A comparison of Foursquare and Instagram to the study of city dynamics and urban social behavior

Thiago H. Silva
Federal Univ. of Minas Gerais
Computer Science
Belo Horizonte, Brazil
thiagohs@dcc.ufmg.br

Pedro O. S. Vaz de Melo
Federal Univ. of Minas Gerais
Computer Science
Belo Horizonte, Brazil
olmo@dcc.ufmg.br

Jussara M. Almeida
Federal Univ. of Minas Gerais
Computer Science
Belo Horizonte, Brazil
jussara@dcc.ufmg.br

Juliana Salles
Microsoft Research
Redmond, WA, USA
jsalles@microsoft.com

Antonio A. F. Loureiro
Federal Univ. of Minas Gerais
Computer Science
Belo Horizonte, Brazil
loureiro@dcc.ufmg.br

ABSTRACT

Social media systems allow a user connected to the Internet to provide useful data about the context in which they are at any given moment, such as Instagram and Foursquare, which are called participatory sensing systems. Location sharing services are examples of participatory sensing systems. The sensed data is a check-in of a particular place that indicates, for instance, a restaurant in a specific location, and also a signal from a user expressing his/her preference. From a participatory sensing system we can derive a participatory sensor network. In this work we compare two different participatory sensor networks, one derived from Instagram, and another one derived from Foursquare. In Instagram, the sensed data is a picture of a specific place. On the other hand, in Foursquare the sensed data is the actual location associated with a specific category of place (e.g., restaurant). Using those social networks we can extract information in many ways. In this work we are interested in comparing two datasets of Foursquare and two datasets of Instagram. We analyze those datasets to investigate whether we can observe the same users' movement pattern, the popularity of regions in cities, the activities of users who use those social networks, and how users share their content along the time. In answering those questions, we want to better understand location-related information, which is an important aspect of the urban phenomena.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavioral Sciences; G.3 [**Mathematics of Computing**]: Probability and Statistics—*Statistical computing*; H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Measurement

UrbComp '13, August 11, 2013, Chicago, Illinois, USA

Keywords

Social media, participatory sensing, Foursquare, Instagram, city dynamics

1. INTRODUCTION

The world continues to be more and more social with social media generating a huge amount of data. Currently, there are tens of different social networking websites[1] such as Facebook, Twitter, Foursquare and Instagram, which can play an important function in urban computing since they have the potential to improve the urban environment, human life quality, and city operation systems.

Some of those social networks share location-related information, which is part of our daily lives. In many situations we would like to know the location of other people to decide our own activities. For instance, when we are organizing a night out with friends, choosing a dish at a new restaurant or looking for a suggestion for what place to visit next in a trip.

Created in 2009, Foursquare is a location-based social network that allows registered users to perform *check-ins* (post their location) at a venue by selecting the location from a list of venues the application locates nearby. Users can also post their check-ins on their accounts on Twitter, Facebook, or both. Users are encouraged to be as specific as possible with their check-ins by indicating and sharing the places they visit, their precise location or activity while at a venue. As a result, Foursquare tries to provide the most of a given place where a user is, supplying personalized recommendations and business deals. Foursquare awards points to users whenever they check into a place according to different rules, i.e., roughly speaking Foursquare can be seen as a game where users can have different status and can "carry" badges. As can be seen, the location of a place plays a fundamental role in Foursquare. As of January 2013, there have been more than 3 billion check-ins with Foursquare from over 30 million people worldwide.

Created in 2010, Instagram is an online photo-sharing and social networking service that allows users to take pictures, apply digital filters to them, and share them on a variety of social networking services, such as Facebook or Twitter. Users can upload pictures, attach their Instagram account to other social networking services, and follow other users' feeds. The user can also associate a location with each picture. Currently, Instagram users can create web

[1] `http://en.wikipedia.org/wiki/List_of_social_networking_websites`

profiles featuring recently shared pictures, biographical information, and other personal details. As of February 2013, Instagram announced that they had 100 million active users.

Social networks and social software have been driven by two aspects: connections between people who use them and the information they share, in particular location-related information [18]. In this work we are interested in comparing two datasets of Foursquare and two datasets of Instagram. We analyze those datasets to investigate whether we can observe the same users' movement pattern, the popularity of regions in cities, the activities of users who use those social networks, and how users share their content along the time. In answering those questions, we want to better understand location-related information, which is an important aspect of the urban phenomena.

The rest of this work is organized as follows. Section 2 describes the related work. Section 3 briefly describes social media as a source of sensing and its importance to urban computing. Section 4 compares the four datasets of the two social networks using location-related information as the main aspect of the analysis. In that section we answer the questions stated above. Finally, Section 5 presents the conclusion and future work.

2. RELATED WORK

The class of social media named Location-based social network is probably the most popular type of participatory sensing system, and they have been receiving a lot of attention recently [20, 21]. Here we discuss some studies that explore this popular type of participatory sensing systems to enable the better understanding of city dynamics and urban social behavior. Cranshaw et al. [3] presented a model to extract distinct regions of a city that reflect current collective activity patterns. Noulas et al. [10] proposed an approach to classify areas and users of a city by using venues' categories of Foursquare. Long et al. [7] used a Foursquare dataset to classify venues based on users' trajectories. Despite not dealing with participatory sensing systems Jiang et al. [6] and Toole et al. [19] also investigated users' activities in regions inside a city, using call records to predict land usage, and a travel diary survey to cluster individuals by daily activity patterns showing that daily routines can be highly predictable, respectively.

