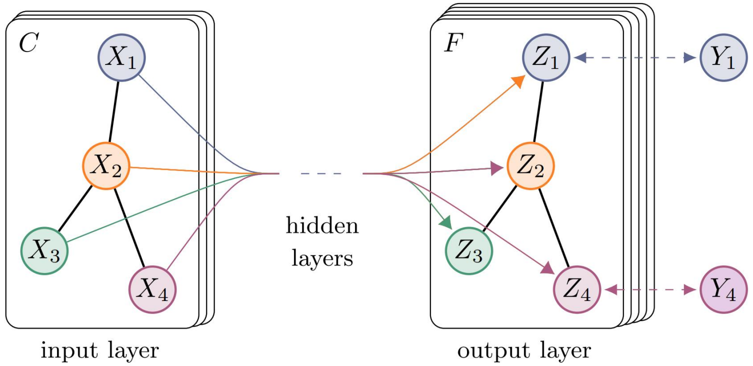
Enhancing the Influence of Labels on Unlabeled Nodes in Graph Convolutional Networks

Jincheng Huang[1], Yujie Mo[1], Xiaoshuang Shi[1], Lei Feng[2], Xiaofeng Zhu[1]*

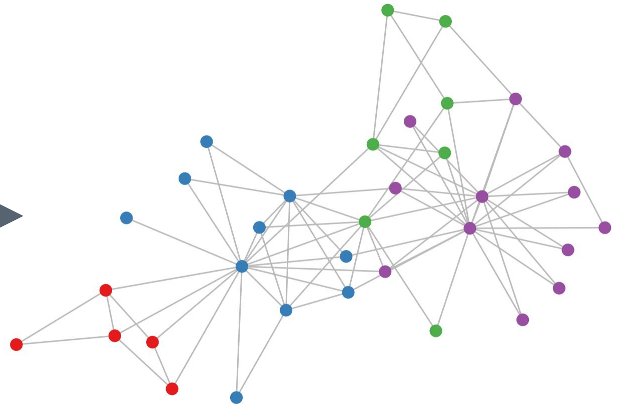
1.University of Electronic Science and Technology of China 2. Southeast University

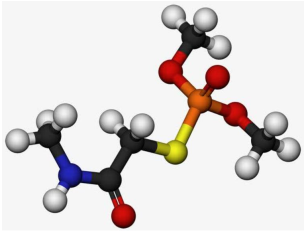
Background

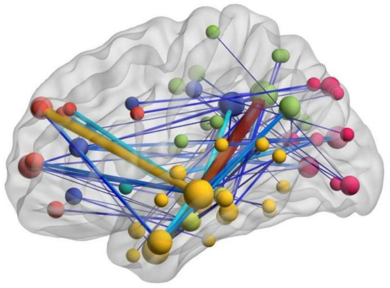
Graph Convolutional Networks (GCNs)



