TEXAS A&M UNIVERSITY
Engineering

On Explainability of Graph Neural Networks via Subgraph Explorations

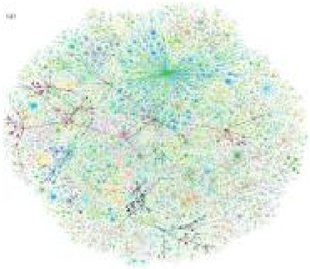
Hao Yuan,[1] Haiyang Yu,[1] Jie Wang,[2] Kang Li,[3] Shuiwang Ji[1]

[1] Texas A&M University
[2] University of Science and Technology of China
[3] West China Biomedical Big Data Center,

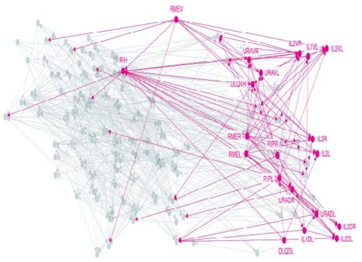
Background

- Graph data

Social networks



Internet



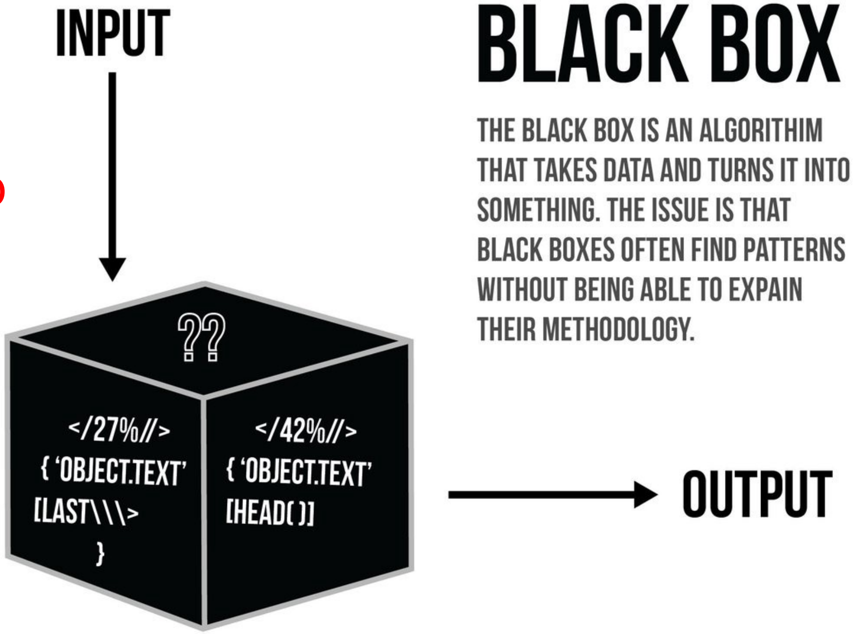
Networks of neurons

- Deep learning on graphs
  - Deep graph models achieve state-of-the-art performance
  - They are popular but lack explainability

Reference: Representation Learning on Networks, snap.stanford.edu/proj/embeddings-www, WWW 2018

