Evaluating the impact of a cultural event on user mobility and preferences by collecting user social media data

Felipe Araújo[1], Fábio Araújo[1], Marcos Seruffo[1], Denis Rosário[1] and Eduardo Cerqueira[1]

[1]Federal University of Pará (UFPA), Belém – PA – Brazil

{felipe.rocha, fabio.araujo}@itec.ufpa.br, {seruffo, denis, cerqueira}@ufpa.br

Abstract. *The popularization of smartphones is a favorable factor to the increase of the use of location sharing services (e.g Foursquare, Instagram and Twitter). It allows the mobile phone user to not only be a consumer of social media data, but also an active producer of those data. Therefore, several patterns of mobility and behavior can be extracted from them. This work uses Participatory Sensor Networks (PSNs) to evaluate the impact of "Círio de Nazaré" (a cultural event that occurs in Amazon region) on user mobility and preferences by analyzing social media data collected before, during and after this cultural event.*

1. Introduction

In the last years, there has been a dramatic change in how we use mobile devices. Since they have came with various embedded sensors (e.g., GPS and accelerometer), the use of location sharing services has increased (e.g., Foursquare, Instagram and Twitter). Mobile devices have been used not only for simple calls but also for sensing or urban planning applications. In this context, an end user has became not only a consumer, but an active producer of information using sensors attached to their devices, enabling a new and powerful source of data [Silva et al. 2014a]. Systems that enable sensed data in this way are named as Participatory Sensing Systems (PSSs) [Silva et al. 2013].

Derived from PSSs, there is the concept of Participatory Sensor Networks (PSNs), where the nodes are autonomous mobile entities and the sensing activity depends on whether they want to participate in the sensing process [Silva et al. 2013]. A PSN is composed of one or more sensing layers, where each layer represents data, with the corresponding attributes, from a given source of data. For instance, **check-ins** layer obtained from Foursquare provides location and category of a place (e.g., residence, food and nightlife spot). In addition, PSN offers a new way of studying human behavior in near real time and planetary scale, since they reflect people routines and preferences, helping governments and enterprises to predict and plan proactive actions to improve the quality of life in smart digital cities [Xavier et al. 2012, Soper 2012, González et al. 2008].

Some large scale events may change the dynamic of cities and, for that reason, understanding these changes is an essential aspect for urban planning applications. Even though many of these large scale events are scheduled and planned in advance and are expected to cause collective changes in the workload, it remains common to notice the congestion of the carrier's resources during them [Xavier et al. 2012]. Therefore, the use of social media data to extract mobility patterns may support the development of better management of urban/network planning. For instance, Círio de Nazaré is the largest Catholic procession in Brazil (Belém, Pará), and it honors Nossa Senhora de Nazaré (Our

Lady of Nazareth). Every second Sunday of October, a wooden image of Nazareth proceeds from Sé Cathedral to Sanctuary Square (about 3.6 km). This event envolves almost every citizen of the state of Pará – Brazil and a large number of people coming from all world, changing the people's movements and the kind of places they go. Therefore, in this work we apply PSN to evaluate the impact of a specific event in the dynamic of a city, enabling a better urban planning.

Given that and considering the importance of mobility to better understand city dynamics, we focus on investigation of the following questions: How a cultural event can impact on mobility patterns of individuals? Where and what are the kind of places people usually visit before, during and after a specific event? In this work, we use the data collected from location social services, such as Foursquare, Instagram or Twitter to study human behavior during the period of an event (in this case Círio de Nazaré) by extracting user movements. For those proposes described above, we divided the city in several zones according to similar characteristics (e.g, proximity, demographic density and neighborhood economy) and the places in some main categories (e.g, Residences, Arts & Entertainment and College & University). We collected and analyzed these social media data in three separate time windows: **BEFORE**, **DURING** and **AFTER** the event. The transition between zones and categories illustrate users' mobility while static graphics of categories show the kind of places these end users prefer.

This paper is organized as follows. In Section II, we present the related works. In Section III, we describe our methodology for extracting user check-ins and evaluate the impact of a specific cultural event on people's mobility and preferences of places. Finally, Section IV presents the concluding remarks and future works.

