Trajectory Data Privacy: Research Challenges and Opportunities

Tarlis T. Portela[1,2], **Francisco Vicenzi**[1], **Vania Bogorny**[1]

[1]Programa de Pós-graduação em Ciência da Computação
Departamento de Informática e Estatística
Universidade Federal de Santa Catarina (UFSC), Florianópolis, SC, Brazil.

[2]Instituto Federal do Paraná (IFPR), Palmas, PR, Brasil.

tarlis@tarlis.com.br, francisco.vicenzi@grad.ufsc.br, vania.bogorny@ufsc.br

Abstract. *With the explosion of trajectory data available from many sources, emerges the problem of data privacy. Trajectory privacy methods have been studied for many years. Data analysis and mining methods can benefit from truthful data sources, but for this purpose, protecting users privacy is crucial. Trajectories have been studied as multidimensional data with space, time and semantic dimensions in which a few works in the literature have considered all of them. The more information that is associated to mobility data, the more sensitive is the user privacy. In this paper we present the basic concepts and the state of the art in trajectory privacy and present the challenges related to mobility data anonymization.*

1. Introduction and Motivation

With the popularization and price reduction of mobile devices, large volumes of mobility data are being collected about our daily routines. In the era of Big Data, movement data can be enriched with information from several sources, such as sensors, internet channels, social networks, etc. With this explosion of enriched data, new technologies and methods are being developed for categorizing, processing, and mining these big data [Ferrero et al. 2016].

The movement data collected by mobile devices are called *moving object trajectories*. The most simple type of trajectory, called raw trajectory, is a sequence of points $T =< p_1, p_2, ..., p_n >$, where each $p_i = (x_i, y_i, t_i)$, with $p_i \in T$, x_i and y_i are the spatial position of the moving object (MO) in space at a time instant t_i.

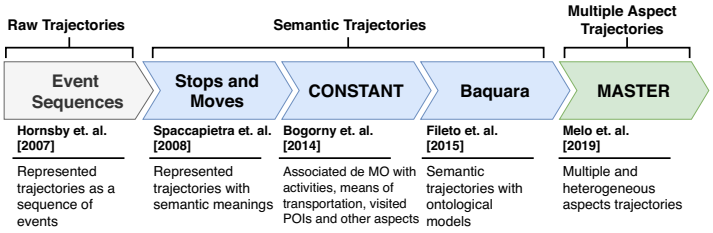

Figure 1. Timeline of the evolution of trajectory data models.

Over the last decade, several data models have been proposed to represent and enrich trajectories with semantic information, which is leading to a without precedent violation of human privacy. The evolution of these data models is shown in Figure 1. In

2007, Hornsby and Cole [2007] started by modelling trajectories as sequences of events in space along time. In 2008, Spaccapietra et al. [2008] proposed the concept of semantic trajectory, which integrates trajectories with geographic information and distinguishes stops and moves. Stops are the parts of a trajectory where the moving object has stayed for a minimum period of time, while the moves represent the movement between stops. In addition to the spatio-temporal attributes of space and time, *Semantic Trajectories* can have each stop associated with semantic information, called Points of Interest (POI). Most commonly, a POI is a place name. Later in 2014, a semantically richer model, called CONSTANT, was proposed by Bogorny et al. [2014], associating the moving object trajectories with the visited POIs, the activities performed at a POI, the means of transportation, the goal of a visit, etc. Fileto et al. [2015] proposed the BAQUARA framework to enrich trajectories with ontologies and linked open data.

Recently, Mello et al. [2019] proposed the model MASTER, which introduces the concept of *multiple aspect trajectory*, allowing the enrichment of trajectories with any type of information, also called aspects. This model solves the problem presented in Ferrero et al. [2016], in which aspects were considered separately. The MASTER model allows the representation of the trajectory with space, time and several aspects, any of which might violate the user privacy. Figure 2 shows a multiple aspect trajectory that follows the definitions given by Mello et al. [2019], with aspects that differ along the trajectory. As can be observed from the figure, a trajectory has very detailed information about the moving object, with several aspects as: (i) at home, the heart and sleeping rates are collected from a smartwatch; (ii) when he moves on foot to work the humor is given by a tweet; (iii) at a smart office sensors collect environment information, as the noise, temperature and pollution; (iv) at night, the characteristics of the places visited by the moving object, as price and rating of a restaurant. In summary, this new type of trajectory reveals the very detailed daily routine of a person, which is more sensitive to privacy than previous models.

