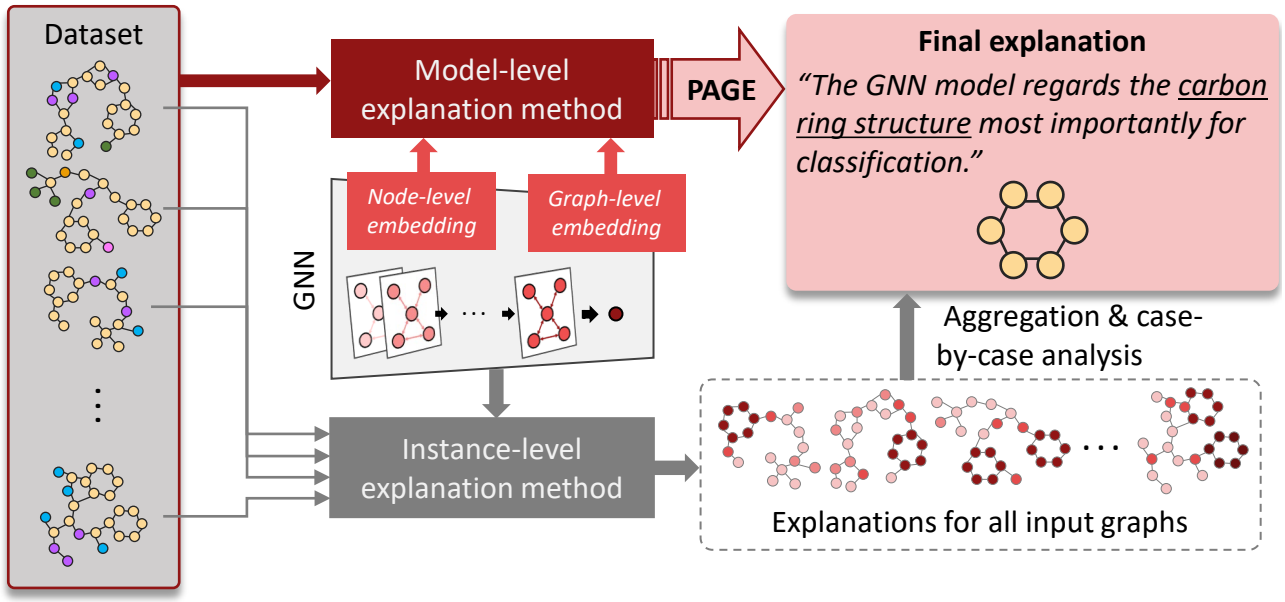
PAGE: Prototype-Based Model-Level Explanations for Graph Neural Networks

Yong-Min Shin, Sun-Woo Kim, and Won-Yong Shin*

*Corresponding author

수학계산학부(계산과학공학)
School of Mathematics and Computing
(Computational Science and Engineering)
MIDaS LAB
Machine Intelligence & Data Science