In our previous work [15], we proposed a technique called City Image for summarizing the city dynamics based on transition graphs, and we show its applicability using eight different cities as examples. In another previous work [16] we performed the first characterization of Instagram using photos shared by users, analyzing them from a sensor network point of view. Moreover, we showed that photo-sharing systems, particularly the Instagram, can also be used to map the characteristics of urban locations at a low cost. Hsieh et al. [5] proposed a time-sensitive model to recommend trip routes based on the information extracted from Gowalla check-ins.

Frias-Martinez et al. [4] used a dataset from Twitter and proposed a technique to determine the type of activities that is most common in a city by studying tweeting patterns. They also proposed another technique to automatically identify landmarks in a city. Quercia et al. [12] studied how social media communities resemble real-life ones. They tested whether established sociological theories of real-life social networks still hold in Twitter. They found, for example, that social brokers in Twitter are opinion leaders who take the risk of tweeting about different topics. Poblete et al. [11] analyzed a twitter dataset aiming the discovery of insights of how tweeting behavior varies across countries, as well as the possible explanations for these differences. Sakaki et al. [13] studied the real-time interaction of events (e.g., earthquakes) in Twitter and proposed an algorithm to monitor tweets to detect a target event.

System	# of data	Interval
Foursquare-OLD	4,672,841 check-ins	Apr/2012 (1 week)
Foursquare-New	4,548,941 check-ins	11 May 13 – 25 May 13
Instagram-OLD	2,272,556 photos	30 Jun 12 – 31 Jul 12
Instagram-New	1,855,235 photos	11 May 13 – 25 May 13

Table 1: Dataset information.

This work differs from previous ones (including ours) because it compares Instagram and Foursquare aiming at understanding whether data from one system could complement the other, or they are compatible regarding the study of city dynamics and urban social behavior.

3. SOCIAL MEDIA AS A SOURCE OF SENSING

Social media systems allow anyone connected to the Internet to provide useful data about the context in which they are at any given moment, such as Instagram and Foursquare, which are called participatory sensing systems (PSSs). A PSS is a concept that originally considers that the shared data is generated automatically, or passively, by sensor readings from portable devices [1]. However, here it is also considered manually or proactively user-generated data. Systems with those characteristics have been called ubiquitous crowdsourcing [8]. The popularity of participatory sensing systems grew rapidly with the widespread adoption of sensor-embedded and Internet-enabled cell phones. Those devices have become a powerful platform that encompasses sensing, computing and communication capabilities, being able to generate both manual and pre-programmed data.

From participatory sensing systems we can derive participatory sensor networks (PSNs). In a PSN, users carry portable devices that are able to sense the environment and make relevant observations at a personal level. Thus, each node in a participatory sensor network consists of a user and his/her mobile device to send data to web services. After that, the data usually can be collected throughout APIs. More details about PSNs can be found in [14, 16]

In this paper we compare two different PSNs, one derived from Instagram, and another one derived from Foursquare. In the PSN derived from Instagram the sensed data is a picture of a specific place. On the other hand, in the PSN derived from Foursquare the sensed data is the actual location associated with a specific category of place (e.g., restaurant). Using those PSNs we can extract information in many ways. For example, we can visualize in near real time how the situation is in a certain area of the city using the network from Instagram. We can also have location labeling using the Foursquare network enabling a better understanding of areas of the city. Another example using both networks is the extraction of popularity of areas in the city.

4. COMPARISON OF DATASETS

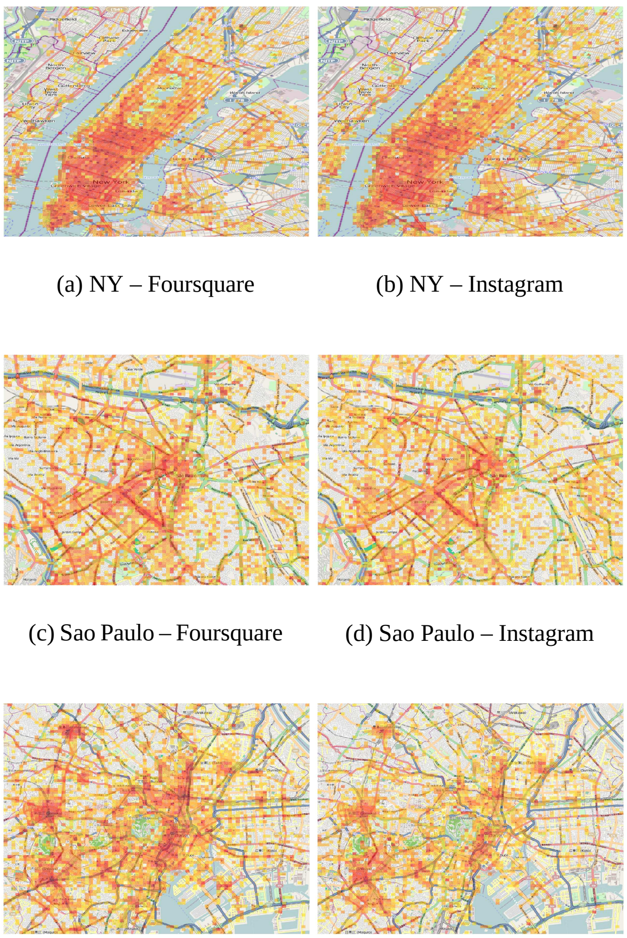
4.1 Dataset description

In this work, we analyze four different datasets as described in Table 4.1. The datasets Instagram-OLD and Foursquare-OLD were previously collected and discussed in [16] and [14], respectively. However, the datasets Instagram-New and Foursquare-New were collected specially for the study presented in this paper. All datasets were collected directly from Twitter[2], since Instagram photos and Foursquare check-ins are not publicly available by default. Note

[2]http://www.twitter.com

that Instagram-New and Foursquare-New have the time of collection in common, which is not the case for Instagram-OLD and Foursquare-OLD datasets. Each content (photo or check-in) consists of GPS coordinates (latitude and longitude) and the time when it was shared.

