Neighbor Contrastive Learning on Learnable Graph Augmentation

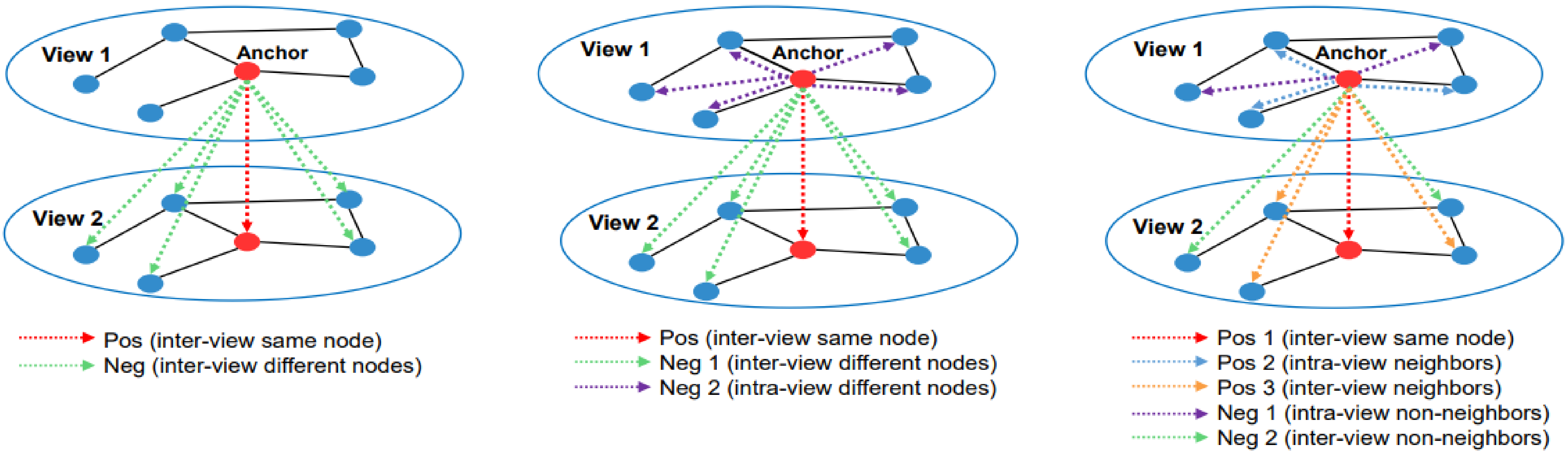
Xiao Shen[1], Dewang Sun[1], Shirui Pan[2], Xi Zhou[3*], Laurence T. Yang[1,4]
[1] School of Computer Science and Technology, Hainan University, Haikou, China
[2] School of ICT, Griffith University, Gold Coast, Australia
[3] College of Tropical Crops, Hainan University, Haikou, China
[4] Department of Computer Science, St. Francis Xavier University, Antigonish, Canada
shenxiaocam@163.com, dwsun@hainanu.edu.cn, s.pan@griffith.edu.au, xzhou@hainanu.edu.cn, ltyang@hainanu.edu.cn

AAAI-2023

Code: https://github.com/shenxiaocam/NCLA.

Reported by Dongdong Hu

Introduction



(a) InfoNCE (b) NT-Xent (c) Neighbor Contrastive Loss (Ours)

However, the existing GCL methods mostly adopt human-designed graph augmentations, which are sensitive to various graph datasets.

However, this is contradictory with the homophily assumption of networks that connected nodes often belong to the same class and should be close to each other

