Extrapolating Sparse Large-Scale GPS Traces for Contact Evaluation

Andrea Hess
University of Vienna
andrea.hess@univie.ac.at

Jörg Ott
Aalto University, Comnet
jo@comnet.tkk.fi

ABSTRACT

Human mobility traces are increasingly used for more realistic evaluation of mobile (opportunistic) communication systems. Although GPS traces yield the most detailed data sets, they are often limited in scale and may be incomplete since they are captured using mobile devices carried by volunteers. In this paper, we explore mechanisms to improve completeness and connectivity patterns of sparse GPS traces and assess their impact by means of the GeoLife data set. We also outline insights into geographic propagation that can be gained through these large-scale location measurements.

Categories and Subject Descriptors

C.2.1 [**Network Architecture and Design**]: Wireless communication; C.4 [**Performance of Systems**]: Modeling techniques

Keywords

Mobility trace analytics, Contact evaluation, Data sparsity

1. INTRODUCTION

Establishing device-to-device networks in challenged environments strongly relies on (predicted) connection opportunities resulting from two or more nodes colocated within radio range for a sufficient period of time. In addition to or instead of synthetic mobility models, researchers increasingly rely on real-world traces as input for evaluating connectivity and communication protocols. Such traces come usually in one of three flavors: 1) *Contact traces* are obtained from mobile devices scanning for peers (e.g., using Bluetooth). 2) In *network traces* WLAN access points, cellular base stations, or fixed Bluetooth scanners record associated devices and those connected to the same network element at the same time are considered to be "in contact"—see, e.g., [2, 6, 8] for contact studies based on contact and network traces. 3) (Assisted) *GPS traces* are recorded by mobile devices as their owners move around.

The first two types of traces share the drawback that exact node locations can hardly be derived and the observation granularity is limited both spatially and temporally: While for pure contact traces the locations of contacts are completely unknown, network traces contain at least IDs of the network infrastructure elements, such as access points and cell towers, and occasionally also their geographic location. However, given the larger transmission ranges of network elements, whether a short-range contact could take place can only be approximated from the data. Evaluations assuming different radio ranges are not feasible at all. The temporal resolution may be limited for (1) by the scanning intervals of the devices and for (2), especially for WLAN access points, by the time it takes to declare a mobile node dissociated after it was not seen anymore. Both may yield fairly coarse results. Moreover, the observations are often biased by the environment (e.g., university campus, theme park), one particular event (e.g., a conference), and/or the community to which devices or software is handed out (e.g., students).

GPS traces overcome spatial and temporal resolution limitations and are not dependent on infrastructure nodes detecting the presence of mobile nodes. GPS trace data are planet-scale and provide fine-grained movement trajectories – both in time (as sampling interval can be set in the range of seconds) and space (as a positioning accuracy of 10m can be assumed). Their main caveat is that positioning in indoor environments is not very accurate (or does not work at all) – and, of course, GPS locations of nodes do not reveal any information about obstacles between them, so that possible contacts remain estimates. Another limitation is that publicly available GPS traces are often collected in a sparse manner: experiments are usually limited to "smaller" user bases, typically with no more than one or two hundred participants; on top of this, the movements are only traced when the GPS device is active; and the readings are only useful when the device was carried by the person[1]. This often implies that traces may exhibit lower node densities than would be observable if traces from all people were available every day.

In this paper, we suggest two simple mechanisms for improving node density and thus contact frequency of GPS traces, which cover longer periods of time (at least several weeks): 1) We interpolate missing trajectories of a user on a given day from observations of the user on the same weekday of one or more other weeks. 2) We introduce "virtual users" that serve as otherwise invisible relays and allow us extrapolating contacts in space and time. We evaluate the impact of those mechanisms using the *GeoLife* trace [18], which was not explored before for direct mobile-to-mobile interactions.

We focus on GPS traces because they provide fine-grained location information, which allows researchers to assess different short-range radio technologies because contacts can be derived from the mobility data as a function of radio range and scanning intervals (but not vice versa). The availability of location data across longer periods of time also allows "filling in" missing pieces, which is not

[1]This actually applies to all data gathered on mobile devices, exceptions are traces of GPS devices in vehicles, such as taxis, for which richer data sets both in number and continuity exist.

