Sensing airports' traffic by mining location sharing social services

John Garofalakis[1], Ioannis Georgoulas[1], Andreas Komninos[2], Periklis Ntentopoulos[1], Athanasios Plessas[1]

[1]Computer Engineering and Informatics Department, University of Patras, Patras, Greece
{garofala, georgoulas, ntentopoul, plessas}@ceid.upatras.gr
[2]Computer and Information Sciences, University of Strathclyde, Glasgow, UK
andreas.komninos@strath.ac.uk

Abstract. Location sharing social services are popular among mobile users resulting in a huge social dataset available for researchers to explore. In this paper we consider location sharing social services' APIs endpoints as "social sensors" that provide data revealing real world interactions, although in some cases, the number of recorded social data can be several orders of magnitude lower compared to the number of real world interactions. In the presented work we focus on check-ins at airports performing two experiments: one analyzing check-in data collected exclusively from Foursquare and another collecting additionally check-in data from Facebook. We compare the two popular location sharing social platforms' check-ins and we show that for the case of Foursquare these data can be indicative of the passengers' traffic, while their number is hundreds of times lower than the number of actual traffic observations.

Keywords: Location sharing services• Foursquare• Facebook• Check-ins• Ubiquitous social computing

1 Introduction

During the last decade, the use of online social networking tools and services became widespread, with mobile devices playing an important role as they allow users to connect with others and share information anytime and anywhere. As social networking platforms engage millions of users, there is a boom in the amount of social data that are produced. These data are proving to be representative of real world phenomena and their analysis allows researchers to get the 'big picture' behind social interactions. As an example, consider the work of Bollen, Mao and Zeng [1] who detected public mood from public tweets and were able, under conditions, to predict stock market behavior. Another example is the real-time detection of events, such as earthquakes [2] or sports events [3], based on social interactions.

Moreover, most social networking platforms allow users to geographically characterize the information they share. All modern mobile devices are equipped with GPS sensors, providing social networking applications with the user's location context. For example, Facebook allows users to check-in (declare one's presence at a location) or

adfa, p. 1, 2011.

geo-tag their posts, while Twitter and Flickr permit users to geo-tag their tweets and photos respectively, and Foursquare to check-in at places, rate them and leave tips for others. These interactions result in large exploitable datasets conveying rich semantics and describing patterns of human interactions with their environment. However, this kind of social data constitute only a small portion of actual social interactions. In this paper we aim our attention especially at check-ins, i.e. the sharing of a user's location in location based social networks, such as Facebook and Foursquare and the geo-annotation of information about a venue using spatial and temporal context.

More specifically, we focus on airports' check-ins and our purpose is to examine whether check-in data from such platforms are adequate in order to get a representative picture of the airports' real passenger traffic. Airport venues belong to the "Travel & Transport" category, which is found to be the most popular, regarding the number of per venue check-ins [4]. We consider location sharing social services APIs' endpoints as sensors providing social data that offer useful insight about real passenger traffic. In the framework of our research we performed two chronologically distinct experiments. During the first one we collected and analyzed approximately 2.000.000 foursquare check-ins from 10 airports in a period of 13 months, wanting to explore possible correlations between these social data and real passenger traffic. During the second experiment we collected around 360.000 foursquare check-ins, 1.155.000 Facebook check-ins and 100.000 Facebook likes from 7 airports for 6 months, while our purpose was to verify or reject our findings when including other location sharing services.

The rest of this paper is organized as follows: in section 2, related work on this area is covered. In section 3, we briefly present the check-in feature of Foursquare and Facebook, while in section 4 we describe how our dataset was collected. In section 5, we present our analysis and finally, in section 6, we discuss our conclusions and our directions for further work.

2 Related Work

Our work is not the first on capturing and analyzing location sharing social data. Several other researchers have already understood the value of these social interactions and how their study may reveal real world patterns and trends. However, as far as we know, this effort is the first that attempts to exploit check-ins at airports and examine possible correlation with actual passenger traffic.

