

Checking out checking in: observations on Foursquare usage patterns

Martin Chorley, Gualtiero Colombo, Matthew Williams,
Stuart Allen, Roger Whitaker

School of Computer Science & Informatics, Cardiff University, UK, CF24 3AA
{m.j.chorley, g.colombo, m.j.williams, stuart.m.allen,
r.m.whitaker}@cs.cardiff.ac.uk

Abstract. Location-based services allow access to a rich store of information on user mobility and behaviour. Such information may reveal indicators towards user personality which can be used to improve models of human mobility and create future recommendation systems. Analysis of Foursquare checkins in venues in two different UK cities over a period of weeks allows us to draw conclusions about user behaviour, checkin patterns and similarity of users.

1 Introduction

Mobile location services such as Foursquare and Gowalla allow users to record spatial and temporal data, keeping a log of the places they have been to and when. These services use social and game features to encourage users to continue to participate, producing an ongoing record of spatio-temporal data that can be mined and analysed to provide a rich picture of user mobility and behaviour. This picture may reveal indicators to user personality and their disposition towards experiences as suggested by the OCEAN [2] personality model, information which could be used to inform the creation and development of mobility models and recommendation systems.

In these location based mobile services, places are represented by ‘venues’, which are described as any place a user can visit, such as a shop, a restaurant, a park, a road and so on. Users ‘check in’ to venues, marking that they have visited a particular venue at a particular time. Venues can be created by users if they are not already present in the service, they have location data describing their position and may also have some user provided data associated with them. This user created data may at the simplest level be a categorisation of the venue into one of several pre-defined categories, or it may contain more complex information such as user tips for the venue, or photographs taken at the venue.

There are two main aims behind the work presented in this paper. It should be possible to use mobility and social data to identify how well checkins reflect predicted user behaviour from social based mobility models. Such mobility models often include drivers that attempt to force social links and groupings to influence individual mobility. Analysis of user provided checkin data should

allow comparison to see how mobility in real world is affected by social links and grouping.

Secondly, we aim to use social, temporal and mobility data to inform the creation of future recommendation systems. We envisage a system where users are recommended content and activities based not just on prior behaviour or the patterns of their friends, but on predicted future behaviour. Such a system should be able to recommend activities or content to users without the users necessarily being aware of their desire for such items. In order to create such a system we must first answer some fundamental questions about how personality and behaviour relate to mobility and social ties.

By monitoring a number of venues across two different cities in the UK, we have captured a snapshot of Foursquare checkin information. This checkin information has been analysed to draw conclusions about user behaviour and the patterns that may be found within. In this work we focus primarily on the analysis of Foursquare data relating to categories of venues, and what this can tell us about user behaviour and similarity.

We analyse the distribution of checkins over users and venues to identify any patterns, as well as examining the geographic distribution of checkins over the two cities. We identify different patterns of behaviour in terms of the checkins made in different categories of venue, and the temporal patterns in the same. Finally we use individual category checkins to compute a measure of similarity between users to assess the level of similarity between friends.

2 Related Work

Social aspects have been increasingly used in recent years to develop computational mobility models. For example, [11] links human mobility with social network theory and uses synthetic mobility traces as good approximations of human movement patterns, while [5] stresses the importance of knowing and understanding the social structure beside the use of realistic pairwise statistics produced by analysing contact and inter-contact time distributions. 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. In this context, the social and spatial temporal dimensions dynamically evolve and interact thanks to the frequent changes in the activity and communication patterns of individuals [4, 13].

The integration of spatial information with social networks has been largely investigated during the last decade by the creation of prototype social networking systems based on physical co-presence. [8] discusses how location-based social networks services such as Foursquare¹, Gowalla², and Brightkite³, present a number of innovative areas of functionality, either not yet available or fully exploited in other commercial mobile social networking platforms. [10] focuses

¹ <https://foursquare.com/>

² <http://gowalla.com/>

³ <http://brightkite.com/>

on how users connect with friends in relation to how they check-in at different places, finding in general that user activity decays faster than exponentially and so users add friends more quickly than they accumulate check-ins. Similarly, [15] analyses user activities and social graph of a mobile social network using data obtained from Nokia Friend View⁴. Results show that user activities are highly correlated and in contrast to from online social networks, where users have a large number of friends but loose weakly-tied subgroups, users of these location-based services tend to have closer, strongly-tied cohesive subgroups. [7] examines the motivations of users of these services, focusing in particular on Foursquare. These include elements of fun, exploration, and coordination with the most popular use being signalling availability to friends, using the history of visited places as a form of presentation of self, and coordinating with the social group.

