Hard Sample Aware Network for Contrastive Deep Graph Clustering

Jianan Zhao[*]
University of Notre Dame
Notre Dame, IN, USA
andy.zhaoja@gmail.com

Qianlong Wen[*]
University of Notre Dame
Notre Dame, IN, USA
qwen@nd.edu

Mingxuan Ju
University of Notre Dame
Notre Dame, IN, USA
mju2@nd.edu

Chuxu Zhang[†]
Brandeis University
Waltham, MA, USA
chuxuzhang@brandeis.edu

Yanfang Ye[†]
University of Notre Dame
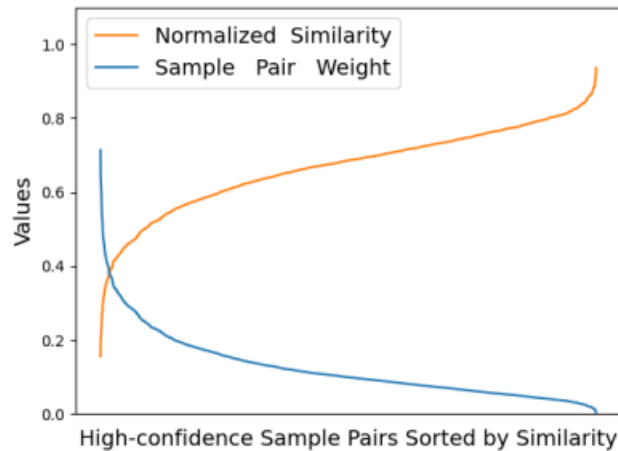Notre Dame, IN, USA
yye7@nd.edu

WSDM-2023

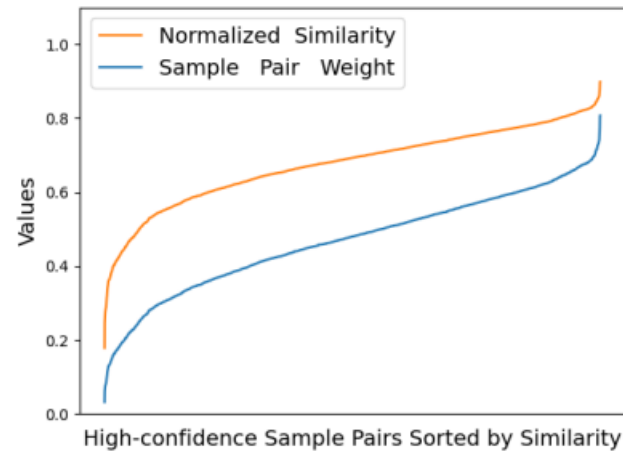Code: https://github.com/andyjzhao/WSDM23-GSR

Reported by Dongdong Hu

Introduction

(a) Positive Sample Pairs

(b) Negative Sample Pairs

1) In the hardness measurement, the important structural information is overlooked for similarity calculation, degrading the representativeness of the selected hard negative samples

2) Previous works merely focus on the hard negative sample pairs while neglecting the hard positive sample pairs. Nevertheless, samples within the same cluster but with low similarity should also be carefully learned

Method
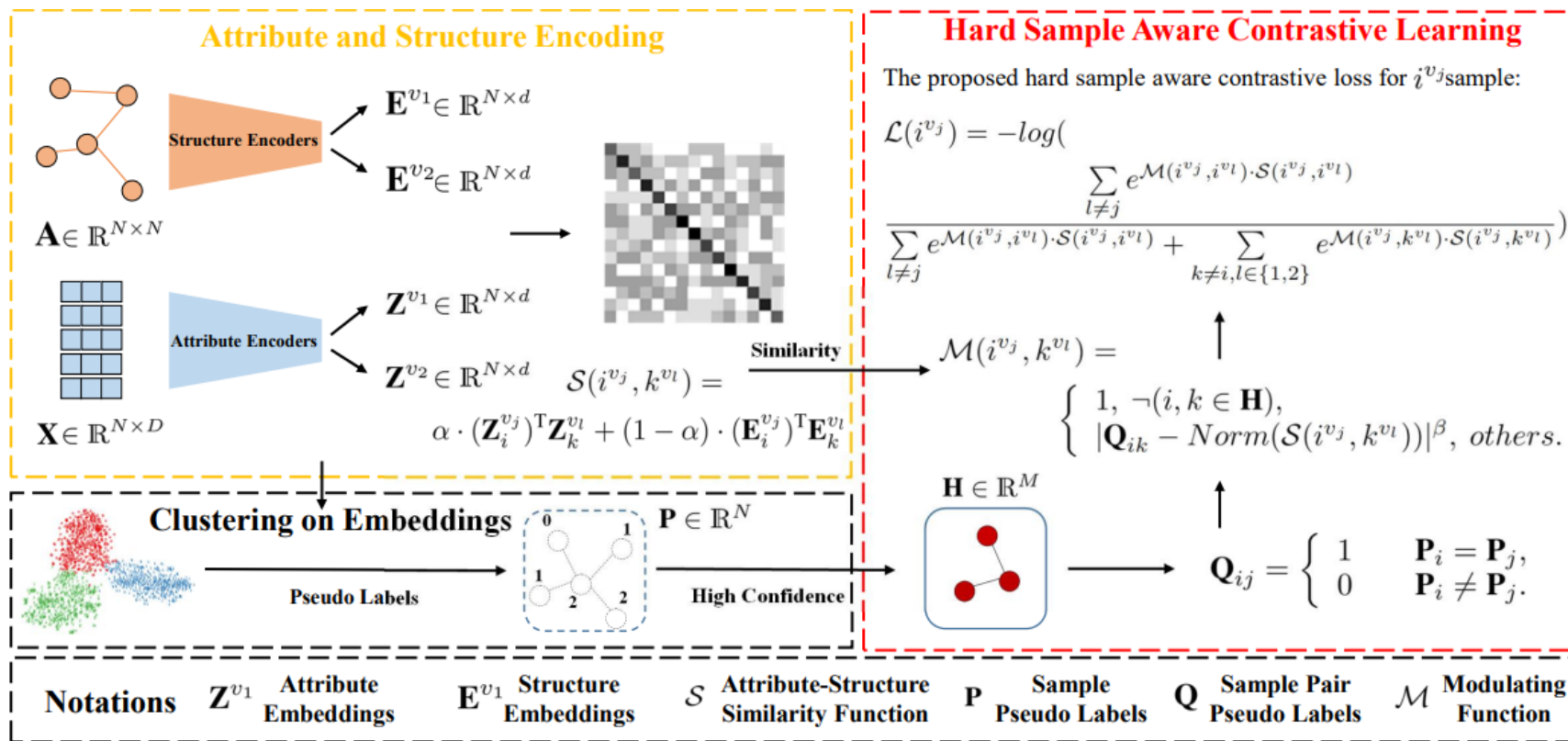
Figure 2: Illustration of our proposed hard sample aware network. In attribute and structure encoding, we embed the attribute and structure into the latent space with the attribute encoders and structure encoders. Then the sample similarities are calculated by a learnable linear combination of attribute similarity and structure similarity, thus better revealing the sample relations. Moreover, guided by the high-confidence information, a general dynamic sample weighting strategy is proposed to up-weight hard sample pairs while down-weighting the easy ones. Overall, the hard sample aware contrastive loss guides the network to focus more on both hard positive and negative sample pairs, thus further improving the discriminative capability of samples.

