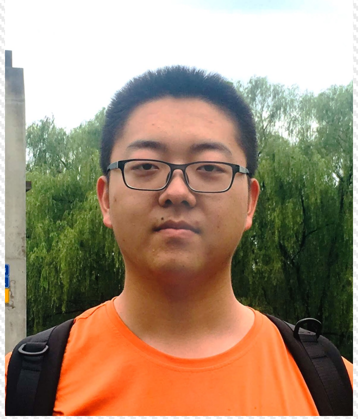
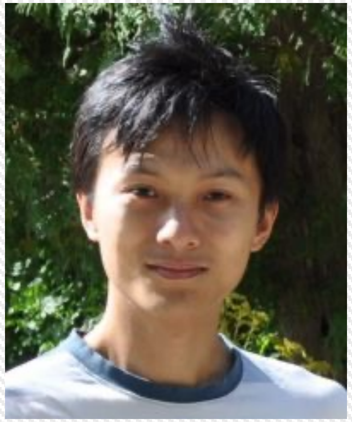
AAAI - 22

Generalized Equivariance and Preferential Labeling for GNN Node Classification


Zeyu Sun Wenjie Zhang Lili Mou Qihao Zhu Yingfei Xiong Lu Zhang

Peking University

PART ONE

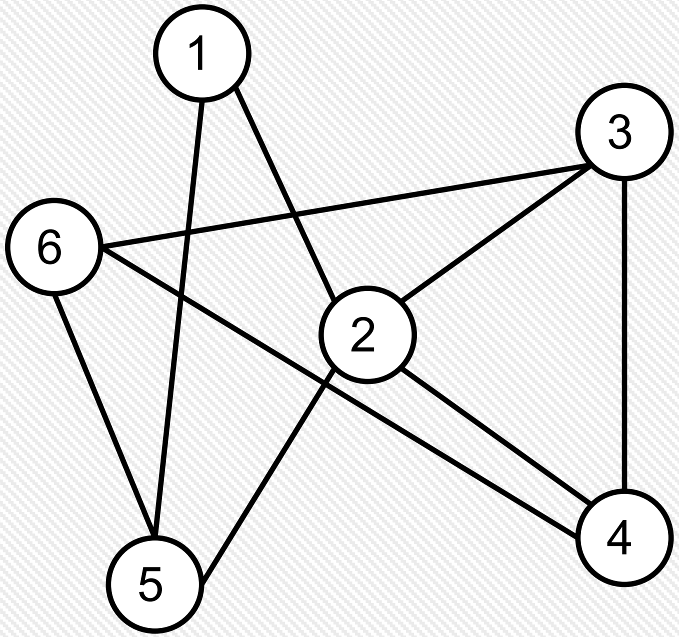
Introduction

Introduction

- Graphs are a widely used type of data structure in computer science.

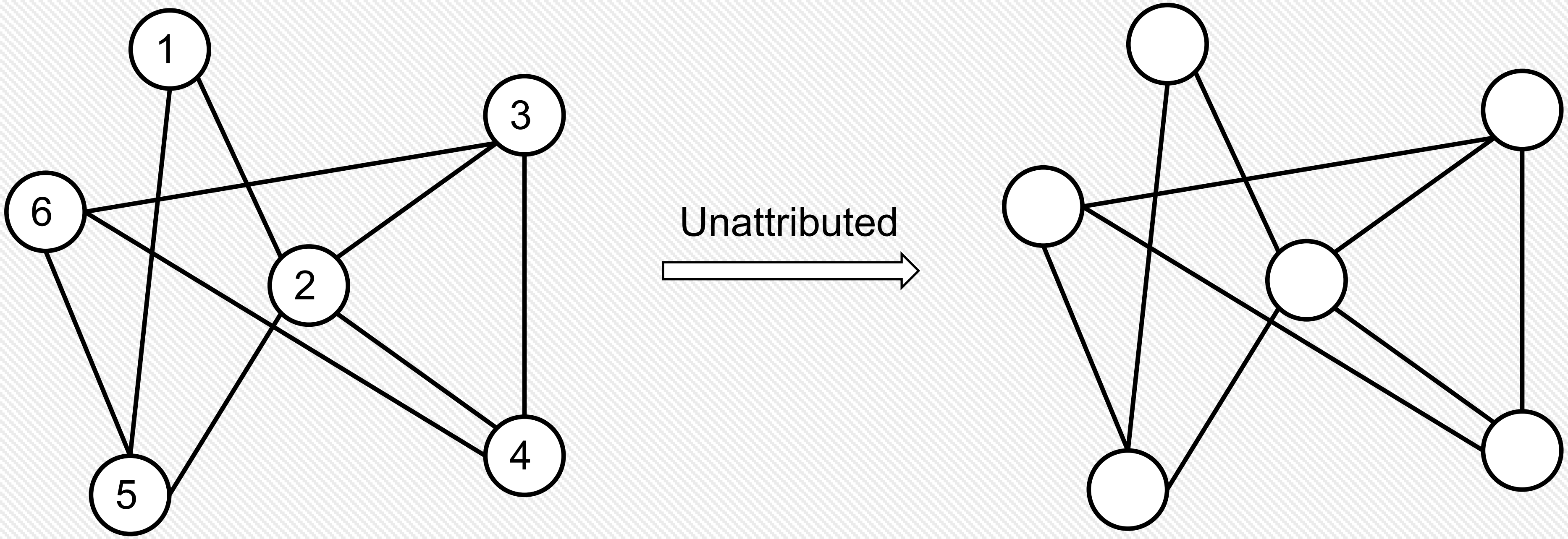
Introduction

- Graphs are a widely used type of data structure in computer science.
  - Existing GNNs highly rely on node embeddings.



