato S, Cardillo A, Belli E, Càrdenas F, Cormenzana B, Latora L, 2012, “Street centrality and the location of economic activities in Barcelona” Urban Studies, 49 (7): 1471-1488

Porta S, Crucitti P, Latora V, 2006, ¨The network analysis of urban streets: a primal approach¨, Environment and Planning B Planning and Design, 33 (5) 705

Preiser W F E, Rabinowitz H Z, White E T, 1988, Post-Occupancy Evaluation, New York: Van Nostrand Reinhold.

Ratti C, 2004, “Space Syntax: some inconsistencies“, Environment and Planning B: Planning and Design 31 487-499.

Ratti C, Pulselli R, Williams S, Frenchman D, 2006, “Mobile Landscapes: using location data from cell phones for urban

analysis" Environment and Planning B: Planning and Design 33(5) 727-748

Reutterer T, Teller C, 2009, "Store format choice and shopping trip types" International Journal of Retail and Distribution Management 37 (8) 695-710

Rowe C, 1976, The Mathematics of the Ideal Villa, MIT Press.

Santi P, Resta G, Szell M, Sobolevsky S, Strogatz S H, Ratti C, 2014, "Quantifying the benefits of vehicle pooling with shareability networks", Proceedings of the National Academy of Sciences, 111 (37) 13290-13294

Schadschneider A, Klingsch W, Klüpfel H, Kretz T, Rogsch C, Seyfried A, "Evacuation Dynamics: Empirical Results, Modeling and Applications" In: Encyclopedia of Complexity and System Science, vol. 5 Springer, Berlin/Heidelberg, 2009, 3142-3176.

Shoval N, 2008, "The GPS Revolution in Spatial Research", Research in Urbanism Series; Vol 1: Urbanism on Track. Application of tracking technologies in urbanism; 15-21, TU Delft, Delft.

Shoval N, Issacson M, 2006, "Application of tracking technologies in the study of pedestrian spatial behavior", The Professional Geographer, 58, 172-183.

Shoval N, McKercher B, Birenboim A, Ng E, 2013, "The application of a sequence alignment method to the creation of typologies of tourist activity in time and space" Environment and Planning B: Planning and Design advance online publication, doi:10.1068/b38065

Sobolevsky S, Sitko I, Grauwin S, Tachet des Combes R, Hawelka B, Murillo Arias J, Ratti R, 2014, "Mining Urban Performance: Scale-Independent Classification of Cities Based on Individual Economic Transactions" arXiv:1405.4301

Sobolevsky S, Szell M, Campari R, Couronné T, Smoreda Z, Ratti R, 2013, "Delineating geographical regions with networks of human interactions in an extensive set of countires", PloS ONE 8(12), e81707

Stiny G, 1980, "Introduction to shape and shape grammars", Environment and Planning B 7 (3) 343-351.

Tan P N, Steinback M, Kumar V, 2005, Introduction to Data Mining (Addison Wesley)

Terzidis K, 2006, Algorithmic Architecture, Routledge.

Timmemans H, 2009, Pedestrian Behavior: Models, Data Collection and Applications, Bingley, Yorks: Emerald.

Timmermans H, 1996, “A stated choice model of sequential mode and destination choice behaviour for shopping trips”, Environment and Planning A, 28 173-184.

Trondle M, Greenwood S, Kirchberg V, Tschacher W, 2014, “An Integrative and Comprehensive Methodology for Studying Aesthetic Experience in the Field: Merging Movement Tracking, Physiology, and Psychological Data“, Environment and Behavior 46 (1) 102-135

Tschacher, W., Greenwood, S., Kirchberg, V., Wintzerith, S., van den Berg, K., & Tröndle, M. (2012). Physiological correlates of aesthetic perception in a museum. Psychology of Aesthetics, Creativity, and the Arts, 6, 96-103.

Turner A, Doxa M, O’Sullivan D, Penn A, 2001, “From isovists to visibility graphs: a methodology for the analysis of architectural space”, Environment and Planning B: Planning and Design 28, 103-121.

Turner A, Penn A, 2002, “Encoding natural movement as an agent-based system: an investigation into human pedestrian behaviour in the built environment”, Environment and Planning B: Planning and Design 29 473-490.

Versichele M, Neutens T, Delafontaine M, Van de Weghe N, 2011, “The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: a case study of the Ghent festivities” Applied Geography 32 208-220

Webb E J, Campbell D T, Schwartz R D, Sechrest L, 2000, Unobtrusive measures: revised edition, Thousand Oaks: Saga Publications Inc.

Weiner E, 1997, Urban Transpotation Planning in the United States: A historical Overview, Fifth Edition, DOT-T-97-24, Technology Sharing Program, U.S. Department of Transportation, Washington, D.C., 1997

Yoshimura Y, Girardin F, Carrascal J P, Ratti C, Blat J, 2012, “New Tools for Studing Visitor Behaviours in Museums: A Case Study at the Louvre” in Information and Communication Technologis in Tourism 2012. Proceedings of the International conference in Helsingborg (ENTER 2012) Eds Fucks M, Ricci F, Cantoni L (Springer Wien New York, Mörlenback) 391-402

Yoshimura Y, Sobolevsky S, Ratti C, Girardin F, Carrascal J P, Blat J, Sinatra R, 2014, “An analysis of visitors’ behaviour in The Louvre Museum: a study using Bluetooth data” Environment and Planning B: Planning and Design 41 (6) 1113-1131

Yoshimura, Y., Sobolevsky, S., Bautista Hobin, J N., Ratti, C., Blat, J (2015a). “Urban Association Rules: uncovering consumer behaviors in urban settings through Transaction data”, Environment and Planning B (submitted).

Yoshimura, Y. Amini, A. Sobolevsky, S, Blat, J, Ratti, C (2015b). "Analysis of pedestrians behaviors through non-invasive Bluetooth monitoring" in Applied Geography (submitted).

Yoshimura, Y. Amini, A. Sobolevsky, S, Blat, J, Ratti, C (2015c). "Analysis of customer' spatial distribution through transaction datasets" in the special issue of Springer's "Transactions on Large-Scale Data-and Knowledge Centered Systems".

Zhu W, Timmermans H, 2008, "Cut-off models for the 'go-home' decision of pedestrians in shopping streets", Environment and Planning B: Planning and Design, 35 248-260.

2. Visitors studies in the Louvre Museum

The next chapter describes analysis of visitors' behaviors in the Louvre Museum. We employed Bluetooth detection technique in order to collect a large-scale datasets of visitors' sequential movement and their length of stay. The unprecedented datasets enables us to uncover unknown aspects of visitors' behaviors in a large-scale museum. This chapter consists of these two papers:

Yoshimura, Y. Girardin, F., Carrascal, J.P., Ratti, C, Blat J (2012), "New tools studing visitor behaviors in museums: a case study at the Louvre" in Information and Communication Technologies in Tourism 2012. Proceedings of the international conference in Helsingborg (ENTER 2012) Eds Fucks M, Ricci F, Cantoni L (Springer Wien New York, Morlenback) 391-402.

Yoshimura, Y. Sobolevsky, S, Ratti, C, Girardin, F, Carrascal J.P., Blat, J, Sinatra, R (2014), "An analysis of visitors' behavior in the Louvre Museum: A study using Bluetooth data" in Environment and Planning B: Planning and Design, 41 (6) 1113-1131.

2.1 New tools studying visitor behaviors in museums: a case study at the Louvre

Yuji Yoshimura, Universitat Pompeu Fabra
Fabien Girardin, Lift Lab
Juan Pablo Carrascal, Universitat Pompeu Fabra
Carlo Ratti, MIT SENSEable City Lab
Josep Blat, Universitat Pompeu Fabra

Abstract

In this paper we discuss the exploitation of data originated from Bluetooth-enabled devices to understand visitor's behaviour in the Louvre museum in Paris, France. The collected samples are analysed to examine frequent patterns in visitor's behaviours, their trajectory, length of stay and some relationships, offering new details on behaviour than previously available. Our work reinforces the emergence of a new methodology to study visitors. It is part of recent lines of investigation that exploit the presence of pervasive data networks to complement more traditional methods in tourism studies, such as surveys based on observation or interviews. However, most past experiments have explored quantitative data coming from mobile phones, GPS, or even geotagged user generated content to understand behaviour in a region, or a city, at a larger scale than that of our current work.

Keywords: Bluetooth sensing; human behavior; museum study; real time management tool

2.1.1. Introduction

In recent decades, tourism has developed to become one of the biggest industries. The World Tourism Organization foresees that the number of tourists will reach 1,600 million around 2020 and the World Travel and Tourism Council predicts that direct/indirect economic impact generated by the touristic industry will amount to

9.6% of Gross Domestic Product (GDP) and generate 9.7% of employment all over the world in 20121.

The increase of its economical, cultural and social impact on urban areas requires more precise and dynamic understanding of tourist behaviours and movements at micro (e.g. district, city) and macro (e.g. region, country) scales. Some emerging technologies make it possible to record and analyse them at city and district level (e.g. GPS, mobile phones with or without GPS (Asakura & Iryob, 2007); the passive mobile positioning data (Ratti, Pulselli, Williams, & Frenchman, 2006, Ahas, Aasa, Roose, Mark, & Silm, 2008); user-generated data (Girardin, Dal Fiore, Ratti, & Blat, 2008, Pereira, Vaccari, Girardin, Chiu, & Ratti, 2012, Girardin, Calabrese, Dal Fiore, Ratti, & Blat, 2008). In museums, the observation, and interview-based surveys have been used mostly to understand the social use of the environment and evaluate its use (see Hooper-Greenhill, 2006 for a review and Yalowitz & Bronnenkant, 2009, Hillier & Tzortzi, 2006). The information collected by these "traditional" methods provides support for the management of the spaces, which have a proved value. However they often provide a snapshot on the life of a built environment, and the interviews and questionnaires can have self-reporting bias. Moreover, they fail to record empirical evidences and measures (e.g. visiting time, sequences of visits, time of stay, density) key to produce a more complete picture on people use of a space.

The purpose of this paper is to discuss a Bluetooth proximity detection approach (previously developed for a traffic data collection system (Sanfeliu, Llácer, Gramunt, Punsola, & Yoshimura, 2010) to gather insights on visiting behaviours in a museum context and to demonstrate its relevance to support the management of environments that must respond to the increasing tourism demand. For instance, the analysis reveals the dynamic description of different use of museum spaces, the visiting profiles and the spatio-temporal patterns of visitors' behaviours.

In section 2 a brief summary of related works, their contributions and their main limitations is provided. In section 3 a Bluetooth proximity approach to detect visitor's presence and sequential

[1] http://www.wttc.org/eng/Tourism_Research/Economic_Research/

movements is proposed. In section 4 the dataset processing for the analysis is discussed and key concepts for our research introduced. Section 5 presents some initial findings from our field trials, the frequent pattern and visitors' spatial uses. Finally, we summarize our on-going work on developing methods and tools for analysing the museum and urban environments.

2.1.2. Strategies to collect empirical visitor data

With the emergence of location technologies, a variety of methodologies have been proposed to locate a person, specifically for the collection of empirical data in the context of tourism (Asakura & Iryo, 2007, Ratti et al., 2006, Ahas et al., 2008, Girardin, Calabrese et al., 2008, Girardin, Dal Fiore et al., 2008, Yalowitz & Bronnenkant, 2009, Hillier & Tzortzi, 2006, Kanda et al., 2007). They are classified into 3 groups, and remark the burden which each method imposes on the persons involved.

The first group of more traditional techniques includes observation, and interviews. With the latter or with user diaries, one can obtain specimens of detailed visitor's behaviour. However, the data can be subjectively biased and the methods are costly requiring a lot of human resources (Girardin, Dillenbourg, & Nova, 2009). Something similar happens with direct observation, which could be difficult to sustain for long periods as it poses a heavy burden to the observer. The representativeness of the sample in interviews and questionnaires can be an issue too.

The second group is based on technologies such as GPS or RFID, which can supply more objective and precise time, location and route data (Asakura & Iryob, 2007, Kanda et al., 2007) than the traditional methods – however, without the motivations which can appear in interviews. Currently, these techniques demand the users to carry specifically enhanced devices that are not widespread. They make the data collection more cumbersome, and they may bias the user behaviour, and thus the collected data.

A third group includes using image sensing devices or passive mobile positioning data, which give little burden to the users – but no motivations are available either. Their main limitation is spatial;

for instance, image sensing devices can record visitor's behaviour with spatio-temporal accuracy (Antonini, Bierlaire, & Weber, 2006), but the recording area covered by a single camera is limited (Yalowitz & Bronnenkant, 2009). Passive mobile positioning data have started to be used in tourism studies (Ratti et al., 2006, Ahas et al., 2008) as it can provide better empirical data on movement of people at global scale; nevertheless, the estimation of the presence and movement of people is limited by the cell size (i.e. the area of coverage of the base station that serves the mobile service).

Table 2.1.1. Data capture techniques showing their main strengths and weaknesses in the context of tourism and urbanism studies

Data capture	Strengths	Weaknesses	Application example
Manual surveys	Capture motivations	Very costly and applied to a limited time period	Timing and Tracking Survey (Yalowitz & Bronnenkant, 2009)
GPS and Cell phone (device-based)	Timely mobility data (potentially augmented with in-situ survey)	Survey limited in time and participants. It does not work inside the buildings	Describe social and spatial characteristics with limited samples (Asakura & Iryob, 2007)
RFID	Precise real-time mobility data	Survey limited in time and participants. Infrastructure deployment needed	Describe social and spatial characteristics with limited samples (Kanda et al., 2007)
Cell phone (aggregated network-based)	Use existing infrastructure to provide real-time mobility data	Does not work at the building and room scale	Real-time urban dynamics (Ratti et al., 2006)
Bluetooth detection	Precise real-time mobility data, non-intrusive to participants	Infrastructure deployment needed	Describe social and spatial characteristics (Kostakos, O'Neill, Penn, Roussos, & Papadongonas, 2010)

This paper presents several contributions in the development of data collection tools and methodologies for the analysis of large samples describing visitor's behaviour at small spatial scale using Bluetooth. The recent wide spread of mobile devices implies that many people have their Bluetooth switched on passively, thus providing an important source of useful data. A variety of projects have exploited Bluetooth data for measuring the social network relationships

between people (Eagle & Pentland, 2005, Paulos & Goodman, 2004, Nicolai, Yoneki, Behrens, & Kenn, 2006), mobility of vehicles (Yalowitz & Bronnenkant, 2009, Barceló, Montero, Marqués, & Carmona, 2010) and mobility of pedestrians and their relationships (O'Neill et al., 2006, Kostakos et al., 2010). However these investigations have not considered a specific analysis of pedestrians and their use of space. This paper aims at reducing this shortcoming.

2.1.3. Data collection settings

A large majority of mobile devices currently on the market embed Bluetooth, and a significant proportion of users have them turned on in passive mode (Kostakos et al., 2010). The presence of these Bluetooth-enabled devices can be detected by means of sensors that scan the wireless spectrum. This section and the following describe the settings of our study and how the collected data were structured to handle privacy issues and allow for the spatial behavioural analysis.

2.1.3.1 Context of the study

The Louvre is the most visited museum in the world with 8.5 million visitors in 2009 and more than 40,000 visitors at peak days2. This context of "cultural enthusiasm" has direct consequences on the quality of the visitor experience as well as on the organization and management of the Museum (e.g. application of flow management strategies and increased stress level of the surveillance staff). In response to the increasing tourism demand and the necessity to setup and evaluate museum strategies, we proposed to collect and analyse empirical data on the flows and occupancy levels of visitors in key areas of the Louvre.

2.1.3.2 Study settings and characteristics of the Bluetooth sensors

Because of the context, our study is particularly focused on one of the busiest areas of the museum identified by Le Louvre officials, namely a trajectory that leads visitors from the entrance (Pyramid)

[2] http://www.theartnewspaper.com/attfig/attfig10.pdf

to the Venus de Milo. 10 Bluetooth sensors were deployed and they were sufficient to gather measures of visiting sequences and staying times at representative locations along the path (Figure 2.2.1). Two were on floor -1 (0 or Hall, 1 or Denon access); five were on floor 0 (2 or Denon 0, 3 or Samothrace 0, 4 or Venus de Milo, o 5 or Caryatides, 6 or Sphinx), and 3 on floor 1 (7 or Big Gallery, 8 or Samothrace 1, 9 or Glass).
The sensors gathered a unique encrypted identifier distinguishing each mobile device that supports Bluetooth and is set to be discoverable, as well as 2 time stamps for check-in and check-out times within the range of each sensor. Assuming that a mobile device belongs to a person, the movement of the device can be related to that of the visitor.

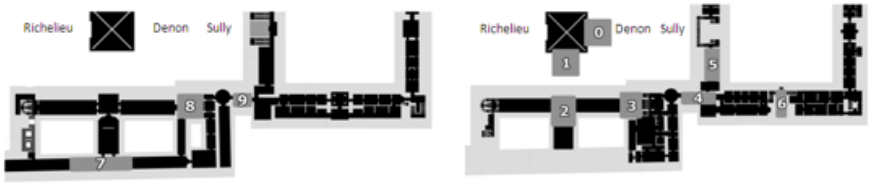


Fig. 2.1.1. Location of 10 sensors (No.0-No.9) indicating their approximate sensing range

The administrative and technical restrictions (e.g. protection against robbery, areas unreachable to visitors, no sources of electrical power, safety and health concerns) guided the deployment of the devices, sometimes preventing the installation in ideal locations for optimal detection. These special circumstances required the use of an ad-hoc battery with 10 days of autonomy for each sensor. This temporal limitation constrained the accumulation of empirical data, but our analyses show that this period is sufficient to extract relevant evidences.

2.1.3.3 Data and privacy issues

Based on a previous research that focused on the privacy issues related to the use of Bluetooth scanners (Sanfeliu et al., 2010), we adopted a solution that 1) does not allow the identification of individuals, 2) keeps the anonymity of trajectory data even after recording and archiving. This is achieved with the application of Secure Hash Algorithm (SHA) to the Bluetooth unique IDs detected by our system.

2.1.3.4 Sensor detectable area and the definition of its node

The spatial definition of the detectable area by a Bluetooth sensor is a critical issue for any research which uses this type of sensors. The shape of the area is similar to a flower with four petals of different length and width. In an optimal setting, the largest petal is an ellipse of almost 40 meters long by 15 meters wide, while the smallest is approximately 15 by 10 meters. The other two have a similar shape and a size of 15 by 10 meters. However, it could be customized for an indoors space with the largest petal dimensions being 20 by 7-8 meters. We identify the area detectable by a sensor as a node which represents the corresponding location, and use this definition through the rest of the paper. The detectable area estimations fluctuate according to the museum settings, due to the location of the sensors (e.g. within wooden boxes or administrative desks) and other factors, but we made sure that they would cover the targeted areas along the studied visitor trail.

Fig. 2.1.2. Conceptual diagram of Bluetooth sensor's detectable area

2.1.4. Collected data and measures

Based on the methodologies proposed in previous sections, the sample data during a specific audit period (more on this in section 4.2) was collected. This section describe how these amounts of collected sample data were organized to extract the desired values, and two measure concepts, length of stay and trajectory are defined.

2.1.4.1 Database

The raw dataset collected from all sensors is huge and requires pre-processing for us to be able to extract meaningful information from it. These data basically consist of a unique encrypted identifier for every mobile device and two timestamps, which correspond to the first and last times such a device has been detected by the sensor. Then a database and a query engine were built to reorganize these data for our analyses (see table 2 for an example). Let us indicate some of the tags used for organizing and analysing the raw data. Rffr is the unique encrypted identifier. Date is the year, month and day when the data were collected. Path indicates the nodes that a mobile device has visited and it is represented by a sequence of node numbers, from 0 to 9, separated by a colon (":"). Nodes represents the total number of nodes that a device has visited during its whole trajectory, while distinct nodes indicates the number of different nodes which a mobile device has visited. Checkin is the moment when the signal of a mobile device is first detected in the museum (i.e. at the first node of the trajectory) and checkout is the moment when it disappears from the last node (i.e., when the device has left the museum). Staylength is the time difference between Checkout and Checkin and it represents the total duration of stay of a mobile device inside the museum.

Table 2.1.2. Example of the dataset.

Rffr	Date	Path	Distinct nodes	Nodes	Checkin	Checkout	Staylength
Unique ID	2010-04-30	0:3:8:7:0	4	5	09:04:35	11:07:52	02:03:17

2.1.4.2 Collected sample

A high frequentation 10-day period in May 2010 was selected to perform a first analysis of visitor's behaviour. During this audit period, our installation recorded the presence of 12,944 unique devices. Through the data cleaning process we removed the logs from security and museum staff by looking at their recurrence, and the time of their presence (e.g. outside visiting times). Also, it was found out that the logs from two sensors had erroneous time synchronization and had to be discarded. Indeed, synchronization is a key element of our approach, for instance to infer the sequence of visit.

2.1.4.3 Measures definition

A sensor log reveals the visitor's presence at a node: once a Bluetooth-enabled mobile device enters the detectable area, the sensor continues to receive the signal emitted from the device until it disappears from its range. Each sensor records the first time the device appears as a check-in time at then records the time when the signal of the device disappears, as the checkout time. The difference between both time stamps is the length of the stay at the node. If the nodes visited are ordered by time, and then the checkin time at the first node in the trajectory and the checkout time at the last are selected, the values of the total duration of the visit to the museum will be obtained. As it can be seen, synchronization of sensors plays a key role for the collected data to be meaningful.

On the other hand when a unique Bluetooth identifier is logged out with time stamp (t1) by sensor A and some time later is logged in with time stamp (t2) by sensor B, the difference between t2 and t1 measures the travel time. The sequential movement of a mobile device detected by a pair of sensors (e.g. A-B) is defined as a trajectory with the travel time of t2-t1 minutes. The concept of trajectory in our research is different from that obtained with GPS systems, which indicate precise locations. The trajectories are obtained from Bluetooth detection through the time stamped sequential transition of a mobile device detected through different nodes (e.g. sequence of A-B-D), while GPS can describe the precise movement of the device. Our measures are, in this way, indirect ones.

2.1.5. Results

Using the Bluetooth data and concepts described previously, a novel approach is developed to analyse the spatial use in the Louvre museum. The following subsections present the on-going analysis efforts built around these concepts and the initial findings in order to obtain indicators for crowd management and to extract the frequent patterns in visitors' behaviour. In 5-1 the representativeness of the sample captured by our sensors is discussed. In 5-2 an analysis of the use of the Pyramid space related to the visitors' trajectories, which may reveal the distribution of

visitors' presence and its basic flow in the museum, as all the visitors use the Pyramid as entrance and exit is presented. In 5-3 an analysis of visitors' trajectories and the time spent in each route is presented, revealing the existence of frequent patterns according to these two parameters.

2.1.5.1 Representativeness of the collected sample

Only a part of the visitors have got devices with Bluetooth, and only a part of them are enabled so that they are detectable. The number of devices detected at the entrance is compared with the official museum head counts and ticket sales to understand the representativeness of our Bluetooth data. The sample represented between 5.9% and 8.7% of the visitors with a strong, positive correlation of +80%, providing support for its representativeness. Figure 3 shows the linear regression fit of the numbers of detected devices and official counts of each day, with data of 101 days.

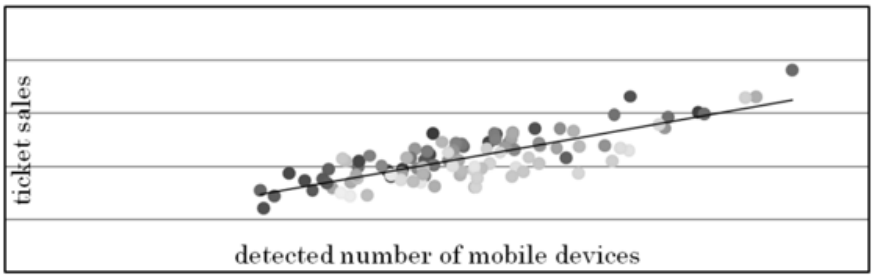


Fig. 2.1.3. Correlation between detected devices and visitors estimations per day

2.1.5.2 Use of the Pyramid space

The Pyramid space serves for distributing the visitors through three museum accesses, named Denon, Sully and Richelieu. As it is the starting point for almost all the museum visitors, it is important to identify and analyse their spatial use in order to devise more efficient and flexible policies for the museum. Since all the sensors are installed on the trails that lead to the Venus de Milo, along the Denon area, mobile devices detected by sensors in nodes 0 to 8 represent people who visited such area. These are the data we deal with in this section.

Sensor 0 is the Hall, and due to the museum's spatial layout, routes 0-3, 0-4, 0-7 and 0-8 mean that the Denon access has been used; routes 0-5, 0-6 and 0-9 mean use of either Sully or Richelieu access; route 0-0 means that only the Sully or Richelieu areas were visited (see Fig.4). Our data indicate that 76% of visitors used the Denon access while only 23% used either of the other two.

However, if one focuses on visitor's exit behaviour, i.e. moving towards node 0 or the Hall, the spatial use tendency changes. The most used route leading to node 0 was the 3-0 (25%), followed by the 7-0 (around 17%) and the 5-0 (around 16%); around 40% of the visitors left the museum through the Sully or Richelieu access, while 60% used Denon as their exit route, which means a decrease of an absolute 16% of the latter.

2.1.5.3 visitor's trajectories

In this subsection, visitors' trajectories, their average length of stay and their relationships are analysed to discover frequent patterns or trends in visitors' behaviours. Sequential pattern mining (Agrawal & Srikant, 1995) has received much attention in the recent decade to find frequent sequences of events in data with a temporal component, with transition probabilities between events. However, extracting meaningful patterns requires appropriate algorithms and parameters. In the following, the grounding work and initial findings of our analysis are presented.

Most used trajectory and visitor's transition rates between nodes. Clarifying visitors' most used trajectory and their nodes transition rates helps to uncover hidden rules behind the seemingly disordered dataset. The data correspond to the route starting by 0-8, which is used by 60.6% of the visitors (7721 devices). Within the people who took this route, 71.1% moved to node 3 (0-8-3), while 13.5% went back to node 0 (0-8-0); 7.2% do not have any further records after node 8 (0-8), and it is assumed that they should have finished their itineraries without being registered by node 0 again. This is because if they would have continued visiting the museum, some of the other sensors should have detected them. Thus, 20.8% of these visitors (12.6% of the total) came to the museum just to take the 0-8-0 route. The popularity of this route might be due to the presence of two major works, Mona Lisa, located between nodes 7 and 8, and

Winged Victory of Samothrace (in node 8). Moreover, the spatial structure of the museum strengthens the link between those two works, and thus an important sequential pattern including a visit to the Mona Lisa, followed by the Winged Victory of Samothrace, and then the Italian Gallery might exist. Let us perform a more detailed analysis considering the objects in the spatial structure, to reveal patterns.

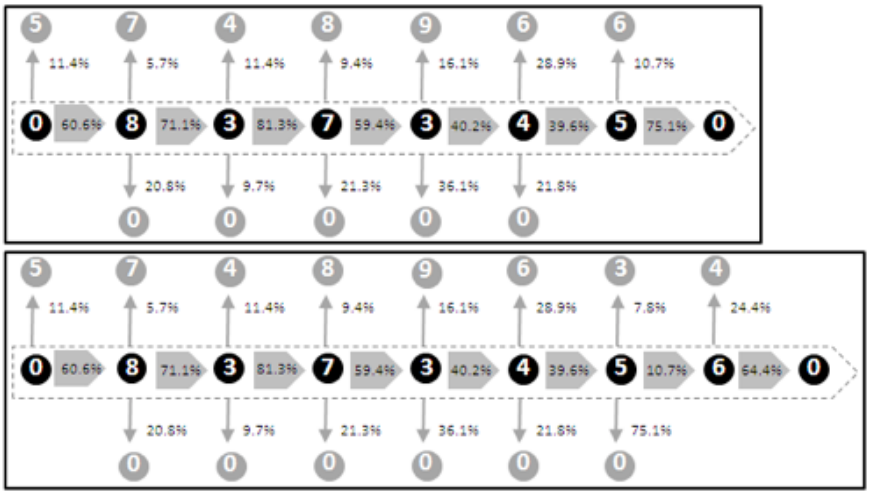


Fig. 2.1.4. Diagram of nodes and percentages of visitors moving between them. Above, the most used trajectory is shown, below, the second most used one.

The visitors' distribution rate from each node to every subsequent node is iteratively computed until the route finished at 0. The percentages that appear with each arrow (Fig.2.1.4) are these transition rates from each node to the next ones. This makes it easier to understand quantitatively the visitors' flows. For instance, the upper part of fig.2.1.4 shows the most used trajectory (0-8-3-7-3-4-5-0) while the lower one shows the second one (0-8-3-7-3-4-5-6-0) both with thick arrows. Thin arrows express the second or third higher transition rates from a node to the following ones.

Various findings from the diagram can be extracted, but the most visible outcome is the strong connection between nodes 8 and 3, and between nodes 3 and 7. While 71.1% of visitors moved to node 3 after visiting node 8, 81.3% of them went to node 7 after visiting node 3. Concerning node 3, the analysis shows that people tend to

use it to make a change of direction as it is in the same way from the Pyramid space. In a similar way as the Pyramid space, which distributes people for three accesses, node 3 also serves for distributing visitors to other places.

All of these findings demonstrate that our methodologies can reveal unknown aspects of visitors' spatial use, which observation and traditional interview-based approaches could not clarify at small scale in spatiotemporal terms.

Relationship between length of stay and number of visited nodes. Next, the relationship between the average stay length and the number of visited nodes of each trajectory is analysed, as it can provide another pattern of visitors' behaviour and spatial use in the museum.

Table 2.1.3. Average length of stay corresponding to each trajectory

Trajectory	Average length of stay
0-8-0	2:46:27
0-8-3-0	2:49:11
0-8-3-7-0	2:45:21
0-8-3-7-3-0	2:35:17
0-8-3-7-3-4-0	2:19:58
0-8-3-7-3-4-5-0 (most used trajectory)	2:19:14

The regression line of the number of nodes in a trajectory versus the average staying time has a negative slope, meaning that the larger the number of nodes in the trajectory, the shorter the visit to the museum lasts (with an r2 value of 0.85) – which opposes to the most obvious assumption.