Several works focus on using location sharing social data to reveal city dynamics. In [5] it was shown that even scarce check-in data are in line with datasets from other sources (e.g. traffic data, air pollution data) and can be used to build a good model of the city's dynamics. Another work on utilizing social media to understand city dynamics is the one presented in [6]. The authors analyzed approximately 18 million Foursquare check-ins (published as tweets) and proposed a methodology that reveals city clusters called "Livehoods", reflecting the dynamic areas that characterize the city. Another effort on using location sharing data for identifying urban neighborhoods and modeling the users' activities is described in [7].

In [8], the authors used Foursquare and Instagram datasets and tried to better understand location related information as an important aspect of urban phenomena. They analyzed these datasets to observe users' movements and activities, popularity of city regions and in general to capture city dynamics. Another work focusing on analyzing urban check-ins is the one presented in [9], where a Foursquare dataset from a German city returned clear patterns separating areas known for different activities, such as nightlife or daily work. In [10], Foursquare check-ins are used in a preliminary effort to sense and analyze the geo-social activity distribution of the cities of London, Paris and New York.

Silva et al. [11] showed a different use of Foursquare check-ins useful for economic purposes and also able to support social and marketing applications. More specifically, the researchers identified cultural boundaries and similarities (such as food and drink habits) across populations by analyzing location sharing data. In [12] check-ins from Foursquare were used to predict the impact of Olympic Games in local businesses. Finally, in [13] the researchers investigated the impact of check-ins on identifying the optimal location for retail shops.

It is apparent that the research community is highly interested in taking advantage of data published in location sharing social networks in order to reveal hidden geo-social patterns. Such works could be beneficial to a wide range of society aspects, including but not limited to marketing, tourism and the financial sector. However, until now, research mainly focuses on understanding urban dynamics, proving useful for urban planning but disregarding potential gains for other sectors.

3 Foursquare and Facebook Check-ins

Foursquare is a popular location sharing service, with over 45 million users as of January 2014 [12], mainly targeting mobile users, as it is available for the most widely adopted mobile operation systems (including Android and iPhone OS). Foursquare users are able to share their location with their friends, in the form of a check-in at a place (venue) from a list of nearby places. Moreover, users are able to leave "Tips" about venues that may prove useful to other users visiting these places.

The approach Foursquare has taken regarding check-ins (allowing users to selectively report their location or not) has changed location from a state (being somewhere) to an action (doing something) [14]. Foursquare users, according to [14], adapt their check-ins to norms of what they perceive as worthwhile check-ins. They tend to check-in at places they find interesting and avoid check-ins at places considered uninteresting. In [15], Lindqvist et al. also examined why people use Foursquare. They categorized it as a primarily social driven application, with much of the motivation coming from the benefit of sharing location check-ins. However, some users are motivated by keeping a record of the places they check-in. At the same study, the researchers found again that users tend to avoid checking in at places that are not considered interesting or where sharing their location would make them feel embarrassed. Cheng et al. [16] and Rost et al. [17] report that airports are among the top places that Foursquare users check-in.

On the other hand, Facebook introduced Facebook Places on 2010, allowing users to check-in via mobile devices and also tag their friends. Location tagging was later expanded as a more general feature, allowing users to share their locations on status updates, photos and wall posts [18]. Similar norms (e.g. increase of self popularity, optimization of self image etc.) as those described earlier exist also about self-presentation via check-ins for Facebook users [18][19].

Foursquare and Facebook offer public APIs, providing information such as the number of check-ins at a place, the number of unique users that have checked-in, the users that are currently present at this place (in Foursquare) etc. However, these check-ins are anonymous and cannot be assigned to the respective users.