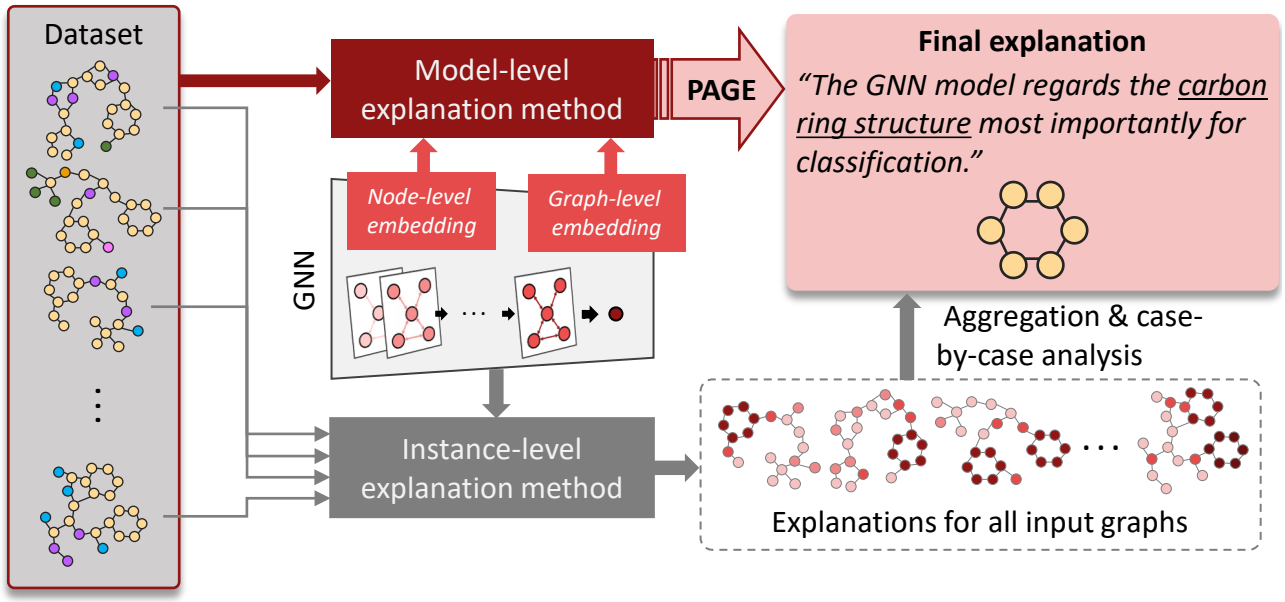
Model-level explanation



- **Model-level explanations** explain the decisions of a deep neural network model by finding the pattern that leads to a specific decision.
- Previous instance-level explanations are not fit for such type of explanations, often requiring aggregating explanations or large-scale case-by-base analysis.

Problem setting

Model-level explanations of graph neural networks (GNNs)

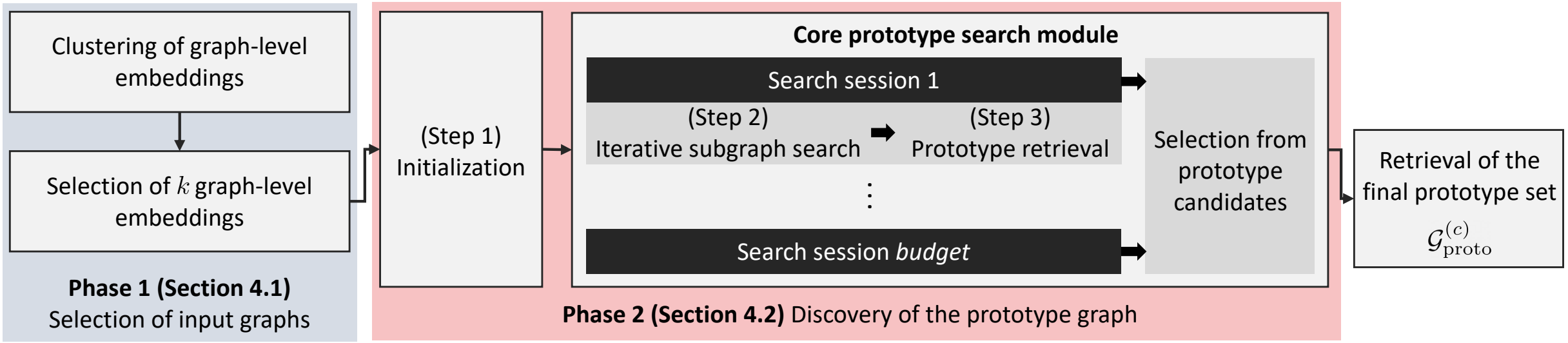


Explanations for GNNs have mostly focused on instance-level explanation

⇩

We propose PAGE, a model-level explanation method for graph classification GNN models.

