Similarity Preserving Adversarial Graph Contrastive Learning

Yeonjun In*, Kanghoon Yoon*, Chanyoung Park

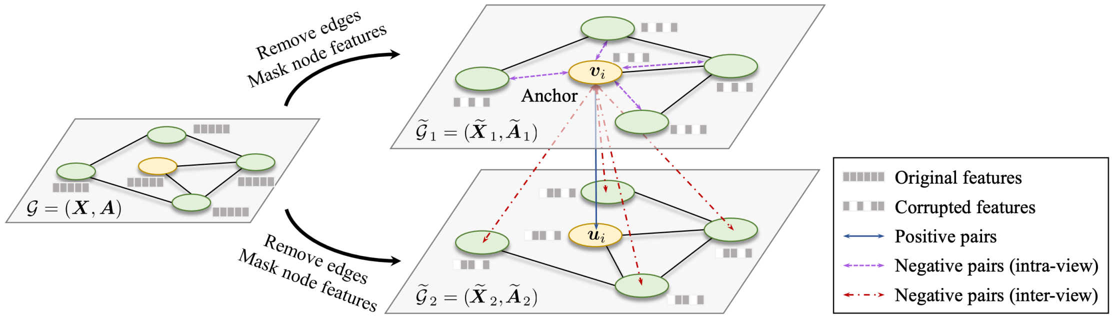
Department of Industrial & Systems Engineering
KAIST
{yeonjun.in, ykhoon08, cy.park}@kaist.ac.kr

Contents

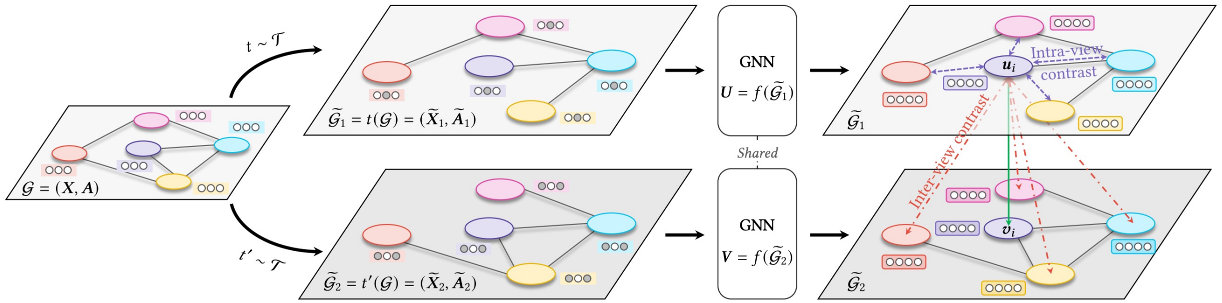
Background

Unsupervised Graph Representation Learning

- Real-world graphs are usually large-scale, and it is difficult to collect labels due to the expensive cost.
- Most recently, the graph contrastive learning (GCL) framework has taken over the mainstream of unsupervised graph representation learning (GRL)
- **Graph contrastive learning (GCL):** pulling together positive samples and pushing apart negative samples.



GRACE, ICML'20

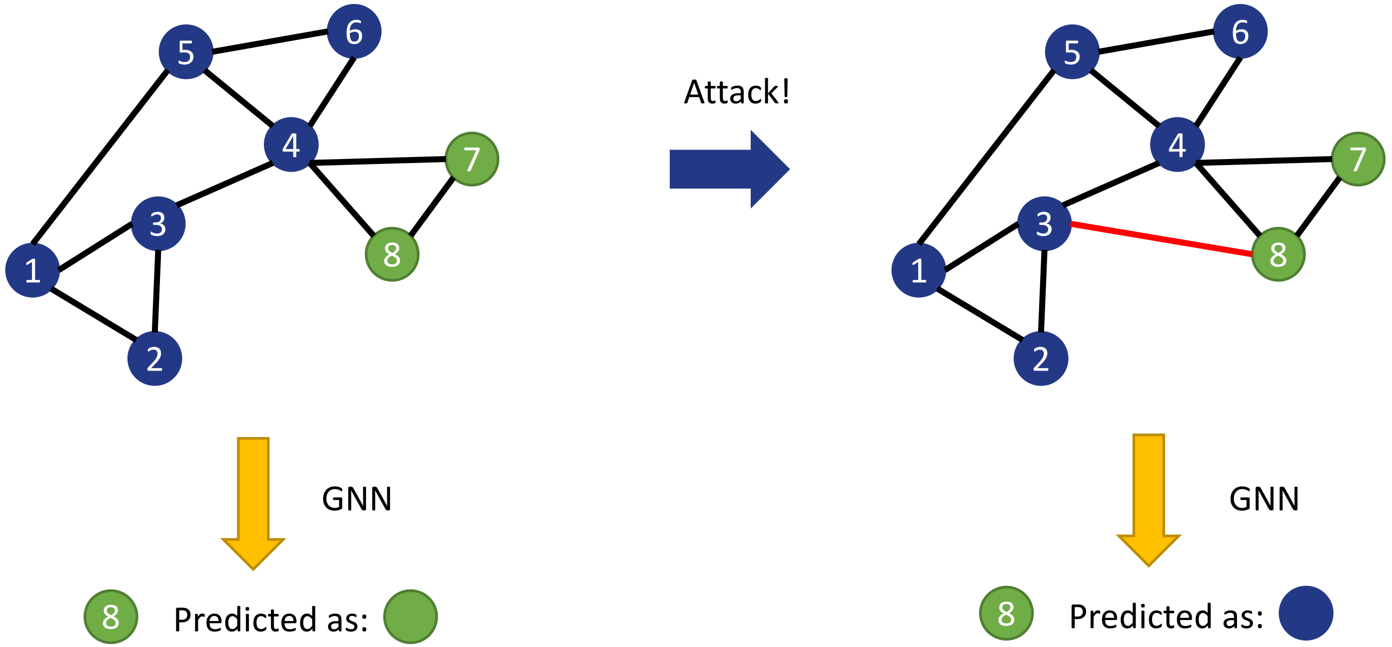


GCA, WWW'21

Figure: Deep Graph Contrastive Representation Learning, ICML'20
Figure: Graph Contrastive Learning with Adaptive Augmentation, WWW'21

Background

Adversarial Attacks on Graph Structures



- Graph Neural Networks are vulnerable to adversarial attacks on graph structures.
- Unsupervised GRL models are also vulnerable to such attacks.

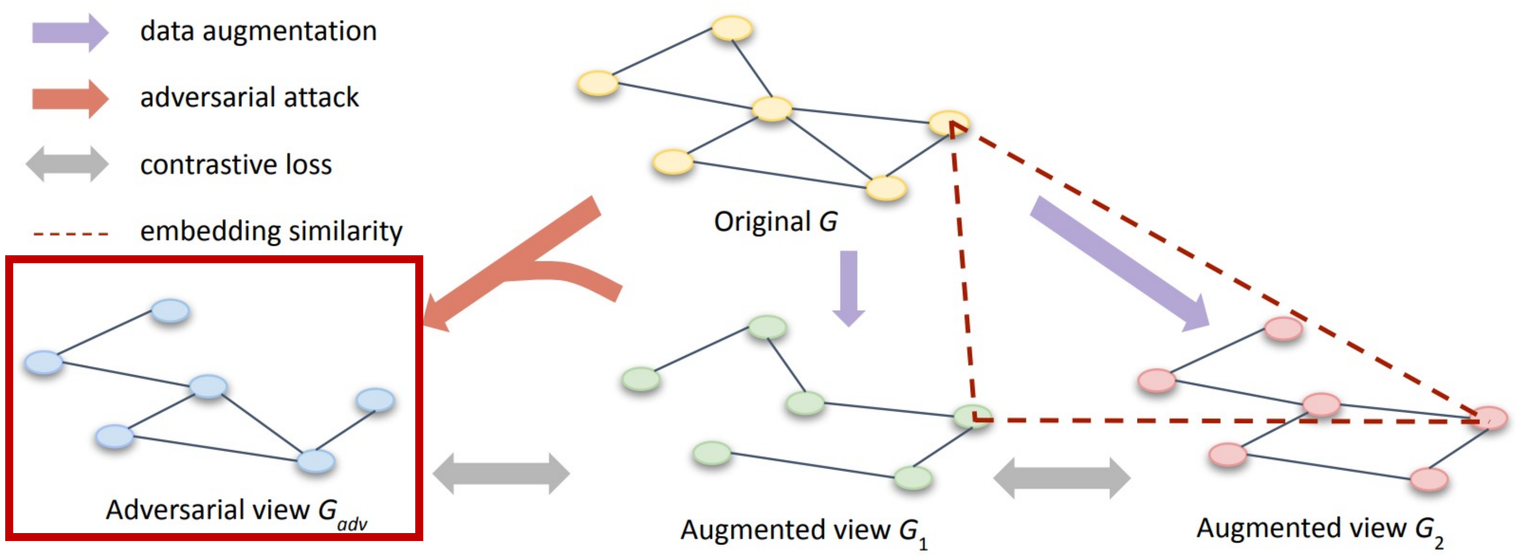
➡ *Leads to the requirements of robust graph representation learning methods*