Method

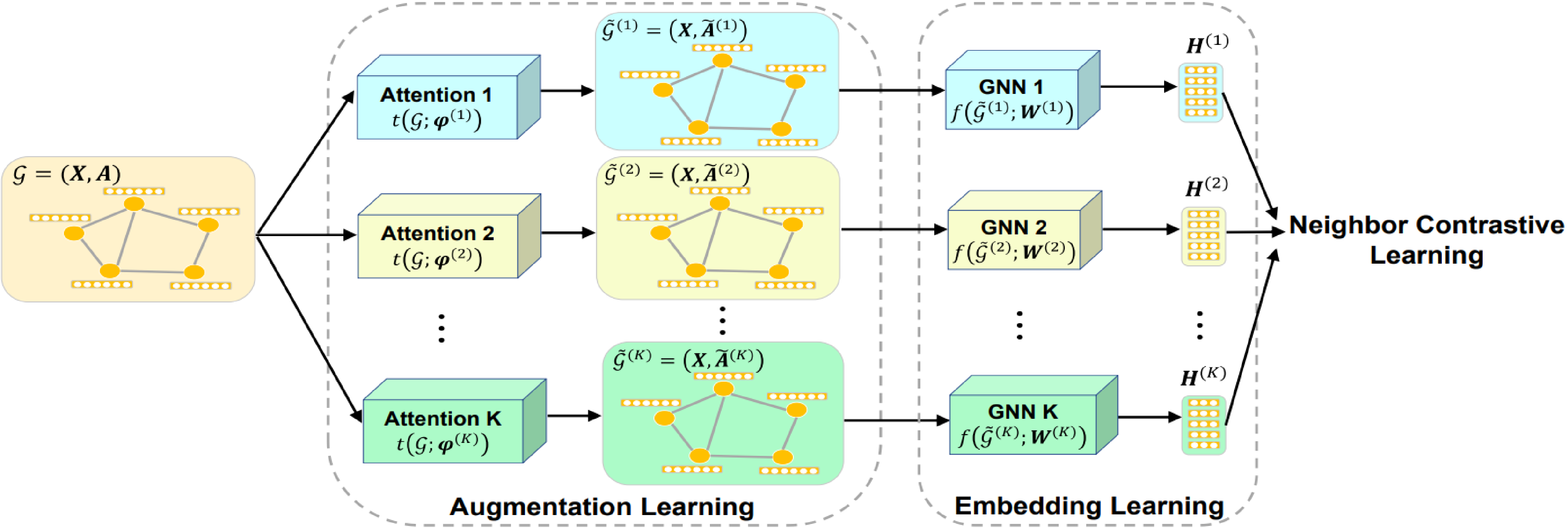
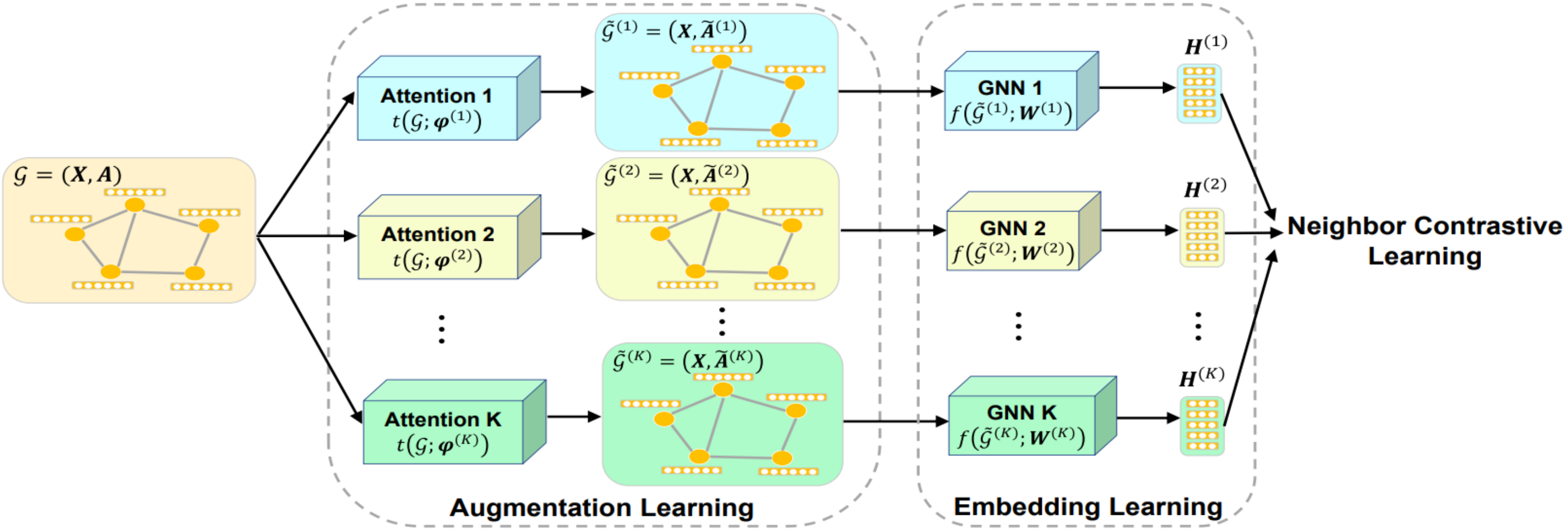


