State of the Geotags: Motivations and Recent Changes

Dan Tasse, Zichen Liu, Alex Sciuto, Jason I. Hong
Human-Computer Interaction Institute
Carnegie Mellon University
dantasse@cmu.edu, zichenl@andrew.cmu.edu, sciutoalex@cmu.edu, jasonh@cs.cmu.edu

Abstract

The widespread adoption of smartphones has made it possible for large numbers of people to geotag their social media posts. Past work has studied the reasons people tag their location and the ways they do so on location-based social networks like Foursquare. But it is unclear how well these findings generalize to other social media not centered on location, such as Twitter or Flickr. Through an analysis of public data and two surveys, we investigate why people geotag their photos, tweets, and other non-location-based social media. We found that their reasons are similar to those in location-based social networks. We also found several surprises due to the different nature of these platforms and the changes since location-based social networks were introduced. For example, people usually consciously geotag, though a significant portion geotags unintentionally; coordinate geotagging is changing to placetagging; and job-posting bots constitute a growing portion of public geotags.

Introduction

Past work has investigated people's behaviors on location-based social networks (LBSNs) like Foursquare, Facebook Places, Dodgeball, and Gowalla, to learn who shares their location and why (Lindqvist et al. 2011; Guha and Birnholtz 2013; Tang et al. 2010). However, many non-location-based social media services–including Facebook, Instagram, Twitter, Snapchat, and Flickr–also allow users to add their location to a photo, video, or text post. One study estimated 600,000 geotagged posts per day from a 10% sample of Twitter (Leetaru et al. 2013). Flickr, meanwhile, has released a publicly available dataset of 49 million geotagged photos (Thomee et al. 2016). A wide field of research has sprung from this wealth of publicly available geotagged posts, such as understanding demographics and social dynamics in cities (Mohammady and Culotta 2014), finding home locations of individuals (Jurgens, Mccorriston, and Ruths 2015), and inferring likely friends (Crandall et al. 2010).

For LBSNs, there is often a clear reason as to why someone checks-in to a location or tags something. However, there is currently little understanding of what people geotag on these popular social media sites not centered around location, and why. Understanding what is being geotagged and why can have implications for research. For example, finding people's home locations based on their check-ins could be very easy if people geotag mostly at home, or very difficult if they only geotag when traveling. As another example, models about people's mobility patterns and social dynamics will be very different if they are based on commute data or weekend shopping and errands data.

It is easy to assume that geotagging in social media is similar to that for LBSNs, but without explicit investigation it can be difficult to know for cetain. To address this problem, we conducted a series of studies to understand whether people use geotagging in social media similarly to the way they use location-based social networks. We analyzed 4 million public tweets and 49 million geotagged Flickr photos, surveyed 78 frequent Twitter users, and followed up by surveying 400 geotaggers across six social media services. We found that most earlier findings in Foursquare ring true in other social media: people geotag consciously and intentionally, they geotag in uncommon places, they primarily do so to communicate and show where they've been, and they geotag soon after being at the place.

However, our analyses uncovered several new findings. We found that most Twitter users geotag consciously and turn geotagging on and off frequently, but many Twitter users were inadvertently geotagging, or geotagging more precisely than they thought. We uncovered a UI change that addressed this issue, while also causing people to add coordinates less frequently and add place names more frequently. We also discovered that the coordinate geotags that remain tend to have more hiring-related spam.

Our findings have several research implications, given how often researchers use geotagged data. It is important that the research community not misunderstand what people are providing when they publish geotagged social media posts, and it is important that we minimize the impact of spam and other quality problems. Knowing why people geotag also helps application developers better customize their software.

To support these research and development implications, this paper offers two contributions. First, we show confirmation and elaboration of earlier findings, and generalization from Foursquare to other social media. Second, we expose a number of changes that have occurred as location sharing has matured.

Copyright © 2017, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Definitions

In this paper, we use the term "geotag" to mean "a location added to a social media post." We use "coordinate geotag" to mean "an exact latitude-longitude coordinate added to a post." A related concept is a "placetag", or a tag referring to a plain-text location. For example, a post at the Eiffel Tower could contain the *placetag* "Eiffel Tower," "Paris," or even "France." It could also or instead include the *coordinate geotag* (48.858, 2.295).

We refer to "checking in", as in Foursquare and other location-based social networks, as a separate but related act. In Foursquare, one opens the app primarily to share one's location. When geotagging a tweet, photo, or other post in social media like Twitter, Flickr, Instagram, and Facebook, however, the content of the post is usually the primary motivation, while location sharing is usually secondary.

Related Work

To motivate our work, we first discuss a small sample of the wide variety of research being done using geotagged social media data. Then, we outline other research on social annotation that informed our inquiries, as well as prior work on geotagging in LBSNs.

Research uses of geotags

A great deal of past work has used geotagged social media for various purposes. Some examples include using geotagged social media to understand demographics, such as race and gender (Mohammady and Culotta 2014), to infer well-being (Schwartz et al. 2013), and to investigate feasibility of delivering packages between Twitter users, based on where the Twitter users are already posting geotagged tweets (Sadilek, Krumm, and Horvitz 2013).

Some work has investigated the potential for using tweets and other geotagged social media data to aid our understanding of city dynamics. LiveHoods (Cranshaw et al. 2012) and HoodSquare (Zhang et al. 2013), for example, used Foursquare checkins to find organic boundaries of neighborhoods. Other examples include studies to understand economic effects of road closures (Zhang, Li, and Hong 2016), find similar neighborhoods in different cities (Le Falher 2014) or understand whether a term belongs to a small area or a large region (Kafsi et al. 2015).

Researchers have also investigated how to infer user locations based on tweet content, homophily of friends, or past geotags (Mahmud, Nichols, and Drews 2012; Pontes et al. 2012; Priedhorsky, Culotta, and Valle 2014). Most of these attempts are able to classify users effectively at the city level; Jurgens (Jurgens, Mccorriston, and Ruths 2015) provides a good overview of this body of work.