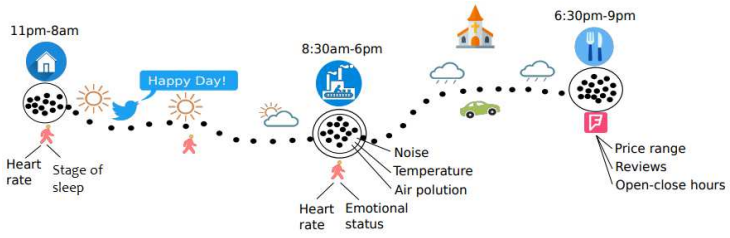

Figure 2. Example of a multiple aspect trajectory [Mello et al. 2019].

The problem with such rich data about human mobility is the sensitivity to human privacy. Many organizations, industries and government often need to publish data for research purposes (e.g. public health) [Chow and Mokbel 2011]. The challenge to researchers around the world is to share the data without revealing sensitive information of the users, and for that they need to protect the information using data anonymization techniques.

Recent concerns in privacy headed to a peak in a Facebook breach that captured 87 million users personal information used to manipulate US elections [Bennett 2018]. Concerns regarding data security motivated countries to implement laws to protect citi-

zens privacy. The Llp [2016] approved The General Data Protection Regulation (GDPR). Brazil published General Law of Data Protection [Cots and Oliveira 2018], in 2018. It means that private sensitive data as trajectories need protection, thus anonymization methods are being developed.

Anonymizing trajectory datasets by simply suppressing or replacing direct identifiers (names and ID numbers) is not enough. Even anonymized records, when joined with external data sources, can still reveal a user identity due to the called quasi-identifiers, that combined with other information can indicate the person in a certain degree of confidence [Sui et al. 2016]. Trajectories can be enriched with lots of information and inferred features (e.g. moving behaviors) that could be quasi-identifiers.

In this paper we survey the state-of-art on trajectory anonymization methods and present some challenges for anonimyzing multiple aspect trajectories. In order to develop trustworthy studies, sensitive information as medical history, relationships, and personal data must be real, so the question is: how to anonymize data without loosing its meaning and without violating users privacy? As the large companies as Facebook, Google, and others generate this new type of trajectories, we believe that multiple aspect trajectory anonymization will become a large research issue in the next decade.

The rest of the paper is organized as follows: Section 2 introduces the basic concepts of privacy. Section 3 presents a comparative study of privacy methods proposed in the literature and their limitations, and the need of new proposals to multiple aspect trajectory privacy. Finally, in Section 4 we discuss our vision of the future research challenges and opportunities in trajectory privacy methods.

2. Privacy and Anonymization Basic Concepts

This section presents the basic privacy concepts, which are essential for the better understanding privacy in the trajectory context. The following subsections describe anonymization objectives (Section 2.1), anonymization techniques (Section 2.2), and the kinds of knowledge an attacker could use to gain private information on the published data (Section 2.3).

2.1. Anonymization Objectives

The premise of privacy-preserving methods is to keep the data usable for research purposes while protecting moving objects identity, thus, allowing data to be shared, publicly released and used in mining studies.

As presented in Figure 3, there are three objectives for anonymization. First, in *Privacy-preserving Data Mining (PPDM)* methods, mining the trajectory datasets is performed before publishing the data and resulting in real statistics. The second, is the *Privacy-preserving Data Querying (PPDQ)* methods used in services that provide portions of data by querying systems. A portion of data is released by retrieving information from a service. Then, to protect a user privacy, the querying system can: (i) filter information that can be retrieved (selective release); (ii) rewrite the query to select more results with nearest neighbors; (iii) translate the resulting data to a higher level of granularity (i.e. translate specific locations into areas); and (iv) generate fake data among results.

The third objective for anonymization is *Privacy-preserving Data Publishing (PPDP)* methods, that results in publishing datasets for mining by third parties [Gramaglia

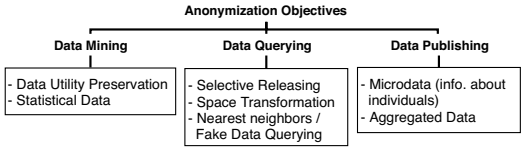

Figure 3. Anonymization objectives.

et al. 2017]. The dataset is modified to anonymize its individuals, generally resulting in releasing their detailed information (microdata) or in an aggregated form. Three are the requisites for publishing datasets: (i) it has to be anonymized, (ii) the published data are the records, and not the results extracted with data mining methods such as classification, association rules or aggregated statistics, and (iii) the records must be truthful, avoiding introduction of fictitious data. The focus of our work relies on PPDP methods that allow data to be used for research, mining studies, and querying systems.

