“Will Check-in for Badges”: Understanding Bias and Misbehavior on Location-based Social Networks

Gang Wang[†], **Sarita Y. Schoenebeck**[‡], **Haitao Zheng**[†] **and Ben Y. Zhao**[†]
[†]Department of Computer Science, UC Santa Barbara
[‡]School of Information, University of Michigan
{*gangw, htzheng, ravenben*}*@cs.ucsb.edu, yardi@umich.edu*

Abstract

Social computing researchers are using data from location-based social networks (LBSN), *e.g.*, “Check-in” traces, as approximations of human movement. Recent work has questioned the validity of this approach, showing large discrepancies between check-in data and actual user mobility. To further validate and understand such discrepancies, we perform a crowdsourced study of Foursquare users that seeks to a) quantify bias and misrepresentation in check-in datasets and the impact of self-selection in prior studies, and b) understand the motivations behind misrepresentation of check-ins, and the potential impact of any system changes designed to curtail such misbehavior. Our results confirm the presence of significant misrepresentation of location check-ins on Foursquare. They also show that while “extraneous” check-ins are motivated by external rewards provided by the system, “missing” check-ins are motivated by personal concerns such as location privacy. Finally, we discuss the broader implications of our findings to the use of check-in datasets in future research on human mobility.

Introduction

Social computing research has increasingly turned to location-based techniques to understand how people move around, interact with one another, and interact with their environments. Capturing traces of human movement has the potential to provide deep insights into human behavior, infrastructures like cities and transportation, and computing applications that support human needs. However, capturing human traces is difficult because of lack of access to proprietary data, privacy concerns about accessing such data, and possible errors in data collection or reporting.

To date, researchers have relied on mobility models (Johnson and Maltz 1996; Jardosh et al. 2003), geotagged data via IP addresses (*e.g.*, Wikipedia), or user check-in data available via APIs (*e.g.*, Foursquare) (Cheng et al. 2011; Noulas and others 2011; Sen and others 2015; Hecht and Stephens 2014). Check-in data is attractive because it is relatively easy to access via APIs or scraping and captures a popular user practice, thus offering relatively high penetration rates. For this reason, a growing body of research has relied on such check-in data to predict human movement (Cho, Myers, and Leskovec 2011; Scellato and others 2011), infer friendships based on visited locations (Allamanis, Scellato, and Mascolo 2012; Scellato, Noulas, and Mascolo 2011), and improve content delivery networks (Scellato et al. 2011).

On the other hand, recent studies have raised questions of validity in users’ check-in activities in location based social networks (Lindqvist and others 2011; Cramer, Rost, and Holmquist 2011; Zhang and others 2013). More specifically, results presented at the Hotnets Workshop raised questions of how valid and representative LBSN datasets are when used as traces of user mobility (Zhang and others 2013). Using a smartphone app, the authors directly compared the GPS locations of a group of participants against their Foursquare check-in activity, and found significant discrepancies. Not only do check-in events only capture a small subset (∼10%) of real locations visited by each user, but nearly 75% of all Foursquare check-ins were found to be events that did not match real mobility.

These significant observations of location misrepresentation prompt further studies to revisit/validate the issue, in order to better understand the sources of bias and misbehavior in user check-in events. In this paper, we do so by performing a crowdsourced user study on behavior in the Foursquare LBSN.[1] We are interested in three key questions. First, is misrepresentation of user locations as common as described in the initial study (Zhang and others 2013), and what role did self-selection bias play in those early results? Second, what are the primary incentives driving users to misrepresent their actual location? Finally, what, if any steps can be taken by Foursquare to curtail these activities, and what are the likely responses to these efforts by current users?

Our study includes two components, a survey study to understand the misrepresented check-in behavior, and a data-driven analysis on false check-ins across two different datasets. First, we conduct a crowdsourced study of Foursquare users, focusing on their check-in behavior and possible misrepresentation of their actual locations. Our survey asks users about their own first-hand creation of false check-ins, and also about secondary observations of such behavior by other users. We ask users to explain their motivation for misrepresenting location in check-ins, and likely responses to potential changes to curtail false check-ins.

[1]Our user study was reviewed and approved by the local IRB.

Second, we obtain the measurement dataset from an earlier Foursquare smartphone study (Zhang and others 2013), and characterize the statistical properties of misrepresented check-ins in Foursquare. Our analysis highlights the correlation between different rewards and misrepresentative behavior. In addition, we extend our user study to participants in the original study. By comparing the results with our crowdsourced user study, we seek to understand how the self-selection bias in the original dataset impacted the rate of false check-in events.

We draw several conclusions from our results. First, relative to the highly active users in the dataset of the initial study (Zhang and others 2013), our broader sample of crowdsourced users report lower levels of misrepresented check-in events. Second, there are still significant levels of misrepresentation both observed and self-reported by users in our broader study. This further confirms the presence of discrepancies between users' check-ins and actual mobility. Third, our survey shows that missing check-ins are generally due to personal decisions regarding location privacy, sharing visits of uninteresting locales, and general forgetfulness to check-in. We also find that extraneous check-ins are most heavily motivated by external rewards including badges, mayorships and financial rewards. Finally, results show that most users believe extraneous check-ins can be reduced by modifying incentives, and that detecting and banning users will not negatively impact their own engagement.

Background: LBSNs and Foursquare

Location-based Social Networks (LBSNs). Today's LBSNs allow users to share social activities along with their locations, *e.g.*, "checking in" to nearby Points-of-Interest (POI). The most popular LBSN is Foursquare. First launched in 2009, Foursquare has gathered over 55 million registered users and 7 billion "check-ins" (as of May 2015). Other related sites include US-based Yelp, Facebook Places, Gowalla (defunct in 2012), and China-based JiePang.

LBSNs incentivize user check-ins using both virtual and financial rewards. For example, the Foursquare user who checks-in to a location most frequently over a 60-day window is designated the "Mayor." In addition, users obtain "badges" for achieving certain requirements, *e.g.*, checked-in to five different coffeeshops. Finally, Foursquare features commercial tie-ins with businesses/stores offering discounts and coupons based on check-ins and mayorships.