Proposed Methodology

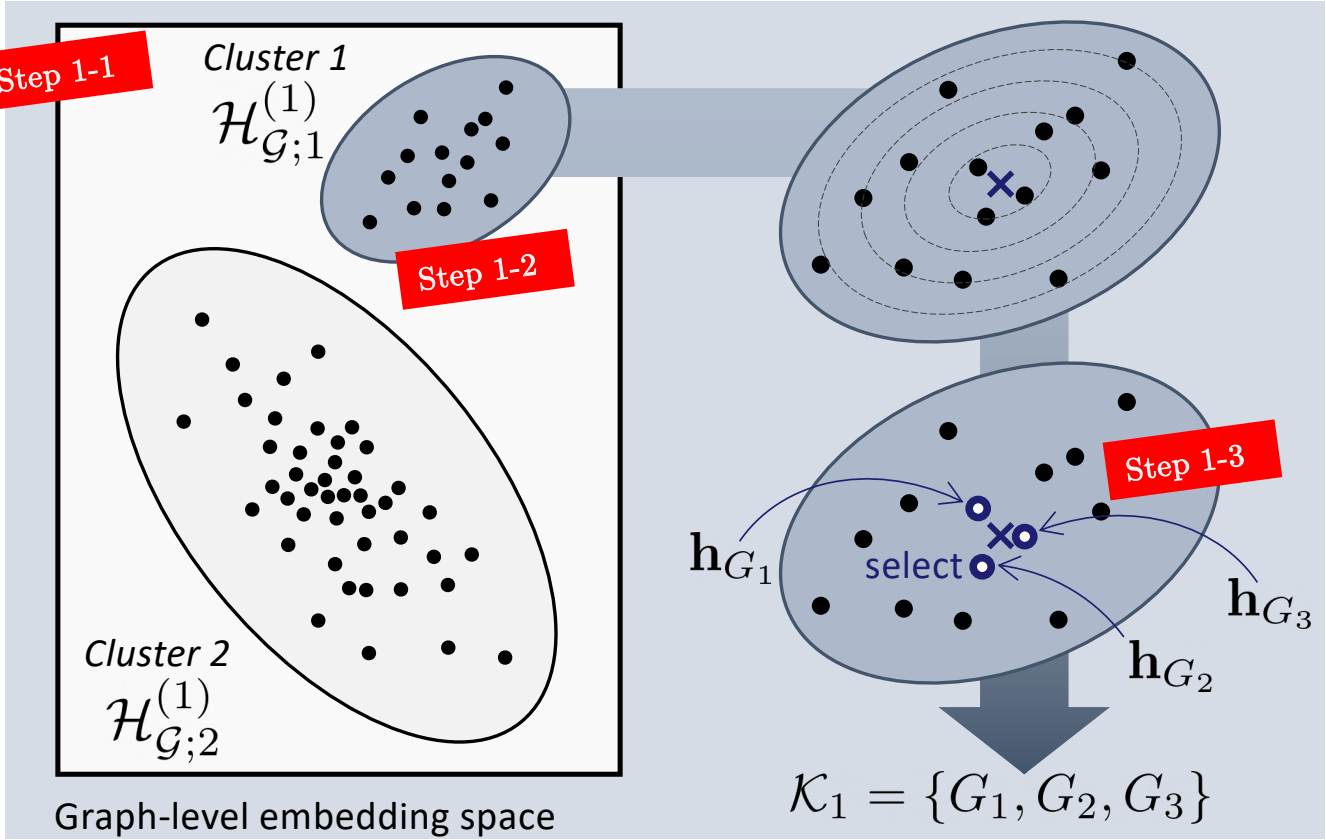


→ Main objective of Phase 1: **Reduce the search space** of the prototype by selecting a small subset of input graphs to work on.

→ In Phase 2, we find the prototype graph from the graphs from Phase 1.
→ Main approach: Search common subgraph **guided by** the embedding / representation vectors from the underlying GNN

1. Step 1: Pre-compute all node-tuple-wise similarity scores
2. Step 2 (**Core prototype search module**): Extract common subgraph pattern via searching for graph patterns with highest overall similarity score

