HomoGCL: Rethinking Homophily in Graph Contrastive Learning

Wen-Zhi Li[1,2], Chang-Dong Wang[1], Hui Xiong[2,3], Jian-Huang Lai[1]

[1]CSE, Sun Yat-sen University

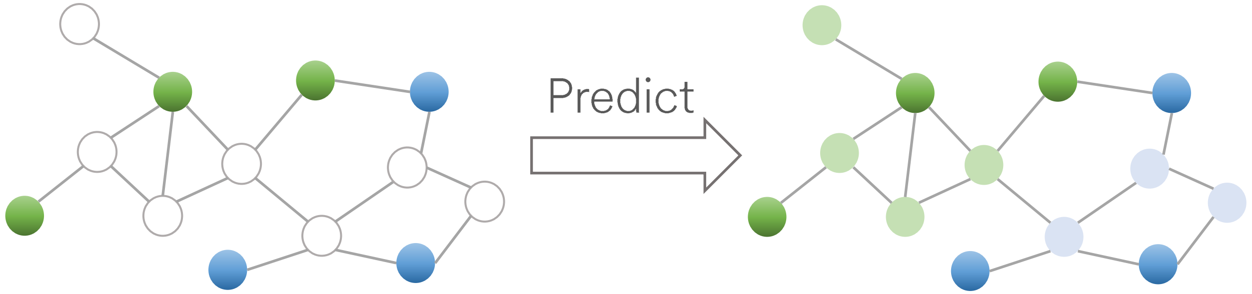
[2]AI Thrust, HKUST(GZ)

[3]CSE, HKUST

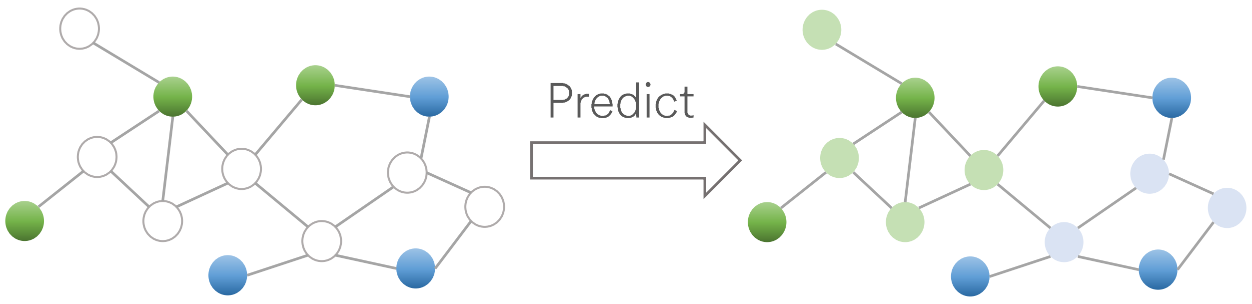
Project Page: https://wenzhilics.github.io/HomoGCL.html

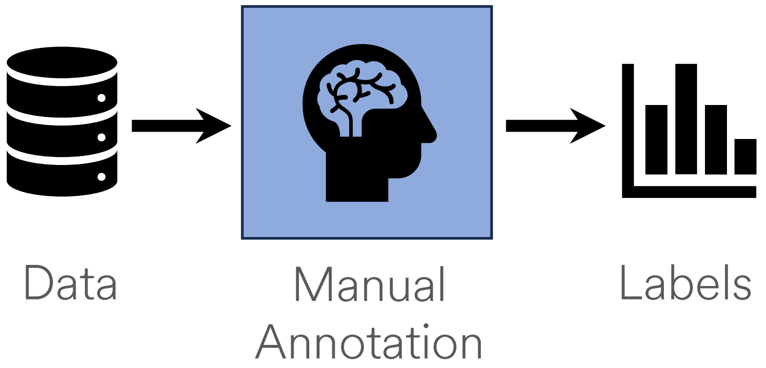
Contact: liwzh63@mail2.sysu.edu.cn

Self-supervised learning on graphs

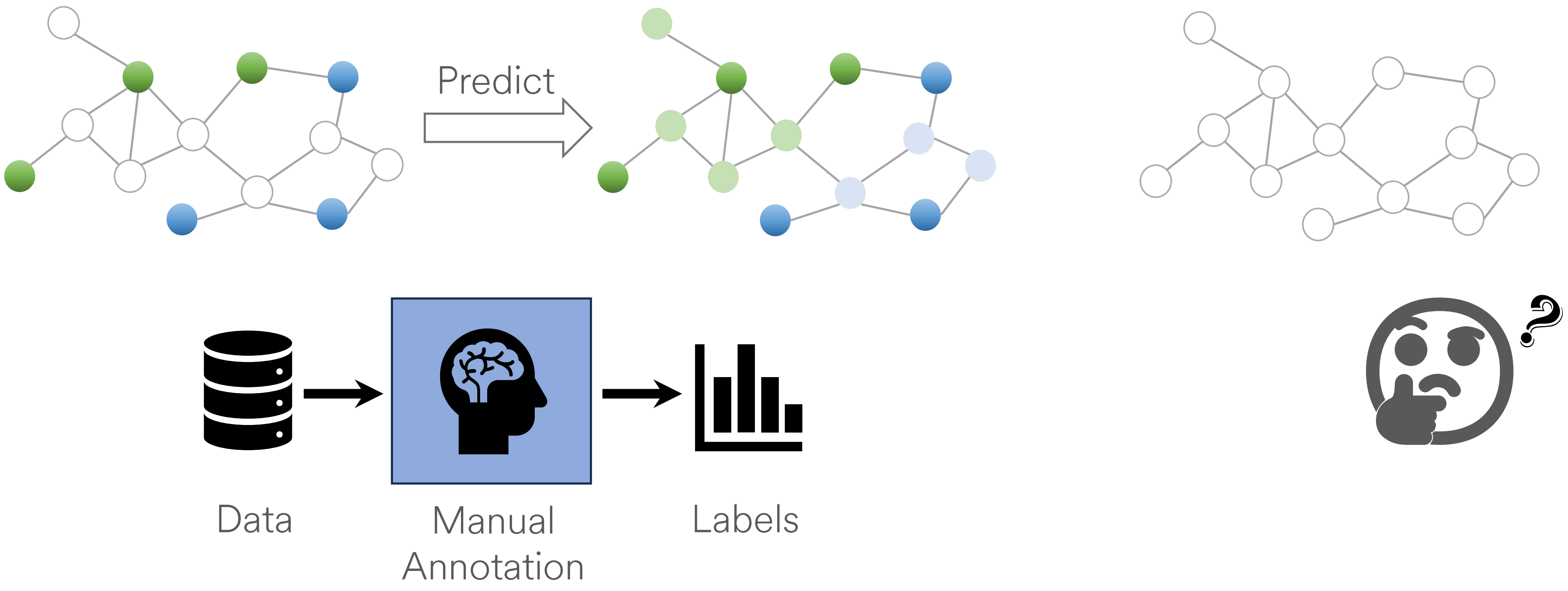


Self-supervised learning on graphs



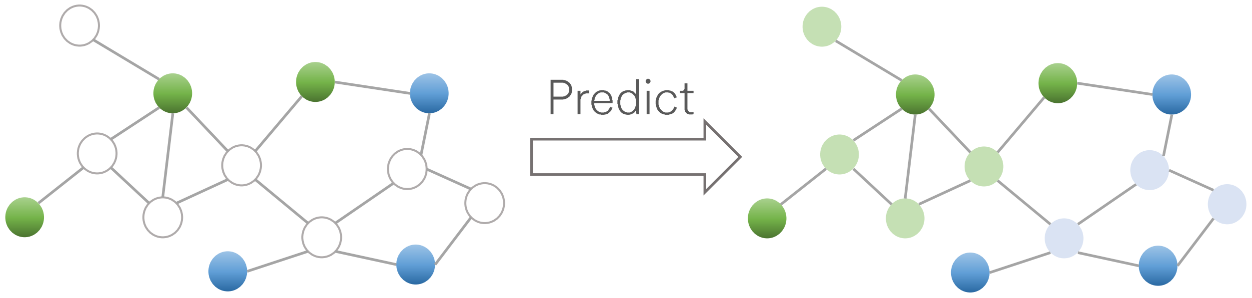


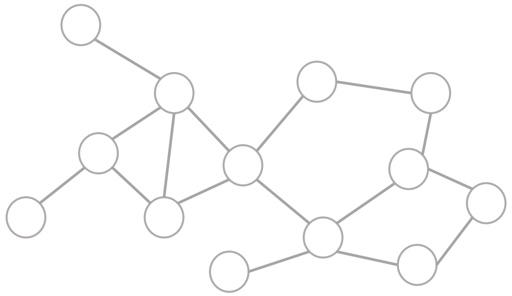
Self-supervised learning on graphs

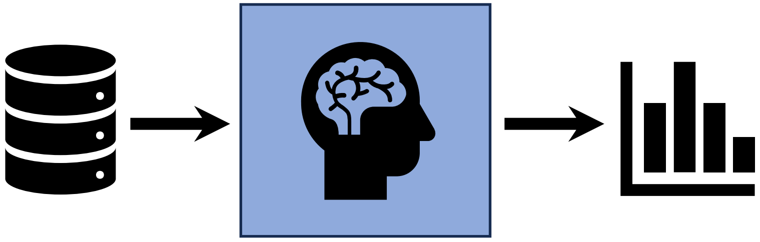


Self-supervised learning: Learning without explicit human annotations.

