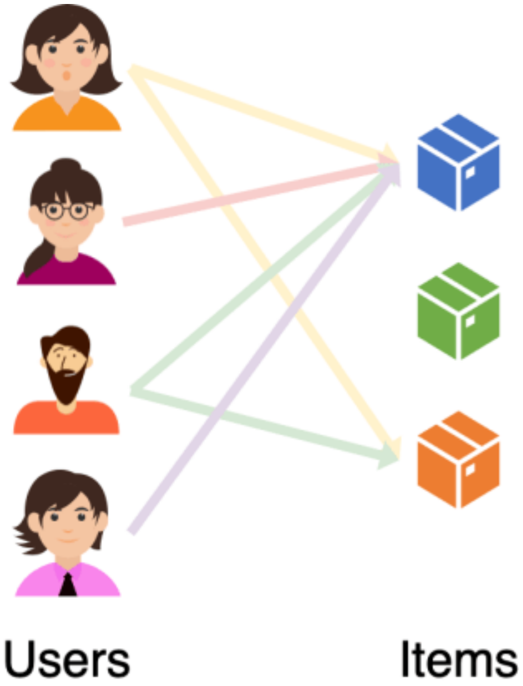
NExT++ Search Centre
A NUS-Tsinghua Joint Centre on Extreme Search

# Should Graph Convolution Trust Neighbors? A Simple Causal Inference Method

Fuli Feng[1,2], Weiran Huang[3], Xiangnan He[4], Xin Xin[5], Qifan Wang[6], Tat-Seng Chua[2]
[1]Sea-NExT Joint Lab, [2]National University of Singapore, [3]The Chinese University of Hong Kong
[4]University of Science and Technology of China, [5]University of Glasgow, [6]Google US
fulifeng93@gmail.com

# Graph & Applications in IR

Graph has become the default choice for relational data modeling in many IR applications.



