Centre for AI-Fundamentals
RAEng Google DeepMind Summer Internship Programme 2025

Project proposal

Project Title	Towards XAI and human-in-the-loop approaches for graph structure learning
Lead supervisor	Julia Handl
Project Description	**Project Background:** Graph neural networks (GNNs) provide an elegant tool for the formulation of machine learning problems involving relational data. Standard GNN benchmarks frequently assume that an underlying graph structure is readily accessible, e.g., in the form of a traffic or social network. Yet, there is increasing interest in the potential of GNNs in application domains where the graph itself is not pre-defined, or may need to be refined, as part of the learning process. Largely, such work on graph structure learning focuses on data-driven, automated approaches, i.e. methodologies that derive plausible relations using heuristics that draw on feature or label information, or optimize the final GNN's generalization performance. Without the introduction of problem-specific insight, the problem of learning the optimal graph structure for a given GNN model is untractable. Furthermore, our current understanding of the interaction between graph structure and learning performance remains poor. Consequently, there is a need to develop explainable AI (XAI) approaches that can support (i) the interactive involvement of domain experts in the process of graph structure learning, (ii) the exploration of the impact of graph structure on learning performance. This project will provide an opportunity to explore this problem space. Specifically, you will implement a GNN, a graph structure learning approach and simple XAI approaches for a given data set. For a particularly ambitious student, there will be the opportunity to explore creative ways in which the impact of graph structure could be communicated to users, or that user feedback could be integrated into the structure learning approach.

Work Plan:

Week 1: Understanding graph neural networks. Development of a first GNN model in a Python package of choice.
Week 2: Debugging; Experimentation with the GNN model on standard benchmark data sets. Interpretation of results.
Week 3: Understanding the problem of graph structure learning Development of a structure learning approach of choice.
Week 4: Debugging; Comparison of the GNN model across different graphs. Interpretation of results.
Week 5: Understanding how simple statistics / visualization approaches can support an understanding of graph structure. Development of a set of techniques of choice.
Week 6: Analysis of the visualization results, aiming to link insight to the models' prediction performance.
Week 7: Write-up, final code documentation and presentation.

Project Objectives:

The core objectives of this project are as follows:

1) Develop and test a GNN model using a benchmark data set with a pre-defined graph structure.
2) Implement and empirically evaluate an existing approach to graph structure learning.
3) Implement simple XAI approaches for graph structure learning, drawing on existing statistics and visualization packages within Python.

Learning Outcomes:

The learning outcomes of this project are as follows:

1) Understand GNNs, as frontier ML approaches, and their similarities and differences to other machine learning paradigms.
2) Develop competency in a specialized Python package for GNNs, e.g. Pytorch or Tensorflow.
3) Understand existing approaches to graph structure definition / learning in a GNN context.
4) Understand the role of XAI in increasing the transparency of machine learning models.