Self-supervised learning on graphs



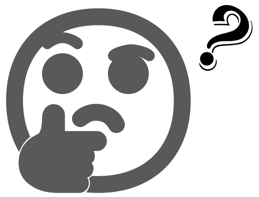




Data

Manual Annotation

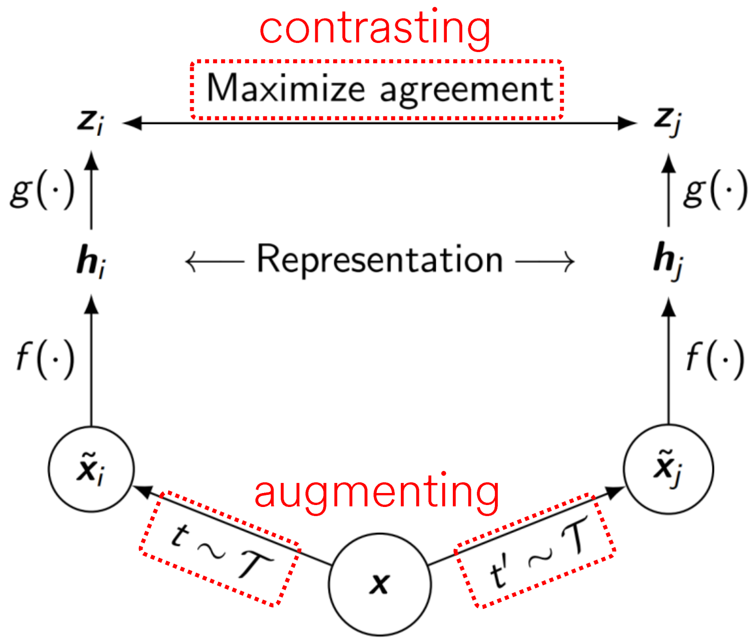
Labels



Self-supervised learning: Learning without explicit human annotations.

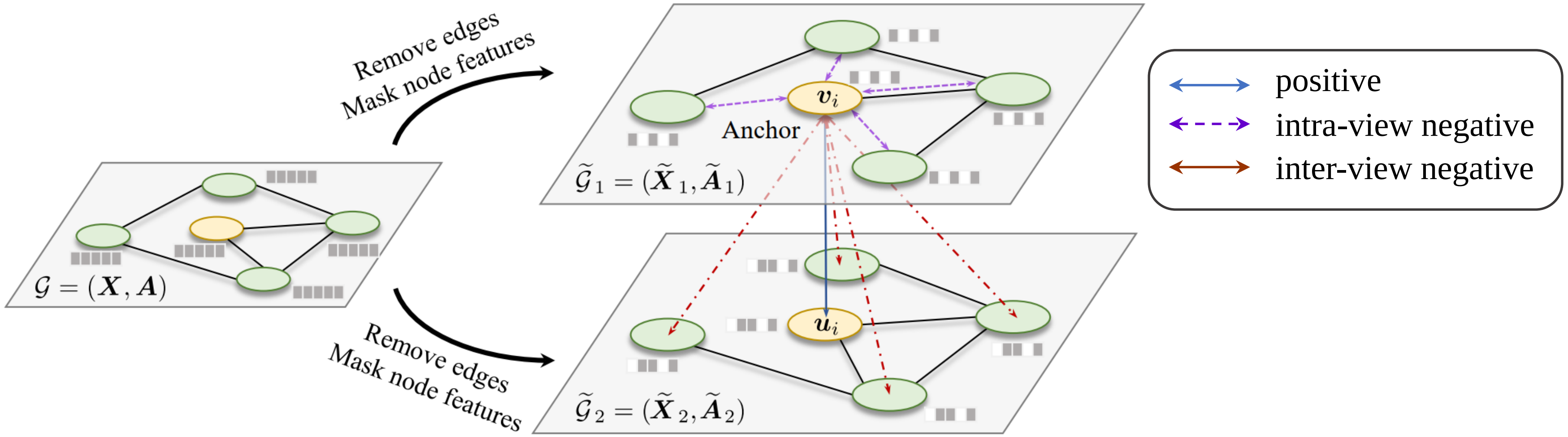
→ Graph contrastive learning!

Contrastive learning



- The "**augmenting-contrasting**" paradigm.
- **Maximizing** the similarities between **positive** pairs;
- **Minimizing** the similarities between **negative** pairs.

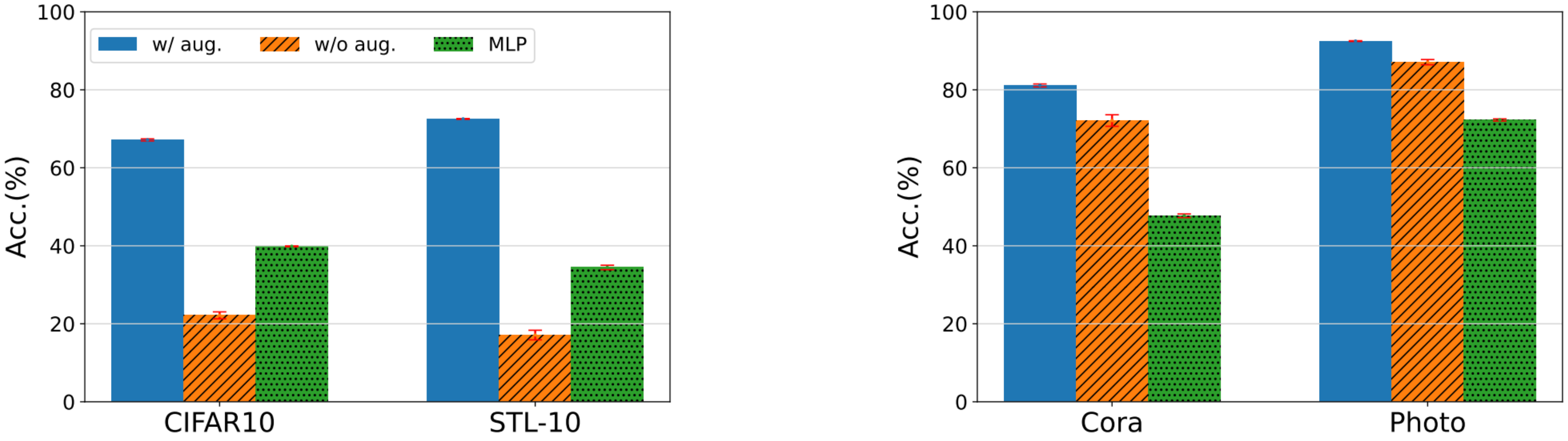
Graph contrastive learning



$$\mathcal{L} = \frac{1}{2N}\sum_{i=1}^{N}\left(\ell\left(\boldsymbol{u}_i, \boldsymbol{v}_i\right) + \ell\left(\boldsymbol{v}_i, \boldsymbol{u}_i\right)\right), \ \text{with}$$

$$\ell\left(\boldsymbol{u}_i, \boldsymbol{v}_i\right) = \log \frac{e^{\theta(\boldsymbol{u}_i, \boldsymbol{v}_i)/\tau}}{\underbrace{e^{\theta(\boldsymbol{u}_i, \boldsymbol{v}_i)/\tau}}_{\textbf{positive}} + \underbrace{\sum_{j \neq i} e^{\theta(\boldsymbol{u}_i, \boldsymbol{v}_j)/\tau}}_{\textbf{inter-view negative}} + \underbrace{\sum_{j \neq i} e^{\theta(\boldsymbol{u}_i, \boldsymbol{u}_j)/\tau}}_{\textbf{intra-view negative}}},$$