2. Related Work

The existing literature shows the usage of location-based social media data to extract common user routines or design new applications.[Silva et al. 2012] proposed a technique called City Image, which exploits urban transition graphs in order to map user movements. This technique shows the transition from a given category and place to another one, providing an easy-to-interpret visualization of the users' routines. Futhermore, [Frias-Martinez et al. 2012] evaluated the use of geolocated tweets as a complementary source of information for urban planning applications. They focused in two urban planning areas: first, based on tweeting patterns, a technique to automatically determine land uses in specific urban areas; and second, a technique to automatically identify urban points of interest as places with high activity of tweets.

[Noulas et al. 2011] proposed a method to model human activity and geographical areas by place categories collected from Foursquare check-ins, allowing the identification of communities with similar categories of places and the comparison of urban neighborhoods within and across cities. The authors used Foursquare data to extract crowd activity pattern in London and New York. [Cranshaw et al. 2012] proposed a clustering model and a research methodology for studying the dynamics, structure, and character of a city on a large scale, presenting new boundaries for neighborhoods.

PSNs can be used to extract information about cultural aspects. In this direction, [Hochman and Schwartz 2012] investigated color preferences in pictures shared in Instagram and found differences from different countries. The study of cultural aspects of a

city is valuable in many fields, enabling various applications. For example, since culture is related to the economy, the identification of similarities between places and areas might be valuable for companies which want to assess the compatibility of preferences across different markets [Silva et al. 2014b]. [Machado et al. 2015] evaluated the urban scenario considering the mobility of users and the weather. Their results showed a phenomenon of behavior transition within a specific temperature range for a group of cities studied.

[Xavier et al. 2012] proposed a methodology for characterizing and analyzing the workload dynamics of a mobile phone network in large scale events. They used data from anonymized mobile phone users to improve the understanding on how users move within an urban area towards the location of a large scale event and how they disperse afterward, helping the management of mobile phone networks. Similar to that, [Silveira et al. 2015] proposed a new model to predict human mobility, called MobDatU, which was designed to use data from mobile calls and data from georeferenced applications.

Although most of the related papers above used data collected from location-based social media (e.g. Foursquare, Twitter and Instagram) for mobility pattern or identification of communities with similar characteristics (similar categories of places), they did not relate them to the impact of a specific event occurring in that area or city. Nevertheless, [Xavier et al. 2012] illustrated the impact of a match soccer occurring in Rio de Janeiro – Brazil, but the authors did not collected data from social medias, they used a mobile phone calls dataset from one of the major cellphone carries operating in Brazil.

3. Evaluating the impact of a cultural event on users' mobility and preferences via PSN

3.1. Dataset and Social Media

Foursquare and Twitter are most popular social media in the world. The first one has more than 55 million users worldwide, who have left more more than 70 million tips and checked in over 7 billion times [Foursquare 2016]. It also created new ways for online interactions based on the physical location of their users, where users can do their "check-ins" with a location selected from a list with nearby venues. These check-ins can be used for describe social, economy or cultural aspects of a city [Cranshaw et al. 2012].

Foursquare check-ins are not public and its API is limited, and thus we gathered data are from Twitter, which contains Foursquare and Instagram data. Each public data consists of a set of geographic coordinates (latitude, longitude), user identifier, tweet identifier, post time, and tweet text. In addition, similarly with existing works, we separated them in ten categories (Arts & Entertainment; College & University; Professional & Other Places; Residences; Outdoors; Shop & Services; Nightlife; Events; Travel & Transport; and Food). These categories were made by grouping subcategories, as shown in Table I, extracted from the tweet text URL in ten different broad categories. For instance, places categorized as Science Museum, Art Gallery or Movie Theater were grouped as Arts & Entertainment, a new broad category.