Foursquare and Swarm. In August 2014, Foursquare relaunched itself with dramatic changes. The Foursquare app was split into two: a new "Swarm" app took over the social networking and location sharing functionality, and the Foursquare app was revamped to focus entirely on venue search and recommendations. The reward mechanisms were changed too. First, "badges" were replaced by "stickers" in Swarm. Users can unlock new stickers by checking-in to different venues. Second, Swarm modified its "mayorship" mechanism so that users compete only within their friend circles, rather than against all other users in the service. However, following significant drops in downloads, traffic and ranking (Hu 2014; VB News 2014), Swarm restored the old (global) mayorship system in June 2015. Our paper focuses primarily on the original Foursquare app, but we briefly discuss the implications of our findings to the new system in our limitations section.

Related Work

Human Mobility and LBSN Check-ins. LBSNs provide a unique source to collect detailed and large-scale human mobility traces, which have been widely used in human mobility research. For instance, researchers have studied the spatial and temporal mobility characteristics using check-in datasets (Noulas and others 2012; Cheng et al. 2011; Noulas and others 2011). Others use check-in traces to build various applications, such as predicting human movements (Cho, Myers, and Leskovec 2011; Scellato and others 2011), inferring friendship (Scellato, Noulas, and Mascolo 2011), predicting customer volume (Georgiev, Noulas, and Mascolo 2014), measuring urban socioeconomic deprivation (Venerandi and others 2015), and even improving the efficiency of content delivery networks (Scellato et al. 2011).

However, recent studies have expressed concerns about the *representativeness* of check-in datasets to model real-world events. A recent study shows that the number of check-ins at a venue (*e.g.*, an airport) and its actual visitors (*e.g.*, airport passengers) can differ by orders of magnitude (Rost and others 2013). More importantly, using check-ins to rank venue popularity would give significantly false results. In addition, biases in other kinds of check-in information have been documented in prior work. For example, Foursquare, Twitter and Flickr are shown to have biases towards urban users vs. rural users (Hecht and Stephens 2014).

An initial study quantifies the severity of this problem by collecting parallel physical mobility (GPS) traces and Foursquare check-in traces for the same set of users (Zhang and others 2013). The result shows high discrepancies between user check-ins and their real movements: among 14000+ collected check-ins, 75% do not match any real visits (*extraneous check-ins*), while 90% of actual visited locations are missing from the check-in trace (*missing check-ins*). While these results raise significant concerns on the representativeness of check-in datasets, little is known about what causes the discrepancies, and what are the user motives behind such check-in behaviors.

User Incentives of using LBSNs. Researchers have studied the general user motivations for location sharing on Foursquare. Earlier studies show that location sharing is not only purpose-driven (*e.g.*, keeping track of places), but also social-driven (*e.g.*, self-presentation, making friends) (Lindqvist and others 2011; Patil et al. 2012a; Bilogrevic and others 2015). Many factors are found to affect user decisions on whether to share their locations via check-ins (*e.g.*, social impression, privacy/safety concerns) (Guha and Birnholtz 2013; Cramer, Rost, and Holmquist 2011; Patil et al. 2012b). These results shed light on why certain location check-ins are omitted by users. Our work looks further into user motivations behind falsified check-ins (e.g., extraneous), and explore how they impact the usability of check-in traces to model human movements.

Dataset	# of Users	Avg. # of days	# of Checkins	# of Visits	# of GPS points
Primary	244	14.2	14,297	30,835	2,600,000

Table 1: Statistics of user mobility dataset.

Our Methodology

In this work, we seek to understand why check-in traces from Foursquare are *not representative* of real user mobility patterns, and the underlying processes that introduce such discrepancies. As suggested in the initial work (Zhang and others 2013), users tend to *miss* significant number of check-ins for their visits, while making *extraneous* check-ins to locations that they did not physically visit. We seek to understand what's the user intent behind such behavior; how is such check-in behavior related to the design choices of Foursquare, and what Foursquare (or researchers) can do in order to address this problem. In this study, we approach the above questions with three steps

- First, we characterize missing and extraneous check-ins on Foursquare, using the complete dataset from the initial study (Zhang and others 2013). We focus on quantifying the discrepancies between check-in traces and users' true mobility patterns, and inferring possible causes.
- Second, we conduct a survey with Foursquare users to gain a deeper understanding on their incentives behind missing and extraneous check-ins. Particularly, our data-driven analysis has identified several key hypotheses, and we use this survey to validate them. Further, we want explore how user incentives correlate to their demographics and activity-levels.
- Third, we discuss the broader implications of our findings to Foursquare and human mobility research. In particular, we consider whether and how we should apply check-in traces to study human mobility.

Missing and Extraneous Check-ins

We start our study by summarizing the "representativeness" problem of Foursquare check-in traces, and characterizing results from the empirical measurements. We use the conclusions of this analysis as background and a baseline of comparison for later results.

Foursquare Check-in Trace

Our analysis is based on the measurement dataset in the initial work on misrepresented check-ins in Foursquare (Zhang and others 2013). Researchers in that study developed a dedicated smartphone app and used it to collect two "parallel" traces from the same set of Foursquare users. The dataset contains (1) a per-minute GPS trace of the user's location; (2) a trace of the user's real-time check-in events polled by Foursquare open API. In total, 244 users installed the app between January and July 2013, and produced two parallel traces (Table 1).

- **Check-in Trace** contains 14,297 check-in events. Each event includes a timestamp, the name of a POI, its venue category and GPS coordinates.
- **GPS Trace** contains a sequence of 2,600,000 GPS coordinates, and a list of 30,835 POI visits.

We refer this dataset as the *Primary* dataset. Note that these users are "organic", *i.e.*, Foursquare users who installed the measurement app developed by (Zhang and others 2013) voluntarily from Google Play and Amazon App store. Since the Primary dataset is a "self-selected" user sample, it is likely to oversample certain subsets of users and produce biased results. Later, we address the issue using our survey study.

Comparing Check-ins to GPS Visits. Now we quantify how well Foursquare check-ins represent a user's GPS mobility patterns. The algorithm in the earlier work (Zhang and others 2013) matches up each Foursquare check-in event to a potential GPS visit, based on their special and temporal closeness. Here, a GPS visit is defined as the user staying at one location for longer than some period of time, *e.g.*, 6 minutes. To minimize the impact of measurement noise due to inaccurate GPS reports and of small time offsets between each check-in and the corresponding physical visit, the algorithm applies thresholds to control the precision of data matching. More specifically, a given check-in event matches a potential GPS visit if and only if it matches the GPS data within α meter range and β minute offset. After experimenting with a wide range of α and β values, the matching results are most consistent (*i.e.*, number of matches converged) for values $\alpha = 500m$ and $\beta = 30min$. Note that the matching algorithm is designed to capture an upper limit on possible event matches. Thus both thresholds are designed to be accommodating and increase the probability of matching check-ins to visits.

