Revisiting Graph Contrastive Learning from the Perspective of Graph Spectrum

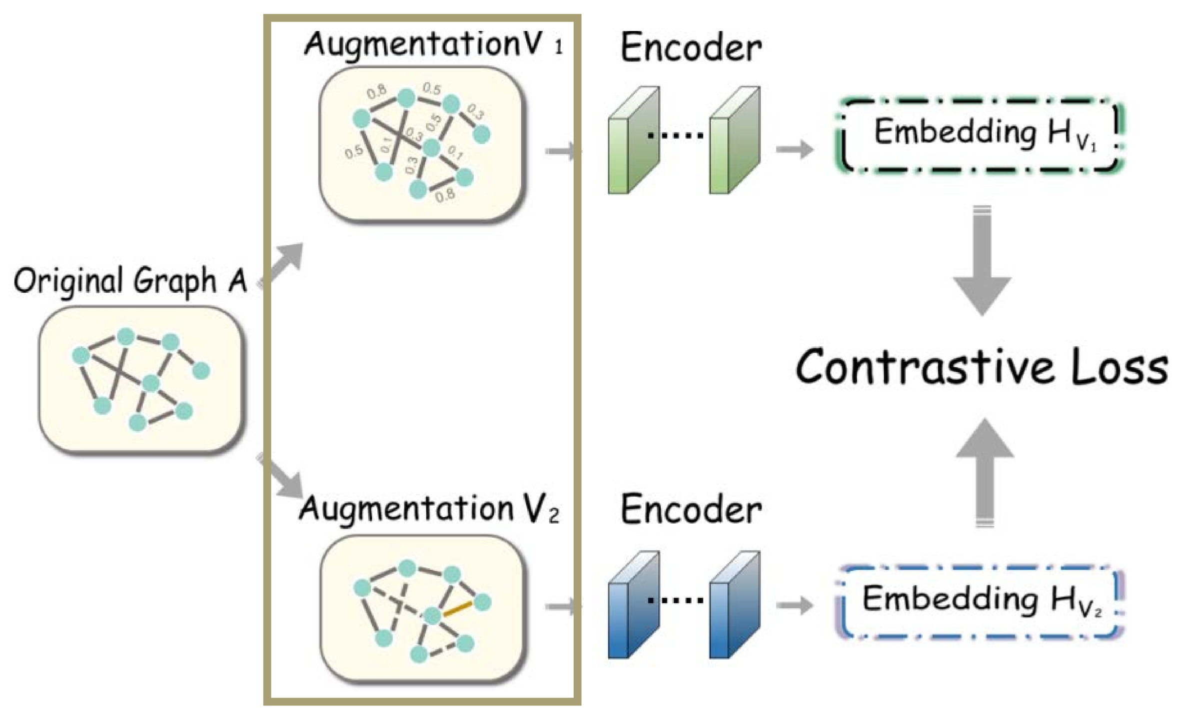
Nian Liu, Xiao Wang, Deyu Bo, Chuan Shi*, Jian Pei