The past work described above are a few examples of research using geotagged social media, to help demonstrate that research on geotagged social media is a flourishing area with both theoretical and practical results. Our research examines the influences and reasons for why people add geotags to their social media data as well as what content is geotagged. Our findings can help other researchers in analyzing or creating algorithms that make use of geotagged data, for example in filtering out certain kinds of unuseful geotagged content or making sure that certain categories of geotagged content are accounted for by their algorithms.

	Organization	*Communication*
Self	Retrieval, directory	Context for self
	Search	Memory
Social	Contribution, attention	Content descriptors
	Ad hoc photo pooling	Social signaling

Table 1: Taxonomy of tagging motivations, from Ames et al, 2007. The same motivations can also apply to geotagging.

Why people add textual tags to photos

Another relevant body of work has looked at why people manually annotate their photos with tags such as "CHI" or "Golden Gate Bridge." Ames and Naaman (2007) describe a 2x2 taxonomy of motivations for tagging photos (see table 1): they can be for themselves or for social purposes, and they can be for organization or for communication. The authors detail what each quadrant means: self/organization tags help the user find photos later, self/communication tags help the user add context like people's names, social/organization tags help other users find their photos, and social/communication tags help involve context and signaling to other people. Nov and Ye expanded their work to show that social presence and motivation to communicate and organize for a public audience correlate with increased tagging (Nov and Ye 2010).

Ames et al's framework and Nov and Ye's findings helped guide our inquiry into why people geotag. Geotagging, like photo tagging, requires some effort; geotagging also requires assessing privacy risks. Organization and communication motivations can begin to explain why people take this effort and risk. If geotagging is like photo tagging, we would expect to see the same: motivations primarily centered around organization for oneself, and organization and communication to the public.

Why people check in

Most closely related to our work here is work done on location-based social networks. These studies provide some hints as to what people geotag and why. For example, Lingel et al describe how "checking in" to places on Facebook can help transnational migrants manage their identity as they move between countries by showing off places that they are or sharing jokes about places with their friends (Lingel, Naaman, and Boyd 2014). Similarly, Lindqvist et al (2011) and Cramer et al (2011) discuss location-sharing on the location-based service Foursquare as a performative and communicative act. These works, drawing on Goffman (1959), reveal the performative and communicative aspects of the service. Our work reveals similar purposes in geotagging throughout different social media platforms.

Another lens for understanding why people geotag is to look at why they would choose not to geotag. Privacy is one clear reason, as has been found in past studies on location sharing in general. For example, Consolvo et al studied

people's willingness to disclose their location when asked (Consolvo et al. 2005). However, as Tang et al note (2010), a location request is a different process than adding a tag to a public post, so some of the previous work on location sharing may not apply to geotagging. Wu and Zhang look at a more directly related process: when people opt to add their location to photos they share (2011). They found that people are more likely to share their location in a public space. This provides one piece of the puzzle, which we elaborate on further in our paper. A few more pieces are provided by Ahern et al (2007), who found that people focus on security, social disclosure, identity, and convenience when thinking about privacy and location sharing.

Sloan and Morgan (2015) address the angle of geotagged tweets, showing that users who tweet in some languages (like Turkish, Portuguese, and Indonesian) are much more likely to geotag than users in other languages (like Korean, Japanese, and Russian). There are also modest effects based on age and gender. They do not posit reasons for these differences, however.

These are all views on angles of location sharing, though. They focus on specific cases (like migrants) or parts of the geotagging decision (like privacy). Our work is most related to the work of Wu and Zhang (2011) and Ahern et al (2007), but we expand on their work in a few ways: we consider real posts that they have geotagged instead of hypothetical cases, we combine large-scale quantitative analysis to deepen the results of surveys, and we more thoroughly investigate the benefits of geotagging in addition to the costs.

Study 1: Analyzing Public Geotagged Data

To understand what and why people geotag, we chose to start by analyzing publicly visible data from Twitter and Flickr, primarily because they offer the largest public data sets of geotagged posts.

We started by collecting geotagged tweets via Twitter's public streaming API. We chose to start in Pittsburgh because it has a wide variety of users and because of our team's high familiarity with the area. We selected all coordinate-geotagged tweets within 0.2 degrees latitude and longitude from the center of Pittsburgh, forming a 34km x 44km rectangle with corners at (40.241667, -80.2) and (40.641667, -79.8). We omitted tweets that listed an area (like "Pittsburgh") but did not contain a latitude-longitude point. We began gathering data in January 2014, and by May 2016 we had about 4 million tweets. We also gathered data in 12 other cities, mostly around the United States[1], to verify any results we found on the Pittsburgh data set. These other cities' data covered a shorter time span (11-23 months) but still the same order of magnitude of tweets, from about 1 million in Austin to 11 million in London, totaling 60 million tweets.

We also examined the YFCC100M dataset (Thomee et al. 2016) to gather information about geotagging on the photo-sharing site Flickr. This data set contains metadata for 100 million photos and videos that are shared publicly with a Creative Commons license. Of these 100 million photos and videos, about 49 million are geotagged.

[1] Austin, Chicago, Cleveland, Dallas, Detroit, Houston, London, Miami, Minneapolis, New York, San Francisco, and Seattle

Study 1 Results

People often toggle geotagging

In Twitter's mobile app, users can choose to geotag or not, but the default is whatever was set last. If a user geotags one post, the next one will be geotagged as well unless the user turns it off. As a result, we had initially assumed that geotagging was a setting people would mostly leave on or off; that they would decide to geotag or not to geotag and then apply that to all of their social media. However, this was not the case.

We selected a random sample of 3406 users from our data set and collected all of their public tweets, geotagged and non-geotagged. We sampled users because Twitter's API has rate limits of 180 requests per 15 minutes, and because it only supports collecting up to 200 tweets per request. As such, collecting all tweets from all 68088 users would have taken prohibitively long.

