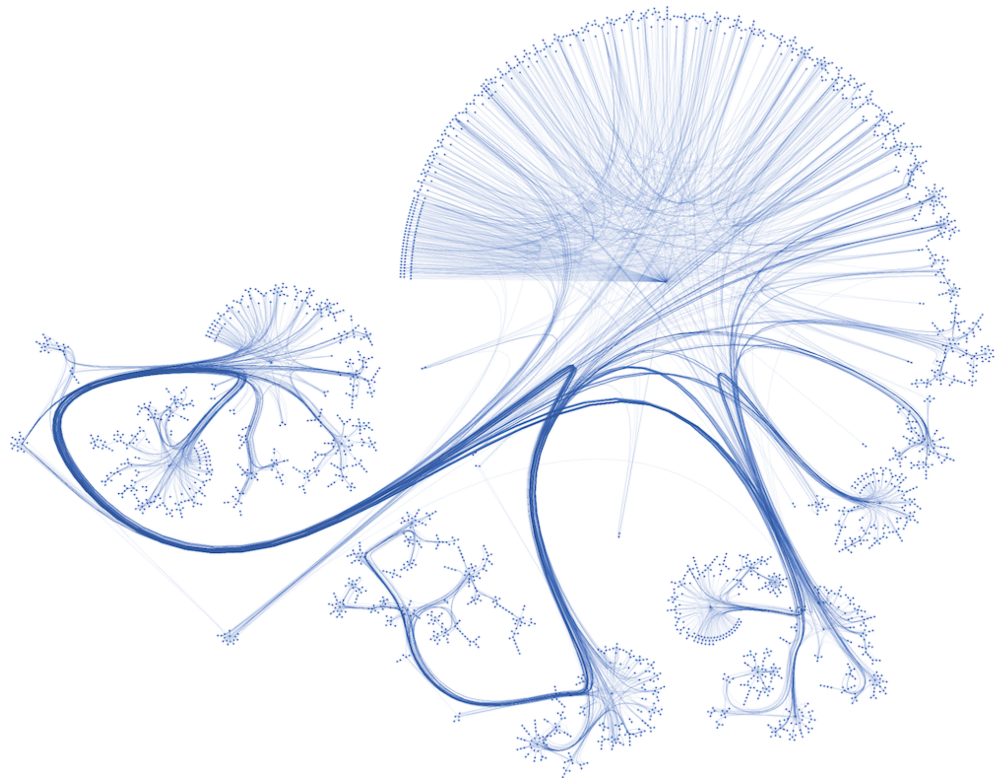
Self Introduction

Jingwei Guo

Email: jingweiguo19@outlook.com

UNIVERSITY OF LIVERPOOL

Xi’an Jiaotong-Liverpool University
西交利物浦大学



Outline

- Basic Information
- Research Experience

Basic Information

University of Liverpool

Dec. 2019 - Jul. 2024

- Ph.D. in Electrical Engineering and Electronic
- Off-based Program in China
- Advisors: Prof. Kaizhu Huang, Prof. Xinping Yi


DKU


SEU

Xi’an Jiaotong-Liverpool University

Sep. 2014 - Aug. 2018

- B.S. in Applied Mathematics (GPA-WES: 3.86/4.00)

Basic Information

Academic Reviewers

- Conference x 5: NeurIPS, KDD, ECML-PKDD, ICPR, ACML
- Journal x 5: TNNLS, NN, PR, IJON, COGN

Awards

- Full Doctoral Scholarship at University of Liverpool (2019)
- Honorable Mention at Interdisciplinary Contest in Modeling (2017)

Research Experience

Graph Learning Against Homophilous Assumptions

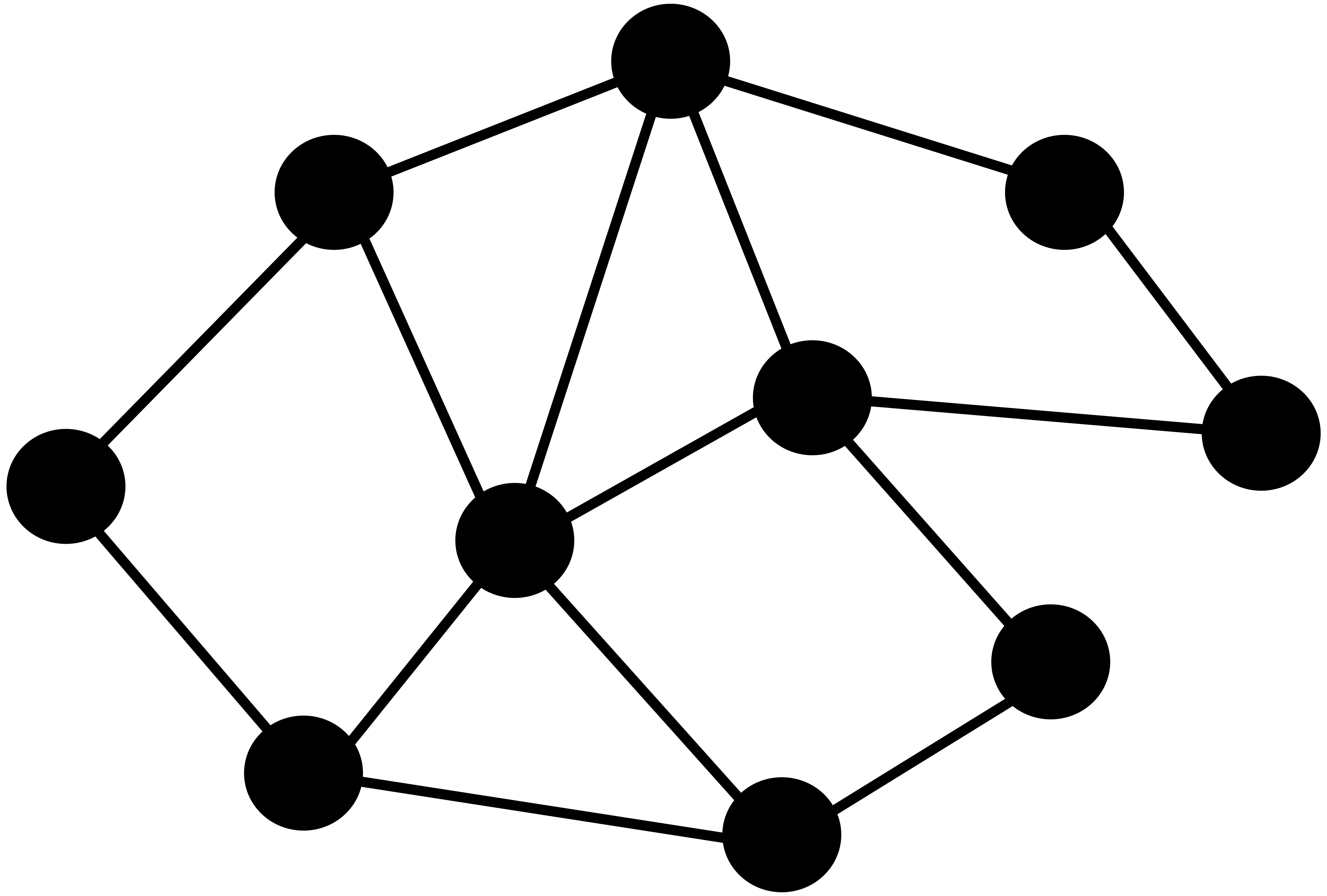
- Primary Contribution
- Output x 4: WWW [CCF A], TPAMI [CCF A], TNNLS [CCF B]

Transfer Learning Under Distribution Shift

- Secondary Contribution
- Output x 2: AAAI [CCF A], A Survey [UE]

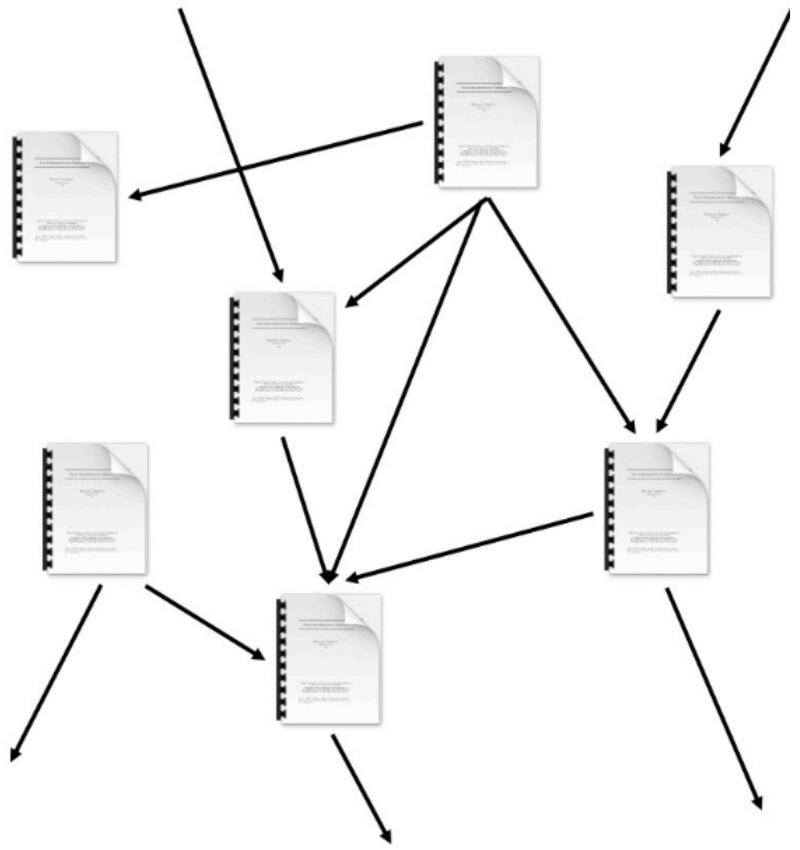
Research Experience — Background

What is Graph?



Graphs are a general language describing and analyzing entities with relations or interactions.

Research Experience — Background

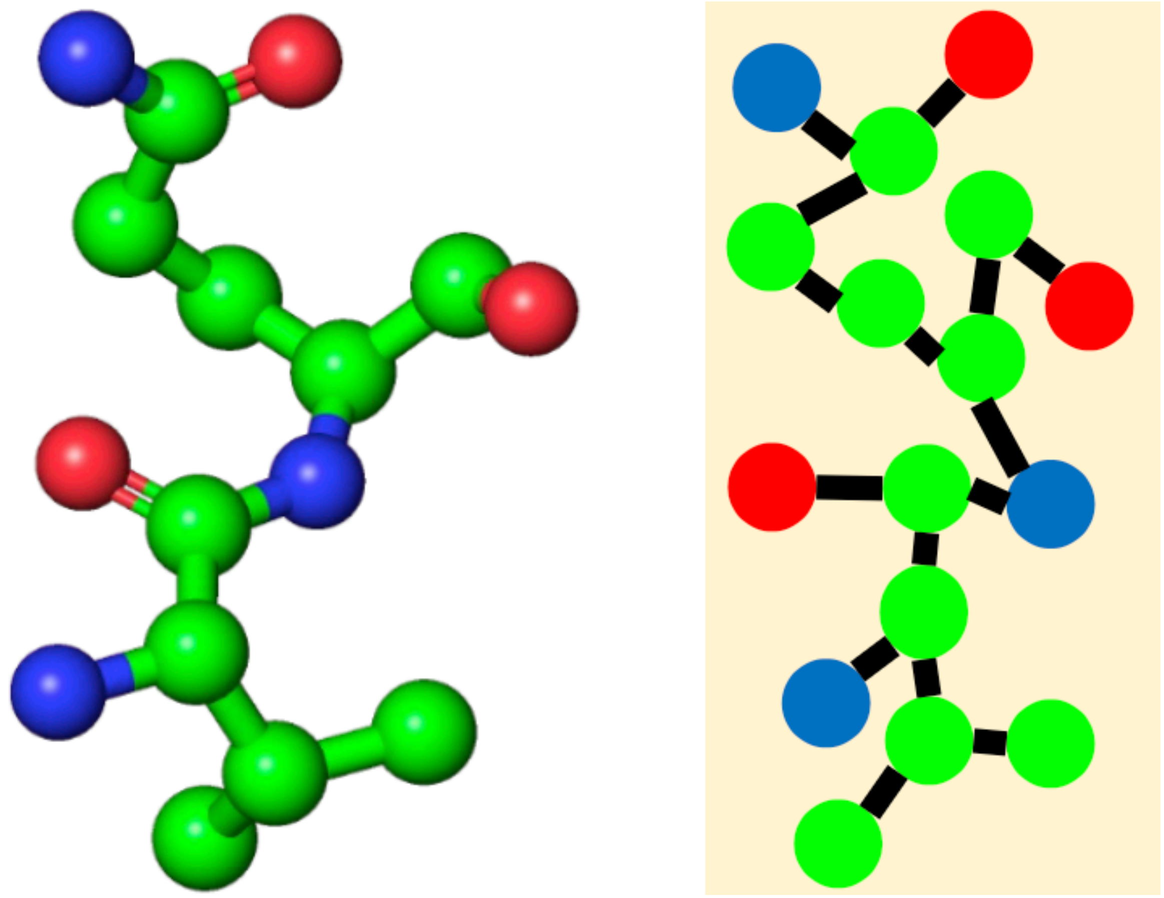


Citation Network

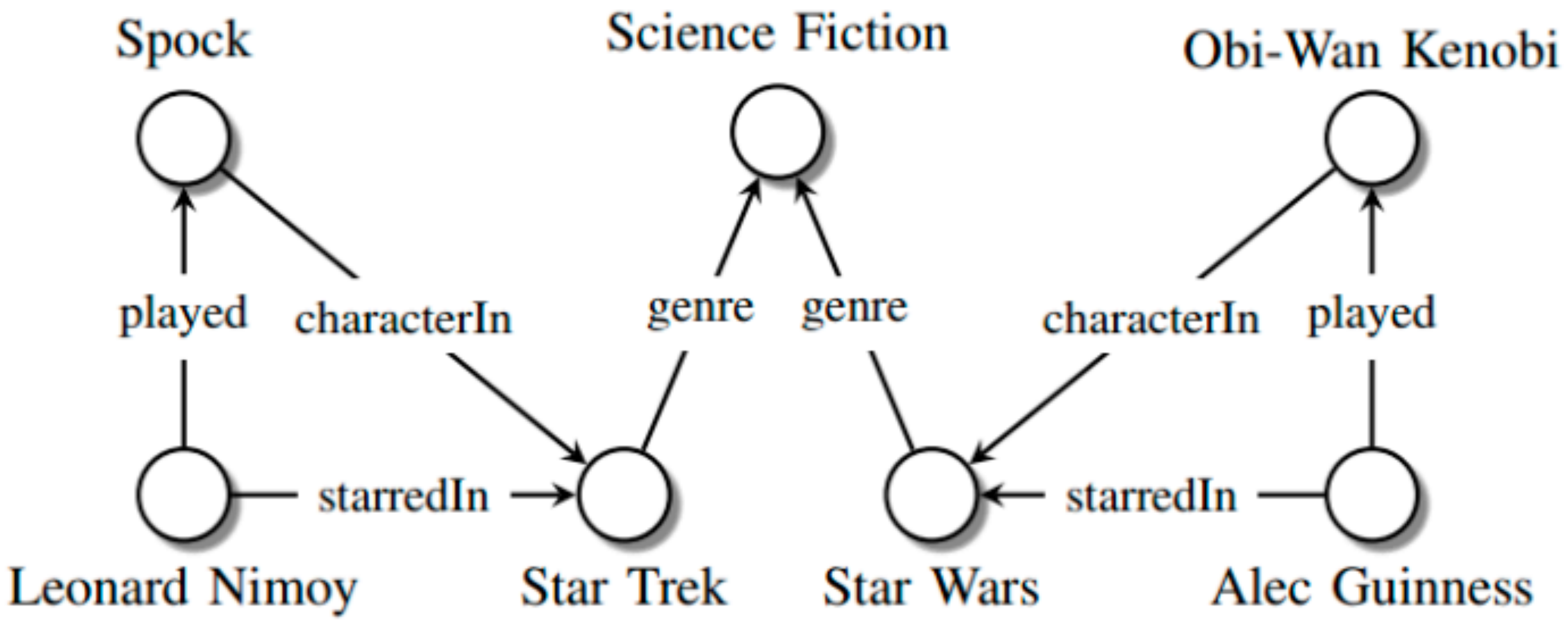
Social Network

Internet (e.g., Webpage)

Brain Network



Molecules



Knowledge Graph

Li, X., et.al. Neural Atoms: Propagating Long-range Interaction in Molecular Graphs through Efficient Communication Channel, *ICLR, 2024*.

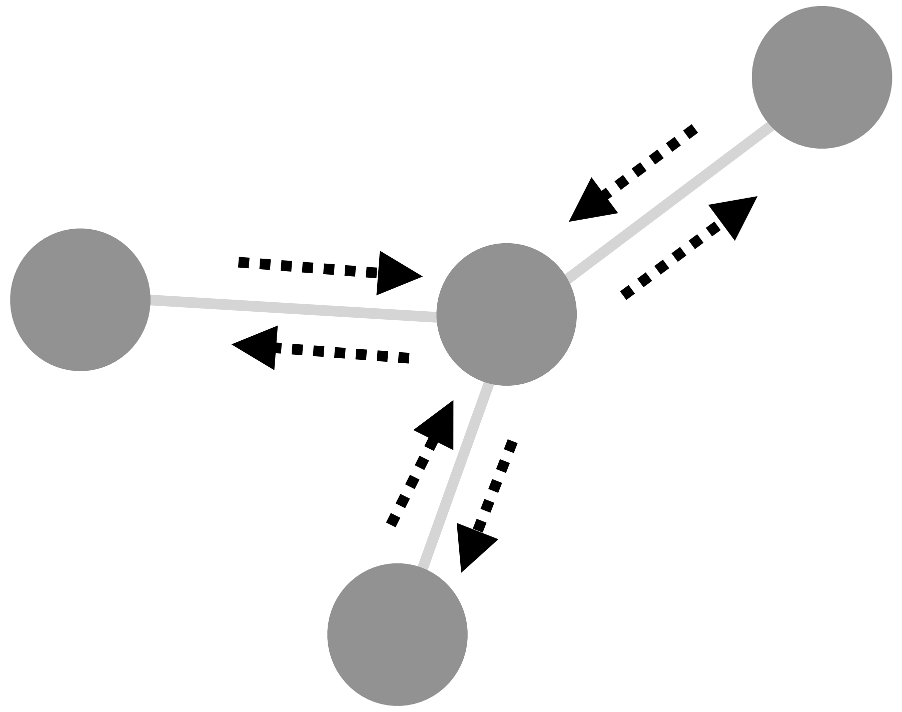
http://cs224w.stanford.edu/

Research Experience — Background

Graph Neural Networks (GNNs)

- Integrate both node features and graph topology via either a message passing framework or a graph filtering operation.
  - Spatial-based Methods

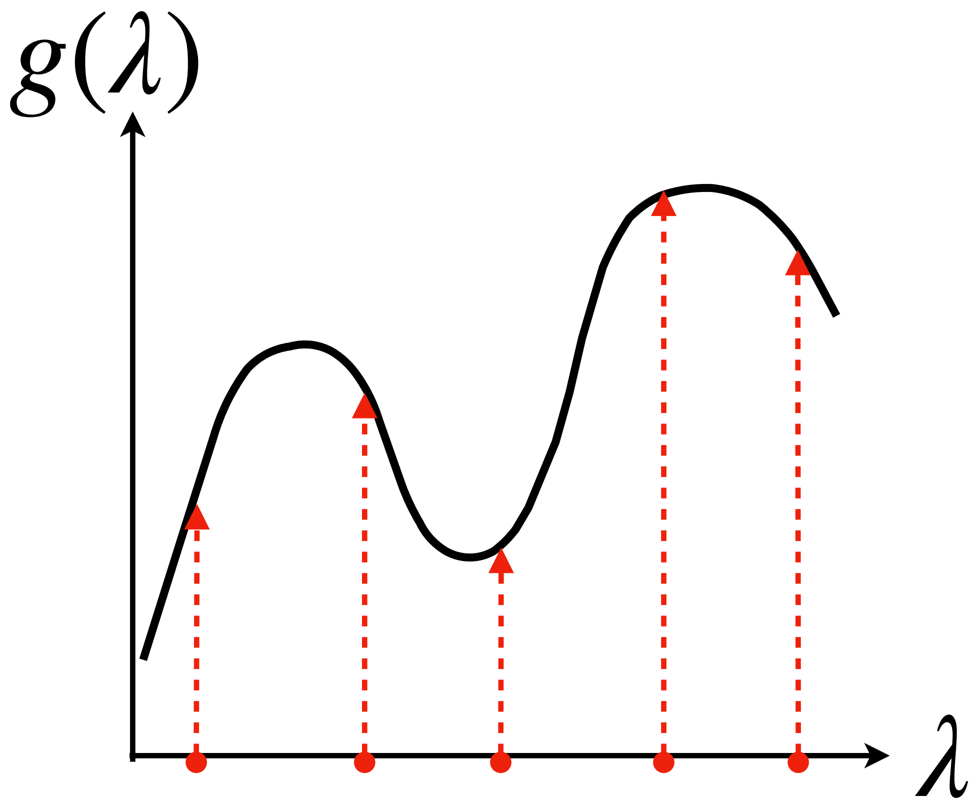
$$\mathbf{z}_i = f_{upd}(\mathbf{x}_i, f_{agg}(\{\mathbf{x}_j \,|\, \forall v_j \in N_i\}))$$



