

You are where you eat:

Foursquare checkins as indicators of human mobility and behaviour

G.B. Colombo, M.J. Chorley, M.J. Williams, S.M. Allen, R.M. Whitaker
School of Computer Science & Informatics, Cardiff University
{g.colombo, m.j.chorley, m.j.williams, stuart.m.allen, r.m.whitaker}@cs.cardiff.ac.uk

Abstract—Location-sharing services such as Foursquare provide a rich source of information about the visits of users to locations. In the case of Foursquare, users voluntarily ‘check in’ to places they visit using a mobile application. An analysis of these data may reveal differences in users personality in terms of their mobility habits, preferred places, and action and location patterns. This knowledge about user behaviour can be used, in addition to information about their preferences, to improve current recommendation systems for mobile platforms.

Keywords-human mobility;location-based services;

I. INTRODUCTION

The widespread adoption of smart mobile devices has led to a growth in usage of social-driven location-sharing applications, such as Foursquare, Google Latitude, and Facebook. Users of these applications can record their visit to a location, referred to as ‘checking in’. This results in a spatio-temporal record of the users’ visits, allowing them to keep track of the places they have been to and when. In this work we focus on the location-based service Foursquare¹, for which we have collected a snapshot of checkin information during a defined period of time. In this service, places are represented by ‘venues’, which are described as any type of location a user can visit e.g. shops, restaurants, parks, railways and bus stations etc.

Foursquare also allows users to explicitly create social links (called ‘friendships’) to other users, resulting in a social graph of users. Users compare their checkins with those of their social contacts, to see where their friends are or have been. Further incentives to participation are provided in terms

of rewards (points, badges and ‘mayorships’ that users can gain by frequently checking in in a specific venue). The Foursquare community is rapidly expanding, reaching the milestone of 10,000,000 users in June 2011². This large community consists of a variety of user types who use the service for different purposes, from being motivated by the social or the gaming aspect to simply using the service as a log of all the places they have visited. A detailed analysis of these checkin data could reveal different users’ habits, preferred places, action and location patterns, thus providing deeper insights into human mobility and behaviour that could transcend the reductive context of a gaming application for smart-phones.

In this paper we analyse the checkins of users. In particular, we are interested in finding differences in the way individuals use the system in order to identify possible commonality of behaviours between different groups. Building a profile of the mobility behaviour of a specific user could, for example, help for the development of recommendation and suggestion systems that take into account user habits and personal preference in terms of mobility and action patterns, instead of solely relying on their preferences of interests. The latter is true in the case of many mobile recommendation applications, including Foursquare. These systems do not take into account any specific personal characteristics in terms of users spatial, temporal, and social mobility. This would help answer questions such as: does a given user prefer visiting the same places regularly or are they more inclined to visit a variety of venues; does the user tend to visit venues

¹<http://foursquare.com>

²blog.foursquare.com/2011/06/20/holysmokes10millionpeople

very frequently and go to places that are close in distance or do they rather prefer occasional visits and maybe travelling significant distances, or do they just prefer going where their friends go?

II. RELATED WORK

Understanding social structure and its relation to mobility is of great relevance for the performance of protocols and algorithms designed for mobile pervasive environments. Recent work suggests that humans follow regular and periodical patterns that can be guided by their social interactions. In fact, often human trajectories show a high degree of temporal and spatial regularity, with each individual presenting a characteristic travel distance and a significant probability to return to a restricted number of highly frequented locations [1], [2]. This regularity of movement has been proved to be strictly related to the complex community structure, capturing highly connected circles of friends, families or professional cliques in a social network. Early studies noticed that with mobile devices becoming location-aware, it is now possible to know when people are physically co-located and incorporate this information into social software in order to incentivise the formation of social groups [3]. These original findings could now be further improved by analysing repeatable mobility patterns that can be reproduced by crawling and analysing the data derived from location-based mobile services [4]. In [5] some of the most successful services such as Foursquare, Gowalla, and the (now defunct) Brightkite are considered and a statistical analysis of the formation of friends and the number of checkins and locations visited is conducted, finding in general that user activity decays faster than exponentially and so users add friends more quickly than they accumulate checkins and places. In [6] a similar analysis is conducted using Flickr⁴ data, investigating the number of co-occurrences in one or more locations (users visiting the same location) within different time intervals. This article also proposes a model to infer the probability of friendship based on the number and frequency of such contemporaneous

⁴<http://www.flickr.com/>

events. The authors previous work [7] emphasises that location-based services allow access to a rich store of information on user mobility and behaviour. Here Foursquare checkins in venues in different cities over a period of weeks are used to analyse preference of interests and similarity of users, suggesting that such information may be used to reveal user personality which can be used to improve analytical models for human mobility and create future recommendation systems.