Method

Attribute and Structure Encoding

$\mathbf{A} \in \mathbb{R}^{N \times N}$

Structure Encoders

$\mathbf{E}^{v_1} \in \mathbb{R}^{N \times d}$

$\mathbf{E}^{v_2} \in \mathbb{R}^{N \times d}$

$\mathbf{X} \in \mathbb{R}^{N \times D}$

Attribute Encoders

$\mathbf{Z}^{v_1} \in \mathbb{R}^{N \times d}$

$\mathbf{Z}^{v_2} \in \mathbb{R}^{N \times d}$

Similarity

$S(i^{v_j}, k^{v_l}) = \alpha \cdot (\mathbf{Z}_i^{v_j})^{\mathrm{T}} \mathbf{Z}_k^{v_l} + (1 - \alpha) \cdot (\mathbf{E}_i^{v_j})^{\mathrm{T}} \mathbf{E}_k^{v_l}$

Clustering on Embeddings

$\mathbf{P} \in \mathbb{R}^N$

Pseudo Labels

High Confidence

Hard Sample Aware Contrastive Learning

The proposed hard sample aware contrastive loss for i^{v_j} sample:

$$\mathcal{L}(i^{v_j}) = -log\left(\frac{\sum\limits_{l \neq j} e^{\mathcal{M}(i^{v_j}, i^{v_l}) \cdot S(i^{v_j}, i^{v_l})}}{\sum\limits_{l \neq j} e^{\mathcal{M}(i^{v_j}, i^{v_l}) \cdot S(i^{v_j}, i^{v_l})} + \sum\limits_{k \neq i, l \in \{1,2\}} e^{\mathcal{M}(i^{v_j}, k^{v_l}) \cdot S(i^{v_j}, k^{v_l})}} \right)$$

$$\mathcal{M}(i^{v_j}, k^{v_l}) = \begin{cases} 1, & \neg(i, k \in \mathbf{H}), \\ |\mathbf{Q}_{ik} - Norm(S(i^{v_j}, k^{v_l}))|^{\beta}, & others. \end{cases}$$

$\mathbf{H} \in \mathbb{R}^M$

$$\mathbf{Q}_{ij} = \begin{cases} 1 & \mathbf{P}_i = \mathbf{P}_j, \\ 0 & \mathbf{P}_i \neq \mathbf{P}_j. \end{cases}$$

Notations | $\mathbf{Z}^{v_1}$ Attribute Embeddings | $\mathbf{E}^{v_1}$ Structure Embeddings | S Attribute-Structure Similarity Function | $\mathbf{P}$ Sample Pseudo Labels | $\mathbf{Q}$ Sample Pair Pseudo Labels | $\mathcal{M}$ Modulating Function

$\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$ $\mathcal{E}$ be a set of edges