Figure 2. The model architecture of NCLA. It generates K learnable augmented views with adaptive topology by multi-head GAT, where each k-th view has its own learnable parameters $\boldsymbol{\varphi}^{(k)}$, $\boldsymbol{W}^{(k)}$ which are not shared with other views. Then, it applies the neighbor contrastive loss to maximize the agreement between the embeddings of positive pairs and minimize that of negative pairs. Both augmentations and embeddings are learned end-to-end in NCLA.

Method



$$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{X}, \boldsymbol{A})$$

$$\mathcal{V} = \{v_1, v_2, \cdots, v_N\} \qquad \mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$$

$$\boldsymbol{X} \in \mathbb{R}^{N\times F} \text{ and } \boldsymbol{A} \in \{0,1\}^{N\times N}$$

$$\mathcal{N}_i = \{v_j | j \neq i, \boldsymbol{A}_{ij} = 1\}$$

$$\tilde{\mathcal{G}}^{(k)} = t(\mathcal{G}; \boldsymbol{\varphi}^{(k)}) = (\boldsymbol{X}, \tilde{\boldsymbol{A}}^{(k)})$$

multi-head GAT

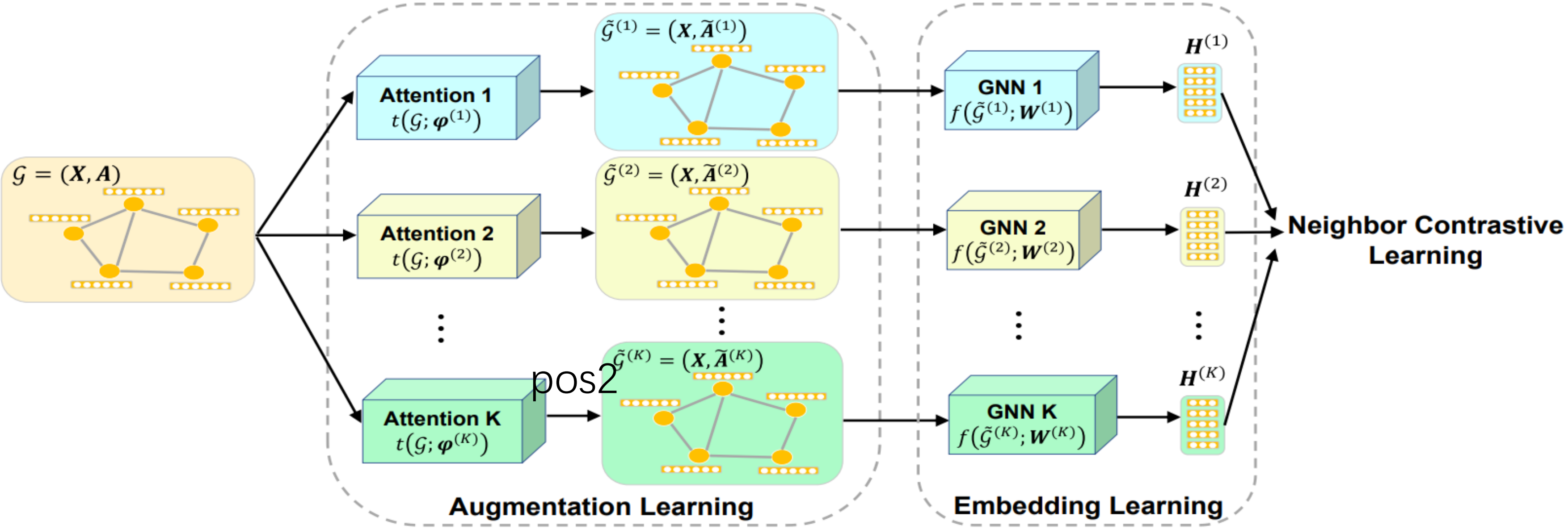
$$\tilde{\boldsymbol{A}}_{ij}^{(k)} = \frac{e^{\mathrm{LeakyReLU}\left(\boldsymbol{\varphi}^{(k)}\left[\boldsymbol{W}^{(k)}\boldsymbol{x}_i \| \boldsymbol{W}^{(k)}\boldsymbol{x}_j\right]\right)}}{\sum_{v_p \in \mathcal{N}_i \cup \{v_i\}} e^{\mathrm{LeakyReLU}\left(\boldsymbol{\varphi}^{(k)}\left[\boldsymbol{W}^{(k)}\boldsymbol{x}_i \| \boldsymbol{W}^{(k)}\boldsymbol{x}_p\right]\right)}} \quad (1)$$

