



OPEN

# Modelling imperfect knowledge via location semantics for realistic privacy risks estimation in trajectory data

Stefano Bennati<sup>✉</sup> & Aleksandra Kovacevic

Mobility patterns of vehicles and people provide powerful data sources for location-based services such as fleet optimization and traffic flow analysis. Location-based service providers must balance the value they extract from trajectory data with protecting the privacy of the individuals behind those trajectories. Reaching this goal requires measuring accurately the values of utility and privacy. Current measurement approaches assume adversaries with perfect knowledge, thus overestimate the privacy risk. To address this issue, we introduce a model of an adversary with imperfect knowledge about the target. The model is based on *equivalence areas*, spatio-temporal regions with a semantic meaning, e.g. the target's home, whose size and accuracy determine the skill of the adversary. We then derive the standard privacy metrics of *k-anonymity*, *l-diversity* and *t-closeness* from the definition of equivalence areas. These metrics can be computed on any dataset, irrespective of whether and what kind of anonymization has been applied to it. This work is of high relevance to all service providers acting as processors of trajectory data who want to manage privacy risks and optimize the privacy vs. utility trade-off of their services.

Trajectories of people and vehicles provide valuable insights into mobility patterns and behaviors that are key resources for location-based services (LBS) such as traffic optimization or fleet management. At the same time, the same insights pose a privacy risk for individuals, as the spatio-temporal nature of location data reveals an additional semantic context about them (e.g. analysis of pick-up and drop-off location and time stamp might reveal home or work addresses). Additionally, location can be linked to large amounts of publicly available data such as phone books or social media; this renders location data especially privacy-sensitive. Therefore, privacy-aware LBS providers must carefully balance the trade-off between the utility value (e.g. accuracy and freshness) and privacy value (e.g. re-identification risk) associated to location data.

The utility of data depends on the characteristics of the LBS to be provided, the service provider has therefore all the information to estimate utility. On the contrary, privacy risks depend as well on the knowledge about one or more target individuals available to an adversary<sup>1</sup>; therefore, the service provider can estimate privacy risks only after defining a model of the adversary's knowledge regarding a specific target.

The most common model in the literature is an adversary with perfect knowledge. Real-world adversaries cannot achieve perfect knowledge because of (i) errors in associating the knowledge about a target with their data, e.g. due to noise in GPS measurements and (ii) availability of additional information, e.g. accessing protected data like medical records.

The aim of this paper is therefore to argue for the replacement of the perfect knowledge model, as it overestimates the privacy risk and leads to over-aggressive anonymization, in favor of adversaries with imperfect knowledge (or *approximate locators*<sup>2</sup>), which are more accurate at re-identifying some locations than others depending on their semantic context.

We distinguish location data in two broad categories: stay-points, e.g. trip origin and destination, and waypoints, e.g. transit on a highway. Staypoints are more likely to be associated to a semantic meaning and purpose than waypoints are, e.g. frequent overnight stays at a specific address or regular visits to a worship center.

The semantic meaning of location is what leads to re-identification and inference of sensitive information; therefore, the paper focuses on the estimation of privacy risks for stay-points<sup>3</sup>. This real-world example demonstrates why the lack of semantic meaning makes re-identification unlikely: Consider the task to re-identify a driver from a waypoint, for example to deliver a fine for an illegal maneuver. First, the maneuver must be

HERE Technologies, Privacy Office, Zürich, Switzerland. ✉email: stefano.bennati@here.com

observed, e.g. by a police patrol. The likelihood of observation is low to begin with, as it is not possible to predict the time and location of an illegal maneuver and therefore to dispatch a patrol to observe the event. Assuming the event is observed, re-identification would require to infer additional information about the driver of the vehicle. This information is either gained directly, e.g. by reading the license plate of the vehicle, or through analysis of the trajectory dataset, e.g. commute patterns can reveal who the driver is. Given that we aim to estimate the privacy risk associated with location data, we only consider the second situation. The next task is to match the observation to the correct data, which is prone to error as both the data and the observation might contain noise in either the spatial and temporal component. Combining the likelihood of successfully completing these steps demonstrates that the likelihood of re-identifying an individual from a waypoint is negligible. A counter-example are waypoints that contain semantic information, e.g. driving patterns, that allows to associate a trajectory to a specific individual. A discussion on how to model such situations in the proposed framework is presented in section “Discussion”.

The success chances of an adversary to re-identify a stay-point are characterized by *equivalence areas*, spatio-temporal regions where the adversary is unable to distinguish between trajectories. The size and shape of equivalence areas reflect the semantic meaning of locations, which might depend on: (i) Time of day. (ii) Information about the map such as type and location of POIs, urban function (residential vs commercial), population density. (iii) The road network that allows access to a location. (iv) Relevant locations for a target, such as favorite places. By properly constructing equivalence areas, different adversaries can be modeled: for example, the NYC taxi re-identification study<sup>4</sup> could be modeled by placing equivalence areas at the location and time of paparazzi photos. In the event where each data point is contained in its own equivalence area, we obtain the perfect knowledge scenario.

The advantages of our approach are: (i) A natural way of modeling imprecision in the data. (ii) Flexibility to model realistic adversaries with different knowledge and precision by changing the approach to constructing equivalence areas. This flexibility allows the practitioner to better tailor the choice of anonymization to the effective privacy risk, thereby achieving a more desirable trade-off between privacy and utility. (iii) Independence on the anonymization algorithm and on whether the data has been anonymized or not. (iv) Compatibility with previous work which relies on a spatio-temporal subdivision of the map, e.g.<sup>5,6</sup>. (v) From equivalence areas we can derive metrics of *k-anonymity*, *l-diversity* and *t-closeness*.

We apply this model to the use-case of pick-up drop-off (PUDO) analysis of trajectory data, which considers the relation between the origin and destination of trajectories. We run an analysis on the NYC TCL Trip Record Data<sup>7</sup> and compute the privacy risks associated with adversaries with different abilities. The risk is computed using standard privacy metrics of *k-anonymity*, *l-diversity* and *t-closeness* which are commonly used with tabular data<sup>8–10</sup> and location data<sup>11–13</sup>; we believe an advantage of using standard metrics over introducing new scenario-specific metrics is the ability to compare risks across use-cases.

To the best of our knowledge, this is the first work to provide a consistent framework to define all these standard metrics for trajectory data, both for non-anonymized and anonymized data. Measuring privacy risks before applying anonymization enables identifying weak, privacy revealing subsets of the data were to apply targeted anonymization. This in turn allows to further optimize the privacy vs. utility trade-off.