Phase 1: Selection from graph-level embeddings

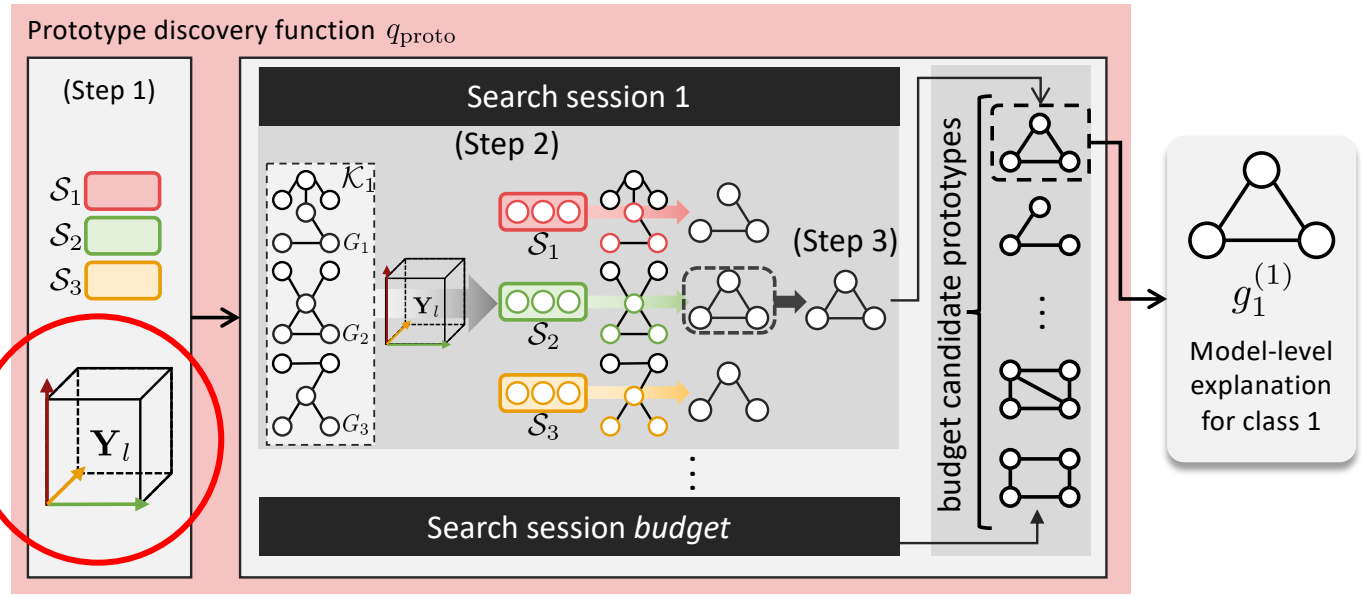


- (Step 1-1) Use the graph dataset to acquire graph-level embeddings for all input, where it contains semantic information learnt from the GNN.
- (Step 1-2) Use clustering to capture different sub-pattern within each class.
- (Step 1-3) Only select nodes that are **near the centroid** for each cluster.

Phase 2: Prototype discovery with Prototype Scoring Function

Definition 1 (Prototype Scoring Function): Given a set of b-dimensional vectors $\{\mathbf{v}_i\}_{i=1}^{k}$, we define the *prototype scoring function* $s : \mathbb{R}^b \times \cdots \times \mathbb{R}^b \to \mathbb{R}$ as follows:

$$s(\mathbf{v}_1, \ldots, \mathbf{v}_k) = \mathbf{1}^T(\mathbf{v}_1 \odot \cdots \odot \mathbf{v}_k), \quad (3)$$

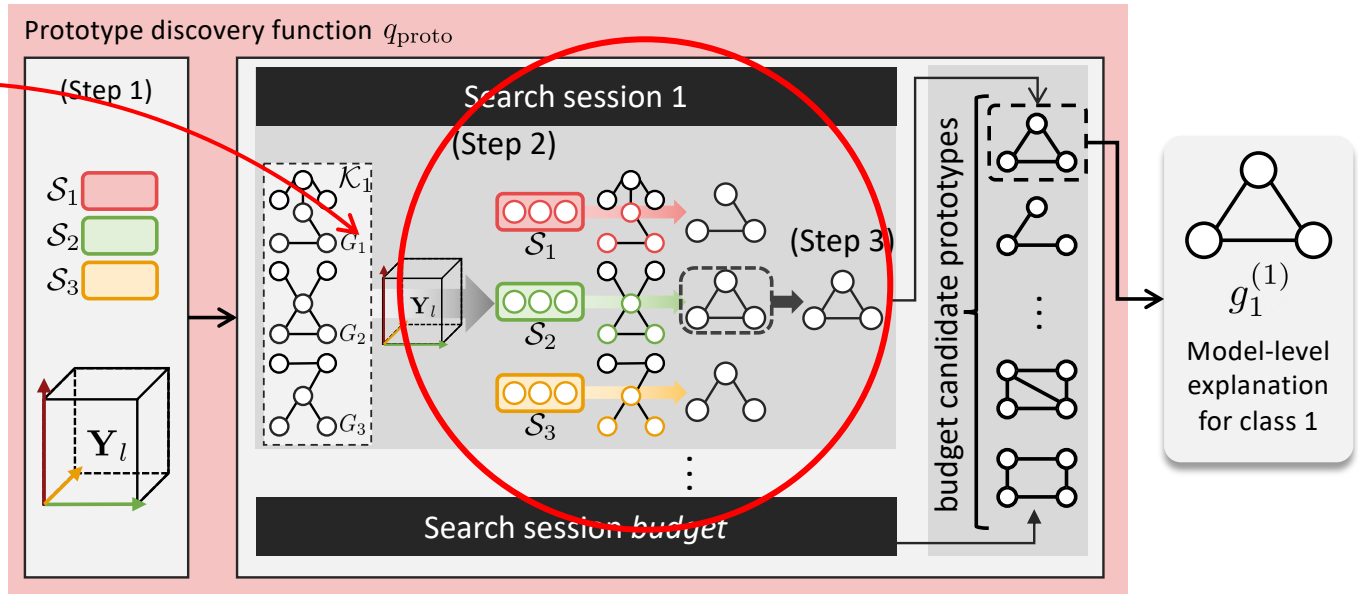


Objective: We want the nodes of the **final prototype graph to have a strong alignment** with the nodes of the k selected graphs.

Question: How can we efficiently capture node-wise similarity between all k selected graphs, which is to be used during the discovery process?

Solution: Define a new Prototype Scoring Function, and pre-compute a similarity tensor that captures all node-wise alignments from node-level embeddings from each k graphs

Phase 2: Prototype discovery with Prototype Scoring Function



Objective: We want the nodes of the **final prototype graph to have a strong alignment** with the nodes of the k selected graphs.