Motivation

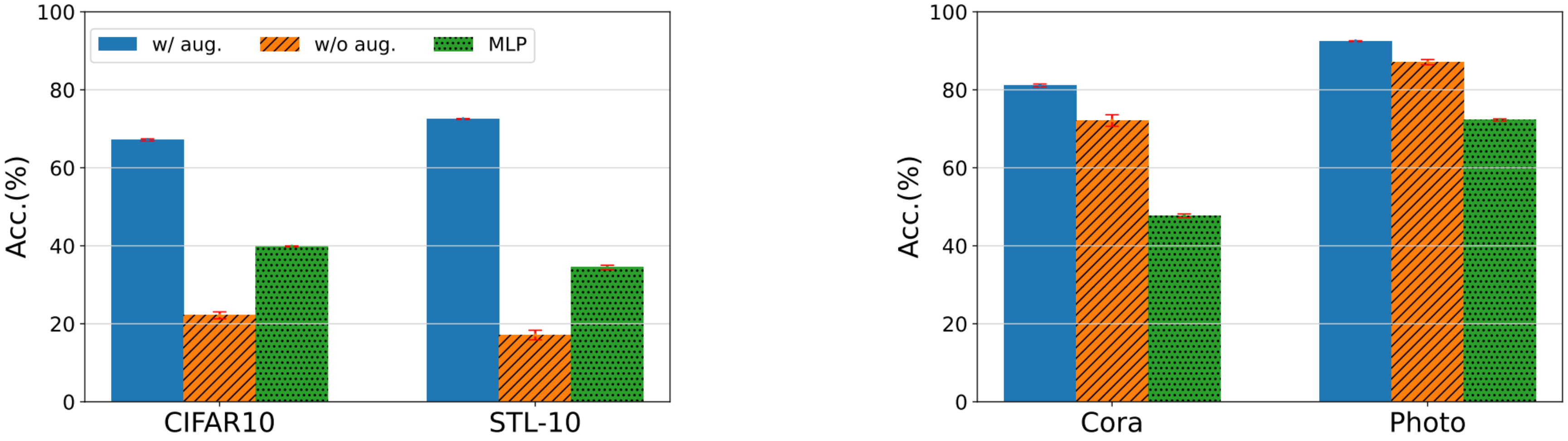


SimCLR GRACE

Performance of CL in vision and graph domains with/without augmentation.

- GCL without augmentation can also achieve decent performance, which is quite different from VCL.

Motivation



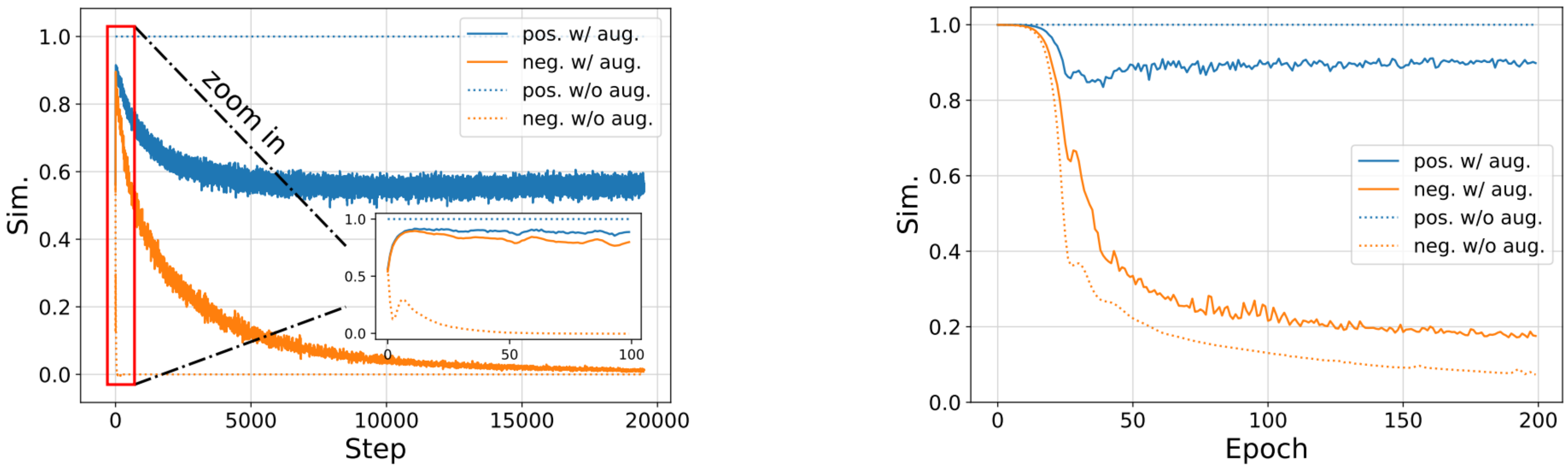
SimCLR GRACE

Performance of CL in vision and graph domains with/without augmentation.

- GCL without augmentation can also achieve decent performance, which is quite different from VCL.

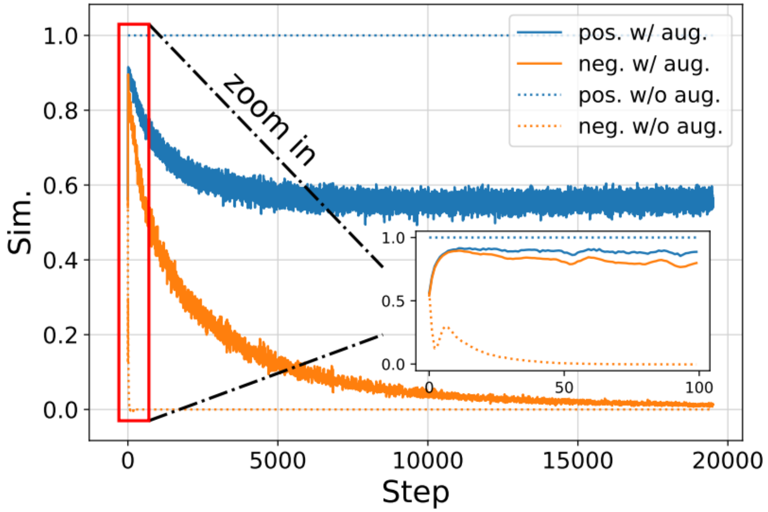
What causes the huge gap between the performance declines of GCL and VCL when data augmentation is not leveraged?

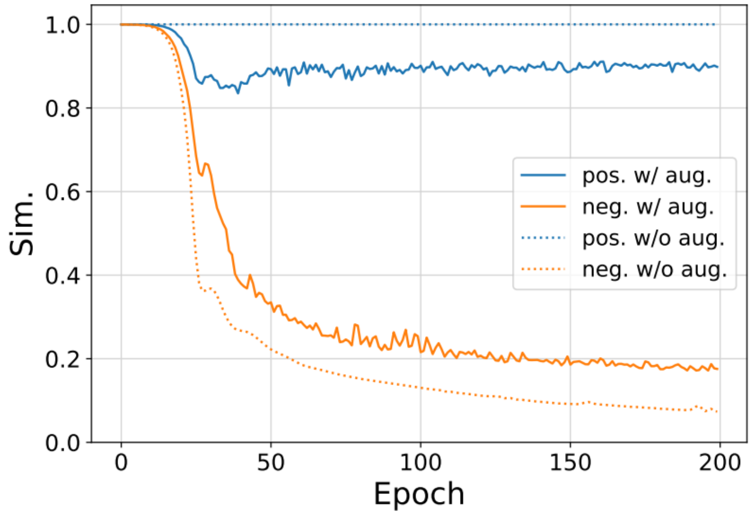
Empirical study



Similarity histogram on CIFAR10 (vision) and Cora (graph).

Empirical study

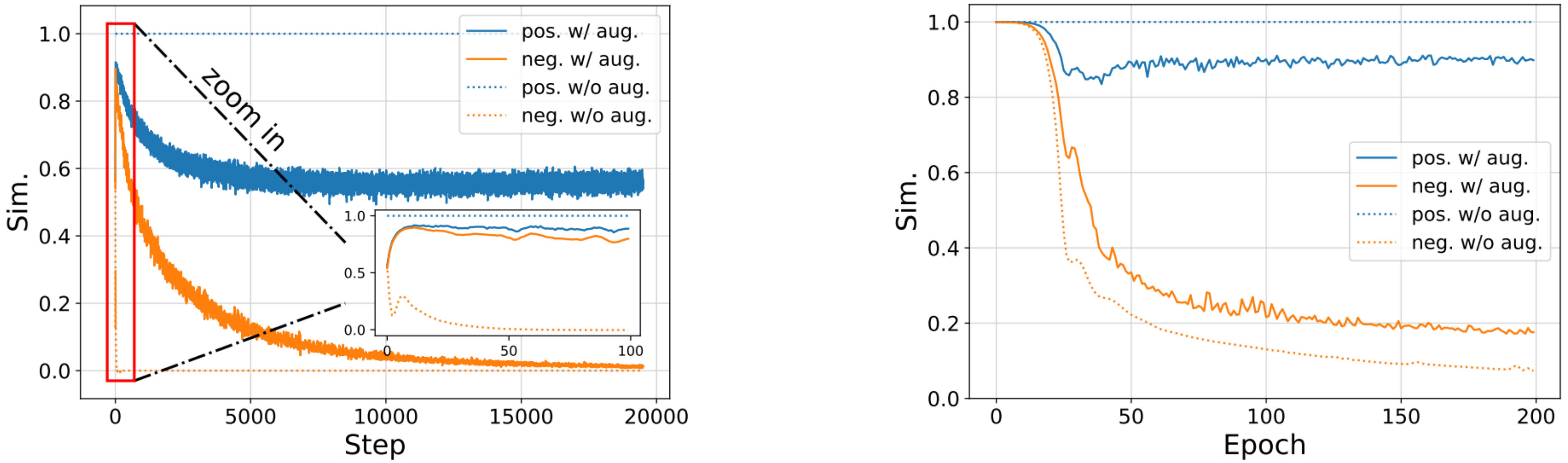




Similarity histogram on CIFAR10 (vision) and Cora (graph).