One of the most distinctive features of these services in comparison with traditional on line social networks is that they allow users to know when their friends are around and provide the ability to meet new people who share their own interests [1]. This could be used for the development of improved recommendation systems in order to suggest places and activities to do at a particular location and time. In fact, current recommender systems are still primarily based on collaborative filtering techniques that measure the similarity between users by considering the rating of common items. Issues such as the social influence on physical proximity in friendship formation can be further introduced to improve these systems beside other factors, such as the similarity of interest and social friendships. [6] proposes three different layers for considering users similarity: a social graph layer (using friendship links from Brightkite), a tags layer based on interests similarity (defined by user profiles), and a spatial layer based on the location based graph of each users updates. [12] uses semantics to propose a clustering method for both the users of location-based services and the corresponding geographic areas. The proposed method is primarily based on the identification of user communities that visit similar categories of places, and on the comparison of urban neighbourhoods within and across cities.

Finally, [9] uses Foursquare as a test bed to conduct a socio-spatial analysis concluding that there exist strong social and geospatial ties among users and their visited locations in the network. The authors propose a recommendation system that, although still based on collaborative filtering with rating, also considers a geospatial variant based on the idea that friends who are close in distance (nearby friends) tend to share more commonly visited locations. The same findings have been used in [3] to propose a mixed-model recommender for leisure activities that uses the well known collaborative filtering model as well as additional ones, such as those based on geographical distance (utility decays with the distance between users) and social learning. An example is the so-called learned preferences model, based on ideas such as that frequent visits to specific categories of places (mexican restaurants in the example) would automatically infer a preference for these locations.

⁴ <http://betalabs.nokia.com/apps/nokia-friend-view/>

3 Foursquare Data

Monitoring of Foursquare checkin data took place in two cities, Cardiff and Cambridge, from Monday 21st March 2011 at 08:49 GMT until Friday 13th May at 11:20 BST. Cardiff is the capital city of Wales with a population of approximately 320,000. It contains campuses for three Universities, with the largest, Cardiff University, having approximately 34,000 students and staff. Cambridge is a relatively small city to the north-east of London, with a population of approximately 130,000. The city contains the University of Cambridge, which has approximately 28,000 students and staff.

An initial search for venues in both Cardiff and Cambridge revealed several thousand registered venues across both cities. A subset of these venues were monitored for checkins, with venues chosen based on geographical location. Each venue was polled for checkins in turn, with the monitoring process cycling through the venue list for each city approximately once an hour, with both cities being monitored in parallel. Although monitoring was carried out continually, there are some gaps in coverage either due to Foursquare API down time or other failures of the monitoring process. Over the monitoring period, checkins were observed in only a proportion of the monitored venues. The numbers of venues found in each city, the number monitored for checkins and the number in which checkins were observed are given in Table 1.

Table 1. Discovered and monitored venues for each city.

	All Venues	Venues Monitored	Active Venues
Cardiff	4,314	2,219	1,234
Cambridge	3,008	1,884	852

Venues in the Foursquare database may be categorised under one or more descriptive categories. The types and names of categories themselves are provided by Foursquare, but the categories applied to venues are user suggested and community moderated, so the venue categories become a restricted form of folksonomy. Categories themselves are arranged in a hierarchical tree, with several top-level parent categories being separated into further sub-categories, and so on.

4 Results and Data Analysis

The data collected during the monitoring period showed Cardiff to be considerably more active, both in absolute terms and checkins per venue or per user (see Table 2).

Table 2. Checkin statistics.