Question: Given the node-level similarity scores, how do we extract the common graph pattern?

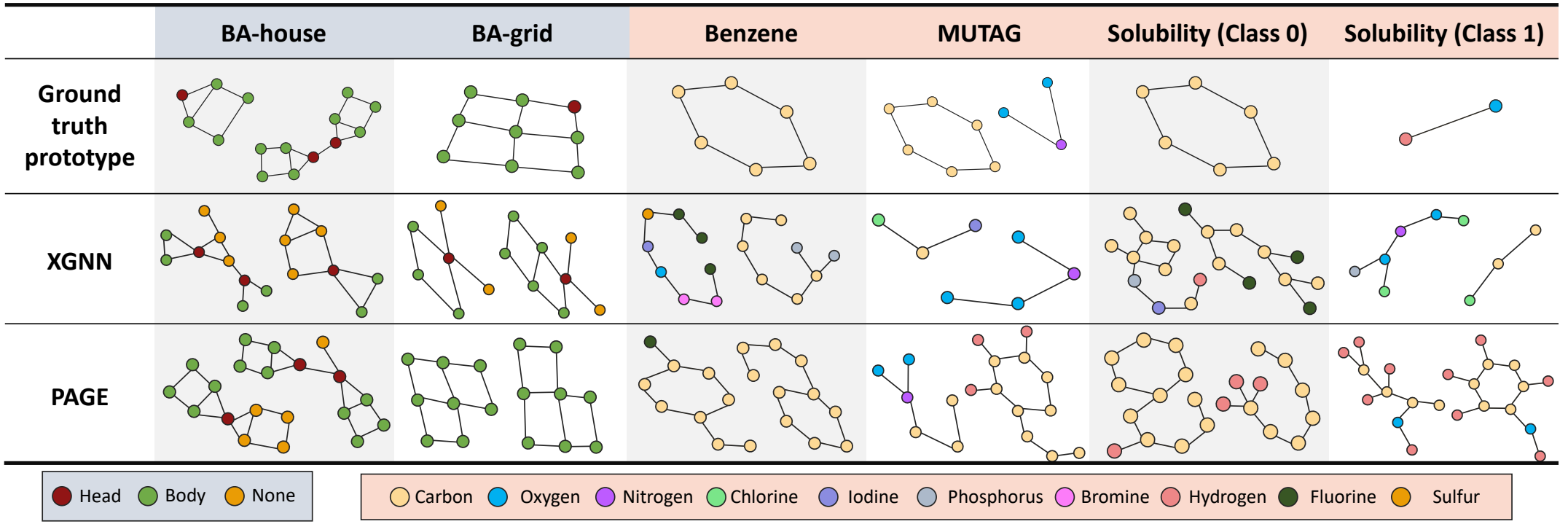
Solution: Perform an **iterative walk for each k graph simultaneously**, selecting the nodes that has the **highest similarity scores as the next node** during the process.

Datasets for explainable AI on graphs

*TP: Test performance of GNN

Dataset	n	$\sum_i \mathcal{V}_i$ (Avg.)	$\sum_i \mathcal{E}_i$ (Avg.)	TP	Ground-truth explanations
BA-house	2,000	21,029 (10.51)	62,870 (31.44)	1.000	Ground-truth explanations: Small subgraph motifs (house-shaped, grid-shaped)
BA-grid	2,000	29,115 (14.56)	91,224 (45.61)	0.9583	
Benzene	12,000	246,993 (20.58)	523,842 (43.65)	0.9444	Chemical molecule groups (Real-world knowledge-based explanations)
MUTAG	4,337	131,488 (30.32)	266,894 (61.54)	0.7247	
Solubility	708	9,445 (13.34)	9,735 (13.75)	0.8717	
MNIST-sp	70,000	5,250,000 (75)	41,798,306 (696.63)	0.7595	Visual semantics

Qualitative evaluation: Comparison against ground-truth explanations



Compared to XGNN [1], PAGE is able to extract explanations that much more resembles the ground-truth explanations (prototypes) for each dataset (synthetic and real-world).

[1] Yuan, H.; Tang, J.; Hu, X.; and Ji, S. 2020. XGNN: Towards model-level explanations of graph neural networks. In KDD.

Quantitative evaluation: Accuracy, Density, Consistency, Faithfulness

Method	BA-house	BA-grid	Solubility	MUTAG	Benzene	MNIST-sp
PAGE	0.5238	0.8571	0.3290	0.9090	0.6667	N/A
XGNN	0.2500	0.3200	0.2341	0.6875	0.2500	N/A

(a) Accuracy (↑).

