Privacy Concerns in Location-based Social Networks

Fatma S. Alrayes, Alia I. Abdelmoty

School of Computer Science and Informatics
Cardiff University
Cardiff, UK
{F.S.Alrayes, A.I.Abdelmoty}@cs.cardiff.ac.uk

Abstract—User location data collected on Location-Based Social Networks (LBSN) can be used to enhance the services provided by those applications. However, it can be potentially utilised for undesirable purposes that can compromise users' privacy. This paper presents a study of the privacy implication of location-based information provision and collection in LBSN. The study is supported by analysis of representative data sets from such applications. The results demonstrate the need for further work on improving the visibility of the information collected to users of the Social Web, to allow them to better assess the implications of their location sharing activities.

Keywords-location privacy; LBSN; location-based inference

I. INTRODUCTION

Massive interest in geographical referencing of personal resources is evident on the Web today. Geographic referencing has evolved to become a natural method of organising and linking information with the aim of facilitating its discovery and use. GPS-enabled devices are enabling individuals to store their mobility tracks, tag photos and events. Embracing these new location-aware capabilities by the social networks has led to the emergence of Geo-Social Networks (GeoSNs) which offer their users the ability to geo-reference their submissions and to share their location with many other users. Subsequently, users can use the location identifier to browse and search for resources. GeoSNs include Location-Enabled Social Networks (LESNs), for example, Facebook, Twitter and Flickr, where users' location is supplementary identification of other primary data sets, and Location-Based Social Networks (LBSNs), for example, Foursquare and Yelp where location is an essential key for providing the service,

GeoSNs collect real-time and large-scale location information on users as well as other contextual information including user relationships and user provided text updates possibly over long periods of time. In particular, LBSNs as opposed to LESNs enable sharing and collection of detailed personal location information, provide significant semantic data associated with the location, such as place name, type and address, as well as allow users to express their opinions and experience in terms of reviews and tips. As a result, users' historical location information can be related to contextual and semantic information publicly available online [1] and can be used to infer personal and sensitive information about users and for constructing comprehensive user profiles. Possible derived information in such profiles can include user activities, relationships, interests and mobility patterns [2][3]. Such enriched location-based profiles can be considered to be useful if used to personalise and enhance the quality of use of the applications. However, they can potentially be used for undesirable purposes and can pose privacy threats ranging from location-based spams to possible physical harm by any adversary [4]. Users may not be fully aware of what location information is being collected, how the information is used and by whom, and hence can fail to appreciate the possible potential risks of disclosing their location information.

In this paper, we study the location privacy of users when using LBSNs. The aim of this study is to investigate the potential privacy implications of LBSNs by examining and demonstrating possible derived information from typical data sets collected by these applications for different types of users. Foursquare was chosen as a representative LBSN application for conducting this study due to its popularity (over 40 million users across the world by September, 2013 [5]).

Firstly, the dimensions of the location privacy problem in LBSNs are examined in terms of the type of data collected, its visibility and accessibility by users of the application, as well as the possible exploitation of these data and the level of security in such services, in order to provide a better understanding of the privacy issues. Secondly, an analytical study is carried out, using a representative data set, to explore the location data content and the range of possible inference that can be made from them. Usage patterns in the dataset are used to guide a classification of users of the application and in the analysis of the data.

Previous studies focused mostly on examining spatiotemporal movement patterns in LBSNs [6][7]. Some studies explored users' privacy concerns and attitude when sharing their location for social purposes, but presented limited evaluations using restricted application scenarios [8][9].

This paper presents a study of the privacy implications of location-based information provision and collection in LBSN. The study is supported by analysis of representative data sets from such applications. The results demonstrate the need for further work on improving the visibility of the information collected to users of the Social Web, to allow them to better assess the implications of their location sharing activities.

The rest of this work is organized as follows: section II provides an overview of related work. Section III discusses the dimensions of the location privacy problem in LBSNs. Section IV describes the experiment conducted. Section V presents the result of the analyses. A conclusion of the work and future directions are given in Section VI.

II. RELATED WORK

Two relevant questions to the problem being studied are: to what extent is location privacy a potential concern for users in LBSNs, and what sort of location-based inference is possible from the data collected by LBSNs. In this section, related work on both issues is reviewed.

A. Users Attitudes and Privacy Concerns in Geo-Social Applications

A growing research interest has been witnessed over the past few years for studying users' attitudes and privacy concerns of their location privacy and investigating how user-empowered location privacy protection mechanism can influence their behaviour. Tsai et al. [8] developed a social location sharing application where participants were capable of specifying time-based rules to share their location and they were then notified of who viewed their locations. Their findings suggested that the control given to user for setting their sharing preferences has contributed to the reduction of the level of privacy concern of the participants.

Similarly, Sadeh et al. [9] enabled users of their People Finder application to set rule-based location privacy controls by determining the where, when and with whom to share their location and were notified when their location information was requested. Participants were initially reluctant to share their location information and then tended to be more comfortable over time. Patil et al. [10] implemented a system that represents actual users' workplace offering live feeds about users and their location, then asked those users to define permissions for their personal information sharing by setting various levels linked to four user categories. They found that these participants were concerned most about their location information and they utilised the permission feature to control it. Another study by Kelley et al. [11] showed that users were highly concerned about their privacy especially when sharing location with corporate-oriented parties. However, their location-sharing with advertising companies can be increased when offering more complex location privacy settings.

Other work was carried out to examine how the employment of visualisation methods may impact users' attitude to location privacy and behaviour. Brush et al. [12] studied users' attitudes towards their location privacy when socially sharing their location or when tracking it using GPS for long periods of time and questioned whether using some obfuscation techniques can address users' concerns. As a result, participants were concerned about revealing their home, identity and exact locations. They visually recognised and chose the best obfuscation techniques they felt protect their location privacy. In addition, Tang et al. [13] investigated to what extent presenting various visualisations of users' location history can influence their privacy concerns

when using location sharing applications. They developed text-, map-, and time-based visualisation methods and considered spatiotemporal properties of sharing historical location. They noted that the majority of participants were concerned about their location privacy including their physical privacy when showing them their visualised location history and consequently preferred text-based visualisation when sharing location with other users, as it was perceived to limit the amount of information exposed.