Message Passing

  - Spectral-based Methods

$$\mathbf{Z} = \mathbf{U}g(\mathbf{\Lambda})\mathbf{U}^T\mathbf{X}$$



Graph Filtering

Research Experience — Background

Graph Neural Networks (GNNs)

- Can be interpreted as different solutions to the same graph denoising problem:

$$\arg\min_{\mathbf{Z}} \alpha||\mathbf{X}-\mathbf{Z}||_2^2 + (1-\alpha)\mathrm{tr}(\mathbf{Z}^T\hat{\mathbf{L}}\mathbf{Z})$$

keep close to the original features

smooth node features across the graph

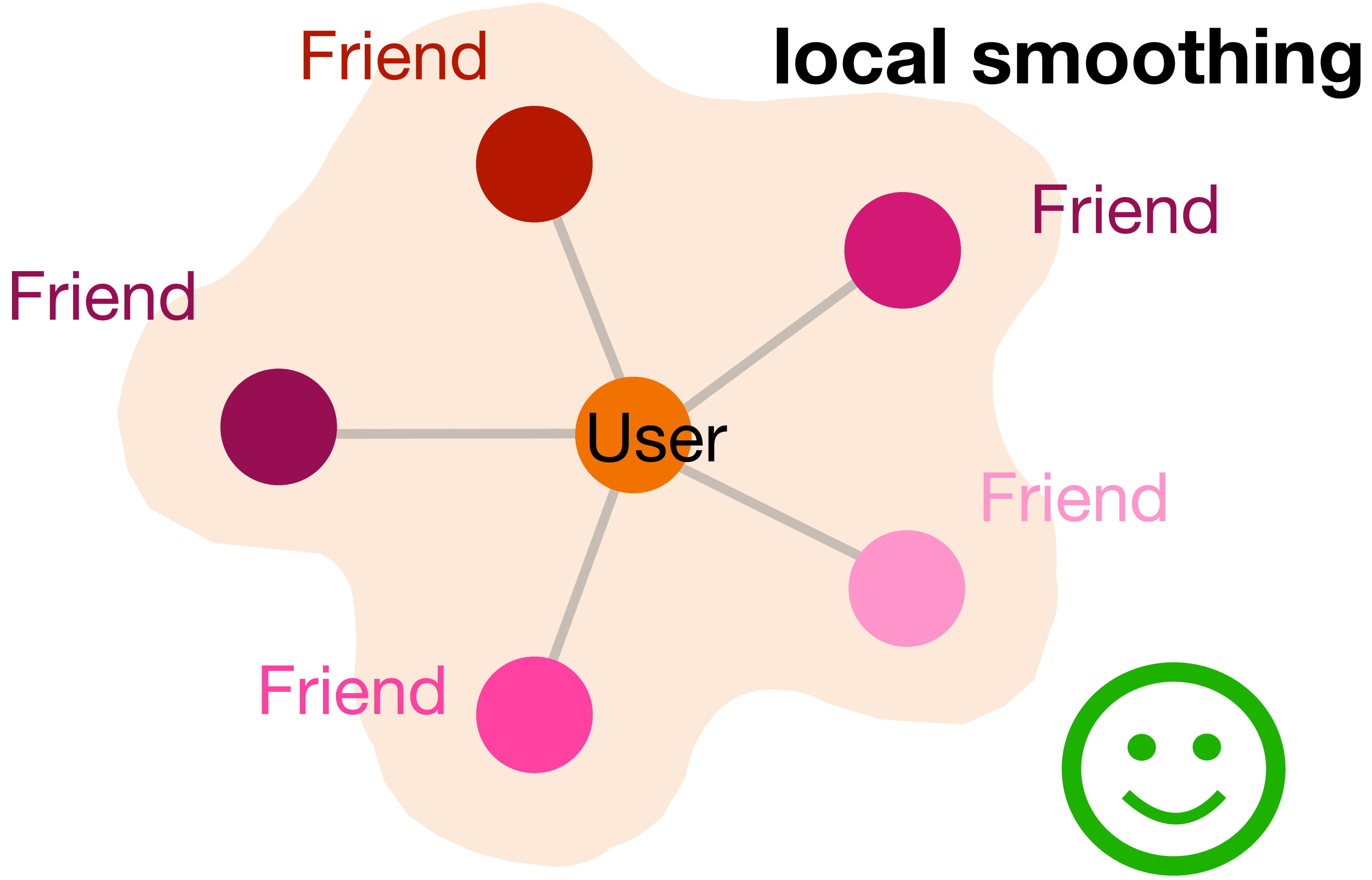
Ma, Y., et.al. A unified view on graph neural networks as graph signal denoising. *CIKM, 2021.* Zhu, M., et.al. Interpreting and unifying graph neural networks with an optimization framework. *WWW, 2021.*

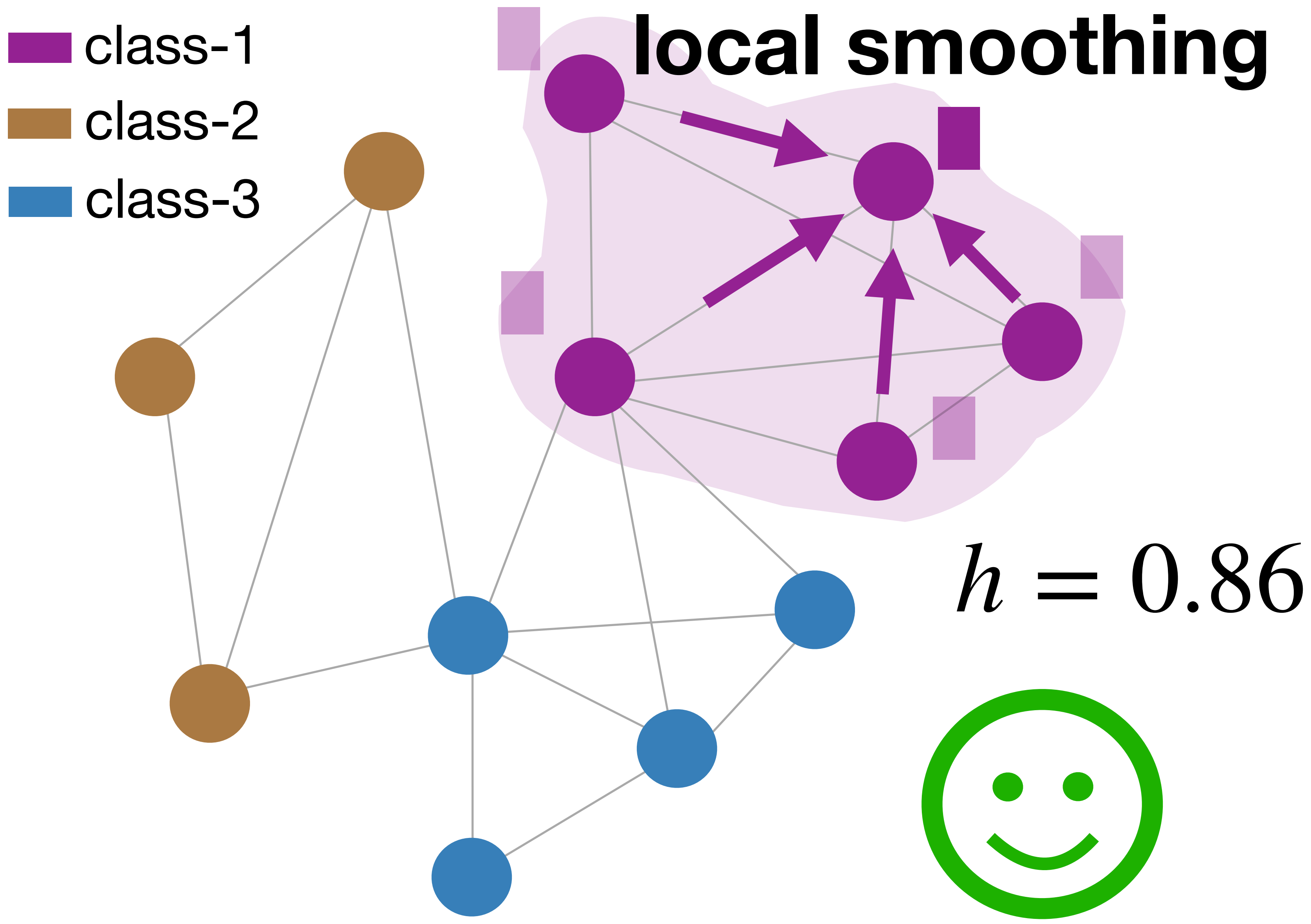
Research Experience — Motivation

Motivation & Challenges

- Most GNNs assume homogenous node interactions and employ neighborhood smoothing.

  - ▶ Homogenous Node Relationships
  - ▶ Homophilic Linking Patterns

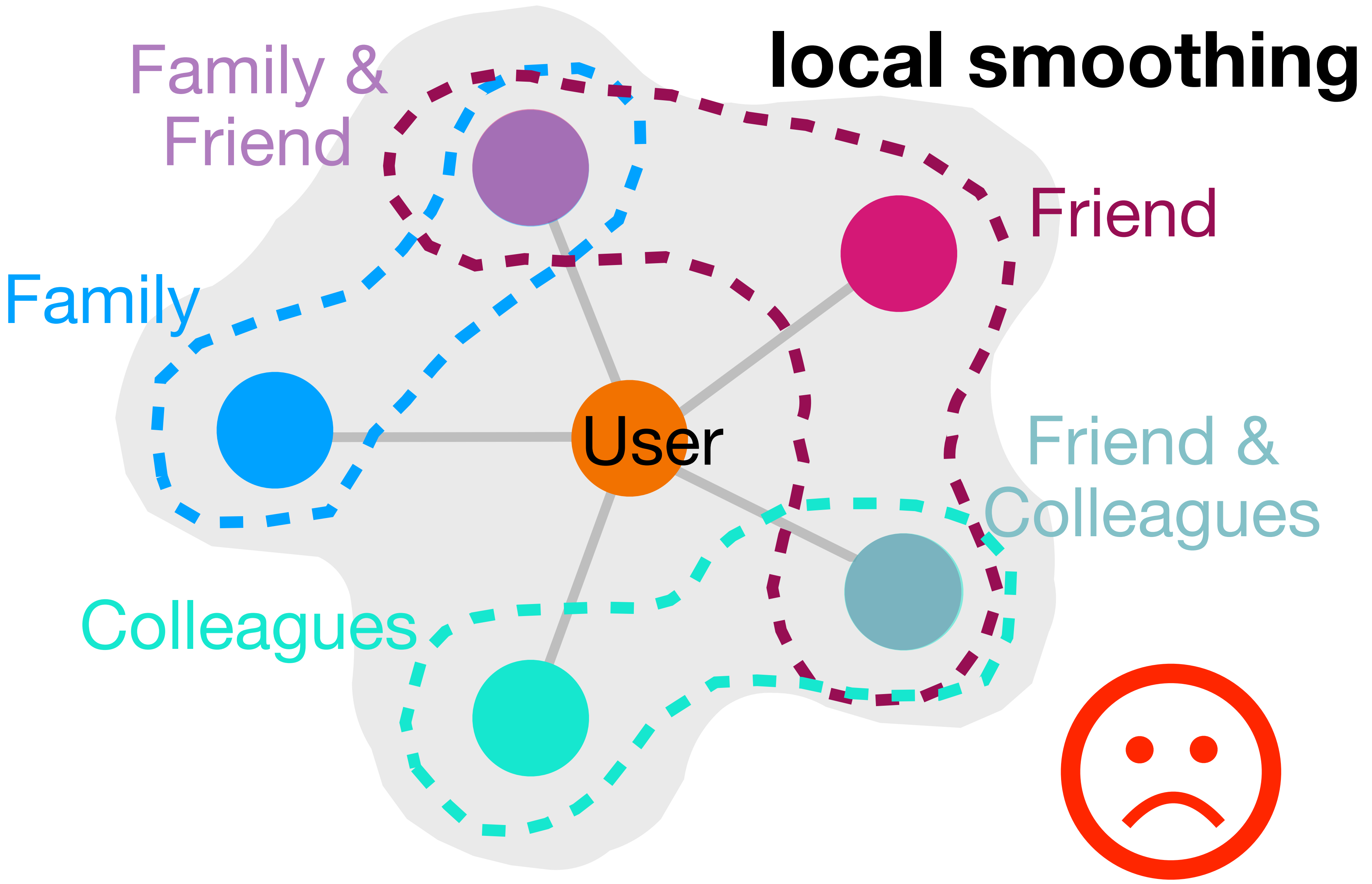


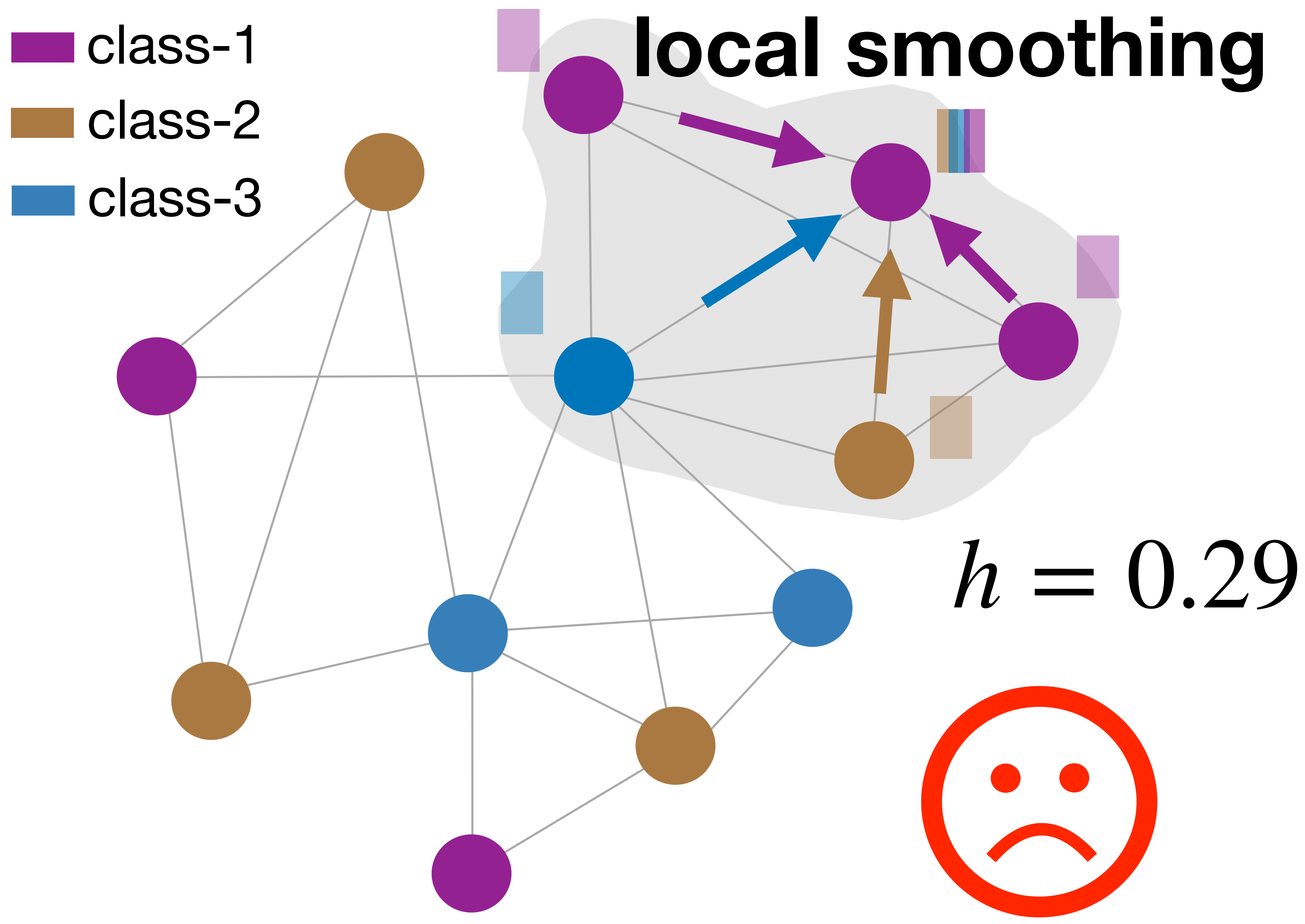


Research Experience — Challenges

Motivation & Challenges

- Most GNNs assume homogenous node interactions and employ neighborhood smoothing.
  - Entangled Node Relationships
  - Heterophilic Linking Patterns





Research Experience — Solutions x 4

Graph Learning Against Homophilous Assumptions

- **Jingwei Guo**, et.al. Learning Disentangled Graph Convolutional Networks Locally and Globally. *TNNLS 2022 [CCF B, IF 10.2].*

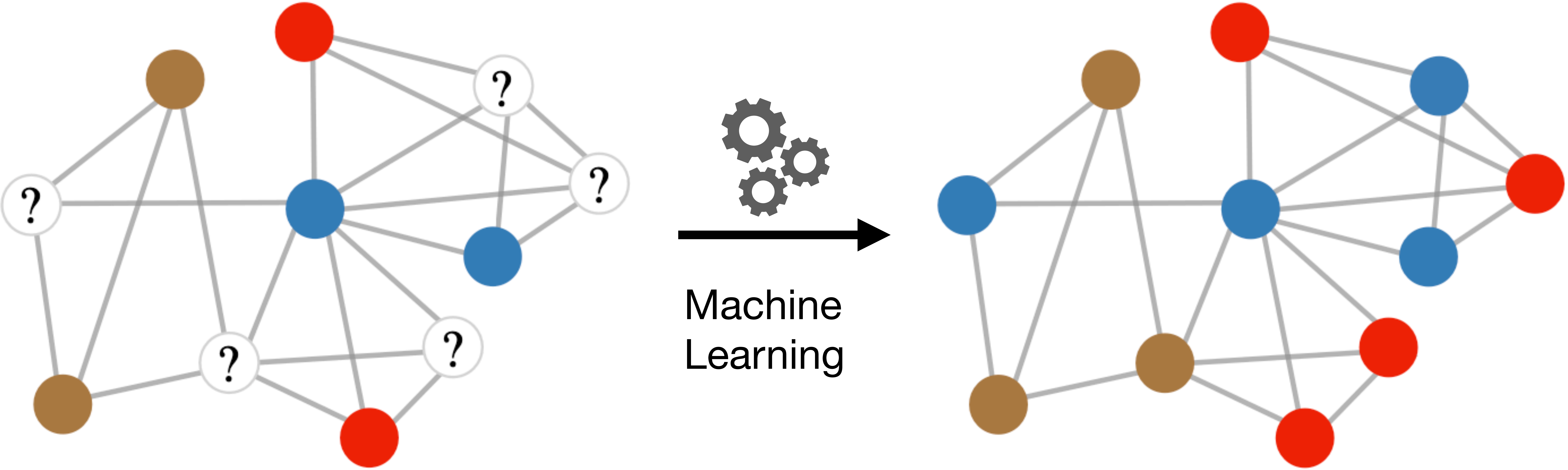
- **Jingwei Guo**, et.al. ES-GNN: Generalizing Graph Neural Networks Beyond Homophily with Edge Splitting. *TPAMI 2024 [CCF A, IF 20.8].*

- **Jingwei Guo**, et.al. Graph Neural Networks with Diverse Spectral Filtering. *WWW 2023 [CCF A]*.

- **Jingwei Guo**, et.al. Rethinking Spectral Graph Neural Networks with Spatially Adaptive Filtering. *TNNLS (Under Review) 2024 [CCF B, IF 10.2]*.

