ProGCL: Rethinking Hard Negative Mining in Graph Contrastive Learning

Jun Xia[1,2], Lirong Wu[1,2], Ge Wang[1,2], Jintao Chen[1,2] and Stan Z. Li[1,2]

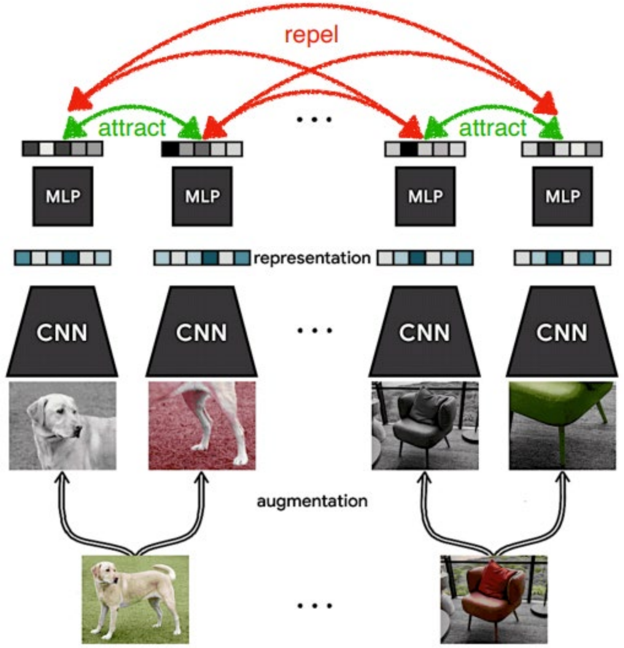
[1]Westlake University, [2]Westlake Institute for Advanced Study

Outline

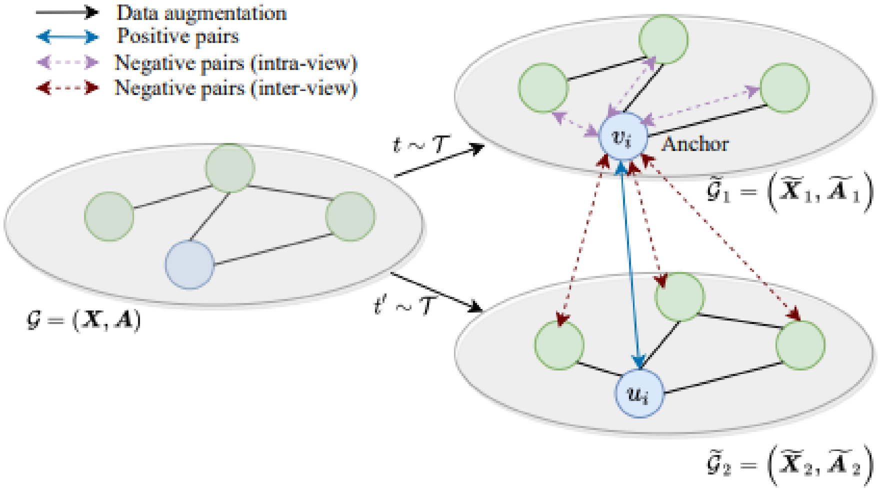
- Preamble
- Analysis
- ProGCL
- Experiments
- Concluding Remarks

Preamble

- Contrastive Learning (CL) & Graph Contrastive Learning (GCL)



