

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:<https://orca.cardiff.ac.uk/id/eprint/70577/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Alrayes, Fatma and Abdelmoty, Alia 2014. No place to hide: a study of privacy concerns due to location sharing on geo-social networks. *International Journal On Advances in Security* 7 (3/4) , pp. 62-75.

Publishers page:

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



No Place to Hide: A Study of Privacy Concerns due to Location Sharing on Geo-Social Networks

Fatma S. Alrayes and Alia I. Abdelmoty

School of Computer Science & Informatics
Cardiff University
Wales, UK

Email: {F.S.Alrayes, A.I.Abdelmoty}@cs.cardiff.ac.uk

Abstract—User location data collected on Geo-Social Networking applications (GeoSNs) can be used to enhance the services provided by such applications. However, personal location information can potentially be utilised for undesirable purposes that can compromise users' privacy. This paper presents a study of privacy implications of location-based information provision and collection on user awareness and behaviour when using GeoSNs. The dimensions of the problem are analysed and used to guide an analytical study of some representative data sets from such applications. The results of the data analysis demonstrate the extent of potential personal information that may be derived from the location information. In addition, a survey is undertaken to examine user awareness, concerns and subsequent attitude and behaviour given knowledge of the possible derived information. The results clearly demonstrate users' needs for improving their knowledge, access and visibility of their data sets as well as for means to control and manage their location data. Future work needs to investigate the current state of personal data management on GeoSNs and how their interfaces may be improved to satisfy the highlighted users' needs and to protect their privacy.

Keywords—location privacy; Geo-social networks; mobility patterns; privacy concerns.

I. INTRODUCTION

The proliferation and affordability of GPS-enabled devices are enabling individuals to accumulate an increasing amount of personal information, such as their mobility tracks, geographically tagged photos and events. Embracing these new location-aware capabilities by social networks has led to the emergence of Geo-Social Networks (GeoSNs) that offer their users the ability to geo-reference their submissions and to share their location with other users. Subsequently, users can use location identifiers to browse and search for resources. GeoSNs include Location-Enabled Social Networks (LESNs), for example, Facebook, Twitter, Instagram and Flickr, where users' locations are 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.

In addition to location data that describe the place the places visited by users, GeoSNs also records other personal information, such as user's friends, reviews and tips, possibly over long periods of time. User's historical location information can be related to contextual and semantic information publicly available online and can be used to infer personal information and to construct a comprehensive user profile

[1], [2]. Derived information in such profiles can include user activities, interests and mobility patterns [3], [4]. Such enriched location-based profiles can be considered to be useful if used to personalise and enhance the quality of the services provided by the social networking applications. For example, by recommending a place to visit on Foursquare and showing local trends on Twitter. However, they can potentially be used for undesirable purposes and can pose privacy threats ranging from location-based spams to possible threats by an adversary [5]. Users may not be fully aware of what location information are being collected, how the information are used and by whom, and hence can fail to appreciate the possible potential risks of disclosing their location information.

In this paper, a study of location privacy of users when using GeoSNs is presented. The aims are to investigate potential privacy implications of GeoSNs, as well as examine users' privacy concerns and attitude when using these networks. We demonstrate the privacy implications by identifying possible derived information from typical data sets collected by LBSNs for different types of users, as was shown in an earlier work [1]. In addition, a survey was undertaken to gauge users' understanding and reaction to possible types of privacy threats resulting from the knowledge of their location information.

Firstly, the dimensions of the problem are examined and the factors that can impact users' privacy are identified. These factors include, the type of data collected, its visibility and accessibility by users, as well as the possible exploitation of these data by the application. Secondly, an analytical study is conducted using a representative data set to explore the location data content and the range of possible inferences that can be made from them. The frequency of usage of the networking application is used to classify users and in the analysis of their behavioural patterns. Finally, a survey was undertaken to examine users' awareness and concerns with respect to privacy implications of their location data and their needs to control access to their data on GeoSNs. Previous studies explored users' privacy concerns and attitude when sharing their location for social purposes, but presented limited evaluations using restricted application scenarios [6], [7]. Questionnaire analysis demonstrate a strong feasibility of inference of users' personal information that may pose a threat to their privacy on these networks. The survey also reveals users' concerns about their location privacy and their motivation to control their location information. The outcomes highlight the need for further work on improving the visibility

of the information collected, to allow users to better understand the implications of their location sharing activities and assess their need to control access to their location data sets.

The rest of this work is organized as follows. Section II gives an overview of related work. In Section III, the dimensions of the location privacy problem in GeoSNs are discussed. Section IV describes the experiment conducted with a realistic data set to explore the spatiotemporal information content explicitly described and that may be inferred from the data. Section V builds on the results of Section IV by designing and deploying a questionnaire that explores users' awareness and attitude towards potential privacy threats. Discussion of the results and conclusions are presented in Section VI.

II. RELATED WORK

Security and privacy of online social networks is a general research area that includes evaluating potential privacy risks, as well as developing privacy-protection methods [8], [9], [10]. This paper focuses on the privacy implications of location-related information in GeoSNs. Two relevant questions to the problem studied are: to what extent is location privacy a potential concern for users in GeoSNs, and what sort of location-based inference is possible from the data collected in GeoSNs. In this section, related works on both issues are reviewed.

A. Users' Attitude and Privacy Concerns in Geo-Social Networks

Much interest has been witnessed over the past few years for studying users' attitude and concerns to location privacy and investigating how user-empowered location privacy protection mechanisms can influence their behaviour. Tsai et al. [6] developed a social location sharing application, where participants were capable of specifying time-based rules to share their location and were then notified of who viewed their locations. Their findings suggested that the control given to users for setting their sharing preferences contribute to the reduction of the level of their privacy concern.

Sadeh et al. [7] 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. [11] developed a system to represent actual users' workplace, offering live feeds about users and their location and asked users to define different levels of permissions for their personal information sharing. They found that participants were concerned most about their location information and that they utilised the permission feature to control this information. Another study by Kelley et al. [12] showed that users were highly concerned about their privacy especially when sharing location information with corporate-oriented parties.

Other works were carried out to examine how the employment of visualization methods may impact users' attitude to location privacy and behaviour. Brush et al. [13] studied users' attitude towards their location privacy when using GPS tracking over long periods of time and questioned whether using some obfuscation techniques can address their concerns.

Participants were concerned about revealing their home, identity and exact locations. They visually recognised and chose the best obfuscation techniques they felt can protect their location privacy. In addition, Tang et al. [14] investigated the extent of presenting various visualizations of users' location history on influencing their privacy concerns when using location-sharing applications. They developed text-, map-, and time-based visualization methods and considered spatiotemporal properties of sharing historical location. They noted that the majority of participants found visualization of location history to be more revealing and tended to prefer text-based presentation methods to limit the amount of data exposed.

With regards to public GeoSNs, there are relatively few research works that examine privacy concerns of users. Lindqvist et al. [15] 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. [16], 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 and 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 [6], [7], [11], [12]. Moreover, as far as we are aware, no studies so far have considered 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 [17], Twitter users' city-level locations were estimated by only exploiting their tweet contents with which it was possible to predict more than half of the sample within 100 miles of their actual place. Similarly, Pontes et al. [18] 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-thirds of the sample within 50 kilometres. Sadilek et al. [19] investigated novel approaches for inferring users' location at any 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. [20] 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 their social network information and geographical distance in LBSNs [21].

