Presented at: IJCAI 2024 workshop on Explainable Artificial Intelligence (XAI), ICC, Jeju Island

On the Feasiblity of Fidelity- for Graph Pruning

Presenter: Yong-Min Shin (jordan7186.github.io)
Supervisor: Prof. Won-Yong Shin

Yonsei University, Seoul, South Korea

IJCAI JEJU 2024

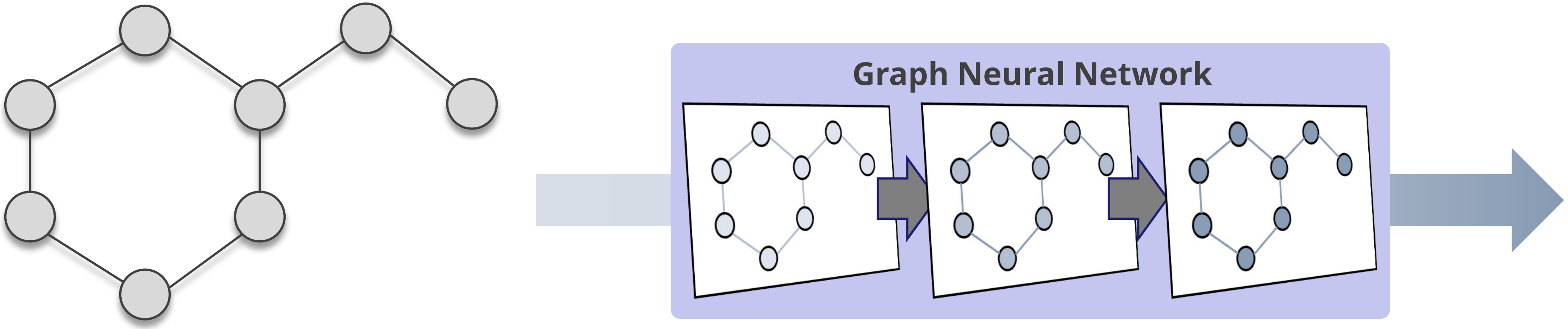
IJCAI 2024
Workshop on Explainable
Artificial Intelligence (XAI)

5th Aug. 2024

수학계산학부(계산과학공학)
School of Mathematics and Computing
(Computational Science and Engineering)

MIDaS LAB
Machine Intelligence & Data Science

Graph Neural Networks have been successfully deployed to learn from graph data and perform various graph tasks.

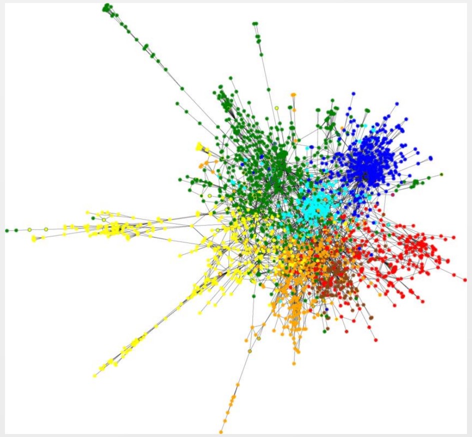


Various graph tasks

Social

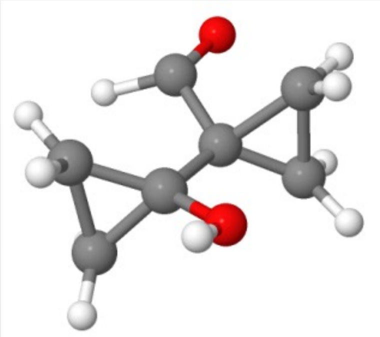
Node: People / Account
Edge: Connection
Node feature: Metadata

