# Praxis der Forschung

Patricia Guerra-Balboa
Chair of IT Security , Karlsruhe Institute of Technology, Karlsruhe, Germany

## 1 Generative graphs models

Social networks, medicine or traffic management are just a few of the innumerable applications of graph theory in data analysis. The great advantage of graph analysis is the ability to save data structure and relation properties. On the other hand, we find the growing of synthetic data as an incredible tool that allows us to simulate not yet encountered conditions, is immune to some common statistic problems and has good properties in terms of scalability. However synthetic data generation finds a new branch of study when we want to generate graphs as these ones need special tools, both in traditional generative approaches and machine learning ones, due to structure and non-independence of these data. Although synthetic data has numerous advantages, in terms of privacy, training databases are susceptible to Membership Inference Attacks so we need to enforce our mechanisms to obtain real privacy protection.

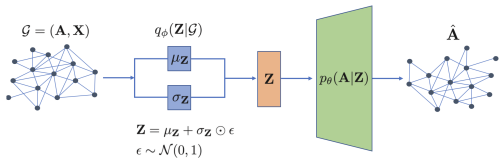


Figure 1: Illustration of a standard VAE model applied to the graph setting

Therefore, the goal of this project is to understand the state of the art in graph data generation both in the traditional approaches and the deep generative models, and which attempts makes sense to try in order to introduce differential privacy mechanisms inside of the generative model. After grasp this content, the next steps would be the development of a DP-generative algorithm for graphs and the performance of some experiments on it.

The use case during all the project is going to be trajectory data management. Depending on how the process unfolds we could present results for more general graphs or experiments more focus on trajectory data graphs.

**Topics:** Graph generative models • Deep learning • Synthetic Data • Differential Privacy

**references:** Chapters 8 and 9 of Hamilton (2020) and Qin et al. (2017)

## 2 Model-based anonymization of GPS trajectories

Everyday, the value of and interest in location and trajectory data increases. Not only can their processing improve our daily lives, for instance through navigation and recommendation, but also various institutional data analysis applications in both the public and private sector. Traffic management, urban planning, transportation system design, routing advice, or homeland security are just a few of the many applications that benefit from trajectory analyses. Yet, it entails extensive privacy risk, as trajectory data is extremely privacy-invasive. Trajectories may reveal accurate behavioral patterns, allowing attackers to infer sensitive aspects of an individual's life, including health status, religious beliefs, social relationships, or sexual preferences. For this reason the not trivial task of anonymize these type of data is the mean proposal of this project.

In order to achieve privacy protection we relay on Differential Privacy as mean notion. Therefore the goal of this project is to understand the s-o-t-a in deferentially private trajectory data release, and

focusing in the ideas of some DP-generative mechanism as the one presented on Gursoy, Liu, Truex, and Yu (2018), develop a mechanism that output a perturbed and protected version of the original database. Finally experiments and evaluations will be perform to test the mechanism.

# References


Gursoy, M. E., Liu, L., Truex, S., & Yu, L. (2018). Differentially private and utility preserving publication of trajectory data. *IEEE Transactions on Mobile Computing*, *18*(10), 2315–2329.

Hamilton, W. L. (2020). *Graph representation learning* (Vol. 14) (No. 3). Morgan & Claypool Publishers.

Qin, Z., Yu, T., Yang, Y., Khalil, I., Xiao, X., & Ren, K. (2017). Generating synthetic decentralized social graphs with local differential privacy. , 425–438.