

**O USO DE REDES DE SENSORIAMENTO
PARTICIPATIVO NO ESTUDO DO
COMPORTAMENTO DE TURISTAS**

ANA PAULA GOMES FERREIRA

**O USO DE REDES DE SENSORIAMENTO
PARTICIPATIVO NO ESTUDO DO
COMPORTAMENTO DE TURISTAS**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

ORIENTADOR: ANTÔNIO ALFREDO FERREIRA LOUREIRO

COORIENTADOR: THIAGO HENRIQUE SILVA

Belo Horizonte

Julho de 2016

ANA PAULA GOMES FERREIRA

**USING PARTICIPATORY SENSING IN THE
STUDY OF TOURISTS BEHAVIOUR**

Dissertation presented to the Graduate Program in Computer Science of the Federal University of Minas Gerais in partial fulfillment of the requirements for the degree of Master in Computer Science.

ADVISOR: ANTÔNIO ALFREDO FERREIRA LOUREIRO
CO-ADVISOR: THIAGO HENRIQUE SILVA

Belo Horizonte

July 2016

© 2016, Ana Paula Gomes Ferreira.
Todos os direitos reservados.

Gomes Ferreira, Ana Paula

F383u Using Participatory Sensing in the Study of Tourists
behaviour / Ana Paula Gomes Ferreira. — Belo
Horizonte, 2016
xxii, 71 f. : il. ; 29cm

Dissertação (mestrado) — Federal University of
Minas Gerais

Orientador: Antônio Alfredo Ferreira Loureiro

1. Partipatory Sensing Networks. 2. Pervasive Social
Computing. 3. Social Media. 4. Mobility. 5. Tourism.
I. Título.

CDU 519.6*22 (043)



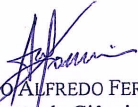
UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

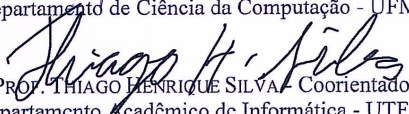
FOLHA DE APROVAÇÃO


Using participatory sensing in the study of tourists behaviour

ANA PAULA GOMES FERREIRA

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:



PROF. ANTONIO ALFREDO FERREIRA LOUREIRO - Orientador
Departamento de Ciência da Computação - UFMG


PROF. THIAGO HENRIQUE SILVA - Coorientador
Departamento Acadêmico de Informática - UTFPR


PROF. ALEXANDRE REIS GRAEML
Departamento Acadêmico de Informática - UTFPR


PROFA. DIOMIRA MARIA CICCI PINTO FARIA
Instituto de Geociências - UFMG


PROF. HEITOR SOARES RAMOS FILHO
Instituto de Computação - UFAL


PROF. PEDRO OLMO STANCIOLI VAZ DE MELO
Departamento de Ciência da Computação - UFMG

Belo Horizonte, 29 de julho de 2016.

Dedico esse trabalho a todos que acreditam em mim.

Acknowledgments

Esse é um dos momentos interessantes na vida: diante da realização de um sonho, começa a passar um filme desde ter uma ideia para algo que se tornou realidade. Algumas pessoas foram fundamentais nesse caminho longo e árduo e por isso hoje eu tenho motivos de sobra para agradecer.

Agradeço ao meu orientador Antônio Loureiro por ensinar sobre colaboração e humildade. Agradeço também ao Thiago Silva pelo tempo e esforço investido em me explicar tantas coisas; sua pesquisa foi uma inspiração. Eu tive a sorte de trabalhar ao lado de dois grandes pesquisadores.

Agradeço a minha família por todo o suporte dado, por acreditarem nas minhas decisões e serem a minha motivação. Agradeço a minha Bahia por sempre manter em mim a garra e o bom humor. Muito obrigada também aos amigos que deixei por lá, sempre corujas e compreensivos nos meus momentos de ausência. A crença em Deus me faz acreditar que o amanhã poderá melhor. Sou grata por esse otimismo incurável!

A comunidade nordestina na UFMG, representada pela Bahia e pelo Ceará, tem a minha gratidão. Foram muitas idas ao Cabral e muitas conversas; algumas felizes e outras tristes, mas sempre boas. Agradeço também aos amigos que Belo Horizonte me deu; vocês tornaram e tornam os meus dias mais leves e felizes. Obrigada ao WISEMAP, o laboratório com o melhor capital humano da UFMG. Obrigada por me ensinarem tanto e pelas risadas ao longo do caminho. Obrigada também ao BOAS Esporte, melhor time de futsal feminino do DCC, pelas amizades que me trouxe e tantos momentos divertidos.

Olhando pra trás e para o tempo presente é incrível poder dizer: eu consegui!!! Muito obrigada a todos que colaboraram para isso ser possível.

*“Meu caminho pelo mundo eu mesmo faço
A Bahia já me deu régua e compasso...”*
(Gilberto Gil)

Abstract

Tourism has become a global economic force, being responsible for approximately 10% of total world GDP. For this reason, offering better services to tourists is indispensable. With this goal in mind, in this work we study how tourists move through time and space and the factors that influence their movements in four major cities: London, Rio de Janeiro, New York and Tokyo. To perform this study we use data from social networking platforms, which are being massively and pervasively used, thanks to mobile devices with powerful networking and computing capabilities. We perform a large scale study of tourists mobility from several aspects. For example, we use a spatio-temporal graph model to study urban mobility of tourists, identifying where and when places are more important to users in the studied cities. In addition, we propose a new methodology, based on a topic model, that enables the automatic identification of mobility pattern themes, which, ultimately, leads to the better understanding of the profile of users. Our results have implications in several segments. In fact, we demonstrate possible uses of our results in a new itinerary recommendation system and how business owners could explore them to offer better service to tourists in different locations that could be culturally distinct.

Keywords: Partipatory Sensing Networks, Location-Based Social Networks, Social Media, Mobility, Tourism, Foursquare.

List of Figures

2.1	Illustration of a Participatory Sensor Network. This image was obtained from [Silva et al., 2014a]	7
3.1	Tweet with check-in information. In 1) the check-in URL, 2) the time and date, 3) geolocalization from Twitter.	14
3.2	Swarm's/Foursquare's page with check-in details. In 1) User name and 2) the name and link of the venue.	14
3.3	Venue page on Swarm's/Foursquare's page	14
3.4	Illustration of the complete process to collect check-in data	15
4.1	Distribution of the number of check-ins performed by tourists (green) and residents (blue)	20
4.2	Distribution of the time interval (in hours) between the check-ins performed by tourists (green) and residents (blue)	21
4.3	Places where tourists (green) and residents (blue) performed check-ins	23
4.4	Temporal check-in sharing pattern throughout the day by tourists and residents during weekdays	26
4.6	Check-ins frequency for each category by tourists and residents by city	26
4.5	Temporal check-in sharing pattern throughout the day by tourists and residents during weekend	27
4.7	Foreigners and Domestic Tourists by city	28
4.8	Places where foreigners (blue) and domestics (red) performed check-ins	29
4.9	Temporal check-in sharing pattern throughout the day by domestic and foreign tourists during weekdays	30
4.10	Check-ins frequency in each category by foreigners and domestics tourists by city	31
5.1	Distribution of Displacement of Tourists and Residents	34
5.2	Distribution of Radius of Gyration of Tourists and Residents	35

5.3	Visualization of the movement of users for different values of radius of gyration in Tokyo.	51
5.4	Illustration of the graph model considered	52
6.1	Distribution of the time interval (in hours) between the check-ins performed by Starbucks (green) and other Coffee Shops (brown) at New York	54
6.2	Subgraph of places visited by New York residents before and after other Starbucks	55
6.3	Subgraph of places visited by New York tourists before and after other Starbucks	56
6.4	Subgraph of places visited by New York residents before and after other Coffee Shops	57
6.5	Subgraph of places visited by New York tourists before and after other Coffee Shops	57
6.6	Yankee Stadium, New York	60
6.7	Jardim Suspenso do Valongo, Rio de Janeiro	61
6.8	Victoria and Albert Museum, London	62
6.9	Kanda Myojin Shrine, Tokyo	63

List of Tables

3.1	Number of check-in and unique users by dataset	15
3.2	Number of check-in by city and dataset	15
3.3	New classification to each category group of Foursquare	16
3.4	Number of tourists identified in each city	16
4.1	Ranking of most popular venues for Tourists	24
4.2	Ranking of most popular venues for Residents	25
5.1	Ranking of degree centrality of New York	38
5.2	Ranking of degree centrality of Rio de Janeiro	41
5.3	Ranking of closeness centrality of London	42
5.4	Ranking of betweenness centrality of Rio de Janeiro	44
5.5	Comparative ranking of degree centrality of the resident’s graph and null model graph of New York	45
5.6	Ranking of closeness centrality of the tourist’s graph and null model graph of London	46
5.7	Ranking of betweenness centrality of the resident’s graph and null model graph of Rio de Janeiro	46
5.8	Profiles of residents in Tokyo according to venues subcategory	47
5.9	Profiles of tourists in Tokyo according to venues subcategory	47
5.10	Profiles of residents in Rio de Janeiro according to venues subcategory	48
5.11	Profiles of tourists in Rio de Janeiro according to venues subcategory	48
5.12	Profiles of residents in New York according to venues subcategory during weekdays	48
5.13	Profiles of residents in New York according to venues subcategory during weekends	49
5.14	Profiles of tourists in New York according to venues subcategory during weekdays	49

5.15	Profiles of tourists in New York according to venues subcategory during weekdays	49
6.1	Profiles of customers of Starbucks who lives in New York	59
6.2	Profiles of customers of Starbucks who visits New York	59
6.3	Recommended places to go based on New York tourists'	60
6.4	Recommended places to go based on Rio de Janeiro tourists'	61
6.5	Recommended places to go based on London tourists'	62
6.6	Recommended places to go based on Tokyo tourists'	63

Contents

Acknowledgments	xi
Abstract	xv
List of Figures	xvii
List of Tables	xix
1 Introduction	1
1.1 Motivation	1
1.2 Objectives	2
1.3 Contributions	2
1.4 Work Organization	4
2 Related Work	5
2.1 Contextualization	5
2.1.1 Human as Sensors	5
2.1.2 Participatory Sensor Networks	6
2.2 Understanding Mobility	7
2.2.1 Studying Mobility Through Traditional Data	7
2.2.2 Studying Mobility Through Social Data	8
2.2.3 Mobility of Tourists and Residents	9
2.3 Applications based on tourist mobility	9
2.4 Discussion	10
3 Data Collection and Processing	13
3.1 Datasets	13
3.2 Identifying tourists and residents	16
3.3 Data limitations	17

4	Behavior of tourists in different cities worldwide	19
4.1	Number and time interval of check-ins	19
4.2	Places visited	21
4.3	Routines	23
4.4	Preferences of tourists	25
4.5	Domestic and foreign tourists	28
4.6	Discussion	32
5	Understanding Tourist’s Mobility	33
5.1	Displacement measures	33
5.1.1	Mean user displacement	33
5.1.2	Radius of Gyration	34
5.2	Centrality metrics on spatio-temporal urban mobility graphs	36
5.2.1	Spatio-temporal urban mobility graphs	36
5.2.2	Popular venues in the city	37
5.2.3	Spreading information	41
5.2.4	Bridge Places	43
5.2.5	Validating the results	44
5.3	Profiles of Tourists Based on Mobility Patterns	45
5.4	Discussion	50
6	Applications	53
6.1	Profile of consumers	53
6.2	Where should I go?	59
6.3	Discussion	64
7	Conclusion	65
	Bibliography	67

Chapter 1

Introduction

1.1 Motivation

We are in an era where social networking platforms are being massively and pervasively used, thanks to mobile devices with powerful networking and computing capabilities. Among these capabilities, resources like GPS has been used widely and in a global level. Some social networks support geolocalization features that allows users to share useful data about urban environments. These networks can be seen as a source of social sensing, called Participatory Sensor Networks, and it enables new research opportunities, such as those related with new patterns of users interactions in the city. For instance, Foursquare, one of the most popular PSN, allows users to share visited locations, enabling unprecedented opportunities for the large scale study of urban social behavior [Silva et al., 2014a].

Tourism truly has become a global economic and social force [Staab et al., 2002]. Tourists may have different desires from those in their home routines. In addition, factors such as cost, distance and personal preferences influence activities a tourist conducts in the visited city. Understanding how tourists move through time and space, and the factors that influence their movements, has important implications in several segments, ranging from transport development to destination planning.

The study of tourist movements is an under-explored facet of tourism scholarship [Lew and McKercher, 2006; Fennell, 1996]. Despite some efforts in the area, very few have attempted to model the actual movement patterns of tourists in large scale [Zheng et al., 2009; Yoon et al., 2010]. In this work, we show how we can use data shared by Foursquare users, the so-called check-ins, to better understand mobility of tourists that would be hard using traditional methods, such as surveys. A check-in is an action performed by the user to register and share his/her location at any given time. It is a

voluntary contribution provided by the user that allows the study of human behavior at different granularities, leading to a better understanding of urban areas, such as the identification of popular places [Silva et al., 2014a].

We consider spatio-temporal aspects of the behavior of tourists and residents. Spatial patterns are related to the different types of places available in the city. It is important to analyze this dimension because, for example, the number of check-ins at a given location may vary according to its popularity and category (i.e., a type of place, for instance, restaurant). Temporal patterns are related to events that occur at certain time slots. This is also another important dimension, since the behavior of users may vary, for instance, during different moments of the day. The joint treatment of these two dimensions is critical to understand the behavior of users and the dynamics of the city where a given person is.

1.2 Objectives

The main objective of this study is to answer the question: is it possible to use participatory sensor networks to study the behavior of tourists? To that end, a fundamental step is to evaluate the potential of using participatory sensor networks to extract useful properties of tourists and residents behavior in a city. Thus, we tackle the main objective of this study answering three different questions:

1. Which and when places are more important to tourists and residents? The temporal property influentiate on spatial choices?
2. Can we find unique properties of behavior of these two classes of users?
3. Can the mobility inside of cities to bring new information about tourists and residents behaviour?
4. Can we explore these properties for new services and applications?

1.3 Contributions

The contributions of this work can be summarized as follows:

- We show that we can have the opportunity to go one step forward in the understanding of tourists' mobility, identifying where and when places are more important to users in different cities. Based on data of Foursquare, we characterize the behavior of tourists and residents, showing, for instance, their preferences

and routines in four popular cities around the world in four continents: London, New York, Rio de Janeiro, and Tokyo. Besides that, we perform a large scale study of tourists mobility from several aspects. For example, we use a spatio-temporal graph model to study urban mobility of tourists of the studied cities. We show that it is possible to find popular transitions among tourists, and typical time that tourists visit certain places. This model also allows to identify central places in the tourist mobility and how they could be explored to evolve the urban computing area;

- We propose a new methodology, based on a topic model, that enables the automatic identification of mobility pattern themes, which, ultimately, leads to the better understanding of users' profile. In this methodology, a user is considered a document, and the categories of places visited by him/her are the words describing the documents. With that, we are able to extract topics that describes typical user movements;
- We demonstrate the applicability of our results in two particular cases: 1) new itinerary recommendation system, based on the mobility pattern themes and the spatio-temporal graph model, that not only suggests a place, but, which place to go after a certain one and a certain time; 2) we also show that our methodology could be used for business owners to understand how to offer a better service for tourist in different locations that could be culturally distinct.

Part of the contributions of this work was reported in the papers:

- FERREIRA, A. P. G.; SILVA, T. H. ; LOUREIRO, A. A. F. . Beyond Sights: Large Scale Study of Tourists' Behavior Using Foursquare Data. In: Workshop on Mobility Analytics from Spatial and Social. Proceedings of IEEE International Conference on Data Mining (ICDM). Atlantic City, United States. 2015;
- FERREIRA, A. P. G.; SILVA, T. H. ; LOUREIRO, A. A. F. . Você é o seu check-in: entendendo o comportamento de turistas e residentes usando dados do Foursquare. In: Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia). João Pessoa, Paraíba, Brazil. 2014;
- A manuscript entitled Tell Me Where You Go and I'll Tell You Who You Are: Studying Tourists Mobility in Large Scale Using Social Media is being prepared to be submitted to the Elsevier Annals of Tourism Research.

Some contributions derived from this study was also explored in the following collaborations:

- Silva, Thiago H.; CUNHA, F. D. ; TOSTES, A. I. J. ; BORGES NETO, J. ; CELES, C. S. F. S. ; MOTA, V.F.S. ; FERREIRA, A. P. G. ; MELO, P. O. S. V. ; Almeida, J. ; Loureiro, A. A. F. . Users in the Urban Sensing Process: Challenges and Research Opportunities. Chapter in Next Generation Platforms for Intelligent Data Collection. 1ed. v.1 , p. 45-95. Elsevier (Amsterdam). 2016;
- SILVA, T. H. ; FERREIRA, A. P. G. ; BORGES NETO, J. ; RIBEIRO, A. I. J. T. ; CELES, C. S. F. S. ; CUNHA, F. D. ; MACHADO, K. L. S. ; MOTA, V. F. S. ; MINI, R. A. F. ; MELO, P. O. S. V. ; LOUREIRO, A. A. F. . Redes de Sensoriamento Participativo: Desafios e Oportunidades. Minicursos / XXXIII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos. Sociedade Brasileira de Computação. V1, p. 266-315. 2015.

1.4 Work Organization

The remaining of this work is organized as follows. Chapter 2 presents the related work. Chapter 3 presents our dataset and the approach we use to identify tourists and residents. Chapter 4 presents the behavioral properties of tourists and residents in different cities worldwide and Chapter 5 shows how we have modeled the mobility. Section 6 demonstrates how we can explore our results for new services and applications. Finally, Section 7 presents the conclusions and future work.

Chapter 2

Related Work

This section is organized as follows. Section 2.1 introduces Human as sensor. In Section 2.2 it discuss related work on behaviour of human mobility. In Section 2.3 it shows some applications related to tourist mobility and touristic places recommendations. Finally, in Section 2.4 we discuss the topics presented and how they are related with our work.

2.1 Contextualization

2.1.1 Human as Sensors

The Internet was one of the greatest revolutions in communication in the history. Since it was a project for academic and military purposes to a means of communication between any computer in the world, we have witnessed what can be achieved when there is collaboration. Through e-mail lists to a collaborative encyclopedia, the Internet helped us to solve many problems of everyday life with collaboration and information. The evolution of the Internet has spread beyond computers: other devices have been added to the lives of people and the connection with the world. Today it is possible to connect with smartphones, watches, cars and televisions, having access to information in real time and anywhere.

