Trajectory Data Privacy: Research Challenges and Opportunities

Tarlis Tortelli 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.

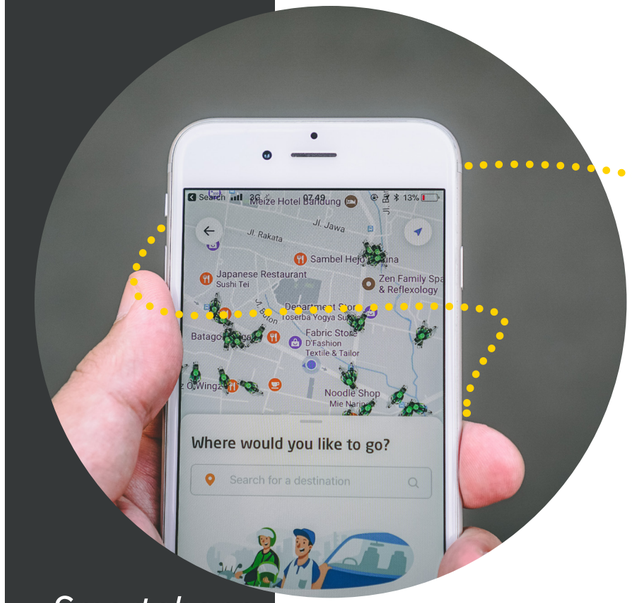
tarlis.portela@posgrad.ufsc.br, francisco.vicenzi@grad.ufsc.br, vania.bogorny@ufsc.br

Agenda

1. Motivation
2. Trajectory Representation Evolution
3. Problem Statement
4. Privacy and Anonymization Basic Concepts
5. Trajectory Anonymization Methods
6. Research Challenges and Opportunities