For each of those users, we sorted their tweets in chronological order, then counted a "toggle" every time they had a geotagged tweet followed by a non-geotagged tweet, as this likely indicated they had made a conscious choice to geotag or not geotag something. We only used tweets from their most frequent tweet source (such as "Twitter for iPhone" or "Twitter Web Client") to avoid counting false "toggles" caused by them, for example, tweeting from their phone then tweeting from their computer. We found that most people in this sample toggled geotagging relatively regularly, and only a minority (40.1%) toggled less than 1% of the time, as we would expect if they were geotagging automatically. Figure 1 shows the distribution of how many of each users' tweets were preceded by a toggle.

We saw similar patterns in the YFCC100M dataset. Out of the 581099 Flickr users in the data set, 214598 (36.9%) had posted at least one geotagged photo. For these users who have geotagged at least once, we counted toggles the same way as for Twitter users; results are shown in Figure 1. Seeing roughly the same pattern as in Twitter gives us confidence that geotagging on social media platforms is a conscious choice, not automatic.

Changes in geotags over time

Although it was not a primary research question, an interesting finding emerged about the pattern of geotags over time. In the YFCC100M dataset (see Figure 3), the overall count of photos is rising in frequency, though geotagged photos are tapering off. However, (Jiang et al. 2017) found the percent of photos with geotags in Flickr is increasing. One possible explanation is that the decline in geotagged photos is an artifact of the process of creating the YFCC100M dataset (or people's willingness to license photos as Creative Commons).

In our Twitter data, however, we noticed a sharp dropoff in the number of geotagged tweets available after about May 2015. This was consistent across all of our cities; Figure 2 shows the dropoff in a few example cities.

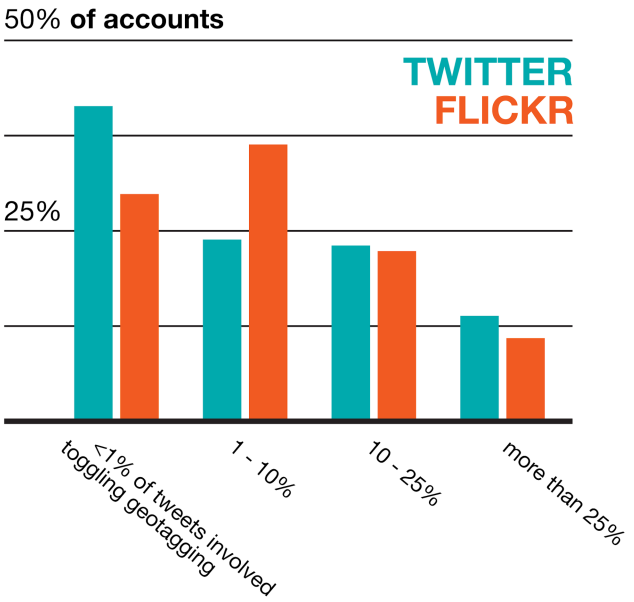


Figure 1: Percent of tweets and photos in which users turned geotagging on or off. Tweets are taken from a random sample of 3406 users out of all 68088 users who had geotagged at least once in Pittsburgh; Flickr photos are taken from the YFCC100M dataset. The "Less than 1%" group can be considered those who always or never geotag; it is a minority, as most people toggle geotagging at least occasionally.

This is due to a UI change Twitter made in April 2015, when announcing a new partnership with Foursquare[2]; placetagging, not coordinate-geotagging, is now the default. As we will discuss in Study 2, while this reduced the data available to researchers, it also removed a source of confusion and accidental privacy leaks.

How many distinct places do people geotag at?

We calculated how widely people's geotag distribution varied. After excluding all accounts with fewer than 20 geotagged tweets, we rounded each geotag to the closest 0.001 degree latitude and longitude (creating bins about the size of 1-2 city blocks), then counted how many places each account had posted a geotagged tweet. We found that most people had between 1 and 35 places (median = 18, 3rd quartile = 35, 95th percentile 103.8; see figure 4), but of course this depends on the number of tweets. We also calculated the number of tweets per place, finding that people usually tweet about 4 times in a place, though this varies widely (1st quartile = 2.3, median = 3.9, 3rd quartile = 7.0, 99th percentile = 102.6).

One surprise we found is that some accounts geotag repeatedly in the same place. We inspected a random sample of 50 of the one-place accounts in the Pittsburgh area and found that about 15 accounts were bots that only posted job listings, 7 accounts had been deleted or protected, 1 was a bot that

[2]https://twittercommunity.com/t/foursquare-location-data-in-the-api/36065

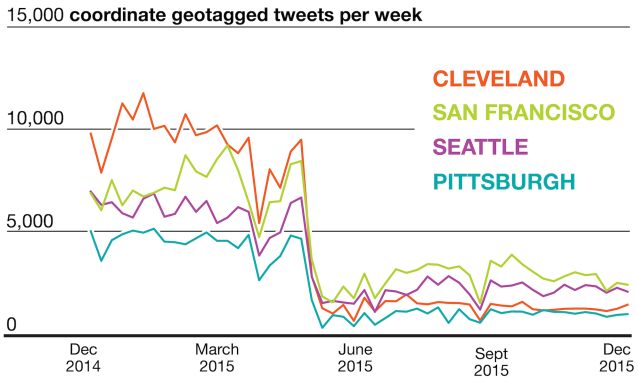


Figure 2: Counts of coordinate geotagged tweets from different cities. The sharp dropoff around May 2015 is due to a Twitter UI change: placetagging, not coordinate geotagging, is now the default.

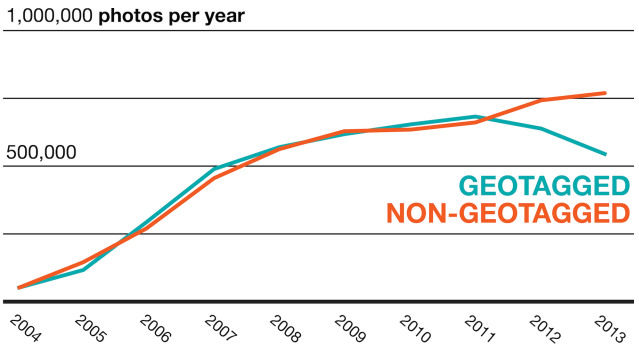


Figure 3: Counts of geotagged and non-geotagged photos from the YFCC100M dataset. Note that, while non-geotagged photos are growing steadily, geotagged photos may be falling.