With regards to public GeoSNs, there is relatively few research works that examines privacy concerns of users. Lindqvist et al. [14] considered users' motivations in using Foursquare and questioned their privacy concerns. Their analysis showed that most of the participants had few concerns about their privacy and users who were more concerned about their privacy chose not to check into their private residence or to delay checking into places till after they leave, as a way of controlling their safety and privacy. A similar observation was noted by Jin et al. in [15], where it was found that users were generally aware of the privacy of their place of residence and tended not to provide full home addresses or blocked access to their residential check-ins to other users.

In summary, it is evident that location privacy presents a real concern to users in location sharing applications, and particularly as they become aware of the data they are providing. Previous studies may have been limited by several factors, including the size and representativeness of the sample user base used in the experiments conducted and the limited features of the proprietary applications used in testing [8][9][10][11]. Moreover, as far as we are aware, there are no previous studies that consider the problem of location privacy on public LBSNs.

B. Location-Based Inference from GeoSNs

There are some studies that utilised publicly available information from GeoSNs in order to derive or predict users' location. In [16], Twitter users' city-level locations were estimated by only exploiting their tweets contents with which it was possible to predict more than half of the sample within 100 miles of their actual place. Similarly, Pontes at al [17] examined how much personal information can be inferred from the publicly available information of Foursquare users and found the home cities of more than two-third of the sample within 50 kilometres. Sadilek at al. [18] investigated novel approaches for inferring users' location at a given time by taking advantage of knowing the GPS positions of their friends on Twitter. Up to 84% of users' exact dynamic locations were derived. Interestingly, Gao et al. [19] formulated predictive probability of the next check-in location by exploiting social-historical ties of some Foursquare users. They were able to predict with high accuracy possible new check-ins for places that users have not visited before, by exploiting the correlation between the social network information and geographical distance in LBSNs [20].

Other works focussed on investigating the potential inference of social relationships between users of GeoSNs. Crandall et al. [21] investigated how social ties between

people can be derived from spatial and temporal co-occurrence by using publicly available data of geo-tagged pictures from Flickr. They found that relatively limited co-occurrence between users is sufficient for inferring high probability of social ties. Sadilek at al. [18] also formulated friendship predictions that derive social relationships by considering friendship formation patterns, messages' content of users and their location. They predicted 90% of friendships with accuracy beyond 80%. Additionally, Scellato et al. [22] investigated the spatial properties of social networks existing among users of three popular LBSNs and found that the likelihood of having social connection decrease with distance. Scellato et al. [23] developed a link prediction system for LBSNs by utilising users' check-ins information and places properties. 43% of the all new links appeared between users that have at least one check-in place in common and especially those who have a friend in common.

Studying and extracting spatiotemporal movements and activities patterns of users on GeoSNs have also attracted much research in recent years. Dearman et al. [24] exploited locations' reviews of Yelp in order to identify a collection of potential activities promoted by the reviewed location. They derived the activities supported by each location by processing the review text and validated their findings through a user questionnaire. Noulas at al. [25] studied user mobility patterns in Foursquare by studying popular places and transitions between place categories. Cheng at al. [6] examined a large scale dataset of users and their check-ins to analyse human movement patterns in terms of spatiotemporal, social and textual information associated with this data. They were able to measure user displacement of consecutive check-ins, distance between users' check-ins and their centre of mass, and the returning probability to venues. They also studied the factors affecting users' movement and found considerable relationship between users' mobility and geographic and economic conditions. More recently, Preotiuc-Pietro et al. [7] investigated the behaviour of thousands of frequent Foursquare users. They analysed users' movements including returning probability, check-ins frequency, inter-event time, and place transition among each venue category. They were also able to group users based on their check-in behaviour such as generic, businessmen or workaholics as well as predicting users' future movement.

The above studies show that there is a significant potential for deriving personal information form GeoSNs and hence imply the possible privacy threats to user of these applications. Whereas previous studies considered mobility and behaviour of large user groups and determined general patterns and collective behaviour, in this work we consider the privacy implications for individual users, with the aim of understanding possible implied user profiles from location data stored in LBSNs.

III. LOCATION PRIVACY ON LBSNs

Four aspects of location privacy on LBSNs can be identified. These are related to the amount of data collected and its quality, its visibility and accessibility, its possible utilisation by potential users, and the level of security offered to the user by the application.

A. Location Data Collection

Location data collection refers to the type of location data collected and stored as well as to its quality.

Foursquare collects and records user location data automatically and continuously, by estimating the user's current latitude and longitude from the device being used. User's check-ins into specific places are verified against their estimated current location and recorded explicitly. Foursquare also states that it collects additional information from third parties services, that communicate with the application, including personal information and activities.

User's visit to a place is recorded by the user intentionally checking-in a place. Check-ins are made against predefined venues. Venues record detailed information about a geographic place, including a name, a type classification, coordinate point representation, and address. In addition, users' check-ins are also time stamped. Depending on the frequency of check-ins, a user (and the system) is able to record a complete and highly specific spatiotemporal track of their mobility.

B. Location Information Accessibility

Location information accessibility is concerned with how much of users' data are available and visible to others including the user, other users and the third parties of the service.

Users' pervious check-in information is provided to them in the form of check-in history where they can view their visited venues, date of the visit and any tips they have made. They are also able to access and download their check-in. However, users seem to have only a limited aspect of accessibility compared with what service provider can collect or exploit these data. For example, Foursquare states in their privacy policy that they record users' location on a continuous bases even without users checking into venues, but users do not have access to this data.

Almost all of the users' information is publicly available by default and can be viewed by other users. This includes profile information, tips, likes, friends list, photos, badges, mayorships, and check-ins. Users are only able to block access to their check-ins and photos by setting their view to 'private'.

