Taylor & Francis
Taylor & Francis Group

Journal of Location Based Services

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tlbs20

User profile visualisation for privacy awareness on Geo-Social Networks

Fatma S. Alrayes, Alia I. Abdelmoty & Waleed El-Geresy

To cite this article: Fatma S. Alrayes, Alia I. Abdelmoty & Waleed El-Geresy (26 Sep 2024): User profile visualisation for privacy awareness on Geo-Social Networks, Journal of Location Based Services, DOI: 10.1080/17489725.2024.2399512

To link to this article: https://doi.org/10.1080/17489725.2024.2399512

Published online: 26 Sep 2024.

JOURNAL OF LOCATION BASED SERVICES
https://doi.org/10.1080/17489725.2024.2399512

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS

User profile visualisation for privacy awareness on Geo-Social Networks

Fatma S. Alrayes[a], Alia I. Abdelmoty[b] and Waleed El-Geresy[c]

[a]Information Systems Department, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia; [b]Cardiff School of Computer Science & Informatics, Cardiff University, Cardiff, Wales, UK; [c]Department of Electrical & Electronic Engineering, Imperial College London, London, UK

ABSTRACT

Geo-Social Networks (GeoSNs) enable user interaction by sharing personal location and contextual information, leading to the accumulation of large databases of users' physical presence and experiences in geographic places. These data form the basis of building (geo)graphic profiles for users on GeoSNs. Users' awareness of the data they share is limited. Without active reminders of the contents of their geoprofiles, users may fail to recognise possible privacy risks associated with their location sharing behaviour. In this work, we argue that users' awareness on GeoSNs should consider the spatial, temporal, and social dimensions of the data, as well as their combinations. We propose a basic strategy for representing and visualizing geoprofiles that aggregates the interaction of users in space over time to summarise the degree of relatedness of users to place and to other users. We conducted a user study with 26 participants to assess the perception of their geoprofiles on a typical GeoSN against the proposed approach. Results demonstrate the potential value of the proposal for improving user awareness of their data and privacy implications, compared to basic search and retrieval functions offered by typical GeoSNs. This work highlights the need for considering and improving user awareness on these platforms.

ARTICLE HISTORY
Received 14 February 2023
Accepted 28 August 2024

KEYWORDS
Location privacy; privacy awareness; usable privacy; privacy on social networks

1. Introduction

Location is a key identifier of a person's profile that is collected by the devices we use and exploited by the services we consume on the internet everyday. Knowing where a user is at any point in time is crucial to many of the types of services and products, both physical and virtual, which are offered and tailored to their contexts and needs. Social networks are

CONTACT Alia I. Abdelmoty ✉ AbdelmotyAI@cardiff.ac.uk

established as primary sources of communication and social interaction for a large percentage of the population. Personal location information is either implicitly determined by these networks, for example, from the IP address of a user's device, or explicitly mentioned and shared by the users themselves when communicating with others on geo-services on these networks. These networks are referred to in this paper as Geo-Social Networks (GeoSNs), to emphasise the fact that a person's location is being collected and used for the benefit of both the user and the application. Over time, these applications are able to compile a detailed historical database of their users' existence, combining their location with other semantic information, such as the types of places visited. Users are rarely aware of the extent of information they share or its potential in revealing personal information about themselves (F. Alrayes and Abdelmoty 2017; Preotiuc-Pietro and Cohn 2013; Rossi and Musolesi 2014). In the case of GeoSNs, sharing location information may be used to infer where a person is at some point in time, what activities they may be doing, with whom, their absence from their home, etc. As applications become more empowered by intelligent algorithms to mine and analyse data, users must be empowered to better know and understand their data and implicit content. Giving individuals more control over their personal data is recognised as a right in emerging legal frameworks, such as the General Data Protection Regulation (EU) 670/2016 (GDPR) (EUR-Lex 2018), and is likely to be enforced and monitored, hence shaping the expectations of users in years to come.

Previous studies have examined the impact of location privacy on users' sharing behaviour (Christin, Michalak, and Hollick 2013; Patil et al. 2014; Sadeh et al. 2009; Tsai et al. 2009), primarily considering the user's awareness of the visibility and accessibility of his or her location information at any point in time. However, less attention has been given to users' comprehension of data mining on accumulated data and the inferences that could be made from combining different types of spatial and semantic data over time. The results of this kind of data analysis are constantly shaping the user's profile on such applications, despite not being made explicit to the user.

In this work, we argue that users' right to control their data also includes, by extension, their right to awareness and comprehension of their data, both that which is shared by the users themselves and that which is inferred by applications over time. This paper focuses primarily on the problem of exposing the implicit information in the location data that are shared by users and its value in improving users' awareness of their information and its privacy implications. We hypothesise that improving the representation of location information as it is presented to users, making potential inferences about personal data more transparent, improves user awareness of the impact of shared data and the associated privacy implications. To test this hypothesis, an experiment with 26 users was designed and conducted to measure how the proposed visualisation of this implicit information impacts users' attitude towards privacy and may lead to changes in behaviour on GeoSNs.

The contributions of this paper are as follows: 1) We explore the question of what aspects of the geoprofiles need to be exposed to users. An analysis is presented of the core elements of personal location data as shared and collected by GeoSNs. In addition to the basic spatial and temporal data, representing the user's visits to places, we propose that geoprofiles should also reveal the social element of user interaction by representing the interests of users in places and the connections between users as a result of their visits to places. 2) We address the question of how to enable the users' awareness of this information. We propose a visualisation approach to make the identified geoprofile information explicit to the users. 3) Finally, we address the question of evaluating users' awareness of their geoprofiles. We designed and conducted an experiment with 26 users of a typical GeoSN application. A prototype was developed that implemented the proposed geoprofile representation and visualisation approach. An analysis was made of the users' perception of their geoprofiles on the GeoSN against their perception of the profiles represented in the prototype. Results of the experiment are presented in detail and demonstrate the enhanced user awareness of privacy risks with the proposed approach.

In section 2, an overview is given of the privacy implications of location disclosure as well as to approaches to feedback and control methods related to location privacy. Analysis of the information content in geoprofile data and the proposed approach to visualisation are presented in section 3. The approach is implemented in a prototype interface in section 4 that is then used in section 5 as the basis for the evaluation experiment. Results are analysed in section 6, followed by a discussion of the study and its limitations. Conclusions and an overview of future work are given in section 7.

2. Related work

Location privacy has been defined as 'the ability of an individual to move in public space with the expectation that under normal circumstances their location will not be systematically and secretly recorded for later use' (Blumbery and Eckersley 2009; Liu et al. 2018). This definition highlights the significance of user awareness of their location information, that is being collected and shared, and its potential in revealing personal and private information. Location information combines spatial, temporal, and user identity data, with the capacity to link disparate datasets and reveal users' activities and associations with places and with other users (Liu et al. 2018). The current state of technological development and the ubiquity of personal devices that almost, by default, assume access to user location information imply a constant risk to user privacy. Location privacy-preserving mechanisms (LPPMs) are methods that can be employed by data collection applications to address the noted privacy risks above. Two approaches are identified: anonymity approaches, in which the user identity is dissociated from personal information, and obfuscation approaches, in which the quality of personal location information is deliberately degraded. While these methods are useful, their use is limited in the case of GeoSNs, whose functions assume precise knowledge of the user's location and the sharing of this information between the users of the application. For example, users notify others of their presence in places and users searching for friends in nearby locations. In this section, we present some research results on the privacy implications of location information disclosure and methods of offering feedback, as well as control, to users on their location privacy.

2.1. Privacy implications of location information disclosure

There has been significant interest in research studying the value and utility of location information on GeoSNs to understand users' behaviour. Studies have considered the connection between GPS trails shared on social networks and users' home locations (Cheng, Caverlee, and Lee 2010; Pontes et al. 2012), the implicit location information shared while driving (Bellatti et al. 2017), and the inference of users' locations from the location of their friends on Twitter (Sadilek, Kautz, and Bigham 2012). Using the user's profile of visited places and socio-historical ties, (Huiji Gao and Liu 2012) demonstrated accurate prediction of future check-in information, and (Gu et al. 2016) were able to identify the user's home location. Other works have investigated the potential inference of social relationships between users of GeoSNs. For instance, users' co-occurrence in place, as extracted from geo-tagged Flickr photos, was sufficient for deducing, with high probability, the nature of their social ties (Crandall et al. 2010) and friendship links (Scellato, Noulas, and Mascolo 2011). Mobility patterns on Foursquare were studied by (Noulas et al. 2011) to identify popular places and to detect transition patterns between place categories, while (Preotiuc-Pietro and Cohn 2013) used the distance between consecutive check-ins of users to compute their probability of returning to venues. Another study demonstrated how geo-tagged content on Flickr can be used to understand landmarks, topics of interest, and active geographic regions of importance to the user and hence can recommend suitable travel routes (Kurashima et al. 2013). Similarly, the identification and clustering of geographic activities by utilising users' history of geo-tagged photos has been used to deduce areas of interest and to enhance the effectiveness of location recommendation (Balby Marinho et al. 2012).

With regard to understanding users, sensitive personal information can be revealed by tracking user check-in information, including gender, educational background, age, and sexual orientation (Rossi and Musolesi 2014; Errounda and Liu 2019). Liu et al. (2018) summarised different methods of attacks that can be used by adversaries on a mobile application to reveal the user's identity and to determine their position and time information, including machine learning methods (Murakami and Watanabe 2016) and collusion of malicious users (Li et al. 2014). On the other hand, simple statistical and visualisation methods were shown to be useful in deriving useful spatio-temporal patterns of mobility from check-in data on Foursquare (F. radAlrayes and Abdelmoty 2014), further highlighting the privacy risk of disclosing this information. Several user studies have revealed that social network users are not fully aware of the privacy threats present concerning the information published on their accounts (F. S. Alrayes et al. 2020; Coppens et al. 2014; Rader 2014) and that applications do not necessarily have the incentive to make users aware of their data collection and processing practices (Coppens et al. 2014; Lindqvist et al. 2011).

The above sample of studies demonstrates the potential value of location information in deducing personal user information and consequently its possible privacy implications. This highlights the pressing need for improving user awareness of their data and its implications.