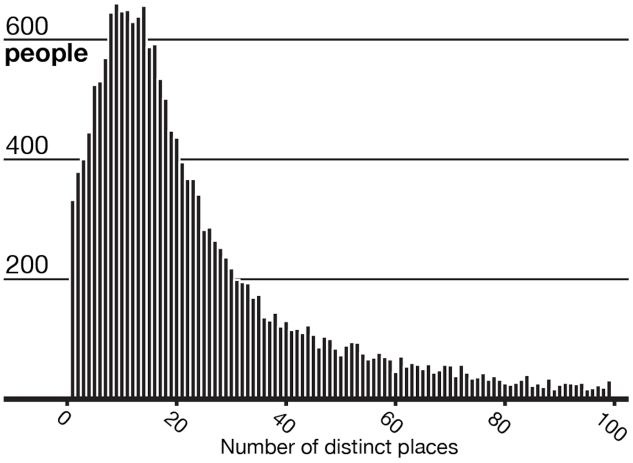


Figure 4: Number of distinct places where each Twitter user in our Pittsburgh data set (with at least 20 tweets) geotags. Notice how many users tweet only in one place.

City	One-place accts	Job bots	%	Multi-place accts	Job bots	%
Pittsburgh	331	73	22.1	19,214	179	0.9
SF	394	103	26.1	31,777	217	0.7
Seattle	259	78	30.1	15,571	203	1.3

Table 2: Distribution of job posting bots on Twitter in Pittsburgh, San Francisco (SF), and Seattle. "One-place accounts" are accounts that post in only one location (rounded to the nearest 0.001 degree latitude and longitude). "Many-place accounts" post in multiple locations. Not all job posting bots post in one place, but a large percentage of one-place accounts are job bots.

tweeted weather reports, and 1 was a bot that tweeted NHL hockey scores. While we could not detect every bot, all of the job posting bots included "job", "career", "work", "join", or "tmj" in their name (and none of the other accounts did), so we could easily scan for other bots in the full data set. We found that a large percentage of one-place accounts were job posting bots; complete statistics are shown in table 2.

Of course, there are plenty of spam accounts on Twitter besides job posting bots. Some of them, like realistic-looking accounts made to promote a product, are difficult to filter out. However, we point out the job bots to show one easy way that researchers who are analyzing geotagged tweets can easily remove a large quantity of tweets that may not be relevant to their purposes.

Study 2: Survey of Twitter Geotaggers

In Study 1, we found that social media users often toggled geotagging, which suggests that people may have nuanced views of privacy and sharing. However, there have also been news articles indicating that people sometimes accidentally shared geotagged media too. We were interested in probing these behaviors more. Towards this end, we conducted a survey of Twitter users.

In November 2014, we compiled a list of our users, sorted by the number of times they tweeted in our data set. After our study was approved by our IRB, we recruited 4119 participants to take a survey by tweeting a link to them. We started from the most prolific tweeters in order to make sure we had active users. A total of 78 responded and were paid with a $5 Amazon.com gift card for participating. (While we wish we could have had a higher response rate, we considered it appropriate for an exploratory survey.) Survey questions are listed in 3. Free-response survey questions were analyzed using affinity diagramming, as described in (Beyer and Holtzblatt 1997). This technique, in which all main points from responses are printed out on post-it notes and grouped iteratively according to main themes, allowed us to find higher-level themes that emerged from the data in a bottom-up manner.

We intended this as a preliminary study; because we had just been studying users' public tweets, hearing from them directly would be helpful. However, we also realized that

1	What is your Twitter username?
2	Did you know that you've posted geotagged tweets in 2014? (answers: yes, no)
3	Describe the first time you geotagged a tweet. What caused you to decide to add your location?
4	Do you still geotag your tweets? (answers: Yes, always; Yes, sometimes; No; I'm not sure)
5	If you currently geotag your tweets sometimes, describe a recent tweet that you decided to geotag.
6	If you currently geotag your tweets sometimes, describe a recent tweet that you specifically decided NOT to geotag.
7	Are you worried about privacy implications of geotagging your tweets? (Answers: yes, very worried; yes, slightly worried; no, not very worried; no, not worried at all)
8	Why or why not?
9	Which Twitter client do you use most often?
10	Did this survey cause you to change your choices about geotagging?

Table 3: Questions in Study 2. All responses without answer choices given were free-response.

only studying on one platform limited our results. While we started from the most prolific users and recruited down the list, this reflects a diverse array of active users: our users had between 57 and 2766 tweets over the course of our one-year time period (median=293).

Study 2 Results

Why they geotag

The most popular reasons people gave for geotagging their tweets were to communicate with and to show off their travels and events to followers. Of our 78 participants, 17 people described geotagging their tweets at an event, 9 described geotagging while traveling, and 18 described a more general desire "To show my followers where I am." This latter set of users described choosing whether to geotag each tweet, rather than simply leaving it on. This diversity of reasons inspired us to look deeper into reasons behind geotagging, which we do in Study 3.

Some users did not know they were geotagging

Surprisingly, nine participants reported being unaware that they were posting geotagged tweets, while six more reported accidentally turning it on at some point and then consciously deciding to leave it on. Four were persuaded to start geotagging by an app and 10 decided to start geotagging on a whim or out of curiosity.

One major reason that people may be unaware of their geotagging is the presence of third-party apps that post geotagged tweets. Two users mentioned that they cross-post geotagged Instagram photos to Twitter, while a third uses a Wordpress plugin that cross-posts blog posts. This user was surprised to learn that she had been geotagging at all. Third-party Twitter clients are also possible causes: one person who used

the Tweetbot client (an app for reading and posting tweets) mentioned that Tweetbot enables geotagging by default.