References

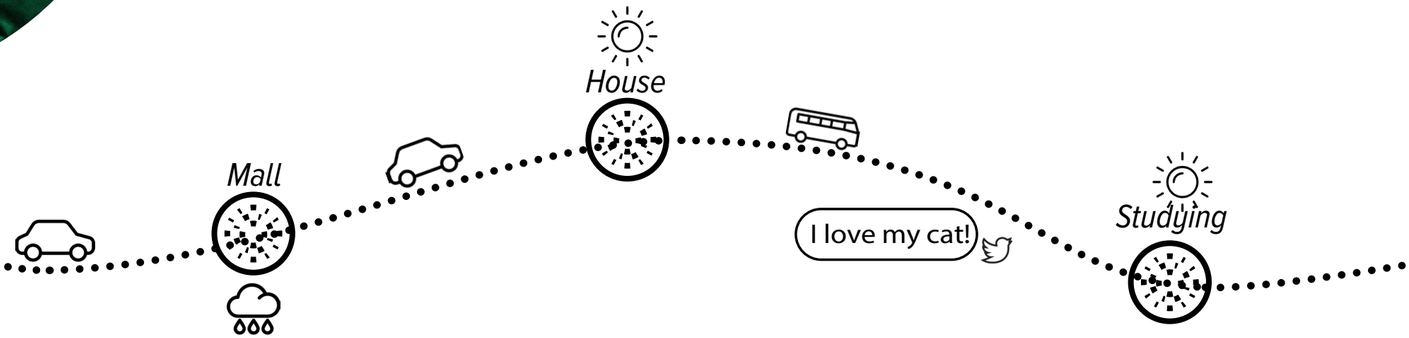
1 . MOTIVATION



Smartphone Application

Popularization and price reduction of mobile devices

- Large volumes of mobility data;
- **Big Data.**



1 . MOTIVATION

Registering traces about our daily routines

Multidimensional and sequential data

Moving Object Trajectories

2. Trajectory Representation Evolution

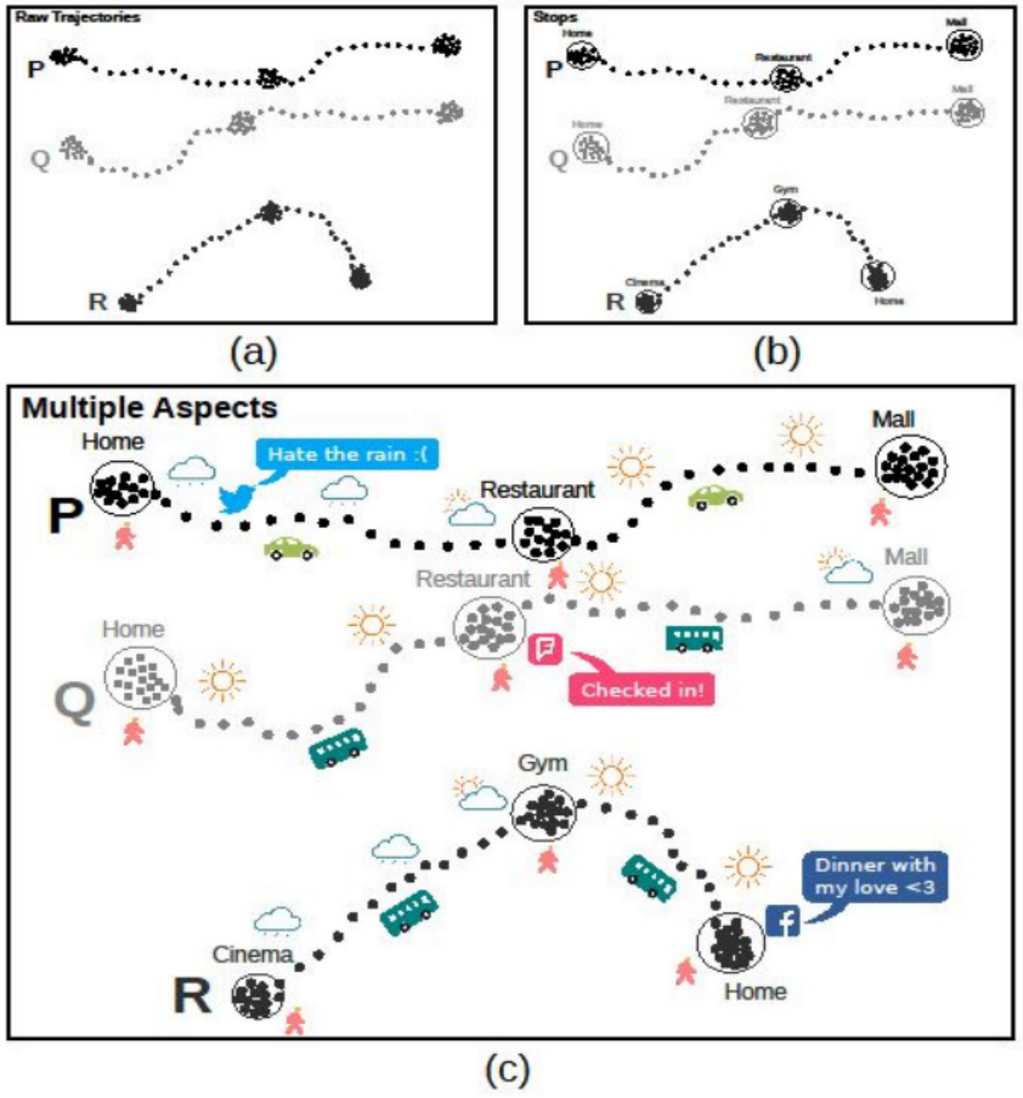


FIGURE (a) < **2007 Raw Trajectories**

FIGURE (b) = **2007 / 2008 Semantic Trajectories**

(Spaccapietra et. al. 2008)

FIGURE (c) **In 2019 Multiple Aspect Trajectories**

MASTER MODEL [Mello et. al. 2019]

