PolyGCL: GRAPH CONTRASTIVE LEARNING via Learnable Spectral Polynomial Filters

Jingyu Chen, Runlin Lei, Zhewei Wei*

Contact: zhewei@ruc.edu.cn

Motivation

- Graph Contrastive Learning (GCL)
  - Graph self-supervised learning: without label y, only $\mathbf{A}$ & $\mathbf{X}$.
  - Output embeddings $\mathbf{Z}$ for downstream tasks.
- Classical GCL paradigms:
  - BCE-based: DGI [Veličković et al., 2019]

$$\mathcal{L} = \frac{1}{N+M}\left(\sum_{i=1}^{N}\mathbb{E}_{(\mathbf{X},\mathbf{A})}\left[\log \mathcal{D}\left(\vec{h}_i,\vec{s}\right)\right] + \sum_{j=1}^{M}\mathbb{E}_{(\widetilde{\mathbf{X}},\widetilde{\mathbf{A}})}\left[\log\left(1-\mathcal{D}\left(\vec{\tilde{h}}_j,\vec{s}\right)\right)\right]\right)$$

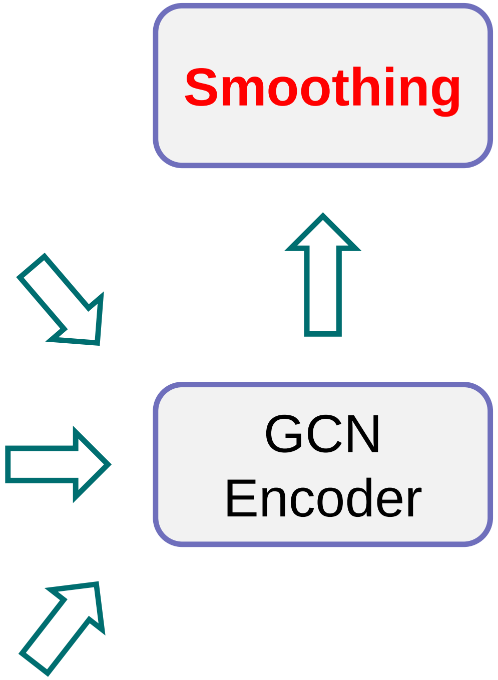
  - InfoNCE-based: GRACE [Zhu et al., 2020b]

$$\ell(\boldsymbol{u}_i,\boldsymbol{v}_i) = \log \frac{e^{\theta(\boldsymbol{u}_i,\boldsymbol{v}_i)/\tau}}{\underbrace{e^{\theta(\boldsymbol{u}_i,\boldsymbol{v}_i)/\tau}}_{\text{the positive pair}} + \underbrace{\sum_{k=1}^{N}\mathbb{1}_{[k\neq i]}e^{\theta(\boldsymbol{u}_i,\boldsymbol{v}_k)/\tau}}_{\text{inter-view negative pairs}} + \underbrace{\sum_{k=1}^{N}\mathbb{1}_{[k\neq i]}e^{\theta(\boldsymbol{u}_i,\boldsymbol{u}_k)/\tau}}_{\text{intra-view negative pairs}}},$$

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

  - Invariance-keeping: CCA-SSG [Zhang et al., 2021]

$$\mathcal{L} = \underbrace{\left\|\tilde{\mathbf{Z}}_A - \tilde{\mathbf{Z}}_B\right\|_F^2}_{\text{invariance term}} + \lambda\underbrace{\left(\left\|\tilde{\mathbf{Z}}_A^\top\tilde{\mathbf{Z}}_A - \mathbf{I}\right\|_F^2 + \left\|\tilde{\mathbf{Z}}_B^\top\tilde{\mathbf{Z}}_B - \mathbf{I}\right\|_F^2\right)}_{\text{decorrelation term}}$$