Figure: Adversarial Attacks and Defenses: Frontiers, Advances and Practice, KDD’20 tutorial

Motivation

Applying Adversarial Training (AT) to Graph Contrastive Learning (GCL)

A graph with the worst-case attack maximizing contrastive loss



Formulation of the adversarial attack in GCL models

$$\delta_{\mathrm{A}}^*, \delta_{\mathrm{X}}^* = \arg \max_{\delta_{\mathrm{A}}, \delta_{\mathrm{X}} \in \Delta} \mathbb{E}\left[\underbrace{\mathcal{L}}_{\text{Contrastive loss}}(f(\mathbf{A}^1 + \delta_{\mathrm{A}}, \mathbf{X}^1 + \delta_{\mathrm{X}}), f(\mathbf{A}^2, \mathbf{X}^2))\right]$$

$$\Delta = \{(\delta_{\mathrm{A}}, \delta_{\mathrm{X}}) \mid \|\delta_{\mathrm{A}}\|_0 \leq \Delta_{\mathrm{A}}, \|\delta_{\mathrm{X}}\|_0 \leq \Delta_{\mathrm{X}}\}$$

Perturbation budgets

- **Goal:** to find the optimal edges and node feature perturbations for the $\mathbf{A}^1, \mathbf{X}^1$ that maximally increase the contrastive loss.
- Since we consider unsupervised adversarial attacks, a contrastive loss is employed instead of a supervised loss.

Figure: Adversarial Graph Contrastive Learning with Information Regularization, WWW 2022

Motivation

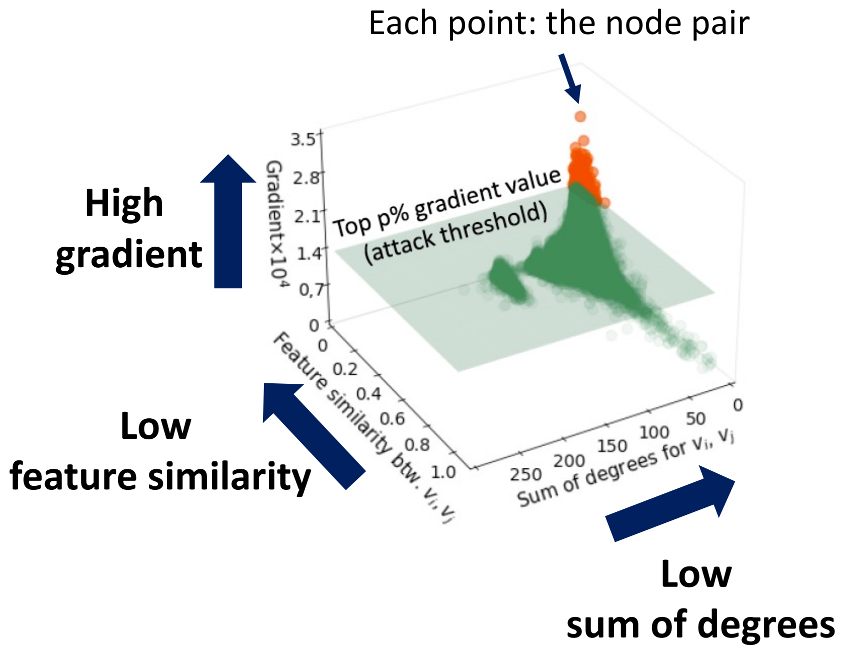
Characteristic of Adversarial Attacks on GCL

Assumption for simplicity
- GCL model with a 1-layer GCN w/o nonlinearity.
- Perturbs only one edge $v_i \rightarrow v_k$.
- Attacked graph $(\mathbf{A}^1 + \boldsymbol{\delta}_A, \mathbf{X}^1)$
- $\mathbf{Z}^{atk} = f\left(\mathbf{A}^1 + \boldsymbol{\delta}_A, \mathbf{X}^1\right)$

- If $\mathbf{z}_i^2 - \mathbf{z}_i^{atk}$ is large, $\boldsymbol{\delta}_A$ is effective perturbation.
- $\mathbf{z}_i^2 - \mathbf{z}_i^{atk}$ is computed as follows:

$$\mathbf{z}_i^2 - \mathbf{z}_i^{\text{atk}} = (\mathbf{z}_i^2 - \mathbf{z}_i^1) + (\mathbf{z}_i^1 - \mathbf{z}_i^{\text{atk}})$$

$$= \mathbf{e}_i + \underbrace{\frac{1}{\sqrt{|\mathcal{N}_{\mathbf{A}^1}^i| + 1}}}_{\text{Degree term}} \underbrace{\left(\sum_{j \in \mathcal{N}_{\mathbf{A}^1}^i \cup \{i\}} \frac{\alpha \mathbf{W} \mathbf{x}_j}{\sqrt{|\mathcal{N}_{\mathbf{A}^1}^i|}\sqrt{|\mathcal{N}_{\mathbf{A}^1}^j|}} - \frac{\mathbf{W}\mathbf{x}_k}{\sqrt{|\mathcal{N}_{\mathbf{A}^1}^k| + 1}} \right)}_{\text{Feature difference term}} \quad (4)$$