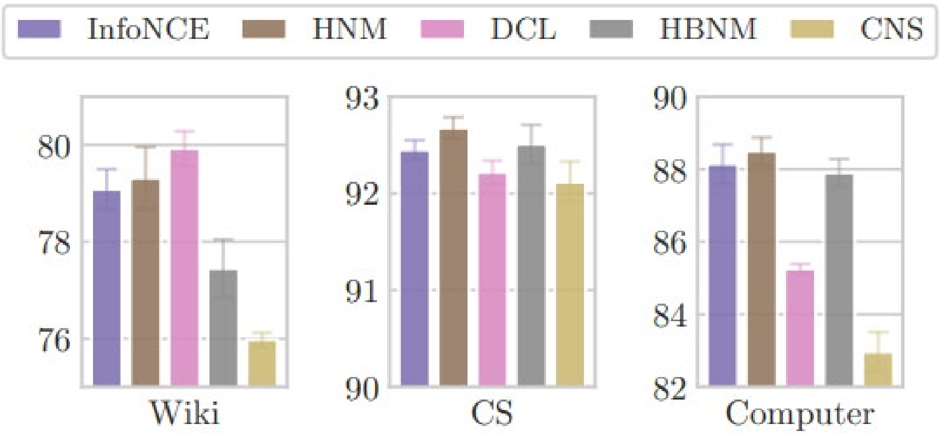
Ref 1. SimCLR, ICML'20



Ref 2. GCA, WWW'21

Analysis

- **Hard Negative Mining Methods Fail in Graph Contrastive Learning**



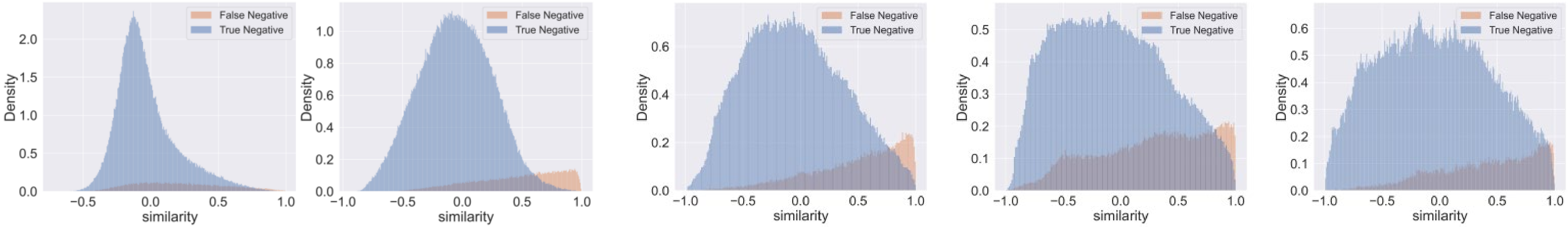
Ref 3. NeurIPS' 21 (Benchmarks Track)

Methods/Datasets	Amazon-Photo	Amazon-Computers	Coauthor-CS
GCA	92.55	87.82	92.40
+DCL	91.02 (↓ 1.53)	86.58 (↓ 1.24)	92.36 (↓ 0.04)
+HCL	91.48 (↓ 1.07)	87.21 (↓ 0.61)	93.06 (↑ 0.66)
+MoCHi	92.36 (↓ 0.19)	87.68 (↑ 0.14)	92.58 (↑ 0.18)
+Ring	91.33 (↓ 1.22)	84.18 (↓ 3.64)	92.48 (↓ 0.08)
+ProGCL-mix	**93.64** (↑ 1.09)	**89.55** (↑ 1.73)	**93.67** (↑ 1.27)

Table 1. Our Results

Analysis

- **Why above phenomena would occur ?**



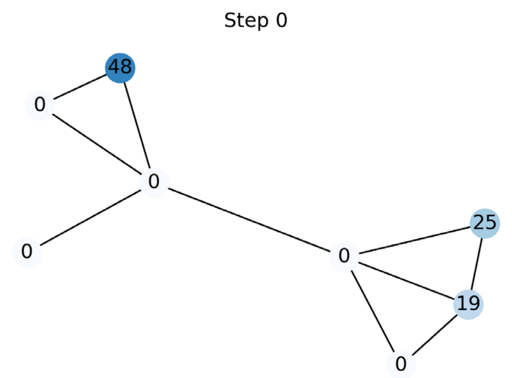
(a) CIFAR-10 (Image) (b) Coauthor-CS (Graph) (a) Amazon-photo (Graph) (b) Amazon-computers (Graph) (c) WikiCS (Graph)

Fig 2. **Similarity histograms of negatives.**

Unlike CL, most negatives with larger similarities to the anchor are false ones in GCL.