Beijing University of Posts and Telecommunications

& Simon Fraser University

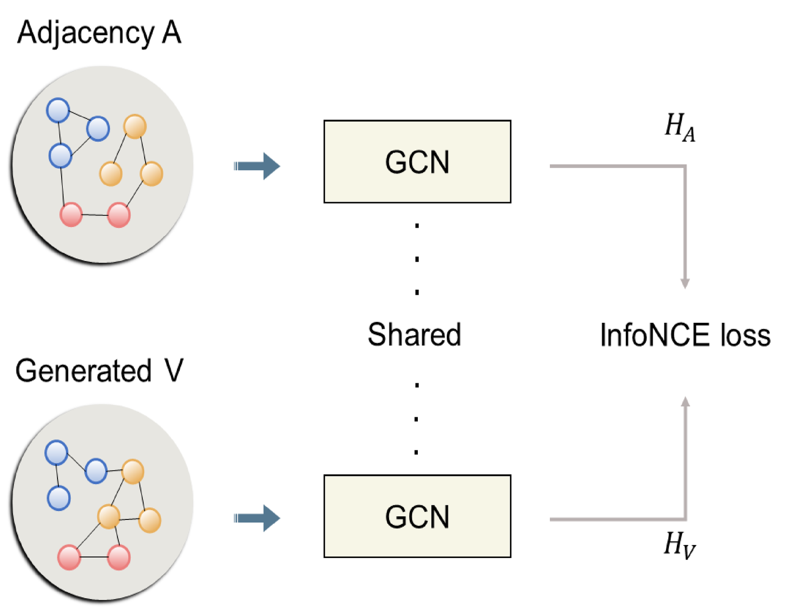
NEURAL INFORMATION PROCESSING SYSTEMS

Rethinking Graph Augmentation in Graph Contrastive Learning (GCL)

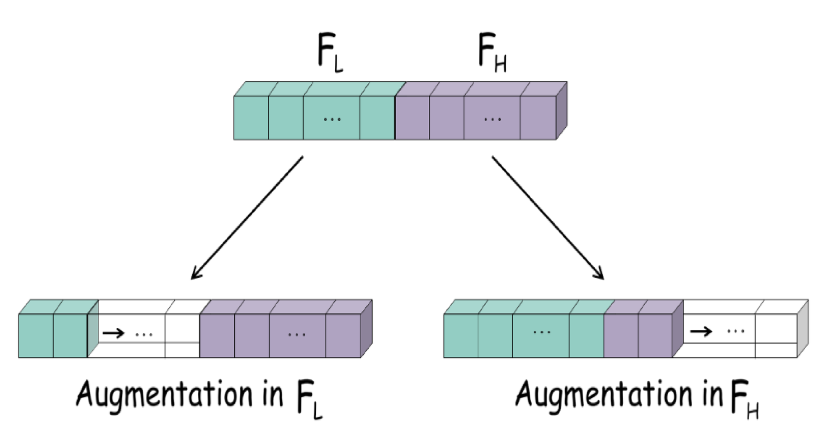


➢ *Our target is to uncover some* ***general rule*** *across different graph augmentation strategies, and use this rule to* ***validate and improve*** *the current GCL methods?*

Impact of Graph Augmentation

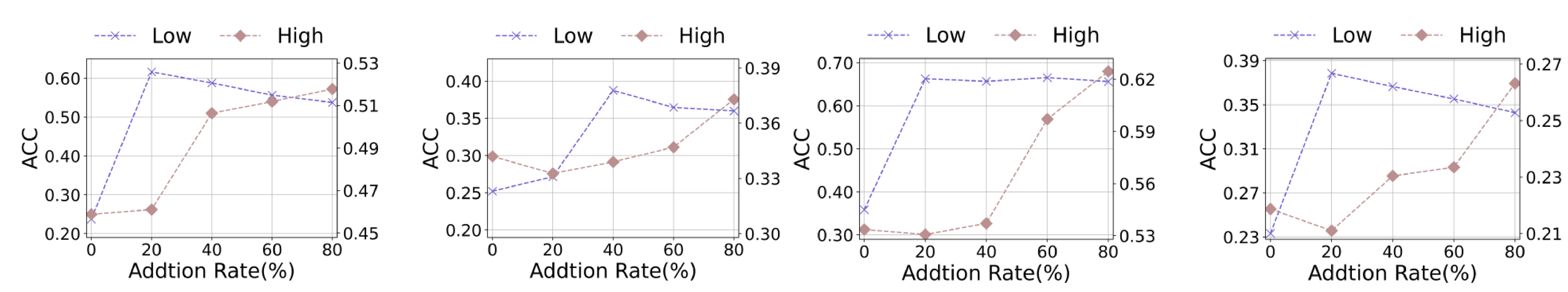


Case study model

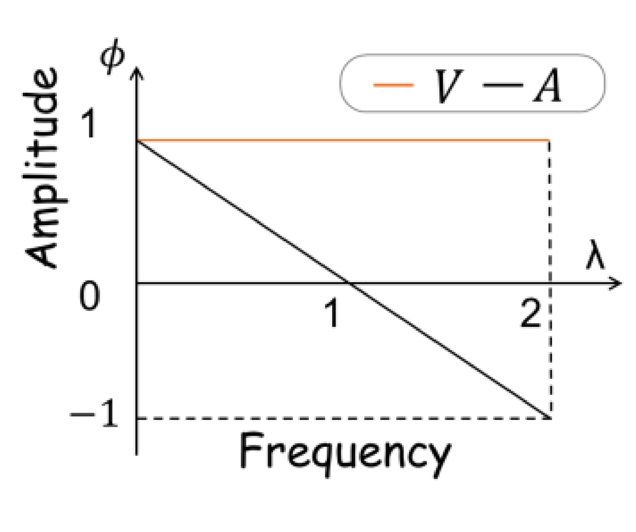


Generating V

➢ **Result & Analysis**:



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



➢ ***Maintain the lowest part*** *of* $\mathcal{F}_\mathcal{L}$

- Performance achieves the best
- Difference in $\mathcal{F}_\mathcal{L}$ is smaller

➢ ***More high frequencies*** *in* $\mathcal{F}_\mathcal{H}$

- Performance generally rises
- Difference in $\mathcal{F}_\mathcal{H}$ is larger

Experience

➢ Node classification

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**

The General Augmentation (GAME rule)

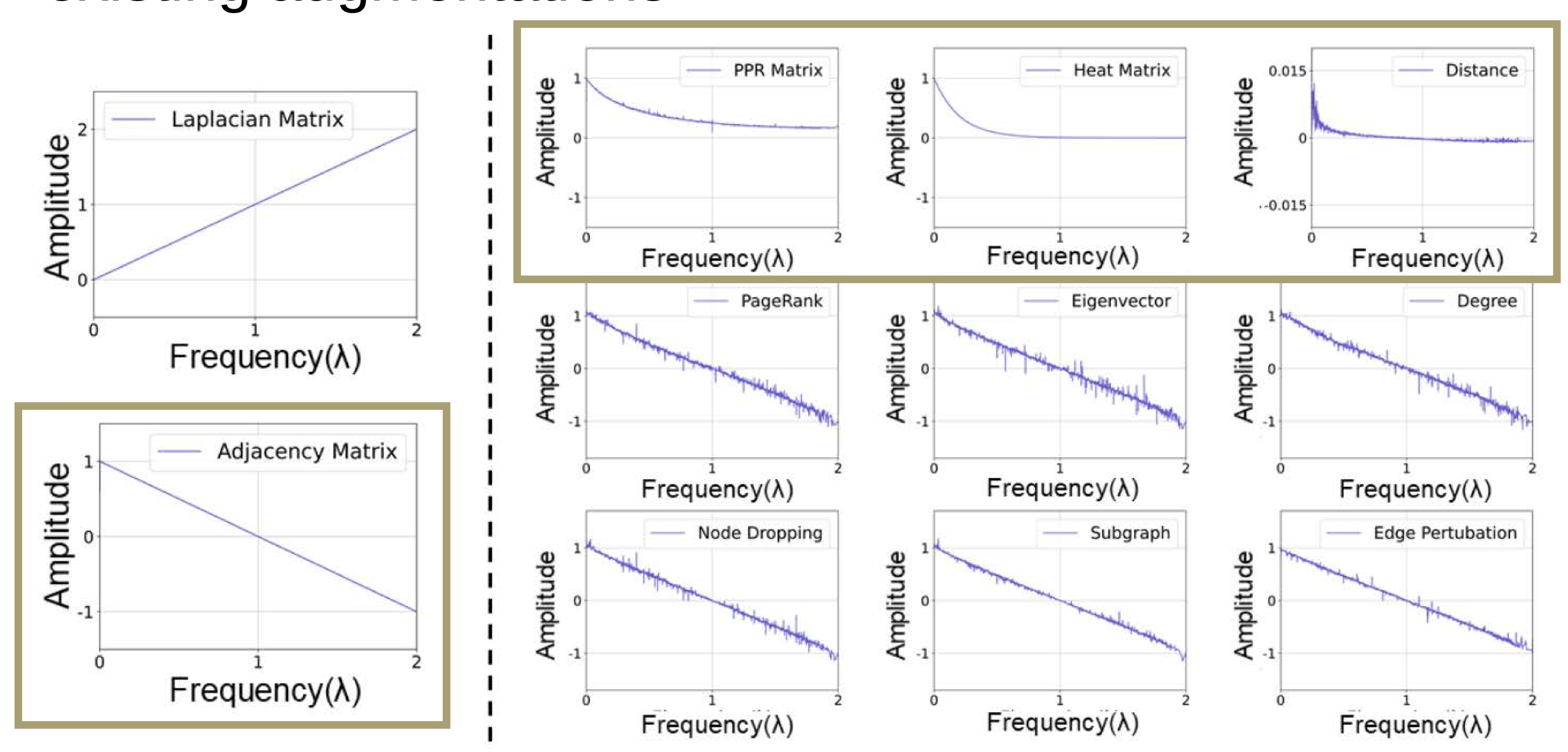
The General Graph Augmentation Rule

Given two random augmentations V_1 and V_2, their graph spectrums are $\phi_{V_1}(\lambda)$ and $\phi_{V_2}(\lambda)$. Then, $\forall\ \lambda_m \in [1,2]$ and $\lambda_n \in [0,1]$, V_1 and 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.