As for information disclosure to third parties including Foursquare API's users, all of the publicly available users' information is accessible by third parties including private users' information such as check-ins in anonymous form that is not linked to individual users. Foursquare also indicates that they will share users' personal information with their business partners and whenever is necessary in some situations, such as enforcement of law.

C. Location Data Exploitation

Location information exploitation refers to how the application or third parties can utilise the data and for which purposes.

Foursquare gives itself absolute privileges over using and manipulating user information as stated in their terms of use.

"By submitting User Submissions on the Site or otherwise through the Service, you hereby do and shall grant Foursquare a worldwide, non-exclusive, royalty-free, fully paid, sublicensable and transferable license to use, copy, edit, modify, reproduce, distribute, prepare derivative works of, display, perform, and otherwise fully exploit the User Submissions in connection with the Site, the Service and Foursquare's (and its successors and assigns') business, including without limitation for promoting and redistributing part or all of the Site (and derivative works thereof) or the Service in any media formats and through any media channels (including, without limitation, third party websites and feeds)."

From the above, it is clear that there are no commitments from the application provider as to how the data may be used or shared by the application or by other parties. Hence, by agreeing to the terms and conditions, users effectively are giving away the data and unconditional rights to use the data to the application.

D. Location Data Security

Location data security refers to the level of data protection provided by the application for securing the user's data against the risk of loss or unauthorized access.

Foursquare declares that the security of users' information is not guaranteed and any "Unauthorized entry or use, hardware or software failure, and other factors, may compromise the security of user information at any time". Without any commitment to responsibility for data security, the application provider is declaring the possible high risk of data abuse by any adversary or even by the application provider themselves.

IV. EXPERIMENT

This analysis is carried out using a real-world dataset from Foursquare for the purpose of demonstrating privacy implications of user activity on LBSNs. The effect of location data density and diversity on the possible inferences that can be made from the data is also analysed.

A. Dataset

The Foursquare dataset used in this analysis is provided by Jin et al. [15]. The dataset contains venue information and public check-ins for anonymised users around the wide area of Pittsburgh, USA from 24 February, 2012 to 22 July, 2012. It contains 60,853 local venues, 45,289 users and 127,6988 public check-ins of these users.

B. Approach and Tools Used

To study the possible impact of location data density on users' privacy, users of the dataset were first classified into groups based on their check-in frequency. A filter was initially imposed to disregard sparse user activity. Hence, users with less than five check-ins per month were removed from the dataset. The rest of the users were categorised into three groups based on their check-in frequency per day, to Moderate, Frequent and Hyper-active user groups, as shown

in Table I. One representative user is selected from each group who has the nearest average check-ins per day to the average check-ins per for the whole group. Table II shows some statistics for the selected users.

The R statistical package was used for analyses and presentation of results. The SQLDF package was used for querying, linking and manipulating the data and the ggplot2 package was used for the presentation of the results of the analysis.

V. RESULTS

Analysis of the dataset questioned the sort of implicit user-related information that can be considered to be private that may be extracted using the location data collected. The user's spatial location history can be extracted, in the form of visits to venues and the exact times of such visits. The places visited are identified and described in detail. For example, user7105 visited 'Kohl's'; a department store, located at latitude 40.51105772555344 and longitude -79.99340577016872 at 9 a.m. Monday 27/2/2012.

The basic information on venue check-ins can be analysed further and combined with other semantic information from the user profile to extract further information that can compromise user's privacy. Analysis will investigate the relationship between users and places visited, their mobility patterns and the relationships between users and other users as follows:

- **Degree of association between user and place.** Relationship with individual place instances as well as with general place types or categories will be studied. Elements of interest will include visit frequency, and possible commuting habits in terms of the association between the visit frequency of places and their location.

- **Spatiotemporal movement patterns.** Visiting patterns to individual places or to groups of places can identify regular movement patterns. In addition, a change of visit patterns can also be a significant pointer to user activity.

TABLE I. STATISTICS OF USERS' GROUPING.

Group Name	Check-ins Range in Total	Users Count	Check-ins Range per Day	Average Check-ins per Day
Moderate	Between 50 and 300	4902	0.3 to 2	1.15
Frequent	Between 301 and 750	880	2 to 5	3.5
Hyper-active	Between 751 and 1303	24	5 to 8.6	6.8

TABLE II. STATISTICS OF THE SELECTED USERS.

Factor	Selected Users		
	User9119	*User7105*	*User2651*
Number of total check-ins	144	511	1019
Average check-ins per day	**0.96**	**3.4**	**6.8**
Number of visited venues	21	99	101
Number of visited venues' categories	17	47	57
Number of visited venues' main categories	10	11	17
Number of friends	20	10	19

- **Degree of association with other users.** Relationship between users can be derived by studying their movement patterns and analysing their co-occurrence in place and time.

A. The Moderate User

The analysis results of user9119 selected from the moderate groups are as follows.

1) Degree of Association Between User and Place

Two frequently visited venues by user9119 are 'Penn Garrison' whose category is 'Home' and 'USX Tower' whose category is 'Office' representing 44% and 36% respectively of the total check-ins. Home and office are highly sensitive places, yet they represent 80% of this user's check-ins. Other visited place types with significantly less frequency include, 'Nightlife Spot':0.5%, 'Travel & Transport': 0.27%, and 'Shop & Service':0.27%. User9119 is also interested in 'Hockey', 'Garden Center' and 'Museum' place types. As could be predicted, the location of venues visited, indicates that most of the visited venues are close to 'Home' and 'Office', whereas this user commutes further away to visit some less frequent venues such as 'Hockey Arena'. Figure 1 shows this user's check-in frequency for different categories of venues classified by the time of day. As can be seen from the figure, user's association with sensitive places like home and place of work can be identified. In addition, a strong association with other place categories is also evident.

