View-based Explanations for Graph Neural Networks

Accepted and Presented at SIGMOD '24

Tingyang Chen[*], Dazhuo Qiu[+], Yinghui Wu[#], Arijit Khan[+], **Xiangyu Ke**[*], Yunjun Gao[*]

[*]Zhejiang University (China),

[+]Aalborg University (Denmark),

[#]Case Western Reserve University (USA)

Santiago
Chile

SIGMOD
PODS
2024

浙江大學 ZHEJIANG UNIVERSITY

AALBORG UNIVERSITET

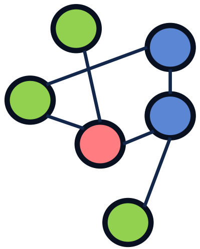
CASE SCHOOL OF ENGINEERING
CASE WESTERN RESERVE UNIVERSITY

Roadmap

- **Introduction**
  - Background
  - Motivation
- **View-based Explanation**
- **Generating Explanation Views**
  - Explain-and-summarize
  - Incremental generation
- **Experiment**
- **Conclusion**

Introduction: Background

Core of GNNs



Input Graph *G*

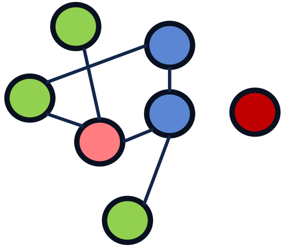
GNN Model *M*

Mutagenetic / Non-mutagenetic?

Output Prediction *y*

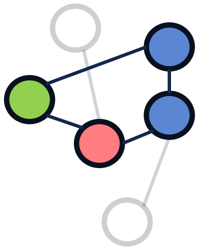
<u>"What Knowledge should/does the black model use to make decisions?"</u>

