# The pretense of residential privacy in geosocial networking data

Grant McKenzie

The Department of Geography
The University of California, Santa Barbara, CA, USA

**Keywords:** Geoprivacy, Location-based Social Networking, Check-ins

## 1 Introduction

One of the top concerns for users of geosocial networking applications is the privacy of the information being shared. While most location-based social networking applications put a great deal of time and effort into voicing how much they care about privacy, not much is known about steps that are taken to ensure the information is kept private and/or anonymous. One of the issues unique to *geo*social networking applications such as *Foursquare* is the concern over the privacy of one's home location. Sharing textual content and photographs is one thing, but sharing the location of your house is something different entirely. Existing work in this area has focused on the different levels of privacy associated with geosocial networking [6,3]. The ability to both preserve the privacy of some data while sharing other parts of the data has proven to be a difficult task for researchers and application developers alike. Residential check-in behavior has been the focus of previous research in this area [5,4] while satirical work *(PleaseRobMe.com)* has shown what is possible given knowledge of an individual's home location [1].

In this work, Points of Interest (POI)[1] of type *Home* are examined on the *Foursquare* platform with the goal of determining how much information is actually accessible and to explore the ways in which the data is kept *private*. The purpose of this work is not to expose individuals' real home locations, but rather examine the data, the privatization methods and the ways in which this privatization is being circumvented by the contributors themselves. This paper presents a snapshot of work-in-progress and should be taken as such. While the results of this paper are promising, further work exploring numerous other privatization practices is being conducted on additional geosocial datasets with varying privacy parameters.

## 2 Data

POI information was accessed in October 2013 via the public *Foursquare* application programing interface (API) for 333,094 venues of type *Home (private)*[2] across the United States. The attribute information available (and of interest for this research) for each venue consists of a *unique identifier*, *venue name*, *Twitter username*, *phone number*, *address*, *city*, *state*, *postal code*, *latitude* and *longitude*. While entry of most of these

[1] Foursquare refers to Points of Interest as *Venues*.
[2] Full set of categories accessible at https://developer.foursquare.com/categorytree

attributes is optional, all of these attributes are contributed by users of the Foursquare application. The original latitude and longitude values are most often determined via the positioning technology (e.g., GPS, WiFi) employed by the contributor's mobile device or on occasion, geocoded based on the user-contributed address of the establishment. Of particular interest to this research is that the latitude and longitude values contributed by the user are altered by Foursquare to preserve privacy before being published via the API. The *Foursquare API* documentation states the following:

> *Some venues have their locations intentionally hidden for privacy reasons (such as private residences). If this is the case... the lat/lng parameters will have reduced precision.*[2]

Figure 1 shows a sample of these venues (with reduced geographic precision) around New York City. The reduced precision of the geographic coordinates, while preserving privacy, do not visually alter the distribution of *Homes* in the region. This is important as we do not see any geographic patterns emerge from the privatized results (e.g., coordinate rounding would lead to a visual grid).

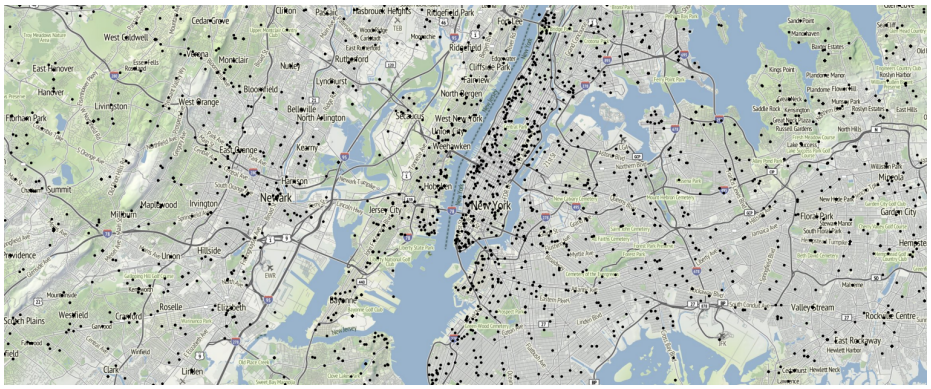


Fig. 1: A sample of venues of type *Home* in and around New York City, NY. The geographic coordinates of the venues have been adjusted (by *Foursquare*) to preserve privacy.
(Basemap ©Stamen Maps and OpenStreetMap contributors)

In discussing privacy concerns with geosocial data it is of value to see what other type of information (besides coordinates) is being shared about these private venues. Of the 333,094 *Home* venues sampled, 3,435 (1.0%) listed a Twitter account and 1,608 (0.5%) showed a phone number (containing 7 or 10 digits). The median number of *check-ins* per venue was 7 with a mean of 61.2 (sd 191.2) and the number of unique *users* showed a median of 1 with a mean of 2.8 (sd 14.5) implying that in most cases only the person that contributed the venue actually *checks in*. Furthermore, users of the platform have the option of *claiming* a venue through a verification processing and in this sample 388 (0.1%) of the venues were claimed.

