NeurIPS | 2022
Thirty-sixth Conference on Neural
Information Processing Systems

Revisiting Graph Contrastive Learning from the Perspective of Graph Spectrum

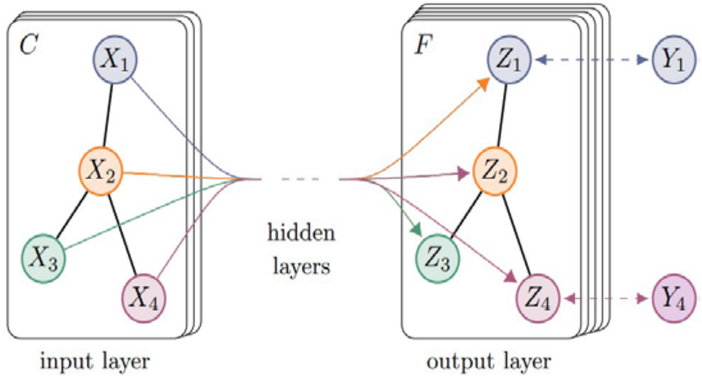
Nian Liu[1], Xiao Wang[1,2], Deyu Bo [1],
Chuan Shi[1,2], Jian Pei [3]

[1]Beijing University of Posts and Telecommunications
[2]Peng Cheng Laboratory
[3]Duke University

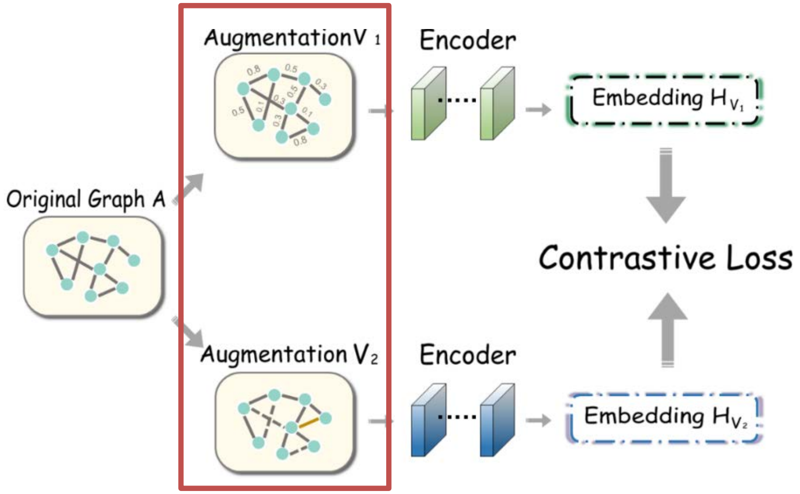
CONTENTS

Overview

Graph Neural Network[1]



Graph Contrastive Learning



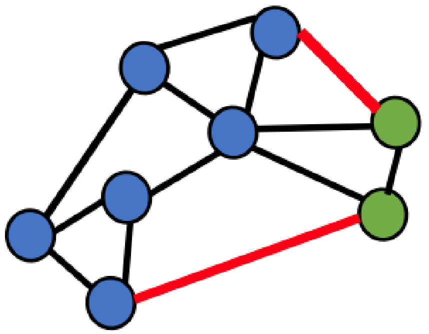
Different graph augmentation strategies

- **Heuristic based**
  - e.g. Node or edge dropping, Feature masking, Diffusion matrix ……
- **Learning based**
  - e.g. InfoMin principle, Disentanglement, Adversarial training ……

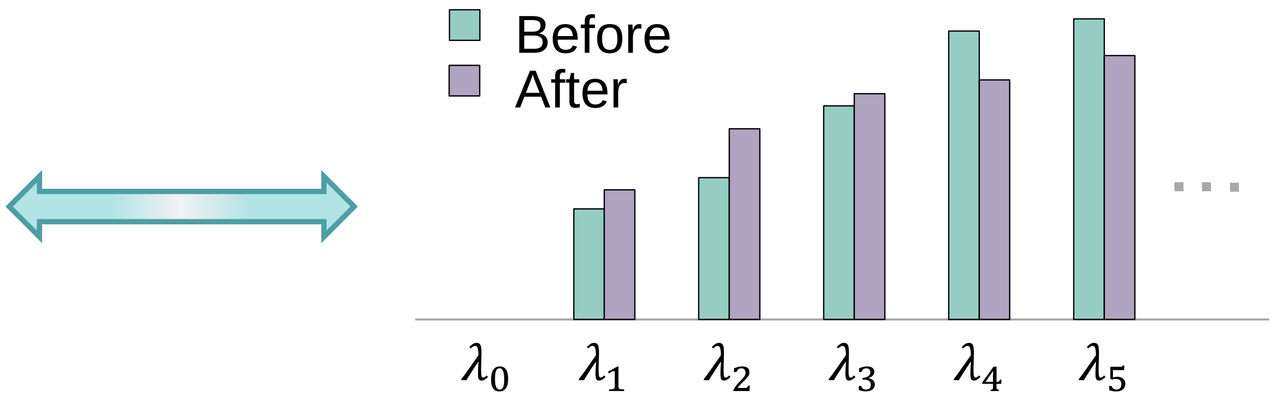
1 T. N. Kipf, and M. Welling. Semi-supervised Classification with Graph Convolutional Networks. ICLR 2017.