Citation / Web



Node: Paper
Edge: Citation
Node feature: Abstract

Molecules

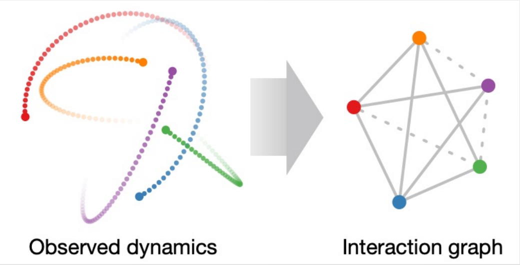


Node: Atom
Edge: Bond
Node feature: Atom type
Edge feature: Bond type

Point Cloud

Node: Data point
Edge: Constructed
Node feature: 3D position

Physical systems

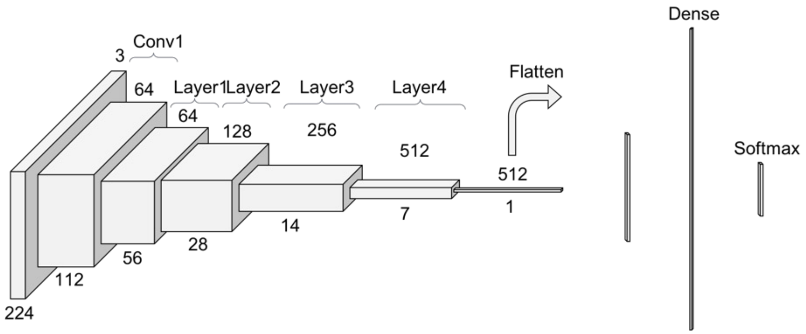


Node: Particles
Edge: Pair-wise interaction
Node feature: Physical properties

Explainable AI is an important research topic, and explanation of Graph Neural Networks is no exception.

Attribution maps are one of the most popular ways, especially in CV and NLP.

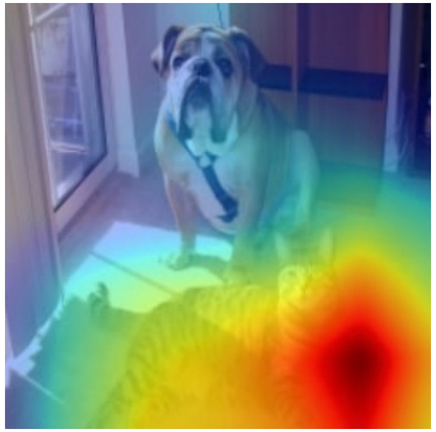
Example: DTD [1], LRP [2], LIME [3], GradCAM [4], …



Input image to ResNet

Output: "Cat"

Result of GradCAM

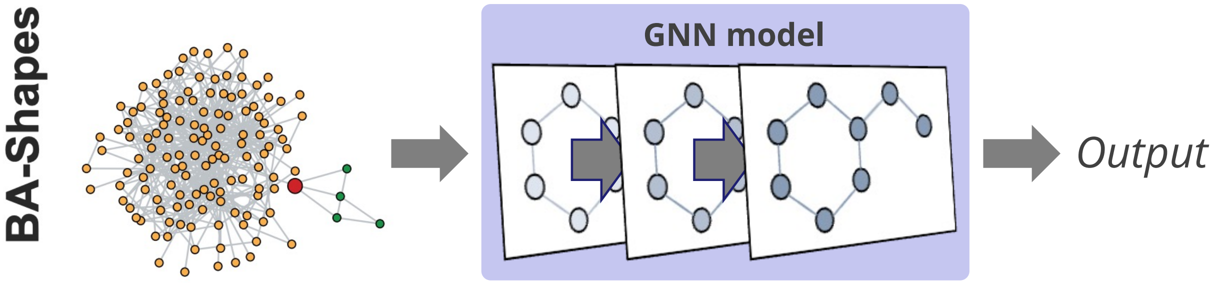


Highlights relevant pixels

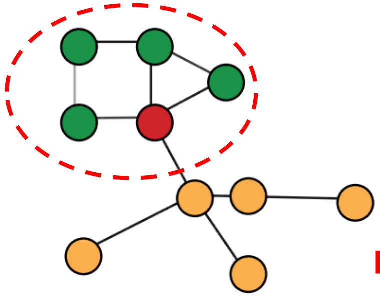
Similar approaches are also popular in **GNN explanations**, too.

Example: GNNExplainer [5], PGExplainer [6], …

Computation graph



Result of Explanation



Highlights relevant subgraph

Early works adopted “general” attribution methods to GNNs, and a plethora of GNN-tailored attribution methods have since been developed.