The mobility in large events is different from other time period. In this way, to better understand changes caused by a given event, in our case Círio de Nazaré, on citizens and non citizens of Belém in terms of mobility and most visited places, check-ins were also separated by zones, as shown in Table II. We created groups of neighborhoods

Table 1. Category and its subcategories

Main category	Subcategories	Acronyms
Arts & Entertainment	Movie Theater and Museum	A & E
College & University	General College & University and Student Center	C & U
Professional & Other Places	Building, Library and Office	P & O
Residences	Home (private) and Residential Building	RE
Outdoors & Recreation	Playground and Recreation Center	O & R
Shop & Services	Business service and Gas station	S & S
Nightlife	Nightclub and Lounge	NL
Events	Music Festival and Parade	EV
Travel & Transport	Bus Station and Hotel	T & T
Food	Brazilian restaurant and Bakery	FO

with similar characteristics, such as, proximity, demographic density and neighborhood economy. For example, Reduto and Nazaré (neighborhoods of Belém) are parts of the same zone, considering the similarity they have (e.g. proximity and demographic density). However, compared to Guamá, a peripheral neighborhood, they are from different zones due to the distance and particularity of each one. Also, in order to increase the data for mobility between zones, we used all geolocated data from twitter, without excluding those which were not from Foursquare. It includes Instagram posts that were shared on Twitter.

Table 2. Zone and its main neighborhoods

Zone	Main neighborhoods
Zone 0	Jurunas and Condor
Zone 1	Umarizal, Nazaré, Reduto and Batista Campos
Zone 2	Telégrafo e Sacramenta
Zone 3	Guamá, Terra Firme and Canudos
Zone 4	Marco, Souza and Pedreira
Zone 5	Val-de-Cães, Marambaia and Mangueirão
Zone 6	Pratinha, Bengui, Tapanã and Tenoné
Zone 7	Curió-Utinga, Castanheira and Aurá

A formal definition of how we extracted human mobility and most visited categorized places from twitter is divided in 4 steps:

- First, we collected from StreamingAPI (an API that offers samples of the public data flowing) a set of users U during the interval d days, $d \in [0 \dots D]$;
- Second, for each user $u_i \in U$, we gathered all geographical tweets from its timeline, using the UserTimelineAPI (an API that returns a collection of most recent tweets posted by the user indicated by the screen-name or user-id parameters);
- Third, tweets were separated into three time windows. The first time window w_1 are tweets posted before the event. The second time window w_2 are tweets posted during the event. Finally, the third time window w_3 are tweets posted after the event.

- Forth and last step, in order to extract human mobility, only tweets of a respective user u_i within interval of post less than 24 hours and more than 5 minutes were used to extract mobility and most visited places. In this way, we have 70952 check-ins collected in $09/2014$ and $09/2015$, before the event; 10763 check-ins collected during the event period in $10/2014$ and $10/2015$; and 66546 check-ins collected in $11/2014$ and $11/2015$ after event.

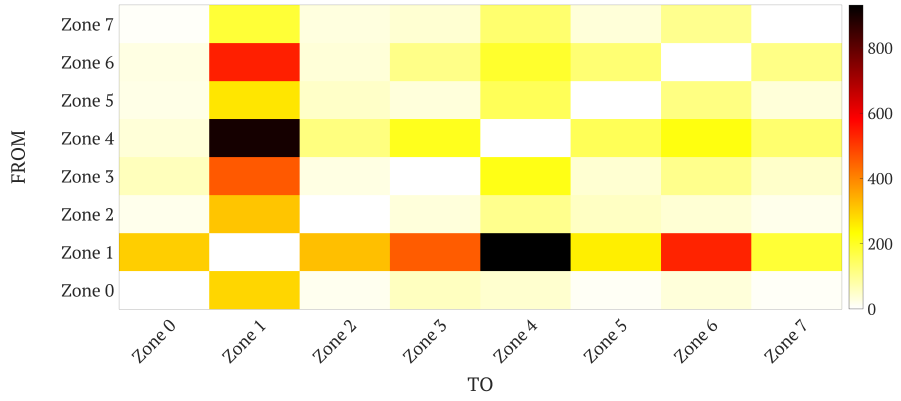
The impact of Círio was evaluated considering user preferences (places categories they visit) as an important factor for the mobility dynamics. For this reason, we investigated the transition between categories of places and zones. Through check-ins shared on social medias, temporal individuals' mobility patterns could be inferred. For that, we created two different matrices: Z and C, where each element of the matrices (Z_{ij} and C_{ij}) can be calculated by summing the number of users who moved from Zone/Category i to the Zone/Category j, as follows:

$$Z = \sum_{i=1}^{n} \sum_{j=1}^{n} z_{ij}, z_{ij} = 0, 1, 2, 3... \tag{1}$$

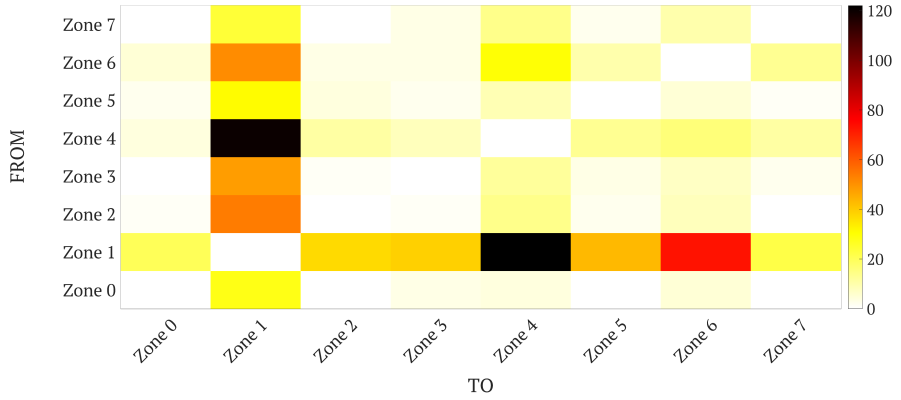
$$C = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}, c_{ij} = 0, 1, 2, 3... \tag{2}$$

3.2. Results and Discussions

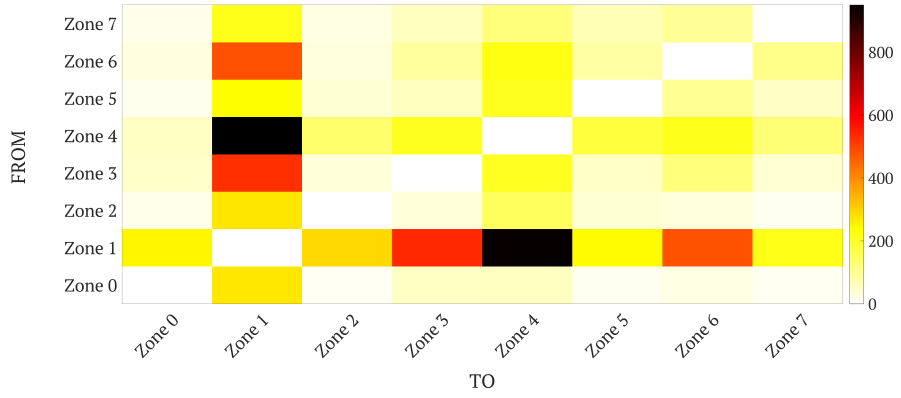
Figure 1 illustrates the matrix Z of approximately 5000 users in three different time intervals w_1, w_2 and w_3. First, it is important to observe that, in all periods, Zone 1 concentrates the highest number of transitions between all zones. It occurs due to this zone being the commercial and work-center of Belém, since it has a considered number of commercial, work and touristic places, with a medium number of residences. Furthermore, the event in question occurs in this specific zone, which makes the difference between three periods lower than comparing with other events. Also, in all the periods, Zone 4 has the highest number of users who moved to the Zone 1. The high number of residences in Marco (Neighborhood that is part of Zone 4) is the main favorable fact for this.