Some did not know their geotags' precision

Most participants expressed some concerns about privacy, including vague feelings that they did not want to tell the world where they were (15 participants), or that they specifically did not want the world to know when they were not at home (8 participants). Several explicitly mentioned potential burglaries. These concerns echo previous location privacy findings (Tsai et al. 2010).

However, 20 participants expressed very little concern about privacy. Worryingly, 12 of these participants expressed belief they were only sharing broad city-level locations, and thought that nobody knew their exact location. However, everyone in our data set, including these 12, had posted public tweets with precise coordinate geotags.

Upon further investigation, we found that the Twitter mobile app showed a confusing user interface: it appeared that users would be posting high-level tags (like "Pittsburgh, PA") when instead the actual latitude-longitude point was stored with the tweet. As noted earlier, the Twitter mobile app's user interface has since been changed to use placetags.

Study 3: Cross-platform geotagger survey

While the survey on Twitter users raised some new interesting questions, it did not fully answer the question of why users geotag. In addition, it focused only on Twitter users. We wanted to increase our sample, as well as broaden it to include other social media users.

Our primary research question for this survey was "Why do people geotag?" Having read many papers about why this topic in location-based social networks (Lindqvist et al. 2011; Guha and Birnholtz 2013; Lingel, Naaman, and Boyd 2014; Cramer, Rost, and Holmquist 2011), we wanted to see if their findings about why people check in can be replicated in geotags in non-location-based social networks. As such, we asked them to pull up their most recent geotagged post in any social media and asked them, "Please explain in detail why you geotagged," with a free response answer. We also asked, more generally, "What are your motivations for geotagging? (check all that apply)" and offered the choices in table 4.

We recruited 406 people from Amazon Mechanical Turk to ask them about their geotagging practices and their most recent geotag on their most used social media service. We recruited participants who were in the United States and had previously geotagged at least once. We asked about the content of their most recent post with geotag: what type of post it was (text, picture, etc), what it was about, why they posted it, and how meaningful it was. We asked about the geotag: what the geotag was, how precise it was, how they would describe the place they geotagged, why they geotagged, what their motivations for geotagging are, how often they go to the place, how far it is from where they live, and how long they waited to post it. The study took about 5 minutes and participants were paid $1. Six people's responses had to be removed as their free responses indicated they were not paying attention, leaving us with 400 valid responses.

Study 3 Results

For reasons of space and conciseness, we will not report on all of the survey questions, instead highlighting some of the most surprising and relevant findings here.

When do they geotag?

Most of our participants reported geotagging in the moment, mostly within an hour of the photo or event they are sharing. If not, they usually geotag by the end of the day: 69.0% geotag within an hour, 89.3% within a day (see table 5 for details). However, a substantial amount waited, for reasons like wanting to 'settle down (at hotel)' , 'find phone service signal', or posting upon friends' request.

This suggests a slight difference between geotags and check-ins, due to the concept of "check-in transience" introduced by Guha and Birnholtz (2013). Check-ins have a short lifespan; it would rarely make sense to check in later in the day or week. However, about 1/3 of our respondents waited over an hour to share their geotagged post.

Where do they geotag?

Our participants mostly geotagged in places far from home, as shown in Figure 5. When we asked about the most recent geotag, only 11.9% of them were in the users' home or neighborhood, and 46.7% were in their home city. Johnson et al (Johnson et al. 2016) found similar findings: that in Twitter, Flickr, and Swarm, only about 75% of posts are from "local" users, but ours are more extreme; depending on how one defines "local," anywhere between 11.9% to 46.7% of geotags are from "local" users. Furthermore, 70.0% of these geotags are from rare places: places they go every year, a few times so far, or this is the first time. This confirms the finding in (Lindqvist et al. 2011) that many people are reluctant to check in at routine places, and the finding in (Cramer, Rost, and Holmquist 2011) that users avoid checking-in to home and work because it can be annoying.

Why do they geotag?

We asked users twice why they geotag: once in general (responses are shown in Figure 6) and once about specifically why they geotagged their last post. The options shown in Figure 6 are based on previous research, as explained in Table 4. We can see that few people chose "other", which suggests that these choices explained people's preferences well. In addition, few people geotag automatically, which corroborates our finding in Study 1. The sharp difference in magnitude between the top two reasons and the rest adds some nuance to our knowledge: the most commonly cited reasons for geotagging are the social-driven ones ("show I was at a cool place" and "keep family/friends updated"), more than the purpose-driven ones.

We analyzed the free response question about why they geotagged their last post using affinity diagramming (as described in (Beyer and Holtzblatt 1997), as we did in study 2.) The categories that emerged almost all fit into one of the six choices shown in Figure 6. The two categories that didn't fit were "No reason" and "Application-driven"; we give some examples in Table 6. We note that these, too, have

Motivation	Theoretical basis
To show that I was at a cool, amazing, special, or popular place	Common in (Lingel, Naaman, and Boyd 2014; Guha and Birnholtz 2013); handles "Impression management" theme without introducing jargon into the survey
To keep track of this place for later on	One of the 5 factors in (Lindqvist et al. 2011)
To promote this place to my social network	Related to the "Place discovery" factor in (Lindqvist et al. 2011); related to self-presentation discussed in (Guha and Birnholtz 2013; Cramer, Rost, and Holmquist 2011)
To coordinate with my friends for activities	Part of the "Social connection" factor in (Lindqvist et al. 2011)
To keep friends/family updated on what I'm up to and where I'm at	Part of the "Social connection" factor in (Lindqvist et al. 2011)
It's automatic/I always have geotag on	This does not appear in LBSN research because it doesn't apply to LBSNs, where checking-in is the app's purpose.
Other: _____ (free response)	

Table 4: Answer choices for "What are your motivations for geotagging?" in our Study 3. Participants could check all options that applied to them. (Only the "Motivation" column was shown to them.)

Time passed	Percent of respondents
Happened at the same time	40.75%
Within an hour	28.75%
Within a day	20.5%
Within a week	6.75%
Within a month	2.0%
More than a month	1.25%

Table 5: Responses to "How long did you wait between having the experience and posting?" in Study 3.