2.2. Anonymization Techniques

There are several techniques employed in data anonymization methods. The simplest approach is by replacing direct identifiers (names and ID numbers) by pseudonyms. Another approach is to suppress trajectories not in a group less then k moving objects [Abul et al. 2008]. From the basic concepts for trajectory anonymization, most of the related works use *suppression, generalization, masking* and *perturbation* as strategies for privacy protection. Considering these concepts, we classify the anonymization works according to Figure 4, in three main categories: *suppression, generalization*, and *masking* to its deriving strategies.

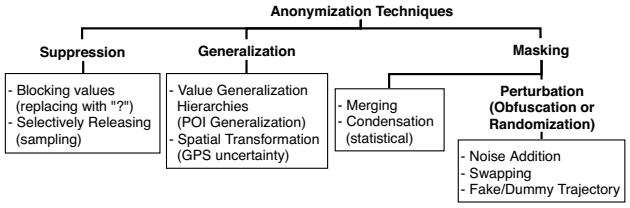

Figure 4. Anonymization techniques.

Suppression is the simplest anonymization way, by removing identifiers and sensitive data from records, which can be replaced by symbols (blocking) or random values. In some cases, the suppression algorithms might remove entire records. However, this operation results in more information loss, which impacts in data utility. In online services for retrieving data (e.g. querying systems), the suppression technique is employed as releasing only the none sensitive data. Suppression is the most common anonymity technique generally employed with k-anonymity and generalization [Ye et al. 2016].

To keep data utility some strategies were developed in order to preserve structure, loosing less information. The most employed technique *Generalization*, which translates granularity on specific data values into a higher level of data category, as for instance, changing a hotel name from *Mercury Hotel* to *Hotel* [Monreale et al. 2011]. With such generalization, it is possible to maintain a certain level of semantics and not revealing the

specific place a person has visited. The generalization technique can be employed as: (i) semantic values (e.g. POI names) by generalization following a hierarchy of categories; (ii) space dimension by adding imprecision or transforming points of the trajectory (e.g. latitude and longitude) into areas like blocks in a grid [Saygin et al. 2009]. The works of Abul et al. [2010], Huo et al. [2012], Gramaglia et al. [2017] and Shaham et al. [2019] discretized the spatial positions of the trajectories into grid cells. According to Pensa et al. [2008] this spatial translation enables to find enough matches of points with respect to any value of k-anonymity, that would be practically impossible with specific spatial granularity.

The third technique, *Masking*, consists on modifying data not to be re-engineered, but it changes the structure. First, the masking technique consists on grouping information by merging similar records [Gramaglia et al. 2017]. The second technique is *Condensation*, that groups data into predefined sizes by transforming them into a certain level of statistical information about original records [Wang et al. 2009]. It is a way of maintaining the true statistics of data, but not keeping its original structure. This suffices to preserve correlations across different dimensions. However, trajectory datasets are not published, thus mining them depends on the data owner. With *Perturbation* techniques, also known as obfuscation or randomization, noise is added in order to hide values in a way that the original data cannot be recovered. In general, this strategy keeps the structure of the data, but looses in its semantic meaning. Employed masking techniques include: (i) the insertion of random noise; (ii) swapping values between records or copying values from one user to another; (iii) inserting fake trajectories into the data. Data scrambling or encryption mechanisms are used as well. We argue that perturbation methods will unlikely keep the trajectories semantic meaning since dummy data is inserted.

Most anonymization methods for trajectory data publishing are based on the concept that an anonymous person cannot be identified in a group of k elements, the called k*-anonymity* [Sweeney 2002]. Hence, data is protected when the information cannot be distinguished from at least $k - 1$ individuals, intending to hide a person in a crowd. Consequently, combining the released records or a subset of its attributes with external sources should not link any individual to match less than k others [Sweeney 2002]. As an example, anonymity methods could cluster similar trajectories in a way that none of the individuals can be distinguished from each other. Similarly, Machanavajjhala et al. [2006] proposed the l*-diversity* concept, where each attribute must have at least l possible values. The more diverse the values in a database are, the lower is the probability for a user to be identified. For example, if a user location is indistinguishable from a set of l different places, then it is less likely to someone guess its location.

2.3. Adversary Knowledge and Attack Model

In general, PPDPs works compare original and anonymized datasets with a quality metric, or model test attacks that foresee what an adversary might previously know. These adversary models describe the capabilities of an attacker is assumed to have [Wagner and Eckhoff 2018]. In this context, attacks on moving objects privacy have two goals (shown in Figure 5): (i) a user location, aiming to disclose a sensitive place, a position at a time or sequential tracking a moving object [Pelekis et al. 2011], and (ii) a user identity aiming to disclose attribute information as personal identification, a meeting or inferring relationships.