Research Experience — Node Classification Tasks

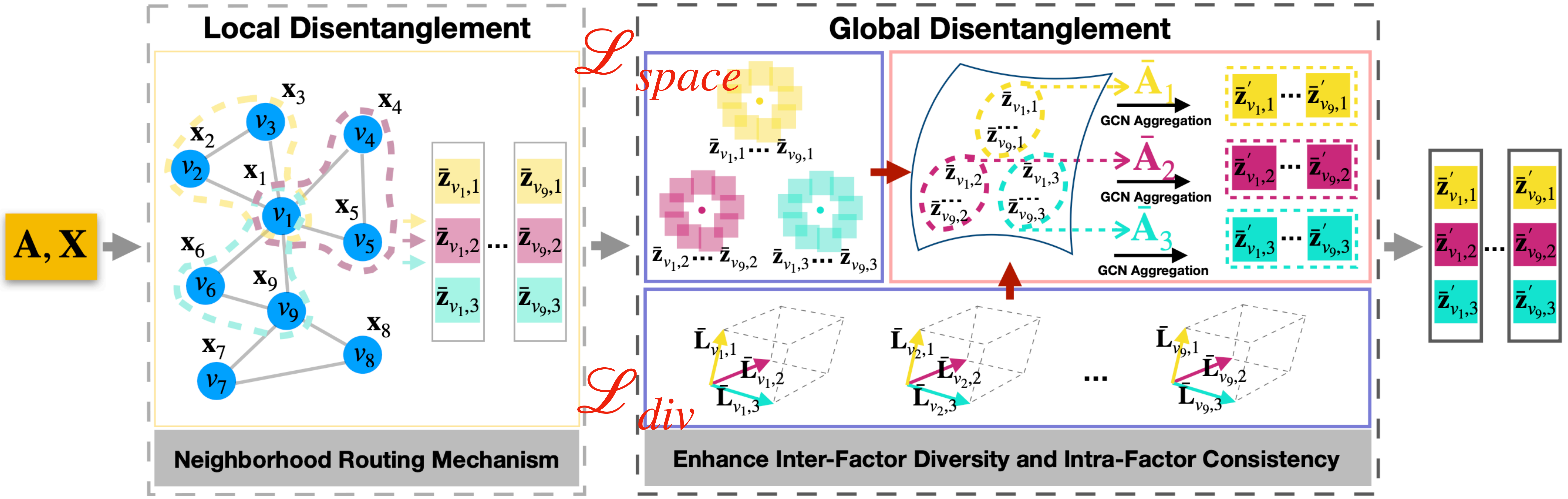
🟦 Class-1 🟥 Class-2 🟫 Class-3



Machine Learning

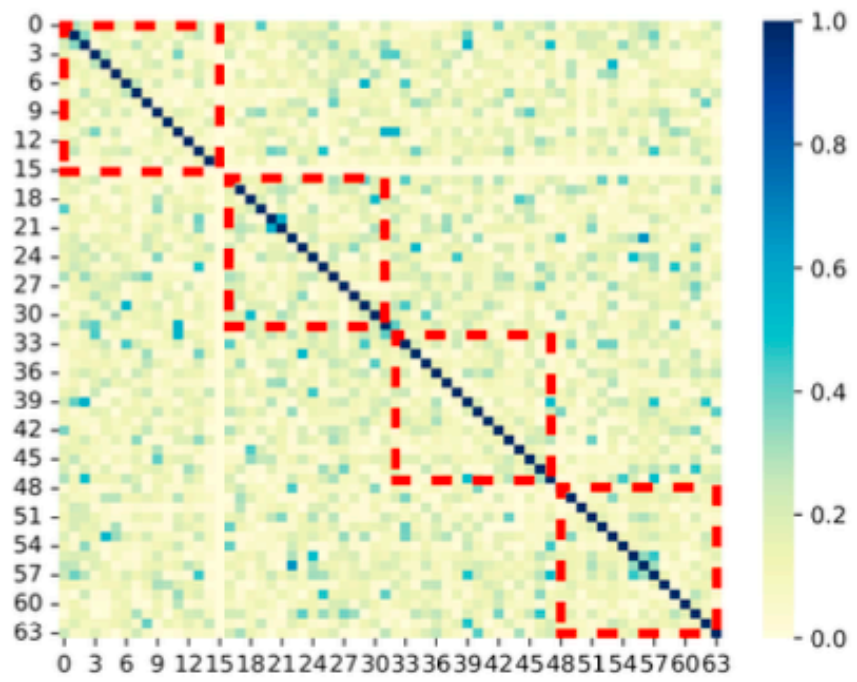
Research Experience — Solution 1

Local-Global Disentanglement

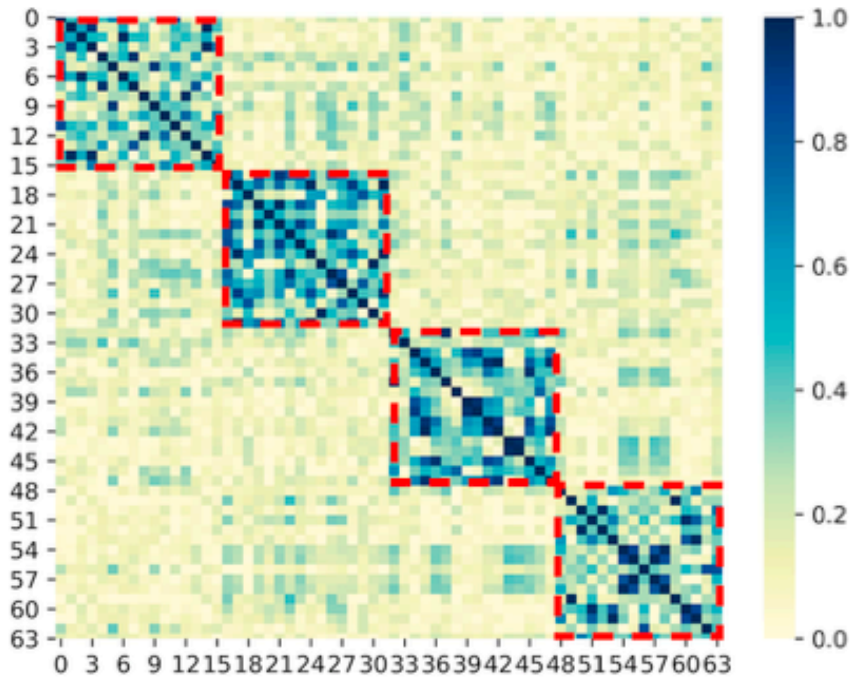


- latent space regularization with GMM
- global message passing on latent graphs

Features Correlation



Baseline

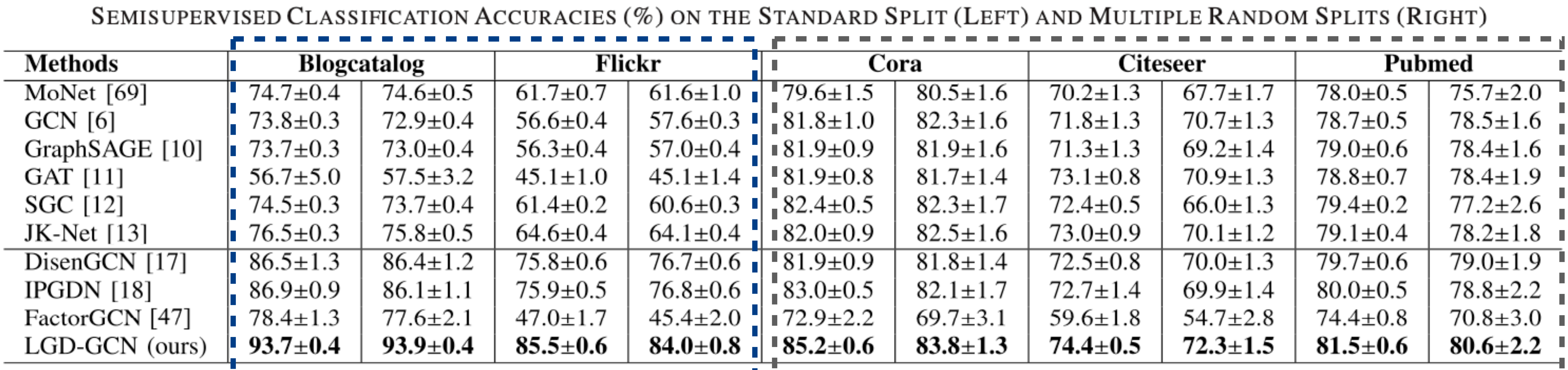


Ours

Jingwei Guo, et.al. Learning Disentangled Graph Convolutional Networks Locally and Globally. *TNNLS, 2022*.

Research Experience — Solution 1

Local-Global Disentanglement



SEMISUPERVISED CLASSIFICATION ACCURACIES (%) ON THE STANDARD SPLIT (LEFT) AND MULTIPLE RANDOM SPLITS (RIGHT)

Methods	**Blogcatalog**		**Flickr**		**Cora**		**Citeseer**		**Pubmed**	
MoNet [69]	74.7±0.4	74.6±0.5	61.7±0.7	61.6±1.0	79.6±1.5	80.5±1.6	70.2±1.3	67.7±1.7	78.0±0.5	75.7±2.0
GCN [6]	73.8±0.3	72.9±0.4	56.6±0.4	57.6±0.3	81.8±1.0	82.3±1.6	71.8±1.3	70.7±1.3	78.7±0.5	78.5±1.6
GraphSAGE [10]	73.7±0.3	73.0±0.4	56.3±0.4	57.0±0.4	81.9±0.9	81.9±1.6	71.3±1.3	69.2±1.4	79.0±0.6	78.4±1.6
GAT [11]	56.7±5.0	57.5±3.2	45.1±1.0	45.1±1.4	81.9±0.8	81.7±1.4	73.1±0.8	70.9±1.3	78.8±0.7	78.4±1.9
SGC [12]	74.5±0.3	73.7±0.4	61.4±0.2	60.6±0.3	82.4±0.5	82.3±1.7	72.4±0.5	66.0±1.3	79.4±0.2	77.2±2.6
JK-Net [13]	76.5±0.3	75.8±0.5	64.6±0.4	64.1±0.4	82.0±0.9	82.5±1.6	73.0±0.9	70.1±1.2	79.1±0.4	78.2±1.8
DisenGCN [17]	86.5±1.3	86.4±1.2	75.8±0.6	76.7±0.6	81.9±0.9	81.8±1.4	72.5±0.8	70.0±1.3	79.7±0.6	79.0±1.9
IPGDN [18]	86.9±0.9	86.1±1.1	75.9±0.5	76.8±0.6	83.0±0.5	82.1±1.7	72.7±1.4	69.9±1.4	80.0±0.5	78.8±2.2
FactorGCN [47]	78.4±1.3	77.6±2.1	47.0±1.7	45.4±2.0	72.9±2.2	69.7±3.1	59.6±1.8	54.7±2.8	74.4±0.8	70.8±3.0
LGD-GCN (ours)	**93.7±0.4**	**93.9±0.4**	**85.5±0.6**	**84.0±0.8**	**85.2±0.6**	**83.8±1.3**	**74.4±0.5**	**72.3±1.5**	**81.5±0.6**	**80.6±2.2**

Social Network

Citation Network

Jingwei Guo, et.al. Learning Disentangled Graph Convolutional Networks Locally and Globally. *TNNLS, 2022*.

Research Experience — Solution 2

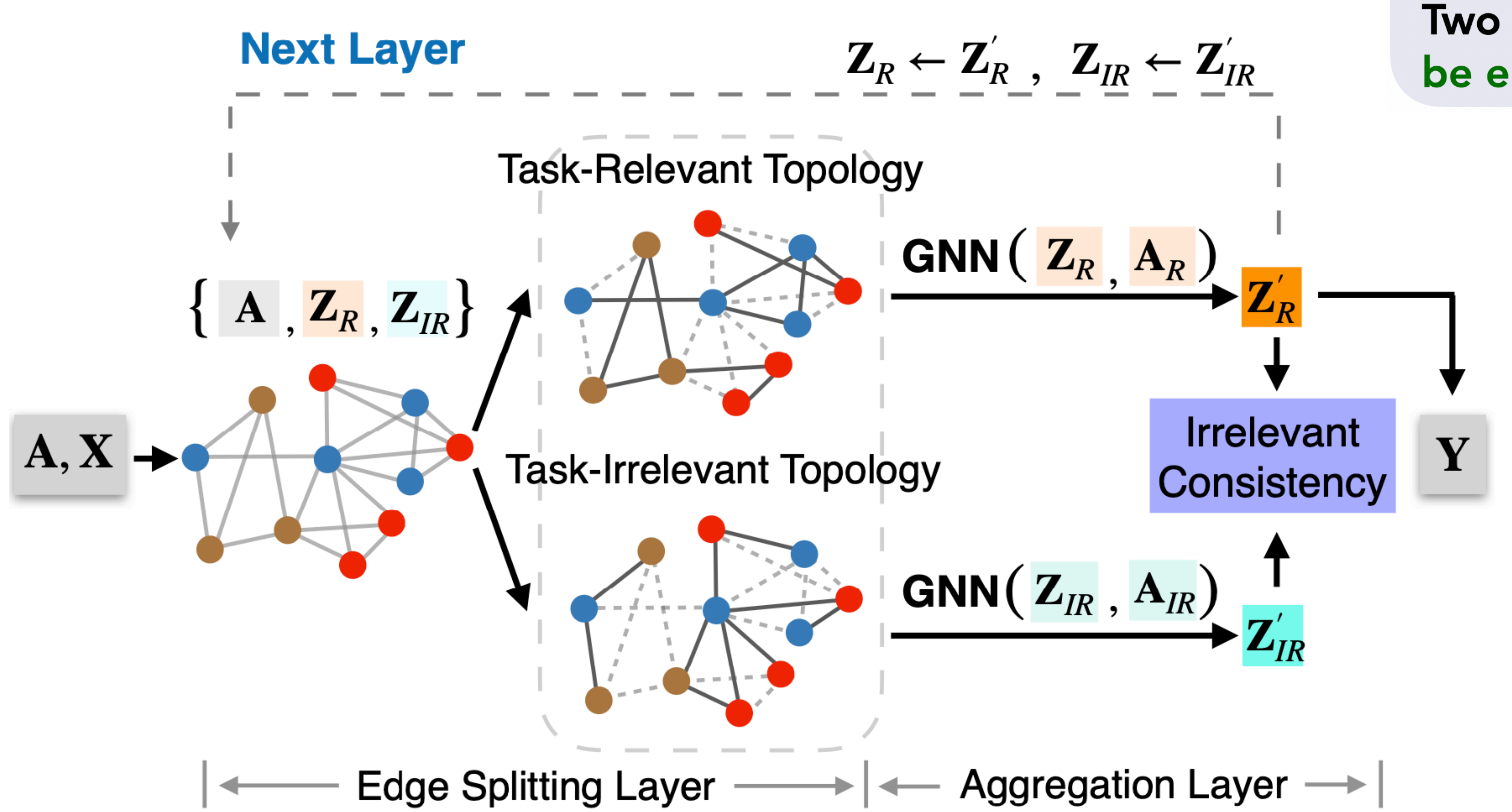
Edge Splitting GNN

Conventional Smoothness Assumption

Two connected nodes mostly share task-beneficial similarity.

Disentangled Smoothness Assumption (Ours)

Two connected nodes share similarity in some features, which could be either relevant or irrelevant (even harmful) to learning tasks.

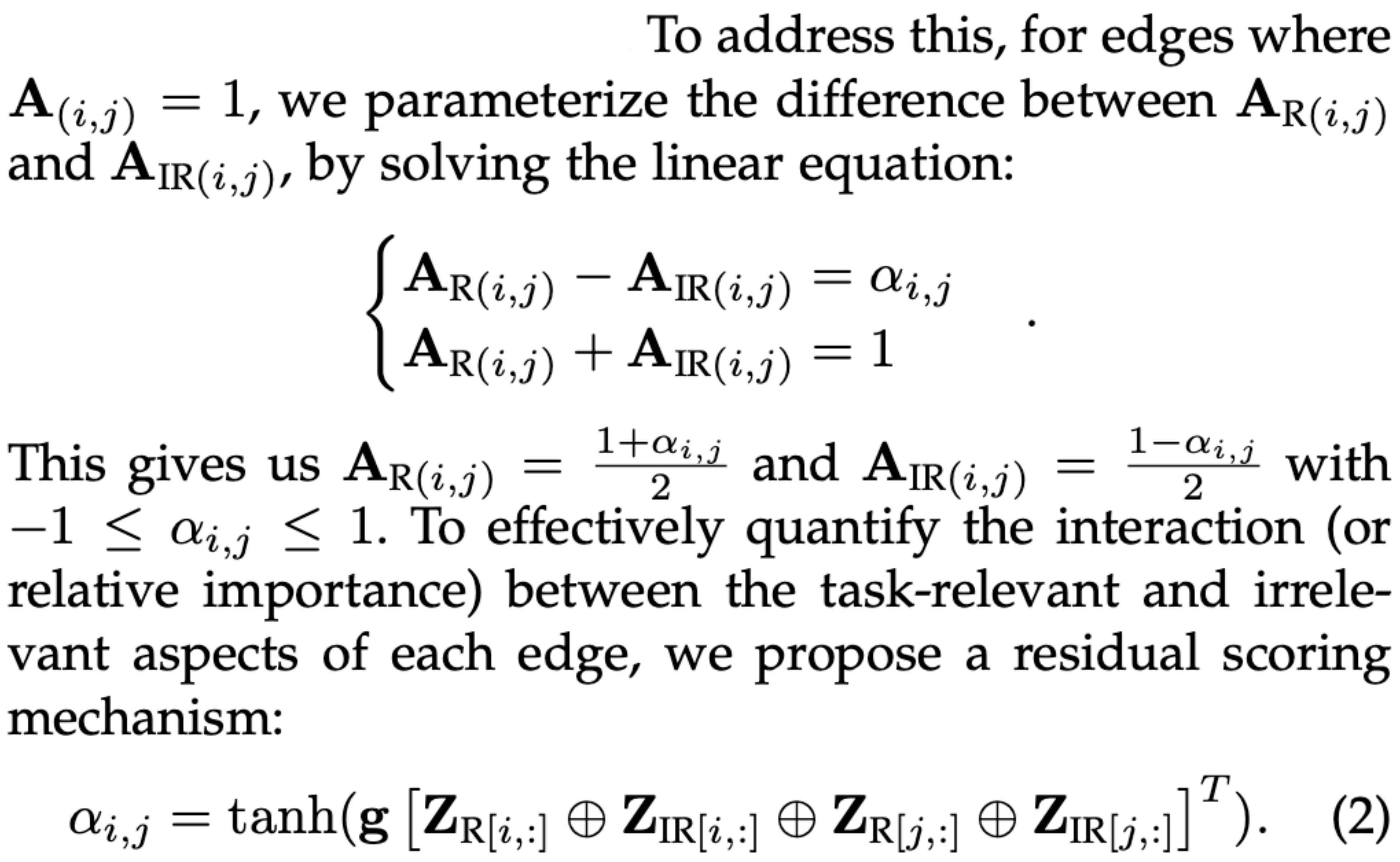


ES-GNN addresses heterophilic graphs by partitioning network topology & disentangling node features.

Jingwei Guo, et.al. ES-GNN: Generalizing Graph Neural Networks Beyond Homophily with Edge Splitting. *TPAMI (Minor Revision), 2024*.

