Chapter 8
An Exploratory Analysis of the Effects of Spatial and Temporal Scale and Transportation Mode on Anonymity in Human Mobility Trajectories

Jennifer A. Miller and Brendan Hoover

8.1 Background

Advancements in location-acquisition technologies such as global positioning systems (GPS), radio frequency identification (RFID), cellular phone networks, and WiFi hotspots have resulted in significant increases in the availability of highly accurate data on moving objects, with unprecedented high spatial and temporal resolution. These location data are often studied as 'trajectories', comprised of a series of time-stamped sequential locations. Moving objects of interest have extended beyond the traditional scale associated with people, animals, and vehicles to include weather events such as hurricanes (Dodge et al. 2012) and eye tracking, where gaze trajectories are compared with computer mouse movement to study human-computer interaction (Demšar et al. 2015).

As a result of the increasingly wide range of types of moving objects studied, several different interdisciplinary communities are now focusing on issues associated with collecting, managing, visualizing, and analyzing spatio-temporal data associated with moving objects. Originating from the relatively long history of animal tracking and telemetry studies, 'movement ecology' has become a rapidly growing subfield in ecology focused on understanding the "causes, mechanisms, and spatiotemporal patterns of (organismal) movement and their role in various ecological and evolutionary processes" (Nathan et al. 2008: 19,052). Within geographic information science (GIScience), 'computational movement analysis' has recently emerged as a subfield that focuses on the development and application of computational techniques for collecting, managing, and analyzing movement data in order to better understand the processes that are associated with them (Gudmundssen et al. 2012). 'Trajectory data mining' harnesses new computing

J. A. Miller (✉) · B. Hoover
Department of Geography and the Environment, The University of Texas at Austin, Austin, TX, USA
e-mail: jennifer.miller@austin.utexas.edu

S.-L. Shaw and D. Sui (eds.), *Human Dynamics Research in Smart and Connected Communities*, Human Dynamics in Smart Cities,
https://doi.org/10.1007/978-3-319-73247-3_8

technologies to discover knowledge from trajectory data (Zheng 2015) with applications mainly in location-based social networks, transportation systems, and urban computing.

Depending upon the technology used to collect the data, the location information can be represented by precise latitude and longitude coordinates (e.g., GPS data from a smartphone or other device) or the catchment area of a single cellular tower (e.g. call detail records (CDR) from cellular phones). These relatively low cost location data have been used to explore human mobility patterns related to, for example, urban planning (Steenbruggen et al. 2015), transportation infrastructure (Wu et al. 2013), disaster planning/evacuation strategies (Ghurye et al. 2016), potential disease spread (Oliver et al. 2015), and many other applications (see review by Becker et al. 2013). Deville et al. (2014) introduced a framework for using mobile phone data to calculate temporally explicit population data at the spatial resolution of cellular tower service areas in order to supplement census data and better understand human dynamics. Mobile phone data have also been used to represent spatiotemporal human mobility dynamics in the context of the spread of diseases such as malaria in Kenya (Wesolowski et al. 2012) and cholera in Senegal (Finger et al. 2016). The spatial resolution of CDR data varies as a function of population density. In their seminal study in an unnamed Western European country, de Montjoye et al. (2013) reported CDR catchment areas ranging from 0.15 km^2 in cities to 15 km^2 in rural areas.

Higher precision GPS data, typically collected with user- or vehicle-carried smartphones, have been used to study route choice behavior (Huang and Levinson 2015), the effects of built environment on physical activity and health (Collins et al. 2012; Carlson et al. 2015), to determine the risk of cycling injury (Strauss et al. 2015), to detect travel model (Xiao et al. 2015), and to provide social itinerary recommendations (Yoon et al. 2011). Sila-Nowicka et al. (2016) explored contextual information linked to GPS locations in order to identify "third places" beyond home and work.

8.1.1 Location Privacy

While studies using mobile phone and GPS data have made important contributions to a better understanding of human mobility and spatiotemporal dynamics in general, there are significant issues associated with the distribution or availability of these data. As often happens with technological advancements, the collection of these data has preceded extensive study on how and what they can (or should) be used for, as well as the privacy implications associated with them.

In particular, there are important privacy issues associated with location or mobility data that can be traced to a single or very few individuals that are often overlooked. Wernke et al. (2012) classify potential location privacy attacks based on the information possessed by the adversary: single or multiple locations, contextual information, and historical information. Ma et al. (2013) provide defensive approaches for scenarios in which an adversary is passive and is given individual mobility data as well as scenarios in which an adversary is active and physically collects contextual and/or ancillary data along with the mobility data. In addition to what might be considered superficially harmless privacy violations, an adversary could use location data for malicious purposes such as stalking or opportunistic criminal activities. Locations or movement patterns could also be used to make inferences that have potentially negative implications, for example, repeated visits to a medical clinic may be a concern for a prospective employer.