$\mathbf{X} \in \mathbb{R}^{N \times D}$ and $\mathbf{A} \in \mathbb{R}^{N \times N}$ $\mathcal{G} = \{\mathbf{X}, \mathbf{A}\}$
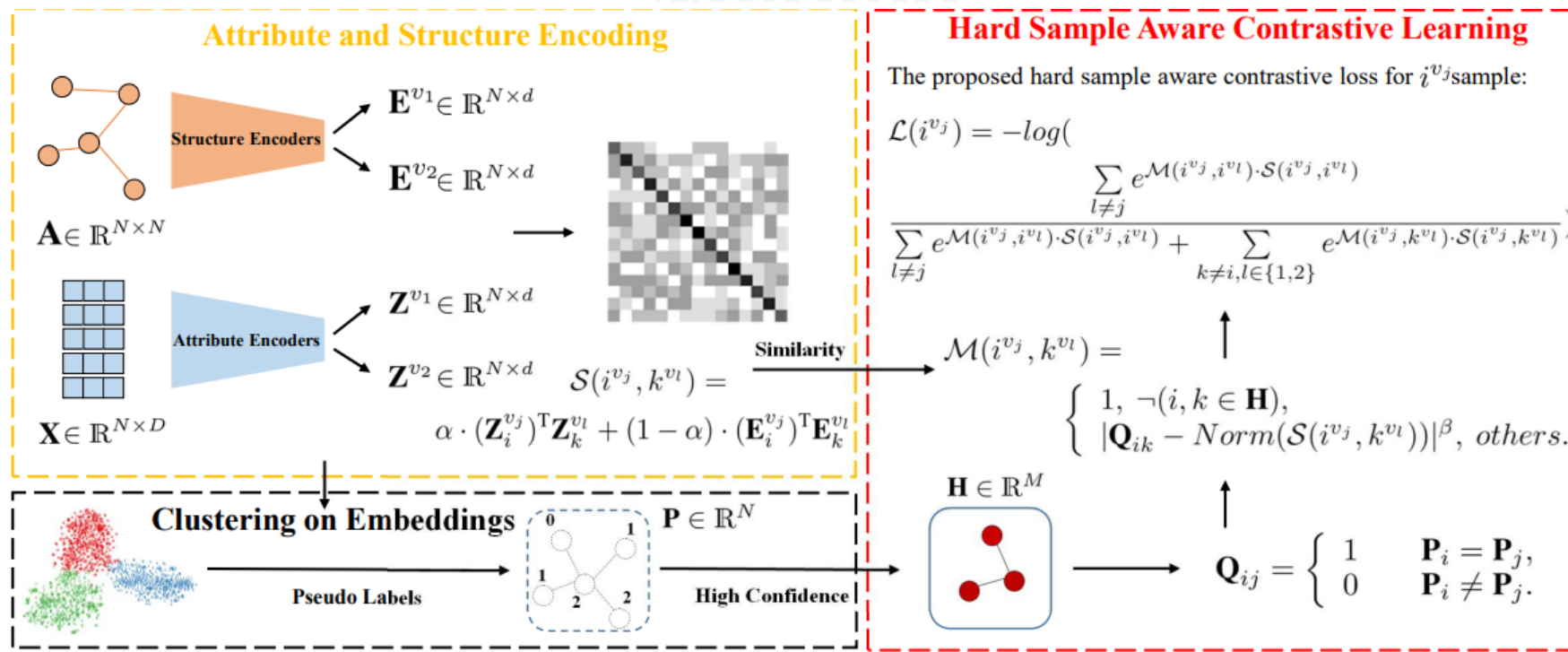
$\mathbf{D} = diag(d_1, d_2, \ldots, d_N) \in \mathbb{R}^{N \times N}$

$\mathbf{E} = \mathcal{F}(\mathbf{A}, \mathbf{X}),$ (1)

$\Phi = \mathcal{C}(\mathbf{E}),$ (2)

$\tilde{\mathbf{L}} = \mathbf{I} - \widehat{\mathbf{D}}^{-\frac{1}{2}} \widehat{\mathbf{A}} \widehat{\mathbf{D}}^{-\frac{1}{2}}$

Method

Attribute and Structure Encoding

$\mathbf{E}^{v_1} \in \mathbb{R}^{N \times d}$

Structure Encoders

$\mathbf{E}^{v_2} \in \mathbb{R}^{N \times d}$

$\mathbf{A} \in \mathbb{R}^{N \times N}$

$\mathbf{Z}^{v_1} \in \mathbb{R}^{N \times d}$

Attribute Encoders

$\mathbf{Z}^{v_2} \in \mathbb{R}^{N \times d}$

$\mathbf{X} \in \mathbb{R}^{N \times D}$

Similarity

$\mathcal{S}(i^{v_j}, k^{v_l}) = \alpha \cdot (\mathbf{Z}_i^{v_j})^{\mathrm{T}} \mathbf{Z}_k^{v_l} + (1 - \alpha) \cdot (\mathbf{E}_i^{v_j})^{\mathrm{T}} \mathbf{E}_k^{v_l}$

Clustering on Embeddings $\mathbf{P} \in \mathbb{R}^N$

Pseudo Labels   High Confidence

Hard Sample Aware Contrastive Learning

The proposed hard sample aware contrastive loss for i^{v_j} sample:

$$\mathcal{L}(i^{v_j}) = -log\left(\frac{\sum\limits_{l \neq j} e^{\mathcal{M}(i^{v_j}, i^{v_l}) \cdot \mathcal{S}(i^{v_j}, i^{v_l})}}{\sum\limits_{l \neq j} e^{\mathcal{M}(i^{v_j}, i^{v_l}) \cdot \mathcal{S}(i^{v_j}, i^{v_l})} + \sum\limits_{k \neq i, l \in \{1,2\}} e^{\mathcal{M}(i^{v_j}, k^{v_l}) \cdot \mathcal{S}(i^{v_j}, k^{v_l})}} \right)$$

$$\mathcal{M}(i^{v_j}, k^{v_l}) = \begin{cases} 1, & \neg(i, k \in \mathbf{H}), \\ |\mathbf{Q}_{ik} - Norm(\mathcal{S}(i^{v_j}, k^{v_l}))|^{\beta}, & others. \end{cases}$$

