On the Validity of Geosocial Mobility Traces

Zengbin Zhang, Lin Zhou, Xiaohan Zhao, Gang Wang, Yu Su, [§]Miriam Metzger,
Haitao Zheng, and Ben Y. Zhao
Department of Computer Science, UC Santa Barbara, Santa Barbara, CA 93106, USA
[§]Department of Communications, UC Santa Barbara, Santa Barbara, CA, 93106, USA
{*zengbin, linzhou, xiaohanzhao, gangw, ysu, htzheng, ravenben*}*@cs.ucsb.edu, metzger@comm.ucsb.edu*

ABSTRACT

Mobile networking researchers have long searched for large-scale, fine-grained traces of human movement, which have remained elusive for both privacy and logistical reasons. Recently, researchers have begun to focus on geosocial mobility traces, *e.g.* Foursquare checkin traces, because of their availability and scale. But are we conceding correctness in our zeal for data? In this paper, we take initial steps towards quantifying the value of geosocial datasets using a large ground truth dataset gathered from a user study. By comparing GPS traces against Foursquare checkins, we find that a large portion of visited locations is missing from checkins, and most checkin events are either forged or superfluous events. We characterize extraneous checkins, describe possible techniques for their detection, and show that both extraneous and missing checkins introduce significant errors into applications driven by these traces.

Categories and Subject Descriptors

H.1.2 [**Information Systems**]: Human factors

General Terms

Human Factors, Measurement

Keywords

Location based social networks, Measurement, Mobility

1. INTRODUCTION

For quite some time, the holy grail quest of mobile networking research has been the search for large-scale, fine-grained mobility traces of human movement. Such traces can provide the basis for a large range of applications, ranging from practical applications like traffic prediction and urban planning, to research applications like guiding the design of cellular network protocols and smartphone energy management systems.

Unfortunately, for both privacy and logistical reasons, access to such mobility traces has been elusive. This in turn led to the rise of numerous alternatives, including synthetic mobility models ranging from random waypoint [14] to obstacle-based models [13], to adoption of unconventional sources of mobility datasets, *e.g.* movement traces from the SecondLife virtual reality platform [16].

Most recently, attention has turned towards geosocial mobility traces, traces of "check-in" events gathered from location based social networks (LBSN) such as Foursquare, Gowalla and Brightkite [8, 19, 21]. These datasets are attractive as mobility traces, because they are relatively easy to obtain, and provide data for relatively large user populations (Foursquare has 30 million users with more than 4 billion checkins, and Facebook Places could provide data on all of its 1 billion users). In fact, researchers are already relying on geosocial mobility traces to predict human movement [9,20,25], infer friendships based on visited locations [4, 26], and improve the efficiency of content delivery networks [24].

So is this it? Have we solved our long running quest for realistic mobility traces, or are we overlooking potentially misleading data in our zealous pursuit of mobility? Just how reflective are "check-in" traces of our true mobility patterns, and how significantly would any potential discrepancies impact applications and systems that rely on these datasets?

In this paper, we take concrete steps towards answering these questions, by performing a large user mobility study, and comparing a ground-truth of user mobility (via GPS data) to a Foursquare dataset for the same users. We compare "checkin" events from Foursquare to stationary events in the GPS trace, and make several surprising findings:

- First, we find that there is only a small subset of common events in the two traces. Foursquare checkins only cover roughly 10% of all locations a user visits. In addition to these missing locations, roughly 75% of all checkins are extraneous events that do not match real mobility.
- Second, we analyze and breakdown extraneous checkins into 3 types of user actions, multiple checkins at a single location, checking in at remote locations, or driveby

Hotnets '13, November 21–22, 2013, College Park, MD, USA.

checkins while moving at high speed. We find strong statistical correlations between each type of action and individual incentives in Foursquare, which hint at the motivation behind these actions.

- Third, we characterize extraneous checkins to identify potential features for use in automated detection.
- Finally, we use both datasets to drive a simulation of a mobile ad hoc network, and show that both missing and extraneous checkins have a significant impact on application outcomes.

To the best of our knowledge, our study is the first to quantitatively evaluate the accuracy and validity of geosocial mobility traces. In addition to identifying significant deviations from true mobility, we offer potential techniques to detect and filter extraneous checkins. We also outline the challenge of recovering "missing" checkins via data extrapolation, a necessary step towards making geosocial mobility traces useful to real applications.