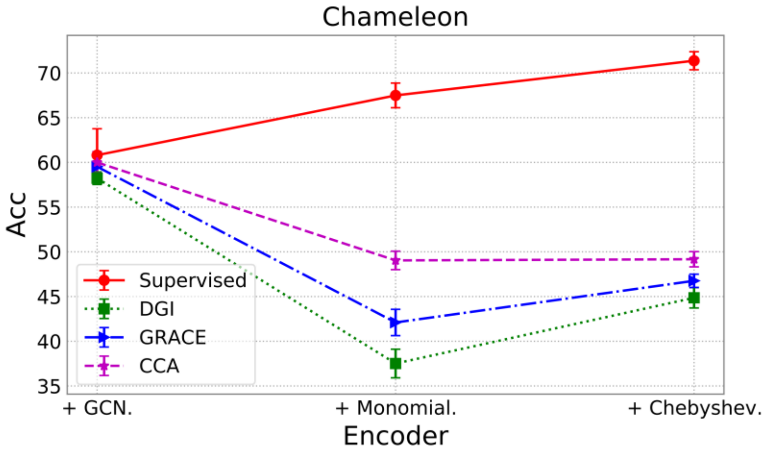
Motivation

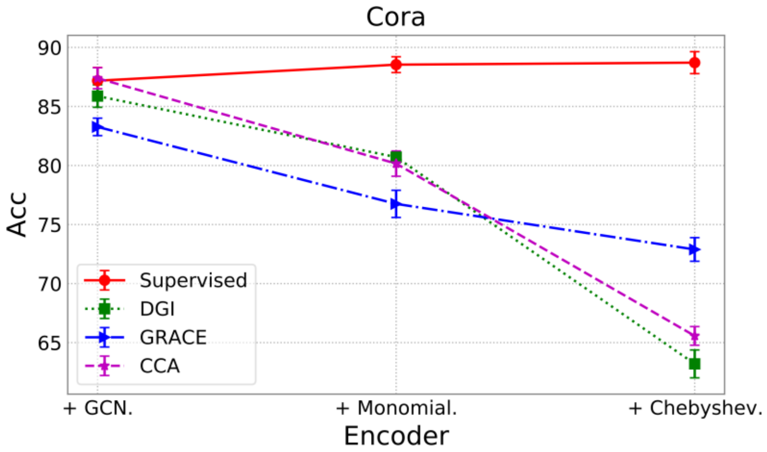
- Spectral GNNs

$$\boldsymbol{y} = \boldsymbol{U}\begin{pmatrix} h(\lambda_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & h(\lambda_n) \end{pmatrix}\boldsymbol{U}^T\boldsymbol{x} \approx \sum_{k=0}^{K} w_k \boldsymbol{L}^k \boldsymbol{x}$$

  - Monomial base: $\boldsymbol{Y} = \sum_{k=0}^{K} \gamma_k \boldsymbol{H}^{(k)}, \boldsymbol{H}^{(k)} = \widetilde{\boldsymbol{P}}\boldsymbol{H}^{(k-1)}$
  - Chebyshev base: $\mathbf{Y} = \frac{2}{K+1}\sum_{k=0}^{K} w_k T_k(\hat{\boldsymbol{L}})\boldsymbol{X}$, $T_k(\hat{\mathbf{L}}) = 2\hat{\mathbf{L}}T_{k-1}(\hat{\mathbf{L}}) - T_{k-2}(\hat{\mathbf{L}})$

- A natural idea: *Can we incorporate the excellent properties of spectral polynomial filters into graph contrastive learning?*





Spectral GNNs in supervised settings.

Plug in GCL setting directly?

Model: PolyGCL

Encoder

- ChebNetII [He et al., 2022]

$$\mathbf{Y}=\frac{2}{K+1}\sum_{k=0}^{K} w_k T_k(\hat{\boldsymbol{L}})\mathbf{X}$$

  - $\tilde{\mathbf{L}}=\mathbf{I}-\tilde{\mathbf{A}}=\mathbf{I}-\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$, $\lambda\in[0,2]$;
  - $\hat{\mathbf{L}}=\frac{2\tilde{\mathbf{L}}}{\lambda_{\max}}-\mathbf{I}$, $\hat{\lambda}\in[-1,1]$;
  - Reparameterize coefficient $w_k=\sum_{j=0}^{K}\gamma_j T_k(x_j)$
    - $x_j=\cos((j+1/2)\pi/(K+1))$ is the Chebyshev Node.
    - γ_j is the learnable parameter, corresponding to filter value $h(x_j)$.