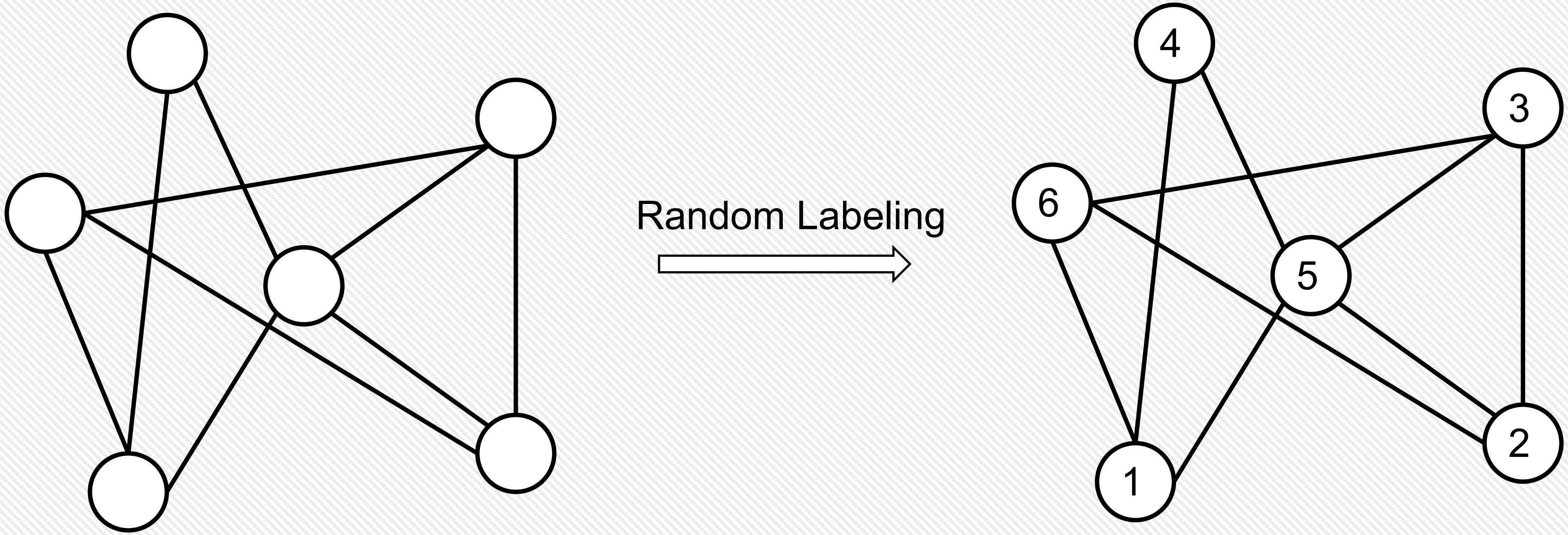
Introduction

- Graphs are a widely used type of data structure in computer science.
  - Existing GNNs highly rely on node embeddings.
  - However, the nodes in a graph may not be attributed.



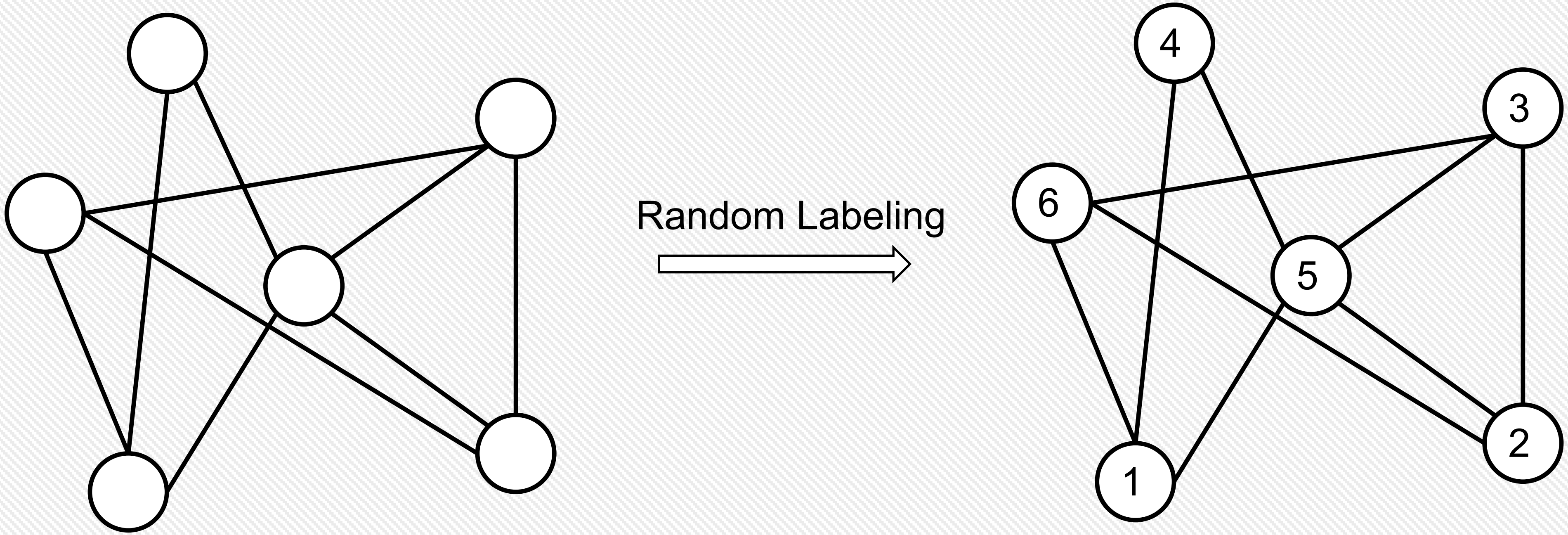
Introduction

- Unattributed graph.
  - Previous methods typically adopt an arbitrary labeling for nodes and represent them by embeddings.