(a) Before (W_1)



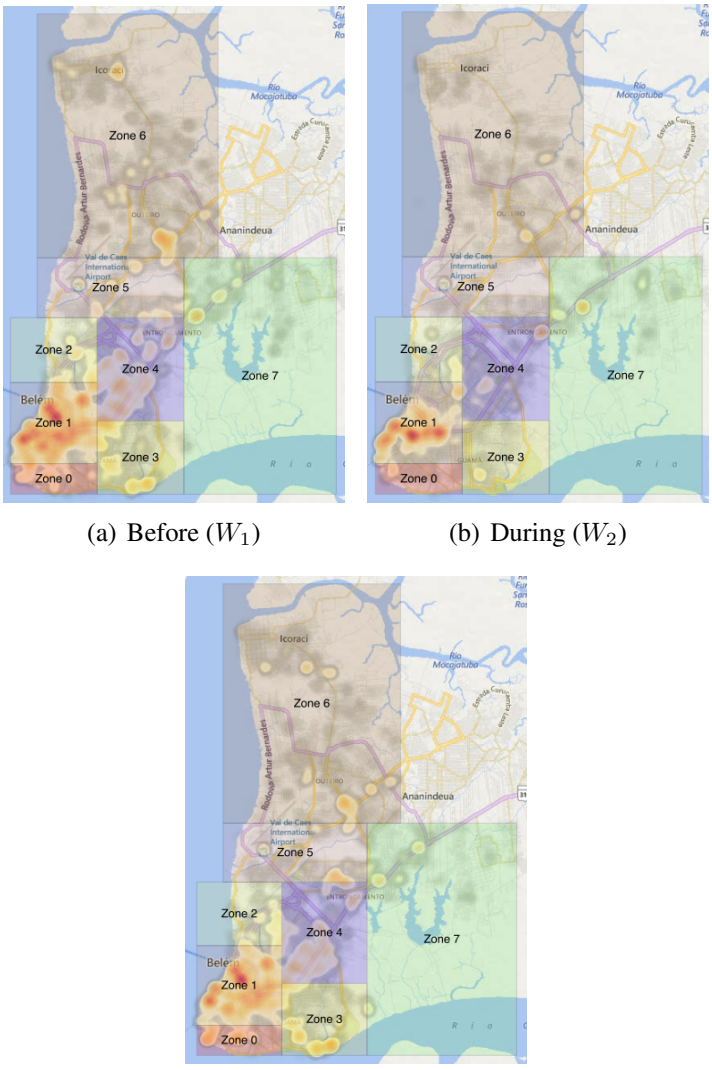
(b) During (W_2)



(c) After (W_3)

Figure 1. Users'transistions between zones.

An alternative form to visualize users' movement is shown in Figure 2, which represents a heat-map of Belem divided by zones. It illustrates a high concentration of people in the center of Belém, but it still has a high number of people in the other zones. Similar patterns can be observed in all three periods w_1, w_2 and w_3. They indicate that the analyzed event does not change the zone of interest. Using the same analysis to other events, the results may be different from that.

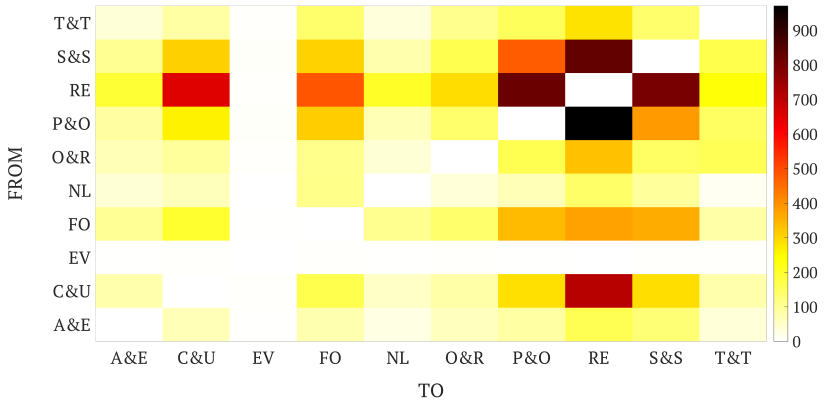


(a) Before (W_1) (b) During (W_2)