III. RESULTS

Checkins were monitored in a set of venues in two UK cities: Cardiff (the capital city of Wales with a population of approximately 320,000) and Cambridge (a university town in the county of Cambridgeshire presenting about a third of the population of Cardiff with its 120,000 inhabitants). The data monitoring occurred from Monday 21st March 2011 until Friday 13th May. The size of the cities was reflected in the number of venues. Cardiff had in that period a total of 1,234 active venues (i.e. with at least one checkin) but only 852 were found for Cambridge. Complete statistics about the number of venues, users and checkins are summarised in Table I.

Table I
CHECKIN STATISTICS.

	Venues	Checkins	Users	Checkins PerVenue	Checkins PerUser
Car.	1,234	13,299	1,701	10.78	7.82
Cam.	852	6,464	1,196	7.59	5.40

A. User and venue checkins

The plots in Figures 1 and 2 show (in a logarithmic scale) the distributions of the number of user checkins and the popularity of venues. The majority of users made very few checkins whereas only a limited number (about 10 users for the Cambridge data and only around 40 users in Cardiff) performed more than a hundred. In particular, the number of users that made only one checkin in the monitoring period was 43% in Cambridge and 31% in Cardiff. Conversely, small percentages of very active users were responsible for the majority of the

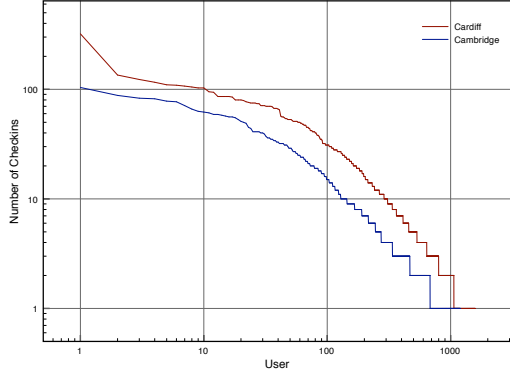


Figure 1. User checkin counts - Cardiff and Cambridge

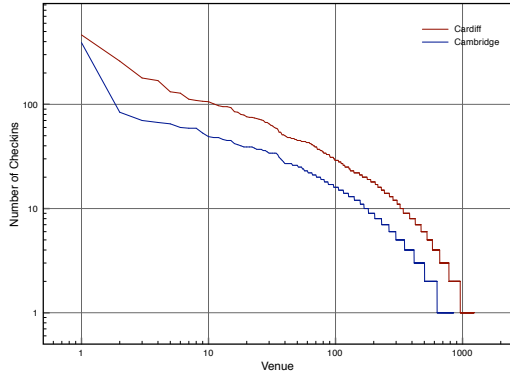


Figure 2. Venues checkin counts - Cardiff and Cambridge

checkins (in Cardiff 1% of users performed 15% of the total amount of checkins). The same pattern is seen by the curve showing the number of checkins per venues (Fig. 2), with many venues showing low popularity while only a limited number of them present a significant number of checkins (> 50). No particular difference in the behaviour can be observed between the two cities, although Cambridge is showing a proportional reduction in the number of users and venues checkins in comparison to the Cardiff results.