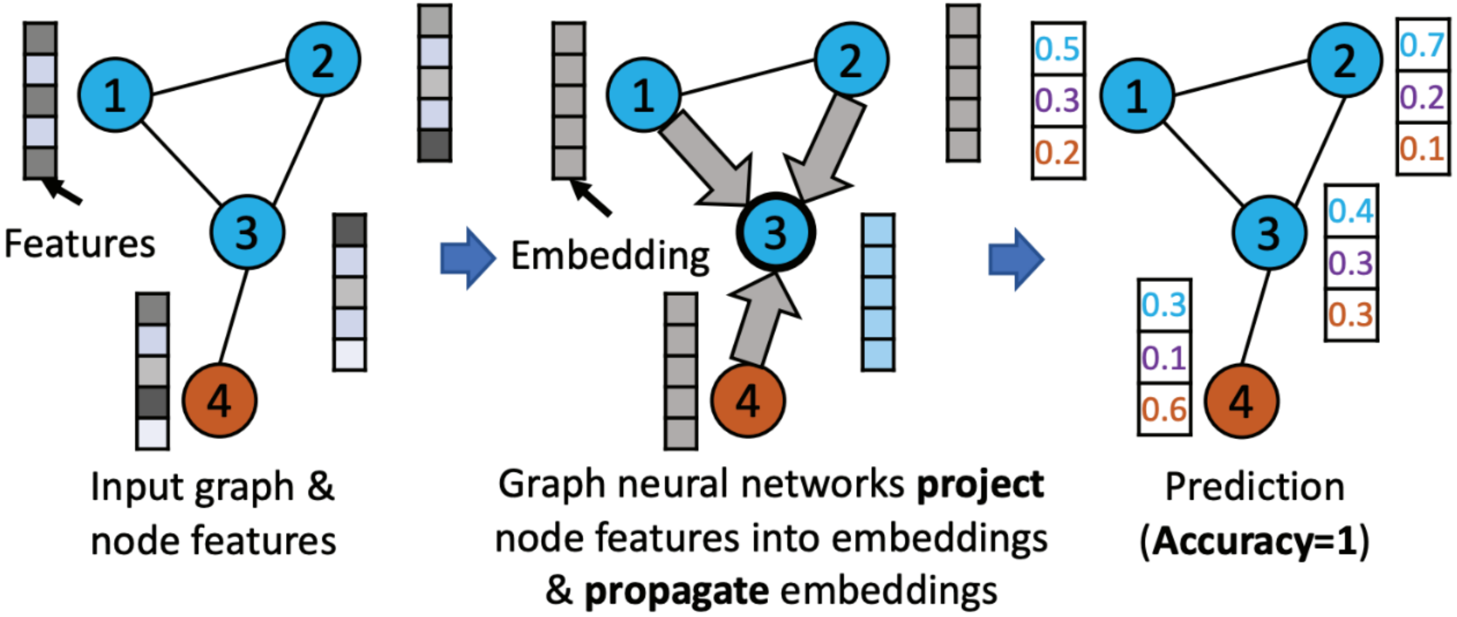

**Social network**: **Node** → user; **Edge** → following
**IR applications:** user profiling, rumour detection, targeted advertising, etc.

**User behavior**: **Node** → user/item; **Edge** → click/buy
**IR applications:** recommendation.

Zhang Munan, et al. *"Inductive Matrix Completion based on Graph Neural Networks."* (2019) ICLR.

# Graph Convolutional Network

**Graph-based learning**: leverage the graph structure to make better predictions.



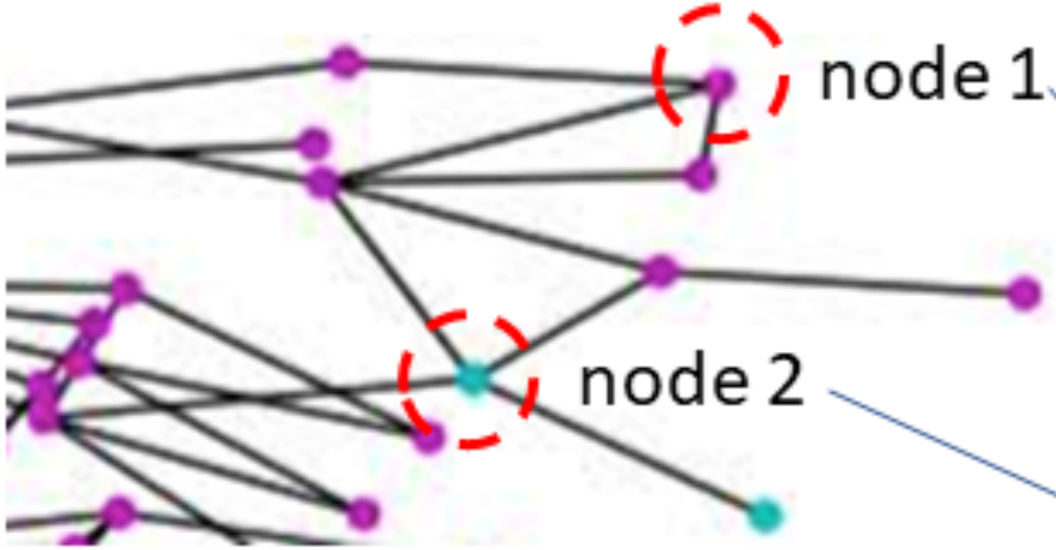

**Graph Convolutional Network (GCN)**

- Node features are propagated over the graph structure.
- ***Node 3*** ← *{Node 1, Node 2, Node 4}* + ***Node 3***
- Node prediction is made after the aggregation.

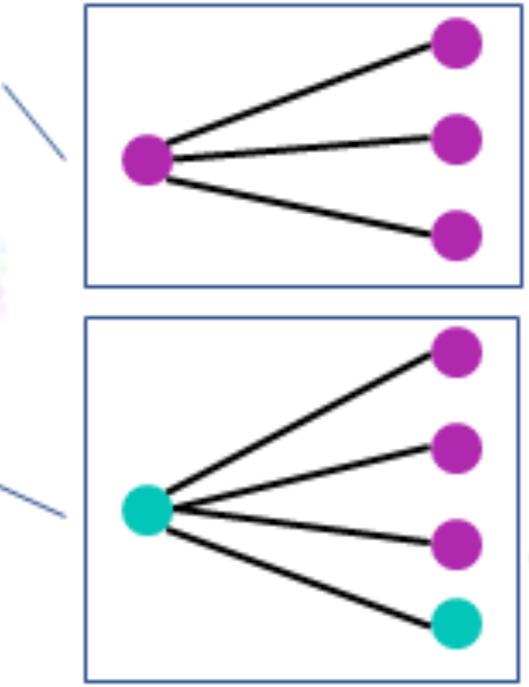
GCN is being increasingly used in IR applications, ranging from search engines, recommender systems to question-answering systems.