Measures that have been implemented to preserve privacy have been shown to be superficial or ineffective. Location data are often released after they have been 'anonymized'—which means that the trajectory has been stripped of any identifying information such as name, address, and phone number. However, as Golle and Partridge (2009) note, anonymity is "a useful but imperfect tool for preserving location privacy" (p. 390). Personal points of interest (home, work) can still be identified by mining trajectory data for movement patterns, and these points of interest are often associated with unique individuals.

Using an extensive dataset of home and work locations in the U.S., Golle and Partridge (2009) showed that at the spatial scale of a census block, the pair of home/work locations is unique for a majority of the working population. At the scale of census tracts, the pair of locations was uniquely identifying for only 5% of the working population but at the much coarser county scale, the 44% of workers who live and work in different counties are considerably more vulnerable to de-anonymization. Zang and Bolot (2011) used anonymized CDR from 25 million individuals across the U.S. to determine the "top N" locations at which calls were recorded for each of three months. They found that when N = 2 (typically corresponding to work and home locations), up to 35% of the individuals could be uniquely identified. When N = 3 (they suggested the 3rd location typically represented a school or shopping related location), 50% could be uniquely identified.

It should be emphasized here that the concept of unicity or the ability to uniquely identify a movement trajectory based on a small subset of locations of which it is comprised is not equivalent to de-anonymization, but it is a requisite first step. Quantifying the uniqueness of locations through which an individual moves is necessary to better understand privacy implications associated with increasingly available human mobility datasets.

In one of the first studies to address quantification of unicity of individual trajectories, de Montjoye et al. (2013) used fifteen months of anonymized mobile phone data (CDR) for 1.5 million individuals in a western European country and found that four randomly selected spatiotemporal points were sufficient to uniquely identify 95% of the individuals. Perhaps more troubling, they found that over 50% of individuals were uniquely identifiable from just two randomly selected locations (typically also corresponding to home and work). Song et al. (2014) found similar results with a dataset of one week of mobility data for 1.14 million people (total 56 million records): 60% of the trajectories were unique using just two random points.

Due to data availability, most of the previous work on measuring 'unicity' or the uniqueness of movement traces or trajectories has been with much coarser scaled cell phone data. However, even relatively coarse spatial resolution location data such as that associated with call detail records (CDR), where 'location' is an area defined by its proximity to a specific cell phone tower, can be used to uniquely identify an individual. Locations of cell phone towers or antennae are based on population density and the area associated with each one varies considerably. In their study in a small (unnamed) European country, de Montjoye et al. (2013) found that the reception or catchment area for an antenna ranged from 0.15 km^2 in urban areas to 15 km^2 in rural areas.

It is important to note that uniqueness does not equate to re-identifiability and the objectives of these studies were to examine how unique individual trajectories were, not to actually de-anonymize them or re-attach an individual's information to a unique trajectory. However, the ability to measure uniqueness of locations on a trajectory is an important prerequisite for re-identification (which would involve correlation with an ancillary dataset) and therefore, represents a potential threat to individual privacy.

8.1.2 Measuring Unicity

While a single widely used or standard measure of unicity has not yet emerged, unicity has been measured in different ways, depending on characteristics of the mobility dataset and research objectives. In studies similar to the research presented here, unicity has typically been quantified by comparing a randomly selected subset of points (either location or location + time) to points in the mobility trajectories; unicity would be high if the subset of points matches very few other trajectories. Unicity would be low (and location privacy less problematic) if the subset of points matched many other trajectories. de Montjoye et al. (2013) measured 'unicity' as the percentage of 2500 random traces that were unique given p random points (p ranged from 2 to 5). Song et al. (2014) defined uniqueness of trajectories as the percentage of all available trajectories that were uniquely associated with p random points, which they varied from 2 to 6. While anonymity (or lack thereof) has been studied with CDR data, as the previous examples show, it has not yet been more extensively addressed with finer spatiotemporal resolution available as GPS locations from, e.g., smartphones (but see Rossi et al. 2015). These datasets could potentially be far more unique and therefore more difficult to anonymize.

In addition to location and location + time, Rossi et al. (2015) also tested how additional movement information derived from three different published mobility datasets could be uniquely associated with individual trajectories. They calculated distance traveled, average speed, and average angle of travel for a specific time window and measured unicity as the average uniqueness over the whole dataset using 1000 subsets of points (number of points ranged from 1 to 5) per individual. They found that direction was the most unique movement parameter for all three datasets, with five points able to uniquely identify 95% of the users.