## Results

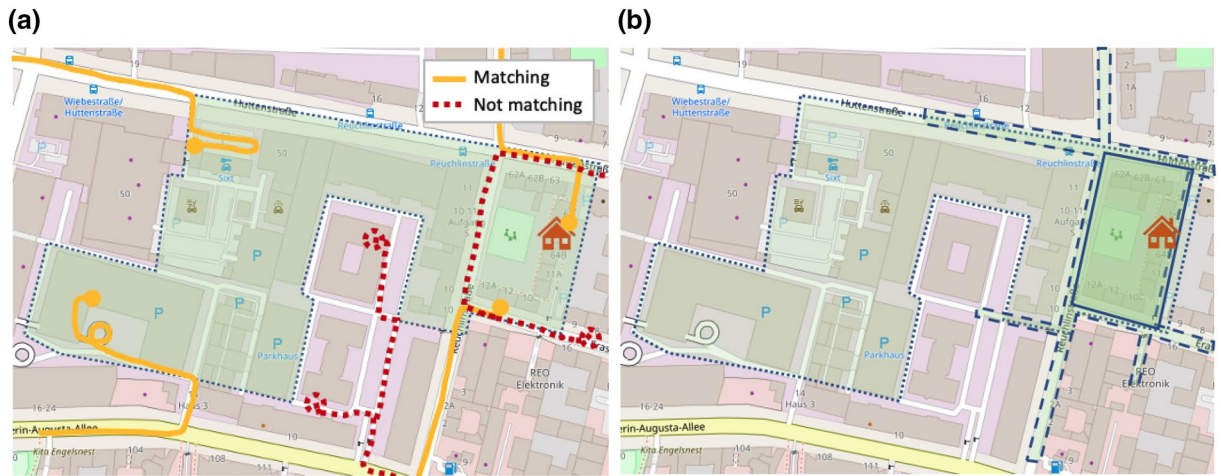
**Definition of equivalence areas.** Estimating the privacy risk associated to location data requires understanding its semantic meaning. We define semantic meanings as polygons in space and time called equivalence areas, such that spatio-temporal points falling into the same equivalence area share the same semantic meaning. Equivalence areas act as *quasi-identifiers*, i.e. attributes whose values can be linked to external information to re-identify the target<sup>1</sup>, such as background knowledge about the target’s habits or behavioral patterns.

At the same time, equivalence areas enable the adversary to infer additional personal information about a target, thus act as *sensitive attributes*. For example, consider commuting from home to work: one specific adversary, e.g. a neighbor, might know the home address and be interested in deducing the work address from the data; another adversary, e.g. a co-worker, might know the work address and be interested to obtaining the home address from the data. In both cases, matching the trajectory data with one equivalence area allows determining the second equivalence area and with it infer personal information about the target. In the remainder of this paper, we assume that the adversary has background knowledge about the origin of a trip and seeks to re-identify trajectories by inferring their destination. This assumption primarily simplifies communication about the results, as described in section “Discussion” the methodology outlined in the following can just as well be used to model the case where the adversary tries to infer origins.

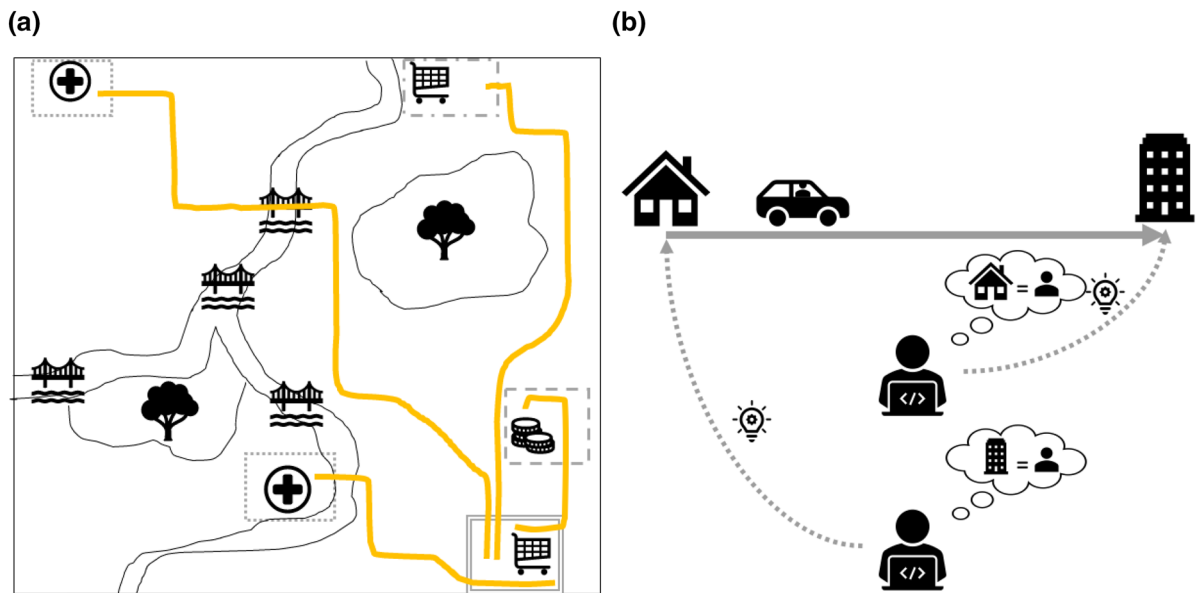
Formally, a *trajectory dataset* is a set of trajectories,  $T = \{\tau_1, \dots, \tau_n\}$ , where a trajectory  $\tau = \{p_1, \dots, p_m\}$  is an ordered set of tuples, or *records*; each trajectory is associated to a pseudonym  $ID_\tau$  that allows to identify all records in that trajectory, e.g. a Vehicle Identifier Number (VIN) or a taxi medallion number. Each record is defined as  $p_i = \langle t_i, s_i, ID_\tau, y_i \rangle$ , where  $t_i$  is the time component, i.e. a timestamp,  $s_i$  is the spatial component, i.e. a set of coordinates,  $ID_\tau$  is the pseudonym of trajectory  $\tau$  to which  $p_i$  belongs, and  $y_i$  is a set of additional attributes of the record, e.g. speed or taxi fare.

We assume that  $y_i$  and  $ID_\tau$  do not contain any personal information about the individual who created the trajectory, nor any information that allows *direct* re-identification or singling out the individual. An ID such as the taxi medallion number allows to directly link the trajectory to a specific taxi driver who owns the medallion; therefore, we assume IDs are pseudonymized, e.g. replaced by cryptographic hashes.