Feng, Fuli, et al. *"Graph adversarial training: Dynamically regularizing based on graph structure."* (2019).

# Local Structure Discrepancy Issue




A part of OGB-arXiv




1) Should GCN always trust the neighbors?
- **Node 1**: Yes
- **Node 2**: No!

2) The distribution of cross-category edges is not consistent over nodes
- Distribution drift

# Existing Solutions

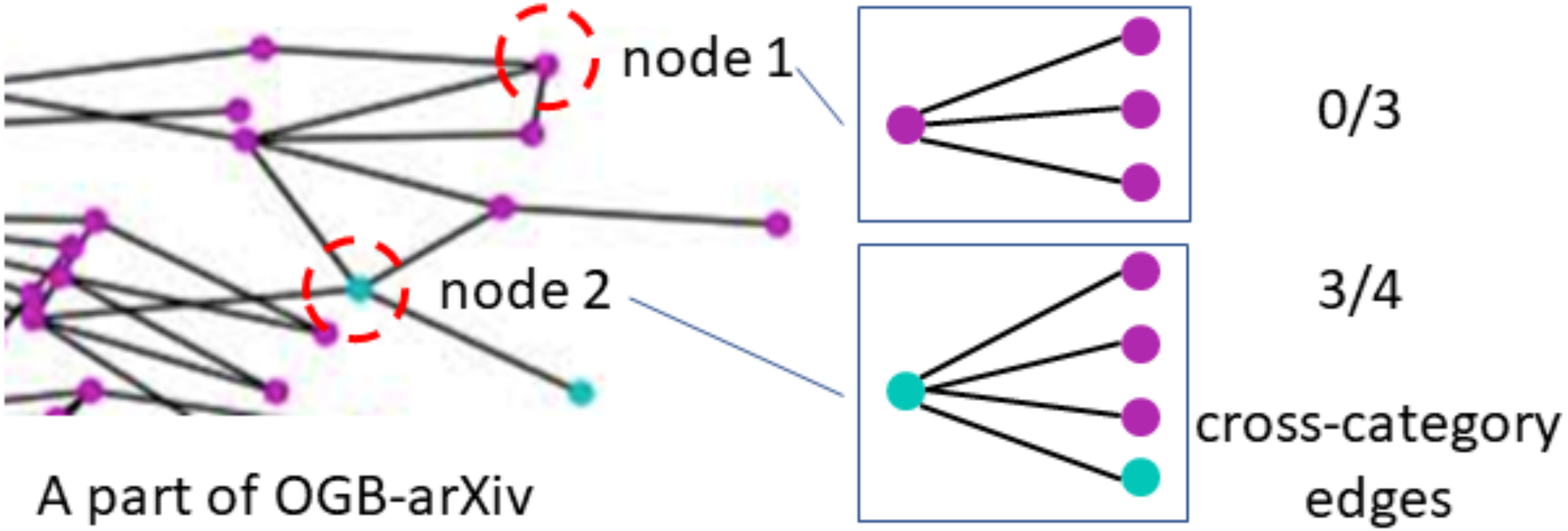
Model Training: Mitigate the impact of the discrepancy issue.

- Denoising
  - Edge classification: identify & remove the cross-category edges
  - Spectral filtering: filter out the high-frequency signal in the adjacency matrix

- Graph attention
  - Neighbor attention: adjust the contribution of neighbors
  - Hop/layer attention: adjust the contribution of neighbors at difference hops

1) Not easy to be trained well in practice; and 2) Hard to generalize well to testing nodes.

# Handling Discrepancy During Model Inference




A part of OGB-arXiv

**Existing method**:
One-pass inference, **indiscriminate** for Node 1 and Node 2.

**Our expectation**:
**Node specific** inference, trust neighbor less when making prediction for Node 2.

How does the neighbors affect the prediction

# Causal Effect & Causal Intervention

- **Causal Graph:**

  Graphical models used to encode assumptions about the data-generating process.

- **Intervention on X [ term: do(X=x) ]**

  Study specific causal relationships between X and the target variable.

