Location-Related Privacy in Geo-Social Networks

Carmen Ruiz Vicente
Dept. of Computer Science
Aalborg University
Denmark
carmrui@cs.aau.dk

Dario Freni
EveryWare Lab, DICo
Università degli Studi di Milano
Italy
dario.freni@unimi.it

Claudio Bettini
EveryWare Lab, DICo
Università degli Studi di Milano
Italy
claudio.bettini@unimi.it

Christian S. Jensen
Dept. of Computer Science
Aarhus University
Denmark
csj@cs.au.dk

Abstract

Geo-Social Networks (GeoSNs) extend social networks by providing context-aware services that support the association of location with users and content. We are witnessing a proliferation of GeoSNs, and indications are that these are rapidly attracting increasing numbers of users. The availability of user location yields new capabilities that provide benefits to users as well as service providers. GeoSNs currently offer different types of services, including photo sharing, friend tracking, and "check-ins." However, the introduction of location generates new privacy threats, which in turn calls for new means of affording user privacy in GeoSNs.

This article categorizes GeoSNs according to the services they offer; it studies three privacy aspects that are central to GeoSNs, namely *location*, *absence*, and *co-location* privacy; and it discusses possible means of providing these kinds of privacy, as well as presents unresolved privacy-related challenges in GeoSNs.

1 Introduction

It is a recent trend in online social networks to enable the publication of geo-located information in real time. Many existing services are designed specifically to enable this functionality, and other services are increasingly assimilating these features. Examples include Facebook, Foursquare, Twitter, Google Latitude, Flickr, Gowalla, Loopt and MyTown. We call such services Geo Social Networks (GeoSNs): they combine real-time location capabilities with traditional social network functionality.

Indeed, the availability of location enables new possibilities, one example being the so-called *check-ins*, where users register when they arrive at locations. A business may offer discounts to those who check-in at their location, thus attracting more customers. In addition to benefiting from the discounts, users are able to identify popular places. The social networks of users enrich the functionality. For instance, a recommendation service that shows the currently popular places in town (by using check-ins) can highlight places that are popular among a user's friends.

Privacy advocates have warned about the dangers of exposing location information. The association of a user with a specific location can reveal health problems, affiliations, habits, etc. The availability of locations in real time even introduces threats such as assault. GeoSNs exacerbate the risks, as the spread of location information occurs more easily and is less controllable.

We use the term *location privacy* to denote the sensitivity of the association between a user's identity and the user's location, be it the user's past, current, or anticipated future locations. Knowing the location of a user also has implications for where the user *cannot* be located. This introduces the risk of burglary of unattended locations such as homes. We use *absence privacy* to denote the privacy of sensitive, unattended places. Absence privacy is ensured by avoiding the release of the association between a user's identity and the user's absence from sensitive locations. Location information can also be used to determine sensitive associations among users. Thus, user location traces may reveal that certain users have been together, possibly frequently and for extended time periods. We use the term *co-location privacy* to refer to the privacy of co-location events. Users enjoy co-location privacy if the co-location information about

the users is not associated with their identities.

GeoSNs should offer their users the possibility to specify when, how, and to whom their information should be exposed. To do so, GeoSNs should support different kinds of relationships and different accuracies of location information. This article offers an overview of existing GeoSNs and their location privacy-related mechanisms, and it describes the requirements of protection techniques that can be applied to GeoSNs.

2 Key Concepts and Classification of GeoSNs

2.1 Overview

GeoSNs extend social networks by enabling the association of location with users and content. Figure 1 depicts the main concepts in GeoSNs, including their relationships.