2. BACKGROUND AND RELATED WORK

As background, we briefly describe today's location-based social networks, followed by existing work on tracking human mobility.

Location-based Social Networks (LBSN). Today's LBSNs allow users to embed current locations into social activities, *e.g.* checking in their nearby Point-of-Interest (POI) via mobile devices. Foursquare is one of the largest LBSNs, with over 30M registered users and 3 billion "checkins" (as of January 2013). Other popular sites include US-based Yelp and Gowalla (now part of Facebook), and China-based Sina Weibo and JiePang.

LBSNs incentivize user checkins using (virtual) rewards. In Foursquare for example, the user who checks in to a location most frequently in the last 60 days is awarded with the "Mayor" designation. In addition, "badges" are given to users for achieving certain checkin requirements, *e.g.* five different coffeeshops. Other LBSNs have similar incentives.

Human Mobility Tracking. Recently, researchers have been working on novel ways to gather detailed, timestamped user mobility traces. While this used to require GPS devices, recent efforts leverage the rich collection of sensors on smartphones to improve location accuracy [11,12,15,18]. But obtaining detailed traces requires significant overhead, thus existing efforts still remain very limited in scale [32,33].

To obtain large-scale human movement traces, others proposed relying on registration data from cellular or WiFi networks [6, 7, 27, 28, 30]. Such data approximates a user's location as the coverage area of her registered cellular basestation or WiFi AP. Unfortunately, these introduce large errors (several Kms for cellular and 100s of meters for WiFi) [31], and are limited in geographical coverage. Since they rely on call registrations, they also sample locations unevenly (*e.g.* centered on home and work locations), leading to biased representations of human mobility [22].

Datasets	# of users	Avg days per user	# of checkins	# of visits	GPS points
Primary	244	14.2	14K	31K	2.6M
Baseline	47	20.8	665	6.3K	558K

Table 1: Statistics of our primary and baseline datasets.

Geosocial Mobility Traces. An increasing number of researchers are using large-scale geosocial data in place of human mobility traces [8, 19, 21]. Researchers have used checkin traces from Gowalla, Brightkite and Foursquare to predict human movement [9,20,25], infer friendships [4,26], and improve content delivery networks [24].

3. DATASETS

To empirically validate geosocial mobility traces, we perform a user study and gather matching GPS traces and Foursquare checkin traces for an identical set of users. In the following, we describe our data collection process and the two datasets used for our study.

Data Collection. Our goal is to collect matching physical mobility (GPS) traces and Foursquare checkin traces[1] for the same set of users. We built a smartphone application (for both Android and iOS platforms), which generates two matching traces: a per-minute GPS trace of the user's location, and a trace of the user's checkin events polled using Foursquare's open API. We process the GPS trace to detect "visits" to points of interest (POI), and define a visit as the user staying at one location for longer than some period of time, *e.g.* 6 minutes. When GPS signals are not available, *e.g.* indoors inside a POI, the app uses the phone's WiFi radio and accelerometer to determine if the user is stationary or moving, similar to [15]. We obtained human subject IRB approval for our study.

Primary Dataset. Our main dataset came from ordinary Foursquare users who installed our application from Google Play and Amazon App stores. We advertised our application specifically to Foursquare users. Between January 2013 and July 2013, we received data from 244 users worldwide, where each user's measurement data covered an average of 14 days. This produced two mobility traces (see Table 1):

- **Checkin Trace** contains 14,297 Foursquare checkins events. Each event includes a timestamp, the name of a POI, its category and GPS coordinates.
- **GPS Trace** contains 2,600,000 sets of GPS coordinates. It captures each user's GPS location on a per-minute basis, and a list of 30,835 POI "visits." Each visit is a period of 6+ minutes when the user remains in one location.

We compared our checkin trace with existing Foursquare checkin traces collected by prior work [3, 21]. They share the same statistical properties, including the distribution of checkin count, inter-checkin time, number of badges, number of friends, and user nationality. This lends support to our belief that our dataset is representative.

[1]We focus on Foursquare checkins due to its popularity.

Baseline Dataset. We collected a smaller dataset (also 2 traces) by recruiting 47 undergraduate volunteers from our university. These students installed our smartphone application and participated as normal users. Later, we will use this dataset as a control group to validate the results of our checkin filtering mechanisms.