Throughout this paper we are going to consider three large and populous cities (New York, Sao Paulo, and Tokyo) in several analyses. Figure 1 shows the heat map of the coverage of the datasets for each city, containing all data from Instagram-New and Foursquare-New. The darker the color[3] in the figure, the higher the number of content shared in that area. The coverage for the datasets Instagram-OLD and Foursquare-OLD can be seen in [16] and [14], respectively.



(a) NY – Foursquare (b) NY – Instagram

(c) Sao Paulo – Foursquare (d) Sao Paulo – Instagram

(e) Tokyo – Foursquare (f) Tokyo – Instagram

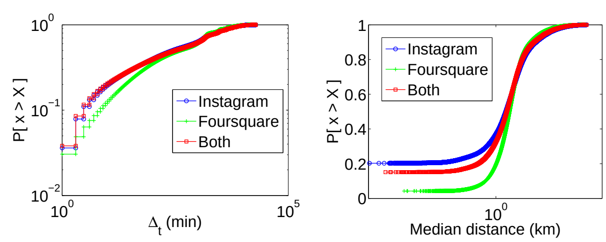
Figure 1: All sensed locations in three populous cities. The number of check-ins in each area is represented by a heat map. The color range from yellow to red (high intensity).

4.2 User behavior

Considering the Instagram-New and Foursquare-New (datasets with a common collection time), we group users in three classes: (1) users that only participated in Instagram; (2) users that only participated in Foursquare; and (3) users that participated in both systems. Figure 2a shows the cumulative density function (CDF) of the frequency of sharing content per class, showing the inter-sharing time in minutes between consecutive content sharing. We can observe that Class 1 (Instagram only), and Class 3 (both systems) contribute more content in shorter intervals than Class 2. For instance, approximately 20% of users in Class 1 and 3 share a consecutive content in an interval up to 10 minutes. In Class 2, the portion of users that share content up to 10 minutes is approximately 12%. This suggests that users tend to share more content in the same place when using Instagram. This was also observed in [16]. The sharing pattern of Class 3 might be dominated by the use of Instagram, explaining the closer similarity among the curves. It is natural to expect a higher volume of content to be shared in the same place through Instagram than in Foursquare. For instance, in a night club users can share a photo of the place, of a drink, and friends.

Figure 2b shows the CDF of the median distance between consecutive uploads for each user. We observe that a significant portion of users from Class 1, around 20%, shared consecutive content at a very short distance, around [1]meter (this was also observed in [16]). This is not observed in the same proportion for the other classes of users. The results for Classes 2 and 3 are 3% and 15%, respectively. This reinforces what was previously observed, i.e., users tend to share content in a shorter distance in Instagram than in Foursquare. For instance, Noulas et al. [9] observed that 20% of the shared locations happen up to [1]km away. Again, the behavior of users that participate in both systems (Class 3), is more similar to Class 1. This closer similarity might be explained by a more intense content contribution in Instagram.

The understanding of user behavior is the first step to model it. With models that explain the user behavior we can make predictions of actions and develop better capacity planning of the system that supports the service.



(a) Contribution – Minute (b) Distance (km)

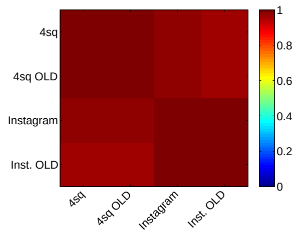
Figure 2: Analysis of classes of users.

4.3 Popularity of areas

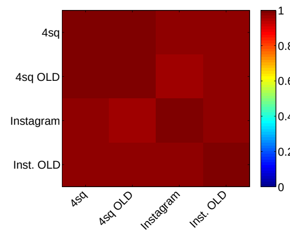
How is the popularity of regions across PSNs derived from Instagram and Foursquare? This is probably one of the main issues in an urban scenario. To answer this question we divided the areas of New York, Sao Paulo, and Tokyo in a 10×10 grid, as shown in Figure 3. After that, we verified the number of content (photo or check-in) shared in each cell of the grid for all four considered datasets. Then, we correlated the number of content in each cell using the Pearson correlation. This result is shown in Figure 4. As we can see the correlation is very high among all datasets. The lowest correlation, although still high, was observed in Tokyo with respect to the correlation of Instagram and Foursquare (both old and

[3]Colors of the heat map for all subfigures are in the same scale.

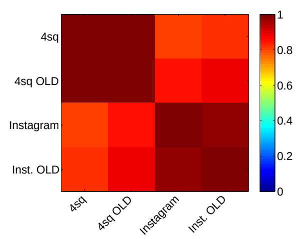
new). This might indicate that the popularity of regions inside cities is consistent regardless of the system, and over the time. Recall that we use two datasets with the same collection time (Instagram-New and Foursquare-New) and two datasets with different collection time (Instagram-OLD and Foursquare-OLD). Besides that, the difference of time between the "new" datasets and the "old" ones are of approximately one year. Maybe what is popular in the city tend to remain popular for a long time and is captured by both systems, since they allow users to express their routines freely.