Rethinking Graph Augmentation

- *What **information** should we **preserve or discard** in an augmented graph?*
- *Are there some **general rules** across different graph augmentation strategies?*
- *How to use those general rules to **validate and improve** the current GCL methods?*



Topological Augmentation



Graph spectrum changes

Explore the effectiveness of augmentations from the graph spectral domain.

The GAME rule

Preliminaries

➢ **Symmetric normalized graph Laplacian**

$$\hat{\mathcal{L}} = \boldsymbol{I_n} - \hat{\boldsymbol{A}} = \boldsymbol{D}^{-\frac{1}{2}}(\boldsymbol{D} - \boldsymbol{A})\boldsymbol{D}^{-\frac{1}{2}} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\top}$$

where $\boldsymbol{\Lambda} = diag(\lambda_1, \ldots, \lambda_N)$ $\boldsymbol{U} = [\boldsymbol{u}_1^{\top}, \ldots, \boldsymbol{u}_N^{\top}] \in \mathbb{R}^{N \times N}$

➢ **Low-frequency & High-frequency components**

Reorder $0 \leq \lambda_1 \leq \cdots \leq \lambda_N < 2$

Low-frequency components $\mathcal{F}_{\mathcal{L}} = \{\lambda_1, \ldots, \lambda_{\lfloor N/2 \rfloor}\}$

High-frequency components $\mathcal{F}_{\mathcal{H}} = \{\lambda_{\lfloor N/2 \rfloor + 1}, \ldots, \lambda_N\}$

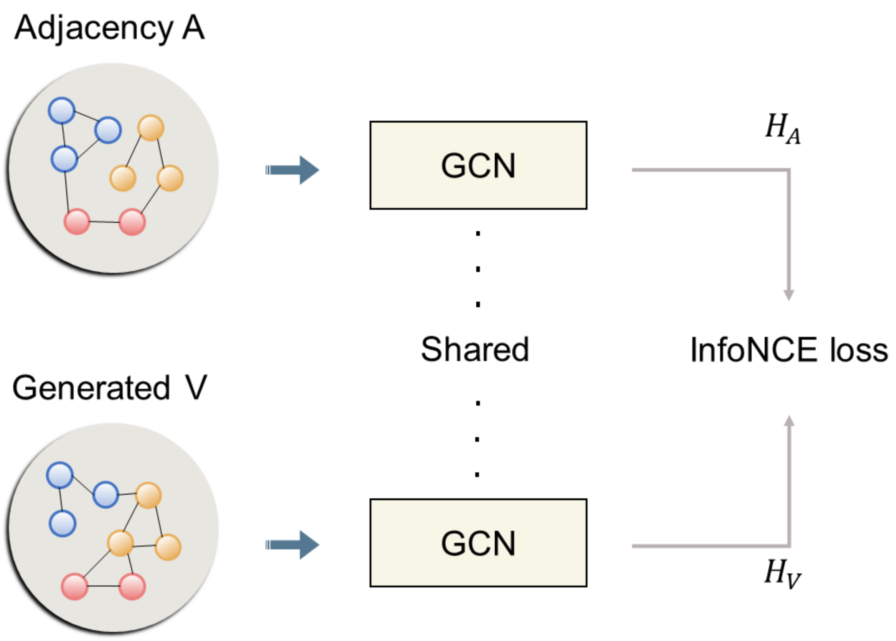
➢ **Graph spectrum**

$\boldsymbol{\phi}(\boldsymbol{\lambda})$: Amplitudes of different frequency components, indicating which parts of frequency are enhanced or weakened.

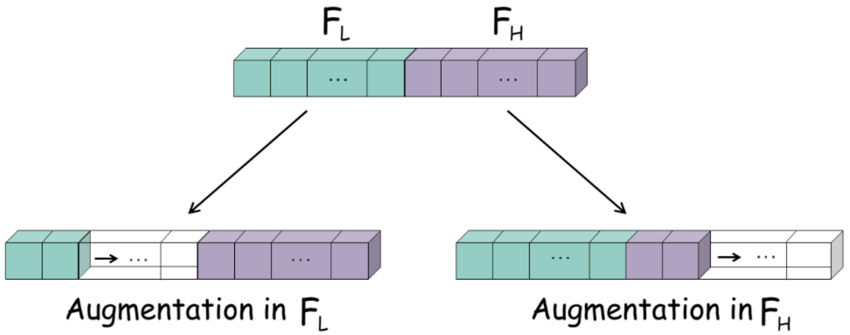
An Experimental Investigation

- **Aim:** In GCL, investigate which part of frequencies should be contrasted.