4 Data Collection

For our first experiment, whose purpose was to examine possible correlations between Foursquare data and real passenger traffic, we have used the Foursquare venue API in order to collect our dataset that includes the number of check-ins and the number of new users that have checked-in at each airport. A web based script was running every hour, querying Foursquare about the following 10 European airports: Athens International Airport, Heathrow Airport, Paris CDG, Moscow Sheremetyevo Airport, Madrid Barajas, Budapest International Airport, Rome Fiumicino, Frankfurt am Main, Amsterdam Schiphol and Munich International Airport. We selected some of the busiest European airports and we also included in our research a couple of airports with less traffic (e.g. Athens, Budapest). Each airport's API endpoint is seen as a social data sensor, used to sense Foursquare users' social interactions. Each query to the API returns the number of check-ins at that time and the number of individual users that have checked-in. As a result, by subtracting each number with that of the previous hour we were able to compute the check-ins and the new users that have checked-in within this hour and this information was saved in our database. Our research took place for a period of 13 months, from May of 2013 until the end of May 2014, with the exception of Budapest Airport that was added later (gathered data for this airport expand from July 2013 to May 2014). In total, there were 2.032.273 check-ins and the new users having checked-in during that period were 757.975.

Our analysis, presented in the next section was performed in a monthly base, as we were able to find only monthly statistics for these airports, to compare with. More specifically, we used the statistics published at Anna Aero[1], which is a website dedicated to airline and airports network news and analysis.

In table 1, we summarize our findings for each airport:

Table 1. Summary of Foursquare airport data from first experiment

Airport	*# Check-ins*	*# New Users*	*# Passengers*
Athens Intl	96.100	26.635	14.255.499
Heathrow	358.155	131.142	79.040.930

[1] http://www.anna.aero

Frankfurt am Main	188.346	75.040	63.636.892
Paris CDG	212.289	105.398	68.393.212
Amsterdam Schiphol	236.206	80.464	58.288.233
Madrid Barajas	153.329	61.948	43.569.777
Munich Intl	161.406	57.773	42.198.301
Moscow SVO	480.295	156.211	32.792.835
Rome Fiumicino	85.007	41.153	39.836.419
Budapest Intl	61.140	22.211	7.922.530
Total	**2.032.273**	**757.975**	**449.934.628**

For our second experiment, where we wanted to confirm our findings using other location sharing services, we expanded our software incorporating the Facebook API and as a result we were able to collect data from both location sharing social services. At first, we had to identify the Facebook pages that correspond to the airport venues from Foursquare that were included in our first experiment. In this case, our research took place for a period of 6 months, from October of 2014 until the end of March 2015. During this period the Facebook pages for Rome Fiumicino and Paris CDG airports that we queried the API for were merged with other existing pages, however we failed to notice this fact. As a result, we were unable to capture check-ins from the merged pages. Moreover, for an unknown reason, the Facebook API returned an abnormally low number of new check-ins for Amsterdam Schiphol airport. Consequently, we were forced to exclude these three airports from our analysis. As shown in table 2, we were left with seven airports, recording in total 357.656 Foursquare check-ins, 86.625 new Foursquare users having checked-in, 1.155.395 Facebook check-ins and 97.454 Facebook likes.

Table 2. Summary of Foursquare and Facebook airport data from second experiment

Airport	***# 4sq Check-ins***	***# 4sq New Users***	***# FB Check-ins***	***# FB Likes***	***# Passengers***
Athens Intl	34.624	6.263	153.673	7.928	5.390.275
Heathrow	77.636	20.640	348.465	37.228	28.122.242
Frankfurt am Main	45.447	15.073	286.256	26.958	21.832.189
Madrid Barajas	34.359	10.980	108.456	2.920	16.516.643
Munich Intl	39.877	11.325	125.521	11.322	14.984.963
Moscow SVO	102.961	15.130	67.130	3.543	11.283.720
Budapest Intl	22.752	72.14	65.894	7.555	4.180.418

Total	**357.656**	**86.625**	**1.155.395**	**97.454**	**102.310.450**

We should note here that at the time of our analysis there were not actual passenger data for all airports (with the exception of the Budapest airport) for March 2015.

