On the Feasibility of Fidelity− for Graph Pruning

Yong-Min Shin, Won-Yong Shin

Yonsei University

Seoul, South Korea

IJCAI JEJU 2024

IJCAI 2024 Workshop on Explainable Artificial Intelligence (XAI)

Paper Homepage

Introduction

XAI for graph neural networks (GNNs)

- Many XAI methods for GNNs **highlight local edges** that are **highly relevant** to the output.
- **Edge attribution**: How much can we attribute the model's output to each input edge?

Quantitative measurement: Fidelity-

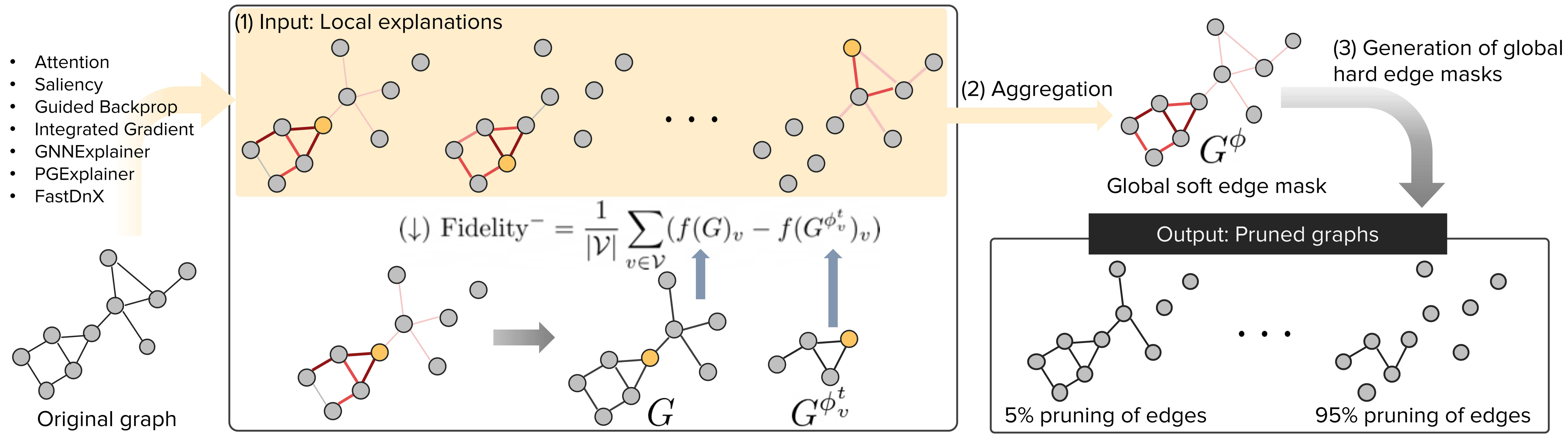
- Output **difference** between two instances when the **original input graph** is used and when the **unimportant edges are removed**.
- If the explanation is valid, then edges deemed less important should have less impact to the model's output after **removal** from the input.

Research Question

Can we improve the efficiency of the GNN model by performing graph pruning by performing graph pruning based on GNN explanations?

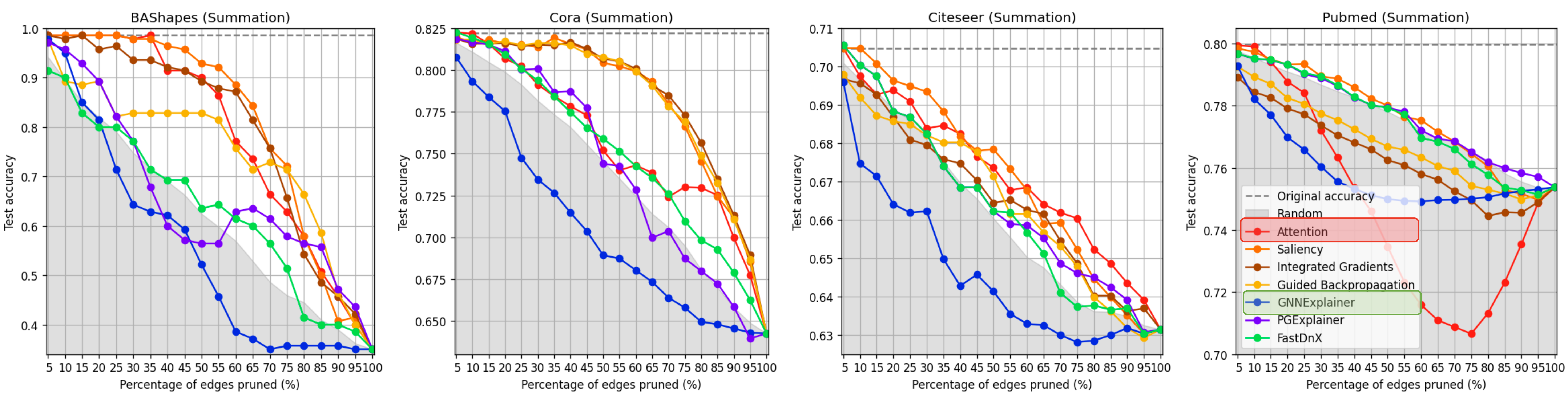
- **Intuition**: If an edge is frequently removed in Fidelity-, it may simply be removed from the original graph.

Overview: FiP (Fidelity-inspired Pruning)



Observations and Discussion

Obs.1: Can Explanation be used for Graph Pruning?



Obs.2: Does Graph Pruning and Fidelity- Scores Translate?

Method	BAShapes	Cora	Citeseer	Pubmed
Att	4.06×10^{-2}	3.67×10^{-2}	2.23×10^{-2}	2.46×10^{0}
SA	3.54×10^{-7}	2.21×10^{-7}	$\mathbf{8.90\times10^{-8}}$	2.46×10^{0}
IG	6.25×10^{0}	1.26×10^{0}	5.68×10^{-1}	$\mathbf{2.25\times10^{0}}$
GB	3.77×10^{0}	1.42×10^{0}	7.04×10^{-1}	2.40×10^{0}
GNNEx	$\mathbf{3.44\times10^{-7}}$	$\mathbf{2.14\times10^{-7}}$	3.52×10^{-1}	2.46×10^{0}
PGEx	3.83×10^{-7}	2.04×10^{-2}	7.11×10^{-3}	2.46×10^{0}
FDnX	1.41×10^{-1}	1.77×10^{-2}	7.05×10^{-3}	2.46×10^{0}

Discussions on Results

1) **Explanations can be used for graph pruning.**
2) Avoid GNN-tailored methods.
3) Graph pruning naturally improve efficiency for GNN.
4) Fidelity does not translate to graph pruning.
5) The problem likely resides in the **practical problem of aggregating local explanation**:
   - Scale of attribution score across nodes
   - Num. of edges for each explanation
6) **Inherent limitation** of graph pruning: Local explanation is unique for each node, **creating a single (pruned) graph is always a lossy compression of information**.