2.2. Feedback and control approaches to location privacy

Feedback and notification tools are used to warn users about security and privacy risks. Several works have attempted to assess the impact of such tools on users' awareness of privacy implications (Malandrino, Scarano, and Spinelli 2013). Rader (2014) observed links between limited awareness of possible privacy violations and the usefulness of policy-based privacy solutions. Other studies noted that increased awareness encouraged users to utilise stricter accessibility options (Emanuel, Bevan, and Hodges 2013). Sadeh et al. (2009) found that methods that increase users' awareness about the way their data are used tend to stimulate users to produce more accurate preferences and increase users' trust in the application. (Malandrino, Scarano, and Spinelli 2013) investigated location disclosure on location-enabled mobile applications and revealed that most participants were unaware of how frequently their location was accessed. Patil et al. (2014) observed that immediate feedback about location disclosures without the ability to control the disclosures evoked feelings of oversharing. They recommended the use of proactive techniques for adjusting recommendations to disclosure settings, especially in the case of socially distant users (those with weak relationships between them) and when visiting atypical (infrequently visited) locations. Ataei et al. (2018) proposed the development of user interface controls for fine-grained management of location privacy settings based on privacy by design and user interface design principles. The design addresses three key issues: whom to share location with, when to share it, and where to share it. The results showed that the proposed interface led to a greater sense of control.

The usability of privacy notices and feedback tools are also relevant here. The complexity of privacy policies and settings and the need for more accessible tools have motivated much research in this area (Gage Kelley et al. 2009; N. Wang, Grossklags, and Xu 2013). Of interest are studies on users' perceptions of privacy risk, where visual cues were shown to be useful (Zhang et al. 2014), particularly when shown in context.

Utility versus privacy is a trade-off that applications face when offering control over location privacy settings. Several studies have considered the effects of visualisation on user awareness. Angulo et al. (2015) proposed a tool that visualises online data disclosure for supporting usable data transparency. Scenario-based usability testing revealed improvement to users' awareness of their data disclosure to web services. Visualisation approaches to privacy warnings were found to be effective in increasing users' awareness of privacy implications (Dang, Dang, and Kung 2020). Fernandez, Nurmi, and Hui (2021) studied users' privacy perceptions when using smart devices and how visualisation can promote users' awareness of information leakage. Anwar and Fong (2012) proposed a visualisation tool to enable users to explore how their profiles are viewed from the perspective of social connections. They found that participants were able to perform a more accurate policy assessment whilst using the visualisation tool.

These studies confirm that users are mostly unaware of the implications of data sharing for their privacy and that their attitudes towards sharing and controlling their data and visibility can change based on increased awareness. To date, little emphasis has been placed in previous works on the data that are shared by the user over time. Rather, the focus was primarily on immediate feedback on what the user is sharing at a particular point in time and its visibility to others. Here, we consider the user's data as a whole, taking into account the implications of sharing location data over time.

3. Geoprofile visualisation approach

In this section, we begin by analysing the information content in geoprofiles that need to be exposed to users and identify core elements that need to be modelled to link the spatial, temporal, and social dimensions of the data. We then propose a visual approach to the presentation of this information.

3.1. Geoprofiles

As in conventional social networks, GeoSNs users maintain profiles that may include demographic data, interests, and preferences. GeoSNs further refine these profiles with user location histories to infer user's relationship to other users and user's relationship to locations (Bao et al. 2015). In this work, these enhanced user profiles are referred to as geoprofiles. An overview of the methods used for collecting and processing data to build these geoprofiles, and their use for recommendation on GeoSNs can be found in (Wei et al. 2022).

Here, a review is presented of the dimensions of the data collected by GoeSNs and their capacity for inferring information to build these geoprofiles.

The raw data shared on GeoSNs are time-stamped locations representing users' visits to specific places, for example, check-ins on Foursquare. A basic geoprofile can thus be modelled as a stream of time-stamped locations that can be plotted along the dimensions of both space and time. Through disambiguation of the location information and the relation of this information to geographic places, as well as the straightforward clustering of location points over time, the application is able to specify visits to specific places. Queries over such a profile can allow a user to issue the following set of sample queries against their dataset:

(1) "Which place did a user visit at a specific point in time?", e.g. "Where was I last Wednesday evening?"
(2) "When did the user visit a specific place?", e.g. "When was I last at the Roasted Coffee shop in Cardiff?"
(3) "Which places of a specific type did a user visit?" e.g. "Which coffee shop did I visit last weekend?"

> This is the range of questions that can be answered by providing simple access to user data, as is done in Foursquare and its associated search application, discussed later in this paper. This basic model of a geoprofile and its associated queries provide limited snapshots of views of the user dataset.

In addition to the spatial and temporal dimensions, users' interactions on a GeoSN can also be modelled and analysed along an additional dimension: the social dimension. As before, the spatial and temporal dimensions represent the time-stamped visits to locations and their corresponding positions in the geographic space. The social dimension is representative of the interaction with other users, namely, by co-location in places, as well as through interaction with geographic places – by doing activities in places or using services provided in places. Analysis of users' visits to places over time can be used to determine (a) the degree of relatedness between users (co-location of users over time), (b) the degree of relatedness between users and places (visits to places over time), and (c) the type of services and activities the user may be interested in (through analysis of the types of places that a user visits). These three key types of information summarise the user's interaction on a GeoSN and encapsulate information provided along the spatial, temporal, and social dimensions.

We assume that these three types of information are used to model user information and build geoprofiles on GeoSNs by mapping them into three distinct classes: places, interests (and activities), and friendships. These classes can be further qualified based on the degree of relatedness to the user. Three qualitative levels of relatedness can be used to describe the instances from each class: 1) all instances – that is, all instances of the class are of similar importance to the user–2) favourite instances – describing the subset of instances with a high degree of relatedness to the user – and 3) routine instances – describing the subset of instances with a recognised pattern of relationship to the user. Note that the degree of relatedness is a temporal function that can be implemented in different manners by the underlying recommendation algorithms, e.g. a favourite activity can be identified as the most frequently related activity in a recent period of time, and a regular activity can be described as one that takes place at specific points in time.

A geoprofile with this information can allow users to issue the following example queries to the GeoSN to extend the three queries on the basic geoprofile.

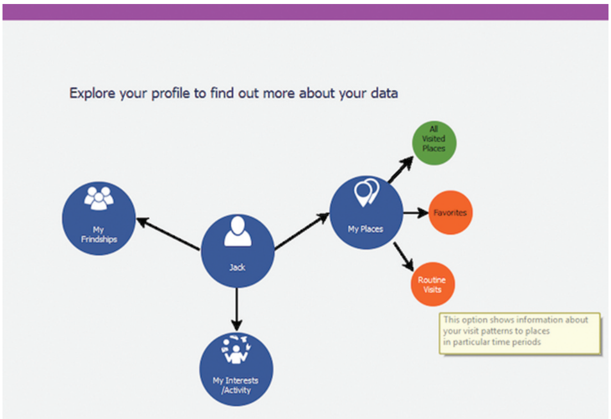
(4) "Which are the user's favourite places?", e.g. "Which are my 5 favourite (most visited) places?"
(5) "Which places does the user regularly visit at specific points in time?, e.g. "Which places do I routinely visit on weekends?"
(6) "Which places of a specific type does a user visit regularly?" e.g. "Which coffee shops do I attend regularly during my lunch hour?"
(7) "Which activities does a user do regularly?" e.g. "What activities am I routinely doing on Saturday mornings?"
(8) "Which other users (friends) visit the same places?", e.g. "Who, on my friend list, visits the Roasted Coffee shop in Cardiff?"
(9) "Which other users (friends) have similar interests or activities in similar time intervals?", e.g. "Who, on my friend list, goes to the park in Cardiff on Saturday mornings?"

The above queries demonstrate the types of analysis used by recommender systems on GeoSNs to build geoprofiles. Answers to these queries should be accessible and comprehensible to users of GeoSNs to enhance their awareness of their data. This work proposes methods for uncovering these fundamental blocks of information to the user. In the rest of this paper, it is assumed that a user geoprofile will contain information on places, interests, and friendships, as described above, to represent the interaction with other users (friendship), interaction with geographic places (interests and activities), and the patterns of these interactions over time.

3.2. Design of the geoprofile visualiser

This work is concerned with evaluating user awareness of the data they share. In this section, an interface is proposed for the visualisation of the geoprofile information described in the previous section. A prototype interface is used in this work to evaluate the value of presenting this information to the user for enhancing their awareness of their data and its privacy implications. Particular design choices and the justification of properties such as the usability of the interface are beyond the scope of the current study and are left for future work.

A *Hub and Spoke* interface design pattern (Tidwell and Brewer 2020) is shown in Figure 1 to visualise the facets of a user's geoprofile. The figure shows how the interface presents to a particular user their geoprofile as three nodes (My Places, My Interests, and My Friendships), which can be individually explored using the degree of relatedness filters described. Table 1 provides details of the node structure and a collection of possible attributes that can be stored at each node. This interface design pattern is chosen because it serves several purposes: 1) it limits the presentation of the user's data through the three clusters of information and thus provides homogeneous filters on the dataset; 2) the graph structure to represent the data and its relationships promotes the visibility of data content and decreases the cognitive workload that would be associated with a search through the geoprofile database; and 3) it allows for



Explore your profile to find out more about your data

This option shows information about your visit patterns to places in particular time periods

Figure 1. The main screen of the geoprofile visualiser when the 'my Places' node is chosen.

Table 1. Hub and spoke architecture of the geoprofile visualiser.

Main node	Subnode	Sample attributes
My Places	All Places	Place Name, Coordinates, Visit Date, Visit Time, Place Type, Address, Visibility
	Favourite	Place Name, Place Type, Address, Number of visits
	Routine	Place Name, Place Type, Time pattern, Day pattern, Number of visits
My Interests/ Activities	All Interests	Interest, Associated Places, Associated Place Types, Timestamp
	Favourite	Interest, Associated Places, Associated Place Types, Number of visits
	Routine	Interest, Associated Place Types, Time pattern, Day pattern, Number of visits
My Friendships	All Friends	Friend, Interests, Co-location Place, Co-location Timestamp
	Favourite	Friend, Place, Place Type, Number of Co-locations
	Routine	Friend, Place, Place Type, Time Pattern, Day Pattern, Number of Co-locations

progressive access to the dataset, both the data shared by the user and the data inferred by the application, focusing the user's attention and thus supporting their awareness of the data elements that are being presented.