Other works focussed on investigating the potential inference of social relationships between users of GeoSNs. Crandall et al. [22] 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 et al. [19] also formulated friendship predictions that derive social relationships by considering friendship formation patterns, content of messages of users and their location. They predicted 90% of friendships with accuracy beyond 80%. Additionally, Scellato et al. [23] 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. In [24], they developed a link prediction system for LBSNs by utilising users' check-ins information and properties of places. 43% of all new links appeared between users with at least one check-in place in common and especially for those who have a friend in common.

Studying and extracting spatiotemporal movement and activity patterns of users on GeoSNs attracted much research in recent years. Dearman et al. [25] exploited location reviews on 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 questionnaire. Noulas et al. [26] studied user mobility patterns in Foursquare by considering popular places and transitions between place categories. Cheng et al. [27] examined a large scale data set 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 between consecutive check-ins, distance between users' check-ins and their centre of mass, as well as the returning probability to venues. They also studied factors affecting users' movement and found considerable relationship between users' mobility and geographic and economic conditions. More recently, Preotiuc-Pietro et al. [28] investigated the behaviour of thousands of frequent Foursquare users. They analysed users' movements including returning probability, check-in 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 predict users' future movement. The above studies show a significant potential for deriving personal information from GeoSNs and hence also 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 GeoSNs.

III. DIMENSIONS OF THE LOCATION PRIVACY PROBLEM ON GEOSNS

Four aspects of the data collected can be identified that can affect location privacy. These are: 1) the amount of data collected and its quality, 2) its visibility and accessibility, 3) its possible utilisation by potential users, and 4) the level of security offered to users by the application. This discussion focuses on the type of privacy-related questions that can be asked and the confidence level in the information that can be derived. Both factors can affect the degree of privacy concern to users. The study considers both LBSNs (Foursquare) and

LESNs (Twitter), the difference in the way location data are acquired in both and the issues implied.

A. Location Data Collection

Here the types of data, its density and quality, as well as the methods of collection and storage are considered.

1) *Method of Collection*: Both LBSNs and LESNs depend on the user device to acquire the user's current location using GPS, wireless access points (WAP) or cellular networks. When using LBSNs, location data are collected automatically since location is mandatory to providing the service. In Foursquare specifically, user's location is implicitly acquired on a continuous basis, even without using the service. User's check-ins into specific places are verified against their estimated current location and recorded explicitly. In LESNs, user's location data are collected only when location-based features are enabled and used. Some features require continuous collection of location data, for example, when tailoring trends to the user's location in Twitter. The mode of data collection, whether continuous or periodic; automatic or manual, will impact the volume of data collected and its accuracy, and hence also the degree of confidence in inferences made from the data.

2) *Types of Data*: The completeness and accuracy of location information are primary factors that determine the possible inferences made based on this information and the possible privacy threats to users. Three types of data can be associated with location data collected in GeoSNs: spatial, non-spatial and temporal.

- **Spatial semantics**: These refer to any type of information that can be used to identify the places visited. In both LBSNs and LESNs, user's location is identified as a point in space with a latitude and longitude. In LBSNs, users identify their locations explicitly, allowing for a rich definition of place identity, including place name, type classification and street address. On the other hand, location in LESNs is determined automatically by reverse geocoding the registered latitude and longitude coordinates, and thus carry a degree of inaccuracy and ambiguity. Increasingly, some LESNs are able to use resources from LBSNs for defining locations. For instance, Instagram allows users to geotag their pictures using the Foursquare API [29]. Twitter also uses Google API for linking users' selected place names with a location on a map. Hence, in both cases it can be assumed that detailed and precise place identities visited by users may be stored by the applications.
- **Non-spatial semantics**: Non-spatial semantics are other types of data about both users and places that may be associated with location information. These include explicit user data, as for example defined in their personal profiles on the application or place-related data, such as reviews, tags and pictures. With the user permission, applications will identify users and share their personal information. Rich place-related semantics may also be mined from resources on the web [30].
- **Temporal semantics**: These represent the time of user's visit to a place and the duration of their visit. In LBSN, the time of visit is registered by the user as they

check-in to a place. The user's physical presence in the place may be validated by comparing their actual GPS coordinates with those of the place they check into. In LESN, a time stamp is encoded with the resource used, for example, a tweet location. However, in this case it is difficult to ascertain whether the user is intentionally visiting the place or happened to be passing by it. In both cases, further processing of the user tracks is needed to estimate the duration of the user's visit.

3) *Data Volume*: The amount of location data collected is another important factor to be considered and is dependent of the user attitude and behaviour when using the application. The pattern of data logging and the frequency of usage will determine the density of the data collected over time and will thus influence the type of information that may be inferred from the data. For example, regular visits to specific places can determine routine mobility patterns, while incidental visits to other places can signify special events or activities.

B. Location Information Accessibility

Location information accessibility represents how much of the user's data are available and visible to others including the user, other users and third parties of the service. In terms of users' accessibility to their collected location and location-related data, GeoSNs provides only limited means for accessing these kinds of information. In Foursquare, users' previous check-in information are available in the form of check-in history, where users can view their visited venues, dates of visits and tips they made. These raw data provide only a limited view of the information content in the data, as discussed in the previous section. In Twitter, users can request to download their tweet history, but location information are not included in this data. As for information visibility, most of the users' information published on GeoSNs are available to their friends and can be visible to other users.

Generally, users of GeoSNs have limited control over the visibility and accessibility of their information by others, since the privacy settings provided to them is not adequate enough to manage all aspects of their information accessibility. In Foursquare, almost all of the user's information is publicly available by default and can be viewed by other users. This include 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. Similarly in Twitter, users' profiles and their tweets are public by default, and can be accessed by others. This means that location information attached with tweets is publicly available as well unless users mark their profile as private, where only followers can view their data. All of the publicly available users' information is accessible by third parties including the geo-social application APIs users. Third parties can also have privileges to access the user's personal information. In the case of Foursquare, third parties can get check-in data in anonymous form, but they also indicate that they will share user's personal information with their business partners and whenever is necessary in some situations, such as enforcement of law. Twitter, on the other hand, states that any content the user submits or displays through the service is available to their third parties without anonymity.

C. Location Data Exploitation

Location information exploitation refers to how the application or third parties can utilise the data and for which purposes. This dimension involves the actual exploitation of user's location and location-related data that lead to posing various levels of privacy threats. It seems that GeoSNs have unlimited rights to utilise their users data in any way, for any purpose as stated in their terms of use. For example, Foursquare gives itself absolute privileges over using and manipulating user information as stated in their terms of use [31].

"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)."

Similarly, Twitter has the right to utilise users data, including location information, in various ways, as stated in their terms of use [32].

"By submitting, posting or displaying Content on or through the Services, you grant us a worldwide, non-exclusive, royalty-free license (with the right to sublicense) to use, copy, reproduce, process, adapt, modify, publish, transmit, display and distribute such Content in any and all media or distribution methods."