Analysis

- Experimental & Theoretical Analysis



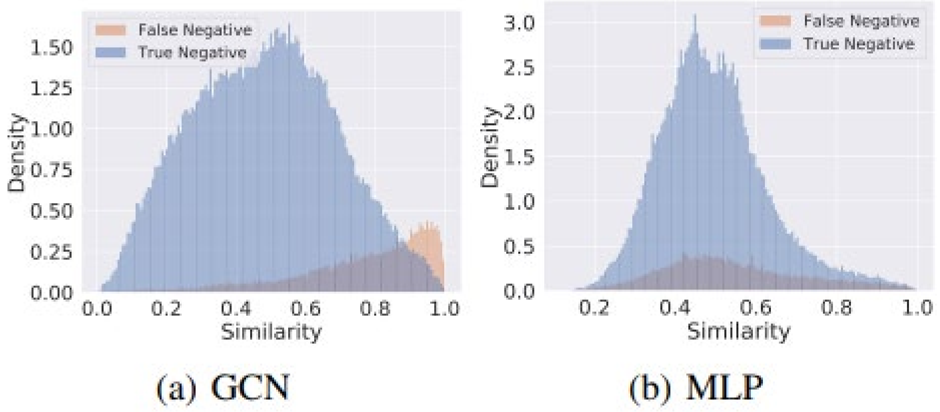


Fig 3. Semantic Diagram of Messaging-Passing.

Fig 4. GCN (with MP) vs. MLP (w/o MP).

Delving into the Role of Message-Passing in GCL

ProGCL

- How to eliminate the bias?

  a. Fit the negatives' distribution with Beta Mixture Model (BMM)

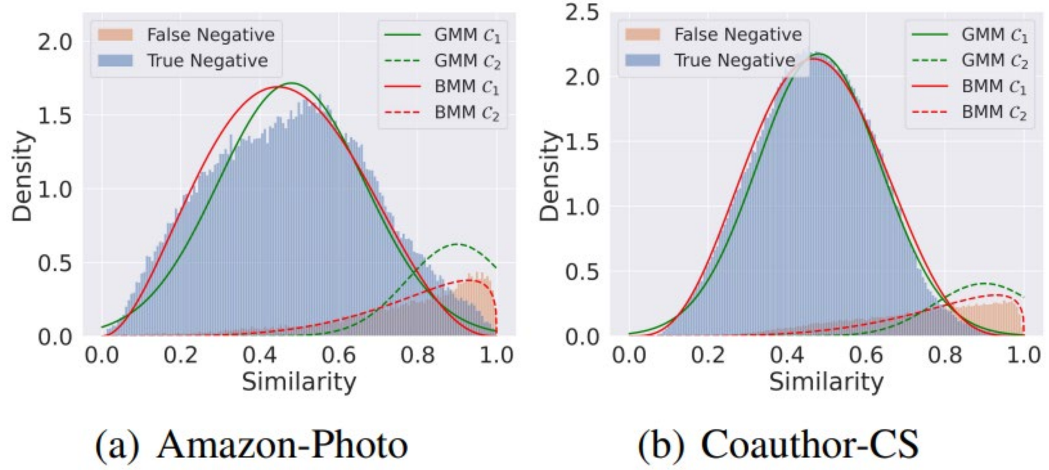


Fig 5. Empirical distribution v.s. estimated distribution.

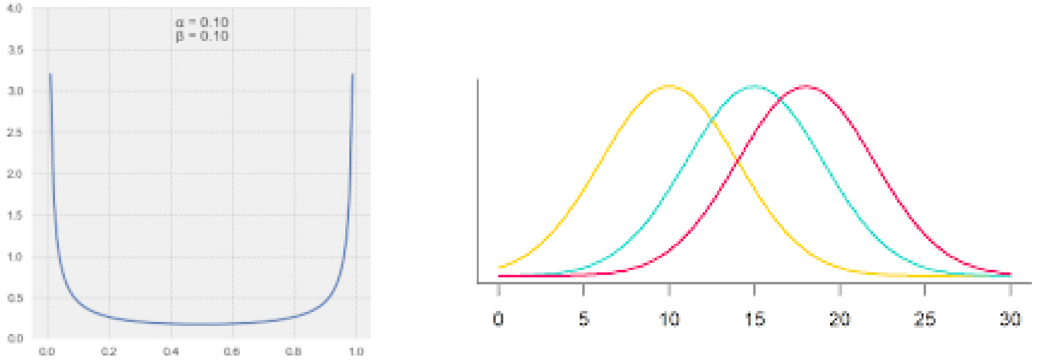


Fig 6. Beta distribution v.s. Normal distribution.