The check-in trace contains 14297 check-in events and the GPS trace has 30835 visits. Here is the matching result:

- **Honest Check-ins**: 3525 check-ins events match up with GPS visits, which show the user was indeed at the physical location of the check-in. This only represents a very small portion of all the check-ins and GPS visits.
- **Extraneous Check-ins**: 10772 check-in events (75% of total check-ins) do not match any visits in the GPS trace.
- **Missing Check-ins (or Unmatched Visits)**: 27310 GPS visits (89% of all visits) do not match any check-ins.

Note that the definition of honest, extraneous and missing check-ins is result-driven. It is based whether the check-in matches a physical GPS visit, not based on user intent. The discrepancies between the check-in and GPS datasets are non-trivial: users miss significant number of check-ins at visited locations, while generating check-ins to locations that they did not physically visit.

Missing Check-ins

We characterize missing check-ins by analyzing which locations are missing and why. Intuitively, users tend not to check-in at specific places they frequent on a daily basis, *e.g.*, home and work. If this is correct, then a small number of places could account for the large majority of missing check-ins. To validate this, we take each user, compute their top-n most visited POIs, and examine the portion of her missing

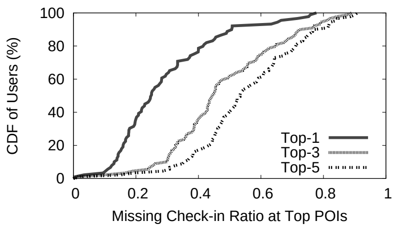


Figure 1: Ratio of missing check-ins at top-k most visited POIs.

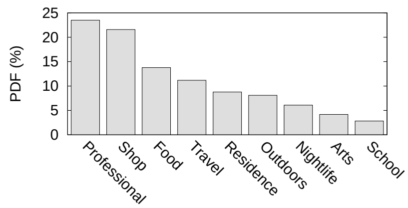


Figure 2: Breakdown of missing check-ins per POI category.

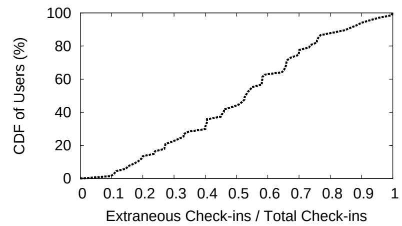


Figure 3: User's ratio of extraneous check-ins.

check-ins attributable to these top POIs. Figure 1 plots the CDF of this ratio across users for their top-5 visits. The results confirm our hypothesis. For roughly 60% of users, 5 locations account for more than half of their missing check-ins. For 20% of users, a single location is responsible for $>$ 40% of the missing check-ins.

We also examine the types of locations for missing check-ins. To do so, we map each missing check-in's GPS to a POI category using *Foursquare Open API*. Figure 2 shows the distribution of missing check-ins locations under 9 Foursquare POI categories. The three categories with the most missing check-ins are *Professional*, *Shop* and *Food*. These are related to people's routine activities: going to work, shopping and meals, and thus are critical parts of a user's mobility pattern.

Extraneous Check-ins

Extraneous check-ins (those without a matching visit) occur when users misrepresent their physical locations. They can be categorized into 3 major types.

- **Superfluous Check-ins**: While visiting one POI, users check-in to multiple nearby POIs. 2176 check-ins (20.2% of all extraneous check-ins) were superfluous.
- **Remote Check-ins**: Check-ins to POIs more than 500 meters away from a user's actual GPS location. 500m is greater than any GPS or POI location error. 5715 check-ins (53.1% of all extraneous check-ins) were remote.
- **Driveby Check-ins**: Users check-in to nearby POIs while moving at moderate or high speeds (speed $>$ 4mph). 1782 check-ins (16.5% of all extraneous check-ins) are driveby.

Extraneous check-ins are highly prevalent among users. As shown in Figure 3, 99% of users have at least one type of the extraneous check-ins; for more than half of the users, extraneous check-ins account for 50% of their check-in events.

Inferring Potential Causes. To better understand potential causes for the misbehavior, we measure correlation between a user's extraneous check-ins and profile features. Here we list the Pearson's correlation score r between features and ratios of check-ins (superfluous, remote, driveby and honest). Here -1 means perfect negative correlation and 1 means perfect positive correlation.

The top correlating feature for superfluous check-in is the user's number of mayorships ($r = 0.34$, $p = 0.02$); remote check-in is most correlated to number of badges ($r = 0.49$, $p = 0.001$); driveby check-in does not have statistically significant correlations with any profile features (with $p > 0.05$). These results show a strong correlation between external rewards (mayorships and badges) and the most obvious types of extraneous check-ins (superfluous and remote check-ins). These account for over 73% of all extraneous check-ins.

Self-selection Bias

One limitation of Primary data is the "self-selected" nature of the user population. Self-selection bias can arise when some people decide to participate in a study and others do not. In the initial study (Zhang and others 2013), the data was collected from Foursquare users who voluntarily installed the measurement app developed by the researchers. Intuitively, these self-selected users are more likely to be highly active users, and can bias results towards higher rates of misbehavior (compared to average Foursquare users). Later, we use our survey results to assess the impact of the self-selection bias in this dataset.

User Study

We conducted a survey to calibrate possible biases in the Primary dataset, and to explore users' incentives for checking-in. The Primary dataset has shown high discrepancies between users' check-in traces and their real mobility patterns. Here, our first goal is to draw on a broader sample of users to examine self-selection biases in Primary dataset. The second goal is to understand user incentives related to missing and extraneous check-ins. We formulate several key hypotheses and examine them using the survey responses. Finally, we explore possible directions to mitigate check-in noises, and examine how acceptable they are to Foursquare users.

In the current section, we describe our hypotheses and survey design. We present the analysis in the next section.

Hypotheses

We first present 5 hypotheses regarding data biases as well as the user incentives of missing and extraneous check-ins. We derive these hypotheses based on the analytical results in last section, existing literature, as well as our intuition.

H1: *The self-selected dataset is biased towards highly active users.*

The dataset collected from users who *voluntarily* installed the third-party Foursquare app (Zhang and others 2013) is a self-selected sample. Our hypothesis is that users who happened to find and install this app are more likely to be highly active on Foursquare, *i.e.*, users who are passionate about location tracking and check-ins. This "self-selected" user group may have different characteristics than typical Foursquare users.