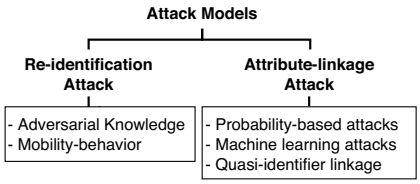

Figure 5. Classification of attack models.

The kind of attacks known as *Re-identification attacks* aim in identifying a user in the trajectory dataset (the published data) or other sensitive information [Dai et al. 2018]. The attackers can use background knowledge, quasi-identifiers or the moving object unique mobility behavior. The background knowledge consists of the information previously known by a malicious opponent, and it might be used to identify someone as, for instance, a place or a sequence of places visited by the moving object at a certain time, his friends, etc. The second kind of attacks, named *Attribute-linkage attacks*, are based on matching trajectories in the database that might reveal the moving object in a level of certainty Dai et al. [2018]. For instance, the values of a sensitive attribute in a group of trajectories are the same, therefore this attribute for that group of individuals can be exactly predicted [Aggarwal and Yu 2008]. In general, sophisticated methods such as probability-based or machine learning models are used to look for patterns in the trajectory data.

3. Trajectory Anonymization Methods

Several works have been proposed to preserve privacy in trajectory databases, i.e., to anonymize the user who is the owner of the trajectory. In this paper we classify trajectory anonymization methods according to Figure 6, by the type of trajectory and the dimensions they are able to treat. As can be observed from the figure, the main problem of most existing works for trajectory data anonymization is that they were developed for raw trajectories, as the works of Pensa et al. [2008], Abul et al. [2010] and Huo et al. [2012], or for trajectories represented as stops and moves, as the works of Monreale et al. [2011], Kopanaki et al. [2016] and Dong and Pi [2018]. Methods for raw trajectories mostly group users with nearest neighbors, distort space or release statistical information of the data. Most methods consider space and time dimensions in anonymization, and a few use only the spatial dimension. There are only a few works for semantic trajectory anonymization, and the recent work of Giotakis and Pelekis [2019] have focused on multiple aspect trajectories for querying systems, that are more sensitive and require more sophisticated privacy protection methods. We believe that this research topic will be the great challenge in the next decade, as Google, Facebook, and other social media generate multiple aspect trajectories.

3.1. Methods for Raw Trajectories Anonymization

The k-*anonymity* is a concept that an anonymous person cannot be identified in a group of k individuals. Methods for k-anonymity were proposed for trajectory datasets with clustering approaches. Indeed, they focused on publishing datasets by anonymizing trajectories of individuals, resulting in a new anonymous dataset. The works of Pensa et al. [2008] and Gurung et al. [2014], for instance, focus on grouping similar trajectories using

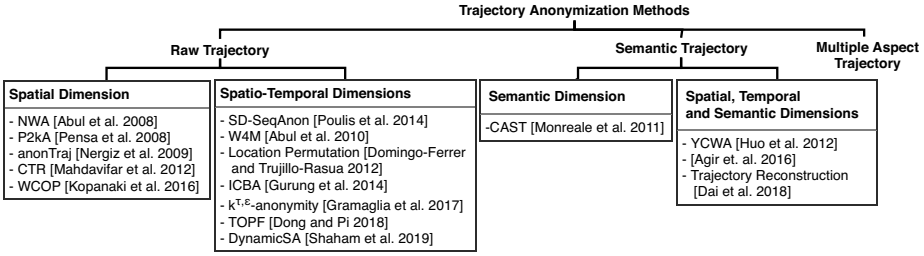

Figure 6. Classification of trajectory anonymization methods.

a measure of similarity. The *P2kA* method proposed by Pensa et al. [2008], uses a prefix tree to anonymize a dataset of spatial locations pruning the tree by the frequency of sequences less than a k threshold. Gurung et al. [2014] proposed the method *ICBA*, which also considers the frequency of spatial locations and remove infrequent subtrajectories within the same time interval.

Abul et al. [2008] extends the k-*anonymity* as (k, δ)-*anonymity* by considering the location of a moving object at a given moment not as a point but a circle of radius δ, and it uses just the spatial dimension. This method, called Never Walk Alone (NWA), groups at least k trajectories in the nearest neighbors, those contained in a $\delta/2$ radius representative cylindrical trajectory. Later, Abul et al. [2010] proposed the method *Wait for Me (W4M)* to obtain higher quality anonymization, considering trajectories near in space that have the same time interval. This method uses randomization and suppression techniques to provide privacy protection.

