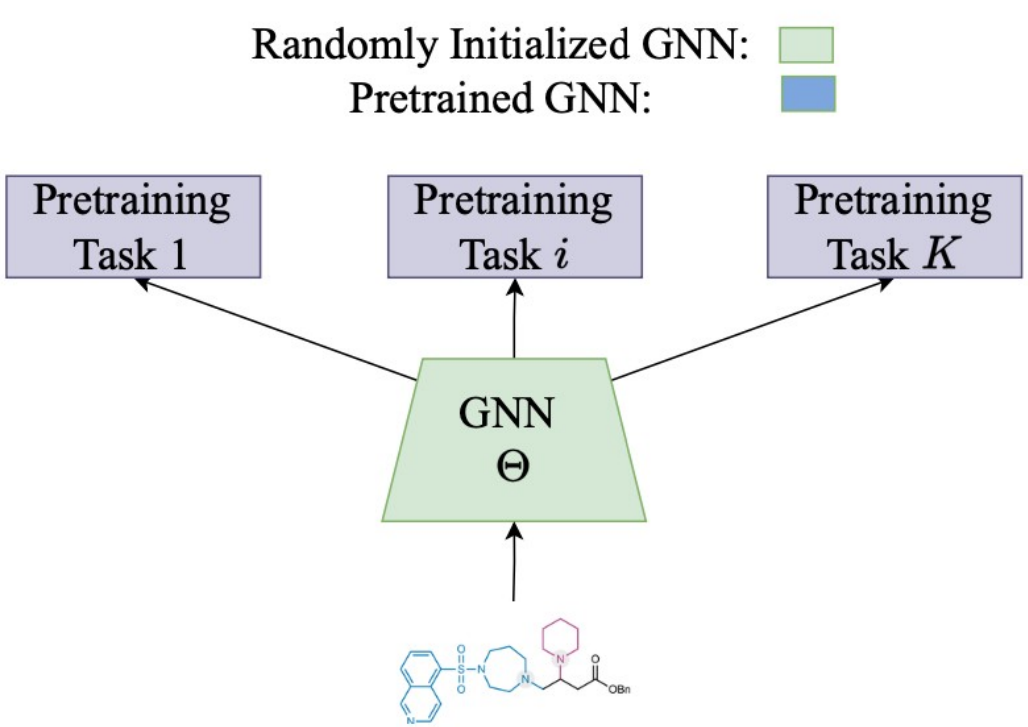
Enhancing Molecular Property Prediction with Auxiliary Learning and Task-Specific Adaptation

Vishal Dey
The Ohio State University, Columbus
dey.78@osu.edu

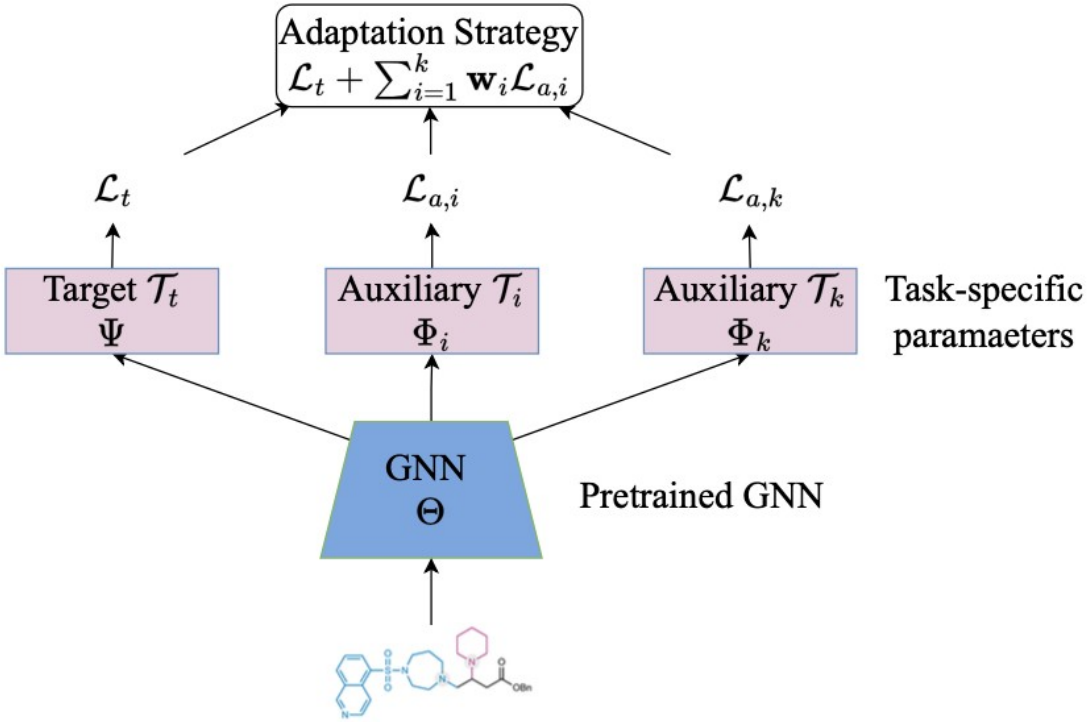
Xia Ning
The Ohio State University, Columbus
ning.104@osu.edu

