Omni-Granular Ego-Semantic Propagation for Self-Supervised Graph Representation Learning

Ling Yang, Shenda Hong

北京大學 PEKING UNIVERSITY

北京大学 健康医疗大数据国家研究院 NATIONAL INSTITUTE OF HEALTH DATA SCIENCE AT PEKING UNIVERSITY

Presenter : Ling Yang

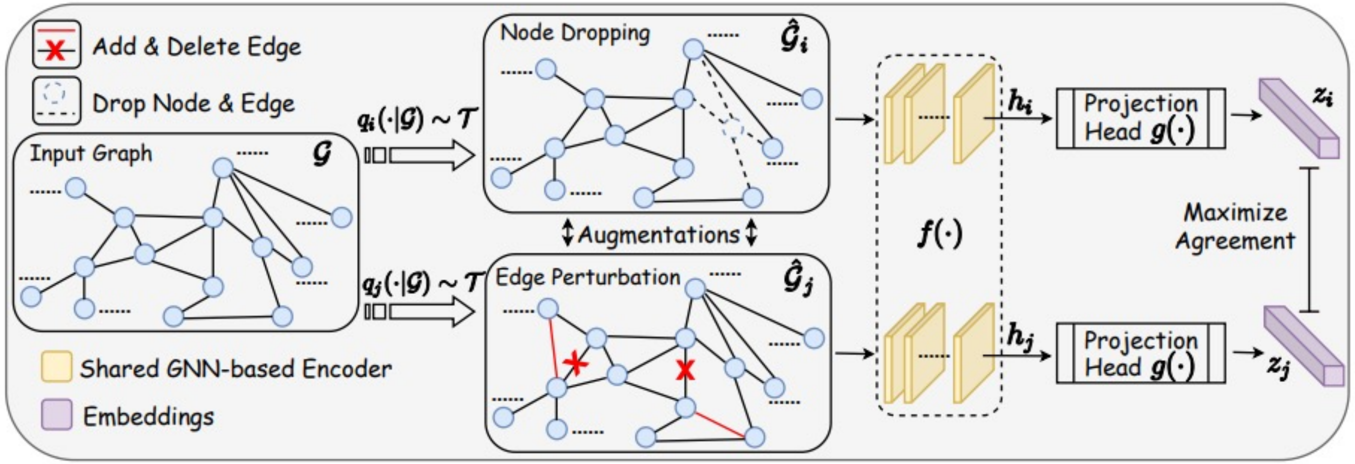
Contact : yangling0818@163.com

Contents

Related Work

- Self-Supervised/Unsupervised Graph Representation Learning

Contrastive

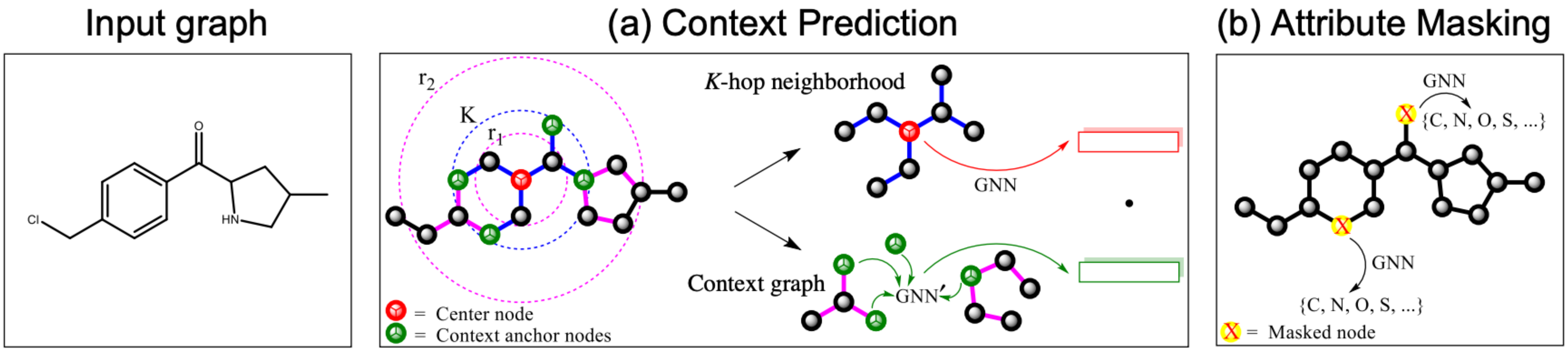


You Y, Chen T, Sui Y, et al. Graph contrastive learning with augmentations[J]. Advances in Neural Information Processing Systems, 2020, 33: 5812-5823.