- Decoupling low-pass and high-pass:
  - **Prefix sum:** $\gamma_i^H=\sum_{j=0}^{i}\gamma_j, i=1,\dots,K$
  - **Prefix difference:** $\gamma_i^L=\gamma_0-\sum_{j=1}^{i}\gamma_j, i=1,\dots,K$
  - **Initialization:** $\gamma_0^H=\gamma_0^L=\gamma_0$
  - $\gamma_i^H\le\gamma_{i+1}^H, \gamma_i^L\ge\gamma_{i+1}^L$

Polynomial Encoder

$$\mathbf{Z}_L=f_\theta\left(\sum_{k=0}^{K} w_k^L T_k(\hat{\mathbf{L}})\mathbf{X}\right),$$
$$\mathbf{Z}_H=f_\theta\left(\sum_{k=0}^{K} w_k^H T_k(\hat{\mathbf{L}})\mathbf{X}\right)$$

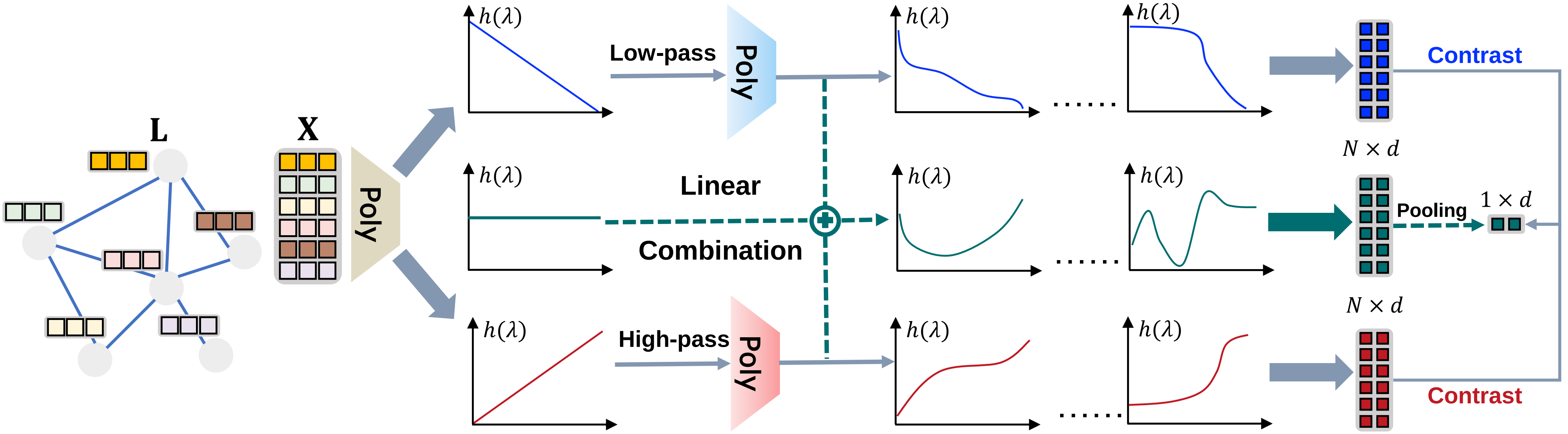
Model: PolyGCL

Optimization

- Final embedding via linear combination
  - $\mathbf{Z} = \alpha\mathbf{Z}_L + \beta\mathbf{Z}_H$
- Binary Cross-Entropy (BCE) Loss
  - $\mathcal{L}_{\mathrm{BCE}} = \frac{1}{4N}\left(\sum_{i=1}^{N} \log \mathcal{D}\left(\mathbf{Z}_L^i, \mathbf{g}\right) + \log\left(1 - \mathcal{D}\left(\tilde{\mathbf{Z}}_L^i, \mathbf{g}\right)\right) + \log \mathcal{D}\left(\mathbf{Z}_H^i, \mathbf{g}\right) + \log\left(1 - \mathcal{D}\left(\tilde{\mathbf{Z}}_H^i, \mathbf{g}\right)\right)\right)$