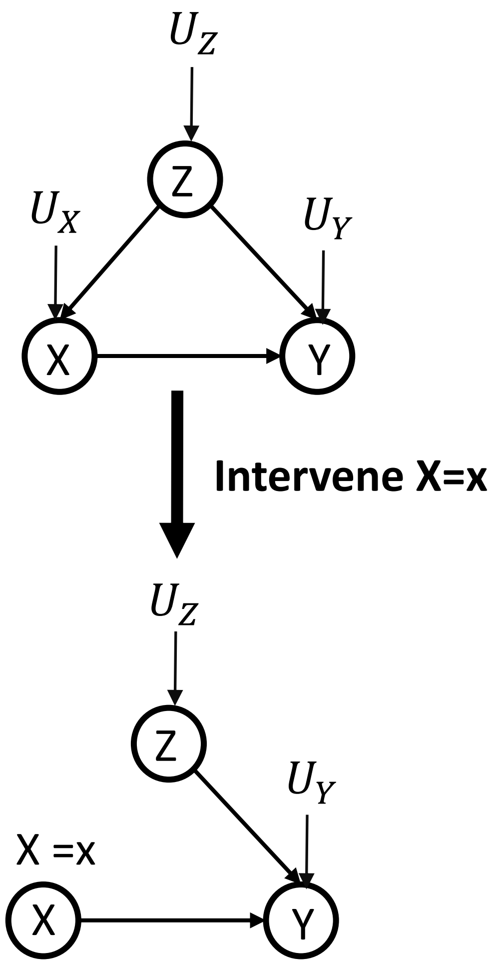
  Randomized controlled trial.

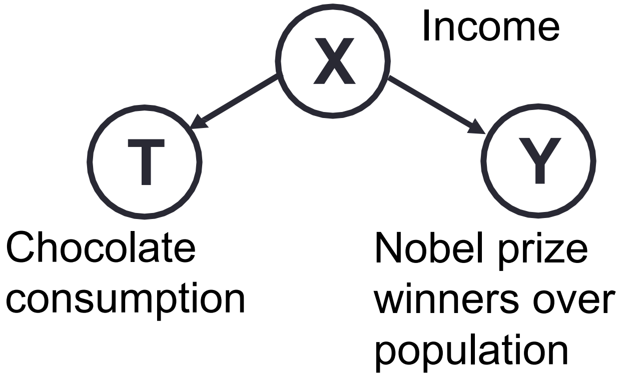
  In graph: Cut off the paths that point into X

- **Causal Effect:**

  P( Y| do(X=$x$)) – P(Y | do(X=$x_{ref}$))

  measures the expected increase in Y as the treatment changes from $X = x$ to X=$x_{ref}$