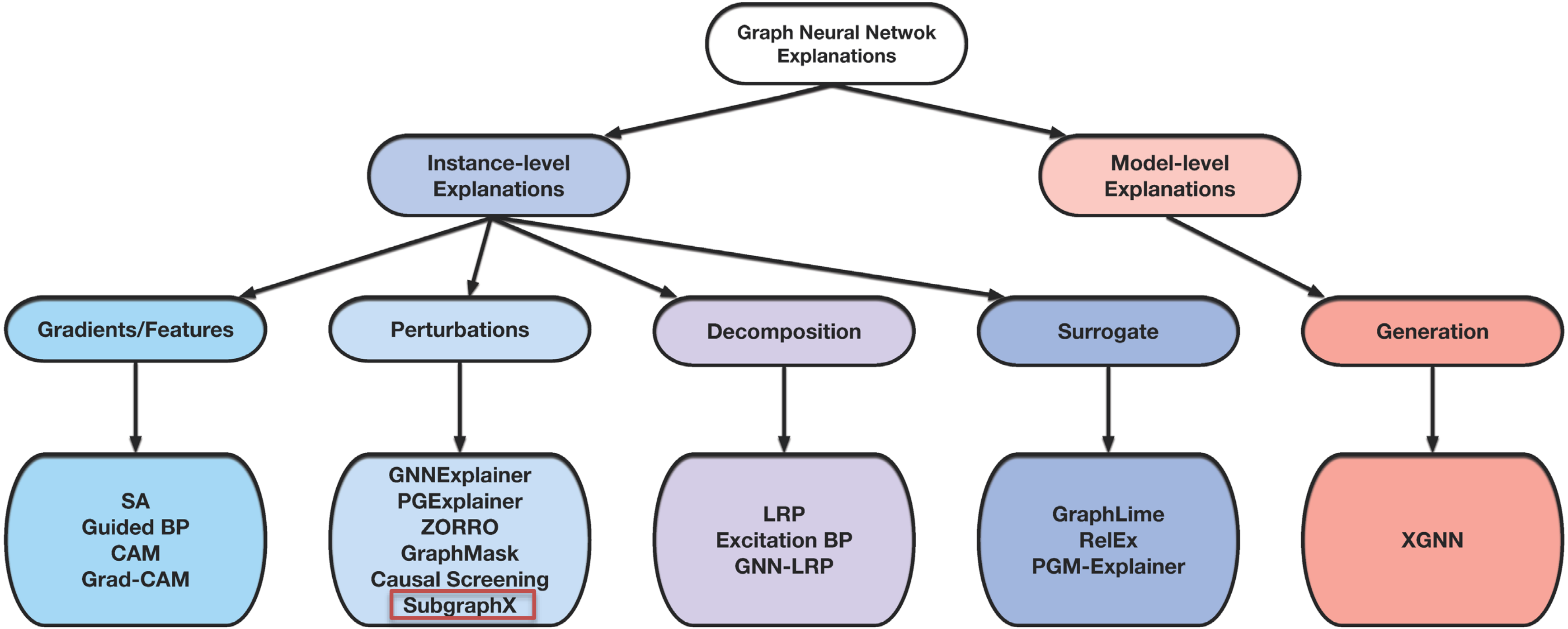
Why Explainability?

- We do not understand deep models!
- They cannot be fully trusted!



Reference: https://towardsdatascience.com/guide-to-interpretable-machine-learning-d40e8a64b6cf

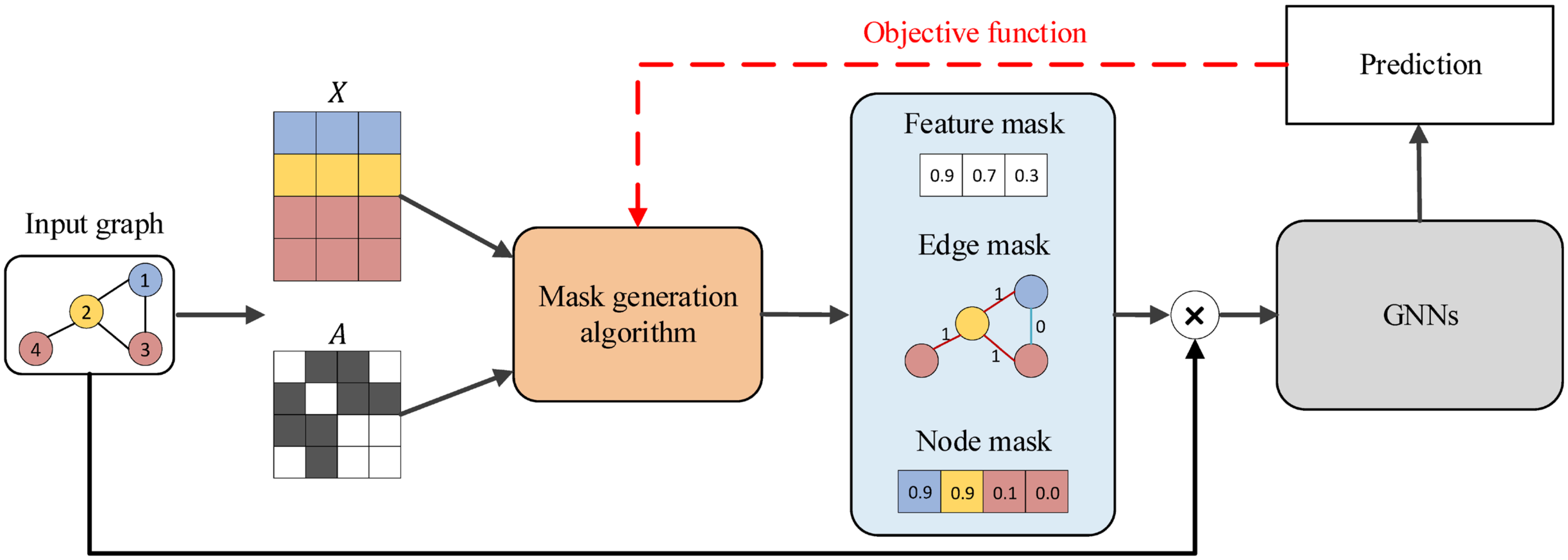
Overview of methods on the GNN Explanations