Research Experience — Solution 2

Edge Splitting GNN

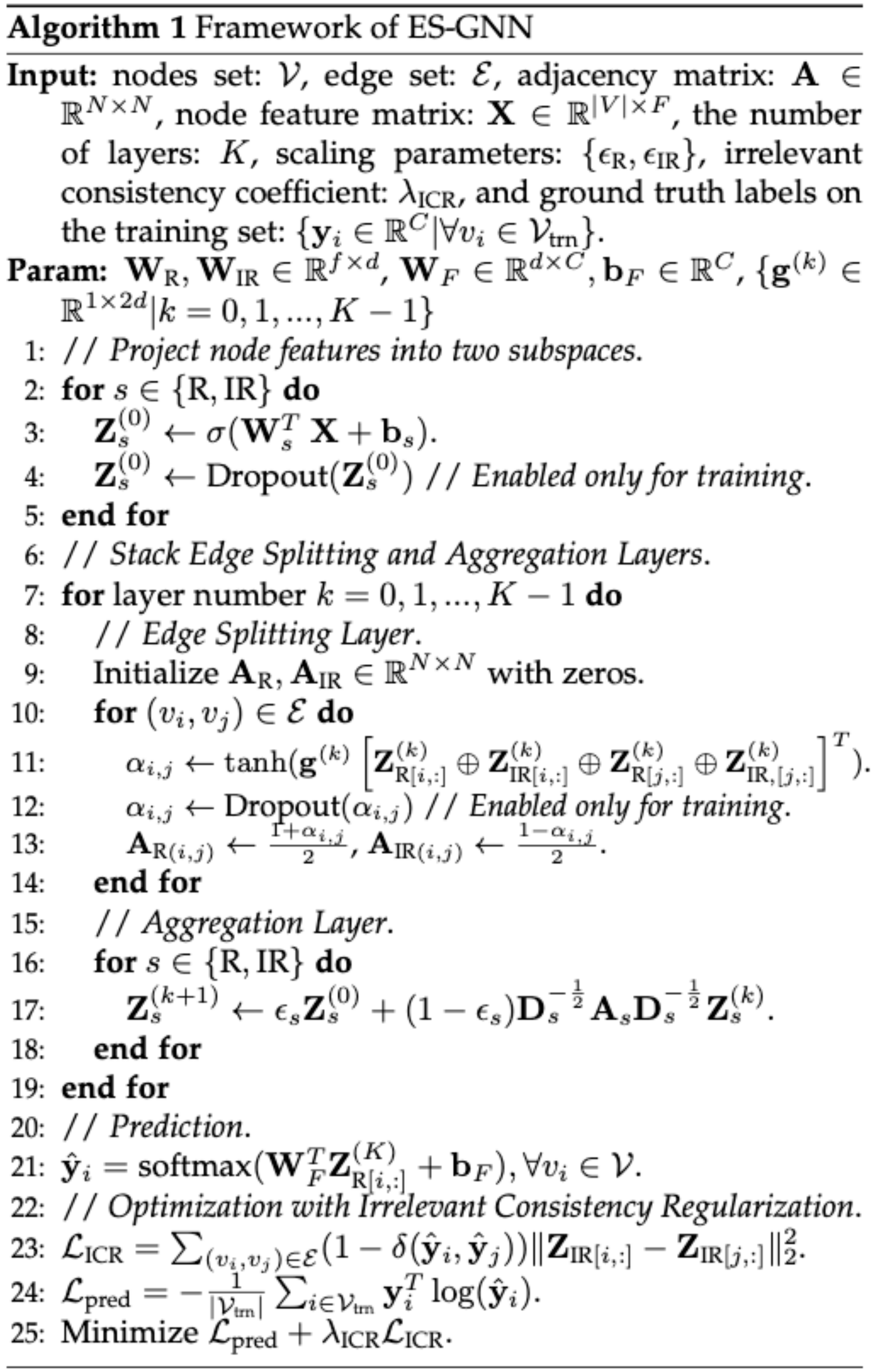


To address this, for edges where $\mathbf{A}_{(i,j)} = 1$, we parameterize the difference between $\mathbf{A}_{\mathrm{R}(i,j)}$ and $\mathbf{A}_{\mathrm{IR}(i,j)}$, by solving the linear equation:

$$\begin{cases} \mathbf{A}_{\mathrm{R}(i,j)} - \mathbf{A}_{\mathrm{IR}(i,j)} = \alpha_{i,j} \\ \mathbf{A}_{\mathrm{R}(i,j)} + \mathbf{A}_{\mathrm{IR}(i,j)} = 1 \end{cases}.$$

This gives us $\mathbf{A}_{\mathrm{R}(i,j)} = \frac{1+\alpha_{i,j}}{2}$ and $\mathbf{A}_{\mathrm{IR}(i,j)} = \frac{1-\alpha_{i,j}}{2}$ with $-1 \leq \alpha_{i,j} \leq 1$. To effectively quantify the interaction (or relative importance) between the task-relevant and irrelevant aspects of each edge, we propose a residual scoring mechanism:

$$\alpha_{i,j} = \tanh(\mathbf{g}\left[\mathbf{Z}_{\mathrm{R}[i,:]} \oplus \mathbf{Z}_{\mathrm{IR}[i,:]} \oplus \mathbf{Z}_{\mathrm{R}[j,:]} \oplus \mathbf{Z}_{\mathrm{IR}[j,:]}\right]^T). \quad (2)$$



Algorithm 1 Framework of ES-GNN

```
Input: nodes set: $\mathcal{V}$, edge set: $\mathcal{E}$, adjacency matrix: $\mathbf{A} \in \mathbb{R}^{N\times N}$, node feature matrix: $\mathbf{X} \in \mathbb{R}^{|V|\times F}$, the number of layers: $K$, scaling parameters: $\{\epsilon_{\mathrm{R}}, \epsilon_{\mathrm{IR}}\}$, irrelevant consistency coefficient: $\lambda_{\mathrm{ICR}}$, and ground truth labels on the training set: $\{\mathbf{y}_i \in \mathbb{R}^C | \forall v_i \in \mathcal{V}_{\mathrm{trn}}\}$.
Param: $\mathbf{W}_{\mathrm{R}}, \mathbf{W}_{\mathrm{IR}} \in \mathbb{R}^{f\times d}, \mathbf{W}_F \in \mathbb{R}^{d\times C}, \mathbf{b}_F \in \mathbb{R}^C, \{\mathbf{g}^{(k)} \in \mathbb{R}^{1\times 2d} | k = 0, 1, ..., K-1\}$
// Project node features into two subspaces.
for $s \in \{\mathrm{R}, \mathrm{IR}\}$ do
    $\mathbf{Z}_s^{(0)} \leftarrow \sigma(\mathbf{W}_s^T \mathbf{X} + \mathbf{b}_s)$.
    $\mathbf{Z}_s^{(0)} \leftarrow \mathrm{Dropout}(\mathbf{Z}_s^{(0)})$ // Enabled only for training.
end for
// Stack Edge Splitting and Aggregation Layers.
for layer number $k = 0, 1, ..., K-1$ do
    // Edge Splitting Layer.
    Initialize $\mathbf{A}_{\mathrm{R}}, \mathbf{A}_{\mathrm{IR}} \in \mathbb{R}^{N\times N}$ with zeros.
    for $(v_i, v_j) \in \mathcal{E}$ do
        $\alpha_{i,j} \leftarrow \tanh(\mathbf{g}^{(k)}\left[\mathbf{Z}^{(k)}_{\mathrm{R}[i,:]} \oplus \mathbf{Z}^{(k)}_{\mathrm{IR}[i,:]} \oplus \mathbf{Z}^{(k)}_{\mathrm{R}[j,:]} \oplus \mathbf{Z}^{(k)}_{\mathrm{IR},[j,:]}\right]^T)$.
        $\alpha_{i,j} \leftarrow \mathrm{Dropout}(\alpha_{i,j})$ // Enabled only for training.
        $\mathbf{A}_{\mathrm{R}(i,j)} \leftarrow \frac{1+\alpha_{i,j}}{2}, \mathbf{A}_{\mathrm{IR}(i,j)} \leftarrow \frac{1-\alpha_{i,j}}{2}$.
    end for
    // Aggregation Layer.
    for $s \in \{\mathrm{R}, \mathrm{IR}\}$ do
        $\mathbf{Z}_s^{(k+1)} \leftarrow \epsilon_s \mathbf{Z}_s^{(0)} + (1-\epsilon_s)\mathbf{D}_s^{-\frac{1}{2}} \mathbf{A}_s \mathbf{D}_s^{-\frac{1}{2}} \mathbf{Z}_s^{(k)}$.
    end for
end for
// Prediction.
$\hat{\mathbf{y}}_i = \mathrm{softmax}(\mathbf{W}_F^T \mathbf{Z}^{(K)}_{\mathrm{R}[i,:]} + \mathbf{b}_F), \forall v_i \in \mathcal{V}$.
// Optimization with Irrelevant Consistency Regularization.
$\mathcal{L}_{\mathrm{ICR}} = \sum_{(v_i,v_j)\in\mathcal{E}} (1 - \delta(\hat{\mathbf{y}}_i, \hat{\mathbf{y}}_j)) \|\mathbf{Z}_{\mathrm{IR}[i,:]} - \mathbf{Z}_{\mathrm{IR}[j,:]}\|_2^2$.
$\mathcal{L}_{\mathrm{pred}} = -\frac{1}{|\mathcal{V}_{\mathrm{trn}}|} \sum_{i\in\mathcal{V}_{\mathrm{trn}}} \mathbf{y}_i^T \log(\hat{\mathbf{y}}_i)$.
Minimize $\mathcal{L}_{\mathrm{pred}} + \lambda_{\mathrm{ICR}} \mathcal{L}_{\mathrm{ICR}}$.
```

Jingwei Guo, et.al. ES-GNN: Generalizing Graph Neural Networks Beyond Homophily with Edge Splitting. *TPAMI (Minor Revision), 2024*.

Research Experience — Solution 2

Edge Splitting GNN

- Conventional GNNs

Graph Denoising Problem

$$\arg\min_{\mathbf{Z}} \quad \|\mathbf{Z}-\mathbf{X}\|_2^2 + \xi \cdot tr(\mathbf{Z}^T\mathbf{L}\mathbf{Z})$$

Our Analysis ↓ $L = L_R + L_{IR}$

$$\arg\min_{\mathbf{Z}} \quad \|\mathbf{Z}-\mathbf{X}\|_2^2 + \xi \cdot tr(\mathbf{Z}^T\mathbf{L}_R\mathbf{Z}) \boxed{+ \xi \cdot tr(\mathbf{Z}^T\mathbf{L}_{IR}\mathbf{Z})}$$

Possible classification-harmful information could be preserved in $\mathbf{Z}$

- Our ES-GNN

Disentangled Graph Denoising Problem

$$\arg\min_{\boxed{\mathbf{Z}_R, \mathbf{Z}_{IR}}} \quad \|\mathbf{Z}_R-\mathbf{X}_{IR}\|_2^2 + \|\mathbf{Z}_{IR}-\mathbf{X}_{IR}\|_2^2$$
$$+ \xi \cdot tr(\mathbf{Z}_R^T\mathbf{L}_R\mathbf{Z}_R) + \xi \cdot tr(\mathbf{Z}_{IR}^T\mathbf{L}_{IR}\mathbf{Z}_{IR})$$
$$\text{where} \quad \mathbf{L}_R = \mathbf{D}_R - \mathbf{A}_R, \mathbf{L}_{IR} = \mathbf{D}_{IR} - \mathbf{A}_{IR}$$
$$\text{s.t.} \quad \mathbf{A}_R + \mathbf{A}_{IR} = \mathbf{A}$$
$$\mathbf{A}_{R(i,j)}, \mathbf{A}_{IR(i,j)} \in [0,1].$$

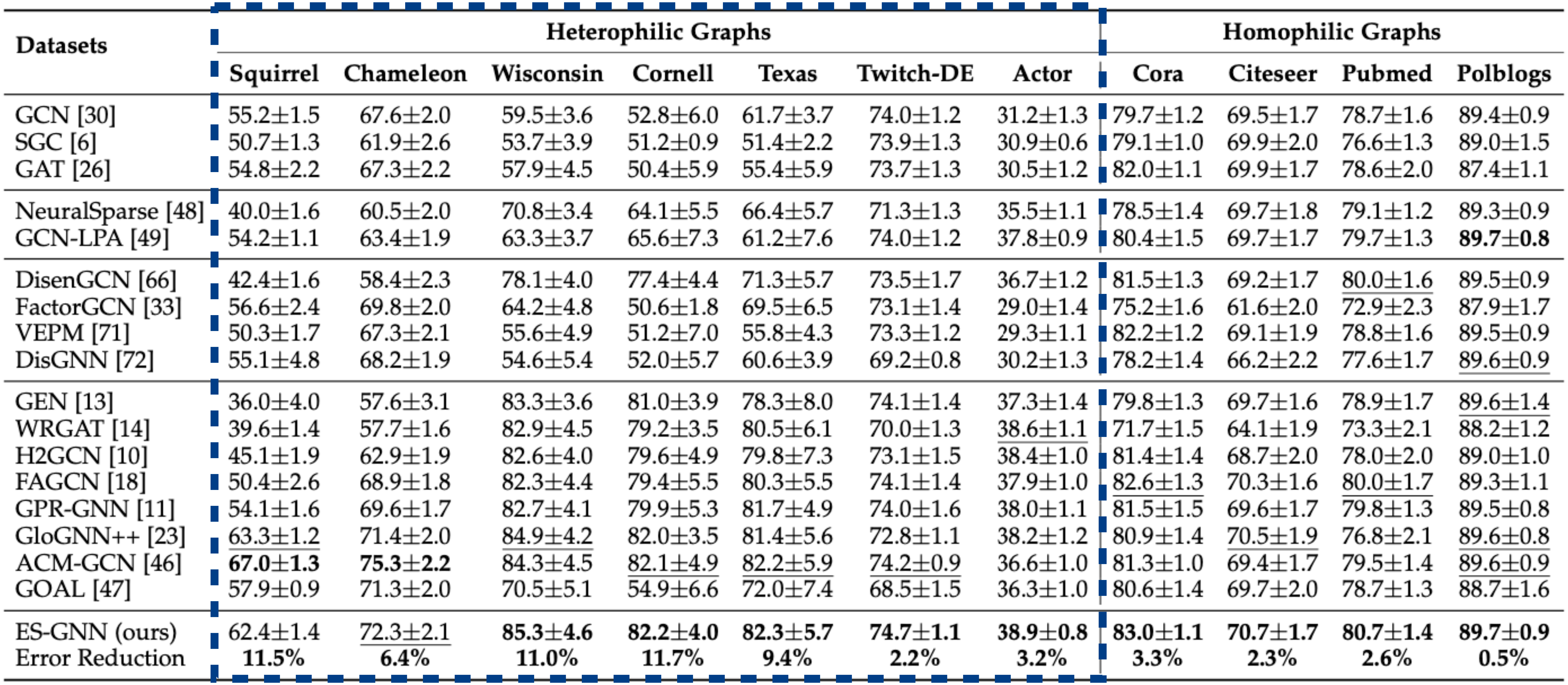
Possible classification-harmful information can be excluded from $\mathbf{Z}_R$ & disentangled in $\mathbf{Z}_{IR}$

Jingwei Guo, et.al. ES-GNN: Generalizing Graph Neural Networks Beyond Homophily with Edge Splitting. *TPAMI (Minor Revision), 2024*.

Research Experience — Solution 2

Edge Splitting GNN

Node classification accuracies (%) over 100 runs. Error Reduction gives the average improvement of ES-GNN upon baselines w/o Basic GNNs.

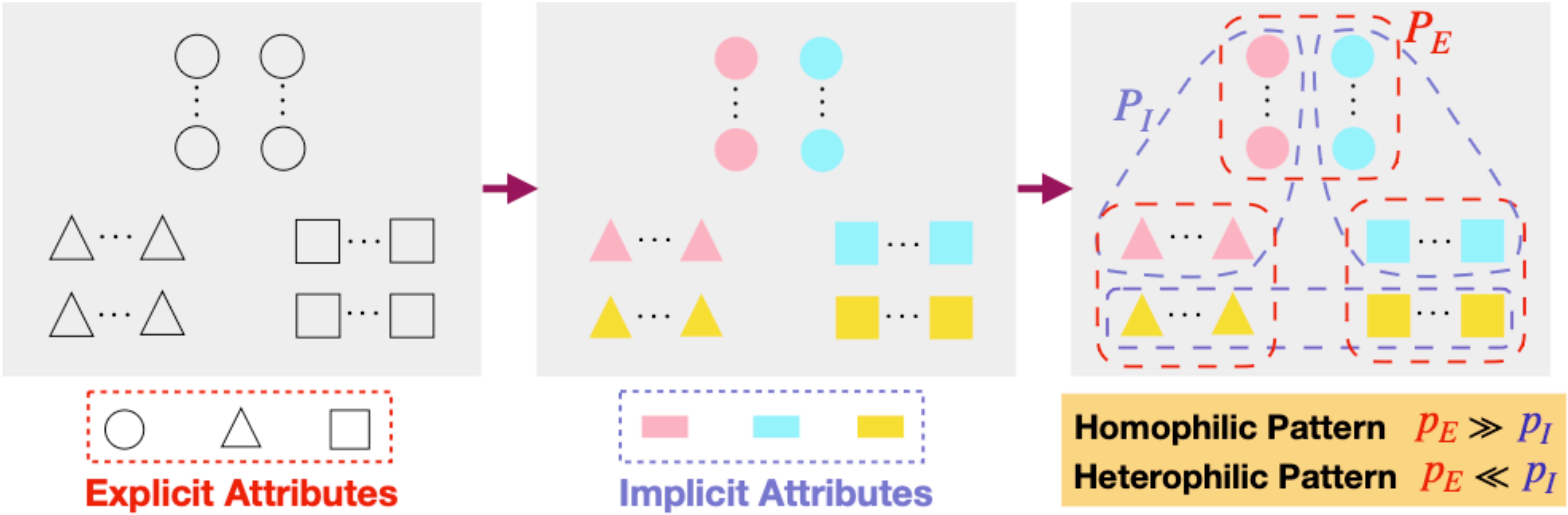


Datasets	Heterophilic Graphs							Homophilic Graphs			
	Squirrel	**Chameleon**	**Wisconsin**	**Cornell**	**Texas**	**Twitch-DE**	**Actor**	**Cora**	**Citeseer**	**Pubmed**	**Polblogs**
GCN [30]	55.2±1.5	67.6±2.0	59.5±3.6	52.8±6.0	61.7±3.7	74.0±1.2	31.2±1.3	79.7±1.2	69.5±1.7	78.7±1.6	89.4±0.9
SGC [6]	50.7±1.3	61.9±2.6	53.7±3.9	51.2±0.9	51.4±2.2	73.9±1.3	30.9±0.6	79.1±1.0	69.9±2.0	76.6±1.3	89.0±1.5
GAT [26]	54.8±2.2	67.3±2.2	57.9±4.5	50.4±5.9	55.4±5.9	73.7±1.3	30.5±1.2	82.0±1.1	69.9±1.7	78.6±2.0	87.4±1.1
NeuralSparse [48]	40.0±1.6	60.5±2.0	70.8±3.4	64.1±5.5	66.4±5.7	71.3±1.3	35.5±1.1	78.5±1.4	69.7±1.8	79.1±1.2	89.3±0.9
GCN-LPA [49]	54.2±1.1	63.4±1.9	63.3±3.7	65.6±7.3	61.2±7.6	74.0±1.2	37.8±0.9	80.4±1.5	69.7±1.7	79.7±1.3	**89.7±0.8**
DisenGCN [66]	42.4±1.6	58.4±2.3	78.1±4.0	77.4±4.4	71.3±5.7	73.5±1.7	36.7±1.2	81.5±1.3	69.2±1.7	80.0±1.6	89.5±0.9
FactorGCN [33]	56.6±2.4	69.8±2.0	64.2±4.8	50.6±1.8	69.5±6.5	73.1±1.4	29.0±1.4	75.2±1.6	61.6±2.0	72.9±2.3	87.9±1.7
VEPM [71]	50.3±1.7	67.3±2.1	55.6±4.9	51.2±7.0	55.8±4.3	73.3±1.2	29.3±1.1	82.2±1.2	69.1±1.9	78.8±1.6	89.5±0.9
DisGNN [72]	55.1±4.8	68.2±1.9	54.6±5.4	52.0±5.7	60.6±3.9	69.2±0.8	30.2±1.3	78.2±1.4	66.2±2.2	77.6±1.7	89.6±0.9
GEN [13]	36.0±4.0	57.6±3.1	83.3±3.6	81.0±3.9	78.3±8.0	74.1±1.4	37.3±1.4	79.8±1.3	69.7±1.6	78.9±1.7	89.6±1.4
WRGAT [14]	39.6±1.4	57.7±1.6	82.9±4.5	79.2±3.5	80.5±6.1	70.0±1.3	38.6±1.1	71.7±1.5	64.1±1.9	73.3±2.1	88.2±1.2
H2GCN [10]	45.1±1.9	62.9±1.9	82.6±4.0	79.6±4.9	79.8±7.3	73.1±1.5	38.4±1.0	81.4±1.4	68.7±2.0	78.0±2.0	89.0±1.0
FAGCN [18]	50.4±2.6	68.9±1.8	82.3±4.4	79.4±5.5	80.3±5.5	74.1±1.4	37.9±1.0	82.6±1.3	70.3±1.6	80.0±1.7	89.3±1.1
GPR-GNN [11]	54.1±1.6	69.6±1.7	82.7±4.1	79.9±5.3	81.7±4.9	74.0±1.6	38.0±1.1	81.5±1.5	69.6±1.7	79.8±1.3	89.5±0.8
GloGNN++ [23]	63.3±1.2	71.4±2.0	84.9±4.2	82.0±3.5	81.4±5.6	72.8±1.1	38.2±1.2	80.9±1.4	70.5±1.9	76.8±2.1	89.6±0.8
ACM-GCN [46]	**67.0±1.3**	**75.3±2.2**	84.3±4.5	82.1±4.9	82.2±5.9	74.2±0.9	36.6±1.0	81.3±1.0	69.4±1.7	79.5±1.4	89.6±0.9
GOAL [47]	57.9±0.9	71.3±2.0	70.5±5.1	54.9±6.6	72.0±7.4	68.5±1.5	36.3±1.0	80.6±1.4	69.7±2.0	78.7±1.3	88.7±1.6
ES-GNN (ours)	62.4±1.4	72.3±2.1	**85.3±4.6**	**82.2±4.0**	**82.3±5.7**	**74.7±1.1**	**38.9±0.8**	**83.0±1.1**	**70.7±1.7**	**80.7±1.4**	**89.7±0.9**
Error Reduction	**11.5%**	**6.4%**	**11.0%**	**11.7%**	**9.4%**	**2.2%**	**3.2%**	**3.3%**	**2.3%**	**2.6%**	**0.5%**

Jingwei Guo, et.al. ES-GNN: Generalizing Graph Neural Networks Beyond Homophily with Edge Splitting. *TPAMI (Minor Revision), 2024*.