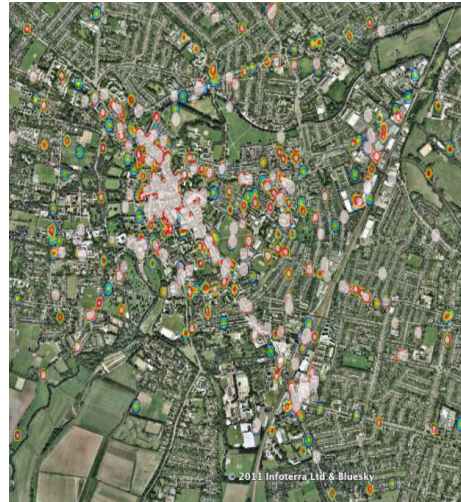
	Active Venues	Checkins	Users	Checkins per Venue	Checkins per User
Cardiff	1,234	13,299	1,701	10.78	7.82
Cambridge	852	6,464	1,196	7.59	5.40

4.1 Geographic Distribution of Checkins

Heatmaps showing the distribution of the checkins observed in the monitoring period over the two cities are given in Figures 1(a) and 1(b). As can be seen, the heavy concentrations of checkins tend to match up to features such as roads and dense areas. Checkins are concentrated within the centre of both cities as may be expected.



(a) Cardiff



(b) Cambridge

Fig. 1. Heatmaps showing the geographic distribution and intensity of checkins in each city.

4.2 User and Venue Checkin Distribution

Figures 2(a) and 2(b) show frequency distributions for the number of user checkins and the popularity of venues. We can see that most users make very few checkins. Indeed, we found that the number of users having made only one checkin in the period was 31% in Cardiff and 43% in Cambridge. While the number of very active users is small, these users account for many of the checkins in their cities. For example, in Cardiff the most-active 1% of users are responsible for

15% of the checkins. A similar relationship can be observed with the popularity of venues (Figure 2(b)); that is, each city contains a large number of less popular venues and a small group of very popular venues.

The approximate straight-line shape in Figure 2(a) (shown on a log-log scale) hints at a power-law relationship in the distribution of users checkin activity. The shape is less pronounced in the case of venue popularity (Figure 2(b)), and in both figures there is a great deal of fluctuation in the tail of the distribution.

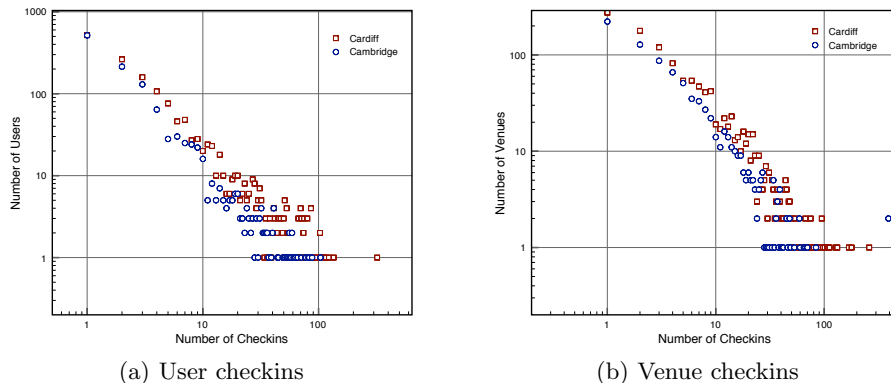
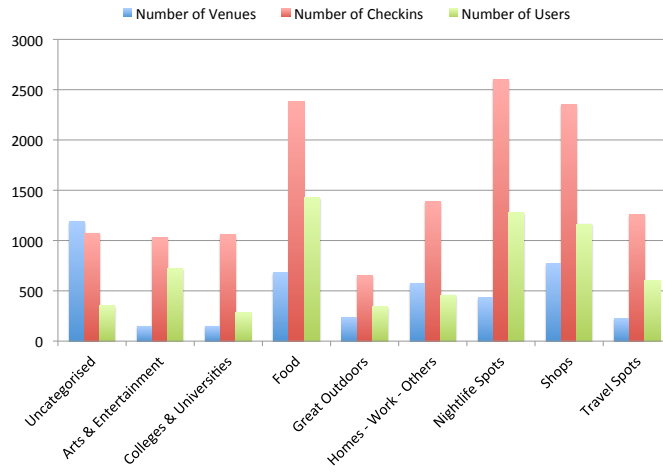


Fig. 2. User and venue checkin counts for both Cardiff and Cambridge. Both plots are on a log-log scale.