The degree of uniqueness of trajectories can vary as a function of factors such as typical commuting patterns, transportation modes, and geographical region (which also affects commuting patterns and transportation modes). While measurement precision and geographic scale have been varied in order to assess their effects on unicity of a mobility dataset, the influence of these other factors has rarely been examined. Information on transportation modes associated with trajectory segments has only recently been provided as a component of some mobility datasets, although previous research has focused on inferring likely transportation modes based on movement characteristics analyzed from the locations (Lin and Hsu 2014; Zheng et al. 2010). Sila-Nowicka and Thakuriah (2016) used travel diaries and mobility data for 358 users in Glasgow, Scotland to compare spatial patterns between the original mobility data and generalized data that resulted from kernel density estimation for four different travel modes (driving, walking, train, bus). They calculated Pearson's correlation coefficient for the original and generalized data and the values ranged from 1.0 for train to 0.852 for walking. No previous studies explicitly examined how different transportation modes affected unicity.

The importance of quantifying the 'anonymity' of a database has been a research focus in information sciences far longer than the relatively new issue of unicity and there are several widely accepted methods. The most commonly used method of k-anonymity was introduced by Sweeney (2002) as a measure to increase anonymity for non-spatial databases. When applied to spatial databases, it ensures that any set of records (locations) for an individual is at least the same as $k - 1$ individuals. Generally, $k = 2$, ensuring that at least two trajectories are equivalent, but as k increases, so too does the anonymity. Extensions of k-anonymity include l-diversity and t-closeness (Li et al. 2007).

These measures are generally used to manage trajectory datasets (e.g., data would be manipulated so that the level of anonymity reached the reported k level), but in order to quantify the actual level of anonymity of trajectory datasets, a rigorous analysis comparing random points from each trajectory to all other trajectories still needs to be conducted. With trajectory datasets now available at one second temporal resolution, the volume of these data can result in computationally intensive analysis.

In addition to removing any identifying information, 'cloaking' or other obfuscation techniques have been used to add noise to or reduce the precision of location data associated with mobility traces (Gambs et al. 2010; Ma et al. 2013). While there has been extensive research on approaches to preserve location privacy (see Kar et al. 2013; Seidl et al. 2016) most of the methods result in significant loss of information and none has been considered to be broadly successful (Narayanan and Felten 2014).

The research presented here explores issues related to privacy and identity associated with more recently available high resolution GPS location data. We quantified the unicity of GPS movement trajectories testing the effect of spatial resolution and temporal resolution. In addition to location, we explored how effective derived movement parameters such as direction could be for uniquely identifying a trajectory. We also calculated unicity for different user-labelled transportation modes and explored how it is affected by spatial resolution.

8.2 Data

We explored these issues using GeoLife Trajectories (Zheng et al. 2008a, b, 2010), a well-known mobility data of individuals in Beijing, China collected by Microsoft Asia. This is an extremely dense dataset, with temporal resolution of $\sim$1–5 s and spatial resolution of $\sim$5–10 m. For the first part of our analysis, we used only one year of data (January 2009–December 2009) and the spatial extent of Beijing (39.6°–40.2°N latitude and 116°–116.8°E longitudes) to remove users who traveled outside of the city during this time period. This resulted in 71 users who had a total of 7243 daily trajectories (number of locations visited within trajectories varied but the mean was 1600). Here we use an individual's daily trajectory (where a day is considered to begin and end at midnight) as the basic trajectory unit; each begins and ends at a time that is dependent on the individual's daily activity.

The second part of our analysis focused on how unicity varied as a function of transportation mode. Individual-labelled transportation modes (bus, subway, train, taxi, car, walk) were only available for 69 individuals in the GeoLife dataset, so we did not confine these trajectories to a certain year and therefore did not include temporal information in the unicity analysis. Transportation mode was attached to the appropriate section of an individual's trajectory and we considered trajectories to be separate if they were at least fifteen minutes apart even if they involved the same transportation mode.

8.3 Methods

The basis of our unicity test involved extracting 500 sets of points of size n ($n = 2, 3, 4$, and 5 points) from each user and counting how many other trajectories contain those points. The percentage of 500 sets of points that matched only one trajectory was calculated and this was done for each of the 71 users for the four different point sizes (n = 2, 3, 4, and 5). Our measure of unicity, u, was the percentage of 500 random points of size n that are contained in only one trajectory averaged across all 71 users. A unicity value close to 100 indicates a highly unique trajectory that could theoretically be de-anonymized, or re-connected with identifying user information more easily; a low unicity value suggests that the random set of points are contained in several different trajectories and therefore would make de-anonymizing trajectories far more challenging. The amount of information from each point was varied—we used just location (x and y), location + time (x, y, and t), and the direction (the absolute angle for point i is measured between the x direction and the step built by relocations i and i + 1).

The original latitude and longitude coordinates for these locations had a spatial precision of six decimal places ($\sim$0.1 m). In order to test how spatial and temporal resolution affected measurement of unicity, the geographic coordinates were coarsened first to four decimal places (~10 m) and the temporal resolution was coarsened to 30 s, then further coarsened to three decimal places (100 m) and 60 s. Additionally, the precision of the absolute angle measure was decreased from the original (five decimal places) to three decimal places.