Proposed Works (2nd) — Edge Splitting GNN

Edge Splitting GNN — Synthetic Graphs



Task-relevant Features

Task-irrelevant Features

CSBM

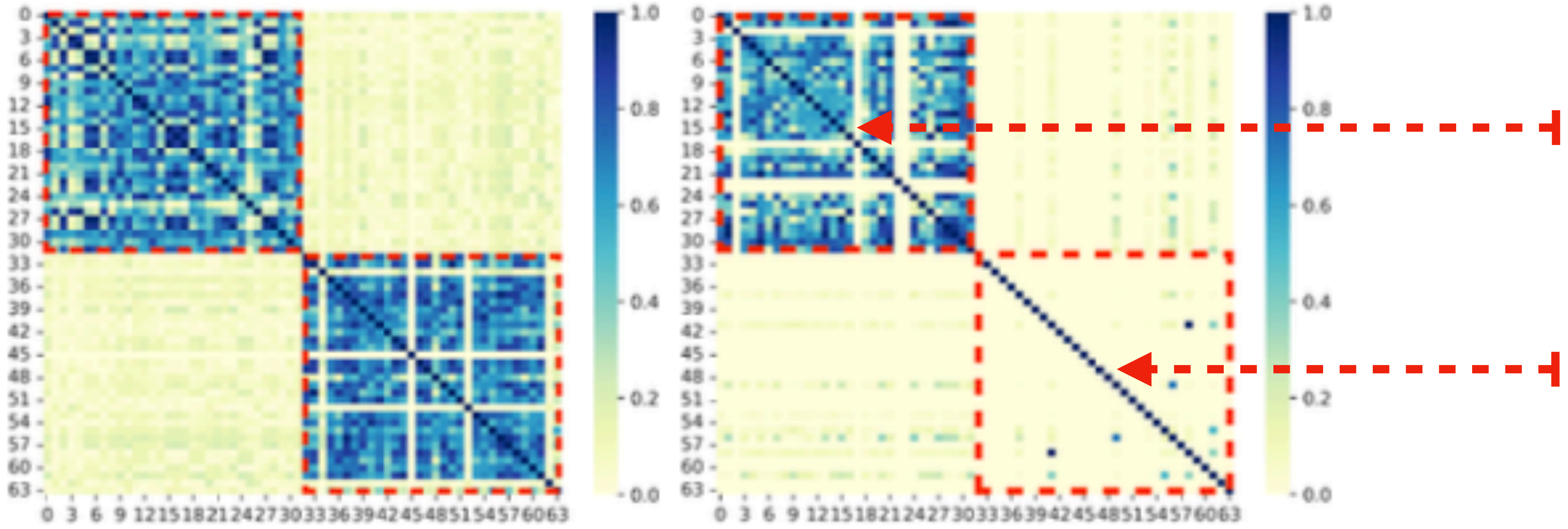
Jingwei Guo, et.al. ES-GNN: Generalizing Graph Neural Networks Beyond Homophily with Edge Splitting. *TPAMI (Minor Revision), 2024*.

Proposed Works (2nd) — Edge Splitting GNN

Edge Splitting GNN — Feature Correlation

- Real-world

Heterophilic vs. Homophilic



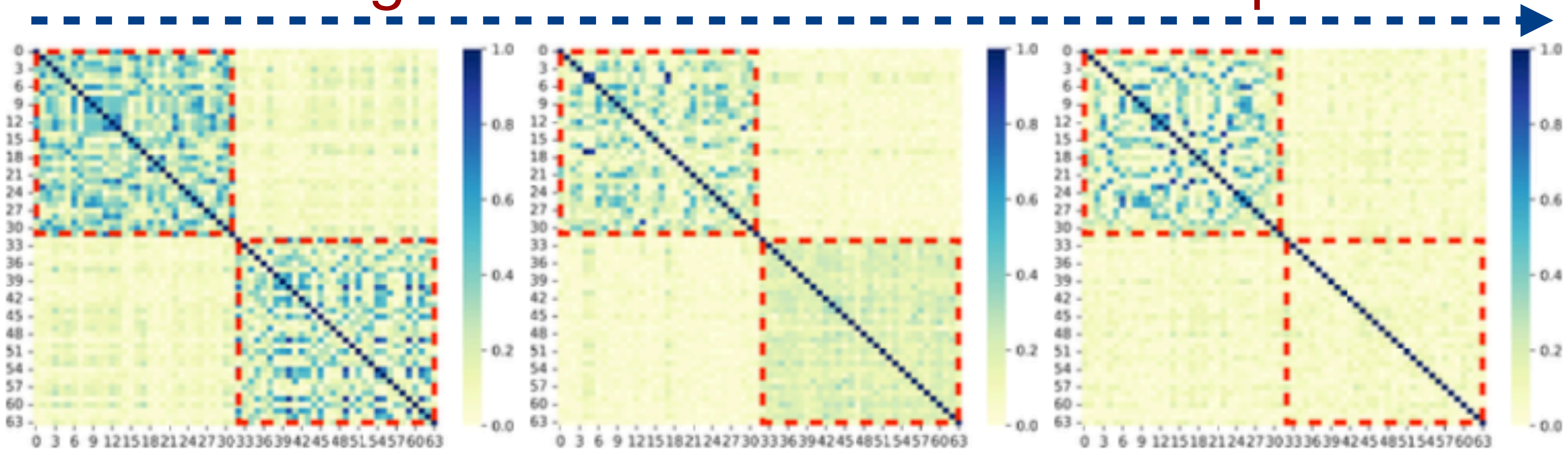
Task-Relevant

Task-Irrelevant

(a) Chameleon (b) Cora

- Synthetic

A reducing trend on the 2nd block-wise pattern



(c) $\mathcal{H}_{\text{syn}} = 0.1$ (d) $\mathcal{H}_{\text{syn}} = 0.5$ (e) $\mathcal{H}_{\text{syn}} = 0.9$

Jingwei Guo, et.al. ES-GNN: Generalizing Graph Neural Networks Beyond Homophily with Edge Splitting. *TPAMI (Minor Revision), 2024*.

Proposed Works (2nd) — Edge Splitting GNN

Edge Splitting GNN

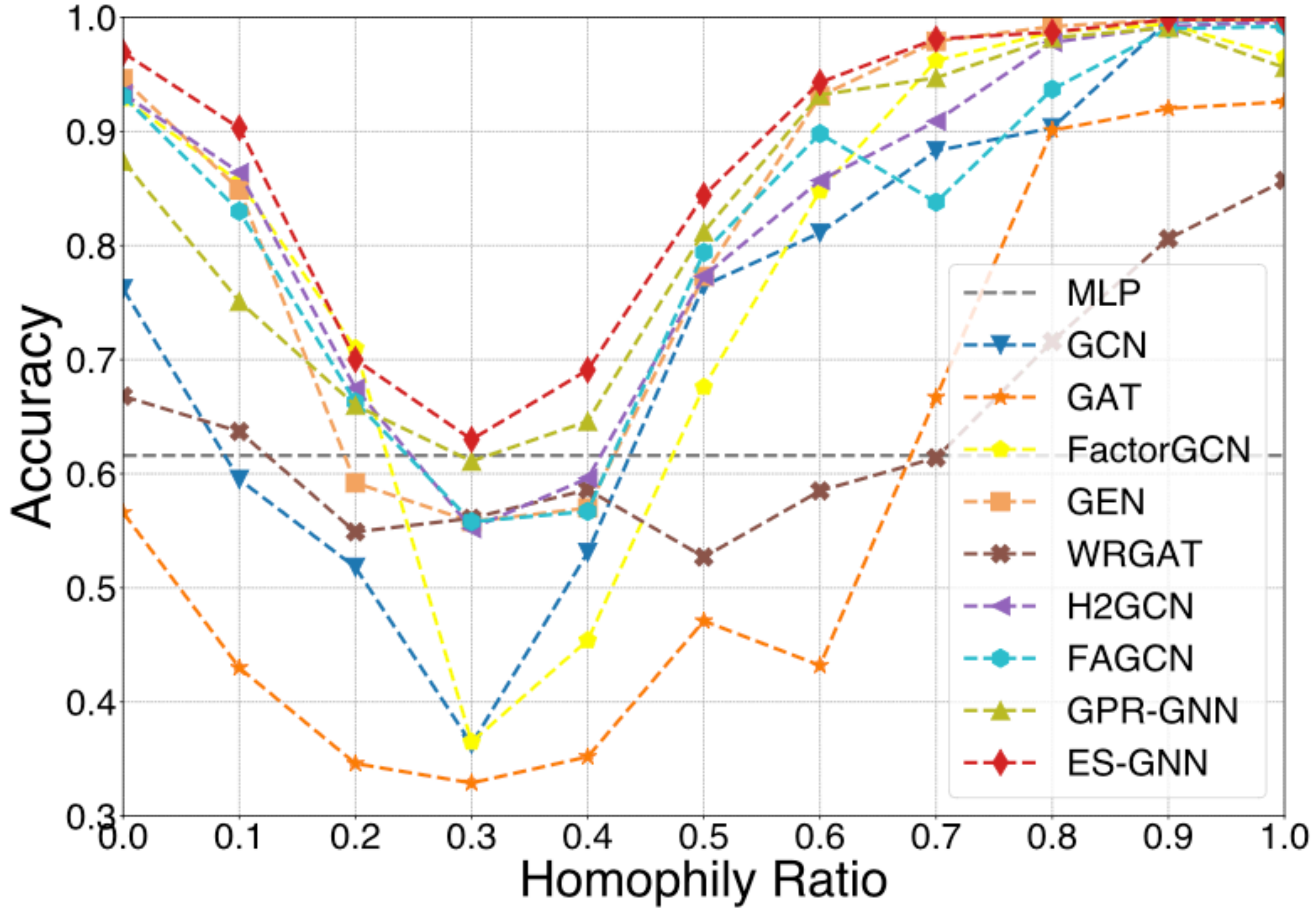
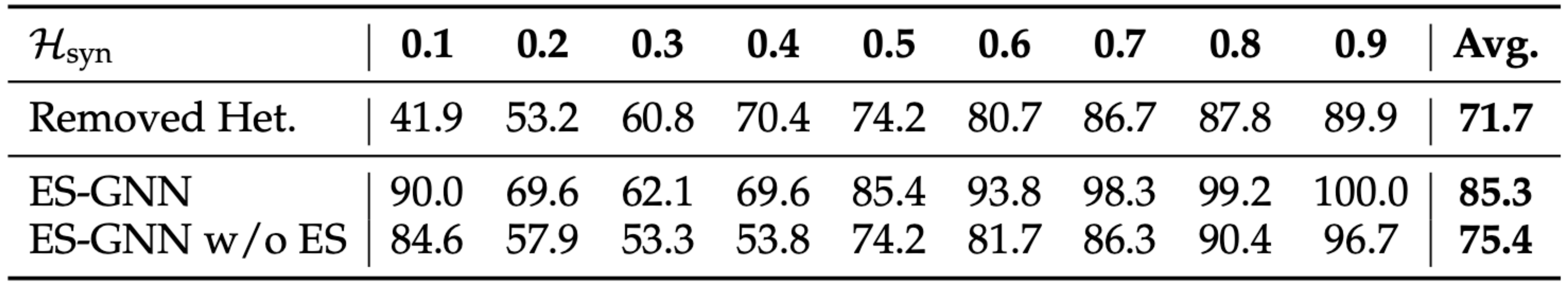


Fig. 5. Results of different models on synthetic graphs with varied homophily ratios, where ES-GNN constantly outperform all the baselines.

TABLE 5

Edge Analysis of our ES-GNN on synthetic graphs with various homophily ratios. "Removed Het." gives the percentage (%) of heterophilic (inter-class) node connections excluded from the task-relevant topology and disentangled in the task-irrelevant topology. The last two rows list the corresponding node classification accuracies (%) of ES-GNN and its variant while ablating ES-layer.



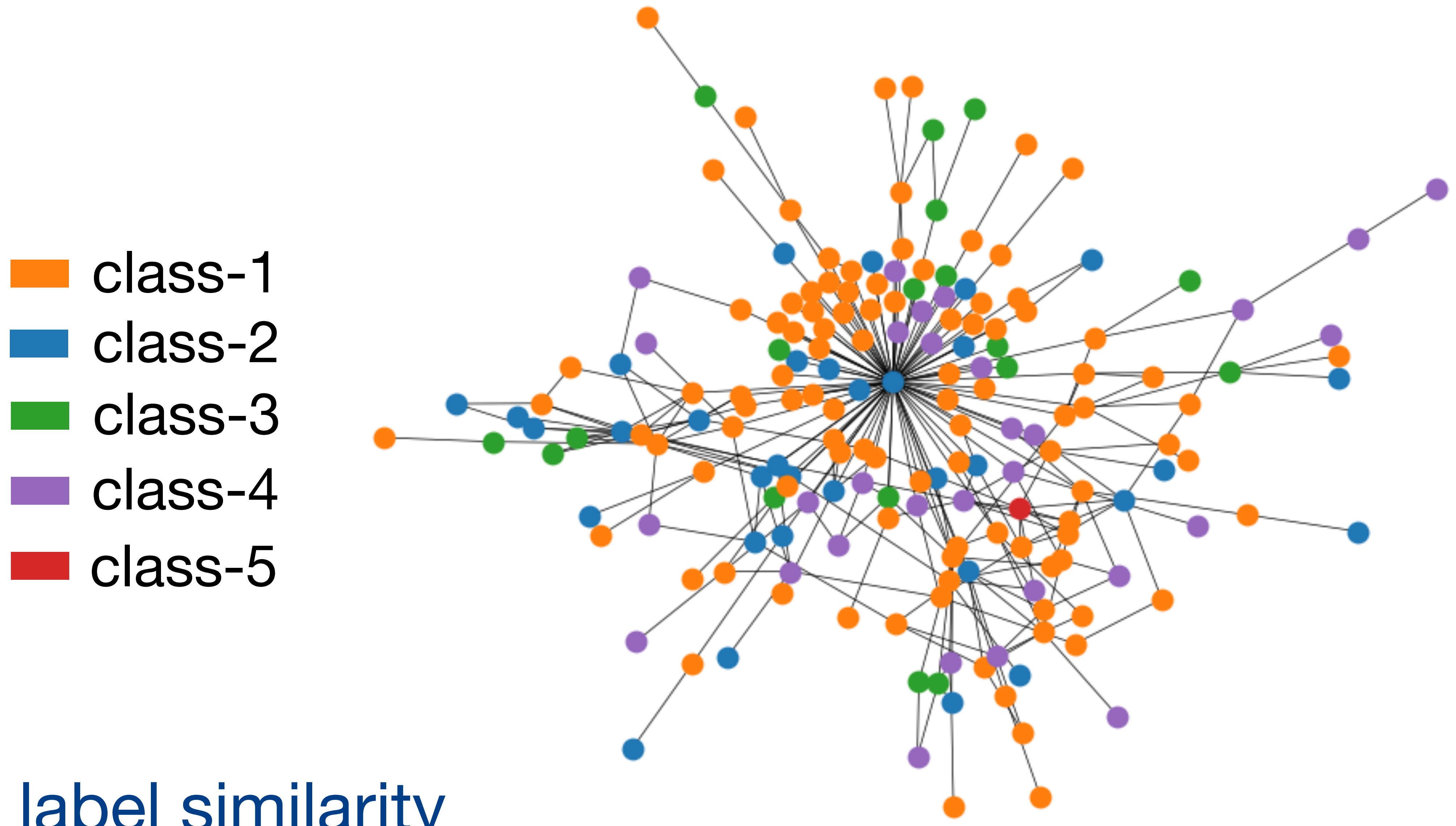
$\mathcal{H}_{\text{syn}}$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**Avg.**
Removed Het.	41.9	53.2	60.8	70.4	74.2	80.7	86.7	87.8	89.9	**71.7**
ES-GNN	90.0	69.6	62.1	69.6	85.4	93.8	98.3	99.2	100.0	**85.3**
ES-GNN w/o ES	84.6	57.9	53.3	53.8	74.2	81.7	86.3	90.4	96.7	**75.4**

Jingwei Guo, et.al. ES-GNN: Generalizing Graph Neural Networks Beyond Homophily with Edge Splitting. *TPAMI (Minor Revision), 2024*.

Research Experience — Solution 3

Address Graph Regional Disparity

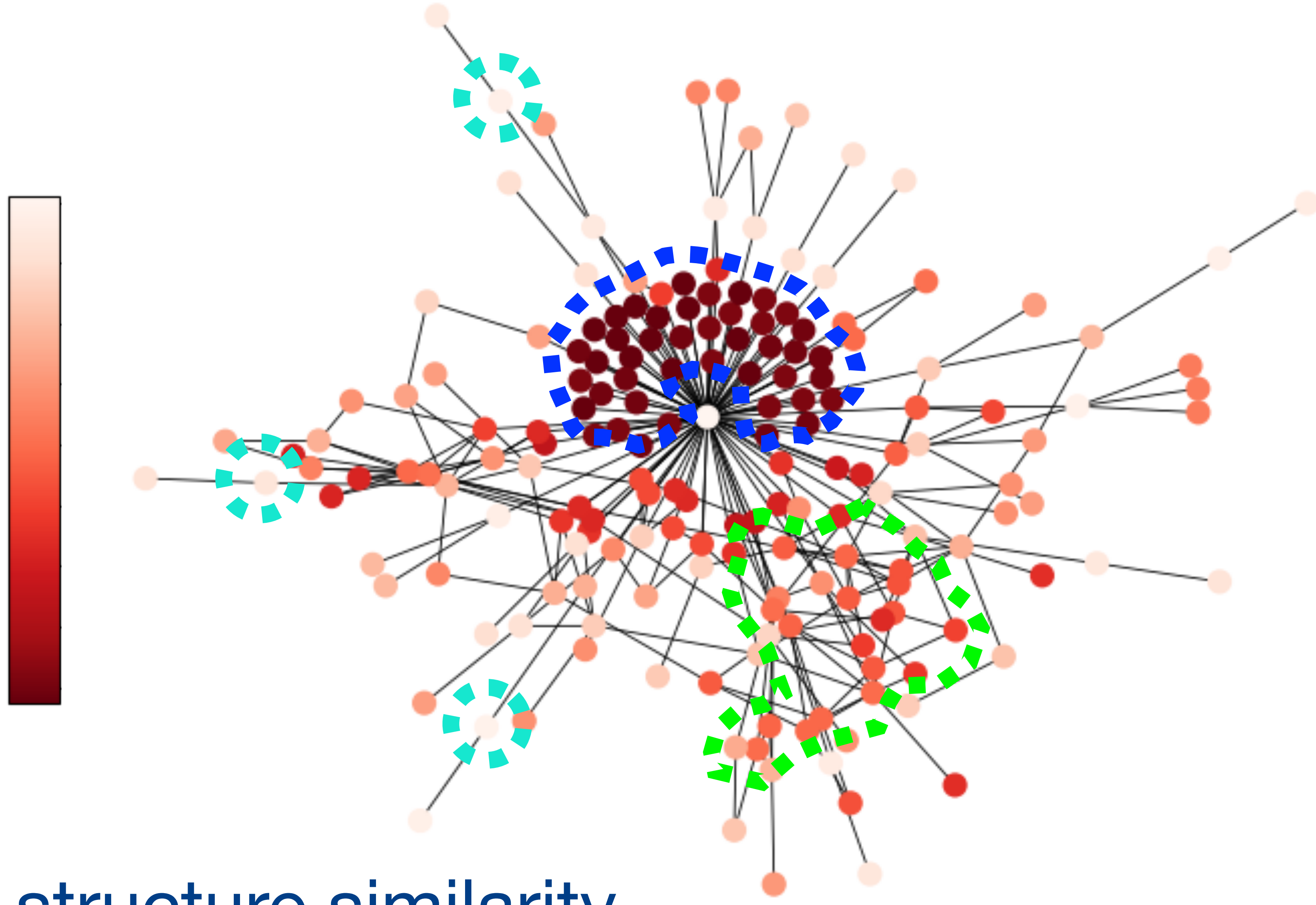
- Edge-level (Previous)



label similarity

Pairwise Distinction

- Subgraph-level (Ours)



structure similarity

Regional Disparity

Jingwei Guo, et.al. Graph Neural Networks with Diverse Spectral Filtering. *WWW, 2023.*