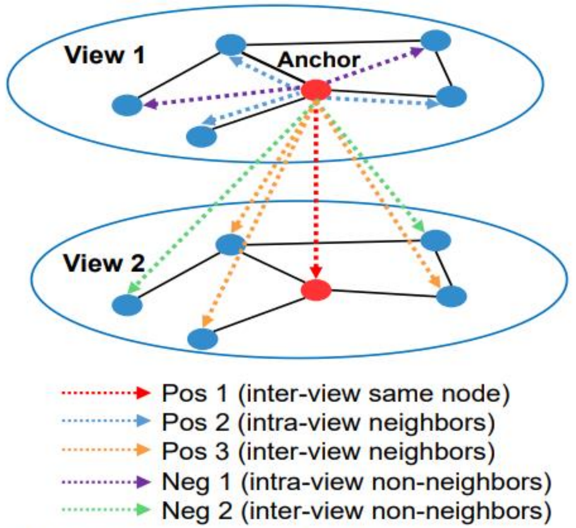
$$\boxed{\alpha_{ij} = \frac{\exp\left(\mathrm{LeakyReLU}\left(\vec{\mathbf{a}}^T[\mathbf{W}\vec{h}_i \| \mathbf{W}\vec{h}_j]\right)\right)}{\sum_{k\in\mathcal{N}_i} \exp\left(\mathrm{LeakyReLU}\left(\vec{\mathbf{a}}^T[\mathbf{W}\vec{h}_i \| \mathbf{W}\vec{h}_k]\right)\right)}}$$

where $\tilde{\boldsymbol{A}}_{ij}^{(k)}$ is set to 0 if $\boldsymbol{A}_{ij} = 0$, $\boldsymbol{W}^{(k)} \in \mathbb{R}^{F'\times F}$

$\boldsymbol{\varphi}^{(k)} \in \mathbb{R}^{1\times 2F'}$ is a learnable weight vector

Method





(c) Neighbor Contrastive Loss (Ours)

GNN encoder $f(\cdot\,;\boldsymbol{W}^{(k)})$

$$\boldsymbol{h}_i^{(k)} = \mathrm{ELU}\left(\sum_{v_j \in \mathcal{N}_i \cup \{v_i\}} \tilde{\boldsymbol{A}}_{ij}^{(k)} \boldsymbol{W}^{(k)} \boldsymbol{x}_j\right) \quad (2)$$

$$\boldsymbol{h}_i = ||_{k=1}^{K} \boldsymbol{h}_i^{(k)} \quad (3)$$

$$\ell(\boldsymbol{H}^{(1)}, \boldsymbol{H}^{(2)}) = \frac{1}{2N}\sum_{i=1}^{N}\left[\ell(\boldsymbol{h}_i^{(1)}) + \ell(\boldsymbol{h}_i^{(2)})\right] \quad (5)$$

$$\mathcal{L} = \frac{1}{K}\sum_{k=1,k\neq l}^{K} \ell(\boldsymbol{H}^{(k)}, \boldsymbol{H}^{(l)}) \quad (6)$$