3 Problem Statement

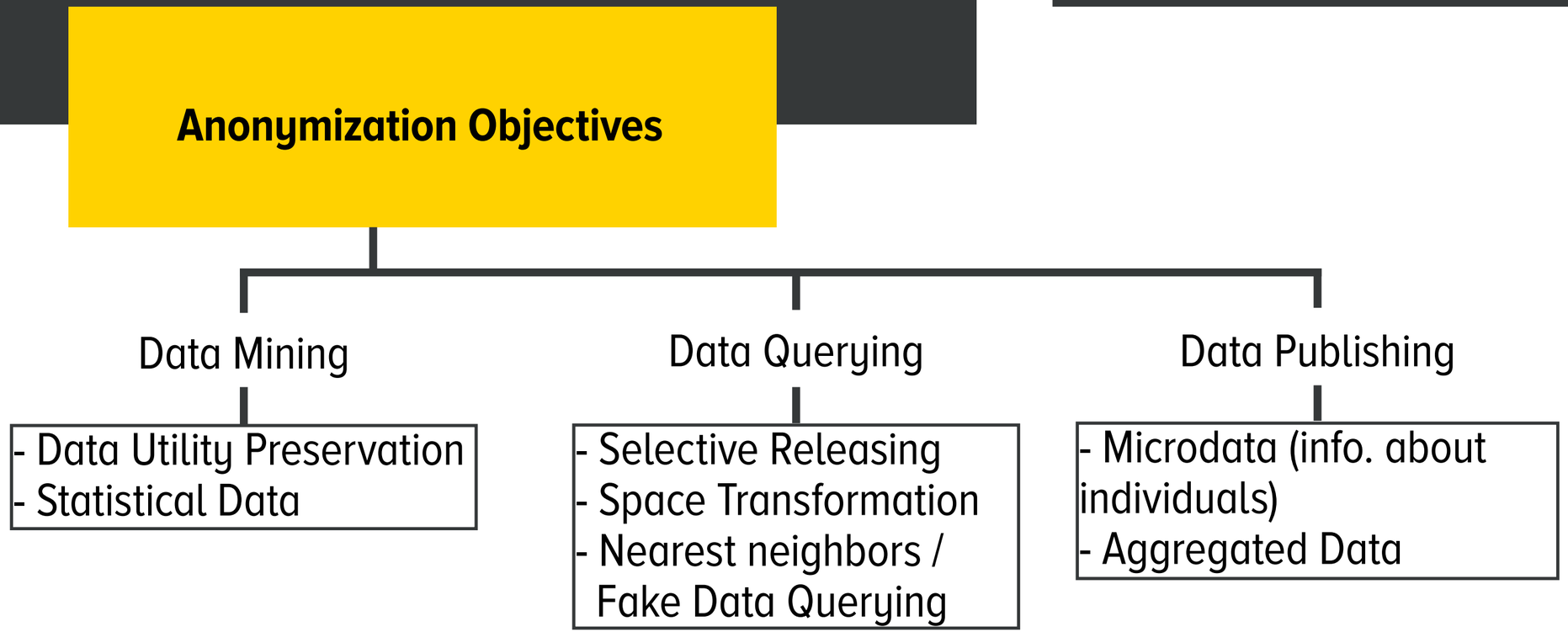
Privacy concerns and **violations**

Re-ID ⟶ **joining external data**

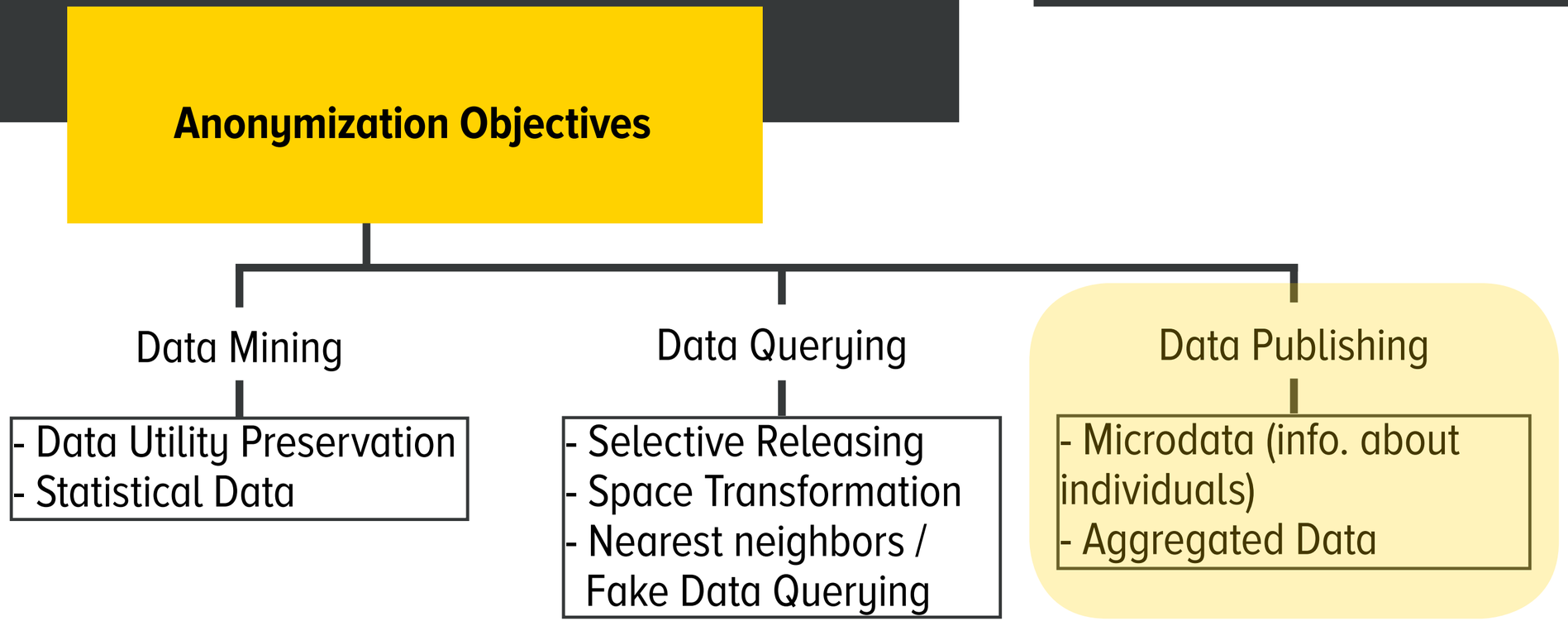
Identifying behaviors:

- A frequently visited place;
- Sequence of two visited places;
- Only the time.

4. Privacy and Anonymization Basic Concepts

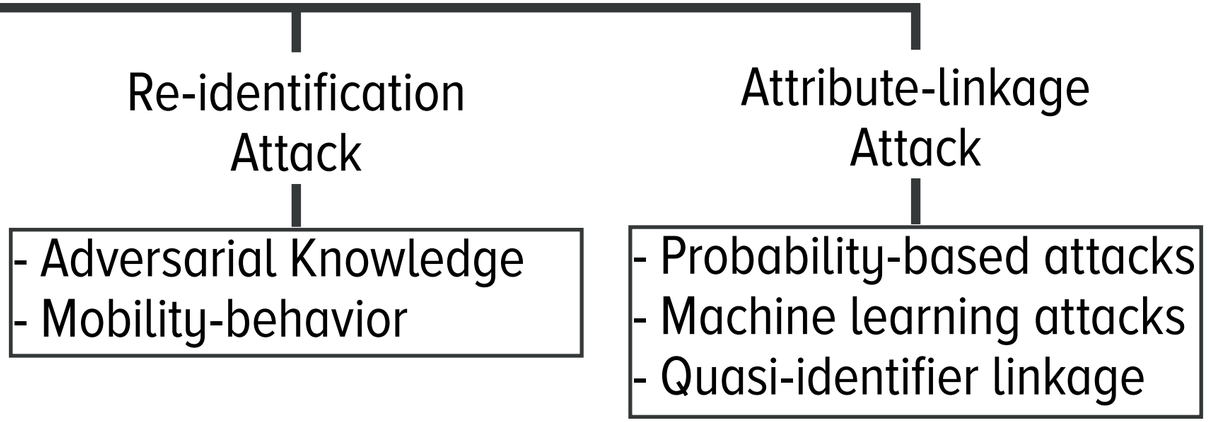


4. Privacy and Anonymization Basic Concepts



4. Privacy and Anonymization Basic Concepts

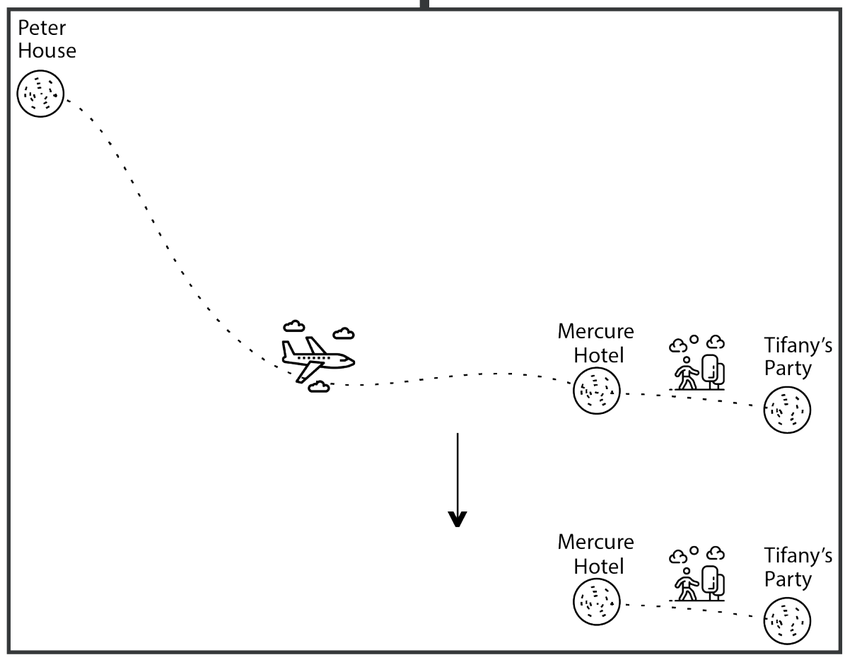
Attack Models



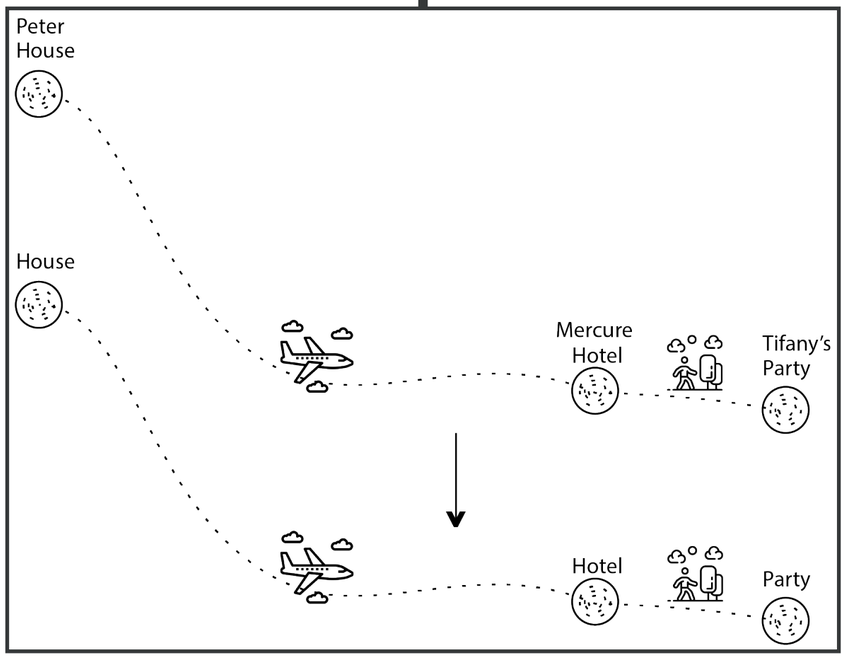
4. Privacy and Anonymization Basic Concepts