Research Experience — Solution 3

Diverse Spectral Filtering

$$\mathbf{Z} = g_\psi(\hat{\mathbf{L}})\mathbf{X} = \sum_{k=0}^{K} \psi_k P_k(\hat{\mathbf{L}})\mathbf{X}$$

Polynomial Approx. with Shared Parameters

Most spectral GNNs assumes homogenous distributions between different graph regions. ☹

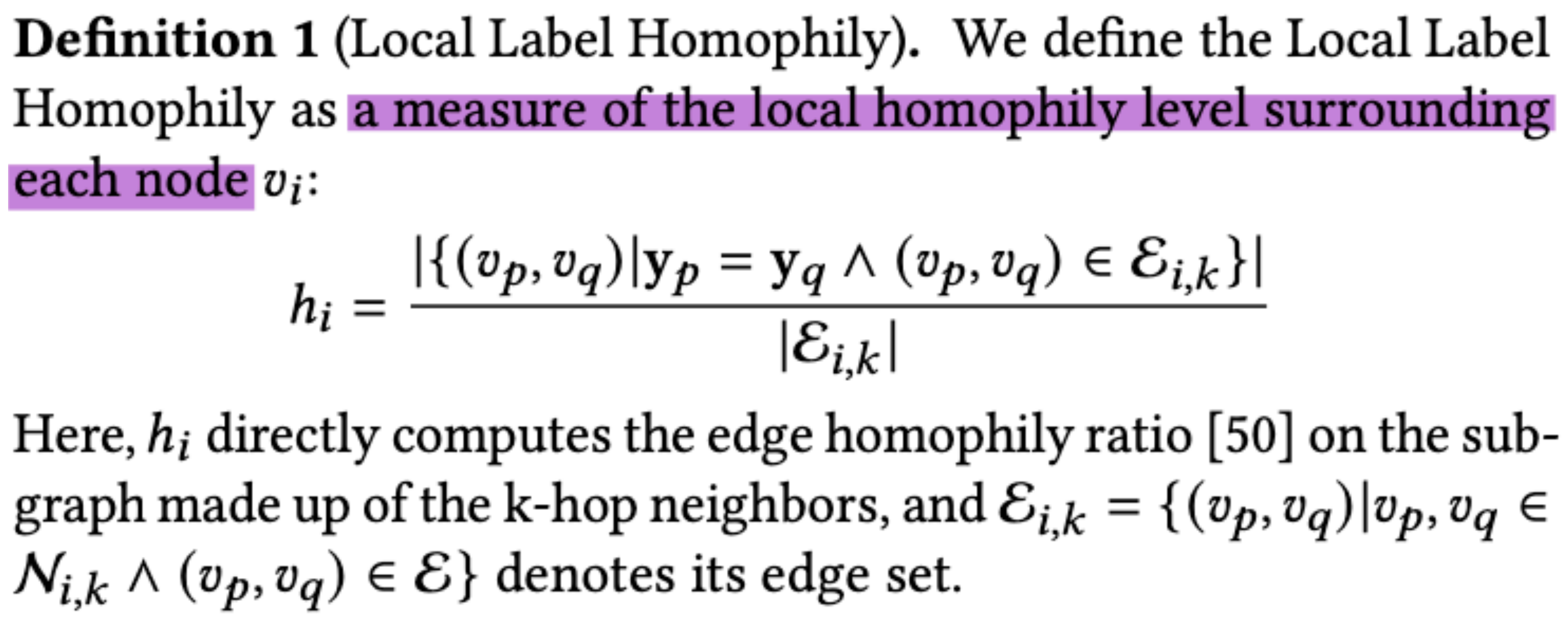
$$\mathbf{Z} = \sum_{k=0}^{K} \begin{pmatrix} \beta_{k,1} & 0 & \cdots & 0 \\ 0 & \beta_{k,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \beta_{k,N} \end{pmatrix} P_k(\hat{\mathbf{L}})\mathbf{X}$$

We augment the original parameters into node-specific filter weights to model diverse regional patterns. ☺

Jingwei Guo, et.al. Graph Neural Networks with Diverse Spectral Filtering. *WWW, 2023.*

Research Experience — Solution 3

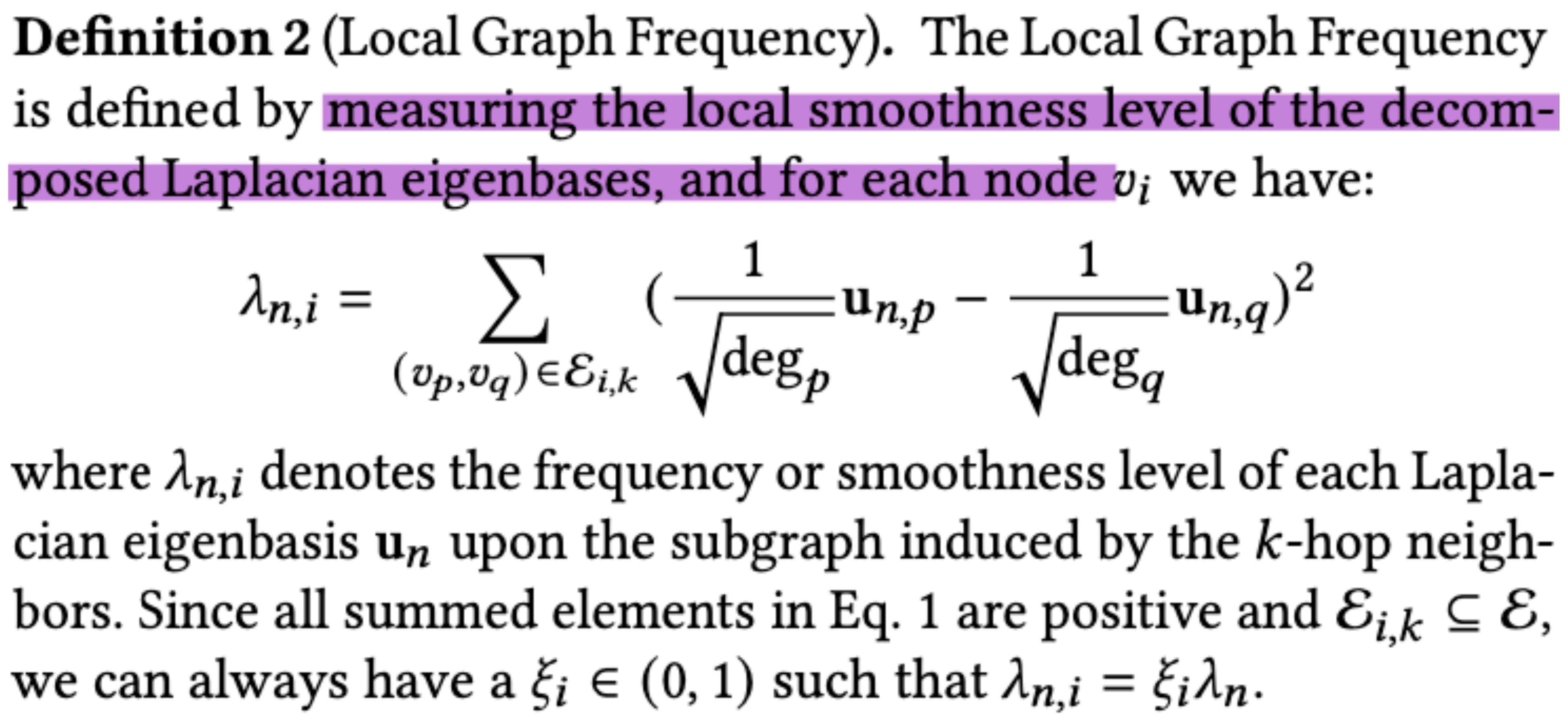
Diverse Spectral Filtering



Definition 1 (Local Label Homophily). We define the Local Label Homophily as a measure of the local homophily level surrounding each node v_i:

$$h_i = \frac{|\{(v_p, v_q)|\mathbf{y}_p = \mathbf{y}_q \wedge (v_p, v_q) \in \mathcal{E}_{i,k}\}|}{|\mathcal{E}_{i,k}|}$$

Here, h_i directly computes the edge homophily ratio [50] on the subgraph made up of the k-hop neighbors, and $\mathcal{E}_{i,k} = \{(v_p, v_q)|v_p, v_q \in \mathcal{N}_{i,k} \wedge (v_p, v_q) \in \mathcal{E}\}$ denotes its edge set.



Definition 2 (Local Graph Frequency). The Local Graph Frequency is defined by measuring the local smoothness level of the decomposed Laplacian eigenbases, and for each node v_i we have:

$$\lambda_{n,i} = \sum_{(v_p,v_q)\in\mathcal{E}_{i,k}} \left(\frac{1}{\sqrt{\deg_p}}\mathbf{u}_{n,p} - \frac{1}{\sqrt{\deg_q}}\mathbf{u}_{n,q}\right)^2$$

where $\lambda_{n,i}$ denotes the frequency or smoothness level of each Laplacian eigenbasis $\mathbf{u}_n$ upon the subgraph induced by the k-hop neighbors. Since all summed elements in Eq. 1 are positive and $\mathcal{E}_{i,k} \subseteq \mathcal{E}$, we can always have a $\xi_i \in (0, 1)$ such that $\lambda_{n,i} = \xi_i\lambda_n$.

Label Homophily

$$h = \frac{|\{(v_i,v_j)|\mathbf{y}_i=\mathbf{y}_j\wedge(v_i,v_j)\in\mathcal{E}\}|}{|\mathcal{E}|}$$

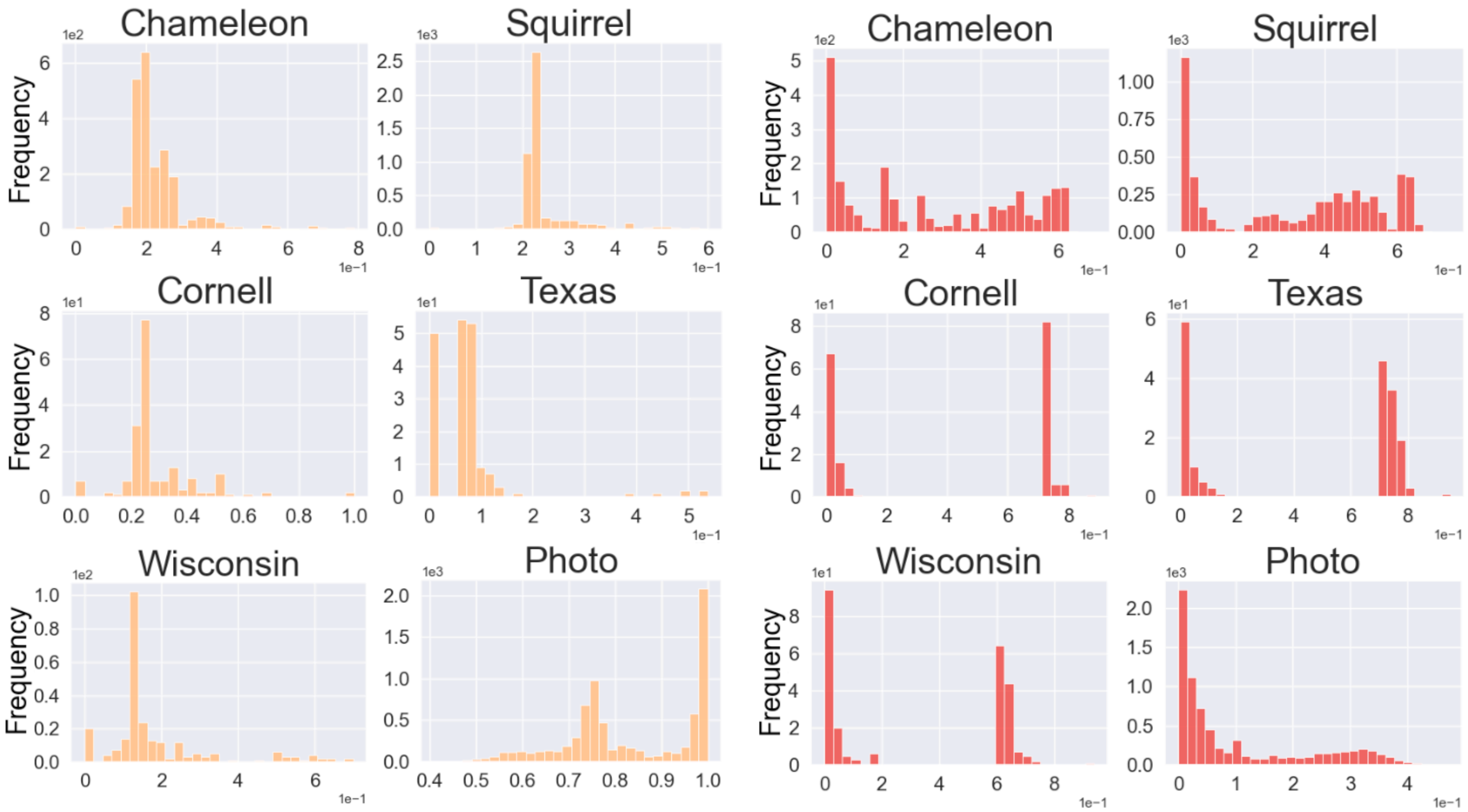
Graph Frequency (Eigenvalue)

$$\begin{aligned}\lambda_n &= \mathbf{u}_n^T\hat{\mathbf{L}}\mathbf{u}_n \\ &= \sum_{(v_p,v_q)\in\mathcal{E}} \left(\frac{1}{\sqrt{\deg_p}}\mathbf{u}_{n,p} - \frac{1}{\sqrt{\deg_q}}\mathbf{u}_{n,q}\right)^2\end{aligned}$$

Jingwei Guo, et.al. Graph Neural Networks with Diverse Spectral Filtering. *WWW, 2023*.

Research Experience — Solution 3

Diverse Spectral Filtering



(a) Local Label Homophily

(b) Local Graph Frequency

Evident Regional Heterogeneity

Jingwei Guo, et.al. Graph Neural Networks with Diverse Spectral Filtering. *WWW, 2023*.

Research Experience — Solution 3

Homogenous Spectral Filtering

Dot product

$$\mathbf{Z} = \sum_{n=1}^{N} \tilde{S}_n \cdot \mathbf{U}_n$$

Scalar coefficient

Graph frequency

$$\tilde{S}_n = \sum_{k=0}^{K} \alpha_k P_k(\lambda_n) \mathbf{U}_n^T \mathbf{X}$$

Diverse Spectral Filtering

Hadamard product

$$\mathbf{Z} = \sum_{n=1}^{N} \tilde{\mathbf{S}}_n \odot \mathbf{U}_n$$

Vector

The i-th element of vector coefficients

$$\tilde{\mathbf{S}}_n(i) = \sum_{k=0}^{K} \alpha_k P_k(\lambda_{n,i}) \mathbf{U}_n^T \mathbf{X}$$

Local graph frequency

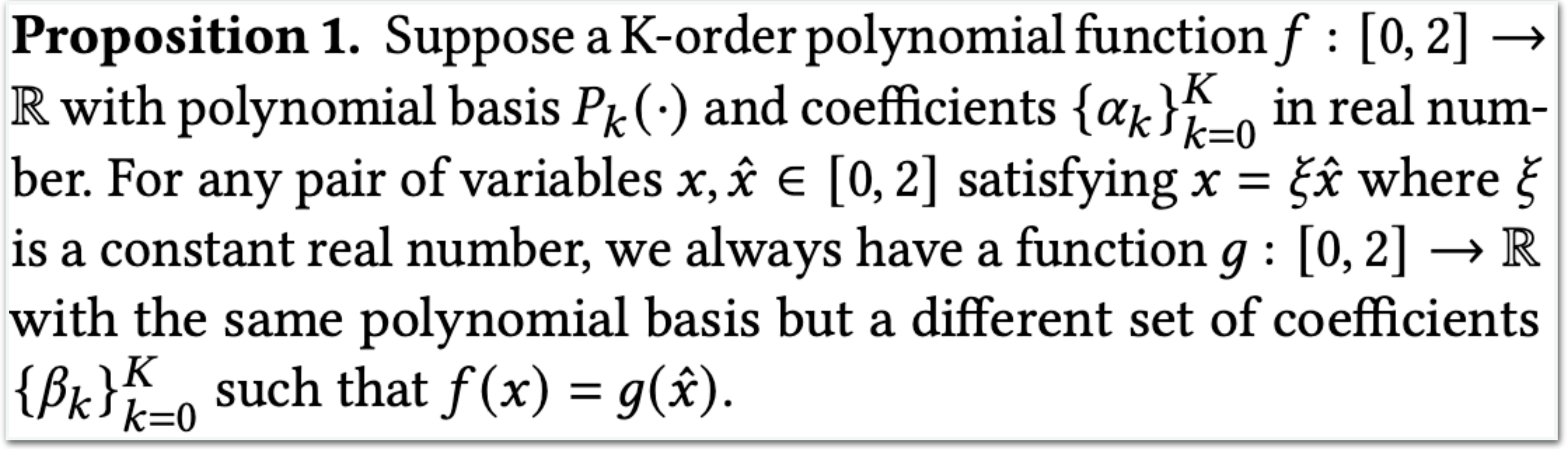
$\mathbf{X}$ is taken as one-dimension as an example

Jingwei Guo, et.al. Graph Neural Networks with Diverse Spectral Filtering. *WWW, 2023*.

Research Experience — Solution 3

Diverse Spectral Filtering

Substitution using $\lambda_{n,i} = \xi_i \lambda_n$ s.t. $0 < \xi_i < 1$



Proposition 1. Suppose a K-order polynomial function $f : [0,2] \rightarrow \mathbb{R}$ with polynomial basis $P_k(\cdot)$ and coefficients $\{\alpha_k\}_{k=0}^{K}$ in real number. For any pair of variables $x, \hat{x} \in [0,2]$ satisfying $x = \xi \hat{x}$ where ξ is a constant real number, we always have a function $g : [0,2] \rightarrow \mathbb{R}$ with the same polynomial basis but a different set of coefficients $\{\beta_k\}_{k=0}^{K}$ such that $f(x) = g(\hat{x})$.