An *equivalence area*  $A$  is one or more (disjoint) polyhedrons in space and time. We denote with  $p_i \in A$  the notion that the coordinates  $t_i$  and  $s_i$  of a record  $p_i$  are contained in the polyhedron (or union of polyhedrons)  $A$ . For simplicity, we assume that the origin of a trajectory acts as a quasi-identifier and the destination acts



**Figure 1.** Illustrations of equivalence areas. Map source: OpenStreetMap. (a) Illustration of trajectories that match (continuous line) and that do not match (dotted line) with the equivalence area (green area). A trajectory matches an area if its starting point (circle) is inside the area. (b) Illustration of equivalence areas obtained by differently skilled adversaries (the temporal aspect of the equivalence areas is not represented): A relatively inaccurate adversary might know the target address but not know their parking habits, thus would consider a large area covering multiple parking spots where the target might likely park (dotted line); a more accurate adversary could know that the target typically parks on the streets and thus define a smaller area that covers nearby roads (dashed line); a skilled adversary might know that the street where the target lives has reserved parking spots for residents, thus, can reduce the area down to a block (continuous line).



**Figure 2.** Illustration of inference on trajectory data and equivalence areas. (a) Illustration of a set of inferences for an equivalence area (square with double continuous border). The temporal aspect of the equivalence area is not represented. Four trajectories match with the considered equivalence area, one of them ends in an equivalence area containing a shop (dash-dotted border), one in an equivalence area defined by a bank (dashed border) and two in an equivalence area related to hospitals that are spatially separated but semantically equivalent (dotted border). These three areas form the set of inferences. (b) Illustration of how location data can take either the role of sensitive attribute or of quasi-identifier, depending on the knowledge of the adversary: the same trajectory could reveal the target’s office (sensitive attribute) to an adversary that knows the target’s home (quasi-identifier), or vice versa.

as a sensitive attribute, we thus denote the first and last point of a trajectory as  $\tau^{QI}$  and  $\tau^{SA}$  respectively. The set of trajectories that match with an equivalence area  $A$ , i.e. whose quasi-identifier is in the area, is defined as  $M_A = \{\tau \in T : \tau^{QI} \in A\}$  (see Fig. 1a), where  $k_A = |M_A|$  is the number of trajectories starting in  $A$ .

The set of inferences, i.e. the set of distinct equivalence areas that match with the sensitive attributes of trajectories in  $M_A$ , is defined as  $I_A = \{A' : \tau^{SA} \in A', \tau \in M_A\}$  (see Fig. 2a), where  $l_A = |I_A|$  is the size of the set of

inferences. Each of those equivalence areas is associated with some semantic information, e.g. type of nearby POIs, which can reveal sensitive information about the target.

**Model of adversary knowledge.** We consider adversaries with imperfect knowledge, also called *approximate locators*<sup>2</sup>, whose ability to re-identify location data varies for each location. A typical example is a journalist who is investigating privacy issues in the data, relying on public data such as the address book to obtain a target's home address; the journalist does not have a physical presence that would allow them to tail a target in the real world as they drive.

We model the adversary's ability and background knowledge as the size of equivalence areas, where a more skilled adversary would be able to produce smaller, more accurate equivalence areas. For example, in a re-identification process, the goal of an adversary is to identify which trajectories in the data are associated to the target with high probability. The skill of an adversary can be defined as the precision in identifying these trajectories, which might depend on the availability and the accuracy of background knowledge about a target, e.g. commute habits. This background knowledge allows the adversary to define one or more areas which delimit what the adversary believes to be likely associated to the target, e.g. knowing the commute habits of a target, the adversary can limit the analysis to all trajectories that leave the environs of a target's home within a specific time frame. A skilled adversary might re-identify all trajectories starting from the street where the target lives and within a 30-minutes time-window of the target's departure, while a less skilled adversary might re-identify all trajectories starting from multiple parking spots within the target's neighborhood (see Fig. 1b) and within a 1 hour time-window from the target's departure. The maximum accuracy for an equivalence area depends on what resources are available to an adversary, e.g. access to information about POI and population density, as well as the intrinsic noise in GPS data. For example, an adversary might pinpoint a target to a specific building but might not be able to distinguish between different semantic meanings associated to the same, e.g. bar vs apartment. An equivalence area with an uncertain semantic meaning models probabilistic inference, e.g. the adversary might infer that the trip's purpose is more likely being 'visiting a friend' than 'going to a bar' except on a Friday night. Our analysis does not consider the extreme case where the adversary has such precise knowledge that each data point, including waypoints, falls in its own equivalence area, as this scenario is already considered in previous work.

**Experimental evaluation.** This section presents analysis that we designed to showcase the use of equivalence areas for privacy risk measurement in a real-world use case. The definitions of the privacy metrics of k-anonymity, l-diversity and t-closeness can be found in Section Methods 5.

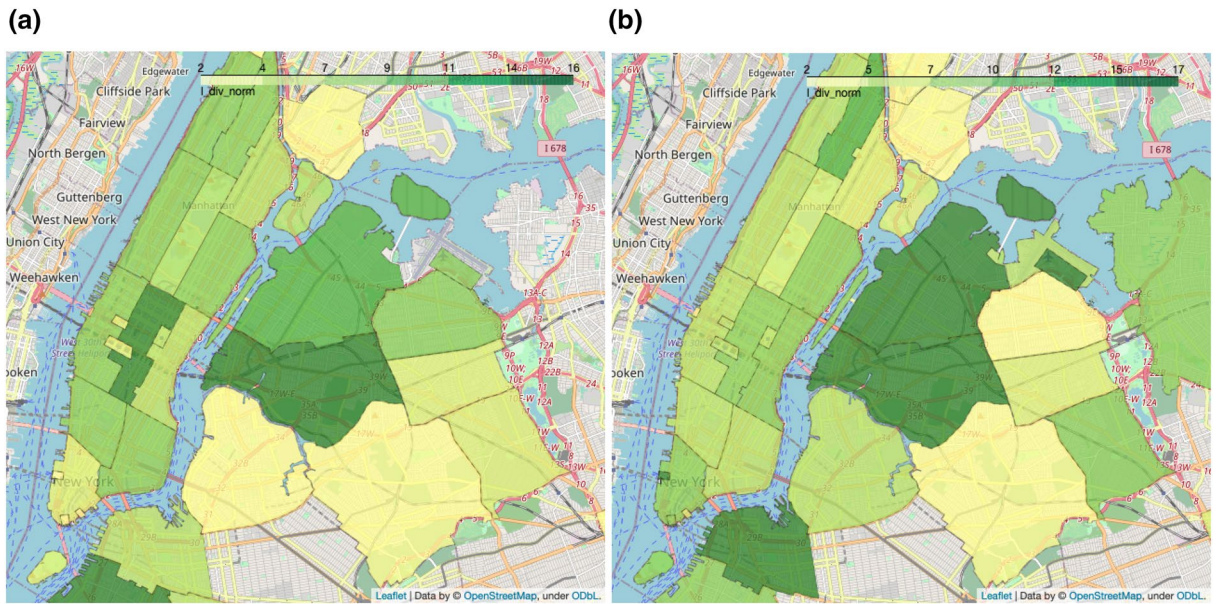
The analysis is based on the NYC TCL Trip Record Data<sup>7</sup>, containing information about the location and time of pick-up and drop-off of yellow cabs in the NYC area. This public dataset contains trajectories with only two stay-points: pick up and drop off, so it might not be representative of other applications like parking availability estimation. We take a subset of data containing the trips that start on Monday 5th of January 2009, although they might end on the next day; around 350,000 trips in total. Additionally, we clean the dataset by removing trips that last less than a minute or with invalid coordinates, i.e. outside the bounding box defined by the coordinates 71° 40' W - 41° 30' N and 70° 45' W - 40° 18' N.