Since not all individuals are equally concerned about their privacy, personalized privacy configurations can be used [Aggarwal and Yu 2008]. Indeed, in many cases, the user may consent sharing limited information. In order to preserve the data utility, anonymity must be carried carefully as the method of privacy leads to loss of information. The *WCOP* method proposed by Kopanaki et al. [2016], used users settings to offer personalization with less data distortion, in which the user chooses to be in a group with a larger or smaller number of other users. This method clusters trajectories near in space, but omitting the time interval. Mahdavifar et al. [2012] proposed the method (*CTR*), that considers different privacy levels to each trajectory, clustering them in a minimum k-anonymity groups from which one cannot be distinguished in space. Both of these methods ignore time dimension, considering only spatial distances.

Night time POIs often represent points of sensibility where users tend to stay most of the time, as their homes [Liu et al. 2018]. These most frequent or infrequent places might characterize user identity. Indeed, distinct movement behaviors like the subtrajectory from "Home" to "Work" are sensitive to users privacy. The *TOPF* method, proposed by Dong and Pi [2018] removes the subtrajectories within the same time interval and less than k individuals, in order to balance usability and privacy [Dong and Pi 2018].

Saygin et al. [2009] and Poulis et al. [2014] proposed methods that use space-based generalization. The former proposed the method *anonTraj*, that replaces geographical points into grid cells that cover two or more generalized locations. The *SeqAnon* method proposed by Poulis et al. [2014] generalizes locations by selecting two nearest

points in space and replaces those with a set containing both. The *Location Permutation* method proposed by Domingo-Ferrer and Trujillo-Rasua [2012] replaces sensitive points (space and time) of the trajectory by others with similar relevance using a perturbation strategy. However, according to Gramaglia et al. [2017] in order to preserve truthfulness of published data, privacy protection mechanisms can not rely on randomized, perturbed, permuted and synthetic data. The $k^{\tau,\epsilon}$-anonymity method proposed by Gramaglia et al. [2017], segments trajectories by time, using generalization and suppression to obtain k-anonymity groups with the same time intervals.

3.2. Methods for Semantic Trajectories

Similarly to the Domingo-Ferrer and Trujillo-Rasua [2012] (Location Permutation) method for raw trajectories, the *Trajectory Reconstruction* method proposed by Dai et al. [2018] considers as semantic dimension the POI name in the process of perturbation, replacing sensitive stops with other points. The method *SD-SeqAnon* proposed by Poulis et al. [2014] uses generalization of locations, replacing each position that is close in space and semantics with a set containing these similar places. Geographic positions and POI names represent the same locations, so if its replacement does not have the same semantic meaning, its utility will be lost.

Huo et al. [2012] use k-anonymity in the method *You Can Walk Alone (YCWA)* method, proposing to hide significant stops instead of the whole trajectory through spatial generalization. The semantic values of POIs are used in the method to define similarity of places according to the number of visitors, duration and the arriving time. Monreale et al. [2011] proposed the method *CAST*, that employs generalization of POI names for semantic trajectories, instead of using k-anonymity. They attempt to maintain the semantic meaning of POIs. Additionally, according to Monreale et al. [2011], hiding a person into a crowd of k individuals is not enough for robust data protection. Generalization is employed by Ağır et al. [2016] with simple privacy mechanisms, using low to high levels of spatial and semantic privacy. They argue that semantic information improves inference of user spatial locations. Evidently, a place name associated with its generalized spatial information has a high risk for inference.

3.3. Summary of Trajectory Anonymization Methods

Table 1 summarizes the state-of-art on trajectory privacy, with the datasets used to validate the method, the kind of trajectory and the used dimensions, the anonymization techniques employed, and compared methods. We observe in Table 1 that the works use several datasets, but only a few works compare their improvements over other methods. Only a few works consider the semantic dimension and no works in trajectory PPDP consider multiple aspect trajectories.

4. Research Challenges and Opportunities

In this section we present some major challenges on multiple aspect trajectory privacy protection and how they lead to new research opportunities. Privacy preserving methods were developed for raw or semantic trajectories. To the best of our knowledge, the work of Giotakis and Pelekis [2019] is the first that supports multiple aspect trajectories for querying systems, rewriting queries in spatial, temporal or semantic dimensions to achieve k-anonymity. For instance, the method proposed by Abul et al. [2008] only

Table 1. Related works of trajectory anonymization methods.