- **Case study model**



- **Generating *V***



Augmenting 20% in $\mathcal{F}_{\mathcal{L}}$

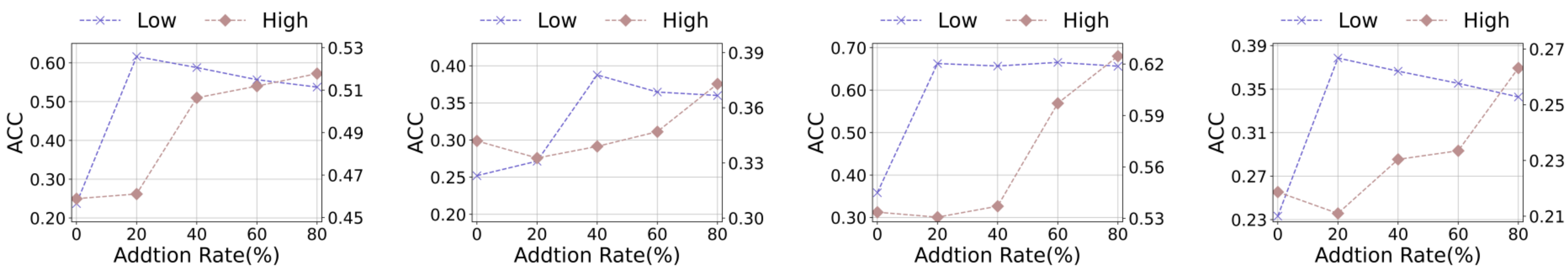
$$\boldsymbol{u}_1\boldsymbol{u}_1^\top + \cdots + \boldsymbol{u}_{0.2*N/2}\boldsymbol{u}_{0.2*N/2}^\top + \boldsymbol{u}_{N/2}\boldsymbol{u}_{N/2}^\top + \cdots + \boldsymbol{u}_N\boldsymbol{u}_N^\top$$

Augmenting 20% in $\mathcal{F}_{\mathcal{H}}$

$$\boldsymbol{u}_1\boldsymbol{u}_1^\top + \cdots + \boldsymbol{u}_{N/2}\boldsymbol{u}_{N/2}^\top + \boldsymbol{u}_{(N+1)/2}\boldsymbol{u}_{(N+1)/2}^\top + \cdots + \boldsymbol{u}_{0.7N}\boldsymbol{u}_{0.7N}^\top$$

Set $\boldsymbol{\lambda}_i = \mathbf{1}$, only consider the effect of eigenspace $\boldsymbol{u}_i\boldsymbol{u}_i^\top$

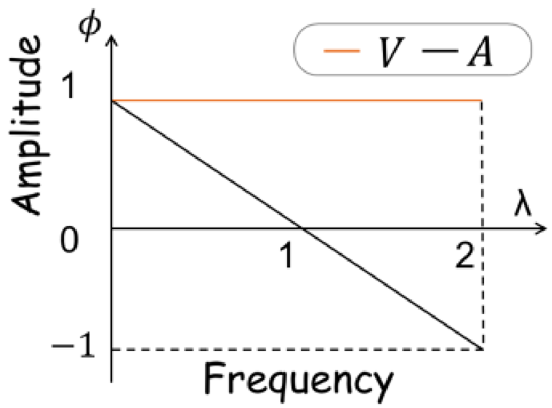
Observation & Results



(a) Cora (b) Citeseer (c) BlogCatalog (d) Flickr

- **Observation 1 --- Purple dash line**
  - The performance achieves the best only when **the lowest part** of $\mathcal{F}_{\mathcal{L}}$ is maintained
- **Observation 2 --- Brown dash line**
  - With **more high-frequency** information in $\mathcal{F}_{\mathcal{H}}$ added, the performance generally rises

Observation & Results



Maintain **the lowest frequency** in V

=> Difference in $\mathcal{F}_{\mathcal{L}}$ becomes **smaller**

Add **more high frequency** in V

=> Difference in $\mathcal{F}_{\mathcal{H}}$ becomes **larger**

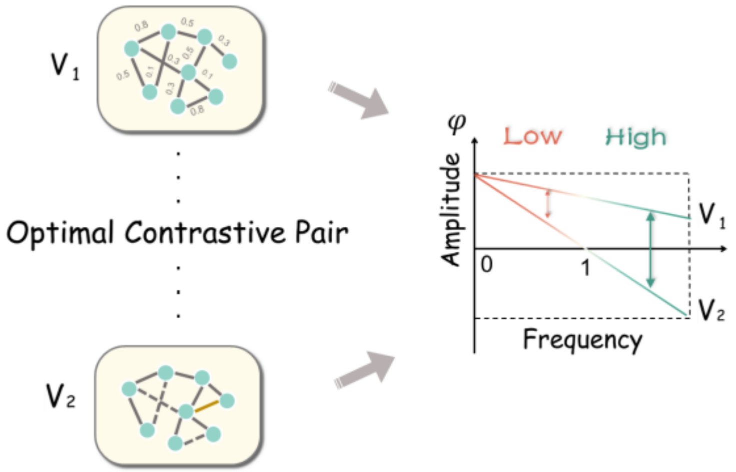
The GAME rule

The General Graph Augmentation Rule

Given two random augmentations $\boldsymbol{V_1}$ and $\boldsymbol{V_2}$, their graph spectrums are $\phi_{V_1}(\lambda)$ and $\phi_{V_2}(\lambda)$. Then, $\forall\ \lambda_m \in [1,2]$ and $\lambda_n \in [0,1]$, $\boldsymbol{V_1}$ and $\boldsymbol{V_2}$ are an effective pair of graph augmentations if the following condition is satisfied:

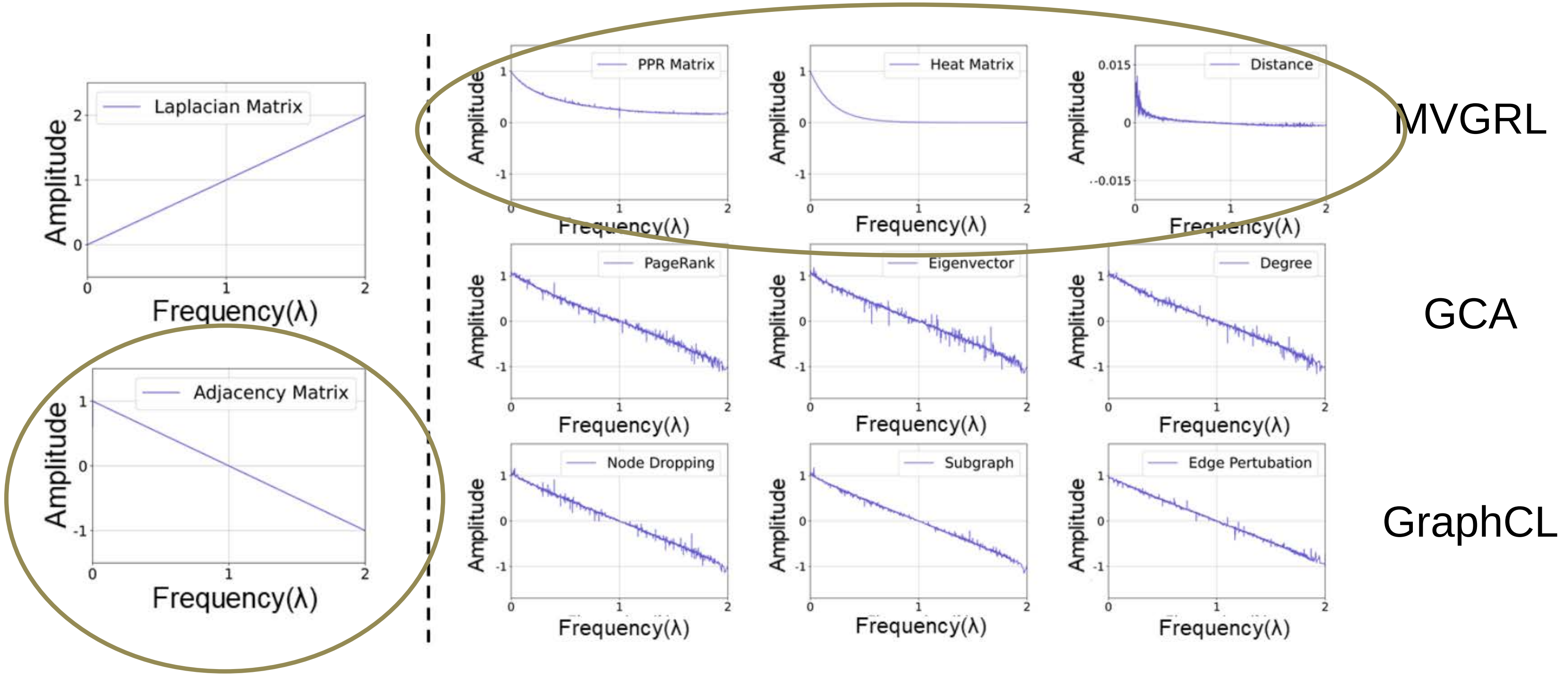
$$|\phi_{V_1}(\lambda_m) - \phi_{V_2}(\lambda_m)| > |\phi_{V_1}(\lambda_n) - \phi_{V_2}(\lambda_n)|.$$

We define such pair of augmentations as optimal contrastive pair.