2.1.6. Discussion, conclusions and future work

The paper shows, through different analyses, that Bluetooth data can throw new light on spatial use and visitors' behaviours at the building scale. Namely, evidence on the use of the different accesses as the entrance route by visitors and an exit pattern, which is different from the entrance pattern, has been given. This evidence would be difficult and costly to obtain from observations and surveys. Again, these traditional methods would have had difficulties to offer estimates of percentages of visitors that have

followed different trajectories, to detect the importance of the 0-8-0 route, or to easily detect, from raw data, the role of node 3 as a crossroads. Another example of the power of the simple analysis on the data is the inverse relation found between length of stay and number of visited nodes.

These initial findings suggest that the methodology proposed has a great potential to clarify the features of the space and its use by visitors in small spatiotemporal scales with unprecedented accuracy. For example, comparing the results of audits of different periods would offer the possibility of obtaining results of seasonal nature; analysing audits of a large amount of data collected, finer detail of patterns and relationships could be obtained – beyond the crude relationship of average visit length and number of nodes. Collecting audits during longer periods requires only very small extra empirical effort besides the one already described; and the analysis would only mean extending and refining the analytical tools.

Before discussing more details of our current work, let discuss some aspects of the data obtained. As seen, the data correspond to a small sample of the visitors – although very large compared to the typical sample used in surveys, and without subjective bias -, but seems to be reasonably valid in terms of the correlation shown. However, more work should be done to clarify the extent to which the sample is representative, as carrying a Bluetooth device set as enabled might be a significant bias. With respect to other data collection strategies involving users carrying specific devices and consequently being aware of them, data appear to be free from potential bias, and the dataset obtained is larger.

Secondly, the data obtained are usually noisy; however, examples of strategies for checking data consistency from the data themselves, and for cleaning it have been given. Larger audits, which would offer larger datasets, can help to strengthen this aspect.

Let turn now to current and future work. Exploring some of the aspects mentioned before based on larger audits of the Louvre has started. And based on the current results, indicators for crowd management and an algorithm for sequential mining are being developed to discover frequent patterns and the underlying association rules. The dynamic estimation of the density and flow of

visitors in and between nodes could be associated with the indicator of the relation between pedestrian flow and its density (Seyfried, Steffen, Klingsch, & Boltes, 2005) for more dynamic crowd management. While several attempts have been made to extract meaningful frequent trajectory patterns and predict further movements of objects at a variety of scales from region and city (Giannotti, Nanni, Pedreschi, & Pinelli, 2007) to retail shop (Larson, Bradlow, & Fader, 2005), improved mining techniques and parameter settings depending on the nature of the data would be needed in order to achieve these goals. Explore similar pedestrian data collected in unconstrained environments has started, and it should help to substantiate the previous statement.

As a final point, one should remark that the understanding of the patterns in visitors' behaviour and its prediction will enable to optimize the spatial layout of objects, human resources and facilities, including advertising and visitor information points, to respond to the increasing tourism demand. It could become a strong management tool not only for museums but also for urban environments in the tourism flourishing age.

References

Agrawal, R. & Srikant, R. (1995). Mining sequential patterns. In S. Yu. Philip, Arbee L.P. Chen (Eds.), ICDE '95: Proceedings of the Eleventh International Conference on Data Engineering, Washington DC. IEEE Computer Society.

Ahas, R., Aasa, A., Roose, A, Mark, U. & Silm, S. (2008) Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. Tourism Management 29(3): 469-486.

Antonini, G., Bierlaire, M. & Weber, M. (2006). Discrete choice models of pedestrian walking behaviour. Transportation Research Part B Methodological 40(8): 667-687.

Asakura, Y. & Iryob, T. (2007). Analysis of tourist behaviour based on the tracking data collected using a mobile communication instrument. Transportation Research Part A: Policy and Practice 41(7): 684-690.

Barceló, J., Montero, L., Marqués, L. & Carmona, C. (2010). Travel Time Forecasting and Dynamic Origin-Destination Estimation for Freeways Based on Bluetooth Traffic Monitoring. Transportation Research Record: Journal of the Transportation Research Board 2175: 19-27.

Eagle, N. & Pentland, A. (2005) Reality mining: sensing complex social systems. Personal and Ubiquitous Computing 10(4):255-268.

Giannotti, F., Nanni, M., Pedreschi, D. & Pinelli, F. (2007). Trajectory Pattern Mining. Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining KDD 07, San Jose: 330-339.

Girardin, F., Calabrese, F., Dal Fiore, F., Ratti, C. & Blat, J. (2008). Digital footprinting: Uncovering tourists with user-generated content. Pervasive Computing, IEEE 7(4): 36-43.

Girardin, F., Dal Fiore, F., Ratti, C., & Blat, J. (2008). Leveraging explicitly disclosed location information to understand tourist dynamics: a case study. Journal of Location Based Services 2(1): 41-56.

Girardin, F., Dillenbourg, P. & Nova, N. (2009). Detecting air travel to survey passengers on a worldwide scale. Journal of Location Based Services 3(3): 210-226.

Hillier, B. & Tzortzi, K. (2006). Space Syntax: The Language of Museum Space. In S. MacDonald, (Ed.), A Companion to Museum Studies, London, Blackwell Publishing.

Hooper-Greenhill, E. (2006). Studying visitors. In S. MacDonald (Ed.), A Companion to Museum Studies, London, Blackwell Publishing.

Kanda, T., Shiomi, M., Perrin, L., Nomura,T., Ishiguro, H. & Hagita, N. (2007). Analysis of people trajectories with ubiquitous sensors in a science museum. Proceedings 2007 IEEE International Conference on Robotics and Automation (ICRA'07): 4846-4853.

Kostakos, V., O'Neill, E., Penn, A., Roussos, G. & Papadongonas, D. (2010). Brief encounters: sensing, modelling and visualizing urban mobility and copresence networks. ACM Transactions on Computer Human Interaction 17(1): 1-38.

Larson, J.S., Bradlow, E.T. & Fader, P.S. (2005). An exploratory look at supermarket shopping paths. International Journal of Research in Marketing, 22: 395-414.

Nicolai, T., Yoneki, E., Behrens, N. & Kenn, H. (2006). Exploring social context with the Wireless Rope. In R. Meersman, Z. Tari, & P. Herrero, (Eds.), On the move to meaningful internet systems 2006: LNCS, vol 4277, Heidelberg: Springer.

O'Neill, E., Kostakos, V., Kindberg, T., Fatah gen. Schieck, A., Penn, A., Stanton Fraser,D. & Jones,T. (2006). Instrumenting the city: Developing methods for observing and understanding the digital cityscape. In P. Dourish and A. Friday, (Eds.), Ubicomp 2006, LNCS 4206: 315-332.

Paulos, E. & Goodman, E. (2004). The familiar stranger: anxiety, comfort, and play in public places. Proceedings of the SIGCHI conference on Human factors in computing systems 6(1): 223-230.

Pereira, F. C., Vaccari, A., Giardin, F., Chiu, C., & Ratti, C. (2012) Crowdsensing in the web: analysing the citizen experience in the urban space. In M. Foth, L. Forlano, C. Satchell, and M. Gibbs (Eds), From Social Butterfly to Engaged Citizen: Urban Informatics, Social Media, Ubiquitous Computing, and Mobile Technology to Support Citizen Engagement, Cambridge (MA): MIT Press.

Ratti, C., Pulselli, R., Williams, S. & Frenchman, D. (2006). Mobile Landscapes: using location data from cell phones for urban analysis. Environment and Planning B: Planning and Design 33(5): 727-748.

Sanfeliu, A., Llácer, M.R., Gramunt, M.D., Punsola, A. & Yoshimura, Y. (2010). Influence of the privacy issue in the Deployment and Design of Networking Robots in European Urban Areas. Advanced Robotics 24(13): 1873-1899.

Seyfried, A., Steffen, B., Klingsch, W. & Boltes, M. (2005). The fundamental diagram of pedestrian movement revisited. Journal of Statistical Mechanics: Theory and Experiment, 2005(10): 10002-10002.

Yalowitz, S. S. & Bronnenkant, Kerry. (2009) Timing and Tracking: Unlocking Visitor Behavior. Visitor Studies 12(1): 47-64.

2.1 AN ANALYSIS OF VISITORS' BEHAVIOR IN THE LOUVRE MUSEUM: A STUDY USING BLUETOOTH DATA

Yuji Yoshimura, MIT SENSEable City Lab
Stanislav Sobolevsky, MIT SENSEable City Lab
Carlo Ratti, MIT SENSEable City Lab
Fabien Girardin, Near Future Laboratory
Juan Pablo Carrascal, Universitat Pompeu Fabra
Josep Blat, Universitat Pompeu Fabra
Roberta Sinatra, Northeastern Univeristy

ABSTRACT

Museums often suffer from so-called "hyper-congestion", wherein the number of visitors exceeds the capacity of the physical space of the museum. This can potentially be detrimental to the quality of visitor's experiences, through disturbance by the behavior and presence of other visitors. Although this situation can be mitigated by managing visitors' flow between spaces, a detailed analysis of visitor movement is required to realize fully and apply a proper solution to the problem. In this paper we analyze visitors' sequential movements, the spatial layout, and the relationship between them in large-scale art museums – The Louvre Museum – using anonymized data collected through noninvasive Bluetooth sensors. This enables us to unveil some features of visitor behavior and spatial impact that shed some light on the mechanism of museum overcrowding. The analysis reveals that the visiting styles of short-stay and long-stay visitors are not as significantly different as one might expect. Both types of visitors tend to visit a similar number of key locations in the museum while the longer-stay visitors just tend to do so more time extensively. In addition, we reveal that some ways of exploring the museum appear frequently for both types of visitors, although long-stay visitors might be expected to diversify much more, given the greater time spent in the museum. We suggest that these similarities and dissimilarities make for an uneven distribution of the number of visitors in the museum space. The findings increase the understanding of the unknown behaviors

of visitors, which is key to improving the museum's environment and visiting experience.

Keywords: Bluetooth tracking, visitor behavior, museum studies, human mobility, building morphology

2.2.1. Mesoscopic research of visitors' sequential movement in an art museum

Falk and Dierking argue that "a major problem at many museums is crowding, and crowds are not always easy to control" (1992, page 145). Museums and their exhibits, along with their own spectacular architecture, become some of the most popular destinations for the tourists, thus triggering "hypercongestion" (Krebs et al, 2007), as the number of visitors often exceeds the capacity of spaces, which results in the museum becoming overcrowded.

Congestion in museums shows, on one hand, high attractiveness and vitality, resulting in positive economic impact. On the other hand, the increased number of visitors implies potential negative effects which are detrimental to the quality of visiting conditions and the visitors' experience can be disturbed by the behavior and presence of other visitors (Maddison & Foster, 2003, page 173-174). In an age when museums play an important role in mass cultural consumption and with urban regeneration and the promotion of the image of cities (Hamnett and Shoval, 2003), museums are expected to achieve seemingly contradictory objectives at the same time; that is, to increase the number of visitors and also enhance the quality of their experience by achieving comfortable visiting conditions through management of the flow of visitors.

Visitors' movement and circulation patterns in museums are recognized as an important topic for research (Bitgood, 2006, page 463). However, most of these studies conducted in art museums have been done for only two extreme cases: (a) visitor patterns at the macroscale to investigate the basic demographic composition of the museum's visitors (Schuster, 1995), along with psychographic factors which influence visit motives and barriers (Hood, 1983); and (b) at the microscale to research visitor circulation in the individual

exhibition rooms, limited galleries or other areas. This often results in revealing that: (1) the visitor's attributive features from a sociocultural point of view (ie, highly educated people and wealthy upper- or middle-class people tend to visit more frequently than people from the lower social classes) (Hein 1998, page 115-116); and (2) there is a local interaction between the layout of the exhibits displayed in the galleries and the visitors' behavior in those spaces (Klein, 1993; Melton, 1935; Parsons and Loomis, 1973; Weiss and Boutourline, 1963). This polarized research resulted in a shortage of mesoscopic empirical analysis of visitors in large-scale art museums, which have different research targets compared with a single exhibition, small or medium-sized museums (Serrel, 1998; Tröndle et al., 2012), or other types of museum (Kanda et al., 2007; Laetsch et al., 1980; Sparacino, 2002).

Space Syntax (Hillier, 1996; Hillier and Hanson 1984) applies a different approach to analyzing the influences of the spatial layout and design of buildings using visitors movement and behavior by describing the overall configuration of the museum setting (for a review, see Hillier and Tzortzi, 2006). This type of knowledge is key to producing patterns of exploration and interaction of visitors, and the copresence and coawareness that exists between visitors in the museum environments as a whole (Choi, 1999).

Yet all of these studies rely on a spatially and temporally limited dataset, which often results in providing just a snapshot of a limited area in the built environment. Even a simulation-based analysis uses a simplification of human behavior to estimate visitors' behavior rather than revealing actual patterns of movement with real-world empirical data.

In this paper we analyze the sequential movement of visitors, the spatial layout, and the relationship between them in order to clarify the behavioral features of visitors in a large-scale art museum – The Louvre Museum. We focus on visitors' circulation from the entrance to an exit as a whole mobility network rather than their movement in particular individual rooms. The way of visiting exhibits is analyzed by means of the visitors' length of stay and the sequences in which they make their visits, because these determine the visitor's perceptions and attentions thus shape their visiting experience (Bitgood, 2006). The length of stay might be thought to

be the key factor that determines the number of places visited and the sequence in which they are visited, which results in a variety of different routes; the more time you are given, the more opportunity you have, and vice versa. The question to be asked is whether this hypothesis is actually true, and by its extension, how the length of stay and the sequence of the places visited make visitors' mobility style different, and how this dissimilarity is seen in the museum. This understanding might be the key to improving the museum environment, as well as to enhancing visitors' experiences.

We employ a systematic observation method relying on Bluetooth proximity detection, which makes it possible to produce large-scale datasets representing visitors' sequential movement with low spatial resolution. "Large-scale datasets" refers to the sample size we used being much larger than those collected in art museums for previous studies [eg, almost 2,000 in Melton (1935); 689 in Serrell (1998); 576 in Tröndle et al (2012); 50 Sparacino (2002)], although each of them contains different types of information with sufficient resolution for their particular objectives and as good as human-based observation, GPS, RFID, or ultra-wideband technology can achieve. In our work we explore the global patterns of visitors' behaviors by increasing the quantity of the data, because "when we increase the scale of the data that we work with, we can do new things that weren't possible when we just worked with smaller amounts" (Mayer-Schönberger & Cukier, 2013, p10).

Thus, we limit our research to dealing with visitors' physical presence in and between places, without questioning the introspective aspects (eg, learning process, making meaning from the experience of the museum), which the previous studies tried to answer by small-scale sample [see Kirchberg and Tröndle (2012) for a review]. However, the superimposition of large amounts of data about individuals' movements over time allows some patterns to appear to be self-organizing in a bottom-up way from seemingly chaotic, disordered, and crowded movement. These results could shed light on the quality of visit conditions derived from overcrowding, not only around the spots where the iconic art works are placed, but also the spaces in the network between them that have dynamic visitor flow. A better understanding of visiting features would help in designing more adequate spatial arrangements and give insights to practitioners on how to manage

visitor flow in a more efficient and dynamic way.

2.2.2. Visitor's sequential movement and analysis framework

The use of large-scale datasets enables us to discover and analyze frequent patterns in human activities. Such analyses have been conducted in the specific spatiotemporal limitations derived from the limited measurement of mobile objects (Miller, 2005), in different contexts and at various scales. These analyses have shed light on unknown aspects of human behavior to discover patterns in human mobility (González et al, 2008; Hoteit et al, 2014; Kung et al, 2014;), communication (Ratti et al, 2010; Sobolevsky et al, 2013), and urban activities (Grauwin et al, 2014; Ratti et al, 2006; Pei et al, 2014) by studying cell-phone usage at the regional scale. Other data like social media (Hawelka et al, 2014) or bank card transactions (Sobolevsky et al, 2014) have also been used. In particular, the sequential patterns of tourists at the local scale has been studied by looking at the number of locations visited, their order, and the length of stay, obtained from GPS data (Shoval et al, 2013), and, for instance, some aspects of customers' purchasing behavior in a grocery store have been disclosed by analyzing the customer's path, length of stay, and the categories of products purchased through RFID data (Hui et al, 2009).

Previous research (Yoshimura et al, 2012) proposed a Bluetooth based data-collection technique in a large-scale art museum at the mesoscopic scale in order to classify visitors' behavior by their most-used paths and their relationship with the length of stay. Bluetooth data collection is based on systematic observation which detects Bluetooth-activated mobile devices, in the framework of "unobtrusive measures", making use of the digital footprint unconsciously left by visitors. A considerable number of studies have employed this method but not in the context of large-scale art museums. Examples include measuring the relationship between peoples' social networks (Eagle and Pentland, 2005; Paulos and Goodman, 2004), analyzing mobility of pedestrians (Delafontaine et al., 2012; Kostakos et al., 2010; Versichele et al., 2012), and estimating travel times (Barceló et al., 2010).

A Bluetooth proximity-detection approach to the analysis of visitor behavior in museums has many advantages. Contrary to the granular mobile-phone tracking (Ratti et al, 2006), the detecting scale using Bluetooth is much more fine grained. In addition, in contrast to RFID tags (Hui et al., 2009; Kanda et al. 2007) and active mobile-phone tracking with or without GPS (Asakura and Iryob, 2007), with Bluetooth previous registration is not required and it is not necessary to attach any devices or tags. The fact that no prior participation or registration is required enables a mass participation of subjects and the collection of an enormous amount of data in the long term, unlike time constrained cases (McKercher et al. 2012; Shoval et al. 2013). Also, the unobtrusive nature of Bluetooth removes bias in the data, which could be created if a subject is conscious of being tracked. Furthermore, Bluetooth proximity detection succeeds inside buildings or in the proximity of tall structures, where GPS connectivity is limited. All these advantages make this method adequate for detecting visitors' sequential movement between key places, without specifying their activities, attributes, or inner thoughts, in a consistent way at the mesoscopic scale in a large-scale art museum.

We identify a visitor's length of stay at a particular location as the indicator for measuring their interest level at that exhibit by merely accounting for their presence without questioning their inner thoughts. We estimate visitors' routes between sensors and time at the place from the collected data.

As our analysis and interpretation of data were conducted within a specific spatiotemporal framework, our approach has some limitations. Firstly the concept of trajectory used in this paper is different from the one usually available when working with data collected by GPS systems. This is because a Bluetooth proximity sensor just provides the time-stamped sequence of individual transitions of a mobile device between nodes (eg, sequence of A-B-D), while a GPS system can track all the movements of a device. However the network of rooms derived from the spatial layout of the museum determines the feasible routes, and this enhances estimation of the paths used by visitors between sensors without observing their exact trajectories and orientations per room (Delafontaine et al., 2012). Secondly, we cannot deal directly with visitors' introspective factors, their expectations, experiences, and

satisfactions (Pekarik, et al. 1999). This results in excluding from our study research questions about "wayfinding", which refers to a visitor's ability to find his or her way within a setting, and "orientation", which indicates an available knowledge in a setting through the use of the hand-held maps and direction signs, because they consist of the complex interaction between environmental cognition and the orientation devices. In addition, a visitor's presence at a specific place is not necessarily related to their time engaging with the exhibits, although previous studies used this to measure visitor interest (Melton, 1935; Robinson, 1928). Finally, our sample is possibly biased in two ways. First, the sample composition is affected by the segments of the mobile-device holders and their decision to activate or not activate the Bluetooth function. Although the latter requires calculating the sample representativeness and is typically conducted by using a short-term manual counting method (Versichele et al, 2012), we employed the long-term (one-month) systematic comparison of the number of devices detected at the entrance with the official museum head count and ticket sales. This method provided us with more comprehensive information compared with previous research.

2.2.3. Concept definitions and data settings

In this section we define the locations of sensors used and the components of the dataset in order to explore our method and data consistency. We collected our dataset during a specific period and processed it into a specific form required for the analysis.

2.2.3.1 Sensors settings in the museum and definition of node

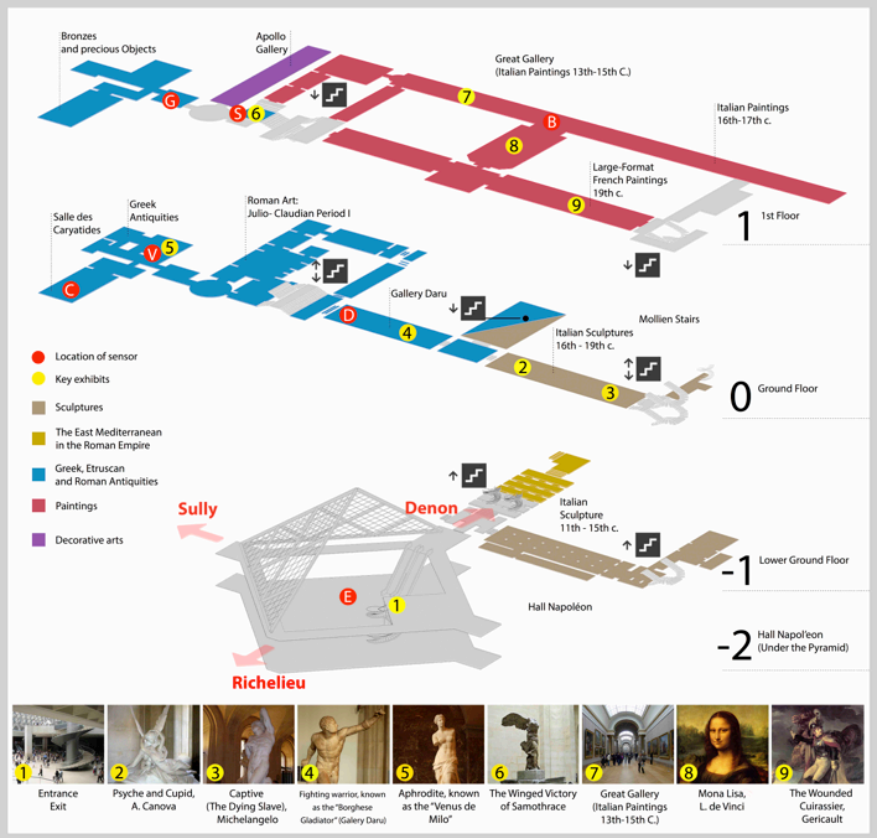


Figure 2.2.1. Location of seven sensors E,D,V,C,B,S, and G, indicating their approximate sensing range

The Figure 2.2.1 shows the location of seven sensors, deployed throughout the museum, covering key places for detecting visitors. They are situated in one of the busiest trails, identified by The Louvre Museum authorities, which lead visitors from the entrance to the Venus de Milo; Entrance Hall (E), Gallery Daru (D), Venus de Milo (V), Salle des Caryatides (C), Great Gallery (B), Victory of Samothrace (S), and Salle des Verres (G).

Each sensor defined a detection area, identified as a node, approximately 20 m long and 7 m wide. The area varied in size, depending on the museum settings and the location of the sensor (eg, inside functional wooden boxes, desks, or in open space). However, all sensors covered targeted areas along the paths to key iconic art works. Once a Bluetooth-activated mobile device enters a detection area, the sensor receives the signal emitted by the mobile

device and the detection continues until the device leaves the area. The sensor registers the time at which the signal from the mobile device first appears, called the check-in time, and when the signal disappears, called the check-out time; the time difference between each mobile device's check-in and check-out time can be calculated to define the length of stay at the node. Similarly, by looking at the first check-in time and the last check-out time for a mobile device over all nodes, provided that the first and last nodes correspond to an entry point and an exit from the museum, respectively, it is possible to calculate how long a visitor stays in the museum. The series of check-in and check-out times registered for a mobile device by all the sensors makes it possible to construct a visitor's trajectory the museum. In addition to the length of the stay, the sensors time stamps allow calculation of the travel time between nodes. The synchronization of all sensors makes it possible to perform fine-grained time-series analysis. All this information can be achieved without invading visitor privacy, because the SHA algorithm (Stallings, 2011, page 342-361) is applied to each sensor where the MACID is converted to a unique identifier (Sanfeliu et al, 2010).

2.2.3.2 Collected sample

We collected data over 24 days; from 30 April to 9 May 2010, 30 June to 8 July 2010, and 7 August to 18 August 2010. We selected data starting and finishing at node E in order to measure the length of stay in the museum. Consequently, 24452 unique devices were chosen to be analyzed for this study. On average, 8.2% of visitors activated Bluetooth on their mobile device device while in The Louvre Museum (Yoshimura et al, 2012).

2.2.3.2.1 Data cleanup

The data collection was performed at different periods by a different number of sensors. We checked for possible synchronization issues arising from a lack of calibration, then adjusted the data to remove any inconsistencies. Finally, we only used data from visitors who started from node E and finished at node E in order to measure the complete length of the visit to the museum – such entries indicate that the visitor was correctly registered when he (or she) entered, moved around inside, and left the museum.

2.2.3.2.2 Data processing

Figure 2.2.2 graphically shows the features of the logged data. It displays all entries in the database for a visitor for one day. Each lettered circle symbolizes detection at the corresponding node. It shows that this particular visitor made a sequential movement, E-S-D-E, and stayed at node E for 3 min 10 seconds, node S for 15 min 20 seconds, node D for 9 min 34 seconds and, again, node E for 6 min 3 seconds. The travel times between corresponding nodes were: 12 min 23 seconds for E-S, 8 min 11 seconds for S-D and 9 min 34 seconds for D-E.

We built a database and designed a query engine to extract and transform the data for the different stages of the analysis. Table 2.2.1 shows an example of components of the transformed dataset. There is one entry per visitor, and it includes the date of the visit, the path followed across the museum, the time of entry (check-in), time of exit (check-out), and the total length of the visit to museum.

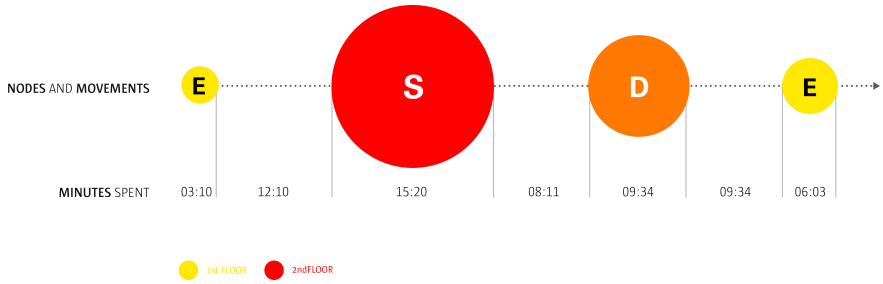


Figure 2.2.2. Visualization of a relationship between the sequential movement and the time of stay of a visitor

Table 2.2.1. Example of the dataset

Rffr Unique ID	Date	Path	checkin	checkout	staylength
	2010-04-30	E-S-D-E	09:04:35	11:07:52	02:03:17

2.2.3.3. Partitioning of Visitors

In order to find the characteristics, the typical patterns of visits, and other determinant features of visitor behavior, we examined two extreme groups. Firstly, we sorted all the visits of our sample (24,452 visits) by their total time spent in the museum. By binning them into deciles, we obtained equally-sized clusters of

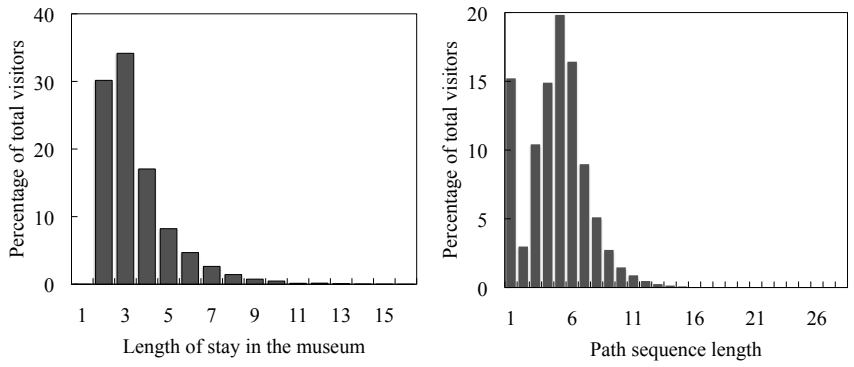
approximately 2446 visits. Referencing all visits, those found in the first decile are called “short visits” and we refer to those visitors as “short-stay visitors”. Similarly, we refer to the visits in the tenth decile as “long visits” and to these visitors as “long-stay visitors”.

2.2.4. Results

In the following subsections, we present an overview of the statistical analysis built around the previously described dataset. We discuss the path sequence length, which is the number of nodes visited, including multiple visits executed without returning to E, the length of the visitors’ paths, and the frequency of the appearance of each path. The distribution of the path sequence length is also presented and analyzed. We reveal visiting patterns, and the similarity and dissimilarity of the behaviors of the long-stay visitors and the short-stay visitors.

2.2.4.1. Basic statistics of visitors’ behavior

We analyzed all visitor data to capture the features of their behavior, focusing on the path sequence length and its relationship with the length of stay in the museum.



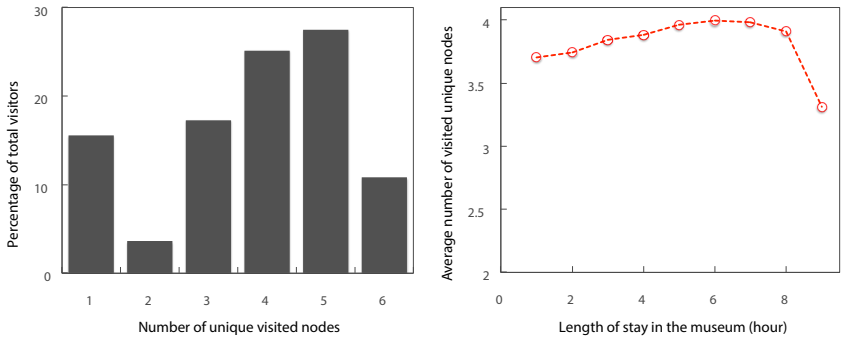
(a) (b)

Figure 2.2.3. (a) The distribution of visits against the length of stay in the museum. (b) The distribution of the path sequence length.