#	Method	Datasets	Trajectory	Dimensions	Anonymization Technique	Compares to
1	NWA [Abul et al. 2008]	Trucks; Brinkhoff's Oldenburg	Raw	Spatio-temporal	Generalization, Suppression	None
2	P2kA [Pensa et al. 2008]	Milan	Raw	Spatial	Generalization, Suppression	None
3	W4M [Abul et al. 2010]	Milan; Brinkhoff's Oldenburg	Raw	Spatio-temporal	Generalization, Suppression, Condensation	NWA [Abul et al. 2008]
4	anonTraj [Saygin et al. 2009]	Brinkhoff's Synthetic Dataset	Raw	Spatial	Generalization, Supression	None
5	CTR [Mahdavifar et al. 2012]	Brinkhoff's Oldenburg	Raw	Spatial	Perturbation	None
6	Location Permutation [Domingo-Ferrer and Trujillo-Rasua 2012]	San Francisco Taxis; Brinkhoff's Oldenburg	Raw	Spatio-temporal	Suppression, Perturbation	NWA [Abul et al. 2008]
7	ICBA [Gurung et al. 2014]	Synthetic dataset; Brinkhoff's generated	Raw	Spatio-temporal	Suppression	P2kA [Pensa et al. 2008]
8	SeqAnon (framework) [Poulis et al. 2014]	Gowalla; Brinkhoff's Oldenburg	Raw and Semantic	POI	Generalization, Suppression, Perturbation	Others for query answering
9	WCOP [Kopanaki et al. 2016]	GeoLife	Raw	Spatial	Suppression	None
10	$k^{\tau,\epsilon}$-anonymity [Gramaglia et al. 2017]	Orange call detail records	Raw	Spatio-temporal	Suppression, Condensation	None
11	TOPF [Dong and Pi 2018]	Brinkhoff's Oldenburg	Raw	Spatial	Generalization, Suppression	NWA [Abul et al. 2008]; ICBA [Gurung et al. 2014]; P2kA [Pensa et al. 2008]
12	DynamicSA [Shaham et al. 2019]	GeoLife	Raw	Spatio-temporal	Generalization, Suppression	$k^{\tau,\epsilon}$-anonymity [Gramaglia et al. 2017]
13	CAST [Monreale et al. 2011]	Milan; Pisa	Semantic	POI	Generalization, Suppression	None
14	YCWA [Huo et al. 2012]	GeoLife	Semantic	Spatio-temporal, POI	Generalization, Suppression	NWA [Abul et al. 2008]
15	Ağır et al. [2016]	Twitter-Foursquare	Semantic	Spatial, POI	Generalization	None
16	Trajectory Reconstruction [Dai et al. 2018]	Synthetic dataset based on GeoLife	Semantic	Spatio-temporal, POI	Generalization, Personalized	None

considers the spatial dimension, and Monreale et al. [2011] only generalize POI names. Anonymization methods must consider these two dimensions together since they refer to the same place. Even an anonymous or generalized POI name is easily revealed by its exact spatial position. In addition, we argue that spatial and temporal dimensions should be associated in privacy methods as they significantly reveal mobility patterns.

Geographical information can not be dissociated of its semantics in anonymization methods. This includes the latitude and longitude, the POI name, and time, which are specific information that are associated with a single point. By anonymizing just one of these dimensions, it can be possible with the other dimensions, to a malicious attacker, infer the original place. This means that time and semantics related to the spatial dimension, i.e., the POI name, compose significant units of user trajectories and anonymizing just one of them is not enough.

In the conceptual model for multiple aspect trajectories proposed by Mello et al. [2019], a point, an entire trajectory or subtrajectory, a moving object and a relationship of moving objects can be enriched with aspects. *Permanent aspects* are associated with a moving object and they hold during the entire life of the moving object (e.g. place and date of birth, gender). When an aspect does not change during an entire trajectory, it is called a *long term aspect* (e.g. the job of a person or a disease), and it is associated to the multiple aspect trajectory. Both *Long term* and *permanent* aspects can be very sensitive to users privacy. These aspects were not considered in previous models, and by

consequence, not in anonymization methods. We believe that these kind of information is very important for many applications, but they should be treated in the anonymyzation process.

In the MASTER model, *Volatile aspects* represent the information related to the points of a trajectory. Only this type of information was considered in existing anonymization works. The big challenge is that now we are not limited to spatial coordinates, time and POI name. With multiple aspect trajectories we can have any kind of information associated to trajectory points (e.g. the mood of the person, the transportation mean he/she is using, the rating and the price range of a POI, a social network post), and this new kind of information can also be used to identify a person.