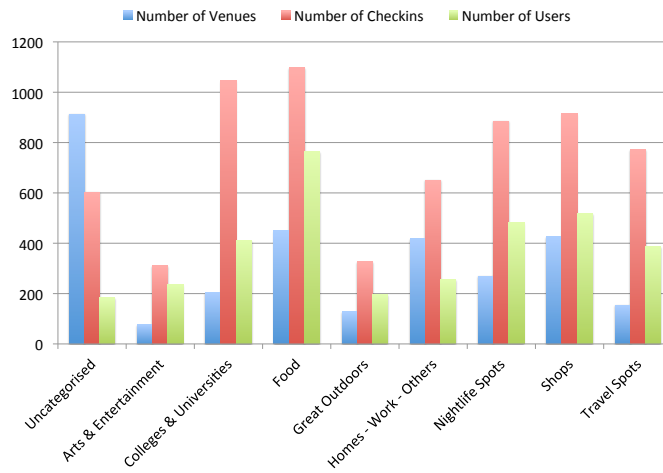
4.3 Checkin Categories

The categories associated with venues make it possible to study the structure of a city and the types of venue in which people tend to make checkins and record their presence, so possibly revealing something about user behaviour in terms of the types of venue visited. Foursquare’s category hierarchy consists of eight top-level categories, each of which can have a number of sub-categories and sub-sub-categories. These sub-categories are too many and too specific to easily examine, so we have grouped venues by their top-level categories, as shown in Figure 3.

It is clear that the three data points for each category give each city a different fingerprint of activity within the city. It may be possible to classify cities based on these values or the ratios between them, and this data may reveal individual characteristics of the city. For instance, Cambridge has almost as many checkins in the *College & Universities* category as the *Food* category, while in Cardiff checkins in the *Food* category outnumber those in the *College & Universities* category by almost 2 to 1. Looking at ratios between the number of venues and the number of checkins reveals further details, for instance in Cambridge the ratio of venues to checkins in the *Nightlife Spots* category is approximately 3 : 1, while in Cardiff this ratio is approximately 6 : 1.



(a) Cardiff



(b) Cambridge

Fig. 3. Checkin statistics for venues grouped by top-level category. In particular: the number of active venues in each category, the number of checkins in the venues in each category, and the number of unique visitors to venues in each category.

4.4 Daily Check-in Behaviour

The per-hour distribution of checkins offers insights into how the behaviour of Foursquare users is influenced by the time of day. The plots in Figure 4 show the total number of checkins during each hour of the day, along with the categories of venues that were visited. To limit potential bias due to interruptions in data collection, a 21-day period with the fewest outages was selected. The number of checkins recorded in this period were 5,226 and 2,609 for Cardiff and Cambridge, respectively. Checkins occurring within the same hour of the day were aggregated to produce the figures. We also differentiate between checkins on weekdays (i.e., Monday to Friday) and checkins on weekends (i.e., Saturday and Sunday), as we expect work patterns to have a strong influence on weekday check-in behaviour.