4. VALIDATING CHECKINS

To understand how well Foursquare checkins correspond to each user's physical mobility patterns, the first step in our validation process is to "match up" visits to points of interest in both the GPS and Foursquare traces.

4.1 Matching Checkins to Visits

Here, we first introduce our algorithm for matching Foursquare checkin events to visits in our GPS trace, then describe the matching results and findings.

Matching Algorithm. Our algorithm matches a user's Foursquare checkins to her GPS visits based on their associated GPS coordinates and timestamps.

- **Step1:** For each checkin event c_i in a user's Foursquare trace, we identify from the same user's GPS trace a set of visits $\{V\}$ whose physical locations are within α meters from c_i's location. Here $\{V\}$ can contain one or multiple visits, or the null set.
- **Step2:** If $\{V\}$ is non-null, find the visit v_j from $\{V\}$ whose timestamp is closest to that of checkin c_i. If the difference between the two timestamps[2] is less than β, then v_j matches c_i.

Ideally, there would be a one-to-one mapping between checkin events and GPS visits. Our algorithm ensures that each checkin event has at most one matching visit. If multiple checkins are matched to the same visit v_j, (*i.e.* a user checks in at multiple POIs when visiting one), we match v_j to the geographically closest checkin event.

We have experimented with a range of α and β values, and found that the matching results are most consistent for values $\alpha = 500m$ and $\beta = 30min$. Thus we select these values for our analysis. Both thresholds are relatively loose to lower the bar for matching checkins to visits. Therefore, our results likely represent an upper limit on event matches.

Matching Results. We first ran the matching algorithm on the Primary dataset with ordinary Foursquare users. Its checkin trace contains 14297 checkins while the GPS trace includes 30835 visits. Figure 1 presents the matching result as a Venn diagram, where we partition the data into three categories.

- **Honest Checkins** – 3525 checkins events match up with GPS visits. These checkins events correspond to GPS readings that show the user was indeed at the physical location matching her Foursquare checkin event. This represents a shockingly small portion of both checkins and GPS visits.
- **Extraneous Checkins** – 10772 checkin events (75% of total checkins) do not match up with any matching visit in the GPS trace.
- **Missing Checkins (or Unmatched Visits)** – These 27310 visits in the GPS trace (89% of all visits) do not match any Foursquare checkin events.

[2]Each visit has a start time T_s and an end time T_e. We calculate Δt, the timestamp difference between v and a checkin with timestamp (T_c) as follows: if $T_s \le T_c \le T_e$, $\Delta t = 0$; Otherwise, $\Delta t = min(|T_c - T_s|, |T_c - T_e|)$.

We wish to validate that our set of "honest checkins" are an accurate representation of real checkin activity. For this, we compare it against our Baseline dataset of undergraduate volunteers. Since these volunteers were participating to satisfy a research requirement, they were much less likely to be influenced by Foursquare rewards. We use several common mobility metrics to compare the two datasets, including inter-arrival time distribution, movement distance distribution, event frequency, speed distribution and POI entropy [8, 9, 19].

Figure 2 plots the inter-arrival time distribution results using the GPS and checkin traces from both datasets, as well as the honest checkins from the Primary dataset. We see that GPS traces from both datasets match up near perfectly. In addition, the entire set of checkin events from the baseline match up perfectly with the honest checkin set from the primary dataset, while the set of all checkins from the primary data shows significant differences. The other metrics led to the same conclusions (results omitted due to space limits), thus validating that our matching algorithm did accurately identify the set of honest Foursquare checkin events.

4.2 Missing Checkins

The large number of missing and extraneous checkins raises serious concerns on whether checkins truly match human mobility. We briefly discuss missing checkins here, and analyze extraneous checkins later in Section 5.

The obvious question is, which locations are users not checking in at, and why? Our intuition is that users typically forget to check in at specific (perhaps boring or routine) places that they visit frequently, *e.g.* home, office, gas stations. If so, then a small number of places will account for the large majority of missing checkins. To validate this, we identify for each user the top-n most visited POIs, and examine the portion of her missing checkins attributable to these top POIs. Figure 3 plots the CDF of this ratio across all users for their top-5 visits. The results confirm our hypothesis, for roughly 60% of all users, 5 locations account for more than half of their missing checkins. For 20% of users, just a single location accounts for more than 40% of their missing checkins.