(a) New York



(b) Sao Paulo



(c) Tokyo

Figure 4: Correlation of popularity of sectors inside cities.

Next, we verified if the popularity of a city is consistent across the systems. Popularity in this case is measured by the number of content shared in the city. For that we considered 29 cities around the world [4]: Latin American cities (Belo Horizonte, Buenos Aires, Mexico City, Rio, Santiago, and Sao Paulo); American cities (Chicago, Los Angeles, New York, San Francisco); European cities (Barcelona, Istanbul, London, Madrid, Moscow, and Paris); Asian cities (Bandung, Bangkok, Jakarta, Kuala Lumpur, Kuwait, Manila, Osaka, Semarang, Seoul, Singapore, Surabaya, Tokyo); and Australian cities (Melbourne, Sydney). We ranked all the cities by the number of content shared on it, then we correlated these ranking using Spearman correlation. Figure 5 displays the correlation results. As we can see the popularity of cities tend to be very correlated over time for the same system, but this is not the case for different systems. This means that users may use Instagram and Foursquare in particular ways on different cities. For instance, Foursquare might be very popular in Tokyo, but Instagram might not be as popular. Cultural differences might help to explain these results.

4.4 Routines and the data sharing

Figure 6 shows the temporal sharing pattern for Instagram and Foursquare considering the old and new datasets. This figure shows the average number of photos shared per hour during weekdays

[4]Chosen by their popularity and representativeness of different regions

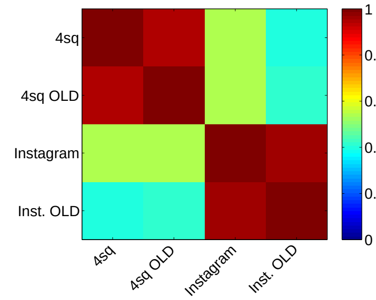


Figure 5: Spearman correlation of popularity between cities.

(Monday to Friday), and also during the weekend (Saturday and Sunday). As previously observed in [16] for the Instagram-OLD dataset, we can also see two peaks of activity throughout the day, one around lunch and the other at dinner time. But, we cannot see a clear peak at breakfast time, as the one observed in Figure 6c and also in [2]. During the weekends we cannot observe clear peaks of activities inherent of routines. Rather, the activity remains intense throughout the afternoon until early evening.

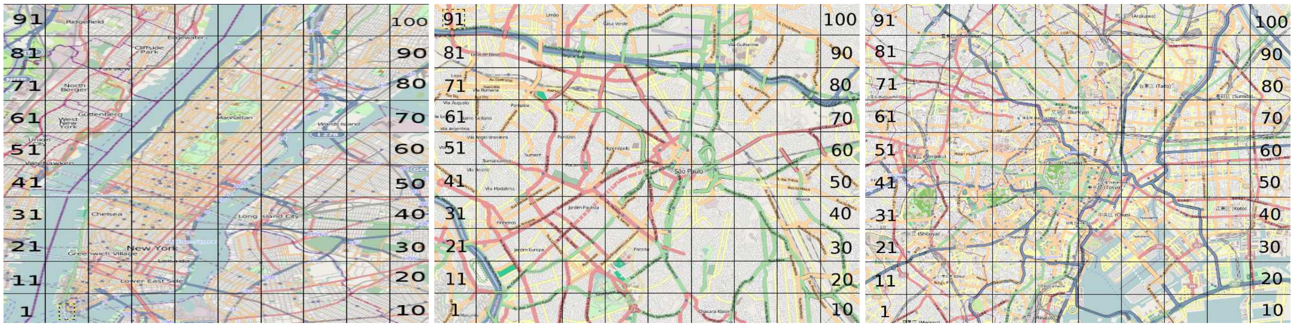
Surprisingly, we see that the sharing pattern for each curve regarding to the old and new datasets, both on Instagram and Foursquare, are very similar, despite the huge gap between collections (approximately one year). This is the case for weekdays and weekends, suggesting that the user behavior in both systems tend to keep consistent over time, reinforcing what was observed in Section 4.3. This is an interesting and important result because it shows how we can use different datasets.

In Figure 7 we show the correlogram for the temporal sharing pattern of Instagram-New and Foursquare-New datasets, during the weekday (Figure 7a) and weekend (Figure 7b). The correlogram plots correlation coefficients on the vertical axis, and lag values (in hours) on the horizontal axis, and it is an important tool for analyzing time series in the time domain. As we can see, the lag of one hour in the time series of Instagram-New dataset provides the highest correlation, however it is not 1 (maximum). Analyzing the cross-correlation for weekend, we observe that a lag of 0 provides a correlation of 1, indicating that the time series is already very correlated. This suggests that users have particular sharing pattern in each system during weekdays, but it is not the case on weekends. The users' routines performed on weekdays may be the explanation for these results. The act of sharing a photo might be more likely to happen in special occasions that are usually out of the routines of people. For example, during breakfast time it is probably uncommon to happen something interesting to share a photo, but, for example, when you go out at night to have a dinner you have more incentives to share photos.

We now study how routines impact the sharing behavior analyzing the sharing pattern during weekdays, considering the datasets Instagram-New and Foursquare-New for New York, Sao Paulo, and Tokyo. The results are shown in Figure 8[5]. In all figures we display two cities from the same country for the new collected datasets, and one city for the old dataset as a reference of comparison.