HotPlanet'13, August 16, 2013, Hong Kong, China.

possible starting from contact data. The fine granularity also allows assessing node encounters "on the move"—while network traces (or "check-in" data from social networks such as FourSquare) won't provide a sufficient spatial or temporal granularity, nor allow for interpolating trajectories between successive data points. Finally, GPS data provide trajectories independent of any local infrastructure or dedicated locations and thus allow tracking the reach of mobile opportunistic networks, e.g., for information propagation.

2. RELATED WORK

A common approach for contact studies is to extract mobility characteristics from mobility trace data and to then adjust the node number in simulations (see, e.g., [6], [7], [9]). However, there have been several approaches previously proposed for adapting node and contact density directly. In [1], the authors present a connectivity trace generator providing synthetic traces with varying connectivity patterns based on real-world WLAN session traces. Characteristics, such as sojourn time at access points, colocation probability for node pairs, and node degree, are extracted from real data, while the connectivity variations are realized by scaling up and down the given node degree distributions. By estimating mobility based on contacts, the algorithm proposed in [15] enables the insertion of additional contacts or new nodes in contact trace-based studies. In [14], city-specific mobility profiles are derived from GPS trajectories, which were traced in various cities during sport activities and made available through public sport track repositories. For studying the connectivity in each city, the traces are densified by randomly sampling the dataset under the assumption that the traces are a representative subset of the overall traffic. W.r.t. spatial distribution analytics based on large-scale data, the study of WiFi data collected planet-scale by an Android application described in [11] is as well related to our work. Here, location cluster patterns for single users are extracted for five large cities on four continents. The GeoLife data set has previously been used for developing travel applications (location interestingness estimation [18] or travel paths recommendation [16]) and location prediction models [5] [10].

3. EXTRAPOLATING CONTACTS

As noted above, GPS traces may be "incomplete" for various reasons, yielding (a) trajectories for fewer users than could be observed and (b) especially for traces of large geographic scale, featuring fewer contacts than desirable for evaluating mobile opportunistic communication applications. We introduce mechanisms to address these two orthogonal issues: For (a), we discuss three mechanisms to "fill in" trajectories of users for which records do not exist for a given day (3.1). Concerning (b), we explore two complementary options to scale up contacts occurring in the traces (3.2). Both classes of measures assist in scaling existing traces to the limited extent: they lead to increased connectivity and thus improve the value of a given traces in protocol evaluation.

We map user locations to cells of a given edge length (we use 100 m) and assume that users located inside the same cell are in contact and thus able to communicate. We are aware that this is a simplification as users close by in adjacent cells won't be considered in contact while users in opposite corners of the same cell will be. However, we consider this approximation sufficient for the purpose of this paper, and the trace analysis as well as the extrapolation mechanisms discussed below may be refined to smaller cell sizes and/or different area shapes to produce more accurate results.

3.1 Trajectory projection

The first question we turn to is filling in trajectories for persons for whom records are missing on a given day. Assume we look at a given date D and a person A, who is active on other days, but does not have any records for D. What would have the person been doing? We assume that the person has a regular day job for a five-day week and works Monday through Friday. In this case, for a weekday, chances are that the behavior will roughly be the same as for the same day in the previous or the next week. In our considerations, we focus on weekdays because weekends with leisurely activities will likely exhibit more diversity.

To project a person's trajectories of several days onto one day, we calculate a path that represents her typical movements for the particular weekday. We account for temporal variations by considering weekday-dependent variations (i.e., only other Mondays are used to fill in for a Monday) as well as seasonal variations (e.g., people commute differently in summer and winter, students' schedules vary each semester). To do so, the trajectories for each person for a certain weekday in a moving time window covering n months (e.g., current, previous, and subsequent month) are included. We suggest three methods for computing a fill-in trajectory representative for a given day from trajectories of one or multiple days in case no trajectory exists for a person on the day in question.

3.1.1 Nearest Trajectory (NT) projection

The Nearest Trajectory method (NT) selects the trajectory closest in time. For example, if there are no movement traces for the particular Monday, the algorithm searches for trajectories traced on the previous Monday (first) and subsequent Monday (second). If none are found on adjacent weeks either, we continue searching two weeks before, two weeks after, and so on. The week count for the search is expanded until either a Monday trajectory is found or the time window limits are reached.