For the transportation mode analysis, unicity was calculated the same way, except the n points were compared only to points labelled with the same transportation mode. The same two levels of coarsening were applied to the geographic coordinates, but temporal information was not used here in order to provide a more general location comparison. The use of different transportation modes varied widely. The 69 individuals collectively used a total of 12,291 transportation-labelled trajectory segments, but the number of labelled trajectories per individual ranged from 1 to ~2800 (i.e., some individuals had only one transportation-labelled daily trajectory, while others had several thousand). The mean use of each transportation mode per individual, along with minimum and maximum, is given in Fig. 8.1 (bars) along with the mean percentage of total trajectory sections per transportation mode (line). Walking was the most frequently reported transportation mode (52% of all trajectory segments) while train was the least frequently used (2%). Cars represented only about 8% of all trajectory segments, but the mean percentage of use across each individual's trajectory segments was 15%, with a minimum of 0% and maximum of 100% (i.e., at the extremes, some individuals never used cars for transportation, while cars were the exclusive transportation model for other individuals).

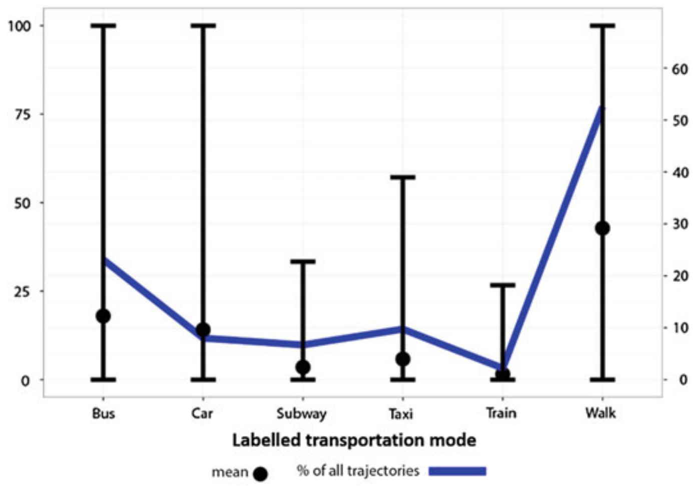

Fig. 8.1 The mean, minimum, and maximum of the percentage of each individual's labelled trajectory segments that was associated with each transportation mode (bars, left axis); and the percentage of total labelled trajectory segments associated with each transportation mode (line, right axis)

8.4 Results

The mean unicity values associated with the size of each random point set and level of coarsening for location, location + time, and direction (absolute angle) are plotted in Fig. 8.2. In general, the locations on a trajectory were highly unique. 90% of the random sets of just two points composed of only location (no timestamp) were associated with only one trajectory. Adding the timestamp increased the unicity of two points to 97%. When five points with location and timestamp were used, the unicity increased to almost 99%. The implications for location privacy are alarming, as these were randomly selected locations and not the 'most visited' that might be associated with potentially more unique work-home pairs of locations. Somewhat surprisingly, the angle of movement alone also had high unicity—when the angles of four points were tested, the unicity (u = 66%) was similar or greater than the unicity of location for CDR using two points as found in de Montjoye et al. (2013) and Song et al. (2014). Five angle values could uniquely identify a trajectory 73% of the time, although coarsening the precision of the angle measurements had a much more negative effect on unicity. More research is needed to address the unicity of derived movement parameters separated from actual locations as a potential privacy issue.

When just two points (no timestamp) were used at the first level of coarsening (spatial precision reduced tenfold to $\sim$10 m), unicity was still almost 69%; when the coarsened time ($\sim$30 s) information was added to location, the unicity was similar to the original resolution (88%).

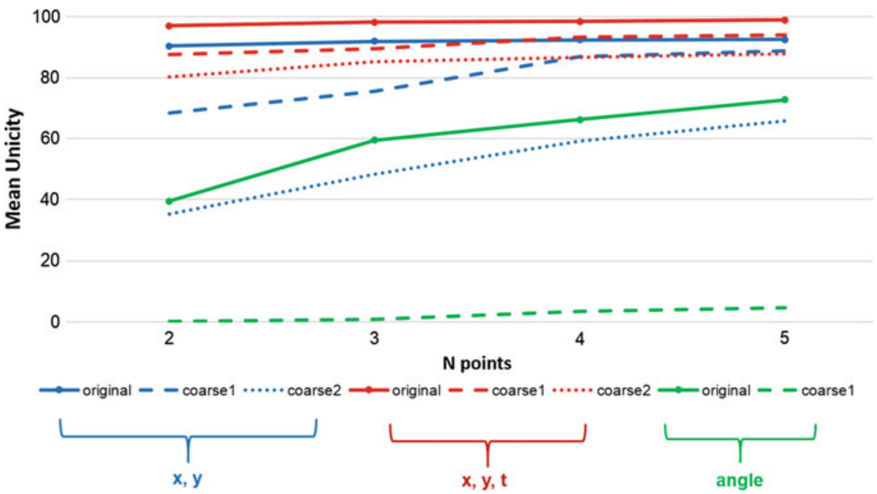

Fig. 8.2 Mean unicity for different types of information (location in blue, location + time in red, and absolute angle (direction) in green) and different measurement precision (solid line is original precision, dashed line is first level of coarsening, and dotted line is second level of coarsening). The number of sample points compared is on the x-axis