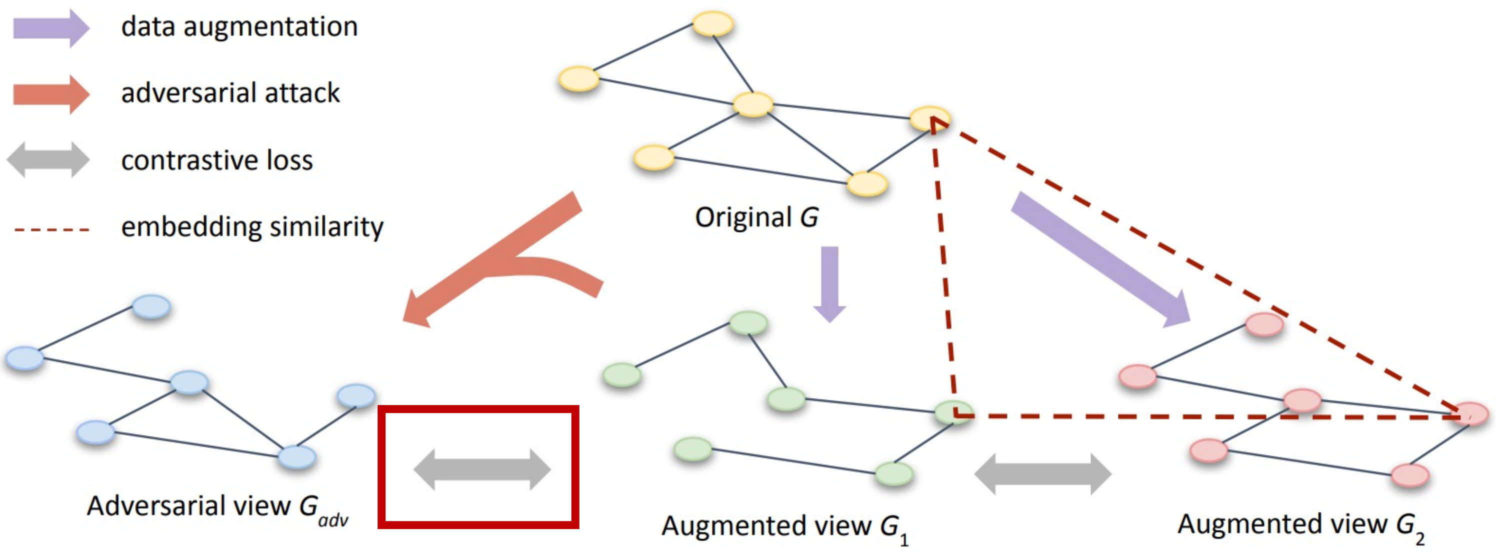


- $\mathbf{z}_i^2 - \mathbf{z}_i^{atk}$ becomes large when degree term ↓ and feature diff. term ↑
    - The degree of v_i is small (*low-degree nodes*)
    - The *features of node v_k (i.e., $\mathbf{x}_k$) is dissimilar* from the aggregation of neighborhood features in a clean graph.

Characteristic of a generated adversarial view by contrastive loss
1. Attack the nodes that have low-degree.
2. **Connect the nodes with dissimilar feature**

Motivation

Applying Adversarial Training (AT) to Graph Contrastive Learning (GCL)



Using the attacked graph as an additional augmentation !

Formulation of Adversarial Graph Contrastive Learning (AGCL)

$$\min_{\Theta} \underbrace{\mathcal{L}(\mathbf{Z}^1, \mathbf{Z}^2)}_{\text{GCL term}} + \lambda_1 \underbrace{\mathcal{L}(\mathbf{Z}^1, \mathbf{Z}^{\mathrm{adv}})}_{\text{AT term}}$$

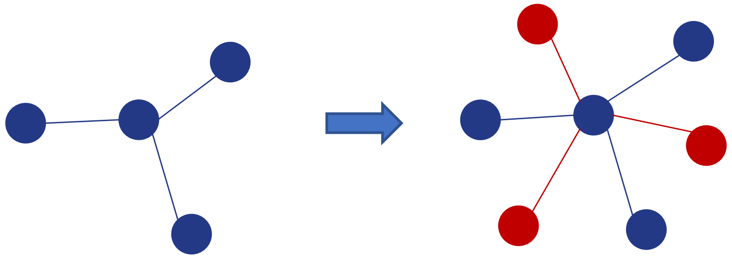
$$\mathbf{Z}^{\mathrm{adv}} = f\underbrace{(\mathbf{A}^1 + \delta^*_{\mathbf{A}}, \mathbf{X}^1 + \delta^*_{\mathbf{X}})}_{\text{Adversarial graph view}}$$

- **Goal:** robust graph representation learning based on adversarial training (AT).
- **Main idea:** to force the representations in the clean graph to be close to those of the attacked graphs.
  - The adversarial graph contrastive learning model minimizes the training objective.

Figure: Adversarial Graph Contrastive Learning with Information Regularization, WWW 2022

Motivation

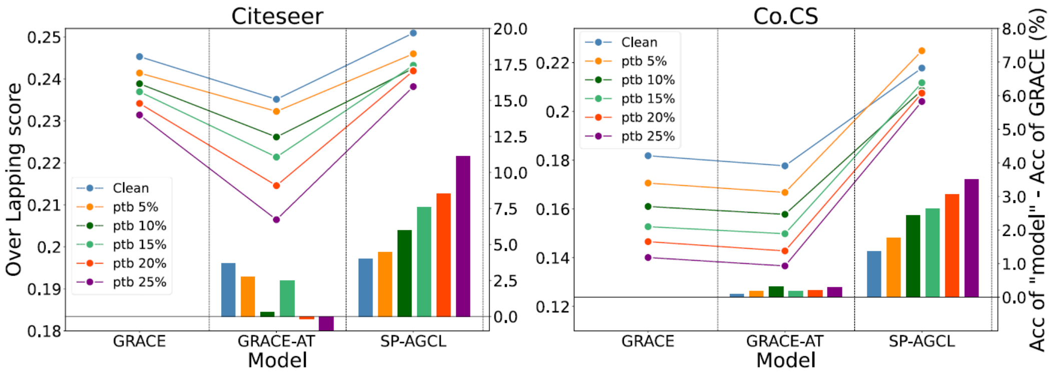
AT fails to preserve node similarity !



- As previously demonstrated, adversarial attacks on graphs tend to connect nodes with dissimilar features.
  - ➢ <u>The neighborhood feature distribution is changed by the adversarial attacks.</u>

- And AGCL reduces the distance between the clean view and the adversarial view to achieve robustness.
  - ➢ <u>Neglecting the changes in the neighborhood feature distributions in the adversarial view.</u>

We argue that existing AGCL models obtain robustness *at the expense of losing the feature information.*



$$OL(\mathbf{A}^{k\mathrm{NN}(\mathbf{Z})}, \mathbf{A}^{k\mathrm{NN}(\mathbf{X})}) = \frac{|\mathbf{A}^{k\mathrm{NN}(\mathbf{Z})} \cap \mathbf{A}^{k\mathrm{NN}(\mathbf{X})}|}{|\mathbf{A}^{k\mathrm{NN}(\mathbf{X})}|}$$

- indicates how much the feature information the representations have

- Solid line: OL score
- Bar plot: performance improvement compared to GRACE

We observe
- GRACE-AT have higher accuracy than GRACE
  - ➢ They obtain robustness.
- GRACE-AT have lower OL score than GRACE
  - ➢ They lose the feature information.

Motivation

Node similarity preservation is crucial !

- As previously demonstrated, existing AGCL models obtain robustness *at the expense of losing the feature information.*
- However, the node feature information is crucial for the robustness against graph structure attacks [1, 2].

We argue that the robustness of AGCL model can be further enhanced
by fully exploiting the node feature information.

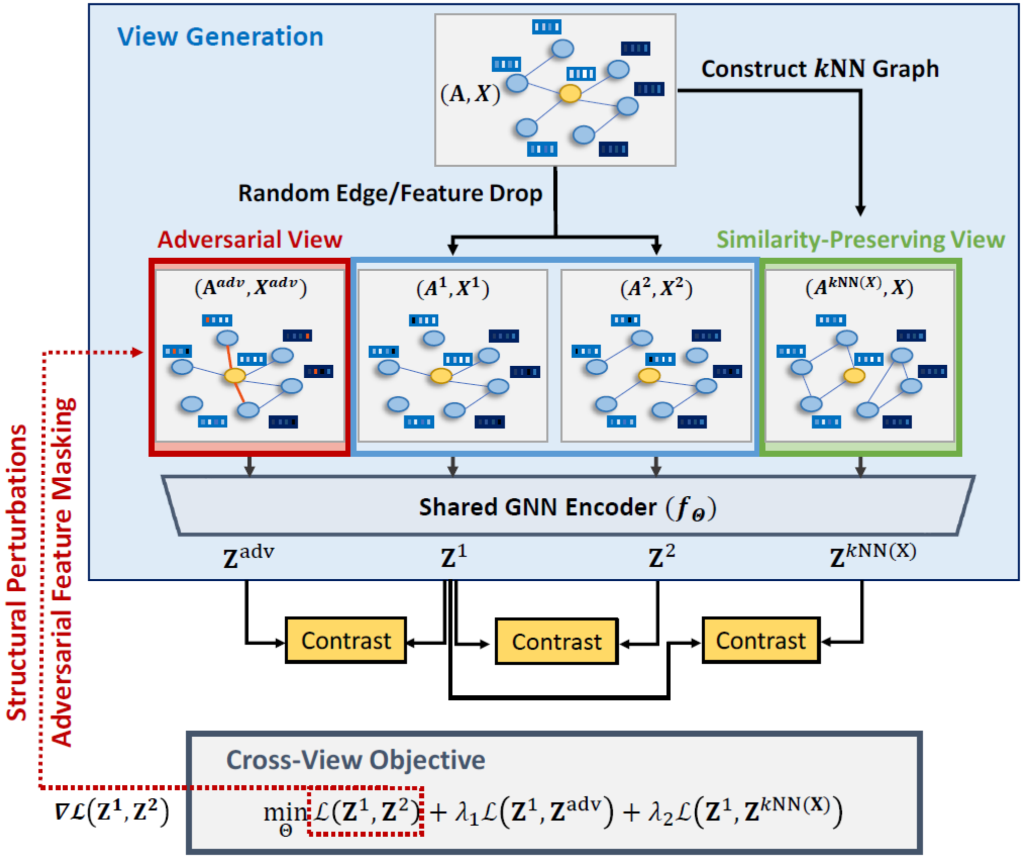
- Moreover, preserving the node feature similarity becomes especially useful for most real-world graphs.
    - Graphs with noisy node labels
    - Graphs with heterophilous neighbors
    - Low-degree nodes

