Compact Graph Structure Learning via Mutual Information Compression

Nian Liu
nianliu@bupt.edu.cn
Beijing University of Posts and Telecommunications
China

Xiao Wang
xiaowang@bupt.edu.cn
Beijing University of Posts and Telecommunications
Peng Cheng Laboratory
China

Lingfei Wu
lwu@email.wm.edu
JD.COM Silicon Valley Research Center
United States

Yu Chen
hugochen@fb.com
Facebook AI
United States

Xiaojie Guo
xguo7@gmu.edu
JD.COM Silicon Valley Research Center
United States

Chuan Shi*
shichuan@bupt.edu.cn
Beijing University of Posts and Telecommunications
Peng Cheng Laboratory
China

WWW 2022
Code: github.com/liun-online/CoGSL

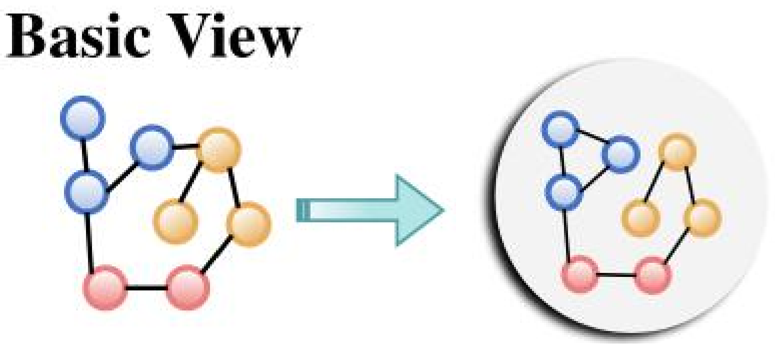
2022.03.20 • ChongQing

Leibniz-Institut
für Sozialwissenschaften

Reported by Chenghong Li

Introduction

- An optimal graph structure should only contain the information about tasks while compress redundant noise as much as possible.