The final level of coarsening decreased the spatial resolution of an x/y pair to ~100 m and the temporal resolution was coarsened to one minute. The spatial resolution here was closer to the resolution of the antenna reception areas used in the de Montjoye et al. (2013) paper (where spatial resolution ranged from 115 m to 15 km), but the coarsened temporal resolution was still much more precise than the one used in the CDR studies. As a result, using location + time for just two points resulted in a high unicity (mean 80.3%), while five points increased the mean unicity to almost 88%. Using just location (no timestamp), the unicity degraded to 32% for two points and 66% for five points.

Using four or five points, the first level of coarsened time + location had similar unicity to the original precision location coordinates (93–94%), while coarsest time + location was similar to the first level of coarsened location coordinates (86–87%). Only one level of coarsening was used for direction, as unicity of the absolute angle degraded substantially—unicity was only 5% when a set of five points was used.

Unicity of trajectories (location only) associated with different transportation modes is plotted in Fig. 8.3. We don't include the results from trajectories labelled with train, as it was only used by 20 individuals and represented only 2% of all labelled trajectory segments. Walking had the highest unicity for all four point sample sizes (*u* ranged from 86% for two points to 93% for five points). Subway

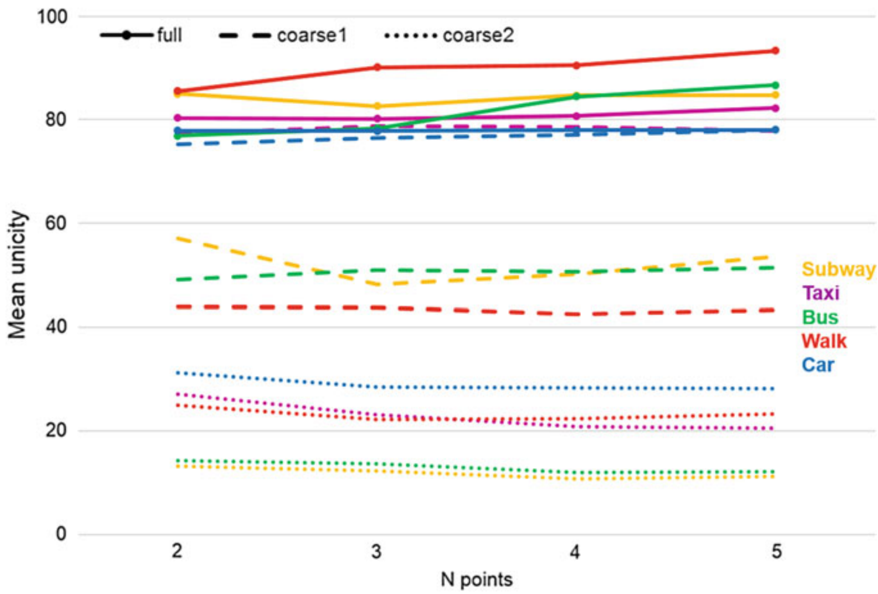

Fig. 8.3 Mean unicity for five transportation modes (subway is yellow, taxi is purple, bus is green, walk is red, and car is blue) and three measurement levels (solid line is original precision, dashed line is first level of coarsening, and dotted line is second level of coarsening). The number of sample points compared is on the x-axis

mode was the next most unique transportation mode, but had slightly higher unicity with two sample points (85%) than for three points (83%) and five points only increased unicity to 85%. Trajectory sections that involved taxis and cars had the most constant unicity across all four sample sizes, ranging from u = 80% (taxis) and 77% (cars) for two points to u = 82% (taxis) and 78% (cars) for five points. These transportation modes were also relatively less frequently used (10% of all trajectory segments involved taxis while 8% involved cars). Unicity associated with buses increased from 77% with two points to 87% with five points.

The spatial resolution of the location coordinates was coarsened to the same two levels as described above (10 and 100 m) in order to assess how unicity associated with transportation modes was affected. Unicity for taxis and cars was only slightly decreased at the first level of coarsening and unicity for both modes was fairly constant across all sample sizes. Unicity did decrease substantially for taxis and cars at the second level of coarsening, although unicity for cars remained higher than for taxis (28–30% and 27–20%, for cars and taxis respectively). Unicity also decreased marginally as the number of sample points increased for taxis and cars, although this could result from their relative scarcity in the dataset.