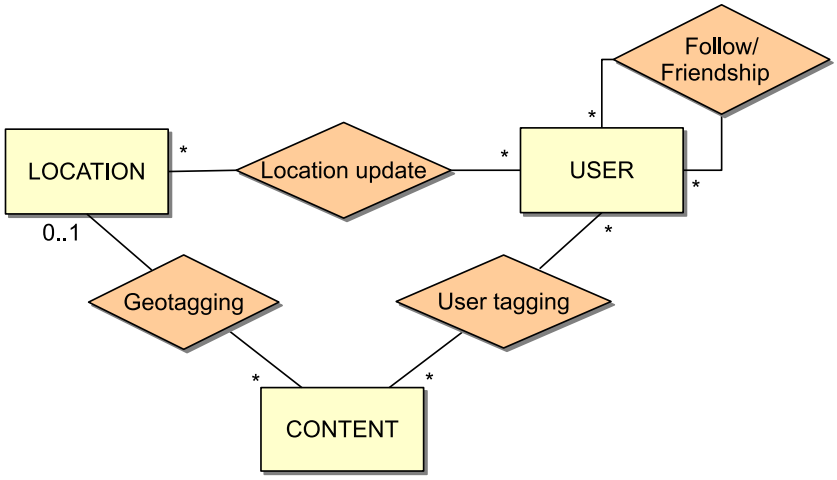

Figure 1: GeoSN Concepts

A *user* is an individual who subscribes to a GeoSN. Users can establish online relationships with other users, and the GeoSN in turn uses such relationships to enrich the services provided. The relationships are established by the users as a way of capturing real-life relations (family, co-worker, friend, etc.) or indicating affinities or common interests; they can be symmetric or asymmetric. Symmetric relationships, e.g., *friendship* in Facebook Places, require both users to approve the association, whereas asymmetric relationships, e.g., one form of *follow* in Twitter, do not require approval from the followed party. In the sequel, we use friendship and follow to denote symmetric and asymmetric relationships, respectively.

A GeoSN often allows its users to publish user-generated *content*, either to all or some of the users. This content may be associated with location, called *geotagging*, and with users, called *user tagging*. When publication of content is the key purpose of a service, we call it a *content-centric* GeoSN.

A user is *located* by means of a *location update*, which takes two forms: in some services, users can update their

location only at certain, predefined places (check-in), whereas other services are capable of continuously tracking a user. We call these *check-in based* and *tracking-based* GeoSNs.

Content and users can be associated with locations, and locations can be associated with content and users. Most GeoSNs are location centric in the sense that they enable convenient retrieval of content according to location. Thus, Flickr shows photos on a map, and Facebook Places discloses the people who are currently in a place.

Figure 2 reviews features of existing GeoSNs and classifies these into content-centric, check-in based and tracking-based GeoSNs. The characteristics of each group are described next.

2.2 Content-Centric GeoSNs

Content-centric GeoSNs allow their users to publish and share content such as notes, status updates, microblogging entries, traffic reports, photos, and videos. Examples include photo sharing services with social network capabilities such as Google Picasa and Flickr, as well as mobile micro-blogging services such as Twitter and Google Buzz. Content can be geotagged, thus enabling location-based search and retrieval and other services.

For some services, as illustrated by Figure 2, accurate location and time information are not strictly necessary. For instance, Twitter does not require publication of an exact location—rather, the geotag can be omitted or can be generalized to a coarser location (e.g., neighborhood, city). However, Twitter requires content to be published in real time. Other services focus on content that characterizes places, like reviews of businesses (e.g. Yelp, Qype). This calls for a location, but publication needs not occur in real time, and can use a pseudonym or be anonymous.

User tags associate content with related users. For instance, when uploading a photo to Picasa, users are allowed to *tag* other users that appear in the photo. User tags establish a link from a user to that content and make it accessible from the tagged users' profiles. The location of a user can be obtained from the geotags of content associated with the user: although a geotag is associated with a piece of content, the geotag implicitly tags the user with which the content is tagged. User tags are typically entered manually. However, the latest improvements in face recognition are starting to enable automatic user tagging (e.g., people suggestions in Picasa).

Users express diverging feelings about automatic user tagging [1]. One concern is that undesired content may become accessible to other users from a user's profile before the user's approval. A possible benefit is that if content related to a user is linked to the user, the user can be aware of its existence. The support in current GeoSNs for privacy preferences for content uploaded by other users is scarce.

2.3 Check-In Based GeoSNs

Check-in based social networks allow users to announce their presence in a particular place, such as a bar, a shop, or a park. The possible places are restricted to a finite set of points of interest and associated businesses, and the available places depend on the scope of the service. Users are in many cases allowed to submit new places. Examples include Facebook Places, Foursquare, Gowalla, Whrrl, and MyTown.

A check-in is normally performed manually by a user, and it is published instantly. Their visibility can be restricted to friends and to other users located in the same place, and it can be hidden from other users. Some services have adopted a popular gaming style in which users must perform certain challenges and are ranked based on performance, or are offered earn virtual rewards or real-world deals. Location-based social gaming services allow the use of pseudonyms instead of real identities.

Some services offer the possibility of doing a single check-in of multiple users (e.g., Facebook Places). However, analogously to user tagged content, users do not usually have control over the check-ins in which they are tagged.