4. Geoprofile generation and prototype implementation

Foursquare[1] and its check-in application, Swarm, were chosen as the platform for this study. Foursquare is a popular GeoSN and has been the subject of several previous studies. Foursquare has millions of users, with more than 12 billion check-ins performed by these users globally.[2] On Foursquare, users are able to check-in to places and provide tips and tags. Friends can see each other's locations and check-in history. A data collection system was developed to collect the participants' data (with their permission) using the Foursquare API.[3] Check-in data on users and friends are collected in JSON format and stored in a database. Figure 2 is the UML class diagram for the basic geoprofile structure for the data collected using the Foursquare API.

A desktop application was developed using Python and PyQt4 for the design of the Graphical User Interface (GUI). Figure 1 shows the main screen of the geoprofile visualiser for the 'My Places' node page. The information in each subnode is also presented in tabular format and can be sorted based on different attributes. Search functionalities are provided for each subnode from which the information can be filtered according to relevant attributes, such as place name, type, and check-in date.

4.1. Building the geoprofile database

Simple data mining and heuristics were used to infer relationships and patterns in the data to construct the geoprofile database. The *Pandas* Python library was used for data analysis and mining. The extracted information is only representative of the possible inferences that can be made with the data and is

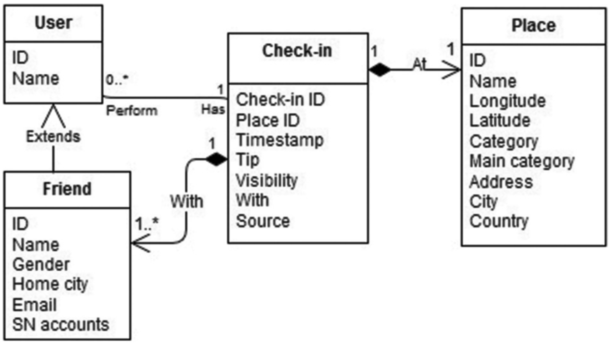


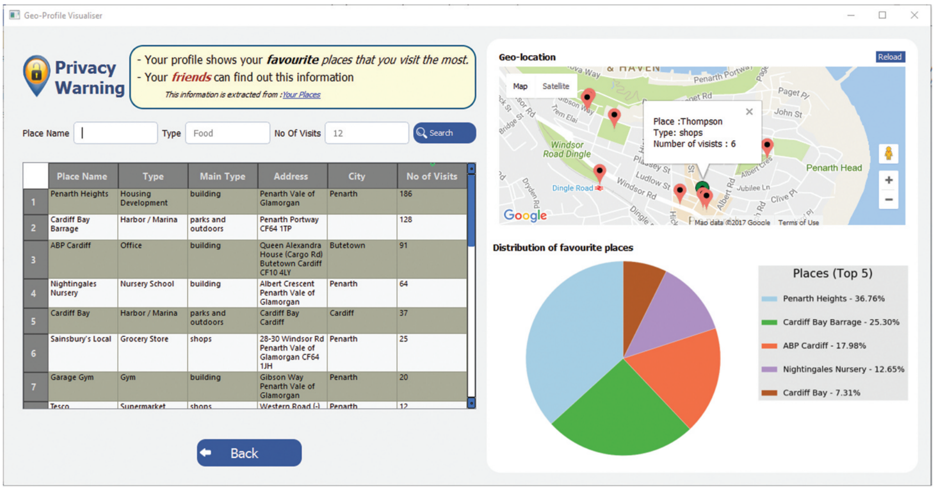
Figure 2. UML class diagram for the basic geoprofile data retrieved from foursquare.

sufficiently representative for the purpose of evaluating the proposal in this paper. The geoprofile database was populated as follows.

(1) Basic information: ‘All Visited Places’, ‘All Interests’, and ‘All Meetings with Friends’ were extracted directly from the collected information.
(2) Place type information is used to represent interests and activities,for example, shopping and travelling.
(3) Simple data mining was used to extract the rest of the information. In particular, ‘Favourite’ instances are defined based on the frequency of occurrence. A threshold of 5 was used. For example, a place is an instance of the favourite set if it is visited by the user a minimum of five times. The top 20 favourite instances are displayed to the user in descending order.

To detect routine instances, regular patterns were detected if 20% of a user’s visits (occurrences) related to places, interests, or co-locations at particular times, days, or both. At least five correlations had to be shown. For instance, if a user visited place X more than 30 times, 6 of them on a Friday morning, then the visit is counted as a routine activity for place X with a time pattern of ‘morning’ and a day pattern of ‘Friday’.

The *Matplotlib* Python library was used for generating the graphs in the ‘Favourites’ and ‘Routines’ subnodes. Figure 3 shows different visualisations of a user’s top five favourite places. More detailed graphs are shown in Figure 4 to represent the `Routine Places’ subnode, where the subject



- Your profile shows your ***favourite*** *places that you visit the most.*
- Your ***friends*** can find out this information

This information is extracted from :Your Places

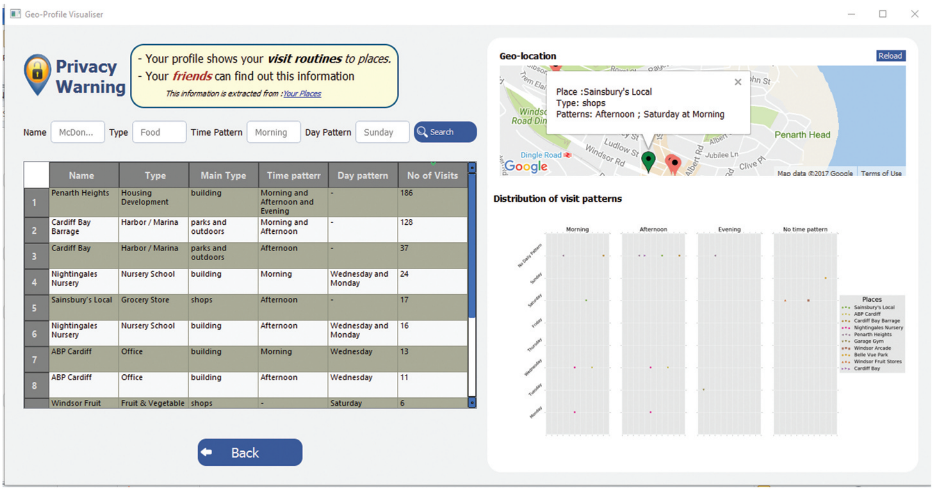
Place Name | Type Food | No Of Visits 12

	Place Name	Type	Main Type	Address	City	No of Visits
1	Penarth Heights	Housing Development	building	Penarth Vale of Glamorgan	Penarth	186
2	Cardiff Bay Barrage	Harbor / Marina	parks and outdoors	Penarth Portway CF64 1TP		128
3	ABP Cardiff	Office	building	Queen Alexandra House (Cargo Rd) Butetown Cardiff CF10 4LY	Butetown	91
4	Nightingales Nursery	Nursery School	building	Albert Crescent Penarth Vale of Glamorgan	Penarth	64
5	Cardiff Bay	Harbor / Marina	parks and outdoors	Cardiff Bay Cardiff	Cardiff	37
6	Sainsbury's Local	Grocery Store	shops	28-30 Windsor Rd Penarth Vale of Glamorgan CF64 1JH	Penarth	25
7	Garage Gym	Gym	building	Gibson Way Penarth Vale of Glamorgan	Penarth	20
	Tesco	Supermarket	shops	Western Road (-)	Penarth	12

Place :Thompson
Type: shops
Number of visists : 6

Distribution of favourite places
Places (Top 5)
Penarth Heights - 36.76%
Cardiff Bay Barrage - 25.30%
ABP Cardiff - 17.98%
Nightingales Nursery - 12.65%
Cardiff Bay - 7.31%

Figure 3. The ‘favourite Places’ subnode of the ‘my Places’ node.



Geo-Profile Visualiser

Privacy Warning

- Your profile shows your ***visit routines*** *to places.*
- Your ***friends*** can find out this information

This information is extracted from :Your Places

Name McDon... Type Food Time Pattern Morning Day Pattern Sunday

	Name	Type	Main Type	Time patterr	Day pattern	No of Visits
1	Penarth Heights	Housing Development	building	Morning and Afternoon and Evening	-	186
2	Cardiff Bay Barrage	Harbor / Marina	parks and outdoors	Morning and Afternoon	-	128
3	Cardiff Bay	Harbor / Marina	parks and outdoors	Afternoon	-	37
4	Nightingales Nursery	Nursery School	building	Morning	Wednesday and Monday	24
5	Sainsbury's Local	Grocery Store	shops	Afternoon	-	17
6	Nightingales Nursery	Nursery School	building	Afternoon	Wednesday and Monday	16
7	ABP Cardiff	Office	building	Morning	Wednesday	13
8	ABP Cardiff	Office	building	Afternoon	Wednesday	11
	Windsor Fruit	Fruit & Vegetable	shops	-	Saturday	6

Geo-location

Place :Sainsbury's Local
Type: shops
Patterns: Afternoon ; Saturday at Morning

Distribution of visit patterns

Figure 4. The 'routine Places' subnode of the 'my Places' node.

of the pattern is displayed as a colour-coded point in relation to a time period on the x-axis and in relation to a day of the week on the y-axis.

The Google Maps API is used to provide a base map view, with basic map zooming and panning functions. The tabular view is linked to the map view, thus allowing the user to move across seamlessly between the two interfaces, providing immediate feedback and visibility of the data. More examples of views for the interests and friendship nodes are shown in the Appendix.

5. Evaluation

Our hypothesis states that better representation of the information contained in geoprofiles and transparency to users will improve user awareness of their shared data and associated privacy implications. To evaluate this hypothesis, we conducted a user study to assess the usefulness of the proposed visualisation of the geoprofiles to support a) users' awareness of the information they share and the possible inferences that may be derived by GeoSNs and b) users' privacy concerns and attitudes towards sharing their data as a result of the proposed approach. The study first starts by alerting participants to the types of information that are represented in the geoprofile before checking their awareness and concerns. To do this, participants are asked to perform a set of predetermined tasks that correspond to a representative set of queries against the geoprofile using the developed prototype and their existing Foursquare app. Foursquare, through its check-in application *Swarm*, offers a basic search facility that allows users to find and retrieve their check-in history using a date filter, as shown in the screenshot in Figure 5. Using existing and proposed approaches for viewing geoprofiles is necessary to allow users to consider the information they currently associate with their geoprofiles against implicit information that can be inferred from their data.