We base our definition of equivalence areas on the NYC Community Districts boundaries<sup>14</sup>, which we enhance by manually drawing additional equivalence areas around the main public transport hubs (airports, train stations, ferry docks). Defining large equivalence areas would likely underestimate the risk caused by a reasonably-skilled adversary, yet community districts have a specific character (business VS residential districts) that allows us to reason about the purpose of trips. For the purpose of risk estimation, we recommend defining smaller equivalence areas that reflect the abilities of a realistic adversary.

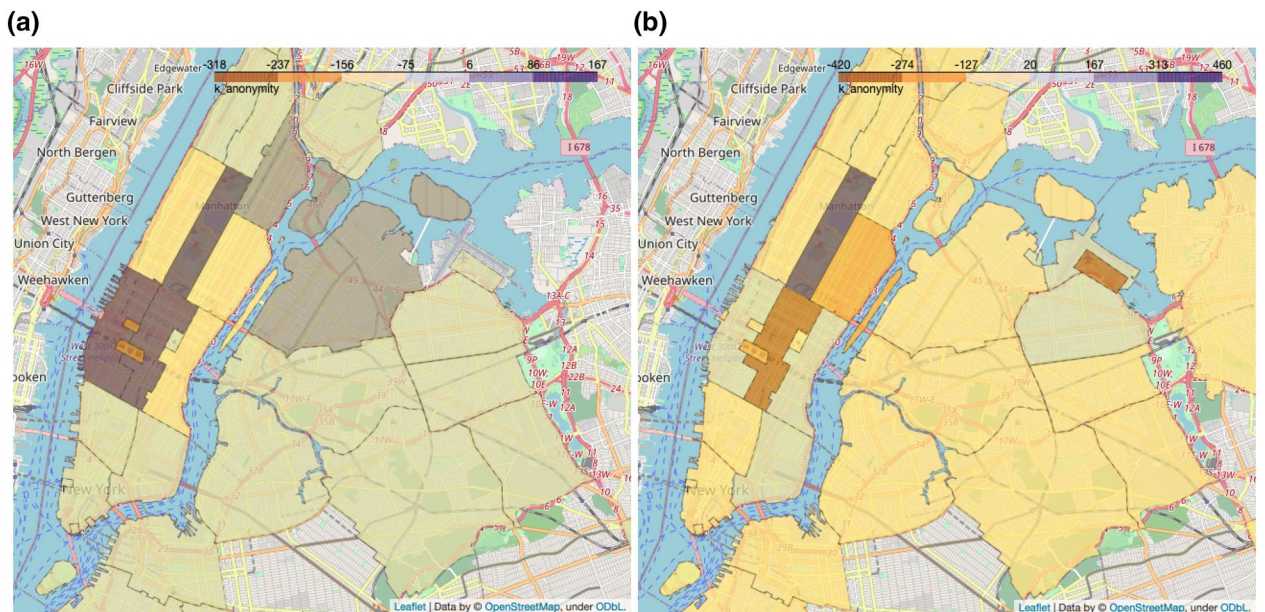
The first analysis looks at the re-identification risk during the morning rush-hour (between 7 and 7:30, for a total of 4,000 trips) and the evening rush-hour (between 17 and 17:30, for a total of 9,300 trips), which are illustrated respectively in Fig. 3a and b. The goal of these figures is to highlight differences in the scores caused by different characteristics of the data, for example different commute patterns; for this reason, l-diversity is chosen as it captures the flux of trips from one origin to a destination. When computing this score, we remove trajectories that start and end in the same equivalence areas, which are especially common in Manhattan, as they do not relate to commute patterns. Lower l-diversity scores mean that fewer destinations are reached by trips starting in that area.

Data shows that a typical commuter pattern is one where taxis transfer workers from the main transit hubs to their offices in the morning and back in the evening. Figure 3a shows that the l-diversity score in the main public transit hubs (Penn Station, Grand Central station, Port Authority bus station) is lower than the surrounding cells, which suggests that most taxi trips that start at these stations end in a few areas, e.g. business districts, while trips starting in the business districts of Manhattan, which might represent business trips, have a more diverse set of destinations. Figure 3b shows an opposite pattern in the evenings, where a lower l-diversity in the business districts in Midtown Manhattan indicates that many trips are headed to the major public transit hubs.

The next analysis showcases how the proposed measures can be used to evaluate the effect of an anonymization algorithm: we chose to perturb the location of pick-ups and drop-offs by adding noise sampled from a Gaussian distribution centered in 0 and with a standard deviation of 500 m; the temporal component has been similarly perturbed with noise sampled from a Gaussian distribution centered in 0 and with a standard deviation of 10 minutes. The analysis has been repeated 3 times and the scores averaged across repetitions. Adding random noise to spatio-temporal coordinates has the effect to move these points to nearby locations. If the map is partitioned, e.g. into equivalence areas, perturbed points that are close to the boundaries of these cells have a high likelihood to move to a neighboring cell. For this reason, we expect anonymization to increase the