Introduction –

- Problems with traditional pre-training fine-tuning approach,
- Their contribution.



(a) Pretraining Stage



(b) Adaptation Stage

Methods –

- Transfer knowledge to off-the-shelf GNNs using existing SSL objectives!
- masked atom prediction (AM), edge prediction (EP), context prediction (CP), graph infomax (IG), and motif prediction (MP),
- Formally,

$$\min_{\Theta,\Psi,\Phi_{i\in\{1..k\}}} \mathcal{L}_t + \sum_{i=1}^{k} \mathbf{w}_i \mathcal{L}_{a,i},$$

- And then,

$$\Theta^{(t+1)} := \Theta^{(t)} - \alpha \left(\mathbf{g}_t + \sum\nolimits_{i=1}^{k} \mathbf{w}_i \mathbf{g}_{a,i} \right)$$

Methods –

- Gradient Cosine Similarity –

$$\Theta^{(t+1)} := \Theta^{(t)} - \alpha \left(\mathbf{g}_t + \sum_{i=1}^{k} \max\left(0, \cos\left(\mathbf{g}_t, \mathbf{g}_{a,i}\right)\right) \mathbf{g}_{a,i} \right)$$

- Gradient Scaling –

$$\Theta^{(t+1)} := \Theta^{(t)} - \alpha \left(\mathbf{g}_t + \sum_{i=1}^{k} \frac{||\mathbf{g}_t||}{||\mathbf{g}_{a,i}||} \mathbf{g}_{a,i} \right)$$

- Bi-level optimizaztion! (also learn ***w***).

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \mathcal{L}_t{}^{(\mathcal{A})}(\Theta^*(\mathbf{w})), \qquad \text{s.t.} \qquad \Theta^*(\mathbf{w}) = \arg\min_{\Theta} \mathcal{L}_f(\Theta, \mathbf{w})$$

Methods –

- Upon simplification,

$$\nabla_{\mathbf{w}} \mathcal{L}_t^{(\mathcal{A})}(\Theta^*(\mathbf{w})) = \nabla_{\Theta} \mathcal{L}_t^{(\mathcal{A})} \cdot \nabla_{\mathbf{w}} \Theta^*(\mathbf{w}) = -\nabla_{\Theta} \mathcal{L}_t^{(\mathcal{A})} \cdot (\nabla_{\Theta}^2 \mathcal{L}_f)^{-1} \cdot \nabla_{\mathbf{w}} \nabla_{\Theta} \mathcal{L}_f$$