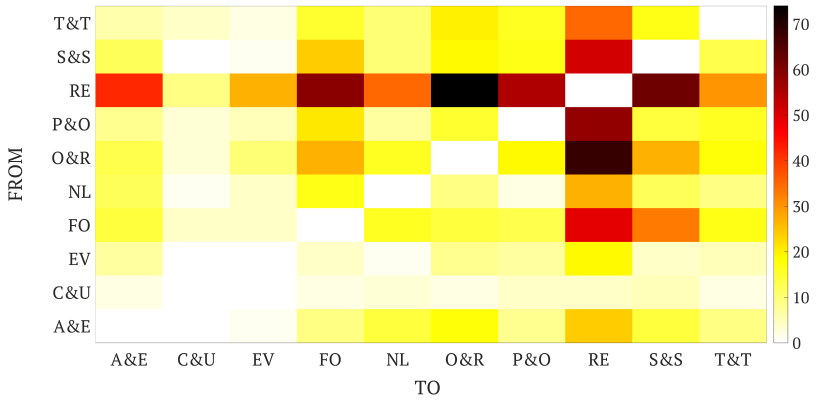
(c) After (W_3)

Figure 2. Heatmap of all users' check-ins.

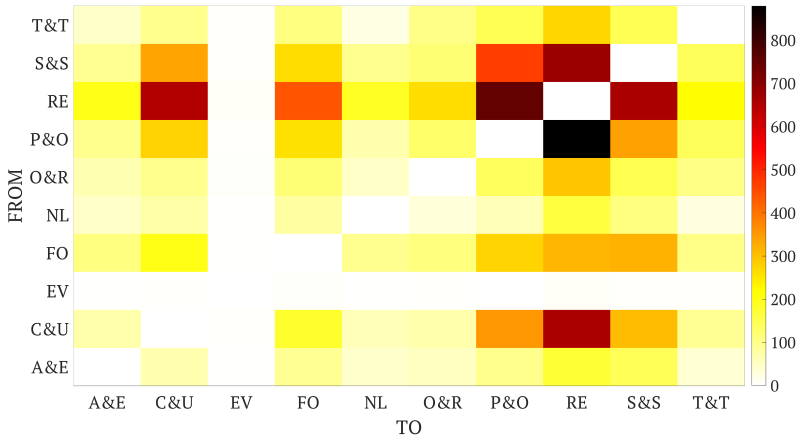
The Figure 3 illustrates the matrix C composed by Foursquare data in three different timestamps: before, during and after Círio de Nazaré. According to the results, we identified transition patterns, such as the high degree of similarity between C_{before} and C_{after}, showing that these users tend to visit the same kinds of places when there is not any large festival occurring in Belém. On the other hand, in the second weekend of October, the result illustrates a different transition matrix compared to the others. For example, during this weekend, the users tend to move from their residences to other kind of places, such as Outdoors & Recreation. In addition, according to C_{during} we can infer that they move back to their residences or go to friend's home since the figure illustrates intense transitions from all categories to residence category.



(a) Before (W_1)



(b) During (W_2)



(c) After (W_3)

Figure 3. Users' transitions between place categories.

Figure 4 illustrates the kind of places the users usually visit and share on the social medias. Based on the data collected from Foursquare, the percentual of individuals who share places of residence category is higher during period of our analyzed event compared to the data collected before and after it. This result can be explained since Círio is a religious festival that people usually go to their friends or relative's residences after the

procession. Moreover, Figure 4 shows that 2.5% of this data is composed by places from event category. It may be explained by the other events that occur in Belém in this period.