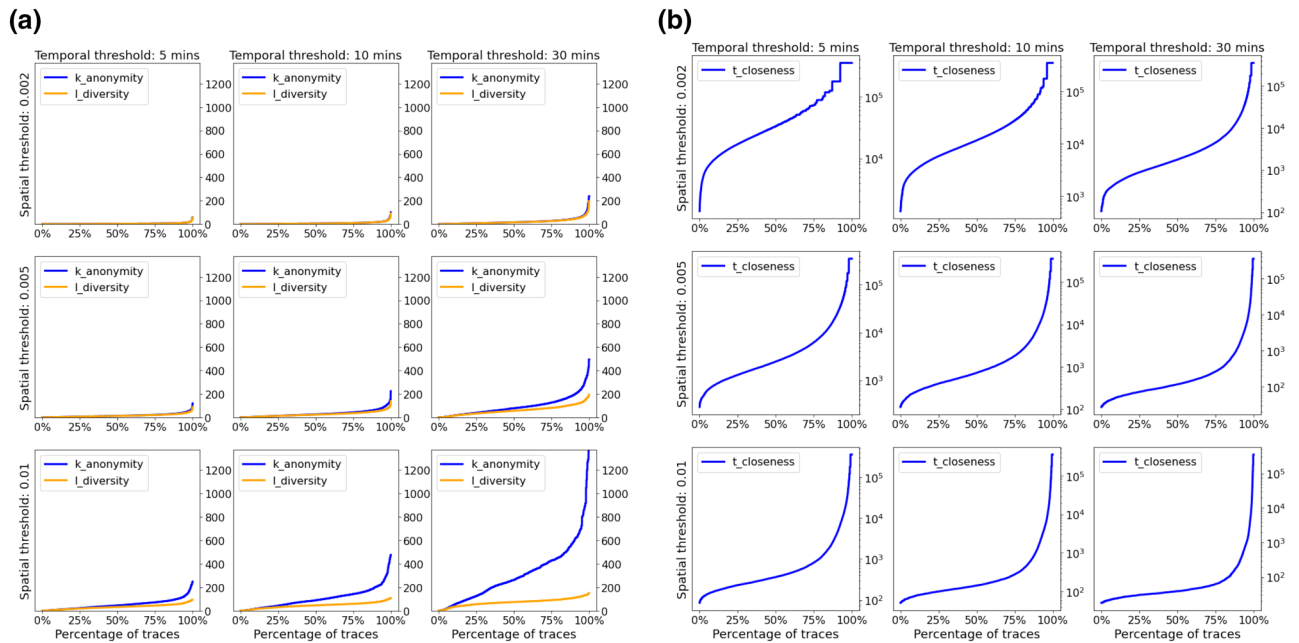
**Figure 3.** L-diversity scores for raw data. Lower values mean that fewer destinations are reached by trips starting in that cell. Scores are consistent with commute patterns to and from the main transit hubs and business districts. (a) Trips during the morning rush-hour (7:00–7:30). (b) Trips during the evening rush-hour (17:00–17:30).



**Figure 4.** Change in k-anonymity scores after applying anonymization to the dataset. (a) Change of k-anonymity values in the anonymized morning data with respect to the raw data. (b) Change of k-anonymity values in the anonymized evening data with respect to the raw data.

re-identification risk of dense areas, while decreasing the risk of neighboring areas. This effect is captured in Fig. 4, which shows that k-anonymity values decrease in dense areas such as the main public transit hubs, while increasing in neighboring areas.

The final analysis looks at how the definition of the equivalence areas affects the scores: For simplicity, we define equivalence areas as cells of a grid with a predefined spatial size (equal for latitude and longitude); the temporal component is similarly partitioned in time-windows of a given size. The analysis defines equivalence areas for all combinations of a spatial size among 0.002 (corresponding to around 200 m), 0.005 (corresponding to around 500 m) and 0.01 (corresponding to around 1 km) and a temporal size among 5 min, 10 min and 30 min. Results are presented on a series of staircase plots, where a score is associated to a percentage of trips in the data. Each row on the plots represents a given spatial size, each column a given temporal size. Figure 5a



**Figure 5.** (a) Parameter exploration across equivalence areas of different spatial (rows) and temporal (columns) sizes. Variation of k-anonymity and l-diversity scores. An increased area size leads to higher privacy. The leftmost part of the curves represents trajectories with values of k or l below a certain threshold (y coordinate), for example the point (20%, 3) indicates that 20% of the traces in the data have less than a score of 3 while 80% have a score larger than 3. (b) Parameter exploration across equivalence areas of different spatial (rows) and temporal (columns) sizes. Variation of t-closeness scores. Increasing the size of the areas leads to higher privacy, i.e. lower t values.

shows how the scores of k-anonymity and l-diversity change with the size of the equivalence areas: larger areas lead to lower k-anonymity values, as the density increases, while the l-diversity values do not grow as much as an increased grid size reduces the number of possible destinations.

In summary, the analysis shows that the proposed metrics reflect an intuitive understanding of privacy risks: for example, residential areas are associated to a higher risk of re-identification than transportation hubs. On the contrary, inference risks in transportation hubs are higher as commuting patterns are easier to predict.

## Discussion

This work introduces a model of privacy risk estimation for trajectory data which assumes adversaries with imperfect knowledge, e.g. caused by noisy GPS signals or inaccurate background knowledge. Previous work defines privacy risk as overlap between spatio-temporal regions, which is either computed over spatio-temporal boxes around individual points<sup>5,15</sup> or over spatio-temporal cylinders that surround trajectories<sup>16–18</sup>.

Both these approaches represent uncertainty as volumes around each data point, while our approach describes uncertainty in terms of *equivalence areas* that correspond to a specific semantic meaning, e.g. associated to a specific concept such as POIs<sup>6,12,19</sup>, road segments<sup>20</sup>. One such definitions is location diversity<sup>12</sup>, an adaptation of l-diversity for location-based queries which measures the diversity in semantic locations where a particular query originates. A limitation of these and of our model is that they do not consider contextual information that is not related to a specific location, e.g. acceleration profile of the driver, which could lead to re-identification of a trajectory. The proposed framework assumes that each trajectory is re-identified via contextual information that is related to one of the stay-points, i.e. the quasi-identifier; in the case the trajectory can be re-identified otherwise, the quasi-identifier should be considered as a sensitive attribute instead. We believe the proposed framework can be applied to re-identified trajectories, e.g. by considering each stay-point as a sensitive attribute for a new virtual trajectory and introducing dummy stay-points to act as quasi-identifier for each virtual trajectory.

Our approach expands on prior work about the measurement of singling-out, linkability and inference in trajectory data. This line of work is based on the framework outlined by<sup>5</sup>, where singling-out risk is modeled as the uniqueness of trajectory data. There, data from telco providers is considered, whose location information is represented by pre-defined areas, i.e. related to cellular tower coverage. Uniqueness of trajectories is then computed for different spatial and temporal aggregation of these areas. Differently from our approach, areas are pre-defined and not based on semantic attribute of the locations; additionally, risks of inference and linkability are not considered in the analysis.