Related Work

- Self-Supervised/Unsupervised Graph Representation Learning

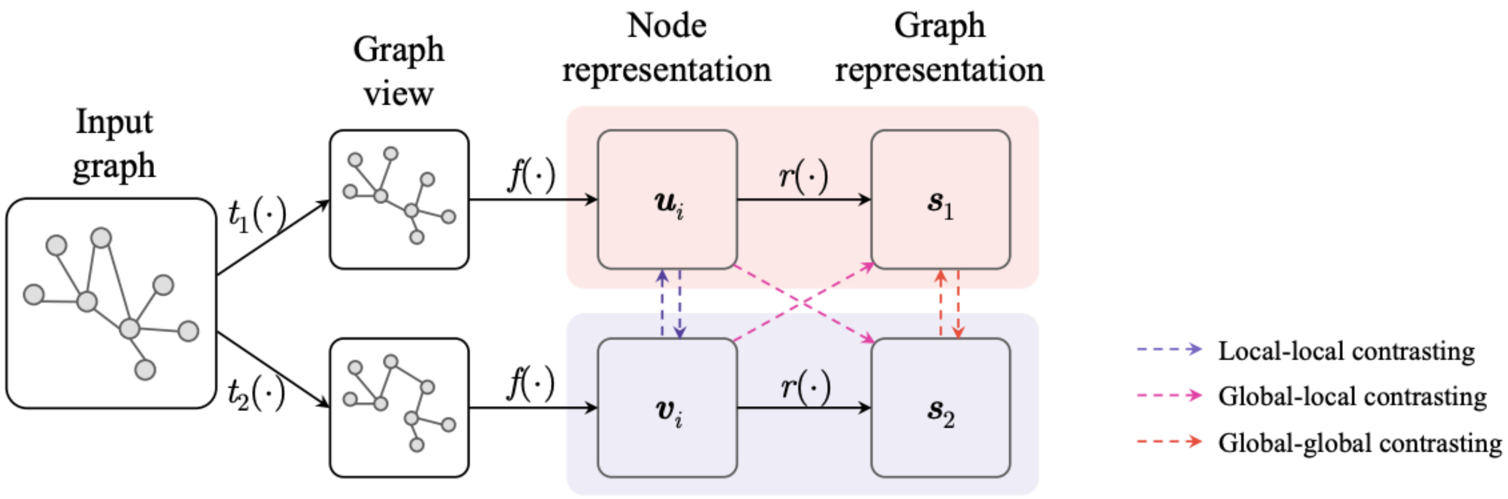
Predictive



Hu W, Liu B, Gomes J, et al. Strategies for Pre-training Graph Neural Networks[C]//International Conference on Learning Representations. 2019.

Related Work

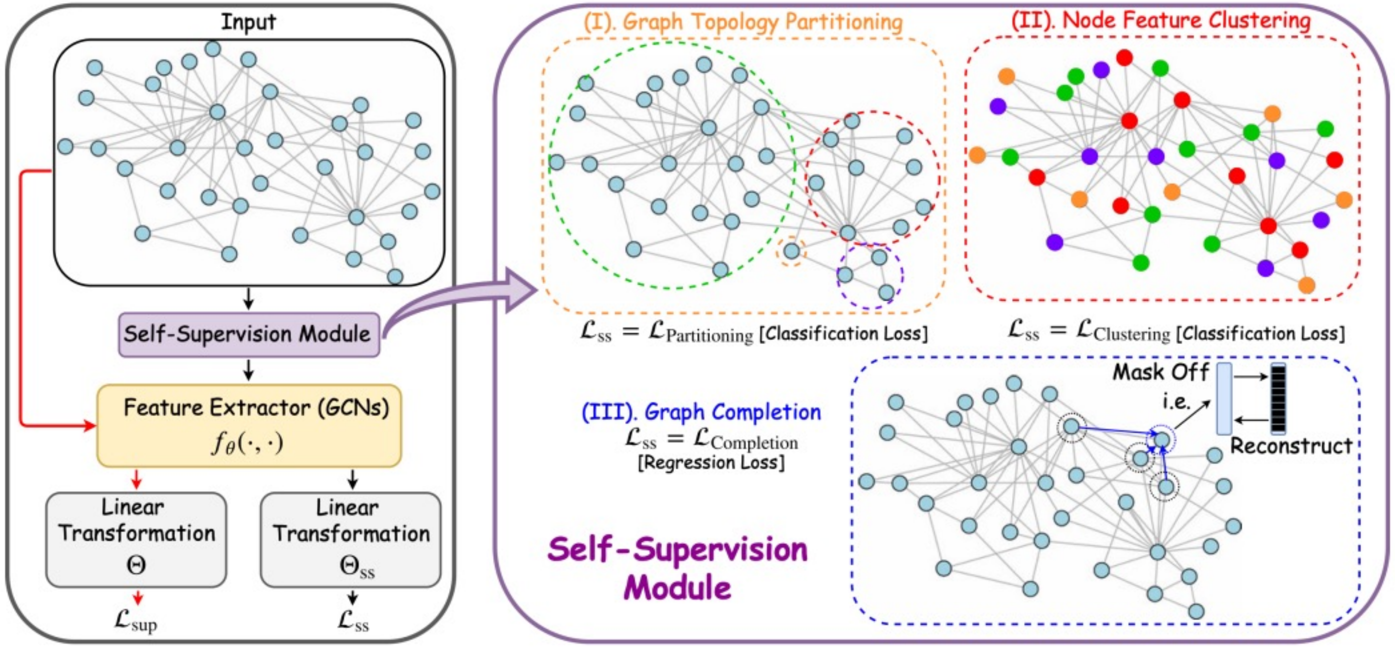
- Global-Level Self-Supervised Objectives