Combined with the devices and increased Internet availability and connection speed, the services also followed the evolution. In addition to sites with listings of jobs and products for sale, for example, we have also websites that connect people: the social networks. Social networks are a structure that connects individuals in specific types of interdependency, such as friendship, common interests and knowledge sharing [Zheng,

2011]. Networks like Facebook¹, G+² and Twitter³ became a channel of communication between people and their friends and between people and their interests. Books, movies, technology forums - many groups that share the same interests were formed and joined with the help of the Internet.

All areas were affected with the use of computers and the Internet, from large to small businesses. While large companies have automated their processes and improved communication through computers, small business and restaurants were able to increase their visibility through the reviews on specialized sites to list places in cities. Through collaboration sites emerged to share photos, reviews and also tourist itineraries.

Considering people and the sharing of different types of data we can say that human beings act as a kind of sensor. To understand more about this subject in Section 2.1.2 we talk about Participatory Sensor Networks.

2.1.2 Participatory Sensor Networks

Social networks allow many people to share information much more quickly. This allows to discover new places and their characteristics from what people post on these networks in real time. Considering the growth in the use of mobile devices such as smartphones, Location-based Social Networks (LBSNs) have become quite popular, especially because they help to reduce the gap between the real world and online services based on social network [Zheng, 2012].

Location-based Social Networks are social networks that include information about the location on the content that is voluntarily shared by users [Roick and Heuser, 2013]. The concept of location can be represented by: 1) geographical position, represented by latitude and longitude; 2) a region (approximate position) and 3) a nominal location (such as home, work, shopping) [Zheng, 2012]. These networks allow users to share information about where they are.

The data from social networks based on location can be seen as a valuable source of sensing, where the sensors are the users (humans as sensors), who share information about their context from their mobile devices. Users send information similarly to sensors in a traditional sensor network on a voluntary basis. In fact, LBSNs are the most popular examples of Participatory Sensor Networks (PSNs), sensing network where sensor nodes are formed by users, who use their mobile devices to send data about their context [Silva et al., 2014b]. Figure 2.1 can view a representation of the PSNs and its interaction with users from the study by [Silva et al., 2014a].

¹<https://www.facebook.com>

²<https://plus.google.com>

³<https://twitter.com>

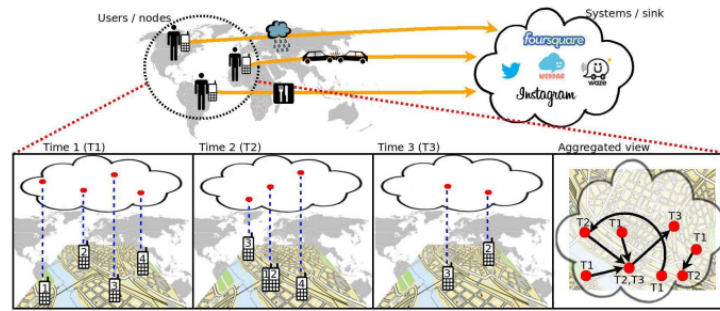


Figure 2.1. Illustration of a Participatory Sensor Network. This image was obtained from [Silva et al., 2014a]

Some examples of PSNs are: Instagram⁴ for photo sharing, Waze⁵ for sharing problems in traffic and Foursquare⁶ for location sharing. All these services use geographic data to provide services / useful information to its users. Data such as weather, photos or the sport a user is practicing can be shared in real time on these networks.

Each shared data in a participatory sensor network is associated with the preferences and habits of users. Such data make it possible to study large-scale urban behavior of people and the dynamics of cities [Silva et al., 2014a]. In this work we consider the temporal and spatial aspects of the data shared in PSNs for our analysis.

2.2 Understanding Mobility

This section discusses studies about the use of PSN data to discover how people behave and their habits inside cities, including patterns of mobility. We divided the work into three groups: mobility studies with traditional data, such as GPS traces (Section 2.2.1); works that study mobility with social data, such as PSNs data (Section 2.2.2); and studies that focus specifically on the study of mobility of tourists in cities (Section 2.2.3).

2.2.1 Studying Mobility Through Traditional Data

This section discusses studies that investigate human mobility and how it works inside of the cities. Human mobility is a fundamental aspect of the dynamics of a city and is the object of study of other areas, such as anthropology and biology. One approach

⁴<https://www.instagram.com/>

⁵<https://www.waze.com>

⁶<https://foursquare.com/>

to perform this type of study is use digital traces from users, such as GPS traces. In the literature, there are several studies about users' habits and routines in a city using digital traces. Some of them analyzed GPS data and cellular footprints of users to understand, for instance, their usual trajectories [Choujaa and Dulay, 2009; González et al., 2008].

Some researchers used the Levy Walk, pattern of movements of the animal kingdom that combines long paths with short random movements, to study human movement with GPS data[González et al., 2008; Karamshuk et al., 2011; Kung et al., 2014]. However, only the drive is not enough to understand the context of the user at that time. According to [Karamshuk et al., 2011], human movements are highly predictable but it is crucial to take into account the spatial and temporal patterns regular. It is difficult to gather traditional data of users' mobility, this work focused on alternative sources. Work related to the use of these sources are discussed in the next section.

2.2.2 Studying Mobility Through Social Data

Other studies used data shared in participatory sensor networks, such as check-ins from Foursquare, to understand several aspects of urban social behavior and mobility [Cheng et al., 2011; Pianese et al., 2013; Long et al., 2013; Preo and Cohn, 2013; Lv et al., 2013; Cho et al., 2011].

Aware that human beings are endowed with similar behaviors in their mobility, the authors [Cheng et al., 2011] used 22 million check-ins shared on Twitter ⁷ to extract a pattern of mobility in shares, and it showed that users adopt periodic behavior and they are influenced by their social, geographical and economic status. In addition to the spatial knowledge discovery potential, messages (e.g. tips) in check-ins can also reveal interests and feelings. In the same direction, [Pianese et al., 2013] used check-ins shared on Foursquare to group and discover communities and places of interest.

Some works are dedicated to the study of habits and user routines in a city. Through GPS records and signals of cellular networks it is possible to understand with good accuracy which way users perform often, as we discussed in Section 2.2.1. Other studies performed studies in this area using data from social networks [Lv et al., 2013; Preo and Cohn, 2013; Pianese et al., 2013]. However, finding patterns from data from social networks brings a greater challenge, since there is an irregularity in the distribution of data over time among users [Pianese et al., 2013] also not always users are encouraged to share data [Lindqvist et al., 2011]. Despite that, several studies have found evidence that the realization of this type of study using data from social

⁷<http://twitter.com/>

networks is possible. For instance, [Pianese et al., 2013] was able to identify patterns in days and times in the activities of users and [Preo and Cohn, 2013] identified user behavior profiles.

2.2.3 Mobility of Tourists and Residents

This section shows some of the main studies about how tourists move and which patterns are recognized in their mobility. Tourism is one of the important economic activities that promotes regional economic growth [Staab et al., 2002]. It is the displacement of their place of residence to a different one, where there is a meeting of cultures and the search for new experiences. A tourist may have different needs than you are used to your routine. In addition, factors such as cost, climate and personal preferences influence the activities to be carried out by the tourist visited city. Thus applied to the tourism economy can understand the factors that influence this decision consumption [Sharpley and J, 2002].

Despite the efforts in understanding urban mobility mentioned in Sections 2.2.1 and 2.2.2 very few studies investigated urban tourist mobility in large scale [Lew and McKercher, 2006; Fennell, 1996]. [Zheng et al., 2009] analyzed 107 GPS logs of users during a period of one year. They concluded that the movement of tourists and residents are different and the behavior of tourists is influenced by their traveling experience and their personal relationships.

There are also many proposals that consider data from social data shared in PSNs. For example, [Silva et al., 2013b] showed how to extract touristic sights using shared photos on Instagram. In addition to the locations, you can also extract information from events that attract tourists to the cities. Besides that, [Hallot et al., 2015] used check-ins performed at the Art Institute of Chicago to show evidences that it is possible to use this source of data to infer the behavior of tourists. In the same direction, [Long et al., 2013] investigated traveler mobility patterns by mining the latent topics of users' check-ins performed in one city in the United States.

[Long et al., 2013] investigated the categories and latent topics related to tourists in a city on Foursquare and identified characteristics of the city that were related to the interests of tourists.

2.3 Applications based on tourist mobility

The study of spatio-temporal tourist mobility in the city and the factors that influence their movements has important implications in several segments, for example, in

smarter destination planning and urban planning to better support tourists. In this section we introduce some applications focused on tourist mobility and places recommendation for this segment of users.

In the same direction of personalized itineraries, [Yoon et al., 2010] proposed an architecture of the recommendation itineraries for tourists, considering the length of the stay and their interest. [Diplaris et al., 2012] created the SocialSensor, a framework that integrates the user's interests and real-time search context. [Zheng, 2014] proposed a recommendation system that exploits the interests of users and similarity between different users, using a collaborative filtering approach and TripAdvisor Data ⁸.

Still following the studies that were based on social data we can quote [Choudhury et al., 2010] and [Majid et al., 2012] who used photos from Flickr⁹ to automatically generate tourist itineraries. [Shi et al., 2011] use the same approach but focused on recommendations of Landmarks and adding data from Wikipedia¹⁰ to enrich the recommendation. [Hsieh et al., 2012] developed TripRec, an application to recommend tourist itineraries based on check-ins.

Exploring more user preferences, [Basu Roy et al., 2011] developed an application where users give feedbacks and iteratively construct their itineraries based on personal interests and available time. [Yerva et al., 2013] proposed an itineraries recommendation system based on user preferences, using data from Lonely Planet, Foursquare and Facebook to suggest locations. [Baraglia et al., 2013] a prediction is made of the next point of interest to the tourist, based on his/her historic, and with this, recommend the next Point of Interest (POI) to the user.

Observing other aspects, such as the perspective of the business area [Karamshuk et al., 2013] have identified the best solution for retail, using social networking data, demonstrating the usefulness of such data for business. From the perspective of events with a tourist view, [Morais and Andrade, 2014] investigated the relevance of messages shared by tourists and residents during a massive event tour on an famous brazilian event.

2.4 Discussion

A differential of our work is consider people as sensors in cities and examine how they behave, either as tourists or as residents, watching in the spatio-temporal perspective. As we discussed, several studies differ from ours since they are based on GPS logs,

⁸<https://www.tripadvisor.com.br>

⁹<https://www.flickr.com>

¹⁰<https://wikipedia.org>

whereas we use sensed data from PSNs and our focus is on better understanding the behaviour of tourists and residents.

Related studies that use social data to study mobility and preferences of tourists focused, typically, on just one city. Our work goes beyond the study of the mobility and preferences of tourists in a place or a city; It is a large-scale study of these aspects in four different cities considering tourists and residents. When considering various different cities different regions of the world, we can visualize the behavior patterns that emerge, and to understand how the cultural traits shape these behaviors.

We investigate how temporal and spatial aspects influence the mobility of tourists and residents of a given city, using check-ins shared by users in Foursquare. This study allows us to include also events and gave us the possibility to recommend places based on urban mobility profile and other tourists behaviour.

Chapter 3

Data Collection and Processing

This section describes the dataset used in this work, as well as how we collect and filter the data, and the procedures used to identify tourists.

3.1 Datasets

We collected check-in from Foursquare, nowadays one of biggest LBSNs with 60 millions of users registered. To retrieve the check-in performed on Foursquare we used the Twitter service, where they are publicly available. This was only possible for Foursquare users who shared their check-in on Twitter, which provides a streaming API¹ to obtain tweets in real time. Today Foursquare company is divided in two apps, one app, called Swarm², is responsible to register just check-in of the users. The other app, which is named Foursquare, is focused on recommend places to users. Because during the time we collected the data this changes in the Foursquare company had not yet happened, we will not mention Swarm anymore in the remainder text. In Figure 3.1 we can see an example of check-in shared on Twitter and in Figure 3.2 we can see the check-in page (accessed through the URL shared on Twitter). In the check-in URL we can also get the venue URL as shown in Figure 3.3.

After retrieved tweets with check-in, we performed an extra collection using the Foursquare API³ to retrieve information about the venue, such as the name and reason of the visit. The complete data collection process is displayed in Figure 3.4.

We gathered data from different cities around the world: London/United Kingdom, New York/United States, Rio de Janeiro/Brazil and Tokyo/Japan. We chose

¹<https://dev.twitter.com/docs/streaming-apis>

²<https://www.swarmapp.com>

³<https://foursquare.com/api>



Figure 3.1. Tweet with check-in information. In 1) the check-in URL, 2) the time and date, 3) geolocalization from Twitter.

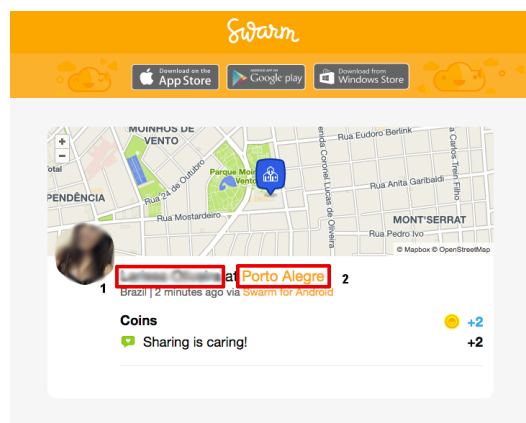


Figure 3.2. Swarm's/Foursquare's page with check-in details. In 1) User name and 2) the name and link of the venue.

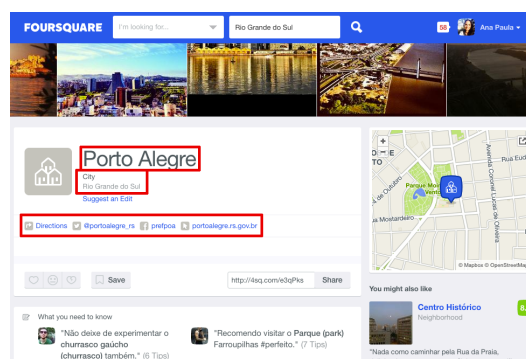


Figure 3.3. Venue page on Swarm's/Foursquare's page



Figure 3.4. Illustration of the complete process to collect check-in data

those cities because they represent distinct regions of the world, representing, potentially, users with cultural differences as well.

We collected data in April, June and July of 2014. Table 3.1 presents more details about number of check-in and users. In table 3.2 the number of check-in is specified by city.

Table 3.1. Number of check-in and unique users by dataset

	# of check-in	# of users
April, June and July of 2014	151.501	13.356

Table 3.2. Number of check-in by city and dataset

London	New York	Rio de Janeiro	Tokyo
5.884	32.554	61.886	51.177

Each check-in has the following attributes: check-in ID, user ID, time and geographic coordinate (latitude and longitude), category and subcategory of the check-in’s location, i.e., the type of place where it occurred. The Foursquare categorizes places in 10 categories: *Arts & Entertainment*, *College & University*, *Food*, *Professional & Other Places*, *Nightlife Spots*, *Residences*, *Great Outdoors*, *Shops & Services*, *Travel & Transport*, *Events*. Each of these categories has subcategories, totaling more than 350 subcategories. The Foursquare categorization sometimes might group subcategories that are very specific subcategories. For example, Travel & Transport contains the subcategories Hotels and Train Stations. In order to have a clear view of the users’ habits, we created a classification of places, grouping subcategories that are more related to each other. For instance, we created the two new categories: transport and travel. Transport contains the subcategories like Airport, Bus Stop and Rental Car Location, and Travel contains the subcategories like Bed & Breakfast, Hostel and Resort. Table 3.3 shows this classification.

To avoid possible noise in our dataset, we eliminate users not found as well as check-in in places that no longer exist. That was verified beyond the Foursquare API. If the user’s ID was not found on the API we eliminate them.

Category by Foursquare	New Category
Arts & Entertainment	arts
College & University	school
Event	entertainment
Food / Nightlife Spot	drink
Food	fastfood
Food	restaurants
Nightlife Spot / Event	entertainment
Outdoors & Recreation	outdoors
Outdoors & Recreation	sports
Professional & Other Places	city
Professional & Other Places	health
Professional & Other Places	professional
Professional & Other Places	religion
Residence	home
Shop & Service	services
Shop & Service	shopping
Travel & Transport	transport
Travel & Transport	travel

Table 3.3. New classification to each category group of Foursquare

3.2 Identifying tourists and residents

After collecting the data we needed to separate data coming from tourists and residents. For this we identified the city where the user spent most time, with at least 21 days of stay, based on check-in intervals⁴. From the check-in sequence performed in each city we check how many days were spent on them. For example, if a user performed a check in on city A on 5/may/2016 and another check-in the same city on 30/may/2016, we assume that he/she stayed 25 days in city A. Eventually, a user may have been in different cities for more than 21 days, in this case we consider the user’s city where he spent most time. If a user give a check-in in a city different of his home he is considered a tourist on that city. This tourist identification process has also been used by other researchers in the work [Paldino et al., 2015; Choudhury et al., 2010]. We used this process in all datasets considered in this work.

City	Tourists	Residents
New York	737	2.584
Rio de Janeiro	498	3.550
London	584	514
Tokyo	629	4.260

Table 3.4. Number of tourists identified in each city

For cities chosen for the analysis we filter all check-in belonging to each of them using the geographic coordinates of check-in. Then we divide the data among tourists and residents, using the residence criteria identified in the process described above. In table 3.4 it is possible to find the number of unique users identified. Users that

⁴All users had their check-in sequence sorted chronologically.

we could not identify his/her resident because of lack of data were excluded from the analysis.

3.3 Data limitations

Conducting research using social networking data allows us to capture what is happening in the world in near real time. The use of this data is proving to be increasingly powerful for the study of urban behavior [Silva et al., 2013a; Zheng et al., 2014], providing advantages, for example faster responses and cheaper cost, over other traditional methods for this purpose, such as surveys and interviews. Although it has many advantages, data from social networks may have limitations. One is the amount of data that can be collected from those services. For example, the Twitter API has a limitation of 1 % of the total volume of data produced, this means that we can not have all the data we want for a given application. In addition, less than 25% of Foursquare users push their check-in Twitter [Long et al., 2013].

Another limitation is the possible bias towards users who have smartphones with Internet access and using this services. This means that what is identified with the use of these data might not represent the entire population of tourists.