A follow-up work takes the same approach to analyze the uniqueness of location data consisting of a series of purchases<sup>15</sup>. Differently from the previous scenario, here location data is associated to a semantic meaning related to the attributes of the transaction, e.g. price, store name. Here, a combination of shop and a time window can be seen as an equivalence area, modeled around an adversary with specific knowledge, e.g. having access to purchase habits of the target but not being capable of standing in the queue behind the target to observe the

purchase. Furthermore, linkability risk is examined by measuring how much the uniqueness of trajectories increases if the adversary knows the approximate price of the transaction, which could be inferred from additional data such as a credit card statement. The approach presented in our work expands on this by explicitly defining equivalence areas and by proposing a method to derive those from trajectory data which does not already possess a semantic meaning, i.e. GPS coordinates not associated to a specific shop. The goal of the analysis also differs, as our work examines the risks of singling-out, inference and linkability as a function of the adversary's background knowledge.

This line of work is further expanded by<sup>6</sup>, where the focus is on measuring matchability, or linkability, between two datasets as the matching points in the second dataset that are within a spatio-temporal radius of the target point. Our proposed method improves on this approach by defining equivalence areas on the semantic properties of the locations, as opposed on a standard value of radius, which enables to estimate singling-out, inference and linkability with a single and uniform processing of the data.

The contributions of this work are: (i) Define a novel threat model for location data privacy in which the adversary has imperfect knowledge about the target. (ii) Introduce the concept of equivalence areas to model the ability of the adversary at discriminating between locations. (iii) Introduce a definition of quasi-identifiers and sensitive attributes for trajectory data that is based on context and semantics of location data. (iv) Define singling-out, linkability and inference risks for trajectory data. (v) Adapt the standard definition of k-anonymity and derivatives to measure privacy risks in trajectory data.

Another advantage of our model is the ability to compute the risk on any dataset, irrespective of the anonymization strategy that was applied to it. This allows the analyst to capture the effect of applying anonymization on the privacy risks. This in turn allows optimizing anonymization by tailoring the anonymization strategy and / or its parameters to specific records or regions based on their intrinsic risks, e.g. apply more aggressive anonymization to the 10% of traces with the lowest k-anonymity value (those represented in the leftmost part of graphs in Fig. 5a).

The purpose of the analysis presented in this work is to showcase the use of our model in a real-world setting. Relying on these results for risk estimation is not recommended due to several limitations of the analysis:

A first limitation is the simplistic definition of equivalence areas, based on static polygons that represent administrative boundaries. This limitation has been mitigated by generating additional polygons around the major public transit hubs, that possess a semantic meaning related to commute patterns. Still, this definition of equivalence areas does not consider map information, e.g. POIs, which is leading to an underestimation of the risk associated to a realistic adversary. An improved definition of equivalence areas could take advantage of existing solutions that define areas around POIs, e.g.<sup>21</sup>. Generally speaking, the risk estimation obtained from the model refers to an adversary with specific skills, as represented by equivalence areas. A privacy practitioner adopting this model should first evaluate how skilled a typical adversary might be in that context, e.g. evaluating which additional information an adversary might have and the availability of such information, and then define equivalence areas accordingly. The risk estimates obtained with that model would then present an upper bound for the risk associated to any less skilled adversary.

A second limitation is caused by the assumption for simplicity that the adversary has background knowledge about the origin of a target's trajectory and is interested in inferring personal data referring to the trajectory's destination. This assumption does not limit the validity of the model because the same process can be applied also if these roles are reversed, although it might lead to different values of risk and thus might require to aggregate those. For example, consider commuting from home to work: one specific adversary, e.g. a neighbor, might know the home address and be interested in deducing the work address from the data; another adversary, e.g. a co-worker, might know the work address and be interested to obtaining the home address from the data. In this example, the home address acts as a quasi-identifier for the first adversary while the work address acts as a sensitive attribute, vice versa for the second adversary (see Fig. 2b).

Accurate risk estimation should consider that each stay-point can take either roles. A similar consideration can apply to trajectories with more than two stay-points, in which case the risk could be obtained with the same method by aggregating estimates for each pair of stay-points in the trajectory. Previous work can guide an extension of this model to other use-cases where trajectories typically contain more than two stay-point, e.g. where re-identification of purchase transaction data typically requires matching multiple points against a user profile<sup>15</sup>. Assuming that each point belongs to a different equivalence area, re-identification would require that a trajectory matches on all equivalence areas simultaneously.

Explicitly modeling privacy risks caused by repeating patterns in mobility data, e.g. frequent visits to a worship facility, could also increase the accuracy of estimation of both re-identification and inference risks. This limitation is driven by the dataset, which lacks information about which trips in subsequent days were associated to the same individual. To model the contribution of patterns to re-identification, we need to explicitly include them in the definition of equivalence areas, e.g. to match all trajectories that start from a specific worship facility at a time that corresponds to the target's habits. Regarding inference risks, the model can already work with repeating patterns, as repeated visits to a specific location would appear as duplicated values in the sensitive attribute, e.g. multiple destinations in the same equivalence area.

## Conclusions

To conclude, our work improves on prior efforts by explicitly modeling the (imperfect) adversary knowledge through the definition of *equivalence areas*. Equivalence areas represent spatio-temporal regions where an adversary, with a given level of knowledge and precision, is unable to re-identify data points related to a target individual, e.g. an adversary with knowledge about the home address of a target might not be able to distinguish the data of the target from that of the neighbors in a given area around the home. The size and shape of equivalence

area reflects the semantic meaning of the location, which depends on features such as nearby POIs, the road network, and so on; as well as the knowledge of the adversary.

Equivalence areas offer flexibility when computing privacy risks, as risks related to adversaries with different capabilities can be computed by adapting how equivalence areas are drawn.

Additionally, we provide a single methodology that works with any type of location data, whether semantic meaning is associated to it in advance or not, whether the data has been previously anonymized or not, and independently of which kind of anonymization algorithm has been used.

Finally, we show how to derive, with a uniform methodology, the standard privacy metrics of *k-anonymity*, *l-diversity* and *t-closeness* from equivalence areas. Using these widely-adopted metrics, it becomes possible to compare privacy risk in trajectory data to risks in other scenarios. We also show that these metrics are effective to quantify risks of singling-out, linkability and inference using a real-world dataset.

This work can be of interest to data providers and providers of location-based services, e.g. Pick-Up Drop-Off (PUDO) analysis, that want to estimate privacy risk associated to trajectory data, or want to evaluate the effectiveness of an anonymization algorithm. Data subjects and end users can also find this work relevant to estimating the privacy cost associated with sharing their trajectory data.