Early work [7]

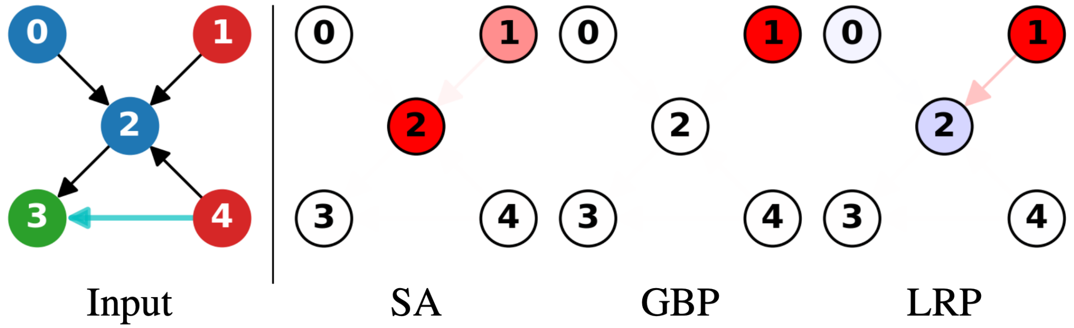
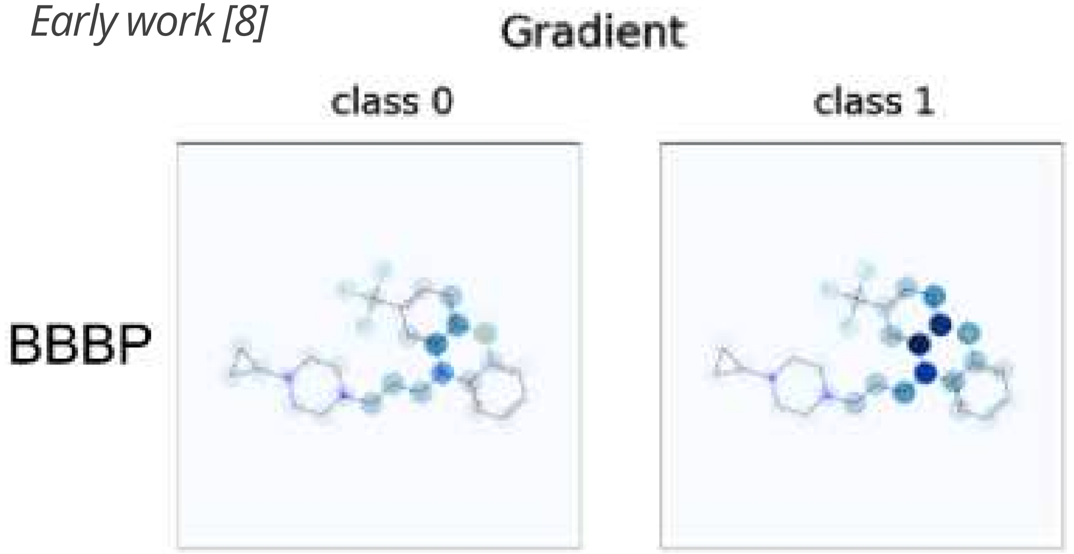


Figure 2: **Explaining why node 2 becomes infected.**

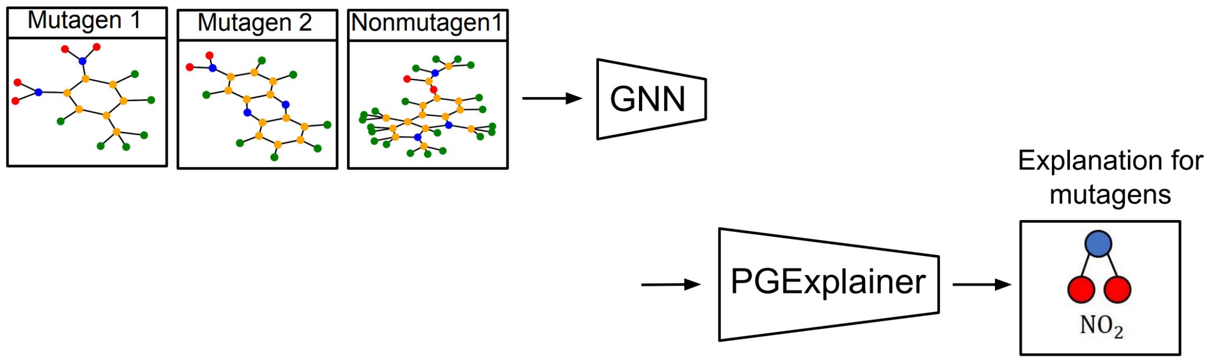
GNNExplainer [5]

$$\max_{G_S} MI\left(Y, (G_S, X_S)\right) = H(Y) - H(Y|G = G_S, X = X_S)$$

Early work [8]