Method

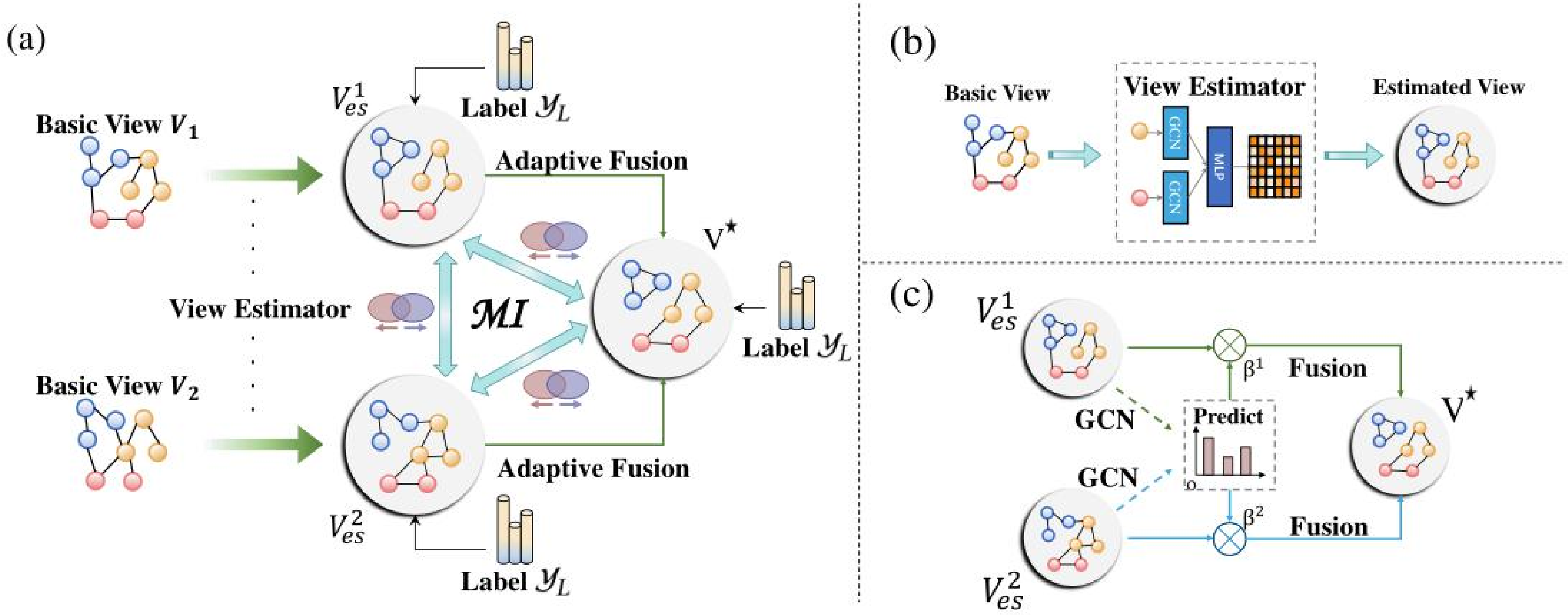


Figure 1: The overview of our proposed CoGSL. (a) Model framework. (b) View estimator. (c) Adaptive fusion.

Basic views:

(1) Adjacency matrix

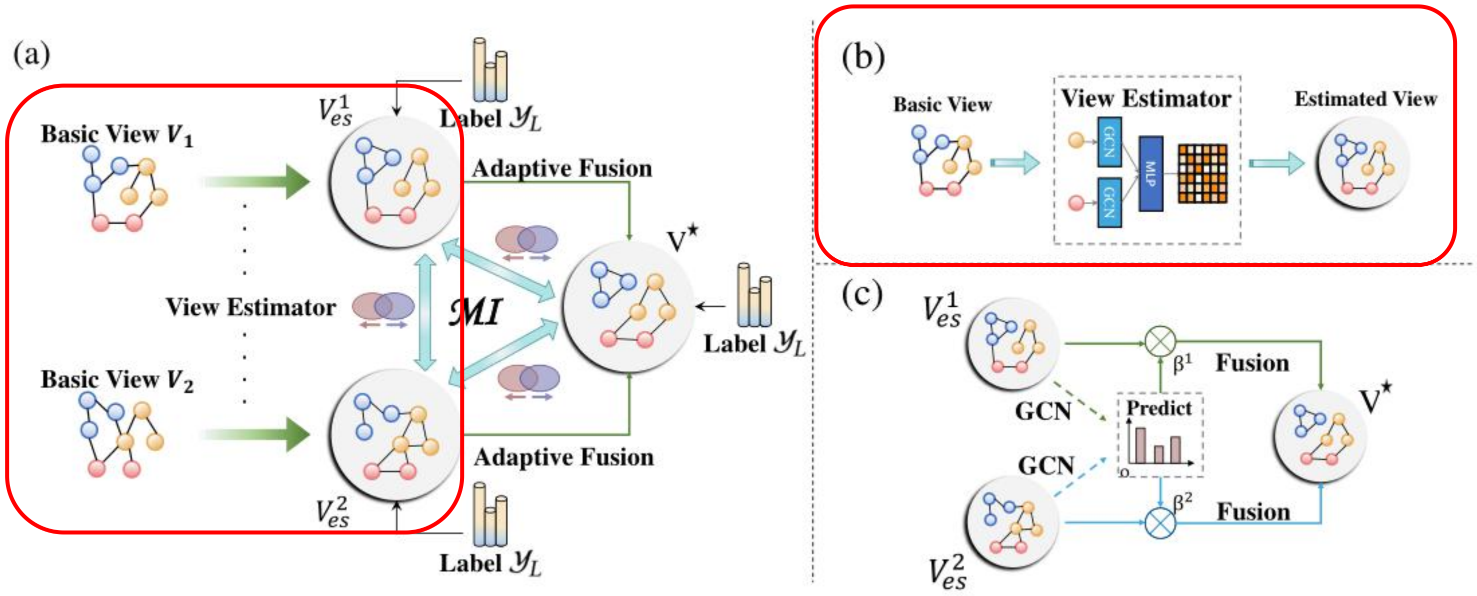
(2) Diffusion matrix $S = \alpha(I - (1-\alpha)D^{-1/2}AD^{-1/2})^{-1}$, where $\alpha \in (0,1]$