Figure 2.2.3 (a) shows the distribution of the number of visits (y axis) to the length of stay in the museum binned for each hour (x axis). Although the maximum length of stay is more than 15 hours,

only 410 visitors stayed for more than 8 hours, which corresponds to 1.6% of the total. Conversely, the minimum length of stay of less than 1 hour was for only one visitor, while more than 30% of visitors stayed for 1-2 hours. Those facts indicate that the extreme visitors, whose length of stay is more than 8 hours or less than 1 hour, can be aggregated for the statistical reliability without substantially affecting the time-sensitive behavioral analysis. The distribution of the length of stay is positively skewed, with the majority of the visitors staying for 4-6 hours.

Next, we look at the distribution of the path sequence length (number of nodes) [see figure 2.2.3. (b)]. Although the maximum length of the path sequence length is thirty nodes, the percentage of visitors who visited more than fifteen nodes was only 0.5%. In general, this plot shows a distribution slightly skewed to the right, but visitors who visited only one node appear quite frequently, covering 15.2% of the total. Very few people visited two nodes (2.9%). However, the length of the sequence by itself does not necessarily reveal the size of the visitor mobility area, because a visitor could easily move between nearby nodes frequently without radially expanding throughout the museum.

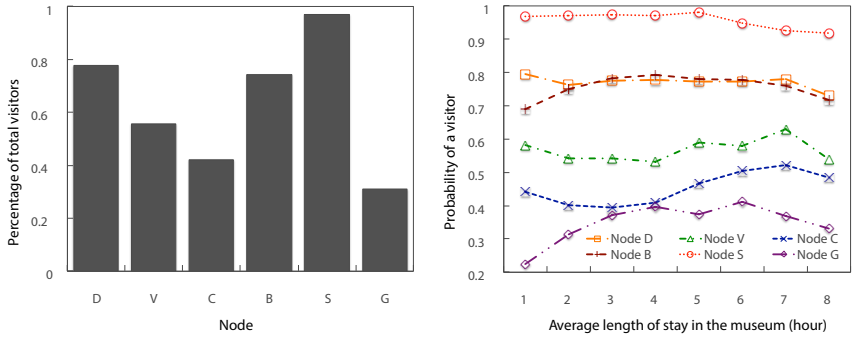


(a) (b)

Figure 2.2.4. (a) Distribution of the number of unique nodes visited other than E. (b) The average number of visited unique nodes visited against the duration of the visit.

Figure 2.2.4 (a) represents the number of unique nodes visitors passed during their stay in the museum. We can observe that visiting two nodes rarely happened, while visiting one and three

nodes have almost the same frequency. The most frequent number of unique nodes visited is four or five nodes, while visiting all six nodes rarely happens. This indicates that in most of the cases some factors prevent the exploration of all the nodes, while all nodes but one could be explored much more often. In addition, figure 2.2.4 (b) reveals that the average number of unique nodes visited against the duration of the visit is almost constant. The correlation coefficient between these two variables (Spearman's correlation=0.072, p-value<2.2e-16) indicates that the unique number of visited nodes is independent of the duration of the visit to the museum, and vice versa. Surprisingly the long-stay visitors usually visit even fewer nodes than the shorter-stay ones.



(a) (b)

Figure 2.2.5. (a) The frequency of visits each node receives. (b) The frequency of visiting different nodes at least once against the duration of stay.

Figure 2.2.5 (a) shows the frequency of visits for each node. 97% of all of visitors passed node S. Nodes D and B are visited frequently (nearly 80% for each). On the other hand, node G is the most rarely visited, with just 30% of all of visitors. Figure 5 (b) presents the attractivity of the nodes depending on the duration of the visit. As we can see, for most nodes the probability of visiting does not depend on the length of stay in the museum as the probability is nearly constant for all nodes. Node G behaves differently from the others, as its probability of attracting visitors increases with the visitors' length of stay in the museum. This shows that short-stay type visitors show a lower tendency to visit node G, while long-stay visitors seem more attracted to visit this node (perhaps having more

time to explore this part of the museum), although its frequency does not surpass 40%, regardless of the visitor type.

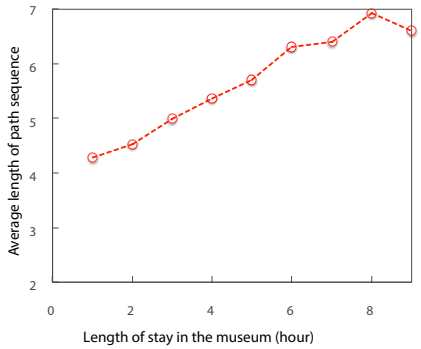
Table 2.2.2. Two types of visitors' transition rate from previous nodes to node G expressed as a percentage. Bold type indicates a substantial increase

Current location/node G	Shorter stay type	Longer stay type	Their difference
D	4.00%	7.17%	3.17%
V	1.38%	3.17%	1.79%
C	4.86%	9.91%	5.05%
B	5.60%	6.30%	0.70%
S	2.53%	5.69%	3.16%

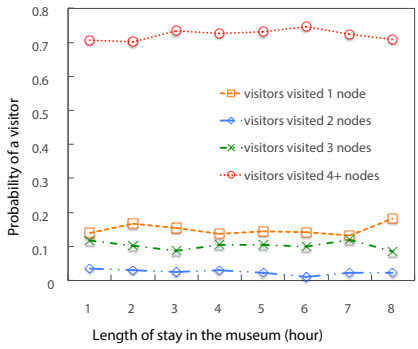
We can observe the difference in the transition rates (probability of moving to the given destination node right after visiting the given origin) from any other node to node G for the two types of visitors (see table 2.2.2). All the transition rates increase as the visitor's length of stay increases; nodes D, C, and S show substantial increases (shown in bold in table 2).

2.2.4.2. Similarity of visitor behaviors

By looking at the path length of the visitors of different stay time we find another surprising effect. Although the path length increases slightly with increase length of visit, the path length of long-stay visitors is not substantially longer than that of the short-stay visitors. In addition, the number of nodes that make up a visit is very similar.



(a)



(b)

Figure 2.2.6. (a) The average length of path sequence (y axis) against the average length of stay in the museum (x axis). (b) The probability of a visitor's path length being 1, 2, 3 or more nodes by their length of stay in the museum (x axis).

Table 2.2.3. The average length of path sequence (number of nodes visited) per hour and its percentage of increase.

Length of visit (hours)	Path sequence length (nodes)	Percentage of increase
1-2	4.28	5.84%
2-3	4.53	9.93%
3-4	4.98	7.63%
4-5	5.36	6.34%
5-6	5.70	10.53%
6-7	6.30	1.43%
7-8	6.39	8.29%
8-9	6.92	4.62%

Table 2.2.4. The probability of a visitor having a path length of 1, 2, 3 or more by the length of their stay in the museum.

Length of stay (hour)	Visitors visited 1 node	2 nodes	3 nodes	More nodes
1-2	0.14	0.03	0.11	0.70
2-3	0.16	0.03	0.10	0.70
3-4	0.15	0.02	0.08	0.73
4-5	0.13	0.03	0.10	0.72
5-6	0.14	0.02	0.10	0.73
6-7	0.14	0.01	0.10	0.74
7-8	0.13	0.02	0.12	0.72
8-9	0.18	0.02	0.08	0.70
Average	0.14	0.02	0.10	0.72
Standard Deviation	0.01	0.00	0.01	0.01

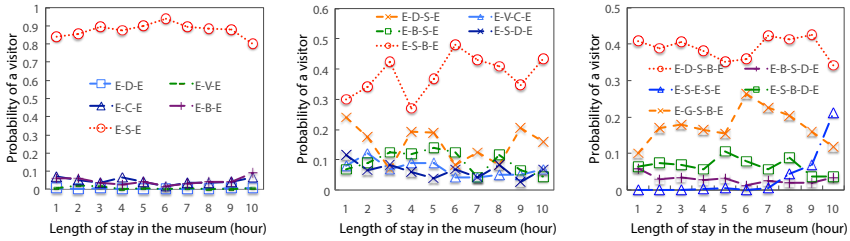
Figure 2.2.6 (a) reveals that, while visitors tend to visit, on average, 4.3 nodes when visiting the museum for 1-2 hours, they are likely to visit only 5.5 nodes when they stay for 3-7 hours. The longer length of stay is three times the shorter, but it results in an increase of only 28% in the sequence length. In addition, the longer stay visitors (ie, 9-10 hours) visited 6.6 nodes on average, which is even less than the 8-9 hour visitors. The path sequence length increases as the duration of the visit increases, but the rate of change is not substantial (see table 2.2.3) especially if compared with the increases in visit times. Figure 2.2.6 (b) presents the probability of

visitors having a certain path length versus their length of stay in the museum. The probability of visiting 1, 2, 3, or more nodes against the length of stay aggregated by each hour appears almost flat, suggesting it is independent of the duration of the visit to the museum. We can also observe this tendency by examining the frequently appearing paths of the short-stay and long-stay visitors (table 2.2.4).

Table 2.2.5. Top five of the frequently appearing paths for paths of four nodes or more and for paths less than four nodes for the long-stay and short-stay visitors.

Path of long-stay visitors Frequency	Path of short-stay visitors Frequency
Visitors whose length of path is more than 4	
E-D-S-B-D-V-C-E; 2.23%	E-D-S-B-D-V-C-E; 8.26%
E-D-S-B-D-E; 1.84%	E-D-S-B-D-E; 6.89%
E-D-S-B-D-V-E; 1.50%	E-D-S-B-V-E; 5.57%
E-D-S-B-D-S-E; 0.89%	E-S-D-V-C-E; 4.67%
E-D-S-B-D-G-E; 0.78	E-C-V-D-S-B-E; 3.71%
Visitors whose length of path is less than 4	
E-S-E; 45.10%	E-S-E; 36.34%
E-D-S-B-E; 12.38%	E-D-S-B-E; 16.62%
E-S-E-S-E; 7.64%	E-V-D-S-E; 4.51%
E-B-E; 5.19%	E-G-S-B-E; 4.12%
E-G-S-B-E; 4.28	E-S-B-E; 3.35%

Table 2.2.5 presents the top five most frequently appearing paths for both the short-stay and long-stay visitors. We counted the number of paths which appear in both groups (visited at least four nodes or visited less than four nodes) and divided by the total number of visitors in the group (ie, 2,445), in order to obtain the frequency of a path appearing. This reveals that both groups have similar frequent path length; the short-stay paths are just slightly shorter compared with the long- stay paths. For both groups, the first and second most frequently appearing paths for the long-stay and short-stay visitors are very similar, otherwise the frequency of the group that visited more than four nodes is much lower than for those who visited less than four nodes. The results show that the behavioral ways of short-stay and long-stay visitors are not as significantly different as one might expect. Both types of visitors tend to visit the same number of popular places but the long-stay visitors just tend to do so more extensively (spending longer studying exhibits).



(a) (b) (c)

Figure 2.2.7. (a) The probability that visitors take particular paths lengths visiting (a) 1 node, (b) 2 nodes, (c) 3 nodes, versus the length of their visit to the museum.

We examine in more detail the visitors whose path length is less than four nodes. Within them, the most frequently appearing path for each category (ie, visited 1, 2 or 3 nodes) coincides well between the groups of short-stay and long-stay visitors. Figure 2.2.7 (a), (b), (c) present the probability of visiting 1 node, 2 nodes, or 3 nodes, respectively. We can observe that in each case only one path has a strong influence on the probability as a whole, especially in Figure 2.2.7 (a), where 89.4% of those visitors took the path E-S-E.

Similarly, visitors who follow the E-S-B-E path, which is the most frequently appearing path for E-D-S-B-E, the most frequently appearing path for those visiting three nodes (38.94%), nodes D and B were added at the beginning and end of their visit, respectively. There is no clear difference between long-stay and short-stay visitors to 1, 2, or 3 nodes; rather, their behaviors seems very similar, other than the substantial difference in the length of the visit to the museum.

2.2.5. Discussion

The previous sections revealed that many features of the behavior of the long-stay and short-stay visitors, including the path sequence length and the unique nodes visited, do not appear to be strikingly different between visits of different duration, and are sometimes even independent or nearly independent of duration. In this section we show that visitors' path and their variations are quite selective, with visitors mostly choosing the same paths in terms of the path sequence length and the sequential order although many other

options exist. This creates an uneven distribution of visitors among spaces, and is possibly one of the main causes of high congestion and vacant spaces in the museum.

2.2.5.1 Uneven spatial distribution of visitors

The interplay between sensor locations and the spatial layout of the museum determines the specific and possible route(s) used by visitors. All sensors were placed logistically in the determinant positions for visitors' route choice in the museum. Therefore, the transition between two places makes it possible to estimate the determinant route that visitors take. Thus, we can clarify the uneven spatial use of the museum accesses for visitors' entry and exit behaviors by analyzing the first two and last two locations, respectively, in their sequence: 71.6% of visitors took E-D, E-B, E-S, meaning that they entered through the Denon access, which indicates that only 28.3% used Sully or Richelieu access (ie, E-V, E-C, E-G). 57.3% of visitors exited from Denon access, 14.3% fewer than entered at Denon. This technique enables us to determine the rooms visited without observing their exact trajectories. Also, this indicates that we could speculate on the volume of visitors and their concentration along specific paths without knowing the exact load per room.

In the previous section it was revealed that 13.5% of all visitors to the museum only visited the Victory of Samothrace (node S – one of the most iconic exhibits in the museum), not visiting any of the other five nodes. Considering the museum's spatial layout, these visitors used the Mollien stairs, which connect the 16th –century to 19th –century Italian sculpture rooms on the ground floor to the 19th-century French painting room on the first floor, to visit node S instead of using the Victory of Samothrace staircase where node D is located (see the orange line at figure 8).

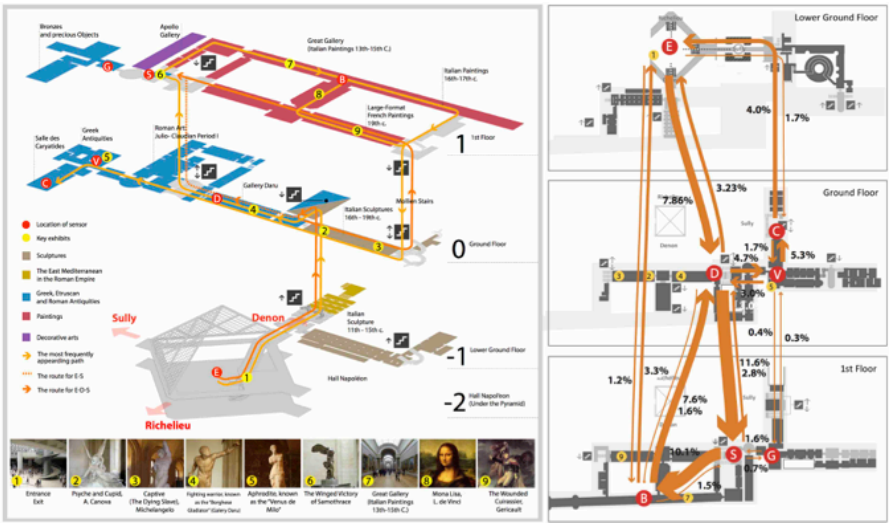


Figure 2.2.8. (a) The map of the spatial layout of the Louvre museum and the used visitors' routes. (b) The transition percentage between locations, which show only major links between each pair of nodes.

From the spatial point of view this is the intriguing result because the shortest path from the entrance (node E) to the Victory of Samothrace (node S) is the one which passes node D, meaning that they turned to the left at the intersection between exhibit No.2 and No.4 (see the orange dotted line at figure 2.2.8). To use the route through the Mollien stairs signifies a detour, both spatially and temporally, to reach the node S.

Table 2.2.6. Three types of visitor transition rate from node E to the subsequent node expressed as a percentage.

Subsequent node from node E	All visitors	Short-stay type	Long-stay type
D	43.32%	42.41%	40.34%
V	11.25%	12.80%	11.38%
C	9.59%	10.02%	11.32%
B	6.80%	9.20%	7.51%
S	21.53%	21.02%	20.82%
G	7.51%	4.54%	8.63%

Table 2.2.6 reveals visitors' route choice more in detail; from node E almost 40% of visitors turned to the left to reach node D (ie, E-D), while around 20% of visitors turned to the right (ie, E-S).

Again, there is no significant difference between the behaviors of the long-stay and short-stay visitors, meaning that both start their museum experience in a similar way. In addition, since nodes D, V, B and G are installed in some key points after exiting the Denon Wing, all those visitors whose path was E-S-E tended to stay in a very confined area of the Denon Wing during their visit. They just explore and stay in the small area during their visit, and this tendency is stronger for the long-stay visitors than the short-stay ones (see Table 2.2.5).

On the other hand, the most frequently appearing path of both groups who visited at least four nodes is E-D-S-B-D-V-C-E; where the visitor visited the Gallery Daru, the Victory of Samothrace, the Great Gallery, and the Venus de Milo (see the yellow line in figure 2.2.8). This path starts from a trail of E-D and finishes with C-E, indicating that the visitor entered the museum from the Denon access, and exited from the Richelieu or Sully access. This suggests that these visitors tend to explore the museum extensively through covering most of the iconic exhibits rather than staying in only one part of the museum. In addition, the frequency of this path for the short-stay visitors is much higher than that of the long-stay visitors. This could indicate that the short-stay visitors might tend to select the most spatially optimized paths to visit all the possible iconic exhibits within their limited available time in the museum.

We believe that short-stay visitors explore fewer of the popular places due to the limited time that they have to spend in the museum. This is intuitive since a visitor's movement and their activities would be limited when the length of their visit to the museum is short. Consequently, the trajectories of the long-stay visitors would be expected to be more complex than those of the short-stay visitors, and vice versa. However, the results show that the behavioral patterns of short-stay and long-stay visitors are not as significantly different as one might expect. Both types of visitor tend to visit a similar number of the popular rooms, but the long-stay visitors tend to do so more time extensively.

The results imply that visitors' trajectories seem to be quite limited in terms of the path sequence length and its order, although there exist a number of possible routes including repeating nodes. More generally, we might say that - and this is partially agrees with

Choi's (1999) statement - the more the number of spaces available, the more the visitor's path tends to be selective. That is, when the number of the rooms with exhibits increases, visitors seem not to visit them all, but visit a few of them selectively. But our findings tell us more; these limited paths and their use are almost independent of the length of the visit to the museum, meaning that most visitors, irrespective of whether their visit is short or long, tend to use the same trajectories.

We speculate that this similarity/dissimilarity of the patterns makes the distribution of the quantity of visitors in the museum space uneven; for instance, the route E-D-S-B-D is frequently observed, independent of the length of the visit, suggesting that there can be a high concentration of visitors in those enclosed areas. In contrast, some spaces can be found to be quite vacant; the sequential pattern between node S and node G is rarely found, especially, in the short-stay visits. This indicates that the topological proximity and the attractivity of a node can be changed depending on the visitor's length of stay (see figure 5). It could be that node G, which tends to be visited when people have more time to explore the museum, is not seen as a necessary or "priority" during the museum visit. Thus, the distribution of visitors is uneven and the number of visits that each room receives varies.

2.2.6. Conclusion

In this study we examined visitors' mobility styles and their respective spatial impacts by analyzing large-scale datasets obtained through Bluetooth proximity detection in a bottom-up methodology. This analysis and the results obtained give a great scientific advancement to improving visiting conditions, which strongly affect the quality of a visitor's experience in the museum.

The results indicate that the behavior of short-stay and long-stay visitors is not as different as one might expect. The path lengths grow at a much slower rate compared with increasing duration of stay. Even more surprisingly, the number of unique nodes visited remains almost constant, independent on the length of the visit. The correlation coefficient between these two variables quantitatively indicates that the unique number of nodes visited is independent of

the duration of the visit to the museum, and vice versa. Both short-stay and long-stay groups visit mostly the same number of sensor locations, while the long-stay visitors just tend to do so more time extensively. Moreover, the probability of the appearance of visitors whose path sequence length is small (<4 nodes), is constant across all time divisions, meaning that there always exists a certain category of visitors who do not try to explore museum space extensively no matter how much time they have to do so. Also we discovered that the frequency of the node visits per hour is almost constant and independent of the length of length of time spent in the museum.

Conversely, we can point out key differences in visitors' behavior within each of two groups – those who visited more than four nodes and those who visited fewer than four. The average number of locations visited, for each of the groups, does not depend on the time people have to spend in the museum (ie, it is independent of a visitor being classified as a short-stay and long-stay visitor). For both short-stay and long-stay visitors the most frequently occurring path in the group that visited at least four nodes is E-D-S-B-D-V-C-E. We might suggest that this path could be one of the most optimized paths, enabling visitors to explore all the interesting places as quickly as possible. Alternatively, the group that visited just a few nodes (less than four), which appears to be of relatively the same size among both short-stay and long-stay visitors, might be interested in just a few of the iconic art works, or just not motivated or informed enough to explore bigger space.

All of this suggests that some routes used to explore the museum appear frequently for both short-stay and long-stay visitors even though the latter might be expected to be much more diverse in their choices given the longer time available. This implies that visitors' sequential movement in The Louvre Museum is quite limited in terms of path sequence length and order, though there are a number of possible routes including repeating the same nodes. We speculate that these similarities/dissimilarities could cause uneven distribution of the number of visitors, resulting in congestion or sparcity in some museum spaces.

These findings present a significant advancement in describing patterns in visitors' activity and behavior in a museum, and might

enable us to foresee visitor movement. This also indicates the possibility of dynamically managing visitor flow and museum congestion, taking into account time-related factors, and the possible advantages of design of the spatial arrangement. In addition, the transition rate and the probability of movement between nodes makes it possible to foresee the specific quantity and flow of visitors at a certain time and space, helping the development of more flexible and dynamic policies for space control. For instance, the similarities/dissimilarities of both types of visitor, which were unknown prior to this study, might make the practitioner reconsider the target of some management techniques that should be applied carefully on the proper and segmented group types (Krebs, et al., 2007; Maddison & Foster, 2003). Also, a dynamic visitor-control system might be developed, based on our findings, by using the audioguides to change suggested visitor routes dynamically depending on the congestion level as calculated by the data gathered from sensors installed throughout the museum.

Finally, these results might enable improvement in the quality of information that can be provided to visitors at an adequate place and time in order to maximize their fulfillment of the social and cultural experience thereby optimizing the museum infrastructure.

References

Asakura Y, Iryob T, 2007, "Analysis of tourist behaviour based on the tracking data collected using a mobile communication instrument" Transportation Research Part A: Policy and Practice 41(7) 684-690

Barceló J, Montero L, Marqués L, Carmona C, 2010, "Travel Time Forecasting and Dynamic Origin-Destination Estimation for Freeways Based on Bluetooth Traffic Monitoring" Transportation Research Record: Journal of the Transportation Research Board 2175: 19-27

Bitgood S, 2006, "An analysis of visitor circulation: Movement patterns and the general value principle" Curator 49 463-475.

Choi Y K, 1999, "The morphology of exploration and encounter in museum layouts" Environment and Planning B: Planning and Design 26(2) 241-250

Delafontaine M, Versichele M, Neutens T, Van de Weghe N, 2012, "Analysing spatiotemporal sequences in Bluetooth tracking data" Applied Geography 34 659-668

Eagle N, Pentland A, 2005, “Reality mining: sensing complex social systems” Personal and Ubiquitous Computing 10(4) 255-268

Falk J H, Dierking L D, 1992, The Museum Experience (Walesback Books, Washington, DC)

González M C, Hidalgo C A, Barabási A L, 2008, “Understanding individual human mobility patterns” Nature 453 779-782

Grauwin S, Sobolevsky S, Moritz S, Gódor I, Ratti C, 2014, ”Towards a comparative science of cities: using mobile traffic records in New York, London and Hong Kong”. arXiv preprint arXiv:1406.4400

Hamnett C, Shoval N, 2003, “Museums as “Flagships” of Urban Development” in Hoffman L M, Judd D, Fainstein S S (eds.) Cities and Visitors: Regulating People, Markets, and City Space (Oxford Blackwell)

Hawelka B, Sitko I, Beinat E, Sobolevsky S, Kazakopoulos P, Ratti C, 2014, “Geo-located Twitter as proxy for global mobility patterns”, Cartography and Geographic Information Science, 41(3) 260-271

Hein G, 1998, Learning in the Museum (Routledge, London)

Hillier B, 1996 Space is the Machine: a configurational theory of architecture (Cambridge University Press, Cambridge)

Hillier B, Hanson J, 1984 The Social logic of space (Cambridge University Press, Cambridge)

Hillier B, Tzortzi K, 2006, “Space Syntax: The Language of Museum Space”, in A Companion to Museum Studies Ed MacDonald S (Blackwell Publishing, London) 282-301

Hood M G, 1983, “Staying away: Why people choose not to visit museums” Museum News, 61(4) 50-57

Hoteit S, Secci S, Sobolevsky S, Ratti C, Pujolle G, 2014, “Estimating human trajectories and hotspots through mobile phone data”, Computer Networks, 64 296-307

Hui S K, Bradlow E T, Fader P S, 2009, “Testing Behavioral Hypotheses Using an Integrated Model of Grocery Store Shopping Path and Purchase Behavior” Journal of Consumer Research 36 478-493

Kanda T, Shiomi M, Perrin L, Nomura T, Ishiguro H, Hagita N, 2007, “Analysis of people trajectories with ubiquitous sensors in a science museum” Proceedings 2007 IEEE International Conference on Robotics and Automation (ICRA’07) 4846-4853

Kirchberg V, Tröndle M, 2012, “Experiencing Exhibitions: A Review of Studies on Visitor Experiences in Museums” Curator:The Museum journal 55(4) 435-452

Klein H, 1993, “Tracking visitor circulation in museum settings” Environment and Behavior 25(6) 782-800

Kostakos V, O'Neill E, Penn A, Roussos G, Papadongonas D, 2010, “Brief encounters: sensing, modelling and visualizing urban mobility and copresence networks” ACM Transactions on Computer Human Interaction 17(1) 1-38

Krebs A, Petr C, Surbled C, 2007, “La gestion de l’hyper fréquetation du patrimoine: d’une problématique grandissante à ses réponses indifférenciées et segmentées” [Managing the hyper-congestion of cultural heritage sites: from a growing problematic to its unspecialized and segmented responses] In 9th International Conference on Arts and Culture Management, University of Valencia. Available from: http://www.adeit.uv.es/aimac2007/index.php [Accessed: 25th August 2013]

Kung K S, Greco K, Sobolevsky S, Ratti C, 2014, “Exploring Universal Patterns in Human Home/work Commuting from Mobile Phone Data”, PLoS ONE 9(6): e96180

Laetsch W, Diamond J, Gottfried J L, Rosenfeld S, 1980, “Children and Family Groups in Science Centers” Science and Children 17(6) 14-17

Maddison D, Foster T, 2003, “Valuing congestion costs in the British Museum” Oxford Economic Papers 55 173-190

Mayer-Schönberger V, Cukier K, 2013, Big Data: A Revolution That Will Transform How We Live, Work and Think (John Murray, London)

Mckercher B, Shoval N, Ng E, Birenboim A, 2012, “First and Repeat Visitor Behaviour: GPS Tracking and GIS Analysis in Hong Kong” Tourism Geographies 14 (1) 147-161

Melton A W, 1935, Problems of Installation in Museums of Art. American Association of Museums Monograph New Series No. 14 (American Association of Museums, Washington, DC)

Miller H J, 2005, “A measurement theory for time geography” Geographical Analysis 37 17-45

Parsons M, Loomis R, 1973 Visitor Traffic Patterns: Then and Now (Office of Museum Programs, Smithsonian Institution, Washington, DC)

Paulos E, Goodman E, 2004, “The familiar stranger: anxiety, comfort, and play in public places” Proceedings of the SIGCHI conference on Human factors in computing systems 6(1) 223-230.

Pei T, Sobolevsky S, Ratti R, Shaw S L, Li T, Zhou C, 2014, "A new insight into land use classification base don aggregated mobile phone data", International Journal of Geographical Information Science, (ahead-of-print), 1-20. arXiv preprint arXiv: 1310.6129

Pekarik A J, Doering Z D, Karns D A, 1999, "Exploring satisfying experiences in museums" Curator: The Museum Journal 42(2) 152-173

Ratti C, Pulselli R, Williams S, Frenchman D, 2006, "Mobile Landscapes: using location data from cell phones for urban analysis" Environment and Planning B: Planning and Design 33(5) 727-748

Ratti C, Sobolevsky S, Calabrese F, Andris C, Reades J, Martino M, Claxton R, Strogatz S, 2010, "Redrawing the Map of Great Britain from a Network of Human Interactions", PLoS ONE 5(12): e14248. doi: 10.1371/journal.pone.0014248

Robinson E S, 1928, The Behavior of the Museum Visitor. American Association of Museums Monograph New Series No. 5 (American Association of Museums, Washington, DC)

Sanfeliu A, Llácer M R, Gramunt M D, Punsola A, Yoshimura Y, 2010, "Influence of the privacy issue in the Deployment and Design of Networking Robots in European Urban Areas" Advanced Robotics 24(13) 1873-1899

Schuster J M, 1995, "The public interest in the art museum's public" In Art in Museums Ed Pearce S (Athlone, London) 109-142

Serrell B, 1998 Paying Attention: Visitors and Museum Exhibitions (American Association of Museums, Washington DC)

Shoval N, McKercher B, Birenboim A, Ng E, 2013, "The application of a sequence alignment method to the creation of typologies of tourist activity in time and space" Environment and Planning B: Planning and Design advance online publication, doi:10.1068/b38065

Sobolevsky S, Sitko I, Grauwin S, Tachet des Combes R, Hawelka B, Arias J M, Ratti C, 2014, "Mining Urban Performance: Scale-Independent Classification of Cities Based on Individual Economic Transactions", arXiv preprint arXiv: 1405.4301

Sobolevsky S, Szell M, Campari R, Couronné T, Smoreda Z, Ratti R, 2013, "Delineating geographical regions with networks of human interactions in an extensive set of countires", PloS ONE 8(12), e81707

Sparacino F, 2002, "The Museum Wearable: real-time sensor-driven understanding of visitors's interests for personalizad visually-augmented museum experiences" in Museums and

the Web 2002: Proceedings Eds Bearman D, Trant J (Archives & Museum Informatics, Tronto)

Stallings W, 2011, Cryptography and Network Security: Principles and Practice, 5th Edition (Prentice Hall, Boston MA)

Versichele M, Neutens T, Delafontaine M, Van de Weghe N, 2011, "The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: a case study of the Ghent festivities" Applied Geography 32 208-220

Weiss R, Boutourline, S, 1963, "The communication value of exhibits" Museum News (Nov.) 23-27

Yoshimura Y, Girardin F, Carrascal J P, Ratti C, Blat J, 2012, "New Tools for Studing Visitor Behaviours in Museums: A Case Study at the Louvre" in Information and Communication Technologis in Tourism 2012. Proceedings of the International conference in Helsingborg (ENTER 2012) Eds Fucks M, Ricci F, Cantoni L (Springer Wien New York, Mörlenback) 391-402

3. Pedestrians analysis in urban settings

The next chapter describes analysis of pedestrian' behaviors in the urban settings. We employed Bluetooth detection technique in order to collect a large-scale datasets of pedestrian sequential movement and their length of stay in the city center of Barcelona. The unprecedented datasets enables us to uncover unknown aspects of pedestrian' behaviors. This chapter consists of one paper:

Yoshimura, Y. Amini, A. Sobolevsky, S, Blat, J, Ratti, C (2015b). "Analysis of pedestrians behaviors through non-invasive Bluetooth monitoring" in Applied Geography (submitted).