3.1.2 Medoid Location (ML) projection

The Medoid Location method (ML) composes the projected trajectory from the spatially most central locations. To do so, the trajectories for each person for one type of weekday within the entire time window are clustered. The Euclidean distance-based medoid location m of this cluster is then computed for every time step t in configurable intervals (e.g., of 30 seconds). We select the medoid (data element closest to the cluster centroid) instead of the centroid to prevent that trajectories embody "artificial" locations a person never visited in reality (and where the person might not even be able to get to). An example of a medoid path p_m resulting from four traced paths is shown in Figure 1(a).

3.1.3 Most Frequent Location (FL) projection

As for the ML, one representative location is computed by the (Most) Frequent Location method (FL) in every time step t. However, for FL, the locations visited most frequently in the time window of n months are selected here to build the frequent path p_f.

3.2 Temporal and spatial extrapolation

Even if we can reconstruct trajectories for all nodes for each weekday in a trace, the observed node density may still be low: when only a few hundred nodes move in a large area, the odds of them getting in contact remain fairly low as we will show later in our evaluation. Nodes may visit the same cells but at different times or may be close to each other in neighboring cells. To overcome this limitation, we introduce the notion of "virtual nodes": they are invisible in the traces but are capable of bridging contacts in time (3.2.1) or space (3.2.2). We assume, not unreasonably at least in urban environments, that further nodes exist that could relay data between the nodes in the traces and thus extend connectivity.

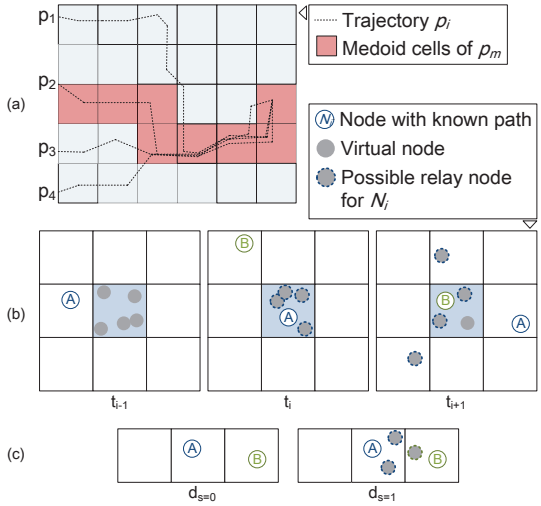

Figure 1: Illustration of (a) trajectory (medoid location) projection, (b) temporal extrapolation for one cell in three time steps, and (c) spatial extrapolation (without extrapolation $d_s = 0$ and with one cell extrapolation $d_s = 1$).

3.2.1 *Temporal interpolation*

In this case, the virtual nodes passing through a cell would replicate data received from one traced node A among each other so that the data survive inside the cell until another traced node B appears some time later. Such an assumption would thus extend the availability of a node in time and support one way information sharing from a node A present earlier in a given cell to a node B entering this cell later (Figure 1(b)). We limit the extension of contacts in time to a period d_t (e.g., $d_t \in \{1, 2, 3\}$h).

The concept of *Floating Content* has shown that geo-based information sharing in constrained areas by replication among mobile nodes is feasible for extended periods of time even under modest node density [13]. A reliable alternative would be fixed relays (similar to throwboxes [17]) to temporarily store data for relaying.

3.2.2 *Spatial extrapolation*

We can also extend the reach of our nodes in space: instead of just counting nodes in the same cell as a node A to be in contact, we also consider nodes such as B in adjacent cells within distance d_s (e.g., $d_s \in \{1, 2, 3\}$) as shown in Figure 1(c). In this case, one or more virtual nodes would be assumed acting as relays between the A and B and thus support establishing (instant) multi-hop paths (possibly using MANET routing protocols) between the two. The same approach can be used to model larger radio ranges.

Note that this approach is subtly different from just increasing the edge length of our cells: in the latter case, nodes A and B would not be considered in contact if the extended cell boundary would happen to be along the line between A and B.

4. EVALUATION

The evaluation of our approach is based on the GeoLife GPS dataset [18], which was mainly collected in China. Its latest version [12] contains movement paths of 182 people for the time period April 2007 to August 2012 in a sampling interval of about five seconds. Since we are interested in the every-day behavior during typical working days we filtered weekends and holidays out. For studying contacts and local distribution we focus on a 10×10 km excerpt of Beijing with high trajectory density, while we select all trajectories starting within Beijing (10 469 out of 17 621 trajectories) for the large-scale spatial distribution study.