➡ To this end, we propose a *similarity-preserving adversarial graph contrastive learning* (SP-AGCL) framework

[1] Graph Structure Learning for Robust Graph Neural Networks, KDD 2020
[2] Node Similarity Preserving Graph Convolutional Networks, WSDM 2021

Proposed Method

Similarity Preserving Adversarial Graph Contrastive Learning (SP-AGCL)



View generation

- *Step 1.* Two stochastically augmented views, $(\mathbf{A}^1, \mathbf{X}^1)$ and $(\mathbf{A}^2, \mathbf{X}^2)$
  - Same as the previous GCL models

- *Step 2.* Adversarial View
  - Structural perturbations

$$\frac{\partial \mathcal{L}}{\partial \mathbf{A}^1} + \frac{\partial \mathcal{L}}{\partial \mathbf{A}^2} = \mathbf{G}_{\mathbf{A}} \in \mathbb{R}^{N \times N}$$

  - Adversarial feature masking

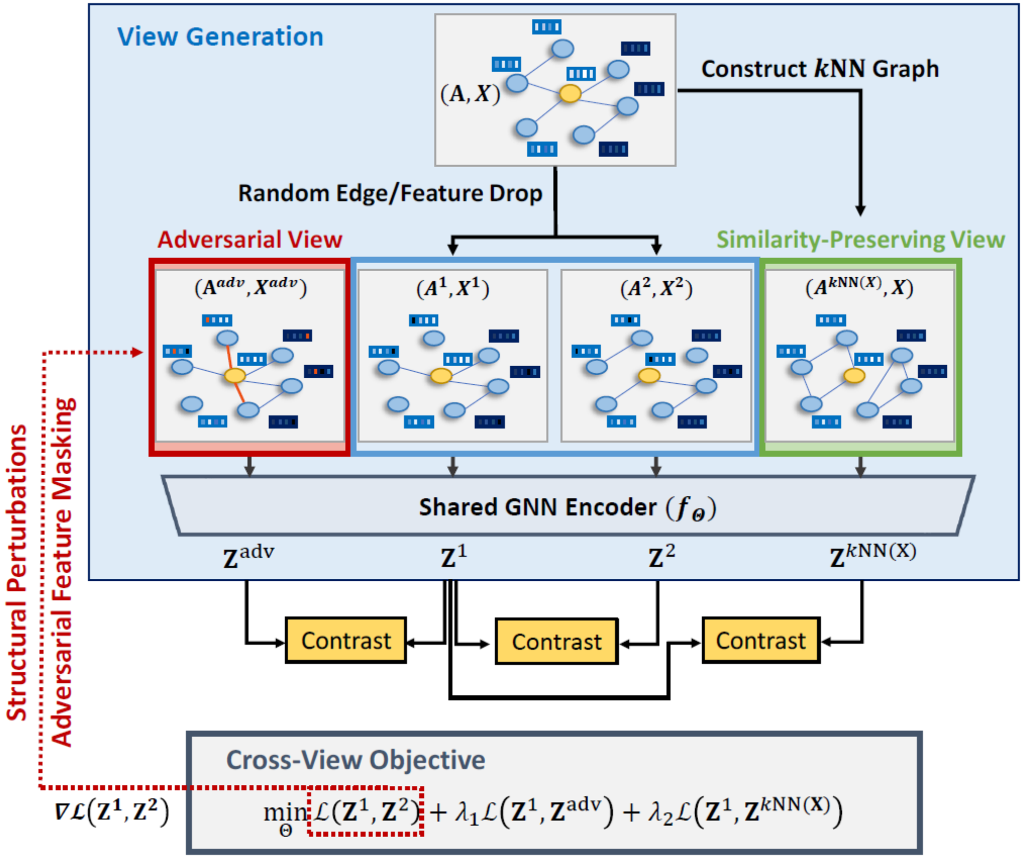
$$\frac{\partial \mathcal{L}}{\partial \mathbf{X}^1} + \frac{\partial \mathcal{L}}{\partial \mathbf{X}^2} = \mathbf{G}_{\mathbf{X}} \in \mathbb{R}^{N \times F}$$

    - Existing works flip the node feature
    - But, it corrupts the co-occurrence/correlation statistics.
    - By masking instead of flipping, we maintaining them.

- *Step 3. Similarity preserving view*
  - Aims to preserve the node feature similarity.
  - kNN graph of node features $\left(\mathbf{A}^{k\mathrm{NN}(\mathbf{X})}, \mathbf{X}\right)$

Proposed Method

Similarity Preserving Adversarial Graph Contrastive Learning (SP-AGCL)



Cross-view Training for Robust GCL

$$\min_\Theta \underbrace{\mathcal{L}(\mathbf{Z}^1,\mathbf{Z}^2)}_{\text{GCL term}} + \lambda_1 \underbrace{\mathcal{L}(\mathbf{Z}^1,\mathbf{Z}^{\text{adv}})}_{\text{AT term}} + \lambda_2 \underbrace{\mathcal{L}(\mathbf{Z}^1,\mathbf{Z}^{k\text{NN}(\mathbf{X})})}_{\text{Similarity-preserving term}}$$

The **representations of nodes with similar features are pulled together,** which in turn preserves the node feature similarity.

Experiment

Experimental settings and datasets

Baselines

- Unsupervised GRL methods
  - GRACE
  - GCA
  - BGRL
- AGCL methods
  - DGI-ADV
  - ARIEL

Various scenarios

- Poisoning attack / evasive attack
- Non-targeted / Targeted attack
- random structure perturbation
- Heterophily graphs
- Noisy node labels

Various downstream tasks

- Node classification
- Link prediction
- Node clustering

Table 7: Statistics for datasets.