Reference: ***Explainability in Graph Neural Networks: A Taxonomic Survey,*** *https://arxiv.org/abs/2012.15445*

Pipeline of Perturbation-based methods

TEXAS A&M UNIVERSITY Engineering

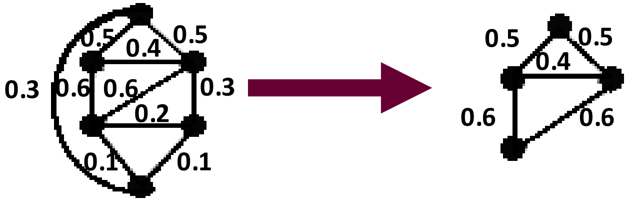


Model-agnostic method, take GNNs as a function f

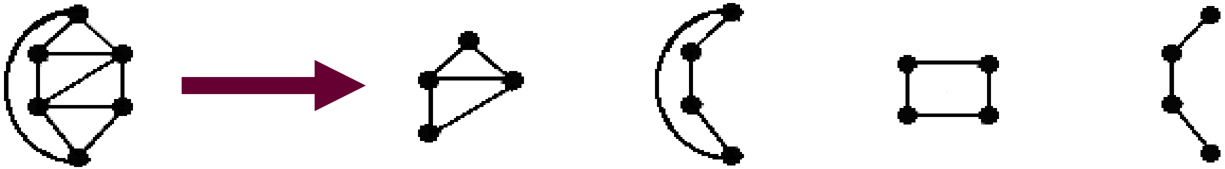
Reference: ***Explainability in Graph Neural Networks: A Taxonomic Survey,*** *https://arxiv.org/abs/2012.15445*

Explain subgraphs

Assigning importance score on edges



We consider the connected subgraphs



Importance score for each subgraph

0.5 0.7 0.5 0.4

Our goal

$$\mathcal{G}^* = \operatorname*{argmax}_{|\mathcal{G}_i| \leq N_{\min}} \text{Score}(f(\cdot), \mathcal{G}, \mathcal{G}_i)$$