Global summary

$$\mathbf{g} = \mathrm{Mean}(\mathbf{Z}) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{Z}_i$$



Revisit GCL From The Spectral View

- Graph filtering in GCL
  - Step 1: $\mathbf{Z} = \mathbf{U}g(\mathbf{\Lambda})\mathbf{U}^T\mathbf{X}$;
  - Step 2: $\sigma(\mathrm{MLP}(\mathbf{Z}))$.
- Spectral Regression Loss (SRL)
  - EvenNet [Lei et al., 2022]
  - $L(\mathcal{G}) = \sum_{i=0}^{N-1}\left(\frac{\alpha_i}{\sqrt{N}} - \frac{g(\lambda_i)\beta_i}{\sqrt{\sum_{j=0}^{N-1} g(\lambda_j)^2 \beta_j^2}}\right)^2$, where $\boldsymbol{\alpha} = \mathbf{U}^\top \Delta \boldsymbol{y}$ and $\boldsymbol{\beta} = \mathbf{U}^\top \mathbf{X}$.

Theorem 1 [Informal]

For the low-pass filter $g_{\mathrm{low}} = c - \frac{c}{2}\lambda \in [0, \mathrm{c}]$ and the high-pass filter as $g_{\mathrm{high}} = \frac{c}{2}\lambda \in [0, \mathrm{c}]$, then a linear combination of the low-pass and high-pass filter $g_{\mathrm{joint}} = x g_{\mathrm{low}} + y g_{\mathrm{high}}$, $x \geq 0, y \geq 0, x + y = 1$ achieves a lower expected SRL upper bound than g_{low} in heterophilic settings.

Theoretical Analysis

Theorem 2 [Informal]

Denote $\mathcal{E}^{L}(\cdot)$ and $\mathcal{E}^{H}(\cdot)$ as the low-pass polynomial filter and high-pass polynomial filter respectively. Let $\mathbf{Z}_i^{(k)}$ be the k-hop neighboring embedding set of node i, $\mathbf{h}_i^{L} = \mathcal{E}^{L}\left(\mathbf{Z}_i^{(k)}\right)$, $\mathbf{h}_i^{H} = \mathcal{E}^{H}\left(\mathbf{Z}_i^{(k)}\right)$, and $\mathbf{h}_i^{agg}$=Linear($\mathbf{h}_i^{L}$, $\mathbf{h}_i^{H}$).Then, $\mathbf{h}_i^{agg}$ that optimising $\mathcal{L}_{\mathrm{BCE}}$ also maximizes the mutual information between $\mathbf{h}_i^{agg}$ and $\mathbf{Z}_i^{(k)}$.

$$\mathcal{L}_{\mathrm{BCE}} = \frac{1}{4N}\left(\sum_{i=1}^{N} \log \mathcal{D}\left(\mathbf{Z}_L^i, \mathbf{g}\right) + \log\left(1 - \mathcal{D}\left(\tilde{\mathbf{Z}}_L^i, \mathbf{g}\right)\right) + \log \mathcal{D}\left(\mathbf{Z}_H^i, \mathbf{g}\right) + \log\left(1 - \mathcal{D}\left(\tilde{\mathbf{Z}}_H^i, \mathbf{g}\right)\right)\right)$$

$$\mathcal{L}_{\mathrm{BCE}} = 2JS\left(P_L^{\mathrm{pos}} \parallel P_L^{\mathrm{neg}}\right) + 2JS\left(P_H^{\mathrm{pos}} \parallel P_H^{\mathrm{neg}}\right) - 4\log 2$$

- **Corollary 2 [Informal].** The downstream classification error bound on learned representation $\mathbf{h}_{agg}$ is lower than both the error upper bound of the low-pass and high-pass representations, $\mathbf{h}_{low}$ and $\mathbf{h}_{high}$.