We believe that the data have been sparsely sampled for several reasons, e.g., the participants might not want to carry GPS devices or activate GPS on their phones continuously every day (due to the power consumption and privacy concerns). The data set does not contain paths to work or university of every workday though, yet people go there largely on a daily basis. Moreover, we select an excerpt of Beijing covering the Haidian District, which is an accumulation zone home to 68 universities/colleges and 231 scientific institutions [3]. This allows to assume a reasonable population density, which is as well indicated by the high trajectory density in the data set. Thus, this area is suitable for contact extrapolation.

In the following, we study the effects of the projection and extrapolation methods proposed along a set of contact and spatial propagation metrics:

1. Node degree: The node degree gives the number of edges of a vertex in the connectivity graph; i.e., the number of nodes each node meets per day is calculated.
2. Number of contacts: The number of contacts gives the sum of contacts between all nodes during a day.
3. Inter-contact time: The inter-contact time gives the time elapsing between two consecutive contacts of a node pair.
4. Contact duration: The contact duration measures the time two nodes stay continuously within communication range of each other.
5. Disconnected islands: This metric gives the number of islands consisting of visited cells in a grid, which are surrounded only by unvisited cells.
6. Propagation speed: The propagation speed is here defined as the fraction of visited cells over time.
7. Flight length and flight time: The flight length is the length of the path traveled between two consecutive pauses (phases without movement), while the flight time is the period between the two pauses.
8. Mobility range: The mobility range is the maximum distance reached from the center of a rectangle covering the whole trajectory.
9. Area coverage: The spatial coverage gives the fraction of sub-areas that has been visited.

The first five metrics are indicators for the connectedness of the network and the frequency of interactions among the nodes. They yield insights into the possible information flow (using opportunistic networking) among the nodes based upon pairwise contacts. The last four metrics address the geographic coverage of the nodes and provide insights on the achievable reach (over time), in terms of density of coverage and distance.

4.1 Contact analysis

For the contact analysis, we choose one week for which the trace data set shows a reasonable participant density for all five days: 20–24 April 2009. Trajectories during this week are complemented by the preceding and following weeks from a total time window of three months as described in the previous section. Table 1 shows the increase in the node density per projection method and the *number of nodes in the connectivity graph* (i.e., nodes meeting at least

one other node) per weekday. The node number is more than doubled for the Nearest Trajectory method (from 5.6 to 13.8 nodes) and about quadrupled for Frequent Location (to 24.4 nodes) and Medoid Location method (to 22.2 nodes).

	Mon	Tue	Wed	Thu	Fri	Avg.
Raw traces	9	6	5	4	4	5.6
NT proj.	13	9	21	16	10	13.8
ML proj.	19	21	27	21	23	22.2
FL proj.	21	26	27	23	25	24.4

Table 1: # nodes in connectivity graph per day and on average.

While the NT projection influences the connectivity to a lesser extent, the FL and ML projections have a similar impact in terms of node number, but improve contact number and node degree notably (Figures 2 and 3(a)). One advantage of the FL over the ML method with respect to large data processing is the faster computation of frequency than of medoid. Therefore, we choose the FL method for comparison to the raw trace results in the remainder of this paper.

4.1.1 Node degree and contact number

Figure 2 shows the *average node degree* for the trajectory projection as well as temporal and spatial extrapolation. The highest values are achieved by the FL method with temporal extrapolation. For example, with $d_t = 1h$ the node degree grows from 2.73 (raw traces) to 13.17 (FL), and without extrapolation from 2.49 (raw traces) to 4.37 (FL). The spatial extrapolation leads to smaller ascents, e.g., with d_s=1c from 2.6 (raw traces) to 7.74 (FL). The *average number of contacts* for all projection and extrapolation methods is depicted in Figure 3(a). By applying FL the number of contacts in the raw traces (221 contacts) is increased with temporal extrapolation to 580 (d_t=1h) or 907 (d_t=3h), and with spatial extrapolation to 400 (d_s=1c) or 665 (d_s=3c) contacts.

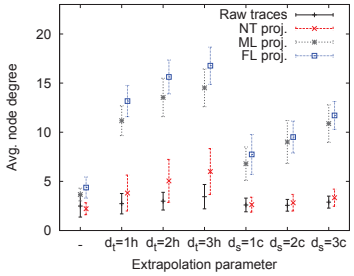

Figure 2: Average node degree (with std.dev.).