One isolated aspect such as time can be an identifier for one user, but to another it may not. For instance, a user that leaves home at a specific time at night. Being the only user to do that in the database and an attacker knowing the time he does it, this behaviour allows inferring his identity. Now consider the combination of multidimensional aspects of a user: the more data are available, the easier it is to re-identify someone. Methods as *Movelets* [Ferrero et al. 2018] and *MASTERMovelets* [Ferrero et al. 2019] can explore all dimensions and reveal the main characteristics that distinguish an individual from the others in the database. Identifying what distinguishes each user is a future challenge to privacy research. In summary, the multiple aspect representation is a big issue in future trajectory data analysis and a challenge for privacy protection researchers.

5. Acknowledgements

This work has been partially supported by the Brazilian agencies CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Finance Code 001), CNPQ (Conselho Nacional de Desenvolvimento Científico e Tecnológico).

References

Abul, O., Bonchi, F., and Nanni, M. (2008). Never walk alone: Uncertainty for anonymity in moving objects databases. In *Proceedings - International Conference on Data Engineering*, pages 376–385.

Abul, O., Bonchi, F., and Nanni, M. (2010). Anonymization of moving objects databases by clustering and perturbation. *Information Systems*, 35(8):884–910.

Aggarwal, C. C. and Yu, P. S. (2008). A General Survey of Privacy-Preserving Data Mining Models and Algorithms. *Privacypreserving data mining*, pages 11–52.

Ağır, B., Huguenin, K., Hengartner, U., and Hubaux, J.-P. (2016). On the Privacy Implications of Location Semantics. *Proceedings on Privacy Enhancing Technologies*, 2016(4):165–183.

Bennett, C. J. (2018). The European General Data Protection Regulation: An instrument for the globalization of privacy standards? *Information Polity*, 23(2):239–246.

Bogorny, V., Renso, C., de Aquino, A. R., de Lucca Siqueira, F., and Alvares, L. O. (2014). Constant-A conceptual data model for semantic trajectories of moving objects. *Transactions in GIS*, 18(1):66–88.

Chow, C.-Y. and Mokbel, M. F. (2011). Trajectory privacy in location-based services and data publication. *ACM SIGKDD Explorations Newsletter*, 13(1):19.

Cots, M. and Oliveira, R. (2018). Lei geral de proteção de dados pessoais comentada.

Dai, Y., Shao, J., Wei, C., Zhang, D., and Shen, H. T. (2018). Personalized semantic trajectory privacy preservation through trajectory reconstruction. *World Wide Web*, 21(4):875–914.

Domingo-Ferrer, J. and Trujillo-Rasua, R. (2012). Microaggregation- and permutation-based anonymization of movement data. *Information Sciences*, 208:55–80.

Dong, Y. and Pi, D. (2018). Novel Privacy-preserving algorithm based on frequent path for trajectory data publishing. *Knowledge-Based Systems*, 148:55–65.

Ferrero, C. A., Alvares, L. O., and Bogorny, V. (2016). Multiple aspect trajectory data analysis: Research challenges and opportunities. *Proceedings of the Brazilian Symposium on GeoInformatics*, 2016-November:56–67.

Ferrero, C. A., Alvares, L. O., Zalewski, W., and Bogorny, V. (2018). MOVELETS: Exploring relevant subtrajectories for robust trajectory classification. *Proceedings of the ACM Symposium on Applied Computing*, pages 849–856.

Ferrero, C. A., Petry, L. M., Alvares, L. O., Zalewski, W., and Bogorny, V. (2019). Discovering Heterogeneous Subsequences for Trajectory Classification. *Data Mining and Knowledge Discovery (accepted for publication)*.

Fileto, R., May, C., Renso, C., Pelekis, N., Klein, D., and Theodoridis, Y. (2015). The Baquara² knowledge-based framework for semantic enrichment and analysis of movement data. *Data and Knowledge Engineering*, 98:104–122.

Giotakis, S. and Pelekis, N. (2019). On preserving sensitive information of multiple aspect trajectories in-house. *The Web Conference 2019 - Companion of the World Wide Web Conference, WWW 2019*, pages 515–522.

Gramaglia, M., Fiore, M., Tarable, A., and Banchs, A. (2017). $k^{\tau,\epsilon}$-anonymity: Towards Privacy-Preserving Publishing of Spatiotemporal Trajectory Data. *arXiv preprint arXiv:1701.02243*, abs/1701.0(iv).

