

Use of semantic pertinence to improve accuracy in differentially private projection processes

Project: This internship is part of project Semantic Networks of Data: Utility and Privacy (SENDUP¹) that studies data privacy in graph databases (e.g., RDF) with underlying semantics.

Laboratory & Team: LIFO, Systems and Data Security team, INSA Centre Val de Loire, 88 boulevard Lahitolle 18022 Bourges

Duration and start: 5 to 6 months, at the candidate's earliest convenience. **Contact:** to apply or request additional information, send a mail and a resume to adrien.boiret@insa-cvl.fr

Requirements:

- Master, Bac +5 in computer science / engineering or equivalent
- Knowledge or interest about databases (especially graph databases, e.g. RDF) and data privacy
- Ability to read and write english documents
- Proficiency in a coding language (preference for Java)
- Willingness to work in autonomy and in a team

Subject:

Data safety and privacy are concerns currently receiving intense attention, notably through the introduction of GDPR reglementations that aim to ensure data collection, treatment, and publication never trespass on a person's right to privacy. The notion of differential privacy (DP) grew popular as a yardstick of privacy for data publication processes, where a database containing sensitive information can still answer queries without compromising privacy.

To facilitate the creation of processes that guaranty DP in graph databases, one method is projection², that limits the degree of graphs by selecting for each node a bounded number of edges. However, if the choice of edges to maintain is arbitrary, the impact of this method on the accuracy of the published results is too variable to guarantee accuracy.

In this internship, we aim to develop methods that allow the priorization of certain edges above others in the projection process, as to improve the accuracy of query results. This optimization would be based on prior knowledge of one or several queries we know to be of interest.

Goals and Objectives:

- Elaborating rules for edge priorizations
- Implementation of projection following those rules
- Experimental and/or theoretical study of their impact on both query accuracy and data privacy

Keywords: Data privacy, Differential privacy, Databases, Graph databases, Path queries

¹https://www.univ-orleans.fr/lifo/evenements/sendup-project/

²https://benjamin-nguyen.fr/papers/adbis2022.pdf