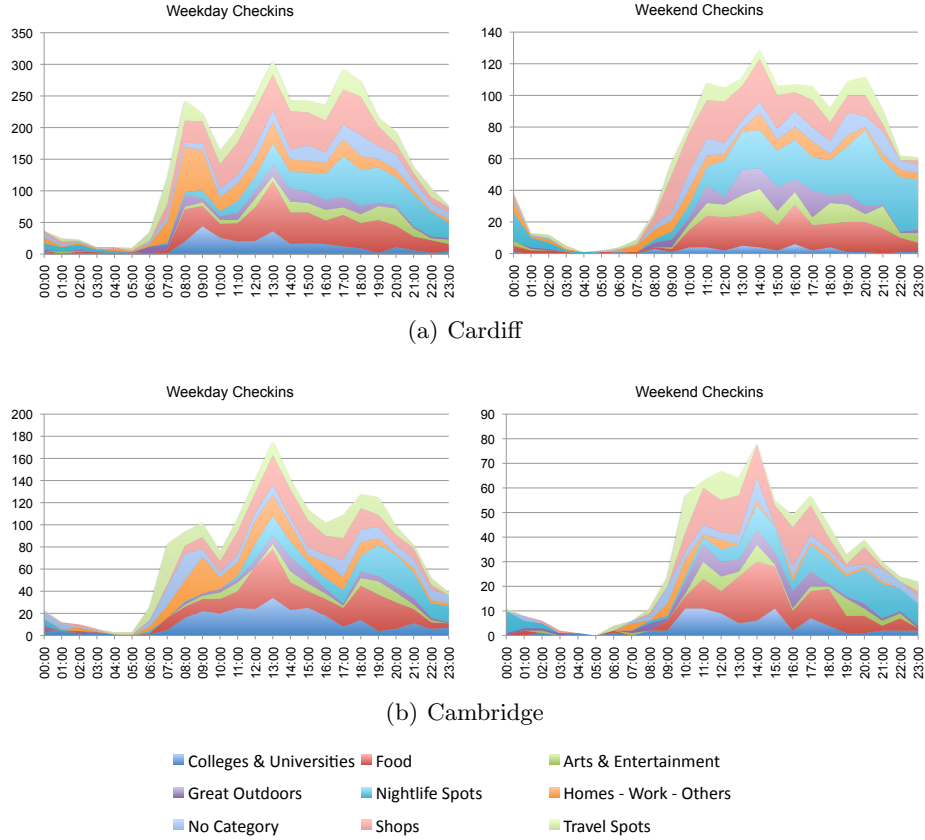


Fig. 4. Number of checkins per hour of the day in Cardiff and Cambridge. The contribution of each venue category to the total checkins for each hour is indicated by colour.

We found that there was a slight increase in the number of checkins per day on weekends. Cardiff averaged $\mu_1 = 246.2$ checkins per day on weekdays (with $\sigma_1 = 57.7$) and $\mu_2 = 255.5$ checkins per day on weekends (with $\sigma_2 = 73.6$). A comparable increase was found in Cambridge, which averaged $\mu_1 = 123.5$ checkins per day on weekdays (with $\sigma_1 = 29.3$) and $\mu_2 = 126.0$ checkins per day on weekends (with $\sigma_2 = 51.6$). Within each city we tested whether the difference in the weekday and weekend means were statistically significant by applying the two-sample T -test with the null hypothesis $H_0 : \mu_1 = \mu_2$. The resulting p -values lead us to accept H_0 for both cities ($p = 0.79$ for Cardiff and $p = 0.91$ for Cambridge) at the 0.05 significance level. Therefore we conclude that there is no statistically significant change in the number checkins on weekdays versus checkins on weekends.

Both cities also appear to have the same ratio of weekday checkins to weekend checkins. The fraction of weekday checkins is $f_1 = 0.704$ for Cardiff and $f_2 = 0.710$ for Cambridge. We tested the statistical significance of this similarity between Cardiff and Cambridge using the two-proportion Z -test with the null hypothesis $H_0 : f_1 = f_2$. The test gave a p -value of 0.55, indicating that there is no statistically significant difference between the two ratios.

Despite the negligible difference in the number of checkins per day, the distribution of checkins on weekdays and weekends are noticeably different. In the case of weekdays, we find that both cities have three bursts of checkins that occur during the day. In Cardiff there is a morning burst from 07:00 to 10:00, an early-afternoon burst from 12:00 to 14:00, and an early-evening burst from 17:00 to 19:00. Cambridge checkins follow a similar pattern, but with the morning and early-evening peaks being smaller relative to the early-afternoon peak.

The morning burst on weekdays is likely due to users waking up and checking in to venues as they travel to work. Indeed, in the 07:00 to 10:00 burst the four most-popular venue categories for Cardiff are *Home - Work - Others* (29%), *Food* (15%), *Shops* (15%), and *Travel Spots* (15%). It is not surprising that the early afternoon (12:00 to 14:00) checkins are predominantly at *Food* (24%) and *Shops* (21%) venues. The percentage of *Home - Work - Others* checkins in the early afternoon (11%) is much smaller than in the morning burst (29%), which may indicate that either many individuals do not leave work for lunch or neglect to check-in on their return to work. A notable change in the early evening checkins is an increase in the percentage of *Nightlife Spots* checkins (21% for 17:00 to 19:00). This category becomes increasingly popular as the evening progresses.