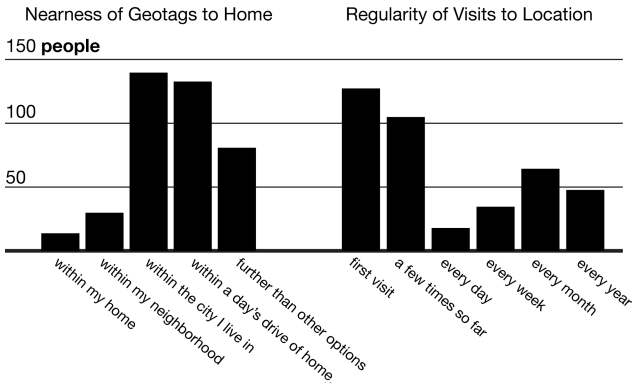


Figure 5: Where participants geotag. Note that most people geotag far from their home: only 11.9% tag in their home neighborhood, and less than half are in their home city. Most people also geotag in places that they visit rarely: 70% of these geotags happen in places that people go annually or less frequently.

No Reason
"No specific reason, just because I wanted to share the dealership where I was buying the car."
"No particular reason other than I tend to enjoy seeing the geotags of my friends so I decided to include it myself."
"Just because I wanted to, no reason in particular, just something to do."
"I didn't really think of it, I just kind of did it...doesn't that explain most of what is on social media? Just because..."
Application-driven
"the geotag put the houston dynamo colors and logo over the picture"
"to get a discount off my melon boba tea."
"To remind myself where I found the pokemon, as well as letting my friends know that it could be a good spot for the future."
"...it's how I find interesting people to follow and I'm sure its how I have gained followers too [on Instagram]"

Table 6: Participants' given motivations for geotagging that fell outside the ones in Table 4

some precedent in the literature; users describe "checking-in" as something to do when bored and are motivated by game elements included in LBSNs (Lindqvist et al. 2011).

Discussion

Geotags are postcards, not ticket stubs

Our studies brought us some unsurprising findings. People geotag to show where they've been; keep their family and friends updated on their travels; record a place for later; or help family, friends, and strangers to find a place. People geotag at rare places more than routine places.

However, our studies add some depth and nuance to our understanding of geotagging. Showing off where they were and keeping family and friends updated rank as the most impor-

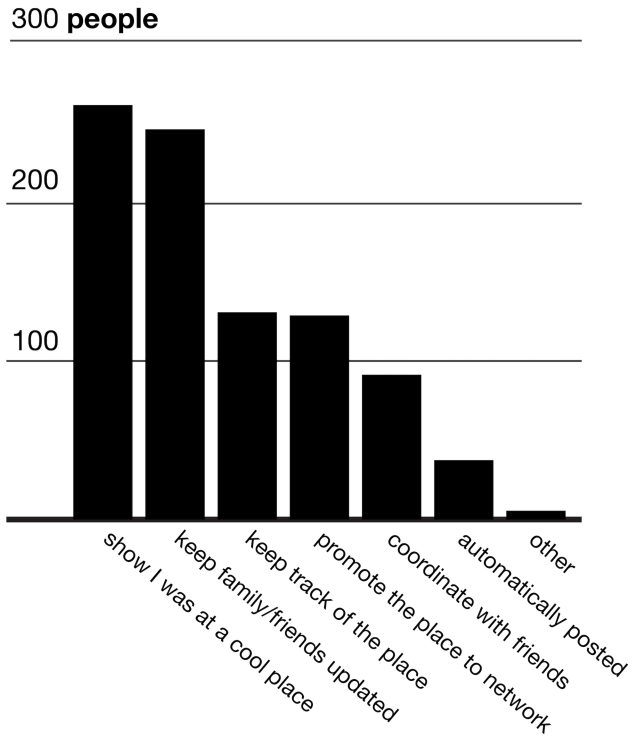


Figure 6: Stated motivations for geotagging. Participants (n=400) could choose multiple options.

tant reasons to geotag. In addition, we document for the first time that most people geotag consciously; they do not simply set their phone to automatically geotag everything, as we saw in Studies 1 and 3. This was not an issue in LBSN research, because using an app like Foursquare without geotagging would not even make sense. But in non-location based social networks, it is important to know whether a geotag is a side effect of another action, like a ticket stub, or a consciously chosen artifact, like a postcard.

Our research shows that a geotag can be seen as a postcard: it shows that a person is at a certain place, it is usually used for social communication, and it is hard (though not impossible) to fake.

In addition, our studies uncovered three new important points:

- The landscape of geotagging is changing, from coordinate geotags to placetags
- Some people may not know that they're geotagging, or how precisely they're geotagging
- New types of spammers are becoming prevalent in public geotagged data

In the rest of this section, we will discuss the implications of these findings for different groups.

Implications for applications and app developers

There are clearly pitfalls to avoid when designing an application that involves users' locations. Many papers have documented the risks, e.g. (Tsai et al. 2010). In this paper, we documented two more risks: the possibility that users are coordinate geotagging when they think they are placetagging (as in Twitter before April 2015), and third-party apps that add geotags with people's knowledge. In this section, we propose ways to avoid these harms and improve users' geotagging experience.

Minimize automatic tags and coordinate geotags To avoid accidental privacy leaks, social media software needs to be more careful when automatically geotagging participants' posts. For some apps, such as the old Foursquare check-in app (now Swarm), geotagging is the main purpose of the app, and so it makes sense to have automatic geotags. However, in Study 2, we also reported on two examples where people's mental models and expectations about geotagging did not match reality.

This point is obvious and straightforward. Less obviously, social media software might consider whether they want geotagging to be a sticky setting at all. That is, if a person geotags one post, should the next post be geotagged by default? Our results suggest no: people largely prefer to make a conscious choice about whether to geotag each post or not. This choice adds only slight overhead, and prevents potentially disastrous privacy leaks.