Analysis of The General Graph Augmentation Rule

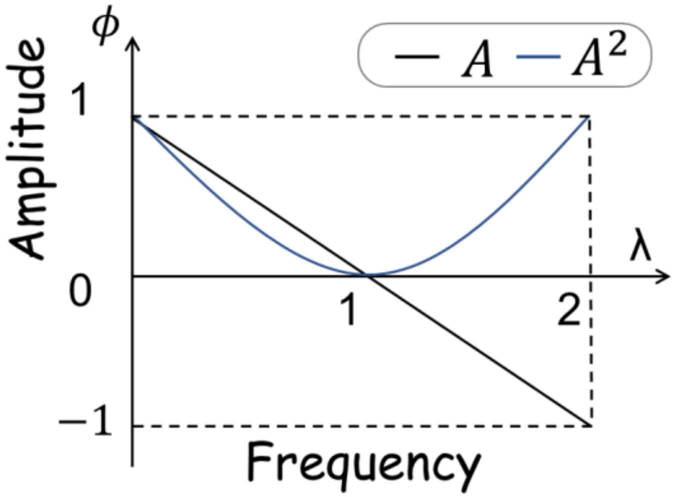
- **Experimental analysis 1 ---** Contrast between ***A*** and 9 existing augmentations



Methods	GraphCL			GCA			MVGRL		
Type	Subgraph	Node dropping	Edge perturbation	Degree	PageRank	Eigenvector	PPR	Heat	Distance
Results	34.9±3.5	29.8±2.3	37.7±4.4	40.2±4.1	38.5±5.0	42.1±4.9	**58.0**±1.6	**49.9**±4.2	**46.1**±7.5

Analysis of The General Graph Augmentation Rule

- **Experimental analysis 2 ---** Contrast between $\boldsymbol{A}$ & $\boldsymbol{A}$, $\boldsymbol{A}$ & $\boldsymbol{A}^2$, $\boldsymbol{A}^2$ & $\boldsymbol{A}^2$



Datasets	$\boldsymbol{A}$ & $\boldsymbol{A}$	$\boldsymbol{A}$ & $\boldsymbol{A}^2$	$\boldsymbol{A}^2$ & $\boldsymbol{A}^2$
Cora	37.0±6.1	**53.7±3.2**	33.3±2.1
Citeseer	35.4±3.9	**44.7±5.0**	35.8±4.1
BlogCatalog	50.6±3.2	**63.1±4.6**	56.2±2.1
Flickr	26.6±2.6	**33.7±2.3**	28.2±1.6

Analysis of The General Graph Augmentation Rule

➢ **Theoretical analysis ---** Why does GAME rule work?

Theorem 1. *(Contrastive Invariance) Given adjacency matrix $\boldsymbol{A}$ and the generated augmentation $\boldsymbol{V}$, the amplitudes of i-th frequency of $\boldsymbol{A}$ and $\boldsymbol{V}$ are λ_i and γ_i, respectively. With the optimization of InfoNCE loss $\mathcal{L}_{InfoNCE}$, the following upper bound is established:*

$$\mathcal{L}_{InfoNCE} \leq \frac{1+N}{2} \sum_i \theta_i \left[2 - (\lambda_i - \gamma_i)^2\right],$$

where θ_i is an adaptive weight of the ith term.