Monte Carlo Tree Search

Brute-Force: Search all the possible subgraphs

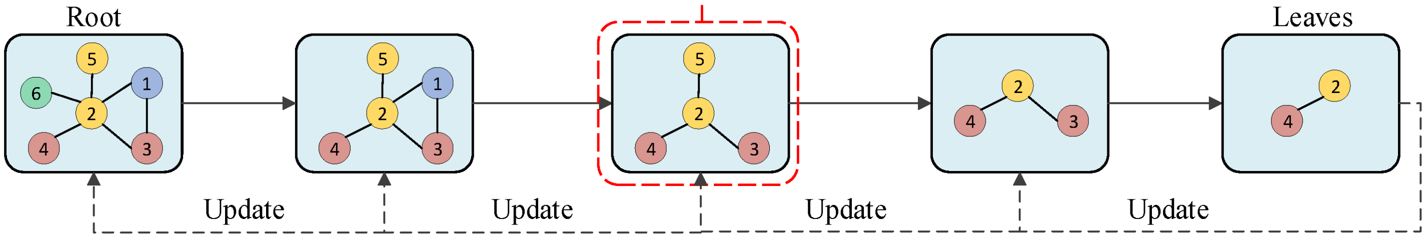
Monte-Carlo Tree Search: Given a whole graph, each time it will cut one node until the nodes in graph is smaller than constrain. At each step, it uses the sum of **importance score** and **exploration score** to decide which node will be cut. Then repeat this rollout procedure while encouraging to explore new states/subgraphs.

A trade-off between the exploration and expedition.

Monte Carlo Tree Search

Monte-Carlo Tree Search:

Build the search tree with connected constrains to make sure that the subgraph is connected.



The heuristic value $a^* = \underset{a_j}{\operatorname{argmax}} Q(\mathcal{N}_i, a_j) + U(\mathcal{N}_i, a_j),$

Q: the importance score for the next subgraph.

U: the exploration score. It becomes lower when we arrive this state/subgraph for many times.

Shapley values

TEXAS A&M UNIVERSITY Engineering

Marginal contribution of coalition subgraph G_i with other players S

$$m(S, \mathcal{G}_i) = f\left(S \cup \{\mathcal{G}_i\}\right) - f(S)$$

The Shapley Value of coalition subgraph G_i

$$\phi(\mathcal{G}_i) = \sum_{S \subseteq P \setminus \{\mathcal{G}_i\}} \frac{|S|!\,(|P| - |S| - 1)!}{|P|!} m(S, G_i)$$

The weighted sum of marginal contribution of coalition subgraph G_i with all different player subset S of the all player sets.

The number of different subset S: $(|P| - |G_i|)!$

Approximation of Shapley Values

Graph Inspired Efficient Computations

Given a L layer GNN, the node will only aggregates the information from its L-hop neighbors.

Therefore, we only take the L-hop neighbors P' to calculate the Shapley value.

$$\phi(\mathcal{G}_i) = \sum_{S \subseteq P' \setminus \{\mathcal{G}_i\}} \frac{|S|!\,(|P'| - |S| - 1)!}{|P'|!} m(S, G_i)$$

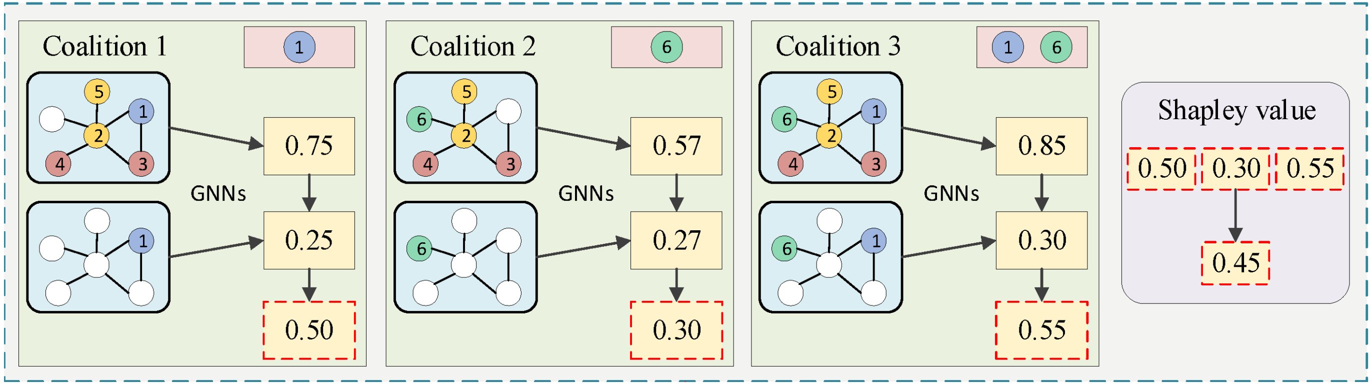
Monte Carlo Sampling estimation:

Instead of elaborating all the combinations, we sample different combination with the weights as the probability to do an unbiased estimation.

$$\phi(\mathcal{G}_i) = \frac{1}{T} \sum_{t=1}^{T} \left(f\left(S_i \cup \{\mathcal{G}_i\}\right) - f(S_i) \right)$$

Shapley values

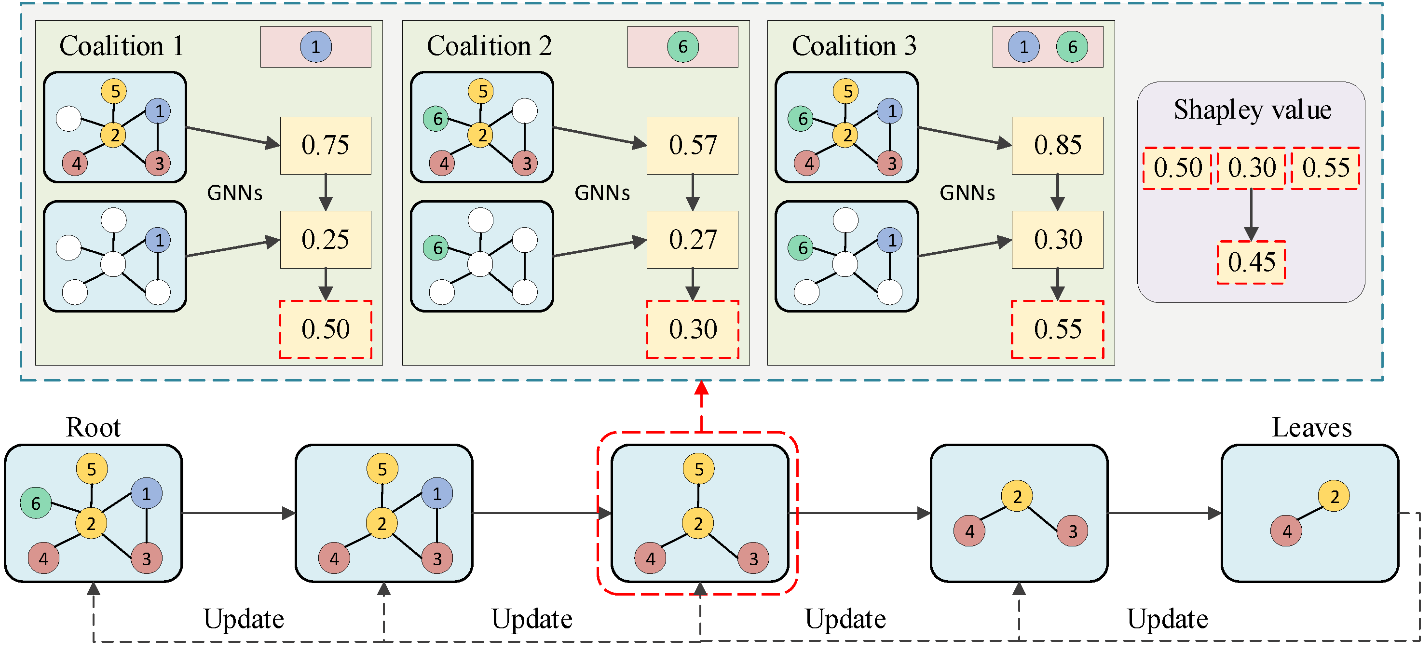
Example of the Shapley value applied in the SubgraphX



Take monte-Carlo sampling to estimate the Shapley value.

Consider the nodes in the l-hop which is the number of layer of the GNNs.