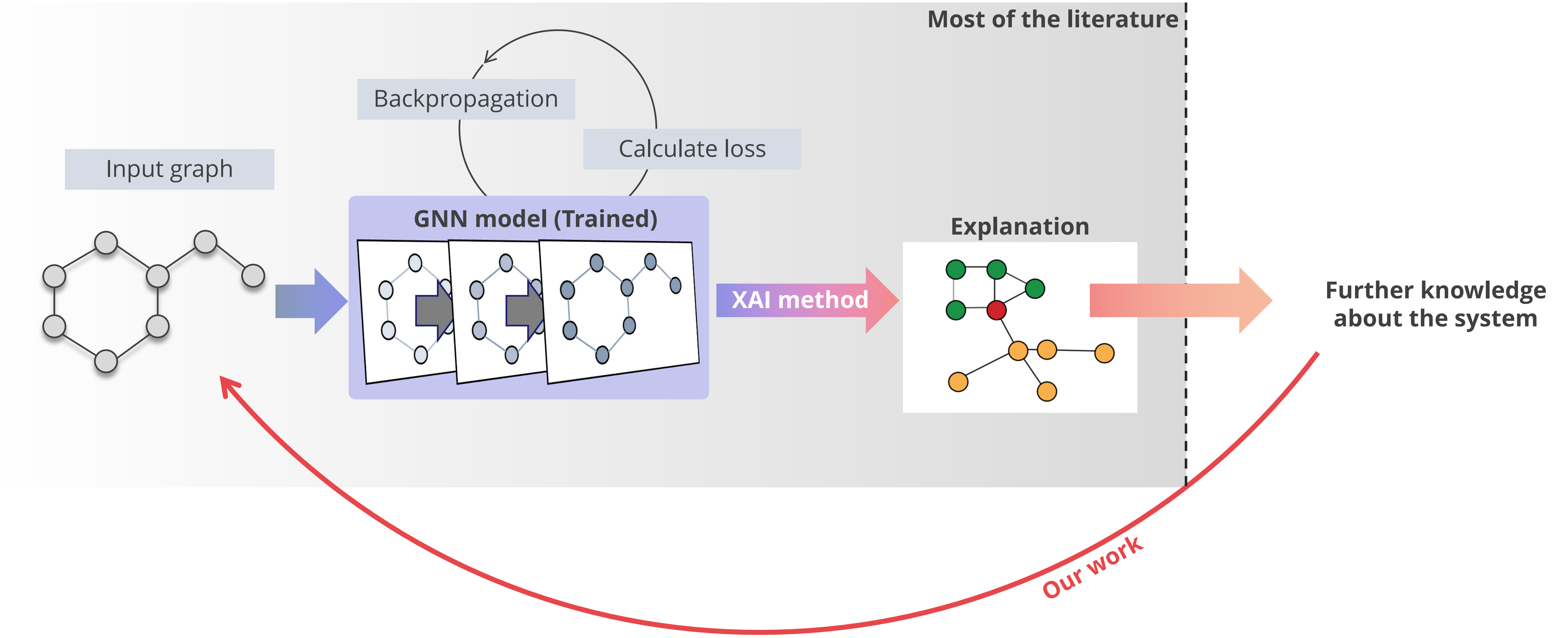


PGExplainer [6]



However, most of the literature focuses on the explanation themselves, but we can go beyond and towards one of the ultimate goal of XAI.

Enhance the performance based on the knowledge gained from the explanation



Our work attempts to observe whether we can directly use node-level explanations in the literature to improve the GNN's efficiency.

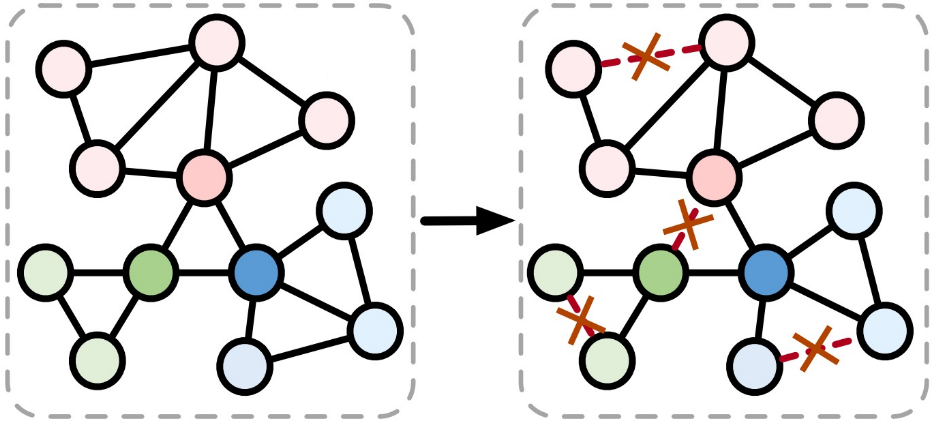
Can we use the local edge attributions for graph pruning?