Gurung, S., Lin, D., Jiang, W., Hurson, A., and Zhang, R. (2014). Traffic information publication with privacy preservation. *ACM Transactions on Intelligent Systems and Technology*, 5(3):1–26.

Hornsby, K. S. and Cole, S. (2007). Modeling moving geospatial objects from an event-based perspective. *Transactions in GIS*, 11(4):555–573.

Huo, Z., Meng, X., Hu, H., and Huang, Y. (2012). You Can Walk Alone Trajectory Privacy-preserving through Stay Point Protection. In *International conference on database systems for advanced applications*, pages 351–366.

Kopanaki, D., Theodossopoulos, V., Pelekis, N., Kopanakis, I., and Theodoridis, Y. (2016). Who cares about others' privacy: Personalized anonymization of moving object trajectories. *Advances in Database Technology - EDBT*, 2016-March:425–436.

Liu, B., Zhou, W., Zhu, T., Gao, L., and Xiang, Y. (2018). Location Privacy and Its Applications: A Systematic Study. *IEEE Access*, 6:17606–17624.

Llp, W. (2016). EU General Data Protection Regulation Finally Adopted. *Official Journal of the European Union*, L119(April):1–3.

Machanavajjhala, A., Gehrke, J., Kifer, D., and Venkitasubramaniam, M. (2006). l-

Diversity: Privacy beyond k-anonymity. In *Proceedings - International Conference on Data Engineering*, volume 2006, page 24. IEEE, ACM Trans.

Mahdavifar, S., Abadi, M., Kahani, M., and Mahdikhani, H. (2012). A clustering-based approach for personalized privacy preserving publication of moving object trajectory data. In *Lecture Notes in Computer Science*, volume 7645 LNCS, pages 149–165.

Mello, R. d. S., Bogorny, V., Alvares, L. O., Santana, L. H. Z., Ferrero, C. A., Frozza, A. A., Schreiner, G. A., and Renso, C. (2019). MASTER: A multiple aspect view on trajectories. *Transactions in GIS*.

Monreale, A., Trasarti, R., Pedreschi, D., Renso, C., and Bogorny, V. (2011). C-safety: A framework for the anonymization of semantic trajectories. *Transactions on Data Privacy*, 4(2):73–101.

Pelekis, N., Gkoulalas-Divanis, A., Vodas, M., Kopanaki, D., and Theodoridis, Y. (2011). Privacy-aware querying over sensitive trajectory data. *International Conference on Information and Knowledge Management, Proceedings*, pages 895–904.

Pensa, R. G., Monreale, A., Pinelli, F., and Pedreschi, D. (2008). Pattern-preserving k-anonymization of sequences and its application to mobility data mining. In *CEUR Workshop Proceedings*, volume 397, pages 44–60.

Poulis, G., Skiadopoulos, S., Loukides, G., Gkoulalas, A., and Gkoulalas-Divanis, A. (2014). Apriori-based algorithms for k^m-anonymizing trajectory data. *Transactions on Data Privacy*, 7(2):165–194.

Saygin, Y., Nergiz, M. E., Atzori, M., and Guc, B. (2009). Towards Trajectory Anonymization:\na Generalization-Based Approach. *Transactions on Data Privacy*, 2(106):47–75.

Shaham, S., Ding, M., Liu, B., Lin, Z., and Li, J. (2019). Machine Learning Aided Anonymization of Spatiotemporal Trajectory Datasets. *arXiv preprint arXiv:1902.08934*, pages 1–6.

Spaccapietra, S., Parent, C., Damiani, M. L., de Macedo, J. A., Porto, F., and Vangenot, C. (2008). A conceptual view on trajectories. *Data and Knowledge Engineering*, 65(1):126–146.

Sui, K., Zhao, Y., Liu, D., Ma, M., Xu, L., Zimu, L., and Pei, D. (2016). Your trajectory privacy can be breached even if you walk in groups. *2016 IEEE/ACM 24th International Symposium on Quality of Service, IWQoS 2016*, pages 0–5.

Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowlege-Based Systems*, 10(5):557–570.

Wagner, I. and Eckhoff, D. (2018). Technical privacy metrics: a systematic survey. *ACM Computing Surveys (CSUR)*, 51(3):57.

Wang, J., Luo, Y., Zhao, Y., and Le, J. (2009). A Survey on Privacy Preserving Data Mining. *2009 First International Workshop on Database Technology and Applications*, pages 111–114.

Ye, H., Cheng, X., Yuan, M., Xu, L., Gao, J., and Cheng, C. (2016). A survey of security and privacy in big data. *2016 16th International Symposium on Communications and Information Technologies, ISCIT 2016*, pages 268–272.