We also looked at the types of POIs responsible for those missing checkins. Figure 4 shows the break down of missing checkins locations into 9 categories based on Foursquare's POI classification. The three POI categories with the most missing checkins are *Professional*, *Shop* and *Food*. Intu-

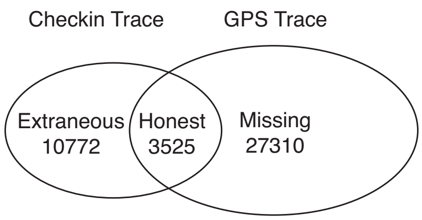


Figure 1: Matching results of the Primary dataset.

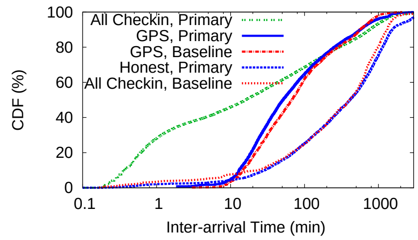


Figure 2: CDF of inter-arrival time.

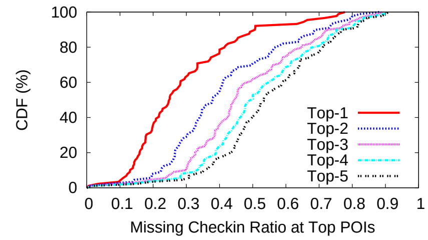


Figure 3: CDF of the portion of missing checkins at top-n most visited POIs.

itively, these places are related to people's routine activities: going to work, grocery shopping and eating. According to existing survey [10, 17], users typically do not checkin at places that they think are "boring" or "private." However, these frequently visited places are critical parts of a user's mobility pattern.

In summary, our analysis shows that missing checkins often cover frequently visited locations in a user's daily life, and are therefore critical components in a human mobility trace. Their absence means geosocial traces are missing a large majority of each user's mobile history.

5. EXTRANEOUS CHECKIN ANALYSIS

While missing checkins are often results of carelessness, extraneous checkins (or checkins without a matching visit) occur when users intentionally misrepresent their physical location. In this section, we study the behaviors and reasons behind extraneous checkins. We first categorize observed extraneous checkins based on user behaviors, and then try to infer possible motivating incentives behind them. We then identify characteristics that can potentially serve to distinguish extraneous checkins.

5.1 Types of Extraneous Checkins

We manually inspected our pool of 10772 extraneous checkins, and found that 90% could be classified into one of three types of user behavior. The remaining 10% do not display any distinctive features.

- **Superfluous Checkins**. When visiting and checking in to one POI, some users also check in to multiple nearby POIs from the same physical location. We found 2176 superfluous checkins in our dataset (15% of all checkins and 20% of extraneous checkins).
- **Remote Checkins.** These are checkins to POIs more than 500 meters away from a user's actual GPS location. 500m is beyond any reasonable GPS or POI location errors, and the user is clearly falsifying her location. Our dataset has 5715 remote checkins (40% of all checkins and 53% of extraneous checkins).
- **Driveby Checkins.** These occur when users checkin to nearby POIs while moving at a moderate speed. Computing speed from our GPS trace, we treat a checkin as driveby if its speed exceeds 4mph. This produces 1782 driveby checkins (13% of all checkins).

Checkin Type	Correlation			
	#Friends	#Badges	#Mayors	#Checkins/Day
Superfluous	0.22	0.07	**0.34**	0.15
Remote	0.18	**0.49**	0.16	0.15
Driveby	-0.10	-0.21	-0.08	**0.21**
Honest	-0.09	-0.42	-0.23	-0.40

Table 2: Correlation between the ratio of each type of checkin and user's profile features.

5.2 Incentives for Extraneous Checkins

There could be many potential causes for extraneous checkins, from collecting mayorships/badges to attracting more friends. To infer the key causes, we measure the correlation between a user's extraneous checkins and her user features, *i.e.* number of friends, number of badges, number of mayorships, and number of checkins per day. Table 2 lists the Pearson's correlation score between each user's four features and the ratios of her checkins (superfluous, remote, driveby and honest). The correlation score is between -1 and 1, where -1 means perfect negative correlation and 1 means perfect positive correlation.