5 Data analysis

In this section we present the results of our data analysis. As already mentioned, in our first experiment our purpose was to examine whether there is a correlation between social data from Foursquare and the actual number of passengers travelled in each airport. Figure 1 depicts for each airport a graph showing the actual number of passengers (orange line), the number of check-ins (blue line) and the number of new users having checked-in (grey line) on a monthly base. The left vertical axis corresponds to the number of passengers (orange line), while the right vertical axis corresponds to the social data gathered (blue and grey lines).

A statistical analysis of the monthly observations for all airports reveals a strong positive correlation between the number of check-ins and the actual number of passengers, which is statistically significant (r_s=0.588 $p<0.01$). Moreover, there is an even stronger positive correlation between the number of new users having checked-in and the number of passengers, which is also statistically significant (r_s=0.666 $p<0.01$). Finally, a strong statistically significant positive correlation was also found between the number of new users and the number of check-ins (r_s=0.956 $p<0.01$). These correlations confirm the similar trends that follow the graphical representations of these variables.

When examining data from each airport separately, results are quite different. Statistically significant correlations between the number of check-ins and the number of passengers were found for the following airports: Athens (r_s=0.736 $p<0.01$), Heathrow (r_s=0.698 $p<0.01$), Frankfurt (r_s=0.648 $p<0.05$), Paris (r_s=0.725 $p<0.01$), while for the airport of Amsterdam the correlation is not marginally statistically significant (r_s=0.523 p=0.066). However, statistically significant correlation is found between the number of new users and the number of passengers when analyzing each airport data, for all cases except for the airport of Munich: Athens (r=0.900 $p<0.01$), Heathrow (r=0.883 $p<0.01$), Frankfurt (r=0.935 $p<0.01$), Paris (r_s=0,841 $p<0.01$), Amsterdam (r=0.605 $p<0.05$), Madrid (r=0.563 $p<0.01$), Moscow (r=0.929 $p<0.01$), Rome (r=0.576 $p<0.05$) and Budapest (r=0.638 $p<0.05$). Again a positive statistically significant correlation was found between the number of new users and the number of check-ins for each airport, except for the international airport of Moscow.

Having found that social data from Foursquare are statistically correlated to actual passenger traffic, we wanted to check if this is also true for other location sharing social networking services. Working with the recorded data from the second experiment we were able to statistically analyze Facebook data in relation to Foursquare data and real-world observations. An important finding of this analysis was that there is no statistically significant correlation between Foursquare check-ins when compared to Facebook check-ins or Facebook likes.

Fig. 1. Data graph for each airport in first experiment

Moreover, the statistical analysis of Facebook data from each airport venue does not reveal statistically significant correlation between Facebook check-ins or likes and real world observations, with the exception of Facebook likes for the Moscow Sheremetyevo airport ($r=0.978$ $p<0.05$). Finally, once again we observe that new Foursquare users having checked in at the airport venues is the best indicator for actual passengers traffic, since we found strong statistically significant correlations for the following airports: Athens ($r=0.928$ $p<0.05$), Heathrow ($r=0.952$ $p<0.05$), Madrid ($r=0.956$ $p<0.05$), Moscow ($r=0.964$ $p<0.01$), while for Frankfurt ($r=0.832$ $p=0.08$) and Munich ($r=0.870$ $p=0.055$) the correlations are not only marginally statistically significant.

6 Discussion and Future Work

In this paper we considered location sharing social networks as sensors providing geo-social data, which we analyzed to reveal correlations with real-world data. We focused on airports, as airport traffic is an important financial factor for national economies and is often used as a tourism indicator [20]. Moreover, as mentioned earlier, venues belonging to the Tourism & Travel category such as airports are the most popular places where location sharing social services' users check-in and as a result we considered them as an interesting starting point for our research. While the analysis of check-ins is popular (as presented in the related work section), as far as we know this is the first study focusing exclusively on airport venues.