Zhu Y, Xu Y, Liu Q, et al. An empirical study of graph contrastive learning[J]. arXiv preprint arXiv:2109.01116, 2021.

Related Work

- Global-Level Self-Supervised Objectives

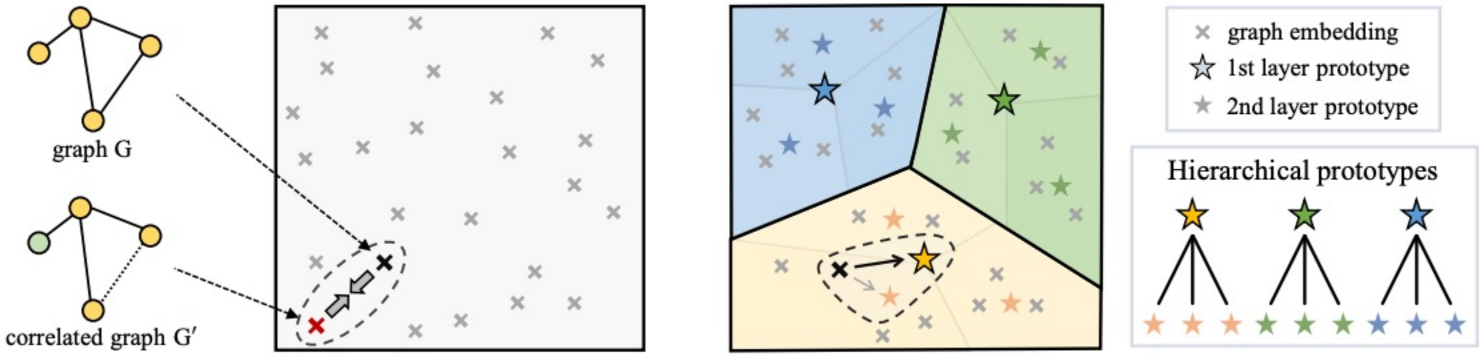


You Y, Chen T, Wang Z, et al. When does self-supervision help graph convolutional networks?[C]//international conference on machine learning. PMLR, 2020: 10871-10880.

PEKING UNIVERSITY

Related Work

- Global-Level Self-Supervised Objectives



(a) Local-instance structure (b) Global-semantic structure

Algorithm 1 Optimization Algorithm of GraphLoG.

Input: Unlabeled graph data set $\mathbf{G}$, the number of learning steps T.
Output: Pre-trained GNN model GNN_{θ_T}.
Pre-train GNN with local objective function (Eq. 9).
Initialize model parameters θ_0 and $\mathbf{C}_0$.
for $t = 1$ **to** T **do**
  Sample a mini-batch $\widetilde{\mathbf{G}}$ from $\mathbf{G}$.
  ♢ *E-step*:
  Sample latent variables $\widetilde{\mathbf{Z}}_{est}$ with $\mathrm{GNN}_{\theta_{t-1}}$ and $\mathbf{C}_{t-1}$.
  ♢ *M-step*:
  Update model parameters:
    $\theta_t \leftarrow \theta_{t-1} - \nabla_\theta(\mathcal{L}_{\mathrm{local}} + \mathcal{L}_{\mathrm{global}})$,
    $\mathbf{C}_t \leftarrow \mathbf{C}_{t-1} - \nabla_{\mathbf{C}}(\mathcal{L}_{\mathrm{local}} + \mathcal{L}_{\mathrm{global}})$.
end for

Xu M, Wang H, Ni B, et al. Self-supervised graph-level representation learning with local and global structure[C]//International Conference on Machine Learning. PMLR, 2021: 11548-11558.