Chapter 4

Behavior of tourists in different cities worldwide

Tourists can behave differently at different cities, depending on the purpose of their visit. Rio de Janeiro, London, Tokyo and New York have particular characteristics, such as different dynamics and local culture, and because of that they may attract tourists with different tastes. Since the behavior of tourists in different cities may be different it is interesting to study tourists' behavior separately for each city considered in this study. For that, in this chapter we show how we can use Foursquare check-ins to understand how tourists behave in different cities.

Analysing tourists and residents in Rio de Janeiro, London, Tokyo, and New York, we observe significant differences between the behavior of those classes of users in several cases. This chapter is organized as follows. Section 4.1 presents a temporal analysis of the check-ins shared by users. Section 4.2 presents different places visited by each class of user. Section 4.3 presents the behavioral properties of tourists and residents in different routines. Section 4.4 shows the preferences of tourists inside the cities. Section 4.5 studies the behavior of domestic and foreign tourists. Finally, Section 4.6 presents the discussion of this chapter.

4.1 Number and time interval of check-ins

Figures 4.1 and 4.2 show the distribution of the number of check-ins and distribution of interval time (in hours) of check-ins made by the same user, tourists and residents, in each city, respectively.

With the help of Figure 4.1 we can see that all the cities analyzed have more tourists than residents performing check-ins. Although there is a difference between the

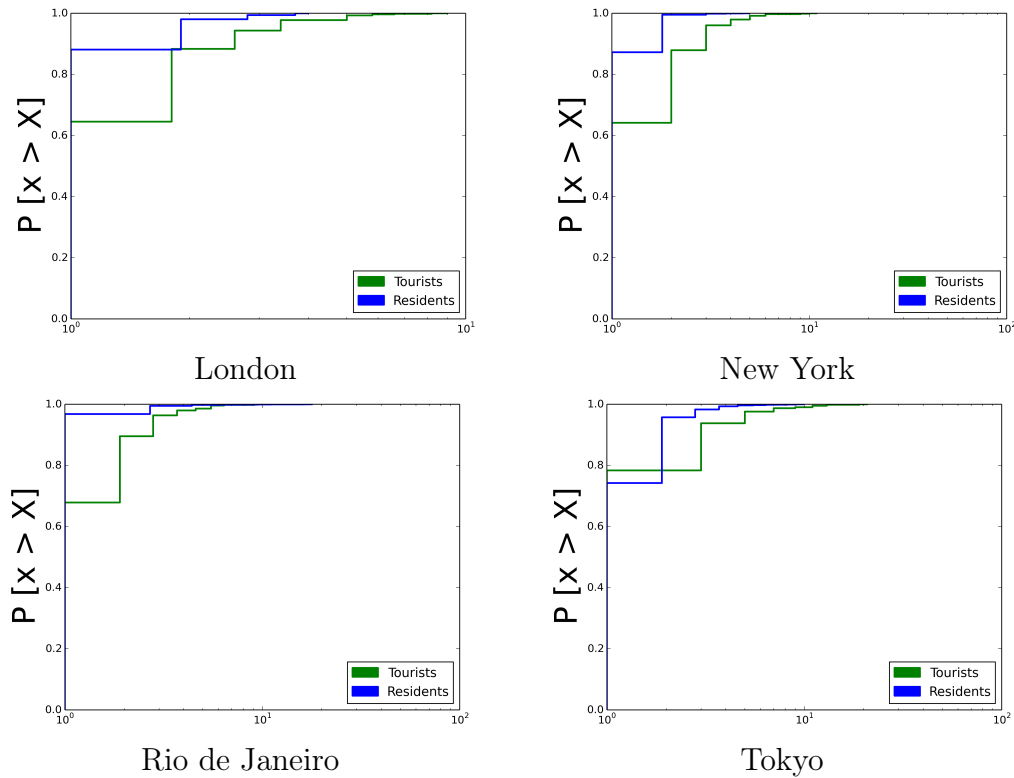


Figure 4.1. Distribution of the number of check-ins performed by tourists (green) and residents (blue)

cities, almost all of them follow the same pattern of behavior of tourists and residents. We believe that this behavior is directly related to people’s motivation to do check-ins during the trips because they are more motivated by the new experiences and places that they are discovering and want to share it with their friends [Bilogrevic et al., 2015].

Unlike the distribution of the number of check-ins, the distribution of check-ins range shown in the Figure 4.2, given in hours, varies between cities. New York and Tokyo have similar behaviors, with tourists and residents share check-ins at similar intervals. Many of the users who check-ins in these cities perform check-ins on a long space. In London and in Rio de Janeiro are similar behavior among themselves, tourists and residents tend to share in a smaller amount of time when compared to Tokyo and New York. However, in London and in Rio de Janeiro tourists share more frequently and in a shorter time interval. These differences can be seen as evidence of the tourist behavior characteristics while they are in these cities.

Observe the distribution of the number of shared check-ins and the interval between them also helps us understand how each user class behaves in these cities. The peculiarities of each city define the profile of tourists that it will receive the resident profile that already live there (as shown in section 5.3).

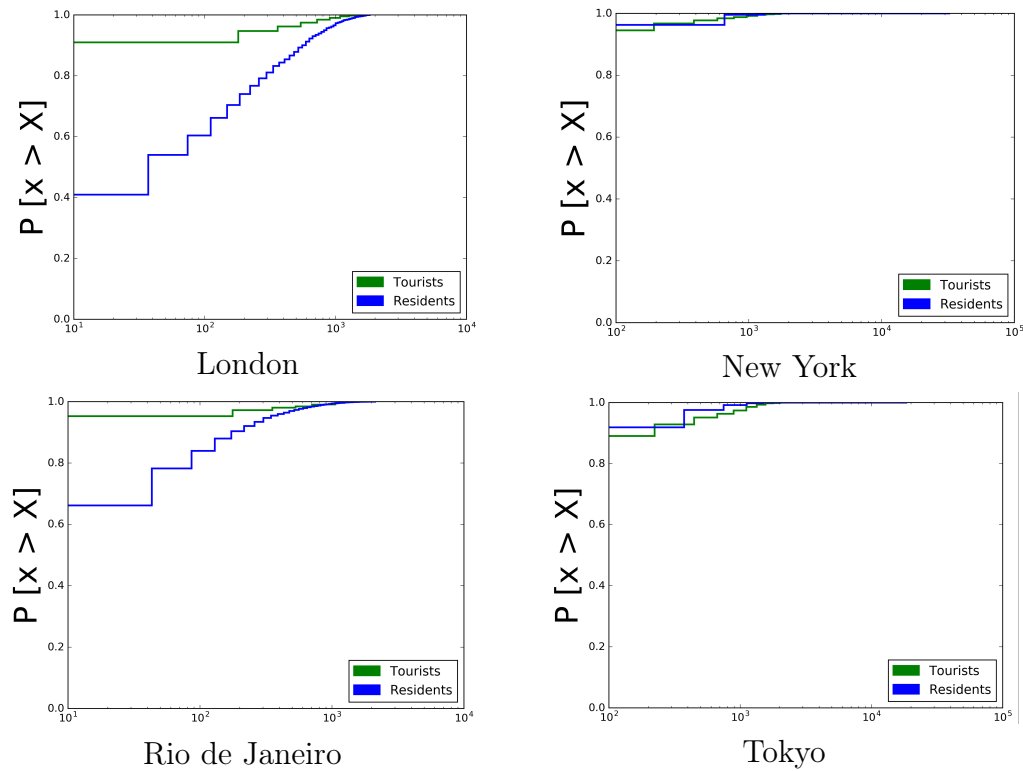


Figure 4.2. Distribution of the time interval (in hours) between the check-ins performed by tourists (green) and residents (blue)

4.2 Places visited

Study the visited sites also helps us to improve our understanding of the differences in these classes of users, which can be useful for government and tourism stakeholders. With this information the government can, for example, know more precisely what places should receive more or less investment or advertising to improve tourism in the city, and what should be improved in the city to allow access to these places.

In this direction, Figure 4.3 shows the places where tourists (green) and residents (blue) performed check-ins in London, New York, Rio de Janeiro, and Tokyo. As we can see, certain areas are more visited by tourists than others. For example, in Rio de Janeiro, most of the tourist activity is concentrated by the sea in a specific area (bottom-right of the figure), where most of the tourist attractions are located, whereas in New York, Manhattan island is the most popular destination of tourists.

Although the map gives us a good sense of where tourists are concentrated in cities, it is important to look deeper where these two types of users tend to go. Tables 4.1 and 4.2 show the ranking of most popular places, according to the number of check-ins, among tourists and residents in the four studied cities. Some places were

expected, such as Times Square and the Empire State Building in New York, Oxford Street and The Buckingham Palace in London. Other places also very popular might not be traditional sights, such as FIFA Fan Fest in Rio de Janeiro, which is a special place created in the city for tourists and residents during the 2014 FIFA World Cup, event that the city hosted, attracting many tourists. This examples illustrates automatically identified dynamic changes in the popularity of places for tourists in the city, including new places and the ones that may exist only for a short period of time.

Looking at the ranking of residents (Table 4.2) we can identify places that are also frequented by tourists, such as airports, shopping malls and parks. However, one can view a different pattern in the types of places. Residents tend to go more in places related to daily routines, such as universities, places to practice sport and restaurants.

Tokyo is a peculiar example in our dataset. The most popular places among tourists and residents are train stations. The rail network in Tokyo is one of the world's largest, which explains the large volume of check-ins of tourists and residents in their stations, using the system either to explore the city or move to perform daily routines. Even though we do not have sights as the top places in that city, those stations give hints of what are the preferences of tourists. Some stations are the same for the two classes, for example, Akihabara, Tokyo and Shinjuku Station. This was expected because around those stations there are several places that attracts tourists and residents. For example, Shinjuku Station arehat world's busiest railway station, handling more than two million passengers every day. Around Shinjuku Station there is large entertainment, business and shopping area. West of the station is Shinjuku's skyscraper district, home to many of Tokyo's tallest buildings, including several premier hotels and the twin towers of the Metropolitan Government Office, whose observation decks are open to the public for free. Besides that, there are stations that are more popular among tourists, such as Ueno Station. Next to this station is Ueno Park, a large public park that attracts thousands of tourists. Today Ueno Park is famous for the many museums sssettled on its grounds, especially the Tokyo National Museum, the National Museum for Western Art, the Tokyo Metropolitan Art Museum and the National Science Museum. It is also home to Ueno Zoo, Japan's first. Additionally, Ueno Park is one of Tokyo's most popular and lively cherry blossom spots with more than 1000 cherry trees lining its central pathway¹.

We have seen that the most visited by tourists and residents provides valuable information for understanding the behavior and motivation in the city. However, there are other factors, such as time, which can provide an additional perspective on this

¹<http://www.japan-guide.com>

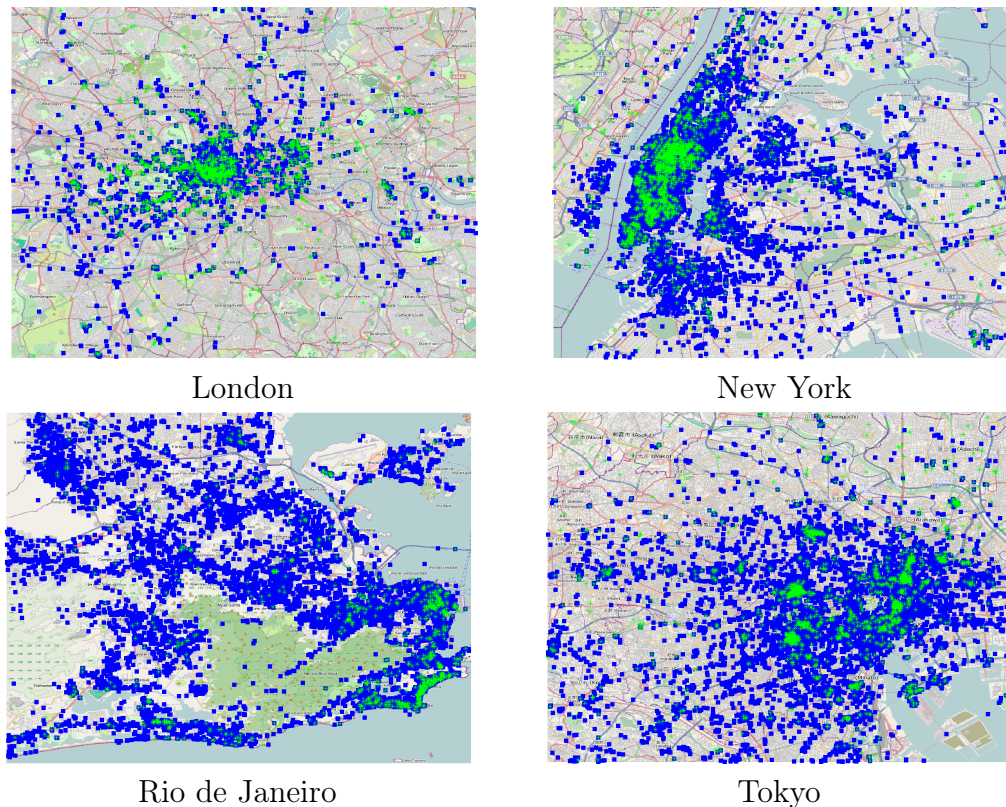


Figure 4.3. Places where tourists (green) and residents (blue) performed check-ins

understanding. And exploring this is the aim of next sections.

4.3 Routines

Tourists and residents perform similar activities in the city, such as eating [Colombo et al., 2012], however there may be differences in the pattern of behavior in performing those activities. Figure 4.4 show the temporal variations of the number of check-ins shared throughout the hours of the day for weekdays and Figure 4.5 show this information for weekends.

Observing the behavior of residents in all cities during the week we can see peaks around the beginning of business hours (8 to 9 hours), lunch time (12 to 13 hours) and at the end of business hours (18 to 19 hours). These schedules clearly show a routine following traditional business hours, followed by residents in their daily lives. While among the tourists we observe a different pattern, not very aligned with traditional daily routines. This could be explained by the freedom that tourists have to perform different activities during the trip. Perhaps Tokyo is the city where the behavior of

Rio de Janeiro	London	New York	Tokyo
Aeroporto do Galeão	Starbucks	John F. Kennedy Airport	秋葉原駅 (Akihabara Sta.)
Aeroporto Santos Dumont	Harrods	Times Square	東京駅 (Tokyo Sta.)
Estádio Maracanã	The London Eye	LaGuardia Airport	新宿駅 (Shinjuku Sta.)
Praia de Copacabana	London	Starbucks	渋谷駅 (Shibuya Sta.)
Rio de Janeiro	Piccadilly Circus	Apple Store	池袋駅 (Ikebukuro Sta.)
Starbucks	Oxford Street	Empire State Building	和光市駅 (Wakoshi Sta.) (TJ-11/Y-01/F-01)
Terminal Rodoviário Novo Rio	London Euston Railway Station	Museum of Modern Art	JR 東海道新幹線 東京駅
FIFA Fan Fest	Hyde Park	American Museum of Natural History	品川駅 (Shinagawa Sta.)
Praia de Ipanema	Buckingham Palace	Yankee Stadium	JR 品川駅
Shopping RioSul	British Museum	The Metropolitan Museum of Art	上野駅 (Ueno Sta.)

Table 4.1. Ranking of most popular venues for Tourists

tourists follows more similarly the behavior of residents, because of the three peaks of activity in common. Note, however, that the activity of tourists tend to be more intense during the day. This might mean that Tokyo attracts a different kind of tourist that tend to perform activities in a regular way, for example having lunch in the same time of residents of Tokyo, helping to explain the observed pattern and understand better the city tourists.

The patterns of when tourists and residents perform activities during weekends are not very different in most cases. We can also observe that both patterns are very different from those observed during weekdays. This could be explained by the fact that during weekends, typically, residents do not have routines (or have different routines), being able to act somehow as tourists in the city, which usually do not have to follow fixed schedules.

Rio de Janeiro	London	New York	Tokyo
FIFA Fan Fest	Cineworld	Starbucks	秋葉原駅 (Akihabara Sta.)
McDonald's	Vue Cinema	Equinox	新宿駅 (Shinjuku Sta.)
BarraShopping	Starbucks	LaGuardia Airport	渋谷駅 (Shibuya Sta.)
Outback Steakhouse	BFI Southbank	John F. Kennedy Airport	池袋駅 (Ikebukuro Sta.)
Universidade Estácio de Sá	Hyde Park	Planet Fitness	東京駅 (Tokyo Sta.)
Aeroporto do Galeão	The O2 Arena	New York Sports Club	東京国際展示場 (東京ビッグサイト/Tokyo Big Sight)
Estádio Maracanã	The King Fahad Academy	Crunch	吉祥寺駅 (Kichijoji Sta.)
Universidade Veiga de Almeida	Harrods	Blink Fitness	ヨドバシカメラマルチメディアAkiba
Starbucks	InMobi	Citi Field	原宿駅 (Harajuku Sta.)
NorteShopping	Soho Square	New York Health & Racquet Club	中野駅 (Nakano Sta.)

Table 4.2. Ranking of most popular venues for Residents

4.4 Preferences of tourists

The categorization of places helps us to better understand the preferences of tourists, because as we showed above, most of the cities have some places that attract more tourists than residents.