Figures 2 and, especially, Figure 3(a) also indicate how many connection opportunities (and thus connectivity) are gained over the raw traces by applying trajectory projection. Looking at the gain from adding the temporal extrapolation on top shows that people follow frequent routes (also shown in our heatmaps discussed below) but just pass at slightly different times. This suggests that more people will likely follow similar (parts of the) trajectories so that the basic idea of extrapolation via invisible users appears sensible. This also hints that adding throwbox-style devices in some places could substantially increase connectivity.

4.1.2 Inter-contact time and contact time

Figure 3(b) and 3(c) show the Complementary Cumulative Distribution Functions (CCDFs) for the inter-contact time and the contact time. Overall, we observe that our projection and extrapolation mechanism largely preserve the basic characteristics while making the traces denser. The most significant shift is shown for the spatial extrapolation, since shorter (inter-)contact times resulting from changing to neighboring cells and back are eliminated.

Looking at the inter-contact times, we observe that applying the temporal or spatial extrapolation to the raw trace does not alter the nature of the distribution, whereas combining those with the FL projection appears to yield a shift towards an exponential distribution. Understanding this aspect better requires further study and experimentation with more traces.

4.2 Spatial analysis

Finally, we investigate spatial propagation characteristics in the local area (4.2.1) and in a greater city and surrounding area (4.2.2).

4.2.1 10×10 km area analysis

Figure 4(a) and 4(b) show disconnected islands and visited cells—100×100m cell grid—during the day (8:00 to 20:00). In both cases the mean of the observed five weekdays is depicted. With respect to *disconnected islands*, the diagram confirms that extrapolation leads to faster connecting islands as one would expect (fewer islands with higher d_s). However, the figure shows that the raw traces do not exhibit the expected decrease when more trips start. It seems that the area segment connectivity does not improve during only one day due to the sparsity. The curve for FL without (w/o) extrapolation initially shows many islands, but after a peak around 10:00 those become more connected over the day.

The curves for the fraction of *visited cells* show a continuous growth over the day for all methods, but the visits during the first hour indicate already the propagation speed differences. At this daytime the visited fraction without extrapolation lies at 2.6% (raw traces) and 4.6% (FL), while with extrapolation of d_s=3c at 16.64% (raw traces) and at 41.88% (FL). The maxima achieved at the end of the day without extrapolation are 8.8% (raw traces) and 21.69% (FL), with d_s=3c 43.65% (raw traces) and 80.27% (FL). The heatmap in Figure 4(c), depicting the number of days each cell is visited in the observed week for the FL method, gives a deeper insight into the propagation results. It can be seen, that some areas are not reached by any trajectory, and only partly if one imagines extrapolation with d_s=1 to d_s=3 on the given map. Further, the map indicates that the large number of islands for FL without extrapolation stems from "cell gaps". Such gaps emerge when the most frequent locations in two consecutive time steps are not in adjacent cells.

4.2.2 Large-scale spatial analysis

In this subsection, we look at characteristics of all trajectories with the starting point in Beijing. These characteristics provide valuable insights into how far (and how fast) a message can be propagated. Figure 5(a) shows the average *flight length* along the corresponding *flight time*. The flight length allows to deduce the distance nodes travel while mobile (not staying at any location). On average 90 km are reached after one hour, while the curve shows then slightly faster ascents as only trajectories covering larger distances reach certain flight times—e.g., 318 km after three hours.

The CCDF for the *mobility range* is depicted in Figure 5(b). 9.94% of the trips show a mobility range smaller than 1 km, the majority of trips (68.14%) exhibit a range between 1 km and 10 km. The largest range values, which are achieved in 1% of the trips, lie between 450 and 980 km. The *area coverage* is exemplarily observed for an area of 200×200 km around Beijing. Figure 5(c) plots the visit frequency for grid cells of size 1×1 km. In total, 4 883 of all 40 000 cells are covered resulting in a coverage of 12.2%. The heatmap shows that the number of visits per cell varies greatly in a range of 1 to above 1 000. The most frequently visited cells are