It is clear therefore that there are no commitments on GeoSNs as to how the data may be used or shared by the application or by other parties. In addition, the reasons for the potential exploitation of users data are vague (e.g., to improve the services) or even not stated. Hence, by agreeing to the terms and conditions, users effectively are giving away their data and unconditional rights to the use of their 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. In general, the fact that data are stored somewhere on servers opens the doors for potential undeclared access and use, and hence it is almost impossible to guarantee the security of the user data. 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. Twitter states that "Twitter complies with the U.S.-E.U. and U.S.-Swiss Safe Harbor Privacy Principles of notice, choice, onward transfer, security, data integrity, access, and enforcement", but give no additional explanation or examples on situations or access methods that these laws apply to.

In the following section, a sample data set from a LBSN is used to explore and analyse the potential information content

that can be derived from the location data.

IV. EMPIRICAL INVESTIGATION

This analysis is carried out using a real-world data set from Foursquare, as a typical example of a LBSN. The purpose is to demonstrate possible privacy implications in terms of personal information inferences and exploitation from user activity on GeoSNs. The effect of location data density and diversity on the possible inferences that can be made is analysed.

A. Dataset

The Foursquare dataset used in this analysis is provided by Jin et al. [16]. 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. Places on Foursquare are associated with pre-defined and structured place categories, e.g., Home, Office, Restaurant, etc. The data set contains 60,853 local venues, 45,289 users and 1,276,988 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 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. Mainly, 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 [33].

C. Results

Analysis of the data set questioned the sort of implicit user-related information that can be considered to be private that may be extracted using the location data collected. 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 40.5111 and longitude -79.9934 at 9 a.m. on 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.

TABLE I: Statistics of user groups in the Foursquare dataset.

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: Profiles of 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 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.
- 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.

1) *The Moderate User*: The analysis results of *user9119* selected from the moderate group are as follows.

- a) 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 them 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, this 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.
- b) Spatiotemporal Movement Patterns: About 40% of this user's total check-ins occurs at 9 am, mostly in the 'Office' and at 7pm, mostly at 'Home'. More than two-thirds of the check-ins are between 10am and 2pm and between 6pm and 11pm, which indicates that this user commutes more frequently during these hours. From the weekly patterns of movement, it can be seen that 71% of the venues were visited after 6pm. Mondays and Thursdays are when this user is most active, representing 41% of the check-ins. *User9119* tends 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

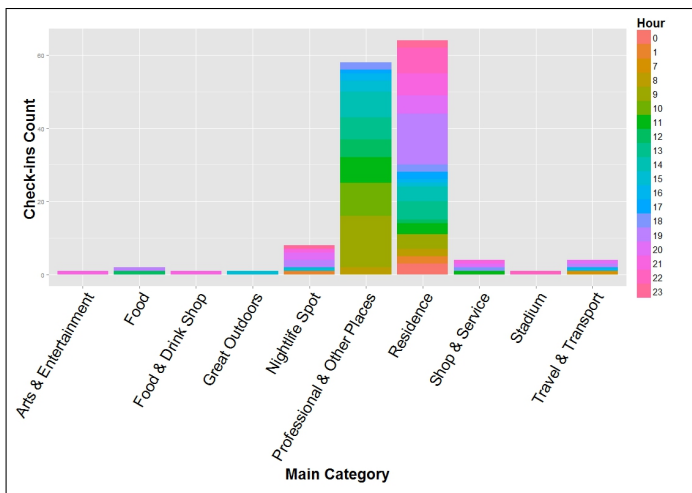


Figure 1: The moderate user's check-ins count, classified by the category of venues for different hours of the day.

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 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 2 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 personal 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 user has not checked in any place for a period of a week between the 21st and 28th of April. *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.

- c) Degree of Association with Other Users: Co-location is used here to denote that users have visited the same venue at the same time. This can be used as a measure of interest in a place and relationships between users. *User9119* was co-located in 6 unique venue categories with two (out of twenty) friends. He shared three co-occurrences with two friends; once with *friend1236* at 'American Restaurant' and twice with *friend15229* at 'Office', which may 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.

2) The Frequent User: Analysis of results of *user7105* from the frequent user group is as follows.

- a) 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 those are to 'Gas Station or Garage' and 25% are to 'Drugstore or Pharmacy'. *User7105* is 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, 'Medical Center'.

- b) Spatiotemporal Movement Patterns: Generally, about 20% of the check-ins occurs from 10am to 12pm, half of which are at 'Home'. In addition, *user7105* tends 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 12am and 3am 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 12am, whereas Monday and Saturday at 4pm. Furthermore, this user has visited more diverse venues on Tuesdays followed by Thursdays and Wednesdays representing 53%, 43% and 38%, respectively of the total visited categories. During the working week, this user tends to visit a 'Bar' (5%), especially on Tuesdays, and 'Gas Station or Garage' (4%). This is 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. *User7105s* check-in patterns were regular over the whole period. However, visits of this user are more frequent and diversified in the month of March. Noticeably, about 28% of the check-ins between 12am and 3am occurred in March, indicating a possible change of lifestyle. Figure 3 presents this user's check-ins count in different categories of venues, classified by day and grouped by month.

- c) 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 *friend38466* 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 *friend38466* at 6 different categories including mostly 'Bar', 'American Restaurant', and 'Sandwich Place', which