- The similarity between negative pairs drops to 0 swiftly on CIFAR10 w/o aug.
- The similarity between negative pairs drops gradually on Cora w/o aug.

Empirical study

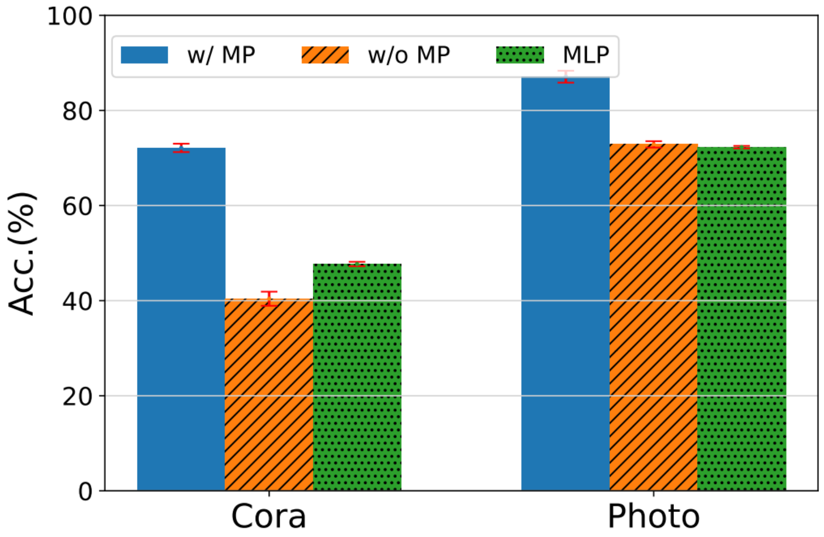


Similarity histogram on CIFAR10 (vision) and Cora (graph).

- The similarity between negative pairs drops to 0 swiftly on CIFAR10 w/o aug.
- The similarity between negative pairs drops gradually on Cora w/o aug.

- Trivial discrimination for CIFAR10 w/o aug.
- Message passing in GNN avoids the trivial discrimination for Cora w/o aug.

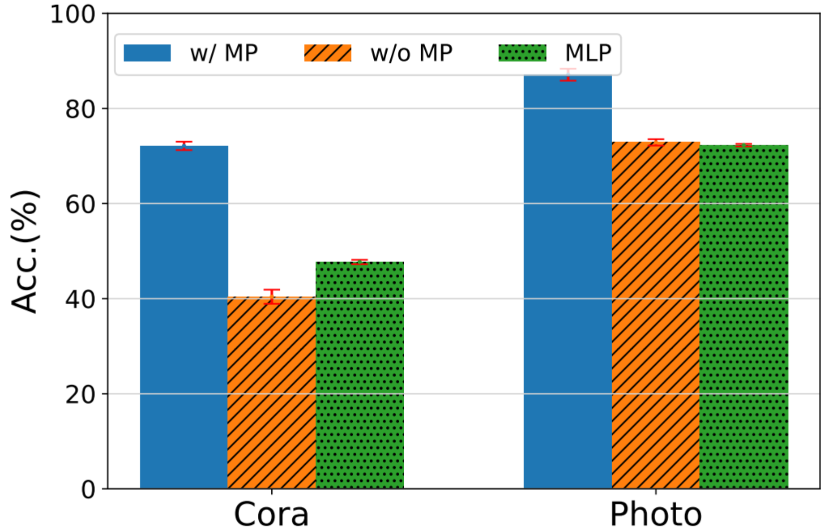
Empirical study



Ablation study on two graph datasets Cora and Photo.

- GRACE (w/o MP) is only on par with or even worse than MLP.
- GRACE (w/ MP) outperforms w/o MP and MLP by a large margin.

Empirical study

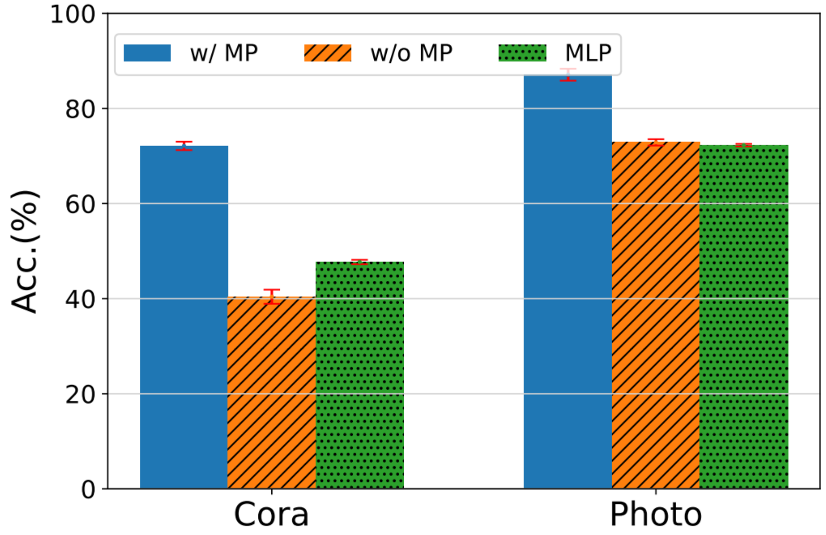


Ablation study on two graph datasets Cora and Photo.

- GRACE (w/o MP) is only on par with or even worse than MLP.
- GRACE (w/ MP) outperforms w/o MP and MLP by a large margin.

- Nodes in GRACE (w/o MP) cannot propagate features to their neighbors, which degenerates them to a similar situation of VCL w/o aug.
- **Message passing** which relies on the **homophily** assumption is the key factor of GCL.

Empirical study



Ablation study on two graph datasets Cora and Photo.

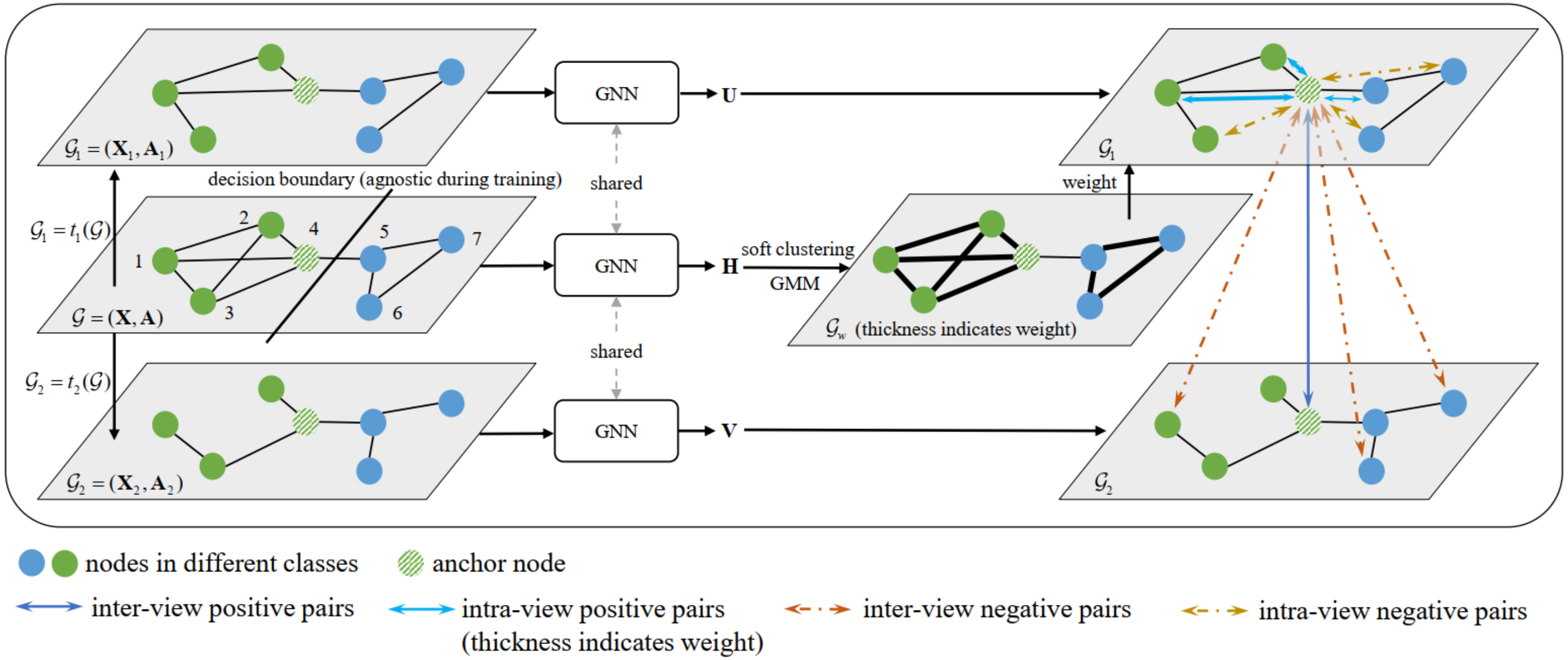
- GRACE (w/o MP) is only on par with or even worse than MLP.
- GRACE (w/ MP) outperforms w/o MP and MLP by a large margin.