## Methods

**Threat model.** The privacy threat comes from an adversary who is looking to extract personal information from a given trajectory dataset. The adversary can either access the data with illegitimate means, e.g. eavesdropping insecure communication (so-called *active* malicious adversary), be authorized to access the data, e.g. the adversary is an *honest-but-curious* service provider<sup>22</sup>, or the dataset is publicly available. An example of such adversary is a data scientist or a privacy researcher at a location-based service provider, or a journalist wanting to track a target individual using public mobility data.

A passive adversary can threaten privacy in two ways: (i) Targeting threats: the adversary aims at obtaining additional information about a specific target, i.e. an individual whose data is contained in the dataset and about whom the adversary possesses some background information. (ii) Observing threats: the adversary aims at obtaining additional information about any target of opportunity; the adversary starts by individuating a target with a high privacy risk and only later scouts for publicly available background knowledge about this individual which allows singling out and inference (both privacy aspects are explained in paragraph “[Privacy Risks](#)”). Our threat model covers both observing and targeting cases of adversary threats.

Therefore, our threat model assumes that: (i) For targeting threats, the adversary knows with certainty that the dataset contains one trajectory created by the target. If this is not the case, the risk of singling-out would be null. For observing threats, we assume that it is possible to find further background knowledge of the (at least one) singled-out individual. (ii) The adversary possesses limited background knowledge about the target. This background knowledge allows the adversary to define a “semantic region” that acts as quasi-identifier, e.g. all roads within 5 minutes walking distance from the target’s home on which parking is possible. (iii) The adversary has access to some additional publicly available data sources that allows them to link this background knowledge to regions of the map, e.g. information about POIs, the road network, population density, address directory. (iv) By correctly re-identifying the target’s trajectory, the adversary can gain personal information about the target, supplementary to the background knowledge.

Both targeting and observing attacks can be modeled in the same way: (i) The adversary defines an equivalence area based on their background knowledge about the target, which acts as quasi-identifier. (ii) The adversary identifies a set of trajectories that match the target’s equivalence area; this result determines the target’s singling-out risk, which covers one aspect of the privacy risk. (iii) The adversary performs inference on the values of the sensitive attributes of the matched trajectories; this result determines the target’s inference risk, which covers another aspect of the privacy risk. (iv) The adversary links the trajectory data with an external dataset via the semantic meaning associated to equivalence areas, e.g. security camera footage, to understand if a trajectory of a specific individual is contained in the dataset.

**Definition of privacy metrics.** This section contains a discussion about the choice of privacy metrics for the experimental evaluation.

The goal of the experimental evaluation is to showcase how a practitioner can use the presented equivalence area concept to measure the privacy risk of location data. For this evaluation, we chose the metrics of *k-anonymity*, *l-diversity* and *t-closeness* because their popularity facilitates comparing our results to different use cases. A further reason to choose these metrics is the clarity and intuitive understanding about data protection, i.e. the concept of ‘crowd to hide’, that they provide to data subjects, practitioners and regulators.

Our choice of metrics does not limit the applicability of the framework, the choice of the most appropriate metric for each use case is left to the practitioner that adopts the concept of equivalence areas. As an example, the metric of *k-anonymity* for location data has been criticized<sup>23</sup> as not representative of the actual privacy risk, so a different metric might be more appropriate, e.g. correctness<sup>24</sup>.

In the remainder of the section, we describe adaptations of the metrics of *k-anonymity*, *l-diversity* and *t-closeness* to scenarios with trajectory data, by relying on the definition of equivalence areas.

In trajectory data, many data points forming one trajectory are associated to the same individual, unlike tabular data such as medical records where typically one record is associated to each patient. Adopting the standard definition of *k-anonymity*, *l-diversity* and *t-closeness* would give a wrong estimate of the privacy risk, as the “crowd to hide”, i.e. data points in one anonymity set, might likely consist of a single individual to which all data points are associated.



**Privacy Risks.** Processing personal data involves the following risks<sup>25</sup>: (i) *Singling out* is a risk that a record or a set of records about an individual in a dataset can be isolated, without using additional information, e.g. other datasets. This risk is measured by *k*-anonymity. (ii) *Inference* is a risk to deduce the value of an attribute from the values of a set of other attributes. For example, access to a healthcare dataset might allow an adversary to deduce the value of diagnosis for a target starting with knowledge about the values of attributes ZIP code and age for that target. Strict *k*-anonymity measures an upper bound of the inference risk. (iii) *Linkability* (or *matchability*<sup>6</sup>) is the risk of connecting, across at least two different datasets, data records that corresponds to the same individual. This way, the adversary can find more information about the individual than what could be gained from a single dataset, possibly being able to single-out the individual. While singling out and inference are privacy risks within the given dataset on its own, linkability describes the privacy risks when combining multiple datasets. *L*-diversity and *t*-closeness measure each a different aspect of linkability.

***K*-anonymity.** The metric of *k*-anonymity quantifies the re-identifiability, or singling-out risk, by counting the number of records that are indistinguishable with respect to some background knowledge, i.e. the “crowd to hide”. If the record identifying the target is in an equivalence area of size *k*, the adversary would not be able to distinguish the correct record from the other *k* – 1 false-positives. We say that each of the *k* trajectories in the equivalence area described by the background knowledge is *k*-anonymous, has *k*-anonymity = *k*, or that the equivalence area is *k*-anonymous.

**Strict *k*-anonymity.** *K*-anonymity can also measure the risk associated with an adversary finding out whether a trajectory from an observed dataset belongs to a specific individual, based on an additional dataset. For example, an adversary might possess security camera footage showing the individual entering a taxi at a particular location and time. This information can help the adversary to determine if the taxi ride is contained in the observed dataset. We cannot know in advance if the adversary has access to data linkable to the equivalence area of the trajectory start, of the trajectory end, or both. Therefore, we take a conservative approach and measure the worst-case scenario of the adversary having knowledge about both origin and destination. We measure this linkability risk by computing the *strict k*-anonymity, which is the number of trajectories that are associated to the same equivalence areas at the origin and at the destination:  $k_\tau = |\{\gamma \neq \tau : \tau^{QI}, \gamma^{QI} \in A, \tau^{SA}, \gamma^{SA} \in A'\}|$ , where  $\tau, \gamma \in T$  are trajectories. This metric measures the “crowd to hide” in a more restrictive way than *k*-anonymity, as it assumes that quasi-identifiers include both the origin and the destination.