Post-hoc Explainability



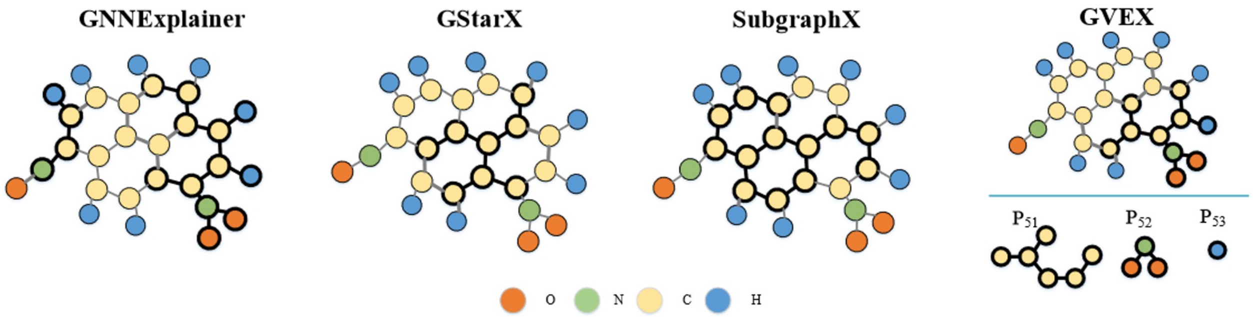
Input Graph *G*
Output Prediction *y*

Which fraction of the input graph is **most critical** to the model's decision



Explanation Subgraph G_s

Introduction: Motivation



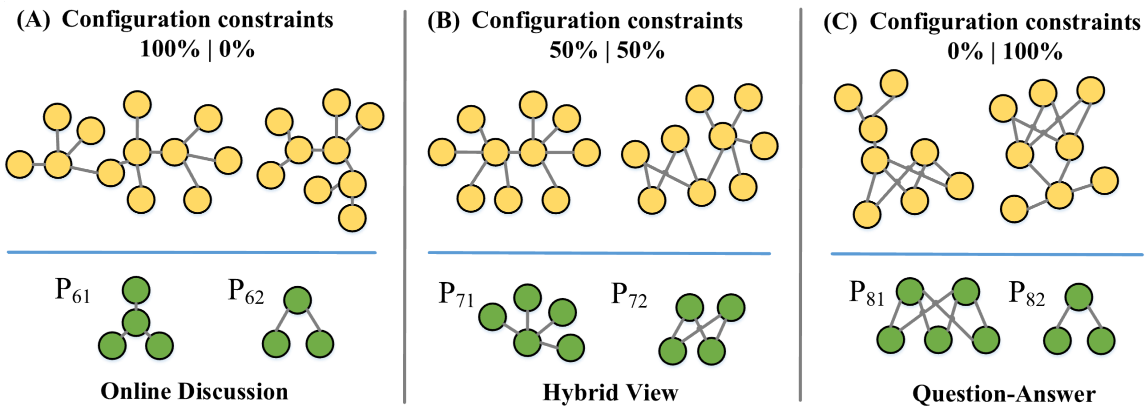
Methods	LEARNING	TASK	TARGET	MA	LS	SB	COVERAGE	CONFIG	QUERYABLE
SubgraphX [68]	✗	GC/NC	Subgraph	✓	✗	✗	✗	✗	✗
GNNExplainer [63]	✓	GC/NC	E/NF	✓	✗	✗	✗	✗	✗
PGExplainer [40]	✓	GC/NC	E	✗	✗	✗	✗	✗	✗
GStarX [73]	✗	GC	Subgraph	✓	✗	✗	✗	✗	✗
GCFExplainer [29]	✗	GC	Subgraph	✓	✓	✗	✓	✗	✗
GVEX (**Ours**)	✗	GC/NC	Graph Views (Pattern+Subgraph)	✓	✓	✓	✓	✓	✓

Comparison with state-of-the-art GNN explanation methods. Here "Learning" denotes whether (node/edge mask) learning is required, "Task" means what downstream tasks each method can be applied to (GC/NC: graph/ node classification), "Target" indicates the output format of explanations (E/NF: Edge/Node Features), "Model-agnostic"(MA) means if the method treats GNNs as a black-box during the explanation stage (i.e., the internals of the GNN models are not required), "Label-specific"(LS) means if the explanations can be generated for a specific class label; "Size-bound"(SB) means if the size of explanation is bounded; "Coverage" means if the coverage property is involved, "Config" means if users can configure the method to generate explanations for designated class labels; "Queryable" means if the explanations are directly queryable.

Introduction: Motivation

Challenges with Existing Methods

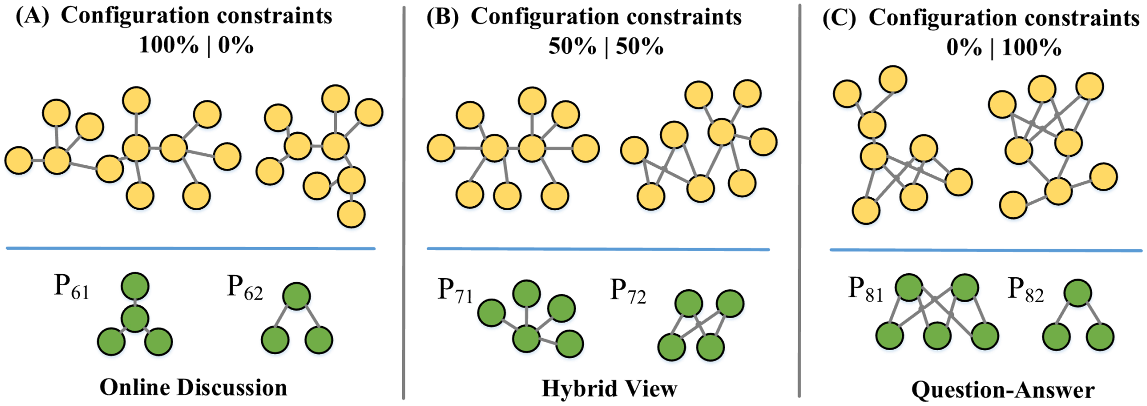
- **Oversized explanation**: Existing methods generate large explanation subgraphs.
- **Lack of meaningful explanations for domain experts**: Not easy to access and inspect with domain knowledge. (Not queryable)
- **Not configurable explanation based on user setting**: Only explaining one class may omit the relevant information between classes. (Not configurable)