5.1. Method

To eliminate possible bias in the experiment (Krol et al. 2016), a between-subjects design was used. Participants were randomly assigned to one of the two study groups: a *non-awareness* (baseline) group and an *awareness* group. In the *non-awareness* group, participants were asked to perform tasks related to finding information about their mobility using the Swarm history search facility. In the *awareness* group, participants performed the same tasks using the geoprofile visualiser. In both groups, participants were regular users of the Foursquare application and used their own accounts in the experiment.

Following the experiment, semi-structured interviews were used to further explore participants' reactions and responses. The experiment is performed in three steps.

(1) Selection of representative participants: users of the Foursquare/Swarm application with a reasonable amount of check-in data (recorded visits to places on the application).

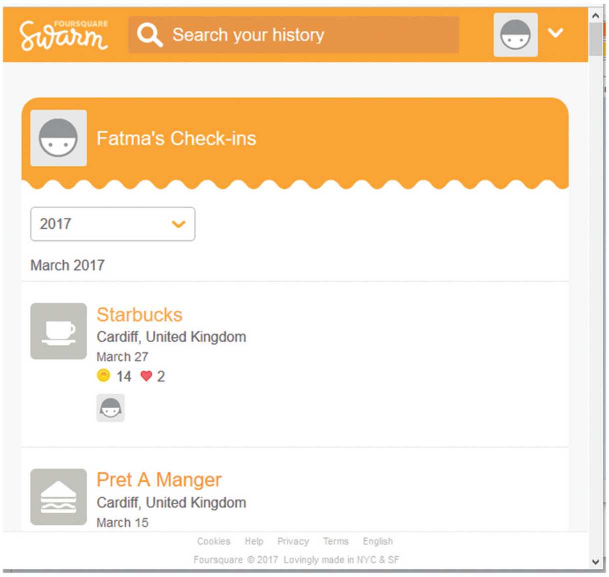


Figure 5. A screenshot of the swarm history search facility.

(2) The check-in data are collected from the user accounts, and the individual geoprofiles are built and stored.
(3) The user study and post-study interviews were conducted for both user groups.

5.2. Recruitment and participants

Participants were solicited by email, on social media (including messages on the Swarm application), and through face-to-face invitations. Participants had to be users of the Swarm application at the time and to regularly check into places to qualify for the study. They had to be willing to provide personal check-in data for use in this study and to be able to attend an interview in person. Enough time was spent recruiting representative participants (28 participants). For practical reasons, we recruited mainly university students and staff. Before carrying out the experiment, pilot tests were first conducted on four volunteers to ensure that participants could easily follow all the study stages and understand the requirements. Two participants did not attend the interview and were assumed to have withdrawn. Twenty-six participants completed the experiment: 13 in the non-awareness group and 13 in the awareness group. Participants checked in, with a range of check-in counts between 36 and 14,911. Seven participants had fewer than 100 check-ins, 11 had between 100 and 1000, and the remaining 8 had more than 1000. Their ages ranged between 19 and 41 ($\mu = 27$, $\sigma = 5.311$). Eleven of them were students in computer science (nine postgraduates and two undergraduates), three were students in the social sciences (two postgraduates and one undergraduate), three were undergraduates in engineering, two were in Business, one was a postgraduate in medicine, one was in Art, one was in mathematics, one was in software engineering, one was a mechanical engineer, one was a commercial manager, and one was a lecturer in computer science. The female participants (Gu et al. 2016) outnumbered the males (Liu et al. 2018). Half of the participants were from Asia, nine from Europe, three from Africa, and one from the Caribbean. It is important to note that the diversity of the recruited group was limited by the availability of participants who met the set criteria.

5.3. Study procedure

The study was carried out as a series of semi-structured interviews that started by ascertaining a brief demographic background. Two Likert-scale questions were then administered to all members to capture the participants' initial sense of safety while using the GeoSN and their level of concern over online privacy. These questions were followed by closed- and open-ended questions to gauge the participants' perception of social networking experience and online privacy in terms of their knowledge of and attitude towards data collection, use, and control.

The second stage involved using the Swarm History app for the non-awareness group and geoprofile visualiser for the awareness group. For each group, the interviewer started with a brief description of the tool's purpose through a demonstration of a think-aloud task that was not related to the subject of the study and gave the participants a few minutes to explore their profiles using the tool. Note that at this point, a personalised geoprofile had already been generated and populated in the prototype using the data for individual participants. Thus, we were able to compute favourite places, routine interests, regular encounters with friends, and other such attributes. The participants were then asked to carry out a set of predefined tasks on the basis of their generated geoprofile (seven tasks on average). The tasks, shown in Table 2, were designed to represent possible information that can be extracted from their geoprofiles. These tasks were individually customised based on the data in the users' profiles. For example, some profiles did not generate routine interests or regular encounters with friends. The think-aloud protocol was useful in providing insight into the user interaction and decision-making process at the interface and the rationale for success or failure of the tasks. It provided useful information that complemented the post-study interviews.

Participants were encouraged to think-aloud and were observed as they used the tool to perform the tasks. The aim of the tasks was to determine whether it was possible to find all of the data that users themselves had shared and whether they were also able to identify the possible inferences that were made by the geoprofile visualiser based on their shared data. During the performance of the tasks, the computer screen was video-recorded. The performance of the tasks was scored as total success, partial success, or failure. The degree of success was measured by the amount of information the user was able to extract in the task. Partial success was considered when they were able to find some of the information that was needed. This could, for example, involve finding a favourite friend but being unable to count the number of times they were co-located in a place, as in question 8 of Table 5.3. The participants were then asked to respond to a series of statements to evaluate their awareness of data sharing, visibility of their data and possible privacy implications.

Following the experiment, members of both groups were asked to answer the same two questions they were given at the start of the experiment to gauge their sense of safety and level of concern over online privacy. This was done to assess the potential impact of the study and to determine whether the results varied. The purpose was to examine any changes that can be attributed to the experiment. They were also asked to rate their willingness to share their location profile. Members of the awareness group were asked three further questions to assess the geoprofile visualiser and its potential utility.

Table 2. Tasks assigned to the participants in the experiment.

No.	Task	Type
1	Find the place, place type and time of any recent check-in of yours.	Place/Basic
2	Find one of your favourite places (most visited), and how many times you visited it.	Place/Favourite
3	Find a place that you visit at a regular time or on a regular day.	Place/Pattern
4	Find an interest/activity that you were involved in and find the place that it occurred.	Interest/Basic
5	Find one favourite interest or activity (that you are often involved in), and determine the number of times it is recorded.	Interest/Favourite
6	Find an interest or an activity that you do at regular times, on regular days or in particular places.	Interest/Pattern
7	Find a place (and address) of a check-in/meeting with one of your friend.	Friend/Basic
8	Find one of your friends that you check-in with a lot and count how many times you checked-in together.	Friend/Favourite
9	Find a friend that you check-in regularly with in particular places or at particular times or days.	Friend/Pattern
10	Find who can see your information, such as places you visited and your favourite places.	Data Visibility
11	Find information you shared directly or indirectly (place of study/work/home).	Data Extraction

6. Results

In this section, the study results are analysed and presented. The Fisher's Exact test (Heumann and Shalabh 2022) is used to determine whether the impact of the type of tool used (e.g. the Swarm history or Geoprofile visualiser) on users' privacy awareness and attitude was statistically significant. The related Cramer's V test was used to examine the strength of this association regardless of sample size by comparing the difference between the means of two samples where 0.1 is considered small, 0.3 is considered medium, and 0.5 is considered large (Cohen 1988). Moreover, the Friedman Chi-Square test and ordinal logistic regression (Heumann and Shalabh 2022) were used to determine the significance and direction of influence of the tool used provided to perform their tasks, on several aspects measured before and after the actual study.

6.1. Pre-study: technological experience, privacy awareness, and attitude

More than half of the participants were experienced users of web applications and technologies (54%). Most felt safe using the Swarm application (3.8 on average out of 5), yet they were generally concerned about their online privacy (4 on average).

Two-thirds of them had deleted a post shared on their social network accounts due to privacy concerns. In total, 65% of the participants had regretted sharing certain information on a social network account, which indicates that users can share information without being conscious at the time of its potential consequences. Moreover, 35% of the participants had made requests to delete their data, for instance, their profiles, from an online service.

Approximately 60% of the participants were not aware of some of the Foursquare/Swarm practices of collecting and sharing users' data. In particular, 85% of them were not aware that the application could share their data with third parties for purposes other than marketing, while 58% of the participants were not aware of their data being used for marketing and advertisement purposes. Approximately 40% of the participants were not aware that the application could obtain and record their location, even if they were not interacting with it.

At this stage, the participants' perceptions about their location, data accessibility, potential use, and control were also queried. When generally asked who they thought could access their data, almost all the participants said 'The application and their friends on it'. Only seven participants were aware that third parties could access their data. This suggests that the participants are not fully aware of all the parties that can view their data. Regarding data control, 65% of the participants (17 of them) said that they had never checked the privacy setting provided by the application. Sixteen of the participants said that they could control who could see their check-ins (private, friends, or public). Some participants believed that the application offered more aspects of control over their data and profiles than actually existed, for example, by controlling the locations they shared (Rossi and Musolesi 2014), controlling the targeted ads in the application (Patil et al. 2014), controlling the collection of location information while using the application (EUR-Lex 2018) and choosing specific people to share information with (Balby Marinho et al. 2012).