***L*-diversity.** *K*-anonymity covers only one aspect of privacy risk; therefore, a large equivalence area does not necessarily guarantee high privacy, as the sensitive attributes could reveal additional information: the sensitive attributes in an equivalence area of size *k* might in practice contain only  $l \ll k$  ‘well-defined’ values. For example, an adversary could infer the work address of a target with certainty if all trajectories leaving from their home address lead to the same office building. We say that any trajectory that originates in a specific equivalence area *A* is *l*-diverse if the *k* trajectories that originate in *A* end into  $l \ll k$  distinct equivalence areas. Similarly, we say that equivalence area *A* is *l*-diverse.

***T*-closeness.** *L*-diversity measures how much sensitive information can be disclosed by the sensitive values in an equivalence area; *t*-closeness applies the same concept across equivalence areas, as differences in the distribution of sensitive values in an equivalence area and in the overall population can reveal sensitive information (via a so-called *skewness attack*<sup>10</sup>). *T*-closeness measures the distance between two distributions: *P*, which indicates the frequency sensitive attribute values in the whole dataset, and *Q<sub>A</sub>*, which indicates the frequency sensitive attribute values *I<sub>A</sub>* in an equivalence area *A*. We say that a dataset is *t*-close if the distance  $D[Q_A, P]$  is smaller than a threshold *t*, for every equivalence area *A*. The higher the distance, the more information the adversary can infer from the data; therefore, contrary to *k*-anonymity and *l*-diversity, a lower value of *t* means higher privacy.

Figure 5b shows how the score of *t*-closeness changes with the size of the equivalence areas: larger areas lead to a steeper curve with more trips having a lower *t*-closeness value. This is consistent with the findings about *k*-anonymity and *l*-diversity (see Fig. 5a) as it indicates a lower diversity between the contents of equivalence areas and therefore higher privacy.

Received: 1 June 2021; Accepted: 22 November 2021

Published online: 07 January 2022

## References

1. Bonchi, F. Privacy preserving publication of moving object data. In *Privacy in Location-Based Applications*, 190–215 (Springer, 2009).
2. Dewri, R. Local differential perturbations: location privacy under approximate knowledge attackers. *IEEE Trans. Mobile Comput.* **12**, 2360–2372 (2012).
3. Golle, P. & Partridge, K. On the anonymity of home/work location pairs. In *International Conference on Pervasive Computing*, 390–397 (Springer, 2009).
4. Public nyc taxicab database lets you see how celebrities tip. <https://gawker.com/the-public-nyc-taxicab-database-that-accidental-ly-track-1646724546>. Accessed 31 May 2021.
5. De Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M. & Blondel, V. D. Unique in the crowd: the privacy bounds of human mobility. *Sci. Rep.* **3**, 1376 (2013).
6. Kondor, D., Hashemian, B., de Montjoye, Y.-A. & Ratti, C. Towards matching user mobility traces in large-scale datasets. *IEEE Trans. Big Data* (2018).
7. Tlc trip record data. <https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>. Accessed 31 May 2021.

8. SWEENEY, L. K.-anonymity. A model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.* **10**, 557–570. <https://doi.org/10.1142/s0218488502001648> (2002).
9. Machanavajjhala, A., Gehrke, J., Kifer, D. & Venkatasubramanian, M. L-diversity: privacy beyond k-anonymity. In *22nd International Conference on Data Engineering (ICDE'06)* (2006). <https://doi.org/10.1109/icde.2006.1>.
10. Li, N., Li, T. & Venkatasubramanian, S. t-closeness: Privacy beyond k-anonymity and l-diversity. In *2007 IEEE 23rd International Conference on Data Engineering* (2007). <https://doi.org/10.1109/icde.2007.367856>
11. Gruteser, M. & Grunwald, D. Anonymous usage of location-based services through spatial and temporal cloaking. In *Proceedings of the 1st international conference on Mobile systems, applications and services-MobiSys '03* (2003). <https://doi.org/10.1145/1066116.1189037>.
12. Xue, M., Kalnis, P. & Pung, H. K. Location diversity: Enhanced privacy protection in location based services. In *International Symposium on Location-and Context-Awareness*, 70–87 (Springer, 2009).
13. Primault, V., Boutet, A., Mokhtar, S. B. & Brunie, L. The long road to computational location privacy: a survey *IEEE Commun. Surv. Tutor.* **10**, 557. (2018).
14. Nyc community district boundaries. <https://data.cityofnewyork.us/City-Government/Community-Districts/yfkn-k7r4>. Accessed 24 Sep 2021.
15. De Montjoye, Y.-A. et al. Unique in the shopping mall: on the reidentifiability of credit card metadata. *Science* **347**, 536–539 (2015).
16. Bettini, C., Wang, X. S. & Jajodia, S. *Protecting Privacy Against Location-Based Personal Identification*, 185–199. Lecture Notes in Computer Science (Springer, 2005). [https://doi.org/10.1007/11552338\\_13](https://doi.org/10.1007/11552338_13).
17. Abul, O., Bonchi, F. & Nanni, M. Never walk alone: Uncertainty for anonymity in moving objects databases. In *2008 IEEE 24th international conference on data engineering*, IEEE, 376–385 (2008).
18. Nergiz, M. E., Atzori, M. & Saygin, Y. Towards trajectory anonymization: a generalization-based approach. In *Proceedings of the SIGSPATIAL ACM GIS 2008 International Workshop on Security and Privacy in GIS and LBS*, 52–61 (2008).
19. Tan, R., Tao, Y., Si, W. & Zhang, Y.-Y. Privacy preserving semantic trajectory data publishing for mobile location-based services. *Wirel. Netw.* 1–10 (2019).
20. Wang, T. & Liu, L. Privacy-aware mobile services over road networks. *Proc. VLDB Endow.* **2**, 1042–1053 (2009).
21. Placekey. <https://www.placekey.io/>. Accessed 24 Sep 2021.
22. Goldreich, O. Foundations of cryptography-a primer. *Found. Trends Theor. Comput. Sci.* **1**, 1–116. <https://doi.org/10.1561/04000001> (2005).
23. Shokri, R., Troncoso, C., Diaz, C., Freudiger, J. & Hubaux, J.-P. Unraveling an old cloak: k-anonymity for location privacy. In *Proceedings of the 9th annual ACM workshop on Privacy in the electronic society*, 115–118 (2010).
24. Shokri, R., Theodorakopoulos, G., Boudec, J.-Y. L. & Hubaux, J.-P. Quantifying location privacy. In *2011 IEEE Symposium on Security and Privacy* (2011). <https://doi.org/10.1109/sp.2011.18>.
25. Article 29 Data Protection Working Party. Opinion 05/2014 on anonymisation techniques (2014).

## Author contributions

S.B. and A.K. conceived the concept and definitions, S.B. conducted the analysis. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to S.B.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022