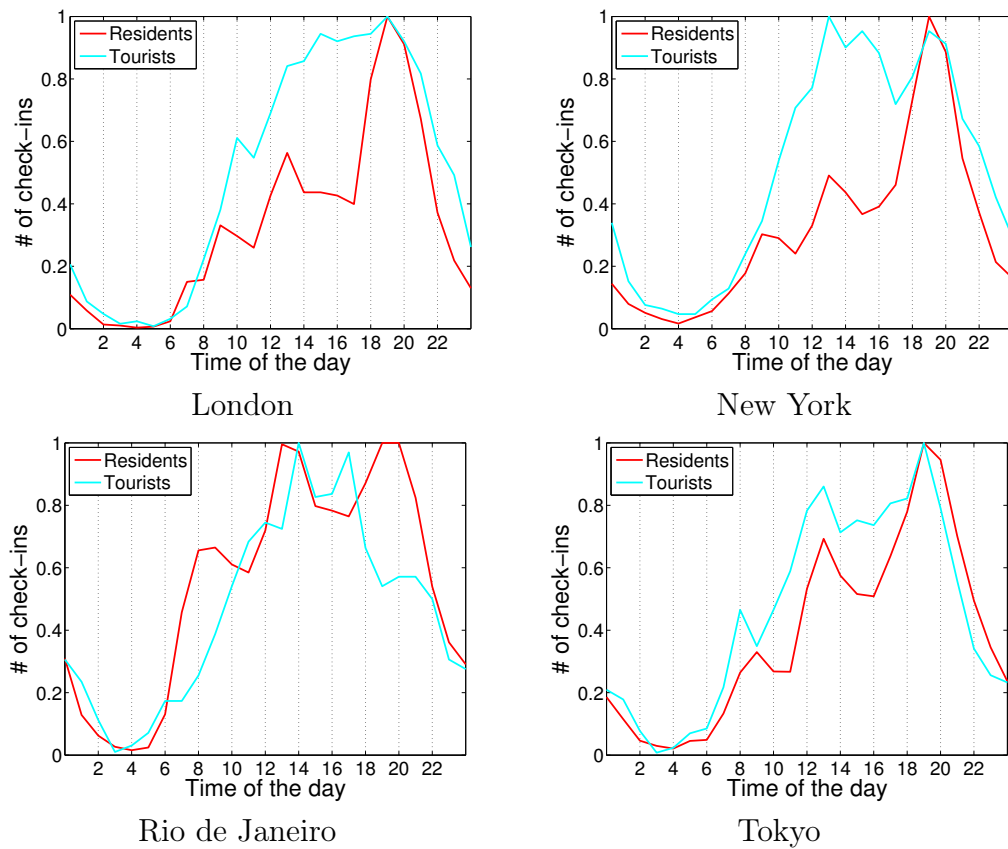


Figure 4.4. Temporal check-in sharing pattern throughout the day by tourists and residents during weekdays

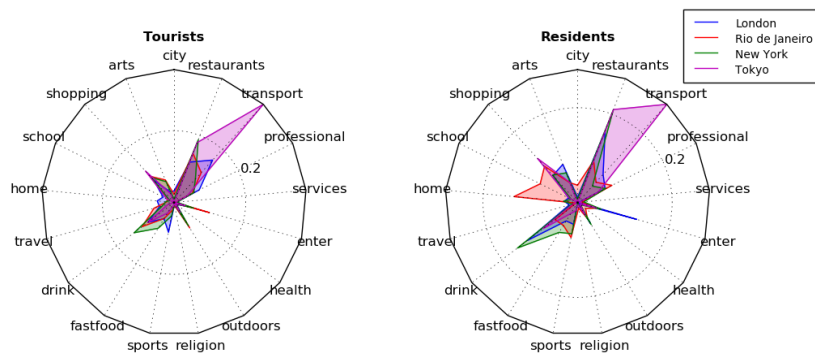


Figure 4.6. Check-ins frequency for each category by tourists and residents by city

To evaluate this point, Figure 4.6 shows a radar chart representing the popularity of categories of places for tourists (left figure) and residents (right figure). To measure the popularity of a category c of place we consider the number of check-ins given in all places that are categorized by c . Some categories were expected to be visited

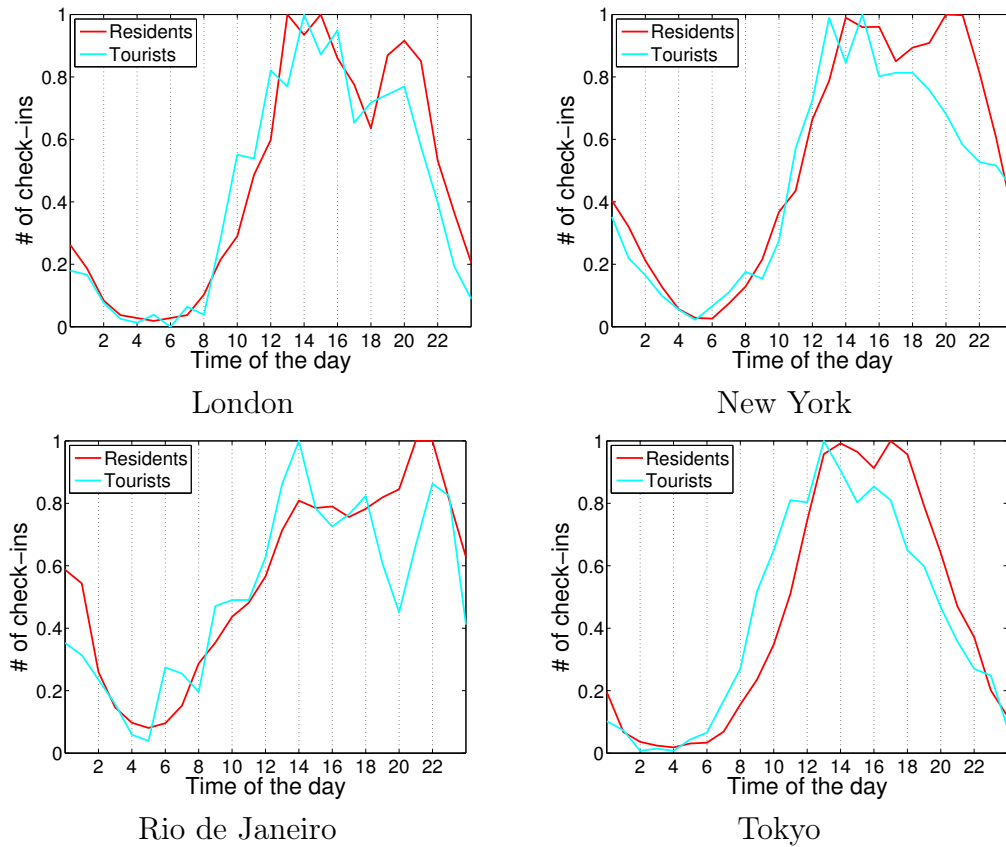


Figure 4.5. Temporal check-in sharing pattern throughout the day by tourists and residents during weekend

by tourists, such as hotels, airports and monuments, whereas others such as houses, markets, colleges and universities were expected to be more popular among residents. Depending on the city, the number and popularity of certain categories vary. In Tokyo, for example, it is not popular for residents to perform check-ins in places such as their *residence*, unlike other cities, where residents typically perform check-ins places that belongs to that category. This is the case of Brazil, where residents perform many check-ins in the category *home*, indicating a minor concern about their privacy. These results could be explained by cultural differences.

Tourists in New York tend to visit several places to drink, while in London tourists tend to attend places related to sports, this category is more common among residents of Rio and New York. In Rio de Janeiro tourists tend to visit places in the category entertainment, such as concert halls, and in the category outdoor, such as beaches. The category outdoor is quite popular in Rio de Janeiro by tourists, unlike tourists in Tokyo. However, the same category is popular among Tokyo residents, which demonstrate their habits in attending outdoor sites and monuments.

Residents in New York City visited significantly baseball stadiums, a very popular sport in that city. In Rio de Janeiro, the subcategory related to barbecue restaurants received many visits, reflecting a typical habit of the local culture. London is known for its pubs and nightlife, beyond the great historic sites, and this behavior is reflected in the category of places visited by tourists and residents. This results are interesting because they reflect typical cultural differences among the studied cities, fact that could be explored, for instance, in new recommendation systems.

4.5 Domestic and foreign tourists

Analyzing the tourists within the cities, we can also separate them into two different classes: domestic tourists and foreign tourists. Domestic tourists are tourists who come from cities in the same country, and foreign tourists come from different countries. Through the process of classification of tourists and residents explained in section 3.2) it is possible to classify where people are originally from. A tourist is classified as domestic if his/her city of origin belongs to the same country of the city he/she is visiting and as foreign if his/her home city is outside the country of the city where he/she is considered a tourist.

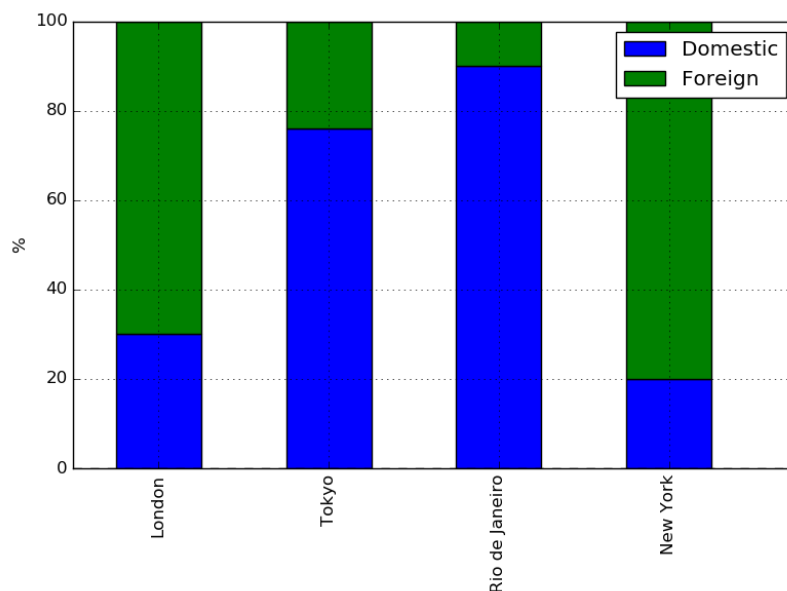


Figure 4.7. Foreigners and Domestic Tourists by city

Figure 4.7 shows the proportion of tourists of each type in all studied cities. Each city has a different representation of these types of tourists. Tokyo and Rio de Janeiro,

for example, receive more domestic tourists, more than 75% in both cities. While in London and New York, most tourists are foreigners.

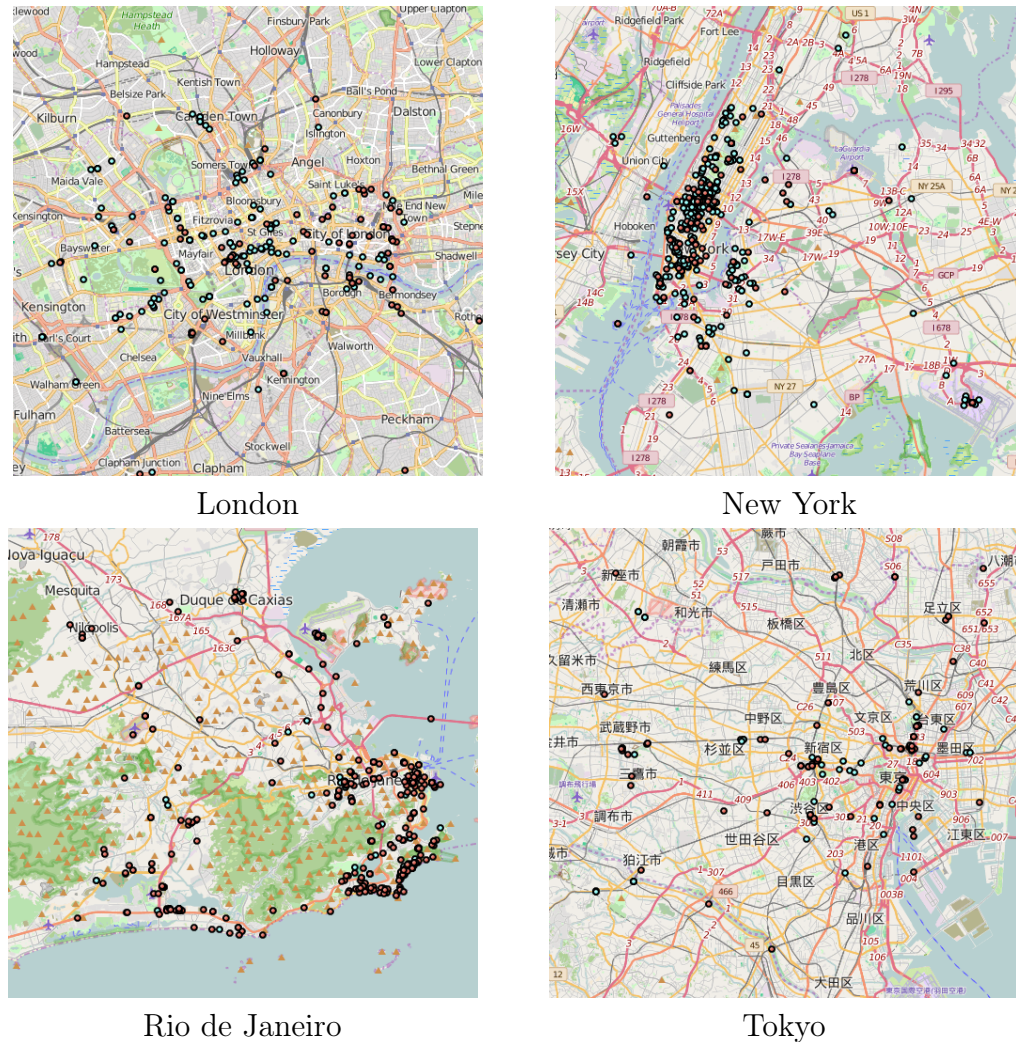


Figure 4.8. Places where foreigners (blue) and domestics (red) performed check-ins

Figure 4.8 shows the locations domestic and foreign tourists visited in all studied cities. In London and New York we can see that foreign tourists venture more to remote areas from the center, while in Tokyo and Rio de Janeiro they are more concentrated in the central areas. This could be explained by some difficulties that some tourists might face to explore those cities, for example the language and exotic local culture. In Tokyo and in Rio not many residents speak English, making many tasks harder to be performed without guidance. On top of that, local habits might not be well understood by foreigners, making more convenient to stay close to the city downtown where is more common to find international places (i.e., common restaurant chains).

All these facts might prevent a portion of foreign tourists to move around by themselves in those cities.

Besides those insights, studying the amount of foreign tourists in the cities gives us a sense of how cosmopolitan a city is. In order to dig further the behavior of those classes we can study the temporal behavior of foreign and local tourists and also their preferences of places.

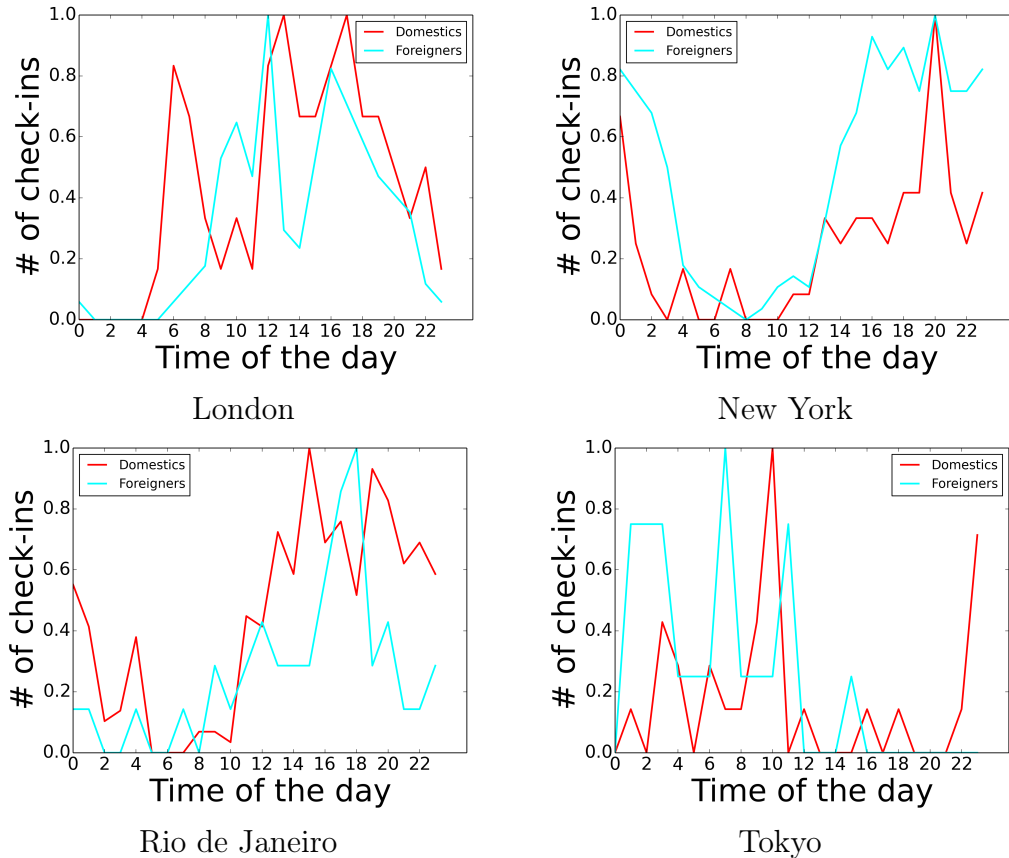


Figure 4.9. Temporal check-in sharing pattern throughout the day by domestic and foreign tourists during weekdays

Figure 4.9 shows the temporal check-in sharing pattern throughout the day by domestic and foreign tourists during weekdays. Analyzing local and foreign tourists regarding temporal aspects we can see some differences. First, by observing Figure 4.9 we can note very distinct behavioral patterns for all cities. For example, observe the figures representing the results for Rio de Janeiro and New York. According to the results, in New York foreigners go out more late at night, while in Rio de Janeiro those at class of tourists is more conservative about when to go out. One possible explanation is the barrier of the local language, as we mentioned above, and also due to security

reasons, in Rio de Janeiro the violence rate is higher than in New York, these facts prevent many tourists from going out at night time in Rio.

Still analyzing the behavior during weekdays we can see that most peaks of activities tend to be displaced for an hour. This is an evidence that foreigners may be bringing their habits to the city where they are. We omitted the results for weekends because the differences are not very significative, and the main message passed analyzing weekdays are still valid.

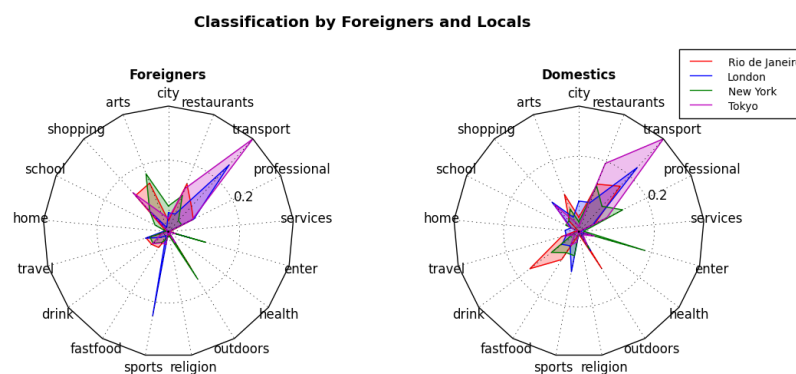


Figure 4.10. Check-ins frequency in each category by foreigners and domestic tourists by city

Understanding which are preferred places for domestic and foreign tourists helps us also to realize what the characteristics of each profile of tourists are in those cities. Figure 4.10 shows a radar chart representing the popularity of category of places for foreigners (left figure) and domestic (right figure). The popularity of categories of places was measured as we explained above.