Unicity for subways, buses, and walking decreased markedly (u was less than 60%) with the first level of coarsening. While buses and walking had similar unicity across all sample sizes, subways had the highest unicity with just two sample points (57%). At the coarsest level, unicity dropped considerably for all five transportation modes, ranging from 31% for cars with two sample points to 12% with five points for buses and subways.

8.5 Conclusion

With the much higher precision and spatial resolution of GPS data currently available, two x/y locations are sufficient to be uniquely associated with a single trajectory 90% of the time, adding the timestamp matches a single trajectory 97% of the time. The three pieces of information—location + time—are so specific that increasing the number of points to match to five increases the unicity very little because it is already so high using just two points. The first level of coarsening for location + time ($\sim$10 m spatial, 30 s temporal) has similar unicity to the original resolution for just location coordinates, and when four or five points are used, the coarsened location + time has slightly higher mean unicity. The location coordinates (no timestamp) show the greatest increase in unicity when additional points are used for matching. This suggests that there is a trade-off between location resolution and amount of information (location points) available.

The relatively unique signature of derived movement information alone highlights potential location privacy issues even when location information (x/y) have been removed from the dataset. Movement parameters such as speed, angle, and step length have rarely been tested as potential identifiers of trajectories, but the case study here focusing on absolute angle highlights their potential importance.

Five absolute angle data points were uniquely associated with a single trajectory 72% of the time. This suggests that individual movement, irrespective of absolute geographic location, can be identifiable with a sufficient level of precision of angle measurements and data points. Future work should focus specifically on how movement parameters could be used singly or together to identify a trajectory.

It was not surprising that unicity varies with transportation mode. Among the five transportation modes used here, walking is the least constricted while subway is likely the most constricted. Consequently, walking was the most unique travel mode at the original precision, although unicity was greatly decreased as precision decreased. This is particularly disconcerting given that walking individuals are perhaps the most vulnerable in the broadest sense of potential privacy attacks. All five transportation modes had unicity values of 78% or higher at original precision for as few as two points. The effect of transportation mode and unicity would also likely vary based on spatial characteristics of a city (ex., extent of sprawl), population characteristics of a city, transportation infrastructure, climate and other factors. Unicity would also be expected to be dependent on the amount of time an individual spends on a public transportation mode—shorter routes would presumably have lower unicity while longer routes might have higher unicity (with exceptions related to well-traveled but more distant destinations such as an airport). The unicity associated with taxi, car, bus, and subway transportation modes was similar at the original resolution, although there were far more occurrences of bus travel in the data. However, at the first level of coarsening, taxi and car mode continued to have high unicity, even higher than walking. This suggests that the spatial scale at which roads are resolved contributes to higher unicity. More research is needed to better understand how unicity varies with transportation mode based on user-provided travel information.

This unicity study has particularly important implications for privacy and the increasing availability of 'anonymized' trajectory datasets. This is one of the few studies to explore unicity with higher resolution GPS data and it should be troubling how unique a set of two location points can be. Coarsening the spatial and temporal resolution reduces the unicity, but five points with x, y coordinates at the coarsest resolution tested here were still uniquely associated with a single trajectory more than 60% of the time. Our results also show an increase in unicity when more than two points are used, presumably the "third place" (see Sila-Nowicka et al. 2016) after points representing e.g., home and work can be an important determinant of unicity. This effect persists with scale coarsening and when just angle of movement is used. We detected a few instances where trajectories were duplicated in this dataset, which could result in conservative estimates of unicity. Conversely, the calculation of unicity will be dependent upon how many individual trajectories are used in the comparison—as we use trajectories for ~ 70 individuals, unicity could potentially be overstated here. More research is needed on how calculation of unicity is affected by dataset characteristics like number of individuals and trajectories, geographic region, infrastructure, and social factors that affect movement patterns.

It is important to note here that the focus of this study was not to re-attach user information to trajectories, it was just to examine how unique trajectories were based on different types and levels of information stored with or derived from them. Problems with anonymizing published mobility datasets have already been highlighted using the relatively coarse spatial resolution of call detail records. These privacy issues will only be exacerbated as higher quality GPS location datasets become increasingly available. The dataset used here represents trajectories for 71 and 69 individuals (transportation mode data), which may be too small to make inferences about unicity. GeoLife was not originally collected for studies focused on location privacy, although it has been used in similar studies (Rossi et al. 2015).

In addition to measuring unicity of location and location + time, further study is needed on how different factors such as transportation mode and movement parameters affect unicity as well as potential implications involving linked datasets from social surveys. While each of the issues addressed here focuses on a single dataset for the case study, we would expect the results to be broadly applicable to other similar mobility datasets.

Acknowledgments This research was funded in part by the Center for Identity at The University of Texas.