We make two interesting observations. First, both remote and superfluous checkins correlate strongly with Foursquare's reward mechanisms (badge and mayorship). This suggests that social rewards are key incentives for these extraneous checkins. For example, to collect badges, Foursquare users need to check in at various new POIs that could be far from their usual locations. Thus they are motivated to submit fake checkins remotely. Similarly, to become mayor of a POI, one must check in more than all other users. This motivates her to check in even when not physically visiting the POI. However, Foursquare does not allow remote checkins to count for mayorships, only badges. Finally, since badges require checkins at specific POIs, superfluous checkins do not help.

Second, honest checkins have negative correlation with all four user features. This indicates that "honest" users tend to be less active (*i.e.* less checkins per day, less badges). Also, driveby checkins look similar to honest ones: all negative correlations except with number of checkins per day. Es-

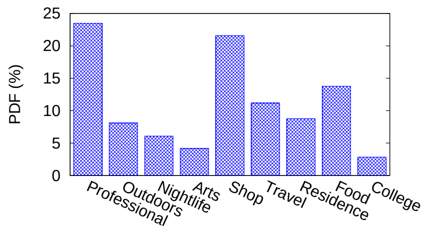


Figure 4: Breakdown of missing checkins by POI category.

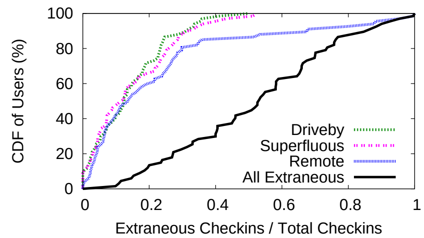


Figure 5: User's ratio of extraneous checkins.

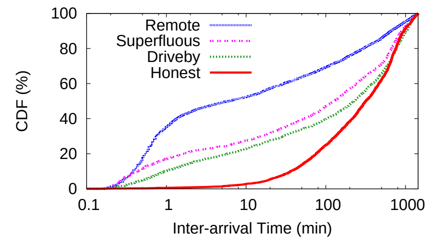


Figure 6: Burstiness of extraneous checkins.

sentially, driveby checkins did not lie; they just did not stay long enough to make a qualified visit. These users seem to be distinct from users who cheat for mayorships and badges.

Our results show that superfluous and remote checkins dominate extraneous checkins, and are likely motivated by Foursquare's user badges and mayorships. Since these rewards play a large role in encouraging user engagement in Foursquare and other LBSNs, these superfluous and remote checkins will likely remain a key component of geosocial mobility traces.

5.3 Distinguishing Characteristics

Finally, we characterize extraneous checkins to identify distinguishing characteristics that may serve to guide detectors or filters. Our initial analysis focuses on *per-user prevalence* and *temporal burstiness* properties.

Per-user prevalence. First, we want to see if extraneous checkins are endemic to specific portions of the user population. If so, then we can focus our attention on removing those users. We compute for each user the portion of her checkins which are extraneous. To our surprise, the CDF in Figure 5 shows that extraneous checkins are widespread. Nearly all users produced extraneous checkins, and for 20% of users, extraneous checkins accounted for up to 80% of their checkin events.

These results raise serious concerns about the prevalence of unreliable checkins in the user population. They also mean we cannot target specific users in an effort to filter extraneous checkins, unless we are willing to sacrifice a significant number of honest checkins. For example, filtering out users who generate 80% of all extraneous checkins would also filter out 53% of honest checkins!

Burstiness. We observe from the dataset that each user's honest checkins are evenly distributed in time, while her extraneous checkins tend to be bursty. Figure 6 plots the CDF of the inter-arrival time for the three types of extraneous checkins as well as that of the honest checkins. The results show that the majority of extraneous checkins arrive within a small interval (less than 10 minutes), and 35% of them arrive within 1 minute! In contrast, the interarrival time for honest checkins is more than 10 minutes. This suggests that we can possibly identify extraneous checkins by looking at their arrival characteristics.