Experiments

- **Downstream task**
  - Node classification
  - Split: 60%/20%/20%
- **Datasets**
  - Synthetic:
    - cSBM [Chien et al., 2021]
    - Parameter $\phi \in [-1,1]$
  - Real-world:
    - Homophilic & Heterophilic

Methods	$\phi=-1$	$\phi=-0.75$	$\phi=-0.5$	$\phi=-0.25$	$\phi=0$	$\phi=0.25$	$\phi=0.5$	$\phi=0.75$	$\phi=1$
DGI	83.04 $\pm$ 0.92	93.24 $\pm$ 0.54	85.75 $\pm$ 0.49	68.41 $\pm$ 0.94	59.95 $\pm$ 0.78	68.70 $\pm$ 0.60	84.04 $\pm$ 0.61	91.53 $\pm$ 0.42	82.68 $\pm$ 0.72
MVGRL	68.80 $\pm$ 1.00	84.35 $\pm$ 0.78	78.81 $\pm$ 0.63	64.14 $\pm$ 1.05	59.09 $\pm$ 1.15	70.74 $\pm$ 0.73	89.91 $\pm$ 0.58	95.95 $\pm$ 0.37	89.13 $\pm$ 0.55
GGD	82.90 $\pm$ 0.83	92.76 $\pm$ 0.63	85.56 $\pm$ 0.58	66.63 $\pm$ 0.66	56.00 $\pm$ 0.51	67.06 $\pm$ 1.06	84.22 $\pm$ 0.61	91.75 $\pm$ 0.45	83.84 $\pm$ 0.76
GMI	54.47 $\pm$ 0.94	54.38 $\pm$ 0.71	50.70 $\pm$ 0.91	50.41 $\pm$ 0.64	51.79 $\pm$ 0.39	59.57 $\pm$ 0.93	82.28 $\pm$ 0.76	93.74 $\pm$ 0.46	96.01 $\pm$ 0.48
CCA-SSG	50.55 $\pm$ 0.75	52.71 $\pm$ 1.08	51.21 $\pm$ 0.98	50.88 $\pm$ 0.85	51.16 $\pm$ 0.67	56.33 $\pm$ 0.90	72.41 $\pm$ 1.20	90.83 $\pm$ 0.62	62.03 $\pm$ 0.91
BGRL	49.86 $\pm$ 0.77	49.47 $\pm$ 0.74	49.95 $\pm$ 0.90	50.21 $\pm$ 0.87	54.58 $\pm$ 0.99	60.80 $\pm$ 0.56	70.79 $\pm$ 1.01	74.46 $\pm$ 0.79	68.69 $\pm$ 0.96
GBT	57.41 $\pm$ 1.43	64.99 $\pm$ 0.53	58.84 $\pm$ 0.80	51.80 $\pm$ 0.87	57.55 $\pm$ 0.69	**72.62 $\pm$ 0.63**	**91.09 $\pm$ 0.37**	97.80 $\pm$ 0.25	96.03 $\pm$ 0.38
GRACE	98.74 $\pm$ 0.28	**97.55 $\pm$ 0.17**	90.06 $\pm$ 0.50	68.74 $\pm$ 1.01	56.85 $\pm$ 1.12	66.70 $\pm$ 0.91	89.50 $\pm$ 0.60	97.41 $\pm$ 0.25	98.78 $\pm$ 0.28
GCA	76.56 $\pm$ 0.92	85.56 $\pm$ 0.40	78.96 $\pm$ 0.43	62.32 $\pm$ 0.89	58.01 $\pm$ 1.07	65.30 $\pm$ 1.15	77.16 $\pm$ 1.03	81.38 $\pm$ 0.59	75.54 $\pm$ 0.76
GraphCL	58.82 $\pm$ 1.06	57.89 $\pm$ 0.68	52.91 $\pm$ 0.70	50.18 $\pm$ 0.59	51.25 $\pm$ 0.76	55.11 $\pm$ 0.56	62.54 $\pm$ 1.13	65.57 $\pm$ 1.17	71.31 $\pm$ 1.01
GREET	50.82 $\pm$ 0.67	58.79 $\pm$ 0.52	59.91 $\pm$ 1.09	63.57 $\pm$ 0.76	65.99 $\pm$ 0.64	71.04 $\pm$ 0.67	80.17 $\pm$ 0.50	83.11 $\pm$ 0.53	75.93 $\pm$ 1.19
PolyGCL	**98.84 $\pm$ 0.17**	94.23 $\pm$ 0.31	**90.82 $\pm$ 0.50**	**75.43 $\pm$ 0.68**	**66.51 $\pm$ 0.69**	69.43 $\pm$ 0.65	88.22 $\pm$ 0.72	**98.09 $\pm$ 0.29**	**99.29 $\pm$ 0.23**