Message passing on graphs





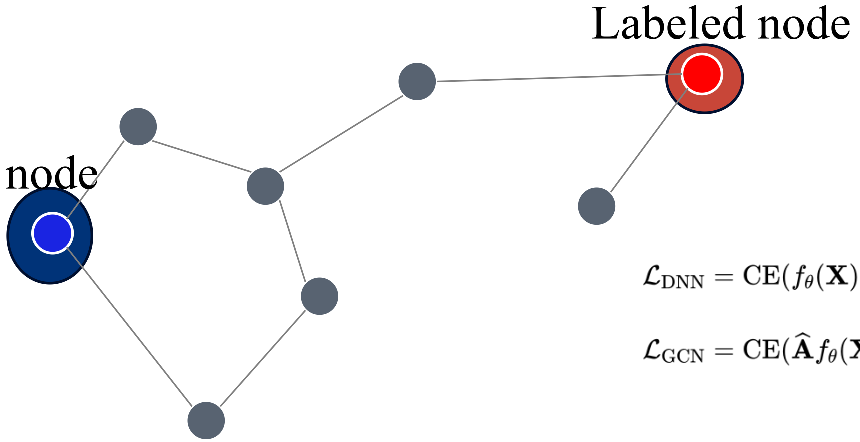


Effectively handle the graph data

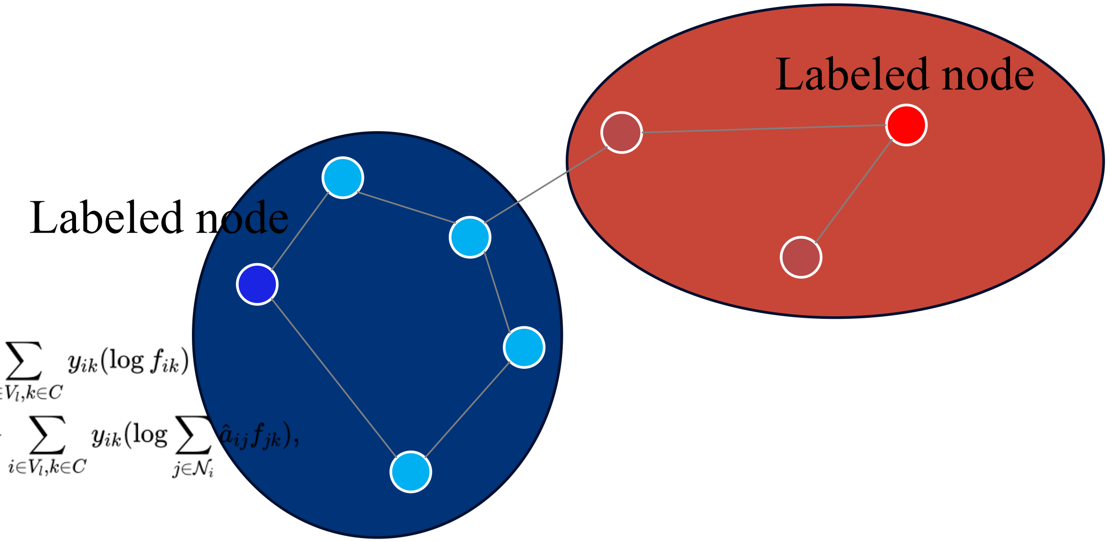
GCNs excel at handling graph-structured data, which relies most on their message passing mechanism

Motivation

Effect of message passing



DNN: One-to-one supervision



$$\mathcal{L}_{\text{DNN}} = \text{CE}(f_\theta(\mathbf{X}), \mathbf{Y}) = -\sum_{i \in V_l, k \in C} y_{ik}(\log f_{ik})$$

$$\mathcal{L}_{\text{GCN}} = \text{CE}(\widehat{\mathbf{A}} f_\theta(\mathbf{X}), \mathbf{Y}) = -\sum_{i \in V_l, k \in C} y_{ik}(\log \sum_{j \in \mathcal{N}_i} \hat{a}_{ij} f_{jk}),$$

GCN: One-to-many supervision

$$\mathcal{L}_{\text{DNN}} = \text{CE}(f_\theta(\mathbf{X}), \mathbf{Y}) = -\sum_{i \in V_l, k \in C} y_{ik}(\log f_{ik})$$

$$\mathcal{L}_{\text{GCN}} = \text{CE}(\widehat{\mathbf{A}} f_\theta(\mathbf{X}), \mathbf{Y}) = -\sum_{i \in V_l, k \in C} y_{ik}(\log \sum_{j \in \mathcal{N}_i} \hat{a}_{ij} f_{jk}),$$

Effect of Message Passing

Message-passing of GCN makes unlabeled nodes can utilizing the label information.

However, is the GCN effectively utilizing label information for unlabeled nodes?