Introduction: Motivation

Key Motivations

- "finer-grained" and user-friendly explanation structures for graph classification problem.
- "queryable" hence are easy for human experts to access and inspect with domain knowledge.
- "configurable" to enable users with the flexibility to obtain comprehensive and detailed explanations tailored to their classes of interest.



View-based Explanation

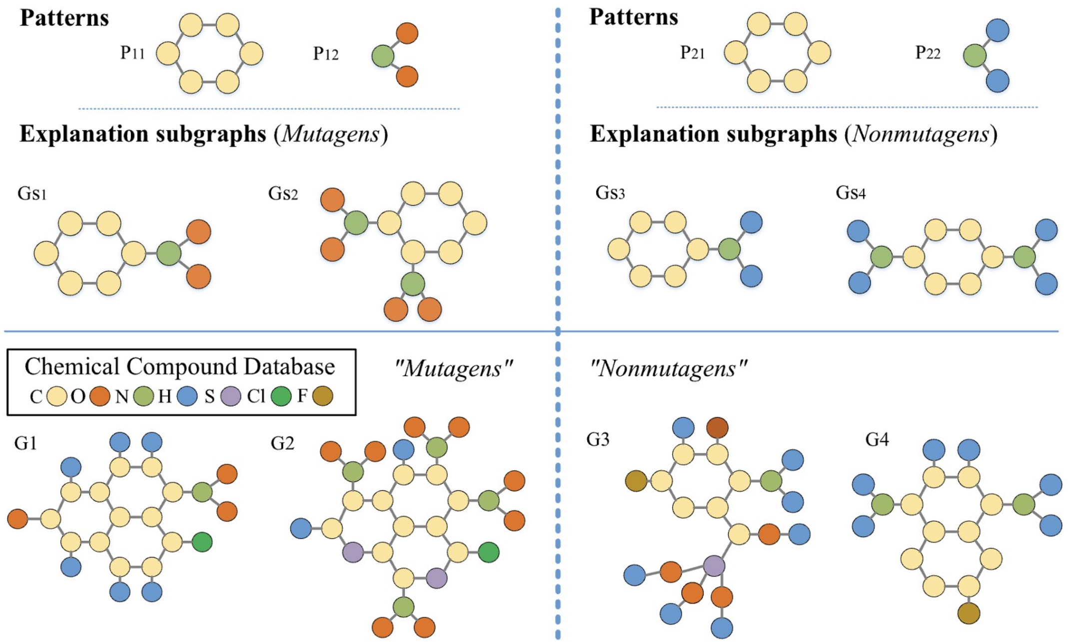
Two-tier Explanation

Lower-tier:

Subgraphs that ensure the same prediction (factual) and its removal changes prediction label (counterfactual).

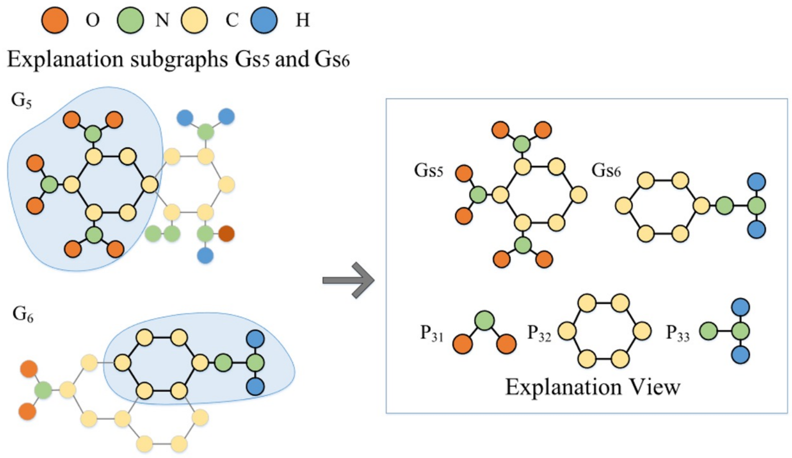
Higher-tier:

Patterns that summarize the explanation subgraphs with coverage guarantee.



View-based Explanation

An explanation view for a single class label: explanation subgraphs and patterns



Quality of Explanation Views

Explainability: An explanation view has better explainability if its explanation subgraphs involve more nodes with features that can maximize their influence via a random walk-based message passing process of GNN.

$$f(\mathcal{G}_{\mathcal{V}}^{l}) = \sum_{G_{si} \in \mathcal{G}_{s}^{l}} \frac{I(V_{si}) + \gamma D(V_{si})}{|V_i|}$$

Feature influence & Neighborhood diversity

Coverage: Besides "lower-tier" explainability, we also expect the "higher-tier" patterns of an explanation view to cover a desirable amount of nodes for each label group of interests.

View-based Explanation

Problem Formulation

Explanation View Generation Problem (EVG):

Given a graph database $\mathcal{G}$, a set of interested labels Ł s.t. $|Ł| = t$, a GNN $\mathcal{M}$, and a configuration C, the *explanation view generation problem*, denoted as EVG, is to compute a set of graph views $\mathcal{G}_{\mathcal{V}} = \{\mathcal{G}_{\mathcal{V}}^{l_1}, \dots, \mathcal{G}_{\mathcal{V}}^{l_t}\}$, such that $(i \in [1, t])$:

- Each graph view $G_V^{l_i} = (P^{l_i}, G_s^{l_i}) \in G_V$ is an explanation view of G for M w.r.t. $l_i \in$ Ł;
- Each $\mathcal{G}_{\mathcal{V}}^{l_i}$ properly covers the label group $\mathcal{G}^{l_i}$; and
- $\mathcal{G}_{\mathcal{V}}$ maximizes an aggregated explainability, i.e.,

$$\mathcal{G}_{\mathcal{V}} = \arg\max \sum_{\mathcal{G}_{\mathcal{V}}^{l_i} \in \mathcal{G}_{\mathcal{V}}} f(\mathcal{G}_{\mathcal{V}}^{l_i})$$

Computational Complexity

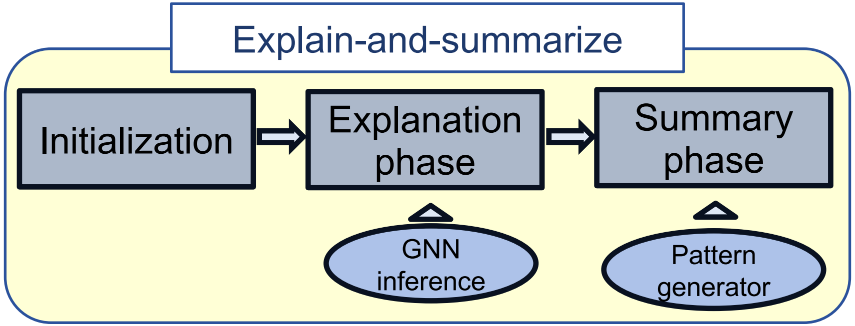
EVG:

For a fixed GNN $\mathcal{M}$, EVG is Σ_P^2-complete, and remains NP-hard even when $\mathcal{G}$ has no edges.