We focus on improving the efficiency of GNNs by graph pruning, i.e., deletion of unimportant edges.

Time & space complexity is dependent on the number of edges.

	GCN [9]	Vanilla SGD	GraphSAGE [10]	FastGCN [11]	VR-GCN [12]	Cluster-GCN [13]
Time complexity	$O(L\|A\|_0F + LNF^2)$	$O(d^LNF^2)$	$O(r^LNF^2)$	$O(rLNF^2)$	$O(L\|A\|_0F + LNF^2 + r^LNF^2)$	$O(L\|A\|_0F + LNF^2)$
Memory complexity	$O(LNF + LF^2)$	$O(bd^LF + LF^2)$	$O(br^LF + LF^2)$	$O(brLF + LF^2)$	$O(LNF + LF^2)$	$O(bLF + LF^2)$

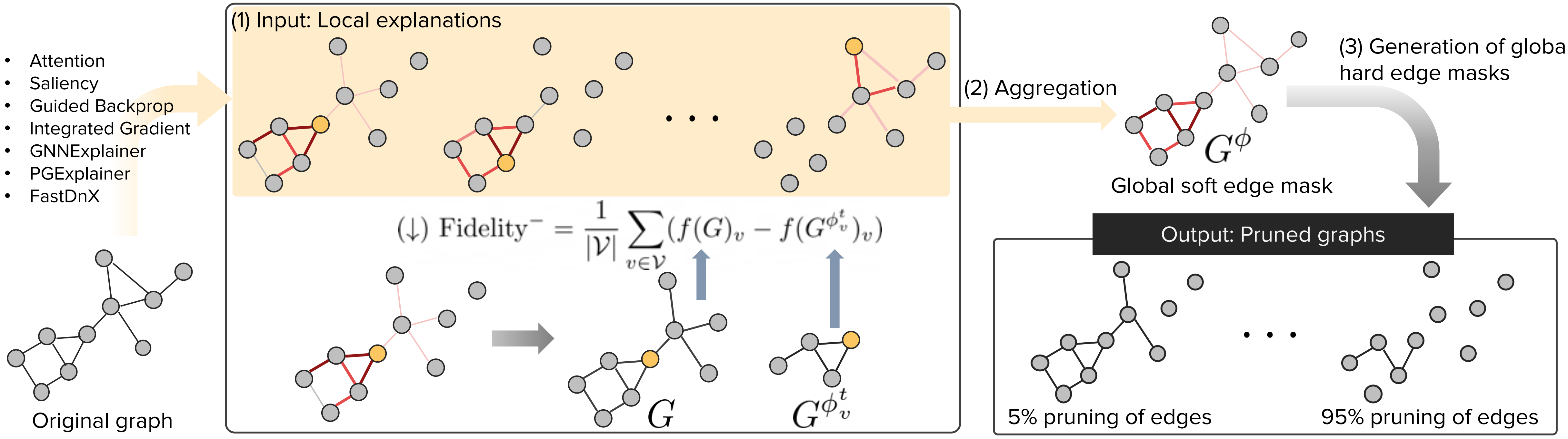
$\|A\|_0$ = 2 X (Num. edges) / d: Average num. edges per node / r: Number of edges to aggregate from



Reduce the number of edges to 1) Increase efficiency & 2) Potentially remove noisy edges