First, observe the distinction between curves of each city in the same system (e.g., Instagram, Figures 8a, c, e) and also across different systems (e.g., Figures 8a and 8b for New York). Next, observe that the sharing pattern for each city in the same country is fairly similar, which might indicate cultural behaviors of inhabi-

[5]Each curve is normalized by the maximum number of content shared in a specific region representing the city.



(a) New York (b) Sao Paulo (c) Tokyo

Figure 3: Grids for the areas of New York, Sao Paulo, and Tokyo.

tants of those countries, presenting somehow the signature of a certain culture.

Note that the sharing pattern in Instagram for American cities (Figure 8a) and Japanese cities (Figure 8e) present peaks that reflect typical lunch and dinner times. This is not the case for the curves that represent the Brazilian sharing pattern in the cities of Sao Paulo and Rio (Figure 8c), where not all peaks represent typical meal times, suggesting that Brazilians share photos in atypical moments. Besides that, in general, the Brazilian activity is more intense late at night. This information was also observed considering only the Instagram-OLD dataset in [16].

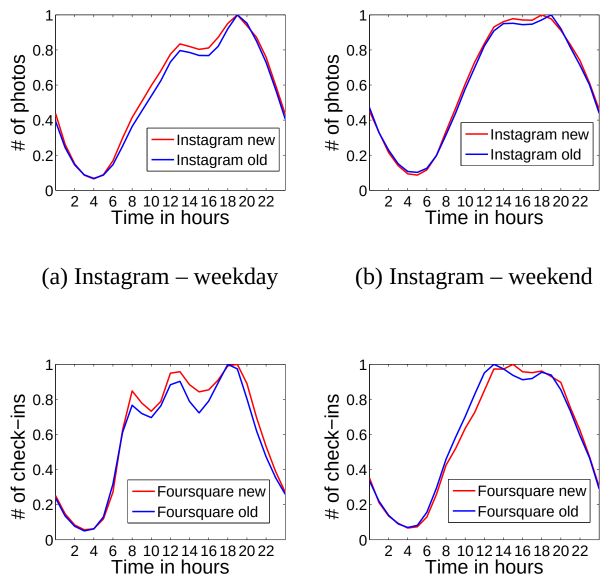
The sharing pattern of the new dataset of Foursquare varied more when compared to the old one (Figures 8b, d, f), than the variation observed in the Instagram datasets (Figures 8a, c, e). Observe also that the sharing pattern in Instagram for each analyzed city is more distinct to each other than the one observed for Foursquare. This suggests that using the sharing pattern from Instagram we might have a more distinguishable "cultural signature" for a certain region, and less susceptible to changes over time.

4.5 Mapping transitions

In a PSN, mobile nodes (users and portable devices) move accordingly to their routines or local preferences sharing data along the way. Looking at data people share it is possible to have a sort of rudimentary location tracking. If we aggregate all transitions performed by all users we can obtain common paths users tend to take in the city.

Given that observation, a question emerges: can we observe a similar movement of people using a PSN derived from Instagram and Foursquare? In order to address this question we create a directed graph $G(V, E)$, where nodes $v_i \in V$ are a cell in the grid a particular city shown in Figure 3. A direct edge (i, j), representing a transition, exists from node v_i to node v_j if at some point in time a user shared a content in cell v_j just after sharing a content in cell v_i. The weight $w(i, j)$ of an edge is the total number of transitions that occurred from cell v_i to cell v_j. Some features of transitions: (i) the content must be shared consecutively and by the same individual; (ii) continuous content sharing at the same considered venue cell represents a self-loop; and (iii) a transition must have occurred at the same day (we only consider transitions occurred from 6:00am to 6:00pm).

Figures 9, 10, and 11 show the transition graphs for New York, Sao Paulo, and Tokyo, respectively. In those figures, for better visualization, we excluded all edges with weight $w = 1$. Nodes' positions in the figure are depicted according to the cell position they represent in the city area. Nodes not displayed mean that no one shared content in that particular area of the city.



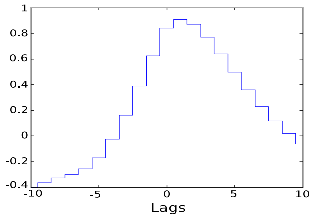
(a) Instagram – weekday (b) Instagram – weekend

(c) Foursquare – weekday (d) Foursquare – weekend

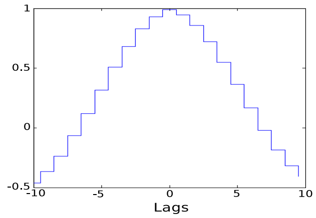
Figure 6: Temporal sharing pattern for Instagram and Foursquare – new and old datasets.

Note that there are few transitions in the city. In other words, typical movements in the city might not be very diverse. It is also interesting to observe that we could capture more transitions with the Foursquare dataset. This means that check-ins might be more effective to track typical routes of users. However this hypothesis needs further investigation, because this result might be due to the large amount of data obtained in the Foursquare dataset. An interesting possibility in this direction is use data mining algorithms, such as [22], on transition graphs to discover movement patterns.

In order to compare the similarity graph, we first discard all self-loops, since those transitions tend to be more likely to happen, then

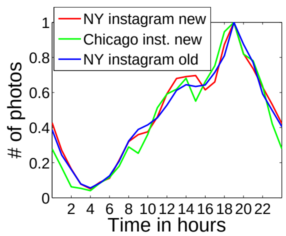


(a) Correlation – weekday

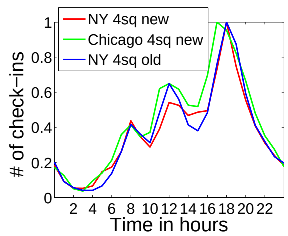


(b) Correlation – weekend

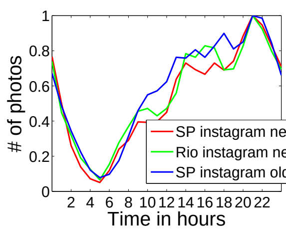
Figure 7: Cross-correlation between Instagram-New and Foursquare-New datasets, during weekday and weekend.