References

Becker, R., Cáceres, R., Hanson, K., Isaacman, S., Loh, J. M., Martonosi, M., et al. (2013). Human mobility characterization from cellular network data. *Communications of the ACM, 56,* 74–82. https://doi.org/10.1145/2398356.2398375.

Carlson, J. A., Saelens, B. E., Kerr, J., Schipperijn, J., Conway, T. L., Frank, L. D., et al. (2015). Association between neighborhood walkability and GPS-measured walking, bicycling and vehicle time in adolescents. *Health & Place, 32,* 1–7. https://doi.org/10.1016/j.healthplace.2014.12.008.

Collins, P., Al-Nakeeb, Y., Nevill, A., & Lyons, M. (2012). The impact of the built environment on young people's physical activity patterns: A suburban-rural comparison using GPS. *International Journal of Environmental Research and Public Health, 9,* 3030–3050. https://doi.org/10.3390/ijerph9093030.

Demšar, U., Buchin, K., Cagnacci, F., Safi, K., Speckmann, B., Van de Weghe, N., et al. (2015). Analysis and visualisation of movement: An interdisciplinary review. *Movement Ecology, 3,* 5. https://doi.org/10.1186/s40462-015-0032-y.

de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports, 3,* 1376. https://doi.org/10.1038/srep01376.

Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., Gaughan, A. E., et al. (2014). Dynamic population mapping using mobile phone data. *PNAS, 111,* 15888–15893. https://doi.org/10.1073/pnas.1408439111.

Dodge, S., Laube, P., & Weibel, R. (2012). Movement similarity assessment using symbolic representation of trajectories. *International Journal of Geographical Information Science, 26,* 1563–1588. https://doi.org/10.1080/13658816.2011.630003.

Finger, F., Genolet, T., Mari, L., de Magny, G. C., Manga, N. M., Rinaldo, A., et al. (2016). Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks. *Proc Natl Acad Sci U S A, 113,* 6421–6426. https://doi.org/10.1073/pnas.1522305113.

Gambs, S., Killijian, M.-O., del Prado Cortez, M. (2014). De-anonymization attack on geolocated data. *Journal of Computer and System Sciences, Special Issue on Theory and Applications in Parallel and Distributed Computing Systems, 80,* 1597–1614. https://doi.org/10.1016/j.jcss.2014.04.024.

Gambs, S., Killijian, M.O., & del Prado Cortez, M.N. (2010). November. Show me how you move and I will tell you who you are. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Security and Privacy in GIS and LBS* (pp. 34–41). ACM.

Ghurye, J., Krings, G., & Frias-Martinez, V. (2016, June). A framework to model human behavior at large scale during natural disasters. In *2016 17th IEEE International Conference on Mobile Data Management (MDM)*, (Vol. 1, pp. 18–27). USA: IEEE.

Golle, P., & Partridge, K. (2009). On the Anonymity of Home/Work Location Pairs. In *Presented at the Proceedings of the 7th International Conference on Pervasive Computing* (pp. 390–397), Berlin: Springer. https://doi.org/10.1007/978-3-642-01516-8_26.

Gudmundsson, J., Laube, P., & Wolle, T. (2012). Computational movement analysis. In W. Kresse & D. M. Danko (Eds.), *Springer handbook of geographic information* (pp. 423–438). Berlin, Heidelberg: Springer.

Huang, A., & Levinson, D. (2015). Axis of travel: Modeling non-work destination choice with GPS data. *Transportation Research Part C: Emerging Technologies, Big Data in Transportation and Traffic Engineering 58, Part B*, 208–223. https://doi.org/10.1016/j.trc.2015.03.022.

Kar, B., Crowsey, R. C., & Zale, J. J. (2013). The myth of location privacy in the United States: Surveyed attitude versus current practices. *The Professional Geographer, 65,* 47–64. https://doi.org/10.1080/00330124.2012.658725.

Li, N., Li, T., & Venkatasubramanian, S. (2007). t-closeness: Privacy Beyond k-Anonymity and l-Diversity, In *2007 IEEE 23rd International Conference on Data Engineering. Presented at the 2007 IEEE 23rd International Conference on Data Engineering* (pp. 106–115). https://doi.org/10.1109/icde.2007.367856.

Lin, M., & Hsu, W.-J. (2014). Mining GPS data for mobility patterns: A survey. *Pervasive and Mobile Computing, 12,* 1–16. https://doi.org/10.1016/j.pmcj.2013.06.005.

Ma, C. Y. T., Yau, D. K. Y., Yip, N. K., & Rao, N. S. V. (2013). Privacy vulnerability of published anonymous mobility traces. *IEEE/ACM Transactions on Networking, 21,* 720–733. https://doi.org/10.1109/TNET.2012.2208983.

Narayanan, A., & Felten, E. W. (2014). No silver bullet: De-identification still doesn't work. *White Paper*, July.

Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., et al. (2008). A movement ecology paradigm for unifying organismal movement research. *PNAS, 105,* 19052–19059. https://doi.org/10.1073/pnas.0800375105.

Oliver, N., Matic, A., & Frias-Martinez, E. (2015). Mobile network data for public health: Opportunities and challenges. *Frontiers in Public Health, 3*.

Rossi, L., Walker, J., & Musolesi, M. (2015). Spatio-temporal techniques for user identification by means of GPS mobility data. *EPJ Data Science, 4,* 11. https://doi.org/10.1140/epjds/s13688-015-0049-x.

Seidl, D. E., Jankowski, P., & Tsou, M.-H. (2016). Privacy and spatial pattern preservation in masked GPS trajectory data. *International Journal of Geographical Information Science, 30,* 785–800. https://doi.org/10.1080/13658816.2015.1101767.

Sila-Nowicka, K., & Thakuriah, P. (2016). The trade-off between privacy and geographic data resolution. A case of GPS trajectories combined with the social survey results [WWW Document]. *ISPRS—International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. URL http://www.isprs.org/publications/archives.aspx (accessed 10.16.16).

Siła-Nowicka, K., Vandrol, J., Oshan, T., Long, J. A., Demšar, U., & Fotheringham, A. S. (2016). Analysis of human mobility patterns from GPS trajectories and contextual information. *International Journal of Geographical Information Science, 30*(5), 881–906.

Song, Y., Dahlmeier, D., & Bressan, S. (2014). Not so unique in the crowd: A simple and effective algorithm for anonymizing location data. *ResearchGate, 1225,* 19–24.

Steenbruggen, J., Tranos, E., & Nijkamp, P. (2015). Data from mobile phone operators: A tool for smarter cities? *Telecommunications Policy, 39*(3), 335–346.

Strauss, J., Miranda-Moreno, L. F., & Morency, P. (2015). Mapping cyclist activity and injury risk in a network combining smartphone GPS data and bicycle counts. *Accident Analysis and Prevention, 83,* 132–142. https://doi.org/10.1016/j.aap.2015.07.014.

Sweeney, L. (2002). K-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10,* 557–570. https://doi.org/10.1142/S0218488502001648.

Wernke, M., Skvortsov, P., Dürr, F., & Rothermel, K. (2012). A classification of location privacy attacks and approaches. *Personal and Ubiquitous Computing, 18,* 163–175. https://doi.org/10.1007/s00779-012-0633-z.

Wesolowski, A., Eagle, N., Tatem, A. J., Smith, D. L., Noor, A. M., Snow, R. W., et al. (2012). Quantifying the impact of human mobility on malaria. *Science, 338,* 267–270. https://doi.org/10.1126/science.1223467.

Wu, W., Cheu, E. Y., Feng, Y., Le, D. N., Yap, G. E., & Li, X. (2013). Studying intercity travels and traffic using cellular network data. Conference on *Mobile Phone Data for Development: Net Mob 2013*. Retrieved from http://perso.uclouvain.be/vincent.blondel/netmob/2013/D4D-book.pdf

Xiao, G., Juan, Z., & Zhang, C. (2015). Travel mode detection based on GPS track data and Bayesian networks. *Computers, Environment and Urban Systems, 54,* 14–22. https://doi.org/10.1016/j.compenvurbsys.2015.05.005.

Yoon, H., Zheng, Y., Xie, X., & Woo, W. (2011). Social itinerary recommendation from user-generated digital trails. *Personal and Ubiquitous Computing, 16,* 469–484. https://doi.org/10.1007/s00779-011-0419-8.

Zang, H., & Bolot, J. (2011). Anonymization of location data does not work: A large-scale measurement study. In *Presented at the Proceedings of the 17th annual international conference on Mobile computing and networking*, (pp. 145–156). USA: ACM. https://doi.org/10.1145/2030613.2030630.

Zheng, Y. (2015). Trajectory data mining: An overview. *ACM Transactions on Intelligent Systems and Technology, 6*(29):1–29:41. https://doi.org/10.1145/2743025.

Zheng, Y., Li, Q., Chen, Y., Xie, X., & Ma, W. -Y. (2008a). Understanding mobility based on GPS data. In *Presented at the Proceedings of the 10th international conference on Ubiquitous computing* (pp. 312–321). USA: ACM. https://doi.org/10.1145/1409635.1409677.

Zheng, Y., Liu, L., Wang, L., & Xie, X. (2008b). Learning transportation mode from raw Gps data for geographic applications on the web. In *Proceedings of the 17th International Conference on World Wide Web, WWW '08*, (pp. 247–256). New York, NY, USA: ACM. https://doi.org/10.1145/1367497.1367532.

Zheng, Y., Xie, X., & Ma, W.Y. (2010). Geolife: A collaborative social networking service among user, location and trajectory. *IEEE Data Eng. Bull., 33*(2), 32–39.