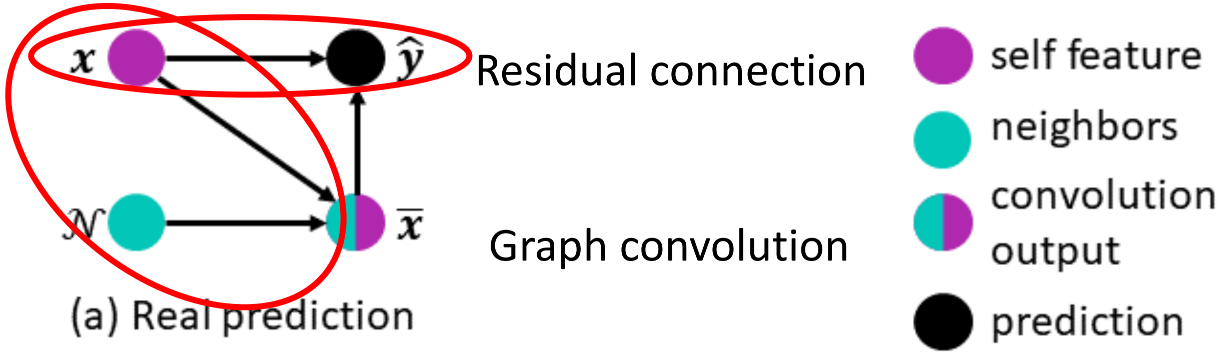

- No causal effect from T to Y

  *Promoting chocolate consumption leads to more Nobel prizes?*

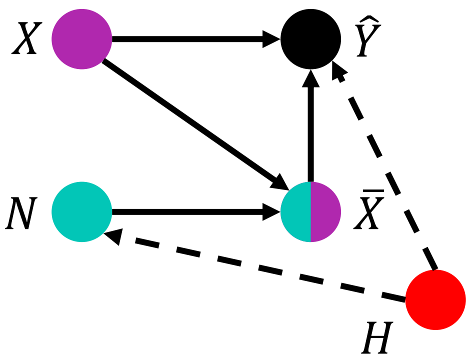
Pearl, J., & Mackenzie, D. "*The book of why: the new science of cause and effect.*" (2018).