$\mathbf{H} \in \mathbb{R}^M$

$$\mathbf{Q}_{ij} = \begin{cases} 1 & \mathbf{P}_i = \mathbf{P}_j, \\ 0 & \mathbf{P}_i \neq \mathbf{P}_j. \end{cases}$$

$$\widetilde{\mathbf{X}} = (\prod_{i=1}^{t}(\mathbf{I} - \widetilde{\mathbf{L}}))\mathbf{X} = (\mathbf{I} - \widetilde{\mathbf{L}})^t \mathbf{X}, \qquad (3)$$

filter the high-frequency noises

$$\mathbf{Z}^{v_1} = \mathrm{AE}_1(\widetilde{\mathbf{X}}); \mathbf{Z}_i^{v_1} = \frac{\mathbf{Z}_i^{v_1}}{||\mathbf{Z}_i^{v_1}||_2}, i = 1, 2, ..., N;$$

$$\mathbf{Z}^{v_2} = \mathrm{AE}_2(\widetilde{\mathbf{X}}); \mathbf{Z}_j^{v_2} = \frac{\mathbf{Z}_j^{v_2}}{||\mathbf{Z}_j^{v_2}||_2}, j = 1, 2, ..., N,$$

(4)

where $\mathbf{Z}^{v_1}$ and $\mathbf{Z}^{v_2}$ denote two-view attribute embeddings

$$\mathbf{E}^{v_1} = \mathrm{SE}_1(\mathbf{A}); \mathbf{E}_i^{v_1} = \frac{\mathbf{E}_i^{v_1}}{||\mathbf{E}_i^{v_1}||_2}, i = 1, 2, ..., N;$$
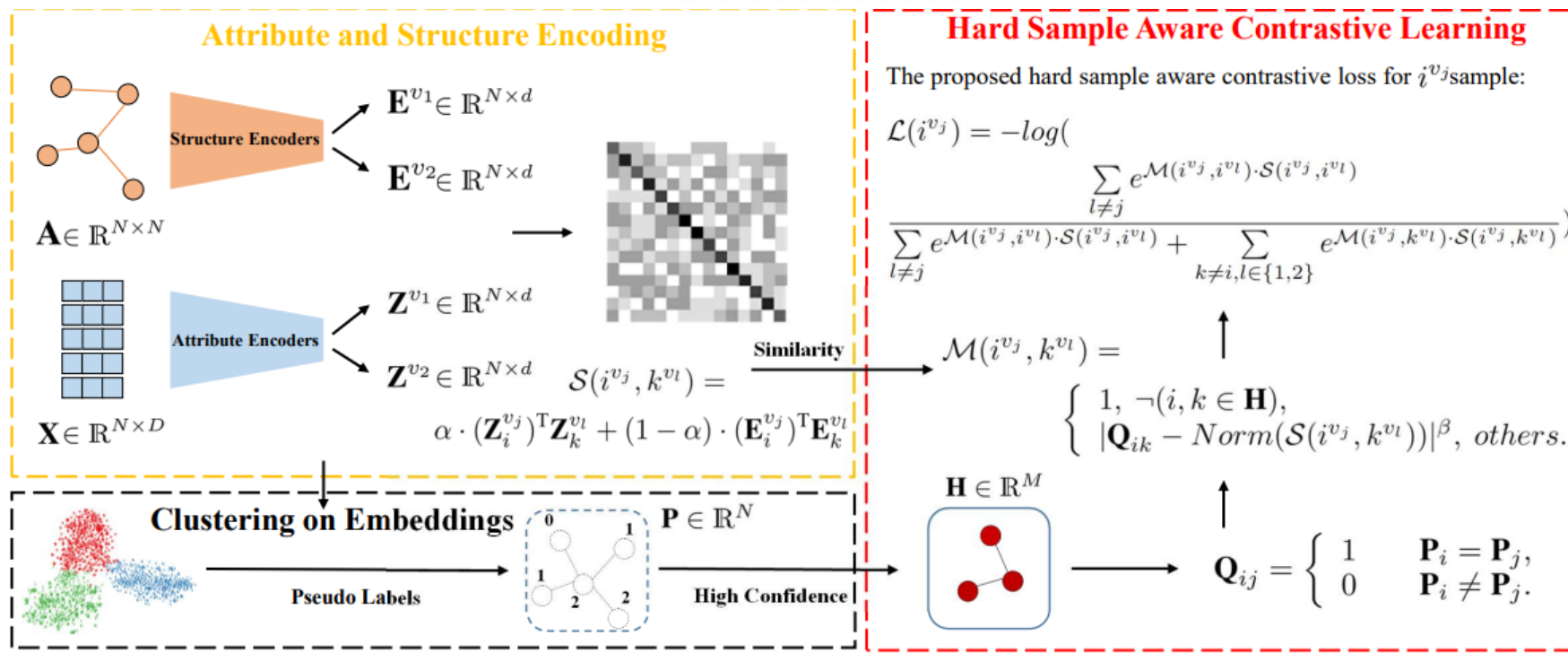
$$\mathbf{E}^{v_2} = \mathrm{SE}_2(\mathbf{A}); \mathbf{E}_j^{v_2} = \frac{\mathbf{E}_j^{v_2}}{||\mathbf{E}_j^{v_2}||_2}, j = 1, 2, ..., N,$$