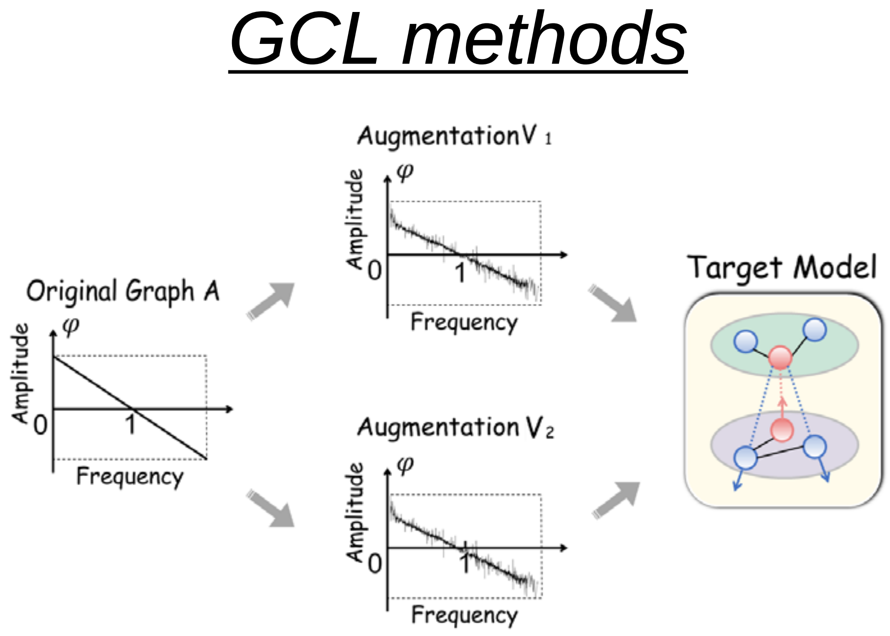
Interpretation

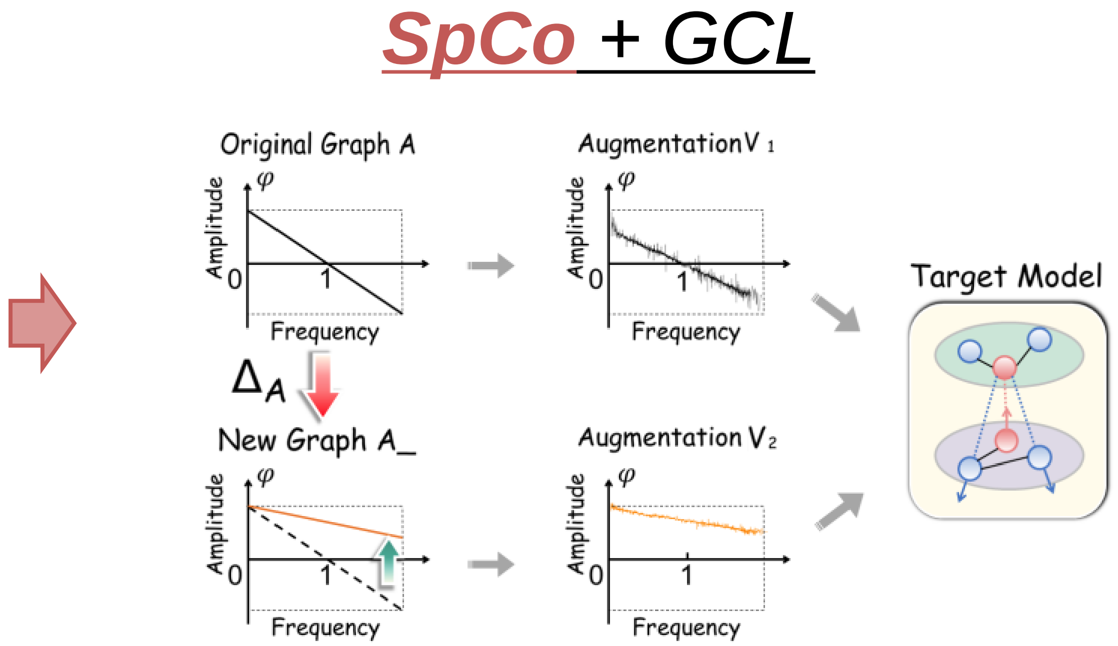
- We find a upper bound for InfoNCE loss.
- Model optimization → Upper bound rising → larger θ_i attaches to smaller $(\lambda_i - \gamma_i)^2$
  → capture invariance between contrasted views
- The GAME rule emphasizes small difference in low-frequency part, so makes model capture low-frequency information.

SpCo

Spectral Graph Contrastive Learning --- A friendly plug-in

- **Target ---** Learn a transformation Δ_A, construct optimal contrastive pair A and $A\_$





Spectral Graph Contrastive Learning --- A friendly plug-in

- **Target ---** Learn a transformation Δ_A, construct optimal contrastive pair A and A_-
- **Optimization objective** (maximization)

$$\mathcal{J} = \underbrace{< \mathcal{C}, \Delta_{A+} >^2}_{\text{Matching Term}} + \underbrace{\epsilon H(\Delta_{A+})}_{\text{Entropy Reg.}} + \underbrace{< f, \Delta_{A+} \mathbb{1}_n - a > + < g, \Delta_{A+}^\top \mathbb{1}_n - b >}_{\text{Lagrange Constraint Conditions}}$$

- $\Delta A = \Delta_{A+} - \Delta_{A-}$
- <**U**, **V**> = $\sum_{ij} U_{ij} V_{ij}$
- $\mathcal{C} = U g(\lambda) U^\top$
- *g(λ)*: monotone increasing

- **Entropy regularization**

  => Expose more edges to optimization

- **Lagrange constraint conditions**

  => The row and column sums meet distribution $\boldsymbol{a}$ and $\boldsymbol{b}$

Spectral Graph Contrastive Learning --- A friendly plug-in

- **Target ---** Learn a transformation Δ_A, construct optimal contrastive pair A and $A\_$
- **Optimization objective** (maximization)

$$\mathcal{J} = \underbrace{< \boldsymbol{\mathcal{C}},\ \Delta_{\boldsymbol{A}+} >^2}_{\text{Matching Term}} + \underbrace{\epsilon H(\Delta_{\boldsymbol{A}+})}_{\text{Entropy Reg.}} + \underbrace{< \boldsymbol{f}, \Delta_{\boldsymbol{A}+}\mathbb{1}_n - \boldsymbol{a} > + < \boldsymbol{g}, \Delta_{\boldsymbol{A}+}^{\top}\mathbb{1}_n - \boldsymbol{b} >}_{\text{Lagrange Constraint Conditions}}$$

- **Solution**

$$\Delta_{\boldsymbol{A}+} = diag(\boldsymbol{u}) \exp\left(2 < \boldsymbol{\mathcal{C}}, \Delta'_{\boldsymbol{A}+} > \boldsymbol{\mathcal{C}} \,/\, \epsilon\right) diag(\boldsymbol{v}) = \boldsymbol{U}_+\boldsymbol{K}_+\boldsymbol{V}_+$$

matrix scaling $\boldsymbol{u} * (\boldsymbol{K}_+\boldsymbol{v}) = \boldsymbol{a}$ and $\boldsymbol{v} * (\boldsymbol{K}_+^{\top}\boldsymbol{u}) = \boldsymbol{b}$