3.1 Analysis of pedestrian behaviors through non-invasive Bluetooth monitoring

Yuji Yoshimura, MIT SENSEable City Lab
Alexander Amini, MIT SENSEable City Lab
Stanislav Sobolevsky, MIT SENSEable City Lab
Josep Blat, Universitat Pompeu Fabra
Carlo Ratti, MIT SENSEable City Lab

Abstract

This paper analyzes pedestrians' behavioral patterns while shopping in the historical center of Barcelona, Spain. We employ a Bluetooth detection technique to capture a large-scale dataset of pedestrians' behavior encompassing a one-month period, including a key sales period. We focus on comparing particular behaviors, before, during, and after the discount sales by analyzing this "real" dataset as opposed to estimating relevant parameters for modeling and simulation through a small-scale sample size. Our results reveal pedestrians actively explore a wider area of the district during a discount period compared to weekdays, giving rise to strong underlying mobility patterns.

Keywords: shopping behavior, pedestrian analysis, impulsive, Bluetooth, Barcelona

3.1.1. Introduction

This paper analyzes pedestrians' mobility patterns during a special event, when their behaviors are believed to differ from those during a normal day. We focus on examining behavioral differences between discount days and other normal sale days, in terms of their paths, consisting of the number of visited nodes, their sequential order, and the length of their stay in the district.

The analysis of pedestrian behaviors during shopping trips is usually performed by manual-based data collection techniques such as through interviews and observations (Flick, 2009). A

questionnaire-based on-site survey method is one of the most often employed methods in pedestrian shopping behavior research (Borgers, Kemperman, Timmermans, Zhu, Joh, Kurose, Saarloos & Ahang, 2008). However, this methodology has some shortcomings. Since the description of pedestrian behavior largely depends on the individual's memory, human errors and mistakes may be introduced into the results. The exact time and place for the activities conducted are quite difficult to answer precisely. Additionally, data collection is frequently performed only on one day or over a few days, and creates an average individual as a representative agent who embodies the behavior of the collective group. To superpose the analysis of an individual behavior does not necessarily result in an accurate uncovering of the collective behaviors (Watts, 2011, p80).

Alternatively, several models of pedestrian movement as well route choice through the shopping area have been proposed and developed (see reviews by Borgers, Kemperman & Timmermans, 2009). This approach is largely based on the strong assumption that the consumer is a rational agent making rational choices (Simon, 1987; McFadden, 1999). In the case of utility-maximizing behavior (McFadden, 1978), models assume that agents maximize their perceived utility, or minimize their perceived cost, having complete knowledge about the stores in a shopping area to be visited, and know exactly the order of their visits in advance. However, all of these assumptions indicate a big difference in the real pedestrian behaviors in the shopping area (Kurose, Borgers & Timmermans, 2001, pp405-406; Zhu & Tmmermans, 2008, pp249), therefore the result might not be appropriate for pedestrian behavior analysis, especially during the sales discount periods. This is because the sales or product discounts tend to lead pedestrians to be more impulsive (Liao, Shen, Chu, 2009; Tinne, 2011; Virvilaite, Saladiene & Bagdonaite, 2009). The impulsive buying is frequently associated with "hedonic motivation" (Teller et al., 2008), which is more emotive and product-independent than task-related and rational (Batra & Ahtola, 1991; Hale, Householder & Greene, 2002). As a result, it is more likely that pedestrians will not have a shopping list of products or items in advance. Rather, the environmental factors significantly affect any non-planned purchase decisions, based on the on-the-spot decision-making (Lee & Kacen, 2008).

This paper proposes the application of a Bluetooth detection technique (Eagle & Pentland, 2005; Kostakos, O'Neill, Penn, Roussos & Panadongonas, 2010; Delafontaine, Versichele, Neutens & van de Weghe, 2012; Nicolai, Yoneki, Behrens & Kenn, 2006; O'Neill, Kostakos, Kindberg, Sciek, Penn, Fraser & Jones, 2006; Paulos & Goodman, 2004; Versichele, Neutens, Delafontaine, van de Weghe, 2011; Yoshimura, Sobolevsky, Ratti, Girardin, Carrascal, Blat & Sinatra, 2014) to monitor pedestrians' sequential movements through the shopping area. This technique enables us to generate a large-scale dataset of human mobility at the district scale, because unannounced tracking methodologies makes it possible to collect them during a longer period (i.e., one month) including weekends and special events, which are not included in most of the previous studies. Thus, we try to examine "real and large-scale empirical data" to uncover the pedestrians' behavioral differences during both sales periods and normal shopping days, in terms of the special trajectory, visited places, and their temporal length of stay in the determined district. This specific application of pedestrian analysis during discount sales makes our research different from the above-mentioned previous studies using a Bluetooth technique. Additionally, this proposed method is different from estimating variables in a model or from validating existing constructed models, based on the collected small-scale samples. Rather, our methodology compensates the modeling approach with different perspectives, shedding light on the significant information about pedestrians' impulsive behaviors, which the modeling approach cannot disclose.

In the following section, we present our methodology and discuss how a Bluetooth detection technique can be a viable alternative to the modeling approach, considering the large-scale datasets to be analyzed. In section 3, we describe the design of the experiment in the historical center of Barcelona, and the features of the obtained datasets. In section 4, we analyze the pedestrians' behavioral differences between the sales period and normal shopping days. The paper concludes with a summary of findings and discussions.

3.1.2. The methodology and limitations for analysis framework

A Bluetooth detection technique is based on the systematic observation method in the framework of the "unobtrusive measures" (Webb, Campbell, Schwartz & Sechrest, 1966), making use of people's unconsciously left digital footprints or "data exhaust" (Mayer-Schönberger & Cukier, 2013, page 113). We selected this technique for our study because we consider it more adequate and reliable for studying pedestrians' behavior during discount sales, compared with other types of data collection techniques together with the analysis framework.

First, this technique enables us to collect a large-scale set of samples of pedestrian behaviors. This is different from active mobile phone tracking with or without GPS (Mckercher, Shoval, Ng, Birenboim, 2012; Shoval, McKercher, Birenboim & Ng, 2013) and RFID-based studies (Larson, Bradlow & Fader, 2005; Hui, Bradlow & Fader, 2009). Although they provide us with more accurate and detailed information of the actual trip made by a person in terms of the location in time and space as well as the duration and distance (Ohmori, Nakazato & Harata, 2005), each of them tend to result in relatively small-scale sample sizes, because they require asking participants to bring some devices in advance.

Second, although the network-based passive mobile detection technique can generate a large-scale dataset of human mobility (González, Hidalgo & Barabási, 2008; Ratti, Pulselli, Williams & Frenchman, 2006), the detection range of Bluetooth is much finer grained. The detection range of the former is based on the antenna's coverage, which cannot identify pedestrians' locations between streets. Finally, Bluetooth detection is based on an unannounced tracking system, therefore subjects are not aware of being tracked, resulting in unbiased behavioral data, as expected. This aspect enables us to collect the relevant dataset for a longer period, including during the weekend or a special event, and is thereby not solely representative of a typical day during the year.

Conversely, this method has some shortcomings, which results in the limitations of our research. First, a Bluetooth proximity sensor only knows the time-stamped sequence of individual transitions between nodes (i.e., sequence of A-B-D) for a given mobile device, resulting in the impossibility of determining the actual paths of activities between consecutive detections. Second, this study does

not aim at revealing a consumer's decision-making process or their value consciousness, because our dataset does not capture their inner thoughts or other subconscious information; typically derived from interviews, questionnaires and participatory observation (Flick, 2009). Finally, our sampling contains some bias in terms of a person's attributions. Bluetooth detection enables us to detect the mobile device, given that the said device has Bluetooth activated, indicating that the dataset consists of only people for whom Bluetooth is on and activated on their device. This requires calculating the sample's representativeness and is typically conducted by using a short-term estimation via manual counting (Versichele et al., 2012). To counter this, Yoshimura et al., (2014) conducted a long-term comparison of ticket sales in a museum and a sensor installed in a same place, revealing 8.2% of visitors activated Bluetooth on their mobile phone.

Within these limitations, we analyze the large-scale dataset regarding pedestrian behaviors in order to reveal their features, similarities and dissimilarities of their spatial impacts, and the urban structure through consumer behavioral patterns.

3.1.3. Data collection settings

The 1st of February marks the start of the second major discount period in Barcelona, Spain and lasts until the end of the month. Furthermore, it is known that the Ciutat Vella around Portal de Angel becomes highly congested during the first Saturday in February, which largely affects the retail shops' sales volume. We collected data from 01/29/2009 to 02/20/2009, resulting in more than 4 million unique devices. The choice of this specific season for the research enables us to analyze the impact of a discount day on the pedestrians' behavior compared with the other normal days.

Study settings

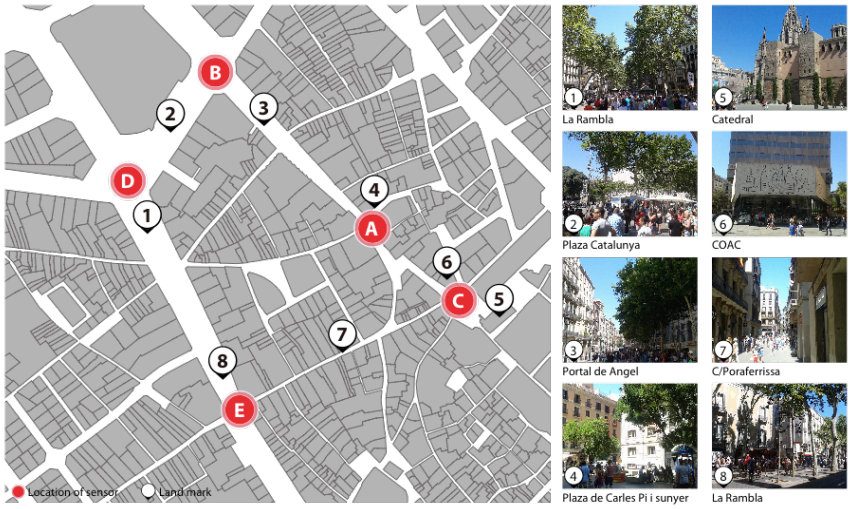


Figure 3.1.1. Location of 5 sensors (red circles) placed in the historical center of Barcelona. Key tourist attractions are numbered by blue circles.

We deployed 5 Bluetooth sensors (nodes) for detecting the presence of pedestrians and their sequential movement between places in the historical center of Barcelona, Ciutat Vella (see Figure 3.1.1). The selected 5 points correspond to locations some of most congested and fluent pedestrian flows in the city of Barcelona. The sensor C is located in front of the Cathedral, which is one of the most important monuments in the city of Barcelona, attracting a large amount of tourists throughout the day. The sensor E is placed in the middle of La Rambla with Portaferrissa street, which is the main street to enter Barri Gotic from La Rambla. The sensor B is placed in the beginning of Portal de Angel street toward the Plaza Carles i Sunny (sensor A), and the sensor D is placed in front of metro station in La Rambla.

Data Preparation

The raw dataset includes several errors and inadequate segments for our analysis. Firstly, this contains data derived from vehicles, because node B and node E face the traffic road, resulting in the collection of the signals from vehicles. Also, there exists data from residents who live near sensor locations. However, we can distinguish the signals from vehicles and pedestrians by looking if the path sequence contains node D, or not. Node D is installed in

the entrance and exit of the metro station where only pedestrians can pass. Thus, we filter them from dataset. Also, we determined the residents, who live near sensor locations by analyzing the recurrence such as devices staying constantly in a node during the whole day and again subsequently removing these logs from dataset.

After clearing the data we identified subsets of the dataset, which start from node D and finish at node D. D-D indicates pedestrians' length of stay in the district, resulting in over 10^5 data samples for the analysis.

The objective of this paper is to reveal the behavioral differences between the first Saturday of the discount period (02/07) represented by F7 and normal Saturday (01/31) by J31, as well as the second Saturday of the discount period (02/14) by F14, and weekdays (Monday to Thursday) during the studied period by W. We exclude all Fridays and Sundays from our analysis, because of the possible bias in pedestrian behaviors: it is clear that people behave differently at Friday during weekdays, and all shops are closed at Sunday.

Within the total dataset of users who started and ended at D, 49.9% (52k samples) occurred from Monday to Thursday, 5.7% (6k samples) occurred during February 7th (Saturday), 5.5% (5.8k samples) during February 14th (Saturday), and 4.2% (4.4k samples) during January 31st.

We aggregate the former groups into one called weekday data for the purpose of comparing against all Saturdays during the studied period.

3.1.4. Results

In the following subsections, we present the results of our analysis built around the previously described dataset. In section 4.1 we discuss the general statistical analysis of the dataset to compare the first Saturday (Feb 7th) with other Saturdays and weekdays, and in section 4.2 we present the most frequently appearing paths taken by pedestrians at each case.

3.1.4.1. General statistical analysis of weekdays and Saturdays

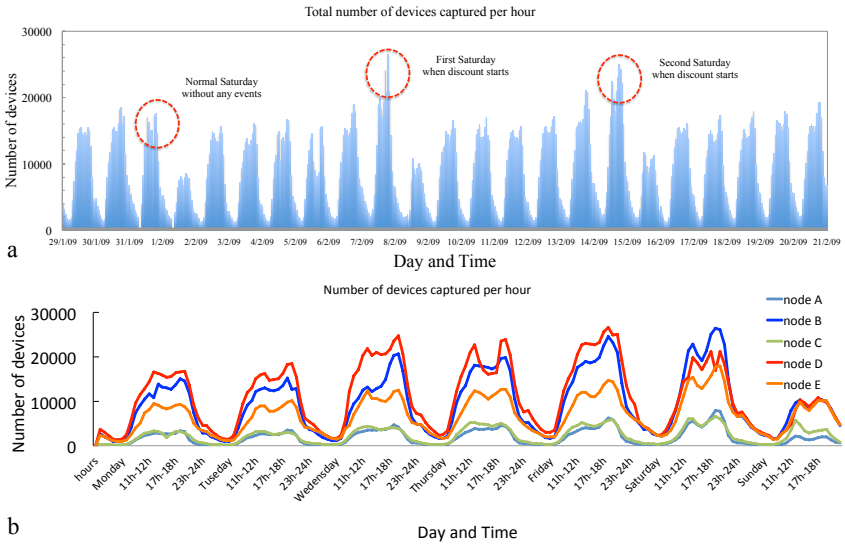
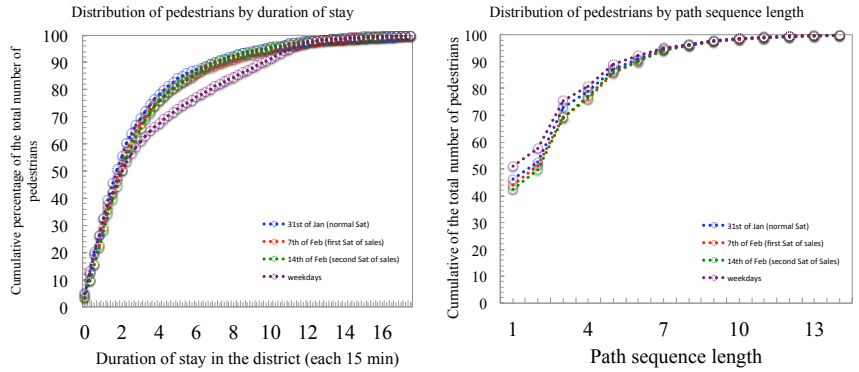


Figure 3.1.2. (a) The number of devices captured per hour over the entire dataset. (b) Weekly patterns for the 5 captured areas during study period.

Figure 3.1.2 (a) presents the number of devices captured per hour over the study period. We can see a consistent pattern of weekday activity and a weaker presence of traffic on Sundays, which amounts to almost half of a weekday. The traffic volume significantly increases on Saturday (F7 and F14). A graph of the weekly activity shows slightly different patterns of activity emerging for the 5 studied areas, with classic morning and afternoon peaks as well as differences between weekdays and weekends. In comparison to other areas, La Ramble (node E) doesn't suffer from a large decrease of activity during the nighttime, and on Sundays, probably due to the streams of tourist present in this area even when shops are closed. Figure 3.1.2 (b) reveals pedestrian patterns during the week: i.e., regular patterns and volumes during the weekdays with sharp decreases on Sundays. Node D always shows the highest volumes of traffic except on Saturdays, where node B overtakes it with increased activity.

These divisional observations motivate us to divide all pedestrians into two groups in order to compare their behaviors in two different groups: pedestrians traveling during the weekdays (Monday to Thursday) and those traveling on Saturdays. We compare the first

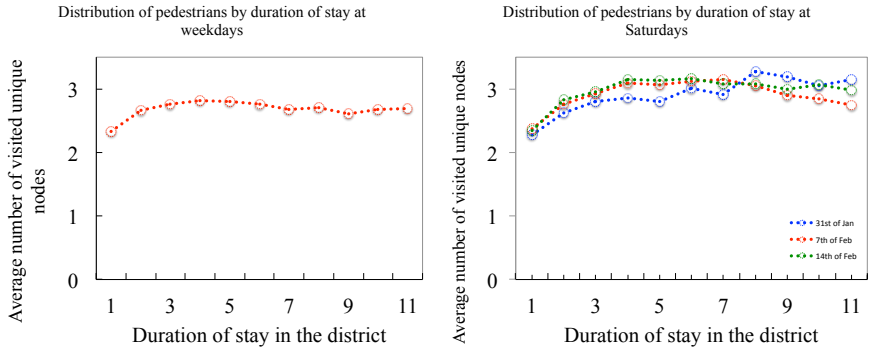
Saturday (F7), when the discount starts, with (1) the normal Saturday (F31) before the discount starts, (2) the second Saturday (F14) after the discount begins, and (3) weekdays (W). Thus, we determine the features of pedestrians' behaviors particularly and distinctly found on Saturdays.



(a) (b)

Figure 3.1.3. (a) The cumulative distribution of pedestrians per the length of stay in the district. (b) The cumulative distribution of path sequence lengths.

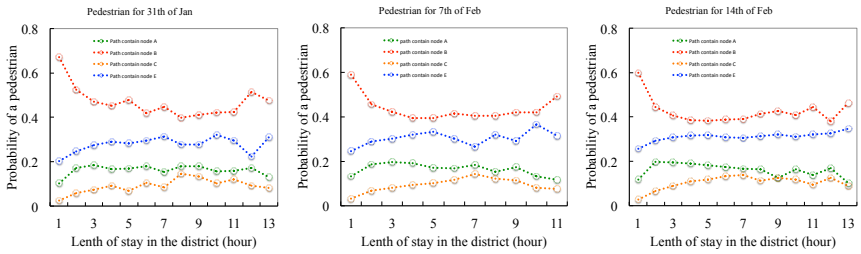
Figure 3.1.3 (a) shows the cumulative distribution of the number of pedestrians against the length of their stay in the district aggregated in 15 minute bins on Saturdays and weekdays. We can observe, while the behaviors during all Saturdays are quite similar, weekdays presents a different distribution. This largely coincides with our intuition that pedestrians' behaviors might be different during weekdays and weekend. Pedestrians during Saturdays tend to rush to leave the district faster than the pedestrians during the weekdays. Figure 3.1.3 (b) shows the cumulative distribution of the path sequence length during Saturdays and weekdays. Again, weekdays show a different distribution, and otherwise all the rest present a similar distribution: pedestrians, who visited only 1 node appear quite frequently, while pedestrians, who visited 2 nodes rarely appear. The average number of a path sequence length during weekdays is shorter than the corresponding average for Saturdays (i.e., 4.56 for W, 4.96 for J31, 5.11 for F7, 5.10 for F14). Although pedestrians for Saturdays tend to stay shorter in the district than on weekdays, they are likely to visit a larger number of nodes within the limited time than pedestrians during the weekdays.



(a) (b)

Figure 3.1.4. (a) The average number of unique visited nodes during weekdays. (b) The average number of unique visited nodes during Saturdays.

However, the path sequence length doesn't correlate to the area or the dimension visited, which the pedestrian explored. This is because they might just visit the same nodes multiple times, resulting in circling only a limited area. Figure 3.1.4 (a) and (b) present the average number of unique visited nodes during weekdays and Saturdays respectively. While pedestrians during weekdays and J31 (normal Saturday) tend to visit less than 3 nodes, those who visit during F7 and F14 surpass more than 3 nodes when they spend more than 3 hours in the area. This is the effect of node C, which pedestrians rarely visit during weekdays and the normal Saturday, but much more frequently during discount Saturdays. Within Saturdays, the number of unique visited node during J31 is always inferior to those during F7 and F14 when the pedestrians' length of stay is less than 7 hours.



(a) (b) (c)

Figure 3.1.5. (a) The probability of pedestrians whose path contain node A, B, C, D for (a) 31st of January (normal Saturday), (b) 7th of February

(first Saturday during sales) and (c) 14th of February (second Saturday during sales).

Let's examine the relationship between the length of stay in the district and the visited node during Saturdays. Almost 70% of pedestrians visit node B within 1 hour during J31 (normal Saturday), while 10% less pedestrians do so during F7 and F14. In addition, the number of pedestrians, who visited node E during F7 and F14, is slightly larger than during the normal Saturday (J31). The probability of a path containing node B is very similar to that of node E, even though the length of stay becomes longer for the former. In addition, the percentage of pedestrians visiting node A is slightly larger during F7 and F14 than the normal Saturday (J31). Considering the sensor locations, this fact indicates that, while pedestrians during J31 tend to stay in a limited area, the ones during F7 and F14 are likely to visit extensive places by walking down La Rambla street (node E) and Portal de Angel street (node A).

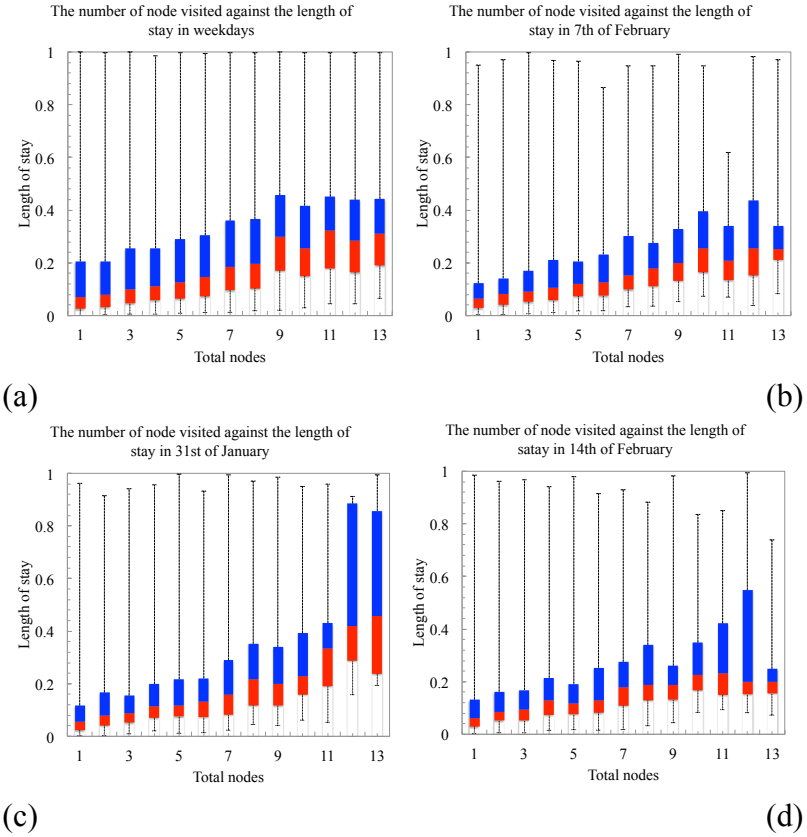


Figure 3.1.6. (a) The total number of visited nodes against the length of stay in the district during (a) W (b) F7 (c) J31 (d) F14.

Table 3.1.1. The ρ and p-value of Spearman's rank correlation coefficient.

	ρ	p-value
Weekdays (W)	0.2740	0
31st of January (J31)	0.4077	9.1164e-179
7th of February (F7)	0.3692	4.1179e-195
14th of February (F14)	0.3707	1.9305e-188

Figure 3.1.6 (a) (b) (c) (d) shows that the relationship between the length of stay and the total number of visited nodes. We utilized a non-parametric correlation analysis (Spearman's rank correlation coefficient), because the variables in question do not follow a Gaussian distribution. We also include a series of boxplots to better explain the relation between these variables. The correlation coefficient for the weekdays, J31, F7, F14 suggests a weak association between two variables ($\rho w = 0.2740$; $\rho j31 = 0.4077$; $\rho F7 = 0.3692$; $\rho F14 = 0.3707$). In addition, the length of stay in the district during the F7 and F14 tends to be shorter than the corresponding length during W and J31, independent from the number of visited nodes. This reveals, on one hand, pedestrians on F7 and F14 circulate more rapidly to a variety of different places and leave the district relatively quickly. Thus, in turn, creates a much higher turnover rate compared to the weekdays and J31. On the other hand, pedestrian behavior during the J31 (Saturday) is much more similar to the weekdays as opposed to Saturdays during the discount period.

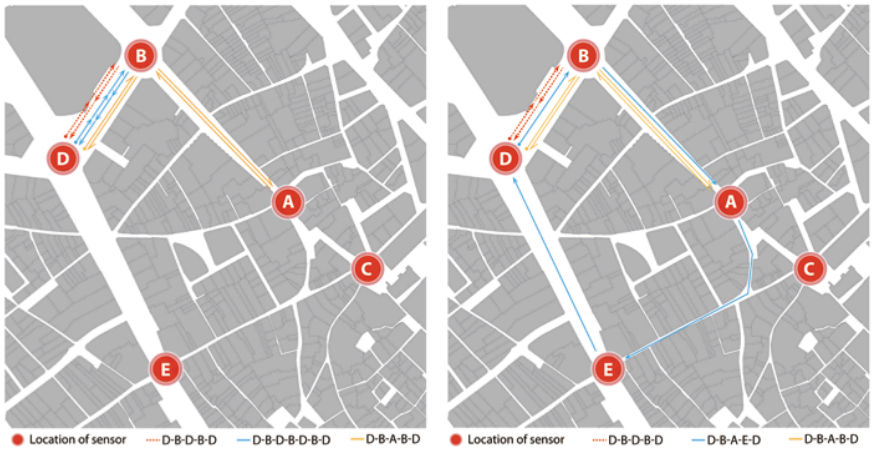
All of these facts reveal that pedestrians during discounted Saturdays (F7 and F14) actively explore different places rather than limiting their stay to a smaller area. In addition, they tend to spend shorter amounts of time in the district than pedestrians during weekdays and normal Saturdays, who tend to visit a larger dimension of the district.

We typically consider that a longer length of stay in the district may lead to an increase in the number of visited nodes, and vice versa. That is, the more time pedestrians spend in the district, the greater the possibility they have of exploring a larger number of nodes. However, our analysis reveals that there is no positive correlation between those two variables, rather our finding indicates a negative correlation which gets stronger during discount Saturdays compared to weekdays and normal Saturdays.

3.1.4.2. Path patterns

Table 3.1.2. Top 5 of the frequently appearing paths.

Pedestrians whose length of path is more than 4	less than or equal to 4
Weekdays	
D-B-D-B-D; 7.8%	D-B-D; 36.5%
D-B-D-B-D-B-D; 2.0%	D-E-D; 13.6%
D-B-A-B-D; 1.8%	D-B-E-D; 4.0%
D-E-D-B-D; 1.4%	D-E-B-D; 0.7%
D-B-D-E-D; 1.2%	D-A-B-D; 0.6%
31st of January	
D-B-D-B-D; 10.4%	D-B-D; 33.8%
D-B-D-B-D-B-D; 2.0%	D-E-D; 11.7%
D-B-A-B-D; 2.0%	D-B-E-D; 3.6%
D-E-D-B-D; 1.5%	D-A-B-D; 0.8%
D-B-D-E-D; 1.3%	D-E-B-D; 0.6%
7th of February	
D-B-D-B-D; 6.5%	D-B-D; 28.6%
D-B-A-B-D; 2.1%	D-E-D; 14.6%
D-B-A-E-D; 2.0%	D-B-E-D; 3.9%
D-B-D-B-D-B-D; 1.7%	D-A-B-D; 0.9%
D-E-A-B-D; 1.6%	D-E-B-D; 0.6%
14th of February	
D-B-D-B-D; 6.8%	D-B-D; 26.9%
D-B-A-E-D; 2.3%	D-E-D; 14.8%
D-B-A-B-D; 1.9%	D-B-E-D; 3.7%
D-E-D-B-D; 1.8%	D-E-B-D; 0.9%
D-B-A-E-A-B-D; 1.5%	D-A-B-D; 0.7%



(a) (b)

Figure 3.1.7. (a) Visualization of the three most frequently appearing paths for W and J31. (b) The three most frequently appearing paths for F7 and F14.

We measured the frequency of path appearance in each case, then normalize by the total number present in each group (Table 3.1.2). The most frequently appearing path for all groups (i.e., D-B-D-B-D) indicates that pedestrians might explore La Rambla street, but they rarely arrive in the Portaferrissa street (node E). Similarly, they are likely to walk down Portal de Angel street, but rarely arrive neither in Plaza de Carles i Sunyer (node A) nor in COAC (node C). Pedestrians tend to remain circulating a limited area around the upper part of Ciutat Vella among the street along with Plaza Catalunya and the beginning of La Rambla. This tendency is stronger for pedestrians during a normal Saturday and weekdays compared to the first Saturday (F7) and second Saturday (F14) of when the discount starts.