(3) Subgraph

(4) KNN graph

2020_ICML_Contrastive Multi-View Representation Learning on Graphs

Method



$$GCN(A, H^{(k)}) = D^{-1/2}AD^{-1/2}H^{(k-1)}W^k, \quad (1)$$

$$Z^1 = \sigma(GCN(V_1, X)), \quad (2)$$

$$Z^1 \in \mathbb{R}^{N\times d_{es}}$$

$$w_{ij}^1 = \cancel{W_1} \cdot [z_i^1 || z_j^1] W_1 + b_1, \quad (3)$$

$$W_1 \in \mathbb{R}^{2d_{es}\times 1}$$

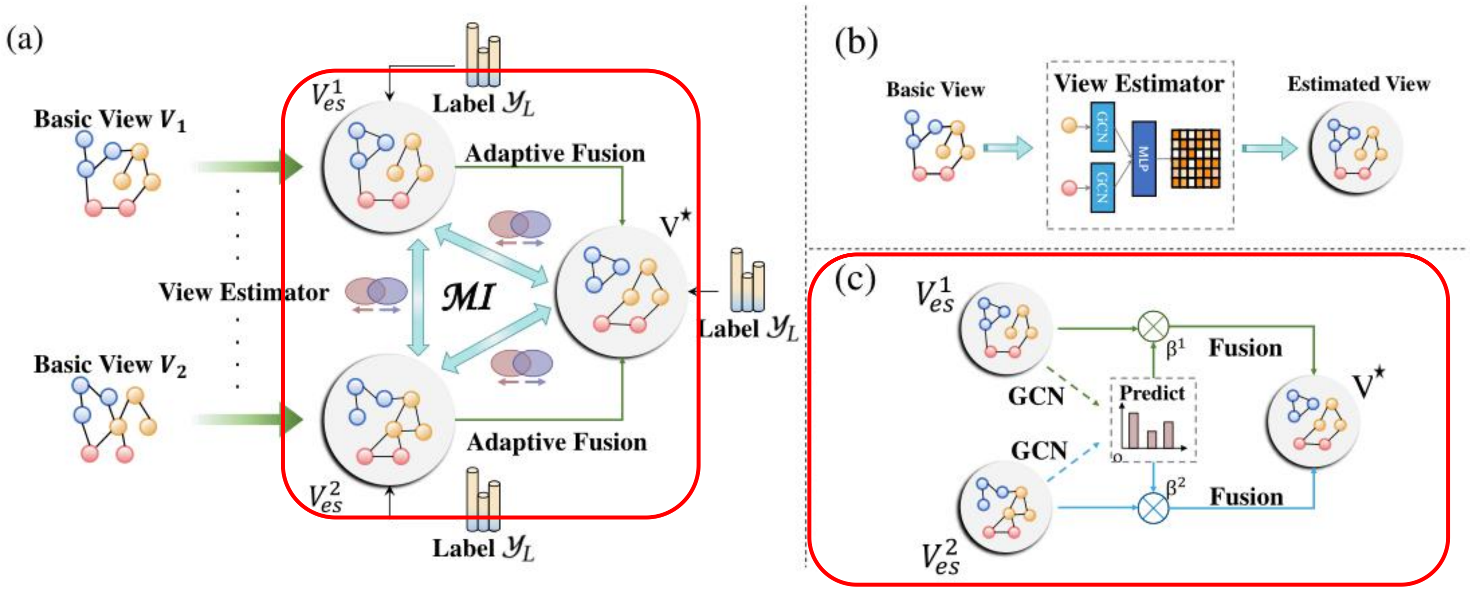
$$b_1 \in \mathbb{R}^{\cancel{2d_{es}\times 1}}$$

$$p_{ij}^1 = \frac{\exp(w_{ij}^1)}{\sum_{k\in S^1} \exp(w_{ik}^1)}. \quad (4)$$

only estimate limited scope S^1

$$V_{es}^1 = V_1 + \mu^1 \cdot P^1, \quad (5)$$

Method



$$O^1 = softmax(GCN(\boldsymbol{V_{es}^1},\ \sigma(GCN(\boldsymbol{V_{es}^1}, \boldsymbol{X})))),$$
$$O^2 = softmax(GCN(\boldsymbol{V_{es}^2},\ \sigma(GCN(\boldsymbol{V_{es}^2}, \boldsymbol{X})))), \tag{6}$$