Algorithm 1 Learning Task Weights with BLO

Input: N, r, α
Initialize $\mathbf{w}$ with $1/k$, Θ from pretrained GNN, Ψ and Φ with default Xavier initializer
for *epoch* from 1 to N **do**
    Compute $\mathcal{L}_f = \mathcal{L}_t + \sum_{i=1}^{k} \mathbf{w}_i \mathcal{L}_{a,i}$
    $\Theta \leftarrow \Theta - \alpha \nabla_{\Theta} \mathcal{L}_f$, $\Phi \leftarrow \Phi - \alpha \nabla_{\Phi} \mathcal{L}_a$, $\Psi \leftarrow \Psi - \alpha \nabla_{\Psi} \mathcal{L}_t$
    if $epoch \% r == 0$ **then**
        $\mathbf{w} \leftarrow \mathbf{w} - \nabla_{\mathbf{w}} \mathcal{L}_t^{(\mathcal{A})}(\Theta(\mathbf{w}))$ ▷ Algorithm 2
    end if
end for
Return Θ, $\mathbf{w}$

Algorithm 2 Computing $\nabla_{\mathbf{w}} \mathcal{L}_t^{(\mathcal{A})}(\Theta(\mathbf{w}))$

Input: $\mathcal{L}_f$, $\mathcal{L}_t^{(\mathcal{A})}$, current $\mathbf{w}$, Θ from Algorithm 1, M, β
Initialize $p = q = \nabla_{\Theta} \mathcal{L}_t^{(\mathcal{A})}|_{(\mathbf{w},\Theta)}$
▷ Hessian inverse approximation
for j from 1 to M **do**
    $p = p - \beta p \nabla_{\Theta}^2 \mathcal{L}_f$
    $q = q + p$
end for
Return $-q \nabla_{\mathbf{w}} \nabla_{\Theta} \mathcal{L}_f|_{(\mathbf{w},\Theta)}$

Experiments –

- Wider set of auxiliary tasks {AM,CP,EP,IG,MP},

Method	SIDER	ClinTox	BACE	BBBP	Tox21	ToxCast	HIV	MUV
FT	**62.05** (0.40)	**71.36** (0.70)	82.68 (1.14)	67.42 (0.66)	**77.55** (0.07)	**66.18** (0.14)	78.60 (0.80)	81.25 (2.17)
GTOT	61.98 (0.12)	**71.48** (0.77)	82.15 (2.20)	**71.34** (0.76)	**77.85** (0.52)	64.90 (0.72)	**80.03** (0.19)	82.38 (1.52)
MTL	56.39 (0.29)	55.69 (2.65)	77.11 (5.11)	64.89 (0.30)	74.31 (0.24)	64.32 (0.25)	76.79 (0.59)	80.81 (0.74)
GCS	59.52 (1.08)	63.10 (2.46)	**85.49** (0.40)	71.23 (0.26)	74.84 (0.36)	65.96 (0.14)	76.69 (0.20)	75.20 (2.42)
GNS	**62.14** (0.37)	68.07 (1.58)	**84.76** (0.30)	**71.60** (0.88)	76.44 (0.16)	**66.24** (0.11)	77.87 (0.13)	**83.77** (1.25)
BLO	58.09 (0.50)	65.33 (2.23)	84.28 (3.51)	69.18 (0.46)	76.04 (0.46)	65.93 (0.30)	**78.62** (0.48)	**82.74** (0.32)

- Smaller set of auxiliary tasks {AM,IG,MP},

Method	SIDER	ClinTox	BACE	BBBP	Tox21	ToxCast	HIV	MUV
FT	62.05 (0.40)	**71.36** (0.70)	82.68 (1.14)	67.42 (0.66)	77.55 (0.07)	66.18 (0.14)	78.60 (0.80)	81.25 (2.17)
GTOT	61.98 (0.12)	**71.48** (0.77)	82.15 (2.20)	71.34 (0.76)	**77.85** (0.52)	64.90 (0.72)	**80.03** (0.19)	82.19 (1.52)
MTL	58.80 (0.38)	60.82 (1.58)	83.67 (0.42)	**72.50** (0.25)	75.22 (0.28)	65.05 (0.27)	78.18 (1.31)	81.26 (2.33)
GCS	**63.10** (0.17)	64.84 (2.21)	84.94 (0.42)	68.53 (5.93)	77.41 (0.13)	**66.63** (0.14)	79.23 (0.15)	**82.56** (1.42)
GNS	62.46 (0.19)	65.44 (1.01)	**85.02** (0.23)	72.38 (0.26)	76.62 (0.30)	65.96 (0.14)	79.18 (0.28)	**82.22** (0.11)
BLO	**63.46** (0.33)	63.06 (0.36)	**85.22** (0.36)	**73.12** (0.50)	**77.94** (0.35)	**66.64** (0.16)	**80.08** (0.44)	81.83 (0.91)