pos1 $\ell(\boldsymbol{h}_i^{(1)}) =$ pos2 ,3

$$-log\frac{\left(e^{\theta(\boldsymbol{h}_i^{(1)},\boldsymbol{h}_i^{(2)})/\tau} + \sum_{v_j \in \mathcal{N}_i}\left(e^{\theta(\boldsymbol{h}_i^{(1)},\boldsymbol{h}_j^{(1)})/\tau} + e^{\theta(\boldsymbol{h}_i^{(1)},\boldsymbol{h}_j^{(2)})/\tau}\right)\right)/(2|\mathcal{N}_i|+1)}{e^{\theta(\boldsymbol{h}_i^{(1)},\boldsymbol{h}_i^{(2)})/\tau} + \sum_{j\neq i}\left(\underline{e^{\theta(\boldsymbol{h}_i^{(1)},\boldsymbol{h}_j^{(1)})/\tau}} + \underline{e^{\theta(\boldsymbol{h}_i^{(1)},\boldsymbol{h}_j^{(2)})/\tau}}\right)} \quad (4)$$

pos1 pos2, neg1 pos3, neg2

$$\underline{\sum_{j\neq i} e^{\theta(\boldsymbol{h}_i^{(1)},\boldsymbol{h}_j^{(1)})/\tau}} = \underbrace{\sum_{v_j \in \mathcal{N}_i} e^{\theta(\boldsymbol{h}_i^{(1)},\boldsymbol{h}_j^{(1)})/\tau}}_{\text{intra-view pos}} + \underbrace{\sum_{v_j \notin \mathcal{N}_i} e^{\theta(\boldsymbol{h}_i^{(1)},\boldsymbol{h}_j^{(1)})/\tau}}_{\text{intra-view neg}}$$

$$\underline{\sum_{j\neq i} e^{\theta(\boldsymbol{h}_i^{(1)},\boldsymbol{h}_j^{(2)})/\tau}} = \underbrace{\sum_{v_j \in \mathcal{N}_i} e^{\theta(\boldsymbol{h}_i^{(1)},\boldsymbol{h}_j^{(2)})/\tau}}_{\text{inter-view pos}} + \underbrace{\sum_{v_j \notin \mathcal{N}_i} e^{\theta(\boldsymbol{h}_i^{(1)},\boldsymbol{h}_j^{(2)})/\tau}}_{\text{inter-view neg}}$$

Experiments

Table 1: Statistics of the datasets.

Datasets	# Nodes	# Edges	# Features	# Labels
Cora	2708	10556	1433	7
CiteSeer	3327	9228	3703	6
PubMed	19717	88651	500	3
Coauthor-CS	18333	163788	6805	15
Amazon-Photo	7650	238162	745	8

Table 2: Hyperparameter Settings of NCLA.

Datasets	# Augmented Views K	# Hidden Layers	# Embedding Dimension F'	Temperature τ	Learning Rate	Weight Decay	# Epochs $\mathcal{T}$
Cora	4	1	32	1	1e-2	1e-4	2000
CiteSeer	4	1	32	5	1e-2	1e-4	2000
PubMed	2	1	32	5	1e-3	5e-5	2000
Coauthor-CS	4	1	32	1	5e-2	1e-4	2000
Amazon-Photo	2	1	32	1	1e-3	1e-4	2000

Experiments

Table 3: Classification accuracy with different label rates on five datasets. The best and second-best results are highlighted in boldface and underlined respectively.