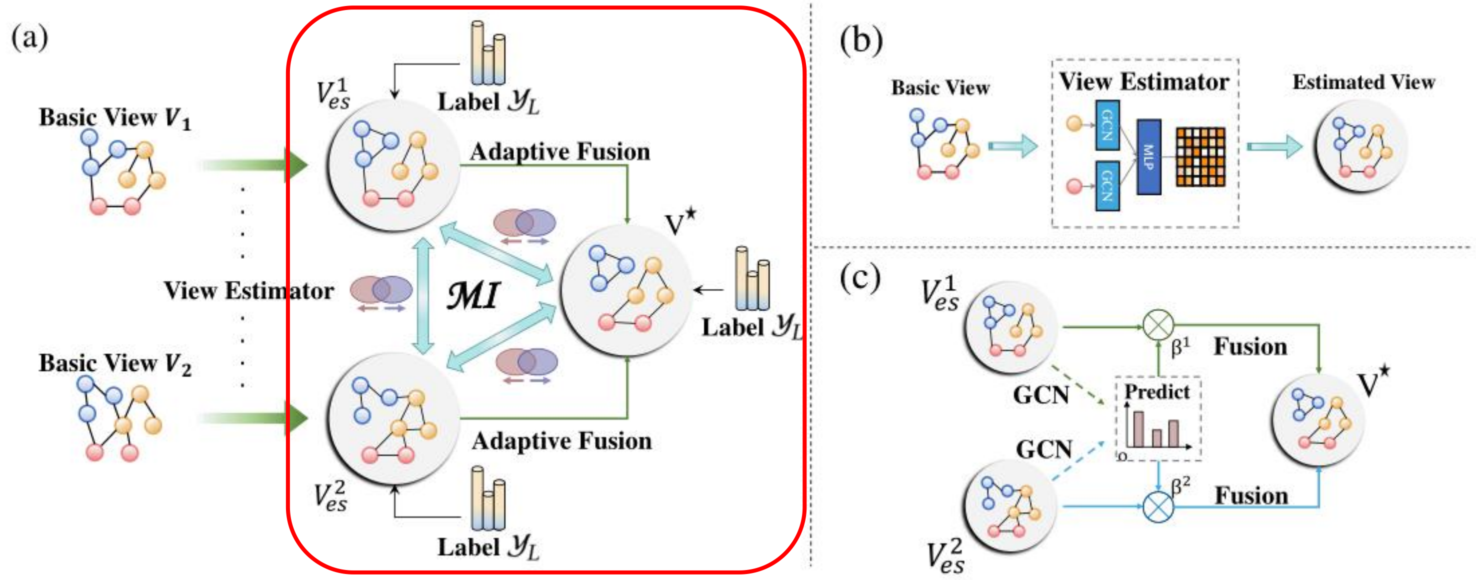
$$\pi_i^1 = e^{\epsilon\left(\lambda \log o_{i,m}^1 + (1-\lambda)\log(o_{i,m}^1 - o_{i,sm}^1)\right)}, \tag{7}$$

$$\beta_i^1 = \frac{\pi_i^1}{\pi_i^1 + \pi_i^2} \quad and \quad \beta_i^2 = \frac{\pi_i^2}{\pi_i^1 + \pi_i^2}. \tag{8}$$

$$\boldsymbol{V_i^\star} = \beta_i^1 \cdot \boldsymbol{V_{es\_i}^1} + \beta_i^2 \cdot \boldsymbol{V_{es\_i}^2}. \tag{9}$$