6.2. Task outcomes

The results showed that the choice of tool had a significant impact on the success of the participants in performing the task of finding privacy-related information about their shared locations (Pearson Chi-Square = 36.737096, $p < .0001$). In addition, this association has a large effect (Cramer's V = 0.742620, $p < .0001$). Each participant's performance in the tasks is presented in Figure 6 for the Swarm History (non-awareness) group and in Figure 7 for the geoprofile visualiser

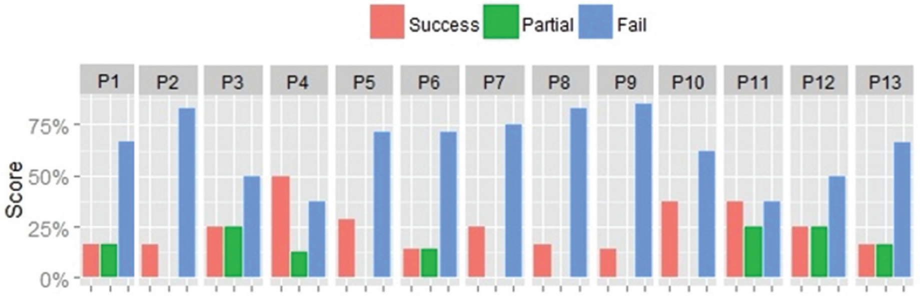


Figure 6. Task results for each participant (P1-P13) in the Swarm History (non-awareness) group.

prototype (awareness) group. On average, the participants in the former group were able to successfully retrieve only 25% of the information they were asked to find, 10% found some of it (partial) and the remaining 65% of the tasks failed to find any. However, the average success score showed a considerable increase to 97% of the latter group; 11 out of these 13 participants were able to perform 100% of their assigned tasks successfully. Not a single failed task was recorded for this group.

In a comparison of the two groups' capacities to find relevant information, the participants of the Swarm History group were able to successfully find 78% of the basic information they shared, including their visited places, interests, and co-locations with friends, as well as 47% of their favourite places, interests, and co-locations with friends as demonstrated in Figure 8. However, 81% of the tasks to

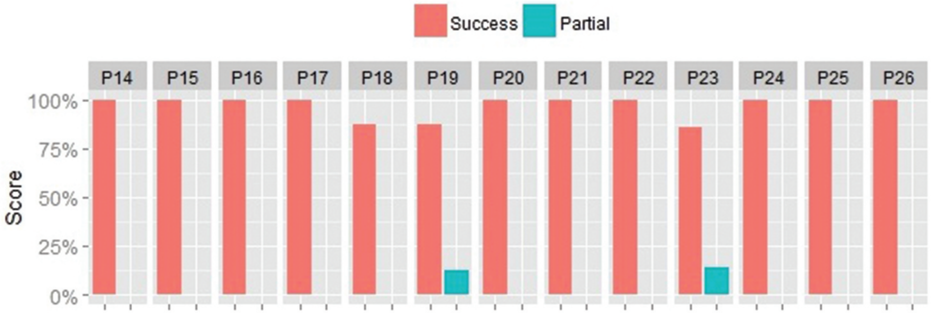


Figure 7. Task results for each participant in the geoprofile visualiser prototype (awareness) group.

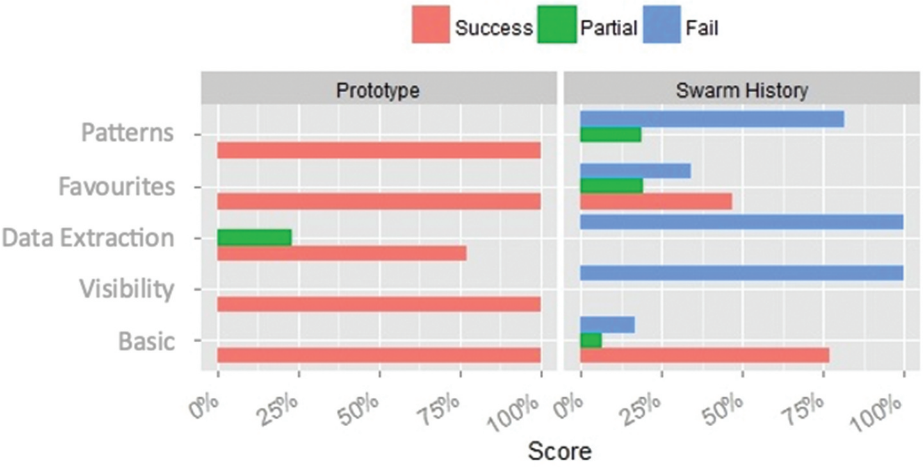


Figure 8. The capacity to find relevant information in both groups based on the type of task.

find patterns ended in failure when no hints were provided. In the group using the geoprofile visualiser, almost all tasks were successful (basic, favourite, pattern, and visibility). Three participants struggled with the data extraction tasks (to find information they shared on their home/work/study place).

The results above demonstrate the difficulty experienced by the Swarm users in carrying out the tasks. Problems were associated with difficulty navigating the interface as well as the large number of error-prone steps involved in accomplishing the tasks. For example, considering the simple task of finding the type of visited place requires guessing and interpretation of icons of place types. This process failed in some cases, with participants resorting to relying on personal memory of the visit to the place to identify its type. In contrast, performing the tasks using the geoprofile visualiser proved to be much simpler and more efficient, involving a small number of predictable steps and thus leading to more successful navigation and completion rates.

6.3. Information awareness and privacy attitudes

This section provides a quantitative evaluation of the participants' views of and attitudes towards their ability to find information related to their privacy. These data are supported qualitatively, by data derived from the open-ended questions, which are discussed in the following subsections. The results of the statistical tests of the quantitative data, including the participants' ratings on a 5-point Likert scale (5 strongly agree to 1 strongly disagree), are shown in Table 3.

6.3.1. Impact on privacy awareness

The type of tool used significantly affects both the participants' ability to view the information they share when they check in and their understanding of it. These two associations also have a large effect, whereby the participants in the prototype group tended to agree more strongly (with an average rating of 4.8), whereas those in the Swarm History group tended to be more neutral about these statements (F. Alrayes and Abdelmoty 2017). When asked whether they found the tool helpful for accessing their collected data, those in the non-awareness group generally reported that Swarm History showed them a basic and general view of their shared check-ins, which some found vague. Almost all of them (with one exception) wanted more explicit details about their check-in data. Some of them said that they had to perform more searches to reach the desired information and that the presentation needed to be improved, including more filtering and ranking features. For example, P8 said 'I need to do a lot of work to find out information'. Others felt that the application had collected more information about them than it had. For instance, P12 mentioned that 'Very limited, the app collects more things'. The awareness-group participants all found that the geoprofile visualiser provided a more detailed and direct view

Table 3. Results of participants' opinions on their ability to find information using the systems in both groups (mean (μ) and standard deviation (σ)).

	Swarm History		Geoprofile		Fisher exact	
Statement	μ	σ	μ	σ	test	Cramer's V
This tool helps me understand the information I share when I check-in	2.8	1.17	4.7	0.46	<.0001	0.858
This tool allows me to view the information I share when I check-in	3.3	1.14	4.9	0.27	0.0001	0.788
This tool helps me understand the possible information that can be extracted about me when I check-in (e.g. patterns of visits, top interests)	1.8	1.03	4.8	0.36	<.000001	1
This tool allows me to view the possible information that can be extracted about me when I check-in (e.g. patterns of visits, top interests)	1.8	0.86	4.8	0.42	<.000001	1
This tool allows me to know who can access my data	1.3	0.61	4.1	0.83	<.00001	0.936
This tool motivates me to be more in control of my online data (e.g. deleting posts, updating privacy settings)	2.5	1.15	4.2	0.86	0.0034	0.704
I am satisfied with the way Foursquare/ Swarm collects and stores my data	3.2	1.12	2.9	1.14	0.058	0.575
This tool makes me more concerned about my privacy	2.6	1.08	4.4	0.74	0.0011	0.741
This tool encourages me to alter the way I share my location information to protect my privacy	2.4	1.08	4.2	0.95	0.0067	0.693

of their location data, which they could easily find. P17 said that it is 'More detailed than swarm app, my profile is grouped in a nice way which makes it easier for me to find information'.

In terms of data inference, the type of tool used also had a significant impact on helping participants to see and understand the information that could be extracted about them when they were checked in (e.g. patterns of visits and top interests). This association has a large effect – the participants in the geoprofile group tended to more strongly agree (4.8), while those in the Swarm History group tended to disagree more strongly with these statements (1.8). When asked about how helpful the tool was in finding the kind of information that could be extracted from their data, all the participants in the non-awareness group reported that Swarm History did not show them this kind of information. They said that 'It gives the bare minimum', and they had to work it out in writing 'Using pen and paper'. All participants showed an interest in finding out about and viewing their information; some were driven by concern for privacy. For example, P2 stated that 'It would be good to have, due to privacy concerns, so I know what I am sharing and what others can know about me'. A few others referred to using the extracted information for other life-management purposes. P5, for example, said 'Patterns will help in day planning and time management'. Examining the feedback on this question from the awareness group participants indicated that all of them found the geoprofile visualiser helpful and thought it interesting to view what could be inferred about them.

They expressed their need to be aware of such information; they said that there was ‘too much’ extracted information and ‘had no idea that such information can be inferred’. They pointed out how the Swarm application encourages users to check into a place by focusing on the game-playing aspect ‘by providing stickers and being a mayor of places’ or hides possibly inferred information ‘so users won’t get freaked out about it’.

The type of tool used has a significant impact on informing the participants about who can access their data. This association also demonstrated a large effect. The participants in the prototype group tended to agree more with the statement (4.1), whereas those in the Swarm History group seemed to disagree more (1.3). Further elaboration on inputs from the non-awareness group revealed that all of them would in fact like to know who can view or has viewed their information. Two of them were particularly interested in knowing how accessible their data were to third parties. P8 pointed out, ‘The app is deliberately not showing who can access my data so I don’t get scared and stop using it’. All the awareness groups found the geoprofile visualiser helpful in revealing the accessibility of their data to protect their privacy. Four of them also wanted to know who, apart from their friends, could see their data.

6.3.2. Impact on privacy attitude

The tool used in each group was shown to have a significant impact on motivating participants to be more in control of their online data, by deleting posts or updating privacy settings, for example. This association also has a large effect. Members of the Swarm History group were, on average, not motivated to perform this action (2.5) compared to the geoprofile visualiser group (4.2). In addition, participants’ satisfaction with how Foursquare/Swarm collected and stored their data significantly differed between the two groups. Approximately half of the participants in the awareness group were not satisfied, compared to only one participant in the non-awareness group, demonstrating the impact of the viewpoint from which the same information was presented on user privacy preferences.