Studying this result we can see that domestic tourists preference restaurants in Tokyo, while foreigners prefer shopping activities. Domestic tourists in Rio de Janeiro prefer outdoors and places to drink. Foreigners in Rio de Janeiro prefer places related to shopping, restaurants and arts. In London, the foreigners preference is more towards sports, while among the locals preference is greater for shopping. While foreigners like to attend places under outdoor category in New York, domestic tourists prefer entertainment places, which might demand greater knowledge of the city.

Information like those obtained in this section are useful to shape marketing strategies focused each type of tourists, as well as understanding what possible tourism products may be relevant for each of them.

4.6 Discussion

Some of the challenges to the understanding of useful properties on the behavior of tourists is to find appropriate metrics. To begin our study, our hypothesis was that tourists have more free time (no predefined routine) while residents have tied behavior to daily routines.

Beyond the time aspect, the places visited say a lot about the tourist and the purpose of visit. As from a spatial analysis of activity in the city we could see which regions the tourists were more concentrated and also the locations of the top most visit sights in each city, which showed what the most visited places in each city are.

These properties of tourist behavior can also be quite useful, as they can allow the modeling of the behavior of tourists according to a specific city, as well as being exploited in activities of recommendation and to planing activities aimed at tourists.

Chapter 5

Understanding Tourist's Mobility

The user mobility within cities can bring rich information about the dynamics of the urban environment, as well as habits of these users on their routines in the city. Using spatial data that implicitly express the preferences of users by specific locations in the city, such as check-ins, we have the possibility to know where people come from and where they go. This, as we show in this chapter, enables us to distinguish the profile of these users mobility within cities, which can be quite distinct between different cities. In this chapter we studied the mobility of tourists from different perspectives. In section 5.1 we analyze the movement of tourists in the city using two well known metrics for this purpose. In section 5.2 we use complex networks centrality metrics in spatial-temporal graphs that capture the movement of users throughout the hours of the day. Finally, in section 5.3 we demonstrate that it is possible to extract different mobility profiles based on observed movement of users.

5.1 Displacement measures

In this section we present an spatial analysis of tourists movements using the mean user displacement as shown in 5.1.1 and radius of gyration as shown in 5.1.2. These two metrics are useful to study human mobility and its implications.

5.1.1 Mean user displacement

Thinking about mobility, it is interesting to analyze the displacement of the users inside the cities. To do that, we start with a study of the mean user displacement. The mean user displacement is the mean of the cumulative distance traveled by an user. To discovery that we calculate the total distance-based displacement of consecutive

check-ins v_n made by users and divide this value by the total number of check-ins N the user has performed. The check-ins was ordered by chronological order performed by the users. The Mean Displacement User is defined by Equation 5.1:

$$d_u = [distance(v_1, v_2) + \dots + distance(v_{n-1}, v_n)]/N, \quad (5.1)$$

where V is the set of visited locations and N is the total number of check-ins. Figure 5.1 shows the cumulative distribution of mean user displacement of tourists and residents. Studying the distance traveled by tourists, we realize that tourists tend to travel a shorter distances, while the probability of residents travel long distances is higher. Although this behavior is different between tourists and residents when we examine each city it is possible to see some variations. In Rio de Janeiro, for example, tourists move more, while in London 80% of the tourists move short distances, up to 5 *km*.

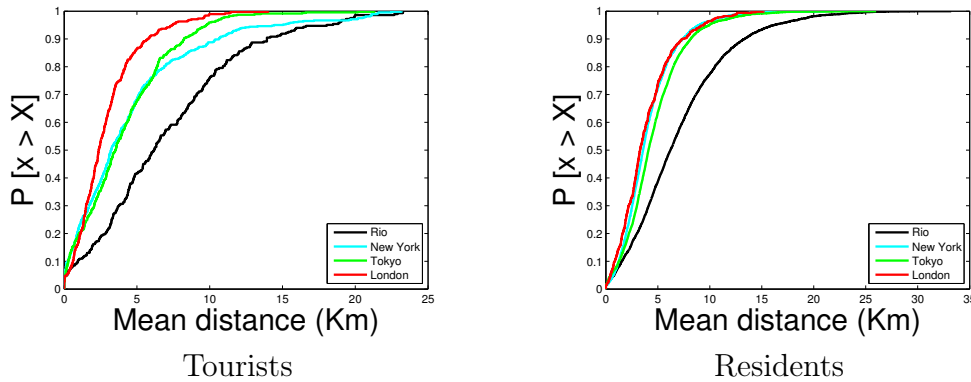


Figure 5.1. Distribution of Displacement of Tourists and Residents

Among the residents we have a higher average displacements, indicating greater distances traveled within the city. Many residents do not reside close to their jobs and usually have greater knowledge of the city they tend to explore it in different ways, including further and hidden places in the city. A possible cause for the smaller displacement among tourists is the concentration in some regions, that it may be resulted by limitation of time and knowledge of the city, making tourists travel less within the city.

5.1.2 Radius of Gyration

The radius of gyration is the typical distance traveled by an individual [?]. While the displacement gives the cumulative distance traveled between all places, the radius of

gyration indicates the area where the user was concentrated according to the points where he/she visited. With this metric we can understand the differences between the area of concentration of tourists and residents in the four studied cities, important information for urban planning and the better understanding of the dynamic of cities under this perspective.

We can calculate the radius of Gyration using Equation 5.2.

$$r_g = \sqrt{\frac{1}{N} \sum_{i \in L} n_i (r_i - r_{cm})^2}, \quad (5.2)$$

where N is the total number of check-ins, L is the set of visited sites, n_i if the number of check-ins at a place i , r_i represents the geographical coordinates, and r_{cm} is the center of mass of the individual (average coordinates). For this analysis were considered users that performed at least 5 check-ins, disregarding users that used the application sporadically.

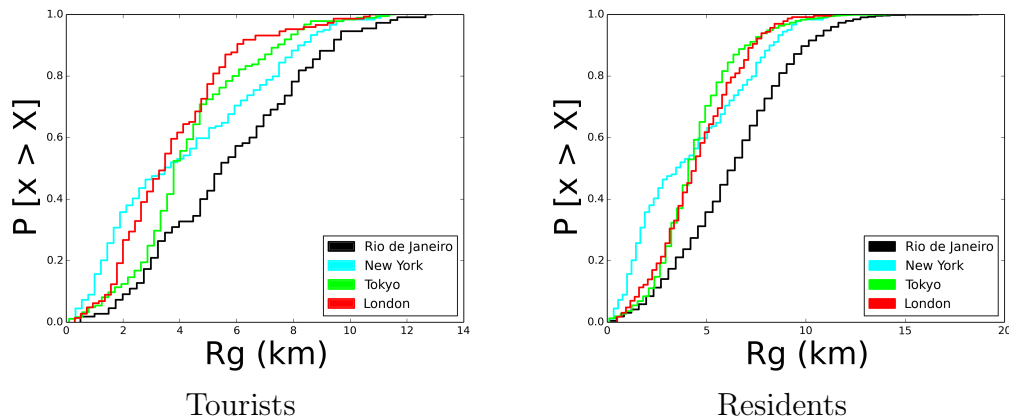


Figure 5.2. Distribution of Radius of Gyration of Tourists and Residents

Figure 5.2 shows a cumulative distribution function of radius of gyration for tourists and residents in the four cities analyzed. For the tourists of the cities studied we observe a smaller radius of gyration than among residents. This means that the area of concentration of tourists tend to be smaller than the area where the residents tend to concentrate. Among the cities there are some differences, which can be explained by geographic features and available transportation infrastructure. Tokyo, for example, has a similar behavior between tourists and residents, while Rio de Janeiro has a difference in the area of concentration of tourists and residents more expressive difference.

With the help of the radius of gyration results we can see that residents go to fewer places, compared to the tourists, but travel longer distances in the city, while tourists

go to more places but with a smaller displacement. The distance of the touristic sites and public transportation available may be a factor that influences the concentration of tourists in a location, as well as the workplace and place of residence may also influence the displacement of residents in cities.

Figure 5.3 shows the movement of users, including tourists and residents, for different values of radius of gyration in Tokyo. The position of the nodes in the graph is in line with the real geographic coordinates of each site. For residents and tourists the smallest radius gyration found was 0.1, but we can see a difference in movement between them. Although they have been moving within the same range, tourists went to different and more places. This intuitively makes sense because tourists tend to visit more places in the new environment where they are.

Studying the largest radius of gyration of tourists and residents, we have 7 different places visited by a resident against 15 visited by a tourist. Meanwhile, the radius of gyration was 14.3 km for the resident and 11.6 km for the tourist. This corroborates with the observation pointed out above, that tourists tend to visit more places, despite not moving longer distances on average compared to residents.

This metric is useful to understand how tourists move within cities and also help to improve recommendation systems for places to tourists. If a tourist have an explorer profile, who likes explorer distant places, we can suggest places in a larger area; following the same idea, if the tourist is more conservative regarding to the distances he usually travels in the city, the suggestion of places should stay in a smaller radius.

5.2 Centrality metrics on spatio-temporal urban mobility graphs

Linked to spatial data, another important factor to understand users mobility is the time. The movement of users might change according to the day of the week and time. For this reason in this section we perform analysis considering the time dimension in the study of mobility of tourists and residents.

5.2.1 Spatio-temporal urban mobility graphs

Graph theory is a powerful tool for representing relationships between entities such as individuals or other entities. In our case, we use a directed weighted graph $G = (V, E)$, where the nodes $v_i \in V$ are specific venues in the city in a certain time (for example, Times Square at 10:00 a.m.) , and a directed edge (i, j) exists from node v_i to v_j if

at some point in time an user performed a check-in at a venue v_j after performing a check-in in v_i .

In our model, we use a 24-hour time interval starting at 5:00 a.m. (instead of 12:00 p.m.). Our goal was to capture nightlife activities using this strategy. The label of vertices follows a simple rule: the name of the location concatenated with the integer hour of the check-in. For instance, a check-in at Times Square at 10:00 a.m. would be “Times Square [10]”. When another user has performed the same trajectory, is incremented one to the weight of the edge. In other words, the weight $w(i, j)$ of an edge is the total number of transitions that occurred from node v_i to node v_j . Isolated vertices were removed from the graph, since there is no movement associated with that particular vertex.

Figure 5.4 depicts our graph model with locations and temporal attributes. It is possible to notice the movement between different locations from the directed edge (with continuous line). The dashed line represents a link between the same location and the temporal distance between consecutive check-ins at that location. The directed edge represents two consecutive check-ins performed by the same user and the weight of the edge the number of users that performed this same tuple of check-ins. For example, in the figure the edge that connects the vertices “Corcovado[10]” and “Maracanã[14]”, both in Rio de Janeiro, represents consecutive check-ins performed at Corcovado at 10:00 a.m. and then at Maracanã Stadium at 2:00 p.m. ten different times.

In a city, there may be thousands of different combination of movements between places, and some of them tend to be more popular than others. Our graph model enables to study the movement of users along the time, and it can also be used to find important places in the cities. The importance of these places can be seen from different perspectives, such as, for instance, popularity in terms of number of visits or best places to disseminate information in the city. For those perspectives there are centrality metric of complex networks that help us understand the importance of the places in the cities, and we discuss some of them next.

5.2.2 Popular venues in the city

The reasons that motivates one person to travel from one place to another might be diverse, for instance: leisure, business, shopping, gastronomy, and academic. The behavior of tourists inside the cities tend to reflect these reasons in many ways. However, we know that some places and needs are common to all kind of tourists, such as food, accommodation, and transportation. In our study we show that we can have the opportunity to go one step forward in the understanding of tourists’ behavior, identifying

where and when places are more important to users (tourists and also residents) in different cities.

In a city might be possible to find thousands of different combination of movements between places, and some of them tend to be more popular than others. Our graph model enables to study the movement of users along the time, and it can also be used to find important places in the cities. It is possible to find central nodes (that represent places) in the graph. To illustrate this idea, we evaluate the degree centrality measure to identify the most important locations in the cities according to this metric.

In a graph G , the degree centrality of a vertex v is the number of incident edges on v normalized by dividing by the maximum degree in the graph. Vertices with a higher degree centrality have a higher number of connections to other vertices of the graph. In the urban mobility graphs of tourists and residents, the higher the degree of vertices, the greater their popularity in the graph.

Table 5.1 (left side) shows the top ten places with the highest degree centrality of the residents’ graph of New York City. Subcategories of places express the behavior of residents, such as neighborhoods, bus station and buildings. We can see that such places are typically visited by people who live in the city and choose places related to their daily activities.

Residents		Tourists	
Venue[time]	Subcategory	Venue[time]	Subcategory
Times Square[16]	Plaza	John F. Kennedy International Airport (JFK)[8]	Airport
Times Square[17]	Plaza	Brooklyn Beer & Soda[19]	Food & Drink Shop
New York Times Building[16]	Office	Wall Street[18]	Street
New York State DMV[18]	Government Building	Times Square[22]	Plaza
Herald Square[17]	Plaza	National September 11 Memorial & Museum[19]	Historic Site
Boi Noodles[16]	Vietnamese Restaurant	LaGuardia Airport (LGA)[6]	Airport
Dunkin’ Donuts[16]	Coffee Shop	New-York Historical Society Museum & Library[13]	Museum
Herald Square[18]	Plaza	Brooklyn Brewery[23]	Brewery
Port Authority Bus Terminal[16]	Bus Station	Charging Bull[18]	Government Building
Herald Square[21]	Plaza	Mike & Tony’s Pizza[19]	Pizza Place

Table 5.1. Ranking of degree centrality of New York

In the first places in the ranking of residents according to the degree centrality, we found Times Square. Although it is defined as square, it is composed of several intersections in downtown Manhattan. It has many shops and offices of large companies, as well as restaurants and sites related to art, attracting many residents. Times Square

is more popular in the late afternoon hours where people begin to prepare to end their shift. These facts could help to explain this result. We can also see in the result table that The New York Times Building, a building which houses major newspapers like The New York Times and International Herald Tribune, is also popular at 4 p.m. (16 hours), perhaps this fact could be explained by the coffee break time of the employees who work there (this place is not among the most popular ones for tourists, which is expected).

Another popular building, also in the evening, is the New York State DMV (Department of Motor Vehicles), where citizens can solve disputes related to vehicle licenses. This is a place that is usually popular in the city, which helps to explain this result. Another popular region, but like Times Square is also classified as Square, is the Herald Square. This region is located in midtown Manhattan and is popular due to its location and high traffic of people. The most popular times are in the late afternoon and at evening: 17, 18 hours (periods that agree with typical hours when people have to return home from their daily routines), and 21 hours.

Dunkin' Donuts is a famous franchise that sells coffee and donuts, popular foods in the United States. Among the residents of NY, this location is most popular in the late afternoon, popular time to make a stop for snacks. Another typical place between people who live in cities are train and bus stations. The Port Authority Bus Terminal is the largest bus terminal in the world and is busy in the late afternoon.

In table 5.1 (right side), we can see the ranking of degree centrality of the tourist's graph of New York. As expected, we have the presence of airports among the most popular places to tourists. The John F. Kennedy International Airport is one of most crowded airports from EUA, bringing people from abroad and also serving like a hub to other cities. Among the tourists, it has more traffic of people at morning, perhaps this is because many intercontinental flights fly overnight and land in the city in the morning.

Studying other popular places among tourists, especially in Manhattan where most tourists tend to visit in NYC, we have as the most popular places Wall Street, National September 11 Memorial & Museum, New-York Historical Society Museum & Library and Times Square. Those places are expected to be popular among tourists because they are well known sights worldwide. Another place frequented by tourists is the Charging Bull, which is a bronze sculpture that stands in Bowling Green Park in the Financial District in Manhattan. The sculpture is featured in the films *For Richer or Poorer* (1997), *Hitch* (2005), *Inside Man* (2006), *The Other Guys* (2010), *The Sorcerer's Apprentice* (2010), *Arthur* (2011) and *The Wolf of Wall Street* (2013). It also appears in the TV series *My Life as Liz* and *Weeds*. This helps to explain its

popularity.

The Brooklyn Beer & Soda and Mike & Tony's Pizza, both located in Brooklyn and offer a typical American menu. Both are popular at night, happy hour and dinner hours, respectively. In Brooklyn we have also Brooklyn Brewery, being most popular at 23 hours. Note that tourists tend to go to Brooklyn to enjoy nightlife in bars and restaurants. One natural question that arises is: what makes Brooklyn instead of Manhattan have the most popular places among tourists at night to enjoy nightlife, since Manhattan also have many of those options?

In the table 5.2 we can see the most important places in the graph of residents and tourists of Rio de Janeiro. Analyzing the behavior of residents in Rio de Janeiro, we observe a considerable interest for shopping malls and municipalities. Rio Sul shopping is located in the south region of Rio de Janeiro, a central region and it is popular at 15 hours. In regions less central there are Barra Shopping (west region), more popular at the end of the day, what helps to explain that is the fact that malls are convenient for going after the working hours, when regular stores, outside the malls, are usually close. Malls are important places to access services and products in the daily routine of people who lives in Rio de Janeiro.

Leme neighborhood, upscale neighborhood of Rio de Janeiro, is quite popular in the early morning. In the same neighborhood, physical activities are common among residents because it is near the beach and have bike and running paths. We also found a university among the most popular places, popular at 18h. In Brazil it is common people study at night in universities, helping to explain this popular time.

Likewise we found for New York, we have evidence that the most popular places among residents are typical of places visited for this class. This helps to validate that our results are capturing common behavior of residents and tourists. Interestingly, in the city of Rio de Janeiro is where we found more check-ins at "home" among residents. This suggests that residents of Rio are less concerned about privacy.