Method

$$O^{\star} = softmax(GCN(V^{\star},\ \sigma(GCN(V^{\star}, X)))). \tag{13}$$



$$\min_{\Theta} \mathcal{L}_{cls} = -\sum_{O \in \{O^1, O^2, O^{\star}\}} \sum_{v_i \in \mathcal{Y}_L} y_i \ln \boldsymbol{o}_i, \tag{14}$$

$$H^{\star} = \sigma(GCN(V^{\star}, X)), \tag{15}$$

$$H_p^{\star} = W^1 \cdot \sigma(W^0 \cdot H^{\star} + b^0) + b^1, \tag{16}$$

$$\begin{aligned} &L(V^{\star}, V_{es}^1) \\ &= -\frac{1}{2|B|} \sum_{i=1}^{|B|} \Bigg[\log \frac{e^{sim(h_{p_i}^{\star}, h_{p_i}^1)/\tau}}{e^{sim(h_{p_i}^{\star}, h_{p_i}^1)/\tau} + \sum_{k \neq i} e^{sim(h_{p_i}^{\star}, h_{p_k}^1)/\tau}} \\ &+ \log \frac{e^{sim(h_{p_i}^1, h_{p_i}^{\star})/\tau}}{e^{sim(h_{p_i}^1, h_{p_i}^{\star})/\tau} + \sum_{j \neq i} e^{sim(h_{p_i}^1, h_{p_j}^{\star})/\tau}} \Bigg], \end{aligned} \tag{17}$$

$$\mathcal{L}_{MI} = L(V^{\star}, V_{es}^1) + L(V^{\star}, V_{es}^2) + L(V_{es}^1, V_{es}^2). \tag{18}$$

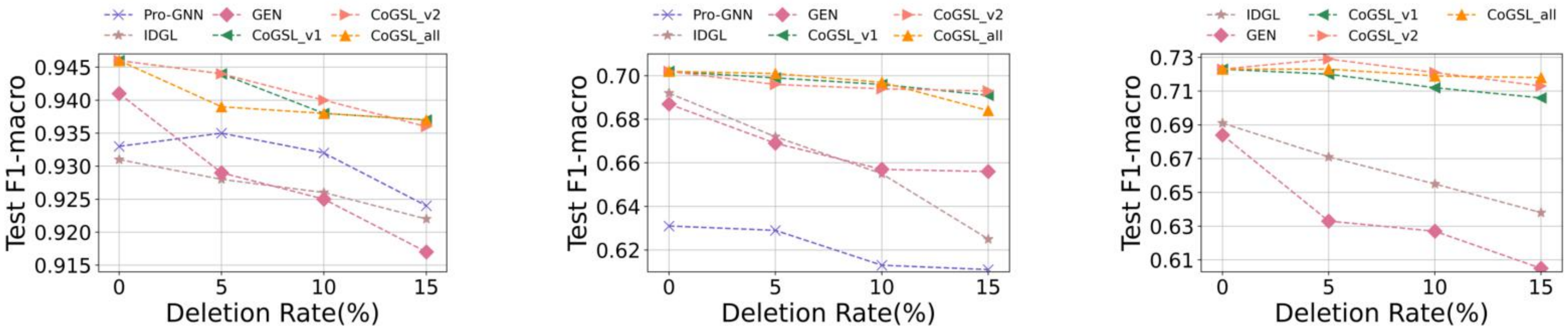
$$\min_{\Omega} \mathcal{L}_{cls} - \eta \cdot \mathcal{L}_{MI}, \tag{19}$$

Experiments

Table 1: Quantitative results (%±σ) on node classification.(bold: best; underline: runner-up)