Domain	Dataset	# Nodes	# Edges	# Features	# Classes
Citation	Cora	2,485	5,069	1,433	7
	Citeseer	2,110	3,668	3,703	6
	Pubmed	19,717	44,338	500	3
Co-purchase	Am.Photo	7,650	119,081	745	8
	Am.Comp	13,752	245,861	767	10
Co-author	Co.CS	18,333	81,894	6,805	15
	Co.Physics	34,493	247,962	8,415	5
Heterohpily	Chameleon	2,277	36,101	2,325	5
	Squirrel	5,201	217,073	2,089	5
	Actor	7,600	33,544	931	5
	Cornell	183	295	1,703	5
	Texas	183	309	1,703	5
	Wisconsin	251	499	1,703	5

Experiment

Node classification on adversarial attack

nettack

Datasets	Methods / # Ptb	Clean	Poisoning (Acc.) 1	2	3	4	5	Evasive (Acc.) 1	2	3	4	5
Cora	GRACE	82.2±2.2	76.9±1.5	70.2±2.4	65.9±3.0	64.6±1.5	58.9±2.1	77.7±2.7	71.1±2.5	67.1±2.5	65.1±2.3	60.2±3.3
	GCA	81.3±1.7	77.7±2.1	71.6±2.2	67.2±2.3	63.9±2.2	59.2±2.0	79.2±1.2	73.0±0.8	69.2±1.2	67.1±1.5	62.3±3.0
	BGRL	**83.0±2.3**	78.6±1.9	73.0±3.9	69.3±2.9	63.7±5.0	60.5±3.1	79.2±2.8	74.5±2.1	70.7±3.2	66.9±2.8	64.1±3.0
	DGI-ADV	81.7±0.7	78.0±2.3	71.1±2.1	69.9±1.1	65.7±1.8	60.7±1.9	78.1±1.6	73.0±2.1	70.6±1.3	66.5±1.2	63.4±1.5
	ARIEL	76.0±1.7	71.9±2.2	64.9±1.3	63.5±1.6	63.0±1.6	53.7±1.7	71.7±2.0	65.4±1.2	63.5±1.5	64.1±1.3	54.6±1.6
	SP-AGCL	82.5±2.0	**79.5±1.8**	**75.3±2.5**	**73.4±1.7**	**67.4±2.3**	**63.6±2.4**	**80.2±2.4**	**78.7±3.1**	**77.1±3.2**	**73.5±3.4**	**72.8±3.7**
Citeseer	GRACE	82.4±0.5	81.8±1.1	77.6±4.2	68.3±4.4	64.3±3.0	59.1±2.7	82.2±0.6	81.1±1.3	78.1±3.2	72.4±4.8	66.4±3.9
	GCA	**82.5±0.0**	82.4±0.5	78.3±2.9	69.4±5.9	65.9±2.0	58.3±4.0	**82.5±0.0**	81.1±1.5	79.2±2.5	77.0±2.6	71.3±4.4
	BGRL	**82.5±0.7**	81.4±1.2	79.7±4.6	75.1±7.3	72.7±7.6	67.3±8.5	81.6±1.1	80.0±3.3	78.9±4.0	76.7±5.6	73.3±6.5
	DGI-ADV	**82.5±0.0**	81.4±0.7	80.2±1.5	74.3±3.7	68.6±1.2	65.6±1.2	82.4±0.5	81.3±1.0	79.7±0.6	78.7±1.3	76.5±1.6
	ARIEL	**82.5±0.0**	81.1±0.9	80.6±0.6	74.3±3.9	66.2±1.6	63.2±1.0	81.9±0.8	81.3±0.6	81.0±0.0	**80.2±0.8**	**78.6±1.5**
	SP-AGCL	**82.5±0.0**	**82.5±0.0**	**81.6±1.1**	**80.0±3.0**	**75.4±6.1**	**72.7±5.3**	82.4±0.5	**82.1±1.0**	**81.6±1.6**	**80.2±4.2**	78.1±4.9
Pubmed	GRACE	87.9±0.6	86.6±0.5	84.3±0.8	81.7±0.9	77.9±1.5	73.0±1.3	86.3±0.4	84.4±0.7	81.8±1.0	78.3±1.2	74.5±1.3
	GCA	**88.0±0.5**	**87.1±0.6**	**84.7±0.7**	82.2±1.5	77.7±1.2	73.6±1.7	**87,0±0.5**	**85.2±0.7**	82.7±1.1	79.2±0.7	75.5±1.2
	BGRL	87.4±0.8	85.8±0.8	83.2±1.0	79.3±1.0	75.4±1.2	70.2±1.2	85.8±1.1	83.7±1.0	80.3±1.2	76.1±1.0	72.3±1.1
	DGI-ADV	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM
	ARIEL	83.9±0.8	82.0±1.0	79.4±1.5	75.2±1.4	72.0±1.3	67.5±2.6	81.9±0.7	78.8±1.6	76.1±1.2	72.0±1.2	68.2±1.6
	SP-AGCL	87.4±0.8	85.9±0.7	84.3±1.0	**82.7±1.1**	**80.1±1.6**	**77.7±2.2**	86.2±0.5	84.5±0.7	**82.9±0.6**	**81.0±1.1**	**78.1±1.6**

Random structure noise

Datasets	Methods / Ptb rate	Clean	Poisoning (Acc.) 20%	40%	60%	80%	100%	Evasive (Acc.) 20%	40%	60%	80%	100%
Cora	GRACE	82.1±1.0	77.5±1.2	74.2±0.9	70.3±1.2	66.9±1.1	65.1±0.9	78.4±1.9	74.8±1.6	71.5±1.9	68.6±2.9	64.1±2.4
	GCA	81.5±0.9	76.8±1.0	72.0±1.2	67.2±1.4	61.9±1.8	53.4±3.2	77.9±1.1	75.0±1.3	72.7±1.3	70.6±1.5	67.8±2.3
	BGRL	82.7±1.0	77.8±1.2	74.8±1.4	72.6±1.4	69.6±0.8	68.0±1.2	79.0±0.9	76.5±1.3	74.2±1.2	73.0±0.7	70.7±0.8
	DGI-ADV	83.7±0.7	78.8±1.0	76.7±0.7	73.8±0.6	69.9±1.1	68.0±1.4	80.6±1.0	78.2±1.1	75.3±1.8	73.2±1.8	70.7±2.4
	ARIEL	80.9±0.5	75.8±0.8	69.8±0.9	64.8±1.3	60.7±1.5	57.6±1.1	76.1±1.0	70.6±1.1	65.4±1.5	60.2±1.5	53.6±1.7
	SP-AGCL	**83.9±0.7**	**81.3±1.3**	**80.2±0.6**	**78.6±0.4**	**76.2±1.3**	**76.8±0.9**	**81.8±1.3**	**80.1±1.1**	**78.7±1.1**	**77.5±1.5**	**76.1±1.3**
Citeseer	GRACE	74.9±0.6	72.0±0.7	68.8±0.9	66.0±0.6	63.6±0.8	61.3±0.7	72.8±0.9	71.4±0.7	70.1±0.7	68.7±0.8	67.7±1.1
	GCA	74.2±0.7	70.8±0.9	67.0±1.6	63.6±1.5	61.1±1.2	57.5±2.2	72.3±0.5	70.9±0.9	69.6±1.1	68.5±0.8	67.6±0.9
	BGRL	73.4±1.0	70.4±1.2	67.7±1.0	65.0±2.2	63.7±1.4	61.4±1.7	71.5±0.9	69.4±0.9	68.1±0.7	66.6±1.2	65.8±1.0
	DGI-ADV	76.6±0.3	73.1±0.4	70.1±0.9	67.4±1.0	66.0±0.6	64.0±0.5	74.7±0.5	72.8±0.6	71.3±0.8	69.6±0.4	68.2±1.3
	ARIEL	**76.7±0.5**	**74.2±0.6**	**72.8±0.8**	70.2±0.4	69.1±0.4	67.6±0.7	**75.0±0.7**	**73.7±0.6**	72.4±0.8	71.1±0.8	**70.7±0.9**
	SP-AGCL	75.9±0.4	74.1±0.7	72.7±0.6	**70.8±0.8**	**69.5±0.4**	**68.3±0.6**	74.8±0.4	73.5±0.6	**72.7±0.7**	**71.7±0.4**	70.6±0.8
Pubmed	GRACE	85.9±0.1	82.1±0.2	80.1±0.3	78.3±0.7	76.7±0.3	75.7±0.2	81.2±0.2	78.9±0.1	77.3±0.3	76.2±0.3	75.5±0.2
	GCA	**86.5±0.2**	**82.6±0.1**	80.4±0.6	78.6±0.7	77.1±0.6	76.0±0.3	81.2±0.2	78.6±0.2	76.8±0.2	75.6±0.3	74.8±0.2
	BGRL	85.1±0.2	81.3±0.6	79.5±0.8	78.3±1.0	77.2±1.2	76.8±0.7	80.6±0.8	78.7±0.9	77.3±1.0	76.3±1.2	75.6±1.0
	DGI-ADV	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM
	ARIEL	83.4±0.1	79.0±0.4	77.2±0.3	76.4±0.3	75.5±0.2	74.8±0.3	78.4±0.5	76.8±0.3	75.7±0.4	74.7±0.2	74.0±0.5
	SP-AGCL	85.5±0.3	82.3±0.2	**80.7±0.2**	**79.9±0.1**	**78.6±0.2**	**78.0±0.2**	**82.1±0.2**	**80.1±0.2**	**78.7±0.5**	**77.9±0.4**	**77.2±0.5**