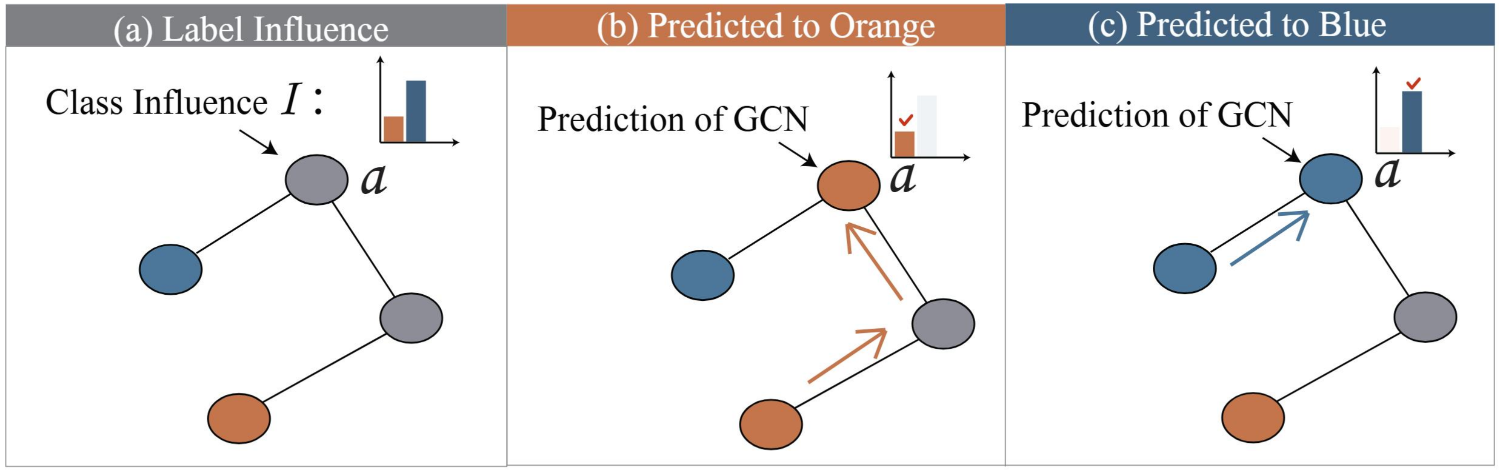
Motivation

电子科技大学
University of Electronic Science and Technology of China

Impact of label on unlabeled nodes under the message passing

LPA can be as a metric

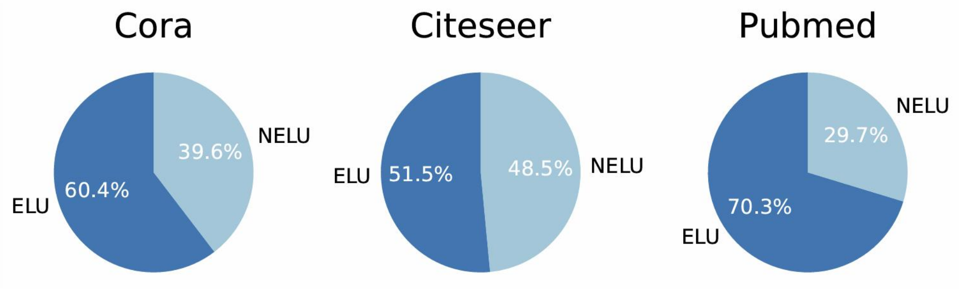
Proposition 1: LPA can be utilized to calculate the probability of every class for unlabeled nodes in the GCN framework.
i.e., The output of LPA represents the class that provides the most label information to a node.



GCN is considered to have effectively utilized the label information if the predicted class aligns with the class that contributes the most label information through message passing.

$$V_{\mathrm{ELU}} = \{V | \mathrm{LPA}(\mathcal{G}) = \mathrm{GCN}(\mathcal{G})\}$$

$$V_{\mathrm{NELU}} = \{V | \mathrm{LPA}(\mathcal{G}) \neq \mathrm{GCN}(\mathcal{G})\}$$



Method

Our goal is

Label influence: $\mathbf{Q} = \mathbf{S}\mathbf{Y}$

GCN prediction: $\hat{\mathbf{Y}} = \mathbf{S}\mathbf{H}, \quad s.t. \quad \mathbf{H} = \mathrm{MLP}(\mathbf{X})$

Objective function: $\min \left\| \mathbf{Q} - \hat{\mathbf{Y}} \right\|_F^2 = \min_{\mathbf{S}} \|\mathbf{S}\mathbf{Y} - \mathbf{S}\mathbf{H}\|_F^2$

We construct a New Traditional graph learning Objective Function

Many existing works are built upon traditional graph learning objective functions, while our approach may offer a new direction.

	Objective Functions	Input
Wang et al.,	$\min_{S} \|\mathbf{S}\mathbf{X} - \mathbf{X}\|_F^2$	X
Daitch et al.,	$\min_{\mathbf{S}} \sum_i \left\| \mathbf{D}_{i,i}\mathbf{X}_i - \sum_j \mathbf{S}_{i,j}\mathbf{X}_j \right\|^2$	X
Zhang et al.,	$\min_{\mathbf{S}} \frac{1}{2} \sum_{i,j} \mathbf{S}_{i,j} \|\mathbf{X}_i - \mathbf{X}_j\|^2$	X
Jiang et al.,	$\min_{\mathbf{L}} \mathrm{tr}\left(\mathbf{X}^\top \mathbf{L}\mathbf{X}\right)$	X
Ours	$\min_{\mathbf{S}} \|\mathbf{S}\mathbf{Y} - \mathbf{S}\mathbf{H}\|_F^2.$	**X,Y**

Method

University of Electronic Science and Technology of China

How to solve our objective function

$$\min \left\| \mathbf{Q} - \hat{\mathbf{Y}} \right\|_F^2 = \min_{\mathbf{S}} \|\mathbf{SY} - \mathbf{SH}\|_F^2$$

We cannot directly optimize this objective function, because:

1. it has a trivial solution: $s_{i,j} = 0, \forall i, \forall j$
2. LPA should reinitialize the labels to avoid modifying the ground-truth.