# Causal GCN Inference Mechanism

## A causal view of generating node prediction

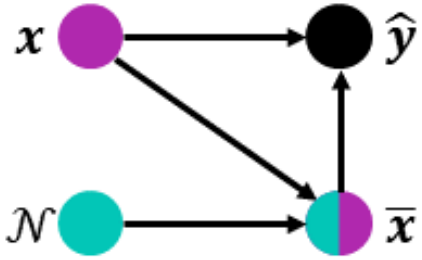



(a) Real prediction




- Unmeasured confounder

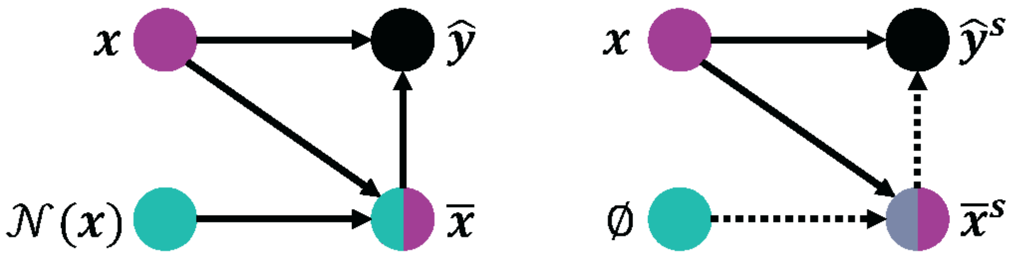
$H$ is the homophily of neighbors




- Drift

$\mathcal{N}$ changes

## Should GCN trust the neighbors?



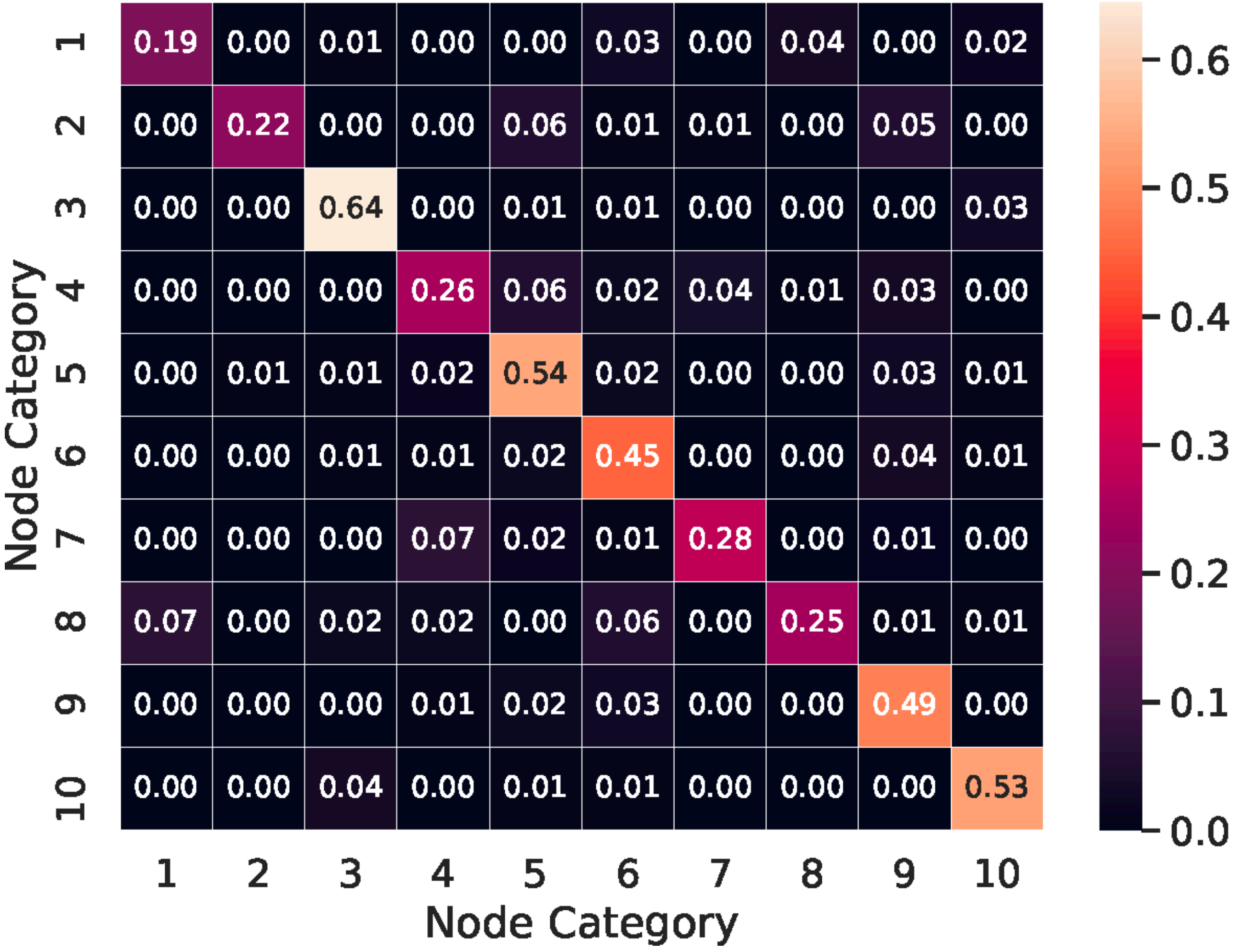

(b) Original prediction (c) Post intervention prediction

$$choose(\underbrace{\hat{y}}_{\text{Original prediction}}, \underbrace{\hat{y}^s}_{\text{Post intervention prediction}})$$

Training a simple binary classifier (choice model) to make choice

# Factors for Making Choice & Causal Uncertainty

Factors for making choice: prediction confidence, category transition, causal uncertainty.




- Causal effect of $\mathcal{N}$

$$\boldsymbol{e} = f(\boldsymbol{x}, \mathcal{N}(\boldsymbol{x})|\hat{\theta}) - f(\boldsymbol{x}, do(N = \emptyset)|\hat{\theta}),$$
$$= f(\boldsymbol{x}, \mathcal{N}(\boldsymbol{x})|\hat{\theta}) - f(\boldsymbol{x}, \emptyset|\hat{\theta}),$$
$$= \hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}^s.$$

- Variance of causal effect

$$\boldsymbol{v} = var(\{f(\boldsymbol{x}, \mathcal{N}(\boldsymbol{x})_k|\hat{\theta})|k \leq K\}),$$

$\mathcal{N}(\boldsymbol{x})_k$ is a sample of the neighbors by randomly dropping edges

# A New Schema for Training GCN

- Training for the choice model
- Two-pass GCN inference