Methods	Cora	Citeseer	Pubmed	Cornell	Texas	Wisconsin	Actor	Chameleon	Squirrel
DGI	85.88 $\pm$ 0.95	76.44 $\pm$ 0.80	82.13 $\pm$ 0.24	70.82 $\pm$ 7.21	81.48 $\pm$ 2.79	75.00 $\pm$ 2.00	32.09 $\pm$ 1.18	58.23 $\pm$ 0.70	38.80 $\pm$ 0.76
MVGRL	87.36 $\pm$ 0.64	78.70 $\pm$ 0.64	86.30 $\pm$ 0.23	67.70 $\pm$ 4.75	73.11 $\pm$ 4.75	74.25 $\pm$ 4.13	32.98 $\pm$ 0.53	57.75 $\pm$ 1.20	40.25 $\pm$ 1.14
GGD	87.21 $\pm$ 1.08	79.25 $\pm$ 0.72	85.38 $\pm$ 0.25	80.33 $\pm$ 1.80	82.62 $\pm$ 3.11	73.25 $\pm$ 2.25	32.27 $\pm$ 1.11	57.64 $\pm$ 1.16	40.87 $\pm$ 0.66
GMI	85.09 $\pm$ 1.13	76.38 $\pm$ 0.70	83.06 $\pm$ 0.24	62.79 $\pm$ 7.54	68.03 $\pm$ 4.10	62.13 $\pm$ 2.88	32.37 $\pm$ 1.01	62.47 $\pm$ 1.55	39.82 $\pm$ 0.93
CCA-SSG	87.39 $\pm$ 0.89	79.60 $\pm$ 0.71	84.96 $\pm$ 0.20	78.69 $\pm$ 3.44	87.87 $\pm$ 1.64	82.88 $\pm$ 1.50	34.86 $\pm$ 0.56	60.00 $\pm$ 1.20	41.50 $\pm$ 0.72
BGRL	84.45 $\pm$ 0.66	74.84 $\pm$ 1.04	83.06 $\pm$ 0.29	59.84 $\pm$ 2.95	69.84 $\pm$ 3.61	62.88 $\pm$ 4.13	32.48 $\pm$ 0.67	64.09 $\pm$ 1.27	47.02 $\pm$ 0.88
GBT	84.89 $\pm$ 1.13	76.59 $\pm$ 0.68	86.10 $\pm$ 0.23	59.18 $\pm$ 9.34	72.79 $\pm$ 6.56	62.38 $\pm$ 3.00	34.34 $\pm$ 0.67	68.77 $\pm$ 1.25	48.86 $\pm$ 0.80
GRACE	83.27 $\pm$ 0.74	73.79 $\pm$ 0.60	81.71 $\pm$ 0.16	60.66 $\pm$ 11.32	75.74 $\pm$ 2.95	72.13 $\pm$ 2.75	31.97 $\pm$ 1.15	59.52 $\pm$ 1.49	42.68 $\pm$ 0.90
GCA	84.09 $\pm$ 0.85	75.23 $\pm$ 0.75	82.01 $\pm$ 0.31	53.11 $\pm$ 9.34	81.97 $\pm$ 2.30	73.50 $\pm$ 3.00	31.13 $\pm$ 0.71	65.54 $\pm$ 1.07	47.13 $\pm$ 0.61
GraphCL	86.54 $\pm$ 0.54	78.99 $\pm$ 0.50	85.16 $\pm$ 0.21	61.48 $\pm$ 5.74	66.07 $\pm$ 6.07	60.63 $\pm$ 3.50	32.45 $\pm$ 1.22	58.49 $\pm$ 1.31	42.92 $\pm$ 0.62
GREET	85.16 $\pm$ 0.77	79.06 $\pm$ 0.44	85.64 $\pm$ 0.24	78.36 $\pm$ 3.77	78.03 $\pm$ 3.94	84.63 $\pm$ 3.88	38.26 $\pm$ 0.87	60.57 $\pm$ 1.03	39.76 $\pm$ 0.74
PolyGCL	**87.57 $\pm$ 0.62**	**79.81 $\pm$ 0.85**	**87.15 $\pm$ 0.27**	**82.62 $\pm$ 3.11**	**88.03 $\pm$ 1.80**	**85.50 $\pm$ 1.88**	**41.15 $\pm$ 0.88**	**71.62 $\pm$ 0.96**	**56.49 $\pm$ 0.72**