☐ **Experimental analysis** --- 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

☐ **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:*

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

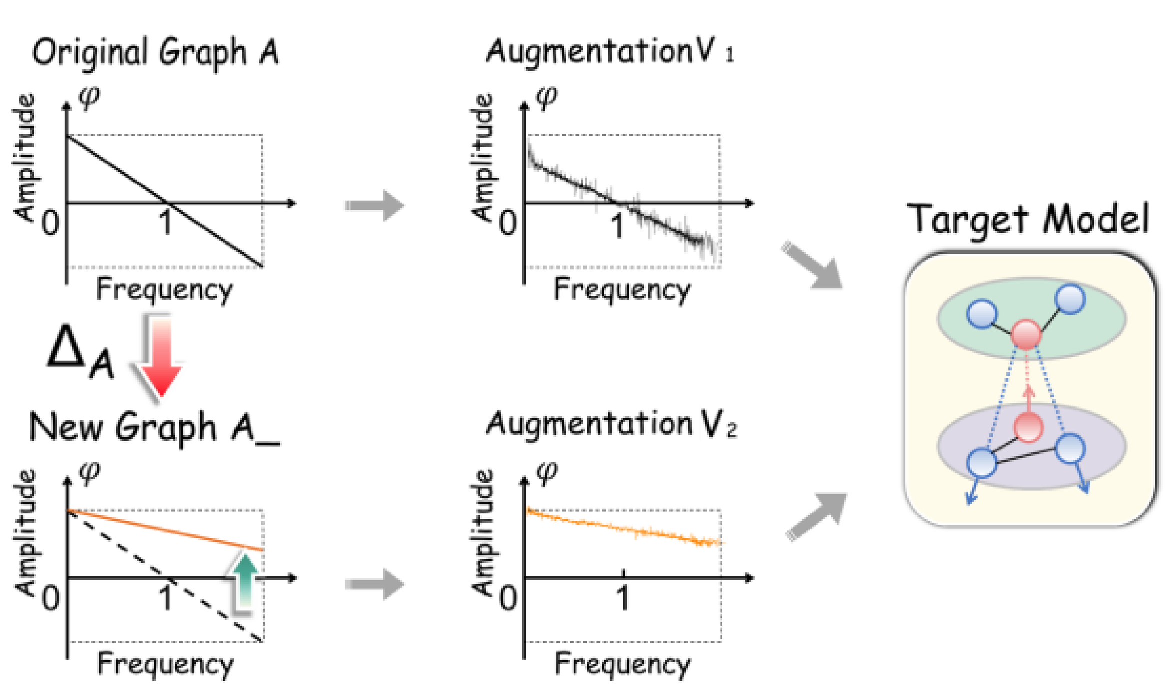
where θ_i *is an adaptive weight of the ith term.*

➢ We are the first to indicates that GCL can make encoder ***capture invariance*** between two contrastive views.

➢ The GAME rule requires smaller difference in low-frequency part → emphasize ***low-frequency information***

Spectral Graph Contrastive Learning

➢ Target: learn a transformation Δ_A from *A* to *A_*



➢ Optimization Objective

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

➢ Solution

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