(5)

$$\mathcal{S}(i^{v_j}, k^{v_l}) = \alpha \cdot (\mathbf{Z}_i^{v_j})^{\mathrm{T}} \mathbf{Z}_k^{v_l} + (1 - \alpha) \cdot (\mathbf{E}_i^{v_j})^{\mathrm{T}} \mathbf{E}_k^{v_l}, \quad (6)$$

where $i, k \in \{1, 2, ..., N\}$ and $j, l \in \{1, 2\}$

Method

Attribute and Structure Encoding

$\mathbf{A} \in \mathbb{R}^{N \times N}$

Structure Encoders

$\mathbf{E}^{v_1} \in \mathbb{R}^{N \times d}$

$\mathbf{E}^{v_2} \in \mathbb{R}^{N \times d}$

$\mathbf{X} \in \mathbb{R}^{N \times D}$

Attribute Encoders

$\mathbf{Z}^{v_1} \in \mathbb{R}^{N \times d}$

$\mathbf{Z}^{v_2} \in \mathbb{R}^{N \times d}$

$\mathcal{S}(i^{v_j}, k^{v_l}) = \alpha \cdot (\mathbf{Z}_i^{v_j})^{\mathrm{T}} \mathbf{Z}_k^{v_l} + (1 - \alpha) \cdot (\mathbf{E}_i^{v_j})^{\mathrm{T}} \mathbf{E}_k^{v_l}$

Similarity

Clustering on Embeddings

$\mathbf{P} \in \mathbb{R}^N$

Pseudo Labels

High Confidence

Hard Sample Aware Contrastive Learning

The proposed hard sample aware contrastive loss for i^{v_j} sample:

$$\mathcal{L}(i^{v_j}) = -\log\left(\frac{\sum_{l \neq j} e^{\mathcal{M}(i^{v_j}, i^{v_l}) \cdot \mathcal{S}(i^{v_j}, i^{v_l})}}{\sum_{l \neq j} e^{\mathcal{M}(i^{v_j}, i^{v_l}) \cdot \mathcal{S}(i^{v_j}, i^{v_l})} + \sum_{k \neq i, l \in \{1,2\}} e^{\mathcal{M}(i^{v_j}, k^{v_l}) \cdot \mathcal{S}(i^{v_j}, k^{v_l})}} \right)$$

$$\mathcal{M}(i^{v_j}, k^{v_l}) = \begin{cases} 1, & \neg(i, k \in \mathbf{H}), \\ |\mathbf{Q}_{ik} - Norm(\mathcal{S}(i^{v_j}, k^{v_l}))|^{\beta}, & others. \end{cases}$$

$\mathbf{H} \in \mathbb{R}^M$

$$\mathbf{Q}_{ij} = \begin{cases} 1 & \mathbf{P}_i = \mathbf{P}_j, \\ 0 & \mathbf{P}_i \neq \mathbf{P}_j. \end{cases}$$

$$\mathbf{Q}_{ij} = \begin{cases} 1 & \mathbf{P}_i = \mathbf{P}_j, \\ 0 & \mathbf{P}_i \neq \mathbf{P}_j. \end{cases} \quad (7)$$

$$\mathcal{L}_{infoNCE}(i^{v_j}) = -\log \frac{e^{\theta(i^{v_j}, i^{v_l})}}{e^{\theta(i^{v_j}, i^{v_l})} + \sum_{k \neq i} (e^{\theta(i^{v_j}, k^{v_j})} + e^{\theta(i^{v_j}, k^{v_l})})}. \quad (8)$$

$$\mathcal{L} = \frac{1}{2N} \sum_{j=1}^{2} \sum_{i=1}^{N} \mathcal{L}(i^{v_j}). \quad (11)$$

$$\mathcal{M}(i^{v_j}, k^{v_l}) = \begin{cases} 1, & \neg(i, k \in \mathbf{H}), \\ |\mathbf{Q}_{ik} - Norm(\mathcal{S}(i^{v_j}, k^{v_l}))|^{\beta}, & others. \end{cases}$$

1) $\mathcal{M}$ can up-weight the hard samples while down-weighting the easy samples.

2) The focusing factor β controls the down-weighting rate of easy sample pairs.