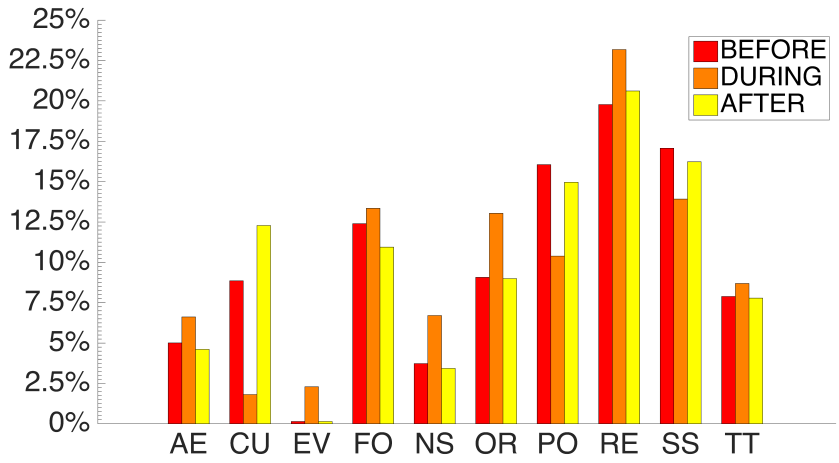
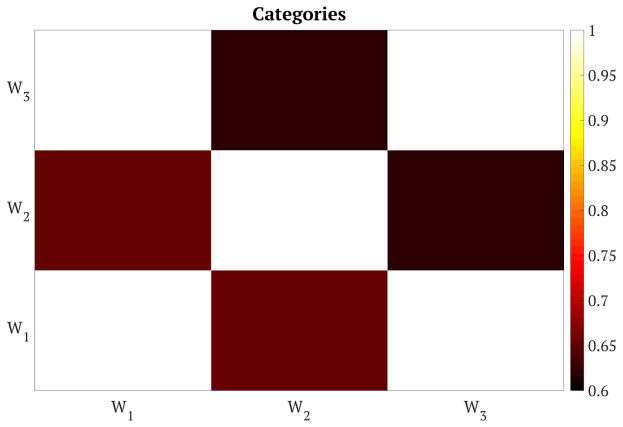
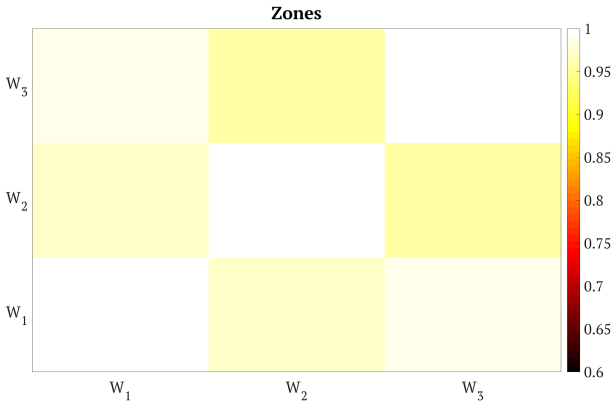


Figure 4. Categories of places Foursquare dataset is composed by.

In the interest of measuring the similarity degree of the transition matrices in the three periods (w_1, w_2 , and w_3), we calculated the correlation coefficients between them. As we can see, the Figure 5(a) illustrates a high correlation value ($r = 0.99$) of the matrix of categories in the period w_1 with the period w_3 and vice-versa. In other words, there is not a large difference between the transition matrices w_1 and w_3. On the other hand, comparing the matrices C_{before} and C_{after} with C_{during}, the Pearson's correlation values were lower than 0.66, showing a higher difference comparing the matrices. However, the correlation values of the transition matrices of zones were extremely high as we can see in the Figure 5(b) (almost 1), indicating a perfect positive linear relationship between them. It can be explained since the event has occurred in the Zone 1 which is the busiest zone of Belém, even when there are not events occurring in the city. A large concentration of people in a specific area may become the traffic or a network slow. Using the same methodology to evaluate impact of several events on the dynamic of a city, it is possible to optimize these resources.



(a) Matrix of Categories



(b) Matrix of Zones

Figure 5. Correlation Coefficients Matrices.

4. Conclusion and future steps

This paper presented a particular analysis to describe a city dynamic, gathering data from social media. The division in zones and categories are favorable factors to understand user movements, what kind of places they tend to visit and where users came from. The main contribution of this paper is to describe the city dynamic by analyzing check-ins from social media that could be applied to other cities and events of large impact on individuals' routine, in order to evaluate how different this routine can become. We observed that on the analyzed event (Círio de Nazaré), the number of movements involving Residence and Event categories is higher than normal. On the other hand, the results show that the number of transitions involving these two categories is significantly lower in periods before and after the event, resulting in a different mobility pattern and preferences of users, in terms of categorized places.