H2: *Self-selection bias can lead to underestimates of missing check-ins and overestimates of extraneous check-ins.*

This hypothesis is under the condition that *H1* holds: the self-selected dataset is biased towards highly active Foursquare users. Intuitively, highly active users are more likely to check-in. Thus the amount of "missing" check-ins are likely to be underestimated by the dataset. In addition, highly active users are more likely to care about badges and mayorships, and more likely to make extraneous check-ins. Thus the extraneous check-in ratio for active users is likely to be overestimated compared to a less biased sample.

This could impact our conclusions in two ways. First, if a user's missing check-in ratio is actually higher, it makes her check-in traces even *less representative* of her actual mobility patterns. Second, we need to examine whether extraneous check-ins are prevalent across Foursquare users, or merely limited to highly active users.

H3: *Users who neglect to perform check-ins do so either because they simply forget, or because they actively avoid checking-in for personal reasons.*

Users could miss check-ins for a number of reasons. First, users may have privacy concerns at certain locations. For example, some people may avoid checking-in at home to avoid stalkers (Patil et al. 2012b). Second, users may choose to not check in at locations they consider uninteresting. Prior work have shown that users share locations on Foursquare to maintain positive social impression (*e.g.*, to appear cool) (Lindqvist and others 2011; Patil et al. 2012a). Intuitively, places that users visit routinely (*e.g.*, home, work places) are less attractive and check-ins would make the user appear boring to their friends. Finally, it is unrealistic to assume users will check-in on Foursquare at every place they visit. Our intuition is users often miss check-ins because they simply forget.

H4: *Extraneous check-ins are primarily motivated by the reward mechanisms in Foursquare.*

Our analysis shows strong correlations between extraneous check-ins and Foursquare rewards (*i.e.*, badges, mayorships. This makes intuitive sense. Many Foursquare users treat check-in as a game to compete with their friends (Cramer, Rost, and Holmquist 2011), and gaining higher rewards is a natural motivation to cheat in a game. Note that the earlier correlation analysis did not consider Foursquare's financial rewards (*e.g.*, coupons, free drinks) due to lack of data. Our hypothesis is all these rewards are key motivations for extraneous check-ins.

H5: *Removing the reward mechanisms from Foursquare would hurt the user engagement of Foursquare.*

To effectively mitigate extraneous check-ins, one approach is to remove the key incentives, that is, the reward mechanisms (if *H6* is true). However, the gaming and reward features are important parts of users' Foursquare experience (Lindqvist and others 2011; Patil et al. 2012a), which are likely to play a big role in encouraging user engagement. We suspect their absence would hurt Foursquare.

Profile Attribute	Profile Statistics: Mean (STD)		
	Turker (108)	P-Replied (23)	P-NoReply (221)
Checkins	603.5 (91)	1770.8 (2521.2)	1730.4 (2994.2)
Badges	13.8 (16.8)	37.7 (29.2)	28.2 (45.2)
Mayors	0.2 (0.4)	14.3 (26.9)	19.1 (38.4)
Friends	23.1 (39.7)	45.4 (58.0)	41.3 (67.0)

Table 2: Profile statistics for Turkers, Primary users who replied (P-Replied) and didn't reply (P-NoReply) survey.

Profile Attribute	Welch t-tests Results: T-statistics (p value)	
	Turker vs. P-Replied	P-Replied vs. P-NoReply
Checkins	-2.09 (0.046)*	0.07 (0.94)
Badges	-3.70 (0.001)*	1.34 (0.19
Mayors	-2.50 (0.020)*	-0.76 (0.45)
Friends	-1.72 (0.09)	0.31 (0.76)

Table 3: Pair-wise comparisons with Welch two-sample t-tests for Turker versus Primary-Replied, and Primary-Replied versus Primary-NoReply. *$p < 0.05$

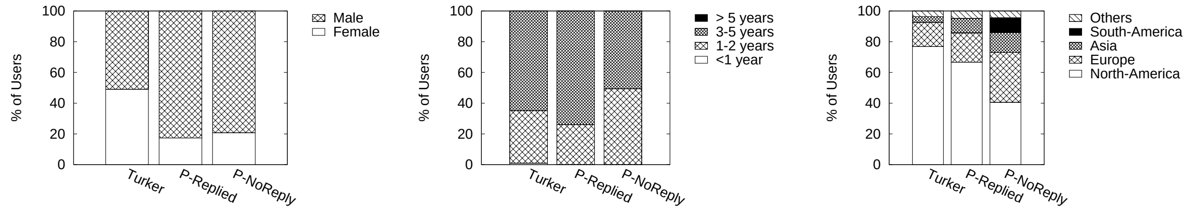
Survey Questions and Participants

To test our hypotheses, we designed a survey containing 12 questions for Foursquare users.[2] At a high-level, Q1–Q4 are about user demographics; Q5–Q7 are about missing check-ins, and Q8–Q12 are about extraneous check-ins. Later, we will discuss each of the questions in detail along with the user responses.

We conducted the survey with two participant groups. The first group was recruited from the *Primary* mobility dataset. In April 2015, we invited each of the 244 Foursquare users to participate in the survey via email. Since we already have their mobility traces (GPS and check-in), we can directly correlate their answers to their behavior. To motivate participation, we offered participants at $1 Amazon gift card for completing the survey and randomly selected one participate in a drawing for an iPad mini. 23 users completed the survey. We note that the survey was only available in English; some of the 244 participants were not native English speakers and were unlikely to participate in the survey.