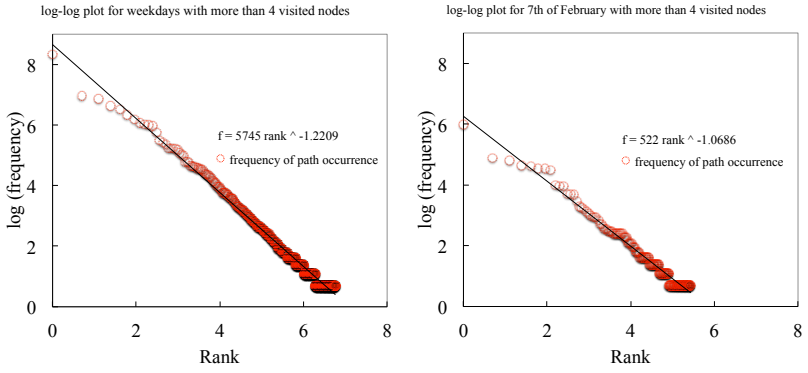
D-B-A-E-D and reverse sequence (i.e., D-E-A-B-D) can be considered a key feature of pedestrians the 7th of February. This path suggests that pedestrians visit the Portaferrissa street (node E) through the middle of Portal de Angel street (node A) from the beginning of Portal de Angel street (node B), then return to the beginning of La Rambla (node D).

The path E-A or A-E is more frequently used by pedestrians during F7 than the ones during W and other Saturdays. We compute the probability for routes upon pedestrian arrival at node A. In cases of

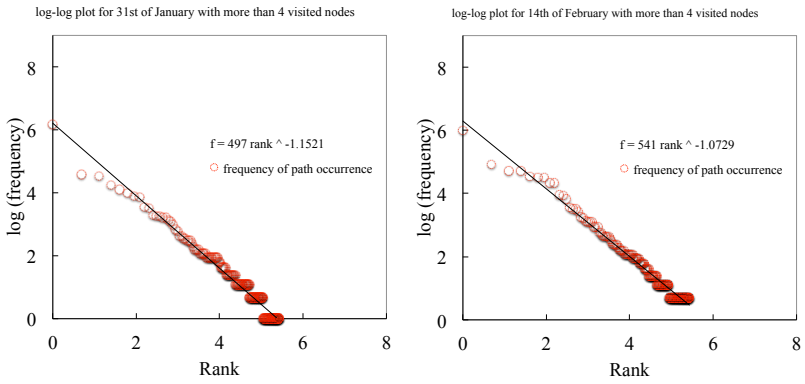
weekdays and J31 (normal Saturday), pedestrians are more attracted to node C than node E. However, during discount periods, they are more attracted to node E than node C. This suggests many more pedestrians during the F7 and F14 select to approach La Rambla through the Portaferrissa street rather than via the Cathedral. At the first and second Saturday in February, pedestrians seem to be drawn more towards the landscape consisting of small retail shops rather than the touristic perspectives of the Cathedral.

Conversely, most of pedestrians during the J31, F14 and W selected to approach to node D upon arriving at node E (i.e., 62.0%, 71.5%, 68.1%), pedestrians during F7 choose to move to node A (26.1%) and node C (23.0%). This indicates that during the first Saturday of the discount period, pedestrians use the Portaferrissa street much more than other Saturdays and weekdays.

This tendency can be also observed when we focus on pedestrian transition time between the pairs of nodes. Pedestrians tend to move slowly when they travel from Portaferrissa street (node E) to Plaza Carles i Sunny (node A), and this tendency is even stronger on the 7th of February compared to weekdays. Conversely, their visit duration becomes significantly longer when they travel the opposite direction (i.e., from node A to node E), with almost at 40 minute difference in travel time. During weekdays, this difference becomes less significant (i.e., 13 minutes), although the trend is reversed, with the transition from E to A taking longer than the one from A to E.



(a) (b)



(c) (d)

Figure 3.1.8. Rank distribution of pedestrians for (a) W (b) F7 (c) J31 (d) F14 whose path length is more than 4.

Table 3.1.3. The slope of the line of best fit in log-log rank plot of path frequencies.

Pedestrians whose length of path is more than 4	less than or equal to 4
Weekdays	
-1.2209	-3.0163
31st of January	
-1.1521	-2.257
7th of February	
-1.0686	-2.8228
14th of February	
-1.0729	-2.0051

We start by examining the strength of pedestrian patterns through their respective paths. We visualize a log-log rank plot of the frequency of the path for weekdays and Saturdays, whose path length is more than 4.

Pedestrians on the F7, who visit 4 or more nodes, show the strongest resemblance to a power law among all groups (slope = -1.0686). This is much stronger than the pedestrians who visited 4 or more nodes on weekdays (-1.2209), and other Saturdays (-1.1521 for the J31 and -1.0729 for the F14). Conversely, the strength of patterns that visit less than 4 nodes is much lower, indicating that these paths have much greater variability in terms of type.

All of these facts indicate that pedestrians on the discount day (F7) tend to explore the district actively by visiting more places rather than staying in a relatively limited area. In addition, we found a correlation between an increased number of pedestrians visiting a given the number of nodes and an increased tendency to use the same path sequence between said nodes. This tendency gets even stronger for the first discount day in February compared to weekdays and other Saturdays. Furthermore, such pedestrians visit these places over a shorter holistic timespan compared to pedestrians on the weekdays, with the exception of several intermediate streets (i.e., A-E, A-B, B-D, D-E). Conversely, the potential area which tends to be explored by pedestrians during the normal Saturday and weekdays is quite limited. Pedestrians generally remain in the upper part of Ciutat Vella without visiting the Cathedral, the middle of La Rambla and Portal de Angel.

3.1.5. Conclusion

This paper analyzes the differences between behavioral patterns of pedestrians on discount days compared to other normal days. The former may be considered more impulsive than the latter, thus our analysis sheds light on unknown aspects of pedestrians' impulsive behaviors, which is difficult to disclose by the standard manually collected small-scale dataset and modeling approach. We installed 5 Bluetooth sensors in the historical center of Barcelona and collected data over the course of one month. This systematic and unobtrusive observation method enables us to obtain a very large-scale dataset of pedestrian mobility in terms of the number of visited nodes, the sequential order of their visits and the length of stay in the district.

Results show that pedestrian behaviors on the Saturdays during the discount periods (F7 and 14F) are quite different from the normal Saturday (J31) and weekdays (W). We intuitively think that pedestrians behave differently during the weekend from weekdays, and our result supports this. However, our analysis reveals that pedestrian behavior during the normal Saturday is more similar to that weekdays than Saturdays during the discount period.

During the Saturdays in a discount period, pedestrians actively explore the district by visiting all nodes including the Cathedral (node C), which is rarely visited by pedestrians on weekdays. We

speculate that pedestrians on discount days might rush to visit the district as much as they can within their limited length of stay. However, we also reveal these types of pedestrians tend to spend longer time than the corresponding ones on weekdays, depending on the streets which they visit, and the direction of transitions between the pair of nodes in said streets. Finally, we reveal that a pedestrian's sequential movement between nodes has underlying patterns in terms of the number of visited nodes and their order. This pattern is much stronger during the Saturdays (F7 and F14) than the normal Saturday (J31) and the weekdays, which makes a relief of the features of pedestrian behaviors during the discount days. By visualizing the number of pedestrians against the path type we see the emergence of the power law distribution, indicating that most pedestrians use only a few path types, and most of path types are used only by a few pedestrians.

These in-depth mobility analyses in the shopping area were not possible prior to our study, because such an adequate dataset was not accessible. Marketing researchers have revealed that many factors may affect the consumers' purchase decisions ranging from a variety perspective such as the psychological, social, cultural to environmental variables like music, lighting or scent (Douce & Janssens, 2013), as well as the customers' mobility aspects in an individual store (Hui et al., 2009). Although Zhuang, Tsang, Zhou & Nicholls (2006) reveals that the number of visited stores has a negative correlation with the purchase behavior in the shopping mall, most of those studies have largely remained in an in-store environment such as the shopping mall, and are rarely applied to the urban district covering several streets and neighhorhoods.

Modeling and simulating pedestrian behaviors were frequently conducted in order to reveal underlying patterns in the shopping area (Borgers et al, 2009). But they are largely based on strong assumptions, where the shopper is rational agent to maximize their utility by minimizing the time and space (Simon, 1987; McFadden, 1999). This assumption does not seems to be appropriately applied to shopping behaviors during the discount sale, because such events can be classified as the impulsive consuming activities related with hedonic ones (Teller, et al., 2008) rather than pre-planned purchased activities. In addition, these research approaches only use a small-scale sample set collected manually to estimate the parameters and

calibrate results of the simulations rather than dealing with the real data about pedestrian behaviors.

Also, our approach and methodologies are different from the previously conducted research: the network-based passive mobile phone detection technique (Gonzales et al., 2009; Ratti, et al., 2006) cannot detect human movement at the street scale, and GPS based tracking (Mckercher et al., 2012; Shoval et al., 2013) and RFID based studies (Larson et al., 2005) is only possible again to make small sample sets. The Bluetooth detection technique is frequently used to collect the pedestrians' sequential movement (Kostakos et al., 2010; Versichele et al., 2010; Delafontaine et al., 2012; Yoshimura et al., 2014), but is rarely applied to reveal the pedestrians' behavioral differences between discount and normal days. Thus, our analysis sheds a new light on unknown aspects of pedestrian behaviors in terms of the number of visited places, the order of their visits and the length of stay in the district, thus making a significant contribution to shopping behaviors.

The obtained results serve as a crucial metric for retailers to enhance their knowledge of hidden aspects of pedestrian behaviors, and understand differences between the discount and normal sale days. As well, these results are extremely helpful for the city authority to make more flexible and efficient urban and security planning, depending on spatial and temporal factors. In addition, the obtained results provide significant information to help formulate pedestrian impulsive shopping behavior models, and this might indicate new perspective for their methodologies. Finally, our analysis becomes a basement for designing new infrastructures and optimizing the use of current infrastructures, in order to manage the crowd and guarantee the safety of people in case of emergency.

Acknowledgements. We would like to thank MIT SMART Program, Accenture, Air Liquide, The Coca Cola Company, Emirates Integrated Telecommunications Company, The ENEL foundation, Ericsson, Expo 2015, Ferrovial, Liberty Mutual, The Regional Municipality of Wood Buffalo, Volkswagen Electronics Research Lab and all the members of the MIT Senseable City Lab Consortium for supporting the research.

References

Batra, R., & Ahtola, O. (1991). Measuring the hedonic and utilitarian sources of consumer attitudes. Marketing Letters, 2, 159-170.

Borgers, A.W.J., Kemperman, A.D.A.M., & Timmermans, H.J.P. (2009). Modeling pedestrian movement in shopping street segments. In H.J.P Timmermans, H. (Ed.), Pedestrian Behavior: Models, Data Collection and Applications, (pp. 87-111). Bingley, UK: Emerald Group Publishing Limited.

Borgers, A.W.J., Joh, C.H., Kemperman, A.D.A.M., Kurose, S., Saarloos, D.J.M., Zhang, J., Zhu, W. & Timmermans, H.J.P. (2008). Alternative ways of measuring activities and movement patterns of transients in urban areas: International experiences. Proceedings 8th international conference on survey in transport (ICTSC), Annecy, France, May 2008, (pp.1-17). Annecy, France.

Delafontaine, M., Versichele, M., Neutens, T., & van de Weghe, N. (2012). Analysing spatiotemporal sequences in Bluetooth tracking data, Applied Geography, 34, 659-668.

Douce, L., & Janssens, W. (2013). The Presence of a Pleasant Ambient Scent in a Fashion Store: The Moderating Role of Shopping Motivation and Affect Intensity. Environment and Behavior, 45, 215-238.

Eagle, N., & Pentland, A.S. (2006). Reality mining: sensing complex social systems. Journal of Personal and Ubiquitous Computing, 10, 255-268.

González, M. C., Hidalgo, C. A., & Barabási, A. L. (2008). Understanding individual human mobility patterns, Nature, 453, 779-782.

Hale, J. L., Householder, B. L., & Greene, K. L. (2002). Theory of reasoned action. In J.P. Dillard., & M. Pfau. (Eds.), The Persuation Handbook: Developments in Theory and Practice (pp. 259-286). Thousand Oaks, CA:Sage.

Hui, S. K., Bradlow, E. T., & Fader, P. S. (2009). Testing Behavioral Hypotheses Using an Integrated Model of Grocery Store Shopping Path and Purchase Behavior, Journal of Consumer Research, 36, 478-493.

Kostakos, V., O'Neill, E., Penn, A., Roussos, G., & Papadongonas, D. (2010). Brief encounters: sensing, modelling and visualizing urban mobility and copresence networks, ACM Transactions on Computer Human Interaction, 17, 1-38.

Larson, J., Bradlow, E., & Fader, P. (2005). An exploratory look at supermarket shopping paths, International Journal of Research in Marketing, 22, 395-414.

Lee, J. A., & Kacen, J. J. (2008). Cultural influences on consumer satisfaction with impulse and planned purchase decisions, Journal of Business Research, 61, 265-272.

Liao, S. L., Shen, Y. C., & Chu, C. H. (2009). The effects of sales promotion strategy, product appeal and consumer traits on reminder impulse buying behaviour, International Journal of Consumer Studies, 33, 274-284.

McFadden, D. (1999). Rationality for economists, Journal of Risk and Uncertainty, 19, 187-203.

Mckercher, B., Shoval, N., Ng, E., & Birenboim, A. (2012). First and Repeat Visitor Behaviour: GPS Tracking and GIS Analysis in Hong Kong, Tourism Geographies, 14, 147-161.

Nicolai, T., Yoneki, E., Behrens, N., & Kenn, H. (2006). Exploring social context with the Wirelss Rope, In R, Meersman., Z, Tari., & P, Herrero. (Eds.), On the move to meaningful internet systems 2006. LNCS, 4277 (pp. 225-242). Springer, Heidelberg Verlag.

O'Neill, E., Kostakos, V., Kindberg, T., Sciek, A.F.g., Penn, A., Fraser, D.S., & Jones, T. (2006). Instrumenting the city: developing methods for observing and understanding the digital cityscape. In P, Dourish & A, Friday. (Eds.), Ubicomp 2006: 8th international conference on ubiquitous computing, LNCS, 4206, (pp. 315-332). Springer Heidelberg.

Ohmori, N., Nakazato, M., & Harata, N. (2005). GPS Mobile Phone-Based Activity Diary Suevey, Proceedings of the Eastern Asia Society for Transportation Studies, 5, 1104-1115.

Paulos, E., & Goodman, E. (2004). The familiar stranger: anxiety, comfort and play in public places. Proceedings SIGCHI Conference on Human Factors in Computing Systems, pp. 223-230.

Ratti, C., Pulselli, R., Williams, S., & Frenchman, D. (2006). Mobile Landscapes: using location data from cell phones for urban analysis, Environment and Planning B: Planning and Design, 33, 727-748.

Shoval, N., McKercher, B., Birenboim, A., & Ng, E. (2013). The application of a sequence alignment method to the creation of typologies of tourist activity in time and space, Environment and Planning B: Planning and Design advance online publication, doi:10.1068/b38065.

Simon, H. (1987). Bounded rationality. In J. Eatewell et al. (Eds.), The New Pargrave: Utility and Probability, New York: Norton.

Teller, C., Reutterer, T., & Schnedlitz, P. (2008). Hedonic and Utilitarian Shopper Types in Evolved and Created Retail

Agglomerations, The International Review of Retail, Distribution and Consumer Research, 18, 283-309.

Tinne, W. S. (2011). Factors affecting impulse buying behaviour of consumers at superstores in Bangladesh, ASA University Review, 5, 209-220.

Tinne W S, 2010, "Impulse Purchasing: A literature Overview", ASA Univeristy Review, 4 (2).

Versichele, M., Neutens, T., Delafontaine, M., & van de Weghe, N. (2011). The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: a case study of the Ghent festivities, Applied Geography, 32, 208-220.

Virvilaite, R., Saladiene, V., & Bagdonaite, R. (2009). Peculiarities of impulsive purchasing in the market of consumer goods, Inzinerine Ekonomika-Engineering Economics, 2, 101-108.

Watts, D. J. (2011). Everything is obvious: once you know the answer, Crown Bussiness, New York.

Webb., E.J., Campbell, D.T., Schwartz, R.D., & Sechrest, L. (1966). Unobtrusive Measures: Nonreactive Research in the Social Sciences, Chicago, IL: RandMcNally.

Yoshimura, Y., Sobolevsky, S., Ratti, C., Girardin, F., Carrascal, J. P., Blat, J., & Sinatra, R. (2014). An analysis of visitors' behaviour in The Louvre Museum: a study using Bluetooth data, Environment and Planning B: Planning and Design, 41, 1113-1131.

Zhuang, G., Tsang, A. S. L., Zhou, N., Li, F., & Nicholls, J. A. F. (2006). Impacts of situational factors on buying decisions in shopping malls: An empirical study with multinational data, European Journal of Marketing, 40, 17-43.

4. Urban Association Rules

The next chapter describes the analysis of consumer' purchasing behaviors and their sequential movement over the city. We examined the subsequent locations, where the consumer purchases, around a three large-scale department stores located in different urban district in the city of Barcelona. We classified all stores into two types: the primary and secondary store. The former largely attracts customers, and distribute the collected customers into the latter. Thus, we measured the attracting power and holding power of each retail shops. As a result, we analyzed their mobility patterns considering their purchasing activities. This chapter consists of these two papers:

Yoshimura, Y. Amini, A. Sobolevsky, S, Blat, J, Ratti, C (2015c). "Analysis of customer' spatial distribution through transaction datasets" in the special issue of Springer's "Transactions on Large-Scale Data-and Knowledge Centered Systems".

Yoshimura, Y., Sobolevsky, S., Bautista Hobin, J N., Ratti, C., Blat, J (2015a). "Urban Association Rules: uncovering consumer behaviors in urban settings through Transaction data", Environment and Planning B (submitted).

4.1 Analysis of customers' spatial distribution through transaction datasets

Yuji Yoshimura, MIT SENSEable City Lab
Alexander Amini, MIT SENSEable City Lab
Stanislav Sobolevsky, MIT SENSEable City Lab
Josep Blat, Universitat Pompeu Fabra
Carlo Ratti, MIT SENSEable City Lab

Abstract

Understanding people's consumption behavior while traveling between retail shops is essential for successful urban planning as well as determining an optimized location for an individual shop. Analyzing customer mobility and deducing their spatial distribution help not only to improve retail marketing strategies, but also to increase the attractiveness of the district through the appropriate commercial planning. For this purpose, we employ a large-scale and anonymized datasets of bank card transactions provided by one of the largest Spanish banks: BBVA. This unique dataset enables us to analyze the combination of visits to stores where customers make consecutive transactions in the city. We identify various patterns in the spatial distribution of customers. By comparing the number of transactions, the distributions and their respective properties such as the distance from the shop we reveal significant differences and similarities between the stores.

Keywords: consumer behaviors, transaction data, human mobility, urban studies, Barcelona.

4.1.1. Introduction

The diversity of a retail shop and its density make an urban district attractive and unique, thereby enhancing the competition between shops and enticing external visitors from other districts both nearby and abroad [1]. Pedestrian exploration and their presence encourage other pedestrians to interact with one another, generating liveliness throughout the neighborhood [2]. Conversely, retailers believe a key

driver of store performance is location [3], which collectively determines the way a customer transitions from shop to shop. This is greatly influenced by geographical accessibility to said shops: a central location is easier to be approached from anywhere, making it more visible and popular to attract both people and goods [4]. “Constraints on mobility determine where we can go and what we can buy” [5].

The objective of this paper is to analyze customers’ spatial distribution considering their consecutive transaction activities through three large-scale department stores in the city of Barcelona, Spain. We study similarities in customers’ origin and destination locations between the same chains of these three stores, which are located in varying urban settings. Essential understanding of this area is largely related to how the power of attraction and distribution for each store affects both the customers as well as the holistic urban environment.

For this purpose, we employ a large-scale transaction dataset provided by one of Spain’s largest banks: Banco Bilbao Vizcaya Argentaria (BBVA). This dataset contains the geographic zip code of a shop where a customer made a transaction, timestamps, and monetary amount of said transaction (see Section 4 for more details). We extracted the combination of retail shops, where customers make consecutive transactions before or after any transactions in one of three large-scale department stores. This approach differs from that in previous studies, which use credit card transactions in the analysis of human behavior [6], [7]. Similary, it is different from analyzing the predictability of human spending activities [5], because the latter utilizes detailed topological analysis whereas we use the physical spatial analysis.

The advantages of our dataset can be summarized as follows: contrary to the point of sales (POS) or the customer loyalty cards [8], BBVA’s credit cards are designed to be used with specific readers installed in over 300,000 BBVA card terminals in Spain [6]. This enables us to analyze spatial distributions of a customer’s sequential purchasing behavior between retail shops over the territory. In addition, the detection scale for the purchase location is smaller than the one for passive mobile phone tracking [9, 10, 11, 12, 13] RFID-based studies [14, 15] or Bluetooth sensing

techniques [16, 17, 18, 19, 20]. This indicates that the attractivity analysis for each shop can be studied at a much finer grain of resolution than in previously recorded studies [21, 22].

Conversely, our research does present several limitations. The dataset consists solely of customers who hold BBVA's credit or debit card and used it for the purchases we analyze. This suggests that our analysis contains a possible bias in terms of the type of customers we study (i.e., highly educated upper and middle class). In addition, our analysis is based on customers' successive order of purchase behaviors between different retail shops, meaning that we cannot deduce their transition path or their purchase decisions when they don't use BBVA's card. Moreover, our dataset cannot reveal customers' decision-making processes or value consciousness because it doesn't contain their inner thought process typically derived from interviews, questionnaires or participatory observation. Furthermore, there is an inherent temporal sparseness present in the data with just a small fraction of all activities being recorded, although this provides enough of sample at the aggregated scale.

Within these limitations, we try to uncover the features of a customer's transaction activities and the similarities of their spatial distribution through the city and the urban structure.

4.1.2. Context of the study: Barcelona

The city of Barcelona is divided into 10 administrative districts, and 73 neighborhoods within those districts, each of which with its own unique identity.

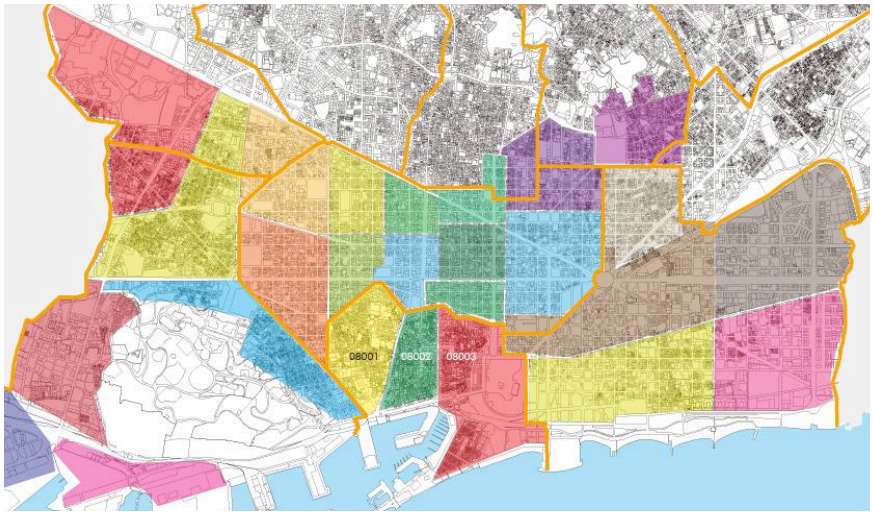


Figure 4.1.1. The map of the city of Barcelona. The zip code, 10 districts and 73 neighborhoods.

Figure 4.1.1 shows the districts, major avenues, and plazas which determine the urban structure of the city of Barcelona. There are approximately 50,000 business entities throughout the city, including department stores, commercial centers, supermarkets, shopping streets with exclusive designer boutiques and international/local brands.

This paper analyzes customer spatial distributions through analysis of their mobility, based on their consecutive activities made before and after visiting the same chain of a large-scale department store. They are located in one of three different neighborhoods in the city. We selected the same chain of large-scale department stores rather than small- and medium-scale shops because (1) we can expect a larger number of customer transactions because of the stores' higher attractivity, (2) customers can be derived from far locations as well as nearby, which enables us to analyze urban structure throughout a larger landscape and (3) the obtained dataset of customers can be more homogeneous rather than distorted and biased.

Each one of these stores attracts a large volume of customers and is therefore able to create expanded distributions of customers to other retail shops in surrounding neighborhoods. They can be considered one of the strongest hubs in the district, triggering a customer's sequential shopping movements. Thus, their presence has great

spatial impact in the district in terms of the volume of attracted customers as well as the associated sequential movements.

The first shop (PC) is located in the city center, Ciutat Vella (old town). Ciutat Vella district is composed of four neighborhoods: El Raval, El Gòtic, La Barceloneta, Sant Pere, Santa Caterina i la Ribera. These neighborhoods are full of retail shops with the most famous brands in the wide commercial area between Pelayo and Portaferrissa streets, and the Portal de l'Àngel. Because of its scenic monuments, architectures and environment, this district attracts tourists as well as locals from all districts of the city.

The second one (AD) is located in Eixample district. This district is divided into six neighbhorhoods (El Fort Pienc, Sagrada Família, Dreta de l´Eixample, Antiga Esquerra de l´Eixample, Nova Esquerra de l´Eixample, Sant Antoni). This area is a business district surrounded by a variety of private companies. Therefore, customers are likely to be workers for these companies as well as people from the wealthy neighborhoods of Pedralbes, Sant Gervasi, and Sarrià.

The last one (PA) is located in Nou Barris district. The shop faces the comer of Sant Andreu and Avenida Meridiana, one of the biggest avenues in Barcelona. This area has a high concentration of immigrants and working-class citizens, as well as a high level of registered unemployment. The specific geographical location is at an entrance to the city of Barcelona and therefore attracts customers traveling from adjacent districts/villages.

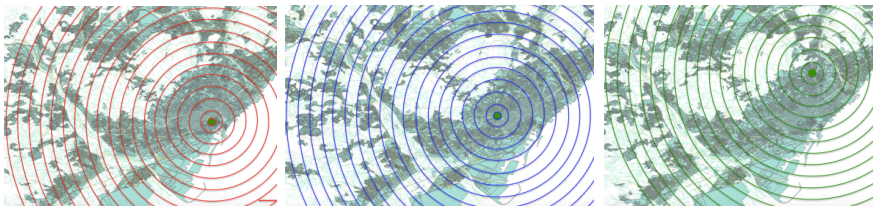
By comparing consumer patterns for the same store located in different regions of the city, our analysis reveals dependencies on neighborhood features more clearly than if different shops has been analyzed.

4.1.3. Methodology

Our goal is to isolate transactions before and after visiting one of three shops in the city of Barcelona within a 24-hour window. We will refer to these three shops (PC, AD, PA) as the focal shops of our study. Specifically, we extracted consecutive sequential credit

and debit transactions as customers moved between stores either before or after visiting the focal shops.

We define an incoming customer as one who makes a transaction in any shop before making a transaction in a focal shop. Similarly, we define a leaving customer as one who makes a transaction in any other shops after doing so in a focal shop.



(a) (b) (c)

Figure 4.1.2. (a). The location of the shop PC with radius of 1km. (b) AD, (c) PA.

Figure 2(a), (b), and (c) show the location of each shop. We aggregate the number of customers within a radius of 1km from each store. This methodology permits us to aggregate customer spending behavior in terms of spatial dimension, where they come from, and where they move to before or after visiting one of those stores.

Within this framework, this paper assesses the spatial distribution based on customers' sequential movement around the large-scale department store located in Barcelona.

4.1.4. Data settings

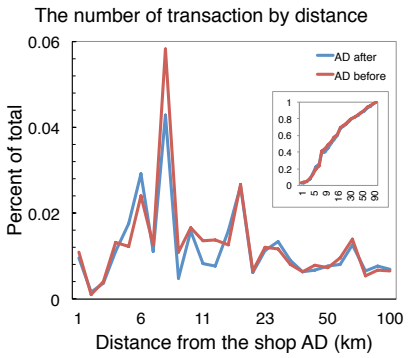
Data for this paper was provided by one of the largest Spanish banks–Banco Bilbao Vizcaya Argentaria (BBVA). The data consists of bank card transactions performed by two groups of card users: direct customers who hold a debit or credit card issued by BBVA and others who made transactions through one of the approximately 300,000 BBVA card terminals. Once customers make transactions with their debit or credit card, the system registers those activities. The information contains the randomly generated IDs of customers, and indication of a customer's residence and a shop where a customer made a transaction at the

level of zip code, a time stamp, and each transaction denoted with its value. The datasets do not contain information about items purchased, and the shops are categorized into 76 business categories such as restaurants, supermarkets, or hotels. In addition, the location where a customer makes transactions is denoted as a zip code rather than the actual street address. The data is aggregated and hashed for anoymization in accordance to all local privacy protection laws and regulations. The total number of customers are around 4.5 million, making more than 178 million transactions totaling over 10 billion euro during 2011 (see [6] for more details).