Another option to reduce privacy risks is to use placetags instead of coordinate geotags. All of our participants' main use cases could usually be handled by placetags, and coordinate geotags are often too precise, revealing more of a user's location than they want. Only the "coordinate with friends" and "keep track of this place for later" cases might require a coordinate geotag, and then only if it's in a wilderness area or other place without well-defined places.

Using placetags could improve the user experience in other ways as well, as coordinate geotags are usually hard for people to understand. Many services, like Twitter and Facebook, are already doing this well: users can select which granularity to placetag, whether it's the building or city that they are in. They can also add their exact location if they want. A minor challenge is that coordinate tags can be generated solely on a smartphone through GPS, whereas placetags require network services and a large database that needs to be kept up to date to do lookups of place names.

Help researchers understand placetags One notable downside of placetags, however, is their interpretability. If a researcher sees a placetag that says "Starbucks" without any finer grained information, how can they know which Starbucks location the user is at? Also, sometimes placetags represent coarse places, like "Singapore," but researchers interpret them as being finer-grained points. They sometimes transform a bounding box into one point at the center, which can have annoying or even disastrous consequences, as Shamma documents (2016). Services need to return geotags at different granularities: cities, polygons, or points. They also need to document why they are returning the granularity that they are: because the user chose it, because the user's GPS could not get an accurate reading, because of the user's default privacy settings, or whatever other reason.

While it would be computationally expensive to send an

entire polygon with every social media point, it would be feasible to publish a gazetteer of places along with a social media API. Services may be tempted to send bounding boxes instead, but this could lead to other problems, such as Mapzen accidentally declaring Copenhagen part of Sweden (Simioni, Oram, and Cope 2016).

Importantly, this is not an appeal to altruism; researchers are internal as well as external. Improving the comprehensibility of placetags for researchers will help a company's own analysts as well as the academic community.

Implications for researchers

Many studies have treated geotag data as sensor data, without much regard for how it came to be. For some use cases, this is sufficient, but in other situations, more care is needed in drawing conclusions from geotagged social media data. Below, we discuss some salient issues for researchers in using geotagged data.

Availability of Coordinate Geotags is Decreasing Publicly visible geotagged social media may seem like an endless source of rich data. However, as we saw in Study 1, Twitter has fewer coordinate geotags available, and even Flickr geotags may be plateauing. Additionally, our participants in Study 3 geotagged primarily on Facebook, Instagram, and Foursquare, but these services do not publish public "firehose" APIs for gathering data.

Furthermore, the geotags that are still present are getting stranger: job posting bots, weather and sports bots, deleted accounts, and other accounts are creating a growing fraction of all public geotagged tweets.

As a result, it is not clear how much more research can be done with coordinate geotags. In addition, for the coordinate geotags that are available, it is important to dig in and filter out whatever spam may be prevalent.

An alternative is that researchers may need to become more comfortable dealing with placetags. For example, it is important to avoid the center-of-rectangle problem mentioned in the previous section. The semantics of a placetag may also vary by application. A few general principles include realizing that the data available at the building or neighborhood level will likely be much smaller than the data at the city level, and the data at the point level smaller still.

Geotag provenance affects research methods It is important to treat geotagging as a performative act, not a passive one. As Study 2 showed, most people consciously decide to geotag each time; as Study 3 showed, people use geotags mostly to show off where they have been, keep family and friends updated, and occasionally coordinate with friends or save a place for later. The content of their tags often reflects vacations or meals at restaurants. As a result, it is a rich data set to study where people go on vacation, eat out, or have places that they want to save for later. However, it does not seem to be a rich data set to study users' everyday lives.

As an example of a use case where the provenance of these tweets matters, we point to (Tasse, Sciuto, and Hong 2016). In this work, the authors tried to find users' home addresses given a sample of their geotagged tweets. If tweets represented a random sample of places the user has been, this would be trivial, because most tweets would occur at users' homes. However, they found that this was impossible for about 15% of Twitter users, because they recently moved, never tweeted at home, or had other complicated use cases. They reference (Krumm 2007), who previously attempted to find home addresses based on GPS sensors on cars. Naturally, with the same methods, the GPS on the car worked much better, because those readings come from a passive and automatic sensor, not a performative act.

But the fact that geotags come from unusual occasions doesn't only limit research; it can add extra context to a post. For example, many geotagged tweets come from Untappd, an app for beer aficionados to share their experiences of fine beers. While Untappd tweets tell us less about the general public's day to day movement, they tell us more about local beer loving communities and, potentially, the sociability of a place. Likewise, the fact that people often geotag on vacation or while out to dinner may provide clues to activity recognition and computer vision algorithms. We encourage researchers to find questions that can take advantage of the rich variety of sources that geotagged tweets provide.

Conclusion

While checking in in location-based social networks has become a widely researched topic, motivations behind geotagging in other social media have not been as fully analyzed. We investigated if people geotagged their social media posts for the same reasons that they checked in on LBSNs, and for the most part found that they do. Geotags are postcards, not ticket stubs; conscious choices, not byproducts. People geotag their social media posts to show off where they are or communicate with family and friends. They geotag in faraway and rare places. However, in our research we found reasons, especially recently, why researchers and developers must be careful. Researchers should be aware that counts of coordinate geotags are shrinking and specific types of spam are rising, while developers should show their users clearly what they are posting, avoid sticky geotagging settings, and prioritize placetags over coordinates. These changes will help us maintain the valuable resource of geotagged post and conduct research wisely and responsibly.

References

Ahern, S.; Eckles, D.; Good, N.; King, S.; Naaman, M.; and Nair, R. 2007. Over-Exposed? Privacy Patterns and Considerations in Online and Mobile Photo Sharing. *CHI '07 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* 357–366.

Ames, M., and Naaman, M. 2007. Why we tag: motivations for annotation in mobile and online media. *Proceedings of the SIGCHI conference on Human factors in computing systems* 971–980.