Datasets	c	GCN	GAT	CGPN	CG3	DGI	GMI	MVGRL	GRACE	GCA	SUGRL	AFGRL	NCLA
Cora	1	42.6±11.6	42.1±9.5	58.6±10.6	55.4±14.3	55.4±11.4	55.9±9.6	59.1±10.9	51.0±9.8	58.4±10.9	55.2±8.9	47.7±7.8	**63.1**±11.2
	2	55.0±7.5	53.2±9.0	67.4±7.1	66.4±7.7	64.9±9.0	65.2±7.6	67.8±8.6	59.7±7.9	66.0±7.8	65.3±6.2	57.8±6.8	**71.8**±6.9
	3	63.1±6.8	63.2±5.3	70.7±4.0	71.5±4.2	71.1±5.6	70.7±5.2	74.5±4.1	64.0±6.6	71.5±4.6	70.5±3.5	64.6±4.7	**75.7**±5.0
	4	66.4±6.4	66.3±5.9	70.7±2.9	72.7±2.4	72.9±4.5	73.3±4.3	76.1±3.2	66.1±5.4	72.9±4.3	73.5±2.9	67.5±4.2	**77.3**±3.8
	20	79.6±1.8	81.2±1.6	74.0±1.7	80.6±1.6	82.1±1.3	79.4±1.2	**82.4**±1.5	79.6±1.4	79.0±1.4	81.3±1.2	78.6±1.3	82.2±1.6
Citeseer	1	33.8±5.9	31.0±7.2	48.6±11.3	48.4±12.8	47.2±9.2	40.8±6.8	32.8±8.4	40.3±7.2	38.7±9.0	46.7±8.4	42.1±7.2	**52.2**±13.5
	2	44.8±5.5	41.1±7.2	58.0±5.1	60.2±6.8	58.6±4.3	50.2±4.1	47.8±7.5	48.5±6.0	49.6±5.3	57.7±4.6	53.3±5.4	**62.2**±6.4
	3	49.2±5.1	48.6±6.7	59.4±5.4	62.1±7.3	63.3±4.3	55.1±2.7	55.2±6.7	52.7±4.6	54.2±4.7	61.8±5.1	58.0±4.4	**65.5**±3.5
	4	51.7±4.5	52.8±6.6	60.6±3.4	65.1±2.5	65.8±2.1	57.9±3.0	59.3±5.5	56.0±3.9	57.3±3.3	65.0±2.6	61.5±2.5	**67.6**±2.1
	20	66.0±1.2	68.9±1.8	63.7±1.6	70.9±1.5	71.6±1.2	66.9±2.2	71.1±1.4	67.0±1.7	65.6±2.4	71.0±1.8	70.8±2.1	**71.7**±0.9
PubMed	1	48.6±7.1	47.9±8.5	53.5±13.4	54.7±8.6	50.0±9.5	53.5±11.9	55.3±9.3	46.5±7.0	57.7±10.5	56.7±8.8	49.7±8.3	**60.2**±12.4
	2	55.8±7.1	54.5±7.7	59.7±10.3	58.9±7.2	58.5±8.7	60.7±9.9	62.7±7.0	53.8±6.9	66.3±7.6	62.9±6.3	56.4±6.4	**66.9**±9.7
	3	62.1±7.3	61.5±6.8	61.8±10.4	65.1±6.5	62.4±7.2	65.5±8.9	68.5±5.8	55.6±7.9	71.9±5.4	67.9±5.7	60.6±5.5	**72.3**±6.2
	4	65.1±5.9	64.2±6.1	62.7±10.3	66.0±5.7	64.1±6.2	67.2±8.1	70.6±6.0	57.7±6.8	73.6±5.4	69.9±5.1	62.4±5.1	**73.8**±4.9
	20	79.0±2.5	78.5±1.8	73.3±2.5	78.9±2.6	78.3±2.4	76.8±2.3	79.5±2.2	74.6±3.5	81.5±2.5	80.5±1.6	76.4±2.5	**82.0**±1.4
Coauthor CS	1	64.8±8.8	64.2±9.0	68.4±8.9	79.8±8.0	71.4±6.3	68.3±7.2	75.4±7.2	60.0±7.7	59.9±7.6	76.9±6.2	75.2±7.6	**83.0**±6.2
	2	79.2±4.2	80.2±4.1	77.7±5.3	85.3±4.0	79.6±5.3	78.1±4.5	84.7±2.7	71.3±4.5	72.5±4.6	85.4±3.1	85.3±2.7	**87.4**±4.1
	3	83.3±4.0	85.0±2.7	80.4±4.4	87.5±3.9	82.3±3.6	80.9±4.4	87.5±2.2	74.8±3.8	77.9±4.1	87.4±2.9	87.7±2.3	**88.3**±3.1
	4	84.2±3.1	86.6±2.1	80.9±3.6	87.1±4.6	84.8±2.8	82.8±2.8	88.5±1.8	77.6±2.8	80.3±3.1	88.2±2.1	88.4±1.9	**88.8**±2.4
	20	90.0±0.6	90.9±0.7	83.5±1.4	90.6±1.0	**92.0**±0.5	88.5±0.8	91.5±0.6	90.0±0.7	90.9±1.1	91.2±0.9	91.4±0.6	91.5±0.7
Amazon Photo	1	60.7±9.3	59.0±11.5	70.4±7.2	69.3±5.8	53.8±10.7	58.2±8.1	59.7±9.0	67.0±9.0	55.3±6.7	71.6±6.2	54.4±9.9	**75.6**±6.0
	2	75.2±7.2	71.7±6.4	75.7±4.3	77.2±3.6	62.7±8.5	68.8±6.2	73.4±6.8	76.6±5.2	68.0±5.6	80.7±3.6	71.3±7.2	**81.6**±3.7
	3	76.9±5.1	75.6±6.3	77.0±4.0	79.4±3.9	66.6±7.7	71.9±5.4	76.8±6.1	78.6±4.8	74.4±5.9	82.2±2.7	75.9±5.7	**83.3**±3.8
	4	81.0±4.6	79.3±5.9	80.1±2.6	81.9±2.9	70.8±6.0	76.2±1.8	82.0±2.3	81.8±1.4	78.8±3.9	84.3±1.6	81.5±2.5	**85.3**±2.0
	20	86.3±1.6	86.5±2.1	84.1±1.5	89.4±1.9	83.5±1.2	86.7±1.5	89.7±1.2	87.9±1.4	87.0±1.9	**90.5**±1.9	89.2±1.1	90.2±1.3