Experiments (from Appendix) –

- Wider set of auxiliary tasks {AM,CP,EP,IG,MP} + GraphMVP-C,

Method	SIDER	ClinTox	BACE	BBBP	Tox21	ToxCast	HIV	MUV
FT	**62.73**(0.40)	68.52(7.19)	**80.07**(0.98)	**70.92**(0.71)	73.71(0.61)	**64.13**(0.68)	**75.01**(1.25)	72.45(2.87)
GTOT	**60.83**(0.55)	**73.17**(5.63)	78.36(2.30)	62.09(1.70)	**74.30**(0.59)	**64.67**(0.11)	74.78(1.94)	72.63(0.83)
MTL	51.31(1.47)	52.84(5.23)	57.20(4.83)	51.08(2.02)	60.53(1.78)	55.02(0.45)	72.12(3.15)	72.26(7.09)
GCS	57.97(0.54)	**69.20**(1.59)	78.00(3.22)	68.27(1.06)	74.14(0.29)	60.87(0.63)	**74.80**(1.19)	73.49(0.21)
GNS	58.37(2.71)	65.91(7.41)	**80.98**(0.61)	**71.82**(0.41)	**74.87**(0.35)	63.97(0.16)	72.88(1.15)	**76.20**(2.28)
BLO	58.89(1.21)	67.26(1.60)	74.88(5.21)	68.18(0.64)	73.63(0.29)	61.93(0.87)	74.29(1.74)	**74.69**(1.34)