- Nodes in GRACE (w/o MP) cannot propagate features to their neighbors, which degenerates them to a similar situation of VCL w/o aug.
- **Message passing** which relies on the **homophily** assumption is the key factor of GCL.

Homophily: The fraction of intra-class edges in a graph.

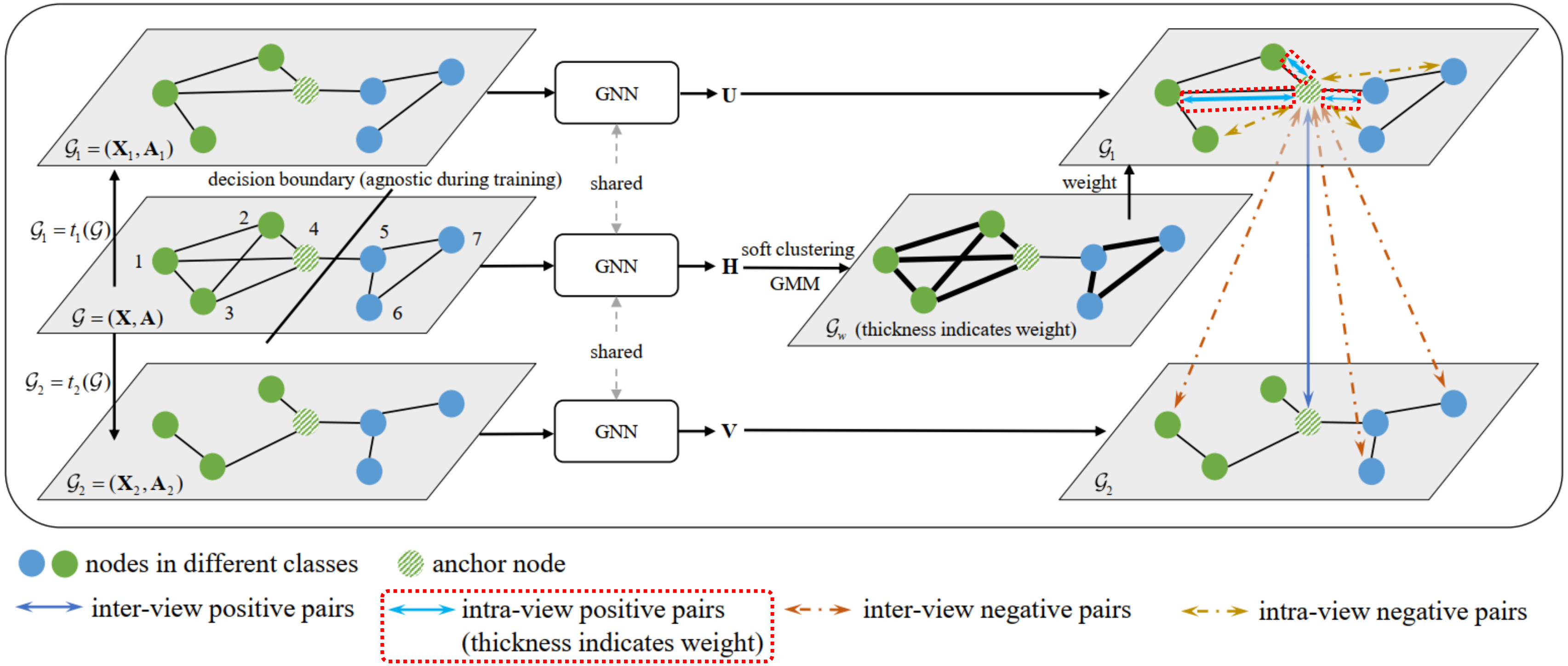
HomoGCL by leveraging homophily

- Overview



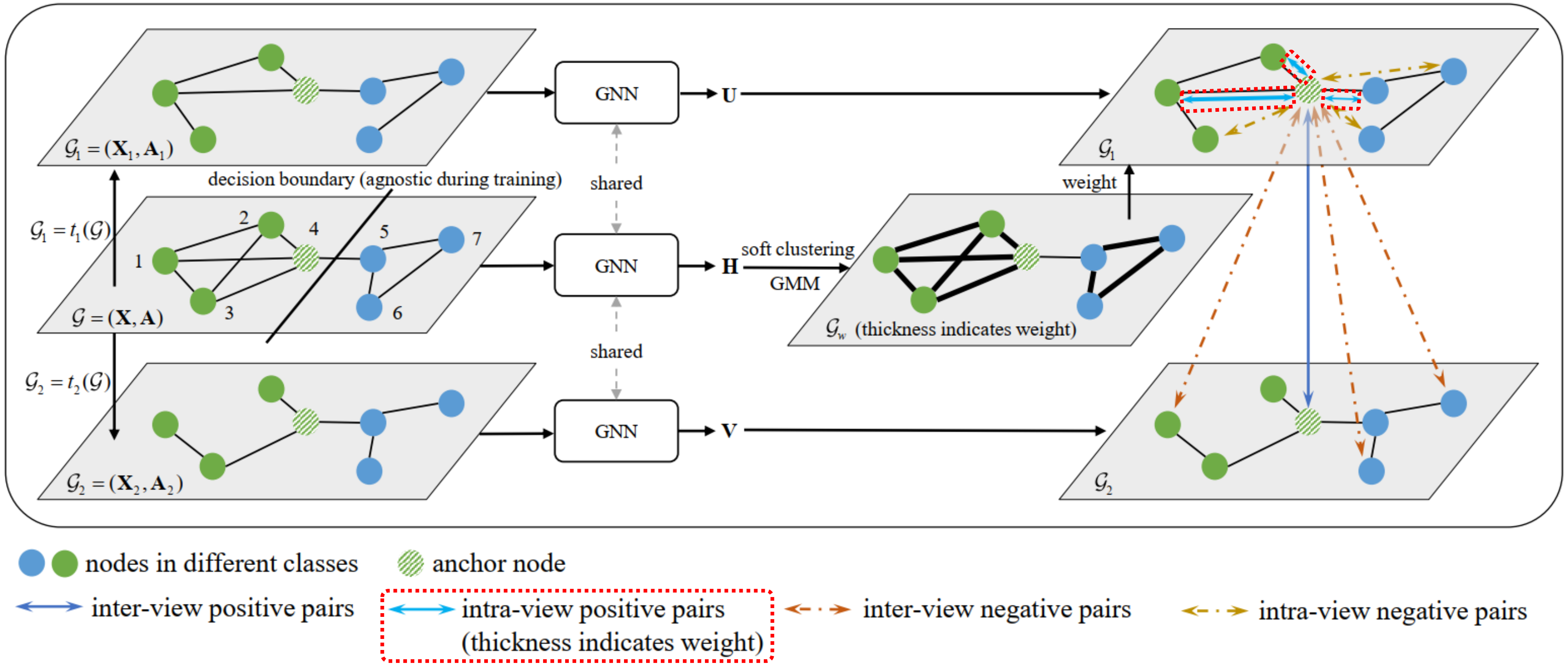
HomoGCL by leveraging homophily

- Overview



HomoGCL by leveraging homophily

- ## Overview



$$\text{pos} = \underbrace{e^{\theta(\boldsymbol{u}_i, \boldsymbol{v}_i)/\tau}}_{\text{inter-view positive pair}} + \underbrace{\sum_{j \in \mathcal{N}_{\boldsymbol{u}}(i)} e^{\theta(\boldsymbol{u}_i, \boldsymbol{u}_j)/\tau} \cdot \underbrace{\mathsf{S}_{ij}}_{\text{saliency}}}_{\text{intra-view positive pairs}}$$