It allows $\mathbf{S}'_n(i) = \sum_{k=0}^{K} \boxed{\alpha_k P_k(\lambda_{n,i})} \mathbf{U}_n^T \mathbf{X} = \sum_{k=0}^{K} \boxed{\beta_{k,i} P_k(\lambda_i)} \mathbf{U}_n^T \mathbf{X}$

$$\mathbf{Z} = \sum_{k=0}^{K} \textbf{diag}(\beta_{k,1}, \beta_{k,2}, \cdots, \beta_{k,N}) P_k(\hat{\mathbf{L}}) \mathbf{X}$$

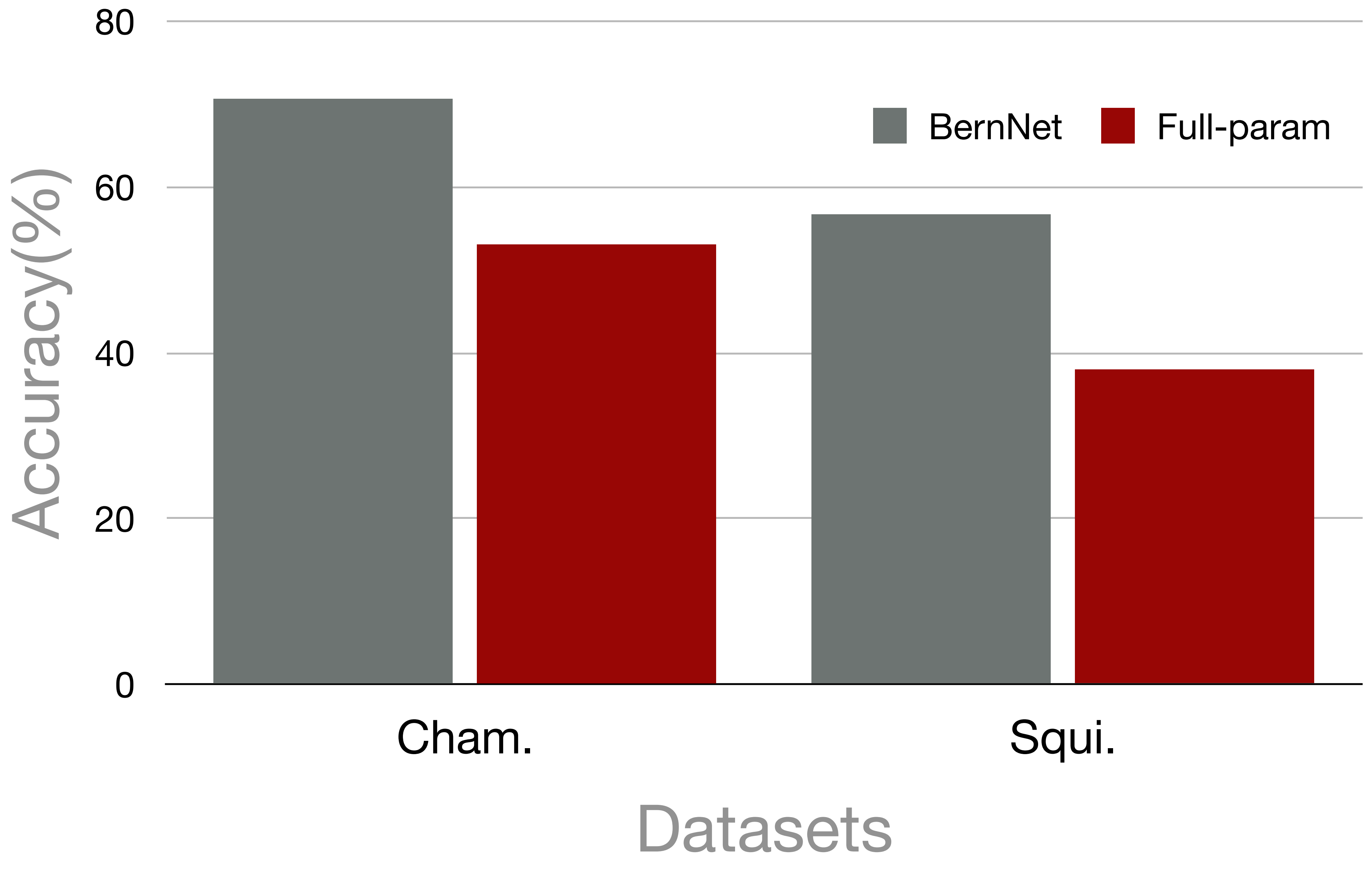
Jingwei Guo, et.al. Graph Neural Networks with Diverse Spectral Filtering. *WWW, 2023.*

Research Experience — Solution 3

Full Parameterization Challenges

- Parameterizing a large number of filter weights (∝ # nodes) would increase model complexity and cause severe overfitting to local noise.

Comparison in Node classification



“A reasonable design should be built upon a shared global model whilst locally adapted to each node with awareness of its graph position.”

Jingwei Guo, et.al. Graph Neural Networks with Diverse Spectral Filtering. *WWW, 2023.*

Research Experience — Solution 3

Diverse Spectral Filtering

- Local and Global Weight Decomposition

$$\beta_{k,i} \leftarrow \gamma_i \cdot \alpha_{k,i}$$

global invariant graph properties

local diverse node contexts

- Position-aware Filter Weights

$$\arg\min_{\mathbf{P}} \|\mathbf{P}^{(0)} - \mathbf{P}\|_2^2 + \kappa_1 tr(\mathbf{P}^T\hat{\mathbf{L}}\mathbf{P}) + \kappa_2\|\mathbf{P}^T\mathbf{P} - \mathbf{I}\|_2^2$$

$\mathbf{P}$ denotes node positional embeddings

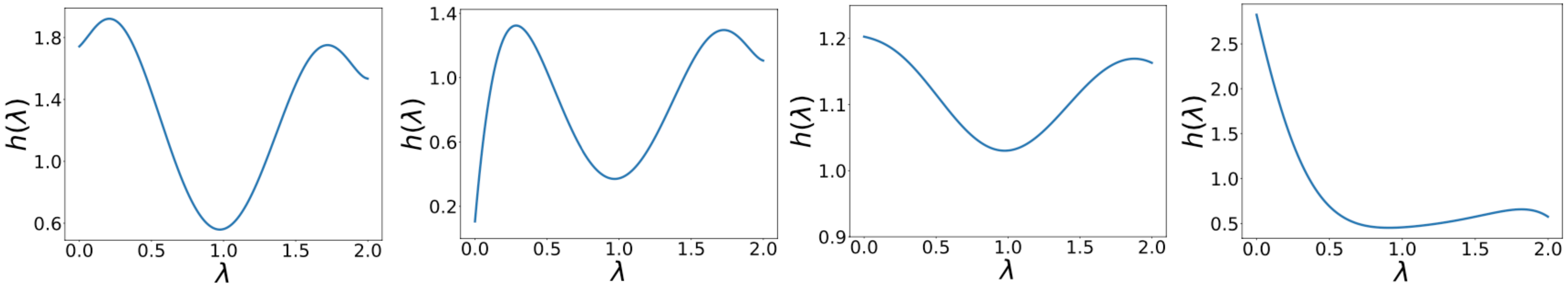
$$\alpha_{k,i} = \sigma_p(\mathbf{W}^{(k)}\mathbf{P}_i^{(k)} + \mathbf{b}^{(k)}) \qquad k = 1,2,\cdots,K$$

Jingwei Guo, et.al. Graph Neural Networks with Diverse Spectral Filtering. *WWW, 2023*.

Research Experience — Solution 3

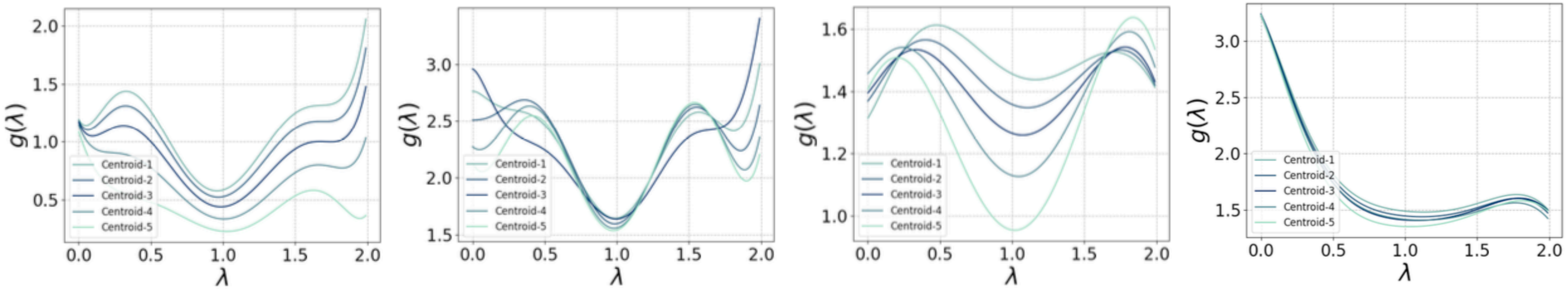
Diverse Spectral Filtering

- Conventional Spectral GNN: BernNet



(f) Chameleon (h) Squirrel (j) Cornell (b) CiteSeer

- Ours: DSF-BernNet

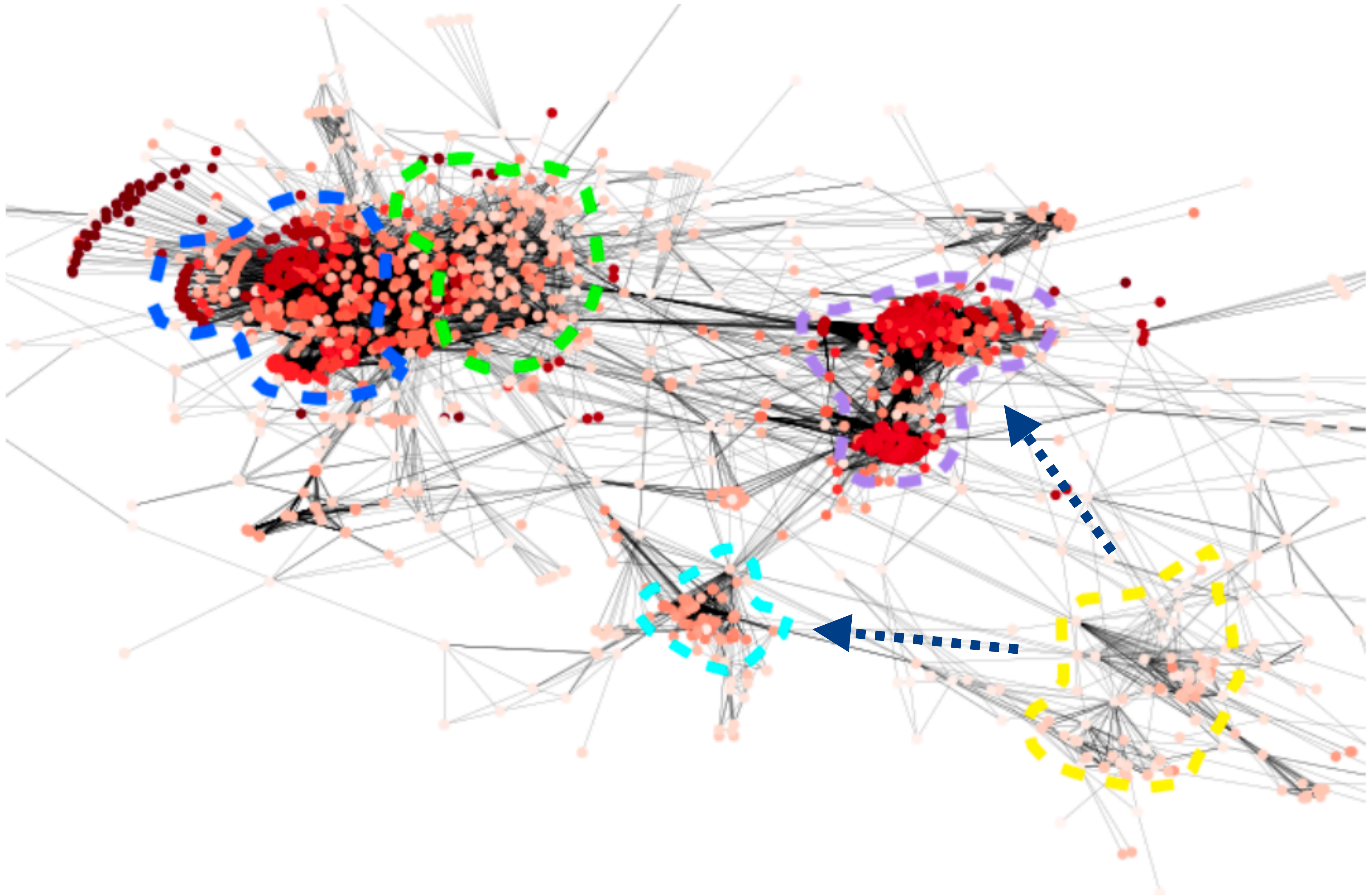


(a) Chameleon (b) Squirrel (c) Cornell (d) Citeseer

Jingwei Guo, et.al. Graph Neural Networks with Diverse Spectral Filtering. *WWW, 2023*.

Research Experience — Solution 3

Diverse Spectral Filtering



Chameleon

Squirrel

DSF captures regional disparity with node-specific filter weights.

Jingwei Guo, et.al. Graph Neural Networks with Diverse Spectral Filtering. *WWW, 2023*.

Research Experience — Solution 3

Diverse Spectral Filtering

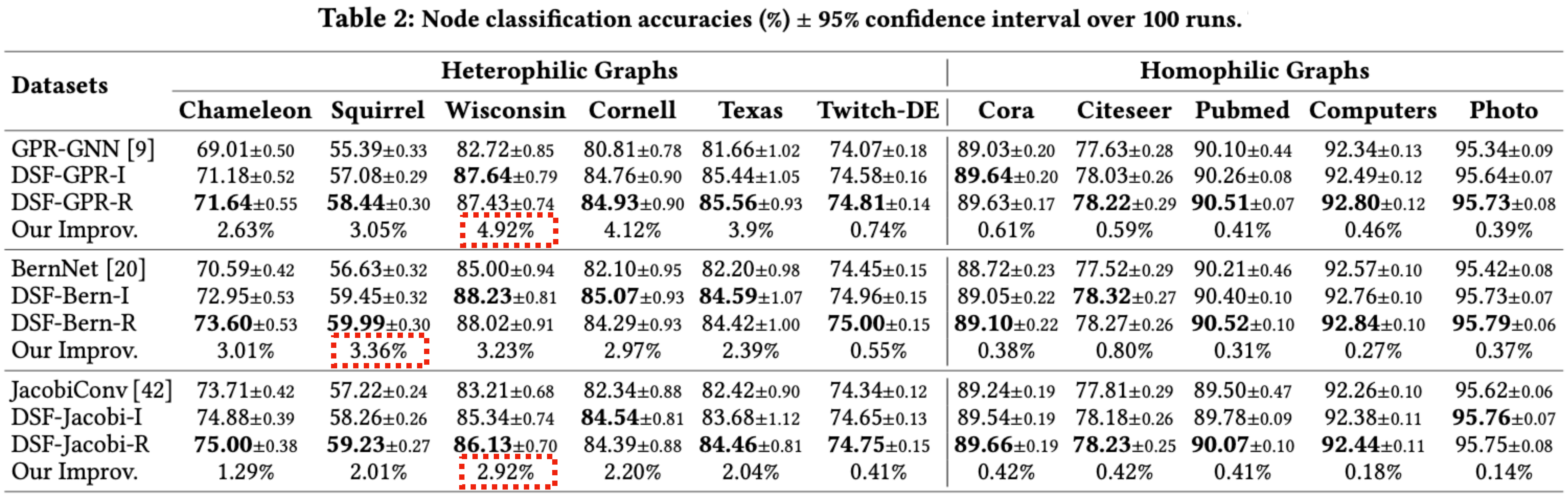


Table 2: Node classification accuracies (%) ± 95% confidence interval over 100 runs.

Datasets	Heterophilic Graphs						Homophilic Graphs				
	Chameleon	Squirrel	Wisconsin	Cornell	Texas	Twitch-DE	Cora	Citeseer	Pubmed	Computers	Photo
GPR-GNN [9]	69.01±0.50	55.39±0.33	82.72±0.85	80.81±0.78	81.66±1.02	74.07±0.18	89.03±0.20	77.63±0.28	90.10±0.44	92.34±0.13	95.34±0.09
DSF-GPR-I	71.18±0.52	57.08±0.29	**87.64**±0.79	84.76±0.90	85.44±1.05	74.58±0.16	**89.64**±0.20	78.03±0.26	90.26±0.08	92.49±0.12	95.64±0.07
DSF-GPR-R	**71.64**±0.55	**58.44**±0.30	87.43±0.74	**84.93**±0.90	**85.56**±0.93	**74.81**±0.14	89.63±0.17	**78.22**±0.29	**90.51**±0.07	**92.80**±0.12	**95.73**±0.08
Our Improv.	2.63%	3.05%	4.92%	4.12%	3.9%	0.74%	0.61%	0.59%	0.41%	0.46%	0.39%
BernNet [20]	70.59±0.42	56.63±0.32	85.00±0.94	82.10±0.95	82.20±0.98	74.45±0.15	88.72±0.23	77.52±0.29	90.21±0.46	92.57±0.10	95.42±0.08
DSF-Bern-I	72.95±0.53	59.45±0.32	**88.23**±0.81	**85.07**±0.93	**84.59**±1.07	74.96±0.15	89.05±0.22	**78.32**±0.27	90.40±0.10	92.76±0.10	95.73±0.07
DSF-Bern-R	**73.60**±0.53	**59.99**±0.30	88.02±0.91	84.29±0.93	84.42±1.00	**75.00**±0.15	**89.10**±0.22	78.27±0.26	**90.52**±0.10	**92.84**±0.10	**95.79**±0.06
Our Improv.	3.01%	3.36%	3.23%	2.97%	2.39%	0.55%	0.38%	0.80%	0.31%	0.27%	0.37%
JacobiConv [42]	73.71±0.42	57.22±0.24	83.21±0.68	82.34±0.88	82.42±0.90	74.34±0.12	89.24±0.19	77.81±0.29	89.50±0.47	92.26±0.10	95.62±0.06
DSF-Jacobi-I	74.88±0.39	58.26±0.26	85.34±0.74	**84.54**±0.81	83.68±1.12	74.65±0.13	89.54±0.19	78.18±0.26	89.78±0.09	92.38±0.11	**95.76**±0.07
DSF-Jacobi-R	**75.00**±0.38	**59.23**±0.27	**86.13**±0.70	84.39±0.88	**84.46**±0.81	**74.75**±0.15	**89.66**±0.19	**78.23**±0.25	**90.07**±0.10	**92.44**±0.11	95.75±0.08
Our Improv.	1.29%	2.01%	2.92%	2.20%	2.04%	0.41%	0.42%	0.42%	0.41%	0.18%	0.14%

DSF can be readily plug-and-play in multiple spectral GNNs and consistently improve their performance.

Jingwei Guo, et.al. Graph Neural Networks with Diverse Spectral Filtering. *WWW, 2023*.

Research Experience — Solution 4

Deep Delve into Spectral GNNs

$$\mathbf{Z} = \mathbf{U} g_{\psi}(\mathbf{\Lambda}) \mathbf{U}^T \mathbf{X}$$

theoretically rooted in spectral domain

$$= g_{\psi}(\hat{\mathbf{L}}) \mathbf{X} = \sum_{k=1}^{K} \psi_k P_k(\hat{\mathbf{L}}) \mathbf{X}$$

practically relying on spatial approximation

- Spatial Interpretability?

$$\hat{\mathbf{A}}^{new} = \mathbf{I} - \frac{\alpha}{1-\alpha}\left(g_{\psi}(\hat{\mathbf{L}})^{-1} - \mathbf{I}\right)$$

Non-locality & Signed Edges

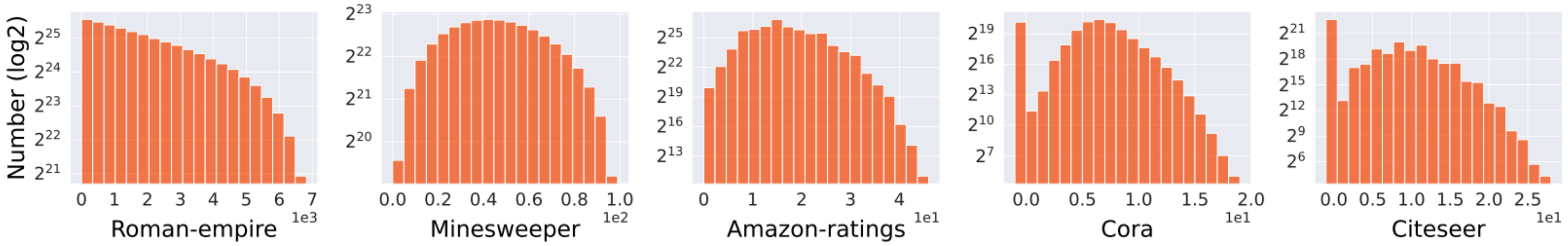


Figure 1: Distributions of connected nodes in the new graph based on their geodesic/shortest-path distance (as $\Delta_{i,j}$) in the original graph. Nodes, distant in the original graph ($\Delta_{i,j} > 1$ in x-axis), can be linked in the new graph (Number > 0 in y-axis).