To collect a larger sample of Foursquare users (for *H1–H2*), we recruited a second group of participants from Amazon Mechanical Turk (Mturk). Mturk is a crowdsourcing platform with workers (known as *Turkers*) who come from relatively diverse backgrounds (Buhrmester, Kwang, and Gosling 2011). Specifically, Turkers are slightly more diverse than other Internet samples and then college student samples, but not as diverse as a representative sample of participants (*e.g.*, Pew Internet data) (Antin and Shaw 2012; Ross and others 2010). Using Turkers who have a variety of activity levels on Foursquare provides a benchmark to compare against the self-selection bias in the Primary dataset

[2]The full question list is available at `https://www.dropbox.com/s/9x3a6yigtttofu9/appendix.pdf`



(a) Gender (b) Foursquare Experience (c) Geographic Origin

Figure 4: Demographics of Turker and Primary respondents, and the rest of Primary users who didn't reply.

of highly active Foursquare users. Considering that Mturk users often provide sloppy or low quality answers in survey studies (Gadiraju et al. 2015), we screen participants using the following criteria. First, participants needed to have a Foursquare account that has been used for at least 6 months. Second, turkers needed a minimum Human Intelligence Task (HIT) approval rate of 90%, and more than 50 approved HITs. This is to prevent non-serious Turkers from participating. Each Turker could only take the survey once, and was rewarded $1.5 (pricing based on the earlier study (Mason and Suri 2012)). During the survey, we also collected the Turker's Foursquare profile data (with their consent). We run the survey in April 2015 and 108 Turkers completed the survey.

Analyzing User Responses

We now analyze the survey responses to test our hypotheses. We first investigate the impact of the sampling bias in the Primary dataset, and analyze how users perceive missing and extraneous check-ins as well as their incentives. Finally, we discuss approaches for mitigating noises in check-in trace.

Data Biases and The Impact

We start with *H1–H2* to examine the self-selection bias in the Primary dataset and the impact. More specifically, we examine the biases by comparing Primary users with Turkers on their profile statistics and demographics (Q1–4). Then we analyze how these biases affect our estimations on the prevalence of extraneous and missing check-ins (Q7–9).

Biases in the Foursquare Profiles. To understand the "self-selection" bias, we first compare the Primary user profiles with Turkers in Table 2. First, Primary users are more active than Turkers in different Foursquare activities. They have 2–3 times more check-ins, badges, mayorships and friends than Turkers, despite they use Foursquare for a similar amount of time (Figure 4(b)). The differences are statistically significant based on Welch t-tests with $p < 0.05$, except for number of friends (Table 3). This confirms *H1*: the customized app did attract highly active Foursquare users, and thereby formed a biased user sample.

Second, within the Primary set there's no clear difference between those who replied our survey (23 users) and those who didn't (221 users). Table 2 shows these two groups have very similar profile statistics. The results of Welch two-sample t-tests (Table 3) show that the differences between

User Group	Extraneous Check-in ratio: Mean (STD)		
	Remote	Superfluous	Driveby
Never/Sometimes	0.09 (0.15)	0.06 (0.12)	0.04 (0.06)
Often/Always	0.62 (0.0)	0.15 (0.15)	0.07 (0.13)

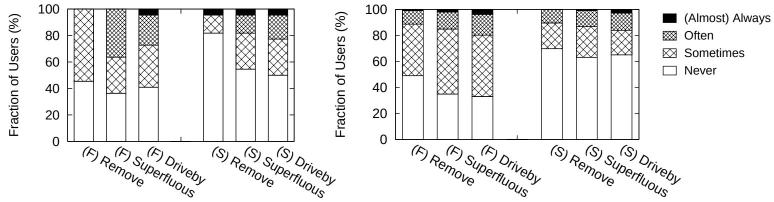
Table 4: Self-reported extraneous check-in versus the actual extraneous check-in ratio based on GPS data. The result shows Primary users who claim "Often" or "Always" do have a higher extraneous check-in ratio than those who claim "Never" or "Sometimes".

Replied and NoReply group are insignificant ($p > 0.05$). Note that getting survey responses from Primary users is also a self-selection process, *i.e.*, it is up to the user to reply or not. This result shows we have a rather representative sample from the Primary set to answer the survey.

User Demographics. The demographics of Primary users offer possible explanations for why users replied (or did not reply) to the survey. As shown in Figure 4(c), the "Replied" group has a much lower coverage in South America, Asia and Europe, where many people do not speak English at all, or as a native language (compared to "NoReply" group). Figure 4(b) shows the vast majority of our participants have been using Foursquare for 1-5 years. "Replied" users have slightly more experience, possibly because earlier adopters were more willing to participate in a study.

Impact to Extraneous Check-ins. Given that Primary datasets are biased towards active users, we next evaluate how this bias impacts the earlier conclusion about extraneous check-ins (*H2*). We investigate whether extraneous check-ins are common behavior across a broader sample of Foursquare users than highly active users.

We ask both Primary users and Turkers how often they perform extraneous check-ins on Foursquare (Q9). Note that asking this type of survey question is subject to social desirability bias: people may mask "dishonorable" behaviors such as lying or cheating (Toma and Hancock 2010). We handle this challenge with three adjustments: First, we set up a baseline question, by asking the participants how often they observe their *friends* performing extraneous check-ins (Q8). Second, we avoid terminology like "superfluous" or "remote check-in" in the questions. Finally, during post-processing, we leverage the 23 Primary respondents' GPS data as ground-truth to estimate how honestly users describe their extraneous check-ins.



(a) Primary

(b) Turker

Figure 5: How often do respondents (claim to) observe extraneous check-ins made by their friends (F) vs. by respondents themselves (S).

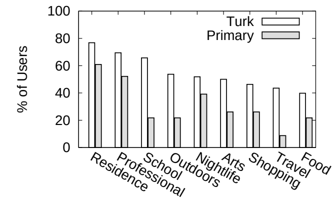


Figure 6: Places where users don't check-in to.

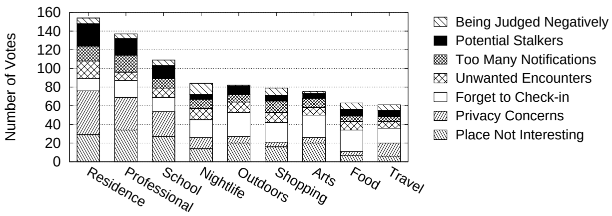


Figure 7: Reasons for not checking-in at certain locations.

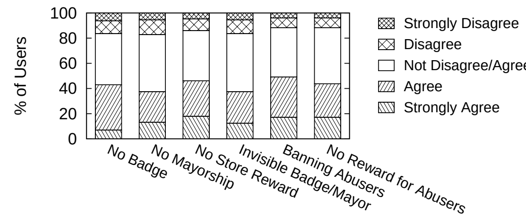


Figure 8: Whether designs help reduce extraneous check-ins.