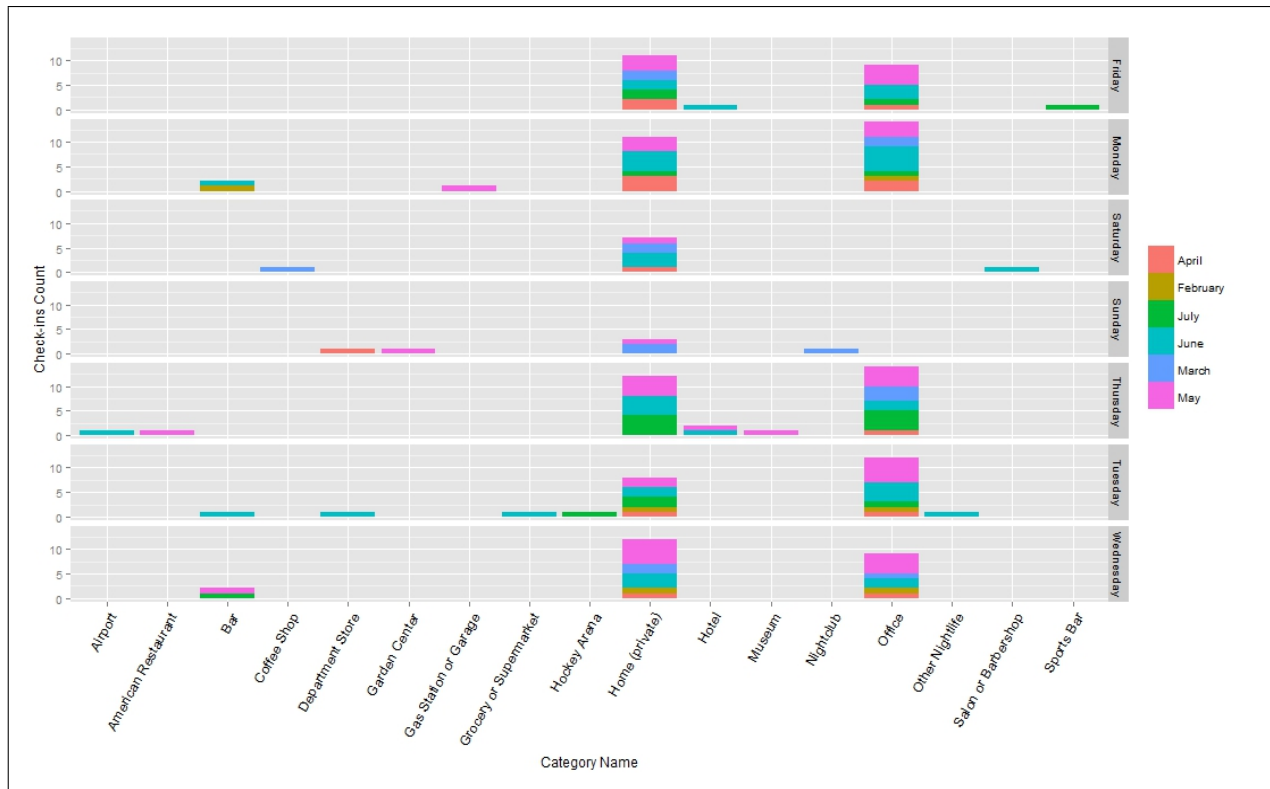


Figure 2: The moderate user’s check-ins count in different categories of venues, classified by day and grouped by month.

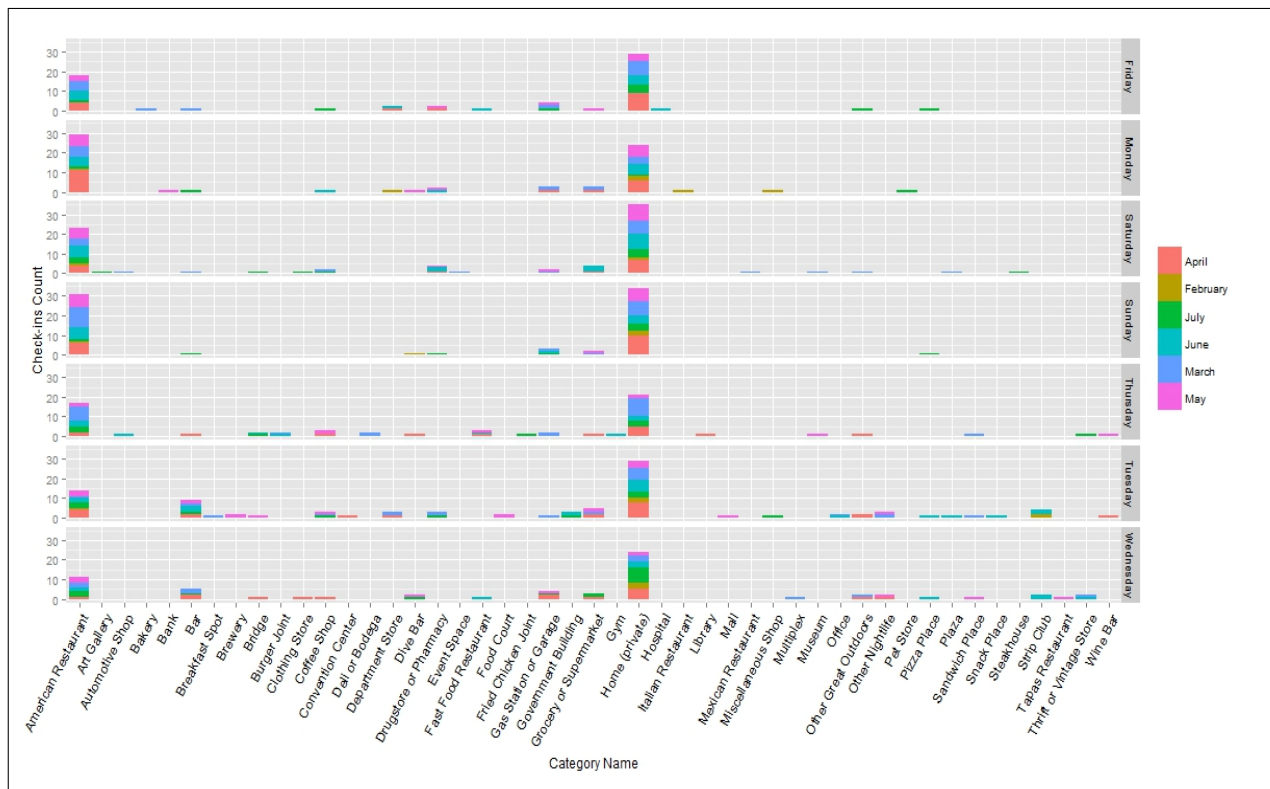


Figure 3: The frequent user’s check-ins count in different categories of venues, classified by day and grouped by month.

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 4. Similarly, this user has 89 co-occurrences with other users, who are not stated as friends, at 29 unique venues, where 38% of these co-occurrences are at 'American Restaurant' and 24% at 'Plaza'.

3) *The Hyper-Active User*: The results of analysis for *user2651* selected from the hyper-active user group are as follows.

a) **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 indicate the user's interests and activities - swimming and Italian food in this case. 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 [34], which may indicate 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%. The location of the visited venues can be clustered into two main areas on a map as illustrated in Figure 5. One area includes 'Home' 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'.

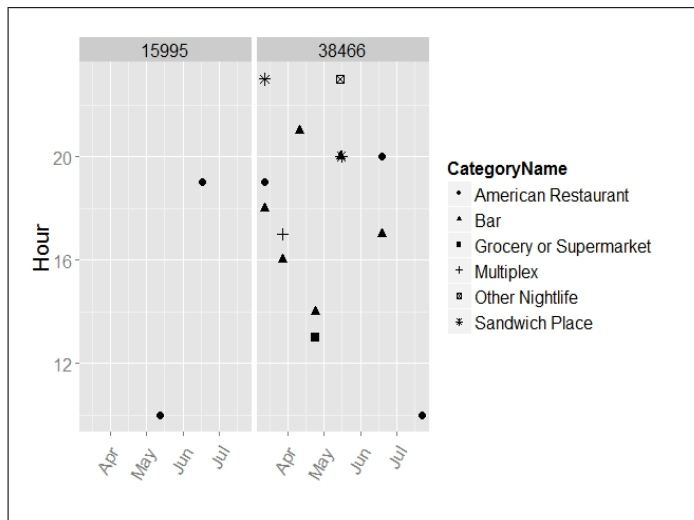


Figure 4: Spatiotemporal tracks of the frequent user co-occurrences with friends.

b) **Spatiotemporal Movement Patterns**: Overall, 53% of residential check-ins occurs between 9am and 12pm. A significant number of check-ins (10%) occur at 2pm, of which almost two-thirds occur in an 'Automotive Shop'. Check-in frequency reaches another peak between 11pm and 12am (18%), of which more than half are in 'Nightlife Spot'. Noticeably, this user tends to be more active at night, where about 70% of the check-ins are registered after 6pm. In his case, weekends have similar check-in frequency as the working week, but Sundays register as the most active day in terms of check-in frequency. Moreover, *user2651* checks in considerably less frequently at the 'Automotive Shop' and the 'Pool' on Wednesdays and Fridays, but checks in the 'Automotive Shop' and 'Nightlife Spot' in weekends. This may indicate that he works shifts on weekends.