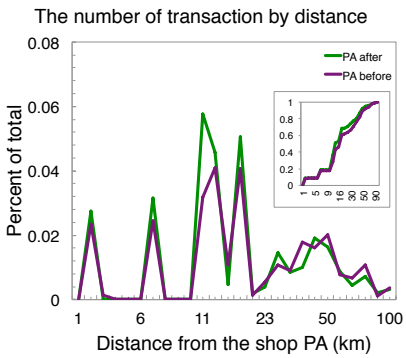
4.1.5. Spatial Analysis

4.1.5.1 Customers distribution in micro scale

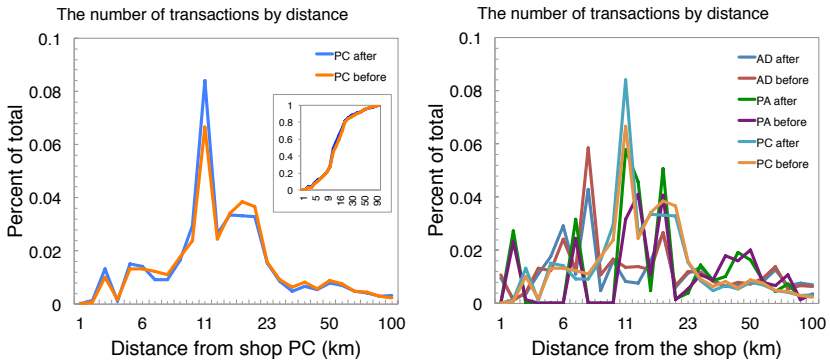
In this section, we analyze the spatial distribution based on customer mobility in the microscopic scale, considering their purchase behaviors. We focus on transactions at shops before or after visiting the three focal shops (AD, PA, PC) around the city of Barcelona. This reveals, on the one hand, each shop's customer mobility in the city of Barcelona, and, on the other hand, the degree of each shop's attracting power and distribution power and their customers' sequential movements around each one.



(a)



(b)



(c) (d)

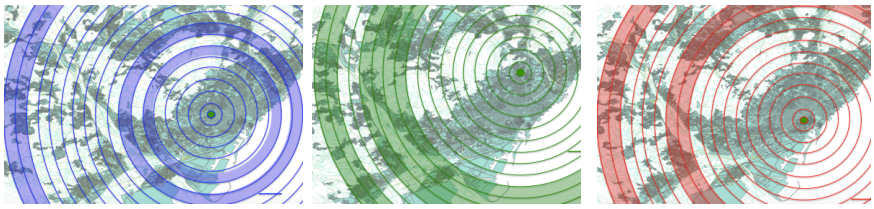
Figure 4.1.3. (a). The distance against the frequency of transactions by the shop AD. (b) PA. (c) PC. (b) All shop.

The volume of transactions against distance for the shop PA can be seen in Figure 4.1.3 (b). PA starts to attract customers from 1 km to 2 km (8.41%), meaning their customers don't make transactions nearby (0-1 km, 0.00%) before/after visiting it. In addition, almost no customers make transactions from proximate locations such as within 2-3 km (0.26%), 3-4 km (0.00%), 4-5 km (0.00%), 5-6 km (0.00%). The hot spot of customers' locations of origin can be found within 6-7 km (9.24%), 10-12 km (14.67%), 12-14 km (14.41%) and 16-18 km (15.12%).

Conversely, the shop AD attracts customers who make transactions nearby (0-1 km, 3.10%). This distribution pattern is unique to AD. The number of customers increases with the distance until 6-7 km (i.e., 3-4 km, 3.71%, 4-5 km, 4.53%, 5-6 km, 8.12%) and is maximized at 7-8 km. In addition, the locations far from the shop tend to show lower percentages of transactions (i.e., 8-9 km, 2.37%, 9-10 km, 4.95%, 10-12 km, 4.58%, 12-14 km, 4.44%, 14-16 km, 4.34%, 16-18 km, 8.05%), indicating that the concentration of transaction volume for AD is intensified in proximal locations.

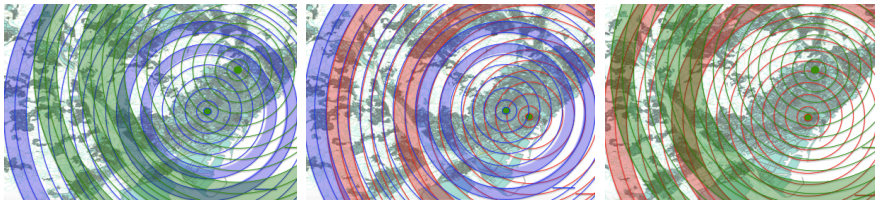
With respect to the shop PC, customer transactions appear within 2-3 km (3.03%); meanwhile, there is almost no customer within 2 km (0-1 km, 0.00%; 1-2 km, 0.29%). The highest concentration of customer transactions occurs within the 10-12 km radius (19.50%) with smaller aggregate transactions intervening (4-5 km, 3.69%; 5-6 km, 3.57%; 6-7 km, 2.80%; 7-8 km, 2.66%; 8-9 km, 4.39; 9-10 km, 6.90%). The customers also increase positively toward 20 km (12-

14 km, 6.60%; 14-16 km, 8.77%; 16-18 km, 9.33%; 18-20 km, 8.98%).



(a) (b) (c)

Figure 4.1.4. (a) The visualization of the peaks of the number of transactions for the shop AD. (b) PA. (c) PC.



(a) (b) (c)

Figure 4.1.5. (a) The visualization of the peaks of the number of transactions for the shop PC and AD. (b) PC and AP. (c) PA and AD.

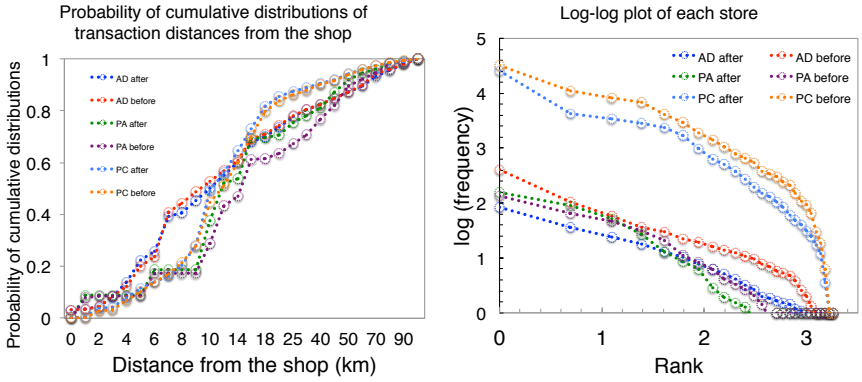
The following is an analysis of the overlap of those geographical locations between the three shops. Figure 4.1.4(a), (b), and (c) visualize the concentration of customer transaction to geographical locations. Figure 4.1.5(a), (b), and (c) show the overlap of those concentrations between PC and AD, and PC and AP, and PA and AD, respectively.

As we can see, PA's trading area is sometimes overlapped with that of shops AD and PC. For the former case, it is southwest of Barcelona, and for the latter case, it is northwest of Barcelona. This indicates that those two shops (i.e., PA and AD, and PA and PC) compete for their trading area rather than complement each other in the city. Conversely, the trading areas between shops AD and PC are nonoverlapping. They are clearly separated, meaning that harmonious operations are achieved by each shop despite the proximity between them.

All these facts uncover the hidden structures of shops' trading areas and their similarities at the micro scale. Each shop has unique concentrations of customer transactions.

4.1.5.2 Customers' spatial distributions in macro scale

This section analyzes the customers' origins and destinations for each store over the wider territory. The goal is to detect the macroscopic trading area through spatial analysis. The difference from the previous section is the scale. While the previous section examined it within the city of Barcelona, this section focuses on the wider territory over the city.



(a) (b)

Figure 4.1.6. (a) The distance from the shop where transactions are made against the cumulative frequency of the normalized number of transactions of leaving/incoming customers. (b) The transaction frequencies for each rank of distance from the shop.

We compute the cumulative number of transactions made by the leaving and incoming customers against the distance from the focal shops. December, January and July show significantly larger number than other months for all three cases. This result coincides with previous studies where those three months mean a high season through a year in Spain. In addition, this result shows that an individual shop's attractivity seems dynamic rather than static depending on the season.

Conversely, we also compute the cumulative distribution of transactions against the distance from the shop (see Figure 4.1.6(a)).

They show that incoming and leaving customers of each shop have a particular pattern in terms of distributions of locations where customers make the consecutive transactions. For instance, shop PC and shop PA present the sudden increase in transactions around 14 km, while shop AD's happened at 7 km. With respect to shop PC, the slope starts to decrease at around 15 km, and 14 km in the case of shop PA. In addition, Figure 4.1.6(b) presents that log-log plot of the number of transactions against the distance from the shop.

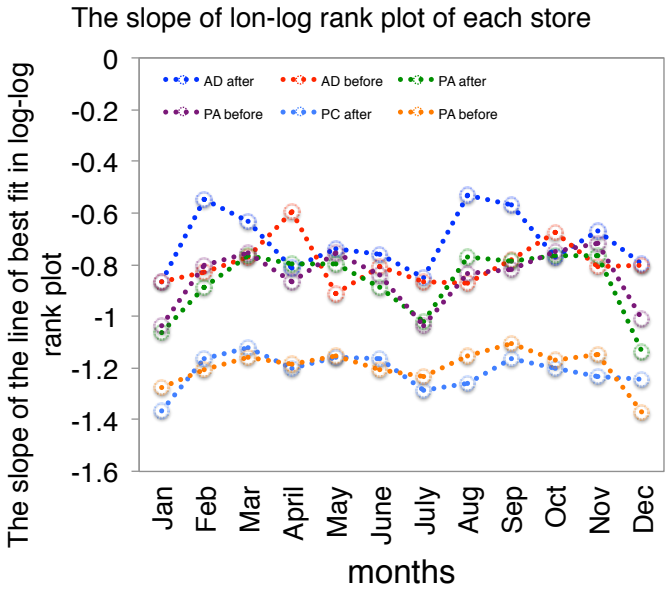


Figure 4.1.7. The change of the log ranked distance bin (slope) by the transactions frequency in each month.

Table 4.1.1. The slope of the line of best fit for each log ranked customers' frequency vs distance during the high seasons.

	January	July	December
AD after	-0.8682	-0.848	-0.7961
AD before	-0.864	-0.864	-0.7998
PA after	-1.0646	-1.0183	-1.1348
PA before	-1.0362	-1.0362	-1.0113
PC after	-1.3679	-1.2859	-1.2432

PC before			
	-1.2729	-1.231	-1.3701

Let's examine the log-log plot of the spatial distribution of the number of transactions in each month. Figure 4.1.6(b) presents transaction frequencies for each ranked distance bin for the entire period, and Figure 4.1.7 presents the change of the slope of its rank plot by each month. We can observe that both pre- and posttransaction in shops AD and PA is nearly -1.0 in January, July, and December, which corresponds with the high seasons. This indicates that few locations have a much higher number of transactions, while most locations have very few transactions. And this tendency is even stronger in shop PA than in shop AD and PC. Most of PA's customers tend to derive from a minimal number of places and subsequently move to few locations. Conversely, the origin and destination shops for PC's customers become largely dispersed in January, July, and December compared to other months.

We can see from these results that customer transaction activities have unique patterns in terms of their spatial distribution, which are unique to each individual shop. We speculate that PA might attract local customers rather than tourists from far away. This explains that the origin as well as the destination of their customers is quite similar, and those few places are the main sources for their customers. Conversely, PC appears to attract tourists rather than local citizens, and this tendency is magnified during the high seasons of the year. The customer origin and destination become more dispersed throughout the discount season.

We tend to consider that high seasons increase the number of transactions since many drastic discounts cause customers to rush to shops even from abroad. Our result partially reveals this phenomenon in the case of shop PC, but this is not a consistent pattern among all stores. On the other hand, we showed that the number of transactions during the high season has the same proportion as the low season, meaning that the former portrays an increase in transaction volume compared to the latter. That is, the spatial distribution of transaction activities is exactly the same between the high and low seasons. However, the cause of this increase varies largely depending on the specific store and its

location. In case of PA, this effect is not due to the increase of customers who come from other places but simply an increase of the quantitative volume from the same places. Contrary to this fact, in the case of PC, this effect is largely due to the ones deriving from other places, indicating that the simple increase of the same customers from the same locations does not apply in this case.

4.1.6. Conclusion

This paper uncovers customers' spatial distributions by analyzing their mobility patterns. We extract locations of consecutive transactions made by customers before and after going to one of the selected three focal shops.

These shops, PC, AD, and PA, are each located in a different urban context across the city of Barcelona, thereby uncovering unique characteristics of their customers as well as the area they are located in. The large-scale and anonymized credit card transaction dataset makes it possible to analyze the successive chains of a customer's purchase history between shops dispersed over the territory rather than an analysis inside a single unique shop.

Our findings reveal that the trading area of each store is largely distributed in a specific way. Customers of shops AD and PC derive from similar places, resulting in competition to attract said customers from each other. Conversely, customers of shops AD and PC share no overlap within the city, allowing them to coexist rather than compete.

In addition, we discover that some distributions of the number of transactions against the distance from the shop follows a power law. This reveals that few locations have higher frequencies of transactions, while most of them have very few transactions. This tendency is amplified even further in shop PA compared to AD or PC. Moreover, our analysis discloses how transaction volumes increase during high and low season. Specifically, customers during high seasons come from similar places rather than from different locations in the case of shop PA. The number of transactions in the former just increases from a similar place in proportion with the ones for the latter, meaning that the customer's spatial distribution

is exactly the same for both. However, in the case of shop PC, the customer's mobility pattern is different. The origin and destination of shop PC's customers become dispersed during the high season rather than converged as in the low season.

The outcome is almost reversed between shops PC and PA, although they are the same chain of the large-scale department store. We speculate that this feature might be due to the geographical and sociocultural context of each store. While shop PA is situated in the suburban area with a higher rate of immigration, shop PC is located at the center of the city, which is one of the most popular touristic places.

We have an intuition that urban contexts and their differences cause the feature of stores and their customers to differ. For instance, the store located at a tourist setting may attract many more tourists compared to one in a business or suburban district, and vice versa. In spite of these beliefs, this paper reveals this difference quantitatively through the spatial analysis based on large-scale dataset.

All of these analyses were not possible prior to our research. The previous researchers have frequently used the Huff model [21, 22] to estimate the trading area of a shop in a macroscopic point of view. This merely reveals the homogeneous distribution of customer home locations and the strength of the shop's attractivity, since the model simply depends on the distance from and the size of the shop. Thus, the result of the analysis doesn't represent heterogeneous customers and their geographical features, or the temporal factors. Also, this information is not possible with active mobile phone tracking with or without GPS [23, 24], or with passive mobile phone tracking [12] and Bluetooth detection techniques [20]. The dataset collected by those methods just provide the users' locations without considering evidence of their purchases. Thus, we are only able to predict when purchases are made with a series of significant assumptions. The combination of RFID [14] and the POS system is proposed to reveal a relationship between sales volumes made by customers and their mobility patterns. However, it is possible only inside a single store or mall.

Our proposed methodologies should address these drawbacks. Our dataset permits us to analyze the customer's consumer behaviors across different retail shops, which are dispersed in the urban area; thus, we reveal subsequent purchase behaviors while considering their mobility aspects when they complete microscopic transaction activities. This means that our current research shows the locations of customer transactions rather than just customers passing through these shops. In addition, our methodology and analysis can reveal the individual shop's attractivity and its influences in the territory as trading areas in the micro scale. Furthermore, our methodology and extracted knowledge are extremely helpful in improving Christaller's urban centrality model [25] and reveal the urban structure as well as its hierarchy. Although spatial structure and hierarchy of cities by size and distance have been well studied [25, 26], "the regularity of the urban size distribution poses a real puzzle, one that neither our approach nor the most plausible alternative approach to city sizes seems to answer" (page 219 in [27]).

These extracted patterns help improve spatial arrangements and services offered to customers. Thus, retail shops and their districts can improve sales as well as their environment, thereby revitalizing the center of the urban districts. In addition, these findings are useful to urban planners and city authorities in revitalizing deteriorated districts or rehabilitating neighborhoods. Understanding customers' sequential movement with transaction activities enables us to identify potential customer groups and their geographical demographics spatially. Finally, city planners can consider optimizing the infrastructures and the locations of the retail shops to make the district more attractive and active by increasing the number of pedestrians. For instance, the customers' sequential movement between different retail shops facilitates collaboration between all shops in a district as a whole rather than individually, to organize planned sale periods. Based on our findings, neighborhood associations can organize discount coupons or advertisements in relevant and adequate places. This can serve as an efficient indicator as to when they are most likely to complete transactions as well as their successive locations.

Acknowledgements. We would like to thank the Banco Bilbao Vizcaya Argentaria (BBVA) for providing the dataset for this study.

Special thanks to Juan Murillo Arias, Marco Bressan, Elena Alfaro Martinez, Maria Hernandez Rubio and Assaf Biderman for organizational support of the project and stimulating discussions. We further thank MIT SMART Program, Accenture, Air Liquide, The Coca Cola Company, Emirates Integrated Telecommunications Company, The ENEL foundation, Ericsson, Expo 2015, Ferrovial, Liberty Mutual, The Regional Municipality of Wood Buffalo, Volkswagen Electronics Research Lab and all the members of the MIT Senseable City Lab Consortium for supporting the research.

References

[1] Jacobs J, 1961, The Death and Life of Great American Cities, Random House, New York.

[2] Gehl J, 2011, Life Between Buildings: Using Public Space, Island Press, Washington-Covelo-London.

[3] Taneja S, 1999, "Technology moves in", Chain Store Age 75 (may), pp. 136-138.

[4] Porta S, Latora V, Wang F, Rueda S, Strano E, Scellato S, Cardillo A, Belli E, Càrdenas F,Cormenzana B, Latora L, 2012, "Street centrality and the location of economic activities in Barcelona" Urban Studies, 49 (7), pp. 1471-1488.

[5] Krumme C, Llorente A,Cebrian M, Pentland A (S), Moro E, 2013, "The predictability of consumer visitation patterns", Sci. Rep. 3, 1645; DOI:10.1038/srep01645.

[6] Sobolevsky S, Sitko I, Grauwin S, Tachet des Combes R, Hawelka B, Murillo Arias J, Ratti R, 2014, "Mining Urban Performance: Scale-Independent Classification of Cities Based on Individual Economic Transactions" arXiv:1405.4301.

[7] Sobolevsky, S., Sitko, I., Tachet des Combes, R., Hawelka, B., Murillo Arias, J., Ratti, C.: Money on the move: Big data of bank card transactions as the new proxy for human mobility patterns and regional delineation. The case of residents and foreign visitors in spain. Big Data (BigData Congress), 2014 IEEE International Congress, 136--143 (2014)

[8] Leenheer J, Tammo H A Bijmolt, 2008, "Which retailers adopt a loyalty program? An empirical study" Journal of Retailing and Consumer Services 15, pp. 429-442.

[9] González M C, Hidalgo C A, Barabási A L, 2008, "Understanding individual human mobility patterns" Nature 453, pp. 779-782.
[10] Hoteit S, Secci S, Sobolevsky S, Ratti C, Pujolle G, 2014, "Estimating human trajectories and hotspots through mobile phone data", Computer Networks, 64, pp. 296-307.
[11] Kung K S, Greco K, Sobolevsky S, Ratti C, 2014, "Exploring Universal Patterns in Human Home/work Commuting from Mobile Phone Data", PLoS ONE 9(6): e96180.
[12] Ratti C, Pulselli R, Williams S, Frenchman D, 2006, "Mobile Landscapes: using location data from cell phones for urban analysis" Environment and Planning B: Planning and Design 33(5), pp. 727-748.
[13] Sobolevsky S, Szell M, Campari R, Couronné T, Smoreda Z, Ratti R, 2013, "Delineating geographical regions with networks of human interactions in an extensive set of countires", PloS ONE 8(12), e81707.
[14] Kanda T, Shiomi M, Perrin L, Nomura T, Ishiguro H, Hagita N, 2007, "Analysis of people trajectories with ubiquitous sensors in a science museum" Proceedings 2007 IEEE International Conference on Robotics and Automation (ICRA'07), pp. 4846-4853.
[15] Larson J, Bradlow E, Fader P, 2005, "An exploratory look at supermarket shopping paths" International Journal of Research in Marketing 22 (4), pp. 395-414.
[16] Delafontaine M, Versichele M, Neutens T, Van de Weghe N, 2012, "Analysing spatiotemporal sequences in Bluetooth tracking data" Applied Geography 34, pp. 659-668.
[17] Kostakos V, O'Neill E, Penn A, Roussos G, Papadongonas D, 2010, "Brief encounters: sensing, modelling and visualizing urban mobility and copresence networks" ACM Transactions on Computer Human Interaction 17(1), pp. 1-38.
[18] Versichele M, Neutens T, Delafontaine M, Van de Weghe N, 2011, "The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: a case study of the Ghent festivities" Applied Geography 32, pp. 208-220.
[19] Yoshimura, Y., Girardin, F., Carrascal, J. P., Ratti, C., Blat, J.: New Tools for Studing Visitor Behaviours in Museums: A Case Study at the Louvre. In: Fucks, M., Ricci, F., Cantoni, L. (eds.) Information and Communication Technologies in Tourism 2012. Proceedings of the International conference in

Helsingborg (ENTER 2012), pp. 391—402. Springer Wien New York, Mörlenback (2012).
[20] Yoshimura Y, Sobolevsky S, Ratti C, Girardin F, Carrascal J P, Blat J, Sinatra R, 2014, "An analysis of visitors' behaviour in The Louvre Museum: a study using Bluetooth data" Environment and Planning B: Planning and Design 41 (6), pp. 1113-1131.
[21] Huff D.L, 1964, "Defining and estimating a trade area" Journal of Marketing 28, pp. 34-38.
[22] Huff D.L, 1966, "A programmed solution for approximating an optimum retail location" Land Economics 42, pp. 293-303.
[23] Asakura Y, Iryo T, 2007, "Analysis of tourist behaviour based on the tracking data collected using a mobile communication instrument" Transportation Research Part A: Policy and Practice 41(7), pp. 684-690.
[24] Shoval N, McKercher B, Birenboim A, Ng E, 2013, "The application of a sequence alignment method to the creation of typologies of tourist activity in time and space" Environment and Planning B: Planning and Design advance online publication, doi:10.1068/b38065.
[25] Christaller, W. (1935). Central Places in Southern Germany; English edition, 1966, translated by Carlisle W. Baskin, Englewood Cliffs, Printice-Hall, New Jersey.
[26] Losch, A. (1954). The Economics of Location (translated by Woglam W H), Yale University Press, New Haven.
[27] Fujita M, Krugman P, Venables A, 1999, The Spatial Economy- Cities, Regions and international Trade, MIT Press, Cambridge.

4.2 Urban association rules: uncovering linked trips for shopping behavior

Yuji Yoshimura, MIT SENSEable City Lab
Stanislav Sobolevsky, MIT SENSEable City Lab
Juan N Bautista Hobin, MIT SENSEable City Lab
Carlo Ratti, MIT SENSEable City Lab
Josep Blat, Universitat Pompeu Fabra

Abstract

In this article, we introduce the method of urban association rules and its uses for extracting frequently appearing combinations of stores that are visited together to characterize shoppers´ behaviors. The Apriori algorithm is used to extract the association rules (i.e., if -> result) from customer transaction datasets in a market-basket analysis. An application to our large-scale and anonymized bank card transaction dataset enables us to output linked trips for shopping all over the city: the method enables us to predict the other shops most likely to be visited by a customer given a particular shop that was already visited as an input. In addition, our methodology can consider all transaction activities conducted by customers for a whole city in addition to the location of stores dispersed in the city. This approach enables us to uncover not only simple linked trips such as transition movements between stores but also the edge weight for each linked trip in the specific district. Thus, the proposed methodology can complement conventional research methods, which are difficult to use because they mostly rely on small-scale samples. Enhancing understanding of people's shopping behaviors could be useful for city authorities and urban practitioners for effective urban management. The results also help individual retailers to rearrange their services by accommodating the needs of their customers' habits to enhance their shopping experience.

Keywords: shopping behaviors, association rule, transaction data, Barcelona

4.2.1. Introduction

In this paper, we explore the applicability of association rules (Agrawal et al., 1993) for extraction of combinations of visited stores for the analysis of linked trips for shopping behavior. The Apriori algorithm (Agrawal & Srikant, 1994), which is widely used and a rather simple yet robust method for market-basket analysis, is applied to our anonymized large-scale transaction dataset. This algorithm was originally designed to extract combinations of other items most likely to be purchased by a customer given a particular item that is already in his or her basket as an input. Instead of the purchased items in a single store, we try to extract the links between the stores in which people make transactions before or after visiting some focal shops, considering all transactions conducted in stores dispersed throughout the city. Thus, we can uncover the edge weight for each trip as the transition probability by comparing it with the general pattern of all other shoppers' behaviors in the given district.

An anonymized bankcard transaction dataset provides us with longer-term evidence that people perform transactions among stores as a digital footprint or "data exhaust" (Mayer-Schönberger & Cukier, 2013, p113). Unobtrusive observations (Webb et al., 1966) can be used to reconstruct customers' sequential movement between the stores in which they make purchases. This point makes our analysis different from previous studies, which have only attempted to "infer" people's activities during their trips and use them for linked trip analysis. For example, the datasets obtained from individual-based global positioning system (GPS)-enabled smartphones (Kazagli et al., 2014) are helpful for imputation processes, together with land-use data, to infer people's activities (Shen & Stopher, 2013). Media access control (MAC) address detection techniques such as Wi-Fi and Bluetooth can also be useful to detect the patterns of users' activities between stable places and infer the type of regular activity depending on place, the day of the time and season (Yoshimura et al., 2014).

Although these analytical methodologies can help alleviate the shortcomings of traditional travel surveys (Rasouli & Timmermans, 2014; Shoval & Issacson, 2006; Stopher & Shen, 2011; Bricka et

al., 2012) and greatly help in behavioral analysis, none of them can generate information regarding actual people's expenditures and their successive movements between stores for a longer term throughout a city.

This paper aims at complementing to the existing methods for analyzing linked shopping trips. Travel survey-based data are not adequate for monitoring the long-term characteristics of non-work/school trip behaviors, but transaction datasets can provide us with continuous and long-term travel information. Conversely, the latter's trip information is quite fragmentary and does not contain sufficient information about the socio-demographic characteristics of travelers, their trip purpose or transportation mode, which the former can typically provide. Alternatively, the combination of GPS technology and interviews or questionnaires can uncover the visited stores and customers' paths, attributions, visit motivation and expenditures, but this methodology can be applied only in spatially and temporally limited terms (e.g., only for a shopping street over a time span of a few days) due to the larger burden for respondents. The mentally quite demanding and time-consuming method results in small-scale samples. Thus, our proposed methodology can shed light on another aspect of shopping behavior that the conventional analysis cannot achieve.

We begin with an introduction of association rules and their potential contribution for linked trip analysis. This introduction is followed by a discussion of the proposed method, which also presents the study area, the data-collection technique, and the analysis methods that were used. Finally, the results are presented and examined. We conclude with a discussion of the implications for future shopping behavior research.

4.1.1.1 Mobility and linked trip study: collecting data from digital footprints

The subject of tracking shopping trips has long been recognized as a critical research area that has been underexplored. Although shopping is one of the most important trip-generating activities, it is a highly varied and flexible activity in both time and space, which make it differ from work/school out-of-home activities (Wang & Miller, 2014). One of the characteristics of shopping behaviors is

"the multipurpose and the purchasing of different items on a single trip" (Arentze et al., 2005). This indicates that traditional household surveys such as person trip surveys or usual single-day trip surveys (see Weiner, 1997 for a review) may have difficulty capturing shopping behaviors for behavioral analysis because the aforementioned methodology is based on trip-based data collection.

Figure 4.2.1 presents the example of linked trips during a day. All arrows represent the real trip of this person, but it is known that people tend not to report all trips, especially short trips around a destination (dotted red lines). Thus, his/her trip is likely to be represented only by the orange and green arrows. The problems regarding the conventional questionnaire-type travel surveys were discussed in Axhausen (1998). To overcome them, the activity-based approach is proposed to better understand people's travel behaviors and the decision structure underlying them, considering time use patterns (for an overview, see Ettema & Timmermans, 1997).

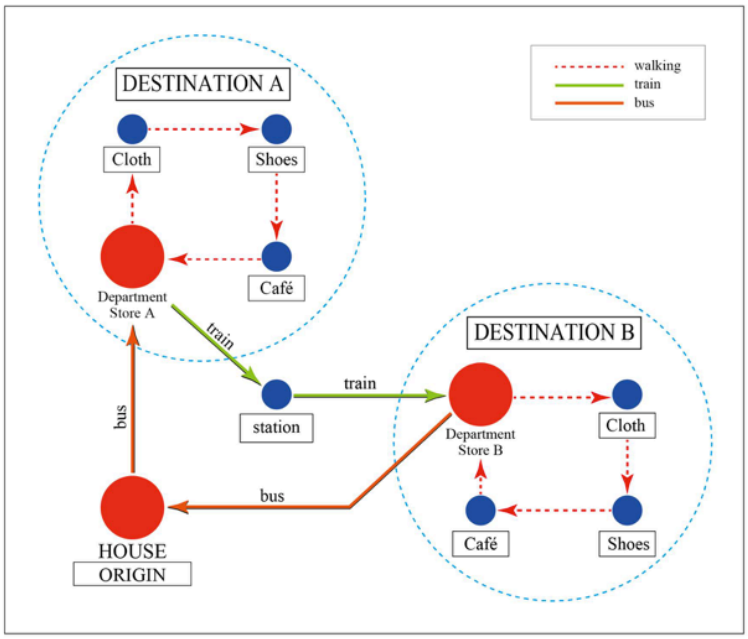


Figure 4.2.1. An example of linked trips for shopping purposes and the model's shortcomings.

Timmermans and his co-workers performed important research regarding linked shopping trips in the framework of modeling and analyzing the dynamics of pedestrian behavior (Borgers & Timmermans,1986a; Borgers & Timmermans, 1986b; Kurose et al., 2001; Borgers & Timmermans, 2005; Zhu & Timmermans, 2008; Timmermans, 2009; Dijkstra et al., 2009; Kemperman et al., 2009; Dijkstra et al., 2014).

Their methodology largely relies on on-site interviews and questionnaires to ask people at the entrances and exits of the shopping district about the following: the locations of the shops that they visited, their route choices, their expenditures, the starting and ending times of their shopping trip, and their mode of transport to the city center. Information about individual socio-economic and demographic variables and psychological factors is also collected. They sometimes complement these traditional qualitative methods with state-of-the-art technology (Axhausen et al., 2002). Comparative studies of GPS and travel survey data are discussed in Bricka & Bhat (2006), and a summary of the modeling of shopping destination choices is presented in Huang & Levinson (2015).

Those combinations enable us to collect human behaviors on a finer granular scale in space and time and greatly improve the quality of datasets for travel analysis (Rasouli & Timmerman, 2014). However, they are also not error-free. We identified the following points as the shortcomings of the previous studies, particularly for linked trips for shopping behaviors.