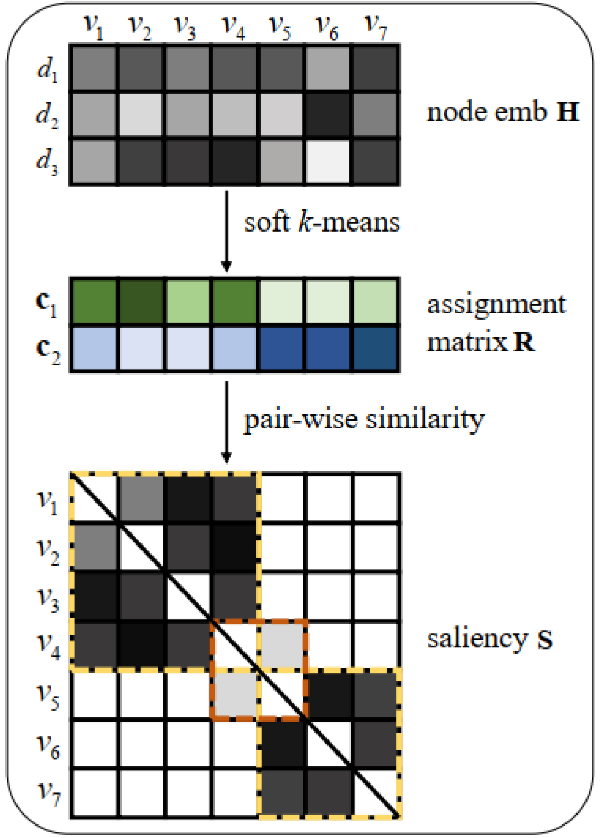
HomoGCL by leveraging homophily

- **Soft clustering**

$$\text{pos} = \underbrace{e^{\theta(\boldsymbol{u}_i,\boldsymbol{v}_i)/\tau}}_{\text{inter-view positive pair}} + \underbrace{\sum_{j\in\mathcal{N}_{\boldsymbol{u}}(i)} e^{\theta(\boldsymbol{u}_i,\boldsymbol{u}_j)/\tau} \cdot \underbrace{\mathsf{S}_{ij}}_{\text{saliency}}}_{\text{intra-view positive pairs}}$$

HomoGCL by leveraging homophily

- Soft clustering



far from the decision boundary

near the decision boundary

$$\text{pos} = \underbrace{e^{\theta(\boldsymbol{u}_i,\boldsymbol{v}_i)/\tau}}_{\text{inter-view positive pair}} + \underbrace{\sum_{j \in \mathcal{N}_{\boldsymbol{u}}(i)} e^{\theta(\boldsymbol{u}_i,\boldsymbol{u}_j)/\tau} \cdot \underbrace{\mathsf{S}_{ij}}_{\text{saliency}}}_{\text{intra-view positive pairs}}$$

$$\begin{cases} p\left(\boldsymbol{h}_i \mid \boldsymbol{c}_j\right) = \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\dfrac{\left\|\boldsymbol{h}_i - \boldsymbol{c}_j\right\|_2}{2\sigma^2}\right), \quad p(\boldsymbol{c}_1) = p(\boldsymbol{c}_2) = \cdots = p(\boldsymbol{c}_k) \\ \mathbf{R}_{ij} = p\left(\boldsymbol{c}_j \mid \boldsymbol{h}_i\right) = \dfrac{p\left(\boldsymbol{c}_j\right) p\left(\boldsymbol{h}_i \mid \boldsymbol{c}_j\right)}{\sum_{r=1}^{k} p\left(\boldsymbol{c}_r\right) p\left(\boldsymbol{h}_i \mid \boldsymbol{c}_r\right)} \\ \mathsf{S}_{ij} = \operatorname{norm}(\mathbf{R}_i) \cdot \operatorname{norm}(\mathbf{R}_j^{\top}) \end{cases}$$

HomoGCL by leveraging homophily

- **Loss function**

Contrastive loss

$$
\begin{cases}
\ell_{cont}(\boldsymbol{u}_i, \boldsymbol{v}_i) = \log \dfrac{\text{pos}}{\text{pos} + \text{neg}} \\
\text{pos} = \underbrace{e^{\theta(\boldsymbol{u}_i, \boldsymbol{v}_i)/\tau}}_{\text{inter-view positive pair}} + \underbrace{\sum_{j \in \mathcal{N}_{\boldsymbol{u}}(i)} e^{\theta(\boldsymbol{u}_i, \boldsymbol{u}_j)/\tau} \cdot \mathbf{S}_{ij}}_{\text{intra-view positive pairs}}, \\
\text{neg} = \underbrace{\sum_{j \notin \{i \cup \mathcal{N}_{\boldsymbol{v}}(i)\}} e^{\theta(\boldsymbol{u}_i, \boldsymbol{v}_j)/\tau}}_{\text{inter-view negative pairs}} + \underbrace{\sum_{j \notin \{i \cup \mathcal{N}_{\boldsymbol{u}}(i)\}} e^{\theta(\boldsymbol{u}_i, \boldsymbol{u}_j)/\tau}}_{\text{intra-view negative pairs}}, \\
\mathcal{L}_{cont} = \dfrac{1}{2N} \sum_{i=1}^{N} \left(\ell_{cont}(\boldsymbol{u}_i, \boldsymbol{v}_i) + \ell_{cont}(\boldsymbol{v}_i, \boldsymbol{u}_i) \right)
\end{cases}
$$

HomoGCL by leveraging homophily

- **Loss function**

Contrastive loss

$$
\begin{cases}
\ell_{cont}(\boldsymbol{u}_i, \boldsymbol{v}_i) = \log \dfrac{\text{pos}}{\text{pos} + \text{neg}} \\
\boxed{\text{pos} = \underbrace{e^{\theta(\boldsymbol{u}_i, \boldsymbol{v}_i)/\tau}}_{\text{inter-view positive pair}} + \underbrace{\sum_{j \in \mathcal{N}_{\boldsymbol{u}}(i)} e^{\theta(\boldsymbol{u}_i, \boldsymbol{u}_j)/\tau} \cdot \mathbf{S}_{ij}}_{\text{intra-view positive pairs}},} \\
\text{neg} = \underbrace{\sum_{j \notin \{i \cup \mathcal{N}_{\boldsymbol{v}}(i)\}} e^{\theta(\boldsymbol{u}_i, \boldsymbol{v}_j)/\tau}}_{\text{inter-view negative pairs}} + \underbrace{\sum_{j \notin \{i \cup \mathcal{N}_{\boldsymbol{u}}(i)\}} e^{\theta(\boldsymbol{u}_i, \boldsymbol{u}_j)/\tau}}_{\text{intra-view negative pairs}}, \\
\mathcal{L}_{cont} = \dfrac{1}{2N} \sum_{i=1}^{N} \left(\ell_{cont}(\boldsymbol{u}_i, \boldsymbol{v}_i) + \ell_{cont}(\boldsymbol{v}_i, \boldsymbol{u}_i) \right)
\end{cases}
$$

$\rightarrow$ Enlarge the positive set

HomoGCL by leveraging homophily

- **Loss function**

Contrastive loss

$$
\begin{cases}
\ell_{cont}(\boldsymbol{u}_i, \boldsymbol{v}_i) = \log \dfrac{\text{pos}}{\text{pos} + \text{neg}} \\
\boxed{\text{pos} = \underbrace{e^{\theta(\boldsymbol{u}_i, \boldsymbol{v}_i)/\tau}}_{\text{inter-view positive pair}} + \underbrace{\sum_{j \in \mathcal{N}_{\boldsymbol{u}}(i)} e^{\theta(\boldsymbol{u}_i, \boldsymbol{u}_j)/\tau} \cdot \mathcal{S}_{ij}}_{\text{intra-view positive pairs}},} \\
\text{neg} = \underbrace{\sum_{j \notin \{i \cup \mathcal{N}_{\boldsymbol{v}}(i)\}} e^{\theta(\boldsymbol{u}_i, \boldsymbol{v}_j)/\tau}}_{\text{inter-view negative pairs}} + \underbrace{\sum_{j \notin \{i \cup \mathcal{N}_{\boldsymbol{u}}(i)\}} e^{\theta(\boldsymbol{u}_i, \boldsymbol{u}_j)/\tau}}_{\text{intra-view negative pairs}}, \\
\mathcal{L}_{cont} = \dfrac{1}{2N} \sum_{i=1}^{N} \left(\ell_{cont}(\boldsymbol{u}_i, \boldsymbol{v}_i) + \ell_{cont}(\boldsymbol{v}_i, \boldsymbol{u}_i)\right)
\end{cases}
$$

Enlarge the positive set

Homophily loss

$$
\mathcal{L}_{homo} = \frac{1}{k|\mathcal{E}|} \sum_{r=1}^{k} \sum_{(v_i, v_j) \in \mathcal{E}} \text{MSE}\left(p\left(\boldsymbol{c}_r \mid \boldsymbol{h}_i\right), p\left(\boldsymbol{c}_r \mid \boldsymbol{h}_j\right)\right)
$$