Anonymization Techniques

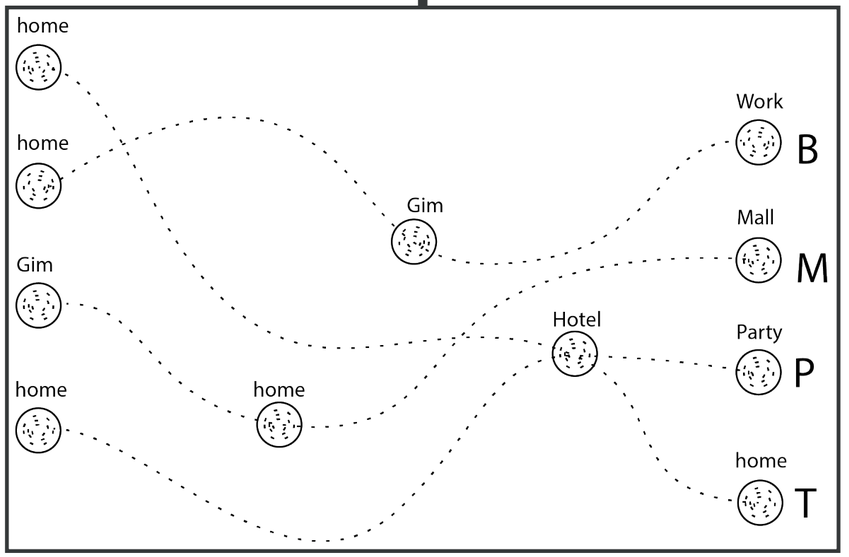
Suppression



Generalization



Masking



5. Trajectory Anonymization Methods

Trajectory Anonymization Methods

Raw Trajectory

Spatial Dimension

- NWA [Abul et al. 2008]
- P2kA [Pensa et al. 2008]
- anonTraj [Nergiz et al. 2009]
- CTR [Mahdavifar et al. 2012]
- WCOP [Kopanaki et al. 2016]