Beyer, H., and Holtzblatt, K. 1997. *Contextual design: defining customer-centered systems.*

Consolvo, S.; Smith, I. E.; Matthews, T.; LaMarca, A.; Tabert, J.; and Powledge, P. 2005. Location Disclosure to Social Relations: Why, When, & What People Want to Share. *CHI*

2005 Conference on Human Factors in Computing Systems 81–90.

Cramer, H.; Rost, M.; and Holmquist, L. E. 2011. Performing a Check-in: Emerging Practices, Norms and 'Conflicts' in Location-Sharing Using Foursquare. In *MobileHCI*.

Crandall, D. J.; Backstrom, L.; Cosley, D.; Suri, S.; Huttenlocher, D.; and Kleinberg, J. 2010. Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences of the United States of America* 107(52):22436–41.

Cranshaw, J.; Schwartz, R.; Hong, J. I.; and Sadeh, N. 2012. The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. *ICWSM*.

Goffman, E. 1959. *The Presentation of Self In Everyday Life*.

Guha, S., and Birnholtz, J. 2013. Can You See Me Now? Location, Visibility and the Management of Impressions on foursquare. In *MobileHCI*, 1–10.

Jiang, L.; Cao, L.; Kalantidis, Y.; Farfade, S.; Tang, J.; and Hauptmann, A. G. 2017. Delving Deep into Personal Photo and Video Search. *WSDM*.

Johnson, I. L.; Sengupta, S.; Schöning, J.; and Hecht, B. 2016. The Geography and Importance of Localness in Geotagged Social Media. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.

Jurgens, D.; Mccorriston, J.; and Ruths, D. 2015. Geolocation Prediction in Twitter Using Social Networks: A Critical Analysis and Review of Current Practice. In *ICWSM*.

Kafsi, M.; Cramer, H.; Thomee, B.; and Shamma, D. a. 2015. Describing and Understanding Neighborhood Characteristics through Online Social Media. In *WWW*.

Krumm, J. 2007. Inference Attacks on Location Tracks. *Pervasive Computing* 10(Pervasive):127–143.

Le Falher, G. 2014. *Finding Similar Neighborhoods Across Cities by Mining Human Urban Activity*. Ph.D. Dissertation.

Leetaru, K. H.; Wang, S.; Cao, G.; Padmanabhan, A.; and Shook, E. 2013. Mapping the Global Twitter Heartbeat: The Geography of Twitter. *First Monday* 18(5):1–12.

Lindqvist, J.; Cranshaw, J.; Wiese, J.; Hong, J.; and Zimmerman, J. 2011. I'm the Mayor of My House: Examining Why People Use foursquare - a Social-Driven Location Sharing Application. In *CHI*.

Lingel, J.; Naaman, M.; and Boyd, D. 2014. City, self, network: transnational migrants and online identity work. *Cscw* 1502–1510.

Mahmud, J.; Nichols, J.; and Drews, C. 2012. Where Is This Tweet From? Inferring Home Locations of Twitter Users. In *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*, 511–514.

Mohammady, E., and Culotta, A. 2014. Using County Demographics to Infer Attributes of Twitter Users. In *Proceedings of the Joint Workshop on Social Dynamics and Personal Attributes in Social Media*, 7–16.

Nov, O., and Ye, C. 2010. Why do people tag?: motivations for photo tagging. *Communications of the ACM* 53:128–131.

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

Priedhorsky, R.; Culotta, A.; and Valle, S. Y. D. 2014. Inferring the Origin Locations of Tweets with Quantitative Confidence. In *CSCW*, 1523–1536.

Sadilek, A.; Krumm, J.; and Horvitz, E. 2013. Crowdphysics: Planned and Opportunistic Crowdsourcing for Physical Tasks. *AAAI* 536–545.

Schwartz, H. A.; Eichstaedt, J. C.; Kern, M. L.; Dziurzynski, L.; Agrawal, M.; Park, G. J.; Lakshmikanth, S. K.; Jha, S.; Seligman, M. E. P.; Ungar, L.; and Lucas, R. E. 2013. Characterizing geographic variation in well-being using tweets. In *AAAI*, number June 2009, 583–591.

Shamma, D. A. 2016. The Social Concerns of Geo-Located Rectangles. Medium. https://medium.com/@ayman/the-social-concerns-of-geo-located-rectangles-9b361f34811d.

Simioni, J.; Oram, J.; and Cope, A. S. 2016. The Assault on Copenhagen. Mapzen Blog. https://mapzen.com/blog/assult-on-copenhagen/.

Sloan, L., and Morgan, J. 2015. Who tweets with their location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on twitter. *PLoS ONE* 10(11):1–15.

Tang, K. P.; Lin, J.; Hong, J. I.; Siewiorek, D. P.; and Sadeh, N. 2010. Rethinking location sharing: exploring the implications of social-driven vs. purpose-driven location sharing. In *CHI*, volume 12, 85–94.

Tasse, D.; Sciuto, A.; and Hong, J. I. 2016. Our House, in the Middle of Our Tweets. In *Proceedings of the 10th International AAAI Conference on Web and Social Media (ICWSM)*, 691–694.

Thomee, B.; Shamma, D. A.; Friedland, G.; Elizalde, B.; Ni, K.; Poland, D.; Borth, D.; and Li, L.-J. 2016. YFCC100M: The New Data in Multimedia Research. *Communications of the ACM* 59(2):64–73.

Tsai, J. Y.; Kelley, P. G.; Cranor, L. F.; and Sadeh, N. 2010. Location-Sharing Technologies: Privacy Risks and Controls. *A Journal of Law and Policy for the Information Society* 6:119–151.

Wu, A., and Zhang, X. 2011. Temporal sensitivity for location disclosure through mobile photo-sharing. *Proceedings of the 1st international workshop on Mobile location-based service - MLBS '11* 67.

Zhang, A. X.; Noulas, A.; Scellato, S.; and Mascolo, C. 2013. Hoodsquare: Modeling and Recommending Neighborhoods in Location-based Social Networks. In *SocialCom*, 1–15.

Zhang, Y.; Li, B.; and Hong, J. 2016. Understanding User Economic Behavior in the City Using Large-scale Geotagged and Crowdsourced Data. In *WWW*, 205–214.