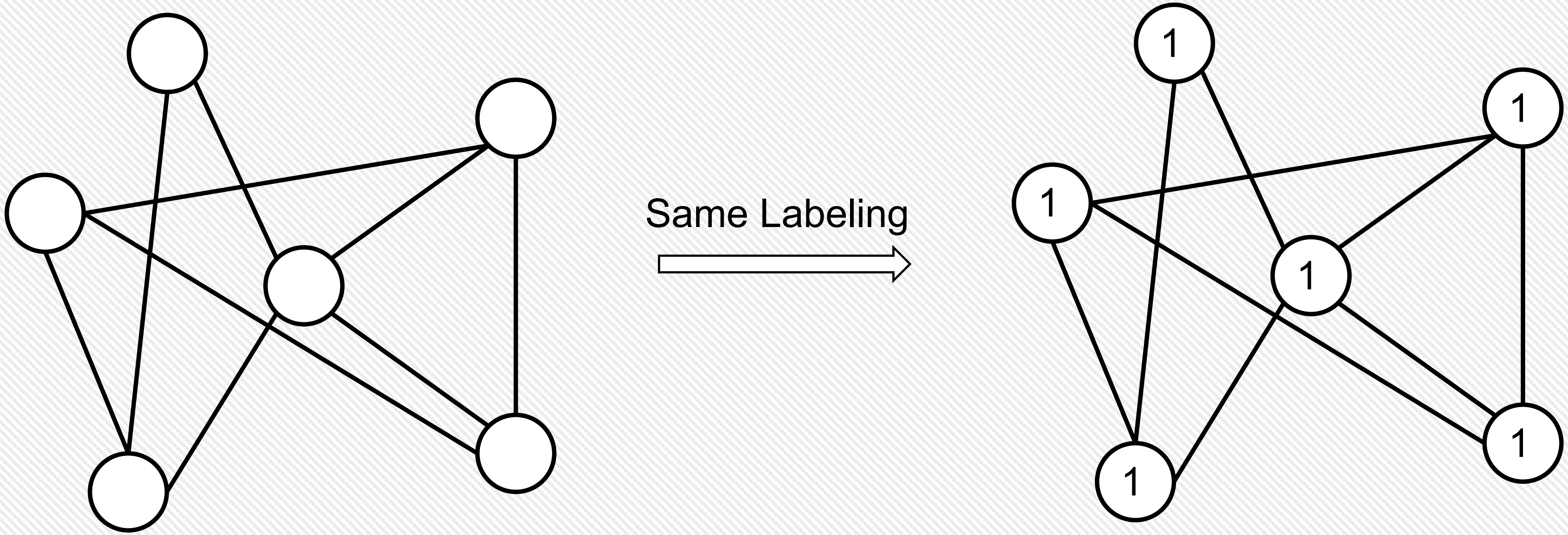
Introduction

- Unattributed graph.
  - Previous methods typically adopt an arbitrary labeling for nodes and represent them by embeddings --> ***artefacts***



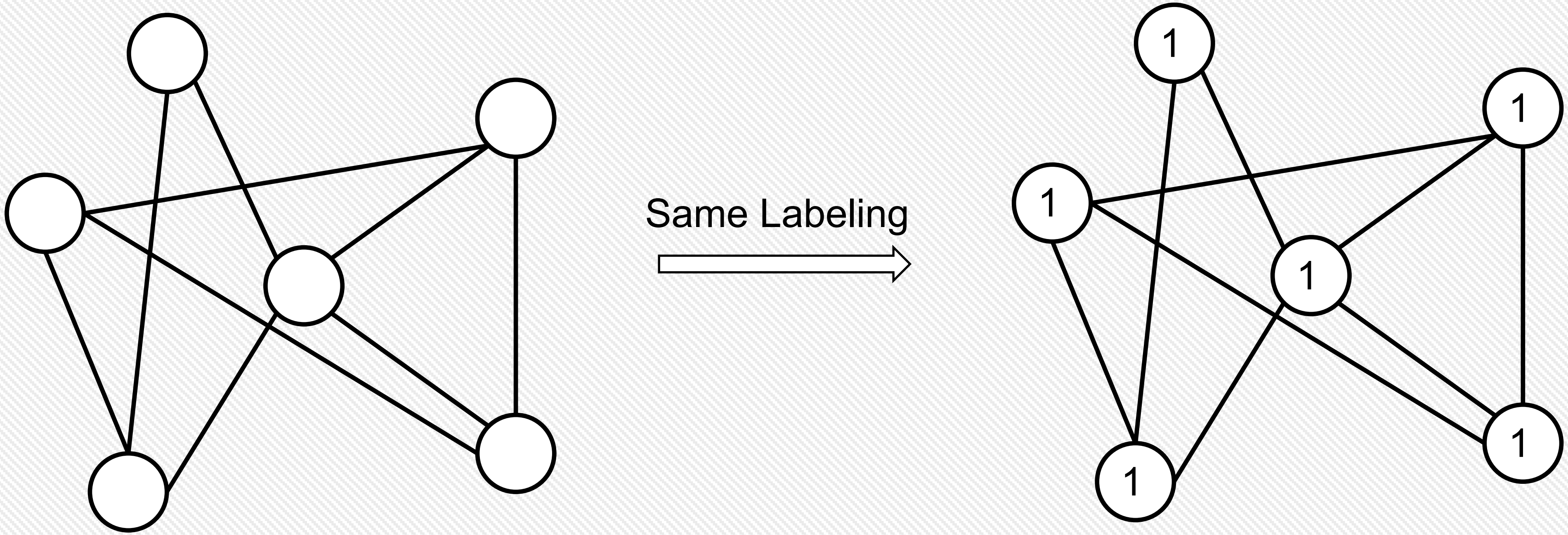
Introduction

- Unattributed graph.
  - Some approaches have realized that such artefacts are undesired, and assign all nodes with the same embedding.



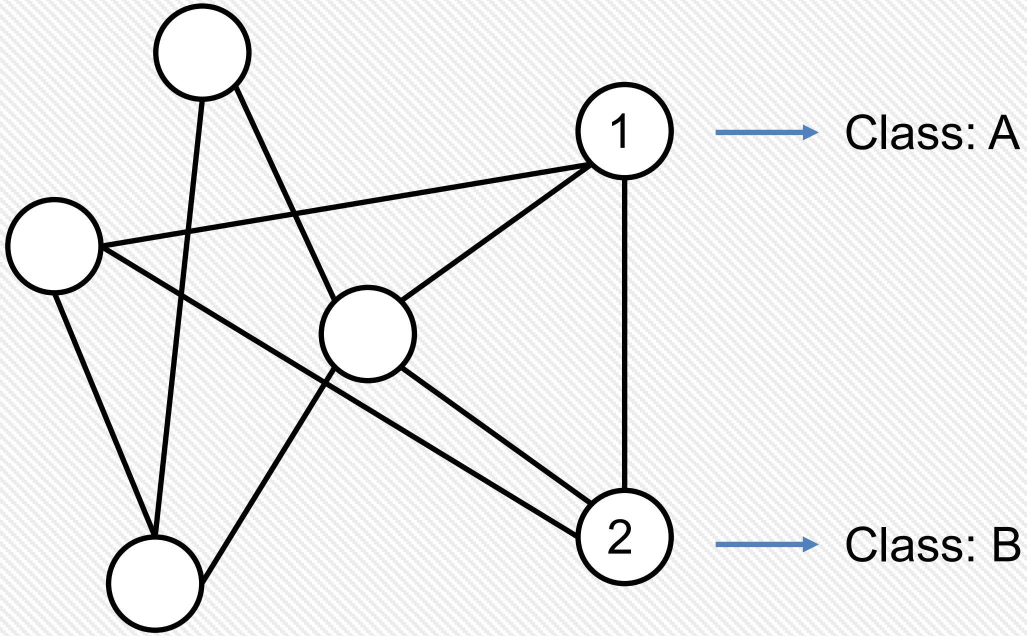
Introduction

- Unattributed graph.
  - Some approaches have realized that such artefacts are undesired, and assign all nodes with the same embedding. --> ***GNN becomes insensitive to the nodes***