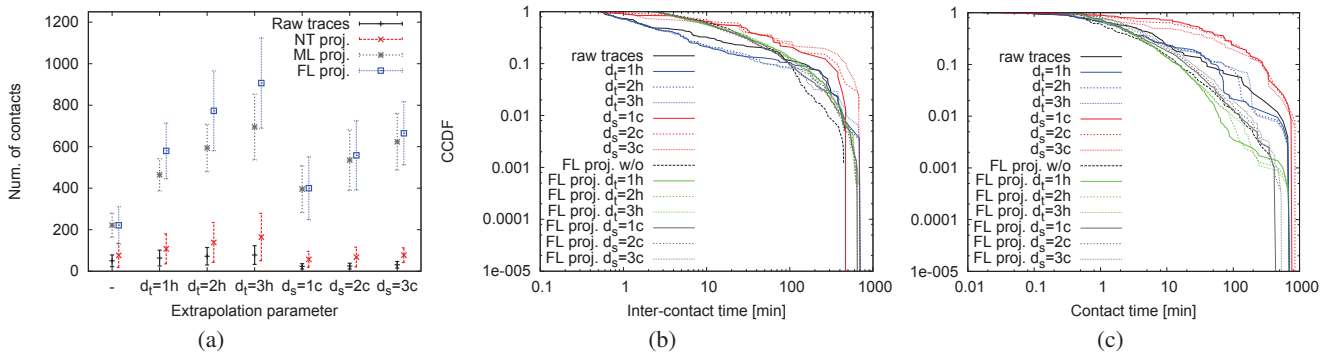

Figure 3: (a) Average number of contacts per day (with std.dev.), (b) inter-contact time CCDF, and (c) contact time CCDF.

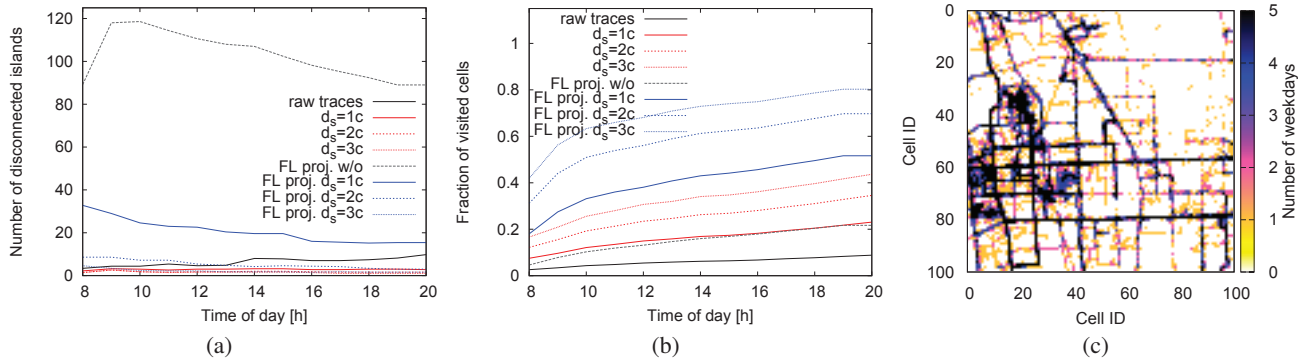

Figure 4: (a) Number of disconnected islands and (b) propagation speed during the course of the day; (c) a heatmap showing number of days each cell of the 10×10 km area is visited.

found in the Beijing city center, where 30 cells are visited more than 1 000 times. Outside Beijing the cell visits picture the course of the road and railway network. When looking at the *number of visited cells per trajectory*, we find that in about 50% of the trips the participants visit less than 10 cells, whereas less than 10% of the trips cover more than 40 cells. This corresponds to the high fraction of trips showing a mobility range of less than 10 km and supports the observation that most trips are short-range trips.

4.3 Preliminary validation

While the above results show promise, they do not capture the accuracy of our proposed methods, since we lack ground truth: how would the filled-in trajectories compare to reality? To provide at least an initial validation of the projection results, we look at the 63 trajectories existing in the data set for the evaluated week and compare them to the generated ones. First, we calculate the fraction of each trajectory's duration where visited cells match, which allows temporal and spatial validation. Table 2 shows fraction of nodes at certain cell distances for exactly the same time and with a two hour tolerance. Known and projected trajectory show good agreement for almost half of the path durations, i.e., nodes are in the same or neighboring cells. These values increase with extended time frame, indicating temporal shifts between similar paths on different days (e.g., traveling earlier or later). In 31% of the durations (or 25%, respectively), the known paths visit cells that are more than 10 cells (1km) away from the projected paths.