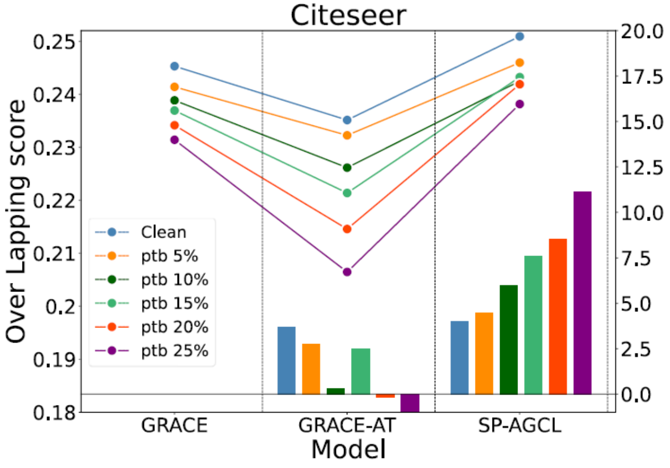
- *Metattack/nettack/random structure noise*
- Poisoning attack / evasive attack

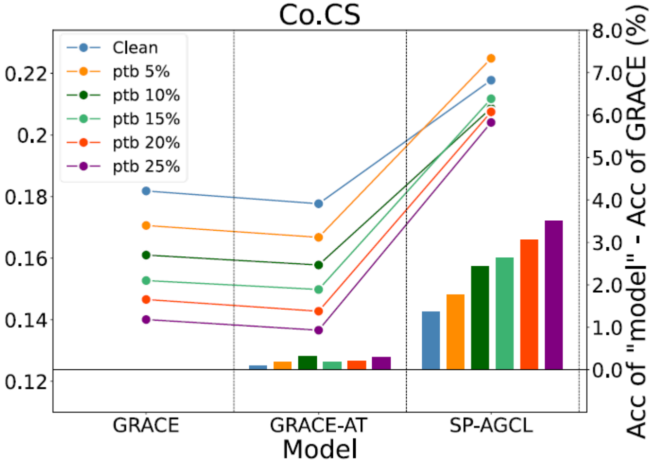
metattack

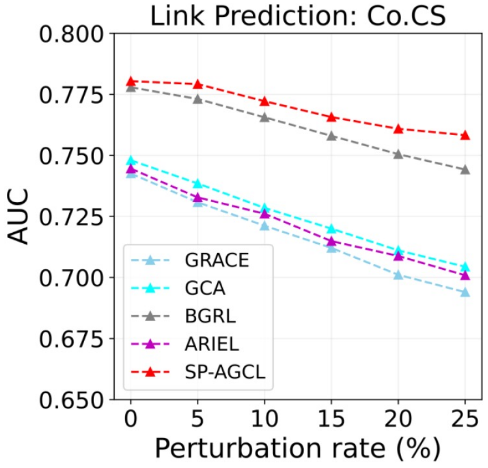
Datasets	Methods / Ptb rate	Clean	Poisoning (Acc.) 5%	10%	15%	20%	25%	Evasive (Acc.) 5%	10%	15%	20%	25%
Cora	GRACE-MLP	63.2±1.7	63.2±1.7	63.2±1.7	63.2±1.7	63.2±1.7	63.2±1.7	63.2±1.7	63.2±1.7	63.2±1.7	63.2±1.7	63.2±1.7
	GRACE	82.1±1.0	78.4±1.5	75.5±1.1	66.1±1.6	55.2±1.8	51.3±2.0	78.9±0.9	75.7±0.9	67.6±1.3	56.5±2.3	51.5±1.8
	GCA	81.5±0.9	79.8±0.8	75.8±0.6	68.4±1.6	53.4±1.7	49.5±1.3	79.7±1.0	76.0±1.1	68.0±1.1	54.7±1.2	49.8±1.3
	BGRL	82.7±1.0	78.2±2.1	74.3±1.8	66.2±1.9	53.8±1.7	50.2±2.3	79.2±1.6	75.2±1.5	67.2±2.0	55.2±1.7	51.2±1.7
	DGI-ADV	83.7±0.7	79.4±0.9	73.3±0.6	63.5±0.6	52.2±0.7	48.1±0.7	79.4±0.9	73.7±0.8	62.9±0.9	53.0±1.0	49.2±1.2
	ARIEL	80.9±0.5	79.2±0.4	77.7±0.6	69.8±0.7	57.7±0.7	52.8±1.0	79.1±0.3	77.8±0.6	70.3±0.9	58.0±1.0	53.2±1.2
	SP-AGCL	**83.9±0.7**	**82.2±0.8**	**79.0±0.6**	**73.25±0.5**	**66.2±2.3**	**65.0±1.5**	**82.0±0.6**	**78.7±1.2**	**73.5±2.8**	**61.5±5.0**	**57.1±5.5**
Citeseer	GRACE-MLP	68.0±1.2	68.0±1.2	68.0±1.2	68.0±1.2	68.0±1.2	68.0±1.2	68.0±1.2	68.0±1.2	68.0±1.2	68.0±1.2	68.0±1.2
	GRACE	74.9±0.6	74.1±0.6	72.5±0.9	71.2±1.3	59.2±1.4	61.2±1.5	74.0±0.7	72.4±1.0	70.4±1.3	59.1±1.9	62.3±1.5
	GCA	74.2±0.7	73.5±0.9	73.0±0.6	71.5±0.9	60.2±1.7	60.1±1.6	73.8±0.7	73.4±0.5	72.0±0.9	59.5±1.8	61.5±1.7
	BGRL	73.4±1.0	72.1±1.1	69.1±1.0	67.5±1.4	57.7±1.3	58.2±2.8	72.5±1.2	69.7±1.3	68.1±1.6	58.5±1.6	60.3±2.0
	DGI-ADV	76.6±0.3	74.8±0.3	71.0±0.5	70.1±0.3	57.9±0.8	60.6±1.2	74.8±0.3	71.3±0.5	69.7±0.5	56.1±0.6	57.4±1.5
	ARIEL	**76.7±0.5**	75.2±0.4	72.8±0.5	70.2±0.5	60.1±1.1	62.7±0.5	**75.3±0.4**	73.3±0.5	70.8±0.4	59.8±0.8	63.6±1.0
	SP-AGCL	75.9±0.4	**75.3±0.5**	**73.5±0.6**	**72.1±1.1**	**66.0±1.5**	**69.6±0.9**	75.0±1.1	**73.5±1.0**	**72.4±1.1**	**60.6±1.1**	**65.6±0.9**
Pubmed	GRACE-MLP	82.4±0.2	82.4±0.2	82.4±0.2	82.4±0.2	82.4±0.2	82.4±0.2	82.4±0.2	82.4±0.2	82.4±0.2	82.4±0.2	82.4±0.2
	GRACE	85.9±0.1	81.3±0.2	78.2±0.4	76.1±1.3	73.9±1.7	71.3±2.6	80.7±0.1	76.8±0.2	73.5±0.1	71.4±0.2	69.0±0.3
	GCA	**86.5±0.2**	81.2±0.5	78.1±0.5	75.9±1.2	74.2±0.4	72.0±1.8	80.7±0.2	76.7±0.3	73.2±0.3	70.9±0.2	68.6±0.3
	BGRL	85.1±0.2	81.3±0.3	79.0±0.4	76.6±0.9	74.8±0.9	73.0±0.5	80.6±0.4	77.5±0.4	74.5±0.6	72.4±0.7	70.3±0.6
	DGI-ADV	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM
	ARIEL	81.2±0.4	77.8±0.3	75.8±0.4	74.0±0.5	72.3±0.5	70.7±0.3	77.8±0.5	75.9±0.5	74.1±0.6	72.3±0.6	70.8±0.5
	SP-AGCL	85.5±0.3	**81.9±0.2**	**79.7±0.3**	**77.4±0.3**	**75.6±0.3**	**73.3±0.3**	**81.9±0.2**	**79.6±0.3**	**77.1±0.4**	**75.1±0.5**	**72.8±0.5**
Am.Photo	GRACE-MLP	87.2±0.8	87.2±0.8	87.2±0.8	87.2±0.8	87.2±0.8	87.2±0.8	87.2±0.8	87.2±0.8	87.2±0.8	87.2±0.8	87.2±0.8
	GRACE	92.0±0.4	89.5±0.5	88.3±1.1	87.6±0.9	87.5±1.2	87.1±1.2	88.6±0.4	87.5±0.8	87.3±0.8	86.6±1.0	85.6±1.1