Limitations of Existing Methods

- Implicit utilization of global semantic with additional supervisions
- Global semantic information is invariant for all nodes/graphs
- Global semantic information cannot be applied in downstream tasks

Our Contributions

- We firstly explicitly characterize the **instance-adaptive global-aware** feature by ego-semantic descriptors
- We propose an omni-granular normalization **over all the hierarchies and scales of ego-semantic**
- Specialized tasks and a **cross-iteration omni-granular momentum update** are both proposed for achieving local-global mutual adaptation in our framework
- Our OEPG substantially outperforms previous in multiple downstream tasks on datasets cross scales and domains, and **generalizes to quantity- and topology-imbalance scenarios**

Our Framework —— OEPG

- Overview of Omni-Granular Ego-Semantic Propagation

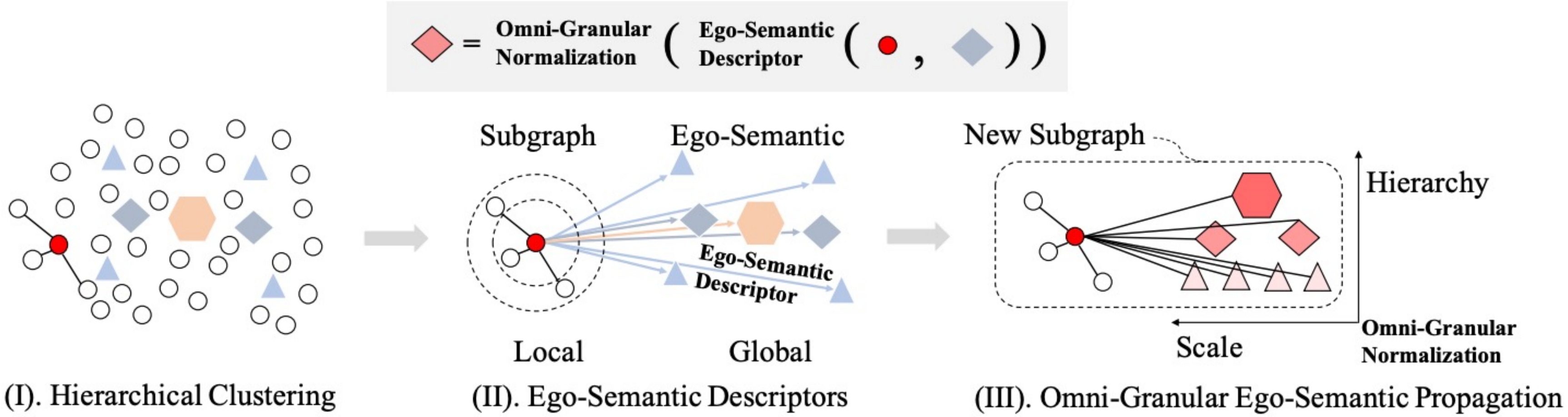


Figure 1. Illustration of OEPG. (I) Obtain hierarchical global clusters. (II) Explicitly define instance-adaptive global-aware ego-semantic descriptors by measuring first- and second-order feature differences between target node (denoted in red) and hierarchical clusters. (III) Perform omni-granular normalization on ego-semantic descriptors and use them to form new subgraph for later feature propagation.

Our Framework —— OEPG

北京大學 PEKING UNIVERSITY

1. Ego-Semantic Descriptor

The First-Order Ego-Semantic Descriptor

Definition 3.1. For s-th cluster in h-th hierarchy, where $h \in [1, H]$, $s \in [1, S_h]$, then the first-order ego-semantic descriptor $\mathbf{D}^{1st}_{s,h} \in \mathbb{R}_d$ for target $\mathbf{V}^{target}$ is defined as:

$$\mathbf{D}^{1st}_{s,h} = (\mathbf{V}^{target} - \mathbf{C}_{s,h})/||\mathbf{V}^{target} - \mathbf{C}_{s,h}||_2, \qquad (2)$$

where we derive $\mathbf{V}^{target}/||\mathbf{V}^{target} - \mathbf{C}_{s,h}||_2 = f^t \in \mathbb{R}^d$, $\mathbf{C}_{s,h}/||\mathbf{V}^{target} - \mathbf{C}_{s,h}||_2 = f^c_{s,h} \in \mathbb{R}^d$, and we specify $\mathbf{D}^{1st}_{s,h}$ as ($[d]$ denotes d-th dimension):