HomoGCL by leveraging homophily

- **Loss function**

Contrastive loss

$$\begin{cases} \ell_{cont}(\boldsymbol{u}_i, \boldsymbol{v}_i) = \log \dfrac{\text{pos}}{\text{pos} + \text{neg}} \\ \boxed{\text{pos} = \underbrace{e^{\theta(\boldsymbol{u}_i, \boldsymbol{v}_i)/\tau}}_{\text{inter-view positive pair}} + \underbrace{\sum_{j \in \mathcal{N}_{\boldsymbol{u}}(i)} e^{\theta(\boldsymbol{u}_i, \boldsymbol{u}_j)/\tau} \cdot \mathcal{S}_{ij}}_{\text{intra-view positive pairs}},} \\ \text{neg} = \underbrace{\sum_{j \notin \{i \cup \mathcal{N}_{\boldsymbol{v}}(i)\}} e^{\theta(\boldsymbol{u}_i, \boldsymbol{v}_j)/\tau}}_{\text{inter-view negative pairs}} + \underbrace{\sum_{j \notin \{i \cup \mathcal{N}_{\boldsymbol{u}}(i)\}} e^{\theta(\boldsymbol{u}_i, \boldsymbol{u}_j)/\tau}}_{\text{intra-view negative pairs}}, \\ \mathcal{L}_{cont} = \dfrac{1}{2N} \sum_{i=1}^{N} \left(\ell_{cont}(\boldsymbol{u}_i, \boldsymbol{v}_i) + \ell_{cont}(\boldsymbol{v}_i, \boldsymbol{u}_i)\right) \end{cases}$$

→ Enlarge the positive set

Homophily loss

$$\mathcal{L}_{homo} = \frac{1}{k|\mathcal{E}|} \sum_{r=1}^{k} \sum_{(v_i, v_j) \in \mathcal{E}} \operatorname{MSE}\left(p\left(\boldsymbol{c}_r \mid \boldsymbol{h}_i\right), p\left(\boldsymbol{c}_r \mid \boldsymbol{h}_j\right)\right)$$

$$\mathcal{J} = \mathcal{L}_{cont} + \alpha \mathcal{L}_{homo}$$

HomoGCL by leveraging homophily

- **Theoretical Insights**

The newly proposed contrastive loss $\mathcal{L}_{cont}$ is a stricter lower bound of MI between raw node features X *and node embeddings* U *and* V *in two augmented views, comparing with the raw contrastive loss* $\mathcal{L}$ *proposed by GRACE. Formally,*

$$\mathcal{L} \leq \mathcal{L}_{cont} \leq I(\mathrm{X}; \mathrm{U}, \mathrm{V}).$$

HomoGCL by leveraging homophily

- **Theoretical Insights**

The newly proposed contrastive loss $\mathcal{L}_{cont}$ is a stricter lower bound

Please refer to the paper for proof!

$$\mathcal{L} \leq \mathcal{L}_{cont} \leq I(X; U, V).$$

Node classification

Model
Raw features
DeepWalk
Node2Vec
GCN
GAT
GAE
VGAE
DGI
GMI
InfoGCL
MVGRL
BGRL
AFGRL
COSTA
CCA-SSG
GRACE
GCA
ProGCL
ARIEL
HomoGCL

Unsupervised methods

Supervised methods

Graph autoencoders

Contrastive learning methods

Node classification

Model
Raw features
DeepWalk
Node2Vec
GCN
GAT
GAE
VGAE
DGI
GMI
InfoGCL
MVGRL
BGRL
AFGRL
COSTA
CCA-SSG
GRACE
GCA
ProGCL
ARIEL
HomoGCL

GRACE-based methods

Node classification

Model	Training Data	Cora	CiteSeer	PubMed	Photo	Computer
Raw features	X, Y	47.7±0.4	46.5±0.4	71.4±0.2	72.27±0.00	73.81±0.00
DeepWalk	A	70.7±0.6	51.4±0.5	74.3±0.9	89.44±0.11	85.68±0.06
Node2Vec	A	70.1±0.4	49.8±0.3	69.8±0.7	87.76±0.10	84.39±0.08
GCN	X, A, Y	81.5±0.4	70.2±0.4	79.0±0.2	92.42±0.22	86.51±0.54
GAT	X, A, Y	83.0±0.7	72.5±0.7	79.0±0.3	92.56±0.35	86.93±0.29
GAE	X, A	71.5±0.4	65.8±0.4	72.1±0.5	91.62±0.13	85.27±0.19
VGAE	X, A	73.0±0.3	68.3±0.4	75.8±0.2	92.20±0.11	86.37±0.21
DGI	X, A	82.3±0.6	71.8±0.7	76.8±0.6	91.61±0.22	83.95±0.47
GMI	X, A	83.0±0.3	72.4±0.1	79.9±0.2	90.68±0.17	82.21±0.31
InfoGCL	X, A	83.5±0.3	**73.5**±0.4	79.1±0.2	-	-
MVGRL	X, A	83.5±0.4	73.3±0.5	80.1±0.7	91.74±0.07	87.52±0.11
BGRL	X, A	82.7±0.6	71.1±0.8	79.6±0.5	92.80±0.08	88.23±0.11
AFGRL	X, A	79.8±0.2	69.4±0.2	80.0±0.1	92.71±0.23	88.12±0.27
COSTA	X, A	82.2±0.2	70.7±0.5	80.4±0.3	92.43±0.38	88.37±0.22
CCA-SSG	X, A	84.0±0.4	73.1±0.3	81.0±0.4	92.84±0.18	88.27±0.32
GRACE	X, A	81.5±0.3	70.6±0.5	80.2±0.3	92.15±0.24	86.25±0.25
GCA	X, A	81.4±0.3(↓0.1)	70.4±0.4(↓0.2)	80.7±0.5(↑0.5)	92.53±0.16(↑0.38)	87.80±0.23(↑1.55)
ProGCL	X, A	81.2±0.4(↓0.3)	69.8±0.5(↓0.8)	79.2±0.2(↓1.0)	92.39±0.11(↑0.24)	87.43±0.21(↑1.18)
ARIEL	X, A	83.0±1.3(↑1.5)	71.1±0.9(↑0.5)	74.2±0.8(↓6.0)	91.80±0.24(↓0.35)	87.07±0.33(↑0.82)
HomoGCL	X, A	**84.5**±0.5(↑3.0)	72.3±0.7(↑1.7)	**81.1**±0.3(↑0.9)	**92.92**±0.18(↑0.77)	**88.46**±0.20(↑2.21)

[1] The results not reported are due to unavailable code.

Node clustering

Dataset	Photo		Computer	
Metric	NMI	ARI	NMI	ARI
GAE	$0.616 \pm \Delta_1$	$0.494 \pm \Delta_1$	$0.441 \pm \Delta_0$	$0.258 \pm \Delta_0$
VGAE	$0.530 \pm \Delta_4$	$0.373 \pm \Delta_4$	$0.423 \pm \Delta_0$	$0.238 \pm \Delta_0$
DGI	$0.376 \pm \Delta_3$	$0.264 \pm \Delta_3$	$0.318 \pm \Delta_2$	$0.165 \pm \Delta_2$
HDI	$0.429 \pm \Delta_1$	$0.307 \pm \Delta_1$	$0.347 \pm \Delta_1$	$0.216 \pm \Delta_6$
MVGRL	$0.344 \pm \Delta_4$	$0.239 \pm \Delta_4$	$0.244 \pm \Delta_0$	$0.141 \pm \Delta_0$
BGRL	$0.668 \pm \Delta_3$	$0.547 \pm \Delta_4$	$0.484 \pm \Delta_0$	$0.295 \pm \Delta_0$
AFGRL	$0.618 \pm \Delta_1$	$0.497 \pm \Delta_3$	$0.478 \pm \Delta_3$	$0.334 \pm \Delta_4$
GCA	$0.614 \pm \Delta_0$	$0.494 \pm \Delta_0$	$0.426 \pm \Delta_0$	$0.246 \pm \Delta_0$
gCooL	$0.632 \pm \Delta_0$	$0.524 \pm \Delta_0$	$0.474 \pm \Delta_2$	$0.277 \pm \Delta_2$
HomoGCL	$\mathbf{0.671} \pm \Delta_2$	$\mathbf{0.587} \pm \Delta_2$	$\mathbf{0.534} \pm \Delta_0$	$\mathbf{0.396} \pm \Delta_0$