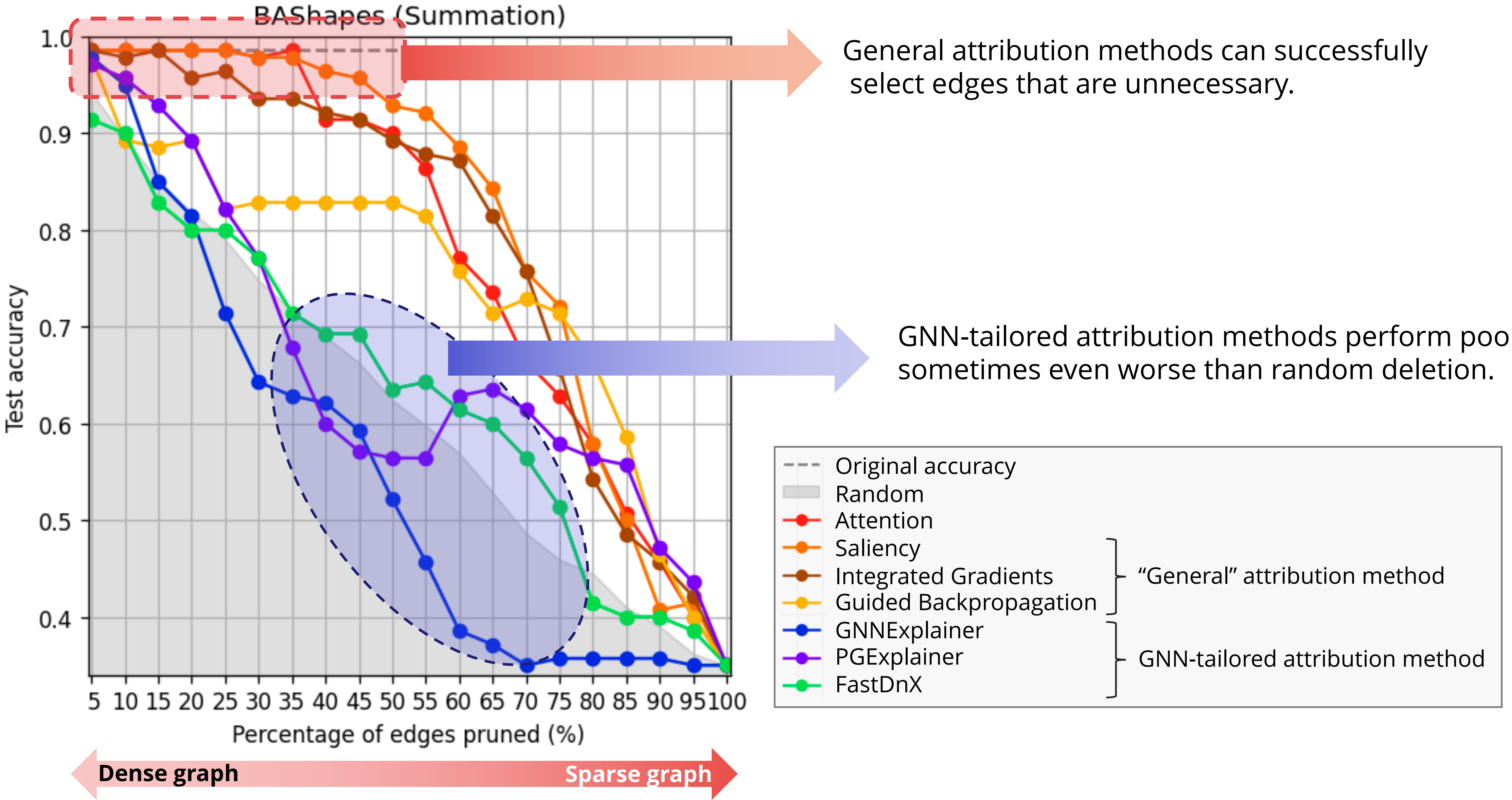
FiP (Fidelity-inspired Pruning)
A framework that can perform graph pruning by taking local explanations as input.

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



*f: GNN model

We found that local graph explanations can be used for graph pruning, but GNN-tailored methods underperform.