2) Spatiotemporal Movement Patterns

About 40% of this user's total check-ins occurs at 9 am, mostly in the 'Office' and at 7 pm, mostly at 'Home'. More than two-thirds of the check-ins are between 10 am and 2 pm and between 6 pm and 11 pm, which indicates that this user commutes more frequently during these hours. From the user's weekly patterns of movement, it can be seen that 71% of the venues were visited after 6 pm. Mondays and Thursdays are when this user is most active, representing 41% of the check-ins. User9119 tend to go to 'Nightlife spots' more frequently during working days, whereas visits to other specific place types occur only at weekends, including, 'Salon or Barbershop', 'Coffee Shop' and 'Garden Centre'. This user typically starts commuting earlier on working days and visits more places than on weekends. Observing the check-ins by month shows that the months of May and June are the most active in terms of the check-in frequency, comprising 60% of total check-ins, as well as diversity of category of venues visited (99% of the total visited categories of venues occurred in those months, including the emergence of new categories such as 'Museum', 'Airport' and 'Hotel'). The user was least active in April. Figure 4 demonstrates this user's check-ins count in different categories of venues, classified by day and grouped by month.

Some changes of this user's habits can be noticed as well, which can suggest a change of the user's circumstances. For example, the user has not visited any Nightlife spots in March and April and has not checked-in in any place on Sundays of June and July including 'Home' and 'Office'. In addition, the us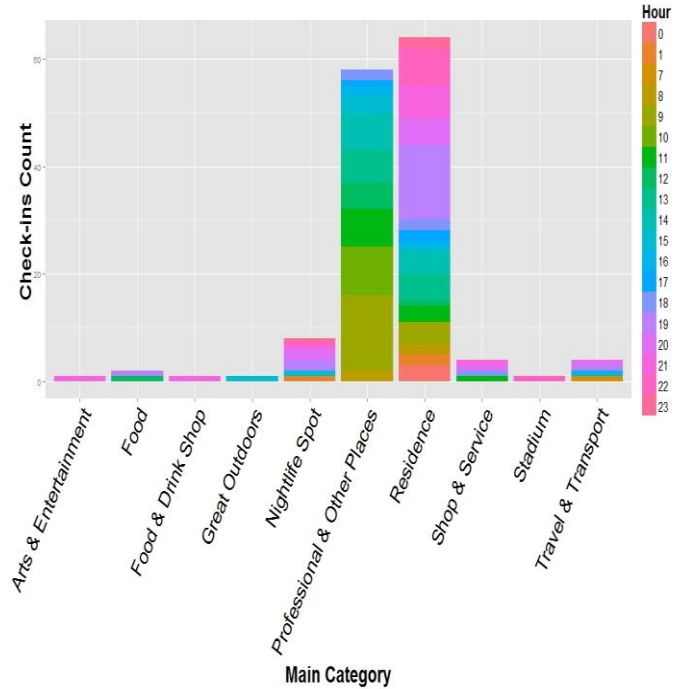er has not checked in any place for a period of a week between the 21st and 28th. User9119 last check-in before this week was on the 20th of April at 'Home'. This may indicate a possible period of time-off work in that week.

Figure 1. The moderate users' check-ins count, classified by the category of venues for different hours of the day.

3) Degree of Association with Other Users

Co-location is used here to denote that users have visited the same venue. This can be used as a measure of uses' interest in a place. User9119 was co-located at in 6 unique venue categories with two (out of twenty) friends.

Spatiotemporal co-occurrence between users is co-location at same place and time. This can be used as a measure of relationship between users. User9119 shared three co-occurrences with two friends; once with friend1236 at 'American Restaurant' and twice with friend15229 at 'Office', which can indicate that friend15229 is a colleague at work. In fact, this user shared 95 co-occurrences with 52 other users, 90% of which were in the 'Office' suggesting the probability of those users being work colleagues.

B. The Frequent User

Analysis of results of user7105 from the frequent user group is as follows.

1) Degree of Association between User and Place

Similar to the moderate user, user7105 most checked-in venue category is 'Home', whose location is identified in detail. However, the second most visited venue is a specific restaurant, whose category is 'American Restaurant', representing 25% of the total check-ins and 28% of category check-ins. This visit pattern may indicate that this is the user's work place.

The third most visited venue category for this user is 'Bar' (4%), that is a subcategory of 'Nightlife Spot', representing about 7% of check-ins. Generally, the third most

visited main category is 'Shop & Service' corresponding to 10% of check-ins where specifically 40% of it to 'Gas Station or Garage' and 25% to 'Drugstore or Pharmacy'. User7105 occasionally interested in visiting places described as 'Great Outdoors', 'Professional & Other Places' and 'Arts & Entertainment'.

The majority of the most frequently visited venues are within close distance to 'Home' and to the 'American Restaurant', whereas user7105 commutes further away for other less frequently visited places, such as, the 'Medical Center'.

2) Spatiotemporal Movement Patterns

Generally, about 20% of the check-ins occurs from 10 am to 12 pm, half of which are at 'Home'. In addition, user7105 tend to move the most between 3pm and 5pm, representing 23% of his total check-ins to 46% of the visited venues' categories. More than half of the check-ins are at 'Atria's', which may indicate that the user starts his work shift in this place at that time. This hypothesis can be ascertained by examining his subsequent check-ins, where 18% of the check-in happens between 12 am and 3 am at 'Home', possibly when the user comes back from work. There is a high correlation in terms of place transition between 'Home' and the 'American Restaurant'.

When examining the weekly mobility, user7105 is more active on Tuesdays followed by Saturdays corresponding to 19% and 16% respectively of total check-ins. Noticeably, the majority of Friday and Tuesday check-ins occurs at 12 am, whereas Monday and Saturday at 4 pm. Furthermore, this user has visited more diverse venues on Tuesdays followed by Thursdays and Wednesdays representing 53%, 43% and 38% respectively of total visited categories.

During the working week, this user tend to visit a 'Bar' (5%), especially on Tuesdays, and 'Gas Station or Garage' (4%). This may be reasonable considering his working shifts. While on weekends, 'Grocery or Supermarket' and 'Drugstore or Pharmacy' venues are among the top four visited categories corresponding to 4% and 5% respectively of weekends' check-ins.