Participants’ privacy concerns were significantly impacted by the tool used, where members of the awareness group were generally concerned (4.4), while those in the non-awareness group were not. The members of the Swarm History group explained why they were not concerned; it showed only the basic check-in information that they had chosen to share. Interestingly, four participants were concerned about who could access their data. For instance, P12 said ‘I am more concerned since I don’t know who can see what of my data because the app collects more but it is not showing it and I should know’. Participants from the geoprofile visualiser group were generally concerned as they realised the possibility of revealing personal information beyond what they had shared. For example, P17 stated, ‘It shows who can see my data. My information can be inferred and I am not aware. Extracted interests and patterns can spoil my privacy'. Moreover, the type of tool used had a significant impact on encouraging participants to protect their privacy by altering the way they shared their location information. This association has a large effect. On average, the geoprofile visualiser group was more willing to change how they shared their location information (4.2) than was the Swarm History group (2.4).

6.4. *Post-study: impact on the sense of privacy and safety*

After finishing the tasks in the experiment to find information from their geoprofiles, the participants were asked to again rate how safe they felt in using Foursquare Swarm and how concerned they were about their online privacy. In the awareness group, using the geoprofile visualiser had a very significant impact on the participants' sense of safety before and after using it (Friedman Chi-Square = 12.000, *p* = .001). Participants felt significantly less safe after using the prototype than before using it (ordinal regression coefficient = −3.975, *p* = 0.000309). Initially, they had felt safe using Foursquare Swarm (3.8 on average), yet their rating dropped to 1.8 on average showing that they no longer felt safe after using the prototype. In the non-awareness group, the Swarm history had a less significant impact on their sense of safety before and after using it (Friedman Chi-Square = 5.000, *p* = .025). In particular, the change in the participants' attitude towards feeling safe in using Foursquare Swarm was not significant either before or after using Swarm History (ordinal regression coefficient = −1.102, *p* = 0.146). They still generally felt safe in using the application (before: 3.8, after: 3.4). Figure 9 presents the results of the participants' ratings of the sense of safety that they felt using Foursquare Swarm before and after the actual experience of using the tool provided for the tasks.

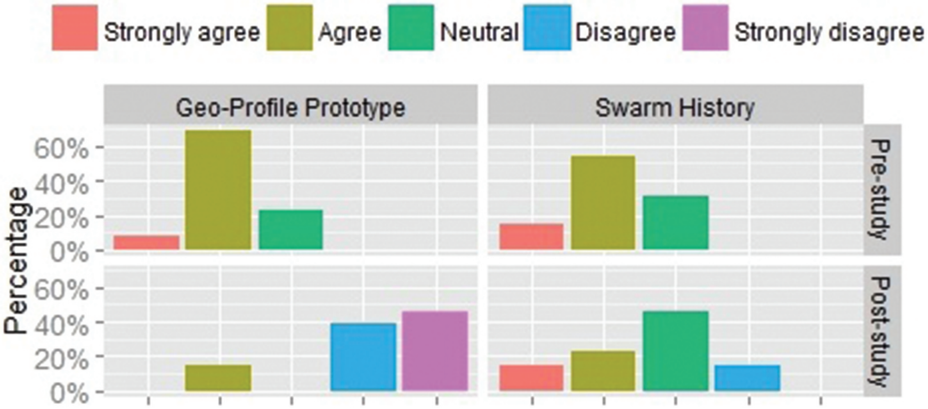


Figure 9. Participants' response to the question "I feel safe using Foursquare/Swarm", before and after the actual experience of using the tool in both groups.

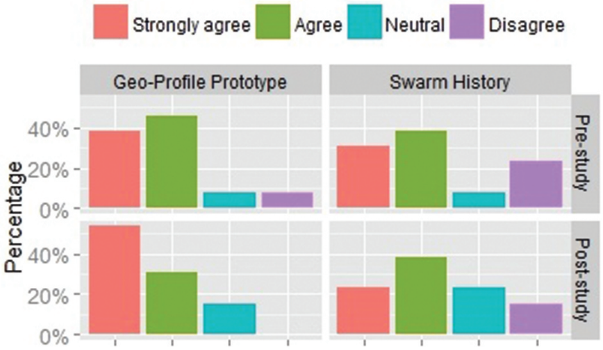


Figure 10. Participants' responses to the question "I am concerned about my online privacy", before and after the actual experience of using the tool in both groups.

There was no significant impact in either group on the level of concern over privacy among the participants either before or after using the tool provided. In other words, the participants were generally still concerned about online privacy. However, those in the awareness group showed a slightly greater increase in their level of concern (before: 4.2 on average, after: 4.4 on average) compared with the members of the non-awareness group (before: 3.8 on average, after: 3.7 on average). Figure 10 shows the participants' ratings regarding their concern over online privacy before and after the actual experience of using the tool for the tasks. Both tools had a significant impact on people's sharing decisions with people other than their friends (Pearson Chi-Square = 13.516, $p = .004$, Cramer's V = .721). The participants who accessed their location profile using the geo-profile visualiser strongly minded sharing it with others (1.1 on average), while those who accessed their profile using Swarm History tended to be generally neutral about sharing it (2.7 on average).

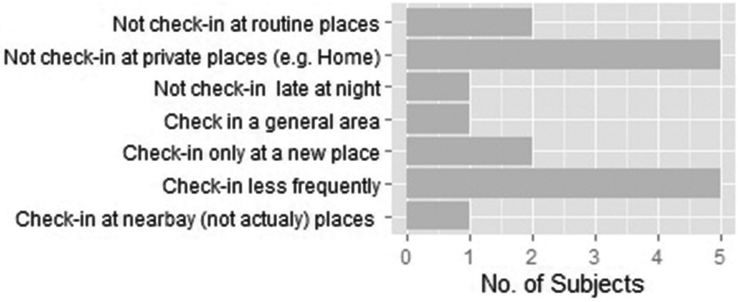


Figure 11. Participants' responses towards approaches to change their sharing behaviour for the awareness group (clustered).

Members of the awareness group were asked specific open-ended questions to discover the impact of the geoprofile visualiser on privacy awareness and location sharing behaviour. All participants agreed that before using the prototype, they had had a limited understanding and awareness of the detailed collection of their data, possible inference power of the information and privacy implications. For example, P14 said 'I was not aware of such detailed data collection and extraction. I thought it was just sharing place and time. It is scary that other people can know the visit pattern at my house'. Some participants criticised the application for not supporting privacy awareness. For instance, P22 said 'It is helpful to improve my awareness and to know better. I am more concerned because people can get my data and know my patterns such as going shopping Sunday morning and here I am. Swarm History is limited and not detailed enough. It is not helping me to be aware'. In addition, almost all the members of the awareness group stated that they would change the way they shared location data. The most often chosen strategy for protecting their privacy was to check-in less frequently and not check into private places such as their homes, as presented in Figure 11. Other strategies mentioned were, not to check into routine places, to check-in only in a general area or if it was their first visit to a place, and to check into a fake location.

Furthermore, all participants showed an interest in using the geoprofile visualiser, mainly to explore their profile by viewing what was collected or inferred about them; they wanted to use this information to manage their privacy by learning what and what not to share. For example, P19 said 'I would use it right at the start of using Swarm to see what they collected and what could be extracted, and to learn about and modify the way I check-in. Then, once in a while after that, to see how my mobility changes and see if I need to change anything'.

7. Discussion

In this section, the implications of the study are discussed in terms of their validity and quality, together with the impact of the information content and presentation on users' privacy awareness and attitudes.

7.1. Validity and limitations

Improving user awareness of their data and the privacy implications of sharing location information is not, so far, a primary use case for the current generation of GeoSNs, as demonstrated with Foursquare in the previous sections. Hence, this work is not a comparison of the Foursquare interface with the proposed approach. Instead, the work evaluated user awareness given direct access to information, as offered by the proposed design, against indirect access to information offered by basic access to raw data in Foursquare.

Foursquare is a widely used GeoSN and is representative of the tasks and experiences of users. Bias was limited by using a between-subjects study design. Tests were performed against the users' own datasets on the GeoSN to ensure full comprehensibility of and familiarity with the data and platform, thus focusing the users' attention on the tasks. The control user group worked with their own data on the GeoSN and thus formed a consistent baseline for the experiment. Participants were carefully selected. Although the number of participants was limited to 26, the think-aloud approach and the qualitative responses provided by the participants in the post-study interviews provided a rich picture of individual attitudes and concerns. For practical reasons of ease of recruitment, the study cohort mainly consisted of university students, mostly from science and engineering backgrounds. This is a limiting factor which could have had an influence on the results.

Further work is needed to study and evaluate the proposed interface design and to assess both the utility and usability of the proposal and how it can be effectively integrated in GeoSNs. This will necessarily require a larger longitudinal study with users from diverse backgrounds to ensure sufficient capture of the user population.

7.2. Impact of geoprofile visualizsation on privacy awareness and attitude

The results show the significant potential of the proposed geoprofile visualiser for enhancing users' awareness of their location data exposure and privacy implications and the users' need for control over their data on GeoSNs.

In particular, recognising the visibility of their data prompted users to choose not to disclose their location information. Users expressed a wish to exercise control over their data accessibility by managing visibility settings, deleting data, and altering the way they used the application altogether.

Similar observations were made in previous studies on data shared directly by users (Angulo et al. 2015; Anwar and Fong 2012; Kani-Zabihi and Helmhout 2011; Tang, Hong, and Siewiorek 2011; Y. Wang et al. 2015). A major challenge in this respect is maintaining a user's awareness over time. Our work addresses this challenge by extracting and presenting both the explicit elements of the data and its implicit content to the user in a manner that promotes awareness and understanding. We note that measuring a full understanding of the proposed visualisation and its impact on user attitudes towards using an application requires a long-term study. This work takes some initial steps into this area and highlights the need for more studies.

Based on our findings, we propose the following set of recommendations to support and improve the privacy awareness of GeoSNs.

- The design of the search and browsing facilities provided for users to explore their data requires the recognition of the spatial, temporal, and social dimensions of the data and their intersections. As with our proposal, raw data as well as implicit data content – as a product of relationships between the different data dimensions – must be made accessible to the user.