- Choice model inference

**Algorithm 1** Applying CGI to GCN

**Input:** Training data $\boldsymbol{X}$, $\boldsymbol{A}$, $\boldsymbol{Y}$.

```
/* Training */
Optimize Equation (4), obtaining GCN (θ̂);                          ▷ GCN training
Construct 𝒟;                                                       ▷ Causal intervention
Optimize Equation (10), obtaining choice model (η̂);                ▷ CGI training
Return θ̂ and η̂.
/* Testing */
Calculate f(x, N(x)|θ̂);                                            ▷ Original prediction
Calculate f(x, ∅|θ̂);                                               ▷ Post-intervention prediction
    Calculate final classification with Equation (8);
```

# EXP1: Semi-supervised Setting (Discrepancy)

Add 10/30/50% cross-category edges on 50% randomly selected nodes

| Dataset | Citeseer(10%) | Citeseer(30%) | Citeseer(50%) |
|---|---|---|---|
| APPNP | 71.0% | 64.4% | 64.2% |
| APPNP_Self | 65.1% | 62.9% | 64.3% |
| APPNP_CGI | **71.8%** | **66.9%** | **68.6%** |
| RI | 1.1% | 3.9% | 7.2% |

**Table 1: Performance of APPNP's original prediction, post-intervention prediction, and CGI prediction on the three synthetic datasets *w.r.t.* accuracy. RI means the relative improvement over APPNP achieved by APPNP_CGI.**

- The causal GCN inference mechanism indeed mitigates the discrepancy issue.
- The relative improvement increases when facing more severe discrepancy.