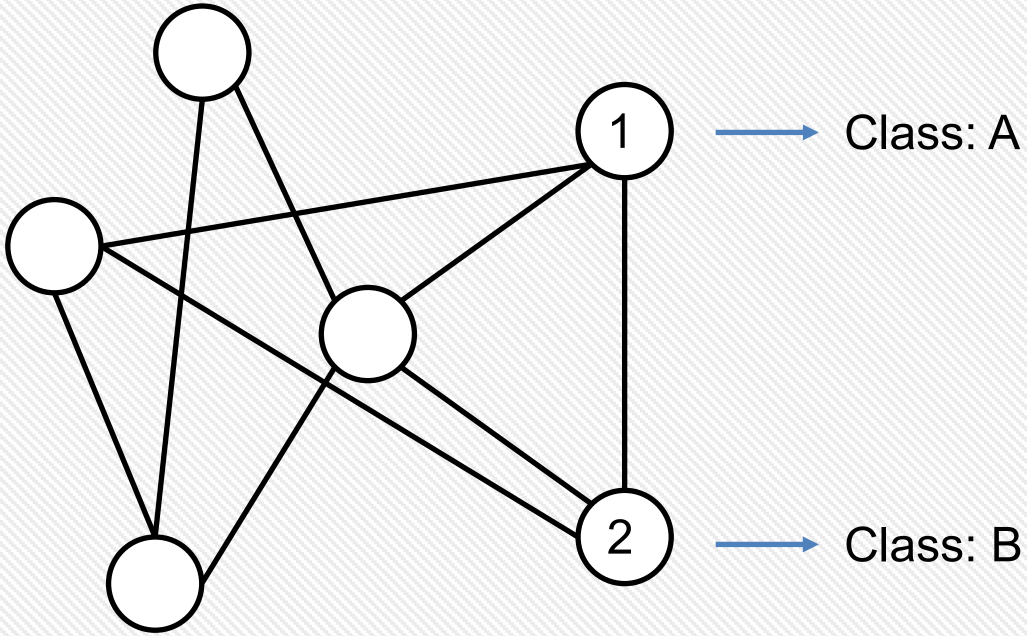
Introduction

- In this work, we analyze unattributed node classification tasks.



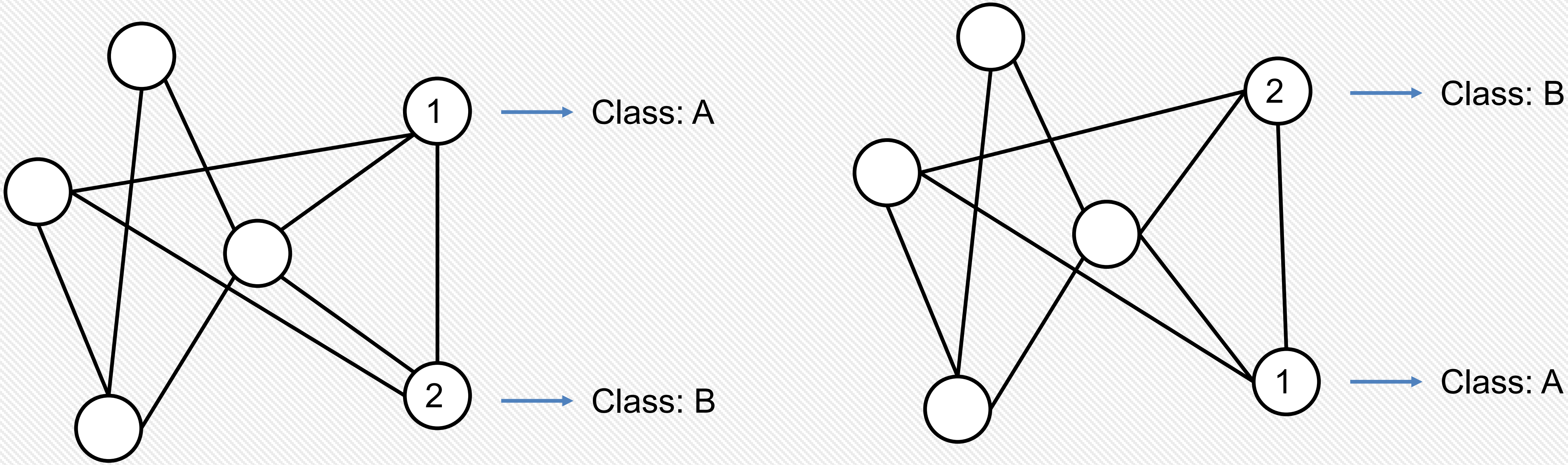
Introduction

- In this work, we analyze unattributed node classification tasks.
  - Such tasks require ***equivariance***, i.e., the change of node labels should be reflected correspondingly in the output.



Introduction

- In this work, we analyze unattributed node classification tasks.
  - Such tasks require ***equivariance***, i.e., the change of node labels should be reflected correspondingly in the output.



