Internship Proposal

AI-based Profiling of Graph Data

Context

Graph-structured data are ubiquitous in modern information systems, arising in domains such as social networks, knowledge graphs, biological networks, software dependency graphs, and communication infrastructures. Profiling such data—i.e., extracting structural, statistical, and semantic characteristics—is a crucial task for data understanding, quality assessment, optimization, and downstream analytics.

Traditional graph profiling techniques rely on handcrafted metrics and exhaustive traversal strategies, which may not scale well or capture complex structural patterns. Recent advances in Artificial Intelligence open new opportunities to automatically learn informative profiles of graph data. This internship focuses on the use of AI techniques for profiling graph data, with an emphasis on graphs stored and managed in graph databases such as Neo4j or GraphDB.

Objectives

- Study existing approaches for graph profiling, graph analytics, and graph data management.

- Explore AI-based methods (e.g., graph embeddings, graph neural networks, or learning-based summarization techniques).

- Design and implement profiling methods that automatically extract relevant characteristics from graph data.

- Integrate and evaluate these methods on graph data stored in graph database systems such as Neo4j or GraphDB.

Expected Outcomes

- Definition of graph profiling tasks suitable for AI-based approaches.

- Prototype implementations mainly developed in Python.

- Experimental evaluation on real-world or synthetic graph datasets.

- A technical report and potential contribution to a research publication.

Required Skills

- Strong background in computer science, data science.

- Good programming skills in Python.

- Knowledge of machine learning; familiarity with deep learning is a plus.

- Prior experience with Neo4j or GraphDB is appreciated but not mandatory.

Working Environment

The internship will take place in the LAMSADE lab, a research-oriented environment with close supervision, regular discussions, and access to relevant datasets, graph database systems, and computational resources.

Supervision

The internship will be supervised by Lyes ATTOUCHE, LAMSADE, pole data science, with possible interaction with PhD students and researchers working on graph data management and AI in the same pole.

Duration and Level

Duration: 5 to 6 months

Level: Master's level (M2) / Engineering school

Application

Interested candidates should submit a CV, a short motivation letter, and academic transcripts of the last three years.

Applications should be sent to: lyes.attouche@dauphine.psl.eu