Experiments

Table 4. Variants of neighbor contrastive loss in NCLA.

Variants	Cora	CiteSeer	PubMed	CS	Photo
NCL*	**63.1**	**52.2**	**60.2**	**83.0**	**75.6**
InfoNCE	60.9	48.5	59.7	80.4	75.3
NT-Xent	60.1	51.6	60.0	80.5	75.2
NCL* w/o Pos 2	62.4	51.0	59.6	82.7	72.5
NCL* w/o Pos 3	62.3	49.5	59.4	81.4	71.2

*NCL is short for neighbor contrastive loss.

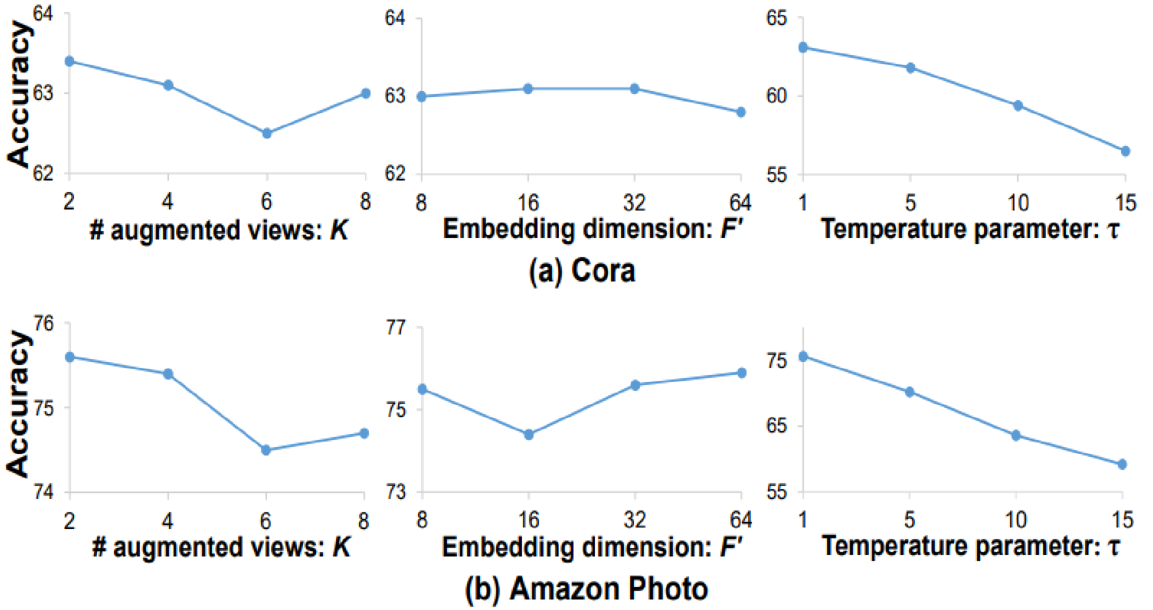


Figure 3. Sensitivity analysis of the hyperparameters K, F' and τ on NCLA.

Thanks