GAIN
Graphs in Artificial Intelligence
and Neural Networks

Seminar or Project Description: Explainable Graph Neural Networks

Graphs are omnipresent in nature and with a rising amount of data available in form of graphs, Graph Neural Networks (GNNs) have become a vibrant research field. Yet, most GNN algorithms are black-boxes, despite the fact that graphs inherently show connections between entities. To increase the possibility of users applying GNN models and to make sure, GNN models deliver meaningful answers, it is therefore necessary to find methods to explain GNN models. In this sense, explainability means that one can find a reason for the prediction of a model, either during the prediction process or post-hoc. For Neural Networks, that work on non-graph data, e.g. Convolutional Neural Networks, many methods for explainability exist. Due to the different architectures of GNNs, these approaches cannot be used for GNNs or have to be adapted in a non-trivial way. The goal of this project or seminar is to get an overview of existing methods to explain GNN models. For a project, additionally several methods of explainability should be compared for a choosen GNN algorithm.

1 Goals

Create an overview of existing methods of explainability for GNN algorithms including their similarities, differences, adventages and disadventages.

2 Prerequisites

- Knowledge of Neural Networks, ideally at least basic knowledge of Graph Neural Networks
- For a project: Python (ideally pytorch geometric)

3 Student's tasks

- Survey different methods to explain GNNs
- For seminar: Write a seminar paper
- For a project: Implement most important methods in python for one GNN algorithm and compare them, write a project report

4 Supervisor's tasks

- Supervision (online or in person if possible)
- Provide a template for project report or seminar

5 Grading criteria

- Seminar: Academic approach (Literature review, approaches, review corrections, review of other students, ...), Seminar paper, Seminar talk
- Project: Academic approach, quality of code, project report

6 Used Languages

- German or English (recommended)

7 Conditions

- Credits: Depending on the examination regulations

8 References

- https://arxiv.org/pdf/2012.15445.pdf
- https://towardsdatascience.com/explainable-graph-neural-networks-cb009c2bc8ea