For ground-truth estimation, we examine whether users' survey answers *correlate positively* with their actual extraneous check-ins identified by GPS data. More specifically, we first group Primary users into two groups based on their answers: "Never"/"Sometimes" versus "Often"/"Always". Then, we compute their average extraneous check-in ratio for each group based on GPS data. As shown in Table 4, the results confirm the positive correlation for all three types of extraneous check-ins. Users who claim "Often"/"Always" do make more extraneous check-ins on Foursquare than those who claim "Never"/"Sometimes". This gives us confidence to the overall quality of Q9's answers.

The final results are shown in Figure 5. About 60% of respondents report that they observe extraneous check-ins from their friends, while about 20%–50% admit making extraneous check-ins themselves. As expected, participants are more likely to describe other people's extraneous check-ins than their own. Our results have two key implications.

First, extraneous check-ins persist in both Primary and Turker users. The 20%–50% of users who self-reporting making extraneous check-ins can serve as a lower-bound, even after we consider the level of dishonesty in the answers. Here we assume respondents only try to hide about extraneous check-ins, but do not claim it if they did not make any. This is still a high number to be a "prevalent" behavior. Even for Turkers, about 40% claim to make extraneous check-ins, suggesting that extraneous check-in is a prevalent behavior across both highly active users and a broader sample of less active users. Note that our result only verifies the prevalence of *users* having engaged in extraneous check-ins. This cannot estimate how prevalent extraneous check-ins are among all the check-in events on Foursquare.

Second, we do observe that a larger portion of Primary users (59%) report to make at least one type of extraneous check-ins than that of Turkers (41%). This supports *H2* that the highly active users are more likely to make extraneous check-ins. The situation is slightly different for each specific type (Figure 5): Primary users report a higher level (50%) of superfluous/driveby check-ins (39% for Turkers), while lower level (20%) of remote check-ins (38% for Turkers).

Impact to Missing Check-ins. Regarding the missing check-ins, our results (Table 2) already show that the Turkers make fewer check-ins than Primary users. This supports *H2* that general Foursquare users would "miss" even more check-ins than estimated by Primary dataset. Therefore check-in traces are unrepresentative for true user mobility.

We also examine this bias in users' responses to Q7 regarding "where" respondents claim to miss check-ins. The results are shown in Figure 6, and the bars in the figure are sorted by the percentage of Turkers under each category. Clearly, Turker's bars are consistently higher than those of the Primary users. Again, we perform Welch two-sample t-tests and find that the difference between Turker and Primary is statistically significant ($t = 3.49$, $p = 0.0032$).This indicates that turkers are more conservative in making check-ins under all listed location categories. Again, the result suggests a higher chance for turkers to miss check-ins, supporting *H2*.

Rationale for Missing & Extraneous Check-ins

Thus far, we have justified that the sampling bias in the Primary dataset does not change our initial conclusions. Now, we analyze the responses from both Turkers and Primary users to understand rationale behind missing check-ins (*H3*) and extraneous check-ins (*H4*).

ID	Motivation	Agreed Turkers	Agreed Primary
1	This place is not interesting	59.2%	34.7%
2	I have privacy concerns	56.4%	39.1%
3	Forget to check-in	47.2%	30.4%
4	Avoid unwanted encounters	36.1%	4.3%
5	Concerned about being stalked	33.3%	0.0%
6	Avoid too many notifications to friends	32.4%	17.3%
7	Avoid being judged negatively	23.1%	8.7%

Table 5: Motivations for missing check-ins, sorted by the % of turkers who agree on each motivation.

Missing Check-ins. We start with the user incentives for missing check-ins (Q6-7). First, Figure 6 shows that for both Turker and Primary users, their top two locations of missing check-ins are *Residence* and *Professional*, representing home and work. This is consistent with the earlier result that users tend to miss check-ins at daily visited locations. Second, Table 5 shows the percentage of Turkers (or Primary users) who agree with each of the listed reasons for missing check-ins. We find that the top 3 reasons are consistent: “this place is not interesting”, “privacy concerns” and “forget to check-in”. This suggests *H3* are generally supported.

A more careful inspection shows that *H3* is supported differently under different location contexts (Figure 7). First, users agree that they miss check-ins because certain locations are not interesting. We notice that the “uninteresting” places cover locations that users daily visit such as home and work. Interestingly, “Arts” and “Outdoors”—places that most people don’t visit routinely—also received high votes for “uninteresting”. This suggests perceptions about what comprises an interesting place are diverse.

Second, “privacy concern” is not only a key reason for missing check-ins, but is also strongly associated with users’ daily visited locations such as home, work, and school. It makes sense that users hesitate to check-in at those locations, since it may violate their privacy preferences without the added benefits of being interesting.

Third, “Forget to check-in” is also among the top reasons for missing check-ins under most POI categories. The exceptions are routinely visited locations (home, work, school), where privacy concerns play a major role. This confirms our intuition that missing check-ins is likely due to simple neglect or disinterest in checking-in at common locations. We believe that missing check-ins are inevitable when active user participation is required. The alternative is passive location tracking, which is “complete,” but any data sharing with the public inevitably produces a loss of privacy without user consent. A later version of the Swarm app used passive user tracking via GPS, but had to limit access to location data to only friends of the user.

Extraneous Check-in Incentives. Next, we examine users’ incentives for making extraneous check-ins (Q10). As shown in Table 6, for both Turkers and Primary users, the top 3 incentives are badges, mayorships, and financial rewards. This supports *H4*, that Foursquare rewards are major incentives for extraneous check-ins. On the other hands, social and gaming related-incentives are less important.

ID	Motivation	Agreed Turkers	Agreed Primary
1	To get extra points to get a badge	29.6%	21.7%
2	To get coupons, discounts or free stuffs	28.7%	13.0%
3	To win the mayorship of those places	22.2%	26.1%
4	To appear cool or interesting	20.3%	8.7%
5	To win a competition among my friends	15.7%	13.0%
6	To impress my friends with my check-ins	14.8%	0.0%
7	To make new friends around those places	10.2%	4.3%

Table 6: Motivations for making extraneous check-ins, sorted by the % of turkers who agree on each motivation.

ID	Motivation	Agreed Turkers	Agreed Primary
1	To record the places that I visited	63.8%	60.8%
2	To tell my friends I like this place	56.4%	30.4%
3	To earn Foursquare badges	51.8%	60.8%
4	To be the mayor of this place	46.2%	52.1%
5	To inform my friends about my location	34.2%	34.7%
6	For coupons, discounts or free stuffs	34.2%	30.4%
7	To appear cool or interesting	32.4%	8.7%
8	Want people/friends to join me here	26.8%	13.0%
9	To share mood or feelings with friends	18.5%	8.7%
10	Hope my geographically distant friends to feel they are part of my daily life	15.7%	8.7%

Table 7: Motivations for regular location check-ins, sorted by the % of turkers who agree on each motivation.