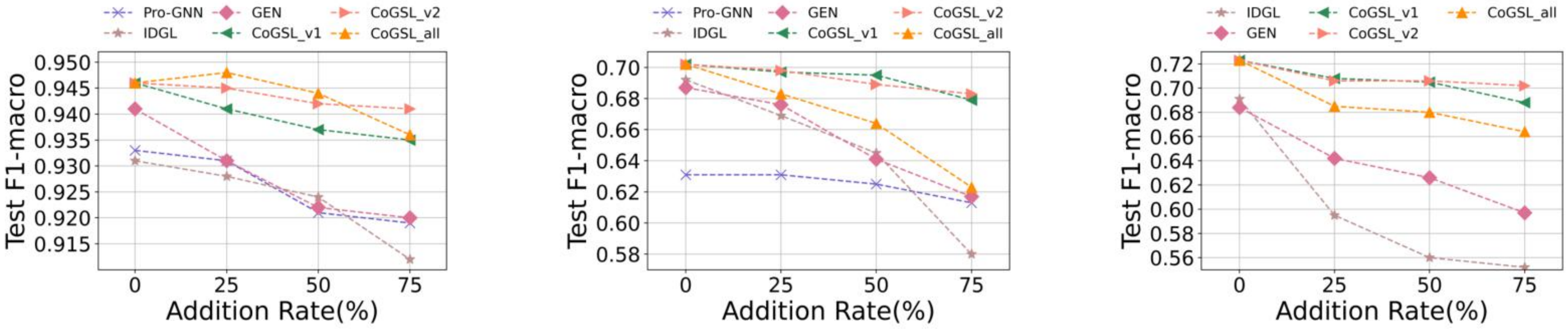
Datasets	Metric	DGI	GCA	GCN	GAT	GraphSAGE	LDS	Pro-GNN	IDGL	GEN	**CoGSL**
Wine	F1-macro	93.6±0.8	94.5±2.7	94.1±0.6	93.6±0.4	96.3±0.8	93.4±1.0	97.3±0.3	96.3±1.1	96.4±1.0	**97.9±0.3**
	F1-micro	93.6±0.8	94.6±2.4	93.9±0.6	93.7±0.3	96.2±0.8	93.4±0.9	97.2±0.3	96.2±1.1	96.3±1.0	**97.8±0.3**
	AUC	99.5±0.1	97.8±1.4	99.6±0.2	97.8±0.2	99.4±0.4	99.0±0.1	99.5±0.1	99.6±0.1	99.3±0.2	**99.7±0.1**
Cancer	F1-macro	85.7±1.9	93.4±1.2	93.0±0.6	92.2±0.2	92.0±0.5	83.1±1.5	93.3±0.5	93.1±0.9	94.1±0.8	**94.6±0.3**
	F1-micro	87.6±1.4	93.8±1.2	93.3±0.5	92.9±0.1	92.5±0.5	84.8±0.8	93.8±0.5	93.6±0.9	94.3±1.0	**95.0±0.3**
	AUC	95.2±2.4	97.9±0.6	**98.9±0.1**	96.9±0.3	96.9±0.5	90.6±0.9	97.8±0.2	98.1±0.3	98.3±0.3	98.5±0.1
Digits	F1-macro	88.9±0.8	89.5±1.4	89.0±1.3	89.9±0.2	87.5±0.2	79.7±1.0	89.7±0.3	92.5±0.5	91.3±1.3	**93.3±0.3**
	F1-micro	89.0±0.8	89.6±1.5	89.1±1.3	90.0±0.2	87.7±0.2	80.2±0.9	89.8±0.3	92.6±0.5	91.4±1.2	**93.3±0.3**
	AUC	99.0±0.1	97.6±0.3	98.9±0.2	98.3±0.4	98.7±0.1	95.1±0.1	98.1±0.2	99.4±0.1	98.4±0.9	**99.6±0.0**
Polblogs	F1-macro	90.9±0.4	95.0±0.2	95.1±0.4	94.1±0.1	93.3±2.5	94.9±0.3	94.6±0.6	94.6±0.7	95.2±0.6	**95.5±0.1**
	F1-micro	90.9±0.4	95.0±0.2	95.1±0.4	94.1±0.1	93.4±2.5	94.9±0.3	94.6±0.6	94.6±0.7	95.2±0.6	**95.5±0.1**
	AUC	96.4±0.3	98.2±0.2	**98.5±0.0**	97.4±0.1	98.1±0.1	98.1±0.4	98.3±0.2	98.2±0.2	98.0±0.6	98.3±0.1
Citeseer	F1-macro	68.1±0.6	60.9±0.9	67.4±0.3	68.4±0.2	67.1±0.8	69.4±0.7	63.1±0.7	69.2±0.9	68.7±0.5	**70.2±0.6**
	F1-micro	72.1±0.6	64.5±1.1	70.1±0.2	72.2±0.2	70.1±0.7	72.2±0.7	65.6±0.8	72.6±0.6	72.5±0.8	**73.4±0.8**
	AUC	90.8±0.1	88.5±0.7	89.9±0.2	90.2±0.1	90.5±0.3	91.3±0.3	88.2±0.3	91.1±0.4	88.4±0.5	**91.4±0.5**
Wiki-CS	F1-macro	56.4±0.1	67.1±1.3	68.8±1.7	70.1±0.1	69.2±0.9	54.6±0.5	63.8±2.0	69.1±1.1	68.4±0.3	**72.3±0.6**
	F1-micro	61.2±0.2	71.3±1.3	70.8±1.8	73.8±0.3	72.2±0.7	53.7±0.5	68.3±1.2	72.7±0.8	71.1±0.9	**75.0±0.3**
	AUC	91.8±0.1	93.2±0.4	95.2±0.3	95.6±0.1	95.0±0.3	88.8±2.1	93.3±0.3	92.0±0.2	91.6±1.2	**96.4±0.2**
MS Academic	F1-macro	88.6±0.2	87.0±1.6	89.4±0.6	86.7±0.6	88.9±0.4	-	-	89.6±0.6	89.8±0.8	**90.5±0.4**
	F1-micro	91.4±0.2	89.8±1.2	91.9±0.5	89.0±0.4	91.1±0.2	-	-	91.9±0.5	92.0±0.5	**92.4±0.5**
	AUC	99.1±0.1	99.3±0.2	99.4±0.1	99.2±0.1	99.4±0.0	-	-	**99.6±0.1**	98.8±0.3	99.4±0.1