User2651 has regular check-in patterns over the whole period. However, in the months of June and July, check-ins into 'Hotel' and 'Pool' significantly increased representing 75% and 60%, respectively of these venues total check-ins. Figure 6 demonstrates this user's check-ins count in different categories of venues, classified by day and grouped by month.

c) **Degree of Association with Other Users**: As with other users, *user2651* was co-located with 23 users at 12 distinct venues, half of these co-occurrences happened in 'Bar', 'Automotive Shop' and 'Grocery or Supermarket'. *User2651* is co-located in 27 unique venues from 19 categories with 9 friends, 13 of which are with friend12432 and 9 with friend12046. Most of the co-locations are in 'Nightlife Spots', 'Gas Station or Garage', 'Pool', 'Flower Shop' and 'Bar'.

The three dimensions analysed above will form the basis of the questionnaire design described in the next section.

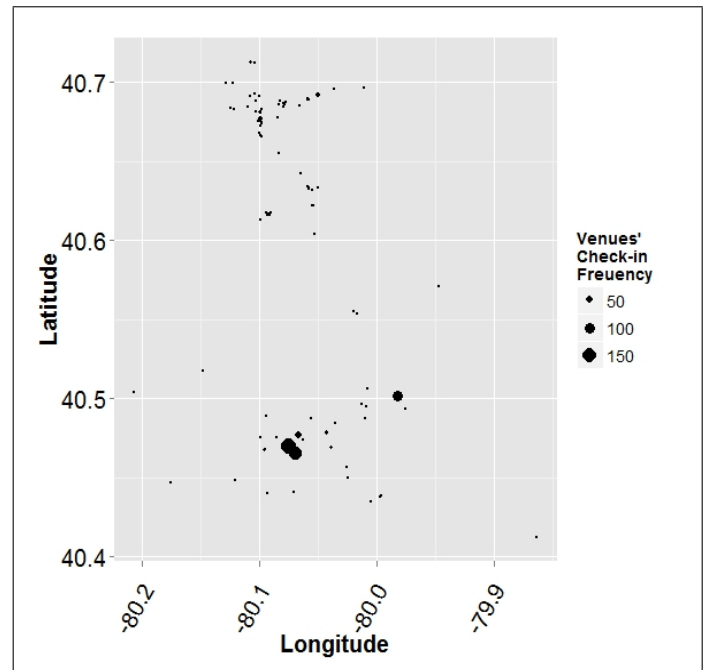


Figure 5: Coordinates of venues visited by the hyper-active user, considering the frequency of visit.

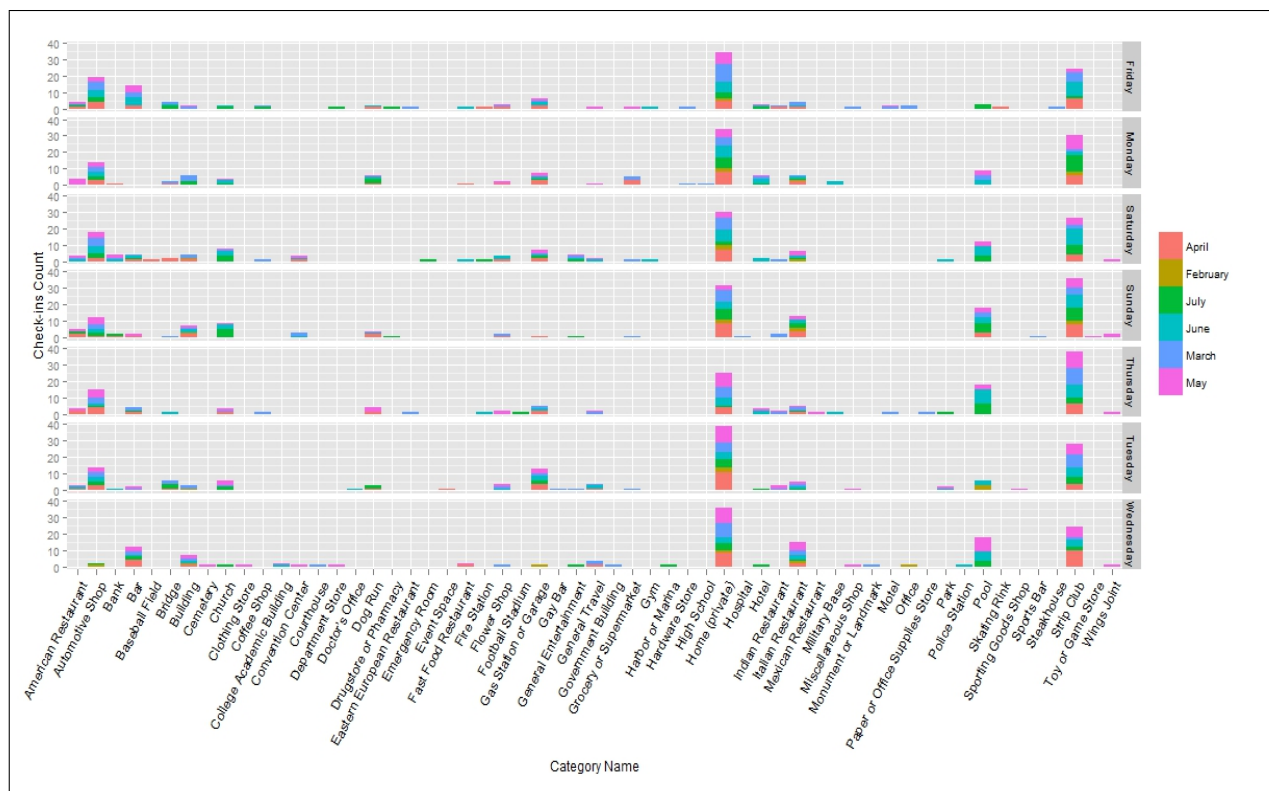


Figure 6: Count of check-ins for the hyper-active user in different categories of venues, classified by day and grouped by month.

V. USER STUDY

None of the related studies reviewed in Section II above has fully explored or focused on improving users' full awareness and understandability of the potential privacy implications when sharing their location information on GeoSNs. Here, a survey is undertaken to examine the privacy concerns and behaviour of users of online social networks, in particular users' concerns towards their location information. Three main aspects are addressed in this study: the extent of users' awareness of the terms of use they sign up to when using these applications, their understanding and attitude to potential privacy implications, and how they may wish to control access to their personal information on these applications.

A. Study Design

The questionnaire was developed using Google Forms. Targeted participants were users of online social networking applications who use location features, e.g., adding location to their posts and photos and checking-in when visiting places. A pilot study was first carried out to ensure the clarity and coherence of the survey. Four volunteers with no specific background completed the survey and provided valuable feedback into the wordings and layout of the questions used. The survey was then disseminated widely within the university to staff and students and was also advertised on social networks through the author's account. A token incentive of £10 Amazon vouchers was offered to ten randomly chosen participants who completed the survey.

The questionnaire consists of four main sections. The first section collects background information on the participants

and their use of GeoSNs. The next section examines users' knowledge of terms of use and privacy policies of the applications, followed by a section on studying perception of possible inferences of personal information. The last section is intended to capture users' attitude to privacy on social networks as well as their attitude to controlling their personal information.