Introduction

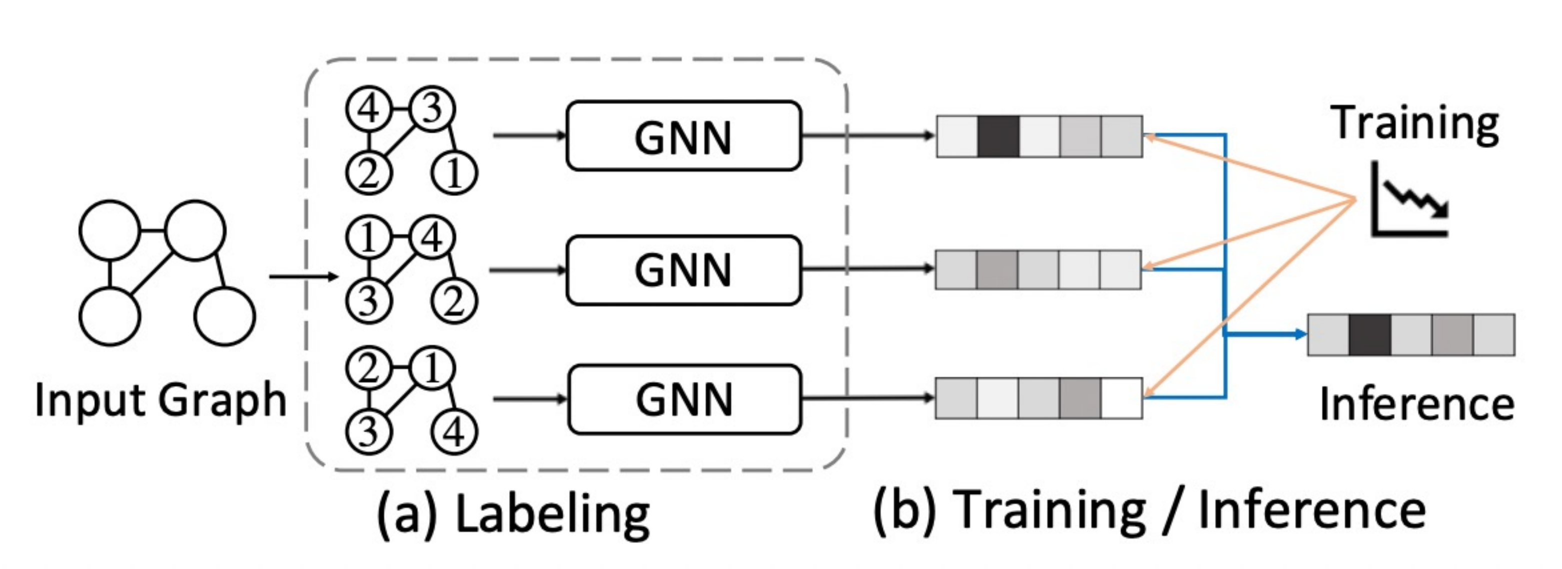
- ***Equivariance.***
  - An equivariant GNN is unable to solve equivariant node classification problems where multiple outputs are appropriate for an input graph.

Introduction

- We propose
  - a ***generalized equivariance property*** that is more suited to unattributed node classification.

Introduction

- We propose
  - a ***generalized equivariance property*** that is more suited to unattributed node classification.
  - a ***Preferential Labeling*** approach that asymptotically achieves our ***generalized equivariance property***.



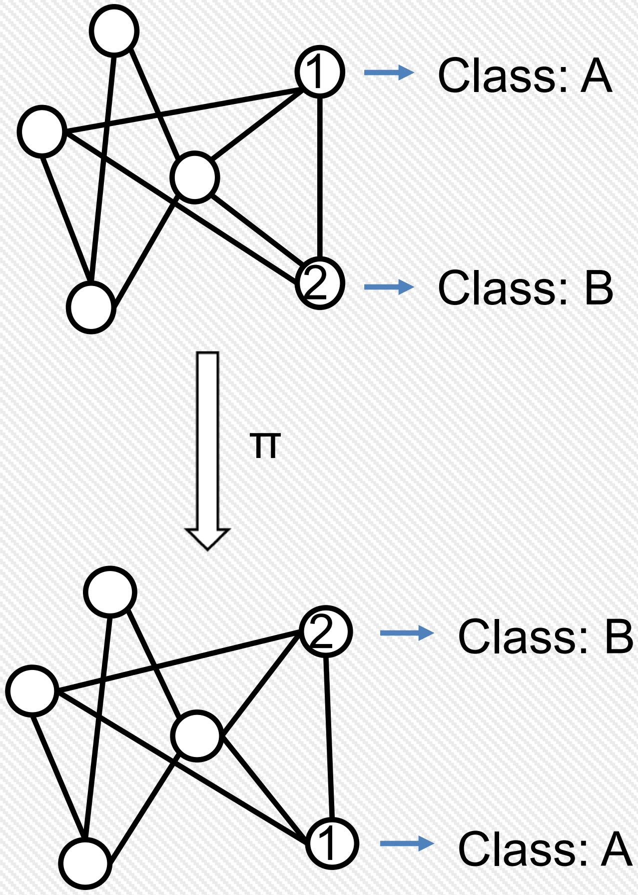
PART TWO

Methodology

Problem Formulation

- Equivariance Property

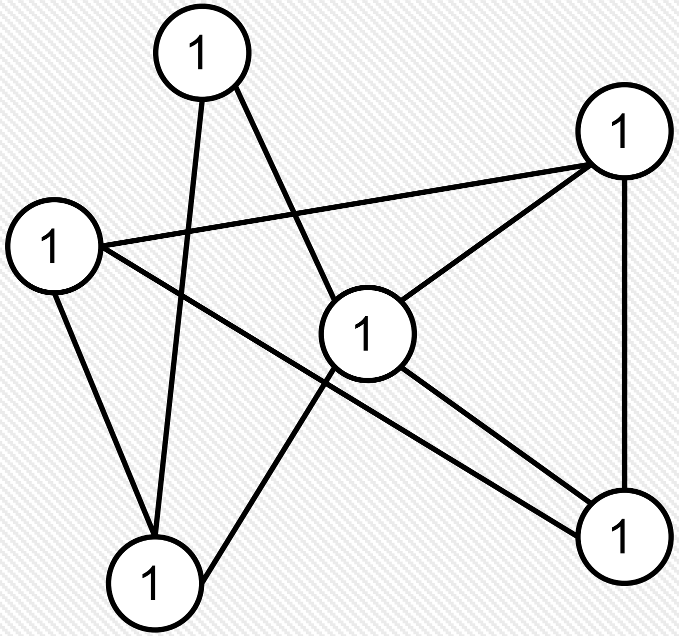
$$f(\pi(X)) = \pi(f(X))$$



- π denotes the permutation on X / f(X).
- f denotes the GNN model.
- X denotes the input graph.

Limitations of Existing GNNs on Unattributed Graphs

- Node Distinction
  - The SOTA approaches assign all nodes with the same embedding.



Limitations of Existing GNNs on Unattributed Graphs

- Node Distinction
- Equivariance Property
  - If the node index changes, the output would change accordingly.

Limitations of Existing GNNs on Unattributed Graphs

- Node Distinction
- Equivariance Property
  - If the node index changes, the output would change accordingly.

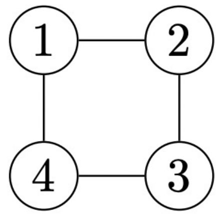


Figure 2: Graph C_4, a circle of length 4. This graph is auto-isomorhpic under $\pi : 1 \mapsto 2, 2 \mapsto 3, 3 \mapsto 4, 4 \mapsto 1$.