- Smaller set of auxiliary tasks {IG,MP} + GraphMVP-C,

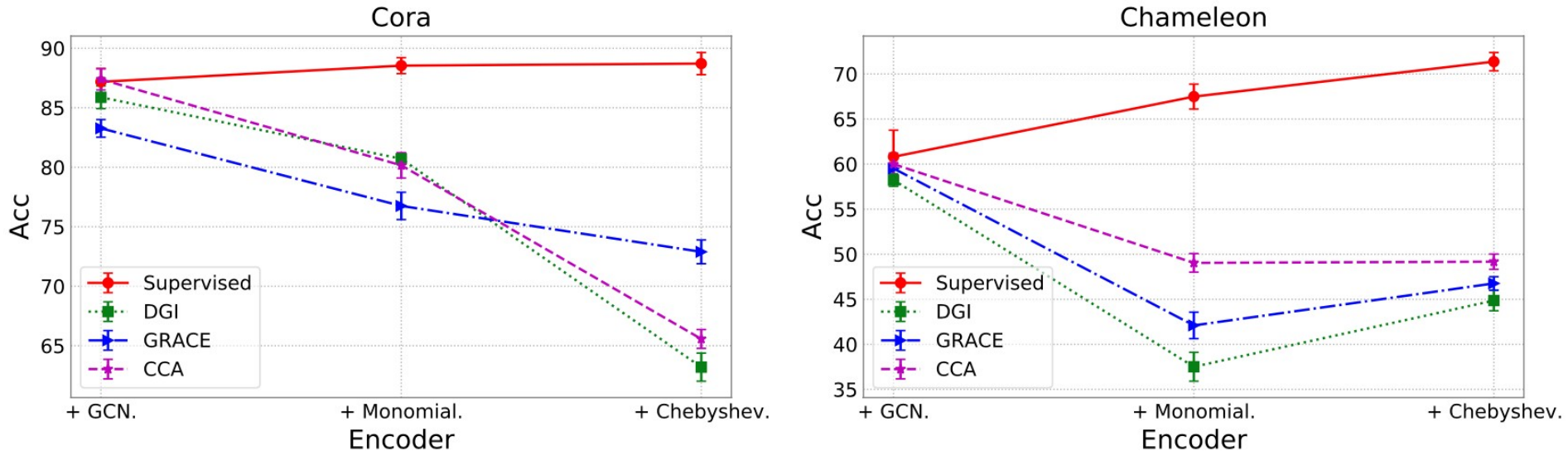
Method	SIDER	ClinTox	BACE	BBBP	Tox21	ToxCast	HIV	MUV
FT	**62.73**(0.40)	68.52(7.19)	80.07(0.98)	70.92(0.71)	73.71(0.61)	**64.13**(0.68)	75.01(1.25)	72.45(2.87)
GTOT	60.83(0.55)	**73.17**(5.63)	78.36(2.30)	62.09(1.70)	74.30(0.59)	**64.67**(0.11)	74.88(1.94)	72.63(0.83)
MTL	60.29(1.87)	61.76(4.14)	78.41(0.50)	72.05(1.67)	73.75(0.40)	61.59(1.07)	75.19(0.53)	**74.18**(4.59)
GCS	**63.34**(0.34)	**71.43**(2.53)	**80.48**(1.73)	71.04(2.33)	**74.73**(0.16)	63.28(1.49)	**77.06**(0.86)	72.42(2.43)
GNS	61.62(1.44)	67.82(2.57)	**81.02**(1.42)	**72.26**(0.45)	**75.12**(0.43)	62.98(0.30)	**76.61**(1.50)	**76.94**(0.18)
BLO	60.94(0.72)	68.54(0.53)	79.92(0.57)	**72.71**(2.01)	73.86(0.63)	62.29(0.60)	74.75(0.76)	73.94(1.27)

PolyGCL: Graph Contrastive Learning via Learnable Spectral Polynomial Filters

Anonymous authors
Paper under double-blind review

Introduction –

- Current SSL/GCL techniques and homo-philicity.,
- Spectral GNNs and heterophilic networks,
- *“How can we effectively introduce the properties of spectral polynomial filters into GCL to ensure the expressiveness on both homophilic and heterophilic settings?”*
- Problems with direct adaptation,



Contributions –

- Introduce polynomial filters into GCL,
- Prove that High-pass filters are necessary in heterophilic settings,
- SOTA performance.

Related Work –

- GCL –
  - Augmentation based,
  - Augmentation free.

- Spectral GNNs.

PolyGCL –

- Directly using polynomial filters + GCL doesn't work!
- Decouple low-pass and high-pass filters to fit only one.

$$\sum_{k=0}^{K} w_k T_k(\hat{\mathbf{L}})\mathbf{X}, \text{ where } \hat{\mathbf{L}} = 2\mathbf{L}/\lambda_{max} - \mathbf{I},$$

$$w_k = \frac{2}{K+1}\sum_{j=0}^{K} \gamma_j T_k(x_j),$$

$$x_j = \cos\left(\frac{j+1/2}{K+1}\pi\right), j = 0, \ldots, K$$

$$\gamma_i^H = \sum_{j=0}^{i} \gamma_j, \quad \gamma_i^L = \gamma_0 - \sum_{j=1}^{i} \gamma_j, i = 1, ..., K,$$

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

PolyGCL –

- Optimization Objective,

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

- Training Algorithm.