Although we can observe a three-burst pattern for both Cardiff and Cambridge during weekdays, there is no strong similarity between their weekend checkin patterns. The number of checkins per hour in Cardiff does not drastically change between 11:00 and 21:00. Cambridge has a high rate of checking in between 10:00 and 15:00; however, unlike Cardiff, this rate is not sustained into the evening. These results indicate that weekday user behaviour is predominantly driven by routine, whereas there is scope for more variation and less predictability in weekend patterns.

4.5 User Similarity

The categorisation of venues allows us to look at user checkins and develop a measure of interest for each category over the monitoring period. If we take the number of checkins n in a category m over all categories M , we describe a user x 's interest profile as a vector $\mathbf{C}^x$, where $C_m^x = n$. Non-zero entries in $\mathbf{C}^x$ for a category m relate to the users relative interest in that category, while a zero entry describes that the user has no relative interest in that category. This vector can then be normalised such that $\sum_{m=1}^M C_m^x = 1$.

Using these interest profiles for each user we can then calculate the proportional similarity $PS(x, y)$ [14] between two users x and y with interest profiles $\mathbf{C}^x$ and $\mathbf{C}^y$ as:

$$PS(x, y) = 1 - \frac{1}{2} \sum_{m=1}^M |C_m^x - C_m^y|. \quad (1)$$

It should be noted that this similarity may change over time as and when (or if) users change their preferences or Foursquare checkin behaviour.

In Figures 5(a) and 5(b) we compare the complementary cumulative distribution of proportional similarity between all pairs of users and between friends only, using the category checkin count interest profiles.

From the distributions it is clear that the probability of having a PS of a certain value or higher is much higher between friends than between all users. For instance, in Cambridge the probability of having a Proportional Similarity of more than 0.1 is about 0.5 for friends, but less than 0.2 for all users. Cardiff shows similar results. This suggests that pairs of friends have a much higher average proportional similarity than pairs of random users. The fact that friends have some level of similarity indicates that users share more similarity in checkins with their friends than with all users.

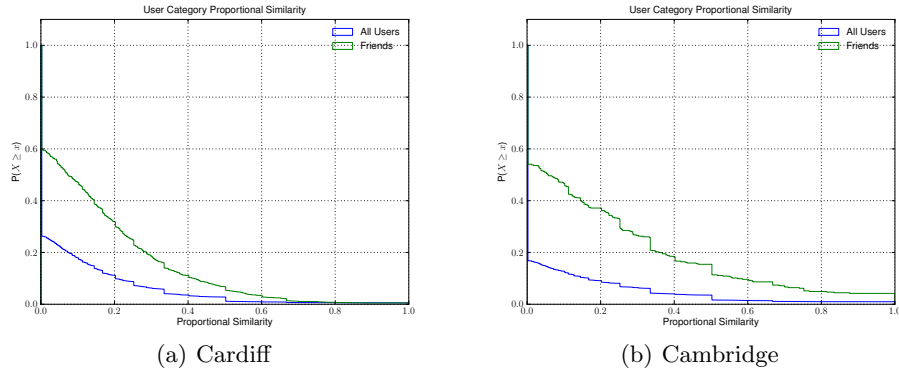


Fig. 5. Proportional Similarity between all users and friends only using category checkin counts to define interests.

5 Conclusions and Future Work

From analysis of Foursquare checkins in two UK cities we have identified several patterns. In general most users have relatively few checkins over a given time period, while a few users have very many checkins. In Cardiff 31% of users have only 1 checkin, while the top 1% of most active users were responsible for 15% of the checkins. The checkin density in each city is closely related to the centre of the cities.

Each city has different behaviour patterns of users in terms of checkins and number of users in top-level categories, and a physical fingerprint comprised of the number of venues in each category. These patterns may reveal information about the city or the users within the city.

A fluctuation in the number of checkins over a daily period has been observed, with weekdays and weekends having different patterns, based on user behaviour. This information may be used to improve future human mobility models.