When examining total observations from all sensors – airports in our first experiment we found that the number of recorded Foursquare check-ins and even more the number of new users performing a check-in are strongly correlated to the actual number of passengers that traveled to and from those airports. Thus, location sharing data seem to be representative of real world data for the airport venue type, which is encouraging for the adoption of geo-social data analysis for other domains apart from urban dynamics, which is mostly referenced in related literature. Consequently, interested parties are able to capture the traffic trend for airports before official data are announced. In addition to this, one could argue that an idea worth to study is this of incorporating airport check-ins into airport traffic prediction models, which is a subject that has attracted considerable attention from the research community [21].

However, a significant finding from our second experiment was that we were not led to the same conclusion when examining Facebook data. As a result, it seems that not all location sharing social networks generate geo-social data indicative of the real-world situations, at least for the case of airport venues. While we report this finding with the reservation of the small number of airports and the short duration of data collection period for our second experiment, we believe that it is an important contribution, since as far as we know it is the first research effort on comparing geo-social data referencing the same places and originating from different social services. Such a comparison of different users' check-ins patterns and attitudes in different location sharing social services is a challenge that we intend to undertake in the future in large-scale experiments.

When analyzing data from each airport separately in our first experiment, we found a correlation between the number of check-ins and airport traffic for 5 out of 10 airports. On the other hand, the analysis showed a correlation between new users and passengers for 9 out of 10 airports. This observation is in line with the graphs in Figure 1, where one can see that the line corresponding to new users approximates more closely the line of airport traffic than that of check-ins. It seems, then, that the number of new users checking-in through Foursquare is a better indicator of airport traffic. This is quite important, in our opinion, since these numbers are hundreds of times lower than the recorded numbers of passengers.

Moreover, another observation from both experiments is that while the extracted Foursquare social data are representative of the general trend regarding airports' traffic, it seems that we cannot use them to directly compare traffic between airports. This observation confirms Rost et al. [17] findings, according to which rankings of airport venues in terms of check-ins does not keep up with rankings in terms of passengers. We confirm that apart from check-ins, this is also true for new users having checked-in at airports. For example, one can see in table 1 that in Moscow airport there were much more check-ins than in Heathrow, but less than half the passengers, while another such case is the comparison between Frankfurt and Amsterdam airports. A possible explanation for this remark could be based on the different location sharing culture of the user types visiting these airports (e.g. youth vs elder or business users). In addition to this, each airport follows a different social media marketing campaign, which may lead to a different user engagement type [22].

In the future, we intend to extend our research to include more airports, in order to ensure that our remarks apply for airports worldwide and are not limited to the European territory. Moreover, we intend to add more venue types in our analysis (e.g. train stations, ports etc.) in order to examine if social data are also representative of passenger traffic for these cases. Finally, we are also currently putting research effort to investigate possible correlation between weather conditions and check-ins in venues such as parks, beaches etc.

7 References

1. Bollen, J., Mao, H. and Zeng, X.: Twitter Mood Predicts the Stock Market. Springer Journal of Computational Science 2 (1), 1-8 (2011).
2. Sakaki, T., Okazaki, M. and Matsuo, Y.: Earthquake Shakes Twitter Users: Real-Time Event Detection by Social Sensors. In Proceedings of the 19th international conference on World Wide Web (WWW2010). ACM, New York, NY, 851-860 (2010).
3. Chakrabarti, D. and Punera, K.: Event Summarization Using Tweets. In Proceedings of the Fifth International Conference on Weblogs and Social Media (ICWSM 11). 66-73 (2011).
4. Li, Y., Steiner, M., Wang, L., Zhang, Z.L., Bao, J.: Exploring venue popularity in Foursquare. In Proceedings of IEEE INFOCOM Workshops (2013).
5. Komninos, A., Stefanis, V., Plessas, A. and Besharat, J.: Capturing Urban Dynamics with Scarce Check-In Data. IEEE Pervasive Computing, 12 (4), 20-28 (2013).
6. Cranshaw, J., Schwartz, R., Hong, J. and Sadeh N.: The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. In Proceedings of the Sixth International Conference on Weblogs and Social Media (ICWSM 12). 58-65 (2012).