$\Delta_x = 0.01x$

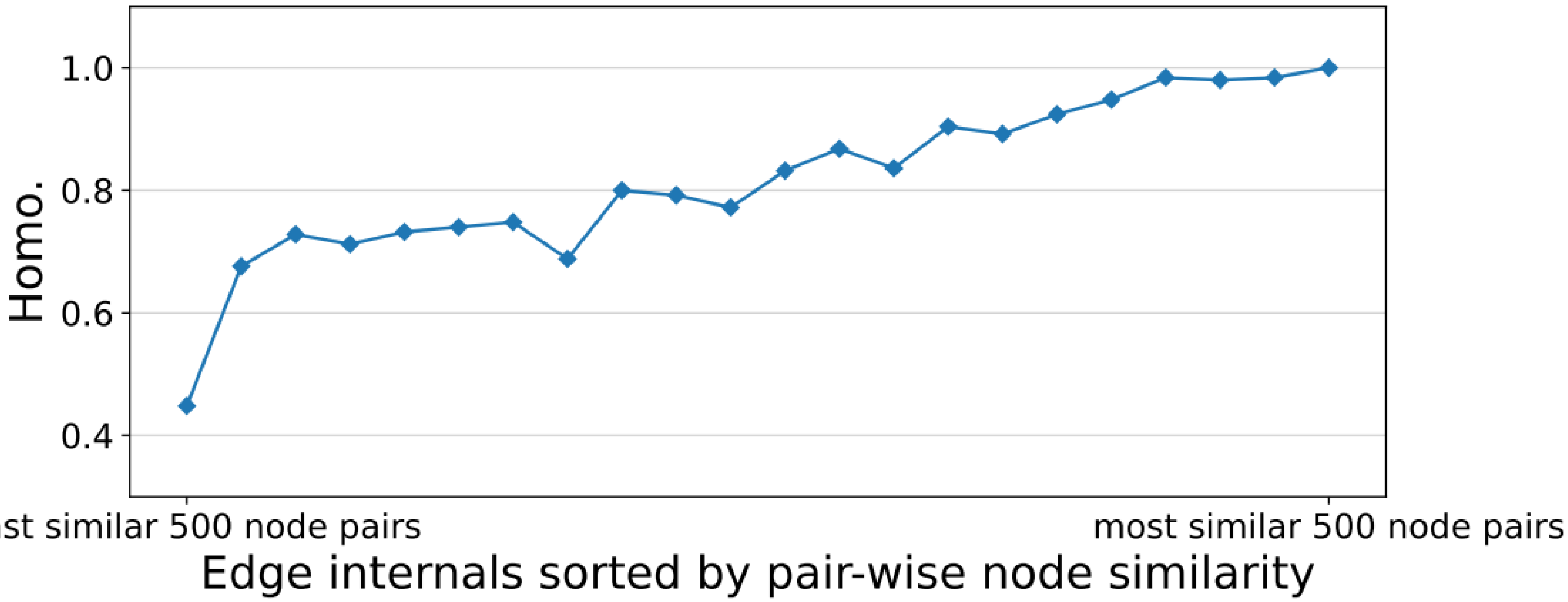
Improving negative-free method BGRL

Model	PubMed	Photo	Computer
BGRL	79.6	92.80	88.23
+**HomoGCL**	80.8(↑1.2)	93.53(↑0.73)	90.01(↑1.79)

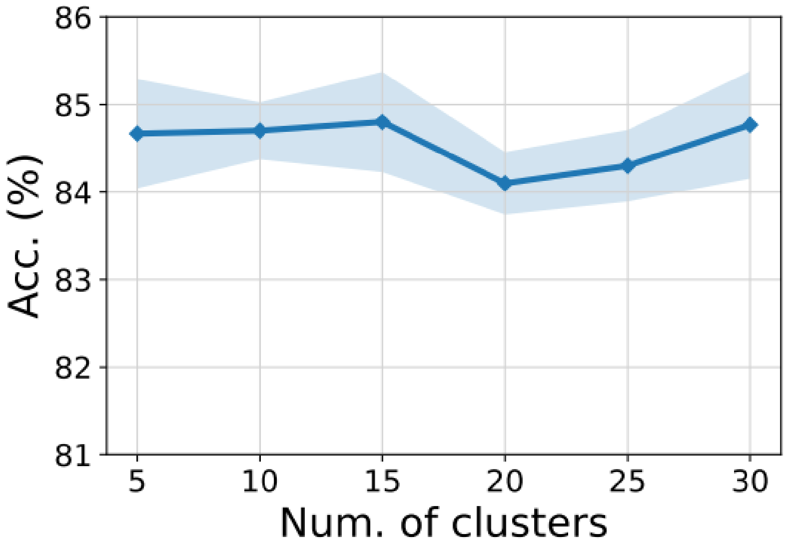
Node classification on large-scale dataset

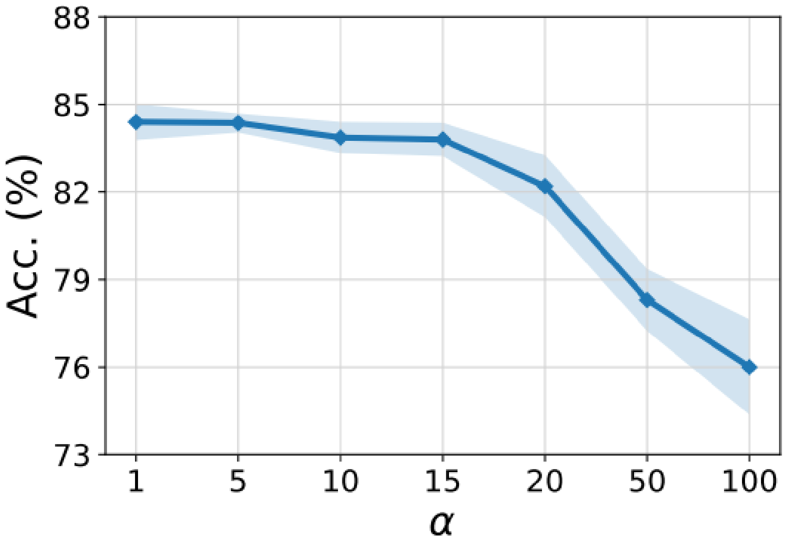
Model	Validation	Test
MLP	57.65±0.12	55.50±0.23
node2vec	71.29±0.13	70.07±0.13
GCN	73.00±0.17	71.74±0.29
GraphSAGE	72.77±0.16	71.49±0.27
Random-Init	69.90±0.11	68.94±0.15
DGI	71.26±0.11	70.34±0.16
G-BT	71.16±0.14	70.12±0.18
GRACE full-graph	OOM	OOM
GRACE-Subsampling (k=2)	60.49±3.72	60.24±4.06
GRACE-Subsampling (k=8)	71.30±0.17	70.33±0.18
GRACE-Subsampling (k=2048)	72.61±0.15	71.51±0.11
ProGCL	72.45±0.21	72.18±0.09
BGRL	72.53±0.09	71.64±0.12
HomoGCL	**72.85**±0.10	**72.22**±0.15

Case study

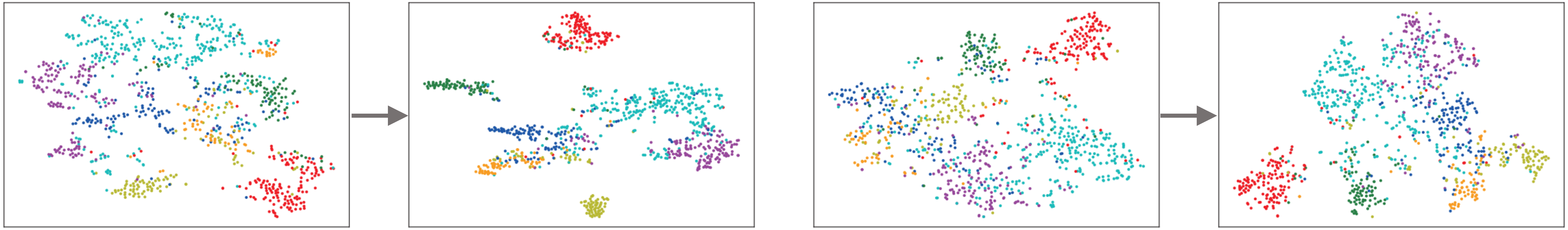

Homo.
1.0
0.8
0.6
0.4
least similar 500 node pairs
most similar 500 node pairs
Edge internals sorted by pair-wise node similarity

Hyper-parameter analysis





Visualization



(a) GRACE (b) GRACE + HomoGCL (c) BGRL (d) BGRL + HomoGCL

Summary

- **Empirical study**: graph homophily plays a key role in GCL.
- **New technique**: HomoGCL to leverage graph homophily explicitly.
- **Experiments**: HomoGCL can be combined with existing GCL models in a plug-and-play way to boost the performance.

Summary

- **Empirical study**: graph homophily plays a key role in GCL.
- **New technique**: HomoGCL to leverage graph homophily explicitly.
- **Experiments**: HomoGCL can be combined with existing GCL models in a plug-and-play way to boost the performance.

Check out our paper and code at...

- **Paper**: https://arxiv.org/abs/2306.09614
- **Project Page**: https://wenzhilics.github.io/HomoGCL.html
- **Code**: https://github.com/wenzhilics/HomoGCL

THANKS

Q&A

Wen-Zhi Li, Sun Yat-sen University & HKUST(GZ)
E-mail: liwzh63@mail2.sysu.edu.cn