$$[f^t[1] - f^c_{s,h}[1], \cdots, f^t[d] - f^c_{s,h}[d]]. \qquad (3)$$

The Second-Order Ego-Semantic Descriptor

Definition 3.2. For s-th cluster in h-th hierarchy, where $h \in [1, H]$, $s \in [1, S_h]$, the second-order ego-semantic descriptor $\mathbf{D}^{2nd}_{s,h} \in \mathbb{R}^{(S_1+S_2+\cdots+S_H)}$ for $\mathbf{V}^{target}$ is:

$$\begin{aligned} \mathbf{X} &= [\mathbf{D}^{1st}_{1,1} \bullet \mathbf{D}^{1st}_{s,h}, \cdots, \mathbf{D}^{1st}_{S_H,H} \bullet \mathbf{D}^{1st}_{s,h}], \\ \mathbf{D}^{2nd}_{s,h} &= \mathbf{X}/||\mathbf{X}||_2, \end{aligned} \qquad (4)$$

where $\bullet$ is the inner product. $\mathbf{X}$ is further specified as:

$$\begin{aligned} &[\sum_{i=1}^{d}(f^t[i] - f^c_{1,1}[i])(f^t[i] - f^c_{s,h}[i]), \cdots, \\ &\sum_{i=1}^{d}(f^t[i] - f^c_{S_H,H}[i])(f^t[i] - f^c_{s,h}[i])] \end{aligned} \qquad (5)$$

Our Framework —— OEPG

1. Ego-Semantic Descriptor

typical local graph convolution

$$\mathbf{V}_l^{target} = \mathcal{F}(\underbrace{\mathbf{V}_{l-1}^{target}, \{\mathbf{V}_{i,l-1}^{local}\}_{i=1}^{N_n}}_{local\ context}),$$

ego-semantic descriptors can be explicitly integrated into the local graph convolutional process

$$\mathbf{V}_l^{target} = \mathcal{F}(\underbrace{\mathbf{V}_{l-1}^{target}, \{\mathbf{V}_{i,l-1}^{local}\}_{i=1}^{N_n}}_{local\ context}, \underbrace{\{\{\mathbf{D}_{s,h}\}_{s=1}^{S_h}\}_{h=1}^{H}}_{global\ ego-semantic}).$$

Our Framework —— OEPG

2. Omni-Granular Normalization

$$\bar{\mathbf{D}}_{s,h}^{1st} = a_{s,h}^{1st} \cdot \mathbf{D}_{s,h}^{1st} = \frac{e^{-\alpha||\mathbf{D}_{s,h}^{1st}||^2}}{\sum_{m=1}^{H}\sum_{k=1}^{S_m} e^{-\alpha||\mathbf{D}_{k,m}^{1st}||^2}} \cdot \mathbf{D}_{s,h}^{1st},$$

$$\bar{\mathbf{D}}_{s,h}^{2nd} = \frac{e^{-\beta||\mathbf{D}_{s,h}^{2nd}||^2}}{\sum_{m=1}^{H}\sum_{k=1}^{S_m} e^{-\beta||\mathbf{D}_{k,m}^{2nd}||^2}} \cdot \mathbf{D}_{s,h}^{2nd}.$$

$$\mathbf{D}_{s,h} = \mathrm{LeakyReLU}(\mathbf{W} \cdot \mathrm{Concat}(\bar{\mathbf{D}}_{s,h}^{1st}, \bar{\mathbf{D}}_{s,h}^{2nd})),$$

Node level

$$\left[\begin{array}{c:c} A[local, local] = A_{ori} & \begin{array}{c} A[target, global] = 1 \\ A[neighbors, global] = 0 \end{array} \\ \hdashline \begin{array}{c} A[global, target] = 1 \\ A[global, neighbors] = 0 \end{array} & A[global, global] = 0 \end{array}\right]$$

Graph level

$$\left[\begin{array}{c:c} A[local, local] = A_{ori} & A[local, global] = 1 \\ \hdashline A[global, local] = 1 & A[global, global] = 0 \end{array}\right]$$

Our Framework —— OEPG

3. Specialized Local-Global Pretext Tasks

$$\mathcal{L}_{contrastive} = -\,\mathbb{E}_{\mathcal{G}\sim p(\mathcal{G})}[sim(\mathcal{F}(T_i(\mathcal{G})), \mathcal{F}(T_j(\mathcal{G})))] + \mathbb{E}_{\mathcal{G}_-\sim p(\mathcal{G}_-)}[sim(\mathcal{F}(T_i(\mathcal{G})), \mathcal{F}(T_j(\mathcal{G}_-)))], \quad (14)$$

$$\mathcal{L}_{predictive} = -\,\mathbb{E}_{\mathcal{G}\sim p(\mathcal{G})}\, log(\sigma(\mathcal{F}(\mathcal{G}_1)\cdot \mathcal{F}_{aux}(\mathcal{G}_2))), \quad (15)$$