7. Zhang, A., Noulas, A.,Scellato, S. and Mascolo, C.: Hoodsquare: Modeling and Recommending Neighborhoods in Location-Based Social Networks. In Proceedings of the 2013 International Conference on Social Computing (SOCIALCOM 13). IEEE Computer Society, 69-74 (2013).
8. Silva, T., Vaz de Melo, P., Almeida, J., Salles, J. and Loureiro, A.: A comparison of Foursquare and Instagram to the study of city dynamics and urban social behavior. In Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing (UrbComp 13). ACM, New York, NY, USA, article No 4 (2013).
9. Rösler, R. and Liebig, T.: Using Data from Location Based Social Networks for Urban Activity Clustering. Geographic Information Science at the Heart of Europe. Lecture Notes in Geoinformation and Cartography. Springer, 55-72 (2013).
10. Phithakkitnukoon, S. and Olivier, P.: Sensing Urban Social Geography. Using Online Social Networking Data. In Proceedings of the Fifth International Conference on Weblogs and Social Media (ICWSM 11). 36-39 (2011).
11. Silva, T., Vaz de Melo, P., Almeida J., Musolesi, M. and Loureiro, A.: You are What you Eat (and Drink): Identifying Cultural Boundaries by Analyzing Food & Drink Habits in Foursquare. In Proceedings of 8th AAAI Intl. Conf. on Weblogs and Social Media (ICWSM 14). (2014)
12. Georgiev, P., Noulas, A. and Mascolo, C.: Where Businesses Thrive: Predicting the Impact of the Olympic Games on Local Retailers through Location-based Services Data. In Proceedings of 8th AAAI Intl. Conf. on Weblogs and Social Media (ICWSM 14). (2014)
13. Karamshuk, D., Noulas, A., Scellato, S., Nicosia, V. and Mascolo, C.: Geo-spotting: mining online location-based services for optimal retail store placement. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD 13). ACM, New York, NY, USA, 793-801 (2013).
14. Cramer, H., Rost, M. and Holmquist, L.E.: Performing a check-in: emerging practices, norms and 'conflicts' in location-sharing using foursquare. In Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI 11). ACM, New York, NY, USA, 57-66 (2011).
15. Lindqvist, J., Cranshaw, J., Wiese, J., Hong, J. and Zimmerman, J.: I'm the mayor of my house: examining why people use foursquare - a social-driven location sharing application. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 11). ACM, New York, NY, USA, 2409-2418 (2011).
16. Cheng, Z., Caverlee, J., Lee, K. and Sui, D.: Exploring Millions of Footprints in Location Sharing Services. In Proceedings of the Fifth International Conference on Weblogs and Social Media (ICWSM 11). 81-88 (2011).
17. Rost, M., Barkhuus, L., Cramer, H. and Brown, B:. Representation and communication: challenges in interpreting large social media datasets. In Proceedings of the 2013 conference on Computer supported cooperative work (CSCW 13). ACM, New York, NY, USA, 357-362 (2013).
18. Wang, S., Stefanone, M.: Showing Off? Human Mobility and the Interplay of Traits, Self-Disclosure, and Facebook Check-Ins. Social Science Computer Review 31 (4), 437-457 (2013).
19. Wang, S.: “I Share, Therefore I Am”: Personality Traits, Life Satisfaction, and Facebook Check-Ins. Cyberpsychology, Behavior, and Social Networking 16 (12), 870-877 (2013).
20. Halpern, N.: Lapland's Airports: Facilitating the Development of International Tourism in a Peripheral Region. Scandinavian Journal of Hospitality and Tourism 8 (1), Taylor & Francis, 25-47 (2008).

21. Wang, M., Song, H.: Air Travel Demand Studies: A Review. Journal of China Tourism Research 6 (1), 29-49 (2010).
22. Vanauken, K.: Using social media to improve customer engagement and promote products and services. Journal of Airport Management 9 (1), 109-117 (2015).