Algorithm 1: Training Algorithm for PolyGCL

Input: Node features $\mathbf{X}$, input adjacency $\mathbf{A}$, initialized encoders $\mathcal{E}$, initialized coefficients α, β, maximum iterations T, polynomial order K.

for *epoch* $= 0, 1, \ldots, T$ **do**
- $\tilde{\mathbf{X}} \leftarrow \text{shuffle}(\mathbf{X})$; % corruption
- $\gamma_0^L = \gamma_0^H = \gamma_0$ % initialize γ_0^L, γ_0^H with γ_0 in $\mathcal{E}$
- **for** $i = 1, \ldots, K$ **do**
  - $\gamma_i^H = \sum_{j=0}^{i} \text{ReLU}(\gamma_j)$; % high-pass encoder
  - $\gamma_i^L = \gamma_0 - \sum_{j=1}^{i} \text{ReLU}(\gamma_j)$;% low-pass encoder
- Obtain $\mathcal{E}^L$ and $\mathcal{E}^H$ via γ_i^L and γ_i^H respectively shown in equation 3
- $\mathbf{Z}_L \leftarrow \mathcal{E}^L(\mathbf{X}, \mathbf{A}), \mathbf{Z}_H \leftarrow \mathcal{E}^H(\mathbf{X}, \mathbf{A})$; % positive embeddings
- $\tilde{\mathbf{Z}}_L \leftarrow \mathcal{E}^L(\tilde{\mathbf{X}}, \mathbf{A}), \tilde{\mathbf{Z}}_H \leftarrow \mathcal{E}^H(\tilde{\mathbf{X}}, \mathbf{A})$; % negative embeddings
- $\mathbf{Z} = \alpha\mathbf{Z}_L + \beta\mathbf{Z}_H$; % linear combination
- Compute loss via equation 4 and update parameters in $\mathcal{E}^L$ and $\mathcal{E}^H$;

Output: $\mathcal{E}^L$ and $\mathcal{E}^H$ with frozen parameters; learned coefficients α, β.

PolyGCL –

- A lot of theoretical analysis!

Experiments –

- Synthetic datasets

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

Experiments –

- Real-world datasets.

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

Experiments –

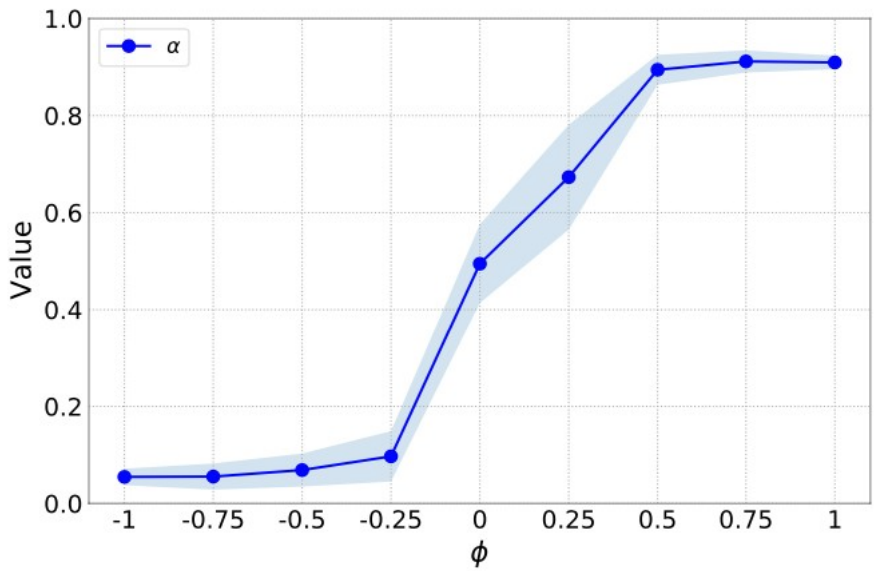
- Real-world datasets (specfically heterophilic).

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

Ablation Studies –

- α and β analysis with ϕ + regularization.

$$\boldsymbol{\alpha} = \bar{\mathbf{U}}^\top \Delta \boldsymbol{y} \text{ and } \boldsymbol{\beta} = \bar{\mathbf{U}}^\top \mathbf{X}$$



	ArXiv-year
DGI	40.60 ± 0.21
GGD	40.86 ± 0.22
MVGRL	-
BGRL	OOM
GBT	41.90 ± 0.26
CCA-SSG	40.76 ± 0.25
GRACE	OOM
PolyGCL	**43.07 ± 0.23**