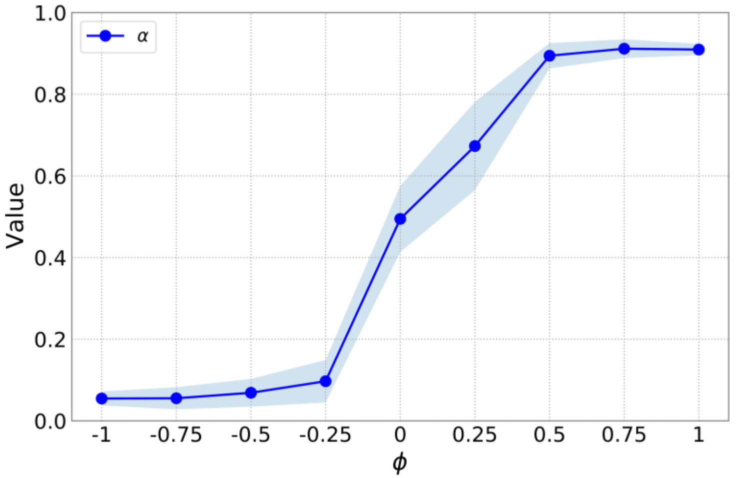
Methods	Roman-empire	Amazon-ratings	Minesweeper	Tolokers	Questions
DGI	58.57 $\pm$ 0.26	42.72 $\pm$ 0.42	68.36 $\pm$ 0.60	76.29 $\pm$ 0.66	74.44 $\pm$ 0.63
MVGRL	70.02 $\pm$ 0.25	42.18 $\pm$ 0.29	**90.07 $\pm$ 0.36**	80.86 $\pm$ 0.63	OOM
GGD	58.04 $\pm$ 0.40	43.15 $\pm$ 0.34	78.15 $\pm$ 0.48	76.43 $\pm$ 0.63	74.63 $\pm$ 0.66
GMI	32.33 $\pm$ 0.27	40.98 $\pm$ 0.30	72.38 $\pm$ 0.63	79.89 $\pm$ 0.62	OOM
CCA-SSG	42.82 $\pm$ 0.24	41.23 $\pm$ 0.25	72.42 $\pm$ 0.60	75.46 $\pm$ 0.75	74.64 $\pm$ 0.57
BGRL	39.34 $\pm$ 0.32	41.17 $\pm$ 0.25	72.82 $\pm$ 0.60	79.73 $\pm$ 0.61	72.27 $\pm$ 0.55
GBT	45.96 $\pm$ 0.34	43.58 $\pm$ 0.28	72.39 $\pm$ 0.56	75.74 $\pm$ 0.78	**75.98 $\pm$ 0.88**
GRACE	59.57 $\pm$ 0.39	43.79 $\pm$ 0.28	68.10 $\pm$ 0.70	76.31 $\pm$ 0.71	74.34 $\pm$ 0.71
GCA	59.77 $\pm$ 0.40	42.57 $\pm$ 0.17	68.11 $\pm$ 0.66	77.26 $\pm$ 0.61	75.09 $\pm$ 0.57
GraphCL	29.92 $\pm$ 0.30	37.81 $\pm$ 0.14	82.15 $\pm$ 0.46	76.88 $\pm$ 0.60	60.51 $\pm$ 1.45
GREET	72.68 $\pm$ 0.31	41.19 $\pm$ 0.25	82.71 $\pm$ 0.51	80.60 $\pm$ 0.56	OOM
PolyGCL	**72.97 $\pm$ 0.25**	**44.29 $\pm$ 0.43**	86.11 $\pm$ 0.43	**83.73 $\pm$ 0.53**	75.33 $\pm$ 0.67