Sinkhorn's Iteration: $\boldsymbol{u}^{(l+1)} = \boldsymbol{a} \,/\, \left(\boldsymbol{K}_+\boldsymbol{v}^{(l)}\right)$ and $\boldsymbol{v}^{(l+1)} = \boldsymbol{b} \,/\, \left(\boldsymbol{K}_+^{\top}\boldsymbol{u}^{(l+1)}\right)$

Spectral Graph Contrastive Learning --- A friendly plug-in

- **Target ---** Learn a transformation Δ_A, construct optimal contrastive pair A and $A\_$
- **Optimization objective** (maximization)

$$\mathcal{J} = \underbrace{< \mathcal{C},\ \Delta_{A+} >^2}_{\text{Matching Term}} + \underbrace{\epsilon H(\Delta_{A+})}_{\text{Entropy Reg.}} + \underbrace{< \boldsymbol{f}, \Delta_{A+}\mathbb{1}_n - \boldsymbol{a} > + < \boldsymbol{g}, \Delta_{A+}^{\top}\mathbb{1}_n - \boldsymbol{b} >}_{\text{Lagrange Constraint Conditions}}$$

- **Solution**

$$\Delta_{A+} = diag(\boldsymbol{u}) \exp\left(2 < \mathcal{C}, \Delta'_{A+} > \mathcal{C} \,/\, \epsilon\right) diag(\boldsymbol{v}) = \boldsymbol{U}_+\boldsymbol{K}_+\boldsymbol{V}_+$$

$$\Delta_{A-} = diag(\boldsymbol{u}') \exp\left(-2 < \mathcal{C}, \Delta'_{A-} > \mathcal{C} \,/\, \epsilon\right) diag(\boldsymbol{v}') = \boldsymbol{U}_-\boldsymbol{K}_-\boldsymbol{V}_-$$

$$\Delta_{A} = \Delta_{A+} - \Delta_{A-}$$

$$\boldsymbol{A}\_ = \boldsymbol{A} + \eta \cdot \mathbb{S} * \Delta_{A}$$

Experiments

Datasets.

Dataset	Nodes	Edges	Classes	Features	Training	Validation	Test
Cora	2708	10556	7	1433	35/70/140	500	1000
Citeseer	3327	9228	6	3703	30/60/120	500	1000
BlogCatalog	5196	343486	6	8189	30/60/120	1000	1000
Flickr	7575	479476	9	12047	45/90/180	1000	1000
Pubmed	19717	88651	3	500	15/30/60	500	1000

BaseLines

01 Classical GNN methods

GCN, GAT

02 GCL models

DGI, MVGRL, GRACE

GCA, GraphCL, CCA-SSG

Tasks.

- Node classification
- Visualization of graph spectrum

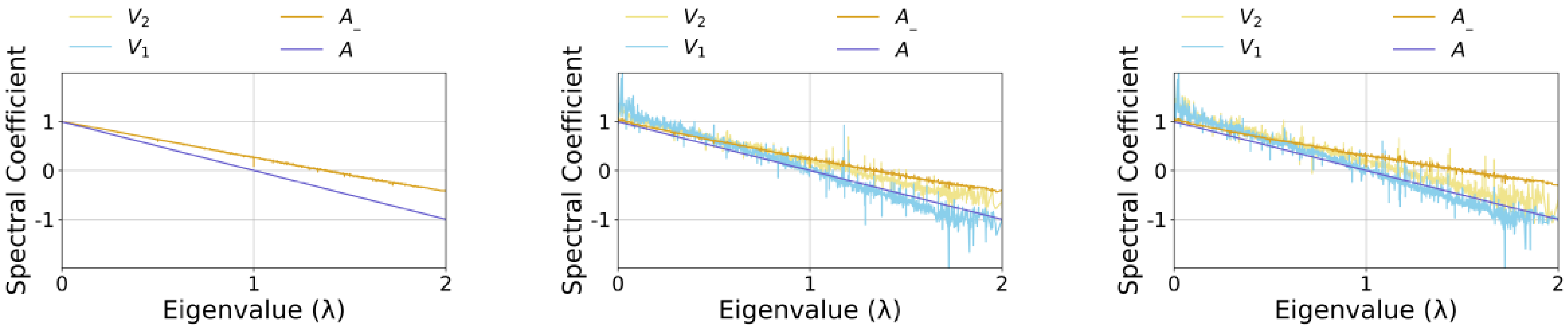
Node classification

Base model: DGI (BCE loss), GRACE (InfoNCE loss), CCA-SSG (CCA loss)