User7105's check-in patterns was regular over the whole period. However, this user's visits are more frequent and diversified in the month of March. Noticeably, about 28% of the check-ins between 12 and 3 am occurred in March, indicating a possible change of lifestyle. Figure 5 presents this user's check-ins count in different categories of venues, classified by day and grouped by month.

3) Degree of Association with Other Users

User7105 had co-locations in 36 unique venues from 19 different categories with 7 friends. In particular, 26 co-locations are shared with the freind38466 at 14 venues categories including 'Coffee Shop', 'Bar', 'Fast Food Restaurant' and 'Other Nightlife'. Co-locations shared with the rest of the friends include 'Bar', 'Mexican Restaurant', 'Hospital' and 'Government Building'.

Moreover, user7105 has 16 spatiotemporal co-occurrences at 14 unique venues from 6 different categories with two friends where 14 co-occurrences with freind38466 at 6 different categories including mostly 'Bar', 'American Restaurant', and 'Sandwich Place', which can denote a close

friendship between them. The other two co-occurrences are with friend15995 at 'American Restaurant' on May 13th and June 17th, 2012. The place and time of this user's co-occurrences with friends are shown in Figure 2. Similarly, this user also has 89 co-occurrences with other users, who are not stated as friends, at 29 unique venues where 38% of these co-occurrences at 'American Restaurant' and 24% at 'Plaza'.

C. The Hyper-Active User

The results of analysis for user2651 selected from the hyper-active user group are as follows.

1) Degree of Association Between User and Place

The first most visited venue by this user is a 'Nightlife Spot' corresponding to 15% of total check-ins. Two 'Home' venues were recorded, 'My Back Yard' and 'La Couch', representing 23% of the check-ins. Both home venues have the same location coordinates, implying that they are actually the same place. 'Automotive Shop', 'Pool' and 'Italian Restaurant', representing 9%, 8% and 5% respectively of this user's total check-ins indicating the user's interests and activities which can be swimming and Italian food. A particular instance with a vague category of 'Building' was among the top 10 most visited venues. Further investigation of this venue using the given place name revealed that this building is a place where an international summit for creative people is held [26]. That indicates that user2651 is possibly an active participant of such an event.

When considering the main category of the visited venues, this user generally visits 'Shop & Service', 'Nightlife Spot', 'Arts & Entertainment' and 'Food' on a regular basis, representing 17%, 14%, 11% and 10% respectively of this user's check-ins. User2651 also usually visits 'Gas Station or Garage': 4%, and 'Church': 3%, which can imply that this user is a person with faith. The location of the visited venues can be clustered into two main areas on a map as illustrated in Figure 3. One area is where the user's 'Home' is location, as well as other frequently visited venues such as 'Nightlife Spots' and 'Gym or Fitness Center'. The other area includes mostly less frequently visited venues such as 'Hospital'.
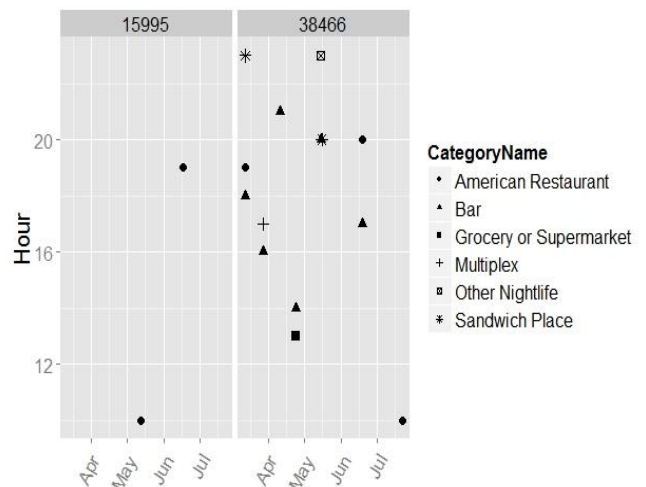

Figure 2. Spatiotemporal tracks of the frequent user co-occurrences with friends.
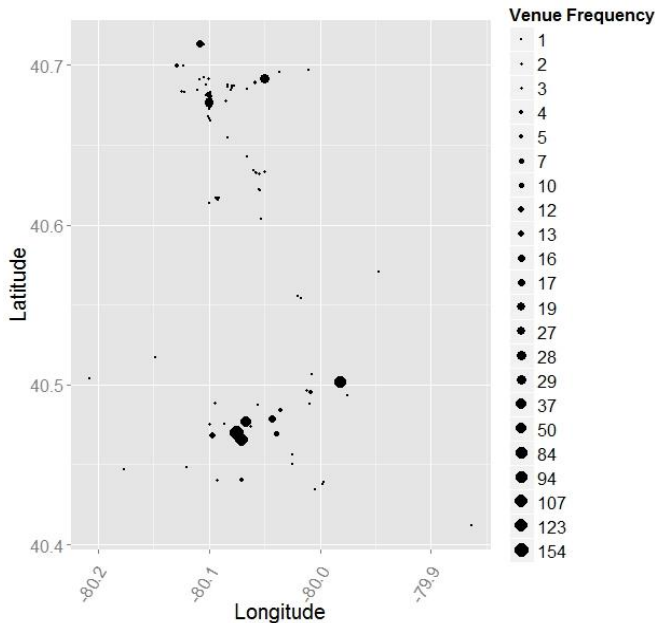
Figure 3. Venues' coordinates visited by the hyper-active user by considering the frequency of visit.

2) Spatiotemporal Movement Patterns

Overall, 53% of residential check-ins occurs between 9 am and 12 pm where user651. This user's check-in frequency reaches the peak at 2 pm where 10% of the check-ins occurs and about two-third of them into the 'Automotive Shop'. Moving towards the night, user2651's check-in frequency reached another peak between 11 and 12 am representing 18% of the check-ins in which more than half is into 'Nightlife Spot' where this user may work at, and a third into 'Home' when potentially returning home. Noticeably, this user tends to be more active at night since about 70% of the check-ins happens after 6 pm.