Beta distribution $$p(s \mid \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} s^{\alpha-1}(1-s)^{\beta-1}$$

Normal distribution $$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

ProGCL

- **Expectation-Maximization for Beta Mixture Distribution**

a. E-Step

$$p(s) = \sum_{c=1}^{C} \lambda_c p(s \mid \alpha_c, \beta_c) \quad p(c \mid s) = \frac{\lambda_c p\left(s \mid \alpha_c, \beta_c\right)}{\sum_{j=1}^{C} \lambda_j p\left(s \mid \alpha_j, \beta_j\right)} \quad \bar{s}_c = \frac{\sum_{i=1}^{M} p(c \mid s_i) s_i}{\sum_{i=1}^{M} p(c \mid s_i)}, \quad v_c^2 = \frac{\sum_{i=1}^{M} p(c \mid s_i)\left(s_i - \bar{s}_c\right)^2}{\sum_{i=1}^{M} p(c \mid s_i)}$$

b. M-Step

$$\alpha_c = \bar{s}_c \left(\frac{\bar{s}_c \left(1 - \bar{s}_c\right)}{v_c^2} - 1 \right), \quad \beta_c = \frac{\alpha_c \left(1 - \bar{s}_c\right)}{\bar{s}_c} \qquad \lambda_c = \frac{1}{M} \sum_{i=1}^{M} p(c \mid s_i) \qquad p\left(c \mid s\right) = \frac{p(c) p\left(s \mid \alpha_c, \beta_c\right)}{p\left(s\right)}$$

ProGCL

- Scheme 1: ProGCL-weight

New Measure: $$w(i,k)=\frac{p\left(c_t \mid s_{ik}\right) s_{ik}}{\frac{1}{N-1}\sum_{j\neq i}\left[p\left(c_t \mid s_{ij}\right) s_{ij}\right]}$$

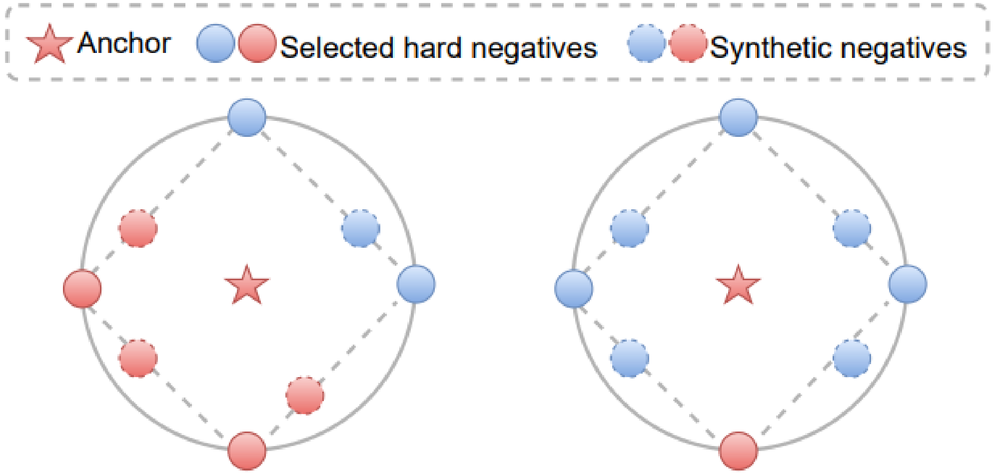
$$\ell_w\left(\boldsymbol{u}_i,\boldsymbol{v}_i\right)=$$