## 3 Privacy Distance

The more interesting contribution to these *Home* venues is in both the *Address* and the *Name* fields. Given the concern for privacy, one might assume that the *Address* field

would most often contain city or neighborhood level information. While this is the most common case, in many instances, a street name is provided and occasionally even a full street address is given. Similarly, the *Name* field provides some unexpected information regarding the location of the contributed private venue. Entries ranging from *"Steve's Party Shack"* to *"The Pad"* show up in this field. It is important to point out that while most of the *Name* fields do no include globally identifiable information, many of them do present some type of personal identifier such as the resident's given name (e.g., *"Grant's Castle"*) or surname (e.g., *"The McKenzie Residence"*). In some cases street addresses such as *"123 Main Street"* are given as the venue's name. Combining these fields with the *State* and *Post Code* fields allows for much of the attribute information attached to these venues to be geocoded.

A powerful feature of the *Google Geocoding API* is that it returns an *Address Type* or what one might call a *precision level* for each of the character strings it geocodes.[3] These *precision levels* range from *street address* (high precision) to *state level* (low precision). Geocoding both the *Address* and *Name* fields of each venue produces the results shown in Table 1. The values reported for each field are shown as percentages of the sample venues. Not surprisingly *City* and *Post Code* make up the majority of the results with *County or State* and *Neighborhood* showing the next highest values. Interestingly 2.2% of the *Home* sample venues can be geocoded to the street address level.

| **Precision Level** | **Address** (%) | **Name** (%) |
|---|---|---|
| Street Address | 2.2 | 2.2 |
| Intersection | 0.8 | 0.1 |
| Street Name | 5.9 | 3.2 |
| Neighborhood | 4.6 | 4.3 |
| Post Code | 24.8 | 0.0 |
| City | 49.8 | 36.1 |
| County or State | 9.1 | 4.0 |
| Other | 1.7 | 1.2 |
| *No geocode results* | *1.1* | *48.9* |

Table 1: The precision levels for geocoded values taken from the *Address* and *Name* fields. Shown here as a percentage of all sample venues of type *Home*.

Provided these 7,328 (2.2%) high precision geocoded results, the distance was calculated between the geographic coordinates returned from the geocoder and the *reduced precision* geographic coordinates provided via the Foursquare API. A histogram showing the results of these distance calculations is shown in Figure 2. By comparison, a continuous fit line (gray) is shown as an overlay on top of the histogram. Based on these findings it appears that the *reduced precision* introduced by the application developers involves adjusting the geographic location of the *Home* venue by drawing a random value from a Gaussian distribution centered 15 meters from the actual geographic location with a standard deviation of roughly 32 meters.

[3] https://developers.google.com/maps/documentation/geocoding/

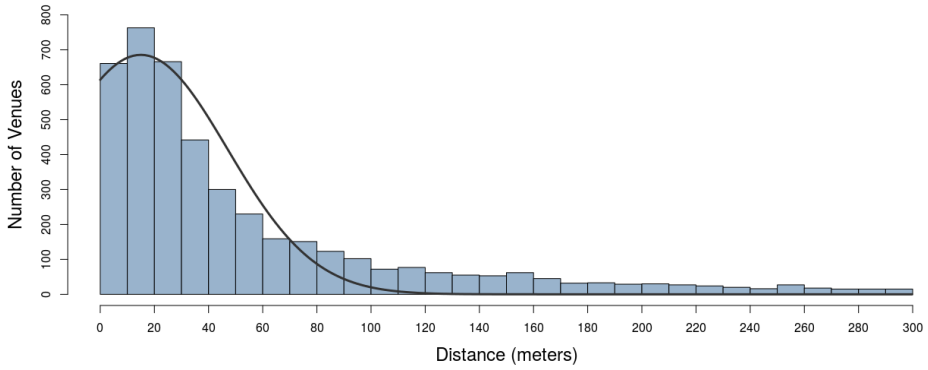


Fig. 2: Number of venues binned by distance between the *reduced precion* geographic coordinates and the *street addressed* geocoded coordinates.

## 4 Conclusions

As the role of location continues to grow in online social networking, concerns over data privacy grow with it. Applications in this field are faced with the difficult task of allowing users to share location-based social information (e.g., Venues and check-ins) while preserving the privacy of information that users do not want shared. In this work, the way in which geosocial data is privatized, namely *Home* locations is explored not with the purpose of exposing private and personal information, but instead to demonstrate the ways in which user-generated geo-content can be privatized. An additional finding of this work is that in many ways, the contributors of this private information are their own worst enemies. Personal identifiers such as first and last names, Twitter usernames, phone numbers and even street addresses are openly published as part of a category of geolocation data that is meant to remain private.

## References

1. Dan Fletcher. Please rob me: The risks of online oversharing. *Time Magazine online*, 2010.
2. Foursquare. Foursquare developer api - venue response documentation. https://developer.foursquare.com/docs/responses/venue. Accessed: 01/03/2015.
3. 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 international conference on Information and knowledge management*, pages 309–318. ACM, 2010.
4. Sébastien Gambs, Marc-Olivier Killijian, and Miguel Núñez del Prado Cortez. Show me how you move and i will tell you who you are. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Security and Privacy in GIS and LBS*, pages 34–41. ACM, 2010.
5. Lei Jin, Xuelian Long, and James BD Joshi. Towards understanding residential privacy by analyzing users' activities in foursquare. In *Proceedings of the 2012 ACM Workshop on Building analysis datasets and gathering experience returns for security*, pages 25–32. ACM, 2012.
6. Carmen Ruiz Vicente, Dario Freni, Claudio Bettini, and Christian S Jensen. Location-related privacy in geo-social networks. *Internet Computing, IEEE*, 15(3):20–27, 2011.