Surprisingly, weekends have similar check-in frequencies as working week, and Sunday has the highest higher check-in frequency among the week days which is not the expected movement habit for average people. Moreover, user2651 checks in considerably less at the 'Automotive Shop' and the 'Pool' on Wednesday and Friday respectively. However, user2651 checks in the 'Automotive Shop' and the 'Nightlife Spot' even in weekends, which may suggest that this user has weekends work shifts. In addition, this user typically has some different priorities of visit between working week and weekend. For example, 'Church' is the sixth most visited venue category in weekends, whereas in working days, 'Bar' is the sixth most visited venue category.

User2651 has regular check-in patterns over the whole period. However, in the months of June and July, the user's check-ins into 'Hotel' and 'Pool' significantly increase representing 75% and 60% respectively of these venues total check-ins. Moreover, other categories or venues are highly visited in certain months. For instance, 35% of total 'Gas Station or Garage' check-ins occurs in April, which can indicate that this user commutes more at that time, and 40% of total 'Church' check-ins occurs in July. Figure 6

demonstrates this user's check-ins count in different categories of venues, classified by day and grouped by month.

3) Degree of Association with Other Users

User2651 shares co-locations in 27 unique venues from 19 categories with 9 friends where 13 co-locations are with friend12432 and 9 with friend12046. Most of co-locations with this user's friends are in 'Nightlife Spots', 'Gas Station or Garage', 'Pool', 'Flower Shop' and 'Bar'. This user also has 16 co-occurrences with three friends where 4 of them with friend12046 and 3 with friend12432 at a 'Nightlife Spots', 'Pool', 'Flower Shop' and with just friend12432 at 'Automotive Shop'. As with other users, user2651 co-occurred with 23 users at 12 distinct venues where half of these co-occurrences happened in 'Bar', 'Automotive Shop' and 'Grocery or Supermarket'.

VI. CONCLOSION AND FUTURE WORK

In this paper, we investigated the privacy implication of location-based information provision and collection in LBSNs. The study is supported by analysis of a representative dataset from Foursquare. The results showed that it is highly feasible to infer rich personal information about users and their mobility. In particular, some of the possible inferences demonstrated are:

- Users' spatiotemporal movement tracks and patterns.
- Users' absence and presence in particular places.
- Visiting frequencies and possible degree of association with specific places or place types.
- Users' commuting habits.
- Co-location patterns with other users and friends.

More work needs to be done to investigate the following issues:

- The relationship between the density of information and the accuracy of the inference.
- The effect of the integration of users' data from different LBSNs.
- The relationship between the amount of information that can be analysed and the users' perception of personal privacy.

The study also demonstrates the need for further work on improving the visibility of the information collected to users of the Social Web to allow them to better assess the implications of their location sharing activities.

REFERENCES

[1] C. Ruiz Vicente, D. Freni, C. Bettini, and C. S. Jensen, "Location-Related Privacy in Geo-Social Networks," IEEE Internet Computing, vol. 15, no. 3, pp. 20–27, May 2011.

[2] D. Riboni, L. Pareschi, and C. Bettini, "Privacy in Georeferenced Context-aware Services : A Survey," in in Privacy in Location-Based Applications, Berlin, Heidelberg: Springer-Verlag, 2009, pp. 151–172.

[3]   S. Gambs, O. Heen, and C. Potin, "A comparative privacy analysis of geosocial networks," in SPRINGL '11 Proceedings of the 4th ACM SIGSPATIAL International Workshop on Security and Privacy in GIS and LBS, 2011, pp. 33–40.

[4]   R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, and J.-P. Hubaux, "Quantifying Location Privacy," 2011 IEEE Symposium on Security and Privacy, pp. 247–262, May 2011.

[5]   Foursquare, "About." [Online]. Available: https://foursquare.com/about. [Accessed: 29-Sep-2013].

[6]   Z. Cheng, J. Caverlee, K. Lee, and D. Sui, "Exploring Millions of Footprints in Location Sharing Services.," in ICWSM, 2011, vol. 2010, no. Cholera, pp. 81–88.

[7]   D. Preotiuc-Pietro and T. Cohn, "Mining User Behaviours: A Study of Check-in Patterns in Location Based Social Networks," in Web Science, 2013.

[8]   J. Y. Tsai, P. Kelley, P. Drielsma, L. F. Cranor, J. Hong, and N. Sadeh, "Who's Viewed You? The Impact of Feedback in a Mobile Location-Sharing Application," in CHI '09 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2009, pp. 2003–2012.

[9]   N. Sadeh et al., "Understanding and capturing people's privacy policies in a mobile social networking application," Personal and Ubiquitous Computing, vol. 13, no. 6, pp. 401–412, Oct. 2008.

[10] S. Patil and J. Lai, "Who gets to know what when: configuring privacy permissions in an awareness application," in Proceedings of the SIGCHI conference on human factors in computing systems (CHI 2005), 2005, pp. 101–110.

[11] P. G. Kelley, M. Benisch, L. F. Cranor, and N. Sadeh, "When are users comfortable sharing locations with advertisers?," in Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11, 2011, pp. 2449–2452.

[12] a. J. B. Brush, J. Krumm, and J. Scott, "Exploring end user preferences for location obfuscation, location-based services, and the value of location," in Proceedings of the 12th ACM international conference on Ubiquitous computing - Ubicomp '10, 2010, p. 95.

[13] K. P. Tang, J. I. Hong, and D. P. Siewiorek, "Understanding how visual representations of location feeds affect end-user privacy concerns," in Proceedings of the 13th international conference on Ubiquitous computing - UbiComp '11, 2011, pp. 207–216.

[14] J. Lindqvist, J. Cranshaw, J. Wiese, J. Hong, and J. Zimmerman, "I'm the mayor of my house: examining why people use foursquare-a social-driven location sharing application," in CHI '11 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2011, pp. 2409–2418.

[15] L. Jin, X. Long, and J. B. D. Joshi, "Towards understanding residential privacy by analyzing users' activities in foursquare," in Proceedings of the 2012 ACM Workshop on Building analysis datasets and gathering experience returns for security - BADGERS '12, 2012, pp. 25–32.