Spatio-Temporal Dimensions

- W4M [Abul et al. 2010]
- Location Permutation [Domingo-Ferrer and Trujillo-Rasua 2012]
- ICBA [Gurung et al. 2014]
- k^t, -anonymity [Gramaglia et al. 2017]
- TOPF [Dong and Pi 2018]
- DynamicSA [Shaham et al. 2019]

Semantic Trajectory

Semantic Dimension (POI)

- CAST [Monreale et al. 2011]
- SeqAnon (framework) [Poulis et al. 2014]

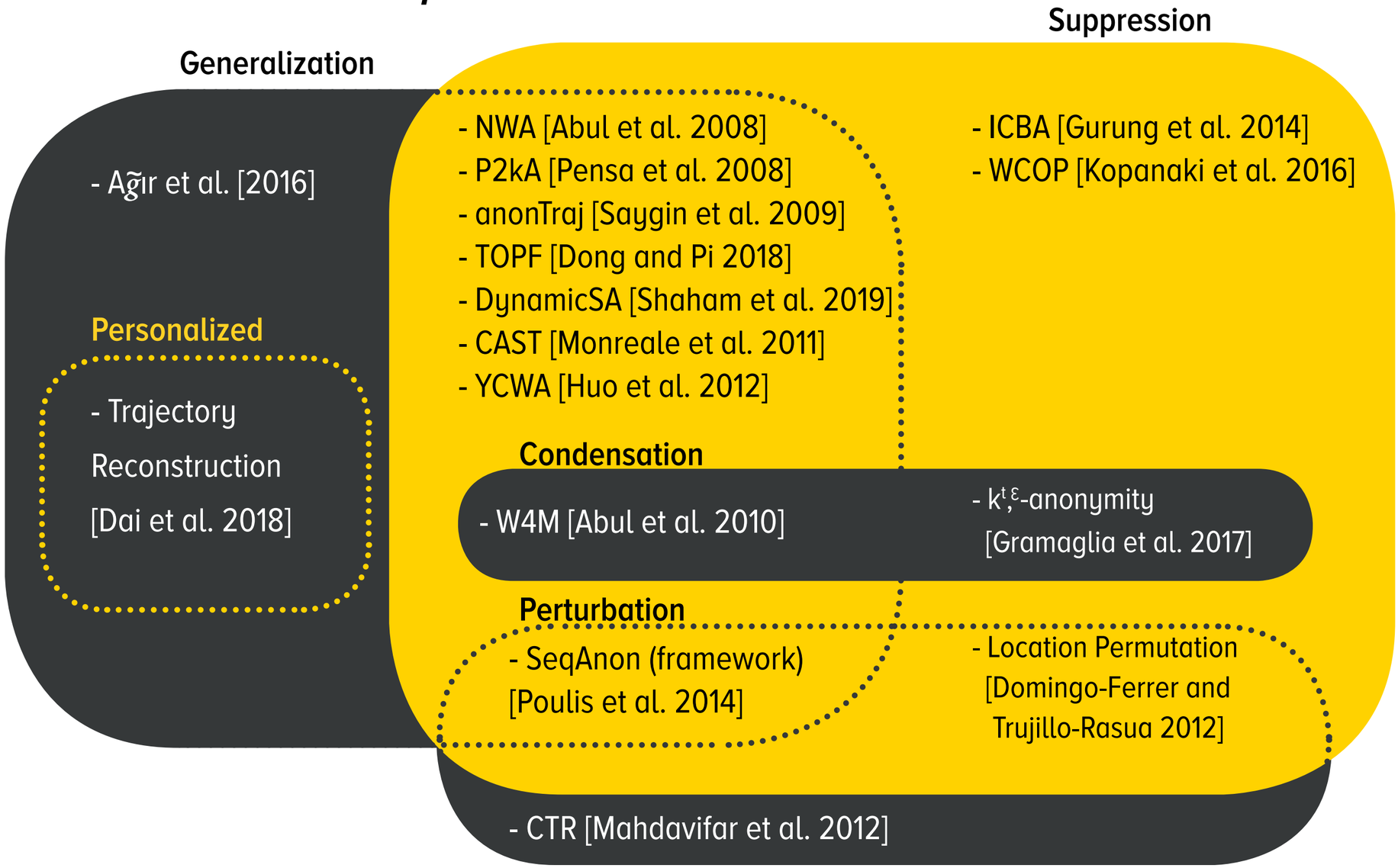
Spatial, Temporal and Semantic Dimensions