Experiments

Dataset	Metric	Classical Deep Graph Clustering					Contrastive Deep Graph Clustering						Hard Sample Mining		
		MGAE CIKM 19	DAEGC IJCAI 19	ARGA IJCAI 19	SDCN WWW 20	DFCN AAAI 21	AGE SIGKDD 20	MVGRL ICML 20	AutoSSL ICLR 22	AGC-DRR IJCAI 22	DCRN AAAI 22	AFGRL AAAI 22	GDCL IJCAI 21	ProGCL ICML 22	HSAN Ours
CORA	ACC	43.38±2.11	70.43±0.36	71.04±0.25	35.60±2.83	36.33±0.49	73.50±1.83	70.47±3.70	63.81±0.57	40.62±0.55	61.93±0.47	26.25±1.24	70.83±0.47	57.13±1.23	77.07±1.56
	NMI	28.78±2.97	52.89±0.69	51.06±0.52	14.28±1.91	19.36±0.87	57.58±1.42	55.57±1.54	47.62±0.45	18.74±0.73	45.13±1.57	12.36±1.54	56.60±0.36	41.02±1.34	59.21±1.03
	ARI	16.43±1.65	49.63±0.43	47.71±0.33	07.78±3.24	04.67±2.10	50.10±2.14	48.70±3.94	38.92±0.77	14.80±1.64	33.15±0.14	14.32±1.87	48.05±0.72	30.71±2.70	57.52±2.70
	F1	33.48±3.05	68.27±0.57	69.27±0.39	24.37±1.04	26.16±0.50	69.28±1.59	67.15±1.86	56.42±0.21	31.23±0.57	49.50±0.42	30.20±1.15	52.88±0.97	45.68±1.29	75.11±1.40
CITE	ACC	61.35±0.80	64.54±1.39	61.07±0.49	65.96±0.31	69.50±0.20	69.73±0.24	62.83±1.59	66.76±0.67	68.32±1.83	70.86±0.18	31.45±0.54	66.39±0.65	65.92±0.80	71.15±0.80
	NMI	34.63±0.65	36.41±0.86	34.40±0.71	38.71±0.32	43.90±0.20	44.93±0.53	40.69±0.93	40.67±0.84	43.28±1.41	45.86±0.35	15.17±0.47	39.52±0.38	39.59±0.39	45.06±0.74
	ARI	33.55±1.18	37.78±1.24	34.32±0.70	40.17±0.43	45.50±0.30	45.31±0.41	34.18±1.73	38.73±0.55	45.34±2.33	47.64±0.30	14.32±0.78	41.07±0.96	36.16±1.11	47.05±1.12
	F1	57.36±0.82	62.20±1.32	58.23±0.31	63.62±0.24	64.30±0.20	64.45±0.27	59.54±2.17	58.22±0.68	64.82±1.60	65.83±0.21	30.20±0.71	61.12±0.70	57.89±1.98	63.01±1.79
AMAP	ACC	71.57±2.48	75.96±0.23	69.28±2.30	53.44±0.81	76.82±0.23	75.98±0.68	41.07±3.12	54.55±0.97	76.81±1.45	OOM	75.51±0.77	43.75±0.78	51.53±0.38	77.02±0.33
	NMI	62.13±2.79	65.25±0.45	58.36±2.76	44.85±0.83	66.23±1.21	65.38±0.61	30.28±3.94	48.56±0.71	66.54±1.24		64.05±0.15	37.32±0.28	39.56±0.39	67.21±0.33
	ARI	48.82±4.57	58.12±0.24	44.18±4.41	31.21±1.23	58.28±0.74	55.89±1.34	18.77±2.34	26.87±0.34	60.15±1.56		54.45±0.48	21.57±0.51	34.18±0.89	58.01±0.48
	F1	68.08±1.76	69.87±0.54	64.30±1.95	50.66±1.49	71.25±0.31	71.74±0.93	32.88±5.50	54.47±0.83	71.03±0.64		69.99±0.34	38.37±0.29	31.97±0.44	72.03±0.46
BAT	ACC	53.59±2.04	52.67±0.00	67.86±0.80	53.05±4.63	55.73±0.06	56.68±0.76	37.56±0.32	42.43±0.47	47.79±0.02	67.94±1.45	50.92±0.44	45.42±0.54	55.73±0.79	77.15±0.72
	NMI	30.59±2.06	21.43±0.35	49.09±0.54	25.74±5.71	48.77±0.51	36.04±1.54	29.33±0.70	17.84±0.98	19.91±0.24	47.23±0.74	27.55±0.62	31.70±0.42	28.69±0.92	53.21±0.93
	ARI	24.15±1.70	18.18±0.29	42.02±1.21	21.04±4.97	37.76±0.23	26.59±1.83	13.45±0.03	13.11±0.81	14.59±0.13	39.76±0.87	21.89±0.74	19.33±0.57	21.84±1.34	52.20±1.11
	F1	50.83±3.23	52.23±0.03	67.02±1.15	46.45±5.90	50.90±0.12	55.07±0.80	29.64±0.49	34.84±0.15	42.33±0.51	67.40±0.35	46.53±0.57	39.94±0.57	56.08±0.89	77.13±0.76
EAT	ACC	44.61±2.10	36.89±0.15	52.13±0.00	39.07±1.51	49.37±0.19	47.26±0.32	32.88±0.71	31.33±0.52	37.37±0.11	50.88±0.55	37.42±1.24	33.46±0.18	43.36±0.87	56.69±0.34
	NMI	15.60±2.30	05.57±0.06	22.48±1.21	08.83±2.54	32.90±0.41	23.74±0.90	11.72±1.08	07.63±0.85	07.00±0.85	22.01±1.23	11.44±1.41	13.22±0.33	23.93±0.45	33.25±0.44
	ARI	13.40±1.26	05.03±0.08	17.29±0.50	06.31±1.95	23.25±0.18	16.57±0.46	04.68±1.30	02.13±0.67	04.88±0.91	18.13±0.85	06.57±1.73	04.31±0.29	15.03±0.98	26.85±0.59
	F1	43.08±3.26	34.72±0.16	52.75±0.07	33.42±3.10	42.95±0.04	45.54±0.40	25.35±0.75	21.82±0.98	35.20±0.17	47.06±0.66	30.53±1.47	25.02±0.21	42.54±0.45	57.26±0.28
UAT	ACC	48.97±1.52	52.29±0.49	49.31±0.15	52.25±1.91	33.61±0.09	52.37±0.42	44.16±1.38	42.52±0.64	42.64±0.31	49.92±1.25	41.50±0.25	48.70±0.06	45.38±0.58	56.04±0.67
	NMI	20.69±0.98	21.33±0.44	25.44±0.31	21.61±1.26	26.49±0.41	23.64±0.66	21.53±0.94	17.86±0.22	11.15±0.24	24.09±0.53	17.33±0.54	25.10±0.01	22.04±2.23	26.99±2.11
	ARI	18.33±1.79	20.50±0.51	16.57±0.31	21.63±1.49	11.87±0.23	20.39±0.70	17.12±1.46	13.13±0.71	09.50±0.25	17.17±0.69	13.62±0.57	21.76±0.01	14.74±1.99	25.22±1.96
	F1	47.95±1.52	50.33±0.64	50.26±0.16	45.59±3.54	25.79±0.29	50.15±0.73	39.44±2.19	34.94±0.87	35.18±0.32	44.81±0.87	36.52±0.89	45.69±0.08	39.30±1.82	54.20±1.84