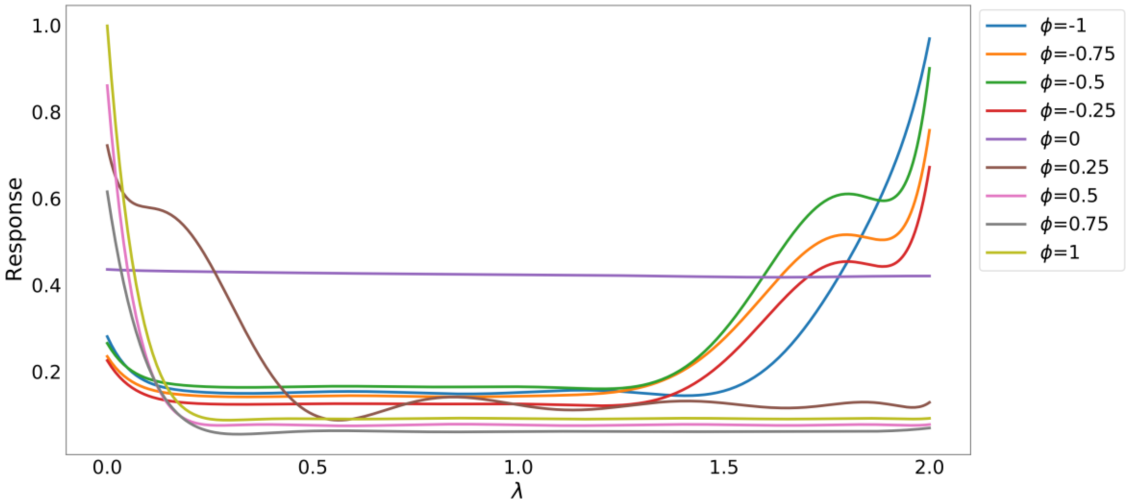
Results:

Baselines: homophily assumption.

PolyGCL: works in all settings.

Experiments





	arXiv-year
DGI	40.60 $\pm$ 0.21
GGD	40.86 $\pm$ 0.22
MVGRL	-
BGRL	OOM
GBT	41.90 $\pm$ 0.26
CCA-SSG	40.76 $\pm$ 0.25
GRACE	OOM
POLYGCL	**43.07 $\pm$ 0.23**

Linear coefficients: $\alpha + \beta = 1$

$\phi < 0$: low-pass information accounts for less.

Normalized learned filters

$\phi < 0$: increasing fuctions;
$\phi = 0$: all-pass property.

Time complexity: $O(KE + N)$

PolyGCL: SOTA performance with satisfactory efficiency.

Conclusion

- PolyGCL: introduces the superior properties of polynomial filters into Graph Contrastive Learning.

- We revisit GCL from the spectral view and theoretically prove the necessity of the high-pass information in heterophilic settings.

- PolyGCL achieves SOTA performance on datasets across homophily without introducing extra complexity.

Thank you!