2.4 Tracking-Based GeoSNs

Tracking-based services allow users to continually share their location with other users and to visualize other users' locations. However, different from check-in based services, the locations are retrieved automatically from a smartphone or similar. Users communicate their location by transmitting their latitude and longitude, and the list of possible locations is not restricted to a set of possible values. Services include Google Latitude and Loopt.

While these services have seen a significant growth in popularity, some users have concerns about constantly releasing their location to a wide audience. Google Latitude supports manual location updates, which gives full control to the user about when and where a location is disclosed. A subcategory of these services are the proximity-based services that allow users to discover nearby users. In some services, like online dating such as MeetMoi and Grindr, the other users are not necessarily friends of the user. Tracking services often allow the use of pseudonyms, especially when the participation of a user in a service is sensitive.

3 Privacy threats in GeoSN

3.1 Threat Categories

A *privacy threat* occurs when an adversary can associate a user's identity with information that the user considers sensitive. GeoSNs expose users to several privacy threats due to the release of spatio-temporal information. As in location-based services in general, two major categories of threats exist: (a) *re-identification through location* and (b) *release of sensitive location information.*

The former refers to the ability of an adversary to reduce the degree of anonymity of a user by considering location information. For example, by knowing that an (anonymous) GeoSN user was in a given place at a given time, an adversary may exclude several candidate individuals and possibly identify the user. If the user considers being a user of the GeoSN as sensitive, the re-identification is a privacy violation.

While we have seen that some GeoSNs allow the use of pseudonyms, this practice does not prevent this type of re-identification. GeoSNs, like proximity-based dating services and location-based social gaming for which anonymity can be particularly important, are exposed to this type of threat.

The latter category applies to cases in which the identity of the user is known and certain location information should not be associated by any adversary with the user. We have seen that some GeoSNs allow users to release location information to other users at a coarser granularity than actually available. This can be considered a way of avoiding this type of threat, based on the idea that the sensitivity of location data decreases at coarser granularities. We will see that current GeoSN privacy protection offers inadequate means of protecting against this category of threats.

We proceed to consider in more detail the threats related to the release of sensitive location information, since it is particularly important in GeoSNs.

3.2 Location Privacy

Revealing one's own exact location to other users is probably one of the most common concerns among GeoSN users. Indeed, the association of the user's identity with a specific *location* at a given time can reveal sensitive information such as health problems, affiliations, and habits. If we consider the user's identity as known to the adversary, protecting this association using anonymity-based techniques is not effective. Instead, it is necessary to obfuscate the location information available to an adversary, so that it is no longer sensitive according to the user.

3.3 Absence Privacy

The publication of a user's location can enable an adversary to infer how far the user is from a certain sensitive location at a given time. It is also easy for an adversary to compute, by considering maximum velocity and other constraints, a minimum interval of time during which the user could not be at the sensitive location. As an example, the adversary may estimate how long a user may be away from home, and may use this information to plan a burglary. As another example, a manager may determine that an employee was far away from work during work hours. In both examples, the release of location information represents an *absence* privacy violation.

3.4 Co-Location Privacy

In a check-in based GeoSN, it may be possible for one user to observe the simultaneous presence of other users, say, Alice and Bob, in the same place. Alice does not consider her presence in the place as sensitive, and neither does she consider it sensitive that Bob is a GeoSN friend. But she does consider it sensitive that she is in the same place as Bob. The disclosure of this information to an adversary that can identify Alice and Bob is a *co-location* privacy violation. The example can be generalized to consider as sensitive the *frequent* co-location with one or more specific users.

Intuitively, the location information that can be obtained by a GeoSN user by observing multiple users may reveal sensitive information about the relationship among those users.

The specification of privacy preferences related to this kind of threat can be particularly challenging. In addition to the frequency of co-location events, the presence of other users at the same location may be of importance. For example,

if Alice and Bob are attending the same event together with many other users, the co-location of Alice and Bob may not be considered as sensitive by Alice.

The co-location privacy threat applies in particular to check-in based and tracking-based GeoSNs, in which users have access to precise location information of several other users, and in which anonymity cannot be achieved easily.

4 Protecting Privacy

4.1 Overview

A number of techniques exist that aim to address the privacy threats covered above, but as they occur in traditional location-based services.