	GCA	92.2±0.4	89.4±0.6	88.3±0.8	87.8±0.7	87.6±1.0	87.5±0.7	88.7±0.6	88.0±0.7	87.8±1.4	87.3±0.9	86.4±1.2
	BGRL	92.1±0.4	89.2±0.6	88.7±0.5	88.8±0.5	89.0±0.7	89.2±0.4	89.4±0.5	88.3±0.6	88.2±0.6	87.6±0.5	87.3±0.6
	DGI-ADV	91.6±0.5	83.5±0.5	80.7±0.6	79.3±0.6	78.1±0.6	77.3±0.6	83.6±0.5	80.8±0.5	79.5±0.5	78.0±0.6	77.4±0.4
	ARIEL	92.5±0.2	90.1±0.4	89.9±0.5	89.9±0.5	89.9±0.5	**89.8±0.6**	89.7±0.4	89.1±0.3	88.6±0.2	**88.6±0.4**	**88.4±0.3**
	SP-AGCL	**93.3±0.3**	**91.4±0.5**	**90.6±0.6**	**90.5±0.8**	**90.2±0.9**	**89.8±1.0**	**90.3±0.4**	**89.3±0.3**	**88.7±0.5**	88.2±0.7	87.6±0.5
Am.Comp	GRACE-MLP	82.7±0.4	82.7±0.4	82.7±0.4	82.7±0.4	82.7±0.4	82.7±0.4	82.7±0.4	82.7±0.4	82.7±0.4	82.7±0.4	82.7±0.4
	GRACE	86.4±0.5	83.7±0.3	82.5±0.6	81.3±0.5	80.3±0.7	78.8±1.0	84.0±0.3	83.6±0.4	82.8±0.3	82.0±0.9	81.7±0.6
	GCA	86.6±0.4	84.6±0.4	83.4±0.3	82.3±0.4	81.4±0.4	80.1±0.5	84.5±0.3	83.8±0.4	82.8±0.2	82.3±0.8	82.0±0.4
	BGRL	88.0±0.4	85.2±0.6	84.2±0.6	83.7±0.6	83.3±0.7	83.4±0.6	85.7±0.7	85.0±0.6	84.1±0.6	83.8±0.7	83.4±0.5
	DGI-ADV	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM
	ARIEL	87.4±0.4	85.4±0.5	84.5±0.4	83.8±0.4	83.7±0.5	83.6±0.5	85.7±0.3	84.6±0.4	83.9±0.5	83.8±0.4	83.6±0.5
	SP-AGCL	**89.1±0.4**	**86.9±0.3**	**85.6±0.5**	**85.1±0.4**	**85.0±0.5**	**84.8±0.7**	**87.2±0.3**	**85.9±0.4**	**85.1±0.5**	**84.4±0.4**	**84.1±0.6**
Co.CS	GRACE-MLP	92.1±0.2	92.1±0.2	92.1±0.2	92.1±0.2	92.1±0.2	92.1±0.2	92.1±0.2	92.1±0.2	92.1±0.2	92.1±0.2	92.1±0.2
	GRACE	92.3±0.2	91.2±0.1	90.6±0.2	90.0±0.2	89.4±0.1	88.9±0.2	91.2±0.3	90.4±0.3	89.7±0.4	89.3±0.4	88.7±0.4
	GCA	92.5±0.1	91.4±0.2	90.7±0.2	90.2±0.2	89.7±0.2	89.3±0.1	91.4±0.2	90.8±0.3	90.0±0.3	89.6±0.3	89.0±0.3
	BGRL	92.4±0.2	91.3±0.1	90.5±0.2	89.9±0.2	89.3±0.2	88.7±0.2	91.3±0.2	90.5±0.2	89.8±0.3	89.4±0.3	88.8±0.2
	DGI-ADV	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM
	ARIEL	92.3±0.2	91.0±0.1	90.2±0.2	89.6±0.3	88.8±0.2	88.1±0.2	90.8±0.1	90.1±0.3	89.2±0.2	88.7±0.2	87.9±0.1
	SP-AGCL	**93.7±0.2**	**92.9±0.2**	**92.8±0.2**	**92.5±0.2**	**92.4±0.1**	**92.3±0.2**	**92.7±0.1**	**91.9±0.2**	**91.2±0.2**	**90.6±0.2**	**89.9±0.2**
Co.Physics	GRACE-MLP	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM
	GRACE	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM
	GCA	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM
	BGRL	95.2±0.1	94.1±0.2	93.2±0.2	92.5±0.1	91.6±0.2	91.0±0.1	94.2±0.1	93.2±0.2	92.5±0.1	91.6±0.1	91.0±0.2
	DGI-ADV	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM
	ARIEL	95.1±0.1	93.2±0.2	92.4±0.2	91.6±0.2	90.7±0.3	90.2±0.2	93.9±0.1	93.3±0.1	92.6±0.2	91.8±0.2	91.4±0.2
	SP-AGCL	**95.8±0.1**	**94.9±0.2**	**94.4±0.1**	**93.6±0.1**	**93.0±0.1**	**92.5±0.1**	**95.0±0.1**	**94.2±0.1**	**93.3±0.2**	**92.4±0.1**	**91.7±0.1**

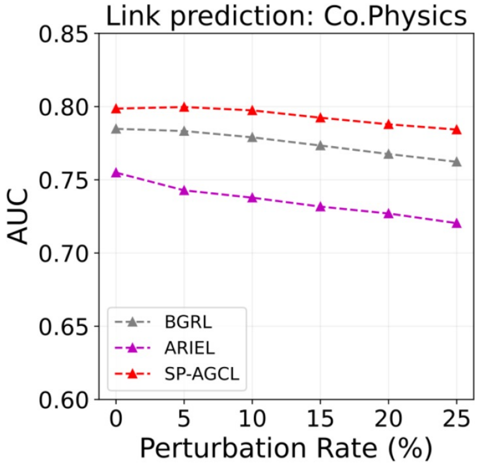
Experiment

Preserving Feature Similarity is beneficial !