Experiments



(a) Cancer (b) Citeseer (c) Wiki-CS

Figure 3: Results of different models under random edge deletion.



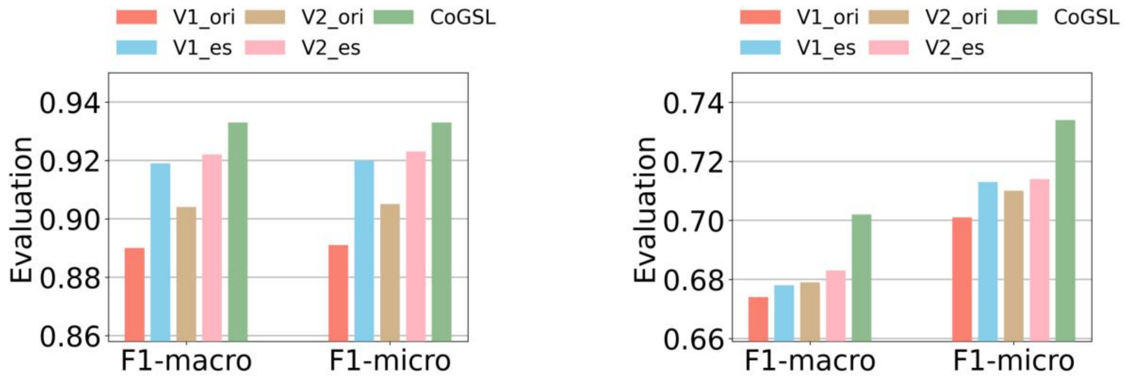
(a) Cancer (b) Citeseer (c) Wiki-CS

Figure 4: Results of different models under random edge addition.

Experiments

Table 2: Quantitative results under feature attack.

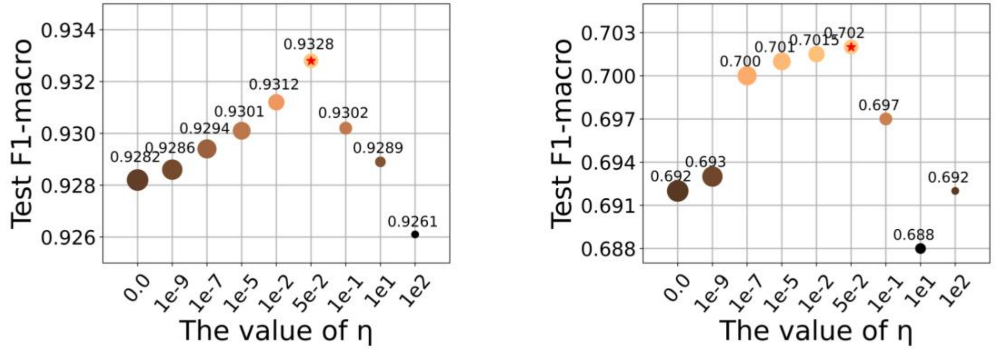
Datasets	~~F1-macro~~ x	Pro-GNN	IDGL	GEN	CoGSLL
Cancer	0.0	93.3	93.1	94.1	**94.6**
	0.1	92.9	91.5	92.9	**94.2**
	0.3	92.6	90.5	91.9	**93.6**
	0.5	92.2	90.2	90.9	**93.4**
Citeseer	0.0	63.1	69.2	68.7	**70.2**
	0.1	55.5	64.1	65.3	**67.8**
	0.3	44.1	22.6	36.1	**49.1**
	0.5	36.8	23.3	29.4	**43.5**
Wiki-CS	0.0	-	69.1	68.4	**72.3**
	0.1	-	63.6	46.8	**70.4**
	0.3	-	41.6	24.2	**46.2**
	0.5	-	12.5	18.5	**24.2**