Data collected from location sharing services could be applied for several applications. For example, it can be used for improving the urban or network planning to support an specific event, due to a slow network or traffic a high concentration of people can cause. This event can be anyone of large-scale, and once known the dynamics of mobility of people during it, this information can be used for better planning of other future

similar events. In future direction, we intend to use a similar methodology for studying the behavior of citizens of other city, also relating big events such as: Rio 2016 Olympics and Presidential Election.

References

Cranshaw, J., Schwartz, R., Hong, J. I., and Sadeh, N. (2012). The livehoods project: Utilizing social media to understand the dynamics of a city. In *International AAAI Conference on Weblogs and Social Media*, page 58.

Foursquare (01/17/2016). Retrieved from: https://foursquare.com/about.

Frias-Martinez, V., Soto, V., Hohwald, H., and Frias-Martinez, E. (2012). Characterizing urban landscapes using geolocated tweets. In *Proc. International Conference on Privacy, Security, Risk and Trust (PASSAT),2012 and International Conference on Social Computing (SocialCom), 2012*, pages 239–248. IEEE.

González, M. C., Hidalgo, C. A., and Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, 453(7196):779–782.

Hochman, N. and Schwartz, R. (2012). Visualizing instagram: Tracing cultural visual rhythms. In *Proceedings of the Workshop on Social Media Visualization (SocMedVis) in conjunction with the Sixth International AAAI Conference on Weblogs and Social Media (ICWSM–12)*, pages 6–9.

Machado, K., Silva, T. H., Vaz de Melo, P. O., Cerqueira, E., and Loureiro, A. A. (2015). Urban mobility sensing analysis through a layered sensing approach. In *Proc. IEEE International Conference on Mobile Services (MS), 2015*, pages 306–312. IEEE.

Noulas, A., Scellato, S., Mascolo, C., and Pontil, M. (2011). Exploiting semantic annotations for clustering geographic areas and users in location-based social networks. *The social mobile web*, 11:02.

Silva, T. H., Vaz de Melo, P., Almeida, J., Viana, A., Salles, J., and Loureiro, A. (2014a). Participatory Sensor Networks as Sensing Layers. In *Proc. of the IEEE Conf. on Social Computing and Networking (SocialCom'14)*, Sydney, Australia.

Silva, T. H., Vaz De Melo, P. O., Almeida, J. M., Loureiro, A., et al. (2014b). Large-scale study of city dynamics and urban social behavior using participatory sensing. *IEEE Wireless Communications*, 21(1):42–51.

Silva, T. H., Vaz de Melo, P. O. S., Almeida, J. M., and Loureiro, A. A. F. (2012). Visualizing the Invisible Image of Cities. In *Proc. IEEE International Conference on Cyber, Physical and Social Computing*, pages 382–389, Besancon, France.

Silva, T. H., Vaz de Melo, P. O. S., Almeida, J. M., and Loureiro, A. A. F. (2013). Challenges and opportunities on the large scale study of city dynamics using participatory sensing. In *IEEE Int. Symp. on Computers and Communications (ISCC'13)*, pages 528–534, Split, Croatia.

Silveira, L. M., Almeida, J. M., Marques-Neto, H., and Ziviani, A. (2015). Mobdatu: A new model for human mobility prediction based on heterogeneous data. In *XXXIII Brazilian Symposium on Computer Networks and Distributed Systems (SBRC), 2015*, pages 217–227. IEEE.

Soper, D. (2012). Is human mobility tracking a good idea? *Communications of the ACM*, 55(4):35–37.

Xavier, F. H. Z., Silveira, L. M., Almeida, J. M. d., Ziviani, A., Malab, C. H. S., and Marques-Neto, H. T. (2012). Analyzing the workload dynamics of a mobile phone network in large scale events. In *Proc. of the first workshop on Urban networking*, pages 37–42. ACM.