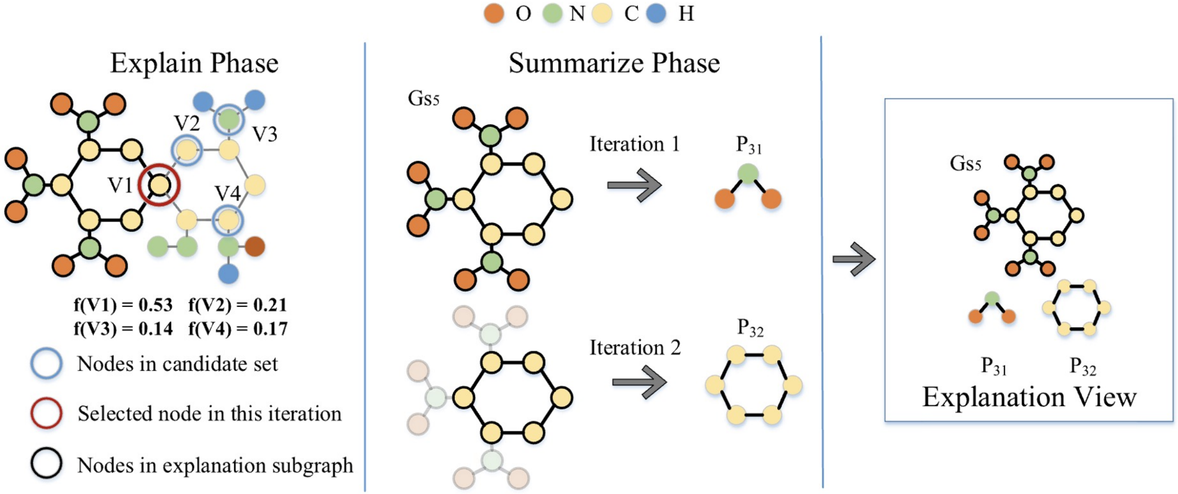
View Verification:

Given a graph database $\mathcal{G}$, configuration C, and a two-tier structure $(\mathcal{P}, \mathcal{G}_s)$, the view verification problem is NP-complete when the GNN $\mathcal{M}$ is fixed.