First, individual-oriented research methods, such as people-centric sensing (Miluzzo et al., 2008) [in contrast with methods conducted manually or that employ emerging technologies] can merely collect information about sample shopping activities in the specific area where data collection is performed. This raises the significant question: whether the observed behavior is unique to the district in which the sample was collected, - e.g., due to the characteristics of the distribution of stores, their types and their number (retail agglomeration) – or is independent of those environmental factors. The conventional research method is not suitable to address this question due to the shortcomings of datasets obtained using this data collection methodology (e.g., interviews or observations).

Second, the dataset used for people's linked trips is both temporally and spatially limited, as we described previously. The sample data collection is typically conducted to intercept customers at the entry/exit points of shopping areas over one day or a few days, resulting in a small number of samples. The results are ineffective for analyzing repetitive choice behavior and assessing the temporal effects of various policies.

Finally, there is a lack of robust tools to analyze the large-scale dataset of people- linked trips. As we described above, the incorporation of emerging technologies in the data collection process enables us to capture human behavior on finer granular scales in space and time and to increase the scale of the quantity of datasets. However, the increasing number of relevant datasets is not only expected to provide new sources for human behavioral analysis but also to cause the problems in the analytical process. We are familiar with analyzing small-scale samples but not yet prepared for large-scale datasets.

We try to complement the shortcomings of the above-mentioned methodology for the analysis of linked shopping trips. We propose to use credit card transaction datasets consisting of more than 100 million unique users and present an adequate methodology for analyzing such datasets. This enables us to analyze the temporally ordered origins and destinations of customers' spending behaviors all over the city rather than in limited areas, such as shopping districts. This unique dataset and its analytical methodology also make it possible to compute the number of stores classified into categories in each area and the number of transactions made in each category in each area. Thus, we can uncover not only people's transition movements between the stores at which they made transactions but also the edge weight as the transition probability in each district.

Our proposed methodology can greatly enhance our knowledge about people's shopping behaviors in the city. The results can provide a distribution map of people's shopping activities and the composition of stores. This is useful for urban managers and city authorities to identify the unevenness of the spatial distributions of people's shopping activities. In addition, our approach can work as a fast, inexpensive and robust filtering system to capture the

significant variables within the large-scale datasets. This greatly reduces costs before starting to explore the causality in a whole dataset. Thus, our proposed methodology is more useful for analysis based on a large-scale dataset rather than analysis based on small-scale samples.

4.2.2. Methodology: from association rules to urban association rules

The analytical framework for this paper is based on association analysis and classification of the obtained results. We analyze the correlations among the visited stores as discrete variables, resulting links between them. This is different from uncovering the causality by relying on randomly chosen small-scale samples, which can be most likely be best obtained from qualitative travel survey data or diary surveys that can elucidate the behavioral process in detail. Such studies also attempt to capture activity scheduling and rescheduling processes (Arentze & Timmermans, 2000; Timmermans, 2001; Miller & Roorda, 2003; Arentze & Timmermans, 2005; Nijland et al., 2009) because the trip could be the consequence of decisions at an earlier stage of behaviors. Hence, activity-based approaches can be used to overcome some of the typical shortcomings and limitations of a personal trip survey (Timmermans, 2005).

Conversely, the present work explores the correlations between the stores that customers visited by increasing the quantity of data because "when we increase the scale of the data that we work with, we can do new things that weren't possible when we just worked with smaller amounts" (Mayer-Schönberger & Cukier, 2013, p10).

Association analysis was proposed by Agrawal et al., (1993) to extract interesting relationships between items purchased by a customer in a market basket by using historical transaction datasets collected in stores (see Tan et al, 2005, pp327-414 for a review). The objective of this analysis is to extract combinations of items that frequently appear in the transaction datasets and seek hidden patterns embedded in the large-scale datasets.

Table 4.2.1. An example of market-basket transactions made based on Table 6.1. from Tan et al. (2005)

TID	Items
1	{Bread, Fruit, Jam}
2	{Diapers, Milk, Beer, Ham, Salad}
3	{Milk, Diapers, Beer, Fish, Salad}
4	{Fruit, Diapers, Bread, Milk, Jam, Salad, Beer}
5	{Bread, Milk, Diapers, Salad, Jam}

Table 4.2.1 presents an example of a set of items purchased by customers in a grocery store. Each row corresponds to a transaction by a given customer, which contains a unique identifier, i.e., a transaction ID (TID), and a set of items. For example, the following rule can be extracted from the table of transactions:

$$\{\text{Diapers}\} \rightarrow \{\text{Beer}\} \quad (1)$$

This is because many customers who purchase diapers also buy beer, thus indicating that there exists a strong relationship between the sales of these two items. The support count, σ (X) for an itemset X indicates the number of transactions that contain a particular itemset. For instance, in Table 4.2.2, the support count for {Diapers, Salad, Beer} is 3 because there are only three transactions that contain all three items.

Association rules are rules that surpass the minimum support and minimum confidence thresholds defined by a user. The strength of an association rule can be measured by the support, confidence and lift.

$$\text{Support, } s(X \rightarrow Y) = \sigma (X \cup Y)/N \quad (2)$$

$$\text{Confidence, } c(X \rightarrow Y) = \sigma (X \cup Y)/ \sigma (X) \quad (3)$$

$$\text{Lift, } (X \rightarrow Y) = \text{supp } (X \cup Y)/ (\text{supp } (X) \text{ supp } (Y)) \quad (4)$$

Support determines the frequency of the transaction that satisfies X and Y out of all of the transactions (N). If a rule has lower support, it might be happening simply by chance. Conversely, confidence is the conditional probability that Y (consequent) occurs given that X (antecedent) has happened (conf (X-> Y) = σ (X U Y)/σ (X). A

higher value of confidence indicates a higher reliability of the inference made by a rule (i.e., X → Y). That is, the higher the confidence, the more likely it is for Y to be present in transactions that contain X. Lift is the fraction of supp (X U Y)/(supp (X) supp (Y)). Greater lift values indicate stronger associations.

The concept and methodology of urban association rules is based on association rules. The significant difference is that whereas the former addresses the items purchased in an individual store, the latter focuses on the retail shops in which a customer completed a purchase and his/her successive purchasing activities among those shops dispersed in the urban settings. In urban association rules, an individual shop is analogous to an item in association rules.

For this paper, CONFIDENCE indicates the ratio of the number of transactions conducted in a particular category of the shop over the total number of transactions in all categories of the shop. LIFT is the fraction of B/A, where B= the number of transactions made in a particular category of the shop over the total number of transactions made in all categories in the same shop and A = the number of transactions made in the category in the district over the total number of transactions conducted in the district. Thus, LIFT reveals the unique feature of the targeted shops' customers' transaction activities compared with the general consumption behavior of people in the district. SUPPORT is the ratio of the number of transactions in a category to the total number of transactions in the district.

4.2.3. Study area and dataset

The present study employs the complete set of bank card transactions recorded by Banco Bilbao Vizcaya Argentatia (BBVA) during 2011 from throughout Spain. This study built on the earlier studies about the nature of datasets obtained through analysis of customers' spending behaviors. Sobolevsky et al., (2015a) examined a relationship between the sociodemographic characteristics of customers (e.g., age and gender) and their spending habits and mobility. Sobolevsky et al., (2015b) also researched how Spanish cities attract foreign customers by analyzing transaction activities.

4.2.3.1 Study area

We analyzed people's linked trips for purchase behaviors in the city of Barcelona, Spain. Barcelona is the capital of the autonomous community of Catalonia, which is located on the Mediterranean coast in northeastern Spain. Approximately 1.6 million people inhabit 100 square kilometers of land with a density of 159 hab/Ha. The city is surrounded by two rivers (the Llobregat and Besòs Rivers), a mountain (Collserolla) and the Mediterranean Sea.

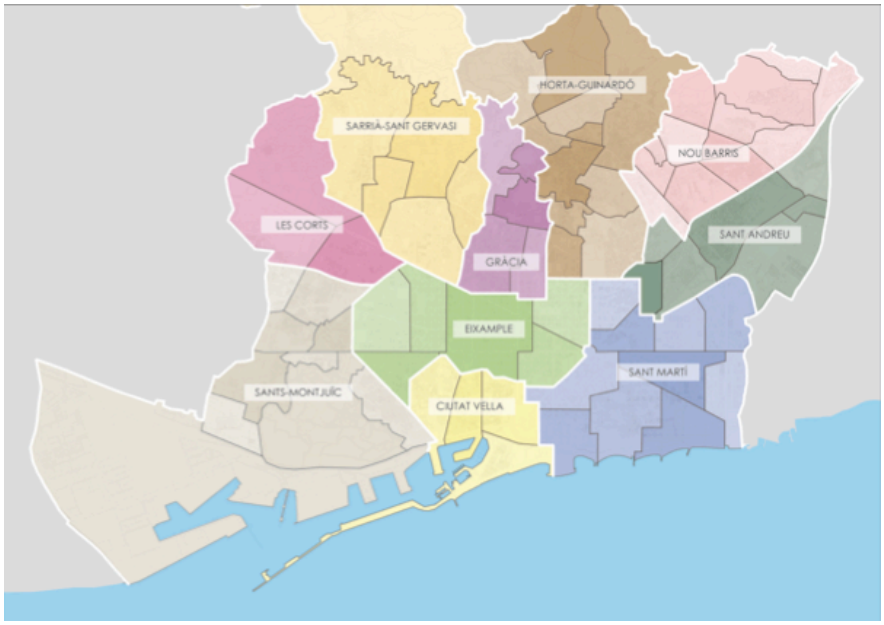


Figure 4.2.2. A map of the city of Barcelona. There are 10 districts, which contain 73 neighborhoods.

The city of Barcelona is divided into 10 administrative districts, and 73 neighborhoods within those districts, each of which has its own identity (see Figure 4.2.2). There are approximately 50,000 locations for economic activity throughout the city, including department stores, commercial centers, supermarkets, and shopping streets with international/local brands, in addition to lesser-known shopping areas such as neighborhood shops, markets and street fairs.

For the analysis in this paper, we chose the customers who made transactions in three large-scale department stores that belong to the

same chain but are located in different neighborhoods of the city (i.e., the shops AD, PA, and PC). We selected the same chain and large-scale department stores as the focal shops for the analysis rather than different and small- or medium-scale shops because (1) each one can be considered to be the strongest hub for shopping in each neighborhood, (2) we can expect a larger number of customer transactions due to the stores' higher attractiveness compared with small- and medium-scale retail shops, and (3) the composition of customers obtained from such shops can be more homogeneously distributed than when considering smaller retail shops.

We define the customers, who made transactions in any other shops before making transactions in any of the shops AD, PA, or PC as incoming customers. Similarly, we define customers, who made transactions in any other shops after making transactions in any of the shops AD, PA, or PC as leaving customers.

The association rule to be tested in this paper can be stated as follows:

$$\{\text{somewhere}\} \rightarrow \{\text{the shop AD or PA or PC}\} \quad (4)$$

$$\{\text{the shop AD or PA or PC}\} \rightarrow \{\text{somewhere}\} \quad (5)$$

Thus, we extract the consecutive transaction activities of incoming/leaving customers for the anchor shops AD, PA and PC.

4.2.3.2 Sample characteristics

The datasets used in this paper consist of anonymized bank card transaction data from two groups of card users: bank direct customers who are residents of Spain and hold a debit or credit card issued by BBVA and foreign customers who hold credit or debit cards issued by other banks and made transactions through one of the approximately 300,000 BBVA card terminals in Spain. For this paper, we mainly focus on this first group, customers who are residents of Spain, and in particular those who made transactions in the city of Barcelona during 2011. This results in 4.9 million transactions, which are the target of our analysis.

The data contain randomly assigned IDs for each customer connected with certain demographic characteristics, an indication of a residence location, the shop ID where a customer made a transaction, the date of the transaction, and the amount of money spent. The data were anonymized prior to sharing in accordance with all local privacy protection laws and regulations. Thus, all information that would enable us to identify a cardholder was removed.

Our dataset is possibly biased in terms of the representativeness of BBVA's customers in terms of the economically active population in the given area owing to the spatial inhomogeneity of the BBVA market share. Thus, before the analysis, pre-processing was required. To compensate for this potential bias, customers' activity was normalized by the respective market share in their residence location, which was provided by the bank at the provincial level. This allows us to estimate the total domestic customer activities (see Sobolevsky et al., 2015a).

The dataset was originally classified into 76 categories, which were further aggregated into 6 general categories with 13 sub-categories (see the Appendix). This was performed based on a previous study about the city of Barcelona (see Porta et al., 2012 for details). For example, the general category A of "retail commerce" was split into three sub-categories: those related to motor vehicles (A4) and those not related, where the latter group is further classified into daily use and non-daily use categories (A1 and A2, respectively). In addition, all of the general categories and sub-categories are grouped into daily and non-daily use categories by the following definition:

The non-daily use category contains shops "which, in themselves, bring people to a specific place because they are anchorages", whereas the daily use category contains "enterprises that grow in response to the presence of primary uses, to serve the people the primary uses draw" (Jacobs, 1961, pp.161-162).

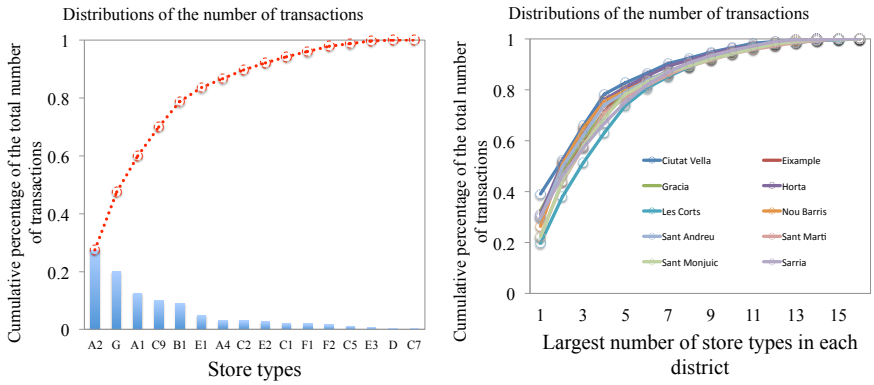
This classification enables us to analyze how customers visit shops in the daily use and non-daily use categories and their consecutive transactions as linked trips. Thus, our methodology permits us to uncover the relationships in customers' purchasing behaviors

between both types of shops, thereby revealing the characteristics of the districts and neighborhoods of the city of Barcelona.

4.2.4. Results

4.2.4.1 Analysis of supply of activity locations

The city of Barcelona is composed of approximately 50,000 stores. In the whole city, 70% of all stores are classified into the non-daily use category and 30% into the daily use category. They are unevenly distributed between districts. For instance, while more than 20% of the total number of stores are concentrated in the Eixample district, only 4.7% are located in the Les Corts district. This distribution is independent of the dimension of the district (r^2=0.0009) but slightly associated with the population of the district (r^2=0.53).

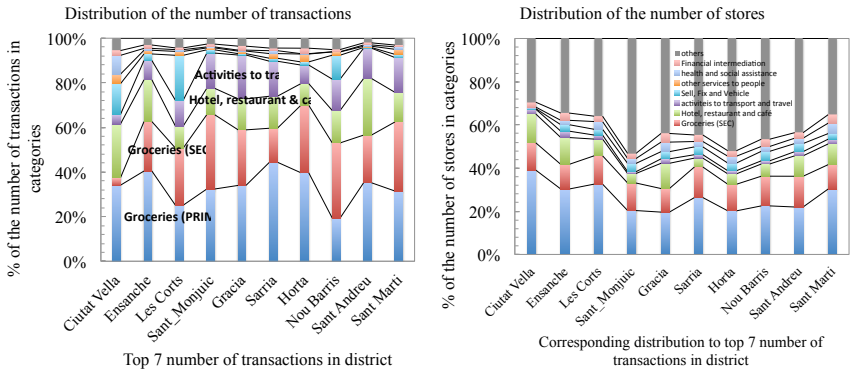


(a) (b)

Figure 4.2.3. (a) The cumulative distributions of the number of classified stores for the entire city. (b) The cumulative distributions of the number of classified stores for each district.

Figure 4.2.3 (a) presents the cumulative distribution of the number of stores classified into the category for an entire city. 27% of all stores are allocated to A2 (Groceries SEC) and 20% are G (vacant stores); thus, almost half of the total stores are composed of only these two categories. Together with three other categories (A1, C9, and B1), the total number of stores belonging to those categories increases to almost 80%. This indicates that the distribution of the number of stores in each district is largely distorted into a few

categories. In addition, the cumulative distribution ordered by the categories that have the larger number of stores is quite similar between districts [see Figure 4.2.3 (b)]. The largest and second largest categories are always the same in every district except Ciutat Vella: A2 and B1 are the largest for Ciutat Vella, while A2 and G are the largest for other districts.



(a) (b)

Figure 4.2.4. (a) Distributions of the number of transactions of the top 7 categories in the districts. (b) Distributions of the number of stores corresponding to the categories in (a).

However, the distribution of the number of transactions made in each district is quite different from the distribution of the number of stores in each district. Figures 4.2.4 (a) and (b) show that a larger number of stores does not necessarily result in a larger number of transactions. We computed the fraction of those two factors: the percentage of the number of stores in each district against all districts/the percentage of the number of transactions made in each district against all districts (Figure 4.2.5). Thus, a higher score (red color) indicates an excess supply of stores, and a lower score (blue color) suggests a shortage of supply compared with the number of realized transactions in each district.

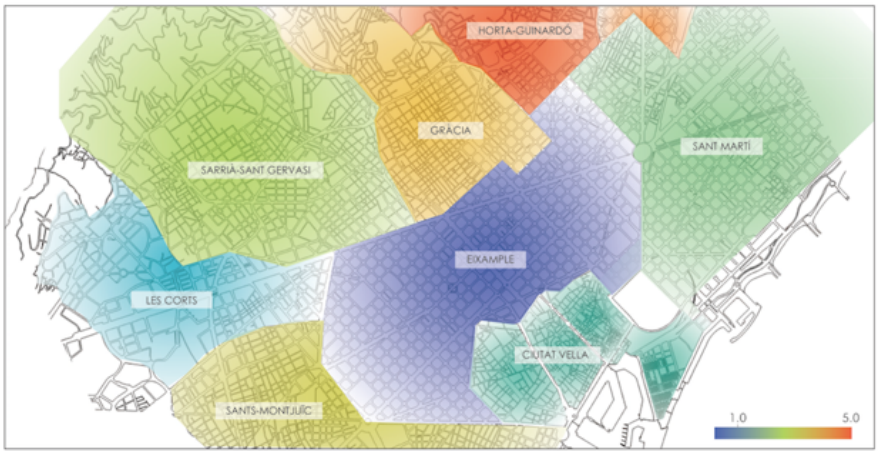


Figure 4.2.5. Visualization of the ratio of the number of stores and the number of transactions made in shops in the city. A higher score is indicated by red

Only the Eixample district presents a value less than 1.0, suggesting the demand (the number of realized transactions) is superior to the supply (the number of stores). Ciutat Vella and Les Corts show $1.0<x<1.5$, indicating that the volume of supply and demand are quite well balanced. However, 7 of 10 districts have much higher scores (i.e., 4.8 for Horta, 4.1 for Nou Barris, 3.4 for Sant Andreu, 2.9 for Gracia, 2.3 for Sant Monjuic, 2.0 for Sarria and 1.7 for Sant Marti), meaning the number of transactions is much smaller than the number of stores that exist in each district.

The obtained results are the basis for the following analysis regarding the focal stores' urban association rules. We attempt not only to extract the individual's linked trips to determine the shopping behaviors around each focal store but also to compare each of them with the general patterns of people's shopping behaviors in each district. Thus, we try to uncover the weight of each linked trip conducted by the individual customer around the focal shop in each district.

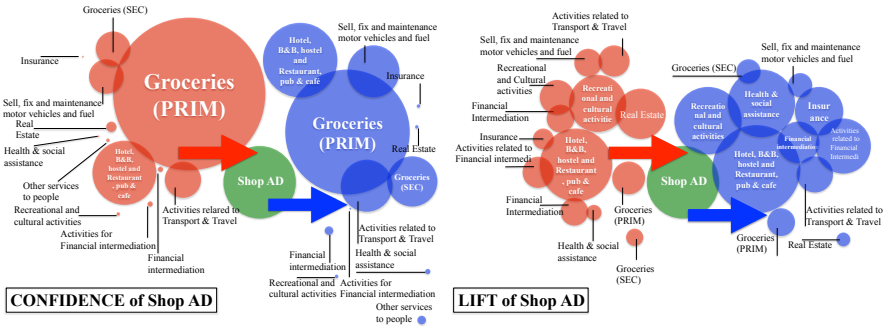
4.2.4.2 Linked trips defined by urban association rules

The linked trips for shopping behaviors in the city of Barcelona were evaluated by three indicators of urban association rules: CONFIDENCE, LIFT, and SUPPORT. Each of them captures one of the aspects of the customers' linked trips for the shopping

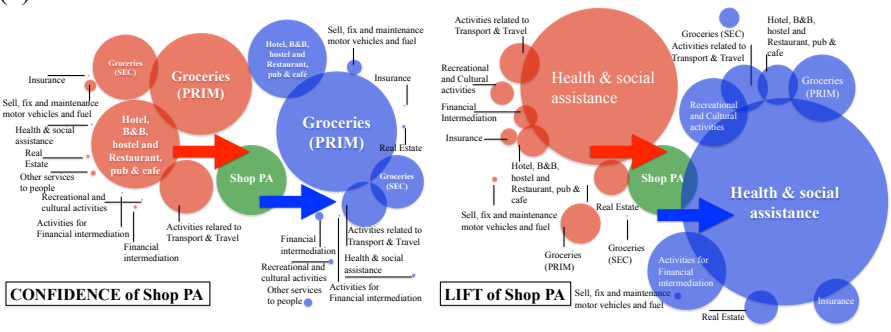
behaviors. Thus, further classification will be conducted by considering all values computed using these three indicators.

Figures 4.2.6 (a), (b) and (c) represent graphic examples of CONFIDENCE, LIFT, and SUPPORT, respectively, for the consecutive transactions made by customers around three focal stores (i.e., shop AD, shop PA, shop PC). These figures can be interpreted as follows: the red circle on the left side represents the categories that each focal shop's customers come from, and the blue circle on the right side describes the categories that the customers transit to make transactions after making the transaction in the focal shop. The size of the circle represents the relative magnitude compared with the green circle (=1.0). Considering the general patterns of people's shopping behaviors in each district in which the focal store is located, this representation allows us to depict the features of customers' behaviors with each focal shop.

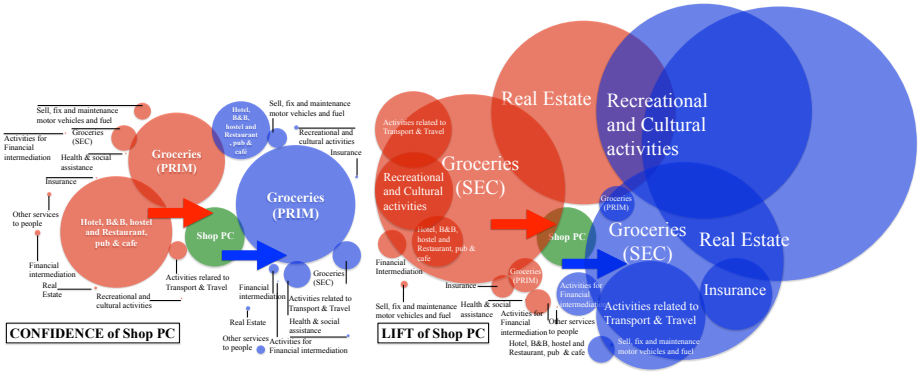
4.2.4.2.1 CONFIDENCE and LIFT



(a)



(b)



(c)

Figure 4.2.6. (a) (b) (c) Visualization of CONFIDENCE, which is the number of customers' consecutive transactions before and after visiting one of three shops. The red color shows the incoming customers, and the blue color presents the leaving customers.

Figure 4.2.6 (a), (b), and (C) illustrate differences in the linked trips for shopping behaviors in terms of CONFIDENCE and LIFT.

The transaction patterns of the customers of shops AD and PA are illustrated in figures 4.2.6 [(a), (b)]. The largest number of the customers of each focal shop derives from the category A2 [the groceries (PRIM)], and moves to the same category. Thus, the largest number of customers of those two focal shops is just moving between the same categories by way of each focal shop. Conversely, we can find a quite different pattern in shop PC's customers' activities [Figure 4.2.6, (c)]. Whereas the largest number of their customers come from B1 (hotel, B&B, hostel and restaurant, pub and café), the largest number of them move to category A2 [groceries (PRIM)]. This is a significant difference in terms of the origin-destination of the customers' linked trip for shopping behavior. For the former group, the focal store functions only as a transit location for their continuous activities of movement between the same category, but for the latter, the shop PC is the place where customers' transaction activities change in terms of their visited categories before and after visiting this focal store.

However, the larger quantity of customers' linked trips showed by CONFIDENCE does not necessarily indicate the unique characteristics of the customer's transaction activities in each store because it might derive from the general pattern of people's

shopping behaviors in the district where the focal shop is located. Thus, we examined LIFT [the right panels of Figure 4.2.6 (a), (b), and (c)].

The visualization of LIFT in Figure 4.2.6 clearly reveals that the PC's customers present extremely higher values of LIFT. This is especially true for the categories of C5 (real estate), A1 (groceries, SEC) and F1 (recreational). Because the higher score of LIFT indicates a greater difference in transaction activities between the customers and people in the district, the above-mentioned categories of the shop PC can be considered unique characteristics of their customers' activities.

Contrary to these facts, the LIFT of the shop AD is almost 1.0 in most of the categories [see the right panel of Figure 4.2.6 (a)]. This indicates that shop AD's customers' transaction activities are quite similar to the people's activities in the district where shop AD is located.

All of these analyses would not be possible if we had focused solely on the transitional probability based on consecutive transactions between stores. Such results are significant when we try to explain shopping behaviors and manage the district.

4.2.4.2.2 Patterns of customers' linked trips

The examination of the scoring of CONFIDENCE and LIFT generated five well-defined groups of customers' behaviors distinguished by the scores of both indicators. The categories that have a higher or lower score were recorded as characterizing the group as a whole. We applied the systematic classification of their behaviors. First, we examine the magnitude of the score of CONFIDENCE and LIFT together. Then, within the identified group, who hold both a higher score of CONFIDENCE and a higher or lower score of LIFT, we study the magnitude of the score of SUPPORT.

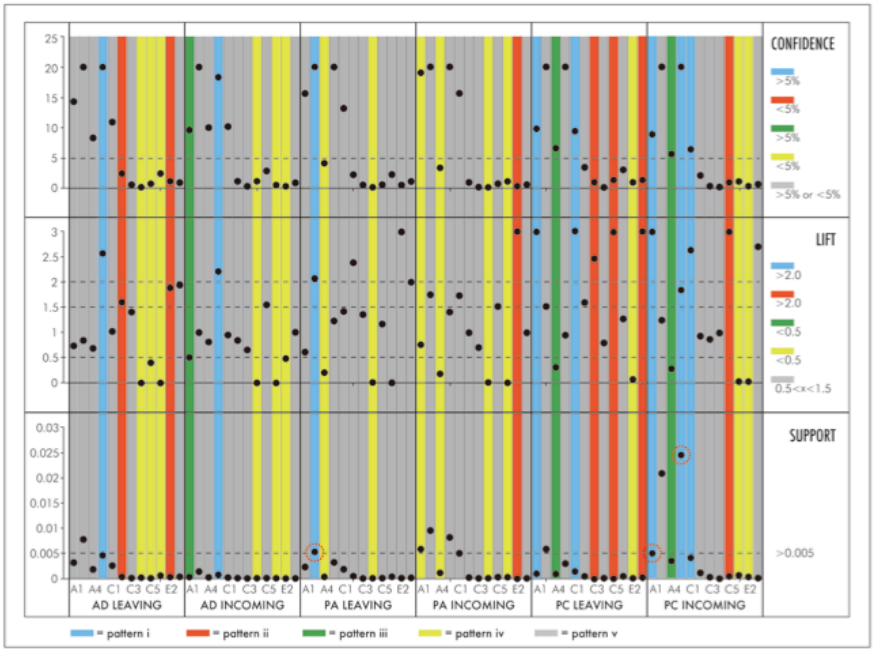


Figure 4.2.7. Visualization of classification of the customers by three indicators. We classify them into 5 types: (i) an extremely high value of LIFT (>2.0) with a high value of CONFIDENCE (>5%), (ii) an extremely high value of LIFT (>2.0) with a lower CONFIDENCE (<5%), (iii) an extremely low value of LIFT (<0.5) with a high value of CONFIDENCE (>5%), (iv) an extremely low value of LIFT (<0.5) with a low value of CONFIDENCE (<5%) and (v) a LIFT value of almost 1.0 (0.5<x<1.5) with higher or lower CONFIDENCE (>5% or <5%).

Figure 4.2.7 is a graphic of the five groups of customers: the blue color is pattern i, the red color is pattern ii, the green color is pattern iii, the yellow color is pattern iv and the grey color is pattern v (see the detailed classification in the Appendix).

Pattern i indicates that people visiting each district rarely made transactions in the indicated categories, but the customers of the focal shop made consecutive transactions in those categories after or before making transactions in the focal shop. In addition, the number of such customers of each focal shop who made the consecutive transactions in said categories is substantially higher. Conversely, the pattern ii indicates that the number of such customers is not significantly greater. Thus, pattern ii cannot be considered as a unique characteristic of the customers' behaviors in each focal shop.

The pattern iii indicates people who made transactions in each district, and the focal stop's customers behave in quite a different manner. In addition, the number of customers of each focal shop who made consecutive transactions in the indicated categories can be considered as not by chance because of the higher CONFIDENCE. This results in a unique feature of each focal shop's customers' behaviors.

The pattern iv indicates that people who made transactions in each district and the focal store's customers behave in different manners, but such correlations may be by chance due to the lower CONFIDENCE. Pattern v indicates that people made transactions in each district and the customers of the focal store behave in a quite similar manner, meaning that the purchase behaviors of each focal shop's customers derive from the features of the shopping behaviors of customers who visited the district rather than the focal shop itself.