Method	BA-house	BA-grid	Solubility	MUTAG	Benzene	MNIST-sp
PAGE	0.1481	0.1563	0.0462	0.1389	0.1111	0.2195
XGNN	0.1235	0.1667	0.1195	0.1875	0.1094	0.3593

(b) Density (↓).

Method	BA-house	BA-grid	Solubility	MUTAG	Benzene	MNIST-sp
PAGE	0.0308	0.0615	0.0846	0.1216	0.0639	0.1025
XGNN	0.2152	0.2705	0.3213	0.1269	0.2227	0.0242

(c) Consistency (↓).

Method	BA-house	BA-grid	Solubility	MUTAG	Benzene	MNIST-sp
PAGE	0.7340	0.5636	0.2164	0.4430	0.2364	0.8182
XGNN	-0.4037	-0.1636	0.0983	0.2504	-0.3091	0.1273

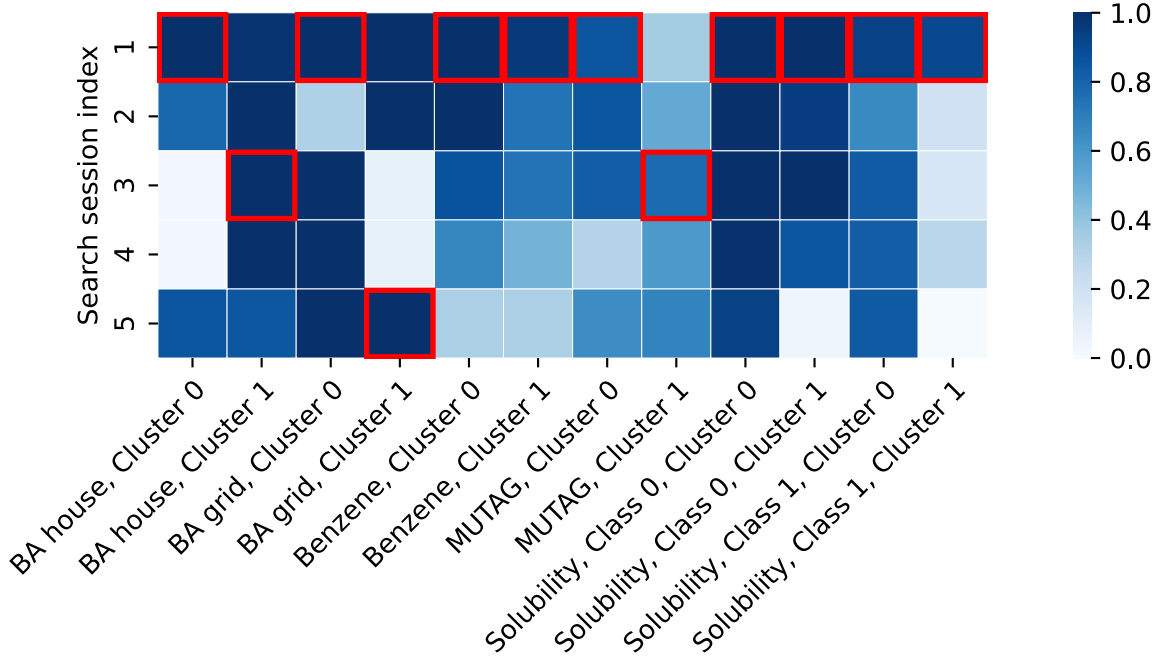
(d) Faithfulness (↑).

For 4 different metrics, PAGE shows clear supiority against XGNN for all 6 datasets.

- **Accuracy**: How much the prototype overlaps with the ground-truth explanation?
- **Density**: How much each method produces explanations with lower graph density? (i.e., sparisity)
- **Consistency**: How much robust is each method for different GNN settings (e.g., # of hidden dimensions)?
- **Faithfulness**: How much the quality of the explanation correlate with the actual performance of the GNN?