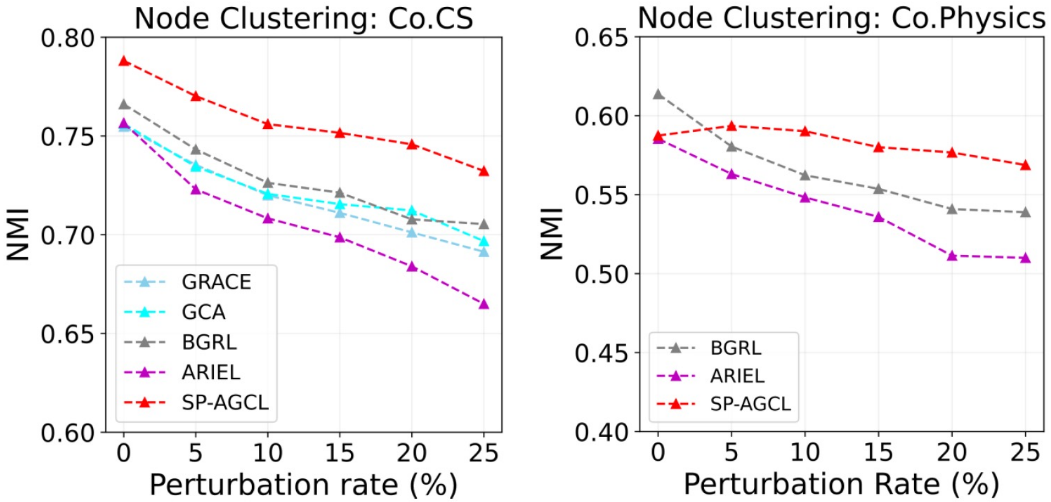


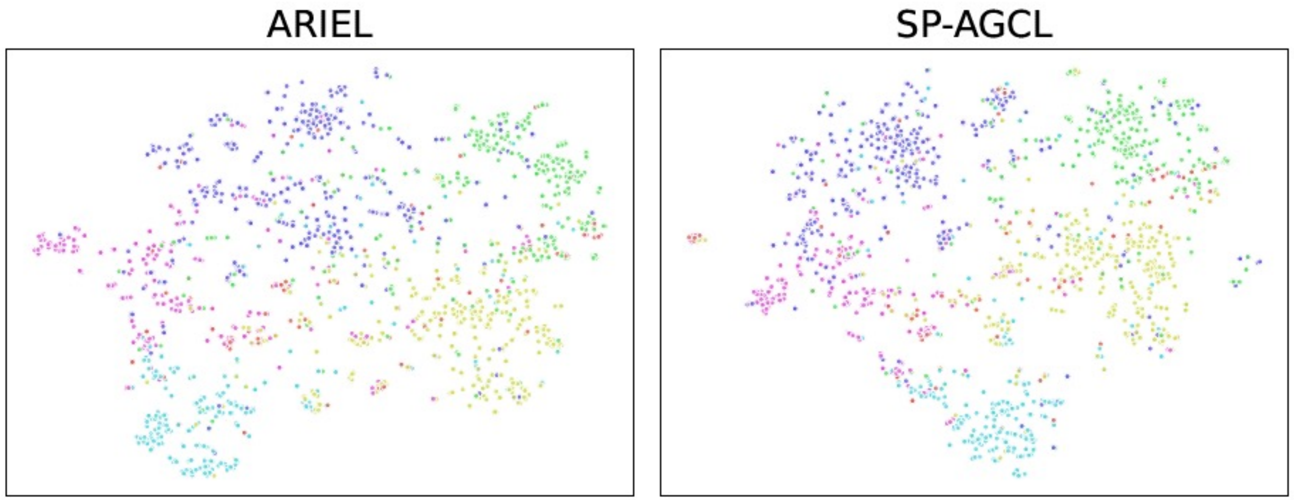
- SP-AGCL preserves the node feature similarity, which results in the robust graph representation.
- SP-AGCL consistently predicts reliable links compared with other baselines across all the perturbation ratios.
    - Moreover, ARIEL, the sota AGCL model, shows the worst performance

Node feature information is beneficial to predicting reliable links
since nodes with similar features tend to be adjacent in many real-world graphs.

Experiment

Preserving Feature Similarity is beneficial !





- SP-AGCL outperforms baselines, especially ARIEL, on the node clustering tasks.
- The representations of ARIEL are separable but widely distributed → vague class boundaries.
- The representations of SP-AGCL are tightly grouped together → more separable class boundaries.
  - ***Reason of the superior performance of node clustering tasks.***
- *Why?*
  - Node feature information is highly related to class information.
  - The AT of ARIEL loses the node feature information, which is preserved in SP-AGCL.

Experiment

The node feature similarity is useful for the real-world graphs

Table 2: Node classification on heterophilous graphs.

	Chameleon	Squirrel	Actor	Texas	Wisconsin	Cornell
GRACE	46.6±2.8	35.2±1.0	29.5±0.5	61.1±6.5	55.3±5.5	61.1±5.0
GCA	50.0±3.0	37.1±1.8	29.3±0.8	60.0±6.3	55.7±8.0	59.5±3.8
BGRL	57.1±3.6	40.6±1.6	31.0±1.2	61.6±6.0	57.7±5.2	57.8±4.7
DGI-ADV	53.4±2.2	40.1±1.6	26.5±0.9	58.4±6.1	57.3±4.9	60.5±5.8
ARIEL	44.3±2.4	36.8±1.2	29.6±0.3	58.4±4.7	53.3±7.2	57.8±4.4
SP-AGCL	**57.5±2.5**	**41.1±1.9**	**32.3±1.3**	**64.9±6.8**	**58.4±5.5**	**64.3±3.6**

- In heterophilous networks, nodes with dissimilar properties (e.g., node features and labels) are connected.
  - ➢ Similar to the properties of the adversarial attacks on graph structures.
- SP-AGCL outperforms the other baselines on heterophilious graphs.
- ARIEL perform worse than GRACE → AT fails to preserve the node feature similarity.

The feature similarity should be preserved
when the given structural information is not informative

Experiment

The node feature similarity is useful for the real-world graphs

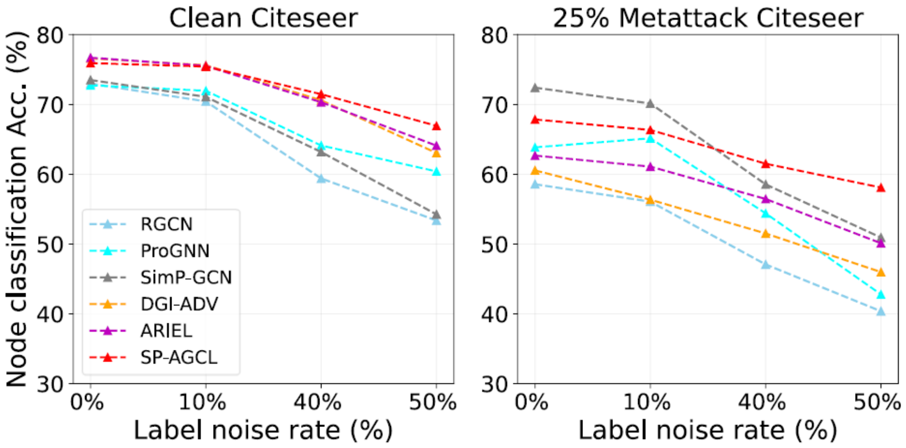


Figure 4: Node classification with noisy label.

- We compare SP-AGCL with both supervised (i.e., RGCN, ProGNN, and SimP-GCN) and AGCL methods.
- We observe that SP-AGCL outperforms both supervised and unsupervised methods.
  - ➢ Supervised methods rely on the noisy supervision information.
  - ➢ Better exploiting feature information results in more robust node representations.

Conclusion

- In this paper, we discover that adversarial GCL models obtain robustness against adversarial attacks at the expense of not being able to preserve the node feature similarity information.

- Based on our findings, we propose SP-AGCL that learns robust node representations that preserve the node feature similarity by introducing the similarity preserving view.

- We verify the effectiveness of SP-AGCL by conducting extensive experiments on thirteen benchmark datasets with multiple attacking scenarios along with several real-world scenarios such as networks with noisy labels and heterophily.

Thank you for listening

Similarity Preserving Adversarial Graph Contrastive Learning

Yeonjun In*, Kanghoon Yoon*, Chanyoung Park

Department of Industrial & Systems Engineering
KAIST
{yeonjun.in, ykhoon08, cy.park}@kaist.ac.kr