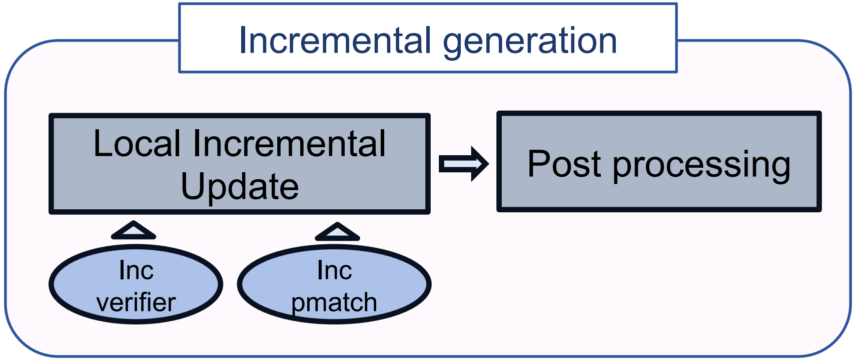
Generating Explanation Views: GVEX



½ Approximation

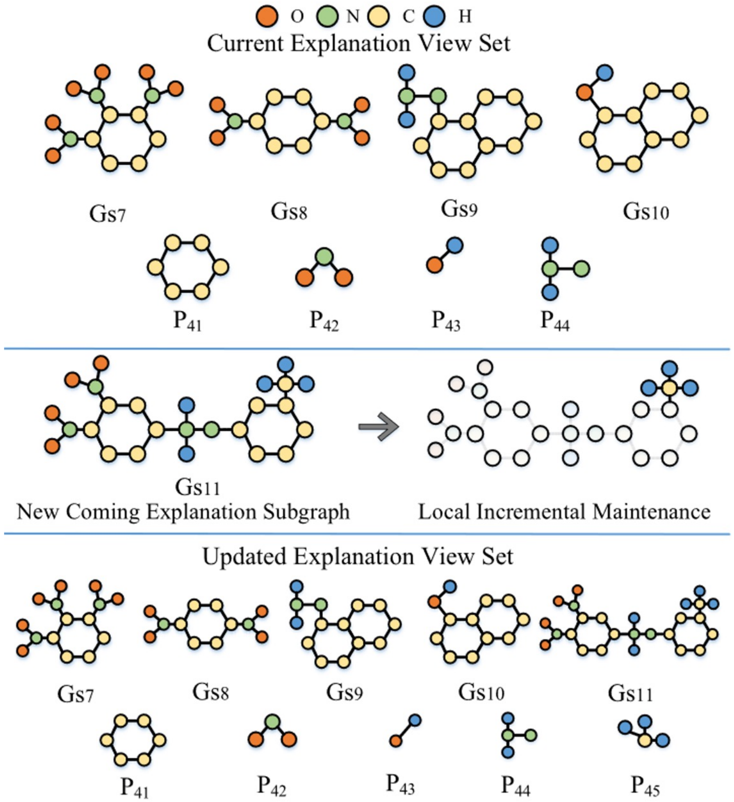


Generating Explanation Views: StreamingGVEX

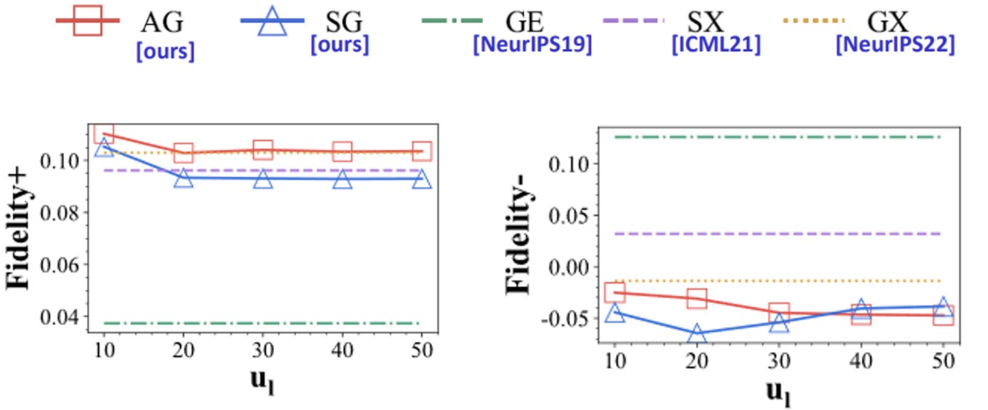


¼ Approximation

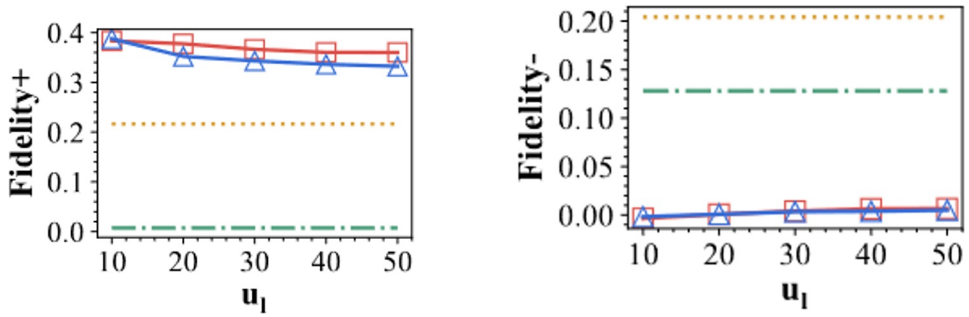
Parallelization: Both algorithms can be parallelized.