Our Framework —— OEPG

PEKING UNIVERSITY

4. Cross-Iteration Omni-Granular Momentum Update

$$s = \operatorname*{argmax}_k(sim(\mathbf{V}, \{\mathbf{C}_{k,h}\}_{k=1}^{S_h})).$$

Instead of updating node queue after each training iteration, we construct a node queue $\{\mathbf{Q}_{s,h,i}\}_{i=1}^{i\in[1,\mathcal{B}]}$ with a budget $\mathcal{B}$ for each cluster. If the number of nodes in a queue equals to the $\mathcal{B}$, the corresponding cluster will be update as follow:

$$\mathbf{C}_{s,h} \longleftarrow m\mathbf{C}_{s,h} + \frac{(1-m)}{\mathcal{B}}\sum_{i=1}^{\mathcal{B}}\mathbf{Q}_{s,h,i}, \qquad (17)$$

Our Framework —— OEPG

北京大學 PEKING UNIVERSITY

4. Full pipeline

Algorithm 1 Algorithm of OEPG

Input: Unlabeled graph dataset $\mathbf{G}$, the number of training steps T.
Output: Pre-trained GNN model GNN_{θ_T}.
Initialize the model parameters θ_0 with local pre-training.
Initialize multi-granular clusters $\{\{\mathbf{C}_{s,h}\}_{s=1}^{S_h}\}_{h=1}^{H}$.
Initialize node queue $\{\mathbf{Q}_{s,h}\}$ for each $\mathbf{C}_{s,h}$.
for $t = 1$ **to** T **do**
  Sample a mini-batch subgraphs $\{\mathcal{G}_i\}_{i=1}^{N} \in \mathbf{G}$.
  for $i = 1$ **to** N **do**
    if $size(\{\mathbf{Q}_{\mathbf{s},\mathbf{h}}\}) = max\ budget\ \mathcal{B}$ **then**
      Update $\mathbf{C}_{s,h}$ with Eq.16,17, empty $\{\mathbf{Q}_{\mathbf{s},\mathbf{h}}\}$.
    else
      Enqueue target node $V_i \in \mathcal{G}_i$ to $\{\mathbf{Q}_{\mathbf{s},\mathbf{h}}\}$..
    end if
    (1). Obtain 1st, 2nd order ego-semantic descriptors $\{\{\mathbf{D}_{s,h}^{1st}\}_{s=1}^{S_h}\}_{h=1}^{H}$, $\{\{\mathbf{D}_{s,h}^{2nd}\}_{s=1}^{S_h}\}_{h=1}^{H}$ for target node according to Eq.2,4.
    (2). Omni-granular norm on $\mathbf{D}_{s,h}^{1st}$, $\mathbf{D}_{s,h}^{2nd}$, and combine them to get $\mathbf{D}_{s,h}$ with Eq.9,10,11.
    (3). Integrate $\{\{\mathbf{D}_{s,h}\}_{s=1}^{S_h}\}_{h=1}^{H}$ with $\mathcal{G}_i$ to form new subgraph $\mathcal{G}_i^{new}$ according to Eq.13.
  end for
  Update model parameters θ_t according to Eq.14, 15.
end for

Results and Analysis

北京大學 PEKING UNIVERSITY

Table 1. Downstream test accuracy (%) in self-supervised learning. The compared results are from the published papers.