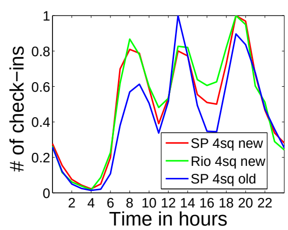
(a) New York – Instagram



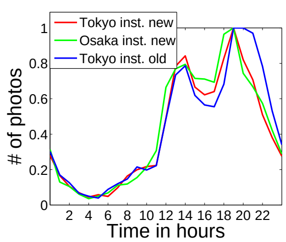
(b) New York – Foursquare



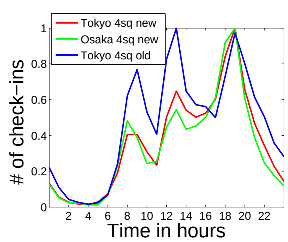
(c) Sao Paulo – Instagram



(d) Sao Paulo – Foursquare



(e) Tokyo – Instagram

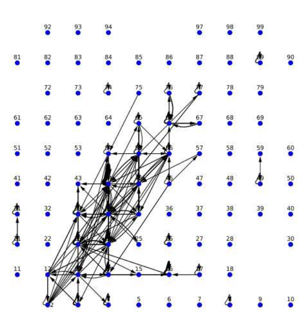


(f) Tokyo – Foursquare

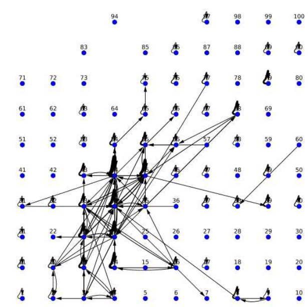
Figure 8: Temporal sharing pattern of Instagram and Foursquare for New York, Sao Paulo, and Tokyo during weekdays.

we rank the resulting transitions and select the top ten from each graph. We compare the groups of top transitions between Instagram and Foursquare graphs of each city, analyzing the number of transitions in common. The results show that approximately 70%, 50%, and 70% of the top transitions are similar for New York, Sao Paulo, and Tokyo, respectively. However, if we take now the top twenty transitions the results for transitions in common are approximately: 59%, 53%, and 50%, for New York, Sao Paulo, and Tokyo, respectively. This indicates that the graphs are not very similar, but popular transitions are more likely to be expressed by both systems.

The similarity graphs are also evaluated in a different way. For this comparison we preserve the graphs without discarding self-loops, then we compute the difference between sets of edges of Instagram and Foursquare for each city. We discovered that graphs for NY have 156 different edges, and these numbers are 237, and 376 for Sao Paulo and Tokyo, respectively. This is a significant difference, and these results are partially explained by the fact that Foursquare graphs captured more transitions.

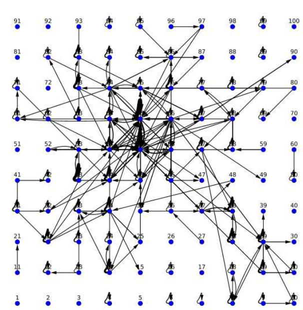


(a) Foursquare – Day

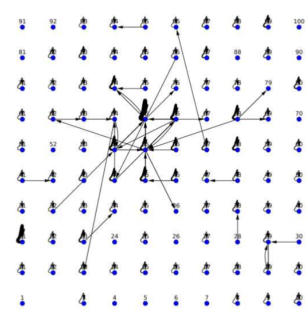


(b) Instagram – Day

Figure 9: Transition graphs – New York.



(a) Foursquare – Day

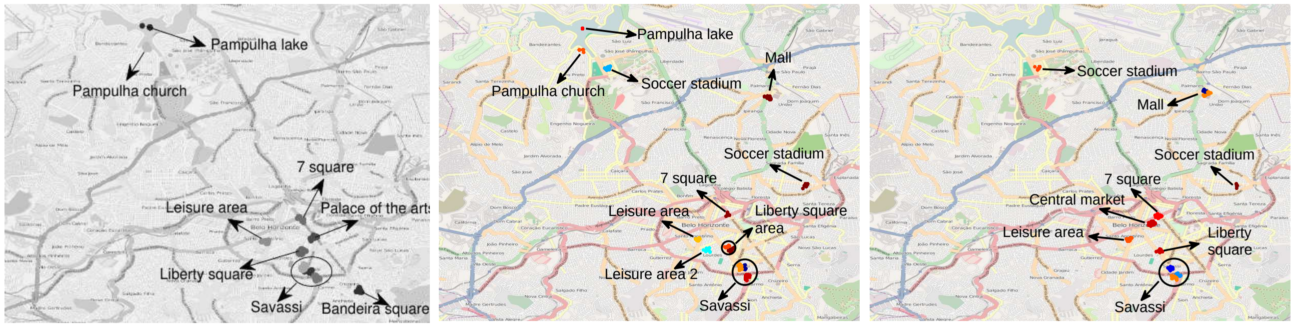


(b) Instagram – Day

Figure 10: Transition graphs – Sao Paulo.