Table 5.2 also shows the highest degree centrality observed in the tourists' graph. There is a huge concentration of tourists visiting the Santos Dummont airport, quite popular in the morning. Some popular sights in Rio appear in this ranking, such as Copacabana Beach and Morro da Urca. These points are one of the most visited by tourists who come to Rio to see its natural beauty. Both show popularity during the day, best time to enjoy the scenery. We also found an university, PUC-Rio, popular among tourists. This is a prestigious university in Brazil and hosts researchers and students from around the world. Besides, it is located in a great location in Rio de Janeiro, nearby Jardim Botânico, another famous spot in Rio. This result could indicate that this university receives academic tourists.

Residents		Tourists	
Venue[time]	Subcategory	Venue[time]	Subcategory
Leme[6]	States & Municipalities	Bob's[11]	Burger Joint
Leme[7]	States & Municipalities	Aeroporto Santos Dumont (SDU)[8]	Airport
Universidade Veiga de Almeida (UVA)[18]	University	Aeroporto Santos Dumont (SDU)[11]	Airport
Companhia do Garfo[7]	Brazilian Restaurant	Aeroporto Santos Dumont (SDU)[7]	Airport
Jr mini pizza[18]	Pizza Place	Praia de Copacabana[17]	Beach
Shopping RioSul[15]	Mall	Aeroporto Santos Dumont (SDU)[9]	Airport
BarraShopping[18]	Mall	Rio de Janeiro[9]	States & Municipalities
Rio de Janeiro[20]	States & Municipalities	boat party[14]	Other Event
Calçadão de Nilópolis[16]	Pedestrian Plaza	Morro da Urca[9]	Mountain
Condomínio Valência e Sevilha[21]	Home (private)	PUC-Rio[8]	University

Table 5.2. Ranking of degree centrality of Rio de Janeiro

5.2.3 Spreading information

Closeness centrality Freeman [1979] of a node v is the reciprocal of the sum of the shortest path distances from v to all $n - 1$ other nodes. Closeness is normalized by the sum of minimum possible distances $n - 1$, since the sum of distances depends on the number of nodes in the graph. In other words, the metric closeness centrality measures how close a vertex v is of all others in a graph G . For that, it is taken into consideration the number of edges separating a node from others. The shorter the distance to all other nodes, the higher is its closeness centrality. With this measure, for example, we can estimate how fast it is possible to reach all vertices in G from v .

In the urban mobility graphs of tourists and residents a node with high closeness centrality indicates an “influential” place at a certain time. In the context we are studying, locations (vertices/nodes) with high value of closeness centrality indicate, for example, strategic locations for the dissemination of information for these two classes of users.

Table 5.3 shows the top ten locations with higher values of closeness centrality of the graph of London residents. According to this ranking, the best places to disseminate information among residents are outdoor locations (such as gardens and piers), supermarkets, coffee shops, and train stations. These are key places for residents as they represent common interests to all (e.g. food and transport). According to the concept of closeness centrality it means that these places are the shortest paths between different routes in the graph.

The Greenwich Market is a big market in London that has art crafts, antiques, fruits and vegetables. At 15 hours it is the best place to spread information and reach residents who pass through it. Another place that also has a good spread of information is the Soho Square, specially at 16 hours. This square attracts many people who are transiting in the region. During the summer there are several outdoor concerts. Next to it and around the same time of most popularity is the Greenwich Pier, public cruises service to central London. Cafes usually attract different people and it is a good place to publicize some things. The time of most popularity, might depend of the region. The Coffee Republic (located on Oxford Street), which is a fairly traditional London franchise has the most popular time at 07 hours.

Chinatown concentrates many Asian restaurants and the best time to reach residents is around dinner time, at 21 hours. This results could be explained by the fact that this area attracts many people to eat Chinese food, during the most important meal for English people, in one of the dozens restaurants of this type available there. It is worth noting that some sites are close, such as Greenwich Market and Greenwich Pier, being an interesting region to reach residents.

Curiously, we found a hotel, JW Marriott Grosvenor House Hotel popular at 19 p.m., among the most popular places to residents according to the studied metric. Investigating possible explanations for that we discover that the Great Room at London's Grosvenor House Hotel is one of Europe's largest hotel conference and banqueting venues. Capable of seating 1,770 people, it is a popular venue for conventions, grand social occasions and televised award ceremonies. One possible explanation for the popularity observed is because this place might attracts many locals due to different events that are hosted in the facilities of this hotel. The time of most popularity indicates that this events could be banquets and ceremonies, typically held at night.

Residents		Tourists	
Venue[time]	Subcategory	Venue[time]	Subcategory
Coffee Republic[7]	Coffee Shop	Urban Outfitters[18]	Clothing Store
Cutty Sark DLR Station[13]	Light Rail	240 Edgware road[13]	Road
Greenwich Market[15]	Market	Light Bar[17]	Cocktail Bar
Greenwich Pier[15]	Pier	National Gallery[9]	Museum
JW Marriott Grosvenor House Hotel[19]	Hotel	Buckingham Palace[12]	Palace
Chinatown[21]	States & Municipalities	Buckingham Palace[10]	Palace
Soho Square[16]	Garden	Duke of Wellington[22]	Gay Bar
Westfield Stratford City[9]	Mall	Wok to Walk[13]	Chinese Restaurant
Cutty Sark[15]	Museum	MacIntyre Coffee[14]	Coffee Shop
TfL Bus 23[15]	Bus Station	Flat Iron[14]	Steakhouse

Table 5.3. Ranking of closeness centrality of London

As expected, some sights are ideal to disseminate information among tourists. In Table 5.3 we can see the rankings of locations with higher values of closeness centrality of tourists graph for London. We identified some famous sights such as Buckingham Palace and the National Gallery, but also other places that are not in traditional itineraries of tourists.

The Urban Outfitters clothing store can be an interesting place to disseminate information late in the evening among tourists who are visiting London. Many tourists like to shop while enjoying the exploration of the city. The 240 Edgware road is a city address that has several buildings with Victorian architecture. The architecture and shopping opportunities in the region can be attractive for tourists. During the afternoon many food sites, such as MacIntyre Coffee, Flat Iron, Wok to Walk and Light Bar are among the most central places to disseminate information among tourists.

The insights that could be extracted using closeness centrality is interesting help in the decision making about choosing places to disseminate information to the two studied classes of users. Such insights can be used by the government to promote more effective public campaigns among residents, and for tourism companies that want to maximize their contact with potential consumers.

5.2.4 Bridge Places

Betweenness centrality of a node v is the sum of the fraction of all-pairs shortest paths that pass through v . Studying this centrality metrics in the mobility graph of tourists and residents we can see which places can make connection between distinct components within the graph. Bringing this to the context of this research, we can look at this metric as an indication of the places that could act as bridges between different groups of places and times. The higher the betweenness, the greater the chance that a user go through that particular location Silva et al. [2014c]. Table 5.4 shows the ranking of the top ten places according to the betweenness centrality in residents and tourists graph of Rio de Janeiro. Among them we have places that are relate to different niche of locations such as subway, bus station and restaurants, popular at the end of the night.

The FIFA World Cup 2014 happened in Brazil and Rio de Janeiro was one of the host cities. An interesting fact to note is the presence of the FIFA Fan Fest, official place of celebration organized by FIFA (Fédération Internationale de Football Association), to gather supporters for all the world cup games. This party was a central location in the routine of people in this city, concentrating more residents at night. Note that these types of location is likely to be a good place to connect different tribes in the

city, fact that could help to justify the result.

Residents		Tourists	
Venue[time]	Subcategory	Venue[time]	Subcategory
City Rio[23]	Bus Station	Pimenta’s Bar[16]	Bar
G & M centro Automotivo[23]	Automotive Shop	Rio de Janeiro[14]	Historic Site
Point dos Amigos[23]	Burger Joint	Bacana Da Gloria[14]	Brazilian Restaurant
FIFA Fan Fest[23]	Festival	Atelie Catherine Hill[14]	Cosmetics Shop
MetrôRio - Estação Irajá[23]	Subway	Emilio’s Bar[16]	Bar
Ki Pizzaria[23]	Pizza Place	Sindicato do Chopp[16]	Bar
Linha 712 - Cascadura / Irajá[23]	Bus Station	boat party[14]	Other Event
MetrôRio - Estação Cardeal Arcoverde[23]	Subway	brother’s bar[16]	Bar
Mista do Léo[23]	Snack Place	Museu Da Policia Militar[16]	Museum

Table 5.4. Ranking of betweenness centrality of Rio de Janeiro

We now turn our attention to tourists in Rio de Janeiro. We can see in Table 5.4 the betweenness the centrality of tourists in that city. We observe that many bars are popular for tourists, according to the betweenness centrality, around 16 hours. Bar is a good option in Rio de Janeiro to eat and drink, going to the bars is a quite common activity for tourists of all profiles in Rio de Janeiro. This helps to justify that this sort of place are interesting places to connect different kinds of tourists, and perhaps this could be explored in the development of new types of application, for example, to improve user interaction in the city.

In addition to bars, we have also sights, such as the Military Police Museum, popular at 16 p.m. The boat party was an event that took place at Gloria Marina during the time of our data collection, which attracted several tourists at that time. We note that all these places have the highest concentration of tourists in the afternoon.

5.2.5 Validating the results

To validate the existing patterns observed in our urban mobility graphs, we generated a null model. The null model consists of a directed weighted graph $G_{Ri}(V, E_{Ri})$, where $i = 1, \dots, 100$. This means that nodes are the same of the original graph that we want to generate a null model to, and the edges are created randomly, respecting to the total sum of the edges weights of the original graph. Edges are distributed between two nodes randomly, up to the number of edge weights. With that, the null model preserves the same number of edges of the original graph. If that edge already exists,

the weight is increased, if not, a new edge is created with weight 1. Thus, this created graph simulates a random walk taken by users in the city.

Tables 5.5, 5.6, and 5.7 show a comparison between the centrality metrics used in the previous sections and their values generated using the null model. We are showing one example for each metric for London, New York, and Rio de Janeiro, for each of them, just to exemplify the results. They represent the key message, for this reason we opted to present the results in a compact way.

To illustrate the results, in Table 5.5 we find that the top two most popular places in the null model in New York for Residents, according to the degree centrality metric, are two regular restaurants in the city. Table 5.6 shows the closeness centrality for the null model for tourists in London. As we can see, the top two places are a community center and a health club, which are not typical choices for tourists. Finally, Table 5.7 shows the betweenness centrality for the null model for the graph of residents in Rio de Janeiro. The results shows that the two most popular places are a regular restaurant and a residential building, places that do not make much sense to be the most popular places among all residents in that that, specially under the metric betweenness.

Studying the results found in the null models, we do not find much temporal or semantic sense for the popularity of places found for tourists and residents. This indicates that our data, and also our original approach (urban mobility graphs) are reflecting typical habits and routines performed by the users in both studied classes and cities.

New York Residents'	
Graph	Null model
Times Square[16]	Mojave NYC[19]
Times Square[17]	Tuck Shop[16]
New York Times Building[16]	LIRR - Atlantic Terminal[14]
New York State DMV[18]	65 Broadway[10]
Herald Square[17]	Corona, NY[3]
Boi Noodles[16]	P&K Food Market[8]
Dunkin' Donuts[16]	Serafina Meatpacking[12]
Herald Square[18]	John F. Kennedy Airport[3]
Port Authority Bus Terminal[16]	Brooklyn Bridge[10]
Herald Square[21]	Plaza

Table 5.5. Comparative ranking of degree centrality of the resident's graph and null model graph of New York

5.3 Profiles of Tourists Based on Mobility Patterns

Related to the motivation of visiting a new city there is the search for experiences that interest tourists (e.g., the local culture of a city). Each city has features, such as

London Tourists’	
Graph	Null model
Urban Outfitters[18]	Better Kings Hall Leisure Centre[19]
240 Edgware road[13]	The Chelsea Club[10]
Light Bar[17]	RBS Bankside[14]
National Gallery[9]	Waterstones[15]
Buckingham Palace[12]	The Harry Potter Shop at Platform 9 3/4 [12]
Buckingham Palace[10]	South Quay DLR Station[22]
Duke of Wellington[22]	Buckingham Palace[10]
Wok to Walk[13]	John Lewis[16]
MacIntyre Coffee[14]	Marble Arch[16]
Flat Iron[14]	Selfridges & Co[16]

Table 5.6. Ranking of closeness centrality of the tourist’s graph and null model graph of London

Rio de Janeiro Residents’	
Graph	Null model
City Rio[23]	Grill Churrascaria e Pizzaria[22]
G & M centro Automotivo[23]	Condominio Jardim Pindorama[9]
Point dos Amigos[23]	Ponto de Onibus[6]
FIFA Fan Fest[23]	Condominio Valencia e Sevilha[2]
MetrôRio - Estação Irajá[23]	Koni Store[0]
Ki Pizzaria[23]	IBEx - Instituto de Biologia do Exército[8]
Linha 712 - Cascadura / Irajá[23]	Praia de Copacabana[22]
MetrôRio - Estação Cardeal Arcoverde[23]	Bar da Bud[1]
Mista do Léo[23]	Terminal 1 (TPS1)[5]
MetrôRio - Estação São Cristóvão[22]	Estadio Jornalista Mario Filho (Maracana)[21]

Table 5.7. Ranking of betweenness centrality of the resident’s graph and null model graph of Rio de Janeiro

religious temples, beautiful scenery and/or a rich local cuisine. Often, cities or countries may be recognized by these features and end up attracting many visitors with related interests. There are some specific purposes of tourisms, such as gastronomic tourism, religious tourism, ecotourism, business. These specific types of tourism exists because we have distinct profile of tourists, based on different interests such as sports, business and cooking.

Through check-ins we can get an idea of how tourists behave in cities, and, therefore, have the opportunity to identify profiles of tourists mobility patterns according to the interests of each tourist. We find profiles based on mobility pattern identifying the set of most visited places by a certain group of people. This is interesting in several cases, in addition to knowing which users belong to a group, useful information to recommend places to other users with similar profile, we can identify what are the characteristics that attract groups of people to that place.

To identify the profiles of tourists based on mobility patterns we use the technique for topic modeling Latent Dirichlet Allocation (LDA) Blei et al. [2003]. This technique is useful to summarize documents in a set of topics, finding words that de-

fine a document, i.e., its subject. The LDA considers a set of documents and a set of words contained in these documents, and the intuition behind this technique is that each document has several topics, and each topic is a distribution of probabilities for a word in the vocabulary.

With the help of check-ins of each user, we can know the number of times each user visited each subcategory (local type). We then consider the subcategory name of the place visited, which can be repeated, as the word of a “document” that represents the user. Subcategories, such as Office and Coffee Shop are examples of document words. With this methodology we are able to get results that indicate, in a way, user profiles.

We can view the topics found for Tokyo residents in Table 5.8. From the subcategories, we classify each topic with a name that represents a profile. The Commuter profile is a topic that use quite a lot urban public transportation, such as train and subway stations. In this group can be included residents of the Tokyo metropolitan area. Of course, in Tokyo there are several Japanese restaurants. The topic representing many bars and restaurants was named Asian Food Lover. In the Academic profile we have a group of people who, in addition to performing routine activities, such as using public transport and eating in restaurants, considerably attend universities.

Profile	Subcategories of most represented places to each group
Commuter	Subway, Train Station, Convenience Store, Bridge
Asian Food Lover	Japanese Restaurant, Ramen / Noodle House, Bar, Chinese Restaurant
Academic	Train Station, Arcade, Ramen / Noodle House, University

Table 5.8. Profiles of residents in Tokyo according to venues subcategory

Profile	Subcategories of most represented places to each group
Electronics Enthusiastic	Electronics Store, Train Station, Café, Ramen / Noodle House
Commuter	Subway, Train Station, Convenience Store, Bus Station
Gammer	Train Station, Arcade, Ramen / Noodle House, Electronics Store

Table 5.9. Profiles of tourists in Tokyo according to venues subcategory

Tourists’ profiles in Tokyo city can be seen in Table 5.9. Many tourists go to Tokyo motivated by technological appeal of the city, as well as motivated by local cuisine. We can view in the profile Electronics Enthusiastic a strong presence of electronics stores, besides other things offered by the city. Following the same line, the Gammer profile is similar to Enthusiastic Electronics, however with a bias to games. Just as the profiles found among residents, we have also a profile called Commuter to tourists. We understand that this profile is composed of users who are visiting Tokyo but just do more check-ins at the train stations they pass than in other types of places visited. As

we pointed out above, a check-in in a train station might be a way to reveal to friends key areas of the city that you are visiting.

Turning now to the other side of the world, in the table 5.10 have the profiles of Rio residents also found it a commuter profile, similar to that found in Tokyo characteristic by the presence of urban transport. Rio has some conurbation cities, making frequent this profile residents. Also, we have the profile Academic, marked by the great presence in educational institutions. Another profile identified, which is quite lives up the city, is the Citizen, marked by the popularity of shopping malls in the city.

Profile	Subcategories of most represented places to each group
Commuter	Home (private), Bus Station, Road, States & Municipalities
Academic	Home (private), School, Mall, University
Citizen	Mall, Subway, Plaza, Road

Table 5.10. Profiles of residents in Rio de Janeiro according to venues subcategory

Profile	Subcategories of most represented places to each group
Business & Academic	Office, University, Restaurant, Pizza Place
Business	Airport, Beach, Government Building, States & Municipalities
Leisure	Airport, Hotel, Bar, Beach

Table 5.11. Profiles of tourists in Rio de Janeiro according to venues subcategory

The Rio de Janeiro, one of the largest cities in Brazil, attracts many tourists for its natural beauty. As a metropolis, also receive different types of tourists. Among the profiles found we can see a leisure tourist, typical of those going to Rio for sightseeing. But we also find business profiles, related work activities, but without ceasing to enjoy the city and what it has to offer attractive, such as restaurants and beaches.

The mobility profile is also influenced by the routine of people on different days of the week. Analyzing travel behavior on weekdays and weekends we noticed a difference in the behavior of people in cities. Tables 5.12 and 5.13 have the profiles of New York residents divided by weekdays and weekends, respectively.

Profile	Subcategories of most represented places to each group
Public transport user	Subway, Home (private), Bus Station, Train Station
Cosmopolitan	Gym / Fitness Center, Performing Arts Venue, Bar, Coffee Shop
Worker	Office, Gym / Fitness Center, Coffee Shop, Building

Table 5.12. Profiles of residents in New York according to venues subcategory during weekdays

Analyzing the New York resident mobility profiles during the week we noticed a pattern related to the routine of work, is the intensive use of means of transport, trips

Profile	Subcategories of most represented places to each group
Nightlife (1)	Bar, Park, American Restaurant, Gay Bar
Nightlife (2)	Bar, Lounge, Gym / Fitness Center, Music Venue
Food Lover	Home (private), Subway, Food & Drink Shop, Train Station

Table 5.13. Profiles of residents in New York according to venues subcategory during weekends

to work or the gym. Observing the residents of mobility standard for the weekend, we noticed a change in the profiles with many activities related to leisure, such as going to bars, park and restaurants.

