

Location sharing services as sensors for analyzing airports' traffic

J. Garofalakis^{1,2}, I. Georgoulas¹, A. Komninos³, P. Ntentopoulos¹, A. Plessas^{1,2}
{garofala, georgoulas}@ceid.upatras.gr, andreas.komninos@strath.ac.uk, {ntentopoul, plessas}@ceid.upatras.gr

1. University of Patras
Patras, Greece
+30 2610 997525

2. CTI Diophantus
Patras, Greece
+30 2610 960430

3. University of Strathclyde
Glasgow, UK
+44 141 548 3092

ABSTRACT

Location sharing services can be seen as “social sensors” that provide data revealing real world interactions. In this paper, we examine the case of foursquare check-ins at airports and we show that these data can be indicative of the passengers' traffic, while their number is hundreds of times lower than the number of actual traffic observations.

Categories and Subject Descriptors

H.2.8 [Database Applications]: Data Mining

General Terms

Human Factors, Verification.

Keywords

Location sharing services, Foursquare, Check-ins, Ubiquitous social computing.

1. INTRODUCTION

During the last decade, the use of Location Based Social Networking (LBSN) services became widespread, with mobile devices playing an important role as they 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 geo-tag their posts, while Twitter permits users to geo-tag their tweets and Foursquare to check-in at places, rate them and leave tips for others.

As these platforms engage millions of users, there is a boom in the amount of geo-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. 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 LBSN services and the geo-annotation of information about a venue using spatial and temporal context. We are using Foursquare as a data source, as it is a popular location sharing service, with over 45 million users as of January 2014 [1].

More specifically, we focus on airports' check-ins and our

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Conference '10, Month 1–2, 2010, City, State, Country.

Copyright 2010 ACM 1-58113-000-0/00/0010 ...\$15.00.

purpose is to examine whether check-in data from Foursquare are adequate in order to get a representative picture of the airports' real passenger traffic. Airport traffic is an important financial factor for national economies and is often used as a tourism indicator [2]. We consider location sharing services API endpoints as sensors providing social data that offer useful insight about real passengers' traffic. In the framework of our research we collected and analyzed approximately 2.000.000 Foursquare check-ins from 10 airports in a period of 13 months.

2. DATA COLLECTION AND ANALYSIS

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 for 9 European airports (Table 1). Each airport API endpoint is seen as a social data sensor, used to sense foursquare users' social interactions. Our research took place for a period of 13 months, from May of 2013 until the end of May 2014. In total, there were 1.971.133 check-ins and the new users having checked-in during that period were 735.764.

Table 1. Summary of airport data

Airport	# Check-ins	# New Users	# Passengers
Athens Intl	96.100	26.635	14.255.499
Heathrow	358.155	131.142	79.040.930
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
Total	1.971.133	735.764	442.012.098

To compare with actual airports' traffic we used the statistics published at Anna Aero¹, which is a website dedicated to airline and airports network news and analysis. 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. Figures 1 and 2 depict graphs showing the actual number of passengers (orange line), the number of check-ins (blue line)

¹ <http://www.anna.aero>

and the number of new users having checked-in (grey line) on a monthly base for the airports Paris CDG and Heathrow. The left vertical axis corresponds to the number of passengers (blue line), while the right vertical axis corresponds to the social data gathered (orange and grey lines).

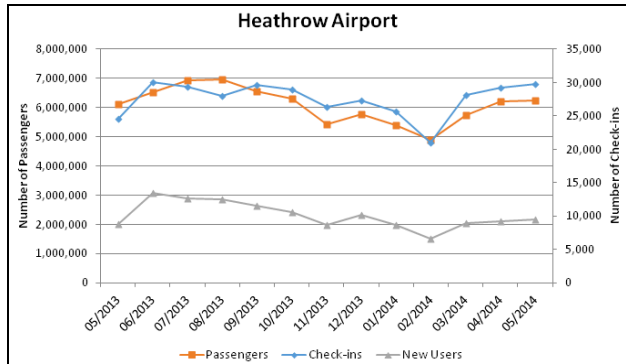


Figure 1. Data graph for Heathrow Airport

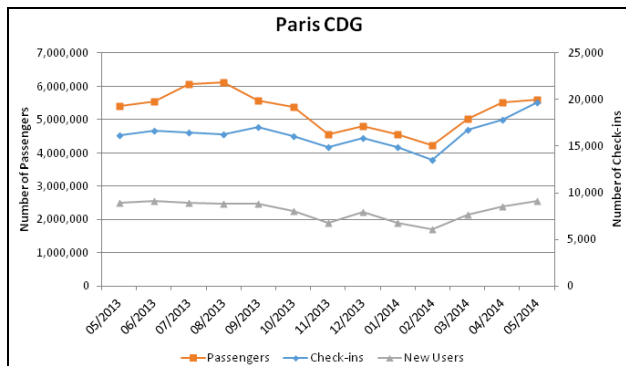


Figure 2. Data graph for Paris Charles de Gaulle Airport

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 ($r = 0.929$ $p < 0.01$) and Rome ($r = 0.576$ $p < 0.05$). 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.

3. 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. When examining total observations from all sensors - airports we found that the number of recorded Foursquare check-ins and 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. When analyzing data from each airport separately, 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. 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, while the extracted 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. [3] 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, 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.

In the future, 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 passengers' traffic for these cases. Finally, we will also put research effort to investigate possible correlation between weather conditions and check-ins in venues such as parks, beaches etc.

4. REFERENCES

- [1] Georgiev, P., Noulas, A. and Mascolo, C. 2014. 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* (Ann Arbor, MI, USA, June, 2-4, 2014). ICWSM 14.
- [2] Halpern, N. 2008. Lapland's Airports: Facilitating the Development of International Tourism in a Peripheral Region. *Scandinavian Journal of Hospitality and Tourism* 8, 1(May 2008), Taylor & Francis, 25-47. DOI= <http://dx.doi.org/10.1080/15022250801987762>
- [3] Rost, M., Barkhuus, L., Cramer, H. and Brown, B. 2013. 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. DOI= <http://dx.doi.org/10.1145/2441776.2441817>