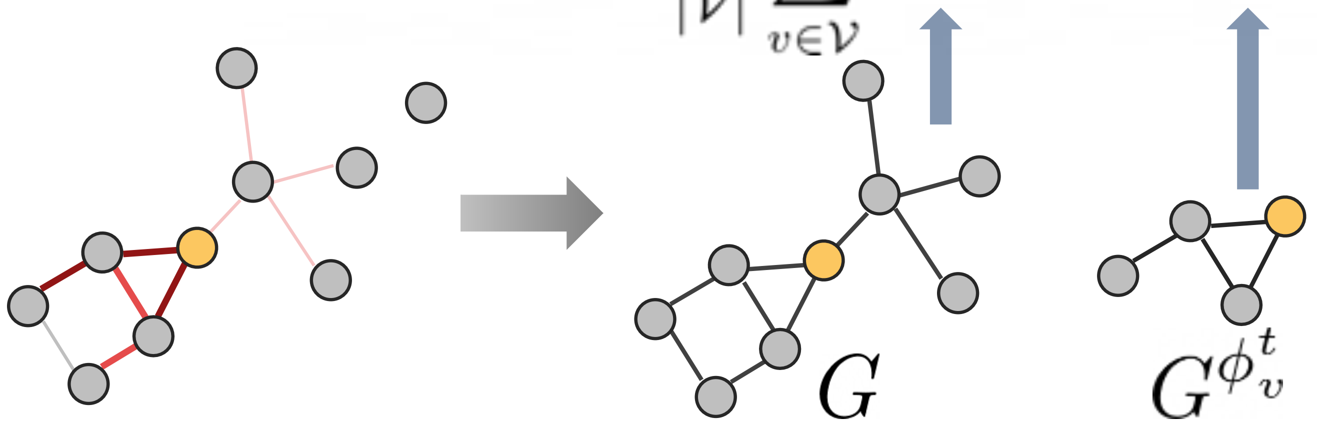


General attribution methods can successfully select edges that are unnecessary.

GNN-tailored attribution methods perform poorly, sometimes even worse than random deletion.

We also found that fidelity- measures does not translate to graph pruning despite logical appeal.

$$(\downarrow)\ \text{Fidelity}^{-} = \frac{1}{|\mathcal{V}|}\sum_{v\in\mathcal{V}}\left(f(G)_v - f(G^{\phi_v^t})_v\right)$$



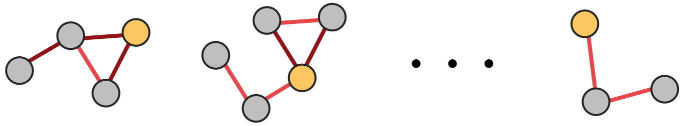
Method	BAShapes	Cora	Citeseer	Pubmed
Att	4.06×10^{-2}	3.67×10^{-2}	2.23×10^{-2}	2.46×10^{0}
SA	3.54×10^{-7}	2.21×10^{-7}	$\mathbf{8.90 \times 10^{-8}}$	2.46×10^{0}
IG	6.25×10^{0}	1.26×10^{0}	5.68×10^{-1}	$\mathbf{2.25 \times 10^{0}}$
GB	3.77×10^{0}	1.42×10^{0}	7.04×10^{-1}	2.40×10^{0}
GNNEx	$\mathbf{3.44 \times 10^{-7}}$	$\mathbf{2.14 \times 10^{-7}}$	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}

Although Attention exhibit poor fidelity- scores, it performs great on graph pruning.

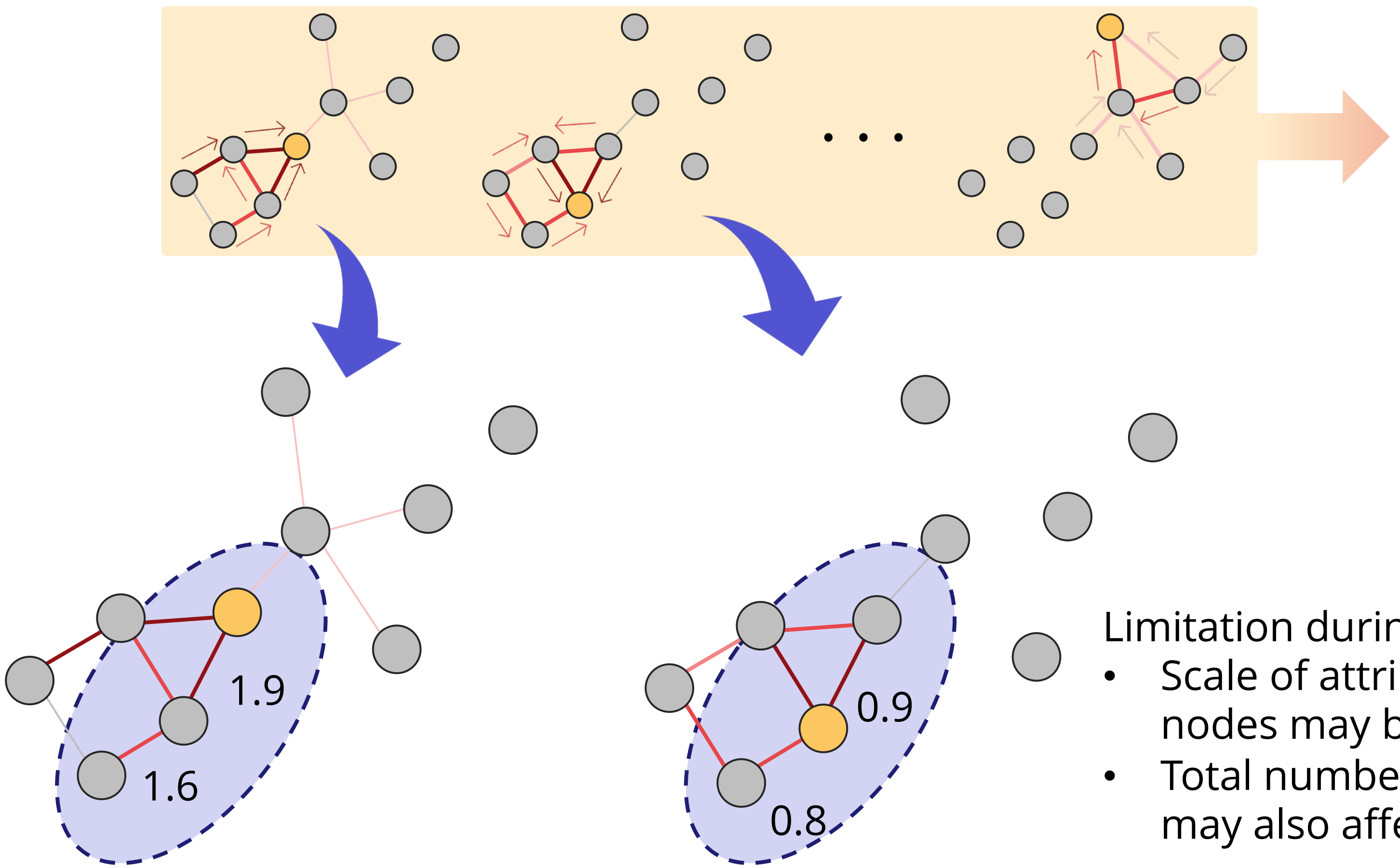
Although GNNExplainer exhibit great fidelity- scores, it results in bad graph pruning results.