B. Results

The questionnaire data were analysed using the R statistical package and the results are presented below. 186 participants completed the survey of which 60% are young adults in the age group 15-24, divided almost equally between males and females. The vast majority of participants (77%) use the services frequently (several times a day) and 72% of participants use the location services in GeoSNs. About 60% use location features on only one application. Adding locations to posts and pictures on Facebook was the most used application, corresponding to 47% of the total number of location services used. This is followed by adding location to tweets on Twitter, photo mapping pictures on Instagram, and checking-in on Foursquare representing 17%, 16% and 10%, respectively as illustrated in Figure 7. In addition, most of the users noted that they 'sometimes' use geosocial applications with almost a fifth of users 'always' using the location services. Foursquare users are more frequent users of the service than other services and 25% of the users have linked their accounts on different social networking applications. The questionnaire is divided up into four sections, was presented to participants in whole and takes roughly about 10 minutes to complete.

In what follows, the results from the different sections of the questionnaire are analysed.

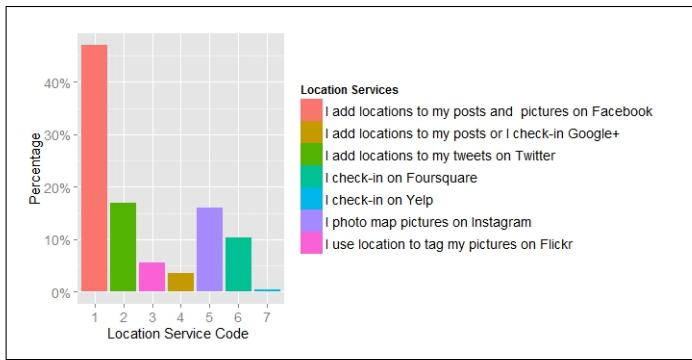


Figure 7: Percentage of the type of location services used by the 186 participants.

1) *Knowledge of Terms of Use and Privacy Policies for Social Networking Applications:* Here, the awareness of the terms of use and privacy policies are examined and analysed against users' profiles. In general, the majority of the users (81%) have not read terms of use or privacy policies of the social networking applications they use. Users were presented with the following typical statements representing the terms of use relating to location information and were asked to indicate whether they are aware of the information in the statements. Note that the following statements are representative of the terms of use of all the GeoSNs in question. The results are shown in Figure 8 grouped by the frequency of use.

- *Term 1: The application collects and stores your precise location (as a place name and/or a GPS point), even if you mark your location as private, for a possibly indefinite amount to time.*
- *Term 2: The application can use your location information in any way possible including sharing it with other applications or partners for various purposes (commercial or non-commercial).*
- *Term 3: If you share your location information, your friends and any other users are able to access and use it in any way possible.*
- *Term 4: The application can collect other personal information, such as your personal profile information and browsing history from other web applications.*

More than half (53%) of users acknowledged awareness of all of the statements and of those 73% have read the terms and policies. Most users (75%) are aware of statement 3, relating to the sharing of information with friends, but are generally unaware of statements 1 and 4, relating to how their location and other information may be collected and stored by the applications. It is interesting to note that frequent users of such application are generally unaware of such statements (49%) as demonstrated in Figure 8. Younger users aged between 15 and 34 tend to be more knowledgeable of these policies (60%), but gender does not seem to be a factor in these results.

2) *Perceptions of Possible Privacy Implications:* In this section, users' attitude towards the inference by the application of personal information is examined. In particular, the questions aim to gauge users' awareness of plausible inferences about their private places, activities at different times, their connections to other users, and possible knowledge of this

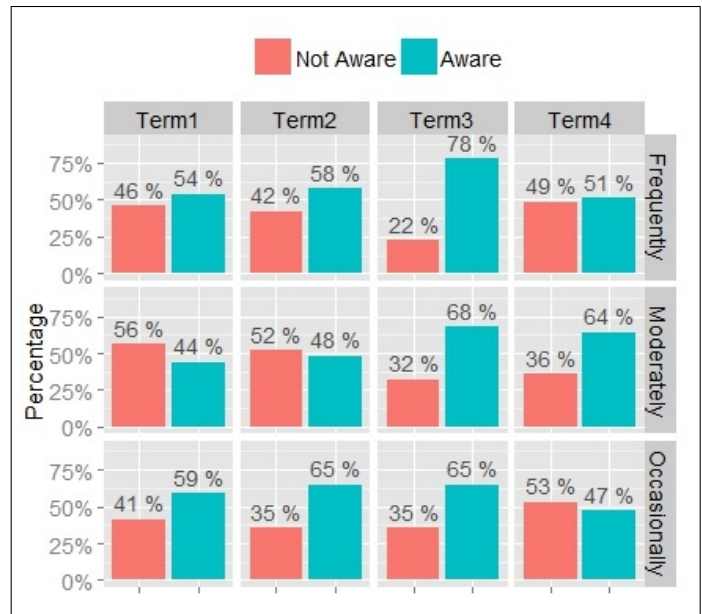


Figure 8: Users' awareness of general terms and policies of GeoSNs (Term1-Term4) grouped by the frequency of use.

information by the application. Participants were presented with 14 statements, shown below. They were then asked to indicate, for each statement, whether they are aware that the statement is possible and to score their reaction to the possibility of this statement as either 'OK', 'Uncomfortable' or 'Very Worried'. The first twelve statements refer to knowledge by the application itself, while the last two statements are reflection of the terms of use that suggest that the application can share the user's data with other users and third parties.

- *S1: I can guess where your home is.*
- *S2: I can guess where your work place is.*
- *S3: I know which places you visit and at what times.*
- *S4: I can tell where you normally go and what you do in your weekends.*
- *S5: I can tell you where you go for lunch or what you do after work.*
- *S6: I know your favourite store (your favourite restaurant, your favourite coffee shop, etc.)*
- *S7: I can guess what you do when you are in a specific place.*
- *S8: I can guess when you are AWAY from home.*
- *S9: I can guess when you are OFF work.*
- *S10: I know who your friends are.*
- *S11: I know when and where you meet up with your friends.*
- *S12: I can guess which of your friends you see most.*
- *S13: Other people can know where you are at any point in time.*
- *S14: Other people can know what you are doing at any point in time.*

In terms of awareness, users seem to be most aware of statements S1, S2 and S10, regarding the location of home, place of work and friends, representing 88%, 89% and 93%, respectively. On the other hand, users are least aware of statements S5, S13 and S14 that relate to other users' knowledge of personal mobility patterns and activities, representing 34%, 37% and 40%. The awareness level of the users is demonstrated in Figure 9 grouped by the frequency of use.

Despite a reasonable level of awareness about the plausibility of these statements, users seemed to be relatively concerned about their privacy. 66% of users' reactions were either uncomfortable (41%) or 'very worried' (25%) as can be seen in Figure 10(a). Over half of the responses to S2 (awareness of workplace-53%) and S10 (awareness of friends-65%) were not concerned.

On the other hand, participants were most concerned with S13 and S14, with the 'Very Worried' category scoring 83% and 84%, respectively. S1 and S11, relating to the location of home and meetings with friends were rated most 'Uncomfortable' corresponding to 53% and 51%, respectively. Statement S8, suggesting the knowledge of user's absence from home and S13, indicating the possible knowledge of this information by other people presented a significant source of worry to users, with 45% and 42%, respectively indicating that they are 'Very Worried' about these statements.