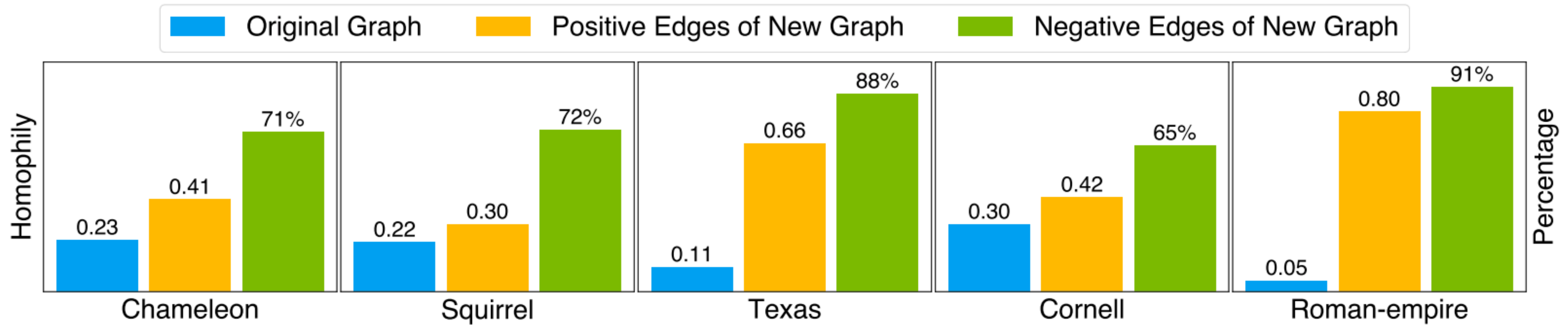


Figure 2: Left y-axis: Homophily comparison between original and new graphs, considering only positive edges (blue and yellow bars). Right y-axis: Percentage of edges connecting nodes from different classes, identified by negative edges (green bar).

Jingwei Guo, et.al. Rethinking Spectral Graph Neural Networks with Spatially Adaptive Filtering. *TNNLS* (Under Review), 2024.

Research Experience — Solution 4

Deep Delve into Spectral GNNs

Cross-Domain Interplay via the Lens of Graph Optimization

$$\arg\min_{\mathbf{Z}} \mathscr{L} = \alpha\|\mathbf{X}-\mathbf{Z}\|_2^2 + (1-\alpha)\mathrm{tr}(\mathbf{Z}^T \boxed{\gamma_\theta(\hat{\mathbf{L}})}\mathbf{Z})$$

Arbitrary Linking Patterns

- $\mathbf{Z}$ refer to node representations
- α is a trade-off coefficient
- $\gamma_\theta(\hat{\mathbf{L}}) = \mathbf{U}\gamma_\theta(\mathbf{\Lambda})\mathbf{U}^T$ determines propagation rate where $\gamma_\theta(\lambda) \geq 0$

Positive Semi-definite Constraint for Convexity Optimization

Jingwei Guo, et.al. Rethinking Spectral Graph Neural Networks with Spatially Adaptive Filtering. *TNNLS* (Under Review), 2024.

Research Experience — Solution 4

Deep Delve into Spectral GNNs

- Closed-form Solution $\frac{\partial \mathscr{L}}{\partial \mathbf{Z}} = 0$

Spectral Filtering

$$\mathbf{Z}^* = (\mathbf{I} + \frac{1-\alpha}{\alpha}\gamma_\theta(\hat{\mathbf{L}}))^{-1}\mathbf{X} = g_\psi(\hat{\mathbf{L}})\mathbf{X} = \mathbf{U}g_\psi(\mathbf{\Lambda})\mathbf{U}^T\mathbf{X}$$

Spectral Filter as a function of $\gamma_\theta(\cdot)$: $g_\psi(\lambda) = (1 + \frac{1-\alpha}{\alpha}\gamma_\theta(\lambda))^{-1}$

- Iterative Solution $\mathbf{Z}^{(k)} = \mathbf{Z}^{(k-1)} - \frac{1}{2}\frac{\partial \mathscr{L}}{\partial \mathbf{Z}}|_{\mathbf{Z}=\mathbf{Z}^{(k-1)}}$

Spatial Aggregation

$$\mathbf{Z}^{(k)} = \alpha\mathbf{X} + (1-\alpha)\hat{\mathbf{A}}^{new}\mathbf{Z}^{(k-1)}$$

New Computation Graph: $\hat{\mathbf{A}}^{new} = \mathbf{I} - \gamma_\theta(\hat{\mathbf{L}}) = \mathbf{I} - \frac{\alpha}{1-\alpha}(g_\psi(\hat{\mathbf{L}})^{-1} - \mathbf{I})$

Jingwei Guo, et.al. Rethinking Spectral Graph Neural Networks with Spatially Adaptive Filtering. *TNNLS* (Under Review), 2024.

Research Experience — Solution 4

Spatially Adaptive Filtering

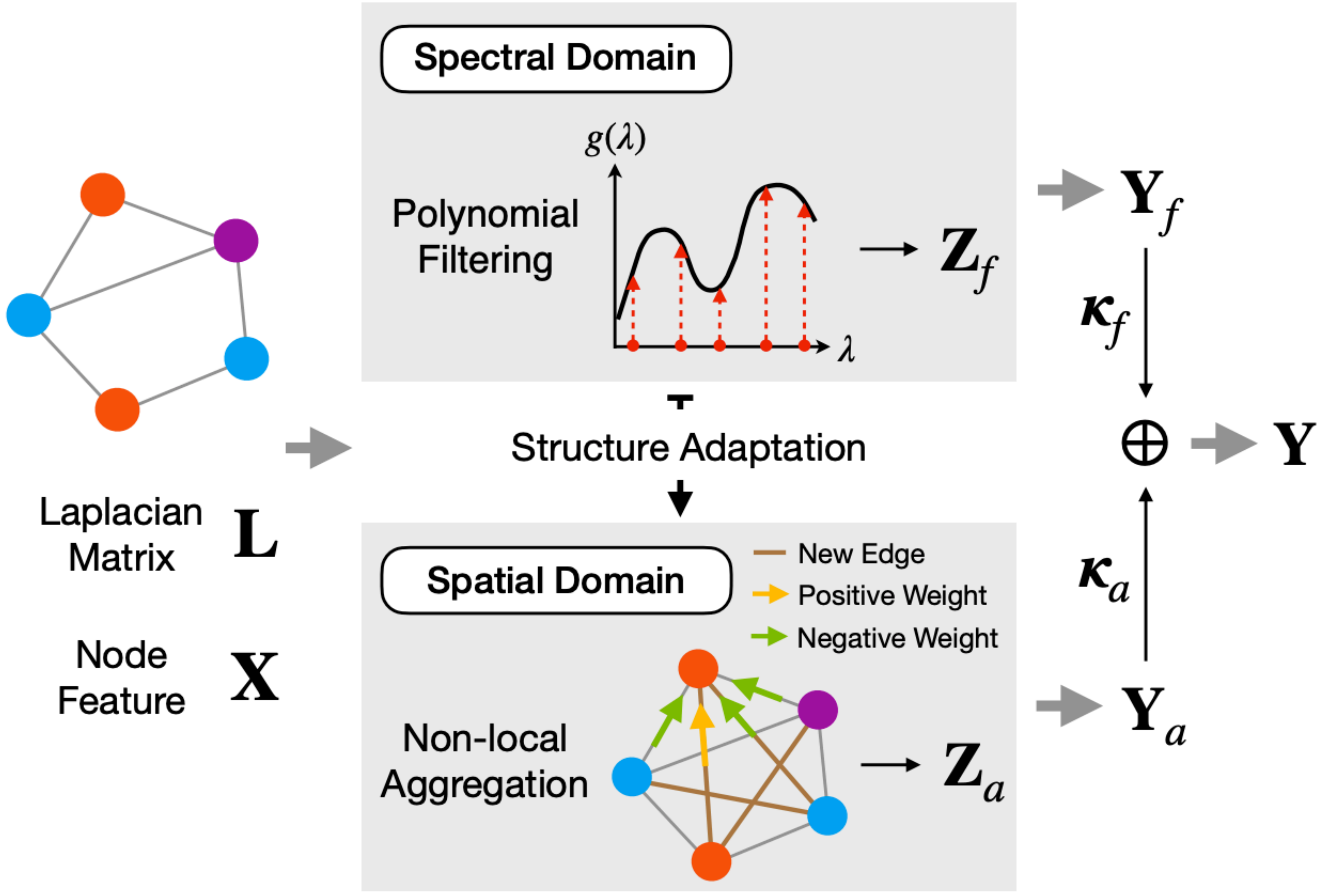


Figure 3: Illustration of the proposed SAF framework, where varying node colors represent different node labels.

SAF leverages the adapted new graph by spectral filtering for non-local aggregation with signed weights.

Address:

- Long-range Dependency
- Heterophilic Linking Patterns

Jingwei Guo, et.al. Rethinking Spectral Graph Neural Networks with Spatially Adaptive Filtering. *TNNLS* (Under Review), 2024.

Research Experience — Solution 4

Spatially Adaptive Filtering

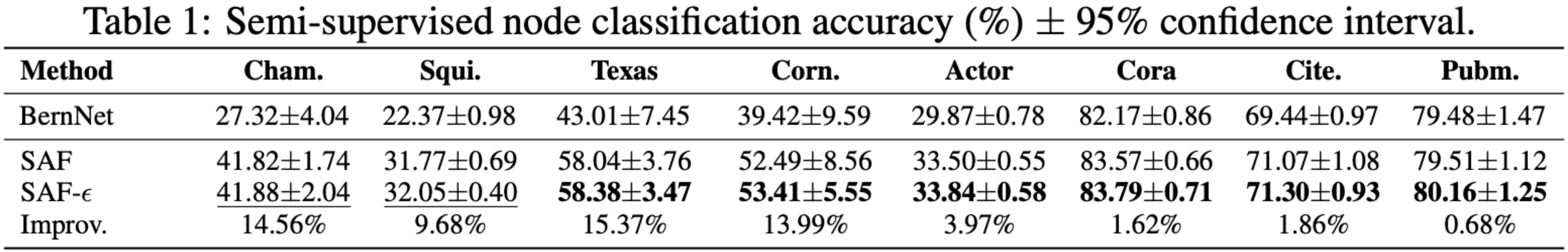


Table 1: Semi-supervised node classification accuracy (%) ± 95% confidence interval.

Method	**Cham.**	**Squi.**	**Texas**	**Corn.**	**Actor**	**Cora**	**Cite.**	**Pubm.**
BernNet	27.32±4.04	22.37±0.98	43.01±7.45	39.42±9.59	29.87±0.78	82.17±0.86	69.44±0.97	79.48±1.47
SAF	41.82±1.74	31.77±0.69	58.04±3.76	52.49±8.56	33.50±0.55	83.57±0.66	71.07±1.08	79.51±1.12
SAF-ϵ	41.88±2.04	32.05±0.40	**58.38±3.47**	**53.41±5.55**	**33.84±0.58**	**83.79±0.71**	**71.30±0.93**	**80.16±1.25**
Improv.	14.56%	9.68%	15.37%	13.99%	3.97%	1.62%	1.86%	0.68%

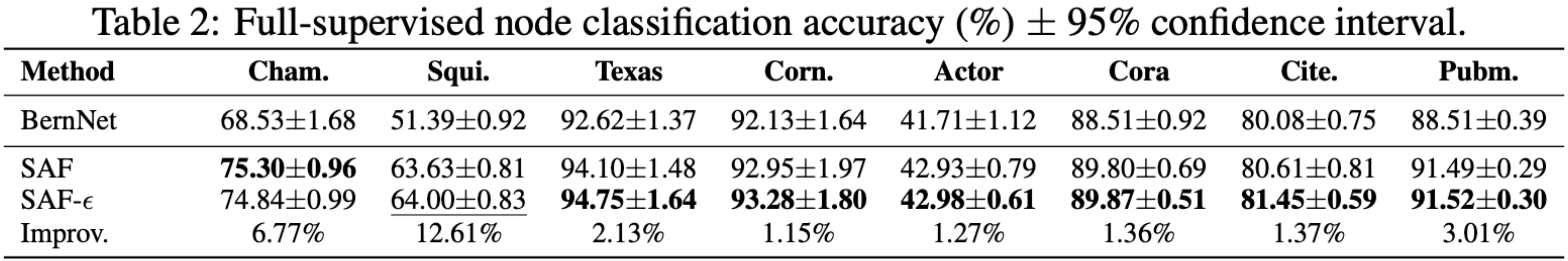


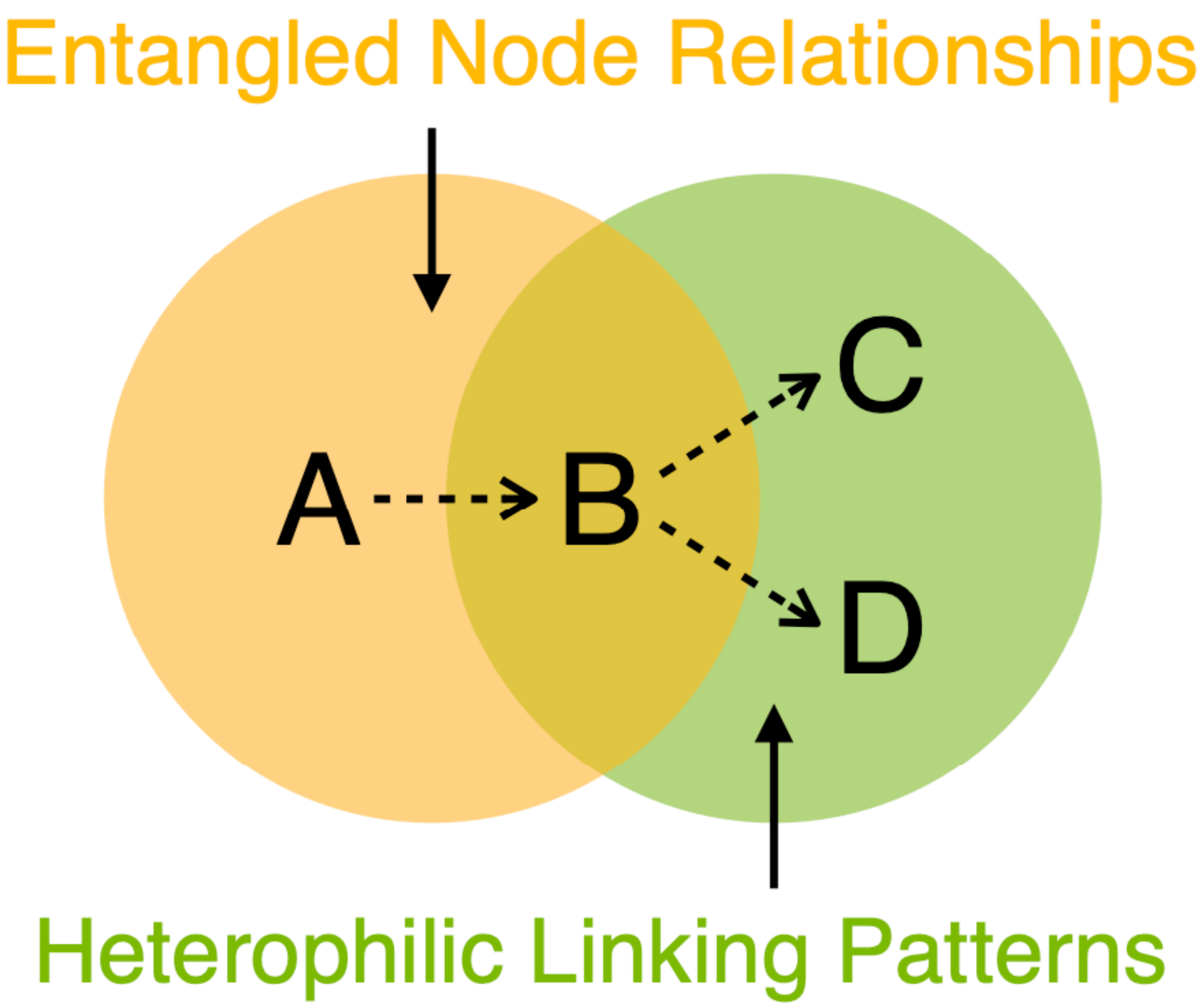
Table 2: Full-supervised node classification accuracy (%) ± 95% confidence interval.

Method	**Cham.**	**Squi.**	**Texas**	**Corn.**	**Actor**	**Cora**	**Cite.**	**Pubm.**
BernNet	68.53±1.68	51.39±0.92	92.62±1.37	92.13±1.64	41.71±1.12	88.51±0.92	80.08±0.75	88.51±0.39
SAF	**75.30±0.96**	63.63±0.81	94.10±1.48	92.95±1.97	42.93±0.79	89.80±0.69	80.61±0.81	91.49±0.29
SAF-ϵ	74.84±0.99	64.00±0.83	**94.75±1.64**	**93.28±1.80**	**42.98±0.61**	**89.87±0.51**	**81.45±0.59**	**91.52±0.30**
Improv.	6.77%	12.61%	2.13%	1.15%	1.27%	1.36%	1.37%	3.01%

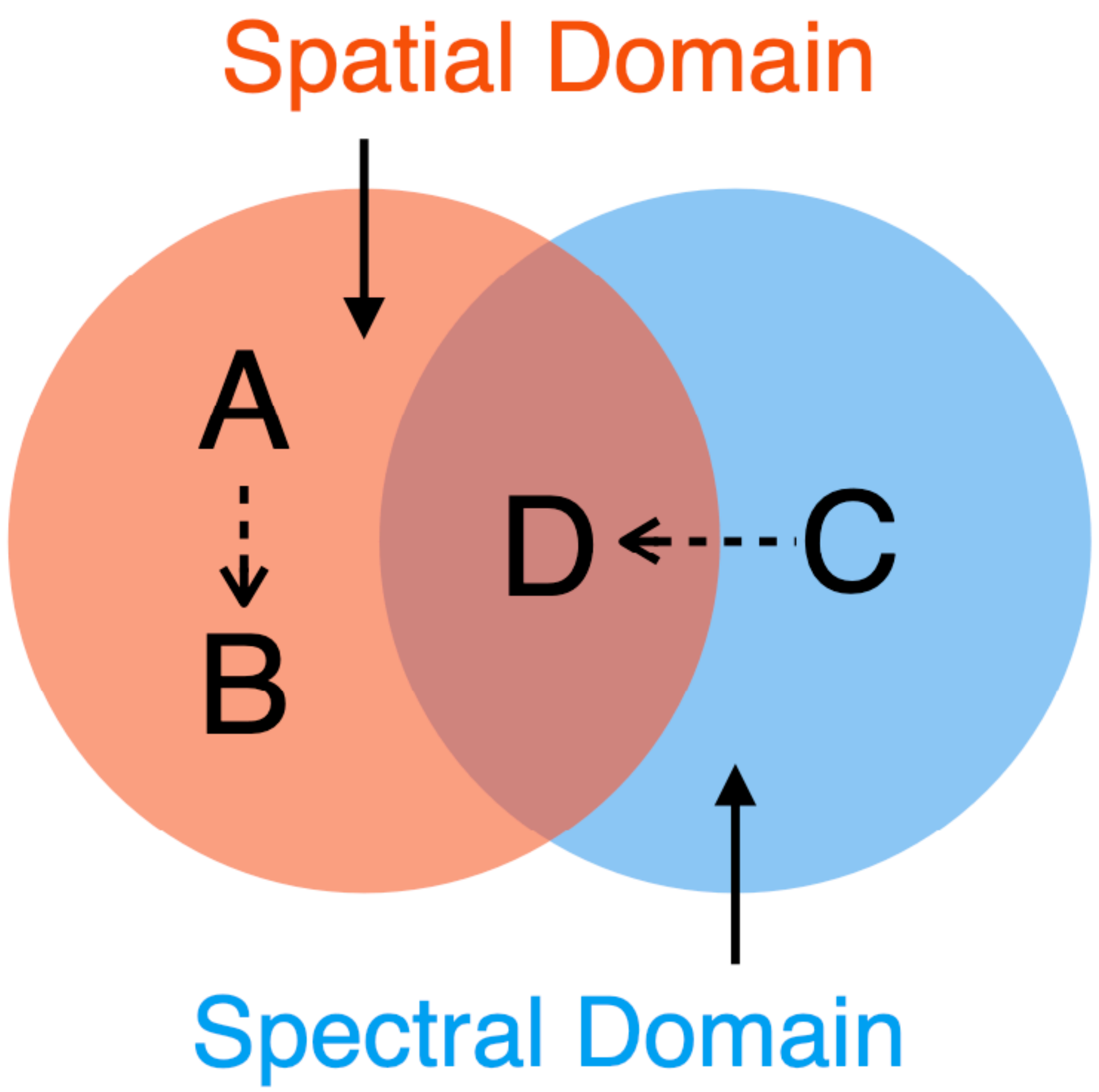
Jingwei Guo, et.al. Rethinking Spectral Graph Neural Networks with Spatially Adaptive Filtering. *TNNLS* (Under Review), 2024.

Research Experience — Conclusion

Relations Among the Developed Models



Problem-Centric



Model-Centric

A: LGD (1st)
B: ES-GNN (2nd)
C: DSF (3rd)
D: SAF (4th)

Thanks!