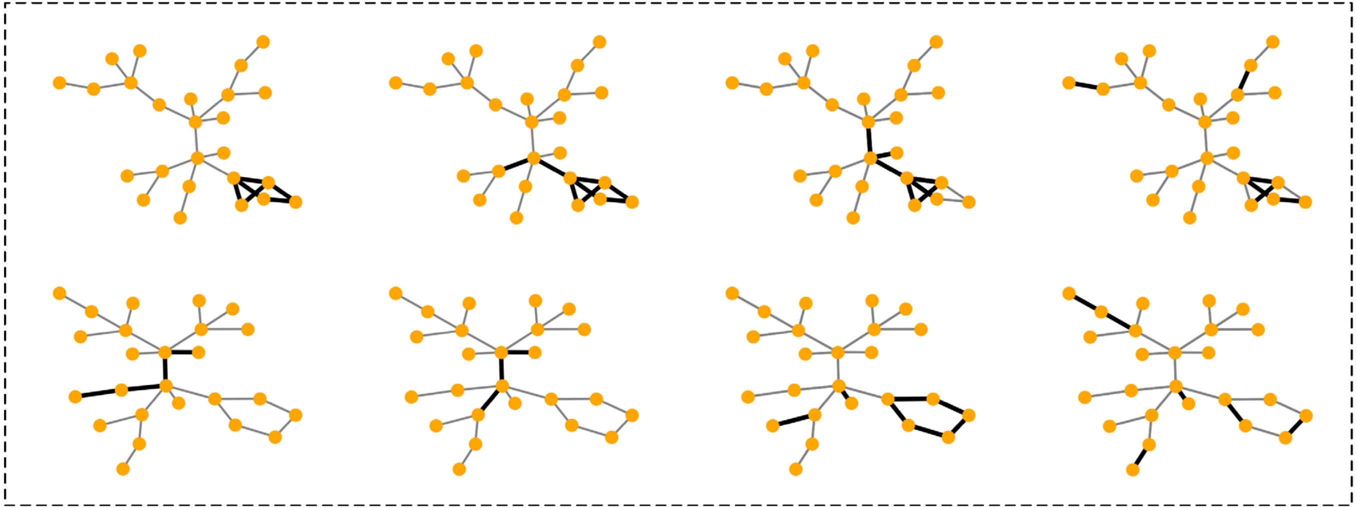
The pipeline of SubgraphX



Experiment Results

TEXAS A&M UNIVERSITY
Engineering

Synthetic Dataset: BA-Motifs

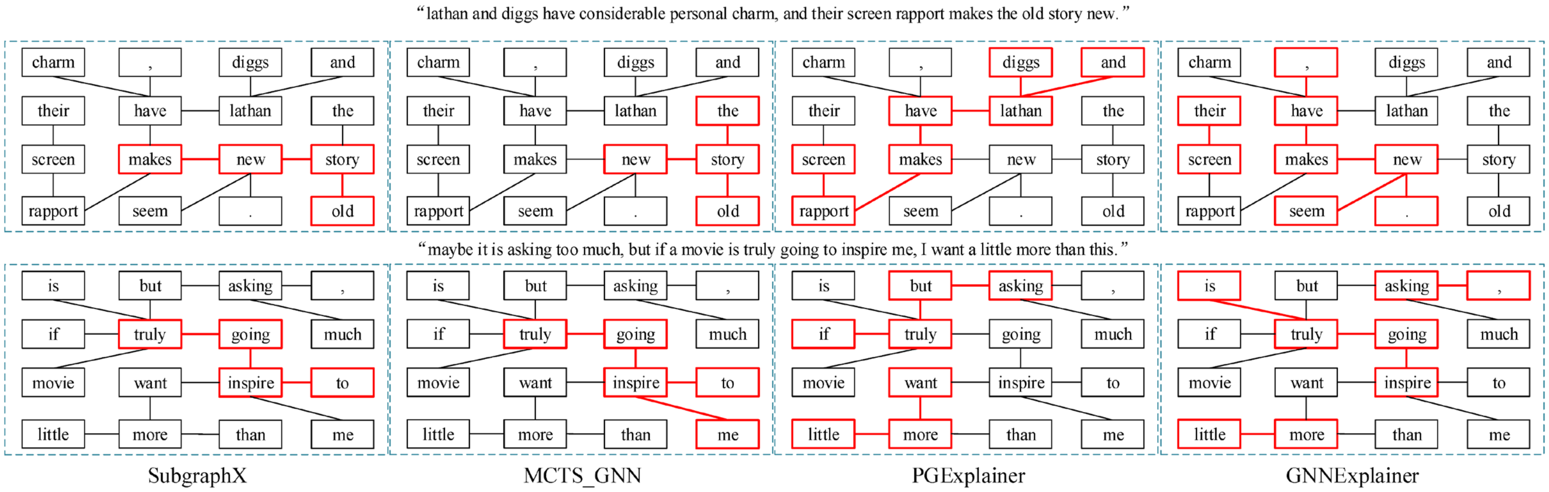