Iteratively update solution

Step 1: Propagate labels

1. We first calculate LPA: $\mathbf{Q}^{(i)} = \mathbf{S}^{(i-1)}\mathbf{Q}^{(i-1)}, (i = 1, \ldots, k)$, where $\mathbf{Q}^{(0)} = \mathbf{Y}$
2. Then, the objective function is changed: $\min_{\mathbf{S}} \left\| \mathbf{Q}^{(i)} - \mathbf{SH} \right\|_F^2 + \beta \sum_{i,j=1} s_{i,j}^2, s.t.\ \mathbf{Q}_l^{(i)} = \mathbf{Y}_l$
   1. Solvable
   2. Reinitialize the labels

Step 2: Re-computing adjacency matrix

1. Get $\mathbf{S}^{(i)}$ by calculating the closed-form solution of the objective function: $\mathbf{S}^{(i)} = \mathbf{Q}^{(i)}\mathbf{H}^T\left(\mathbf{HH}^T + \beta\mathbf{I}_N\right)^{-1}$

Iterate steps 1 and 2 to solve the final S

Algorithm 2 Pseudo code of calculating $\mathbf{S}^*$.

```
1: for i ← 1, 2, ··· , k do
2:     Calculate Q^(i) by Step 1;
3:     Calculate S^(i) by Step 2;
4: end for
```

Method

University of Electronic Science and Technology of China

Model Efficiency

The time complexity of calculate $\mathbf{S}^{(i)} = \mathbf{Q}^{(i)}\mathbf{H}^T\left(\mathbf{H}\mathbf{H}^T + \beta\mathbf{I}_N\right)^{-1}$ is $O(n^3)$!!!

Calculation Trick

We can avoid to calculation $\mathbf{S}^{(i)}$, directly substitute $\mathbf{S}^{(i)}$ in the second step into the first step to avoid the output being an n*n matrix and use the Woodbury identity trick, the finally time complexity change to $O(nc^2 + c^3)$, $c \ll n$.

$$\mathbf{Q}^{(i)} = \mathbf{S}^{(i-1)}\mathbf{Q}^{(i-1)} = \mathbf{Q}^{(i-1)}\mathbf{H}^T\left(\frac{1}{\beta}\mathbf{I}_N - \frac{1}{\beta^2}\mathbf{H}\left(\mathbf{I}_c + \frac{1}{\beta}\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\right)\mathbf{Q}^{(i-1)}, s.t.\ \mathbf{Q}_l^{(i)} = \mathbf{Y}_l,$$

Solution Process Changes

Algorithm 2 Pseudo code of calculating $\mathbf{S}^*$.

1: **for** i $\leftarrow 1, 2, \cdots, k$ **do**
2: Calculate $\mathbf{Q}^{(i)}$ by **Step 1**;
3: Calculate $\mathbf{S}^{(i)}$ by **Step 2**;
4: **end for**

→

Algorithm 2 Pseudo code of calculating $\mathbf{S}^*$.

1: **for** i $\leftarrow 1, 2, \cdots, k$ **do**
2: Calculate $\mathbf{Q}^{(i)}$ by above equation;
3: **end for**
4: Calculate $\mathbf{S}^* = \mathbf{Q}^{(k)}(\frac{1}{\beta}\mathbf{H}^T - \frac{1}{\beta^2}\mathbf{H}^T\mathbf{H}\left(\mathbf{I}_c + \frac{1}{\beta}\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T).$;

Contrastive Constrain

How to fully use the original graph information and the ELU-graph?