In addition, we investigate whether Foursquare rewards also play a role in incentivizing *regular* non-extraneous check-ins. Table 7 summarizes users’ responses to Q6. Foursquare rewards are still ranked high in the list (top 5 out of 10), indicating their importance to Foursquare users. Interestingly, for regular check-ins, neither Turker nor Primary users vote Foursquare rewards as the top incentive. Turkers in particular, place more value on Foursquare social functionalities, such as “recording visited places” and “recommending this place to friends”.

Rewards and User Engagements

Finally, we explore the ways to mitigate extraneous check-ins and validate *H5*. We describe 6 different approaches, and ask users for their opinions (Q11–12). These approaches include canceling the reward mechanism of badges, mayorships, or financial rewards; making user badges (mayors) invisible to their friends to mitigate competition-motivated extraneous check-ins; or punishing abusers who make extraneous check-ins by banning their accounts or canceling the rewards earned from extraneous check-ins.

As shown in Figure 8, most respondents agree that these approaches could help to mitigate extraneous check-ins, though they show no preferences towards a particular option. The user preferences on the six approaches have insignificant differences ($p>0.05$ for the pairwise Welch t-tests). However, Figure 9 shows that respondents believe different approaches would have different impact on their

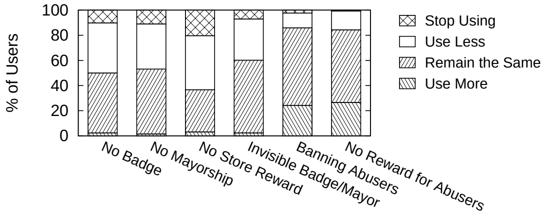


Figure 9: Impact of design choices on user engagement.

usage of Foursquare. Canceling the reward systems would damage user engagement. About 40% of respondents claim they would use Foursquare less and 10%–20% believe they would stop using it if the rewards were gone. This supports *H5*, indicating Foursquare rewards are playing an important role in engaging users. Other alternative approaches look more promising. For example, the last two approaches that punish the responsible users of extraneous check-ins are significantly more acceptable (Welch t-tests $p<0.05$), and could even increase some users' engagement.

We note that Q11 and Q12 both pose *hypothetical* questions. Prior research has shown that people often do poorly at predicting their feelings and behaviors in hypothetical scenarios where they don't have real experiences (Fowler 1995). Recall that our survey participants have been using Foursquare for a long time ($>$3 years). This may reduce some of the problems because related experience and knowledge can reduce prediction biases (Fowler 1995).

Implications and Limitations

Implications. First, compared to highly active users in the Primary dataset, our crowdsourced Foursquare users report relatively lower levels of misrepresented check-in events. However, the level of misrepresentation from crowdsourced users is still quite high in terms of both observed and self-reported results. This suggests that the discrepancies between Foursquare check-ins and actual user mobility persist. Second, results suggest that missing check-ins are caused by reasons including privacy concerns, lack of interest, and the attention required to check-in at every location. Thus missing check-ins seem unavoidable. Extraneous check-ins are primarily motivated by rewards, but simply removing them is a poor solution as it could hurt user engagement.

Our work has direct implications on research efforts that apply check-in data to study human mobility without careful consideration of biases and limitations. For example, Cheng et al. analyze Foursquare check-ins, and use their results to report a strong *periodic* pattern in human movements (Cheng et al. 2011). Later work further leverage these patterns to predict users' future movement (Cho, Myers, and Leskovec 2011; Scellato and others 2011). But given the significant discrepancies between check-ins and real mobility patterns, it is unclear how these inferences are influenced by deviations from actual mobility patterns.

In addition, researchers use check-in traces to validate human mobility theories. Noulas et al. describe a universal law for human mobility based on Foursquare check-ins (Noulas and others 2012), that is, the probability of transiting from one place to another is inversely proportional to Stouffer's distance (Stouffer 1940) rather than physical distance. Again, such specific claims would likely look different if significant portions of the check-in trace were filtered out as extraneous. Finally, studies have shown that correlating check-in events to real visitors often gives misleading results, because the rate of missing check-ins is unevenly distributed between different venues (Rost and others 2013). Despite this, studies continue to use check-in events to estimate real-world visitors and popularity of POIs (Georgiev, Noulas, and Mascolo 2014).

On the positive side, large-scale check-in traces can help studies that focus on the social and contextual aspect of check-in events. Scellato et al. introduce geo-location features to improve link prediction algorithms, which offers better friend recommendations (Scellato, Noulas, and Mascolo 2011). Venerandi et al. leverage the rich contextual information of POIs and check-ins to classify urban socioeconomic deprivation (Venerandi and others 2015).

Our observations echo prior work that surfaces other kinds of biases (*e.g.*, urban versus rural (Hecht and Stephens 2014)) on Foursquare as well as Twitter and Flickr. Together, this body of literature points to the powerful opportunities offered by location-based check-ins for understanding human mobility and behavior, but also the importance of careful scrutiny of biases in the data.

While it is exciting to facilitate new scientific findings using social media data, we believe it is equally important to replicate and validate existing findings. Our work has revisited (and adjusted) prior conclusions on the severity of check-in biases, and point out different aspects of existing research that can be affected by such biases. Further efforts are still needed to understand biases in user-generated traces, and how they change our views towards human behavior.

Limitations. First, using Mechanical Turk to recruit survey participants inevitably introduces sampling biases (*e.g.*, bias towards Foursquare users who also use MTurk). However, the fact that two very different groups of users (Turkers and users in the Primary study) reach relatively consistent results lends confidence to these conclusions.

Second, our study focuses on the original Foursquare app, before the introduction of Swarm. We note that changes to Swarm over time actually serve to confirm some of our findings. When first introduced, Swarm deemphasized reward and gaming mechanisms by removing badges and then limiting mayorships to competition between friends. However, evidence showed that users reacted negatively to the loss of these incentives with significant drops in app downloads, review scores and web traffic (VB News 2014; Hu 2014). As a result, Swarm recently restored the old mayorship system (Foursquare Blog 2015a), and began to reintegrate the old badge system into the new system (Foursquare Blog 2015b). This backtracking in design further validates our hypothesis (*H5*) that reward mechanisms are indeed critical to driving engagement by Foursquare users.