It appears that users who read the terms and policies are more aware (by 9%) of the statements, while users who have not read the terms and policies were significantly 'Very Worried' (by 21%) than other users. Moreover, there is a positive correlation between the age of the participant and their level of awareness; level of awareness considerably increases with increase in age group, with the oldest active age group (35 to 44 years) scoring 89%. Yet, younger users, in the age group 15 to 34 years, tend to be relatively less concerned than older users (by 4%). The level of users' concern increases with the decrease in the frequency of use of the applications, where 76% of occasional users are concerned compared to 63% of frequent users. Users of Facebook and Instagram registered the highest degree of concern among all users of GeoSNs scoring 63% and 62%, respectively as shown in Figure 10(b). Again, gender does not seem to have any significant influence in this study.

3) *Attitude to Privacy on Social Networks:* The aim of this section of the questionnaire is to understand the users' reaction with regards to using the applications, given the knowledge of potential implications on privacy from the previous section.

61% of users stated that they would change the way they share their location information, 55% of whom are willing to stop sharing their location information completely, with the rest of the group indicating they would share it less often. Frequent users seem to be the most motivated to change their sharing behaviour (13% more than infrequent users), as illustrated in Figure 11, but they are also less willing to stop sharing the information and would prefer to share less frequently than the infrequent users (by 47%). Interestingly, users of location services are more tempted (by 10%) to change how they disclose their location information compared to users who have not used them. 57% of the first group of users want to share their location less frequently and 43% are willing to



Figure 9: Users' awareness about potential information inferences (S1-S14) grouped by frequency of use of GeoSNs.

discontinue disclosing their location data. Younger users (15-34) are more willing to change their usage behaviour (by an average of 18%) and are even more willing to stop sharing location information completely (by an average of 10%) than older users. In this case, it seems that female users are more motivated to change their attitude regarding location disclosure (by 11%) than males, yet 60% of male participants suggested their willingness to discontinue using location services.

4) *Managing Personal Information:* In this section, users' views on managing and controlling access to their location information are explored. This includes several aspects related to what information is stored, how it is shared or viewed by the application and by others, and whether users need to manage access to their information. The following statements were presented to the participants who were asked to rate how often they would use them: 'All the time', 'Occasionally' or 'Never'.

- C1: I would like to be able to turn off location sharing for specific durations of time.
- C2: I would like to turn off location sharing when I visit specific types of places.
- C3: I would like to decide how much of my location information history is stored and used by the application for example use only my check-in history for the last 7 days.
- C4: I would like to see the predicted personal information that the application stores about me based on my location information.
- C5: I would like to decide how people see my current location for example, exact place name, or a rough indication of where I am.
- C6: I would like to decide who can download my location information data.



Figure 10: (a) Users’ reaction towards potential inferences grouped by the inference statements (S1-S14). (b) Data in (a) grouped by the GeoSNs used.

- C7: I would like to know, and control, which information can be shared with other Web applications.
- C8: I would like to make my location information private seen only by myself and by the people I choose.

Results are given in Figure 12(a) and show a significant desire to use these controls for location privacy. Overall, 76% of participants would like to apply those controls 'All the time', 20% are happy to apply them 'Occasionally', and only 4% of users will not consider these controls.

In general, C2, C6, C7 and C8 were most favoured controls, scoring over 97% each of users’ responses. Controls C1, C6 and C7 were the most chosen controls to be applied all the time, representing 91%, 88% and 86% of users’ responses, respectively. It is worth noting that users of different location services have similar acceptance rate for these control. Foursquare and Facebook users have the highest preference for applying the controls 'All the Time', corresponding to 76% and 75%, respectively.

A negative correlation appears to exist between users’ tendency to use these privacy controls all the time and their age group. The youngest active age group of 15-24 years old has the highest desire for all-the-time application of controls representing 78% of this group’s responses.

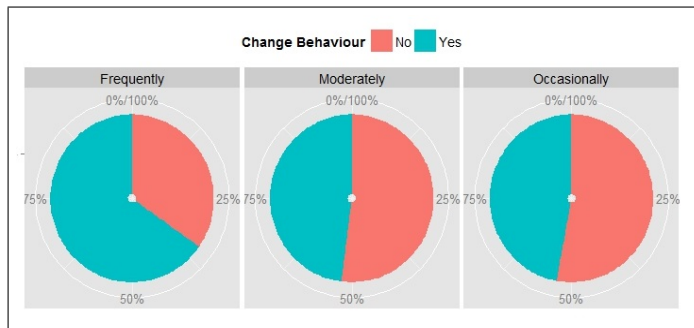


Figure 11: Users’ attitude to location privacy risk, grouped by frequency of use of GeoSNs.

As expected, users who are tempted to change their location sharing behaviour have relatively higher motivation to use these controls representing 97% of this group’s responses (4% higher than users who are reluctant to change). The factors of gender, whether users read the applications’ terms or how frequent they use the social networks, as shown in Figure 12(b), seem to have minimal influence on their willingness to use these controls. In the future it will be useful to undertake a longitudinal study that tracks user behaviour over time to understand the factors that may influence their attitude to location privacy, for example the impact of friends and age group.

VI. DISCUSSIONS AND CONCLUSIONS

The proliferation of location-based GeoSNs and the large-scale uptake by users suggest the urgency and importance of studying privacy implications of personal information collected by these networks. Identifying user profiles is a goal of many businesses that is now commonly accepted by users for the purpose of improving the quality of service. However, GeoSNs do not explicitly present similar business goals and thus their motivations for collecting and sharing personal location information are not clear. Also, the issue is complicated as the data collected may be shared or accessed by other users and applications. The results of this study highlight the possible implications to user privacy and the need for developing means for raising the user awareness of these issues, and possibly also giving the user control on managing access to their data.

The data analysis experiment conducted here shows the amount and types of personal information that can be inferred using location data. Users’ spatiotemporal mobility tracks can be analysed to identify where they are, where they are likely to be, and sometimes more significantly, where they are not present. Tracking user location data may also give indications to their preferred activities, places, habits and friendship community.