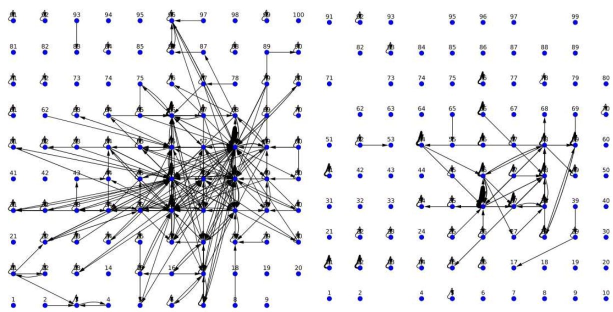
4.6 Sights extraction

In a previous work [16] we showed that with a PSN derived from Instagram it might be possible to identify points of interest (POI) in a city, which are particular areas that attract more attention of residents and visitors. We also showed that out of the POIs it might be possible to extract sights of the city. This is possible because each picture represents, implicitly, an interest of an individual at a given moment.



(a) Sights – Instagram-OLD (b) Sights - Instagram-New (c) Sights – Foursquare-New

Figure 12: Sights identified in different datasets.



(a) Foursquare - Day (b) Instagram - Day

Figure 11: Transition graphs – Tokyo.

Here we have two goals: (i) verify whether with the Instagram-New dataset we can identify the same sights showed in [16], which used the Instagram-OLD dataset; and (ii) verify whether the Foursquare dataset can also be used for this purpose, by using the Foursquare-New dataset. Following the steps described in [16], we formalize the process of identifying sights in the following manner:

1. Associate with a point p_i each pair i (photo or check-in) of coordinates (longitude, latitude) $(x, y)_i$;

2. Calculate the distance [17] between each pair of points (p_i, p_j);

3. Group all points p_i that have a distance smaller than [250]m into a cluster C_k[6];

4. Exclude clusters that may have been generated by random situations, i.e., those that do not reflect the dynamics of the city. In order to perform that, for each cluster C_k, we create an alternative cluster C_r. Then, for each photo f_i, we randomly choose an alternate cluster C_r and we assign f_i to C_r. After that, from the original clusters C_k found in the previous step, we exclude those in which the number of content (photo or check-in) is within the distance 2σ from the average μ, or is in the range $[\mu - 2\sigma; \mu + 2\sigma]$ (number of content assigned to each alternative cluster follows a normal distribution with mean μ and standard deviation σ);

5. Generate a graph $G(V, E)$, where the vertices $v_i \in V$ are all POIs and there is an edge (i, j) from vertex v_i to vertex v_j if in a given time a user shared a content on a POI v_j, after having shared a content on POI v_i. The weight $w(i, j)$ of an edge of G represents the total number of transitions performed from POI v_i to POI v_j considering transitions of all users. Following the same procedure mentioned in [16], we exclude from G all edges (i, j) with weights $w(i, j)$ smaller than a threshold t ($t = 4$ and $t = 8$ for the Instagram-New and Foursquare-New datasets, respectively), which are given by the probability of generating $w(i, j)$ randomly in a random graph $G_R(V, E_R)$. The idea is to preserve edges (i, j) with high weights $w(i, j)$, because according to the conjecture they denote these frequent transitions from one sight to another.

Figure 12 shows sights identified for different PSNs. As a baseline of comparison, Figure 12a shows the sights previously identified using the dataset Instagram-OLD and presented in [16]. Figures 12b and 12c show the sights identified for Instagram-New and Foursquare-New datasets, respectively. During the collection of Instagram-OLD Belo Horizonte was not receiving soccer games. This explains why no soccer stadium was identified. Apart of that, we can see that many of the sights identified are in common in all three datasets, for example, Liberty Square, one of the most important sights of Belo Horizonte. The sights that were only previously identified, Palace of the Arts, and Bandeira Square, might not have been identified in the new datasets because no special event happened in those places. Palace of the Arts is a gallery with itinerant expositions, and Bandeira Square is not a spot that attracts naturally many people, especially tourists. It is interesting to note that, all social networks identified relevant sights of the city of Belo Horizonte, and they might be able to complement each other, since no one found all sights.

5. CONCLUSIONS AND FUTURE WORK

Data available in Instagram and Foursquare can naturally complement each other. For instance, a user using Foursquare could check-in in a location X and label this place as a restaurant. From this we know that in location X there is a restaurant. The same user could also share photos in location X using Instagram. Now,

[6] for each cluster C_k, we consider only one point (photo or check-in) per user.

besides knowing that location X is a restaurant we have the opportunity to visually check the environment by looking at the shared pictures. In this work we do not take into account these pieces of data that complement Instagram or Foursquare. Instead we compare Instagram and Foursquare considering just the time and location where the content (photo or check-in) were shared. We aim to understand whether this data from one system could complement the other, or they are compatible regarding the study of city dynamics and urban social behavior.

In the following, we summarize our findings:

- both Instagram and Foursquare datasets might be compatible in finding popular regions of cities;
- the temporal sharing pattern did not vary considerably over time for the same system. However, the sharing pattern for each system during weekdays are distinct;
- both Instagram and Foursquare might be used to capture particular signatures of cultural behaviors, but apparently Instagram offers a more distinguishable "cultural signature", and is less susceptible to changes over time;
- Foursquare is apparently better to express typical routes of people inside cities;
- and both Instagram and Foursquare are good in sights identification, and since the results might be complementary it is recommended to use both of them in this task.