Profile	Subcategories of most represented places to each group
Business	Office, Gym / Fitness Center, Coffee Shop, Food & Drink Shop
Nightlife	Bar, Stadium, Pub, Sports Bar
Arts	Plaza, Hotel, Performing Arts Venue, Clothing Store

Table 5.14. Profiles of tourists in New York according to venues subcategory during weekdays

Profile	Subcategories of most represented places to each group
Parks	Hotel, Airport, Train Station, Plaza
Nightlife	Hotel, Park, Nightclub, General Entertainment
Shopping	Coffee Shop, Bar, Clothing Store, Gay Bar

Table 5.15. Profiles of tourists in New York according to venues subcategory during weekdays

Watching the tourists from New York in Tables 5.15 and 5.13, respectively, we see how they behave tourists on weekdays and weekends. During the week we can see the appearance of the Business profile, a pattern of mobility which involves tourists who have way trip to work. Although during the week realize activities related to leisure, through the profiles and Nightlife Arts, weekends this activity is intensified.

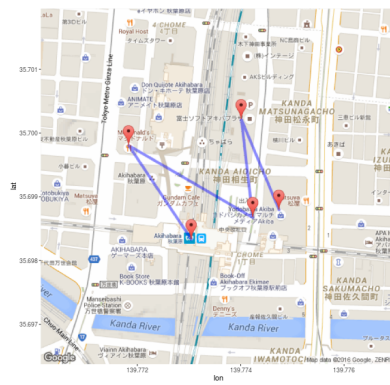
The identification of these mobility profiles is extremely important to identify the behavior of the mobility of different groups of people. The temporal aspect, as weekdays and weekends, helps to segment and better visualize how the mobility behavior are distributed within the two classes of users. This information can serve as input for urban planning, seeking to better serve the people visiting the cities and their displacement, and also to recommendation systems suggest locations that have higher affinity with the person's interests.

5.4 Discussion

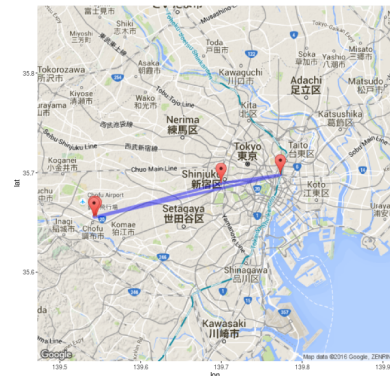
To do a better study on mobility in physical spaces we use metrics such as Displacement and Radius of Gyration. From these two metrics we analyze how tourists and residents are moving differently within cities and between them. We were able to verify, for instance, that tourists tend to stay in a region, moving less, while residents exploits more the city.

In addition to displacement, we also explore how places are related from a graph of urban mobility, where we utilize centrality metrics to understand which are the best places to be and to disseminate information. The use of these centrality metrics in the studied context is a very powerful tool that could be explored in several ways. For example, using these measures of centrality it is possible to know which places have the higher influence on a region. Exploring the insights that could be obtained using those metrics, several new applications could, potentially, be created.

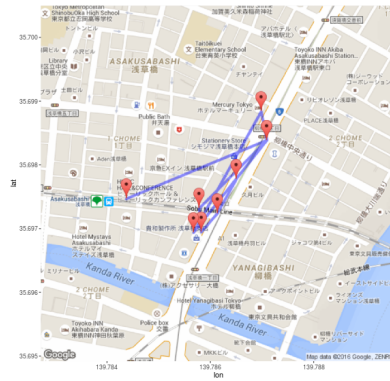
After analyzing the displacement in these perspectives, we also analyzed the mobility profile of tourists and residents, where we could identify business and leisure travelers, for example, from the presented mobility pattern.



Smallest- Resident



Biggest - Resident



Smallest - Tourist



Biggest - Tourist

Figure 5.3. Visualization of the movement of users for different values of radius of gyration in Tokyo.

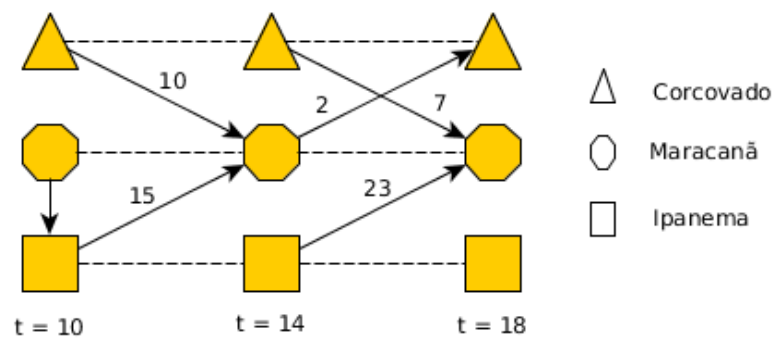


Figure 5.4. Illustration of the graph model considered

Chapter 6

Applications

The study of the behavior of tourists and residents aims to understand which properties influence/help to explain the activities performed by users of those classes, which tend to be different, as we showed in the previous chapters. This understanding enables opportunities for useful applications to everyday life of these users. In this chapter we show some possible applications that use the information obtained with the methodologies considered in this study. Section 6.1 discuss how to use our methodologies to discover behavior of consumers. Section 6.2 describes the use of urban mobility graphs to generate suggestions of tourist itineraries.

6.1 Profile of consumers

In our research, we can also analyze the locations people choose most often as a starting point to other places or as a final destination, including the corresponding period of the day. This kind of information allows us to study the behavior of people in the cities, propose a better urban planning, and create business strategies. As an example, we analyze Starbucks, a coffee shop present in several places in the world, using data from Rio de Janeiro and New York.

Analyze preferences and behavior of consumers of one particular business franchise is interesting because, typically, the purpose is to reach a diverse audience in different locations, always with the aim to expand the options of products and loyalty of customers. It is also interesting to compare this behavior with other consumers of coffee shops, in general, i.e. not only of the franchise, to answer the question: is there any difference between the behavior of consumers who attends a particular establishment and consumers attending all establishments of the same category?

Figure 6.1 shows the popularity of visits to Starbucks throughout the day in New York, for residents (Figure 6.1a) and tourists (6.1b). The franchise is known to have many units in the city of New York, which has 283 Starbucks units¹. Studying the residents visits in Starbucks, we note that there is a greater popularity at lunch time, around 16 hours and considerable popularity from 19 to 21 hours. Other coffee shops in the city are also very popular around lunch time, however we do not observe a peak of popularity in the afternoon, and they are not very popular at night as Starbucks. Studying the behavior of tourists we can see that there is no regular pattern influenced by typical routines of inhabitants of the city. Among tourists, Starbucks and other coffee shops are popular around lunch time, however, the popularity of Starbucks among tourists drops around 16 hours, while there is a growth of popularity for other coffee shops in town. The highest peak of popularity among tourists in Starbucks is at night, differently of residents which is around lunch time. This behavior can be an indication that residents use more Starbucks for practical reasons during their daily activities, while tourists give a different value to the place, perhaps considering it a kind of attraction.

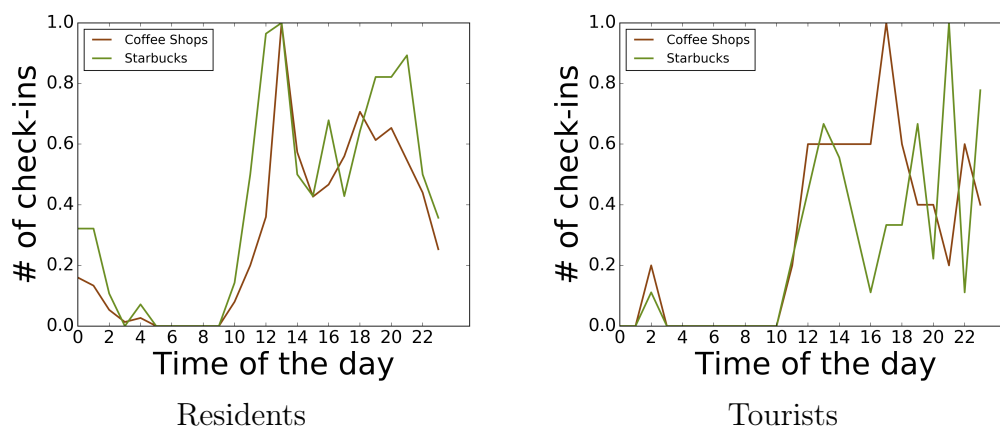


Figure 6.1. Distribution of the time interval (in hours) between the check-ins performed by Starbucks (green) and other Coffee Shops (brown) at New York

Analyzing the visits in these two types of places we can better understand the dynamics of establishments in the city and how to differentiate themselves from their competitors. Another important aspect to study about the behavior of consumer is the places users go before and after being a in certain place.

We continue with our case of example: Starbucks. It is interesting to identify the places where people come from before visiting Starbucks and going to after visiting Starbucks to understand the characteristics of Starbucks' customers. Using the urban

¹<https://nycfuture.org/research/publications/state-of-the-chains-2013>

mobility graph G described on Section 5, we can track the most popular places where people was before and after to go to some place. This is possible by looking through the edges $e_i(v_{n-1}, v_n)$ related to the correspondent node v_n of the venue. Here we also consider Starbucks and other Coffee Shops. Figures 6.2 and 6.3 show the most popular subcategories of places that precede and succeed the visit to Starbucks of residents and tourists of New York, respectively. Yellow nodes represent venues where people were before going to Starbucks. Green nodes represent venues where people went after going to Starbucks.

Note that, in New York, where the franchise is quite popular, we observe typical destinations for residents, such as school (before going to Starbucks), and drugstore (after going to Starbucks).

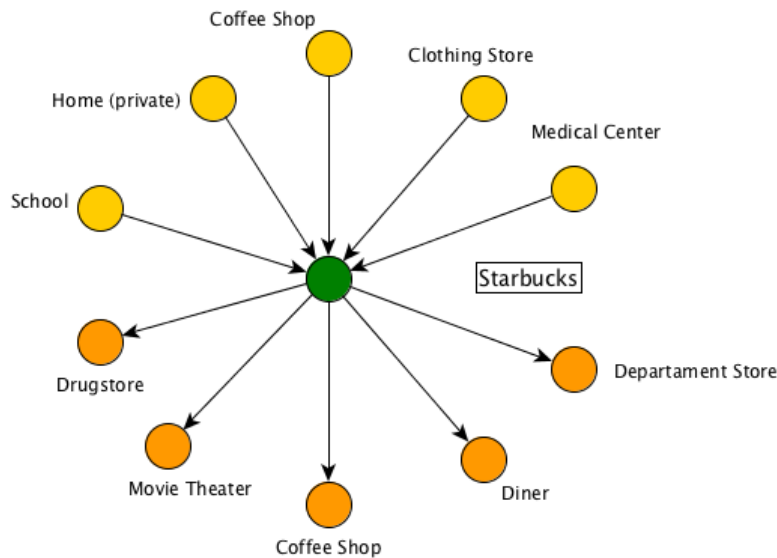


Figure 6.2. Subgraph of places visited by New York residents before and after other Starbucks



Figure 6.3. Subgraph of places visited by New York tourists before and after other Starbucks

Analyzing the subgraph of the venues visited by tourists, Figure 6.3, we see some differences from the graphs of residents. Among the tourists, we observe that users who frequent Starbucks tend to frequent before plazas and parks (very common in New York). After visiting Starbucks, it is common to visit places to shop and sightseeing (some historic buildings are classified as Building).

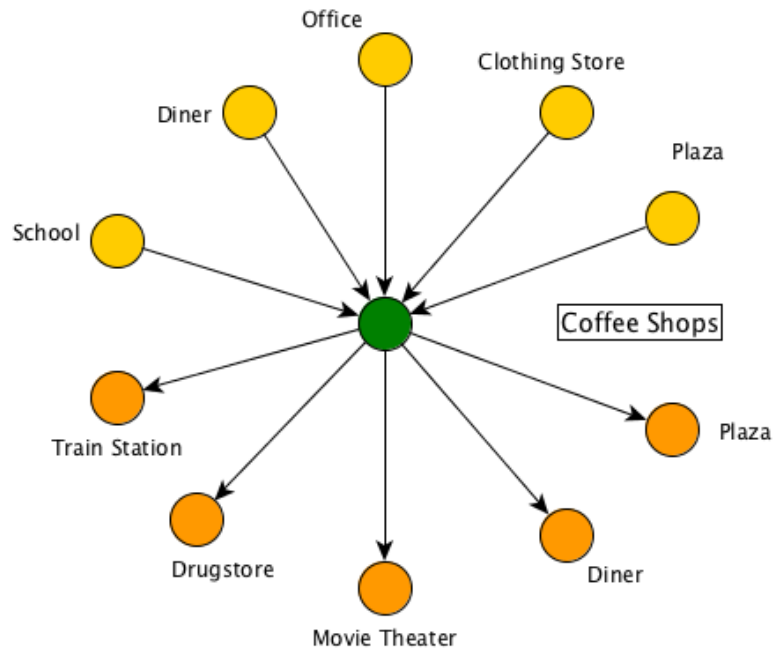


Figure 6.4. Subgraph of places visited by New York residents before and after other Coffee Shops

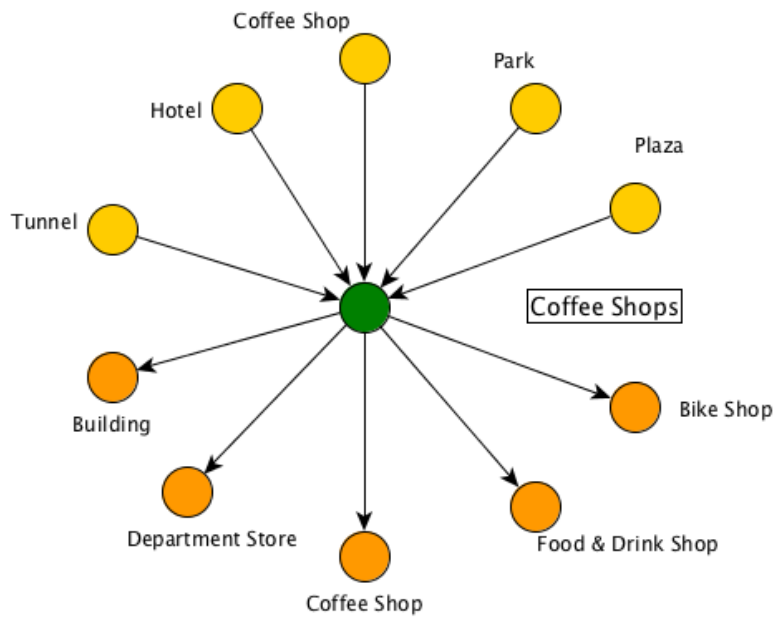


Figure 6.5. Subgraph of places visited by New York tourists before and after other Coffee Shops

To better understand what differentiates consumers of Starbucks and other coffee

shops, we observe what are the places that consumers in other coffee shops tend to go before and after. Figures 6.4 and 6.5 show the locations visited before and after the visit of consumers in other coffee shops in New York for residents and tourists, respectively.

Studying Figure 6.4 we can understand what the residents of New York who visit other coffee shops do. The behavior between users who visit Starbucks and other coffee shops is very similar. Places like School and Clothing Store are common in both types of consumers. However, there is a higher frequency of work (Office) to Coffee Shops. Looking at the times of highest peaks of this type of consumer, we can see that after work (around 20 hours) there is a higher popularity among consumers. After leaving the coffee shops, visits to places such as Train Station and Movie Theater are common, as in Starbucks. However, there is a pattern of going plazas only observed among those consumers.

Among tourists, we can see that, as in Starbucks, it is common to visit Parks and Plazas before going to coffee shops. But we can also see that it is also common to go to other coffee shops after check-ins in Hotels. When leaving coffee shops, tourists go to places of the class historic buildings and places to shop, such as Department Store and Bike Shop. Compared to the Starbucks case, we can see that tourists customers of other coffee shops visit more locations to eat, while Starbucks' customers prefer to shop. Analyze the profile of each of them allows us to visualize where companies can identify their competitive advantages and how they can use other places to boost their sales.

In addition to the analysis performed above, we can go a little deeper and understand what are the different profiles of consumers who visit Starbucks and other coffee shops. In this analysis we extract 3 profiles of each of these groups analyzed in the city of New York, where we observe the highest frequency of visits. The same process was shown in Section 5.3.

Table 6.1 shows the profile of Starbucks' consumers who reside in New York. We find profiles related to Sports, Arts and Movies. We note that in all profiles there is the presence of categories related to local food, such as Pizza Place and American Restaurant. We also note the presence of places related to art, such Movie Theater and Performing Arts Venue, where there are performances of dance and theater. This information shows that customers of Starbucks and residents of New York enjoy the local food and also events related to art. This information can serve as an input to marketing and creation of new products campaigns.

Table 6.2 shows profiles of customers of Starbucks who were visiting New York. We found profiles related to Business, Shopping and Arts. As for the residents' case, we observe a preference of sites related to art among a group of tourists who go to

Profile	Subcategories of most represented places to each group
Sports	Gym / Fitness Center, Park, Bar, Home (private)
Arts	Pizza Place, Bar, American Restaurant, Performing Arts Venue
Movies	Pizza Place, Bar, Coffee Shop, Movie Theater

Table 6.1. Profiles of customers of Starbucks who lives in New York

Starbucks. Among the tourists, we also have two groups related to work, a very common profile in the city of New York. We observe that Home category appeared in tourist's profile. Analyzing what places are attributed to this category in New York, we find places such as Manhattan, Wall Street and Queens. Those places are typical places to tourists check-in, specially when arriving in the city, helping to explain that result.

Profile	Subcategories of most represented places to each group
Business & Shopping	Office, Clothing Store, Bar, Home (private)
Arts	Food & Drink Shop, Bar, Coffee Shop, Performing Arts Venue
Business	Park, Bar, Coffee Shop, Office

Table 6.2. Profiles of customers of Starbucks who visits New York

6.2 Where should I go?

Using the urban mobility graph of tourists in New York, Rio de Janeiro, London and Tokyo we can see which are the most relevant places, which are the best to disseminate information and at what time (time) they take on these characteristics.