$$\log \frac{e^{\frac{\theta\left(\boldsymbol{u}_i,\boldsymbol{v}_i\right)}{\tau}}}{\underbrace{e^{\frac{\theta\left(\boldsymbol{u}_i,\boldsymbol{v}_i\right)}{\tau}}}_{\text{positive pair}}+\underbrace{\sum_{k\neq i} w(i,k) e^{\frac{\theta\left(\boldsymbol{u}_i,\boldsymbol{v}_k\right)}{\tau}}}_{\text{inter-view negative pairs}}+\underbrace{\sum_{k\neq i} w(i,k) e^{\frac{\theta\left(\boldsymbol{u}_i,\boldsymbol{u}_k\right)}{\tau}}}_{\text{intra-view negative pairs}}}$$

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

ProGCL

- Scheme 2: ProGCL-mix



(a) MoCHi (b) ProGCL-mix

Fig 7. MoChi (NeurIPS' 21) v.s. ProGCL-mix.

$$\tilde{\boldsymbol{u}}_k = \alpha_k \boldsymbol{v}_p + (1-\alpha_k)\,\boldsymbol{v}_q,$$

$$\alpha_k = \frac{p(c_t \mid s_{ip})}{p(c_t \mid s_{ip}) + p(c_t \mid s_{iq})}$$

$$\ell_m(\boldsymbol{u}_i, \boldsymbol{v}_i) = \log \frac{e^{\frac{\theta(\boldsymbol{u}_i,\boldsymbol{v}_i)}{\tau}}}{\underbrace{e^{\frac{\theta(\boldsymbol{u}_i,\boldsymbol{v}_i)}{\tau}}}_{\text{positive pair}} + \underbrace{\sum_{k\neq i} e^{\frac{\theta(\boldsymbol{u}_i,\boldsymbol{v}_k)}{\tau}}}_{\text{inter-view negative pairs}} + \underbrace{\sum_{k\neq i} e^{\frac{\theta(\boldsymbol{u}_i,\boldsymbol{u}_k)}{\tau}}}_{\text{intra-view negative pairs}} + \underbrace{\sum_{k=1}^{m} e^{\frac{\theta(\boldsymbol{u}_i,\tilde{\boldsymbol{u}}_k)}{\tau}}}_{\text{synthetic negative pairs}}}$$

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

Experiments

- Results in Transductive Setting

Method	Available Data	Amazon-Photo	Amazon-Computers	Coauthor-CS	Wiki-CS
Raw features	$\boldsymbol{X}$	78.53 ± 0.00	73.81 ± 0.00	90.37 ± 0.00	71.98 ± 0.00
node2vec	$\boldsymbol{A}$	89.67 ± 0.12	84.39 ± 0.08	85.08 ± 0.03	71.79 ± 0.05
DeepWalk	$\boldsymbol{A}$	89.44 ± 0.11	85.68 ± 0.06	84.61 ± 0.22	74.35 ± 0.06
DeepWalk + features	$\boldsymbol{X}, \boldsymbol{A}$	90.05 ± 0.08	86.28 ± 0.07	87.70 ± 0.04	77.21 ± 0.03
GAE	$\boldsymbol{X}, \boldsymbol{A}$	91.62 ± 0.13	85.27 ± 0.19	90.01 ± 0.17	70.15 ± 0.01
VGAE	$\boldsymbol{X}, \boldsymbol{A}$	92.20 ± 0.11	86.37 ± 0.21	92.11 ± 0.09	75.35 ± 0.14
DGI	$\boldsymbol{X}, \boldsymbol{A}$	91.61 ± 0.22	83.95 ± 0.47	92.15 ± 0.63	75.35 ± 0.14
GMI	$\boldsymbol{X}, \boldsymbol{A}$	90.68 ± 0.17	82.21 ± 0.31	**OOM**	74.85 ± 0.08
MVGRL*	$\boldsymbol{X}, \boldsymbol{A}$	92.08 ± 0.01	87.45 ± 0.21	92.18 ± 0.15	77.43 ± 0.17
BGRL*	$\boldsymbol{X}, \boldsymbol{A}$	92.95 ± 0.07	87.89 ± 0.10	92.72 ± 0.03	78.41 ± 0.09
MERIT*	$\boldsymbol{X}, \boldsymbol{A}$	92.53 ± 0.15	88.01 ± 0.12	92.51 ± 0.14	78.35 ± 0.05
GCA*	$\boldsymbol{X}, \boldsymbol{A}$	92.55 ± 0.03	87.82 ± 0.11	92.40 ± 0.07	78.26 ± 0.06
ProGCL-weight	$\boldsymbol{X}, \boldsymbol{A}$	93.30 ± 0.09	89.28 ± 0.15	93.51 ± 0.06	$\mathbf{78.68 \pm 0.12}$
ProGCL-mix	$\boldsymbol{X}, \boldsymbol{A}$	$\mathbf{93.64 \pm 0.13}$	$\mathbf{89.55 \pm 0.16}$	$\mathbf{93.67 \pm 0.12}$	78.45 ± 0.04
Supervised GCN	$\boldsymbol{X}, \boldsymbol{A}, \boldsymbol{Y}$	92.42 ± 0.22	86.51 ± 0.54	$\underline{93.03 \pm 0.31}$	77.19 ± 0.12
Supervised GAT	$\boldsymbol{X}, \boldsymbol{A}, \boldsymbol{Y}$	$\underline{92.56 \pm 0.35}$	$\underline{86.93 \pm 0.29}$	92.31 ± 0.24	$\underline{77.65 \pm 0.11}$