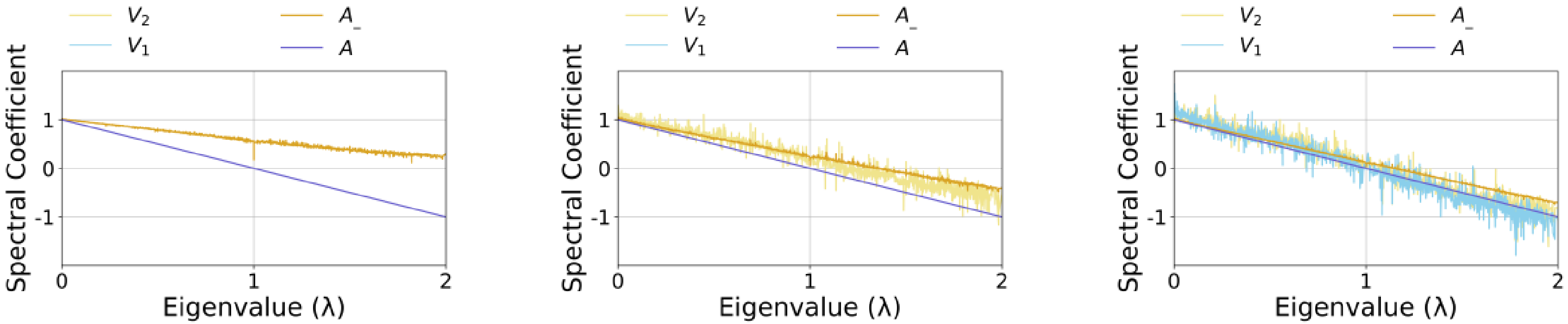
Datasets	Metrics	GCN	GAT	DGI	**DGI+SpCo**	MVGRL	GRACE	**GRACE+SpCo**	GCA	GraphCL	CCA-SSG	**CCA+SpCo**
Cora	Ma-F1	79.6±0.7	81.3±0.3	80.4±0.7	**81.1±0.5**	81.5±0.5	79.2±1.0	**80.3±0.8**	79.9±1.1	80.7±0.9	82.9±0.8	**83.6±0.4**
	Mi-F1	80.7±0.6	82.3±0.2	82.0±0.5	**82.8±0.7**	82.8±0.4	80.0±1.0	**81.2±0.9**	81.1±1.0	82.3±0.9	83.6±0.9	**84.3±0.4**
Citeseer	Ma-F1	68.1±0.5	67.5±0.2	67.7±0.9	**68.3±0.5**	66.8±0.7	65.1±1.2	**65.1±0.8**	62.8±1.3	67.8±1.0	67.9±1.0	**68.5±1.0**
	Mi-F1	70.9±0.5	72.0±0.9	71.7±0.8	**72.4±0.5**	72.5±0.5	68.7±1.1	**69.4±1.0**	65.9±1.0	71.9±0.9	73.1±0.7	**73.6±1.1**
BlogCatalog	Ma-F1	71.2±1.2	67.6±2.2	68.2±1.3	**71.5±0.8**	80.3±3.6	67.7±1.2	**68.2±0.4**	71.7±0.4	63.9±2.1	72.0±0.5	**72.8±0.3**
	Mi-F1	72.1±1.3	68.3±2.2	68.8±1.4	**72.3±0.9**	80.9±3.6	68.5±1.3	**69.4±1.3**	72.7±0.5	64.6±2.1	73.0±0.5	**73.7±0.3**
Flickr	Ma-F1	48.9±1.6	35.0±0.8	31.2±1.6	**33.7±0.7**	31.2±2.9	35.7±1.3	**36.3±1.4**	41.2±0.5	32.1±1.1	37.0±1.1	**38.7±0.6**
	Mi-F1	50.2±1.2	37.1±0.3	33.0±1.6	**35.2±0.7**	33.4±3.0	37.3±1.0	**38.1±1.3**	42.2±0.6	34.5±0.9	39.3±0.9	**40.4±0.4**
PubMed	Ma-F1	78.5±0.3	77.4±0.2	76.8±0.9	**77.6±0.6**	79.8±0.4	80.0±0.7	**80.3±0.3**	80.8±0.6	77.0±0.4	80.7±0.6	**81.3±0.3**
	Mi-F1	78.9±0.3	77.8±0.2	76.7±0.9	**77.4±0.5**	79.7±0.3	79.9±0.7	**80.7±0.2**	81.4±0.6	76.8±0.5	81.0±0.6	**81.5±0.4**

SpCo can generally improve performances compared with base models

Visualization of graph spectrum



(a) DGI: Cora (b) GRACE: Cora (c) CCA-SSG: Cora



(a) DGI: Citeseer (b) GRACE: Citeseer (c) CCA-SSG: Citeseer

***A* and $A\_$ is optimal contrastive pair, so boosting the final results.**

Conclusion

01 Augmentation strategies & Graph spectrum.

Reveal the general graph augmentation rule (The GAME rule)

Explain why GCL works (Contrastive Invariance Theorem)

02 Optimal contrastive pair & SpCo.

Propose a novel concept -- optimal contrastive pair

Theoretically derive a general GCL framework -- SpCo

03 Extensive experiments.

DGI/GRACE/CCA-SSG + SpCo, validate the effectiveness of SpCo

Thank you!

Contact:

Nian Liu, nianliu@bupt.edu.cn

More Information:

http://www.shichuan.org/

Our official account:

图数据挖掘和机器学习