Problem Statement and Goals

Yuanqi Xue

Table 1: Revision History

Date	Developer(s)	Change
Jan.22, 2025	Yuanqi Xue	First Draft
Jan.22, 2025	Yuanqi Xue	Updated Based on Feedback (Issue 1)
Jan.30, 2025	Yuanqi Xue	Second Draft

1 Problem Statement

1.1 Problem

Graph Neural Networks (GNNs) have shown strong performance in node classification, graph classification, and link prediction tasks. However, their black-box nature makes it difficult to understand their decision-making process, limiting their usage in critical areas such as medical diagnosis. To address this, the paper *ProtGNN: Towards Self-Explaining Graph Neural Networks* [1] proposes ProtGNN, a GNN model with built-in interpretability.

1.2 Inputs and Outputs

To maintain consistency with [1], we will be using the *MUTAG* [2] dataset.

The input is a graph dataset. The output includes a trained model, classification results for the input graphs, and a set of learned prototypes, representing the key structural characteristics of each class.

1.3 Stakeholders

Researchers or students interested in the reproducibility and validation of [1].

1.4 Environment

For our implementation, we will use an NVIDIA GeForce RTX 3060 GPU for training and testing. However, testing should also be feasible on a personal laptop.

2 Goals

- Implement the ProtGNN model as proposed in [1].
- Reproduce the classification accuracy on the MUTAG [2] dataset, and ensure that the identified prototypes (i.e., key structural characteristics of each class) have comparable quality to the original study.

3 Stretch Goals

Examine the reproducability of the paper [1] and validate its results.

4 Challenge Level and Extras

Challenge Level: Research Project
Extras: None

5 Reference

[1] Z. Zhang, Q. Liu, H. Wang, C. Lu, and C. Lee, "ProtGNN: Towards self-explaining graph neural networks," in Proc. AAAI Conf. Artif. Intell., vol. 36, no. 8, pp. 9127–9135, Jun. 2022.

[2] A. K. Debnath, R. L. Lopez de Compadre, G. Debnath, A. J. Shusterman, and C. Hansch, "Structure-activity relationship of mutagenic aromatic and heteroaromatic nitro compounds. Correlation with molecular orbital energies and hydrophobicity," J. Med. Chem., vol. 34, no. 2, pp. 786–797, Feb. 1991.