Second, we compute the Modified Hausdorff Distance (MHD) [4], a metric measuring the similarity between two point sets (without considering time). The CDF for MHD in Figure 6 shows a high fraction of paths with small MHD, i.e., high similarity. For example, 25.4% of the trajectories have a MHD≤2.99 Euclidean cell distance[2] and 49.3% MHD≤8.78, which also points to the regularities in the daily paths.

	Distance [cells]						
	0	1	2	3	4	5-10	>10
±0h	31.46	15.12	4.01	2.30	2.49	13.79	30.83
±2h	42.62	15.56	4.42	2.60	2.56	7.40	24.84

Table 2: Temporal/spatial matching: known vs. FL paths.

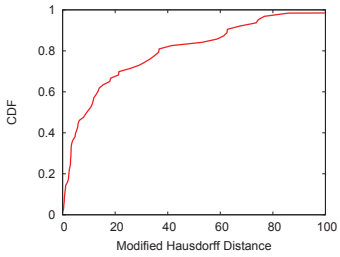

Figure 6: Spatial similarity: MHD for known vs. FL paths.

4.4 Discussion

The above evaluation of our simple mechanisms to "repair" and scale GPS traces shows some initial promise. However, the main issue is still that we do have only limited *ground truth* to compare

[2]Euclidean distance means here the distance between the center points of two cells measured in number of cells.

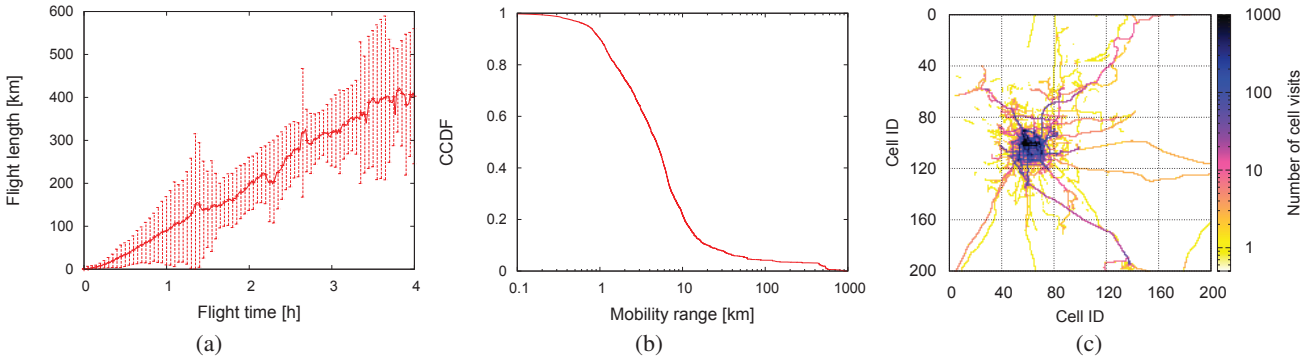

Figure 5: (a) Average flight length (with std.dev.) as a function of flight time, (b) mobility range CCDF, and (c) heatmap for visited cells in a grid with 1×1 km cell size for all trajectories starting in Beijing.

against: we simply do not know what a person really did on a given day if we are lacking trace data. Our initial evaluation in Section 4.3 is limited to some 60 trajectories and so far we explored only one GPS trace data set.

Another caveat are the fill nodes: we presently assume a uniform availability of the fill nodes across the area. Given the otherwise uneven distribution of nodes this would obviously not hold. Hence, the temporal extrapolation would need to be adapted dynamically based upon the typical node density in a given cell (at a given time). This is subject to future work.

The extrapolated traces would yield three classes of connectivity: 1) The contacts observed in the raw data plus the projections would be bidirectional as the nodes can interact directly with each other. 2) Virtual nodes supporting spatial extrapolation would yield multi-hop paths between two nodes and thus lead to increased communication latency and/or lower data rate. 3) Assistance of virtual nodes from temporal extrapolation, in contrast, would yield unidirectional contacts during which a node visiting the cell earlier would be able to send data to a node visiting later but not vice versa. Extrapolated traces would need to make those differences explicit so that they could be considered in simulations.