- Users need to have fine-grained control over the visibility of their location data; Visibility control on an individual user basis, in addition to control of visibility for groups of users.
- Users need to provide consent when data, or products of their data, are to be used by the application or to be shared with third parties.

Asynchronous utilities for exploring the data, such as the SWARM application for Foursquare and the interface proposed here can be considered 'on-demand' user awareness. That is, users actively seek to find the information when needed. The amount of information required by service providers to run a sustainable service needs to be made transparent to users. It is proposed that the design of the GeoSN should integrate awareness-enhancing utilities such that information and feedback are continuously provided to the user while using the network directly, i.e. switching to a 'live' user awareness mode, where the information is pushed continuously to the user. Further work needs to be carried out to understand the users' needs for information, their level of awareness when performing different tasks, and the impact of the information projected to them on their awareness and attitude towards using the networks. Additionally, the interests of the service providers, their business models, and their needs to utilise and share the users' data need to be studied, while considering the trade-off between privacy and utility (Errounda 2019).

8. Conclusions

This paper addresses the issue of user awareness of the shared and collected data on GeoSNs. Giving the users access to their raw data, or the presentation of the user's history on a timeline, as common in typical GeoSNs, offers a limited view of the user's profile and thus also limits users' awareness of privacy implications. We explore the question of what aspects of the geoprofiles need to be exposed to users. We propose that in addition to basic spatiotemporal information about visits to places, geoprofiles should also reveal the social elements of user interaction by representing the interests of users in places and the connections between users as a result of their visits to places. We then address the question of how to present this information to users and propose a visualisation approach to represent the spatio-social dimensions of the data and use the temporal dimension to express the degree of relatedness between the data elements. To evaluate the effectiveness of the proposal, we conducted an experiment to test user's perception of their geoprofiles on a popular GeoSN application and their perception of their profiles when using the proposed geoprofile visualisation approach. The focus of the experiment was to assess user awareness of their data and privacy implications when given the opportunity to perceive the rich information content of their data, and how this impacts their attitude towards data sharing on GeoSNs. The results show a significant impact from the increased awareness of the information content in the geoprofiles that may indeed lead to change in users' attitudes towards sharing and use of these applications. The question still remains on how to integrate this approach to increased awareness with GeoSN applications while also maintaining effective utility. More work is needed to consider and evaluate alternative approaches to address these concerns.

This research was funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R319), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Notes

1. https://foursquare.com/city-guide.
2. https://location.foursquare.com/company/who-we-are/ [Accessed. on 05-09-2024].
3. Ethical approval for data collection was formally obtained for the experiment.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the Princess Nourah Bint Abdulrahman University [PNURSP2024R319].

Ethical approval

Ethical approval was obtained from Cardiff University, and appropriate informed consent was obtained from all subjects involved in the study.

References

Alrayes, F., and A. I. Abdelmoty. 2017. “Towards Understanding Location Privacy Awareness on Geo-Social Networks.” *ISPRS International Journal of Geo-Information* 6 (4): 109. 04. https://doi.org/10.3390/ijgi6040109.

Alrayes, Fatma, and Alia Abdelmoty. 2014. “Privacy Concerns Due to Location Sharing on Geo-Social Networks.” *International Journal on Advances in Securityl* 7 (3 and 4): 62–75.

Alrayes, Fatma S., A. I. Abdelmoty, W. B. El-Geresy, and G. Theodorakopoulos. 2020. “Modelling Perceived Risks to Personal Privacy from Location Disclosure on Online Social Networks.”

International Journal of Geographical Information Science 34 (1): 150–176. https://doi.org/10.1080/13658816.2019.1654109.

Angulo, Julio, Simone Fischer-Hübner, Tobias Pulls, and Erik Wästlund. 2015. "Usable Transparency with the Data Track: A Tool for Visualizing Data Disclosures." *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, Seoul, Republic of Korea, 1803–1808. ACM.

Anwar, Mohd, and Philip W. L. Fong. 2012. "A Visualization Tool for Evaluating Access Control Policies in Facebook-Style Social Network Systems." *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, Trento, Italy, 1443–1450. ACM.

Ataei, M. A., A. Debgelo, and C. Kray. 2018. "Privacy Theory in Practice: Designing a User Interface for Managing Location Privacy on Mobile Devices." *Journal of Location Based Services* 12 (3–4): 141–178. https://doi.org/10.1080/17489725.2018.1511839.

Balby Marinho, Leandro, Iury Nunes, Thomas Sandholm, Caio Nóbrega, Jordão Araújo, Carlos Eduardo Santos Pires, and ACM. 2012. "Improving Location Recommendations with Temporal Pattern Extraction." *Proceedings of the 18th Brazilian symposium on Multimedia and the web*, São Paulo/SP Brazil, 293–296.

Bao, Jie, Yu Zheng, David Wilkie, and Mohamed Mokbel. 2015. "Recommendations in Location-Based Social Networks: A Survey." *GeoInformatica* 19 (3): 525–565. https://doi.org/10.1007/s10707-014-0220-8.

Bellatti, Jacob, Andrew Brunner, Joseph Lewis, Prasad Annadata, Wisam Eltarjaman, Rinku Dewri, and Ramakrishna Thurimella. 2017. "Driving Habits Data: Location Privacy Implications and Solutions." *IEEE Security & Privacy* 15 (1): 12–20. https://doi.org/10.1109/MSP.2017.6.

Blumbery, A. J., and P. Eckersley. 2009. "On Location Privacy and How to Avoid Losing it Forever." *Electronic Frontier Foundation* 10 (11): 1–7.

Cheng, Z., J. Caverlee, and K. Lee. 2010. "You are Where You Tweet: A Content-Based Approach to Geo-Locating Twitter Users." *Proceedings of the 19th ACM international conference on Information and Knowledge Management CIKM '10*, Toronto, ON, Canada, 759–768.

Christin, Delphine, Martin Michalak, and Matthias Hollick. 2013. "Raising User Awareness About Privacy Threats in Participatory Sensing Applications Through Graphical Warnings." *Proceedings of International Conference on Advances in Mobile Computing & Multimedia*, Vienna Austria, 445. ACM.

Cohen, Jacob. 1988. "Statistical Power Analysis for the Behavioural Sciences. Hillside." *NJ: Lawrence Earlbaum Associates*.

Coppens, Paulien, Laurence Claeys, Carina Veeckman, and Jo Pierson. 2014. "Privacy in Location-Based Social Networks: Researching the Interrelatedness of Scripts and Usage." *Proceedings of the Symposium on Usable Privacy and Security*, Menlo Park, CA.

Crandall, David J, Lars Backstrom, Dan Cosley, Siddharth Suri, Daniel Huttenlocher, and Jon Kleinberg. 2010. "Inferring Social Ties from Geographic Coincidences." *Proceedings of the National Academy of Sciences* 107 (52): 22436–22441. https://doi.org/10.1073/pnas.1006155107.

Dang, T. T., T. K. Dang, and J. Kung. 2020. "Interaction and Visualization Design for User Privacy Interface on Online Social Networks." *SN Computer Science* 1 (5): 297. https://doi.org/10.1007/s42979-020-00314-9.

Emanuel, Lia, Chris Bevan, and Duncan Hodges. 2013. "What Does Your Profile Really Say About You?: Privacy Warning Systems and Self-Disclosure in Online Social Network Spaces." In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, edited by Patrick Baudisch, Michel Beaudouin-Lafon, and Wendy E. Mackay, 799–804. New York, NY, USA: Association for Computing Machinery.

Errounda, F. Z., and Y. Liu. 2019. "An Analysis of Differential Privacy Research in Location Data." *Proceedings of 2019 IEEE 5th International Conference on Big Data Security on Cloud (BigDataSecurity)*, Washington, DC, USA, 53–60. IEEE.

EUR-Lex. 2018. "Regulation (Eu) 2016/679 of the European Parliament." June. https://eur-lex.europa.eu/eli/reg/2016/679/oj.

Fernandez, C. B., P. Nurmi, and P. Hui. 2021. "Seeing is Believing?: Effects of Visualization on Smart Device Privacy Perceptions." *Proceedings of ACM multimedia conference*, Virtual Event, China, 4183–4192. ACM.

Gage Kelley, Patrick, Joanna Bresee, Lorrie Faith Cranor, and Robert W Reeder. 2009. "A Nutrition Label for Privacy." *Proceedings of the 5th Symposium on Usable Privacy and Security*, Mountain View, California, USA, 4. ACM.

Gu, Y., Y. Yao, W. Liu, and J. Song. 2016. "We Know Where You Are: Home Location Identification in Location-Based Social Networks." *Computer Communication and Networks (ICCCN), 25th International Conference*, Waikoloa, HI, USA, 1–9. IEEE.

Heumann, C., and M. S. Shalabh. 2022. *Introduction to Statistics and Data Analysis with Exercises, Solutions and Applications in R*. Switzerland: Springer Cham.

Huiji Gao, J. Tang, and H. Liu. 2012. "gScorr: Modeling Geo-Social Correlations for New Check-Ins on Location-Based Social Networks." *Proceedings of the 21st ACM international Conference on Information and Knowledge Management, CIKM '12*, Maui Hawaii, USA, 1582–1586.

Kani-Zabihi, Elahe, and Martin Helmhout. 2011. "Increasing Service Users Privacy Awareness by Introducing On-Line Interactive Privacy Features." *Nordic Conference on Secure IT Systems*, Karlskrona, Sweden, 131–148. Springer.

Krol, Kat, Jonathan M. Spring, Simon Parkin, and M. Angela Sasse. 2016. "Towards Robust Experimental Design for User Studies in Security and Privacy." In *The LASER Workshop: Learning from Authoritative Security Experiment Results (LASER 2016)*, San Jose, CA, 21–31. USENIX Association.

Kurashima, Takeshi, Tomoharu Iwata, Takahide Hoshide, Noriko Takaya, and Ko Fujimura. 2013. "Geo Topic Model: Joint Modeling of user's Activity Area and Interests for Location Recommendation." *Proceedings of the sixth ACM international conference on Web search and data mining*, Rome, Italy, 375–384. ACM.