Further analysis I

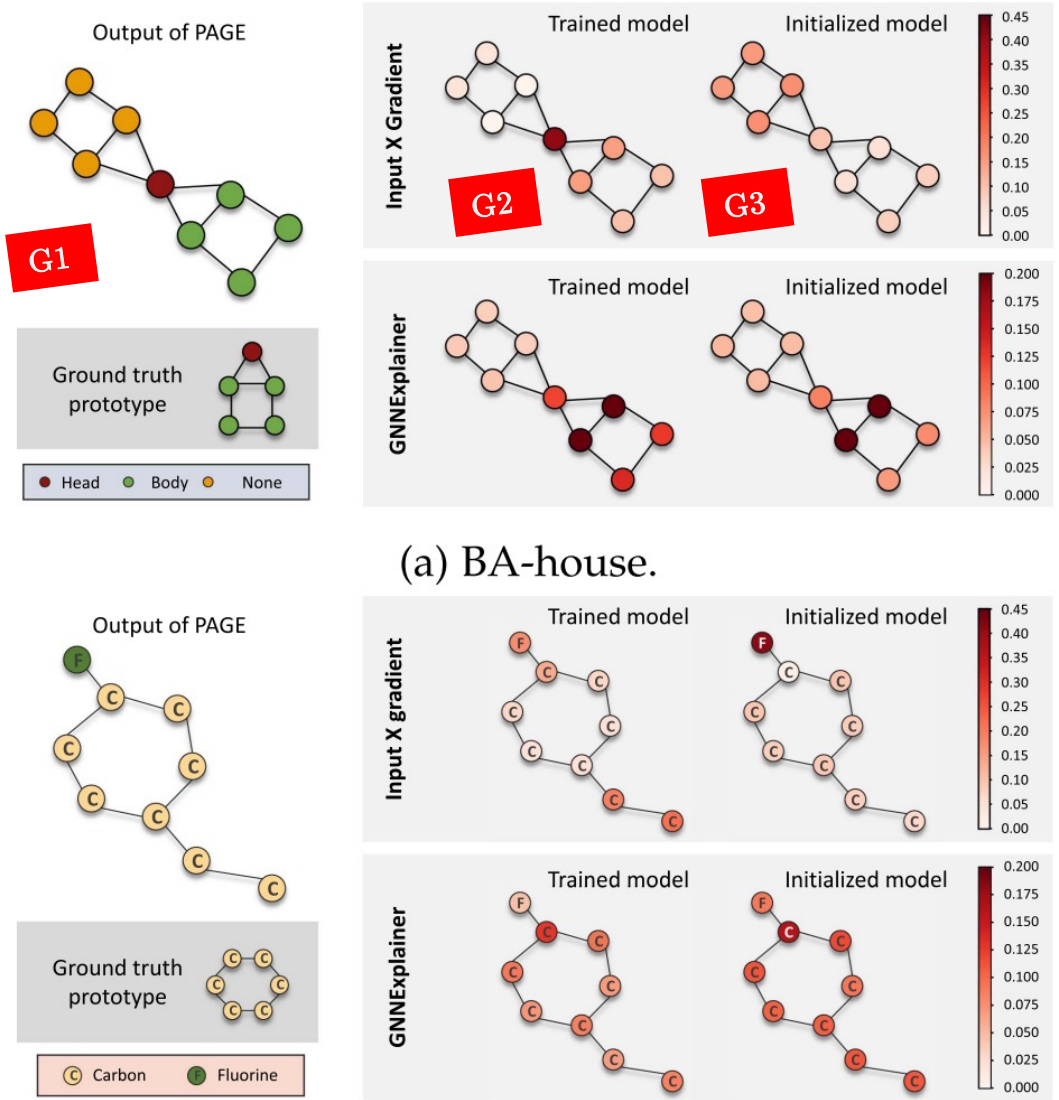
How much search trials are required?



- During phase 2, we implement a strategy that involves performing multiple attempts of extraction (with different starting points).
- Empirically, we find that PAGE can find the **best result** in the **first try** for most of the time.

Further analysis II

How much does instance-level methods agree with our results?

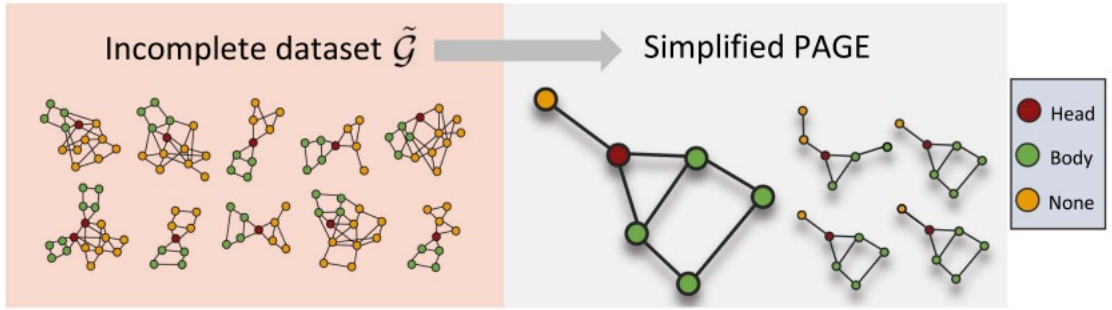


- We also explore the alignment between PAGE and other well-known instance-level explanation methods (i.e., Input X Gradient, GNNExplainer)
- We ask each instance-level explanation: How much do you think the explanation of PAGE includes the important parts in terms of explanation?
- Procedure:
  - Get G1 from PAGE.
  - Select an instance-level explanation method.
  - Ask to explain G1 as a heatmap (G2).
  - Also ask the same when the GNN is not trained (G3), which is considered as a 'baseline'.