Currently, we are investigating in further details the use of Instagram and Foursquare as a tool to identify cultural differences. We are also using data from those systems to better understand the city dynamics, and, thus, offering smarter services in the cities. As a future step we intend to analyze other PSNs and develop new applications that exploit these networks.

6. REFERENCES

[1] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava. Participatory sensing. In *In: Work. on World-Sensor-Web (WSW'06): Mobile Device Centric Sensor Networks and App.*, pages 117–134, 2006.

[2] Z. Cheng, J. Caverlee, K. Lee, and D. Z. Sui. Exploring Millions of Footprints in Location Sharing Services. In *Proc. of the Fifth Int'l Conf. on Weblogs and Social Media (ICWSM'11)*, 2011.

[3] J. Cranshaw, R. Schwartz, J. I. Hong, and N. Sadeh. The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. In *Proc. of the Sixth Int'l Conf. on Weblogs and Social Media*, 2012.

[4] V. Frias-Martinez, V. Soto, H. Hohwald, and E. Frias-Martinez. Characterizing urban landscapes using geolocated tweets. In *Proc. of the 2012 ASE/IEEE Int. Conf on Social Computing*, pages 239–248, Washington, DC, USA, 2012. IEEE.

[5] H.-P. Hsieh, C.-T. Li, and S.-D. Lin. Exploiting large-scale check-in data to recommend time-sensitive routes. In *Proc. of the ACM SIGKDD Int Workshop on Urban Computing*, UrbComp '12, pages 55–62, Beijing, China, 2012. ACM.

[6] S. Jiang, J. Ferreira, Jr., and M. C. Gonzalez. Discovering urban spatial-temporal structure from human activity patterns. In *Proc. of the ACM Int. Work. on Urb. Comp.*, UrbComp'12, pages 95–102, Beijing, China, 2012. ACM.

[7] X. Long, L. Jin, and J. Joshi. Exploring trajectory-driven local geographic topics in foursquare. In *Proc. of the 2012 ACM Conf on Ubiquitous Computing*, UbiComp '12, pages 927–934, New York, NY, USA, 2012. ACM.

[8] A. J. Mashhadi and L. Capra. Quality Control for Real-time Ubiquitous Crowdsourcing. In *Proc. of the 2nd Int'l Work. on Ubiq. Crowdsouring (UbiCrowd'11)*, pages 5–8, 2011.

[9] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil. An Empirical Study of Geographic User Activity Patterns in Foursquare. In *Proc. of the Fifth Int'l Conf. on Weblogs and Social Media (ICWSM'11)*, 2011.

[10] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil. Exploiting Semantic Annotations for Clustering Geographic Areas and Users in Location-based Social Networks. In *Proc. of the 5th Int'l Conf. on Web. and Soc. Med.(ICWSM'11)*, 2011.

[11] B. Poblete, R. Garcia, M. Mendoza, and A. Jaimes. Do all birds tweet the same?: characterizing twitter around the world. In *Proc. of the 20th ACM int. conf. on Inf. and knowl. manag.*, pages 1025–1030. ACM, 2011.

[12] D. Quercia, L. Capra, and J. Crowcroft. The social world of twitter: Topics, geography, and emotions. In *The 6th int. AAAI Conf on weblogs and social media, Dublin*, 2012.

[13] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proc. of the 19th int. conf. on World wide web*, WWW '10, pages 851–860, New York, NY, USA, 2010. ACM.

[14] T. H. Silva, P. O. S. Vaz de Melo, J. M. Almeida, and A. A. F. Loureiro. Challenges and opportunities on the large scale study of city dynamics using participatory sensing. In *18th IEEE ISCC'13*, Split, Croatia, July 2013.

[15] T. H. Silva, P. O. S. Vaz de Melo, J. M. Almeida, J. Salles, and A. A. F. Loureiro. Visualizing the invisible image of cities . In *Proc. of IEEE Int. Conf on Cyber, Physical and Social Computing (CPScom'12)*, November 2012.

[16] T. H. Silva, P. O. S. Vaz de Melo, J. M. Almeida, J. Salles, and A. A. F. Loureiro. A picture of Instagram is worth more than a thousand words: Workload characterization and application. In *Proc. of the IEEE Int. Conf on Dist. Comp. in Sensor Sys. (DCOSS'13)*, Cambridge, MA, USA, May 2013.

[17] R. W. Sinnott. Virtues of the Haversine. *Sky and Telescope*, 68(2):159+, 1984.

[18] I. Smith, S. Consolvo, A. Lamarca, J. Hightower, J. Scott, T. Sohon, J. Hughes, G. Iachello, and G. D. Abowd. Social Disclosure of Place: From Location Technology to Communication Practices. In *Proc. of the 3rd Int. Conf on Perv. Comp. (Pervasive '05)*, pages 134–151, 2005.

[19] J. L. Toole, M. Ulm, M. C. González, and D. Bauer. Inferring land use from mobile phone activity. In *Proc. of the ACM SIGKDD Int. Workshop on Urban Computing*, UrbComp '12, pages 1–8, Beijing, China, 2012. ACM.

[20] Y. Zheng. Location-based social networks: Users. In *Computing with spatial trajectories. Zheng, Yu and Zhou, Xiaofang*. Springer press, 2011.

[21] Y. Zheng. Tutorial on Location-Based Social Networks. In *Proc. of Int. Conf. on World Wide Web (WWW'12)*, Lyon, France, 2012.

[22] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma. Mining interesting locations and travel sequences from GPS trajectories. In *Proc. of Int. Conf. on World Wide Web (WWW'09)*, Madrid, Spain, 2009.