5. CONCLUSIONS

We have investigated two types of improvements for mobility traces: filling in data points missing for certain nodes by projections from different weeks to "repair" parts of a trace and scaling up connectivity between nodes. Both may increase the value of traces for evaluating mobile opportunistic communication scenarios. We observe quite substantial increases in the number of contacts. While the contact characteristics remain of similar nature for our scaling mechanisms, they may exhibit some changes for the projections, which requires further study. Currently, we are working on ways to better understand the accuracy of the proposed mechanisms. Both entails applying the proposed extrapolation approach to further sparse location trace data. Our future work will use the resulting traces in DTN routing protocol evaluations.

6. ACKNOWLEDGMENTS

This work has been partially supported by the European Community's Seventh Framework Programme under grant agreement no. 258414 (SCAMPI).

7. REFERENCES

[1] R. Calegari, M. Musolesi, F. Raimondi, and C. Mascolo. CTG: a connectivity trace generator for testing the performance of opportunistic mobile systems. ESEC-FSE, 2007.

[2] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Impact of Human Mobility on Opportunistic Forwarding Algorithms. *IEEE Transactions on Mobile Computing*, 6:606–620, 2007.

[3] X. Cheng, Y. Zhao, and Y. Xie. Innovation Capacity and Development Potential of Haidian District. In *Conf. on Micro Evidence on Innovation in Developing Economies*, 2009.

[4] M.-P. Dubuisson and A. Jain. A modified Hausdorff distance for object matching. In *IAPR Int. Conf. on Pattern Recognition*, 1994.

[5] S. Gambs, M.-O. Killijian, and M. N. n. del Prado Cortez. Next place prediction using mobility Markov chains. In *Workshop on Measurement, Privacy, and Mobility (MPM)*. ACM, 2012.

[6] W. Hsu, T. Spyropoulos, K. Psounis, and A. Helmy. Modeling time-variant user mobility in wireless mobile networks. In *IEEE INFOCOM*, 2007.

[7] K. A. Hummel and A. Hess. Movement activity estimation and forwarding effects for opportunistic networking based on urban mobility traces. *Wireless Communications and Mobile Computing*, 13(3):343–360, 2013.

[8] T. Karagiannis, J.-Y. L. Boudec, and M. Vojnovic. Power law and exponential decay of intercontact times between mobile devices. *IEEE Transactions on Mobile Computing*, 9(10):1377–1390, 2010.

[9] K. Lee, S. Hong, S. J. Kim, I. Rhee, and S. Chong. SLAW: A New Mobility Model for Human Walks. In *IEEE INFOCOM*, 2009.

[10] M. Lin, W.-J. Hsu, and Z. Q. Lee. Predictability of individuals' mobility with high-resolution positioning data. In *ACM UbiComp*, 2012.

[11] R. Meikle and J. Camp. A global measurement study of context-based propagation and user mobility. In *ACM HotPlanet*, 2012.

[12] Microsoft-Research. GeoLife GPS trajectory data set, version 1.3, 2012. http://research.microsoft.com/en-us/downloads/b16d359d-d164-469e-9fd4-daa38f2b2e13/.

[13] J. Ott, E. Hyytiä, P. Lassila, T. Vaegs, and J. Kangasharju. Floating Content: Information Sharing in Urban Areas. *Elsevier Personal Wireless Communications (PMC)*, 7(6):671–689, 2011.

[14] M. Piórkowski. Sampling urban mobility through on-line repositories of GPS tracks. In *ACM HotPlanet*, 2009.

[15] J. Whitbeck, M. D. de Amorim, V. Conan, M. Ammar, and E. Zegura. From encounters to plausible mobility. *Pervasive and Mobile Computing*, 7(2):206–222, 2011.

[16] H. Yoon, Y. Zheng, X. Xie, and W. Woo. Smart itinerary recommendation based on user-generated GPS trajectories. In *Int. Conf. on Ubiquitous intelligence and computing (UIC)*, 2010.

[17] W. Zhao, Y. Chen, M. Ammar, M. Corner, B. Levine, and E. Zegura. Capacity Enhancement using Throwboxes in DTNs. In *IEEE Int. Conf. on Mobile Adhoc and Sensor Systems (MASS)*, 2006.

[18] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma. Mining interesting locations and travel sequences from GPS trajectories. In *WWW*. ACM, 2009.