Conclusion: Explanation as graph pruning looks promising, but many challenges remain.

The problem likely lies in the aggregation of local explanations & Limitation of graph pruning



Limitation of graph pruning approach
- **Every node has a different explanation**
- A single graph <u>cannot</u> fully express all local explanation (lossy compression)



Limitation during aggregation
- Scale of attribution score across nodes may be different
- Total number of edges for each explanation may also affect how should we normalize

Takeaway messages

- ***Explainable AI*** *is an important research topic, and* ***GNNs are no exception****.*
- *Previous literature mainly focus on explanations itself*
- *Ultimate goal of XAI:* ***Enhance the original system using knowledge from XAI***
- ***Graph explanations can be effectively used for graph pruning***
- *However,* ***good fidelity does not translate well into graph pruning***
- *The main limitation may be caused during aggregation, and the approach of graph pruning itself.*

Paper Homepage

Thank you!

List of References

"General" attribution methods

[1] Montavon et al., "Explaining nonlinear classification decisions with deep Taylor decomposition", Pattern Recognit. 65: 211-22 (2017)
[2] Bach et al., "On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation", PLOS ONE 10(7): e0130140.
[3] Riberiro et al, ""Why Should I Trust You?": Explaining the Predictions of Any Classifier", KDD 2016
[4] Selvaraju et al., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization", ICCV 2017

GNN-tailored attribution methods

[5] Ying et al., "GNNExplainer: Generating Explanations for Graph Neural Networks", NeurIPS 2019
[6] Luo et al., "Parameterized explainer for graph neural network", NeurIPS 2020

Early works in GNN attribution

[7] Baldassarre & Azizpour, "Explainability Techniques for Graph Convolutional Networks", ICML 2019 Workshop
[8] Pope et al., "Explainability methods for graph convolutional neural networks.", CVPR 2019

GNN models

[9] Kipf & Welling, "Semi-Supervised Classification with Graph Convolutional Networks", ICLR 2017
[10] Hamilton et al., "Inductive Representation Learning on Large Graphs", NIPS 2017
[11] Chen et al., "FastGCN: Fast Learning with Graph Convolutional Networks via Importance Sampling", ICLR 2018
[12] Chen et al., "Stochastic Training of Graph Convolutional Networks with Variance Reduction", ICML 2018
[13] Chiang et al., "Cluster-GCN: An Efficient Algorithm for Training Deep and Large Graph Convolutional Networks", KDD 2019

Miscellaneous

(Figure at page 7) Liu et al., "Comprehensive Graph Gradual Pruning for Sparse Training in Graph Neural Networks", arXiv (2022)