(a) Digits (b) Citeseer

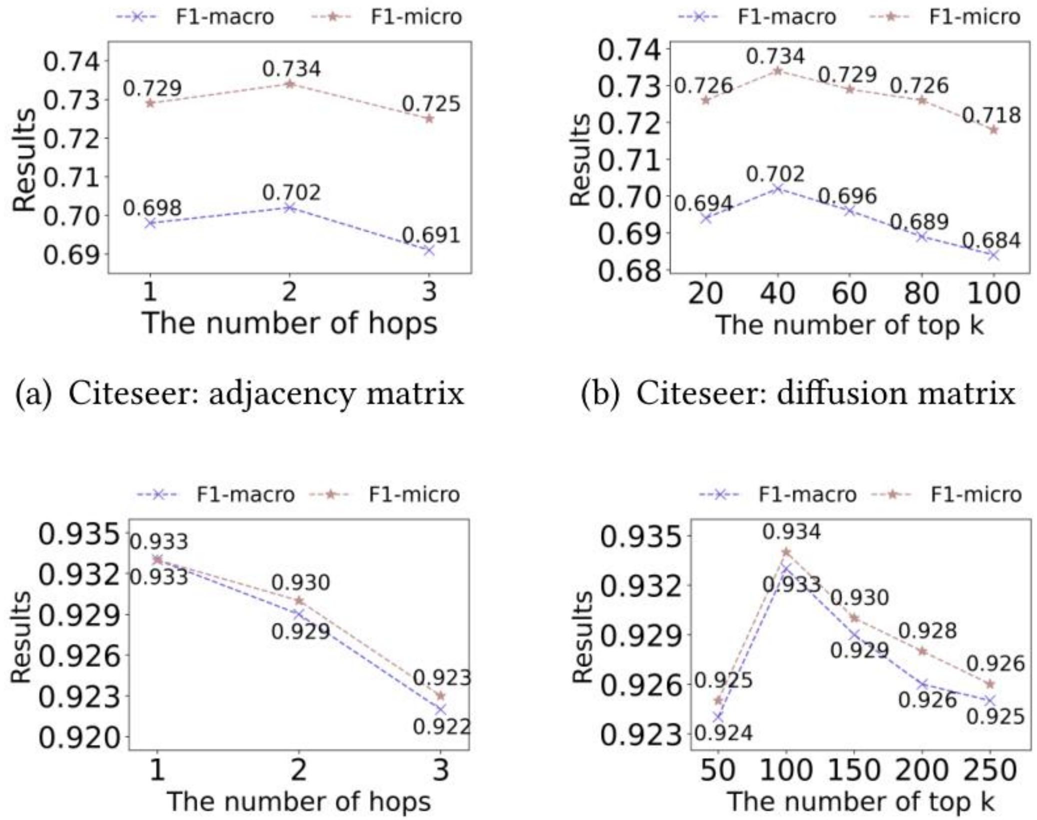
Figure 5: Test on the effectiveness of view estimator.

Experiments



(a) Digits (b) Citeseer

Figure 6: The investigation of change of MI.



(a) Citeseer: adjacency matrix (b) Citeseer: diffusion matrix

(c) Digits: adjacency matrix (d) Digits: diffusion matrix

Figure 7: Impact of hyper-parameter scope.

Thanks