Table 2: The average clustering performance of ten runs on six benchmark datasets. The performance is evaluated by four metrics with mean value and standard deviation. The red and blue values indicate the best and the runner-up results, respectively.
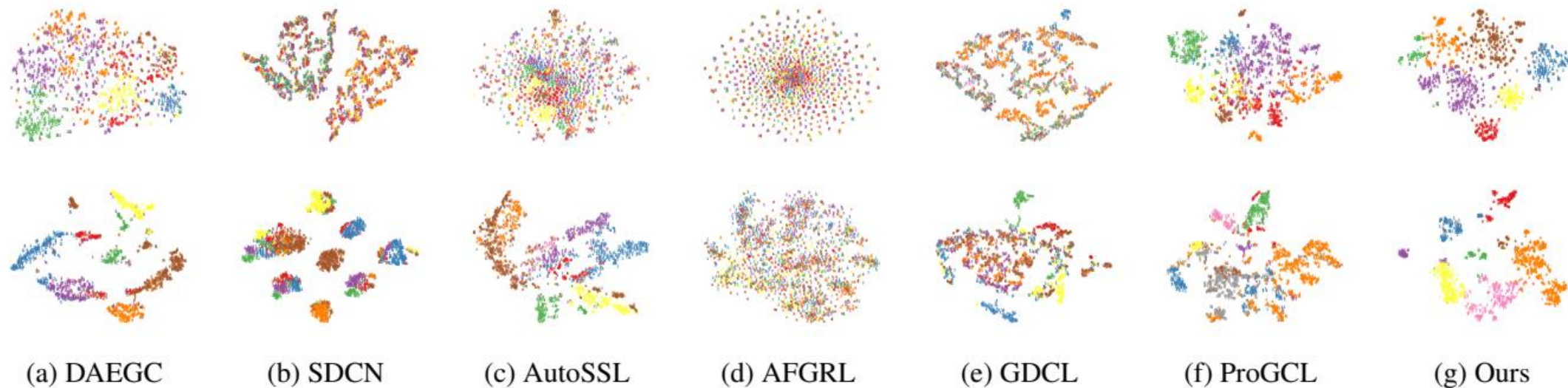
Experiments

(a) DAEGC   (b) SDCN   (c) AutoSSL   (d) AFGRL   (e) GDCL   (f) ProGCL   (g) Ours

Figure 3: 2D t-SNE visualization of seven methods on two benchmark datasets. The first row and second row corresponds to CORA and AMAP dataset, respectively.
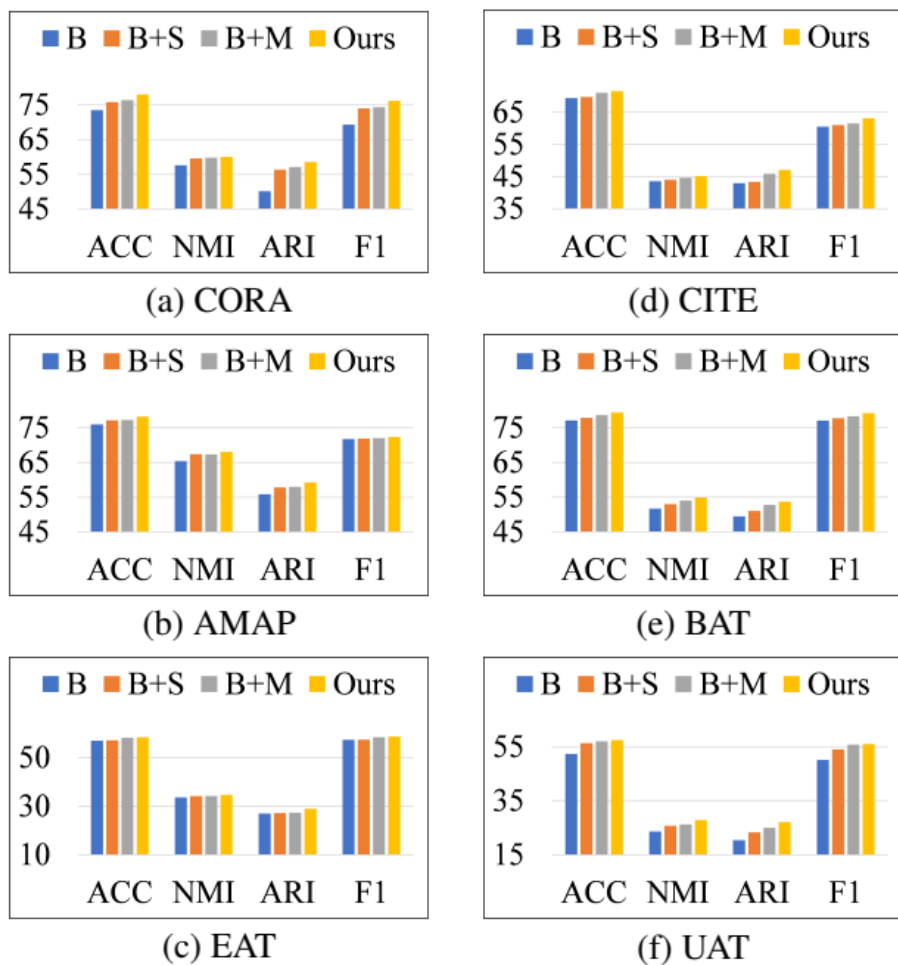
Experiments