GeoSNs are more challenging than location-based services because they provide new ways to communicate location data through the publication of geotagged and user tagged content and multiple-user check-ins. Also, GeoSNs are conceived to expose location-related data, often in real time, to a large number of users, each one being a possible adversary. Partly for these reasons, absence and co-location privacy are particularly important concepts in GeoSNs. Therefore, existing techniques may need to be re-engineered to be of use in GeoSNs.

We proceed to briefly report the general principles of techniques from location-based services and then consider their application to the threats we have identified.

4.2 Spatio-Temporal Data Transformation Techniques

The main techniques can be grouped into spatial and temporal cloaking and encryption. In *spatial cloaking* a location is generalized to a region. The idea is that an adversary then only knows that the user is located somewhere within that region. Techniques differ in the regions they use and how they compute them. Similarly, *temporal cloaking* consists in altering the temporal data reported by a user (possibly together with a location). This is usually done by delaying a service request, or, in the context of content-based GeoSNs, by delaying the publication of a piece of content, so that an adversary has uncertainty about the actual time associated with the content. Finally, encryption based techniques are based on the encryption of the location information (or of the whole information being exchanged). Recent proposals allow a service request to be answered by a service provider without revealing to anybody, including the provider itself, the location data in the request [2]. Other proposals use multi-party secure computations to provide location privacy in proximity services [3]. Techniques differ in the encryption functions, in the protocols, and consequently in their costs.

4.3 Selecting and Adapting Techniques to Specific Threats

We consider the applicability of the above general techniques to the main threats that we have identified. For each threat we refer to our classification of GeoSN to identify general constraints on the applicability of the techniques. A summary

of the characteristics of each group and the applicable protection techniques is found in Table 1. The information is based on general properties of the groups and some exceptions may appear in particular services.

		content-centric	check-in based	tracking-based
Features	**Content as main information**	Yes	No	No
	Continuous location update	No	No	Yes
	Set of locations	Some services	Yes	No
	Multiple user tagging	Yes	Yes	No
	Real time	Some services	Yes	Yes
Applicable techniques	**Spatial cloaking**	Yes	No	Limited
	Temporal cloaking	Yes	Limited	No
	Encryption	Limited	Limited	Yes

Table 1: Features and Protection Techniques for GeoSNs

Avoiding re-identification through location

Whenever a user is interested in participating anonymously or through a pseudonym in a GeoSN, the user should be concerned about this threat. The actual risk of re-identification through location depends on the external information that an adversary can acquire to match the anonymous users that are in a location at a given time with their identities. This risk can be reduced with the above techniques, including spatial cloaking where a precise location of a user contained in a service request (or content publication) is transformed into a region covering the locations of at least k users, so that any adversary would not be able to associate the request (or content) with a specific user among the k candidates.

Spatial cloaking can be coupled with temporal cloaking in order to find smaller regions with respect to the ones identified by spatial cloaking alone. Intuitively, by allowing temporal uncertainty within an interval, it is sufficient to find a region that contained k users as the overall number of users that visited the region within that interval of time. While a temporal delay may not be tolerable in real-time services, it may be reasonable in some GeoSNs, content-based GeoSNs.

Protecting location privacy

When the location information is sensitive, spatial cloaking may be applied to obtain regions large enough to lower the sensitivity of the information to an acceptable level.

Temporal cloaking can also be used to lower the sensitivity of spatio-temporal information. Suppose a user does not want to reveal that she was located in a particular place at a certain time (maybe because she was expected to be at another location at that time). This can be classified as a location privacy violation. The threat can be eliminated by generalizing the temporal information to a larger interval such as an entire day.

In a recent article [4], we propose two solutions based on spatial and temporal cloaking to preserve location privacy in the context of a content-centric GeoSN that supports user tagging. This setting calls for careful adaptation of techniques known from traditional location-based services. Users specify their privacy preferences in terms of *minimum uncertainty regions* (MURs), which are spatio-temporal regions in which an adversary cannot exclude any of the points as possible locations of the users. For example, Alice specifies that any resource in which she is tagged should not report the specific

campus building where she was at 10:30 a.m. today. One solution is to define the combination of the entire campus region and the time period "this morning" as one of the MURs. When a resource is tagged with multiple users, that resource should be altered so that the different privacy requirements of all the tagged users are satisfied.

In check-in based services, spatial cannot normally be applied as it would eliminate the significance of the check-in. Temporal cloaking may be used if the service functionality tolerates a certain time delay. Encryption can also be used for check-in and content-centric GeoSNs, since the location information can only be decrypted by authorized trusted recipients. However, encryption introduces additional computational costs are is likely to reduce the utility of the data to the service providers.