B. Time and distance between checkins

In this section we take a closer look at the behaviour of users in order to identify patterns of behaviour between consecutive checkins in terms

of intervals of time and distance. For each individual user we select all pairs of consecutive checkins that are made within a given time threshold, and look at the exact time between the checkins and the distance between the venues checked in to (we consider time intervals between three hours and one month). Subsequently, we have computed and plotted the average time and distance between each pair of consecutive checkins among each group of selection. Figures 3 and 4 present the Cardiff results showing the cumulative distribution function for average time and distance intervals between pairs of checkins for each of the different thresholds considered. Looking at the spatial lengths (Figure 3) between checkins we can find that users tend to perform checkins within short distances (around 1km) with a significant probability of ≈ 0.6 . The Cambridge findings (here omitted due to space restrictions) showed a similar, yet slightly higher probability value.

In addition, the slope of the curve showing the distribution for checkins within a given distance also presents some differences: the curve appears to be almost linear with a sharp increase in its first part whereas for longer distances it becomes smoother with significant smaller gradients, thus showing that checkins separated by significant space are far more rare. The behaviour described above does not appear to be heavily influenced by the time threshold of the checkins considered.

If we consider the similar plot showing the percentage of users in relation to time interval between different checkins (Figure 4) we can draw similar conclusions: users check-in consecutively within a relatively short time interval with significant probability (≈ 0.5 if we consider all pairs of checkins that occurred within one day, higher for the other cases). However, because of filtering this effect also depends on the specific time threshold between checkins (we here show intervals between three hours and one day).

C. Patterns within checkins

The aim of this short section is the identification of patterns of repeated behaviour among the checkins of individual users. Users appear to

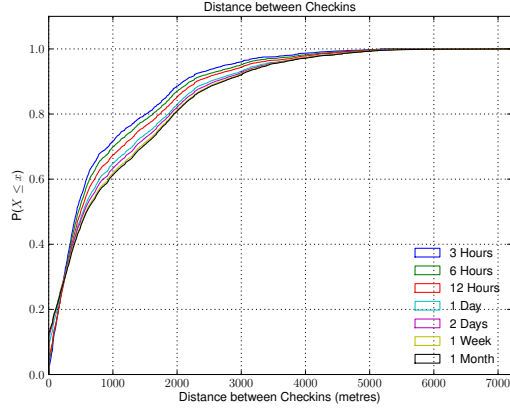


Figure 3. Cumulative distribution function for distance between checkins - Cardiff

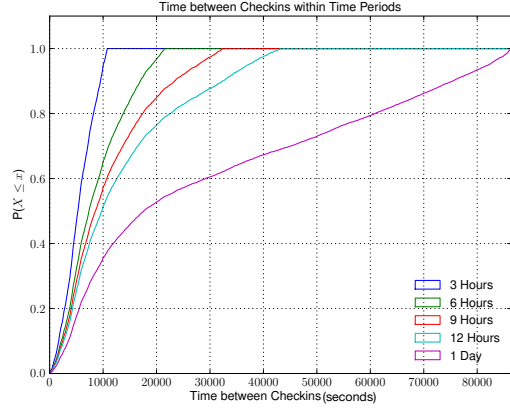


Figure 4. Cumulative distribution function for time between checkins - Cardiff

often repeatedly check in within the same venues, in a similar or the same order, and it is these repetitions of behaviour that we are interested in. So, for instance, if a user checks in to the ‘bank’ and then the ‘supermarket’ twelve times during the monitoring period, we have a tuple of two venues that is repeated twelve times. (This repeating behaviour has an extreme case: that of users checking-in exclusively in a very limited number of venues, i.e. checking in only at home and at work for instance). We examine these patterns by firstly considering ordered unique tuples representing a consecutive sequence of checkins

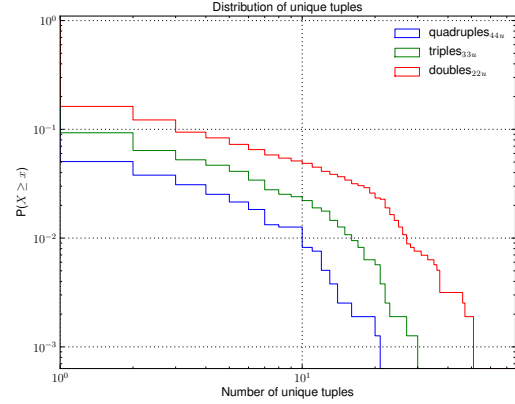


Figure 5. CCDF for number of unique ordered sequences of checkins (two, three and four venues) - Cardiff