[16] Z. Cheng, J. Caverlee, and K. Lee, "You are where you tweet: a content-based approach to geo-locating twitter users," in CIKM '10 Proceedings of the 19th ACM international conference on Information and knowledge management, 2010, pp. 759–768.

[17] T. Pontes, M. Vasconcelos, J. Almeida, P. Kumaraguru, and V. Almeida, "We Know Where You Live : Privacy Characterization of Foursquare Behavior," in UbiComp '12 Proceedings of the 2012 ACM Conference on Ubiquitous Computing, 2012, pp. 898–905.

[18] A. Sadilek, H. Kautz, and J. Bigham, "Finding your friends and following them to where you are," in WSDM '12 Proceedings of the fifth ACM international conference on Web search and data mining, 2012, pp. 723–732.

[19] H. Gao, J. Tang, and H. Liu, "Exploring Social-Historical Ties on Location-Based Social Networks.," in Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media, 2012, pp. 114–121.

[20] H. Gao, J. Tang, and H. Liu, "gSCorr: modeling geo-social correlations for new check-ins on location-based social networks," in Proceedings of the 21st ACM international conference on Information and knowledge management, 2012, pp. 1582–1586.

[21] D. J. Crandall, L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. Kleinberg, "Inferring social ties from geographic coincidences.," Proceedings of the National Academy of Sciences of the United States of America, vol. 107, no. 52, pp. 22436–41, Dec. 2010.

[22] S. Scellato, A. Noulas, R. Lambiotte, and C. Mascolo, "Socio-Spatial Properties of Online Location-Based Social Networks.," in ICWSM, 2011, pp. 329–336.

[23] S. Scellato and C. Mascolo, "Exploiting Place Features in Link Prediction on Location-based Social Networks Categories and Subject Descriptors," in Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, 2011, no. Section 3, pp. 1046–1054.

[24] D. Dearman and K. Truong, "Identifying the activities supported by locations with community-authored content," in Proceedings of the 12th ACM international conference on Ubiquitous computing, 2010, pp. 23–32.

[25] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil, "An Empirical Study of Geographic User Activity Patterns in Foursquare.," in ICWSM, 2011, pp. 70–573.

[26] "World Domination Summit." [Online]. Available: http://worlddominationsummit.com/faq/#primary-content. [Accessed: 05-Nov-2013].
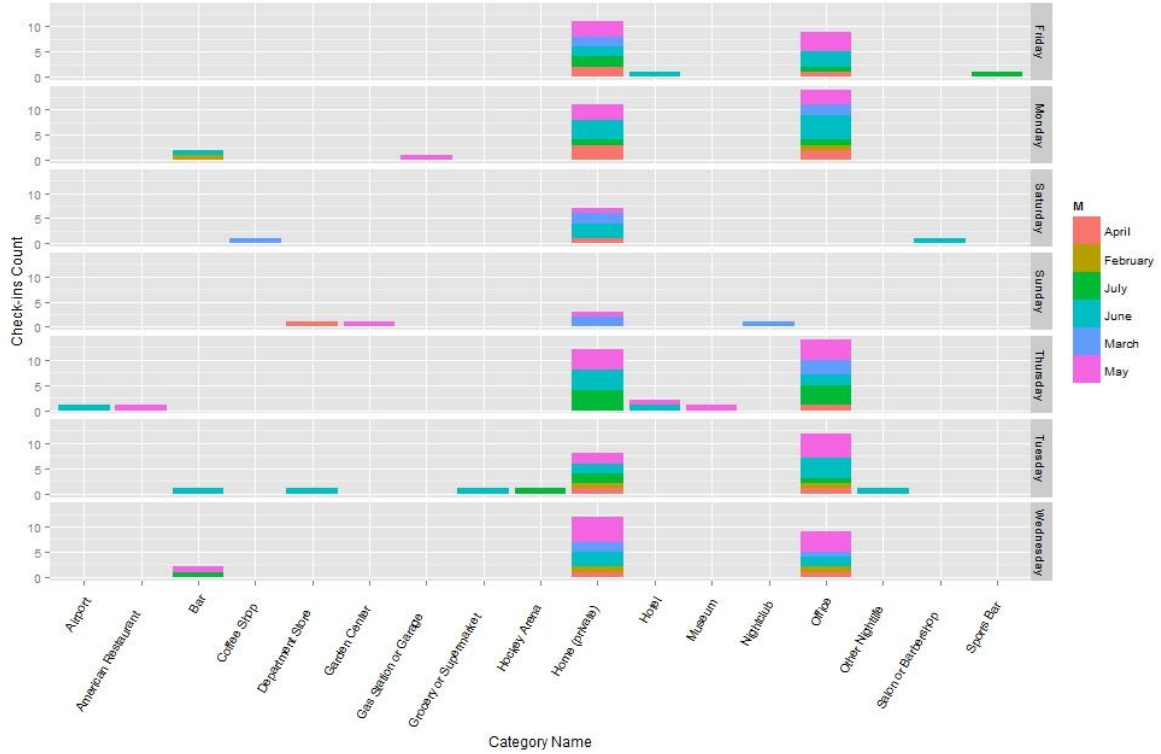
Figure 4. The moderate user's check-ins count in different categories of venues, classified by day and grouped by month.
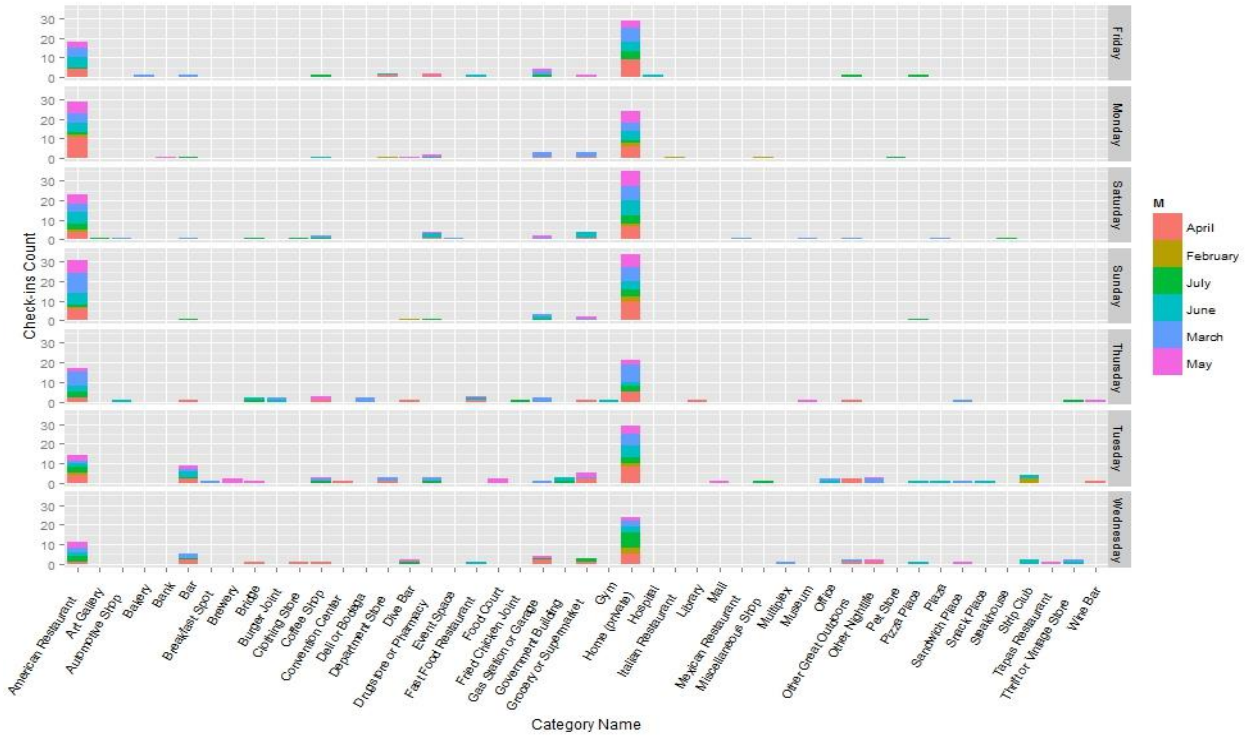