  - Consider the maximum independent set (MIS) problem.
    - {1,3} and {2,4} --> {1,3} and {2,4}
    - f(X) = π (f(X))

Our Solution

- A naive attempt
  - For node distinction, a naive idea is to assign embeddings by random labeling.
    - For training, node labels are assigned randomly.
    - During inference, it assigns multiple random labels and uses an average ensemble for prediction.

Our Solution

- A naive attempt
  - For node distinction, a naive idea is to assign embeddings by random labeling.
    - For training, node labels are assigned randomly.
    - During inference, it assigns multiple random labels and uses an average ensemble for prediction.

  - We found it still suffer from the limitation of the Equivariance Property.

Our Solution

- To address this issue, we propose a desired ***generalized equivariance property***.

Our Solution

- To address this issue, we propose a desired ***generalized equivariance property***.

$$X = \pi(X_*)$$

$$\gamma(X_*) = X_* \quad \text{and} \quad f(\pi(X_*)) = \pi\gamma(f(X_*)).$$

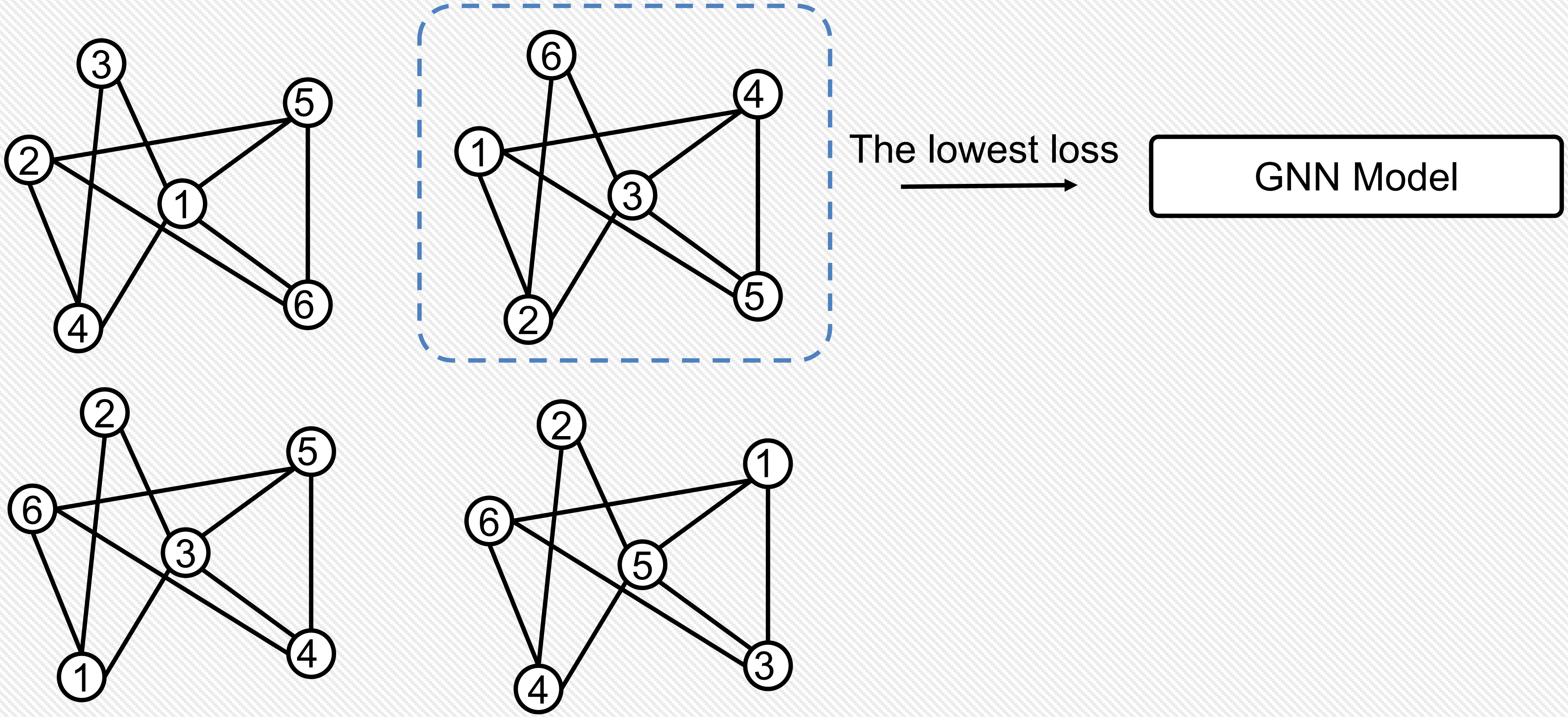
  - π and γ denote the permutation on X / f(X).
  - f denotes the GNN model.
  - X denotes the input graph.

Our Solution

- We further propose a simple yet effective approach, ***Preferential Labeling***, which asymptotically satisfies ***generalized equivariance property.***