Experiments

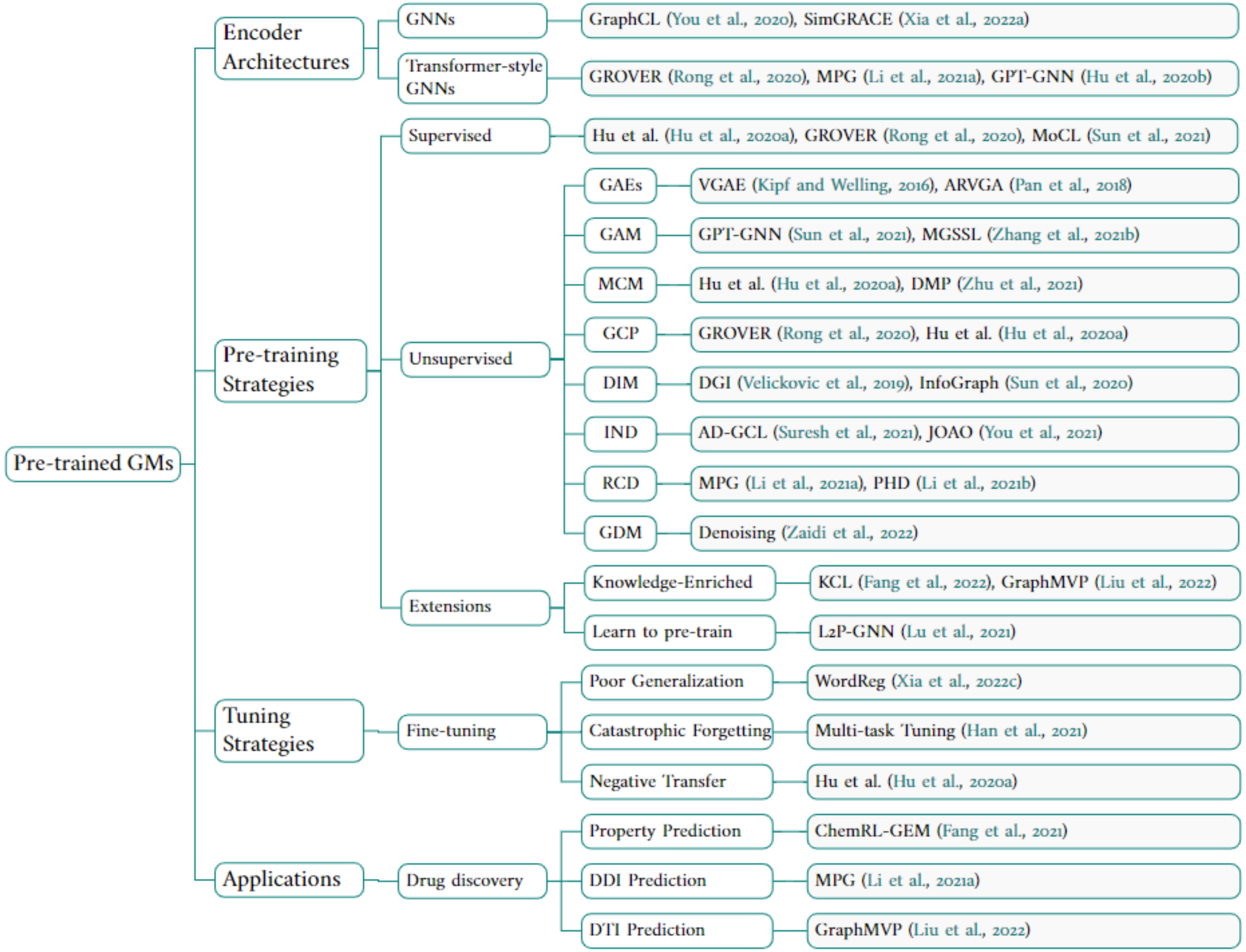
- Results in Inductive Setting

Method	Available Data	Flickr	Reddit
Raw features	$\boldsymbol{X}$	20.3	58.5
DeepWalk	$\boldsymbol{A}$	27.9	32.4
GraphSAGE	$\boldsymbol{X}, \boldsymbol{A}$	36.5	90.8
DGI	$\boldsymbol{X}, \boldsymbol{A}$	42.9±0.1	94.0±0.1
GMI	$\boldsymbol{X}, \boldsymbol{A}$	44.5±0.2	94.8±0.0
COLES-S^{2}GC	$\boldsymbol{X}, \boldsymbol{A}$	46.8±0.5	95.2±0.3
GRACE	$\boldsymbol{X}, \boldsymbol{A}$	48.0±0.1	94.2±0.0
ProGCL-weight	$\boldsymbol{X}, \boldsymbol{A}$	49.2±0.6	95.1±0.2
ProGCL-mix	$\boldsymbol{X}, \boldsymbol{A}$	**50.0±0.3**	**95.6±0.1**
Supervised FastGCN	$\boldsymbol{X}, \boldsymbol{A}, \boldsymbol{Y}$	48.1±0.5	89.5±1.2
Supervised GraphSAGE	$\boldsymbol{X}, \boldsymbol{A}, \boldsymbol{Y}$	50.1±1.3	92.1±1.1

	Validation	Test
MLP	57.65 ± 0.12	55.50 ± 0.23
node2vec	71.29 ± 0.13	70.07 ± 0.13
Random-Init	69.90 ± 0.11	68.94 ± 0.15
DGI	71.26 ± 0.11	70.34 ± 0.16
GRACE-Subsampling	72.61 ± 0.15	71.51 ± 0.11
BGRL	72.53 ± 0.09	71.64 ± 0.12
COLES-S^{2}GC	–	72.48 ± 0.25
ProGCL-weight	72.45 ± 0.21	72.18 ± 0.09
ProGCL-mix	$\mathbf{72.82 \pm 0.08}$	$\mathbf{72.56 \pm 0.20}$
Supervised GCN	73.00 ± 0.17	71.74 ± 0.29

Concluding Remarks

- **Pretrained Graph Models for Molecular Representations: Retrospect and Prospect**



Concluding Remarks

- Useful Resources

a. The first comprehensive survey of pre-training on molecular graphs.

✓ https://bit.ly/PGMs_survey

✓ Journal version is under review.

b. A curated list of must-read papers, open-source pre-trained models and pre-training datasets.

✓ https://bit.ly/PGM_resources

西湖大學 WESTLAKE UNIVERSITY

ICML International Conference On Machine Learning

Thank you!

Code

Paper

Homepage

xiajun@westlake.edu.cn

JunXia_Westlake