Experiment



ENZYMES

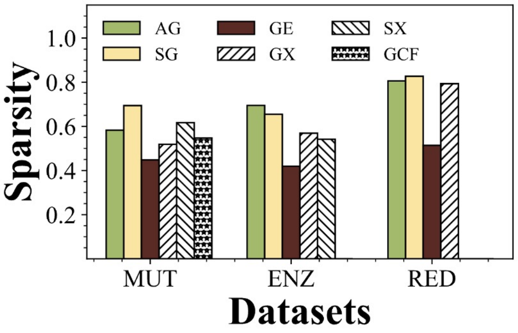


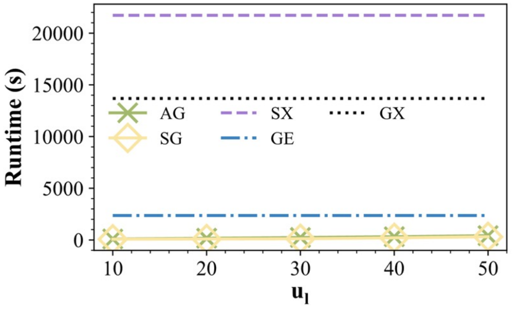
REDDIT-BINARY

Fidelity+ quantifies the deviations caused by removing the explanation substructure from the input graph. Higher is better.

Fidelity- measures how close the prediction results of the explanation substructures are to the original inputs. Lower is better.

Experiment

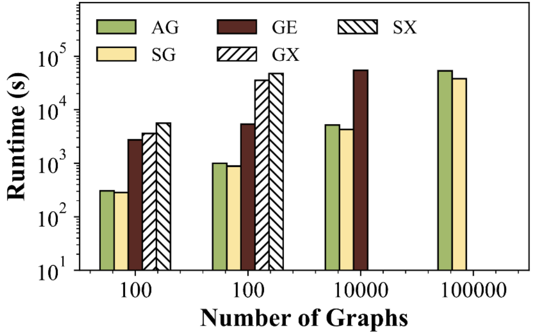


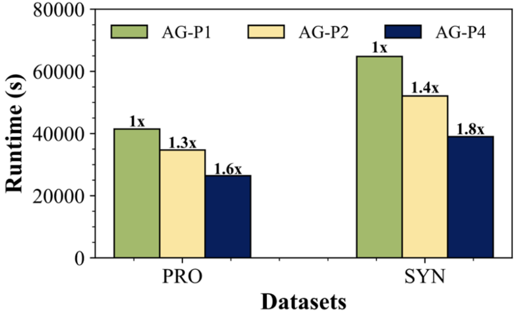


ENZYMES

Sparsity quantifies how compact is the explanation. Higher is better.

Efficiency experiment shows superior low time cost over learning-based methods.





Scalability experiment indicates that both algorithms is capable of handling large-scale datasets.

Parallelization of approximate algorithm enable significant speed-up.

Conclusion

Conclusion

A novel class of explanation structure for GNN-based graph classification

A set of quality measures in terms of explainability

An algorithmic solution for generating graph views to explain GNNs

- Codebase link: https://github.com/ZJU-DAILY/GVEX
- Demo video link: https://www.youtube.com/watch?v=q9d7IdulIuQ
- Demo paper (SIGMOD 2024):
- User-friendly, Interactive, and Configurable Explanations for Graph Neural Networks with Graph Views
- Full paper in SIGMOD24: https://dl.acm.org/doi/10.1145/3639295

THANK YOU !

Acknowledgment: Chen, Ke, and Gao are supported in part by the NSFC under Grants No.(62025206, U23A20296) and Yongjiang Talent Introduction Programme (2022A-237-G). Qiu and Khan are support from the Novo Nordisk Foundation grant NNF22OC0072415. Wu is supported in part by NSF under CNS-1932574, ECCS-1933279, CNS-2028748 and OAC-2104007.

浙江大學 ZHEJIANG UNIVERSITY

AALBORG UNIVERSITY

CWRU

Santiago Chile

SIGMOD PODS 2024