- YCWA [Huo et al. 2012]
- [Agir et. al. 2016]
- Trajectory Reconstruction [Dai et al. 2018]

Multiple Aspec Trajectory

5. Trajectory Anonymization Methods

Anonymization Technique



5. Trajectory Anonymization Methods

Datasets

Orange CDR (call detail records)

- $k^{t},^{\varepsilon}$-anonymity [Gramaglia et al. 2017]

Twitter-Foursquare

- Ağır et al. [2016]

GeoLife

- WCOP [Kopanaki et al. 2016]
- DynamicSA [Shaham et al. 2019]
- YCWA [Huo et al. 2012]
- Trajectory Reconstruction [Dai et al. 2018]*

GeoLife

- P2kA [Pensa et al. 2008]
- CAST [Monreale et al. 2011]

Pisa

Brinkhoff's Oldenburg / Synthetic

- anonTraj [Saygin et al. 2009]*
- CTR [Mahdavifar et al. 2012]
- ICBA [Gurung et al. 2014]*
- TOPF [Dong and Pi 2018]
- W4M [Abul et al. 2010]

San Francisco Taxis

- Location Permutation [Domingo-Ferrer and Trujillo-Rasua 2012]

Gowalla

- SeqAnon (framework) [Poulis et al. 2014]

Trucks

- NWA [Abul et al. 2008]

* Uses synthetic data

5. Trajectory Anonymization Methods

Comparison with Other Methods

Nome

- **P2kA [Pensa et al. 2008]**

- **NWA [Abul et al. 2008]**

- anonTraj [Saygin et al. 2009]
- CTR [Mahdavifar et al. 2012]
- WCOP [Kopanaki et al. 2016]
- CAST [Monreale et al. 2011]
- Agır et al. [2016]
- Trajectory Reconstruction [Dai et al. 2018]

- **$k^t,^\varepsilon$-anonymity [Gramaglia et al. 2017]**

- **ICBA [Gurung et al. 2014]**

- TOPF [Dong and Pi 2018]

- W4M [Abul et al. 2010]
- Location Permutation [Domingo-Ferrer and Trujillo-Rasua 2012]
- TOPF [Dong and Pi 2018]
- YCWA [Huo et al. 2012]

- DynamicSA [Shaham et al. 2019]

Others for query answering

- SeqAnon (framework) [Poulis et al. 2014]

More Details
Easier to ID

6. Research Challenges and Opportunities

PPDP ⟶ raw or semantic trajectories

Together, not dissociated:

- Latitude and longitude, the POI name, and time ⟶ single point

Exploring dimensions (MASTER model)

Challenge and Opportunity to privacy research

6. Research Challenges and Opportunities

Time alone

Each user ⟶ **different**

Movelets [Ferrero et al. 2018] and **MASTER Movelets** [Ferrero et al. 2019]

- Explore all dimensions
- Characteristics that distinguish

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. Privacy preserving 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.

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.

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 pathfor 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). TheBaquara knowledge-based framework for semantic enrichment and analysis of movement data. Data and Knowledge Engineering, 98:104–122.

References

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). $k^{t},^{\varepsilon}$-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.

References

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 km-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: a Generalization-Based Approach. Transactions on Data Privacy,2(106):47–75.

Shaham, S., Ding, M., Liu, B., Lin, Z., and Li, J. (2019).Machine Learn-ing 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.

Thank You!

Trajectory Data Privacy: Research Challenges and Opportunities

Tarlis Tortelli 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.portela@posgrad.ufsc.br, francisco.vicenzi@grad.ufsc.br, vania.bogorny@ufsc.br