For tracking based services, some solutions based on encryption and/or spatial cloaking have been proposed [3, 5]. However, these techniques are designed for the specific subcategory of proximity-based services.

Protecting absence privacy

The notion of absence privacy is relatively new, and only one proposal for absence privacy protection exists, presented for content-centric GeoSNs [4]. The privacy preferences expressed by the users are called *absence privacy regions* (APRs). The semantics of these regions is that no location information should be disclosed about a user such that it can lead to the exclusion of any of the points of any of the APRs as possible current location of the user. As already mentioned, user tags are also considered, and so the publication of a resource should satisfy all users' privacy preferences. The proposed technique determines a publication delay for each submitted resource so that it is safe to publish it.

The same principle of delaying the publication of the location information could be applied to check-in based services. Note however, temporal delay must be very limited to preserve the real-time functionality. For the same real-time requirement, delaying is not desirable in tracking-based services. Then it is still an open issue to devise an appropriate technique for absence privacy protection.

Protecting co-location privacy

To the best of our knowledge, the problem of protecting co-location privacy has not yet been investigated in the context of GeoSNs. In principle, any of the general spatio-temporal data transformation techniques can be applied. Apart from encryption for which the same arguments given for other threats apply, the other techniques should be adapted and evaluated with respect to specific co-location privacy preferences. Cloaking may be applied to one or more of the reported locations, and it may be used both to ensure that co-location involves sufficiently many people and to generalize the co-location region so that the two locations cannot be considered close to each other with sufficient confidence. The formalization of co-location as well as the design of protection techniques are actually relevant research open issues.

5 Challenges

There are a number of open problems regarding privacy preservation in GeoSN. In the following we briefly illustrate some challenges in this direction.

5.1 Formalization of Threats

The formalization of some of the privacy threats we mentioned in this paper is still an open task. While some initial work has been done for specific combinations of privacy threats and categories of services, there is still need for a more expressive formal model that could apply to all the scenarios we addressed. The formalization of the problem of co-location privacy, that to the best of our knowledge has not been investigated yet, is particularly challenging.

5.2 Specification of Privacy Preferences

The specification of a simple yet flexible way for users to express their privacy preferences in the context of GeoSN, considering the new threats we have identified, is another challenging task. Appropriate absence and co-location privacy preferences may not be easy to identify and to express since, as we have illustrated in Section 3, can involve presence or absence of multiple users, frequency of co-location events, and other aspects.

5.3 Managing Historical Data

Since GeoSNs have been recently introduced, the importance of aspects related to historical geo-referenced data that is continuously acquired by the service providers and by GeoSN users has not been fully recognized. Historical data can enable very interesting services as, for example, reminding when two users happened to be in the same place or attended the same event in the past; or being used to predict future locations of users. However, it can also expose users to new types of threats; For example, inferences on historical location data can lead to loss of anonymity [6]. This challenge includes modeling and managing approximate trajectories, devising new protection techniques, as well as possibly introducing policies enforcing data expiration.

6 Conclusion

Several sites and mainstream articles are alerting users about the risks of oversharing in GeoSNs without appropriate privacy controls [7]. Indeed, in many cases users are not aware that their publishing location information, like when uploading a picture from a smartphone that automatically geotags photos. The lack of standard privacy controls and privacy preserving techniques among GeoSNs makes it difficult for users to understand the risks and make appropriate actions to protect their privacy.

In this paper we provided a classification of current GeoSNs and identified the main privacy threats that users are exposed to. We have illustrated the state of the art privacy protection techniques that can be applied to prevent these threats and have analyzed their applicability depending on the GeoSN service being considered. The privacy issues involved in the participation to a GeoSN are still far to be fully understood and solved, but we hope that this work can provide a solid background and the identified technical challenges can stimulate further research.

References

[1] Andrew Besmer and Heather Richter Lipford. Moving beyond untagging: photo privacy in a tagged world. In *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems*, pages 1563–1572. ACM, 2010.

[2] Gabriel Ghinita, Panos Kalnis, Ali Khoshgozaran, Cyrus Shahabi, and Kian-Lee Tan. Private queries in location based services: Anonymizers are not necessary. In *Proc. of SIGMOD*, pages 121–132. ACM Press, 2008.