# EXP2: Semi-supervised Setting (Random Split)

| Dataset | Cora | Citeseer | Pubmed |
|---|---|---|---|
| APPNP | 81.8% | 72.6% | 79.8% |
| APPNP_Self | 69.3% | 66.5% | 75.9% |
| APPNP_Ensemble | 78.0% | 71.4% | 79.2% |
| APPNP_CGI | **82.3%** | **73.7%** | **81.0%** |
| RI | 5.5% | 2.8% | 2.3% |

**Table 2: Performance of APPNP with different inference mechanisms on three semi-supervised node classification datasets *w.r.t.* the classification accuracy. RI means the relative improvement of APPNP_CGI over APPNP_Ensemble.**

- The causal GCN inference mechanism is effective in the conventional setting.
  - Insufficient labels: 20-shot per class

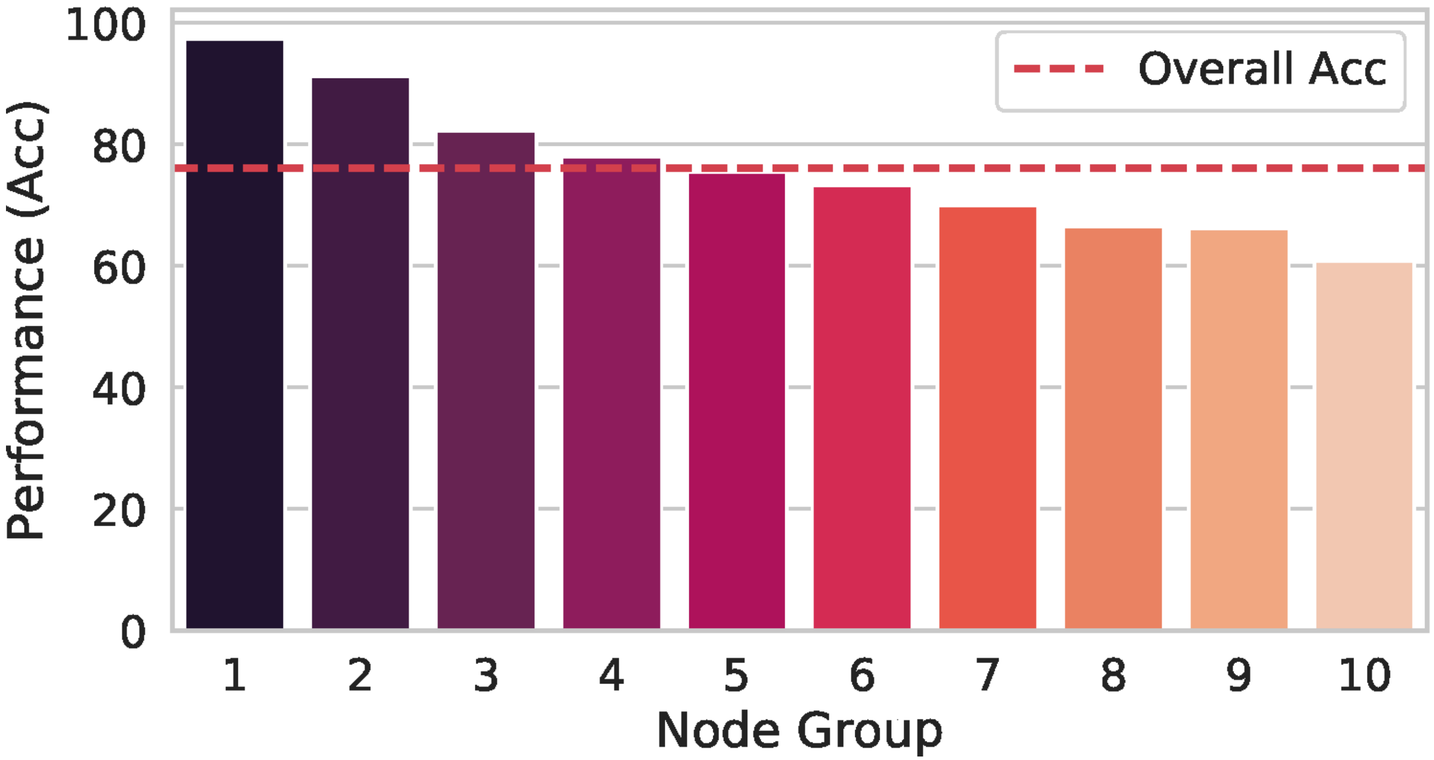
# EXP3: Full supervised Setting

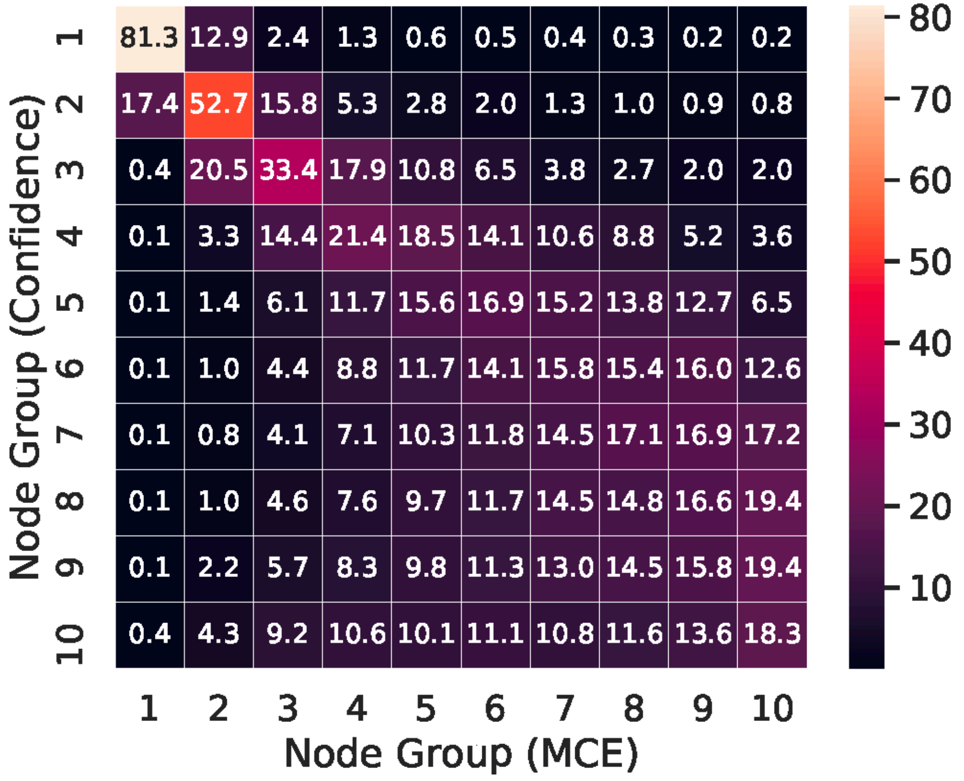
| | | | |
|---|---|---|---|
| RoBERTa (768) | JKNet | 75.59% | 75.54% |
| | MLP | 72.26% | 72.26% |
| | DAGNN | 74.93% | 74.83% |
| | APPNP | 75.74% | 75.61% |
| | APPNP_Self | 73.43% | 73.38% |
| | APPNP_Ensemble | <u>76.26%</u> | <u>75.86%</u> |
| | APPNP_CGI | **76.52%** | **76.07%** |

**Table 3: Performance comparison under full-supervised settings. We use bold font and underline to highlight the best and second best performance under each setting.**

- The causal GCN inference mechanism is effective in the conventional setting.
  - Chronological split

# EXP4: Causal Uncertainty




- The causal uncertainty reveals the correctness of a prediction
- Causal uncertainty is complementary to classification confidence

# Conclusion & Future Work

- Solving the local structure discrepancy issue during GCN inference
- The one-pass model inference might be insufficient
- Incorporating causal intervention is beneficial

- More causal inference techniques, e.g., counterfactual inference
- Eliminating the bias in GCN, e.g., degree bias