METHODS	NCI1	PROTEINS	DD	MUTAG	COLLAB	RDT-B	RDT-M5K	IMDB-B
NODE2VEC (GROVER & LESKOVEC, 2016)	54.9±1.6	57.5±3.6	-	72.6±10.2	-	-	-	-
SUB2VEC (ADHIKARI ET AL., 2018)	52.8±1.5	53.0±5.6	-	61.1±15.8	-	71.5±0.4	36.7±0.4	55.3±1.5
GRAPH2VEC (NARAYANAN ET AL., 2017)	73.2±1.8	73.3±2.1	-	83.2±9.3	-	75.8±1.0	47.9±0.3	71.1±0.5
GAE (KIPF & WELLING, 2016B)	-	-	-	87.7±0.7	-	87.1±0.1	52.8±0.2	70.7±0.7
MVGRL (HASSANI & KHASAHMADI, 2020)	-	-	-	75.4±7.8	-	82.0±1.1	-	63.6±4.2
INFOGRAPH (SUN ET AL., 2019)	76.2±1.1	74.4±0.3	72.9±1.8	89.0±1.1	70.7±1.1	82.5±1.4	53.5±1.0	73.0±0.9
GRAPHCL (YOU ET AL., 2020A)	77.9±0.4	74.4±0.5	78.6±0.4	86.8±1.3	71.4±1.2	89.5±0.8	56.0±0.3	71.1±0.4
JOAO (YOU ET AL., 2021)	78.4±0.5	74.1±1.1	77.4±1.2	87.7±0.8	69.3±0.3	86.4±1.5	56.0±0.3	70.8±0.3
ADGCL (SURESH ET AL., 2021)	69.7±0.5	73.8±0.5	75.1±0.4	89.7±1.0	73.3±0.6	85.5±0.8	54.9±0.4	72.3±0.6
INFOGCL (XU ET AL., 2021A)	80.2±0.6	-	-	91.2±1.3	80.0±1.3	-	-	75.1±0.9
DGCL (LI ET AL., 2021)	81.9±0.2	76.4±0.5	-	92.1±0.2	81.2±0.3	92.7±0.2	56.1±0.2	75.9±0.7
OEPG (OURS)	**84.8**±0.4	**79.6**±0.7	**81.4**±0.9	**95.3**±0.6	**84.7**±0.7	**96.3**±0.9	**60.5**±0.3	**78.5**±0.6

Table 2. Downstream test ROC-AUC (%) in transfer learning. The compared results are from the published papers.

METHODS	BBBP	TOX21	TOXCAST	SIDER	CLINTOX	MUV	HIV	BACE
EDGEPRED (KIPF & WELLING, 2016B)	67.3±2.4	76.0±0.6	64.1±0.6	60.4±0.7	64.1±3.7	74.1±2.1	76.3±1.0	79.9±0.9
INFOGRAPH (SUN ET AL., 2019)	68.2±0.7	75.5±0.6	63.1±0.3	59.4±1.0	70.5±1.8	75.6±1.2	77.6±0.4	78.9±1.1
ATTRMASKING (HU ET AL., 2020B)	64.3±2.8	76.7±0.4	64.2±0.5	61.0±0.7	71.8±4.1	74.7±1.4	77.2±1.1	79.3±1.6
CONTEXTPRED (RONG ET AL., 2020)	68.0±2.0	75.7±0.7	63.9±0.6	60.9±0.6	65.9±3.8	75.8±1.7	77.3±1.0	79.6±1.2
GRAPHPARTITION (YOU ET AL., 2020B)	70.3±0.7	75.2±0.4	63.2±0.3	61.0±0.8	64.2±0.5	75.4±1.7	77.1±0.7	79.6±1.8
GRAPHCL (YOU ET AL., 2020A)	69.5±0.5	75.4±0.9	63.8±0.4	60.8±0.7	70.1±1.9	74.5±1.3	77.6±0.9	78.2±1.2
JOAO (YOU ET AL., 2021)	71.4±0.9	74.3±0.6	63.2±0.5	60.5±0.7	81.0±1.6	73.7±1.0	77.5±1.2	75.5±1.3
GRAPHLOG (XU ET AL., 2021B)	72.5±0.8	75.7±0.5	63.5±0.7	61.2±1.1	76.7±3.3	76.0±1.1	77.8±0.8	83.5±1.2
OEPG (OURS)	**75.7**±0.6	**79.2**±0.7	**66.2**±0.4	**64.1**±0.9	**84.5**±1.7	**81.6**±1.4	**81.3**±0.9	**85.2**±1.3

Results and Analysis

PEKING UNIVERSITY 北京大學

Table 3. Downstream test accuracy (%) in semi-supervised learning. The compared results are from the published papers.

Methods	WikiCS	Amazon Computers	Amazon Photos	Coauthor CS	Coauthor Physics
DGI (Veličković et al., 2018)	75.4±0.1	84.0±0.5	91.6±0.2	92.2±0.6	94.5±0.5
GMI (Peng et al., 2020)	74.9±0.1	82.2±0.3	90.7±0.2	OOM	OOM
MVGRL (Hassani & Khasahmadi, 2020)	77.5±0.1	87.5±0.1	91.7±0.1	92.1±0.1	95.3±0.1
GBT (Bielak et al., 2021)	77.3±0.6	88.0±0.3	92.2±0.4	92.9±0.3	95.2±0.1
GRACE (Zhu et al., 2020)	80.1±0.5	89.5±0.4	92.8±0.5	91.1±0.2	OOM
GCA (Zhu et al., 2021b)	78.4±0.1	88.9±0.2	92.5±0.2	93.1±0.1	95.7±0.1
CCA (Zhang et al., 2021a)	-	88.7±0.3	93.1±0.1	93.3±0.2	95.4±0.1
BGRL (Thakoor et al., 2021)	79.4±0.5	89.7±0.3	92.9±0.3	93.2±0.1	95.6±0.1
OEPG (ours)	**83.3**±0.3	**91.9**±0.5	**95.1**±0.4	**95.4**±0.1	**97.3**±0.1