SubgraphX MCTS_GNN PGExplainer GNNExplainer

Experiment Results

TEXAS A&M UNIVERSITY Engineering

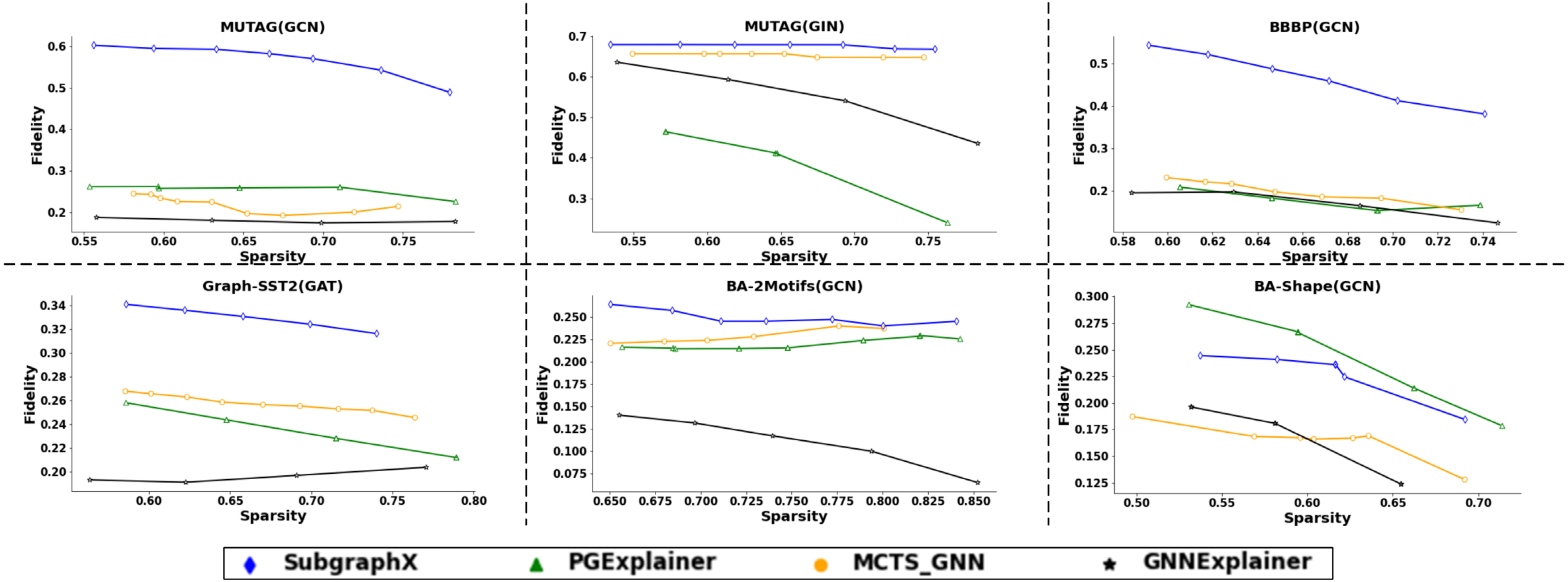
Human understandable explanation result on the Graph-SST2 dataset.