Figure 5. The frequent user's check-ins count in different categories of venues, classified by day and grouped by month.
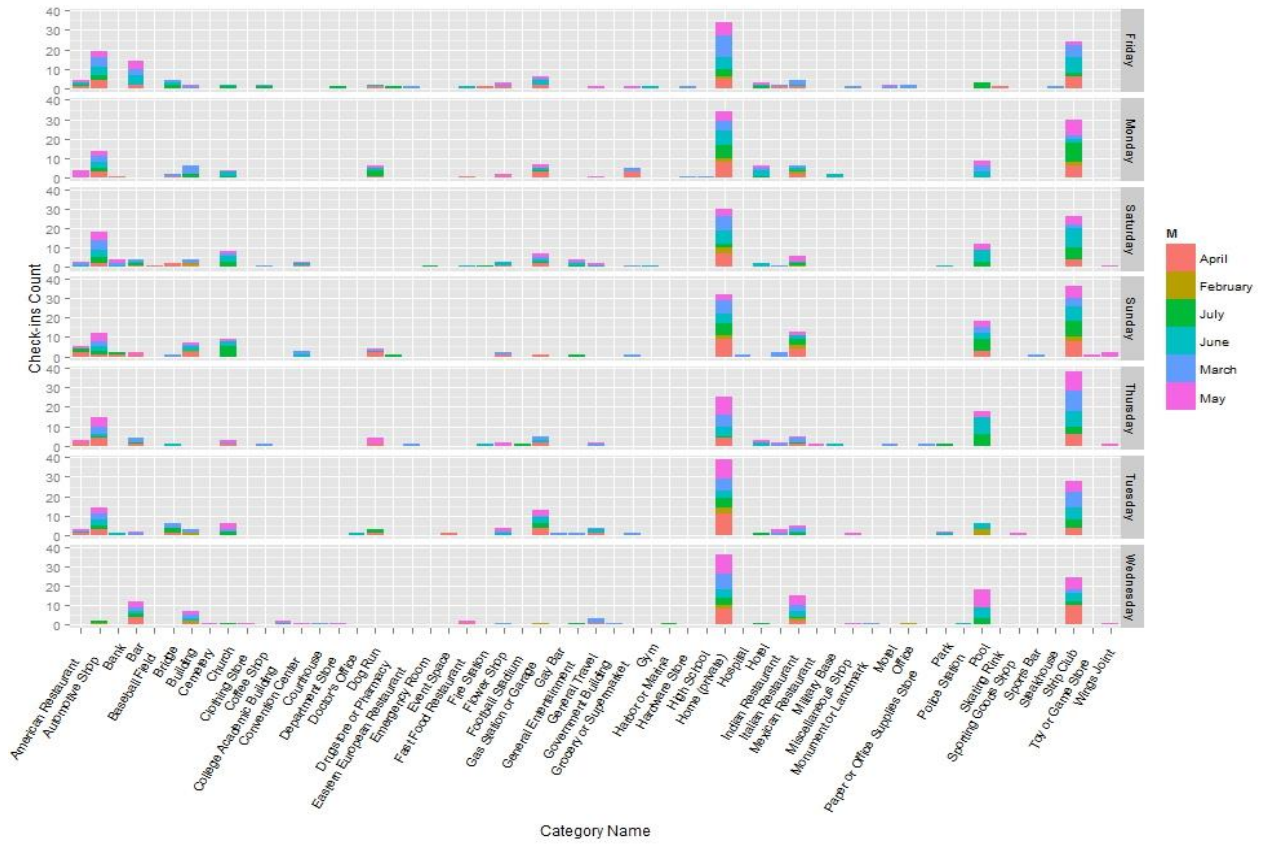
Figure 6. The hyper-active user's check-ins count in different categories of venues, classified by day and grouped by month.