Li, M., H. Xhu, X. Gao, S. Chen, L. Yu, H. Shanggian, and K. Ren. 2014. "All Your Location are Belong to Us: Breaking Mobile Social Networks for Automated User Location Tracking." *Proc. ACM Int. Symp. Mobile Ad Hoc Networks*, Philadelphia, Pennsylvania, USA, 43–52.

Lindqvist, J., J. Cranshaw, J. Wiese, J. Hong, and J. Zimmerman. 2011. "I'm the Mayor of My House: Examining Why People Use Foursquare-A Social-Driven Location Sharing Application." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11*, Vancouver, BC, Canada, 2409–2418.

Liu, B., W. Zhou, T. Zhu, L. Gao, and Y. Xiang. 2018. "Location Privacy and Its Applications: A Systematic Study." *Institute of Electrical and Electronics Engineers Access* 6:17606–17624. https://doi.org/10.1109/ACCESS.2018.2822260.

Malandrino, Delfina, Vittorio Scarano, and Raffaele Spinelli. 2013. "Impact of Privacy Awareness on Attitudes and Behaviors Online." *Science* 2 (2): –65.

Murakami, T., and H. Watanabe. 2016. "Localization Attacks Using Matrix and Tensor Factorization." *IEEE Transactions on Information Forensics and Security* 11 (8): 1647–1660. https://doi.org/10.1109/TIFS.2016.2547865.

Noulas, A., S. Scellato, C. Mascolo, and M. Pontil. 2011. "An Empirical Study of Geographic User Activity Patterns in Foursquare." *ICWSM* 5 (1): 70–73. https://doi.org/10.1609/icwsm.v5i1.14175.

Patil, Sameer, Roman Schlegel, Apu Kapadia, and Adam J. Lee. 2014. "Reflection or Action? How Feedback and Control Affect Location Sharing Decisions." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Toronto, Ontario, Canada, 101–110. ACM.

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

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

Rader, Emilee "Awareness of Behavioral Tracking and Information Privacy Concern in Facebook and Google." *Proc. of Symposium on Usable Privacy and Security (SOUPS)*, *Menlo Park, CA, USA*, 2014.

Rossi, Luca, and Mirco Musolesi. 2014. "It's the Way You Check-In: Identifying Users in Location-Based Social Networks." *Proceedings of the second edition of the ACM conference on Online social networks*, Dublin Ireland, 215–226. ACM.

Sadeh, Norman, Jason Hong, Lorrie Cranor, Ian Fette, Patrick Kelley, Madhu Prabaker, and Jinghai Rao. 2009. "Understanding and Capturing people's Privacy Policies in a Mobile Social Networking Application." *Personal and Ubiquitous Computing* 13 (6): 401–412. https://doi.org/10.1007/s00779-008-0214-3.

Sadilek, A., H. Kautz, and J. Bigham. 2012. "Finding Your Friends and Following Them to Where You are." *Proceedings of the fifth ACM international conference on Web Search and Data Mining, WSDM '12*, Seattle, Washington, USA, 723–732.

Scellato, S., A. Noulas, and C. Mascolo. 2011. "Exploiting Place Features in Link Prediction on Location-Based Social Networks Categories and Subject Descriptors." *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, San Diego, California, USA, 1046–1054.

Tang, Karen P., Jason I. Hong, and Daniel P. Siewiorek. 2011. "Understanding How Visual Representations of Location Feeds Affect End-User Privacy Concerns." *Proceedings of the 13th international conference on Ubiquitous computing*, Beijing, China, 207–216. ACM.

Tidwell, J., and C. Brewer. 2020. *Designing Interfaces: Patterns for Effective Interaction*. O'Reilly Media, Inc.

Tsai, Janice Y, Patrick Kelley, Paul Drielsma, Lorrie Faith Cranor, Jason Hong, and Norman Sadeh. 2009. "Who's Viewed You?: The Impact of Feedback in a Mobile Location-Sharing Application." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Boston, MA, USA, 2003–2012. ACM.

Wang, Na, Jens Grossklags, and Heng Xu. 2013. "An Online Experiment of Privacy Authorization Dialogues for Social Applications." *Proceedings of the 2013 conference on Computer supported cooperative work*, San Antonio, Texas, USA, 261–272. ACM.

Wang, Yang, Liang Gou, Anbang Xu, Michelle X. Zhou, Huahai Yang, and Hernan Badenes. 2015. "Veilme: An Interactive Visualization Tool for Privacy Configuration of Using Personality Traits." *CHI '15: CHI Conference on Human Factors in Computing Systems*, 817–826. ACM. https://doi.org/10.1145/2702123.27022.

Wei, Xuemei, Yang Qian, Chunhua Sun, Jianshan Sun, and Yezheng Liu. 2022. "A Survey of Location-Based Social Networks: Problems, Methods, and Future Research Directions." *GeoInformatica* 26 (1): 159–199. https://doi.org/10.1007/s10707-021-00450-1.

Zhang, Bo, Mu Wu, Hyunjin Kang, Eun Go, and S. Shyam Sundar. 2014. "Effects of Security Warnings and Instant Gratification Cues on Attitudes Toward Mobile Websites." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Toronto, Ontario, Canada, 111–114. ACM.

Appendix . Interview questionnaire

Pre-study

Demographics

- How old are you?

- What is your gender?

  - Male
  - Female

- What do you work/study?
- Where are you from?

  - North America
  - South America
  - Europe
  - Africa
  - Asia
  - Australia

Web and social networking background

- What is your experience in web applications and technologies?

  - Not too experienced
  - Somewhat experienced
  - Experienced
  - Very experienced

- I feel safe using Foursquare/Swarm

  - *5-point Likert Scale of Strongly agree to Strongly disagree*

- I am concerned about my online privacy

  - *5-point Likert Scale of Strongly agree to Strongly disagree*

- Have you ever regretted sharing certain information online?

  - Yes
  - No

- Have you ever deleted a post (comment, picture location) due to privacy concern?

  – Yes
  – No

- Have you ever requested to delete your data from a service before (e.g. request to delete the background location in Foursquare or all location information in Twitter)?

  – Yes
  – No

- Have you ever checked your shared check-ins using Foursquare History?

  – Yes
  – No

- Have you ever used tools/applications that help you manage your online privacy? (e.g. browser add-ons)

  – Yes
  – No

- If yes, like what?

- Can Foursquare/Swarm collect your location even if you are not using the application

  – Yes
  – No
  – I do not know

- Can Foursquare/Swarm share your data with third-party agencies for targeted-marketing/ advertising purposes

  – Yes
  – No
  – I do not know

- Can Foursquare/Swarm shares your data with third-party agencies to be used for other purposes (other than marketing)

  – Yes
  – No

– I do not know

- If yes, what do you think the other purposes are?

- Who do you think can access your check-in data?

- Who do you think can access your Swarm check-in data?

  – No one (Private)
  – The application (Foursquare/Swarm)
  – My friends on the application
  – Other users of the application
  – Third parties
  – I do not know

- When you share your location on Social Networks, what sort of information can be known about you?

- Have you ever checked your privacy settings in Foursquare/Swarm?

  – Yes
  – No

- How often do you update your privacy settings?

  – Rarely
  – Often
  – Always

- What aspects of your data can you control on a Foursquare/Swarm application?

- What of these listed options can you control on the Foursquare/Swarm application?

  – Who can see my contact information
  – Visibility of my check-ins to the place managers
  – Enabling my friends to check me in and including my name on their social media accounts

– Whether the application can collect my location when I am not using it (application is closed)
– Whether the application can collect my location while I am using it
– Check into a place privately (not seen by my friends)
– Getting behavioural targeted ads outside the application
– Getting behavioural targeted ads inside the application
– Deleting all of your check-ins
– Deleting your profile
– None

The actual study

This section involved the use of a location-data access tool specified for each group: Swarm History for the no-awareness group and a geoprofile visualiser tool for the awareness group. For each group, the interviewer started with a brief description of what the tool shows or provides and for a few minutes allowed the participant to explore his/her profile using the tool. Then, the participants were asked to carry out pre-defined tasks that were personalised for them on the basis of their generated geoprofile (seven tasks on average).

Information awareness and privacy attitude

• This tool helps me understand the information I share when I check-in
 - *5-point Likert Scale of Strongly agree to Strongly disagree*

• This tool allows me to view the information I share when I check-in
 - *5-point Likert Scale of Strongly agree to Strongly disagree*

• This tool helps me understand the possible information that can be extracted about me when I check-in (e.g. patterns of visits and top interests)
 - *5-point Likert Scale of Strongly agree to Strongly disagree*

• This tool allows me to view the possible information that can be extracted about me when I check-in (e.g. patterns of visits and top interests)
 - *5-point Likert Scale of Strongly agree to Strongly disagree*

• This tool allows me to know who can access my data
 - *5-point Likert Scale of Strongly agree to Strongly disagree*

• This tool motivates me to be more in control of my online data (e.g. deleting posts and updating privacy settings)
 - *5-point Likert Scale of Strongly agree to Strongly disagree*

• I am satisfied with the way Foursquare/Swarm collects and stores my data
 - *5-point Likert Scale of Strongly agree to Strongly disagree*

• This tool makes me more concerned about my privacy
 - *5-point Likert Scale of Strongly agree to Strongly disagree*

• This tool encourages me to alter the way I share my location information to protect my privacy

- *5-point Likert Scale of Strongly agree to Strongly disagree*

• Can the tool help you see/access your data collected/stored? If yes, to what extent?
• Can the tool help you find out who has access to your data? If yes, to what extent? in which ways?
• Can the tool help you find out the kind of information that can be extracted from your data? If yes, like what?
• After carrying out the tasks, are you concerned about your privacy? If yes, what triggers your concern?

post-study

• I feel safe using Foursquare/Swarm

- *5-point Likert Scale of Strongly agree to Strongly disagree*

• I am concerned about my online privacy

- *5-point Likert Scale of Strongly agree to Strongly disagree*

• I do NOT mind sharing my geoprofiles with others

- *5-point Likert Scale of Strongly agree to Strongly disagree*

The members of the awareness group were asked three further questions:

• Do you think that your initial understanding of your location data collection was limited? How?
• Do you think that your initial understanding of the possible utilisation and privacy implications of your shared location data was limited? How?
• Would you change the way you share location data after using this tool? How?
• Would you use such an application? Why?