Conclusion

Capturing human movement through online traces has the potential to positively impact a wide variety of application areas. Yet, discrepancies between users' check-in behaviors and their actual movement patterns persist. This paper details our efforts to understand the source of discrepancies between Foursquare check-in traces and users' real mobility patterns. Our survey study shows that both active and regular Foursquare users misrepresent their behavior. Specifically, we find that Foursquare users misrepresent their own locations in LBSNs and observe other users misrepresenting their locations as well. While missing check-ins are explained by personal interests such as privacy concerns, the place is uninteresting, and simply forgetting, extraneous check-ins are the result of extrinsic motivators such as badges, mayorships and financial rewards.

We present an effort to replicate and expand existing research findings using social media data. In addition to quantifying discrepancies between check-ins and actual mobility, we use a new crowdsourced user study dataset to explore motivations behind user check-in behavior, and discuss its broader implications to human mobility research. Future work could expand these results to a broader demographic of Foursquare users, and also investigate location misrepresentations in other types of user-generated traces online.

Acknowledgments

We thank the anonymous reviewers for their constructive comments. This project was supported by NSF grants CNS-1527939 and IIS-1321083.

References

Allamanis, M.; Scellato, S.; and Mascolo, C. 2012. Evolution of a location-based online social network: analysis and models. In *Proc. of IMC*.

Antin, J., and Shaw, A. 2012. Social desirability bias and self-reports of motivation: A study of amazon mechanical turk in the us and india. In *Proc. of CHI*.

Bilogrevic, I., et al. 2015. Predicting users' motivations behind location check-ins and utility implications of privacy protection mechanisms. In *Proc. of NDSS*.

Buhrmester, M. D.; Kwang, T.; and Gosling, S. D. 2011. Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data? *Perspectives on Psychological Science* 6:3–5.

Cheng, Z.; Caverlee, J.; Lee, K.; and Sui, D. 2011. Exploring millions of footprints in location sharing services. In *Proc. of ICWSM*.

Cho, E.; Myers, S. A.; and Leskovec, J. 2011. Friendship and mobility: user movement in location-based social networks. In *Proc. of SIGKDD*.

Cramer, H.; Rost, M.; and Holmquist, L. E. 2011. Performing a check-in: Emerging practices, norms and "conflicts" in location-sharing using foursquare. In *Proc. of MobileHCI*.

2015a. Claim that crown: Mayorships are back. Foursquare Blog.

2015b. Your old Foursquare badges and mayorships, evolved into a brand new game on Swarm. Foursquare Blog.

Fowler, F. J. 1995. *Improving Survey Questions: Design and Evaluation (Applied Social Research Methods)*. Sage Publications, Inc, 1 edition.

Gadiraju, U.; Kawase, R.; Dietze, S.; and Demartini, G. 2015. Understanding malicious behaviour in crowdsourcing platforms: The case of online surveys. In *Proc. of CHI*.

Georgiev, P.; Noulas, A.; and Mascolo, C. 2014. Where businesses thrive: Predicting the impact of the olympic games on local retailers through location-based services data. In *Proc. of ICWSM*.

Guha, S., and Birnholtz, J. 2013. Can you see me now?: Location, visibility and the management of impressions on foursquare. In *Proc. of MobileHCI*.

Hecht, B., and Stephens, M. 2014. A tale of cities: Urban biases in volunteered geographic information. In *Proc. of ICWSM*.

Hu, E. 2014. Some loyal foursquare users are checking out after swarm spinoff. NPR.

Jardosh, A.; Belding, E. M.; Almeroth, K. C.; and Suri, S. 2003. Towards realistic mobility models for mobile ad hoc networks. In *Proc. of MobiCom*.

Johnson, D. B., and Maltz, D. A. 1996. Dynamic source routing in ad hoc wireless networks. *Mobile Computing* 5:153–179.

Lindqvist, J., et al. 2011. I'm the mayor of my house: Examining why people use foursquare - a social-driven location sharing application. In *Proc. of CHI*.

Mason, W., and Suri, S. 2012. Conducting behavioral research on amazons mechanical turk. *Behavior Research Methods* 44(1):1–23.

Noulas, A., et al. 2011. An empirical study of geographic user activity patterns in foursquare. In *Proc. of ICWSM*.

Noulas, A., et al. 2012. A tale of many cities: universal patterns in human urban mobility. *PloS one* 7(5).

Patil, S.; Norcie, G.; Kapadia, A.; and Lee, A. 2012a. "check out where i am!": Location-sharing motivations, preferences, and practices. In *Proc. of CHI (Extended Abstracts)*.

Patil, S.; Norcie, G.; Kapadia, A.; and Lee, A. J. 2012b. Reasons, rewards, regrets: Privacy considerations in location sharing as an interactive practice. In *Proc. of SOUPS*.

Ross, J., et al. 2010. Who are the crowdworkers?: Shifting demographics in mechanical turk. In *Proc. of CHI (Extended Abstracts)*.

Rost, M., et al. 2013. Representation and communication: Challenges in interpreting large social media datasets. In *Proc. of CSCW*.

Scellato, S., et al. 2011. Nextplace: a spatio-temporal prediction framework for pervasive systems. In *Proc. of Pervasive*.

Scellato, S.; Mascolo, C.; Musolesi, M.; and Crowcroft, J. 2011. Track globally, deliver locally: improving content delivery networks by tracking geographic social cascades. In *Proc. of WWW*.

Scellato, S.; Noulas, A.; and Mascolo, C. 2011. Exploiting place features in link prediction on location-based social networks. In *Proc. of SIGKDD*.

Sen, S. W., et al. 2015. Barriers to the localness of volunteered geographic information. In *Proc. of CHI*.

Stouffer, S. 1940. Intervening opportunities: A theory relating mobility and distance. *American Sociological Review* 845–867.

Toma, C. L., and Hancock, J. T. 2010. Reading between the lines: Linguistic cues to deception in online dating profiles. In *Proc. of CSCW*.

2014. Will the check-in survive? Foursquares controversial relaunch, by the numbers.

Venerandi, A., et al. 2015. Measuring urban deprivation from user generated content. In *Proc. of CSCW*.

Zhang, Z., et al. 2013. On the validity of geosocial mobility traces. In *Proc. of HotNets*.