Figure 4: Ablation studies of the proposed similarity function $\mathcal{S}$ and weight modulating function $\mathcal{M}$ on six datasets.
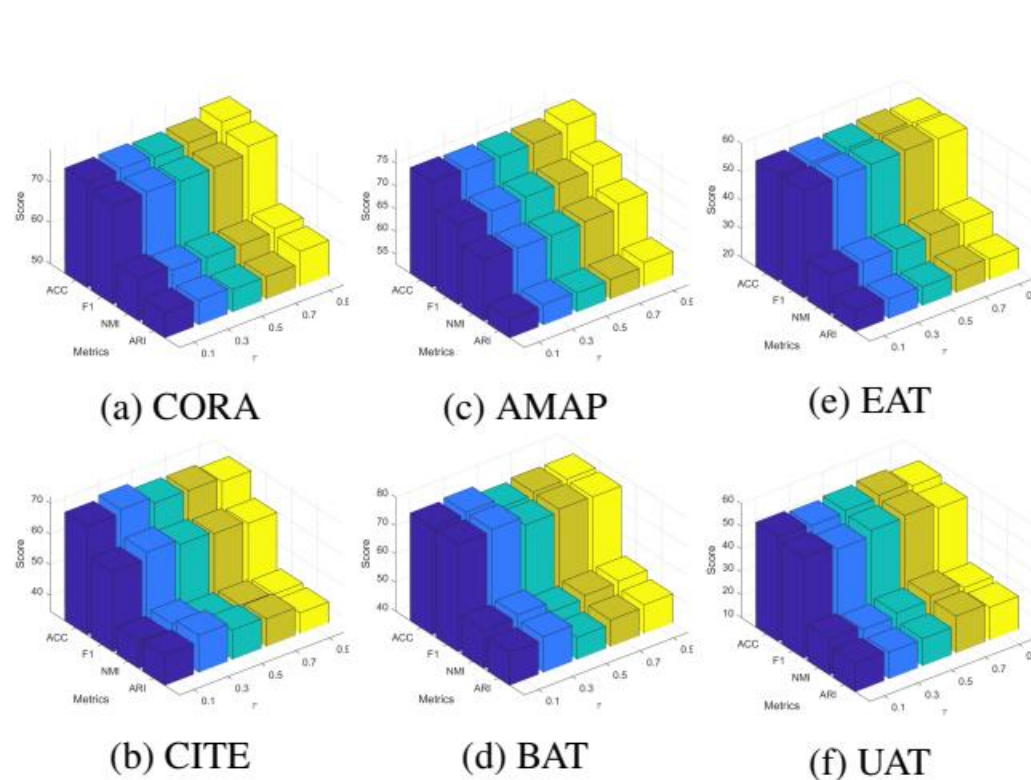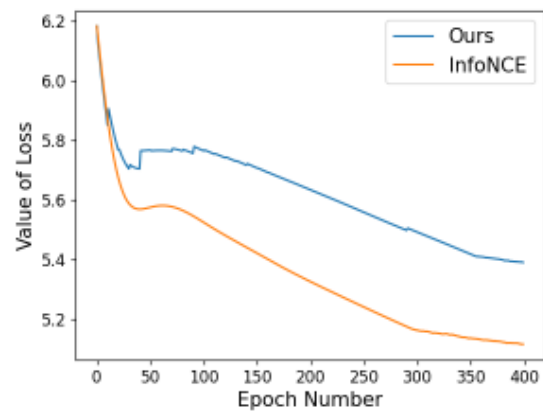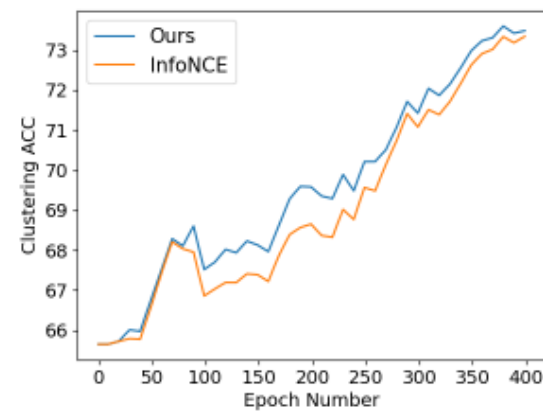

Figure 5: Analysis of the confidence hyper-parameter τ.

Experiments

(a) Loss on BAT

(b) ACC on BAT

Figure 6: Convergence analysis on BAT dataset.

Thanks