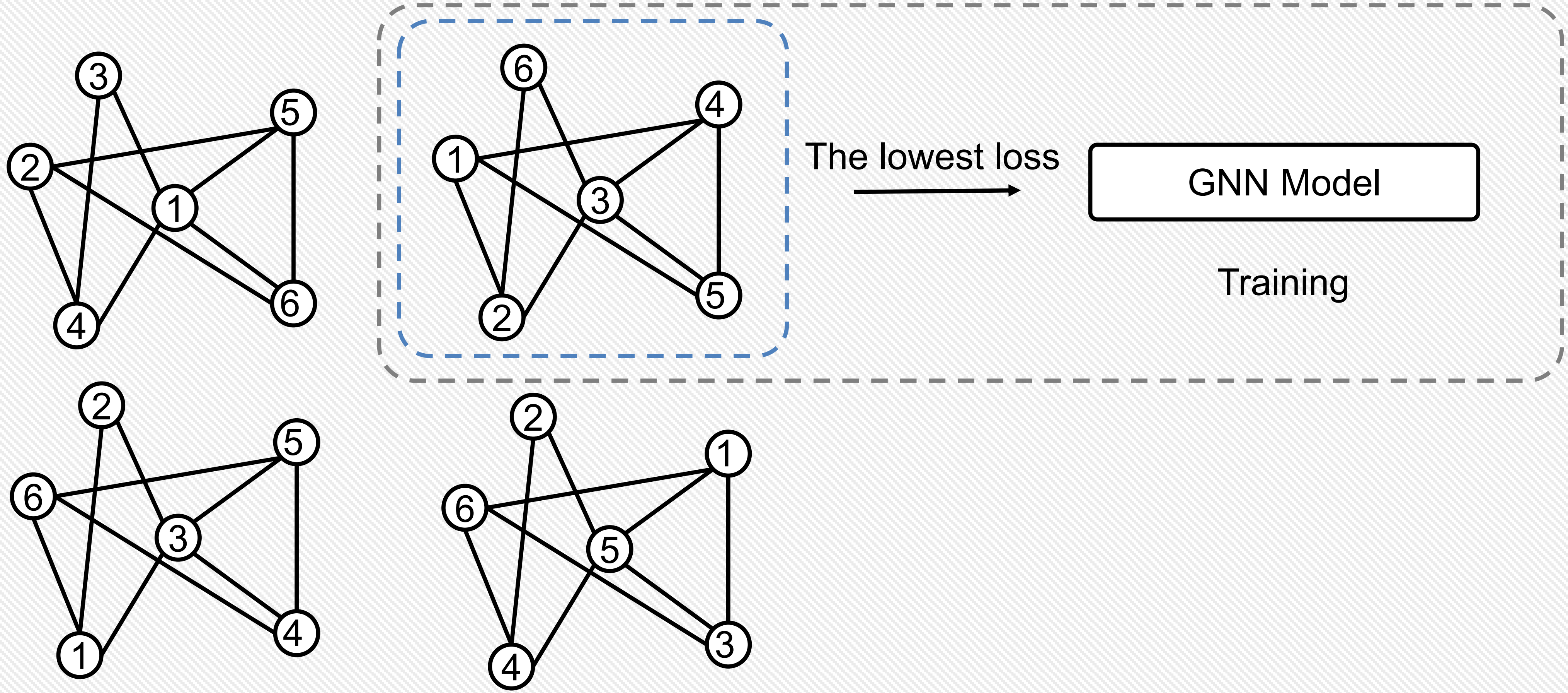
Our Solution

- We further propose a simple yet effective approach, ***Preferential Labeling***, which asymptotically satisfies ***generalized equivariance property.***



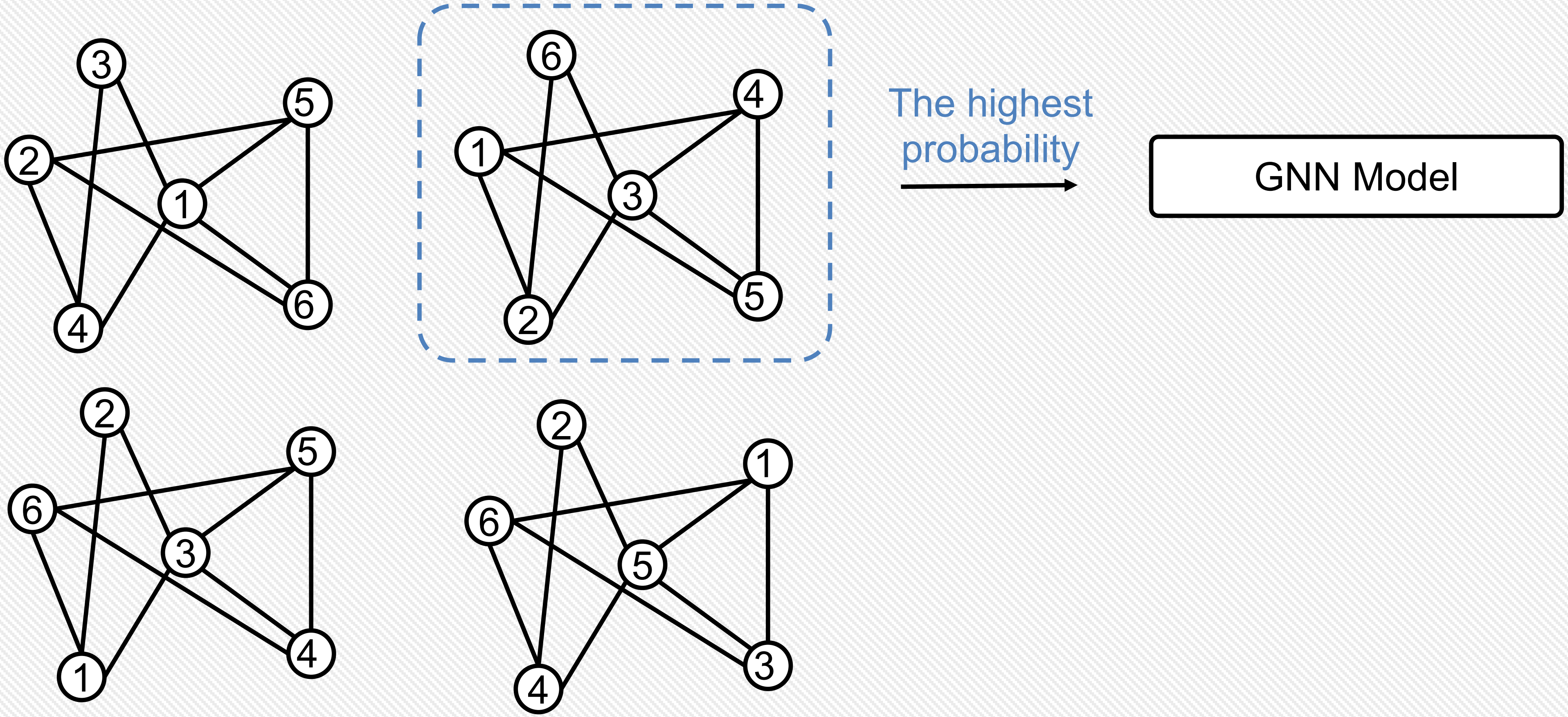
Our Solution

- We further propose a simple yet effective approach, ***Preferential Labeling***, which asymptotically satisfies ***generalized equivariance property.***