Most common used and efficient strategy:

$$\widehat{\mathbf{Y}} = \mathrm{Softmax}((1-\eta)\mathbf{H}$$

On this basis, we fist learn a projection head p_{θ}, $\mathbf{P}$

$$\mathcal{L}_{\mathrm{con}} = -\log \frac{\mathrm{Pos}}{\mathrm{Pos}+\mathrm{Neg}} - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{c}\hat{y}_{i,j}\log\hat{y}_{i,j} \quad s.t. \left\{ \mathrm{Pos} = \frac{1}{|V_{\mathrm{ELU}}|}\sum_{i=0}^{V_{\mathrm{ELU}}}\exp(d(\mathbf{P} \right.$$

The final objective function:

$$\mathcal{L} = CE(\widehat{\mathbf{Y}}, \mathbf{Y}) + \lambda\mathcal{L}_{con}$$

Theoretical Analysis

Theorem 3.3. *Given a graph $\mathcal{G}$ with its adjacency matrix $\mathbf{A}$, the label matrix in the training set $\mathbf{Y}$ and the label matrix of the ground truth $\mathbf{Y}_{\text{true}}$, for any unlabeled nodes, if a graph structure makes the labels in training set be consistent to the ground truth,* i.e., $\mathbf{Y}_{\text{true}} = \mathbf{A}\mathbf{Y}$, *then the upper bound of the generalization ability of the GCN is optimal.*

Theorem 3.4. *The optimization Eq. (5) is equivalent to an approximate optimization of* $\min_{\mathbf{A}} \|\mathbf{A}\mathbf{Y} - \mathbf{Y}_{\text{true}}\|_F^2$.

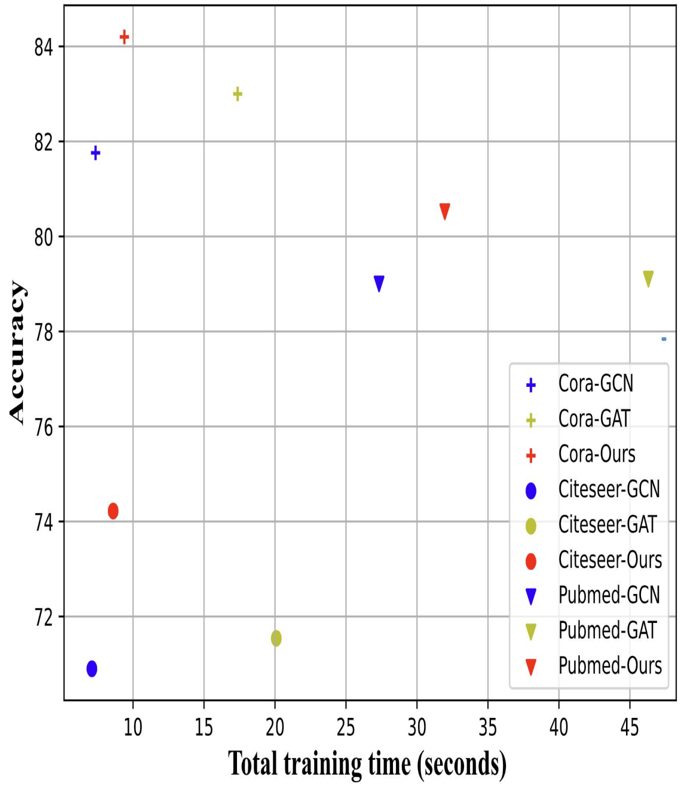
Experiments

University of Electronic Science and Technology of China

Effectiveness and Efficiency

Table 1. Performance on node classification task. The highest results are highlighted in bold. ”OOM” denotes out of memory.

Method	Cora	Citeseer	pubmed	Computers	Photo	Chameleon	squirrel
GCN	81.61±0.42	70.35±0.45	79.01±0.62	81.62±2.43	90.44±1.23	60.82±2.24	43.43±2.18
GAT	83.03±0.71	71.54±1.12	79.17±0.38	78.01±19.1	85.71±20.3	40.72±1.55	30.26±2.50
APPNP	83.33±0.62	71.80±0.84	80.10±0.21	82.12±3.13	88.63±3.73	56.36±1.53	46.53±2.18
GPRGNN	80.55±1.05	68.57±1.22	77.02±2.59	81.71±2.84	91.23±2.59	46.85±1.71	31.61±1.24
PCNet	82.81±0.50	69.92±0.70	80.01±0.88	81.82±2.31	89.63±2.41	59.74±1.43	48.53±1.12
GCN-LPA	83.13±0.51	72.60±0.80	78.64±1.32	83.54±1.41	90.13±1.53	50.26±1.38	42.78±2.36
N.S.-GCN	82.12±0.14	71.55±0.14	79.14±0.12	81.16±1.53	89.86±1.86	55.37±1.64	46.86±2.02
PTDNet-GCN	82.81±0.23	72.73±0.18	78.81±0.24	82.21±2.13	90.23±2.84	53.26±1.44	41.96±2.16
CoGSL	81.76±0.24	72.79±0.42	OOM	OOM	89.63±2.24	52.23±2.03	39.96±3.31
NodeFormer	80.28±0.82	71.31±0.98	78.21±1.43	80.35±2.75	89.37±2.03	34.71±4.12	38.54±1.51
GSR	83.08±0.48	72.10±0.25	78.09±0.53	81.63±1.35	90.02±1.32	62.28±1.63	50.53±1.93
BAGCN	83.70±0.21	72.96±0.75	78.54±0.72	79.63±2.52	**91.25±0.96**	52.63±1.78	42.36±1.53
ELU-GCN	**84.29±0.39**	**74.23±0.62**	**80.51±0.21**	**83.73±2.31**	90.81±1.33	**70.90±1.76**	**56.91±1.81**