6. APPLICATION-LEVEL IMPACT

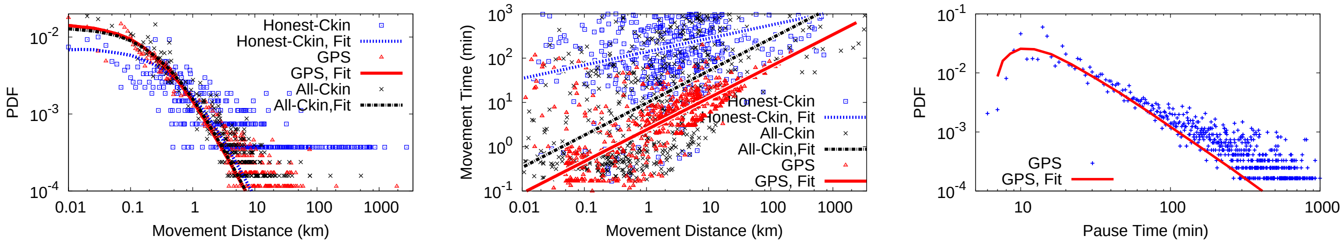
Having quantified some of the differences between Foursquare checkins and physical mobility, we now seek to quantify the impact on applications that rely on geosocial mobility traces as mobility datasets. We use GPS and Foursquare traces to drive a simulation of a mobile ad hoc network, and show that both missing and extraneous checkins significantly impact application outcomes.

6.1 Mobility Model Training

Simulations of mobile ad hoc networks rely on a mobility model to generate movement patterns for arbitrary sized networks [5]. In our experiments, we use our GPS and Foursquare checkin traces to drive a mobility model, and evaluate the net impact on mobile ad hoc network performance. For our model, we choose Levy Walk [23, 29], the most popular model able to generate mobility predictions by fitting to GPS data. To understand the impact of extraneous and missing checkins, we use three traces to train the mobility model: all-checkins (honest+extraneous), honest-checkins, and GPS visits.

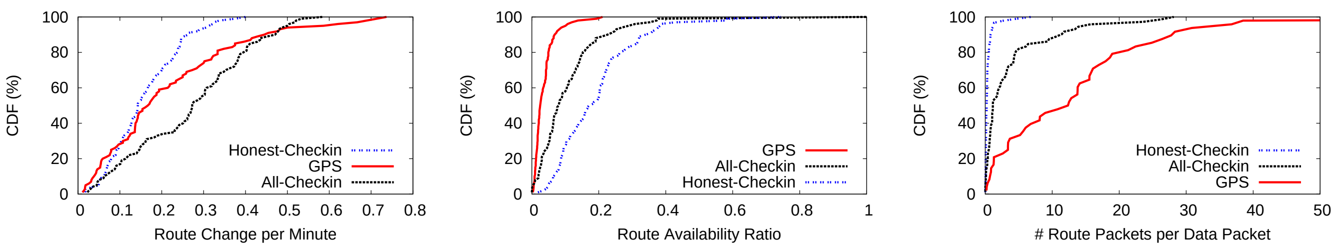
The Levy Walk model captures human movements as a sequence of trips with pauses in between. It has three inputs: distributions of movement distance, movement time and pause time. Following prior work [23], we fit the movement distance and pause time to the Pareto distribution, and the movement time distribution to $t = ke^{(1-\rho)}$, where t is the movement time, d is the movement distance, and k and ρ are constants. For the two checkin datasets that do not provide pause time, we take a conservative approach and use the distribution from GPS in their model training.

Fitting Results. Figure 7 shows the fitting results. We observe large differences among the three datasets. Comparing GPS with honest-checkin, we see that missing checkins are responsible for lowering both moving distance (Figure 7(a)) and moving time (Figure 7(b)). This is not surprising. because missing checkins offer a more detailed sample of user mobility. Next, we can observe the impact of extraneous checkins by comparing all-checkin to honest-checkin: they lower the movement distance and produce many more fast moving segments. Next, we will examine how these model differences translate into application performance results.



(a) Movement Distance (b) Movement Time vs. Distance (c) Pause Time

Figure 7: Mobility model fitting using three datasets: honest-checkin, all-checkin, and GPS.



(a) Route Change Frequency (b) Route Availability Ratio (c) Routing Overhead

Figure 8: MANET performance.

6.2 MANET Simulation Results

We use the three fitted models in simulations of a mobile ad hoc network (MANET). We generate synthetic mobile movement traces using the fitted models and feed them into a NS-2 AODV simulator [2]. We place 200 mobile nodes in a $100km \times 100km$ area, each with a $1km$ communication range. They form 100 random node pairs and each node pair communicates using constant bit rate (CBR) streams. The simulation outputs three application metrics: route change frequency, route availability ratio, and routing overhead.