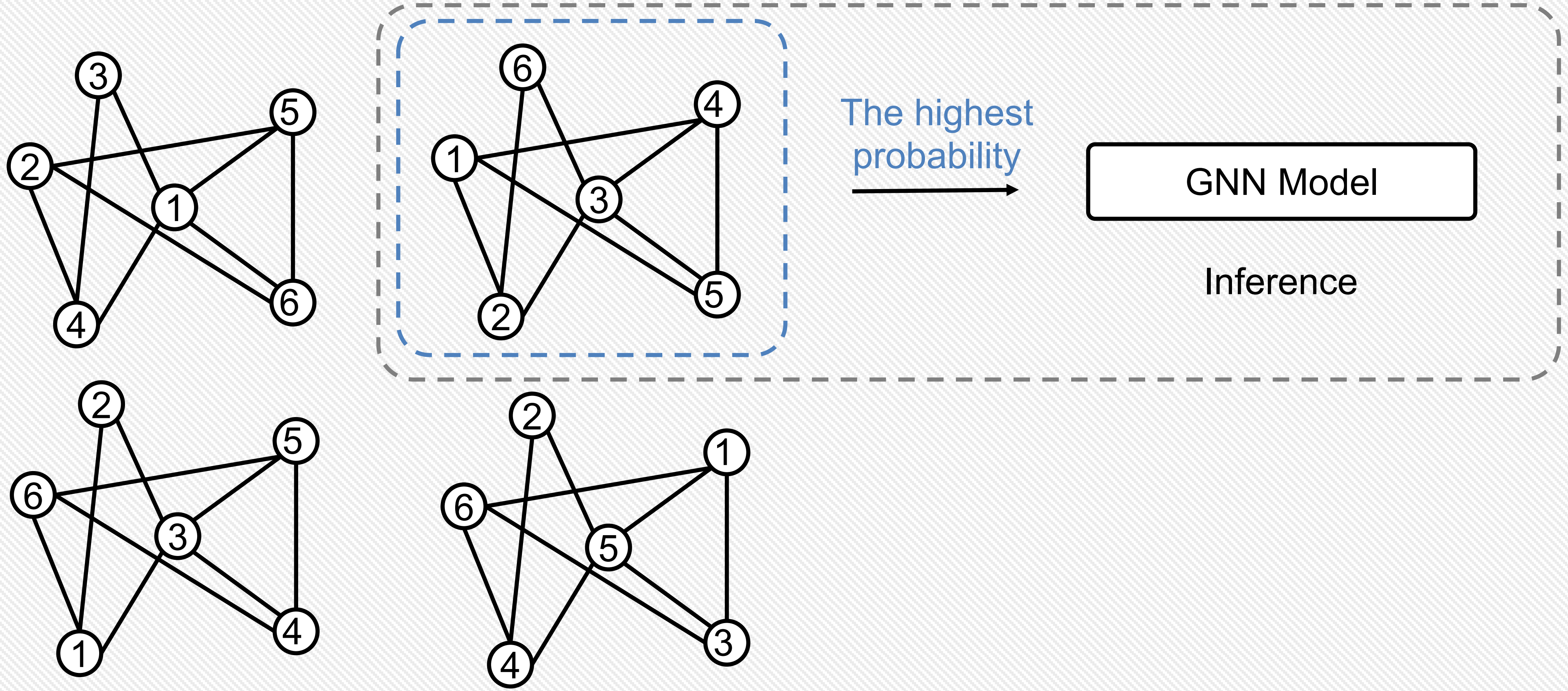
Our Solution

- We further propose a simple yet effective approach, ***Preferential Labeling***, which asymptotically satisfies ***generalized equivariance property.***



Our Solution

- We further propose a simple yet effective approach, ***Preferential Labeling***, which asymptotically satisfies ***generalized equivariance property.***



Our Solution

- We further propose a simple yet effective approach, ***Preferential Labeling***, which asymptotically satisfies ***generalized equivariance property.***
  - Theoretical Analysis

PART THREE

Experiments

Competing Methods

- **Static Labeling**. The static labeling assigns an embedding based on the identity of a node.
- **Same Embedding**. This baseline assigns all nodes in the unattributed graph with the same embedding.
- **Random Labeling**. The random labeling assigns an embedding randomly during training and inference.
- **Degree Feature**. We use 1/(d + 1) as a one-dimensional, non-learnable embedding feature.
- **Degree Ranking Embedding**. We sort all nodes by the degrees in descending order, and a node having i-th largest degree is encoded by i-th embedding vector.

MIS Solving

- Solving the maximum independent set (MIS).
- **Model**. In this experiment, we adopt the state-of-the-art model GCN.

MIS Solving

- Solving the maximum independent set (MIS).
- **Model**. In this experiment, we adopt the state-of-the-art model GCN.

Row #	GCN (Li, Chen, and Koltun 2018)	Accuracy
1	Same	75.59%
2	Degree Feature	73.22%
3	Degree Ranking Embedding	71.58%
4	Static Labeling	74.57%
5	Random Labeling	75.28%
6	Preferential Labeling-10	**85.04%**

SAT Solving

- Solving the propositional satisfiability problem (SAT).
- **Model.** The GNN model and settings are generally adopted from the state-of-the-art NLocalSAT.

SAT Solving

- Solving the propositional satisfiability problem (SAT).
- **Model.** The GNN model and settings are generally adopted from the state-of-the-art NLocalSAT.

Row #	NLocalSAT (Zhang et al. 2020)	Error Rate Test-5	Test-10	Test-20	Test-40	Avg.
1	Same	5.26%	8.17%	15.03%	27.62%	14.02%
2	Degree Feature	5.31%	8.37%	14.25%	24.94%	13.22%
3	Degree Ranking Embedding	5.45%	10.23%	16.17%	28.04%	14.97%
4	Static	6.11%	9.86%	16.89%	28.88%	15.44%
5	Static & Inference-10 (Averaging)	5.00%	8.77%	15.74%	29.70%	14.80%
6	Static & Inference-10 (Max Prob.)	1.77%	3.65%	7.86%	16.22%	7.38%
7	Random	3.38%	6.17%	12.70%	23.66%	11.48%
8	Random & Inference-10 (Averaging)	3.39%	6.07%	12.42%	23.34%	11.31%
9	Random & Inference-10 (Max Prob.)	2.72%	5.03%	11.37%	22.06%	10.30%
10	Preferential Labeling-10 (Max Prob.)	**1.13%**	**1.68%**	**1.81%**	**5.24%**	**2.47%**

Conclusion

- We analyze the limitations of existing GNNs.
- We propose a ***generalized equivariance property*** and ***Preferential Labeling***.

Thanks

Generalized Equivariance and Preferential Labeling for GNN Node Classification

Zeyu Sun, Wenjie Zhang, Lili Mou, Qihao Zhu, Yingfei Xiong, Lu Zhang

szy_@pku.edu.cn