in distinct venues. Figure 5 shows (in logarithmic scale) the complementary cumulative distribution of the number of unique distinct tuples within the whole population (we considered ordered tuples of two, three and four venues). We can see how the probability of having even a small number of repeated pairs of consecutive ordered checkins is rather low (on the order of a few percent) and sharply decreases when considering tuples of three and four venues or a greater number of tuples. However, these probabilities rapidly increase if we consider a fuzzy representation of tuples allowing sequences of checkins to occur in any order (so for example allowing the ‘supermarket’ followed by the ‘bank’ to appear the same as the ‘bank’ followed by the ‘supermarket’) and also allow venues to be spread among a larger number of checkins (so allowing the ‘bank’ then the ‘barbers’ then the ‘supermarket’ to also count as the same). Figure 6 shows the complementary cumulative probability distribution for the number of unordered checkins for pairs of distinct venues (doubles) when the fuzzy representation is defined over a sequence of up to ten consecutive checkins. We can see from the plots that the probability of having up to fifty repeated pairs of checkins for example rises from about 0.003%, when considering unordered pairs, to about 0.3% if we allow the pairs to be spread among five and ten consecutive checkins.

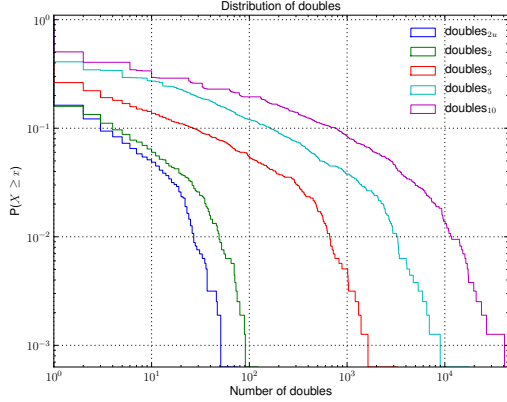


Figure 6. CCDF for number of unique unordered pairs of checkins with fuzzy representation - Cardiff

D. Co-located Checkins

If two users checkin in the same venue within a given time window, we consider this to be a co-located checkin. In this final section we consider these co-located checkins, with the aim of discovering how they are distributed throughout the dataset and what trends or patterns they reveal. In addition we look at how these statistics may be different when considering co-located checkins between generic pairs of users or between users that are linked with a friendship link in Foursquare: do users tend to check in at the same venues as their friends? We also consider the time difference between these checkins: do friends tend to go to the same venues at the same time or within short intervals?

Number of users with co-locations. We first look at the number of co-located checkins present in the data for each user. Figure 7 shows the cumulative distribution function for the number of users having any co-located checkins, considering co-located checkins with all users (similar results are obtained if we look at friends only). We see that while a few users have co-located checkins with a large number of other users (up to 350 unique user co-locations in Cardiff and 160 in Cambridge), the majority of users have fewer unique users with which they have co-located. In Cardiff, the probability that an individual user has co-located

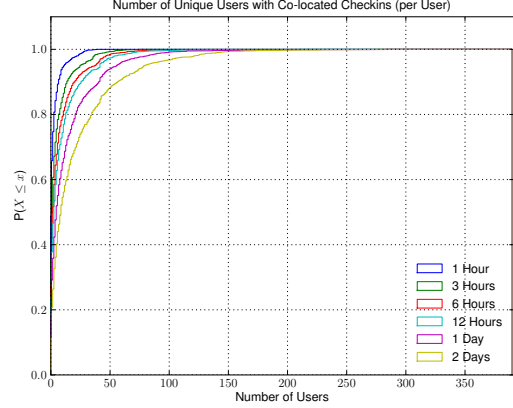


Figure 7. Cumulative distribution function for number of total users with co-located checkins - Cardiff

with 50 users or fewer is above 0.8 for all time windows. Despite the high probability for small numbers of co-located checkins, the results show that users in fact do have some co-located checkins, and that they have co-located checkins with both friends and other users.