Figure 8 compares the resulting application metrics for traces generated based on the three mobility traces. We can clearly see the difference. First, compared to the groundtruth results using the GPS trace, routes in all-checkin have higher update frequency, availability and lower routing overhead. This is because the all-checkin mobility model produces much lower moving speeds. As users move faster, routes are less stable, incurring more overhead. This demonstrates the compound effect of both missing and extraneous checkins.

We also look at the results after removing extraneous checkins, which are shown by the honest-checkins line. We see that it still has significant deviations from the groundtruth GPS results. Compared to GPS, routes in honest-checkin are updated less frequently, incur much less overhead, and yet have almost 2x higher availability.

Summary. The key takeaway result is that our MANET experiment shows significant deviations when relying on geosocial mobility traces (all-checkin traces). Once we remove all extraneous checkins, the resulting trace still produces significant errors. This means that to achieve accurate results, we need to both remove extraneous checkins and add data points to account for the missing checkin events. Finally, we believe the same issues apply to a variety of applications. For example, city planning applications [1] will under-estimate traffic on routes between residential areas and offices, due to fewer checkins in these places. Similarly, friendship recommendation applications [4, 26] leverage user physical proximity to suggest social connections. Using data including fake checkins will lead to wrong inferences on user proximity, and lead to incorrect suggestions.

7. CONCLUSION & OPEN PROBLEMS

In this paper, we used ground-truth GPS traces from a user study to validate the ability of geosocial traces to capture human mobility. We find that 75% of events in Foursquare checkin traces are extraneous checkins generated by users to achieve in-system rewards, and checkin events only capture 10% of actual visited locations from real physical mobility traces. We also show that these discrepancies translate to significan deviations in results of applications relying on these traces. Looking forward, we see two major challenges.

Detecting Extraneous Checkins. Identifying extraneous checkins is the first step towards a trace that more accurately captures real mobility patterns. Our preliminary work identified temporal burstiness as one potential feature for detection, but a more thorough analysis (perhaps applying machine learning techniques) is necessary.

Recovering Missing Locations. A more difficult challenge is to fill in the missing locations visited by but not reported by users. Our work shows that even approximations of 1 or more key locations (home, work) will go a long way towards improving accuracy. One approach is up-sampling observed checkins to based on statistical models of real user mobility. Another is to fill in locations based on models of user checkin rates for different POI categories.

8. REFERENCES

[1] Tampa master plan to integrate Foursquare data. `http://www.bizjournals.com/tampabay/blog/socialmadness/2012/04/tampa-master-plan-foursquare.html`.

[2] The network simulator: ns-2. `http://www.isi.edu/nsnam/ns/`.

[3] AGGARWAL, A., ALMEIDA, J., AND KUMARAGURU, P. Detection of spam tipping behaviour on foursquare. In *Proc. of WWW* (2013).

[4] ALLAMANIS, M., SCELLATO, S., AND MASCOLO, C. Evolution of a location-based online social network: analysis and models. In *Proc. of IMC* (2012).

[5] BAI, F., SADAGOPAN, N., AND HELMY, A. The important framework for analyzing the impact of mobility on performance of routing protocols for adhoc networks. *Ad Hoc Networks 1*, 4 (2003), 383–403.

[6] BALAZINSKA, M., AND CASTRO, P. Characterizing mobility and network usage in a corporate wireless local-area network. In *Proc. of MobiSys* (2003).

[7] BHATTACHARJEE, D., RAO, A., SHAH, C., SHAH, M., AND HELMY, A. Empirical modeling of campus-wide pedestrian mobility observations on the usc campus. In *Proc. of VTC* (2004).

[8] CHENG, Z., CAVERLEE, J., LEE, K., AND SUI, D. Z. Exploring millions of footprints in location sharing services. In *Proc. of ICWSM* (2011).

[9] CHO, E., MYERS, S. A., AND LESKOVEC, J. Friendship and mobility: user movement in location-based social networks. In *Proc. of SIGKDD* (2011).

[10] CRAMER, H., ROST, M., AND HOLMQUIST, L. E. Performing a check-in: emerging practices, norms and "conflicts" in location-sharing using foursquare. In *Proc. of MobileHCI* (2011).

[11] HARIHARAN, R., AND TOYAMA, K. Project lachesis: parsing and modeling location histories. In *Proc. of GIScience* (2004).

[12] HIGHTOWER, J., CONSOLVO, S., LAMARCA, A., SMITH, I., AND HUGHES, J. Learning and recognizing the places we go. In *Proc. of UbiComp* (2005).