Quantitative Studies

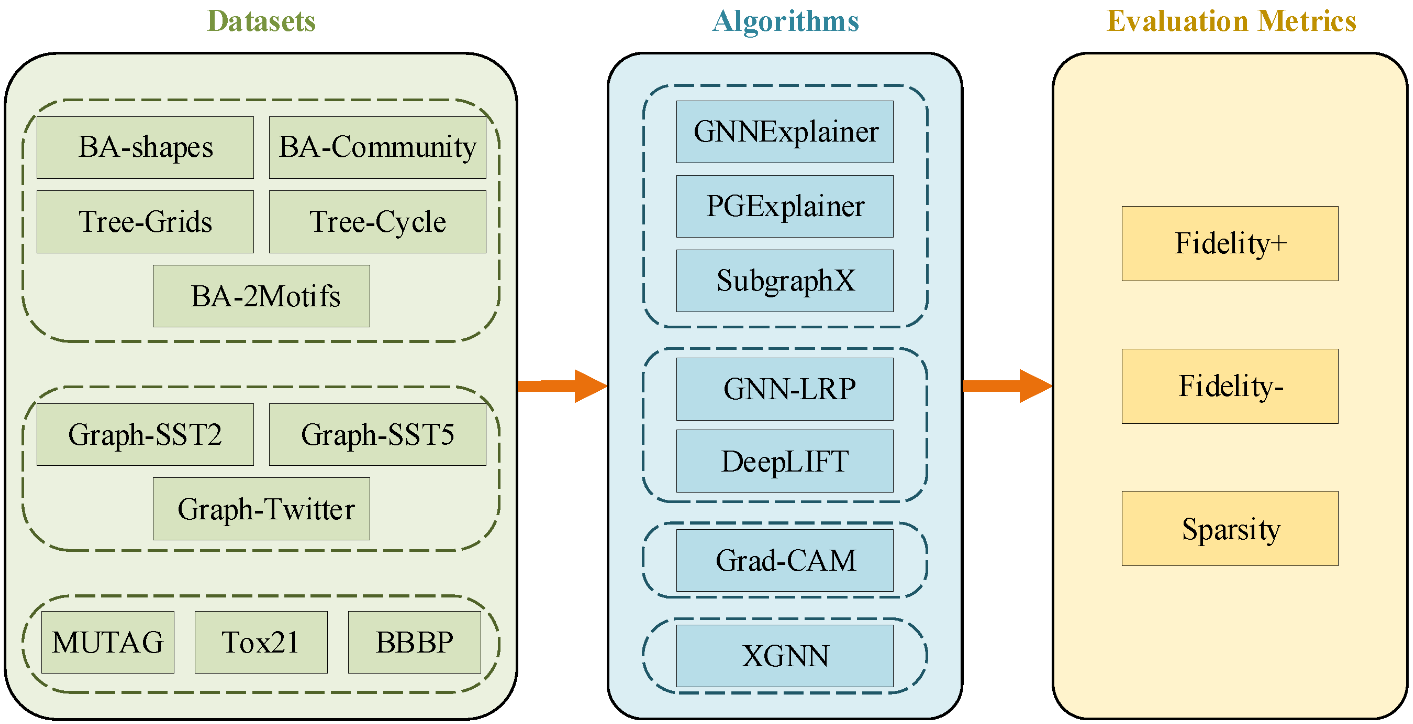
TEXAS A&M UNIVERSITY
Engineering

Fidelity and Sparsity



Package

DiveIntoGraphs



Reference: ***Explainability in Graph Neural Networks: A Taxonomic Survey***, *https://arxiv.org/abs/2012.15445*
DIG: A Turnkey Library for Diving into Graph Deep Learning Research, *https://github.com/divelab/DIG, https://arxiv.org/abs/2103.12608*

TEXAS A&M UNIVERSITY
Engineering

On Explainability of Graph Neural Networks via Subgraph Explorations

Thank you