Time between co-locations. We now look at the time between checkins at the same location, again examining co-locations within given time windows. Figure 8 shows the cumulative distribution functions for the times between co-located checkins in Cardiff within 1 hour and 3 hours, for co-locations with all users and co-locations with friends only. These results show a peculiarity of behaviour when we consider users that are also friends. In fact, these plots illustrate that if a co-located checkin occurs within a short time frame of the original checkin, the probability that the co-located checkin is with a friend is much higher than the probability that it is with any other user. For instance in Cardiff the probability that a co-located checkin with a friend occurs within 2000 seconds (33 minutes) is 0.8, while for all users the probability is 0.7, if the checkins examined are restricted to those occurring within an hour of the original. If this time window is increased to 3 hours, any co-located checkin that occurs has a probability of 0.55 of occurring between friends within 2000 seconds, but a probability of only 0.35 of occurring

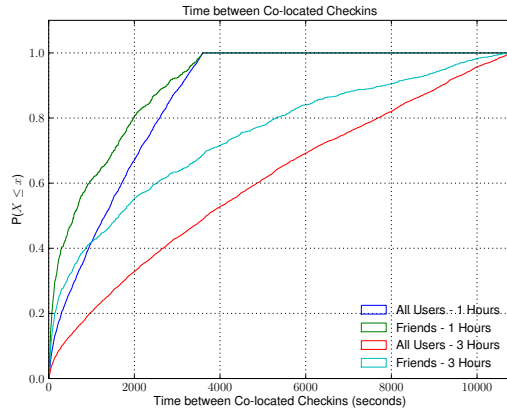


Figure 8. Time between Co-located Checkins, for co-locations occurring within 1 and 3 hours, between friends only and all users - Cardiff

within 2000 seconds with any user from all other users. This suggests that co-locations with friends tend to occur within shorter time frames than co-located checkins with other users, which may stem from users visiting locations with their friends and checking in at the same time.

IV. CONCLUSION

In this article we have investigated and discussed how users of location based mobile services present different behaviour in terms of their mobility patterns, using Foursquare checkins as a test-bed. In particular, we have observed that a significant percentage of users tend to make frequent and regular visit to a limited number of venues, as well as showing an inclination for visiting places frequented by their social friends. Patterns of repeated behaviour can be observed, and the use of specific time and distance intervals between checkins can be identified, implying relevant differences in their mobility behaviour. Future planned work will consider closely the relation and interaction between the preference of interests of the users and their behavioural characteristics. If a user checks in regularly in a number of places does this also reflect his/her preference in terms of types of venues? And does he/her tend to frequent places where his/her social friends go, even if

they are places of different types than those they favour? The further knowledge that can be derived from combining information about the mobility behavioural characteristics of individuals with that of their interest preferences can be of fundamental importance for building more realistic and effective personal profiles of users. Thus could be used for the development of personalised recommender application systems for mobile platforms.

ACKNOWLEDGEMENTS

This research was funded by RECOGNITION grant 257756, an EC - FP7 Future Emerging Technologies project concerning Self-Awareness in Autonomic Systems.

REFERENCES

- [1] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, pp. 779–782, June 2008.
- [2] M. Zignani and S. Gaito, "Extracting human mobility patterns from gps-based traces," in *Proceedings of the 2nd. IFIP Wireless Days 2009*, (Paris, France), IEEE, December 2009.
- [3] S. Counts and J. Geraci, "Incorporating physical co-presence at events into digital social networking," in *Proc. of CHI '05*, (New York, NY, USA), pp. 1308–1311, ACM, 2005.
- [4] J. Cranshaw, E. Toch, J. Hong, A. Kittur, and N. Sadeh, "Bridging the gap between physical location and online social networks," in *Proceedings of Ubicomp '10*, (New York, NY, USA), pp. 119–128, ACM, 2010.
- [5] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil, "Measuring user activity on an online location-based social network," in *Proc. of NetSciCom 2011*, (Shanghai, China), 15, April, 2011.
- [6] D. J. Crandall and et al., "Inferring social ties from geographic coincidences," *Proc. of the National Academy of Sciences*, vol. 107, Dec. 2010.
- [7] M. Chorley and et al., "Checking out checking in: observations on foursquare usage patterns," in *Finding NEMO workshop*, (Athens, Greece), 9, September, 2011.