[13] JARDOSH, A., BELDING, E. M., ALMEROTH, K. C., AND SURI, S. Towards realistic mobility models for mobile ad hoc networks. In *Proc. of MobiCom* (2003).

[14] JOHNSON, D. B., AND MALTZ, D. A. Dynamic source routing in ad hoc wireless networks. *Mobile Computing 5* (1996), 153–179.

[15] KIM, D. H., KIM, Y., ESTRIN, D., AND SRIVASTAVA, M. B. Sensloc: sensing everyday places and paths using less energy. In *Proc. of SenSys* (2010).

[16] LA, C.-A., AND MICHIARDI, P. Characterizing user mobility in second life. In *Proc. of WOSN* (2008).

[17] LINDQVIST, J., CRANSHAW, J., WIESE, J., HONG, J., AND ZIMMERMAN, J. I'm the mayor of my house: examining why people use foursquare-a social-driven location sharing application. In *Proc. of SIGCHI* (2011).

[18] MONTOLIU, R., BLOM, J., AND GATICA-PEREZ, D. Discovering places of interest in everyday life from smartphone data. *Multimedia Tools and Applications 62*, 1 (2013), 179–207.

[19] NOULAS, A., SCELLATO, S., LAMBIOTTE, R., PONTIL, M., AND MASCOLO, C. A tale of many cities: universal patterns in human urban mobility. *PloS one 7*, 5 (2012), e37027.

[20] NOULAS, A., SCELLATO, S., LATHIA, N., AND MASCOLO, C. Mining user mobility features for next place prediction in location-based services. In *Proc. of ICDM* (2012).

[21] NOULAS, A., SCELLATO, S., MASCOLO, C., AND PONTIL, M. An empirical study of geographic user activity patterns in foursquare. In *Proc. of ICWSM* (2011).

[22] RANJAN, G., ZANG, H., ZHANG, Z.-L., AND BOLOT, J. Are call detail records biased for sampling human mobility? *ACM SIGMOBILE Mobile Computing and Communications Review 16*, 3 (2012), 33–44.

[23] RHEE, I., SHIN, M., HONG, S., LEE, K., KIM, S. J., AND CHONG, S. On the levy-walk nature of human mobility. *IEEE/ACM Trans. on Networking 19*, 3 (2011), 630–643.

[24] SCELLATO, S., MASCOLO, C., MUSOLESI, M., AND CROWCROFT, J. Track globally, deliver locally: improving content delivery networks by tracking geographic social cascades. In *Proc. of WWW* (2011).

[25] SCELLATO, S., MUSOLESI, M., MASCOLO, C., LATORA, V., AND CAMPBELL, A. T. Nextplace: a spatio-temporal prediction framework for pervasive systems. In *Proc. of Pervasive* (2011).

[26] SCELLATO, S., NOULAS, A., AND MASCOLO, C. Exploiting place features in link prediction on location-based social networks. In *Proc. of SIGKDD* (2011).

[27] SONG, C., QU, Z., BLUMM, N., AND BARABÁSI, A.-L. Limits of predictability in human mobility. *Science 327*, 5968 (2010), 1018–1021.

[28] TUDUCE, C., AND GROSS, T. A mobility model based on wlan traces and its validation. In *Proc. of INFOCOM* (2005).

[29] WANG, P., AND AKYILDIZ, I. F. Effects of different mobility models on traffic patterns in wireless sensor networks. In *Proc. of GLOBECOM* (2010).

[30] YOON, J., NOBLE, B. D., LIU, M., AND KIM, M. Building realistic mobility models from coarse-grained traces. In *Proc. of MobiSys* (2006).

[31] ZANDBERGEN, P. A. Accuracy of iphone locations: A comparison of assisted gps, wifi and cellular positioning. *Trans. in GIS 13*, s1 (2009), 5–25.

[32] ZHENG, Y., ZHANG, L., XIE, X., AND MA, W.-Y. Mining interesting locations and travel sequences from gps trajectories. In *Proc. of WWW* (2009).

[33] ZHOU, C., FRANKOWSKI, D., LUDFORD, P., SHEKHAR, S., AND TERVEEN, L. Discovering personally meaningful places: An interactive clustering approach. *ACM Trans. on Information Systems 25*, 3 (2007), 12.