Similarity of users can be calculated using the number of checkins in each category, and it has been shown that users are more likely to be similar to their friends than other users.

5.1 Future Work:

We intend to expand analysis to further regions to allow further comparison of different cities. Although visual inspection suggests a potential power law relationship in the distribution between number of checkins for venues and users, we intend to carry out further investigation to validate this.

The weekday patterns identified by the analysis seems to be driven by routine, however our analysis is carried out on aggregate behaviour, individual behaviour of routine and regularity also requires study. The diversity of checkin behaviour across venues and categories will be investigated, and checkin sequences analysed for predictability.

Co-located checkins, where two users register a checkin at the same venue at the same time can be mined from the data to identify the role that social relationships have on mobility, to discover if friends check in together frequently, or to discover at what types of venues co-located checkins occur. Aggregate checkin behaviour over many factors (number of co-located checkins, categories of checkins, diversity of venue choice and so on) may allow classification of users or personalities which can aid the development of future recommendation systems.

6 Acknowledgements

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

References

1. Backstrom, L., Sun, E., Marlow, C.: Find me if you can: improving geographical prediction with social and spatial proximity. In: Proceedings of the 19th international conference on World wide web. pp. 61–70. ACM, New York, NY, USA (2010)
2. Digman, J.: Personality structure: Emergence of the five-factor model. *Annual review of psychology* 41(1), 417–440 (1990)
3. Ducheneaut, N., Partridge, K., Huang, Q., Price, B., Roberts, M., Ch, H., Bellotti, V., Begole, B.: Collaborative filtering is not enough? experiments with a mixed-model recommender for leisure activities. *Lecture Notes in Computer Science* 5535, 295–306 (2009)
4. Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.L.: Understanding individual human mobility patterns. *Nature* 453(7196), 779–782 (Jun 2008)
5. Hossmann, T., Spyropoulos, T., Legendre, F.: A complex network analysis of human mobility. In: Third International Workshop on Network Science for Communication Networks (NetSciCom 11). IEEE, Shanghai, China (Apr 2011)
6. Li, N., Chen, G.: Multi-Layer Friendship Modeling for Location-Based Mobile Social Networks. In: *MobiQuitous '09: Proceedings of Mobiquitous 2009*. pp. 1–10 (Jul 2009)
7. Lindqvist, J., Cranshaw, J., Wiese, J., Hong, J., Zimmerman, J.: I'm the mayor of my house: examining why people use foursquare - a social-driven location sharing application. In: Proceedings of the 2011 annual conference on Human factors in computing systems. pp. 2409–2418. ACM, New York, NY, USA (2011)
8. Lubke, R., Schuster, D., Schill, A.: Mobilisgroups: Location-based group formation in mobile social networks. In: 2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops). pp. 502–507. IEEE, Seattle, WA, USA (21–25 March 2011)
9. Mao, Y., Peifeng, Y., Wang, L.C.: Location recommendation for location-based social networks. In: Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems. pp. 458–461. GIS '10, ACM, New York, NY, USA (2010)
10. Mascolo, C., Scellato, S.: Measuring user activity on an online location-based social network. In: Proceeding of 3rd Workshop Social Mobile Web (SMW'11 (2011)
11. Musolesi, M., Mascolo, C.: Designing mobility models based on social network theory. *ACM SIGMOBILE Mobile Computing and Communication Review* 11, 59–70 (2007)
12. Noulas, A., Scellato, S., Mascolo, C., Pontil, M.: Exploiting semantic annotations for clustering geographic areas and users in location-based social networks. In: Proceedings of Third International Workshop on Network Science for Communication Networks (NetSciCom 2011) (2011)
13. Palla, G., Işzıl Barabási, A., Vicsek, T., Hungary, B.: Quantifying social group evolution. *Nature* 446, 2007 (2007)
14. Vegelius, J., Janson, S., Johansson, F.: Measures of similarity between distributions. *Quality and Quantity* 20(4), 437–441 (December 1986)
15. Wang, H., Chin, A.: Evolution analysis of a mobile social network. In: Proceedings of the 6th international conference on Advanced data mining and applications: Part I. pp. 310–321. ADMA'10, Springer-Verlag, Berlin, Heidelberg (2010)