[3] Sergio Mascetti, Claudio Bettini, Dario Freni, X. Sean Wang, and Sushil Jajodia. Privacy-aware proximity based services. In *Proc. of the 10th International Conference on Mobile Data Management*, pages 31–40. IEEE Computer Society, 2009.

[4] Dario Freni, Carmen Ruiz Vicente, Sergio Mascetti, Claudio Bettini, and Christian S. Jensen. Preserving location and absence privacy in geo-social networks. In *Proceedings of the 19th ACM Conference on Information and Knowledge Management (CIKM 2010)*, 2010, to appear.

[5] Ge Zhong, Ian Goldberg, and Urs Hengartner. Louis, Lester and Pierre: Three protocols for location privacy. In *Privacy Enhancing Technologies*, volume LNCS 4776, pages 62–76. Springer, 2007.

[6] Claudio Bettini, X. Sean Wang, and Sushil Jajodia. Protecting privacy against location-based personal identification. In *Proc. of the 2nd VLDB workshop on Secure Data Management*, volume 3674 of *LNCS*, pages 185–199. Springer, 2005.

[7] Gerals Friedland and Robin Sommer. Cybercasing the joint: On the privacy implications of geo-tagging. In *USENIX Workshop on Hot Topics in Security*, 2010.

7 Biographies

Carmen Ruiz Vicente is a PhD student in the Computer Science Department at Aalborg University, Denmark. Her research interests include location privacy in location-based services and social networks. She has a Computer Engineering degree from the Universidad de Valladolid, Spain.

Dario Freni received his M.Sc. degree in computer science from the Università degli Studi di Milano in 2007 and is currently a Ph.D. student at the DICo department of the same university.

Claudio Bettini is Professor of Computer Science in the Dipartimento di Informatica e Comunicazione (DICo) of the Università degli Studi di Milano, Italy, where he leads the EveryWare Laboratory. He is also a member of the Center for Secure Information Systems at George Mason University, VA. His main research interests are Temporal and Spatial Data Management, Mobile and Pervasive Computing, Knowledge Representation and Reasoning, Security and Privacy. He is a member of ACM SIGMOD.

Christian S. Jensen, Ph.D., Dr.Techn., FIEEE is a Professor of Computer Science at Aarhus University, Denmark. From September 2008 to August 2009, he was on sabbatical at Google Inc., Mountain View. His research concerns data management and spans semantics, modeling, indexing, and query and update processing. During the past decade, his focus has been on spatio-temporal data management. He is a member of Royal Danish Academy of Sciences and Letters, the Danish Academy of Technical Sciences and the EDBT Endowment, and a trustee emeritus of the VLDB Endowment. He is vice president of ACM SIGMOD and an editor-in-chief of the VLDB Journal and has served on the editorial boards of ACM TODS and IEEE TKDE.

	Publication of resources	Relationships (Symmetric / Asymmetric)	Continuous communication of location	Multiple user tagging	Exact location required	Real time	User identity ((Real, Pseudo., Anony.)	Group [3]
Facebook Places		Friends (S)		Yes	Yes	Yes	R	CH
Foursquare		Friends (S)			Yes	Yes	P	CH
Twitter	Yes	Followers (A)				Yes	P	CO
Google Latitude		Friends (S)	Yes[2]				R	T
Gowalla	Yes[1]	Friends (S)			Yes	Yes	P	CH
MyTown					Yes	Yes	P	CH
SCVNGR		Friends (S)		Yes	Yes	Yes	P	CH
Whrrl					Yes	Yes	P	CH
MeetMoi			Yes		Yes	Yes	P	T
Flickr	Yes	Friends (S)		Yes			P	CO
Picasa	Yes	Followers (A)		Yes			P	CO
Brightkite	Yes[1]	Friends (S)			Yes	Yes	P	CH
Google Buzz	Yes	Followers (A)				Yes	R	CO
Yelp	Yes				Yes		A	CO
Qype	Yes				Yes		P	CO
Grindr			Yes		Yes	Yes	P	T
Loopt	Yes[1]	Friends (S)	Yes		Yes	Yes	R	T
Gbanga		Friends (S)			Yes	Yes	P	CH
Geocaching	Yes				Yes		P	CH
Waze	Yes	Groups (S)	Yes		Yes	Yes	A	CO
Trapster	Yes	Friends (S)			Yes	Yes	A	CO

[1] The publication of content is supported but it is not indispensable for the service
[2] Also manual update is supported
[3] Checking-based (CH), Tracking-based (T) and Content-centric (CO)

Figure 2: Features of Existing Services