As can be expected, the more frequent the applications are used, the more dense the spatiotemporal history of user data collected and the more certainty in the derived information extracted from this data. Whilst the statistical analysis carried

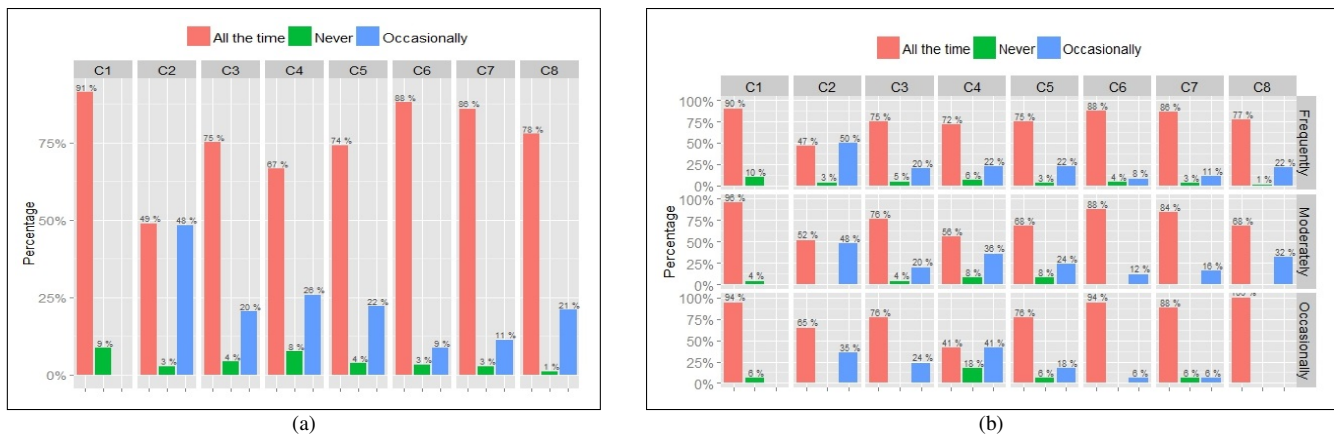


Figure 12: (a) Users' desire to use location privacy controls grouped by statement of controls C1-C8 (see section V-B4). (b) Data in (a) grouped by frequency of use of GeoSNS.

out in this study highlight some of the basic and interesting inferences that can be made, more sophisticated location-based inference methods can be developed to infer, for example, the probability of future movements, methods of transport and places visited. The now common practice of linking user accounts in several GeoSNSs increases the availability of data and compounds the privacy risks to users, who sign up to different, possibly contradicting, terms of use and policies of different applications. For example, developers now use the Twitter API to collect user check-ins in Foursquare.

The questionnaire conducted in Section V provides valuable insights that convey many aspects of location privacy on the Social Web from the perspective of the end user. The main and (possibly only) means of communicating how the collected user information may be used and exploited by the application is described in the application's terms of use. It is clear from the results of the questionnaire undertaken that the majority of users, especially those who use location services, do not read the terms of use and policy documents. The findings also indicate that users are aware of the potential information, and possible derivatives thereof, stored by the application. However, it appears that they are also quite concerned about the privacy implications. This apparently contradicting findings may be due to that such awareness and concerns are evident when users are actively questioned about these issues, but are somewhat screened from the users' minds during the continuous use of the application. The study also suggests that users may not fully understand the privacy implications, where their level of concern was much more pronounced when faced with statements that indicate that other people may be aware of their location information in comparison to statements indicating that the application holds such information.

The study reveals that there is a strong need for the users to be continuously aware of their data, how it is stored and to have the ability to control access to and visibility of their location data sets. Further research is needed into methods that enhance the communication of the information by the applications as well as method to allow users to better understand and control their personal profiles on such networks.

REFERENCES

- [1] F. Alrayes and A. Abdelmoty, "Privacy concerns in location-based social networks," in *GEOProcessing 2014: The Sixth International Conference on Advanced Geographic Information Systems, Applications, and Services*. IARIA, 2014, pp. 105–114.
- [2] C. Vicente, D. Freni, C. Bettini, and C. Jensen, "Location-related privacy in geo-social networks," *IEEE Internet Computing*, vol. 15, no. 3, 2011, pp. 20–27.
- [3] D. Riboni, L. Pareschi, and C. Bettini, "Privacy in georeferenced context-aware services: a survey," in *Privacy in Location-Based Applications*. Springer Verlag, 2009, pp. 151–172.
- [4] 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.
- [5] R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, and J.-P. Hubaux, "Quantifying location privacy," in *IEEE Symposium on Security and Privacy*, 2011, pp. 247–262.
- [6] J. Tsai, P. Kelley, P. Drielsma, L. 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.
- [7] N. Sadeh, J. Hong, L. Cranor, I. Fette, P. Kelley, M. Prabhakar, and J. Rao, "Understanding and capturing peoples privacy policies in a mobile social networking application," *Personal and Ubiquitous Computing*, vol. 13, no. 6, 2008, pp. 401–412.
- [8] H. Gao, J. Hu, T. Huang, J. Wang, and Y. Chen, "Security issues in online social networks," *Internet Computing, IEEE*, vol. 15, no. 4, 2011, pp. 56–63.
- [9] J. Nagy and P. Pecho, "Social networks security," in *Emerging Security Information, Systems and Technologies, 2009. SECURWARE'09. Third International Conference on*. IEEE, 2009, pp. 321–325.
- [10] P. Joshi and C.-C. Kuo, "Security and privacy in online social networks: A survey," in *Multimedia and Expo (ICME), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1–6.
- [11] 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.
- [12] P. Kelley, M. Benisch, L. 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.
- [13] A. 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, pp. 95–104.

- [14] K. Tang, J. Hong, and D. 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.
- [15] 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 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11, 2011, pp. 2409–2418.
- [16] L. Jin, X. Long, and J. 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.
- [17] Z. Cheng, J. Caverlee, and K. Lee, "You are where you tweet: a content-based approach to geo-locating Twitter users," in Proceedings of the 19th ACM international conference on Information and Knowledge Management CIKM '10, 2010, pp. 759–768.
- [18] 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.
- [19] A. Sadilek, H. Kautz, and J. Bigham, "Finding your friends and following them to where you are," in Proceedings of the fifth ACM international conference on Web Search and Data Mining, WSDM '12, 2012, pp. 723–732.
- [20] 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. 1140–121.
- [21] —, "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, CIKM '12, 2012, pp. 1582–1586.
- [22] D. Crandall, L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. Kleinberg, "Inferring social ties from geographic coincidences," in Proceedings of the National Academy of Sciences of the United States of America, vol. 107, no. 52, 2010, pp. 22436–22441.
- [23] S. Scellato, A. Noulas, R. Lambiotte, and C. Mascolo, "Socio-spatial properties of online location-based social networks," in ICWSM, 2011, pp. 329–336.
- [24] S. Scellato, A. Noulas, 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, pp. 1046–1054.
- [25] 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.
- [26] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil, "An empirical study of geographic user activity patterns in foursquare," in ICWSM, 2011, pp. 70–73.
- [27] Z. Cheng, J. Caverlee, K. Lee, and D. Sui, "Exploring millions of footprints in location sharing services," in ICWSM, vol. 2010, 2011, pp. 81–88.
- [28] D. Preotiuc-Pietro and T. Cohn, "Mining user behaviours: a study of check-in patterns in location based social networks," Web Science, 2013.
- [29] "Instagram location endpoints," Nov. 2014. [Online]. Available: <http://instagram.com/developer/endpoints/locations/#>
- [30] A. Popescu, G. Grefenstette, and P.-A. Moëllic, "Gazetiki: automatic creation of a geographical gazetteer," in ACM/IEEE Joint Conference on Digital Libraries. ACM, 2008, pp. 85–93.
- [31] "Foursquare terms of use," Nov. 2014. [Online]. Available: <https://foursquare.com/legal/terms>
- [32] "Twitter terms of use," Nov. 2014. [Online]. Available: <https://twitter.com/tos>
- [33] "R project," Nov. 2014. [Online]. Available: <http://www.r-project.org>
- [34] "World domination summit," Nov. 2014. [Online]. Available: <http://worlddominationsummit.com/faq/#primary-content>