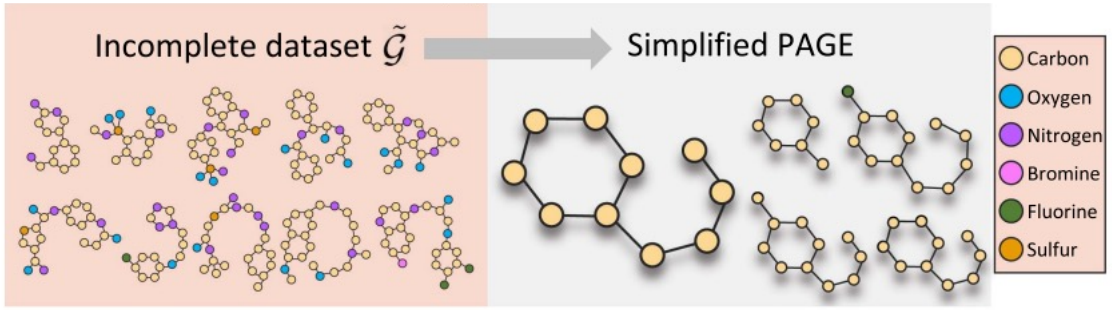
- Results show that instance-level methods generally **agree** that the prototypes of **PAGE does include essential subgraph patterns learned by the GNN**.

Further analysis III

How does PAGE perform when only a subset of dataset is available?



(a) BA-house.



(b) Benzene.

Empirical results clearly indicate PAGE can reliably run when the dataset is incomplete.

- We generate an incomplete dataset with only 10 graphs.
- We then run a simplified version of PAGE, where it skips Phase 1 (selection of k graphs via clustering) and replaces by random selection.
- We then run PAGE multiple times and observe the output.

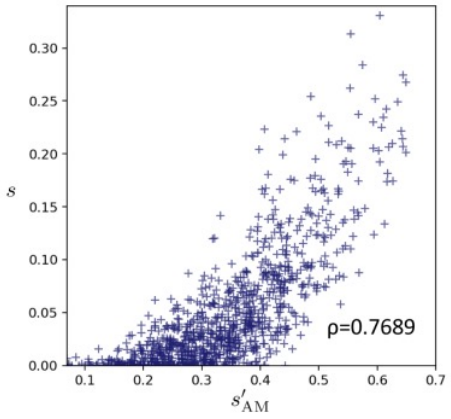
Further analysis IV

On the efficiency and quality of the Prototype Scoring Function

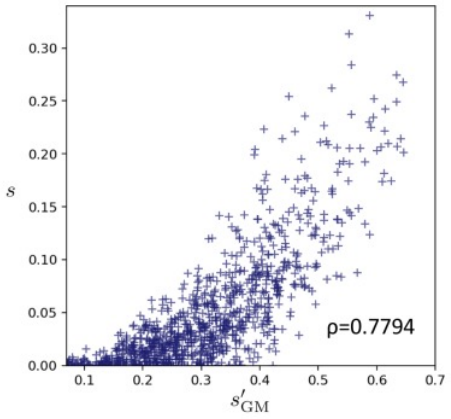
We compare our Prototype Scoring Function with naïve baselines, including a pairwise average of the arithmetic (AM) & geometric mean (GM).

TABLE IV
COMPARISON OF DIFFERENT PROTOTYPE SCORING FUNCTIONS IN TERMS OF THE RUNTIME COMPLEXITY IN MICROSECONDS (MEAN ± STANDARD DEVIATION)

Prototype scoring function	s	s'_{AM}	s'_{GM}
Time (μs)	14.42 ± 12.65	38.34 ± 17.81	31.39 ± 17.81



(a) s'_{AM} versus s



(b) s'_{GM} versus s

The runtime comparison shows that our proposed Prototype Scoring Function is faster than alternatives.

We also show that there are a high correlation between our Prototype Scoring Function and the alternatives, providing an intuitive understanding of what it calculates.

Conclusion and Outlook

- In our work, we propose PAGE, a novel model-level explanation of a GNN model that performs graph classification.
- In contrast to XGNN, which relies on reinforcement learning that requires carefully designed reward functions along with domain knowledge, our method discovers explanations within the dataset.
- Our future avenues include expanding our method to GNNs with node-level or edge-level tasks.

Thank you for your attention!

jordan3414@yonsei.ac.kr