From this graph we can provide information for urban planning, consumer behavior research and also to offer suggestions of places so that other tourists can visit. Looking more temporal information we have through this graph, we can suggest that the best places based on time, based on previous experiences of other users.

To further explore the information on relevant sites and when these places are most visited, we created an application called DayTrip. This application receives a number of places that users want to visit and generate recommendations that can be followed in one day (24 hours). The relevance of seats is based on the weight of the graph edges. This means that we consider the couple of places that were most visited by users.

To build this application, first we order the edges with greater weight in the graph, which enabled display pairs of places (transitions) more frequented by users. To filter the places that would be better for tourists we exclude the places with the

classifications of the types: *City, Fast Food, Food, Home, Professional, School, Services, Shopping, Transport e Travel*. Thus, the sites considered are more related sights.

To illustrate the operation of the application, we choose 5 people in each city, based on visits by tourists. We assume that tourists have only one full day in the city, just to illustrate the application potential. Table 6.3 can view the simulation result of this application. We have a concentration of visits during the afternoon and evening. We can see the Times Square and Yankee Stadium (Figure 6.6²), places well known by tourists and indicated by places recommendation sites like TripAdvisor³. We can also see other places to drink, like Lincoln Park Tavern and Brooklyn Brewery, and Chinese food bistro SoHo. These places appear in the recommendation because they fall under the category Nightlife Spots and are quite frequented by tourists.

Venue[time]
Times Square[13]
SoHo Bistrot[16]
Yankee Stadium[18]
Lincoln Park Tavern[22]
Brooklyn Brewery[23]

Table 6.3. Recommended places to go based on New York tourists⁷



Figure 6.6. Yankee Stadium, New York

Table 6.4 can visualize the places in Rio de Janeiro, with more daytime programming, according to the script provided by the application. Recommended to go in

²<http://newyork.yankees.mlb.com/nyy/ballpark/information/index.jsp>

³<https://www.tripadvisor.com>

the morning, we have the Conselho Espirita do Estado do Rio de Janeiro - CEERJ, a religious center that carries out activities such as conferences and meetings for his followers. Thinking about the places frequented by tourists seeking religious tourism kept the category *religion* in the recommendation. Other places well frequented by tourists and that are also popular are the Jardim Botânico do Rio de Janeiro[10] and Jardim Suspenso do Valongo[18].

The Jardim Botânico is quite extensive, so it is recommended to arrive in the morning and spend the whole day there. The Jardim Suspenso do Valongo (Figure 6.7⁴) is open to visitors daily and is home to archaeological finds. Although it is a very interesting place, it is not popular among most visited places by tourists in sites or tourist books, which shows that the application can point to trends and what ceases to be sought more quickly than traditional methods. For activities of nightlife type, we have the Vienna Express and Bar Oswaldo night.

Venue[time]
Conselho Espirita do Estado do Rio de Janeiro - CEERJ[9]
Jardim Botânico do Rio de Janeiro[10]
Jardim Suspenso do Valongo[18]
Viena Express[19]
Bar do Oswaldo[20]

Table 6.4. Recommended places to go based on Rio de Janeiro tourists'



Figure 6.7. Jardim Suspenso do Valongo, Rio de Janeiro

⁴www.panoramio.com

Observing the recommendations for London, we have some very popular places among tourists, such as the Victoria and Albert Museum (Figure 6.8⁵), Buckingham Palace and The Courtauld Gallery. Among the recommendations, we can also see The Big Bang London & South East, a fair of science and engineering. This event took place during the collection period and may illustrate the ability to include events in real time using the approach of mobility temporal graph. In addition to the above places, we also have the Golden Dragon in *nightlife* category, representing the Oriental cuisine in London.

Venue[time]	
Victoria and Albert Museum (V	A)[10]
Buckingham Palace[11]	
The Courtauld Gallery[13]	
The Big Bang London & South East[19]	
Golden Dragon[20]	

Table 6.5. Recommended places to go based on London tourists⁷



Figure 6.8. Victoria and Albert Museum, London

In Tokyo we have a great diversity on the recommendation of the places. The Kanda Myojin Shrine (神田明神[2], Figure 6.9⁶) is a sanctuary and receives many tourists during the night. Another recommendation is the Taito Game Station, a major center games. Tokyo is known for squandering technology and have several sites for those who want to buy technology products but also to play. For the afternoon, we have the recommendation of the Jingu Stadium, a large stadium in Tokyo, quite busy at this time⁷. At night, we have たん清[18] and 吉野家 浅草駅前店[23], both directed to nightlife and Japanese cuisine.

⁵www.visitlondon.com

⁶<http://www.gotokyo.org>

⁷<http://www.jingu-stadium.com/english/schedule.html>

Venue[time]
神田明神[2]
GAME TAITO STATION[11]
明治神宮野球場 (JINGU STADIUM)[13]
たん清[18]
吉野家 浅草駅前店[23]

Table 6.6. Recommended places to go based on Tokyo tourists'



Figure 6.9. Kanda Myojin Shrine, Tokyo

The recommendation of places using mobility graphs bring a different perspective compared the list of most popular places in each city. Through mobility temporal graph, we can see the places importance and also in the temporal aspect.

6.3 Discussion

Based on the findings presented, we can see that the understanding of how tourists and residents behave in cities open many opportunities in different areas. To observe the temporal aspect helps us to understand the habits and, combined with local which precede and succeed visits enrich the analysis of how tourists interact with local as well as we can recommend sites that reflect the most relevant aspects of the city.

Under the government's perspective of cities, we can see the strengths of tourism within them and work for them has more focus and tourism to be better exploited. Examine areas and times that are more expressive, and the events that happen, offers a new perspective on tourism.

For companies, this information is relevant to businesses that want to better understand how their consumers tourists behave and how to differentiate themselves from the competition. There are opportunities from targeting marketing campaigns to the creation of new services / products specifically for tourists. The study on Starbucks, for example, shows us that to understand how consumers interact with the franchise can be a competitive advantage in sales.

With urban mobility graph used in this work, in addition to the aforementioned advantages, we can recommend places to the end user, considering the relevance of it to others and also the temporal aspect. The recommendation focuses on offering suggestions for places according to their spatial and temporal popularity elected by other city tourists. We believe this combination offers an interesting perspective for the user who wants to experience the city from the collective intelligence.

Chapter 7

Conclusion

The use of social data to conduct behavioral studies has great potential, as demonstrated in several studies. Its use covers various areas, such as tourism. From the study of tourists and residents we can see that there are behavioral differences between these two types of users and that cities have a very important role in this behavior. We found that the spatio-temporal aspects are fundamental for each class of users and that the cultural aspect is a factor of great influence. The information obtained from this study are an important input for the planning of cities, allowing the responsables for tourism promotion think new strategies to foster this economic activity and prepare the city in case of events and changes in the behavior of tourists. In addition to the user for the government to better plan cities for tourism, companies also benefit from this information with the possibility of creating new touristic products. One can also create more personalized recommendations systems, encouraging visits in places that have a profile more similar with the user.

In this research, we conducted the analysis using data of Foursquare check-ins. We study the movement of tourists through the city, considering various metrics, such as analysis of displacement and radius of gyration. Still focusing on mobility, we explored closeness centrality and betweenness on a spatial-temporal graph model, able to provide us with relevant information on dissemination of information and places that are more likely to have touristic activities. In addition, we analyzed different profiles found in cities, comparing consumers of a franchise with consumers from other business in the same segment.

In our work, we consider spatio-temporal aspects of tourists and residents behavior. The spatial patterns are related to available places in the city. It is important to analyze this dimension since, for example, the number of check-ins at a particular location can vary depending on the popularity and the category of it. The temporal

patterns are related to events occurring at certain time intervals. This is another dimension of utmost importance, since the behavior of users may vary, for example, with different shifts of the day. Consider these dimensions is critical to understand user behavior and the dynamics of the city in which he finds himself.

From the analyzes performed in this study, evaluated in four cities, London, Tokyo, New York and Rio de Janeiro, we also find that cultural aspects are extremely relevant in understanding human behavior. In the context of tourism, it is like that the behavior of tourists are guided by characteristics of the city.

Conduct research using social networking data allows us to capture what is happening in the world in near real time. The use of this data is proving to be increasingly powerful for the study of urban behavior [Silva et al., 2013a; Zheng et al., 2014], providing advantages, for example faster responses and cheaper cost, over other traditional methods for this purpose, such as surveys and interviews. Although it has many advantages, data from social networks may have limitations. One is the amount of data that can be collected from those services. For example, the Twitter API have a limitation of 1 % of the total volume of data produced, this means that we can not have all the data we want for a given application. In addition, less than 25% of Foursquare users push their check-ins to Twitter [Long et al., 2013]. Another limitation is the possible bias towards users who have smartphones with Internet access and despite that users that are using these apps. This means that what is identified with the use of these data might not represent the entire population. Users with smartphones and Internet access might represent more privileged people, a factor that could bring an income bias.

We believe that our work enables opportunities to perform new studies in the same area and also in other domains. When considering two layers (dimensions), time and space, we realized that we can introduce others such as weather, traffic and feelings (from textual analysis of the tips, for example). Multiple layers has a great potential not only for the study of tourists and residents but for the study of behavior as a whole.

Another area that we can go deeper is the analysis of consumer and business behavior. In addition to the characteristics of consumers, as presented in our work, we can also analyze the variations within the different areas of the cities and the competitive advantages among competition. Other information such as average income of residents of a region and demographics can be rich for this study.

Besides these research opportunities, another interesting possibility is the study of big events. Many tourists travel to other regions motivated by participation in big special events such as Carnival and music festivals, such as Rock in Rio. Analyze the dynamics of the city before, during and after the events can be very useful for urban planning and the business organization, such as hotel chain.

Bibliography

- Baraglia, R., Muntean, C. I., Nardini, F. M., and Silvestri, F. (2013). Learnnext: learning to predict tourists movements. *CIKM '13*.
- Basu Roy, S., Das, G., Amer-Yahia, S., and Yu, C. (2011). Interactive itinerary planning. In *Proceedings of the 2011 IEEE 27th International Conference on Data Engineering, ICDE '11*, pages 15--26, Washington, DC, USA. IEEE Computer Society.
- Bilogrevic, I., Huguenin, K., Mihaila, S., Shokri, R., and Hubaux, J.-P. (2015). Predicting users' motivations behind location check-ins and utility implications of privacy protection mechanisms. In *22nd Network and Distributed System Security Symposium (NDSS' 15)*, number EPFL-CONF-202202.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993--1022. ISSN 1532-4435.
- Cheng, Z., Caverlee, J., Lee, K., and Sui, D. Z. (2011). Exploring millions of footprints in location sharing services. *Proc. of ICWSM*, pages 81--88.
- Cho, E., Myers, S. A., and Leskovec, J. (2011). Friendship and mobility: User movement in location-based social networks. *Proc. of KDD*, pages 1082--1090.
- Choudhury, M. D., Feldman, M., and Amer-Yahia, S. (2010). Automatic construction of travel itineraries using social breadcrumbs. *Hipertext'10*.
- Choujaa, D. and Dulay, N. (2009). Activity Recognition from Mobile Phone Data: State of the Art, Prospects and Open Problems. *Imperial College London*.
- Colombo, G., Chorley, M., Williams, M., Allen, S., and Whitaker, R. (2012). You are where you eat: Foursquare check-ins as indicators of human mobility and behaviour. *2012 IEEE International Conference on Pervasive Computing and Communications Workshops*, pages 217--222.

- Diplaris, S., Papadopoulos, S., Kompatsiaris, I., Goker, A., Macfarlane, A., Spangenberg, J., Hacid, H., Maknavicius, L., and Klusch, M. (2012). Socialsensor: Sensing user generated input for improved media discovery and experience. In *Proceedings of the 21st International Conference on World Wide Web, WWW '12 Companion*, pages 243--246, New York, NY, USA. ACM.
- Fennell, D. A. (1996). A tourist space-time budget in the shetland islands. *Annals of Tourism Research*, 23(4):811--829.
- Freeman, L. C. (1979). Centrality in networks: I. conceptual clarification. *PLoS ONE*, (6):215--239.
- González, M. C., Hidalgo, C. A., and Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, 453:779--782.
- Hallot, P., Stewart, K., and Billen, R. (2015). Who are my Visitors and Where do They Come From? An Analysis Based on Foursquare Check-ins and Place-based Semantics. In Biljecki, F. and Turre, V., editors, *Eurographics Workshop on Urban Data Modelling and Visualisation*. The Eurographics Association. ISSN 2307-8251.
- Hsieh, H.-P., Li, C.-T., and Lin, S.-D. (2012). Triprec: Recommending trip routes from large scale check-in data. In *Proceedings of the 21st International Conference on World Wide Web, WWW '12 Companion*, pages 529--530, New York, NY, USA. ACM.
- Karamshuk, D., Boldrini, C., Conti, M., and Passarella, A. (2011). Human mobility models for opportunistic networks. *IEEE Communications Magazine*, 49:157--165.
- Karamshuk, D., Noulas, A., Scellato, S., Nicosia, V., and Mascolo, C. (2013). Geospotting: Mining online location-based services for optimal retail store placement. *CoRR*, abs/1306.1704.
- Kung, K. S., Greco, K., Sobolevsky, S., and Ratti, C. (2014). Exploring universal patterns in human home-work commuting from mobile phone data. *PLoS ONE*, 9(6).
- Lew, A. and McKercher, B. (2006). Modeling tourist movements: A local destination analysis. *Annals of tourism research*, 33(2):403--423.
- Lindqvist, J., Cranshaw, J., Wiese, J., Hong, J., and Zimmerman, J. (2011). I'm the Mayor of My House: Examining Why People Use foursquare - a Social-Driven Location Sharing Application. *Proc. of CHI*, pages 2409--2418.

- Long, X., Jin, L., and Joshi, J. (2013). Towards Understanding Traveler Behavior in Location-Based Social Networks. *Globecom*.
- Lv, M., Chen, L., and Chen, G. (2013). Mining user similarity based on routine activities. *Information Sciences: an International Journal*, 236:17–32.
- Majid, A., Chen, L., Chen, G., Mirza, H. T., and Hussain, I. (2012). Gothere: Travel suggestions using geotagged photos. In *Proceedings of the 21st International Conference on World Wide Web, WWW '12 Companion*, pages 577–578, New York, NY, USA. ACM.
- Morais, A. M. and Andrade, N. (2014). The relevance of annotations shared by tourists and residents on a geo-social network during a large-scale touristic event: the case of são joão. *Proc. of COOP*.
- Paldino, S., Bojic, I., Sobolevsky, S., Ratti, C., and González, M. C. (2015). Urban magnetism through the lens of geo-tagged photography. *EPJ Data Science*, 4(1):1–17.
- Pianese, F., Kawsar, F., and Ishizuka, H. (2013). Discovering and predicting user routines by differential analysis of social network traces. *Proc. of WoWMoM*, pages 1–9.
- Preo, D. and Cohn, T. (2013). Mining user behaviours: A study of check-in patterns in location based social networks. *Proc. of WebSci*.
- Roick, O. and Heuser, S. (2013). Location based social networks—definition, current state of the art and research agenda. *Transactions in GIS*, 17(5):763–784.
- Sharpley, R. and J, T. D. (2002). *Tourism and development: concepts and issues*. Channel View Publications. ISBN 9781873150351.
- Shi, Y., Serdyukov, P., Hanjalic, A., and Larson, M. A. (2011). Personalized landmark recommendation based on geotags from photo sharing sites. In *ICWSM '11: the 5th AAAI Conference on Weblogs and Social Media*, pages 622–625, Barcelona, Spain. AAAI, AAAI.
- Silva, T., Vaz De Melo, P., Almeida, J., and Loureiro, A. (2014a). Large-scale study of city dynamics and urban social behavior using participatory sensing. *Wireless Communications, IEEE*, 21(1):42–51.

- Silva, T. H., de Melo, P. O. S. V., Almeida, J. M., Vianna, A. C., Salles, J., and Loureiro, A. A. F. (2014b). Definição, modelagem e aplicações de camadas de sensoriamento participativo. *32º Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*.
- Silva, T. H., de Melo, P. O. S. V., M, J., and Loureiro, A. A. F. (2013a). Challenges and opportunities on the large scale study of city dynamics using participatory sensing. In *IEEE symposium on Computers and Communications*, page 7. IEEE.
- Silva, T. H., Vaz de Melo, P. O. S., Almeida, J. M., Salles, J., and Loureiro, A. A. F. (2013b). A picture of Instagram is worth more than a thousand words: Workload characterization and application. In *Proc. of the IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS'13)*, pages 123--132, Cambridge, MA, USA.
- Silva, T. H., Vaz de Melo, P. O. S., Almeida, J. M., Salles, J., and Loureiro, A. A. F. (2014c). Revealing the city that we cannot see. *ACM Trans. Internet Technol.*, 14(4):26:1--26:23. ISSN 1533-5399.
- Staab, S., Werthner, H., Ricci, F., Zipf, A., Gretzel, U., Fesenmaier, D. R., Paris, C., and Knoblock, C. (2002). Intelligent Systems for Tourism. *IEEE Intelligent Systems*, 17(6):53--64.
- Yerva, S. R., Grosan, F., Tandrau, A., and Aberer, K. (2013). Tripeneer: User-based travel plan recommendation application. In *ICWSM*. The AAAI Press.
- Yoon, H., Zheng, Y., Xie, X., and Woo, W. (2010). Smart Itinerary Recommendation based on User-Generated GPS Trajectories. *Ubiqu. Intel. and Computing*, pages 19--34.
- Zheng, Y. (2011). *Computing with Spatial Trajectories*, chapter Location-Based Social Networks: Users, pages 243--276. Springer New York, New York, NY.
- Zheng, Y. (2012). Tutorial on location-based social networks. In *WWW 2012*. ACM.
- Zheng, Y. (2014). Para onde devo viajar: Recomendação de cidades baseada em comunidades de usuários. In *BraSNAM 2014*. USP.
- Zheng, Y., Capra, L., Wolfson, O., and Yang, H. (2014). Urban computing: Concepts, methodologies, and applications. *ACM Transaction on Intelligent Systems and Technology*.

Zheng, Y., Zhang, L., Xie, X., and Ma, W.-Y. (2009). Mining interesting locations and travel sequences from GPS trajectories. *Proc. of WWW*.