Table 4. Test Macro-F1 (%) in semi-supervised learning. The imbalance ratio (I.R.) is set to different levels (5%, 10%) to test under different imbalance intensities. The supervised method is Chen et al. (2021)).

Datasets	Coauthor CS		Coauthor Physics	
I.R.	5%	10%	5%	10%
Supervised	83.9±2.1	81.3±3.2	72.4±2.6	70.2±2.8
GraphCL	75.3±4.1	72.2±5.3	68.2±4.6	66.5±4.9
JOAO	77.8±3.9	74.6±4.5	67.4±4.1	66.0±4.8
DGCL	80.4±3.2	78.1±3.5	69.5±3.9	67.1±3.9
OEPG	**85.1**±1.8	**83.4**±2.1	**75.9**±2.2	**72.6**±2.5

Table 5. Semi-supervised learning on large-scale OGB datasets on (accuracy in % on ogbg-ppa, F1 score in % on ogbg-code, ROC-AUC in % on ogbg-molhiv). L.R. denotes the label ratio.

L.R.	Methods	ppa	code	molhiv
1%	GraphCL	40.8±1.3	7.6±0.3	67.6±1.6
	JOAO	47.2±1.3	6.8±0.3	-
	DGCL	-	-	69.0±1.7
	OEPG	**52.7**±1.2	**9.1**±0.4	**74.2**±1.6
10%	GraphCL	57.8±1.3	22.5±0.2	70.6±1.6
	JOAO	60.9±0.8	22.1±0.3	-
	DGCL	-	-	73.6±1.5
	OEPG	**64.4**±0.7	**24.8**±0.3	**77.5**±1.7

Results and Analysis

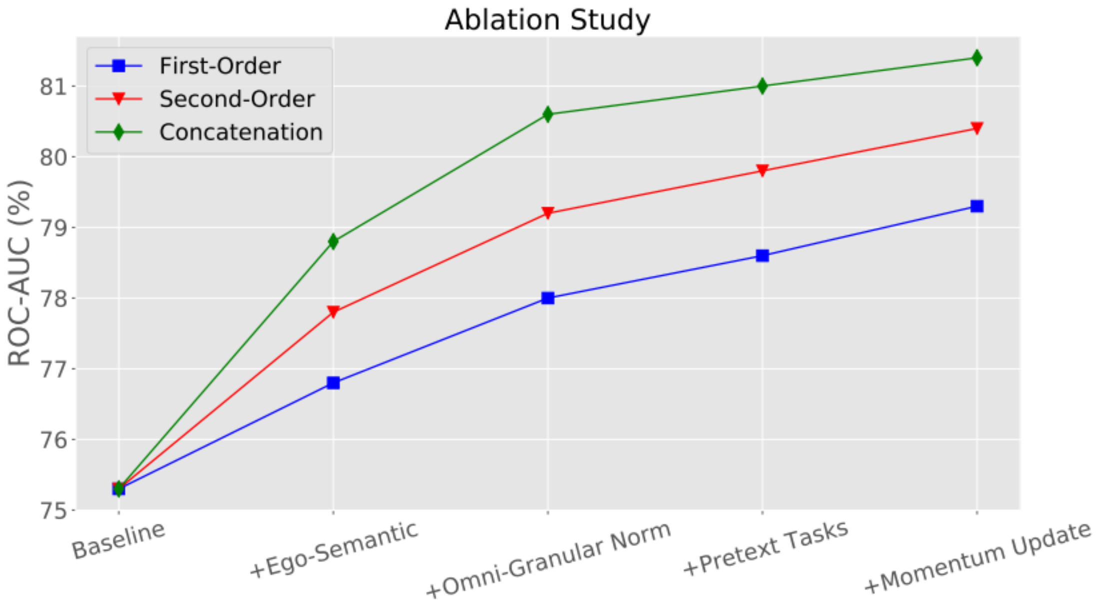


Figure 2. Ablation study on MUV dataset.

Thank You!