The next stage in analyzing the data was to attempt to discover the higher value of SUPPORT (>0.005) within patterns i and iii, which show both higher values of CONFIDENCE and the extremely high or low value of LIFT. SUPPORT generates the probability of a detected correlation over the total number of transactions made for an entire city. Thus, the score of SUPPORT indicates the strength of the pattern, considering the whole city rather than only a district.

The three categories with the red dots in the SUPPORT area in Figure 4.2.7 present the final identified unique patterns among stores. The categories include shop PC's incoming customers in B1 and A2 and the shop PA's outgoing customers in A2.

4.2.5. Discussion and conclusion

Urban association rules provide us with more insight into linked trips for shopping all over the city. Due to the introduction of the state-of-the-art-technologies in retail stores, the retail planning and marketing strategies inside stores are greatly improved (Pantano & Timmermans, 2011). However, the area development based on the analysis of linked trips for shopping is still limited. In addition,

there is a lack of robust tools to enable us to analyze the large-scale datasets of people's purchase behaviors. Our proposed methodology can fill in these gaps.

The current study suggests the following direct benefits for transport and urban studies researchers:

(1) Five of six groups from three focal shops present similar behaviors with regard to the category B1 (hotel, hostel B&B and restaurant, pub and café), but the corresponding LIFT values vary greatly. For example, the analysis of LIFT indicates that there is a large difference in purchasing activities in this category between shop AD's customers and people in the district where shop AD is located. The former made many more transactions than the latter. This knowledge is useful for urban planners and neighborhood associations in their efforts to economically develop regular areas, revitalize deteriorated districts or re-habilitate neighborhoods. For example, because shop AD's customers' origin and destination is known in terms of the shop's category, coupons or royalty cards can be introduced between the relevant stores to increase the number of transactions in this district. Conversely, shop AD can provide similar coupons for their customers to use in stores of category B1 in other districts. Thus, city planners together with the commercial entities can attract or distribute more customers depending on the strategy of the district.

(2) Similarly, our proposed methodology is helpful for city authorities to evaluate the commercial planning for a specific district, overviewing its balance of an entire city. For example, in the district of shop PA, 34.8% of existing shops are classified into the category A2. However, people, who visited this district did not make many transactions in this category. In contrast, customers of shop PA were likely to make transactions in this category within the area. The fraction between the number of stores of category A2 and the number of transactions conducted in this category is 2.15. This reveals that the number of stores was almost two times greater than the number of transactions conducted in this category. This analysis is useful for city authorities to determine whether they should allow or limit the opening of new stores of the same or different categories within the area. This makes the city more efficient and well balanced.

(3) More generally, the application of LIFT to a transaction dataset is helpful for creating an overview the situation of a city and the differences between districts. In the beginning of the planning process, the analysis of the districts that comprise a whole city is more important than the deep analysis of the districts in detail. That is, the quantity from making a larger number of proposals is more useful than the quality of each one in the early stages. Although causality must be explored, a quick draft should be made before performing a deep analysis. This greatly reduces the cost in terms of people and time for urban planners and city authorities. In this manner, this analysis enables us to evaluate several drafts and compare them for further more-elaborate planning.

(4) The dataset used for the proposed methodology derives from secondary data collected for other purposes, i.e., is a byproduct of people's activities. The transaction datasets are obtained when people make a transaction for the purchase of items in the store as a digital footprint or "data exhaust" (Mayer-Schönberger & Cukier, 2013, p113). This signifies that there is no need for further costs to collect the data for the linked trip analysis. The proposed method is highly helpful for transport and urban studies researchers or practitioners to continuously monitor and overview a person's cash flow in purchasing behaviors as a supplementary dataset without incurring additional costs.

Despite the obvious value of this method, it does have several limitations. Our analysis is based on the successive order of customer purchases between different shops in which the customer uses BBVA's credit card, meaning that we cannot detect transactions and interactions at locations where the customer does not use BBVA's credit card. In addition, our dataset cannot reveal the customer's decision-making process or value consciousness, which underlies the organization of activities in time and space, because it does not contain information about his/her inner thoughts, which is typically derived from interviews, questionnaires and participatory observation. Thus, the scope of this paper excludes generation of the daily activity-travel patterns of individuals, considering their socio-demographic characteristics. This indicates that we need to enrich the models via traditional data collection means such as household surveys. Moreover, our dataset

contains a possible bias in terms of the representativeness of BBVA's customers relative to the economically active population in the given area. This potential bias is associated with the spatial inhomogeneity of BBVA market share.

Thus, the obtained results in this paper are subject to further research or validation. It is also clear that they are likely to be more useful in combination with other types of data through data fusion rather than in substitution. Hence, data fusion is one of the key approaches to use the results for urban and transportation planners for longer-term planning.

The application of association rules in the context of the city would allow researchers to create linked trips for shopping behaviors and thus create typologies of their activities, considering a particular case of three shops in Barcelona. Although the methodology presented herein was applied to these particular sources of data, it could easily be generalized to other focal stores in an arbitrary urban context, and also to other equivalent types of data.

Overall, the method offers an effective means to extract the patterns of the linked trip for the shopping behaviors from bank card transactions and by doing so may also present new methods of analyzing such data. Our research considered the locations where customers made transactions rather than considering locations that the customers simply passed by. Most of the previous research, which used human movement datasets obtained using state-of-the-art-technologies, largely depended on inferring people's activities by combining other types of data (e.g., land use data) with the reconstructed paths (Alexander et al., 2015). Such datasets are not likely to contain people's expenditures during their trip. This is a piece of critical information that was not obtainable prior to this study.

Acknowledgements. We would like to thank the Banco Bilbao Vizcaya Argentaria (BBVA) for providing the dataset for this study. Special thanks to Juan Murillo Arias, Marco Bressan, Elena Alfaro Martinez, Maria Hernandez Rubio and Assaf Biderman for organizational support of the project and stimulating discussions. We further thank MIT SMART Program, Accenture, Air Liquide, The Coca Cola Company, Emirates Integrated Telecommunications

Company, The ENEL foundation, Ericsson, Expo 2015, Ferrovial, Liberty Mutual, The Regional Municipality of Wood Buffalo, Volkswagen Electronics Research Lab and all the members of the MIT Senseable City Lab Consortium for supporting the research.

References

Agrawal R, Imielinski T, Swami A, 1993, "Mining association rules between sets of items in large databases" In Proceedings of the 1993 ACM SIGMOD conference Washington, DC, USA, May 1993 207-216

Agrawal R, Srikant R, 1994, "Fast algorithms for mining association rules" In VLDB'94 487-499

Alexander L, Jiang S, Murga M, González M C, 2015, "Origin-destination trips by purpose and time of day inferred from mobile phone data", Transportation Research Part C, 58, 240-250

Arentze A A., Oppewal H., Timmermans H J P., 2005, "A Multipurpose Shopping Trip Model to Assess Retail Agglomeration Effects", Journal of Marketing Research vlo. XLII, 109-115

Arentze T A, Timmermans H J P, 2000, Albatross: A Leraning-Based Transportation Oriented Simulation System, European Institute of Retailing and Service Studies, Eindhoven

Arentze T A, Timmermans H, 2005, "An analysis of context and constraints dependent shopping behaviour using qualitative decision principles", Urban Studies, 42, 435-448

Axhausen K W, 1998, "Can we ever obtain the data we would like to have?", in T Garling, T Laitila, K Westin (eds), Theoretical foundations of travel choice modelling, 305-333, (Oxford, UK: Pergamon Press)

Axhausen K W, Zimmermann A, Schönfelder S, Rindsfüser G, Haust T, 2002, "Observing the rhythms of daily life: a six-week travel diary", Transportation 29 (2), 95-124

Borgers, A.W.J., & Timmermans, H.J.H. (2005). Modelling pedestrian behaviour in downtown shopping areas. Proceedings of CUPUM 05, Computers in Urban Planning and Urban Management, 30-Jun-2005, London, (pp.83-15). London: Center for Advanced Spatial Analysis- University College London.

Borgers, A.W.J., & Timmermans, H.J.P. (1986a). A model of pedestrian route choice and demand for retail facilities within inner-city shopping areas. Geographical Analysis, 18, 115-128.

Borgers, A.W.J., & Timmermans, H.J.P. (1986b). City center entry points, store location patterns and pedestrian route choice behaviour: A microlevel simulation model. Socio-Economic Planning Sciences, 20, 25-31.

Borgers, A.W.J., Kemperman, A.D.A.M., & Timmermans, H.J.P. (2009). Modeling pedestrian movement in shopping street segments. In H.J.P Timmermans, H. (Ed.), Pedestrian Behavior: Models, Data Collection and Applications, (pp. 87-111). Bingley, UK: Emerald Group Publishing Limited.

Bricka S, Bhat C R, 2006, "A comparative analysis of GPS-based and travel survey-based data", In Proceedings of the 85th Annual Meeting of the Transportation Research Board, Washington, DC: Academic Press

Bricka S, Sen S, Paleti R, Bhat C R, 2012, "An analysis of the factors influencing differences in survey-reported and GPS-recorded Trips", Transportation Research Part C, Emerging Technologies, 21 (1), 67-88

Dijkstra, J., Timmermans, H.J.P., & Jessurun, A.J. (2014). Modeling planned and unplanned store visits within a framework for pedestrian movement simulation, Transporation Research Procedia, 2, 559-566.

Dijkstra, J., Timmermans, H.J.P., & Vries, B. de (2009). Modeling Impulse and Non-Impulse Store Choice Processes in a Multi-Agent Simulation of Pedestrian Activity in Shopping Environments. In H.J.P. Timmermans (Ed.), Pedestrian Behavior: Models, Data Collection and Applications (pp. 63-87). Bingley: Emerald Group Publishing Limited.

Ettema D, Timmermans H, 1997, Activity-based approaches to travel analysis, Oxford, Pergamon.

Huang A., Levinson D., 2015, "Axis of travel: Modeling non-work destination choice with GPS data", Transportation Research Part C 58, 208-223

Kazagli E, Chen J, Bierlaire M, 2014, "Individual Mobility Analysis Using Smartphone Data",in Rasouli S, Timmermans H (eds.) Mobile Technologies for Activity-Travel Data Collection and Analysis, 187-208

Kemperman, A.D.A.M., Borgers, A.W.J., & Timmermans, H.J.P. (2009). Tourist shopping behavior in a historic downtown area, Tourism Management, 30, 208-218.

Kurose, S., Borgers, A.W.J., & Timmermans, H.J.P. (2001). Classifying pedestrian shopping behaviour according to implied heuristic choice rules, Environment and Planning B: Planning and Design, 29, 405-418.

Mayer-Schönberger V, Cukier K, 2013, Big Data: A Revolution That Will Transform How We Live, Work and Think (John Murray, London)

Miller E J, Roorda M J, 2003, "A prototype model of 24-h household activity scheduling for the Toronto Area", Transportation Research Record number 1831, 114-121

Miluzzo E, Lane N D, Fodor K, Peterson R, Lu H, Musolesi M, Eisenman S B, Zheng X, Campbell A T, 2008, "Sensing meets mobile social networks: The design, implementation and evaluation of the cenceme application", In Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems, 337-350, Berlin, Germany: ACM

Nijland E W L, Arentze T A, Borgers A W J, Timmermans H J P, 2009, "Individuals activity-travel rescheduling behaviour: experiment and model-based analysis", Environment and Planning A, 41, 1511-1522

Pantano E, Timmermans H, 2011, Advanced Technologies Management for Retailing: Frameworks and Cases (IGI Global, USA)

Porta S, Latora V, Wang F, Rueda S, Strano E, Scellato S, Cardillo A, Belli E, Càrdenas F, Cormenzana B, Latora L, 2012, "Street centrality and the location of economic activities in Barcelona" Urban Studies, 49 (7): 1471-1488

Rasouli S, Timmermans H, 2014, Mobile Technologies for Activity-Travel Data Collection and Analysis (IGI Global, USA)

Shen L, Stopher P R, 2013, "A process for trip purpose imputation from global positioning system data", Transportation Research Part C, Emerging Technologies, 36, 261-267. doi:10.1016/j.trc.2013.09.004

Shoval, N., & Issacson, M. (2006). Application of tracking technologies in the study of pedestrian spatial behavior, The Professional Geographer, 58, 172-183.

Sobolevsky S, Bojic I, Belyi A, Sitko I, Hawelka B, Arias J M, Ratti C, 2015, "Scaling of city attractiveness for foreign visitors through big data of human economical and social media activity", arXiv preprint arXiv:1504.06003. IEEE Big Data Congress'2015.

Sobolevsky S., Sitko I., Tachet des Combes R., Hawelka B., Murillo Arias J., Ratti C, 2015, "Cities through the Prism of People's Spending Behavior", arXiv preprint arXiv:1505.03854

Stopher P, Shen J, 2011, "In-depth comparison of global positioning system and diary records", Transportation Research Record, 2246 (1), 32-37

Tan P N, Steinback M, Kumar V, 2005, Introduction to Data Mining (Addison Wesley)

Timmemans, H.J.P. (2009). Pedestrian Behavior: Models, Data Collection and Applications, Bingley, UK: Emerald Group Publishing Limited.

Timmermans H J P, 2001, Models of activity scheduling behaviour, Stadt, Region, Land, No. 71, pp.33-47, Institut Fur Stadtbauwessen und Stadtverkehr, RWTH Aachen, Aachen

Timmermans H, 2005, Progress in Activity-Based Analysis, Oxford, Elsevier

Wang J., Miller E.J. 2014, "A prism-based and gap-based approach to shopping location choice", Environment and Planning B: Planning and Design 41 977-1005

Webb., E.J., Campbell, D.T., Schwartz, R.D., & Sechrest, L. (1966). Unobtrusive Measures: Nonreactive Research in the Social Sciences, Chicago, IL: RandMcNally.

Weiner E, 1997, Urban Transpotation Planning in the United States: A historical Overview, Fifth Edition, DOT-T-97-24, Technology Sharing Program, U.S. Department of Transportation, Washington, D.C., 1997

Yoshimura Y, Sobolevsky S, Ratti C, Girardin F, Carrascal J P, Blat J, Sinatra R, 2014, "An analysis of visitors' behaviour in The Louvre Museum: a study using Bluetooth data" Environment and Planning B: Planning and Design 41 (6) 1113-1131

Zhu, W., & Timmermans, H.J.P. (2008). Cut-off models for the 'go-home' decision of pedestrians in shopping streets. Environment and Planning B: Planning and Design, 35, 248-260

5. Conclusions

This dissertation has dealt with spatial analysis through the large-scale datasets of human mobilities and their behaviors in the framework of architectural programing. As stated in the introductory chapter, this dissertation aimed to address three research sub-questions. We summarize the results, i.e. answers provided throughout the dissertation (mainly in Chapters 2 to 4), and the contributions.

5.1 Conclusions in relation to Research Questions

The human movement in the context of inside architecture

Question 1: Which are the factors that affect people's behaviors in architecture? How can we use those factors to improve people's experiences in architecture?

We discovered that the visiting style of short (less than 1:30 min) and long (more than 6 hours) are not as significantly different as one could expect. Both types of visitors tend to visit a similar number of key locations in a museum while the longer stay type visitors just tend to do so more extensively. This indicates that visitors' trajectories and the number of visited nodes are independent from the length of stay in a museum. We speculate that they could cause uneven distribution of the quantity of visitors, resulting in the congestion/vacancies in the museum spaces.

This finding clarified the spatial impacts on architectural spaces, which is not typical and traditional ways that architects and their community have been used. While the former tends to consider the hardware (i.e., building itself), our vision focuses on the software (i.e., human activities), in order to clarify features and drawbacks of architecture as a bottom-up.

These drawbacks largely derive from the lack of adequate datasets for the quantitative analysis in the field of architecture. There have been few means to capture and collect their behaviors in a quantitative way such as large-scale datasets. As a consequence, the

scientific method is rarely applied for the design process of architecture.

Our methodology and the obtained results present the alternative way to the conventional design approach for the architecture, and help to uncover the spatial problems deriving from people's activities. Indeed, we showed that the obtained results started to shed light on some of the unknown aspects of human behaviors, which we described above. They can be basic information for re-designing, renovating or rehabilitating architecture.

The human movement in the context of urban district

Question 2: Which are the factors that affect people's behaviors in the urban district? How can we use those factors in urban planning to enhance their experiences?

We uncovered that pedestrians in a discount day tend to stay shorter than the ones in the normal days, but the former visit the larger number of nodes than the latter. In a discount day, they actively explore the district by visiting all nodes including the Cathedral, which is rarely visited by pedestrians in other normal days. We found, however, pedestrians in a discount day tend to spend longer time in some streets than the ones in other normal days, depending on the street where they visit. Finally, we found that pedestrians' sequential movements between nodes have patterns in terms of the number of visited nodes and their order. The most of pedestrians use a few path types, and most of path types are used only by a few pedestrians. This tendency is getting stronger in a discount day than in other normal days.

These findings were not possible prior to our study, largely, due to the shortage of the adequate datasets. Therefore, architects tend to make a urban planning for the commercial districts based on their previous experiences, intuitions, imaginations and artistic drawings, rather than relying on the scientific analysis of human behaviors.

For instance, the interview-based research tends to remain just at small-scale samples, which are meant to represent "typical" pedestrian behavior. The modeling and simulation approaches (Borger and Timmermans, 1986) use small-scale sample collected

manually for the purpose of validating their models. The passive mobile phone detection (Gonzales et al., 2009; Ratti, et al., 2006) cannot detect human movement in fine grade such as the street scale, and GPS is only possible to make small samples due to the necessity of providing the mobile devices to experimenters in advance. Bluetooth detection is frequently used to collect pedestrian sequential movement (Kostakos et al., 2010; Delafontaine et al., 2012), but they never apply to reveal pedestrians' behavioral differences between the discount day and normal days.

Against these situations, our findings shed some light on pedestrians' behaviors and their differences during discount sale and normal days, in terms of their sequential movement and length of stay in the district. These findings show that collecting relevant datasets about pedestrians' behaviors, could be a basic information for use when regulating pedestrians' incoming and outcoming flow in a district. This would be helpful for architects in renovating and rehabilitating processes for a district. The rehabilitation of a district is a significant architect task for, but this planning is normally conducted without the information we show can be obtained. Perhaps rearranging the stores location in a district, pedestrians' behaviors could be altered in some ways (i.e, length of stay in the district), resulting in enhancing the quality of the district.

The human movement in the context of over the city

Question 3: Which are the factors that generate human mobility over the city? How can we use those factors to improve the attractiveness of the district?

We discovered that five of six groups from three focal shops selected in the city of Barcelona present similar behaviors with regard to the category B1 (hotel, hostel B&B and restaurant, pub and café), but the corresponding LIFT values vary greatly. For example, the analysis of LIFT indicates that there is a large difference in purchasing activities in this category between shop AD's customers and people in the district where shop AD is located. The former made many more transactions than the latter. Also, we found that, in the district of shop PA, 34.8% of existing shops are classified into the category A2, although people, who visited this district did not make many transactions in this category.

The fraction between the number of stores of category A2 and the number of transactions conducted in this category is 2.15. This reveals that the number of stores was almost two times greater than the number of transactions conducted in this category.

These knowledge are useful for urban planners and neighborhood associations in their efforts to economically develop regular areas, revitalize deteriorated districts or re-habilitate neighborhoods. For example, because shop AD's customers' origin and destination is known in terms of the shop's category, coupons or royalty cards can be introduced between the relevant stores to increase the number of transactions in this district. Conversely, shop AD can provide similar coupons for their customers to use in stores of category B1 in other districts. Thus, city planners together with the commercial entities can attract or distribute more customers depending on the strategy of the district. Also, our proposed methodology is helpful for city authorities to evaluate the commercial planning for a specific district, overviewing its balance of an entire city. This analysis can be a basic information for city authorities determine whether they should allow or limit the opening of new stores of the same or different categories within the area. This makes the city more efficient and well balanced.

The answer to the main research question

Question: How can the scientific analysis through datasets improve architecture and urban planning?

We answer to this main research question by considering the findings obtained from above-mentioned three research sub-questions.

In case of the museum, we discovered behavioral differences between the shorter stay type visitors and the longer stay type visitors. The result shows the similarity of both types of visitors in terms of their paths, although their length of stay is different. In case of urban space, we found, in a discount day, they explore more places than in the normal days, but their length of stay in the district is shorter than the others. Also, we discovered that, in the former case, the shorter stay type visitor tends to visit the similar number of nodes as the longer stay type visitor does. This indicates that the

shorter stay type visitors tend to explore the museum by visiting the same number of nodes within the limited length of stay in the museum. Conversely, in the latter case, we discovered that pedestrians at a discount day tend to explore more nodes, but their length of stay is shorter than the ones in the normal days.

Regarding the customer's behaviors over the city, we discovered that there are similarities and dissimilarities of purchasing behaviors between the local people in the district and the customers in the specific focal shop located in the said district. For instance, the five of six groups from three focal shops selected in the city of Barcelona present similar behaviors with regard to the category B1 (hotel, hostel B&B and restaurant, pub and café). Also, we discovered the distribution of the number of transactions made in each district is quite different from the distribution of the number of stores in each district. We computed the fraction of those two factors: the percentage of the number of stores in each district against all districts/the percentage of the number of transactions made in each district against all districts (see Figure 4.2.5). Thus, a higher score (red color) indicates an excess supply of stores, and a lower score (blue color) suggests a shortage of supply compared with the number of realized transactions in each district.

Our findings provide important insights for understanding human mobility in architecture and urban areas. All of these findings and obtained results can support that our proposed methodology is helpful for improving the architecture and urban planning. We showed that they enable us to uncover some of the unknown aspects of human behaviors in architecture and urban district.

This indicates that these scientific analyses of human behaviors can improve the quality of architecture and urban planning, especially, in the framework of Architectural Programing (Cherry, 1998). The science-based software design process, rather than the intuition or art-based hardware design, can improve the second step of Architectural Programing (Collect and Analyze Facts) through the scientific analysis of human mobility in the built environment. The presented methodologies can add different perspectives and approaches to the traditionally conducted ones by architects for their pre-design process for the architecture.

Finally, they enable us to provide the basic information for the pre-design process for architecture and urban planning, and the scientific analysis and its results makes it possible to increase the quality of architecture and urban planning.

5.2 Limitations

The obtained results of this dissertation revealed the features of urban spaces and their differences through analyzing the large-scale datasets of human movements and behaviors. Thus, we clarified unknown aspects of spatial impacts derived from the analysis of human activities in a built environment. However, this doesn't indicate that the employed methodologies and results are free from the limitations and drawbacks.

Firstly, this research and obtained result contains the bias due to the selected and applied methodology for the data collection (i.e., Bluetooth detection technique, transaction data registered by credit card usage). In the former case, the elder and children is not likely to possess mobile devices and have Bluetooth activate on them. In the latter case, although the credit cards are largely penetrated in Spanish society, many of them might not be familiar with using them yet, because of their traditional and cultural habits. Thus the analytical targets and the scope of the results to be able to cover for either case can be limited in specific persons and groups, who have unique attributes and habits.

Secondly, this research cannot clarify neither the inner thoughts nor attributes of people, indicating that we cannot reveal the cause of their behaviors and activities. This drawback largely derives from the methodology applied for this research. We employed systematic observation in the framework of the "unobtrusive measures" (Webb et al. 2000), making use of unconsciously left visitors' digital footprint. This enables us to obtain people's sequential movements between key places and their length of stay at these places in a consistent way. This is a very different approach to traditionally conducted qualitative research methods. The qualitative methods largely depend on observations, interviews and questionnaires, which disclose the inner thoughts and introspective aspects of pedestrians.

This dissertation focused on dealing with the physical presences of human movements with a large-scale datasets rather than their attributes and inner thoughts with the small-scale datasets.

However, even with these shortcomings, the contribution made by this dissertation suggests that there is space for continue exploring research methods and techniques to find other aspects of human behaviors, and applying them for other large-scale datasets.

5.3 Future work

With respect to the museum management, we are now developing indicators for crowd management. The dynamic estimation of the density in a specific art work (i.e., Venus de Milo) could be associated with the length of stay around there. This uncovers visitors' comfortability depending on the change of the density such as higher or lower. The better understanding of visitors' behaviors and its prediction based on the extracted patterns will enable us to optimize the spatial layout of objects, human resources and facilities including advertising. It could become the strong management tools not only for museum but also for urban environment.

With respect to the analysis of shopping behaviors, we try to extend data collection period from a month to several months covering different seasons. This is because the quantity of pedestrians largely depends on the seasons and days. The current research remains to uncover the similarity and dissimilarity of their behaviors by comparing the first Saturday of the discount period and other Saturdays for normal weeks during a month. The extension of data collection periods and datasets to be obtained will shed light on whether their behaviors might be specific to that period, or might change depending on the seasons, or not.

With respect to the analysis of shops' attracting power, we plan to analyze the distribution patterns of those retail shops over the city. Because the attracting power and distribution power might be different how those shops are agglomerated and distributed in each district. Thus, customers' mobility patterns are determined by those distributions of retail shops.

References

Borgers A, Timmermans H J P, 1986, "A model of pedestrian route choice and demand for retail facilities within inner-city shopping areas", Geographical Analysis, 18 (2) 115-128.

Cherry E, 1998, Programming for Design: From Theory to Practice, Wiley.

Delafontaine M, Versichele M, Neutens T, Van de Weghe N, 2012, "Analysing spatiotemporal sequences in Bluetooth tracking data" Applied Geography 34 659-668

González M C, Hidalgo C A, Barabási A L, 2008, "Understanding individual human mobility patterns" Nature 453 779-782

Kostakos V, O'Neill E, Penn A, Roussos G, Papadongonas D, 2010, "Brief encounters: sensing, modelling and visualizing urban mobility and copresence networks" ACM Transactions on Computer Human Interaction 17(1) 1-38

Ratti C, Pulselli R, Williams S, Frenchman D, 2006, "Mobile Landscapes: using location data from cell phones for urban analysis" Environment and Planning B: Planning and Design 33(5) 727-748

Webb E J, Campbell D T, Schwartz R D, Sechrest L, 2000, Unobtrusive measures: revised edition, Thousand Oaks: Saga Publications Inc.

6. ANNEX. Research activities surrounding this thesis

Through some national projects such as In4mo and Mobitrans (http://www.mobitrans.es/) as well as European Project such as ICING (Innovative Cities for the Next Generation), we implemented Bluetooth sensors in the city of Barcelona, Valencia and Zaragoza, and demonstrated its efficiency, compared with the existing data collection for the human mobility (Barcelo et al., 2012). Finally, Yoshimura et al., (2014) and Yoshimura et al., (2012) use Bluetooth proximity detection technique in a large-scale museum - the Louvre Museum - where data collections of visitors' behaviors is extremely difficult.

The author of this thesis involved in URUS project, which is European Union project. The aim of this project is to propose the networked connected autonomous robot in the urban setting. The author of this thesis contributed to analyze the relevant legal issues and legislations in European urban settings. The legal issue related with the privacy is frequently ignored by urban researchers, although the collected sample data may contain the sensitive information such as personal data. Thus, this paper uncovers the legal framework and the possibility of the implementation of sensors in urban settings.

Sanfeliu A., Llacer M.R., Gramnunt M.D., Punsola, A., Yoshimura, Y. (2010), "Influence of the privacy issue in the Deployment and Design of Networking Robots in European Urban Areas" Advanced Robotics 24 (13) 1873-1899.

Organized projects

In parallel, the author of this dissertation organized and involved in the research projects of my domains of investigations.

STRAIGHTSOL (STRAtegies and measures for smarter urban freIGHT SOLutions)(2011): Transport (Including Aeronautics): Call ID FP7-SST-2011-RTD-1, Proposal N. 285295.

CITYSOLVER: Pre-Trip&in-Trip integrated multimodal route plannerand traffic manager for sustainable cities(2010): ICT-2010.6.2: ICT for Mobility for the Future.

In4Mo- Sistema avanzado de información sobre la movilidad de personas y vehículos(2010): subprograma Avanza I+D, Ministerio de Industria Turismo y Comercio.

European STREP URUS Project (Ubiquitous Networking Robotics in Urban Settings)(2008-2010): EU Sixth framework program, priority 2, information society technologies, project number: IST-045062.

MOBITRANS- Tecnologias de informacion al viajero para la movilidad urbana sostenible (2009) Innovation of information technologies for travellers to foster the sustainable urban mobility, subvencion del Ministerio de Ciencia e Innovacion num.ref.del proyecto:MOBITRANS-E06/8

Organized Conference

In parallel, the author of this dissertation organized venues and setup events for researchers and professionals of my domains of investigation to share and discuss their works and thoughts.

Yoshimura, Y. & Kobayashi, I. (2015), Mobility and Opendata in Smart Cities, Yokohama City Council Economical Development Affairs, Yokohama, Japan, August, 08, 2015.

Yoshimura, Y. & Kobayashi, I. (2015), Tourism and Opendata in Smart Cities, Yokohama City Council Economical Development Affairs, Yokohama, Japan, December, 21, 2015.

Book Chapter

Yoshimura, Y. (2015) “Designing the technology and mobility” in Architects working abroad- the world is filled with opportunity, pp48-70, Gakugei publisher, Kyoto, Japan.

Other medias (non-scientific papers)

Yoshimura, Y. (2015), “The possibility of Big Data in visitors studies in the museum: a case study of the Louvre Museum“ in Japan Museum Management Academy (JMMA) p7-11, 75 20-2 (Japanease).

Yoshimura, Y. (2011), “Pedestrian Planning in Gracia as a central strategy for urban regeneration” in Sustainable Urban Regeneration vol. 14 p43-45, The University of Tokyo (Japanese).

Yoshimura, Y. (2011), “The light and shadow of Barcelona’ urban planning” in FORE, Future of Real Estate 69 (5): 8-9 (Japanese).

Interviews with Yoshimura Y. (2016), “a Japanese architect, who design the city of Barcelona“, Nippon.com, Your Doorway to Japan,

Interviews with Yoshimura Y. (2010), “New Generation of Japanese architects”, Nikkei Architecture 932, 8-9 p48 (Japanese).

A blog as a research tool

From September 2006, before my Ph.D coursework, to the present (September 2014), the author published a personal research blog (http://blog.archiphoto.info/). It started as an attempt to move my research notebook online to forge new connections and keep track of my thoughts. My blog became one of the most popular ones in Japan, having almost 5,000 visitors per a day.