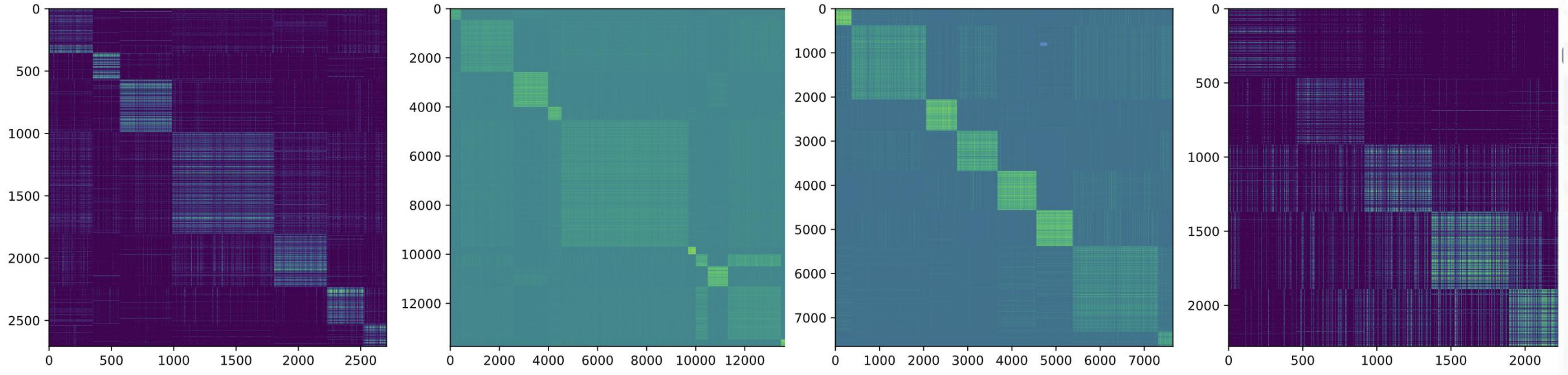
Experiments

电子科技大学
University of Electronic Science and Technology of China

Ablation Study and Visualization of ELU Graph

Table 2. Classification performance of each component in the proposed method on all datasets.

Method	Cora	Citeseer	pubmed	Computers	Photo	Chameleon	squirrel
GCN	81.61±0.42	70.35±0.45	79.01±0.62	81.62±2.43	90.44±1.23	60.82±2.24	43.43±2.18
+ELU graph	83.49±0.55	72.02±0.36	80.25±0.79	82.56±1.23	90.52±1.33	65.12±1.43	54.12±1.32
+$\mathcal{L}_{con}$	**84.29**±0.39	**74.23**±0.62	**80.51**±0.21	**83.73**±2.31	**90.81**±1.33	**70.90**±1.76	**56.91**±1.81



(a) Cora (b) Computers (c) Photo (d) Chameleon

Figure 3. Visualization of the adjacency matrix of the ELU graph on Cora, Computers, Photo, and Chameleon datasets. The rows and columns are nodes that are reordered based on node labels, the lighter a pixel, the larger the value of the ELU graph matrix weight.

Summary

Conclusion

- We propose a new objective function that optimizes the graph structure to ensure that GCNs effectively utilize label information for unlabeled nodes.
  - This objective function may inspire more supervised graph structure learning method.
- we design a contrastive loss to capture consistency or mutually exclusive information between the original graph and the ELU graph.
  - The proposed contrasting form of one-to-one and the traditional form of one-to-many can be further explored and interacted.
- Theoretical contribution: Provides guidance on improving generalization ability through graph structure learning.
  - The accuracy of label propagation is an important indicator of graph structure quality.

电子科技大学
University of Electronic Science and Technology of China

T h a n k s !