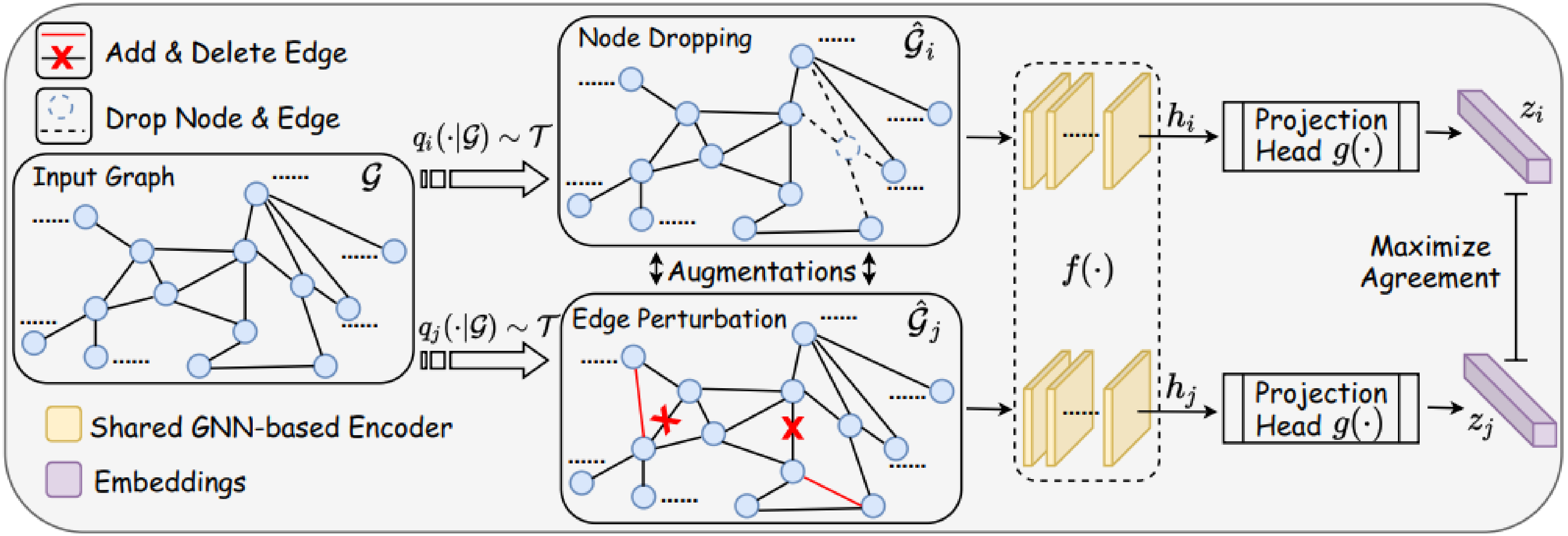
Coloring Learning for Heterophilic Graph Representation

Miaomiao Huang[1], Yuhai Zhao[1], Zhengkui Wang[2], Fenglong Ma[3], Yejiang Wang[1], Meixia Wang[1], Xingwei Wang[1]

[1] Northeastern University, China
[2] Singapore Institute of Technology, Singapore
[3] Pennsylvania State University, United States

Graph Contrastive Learning (GCL) [1]

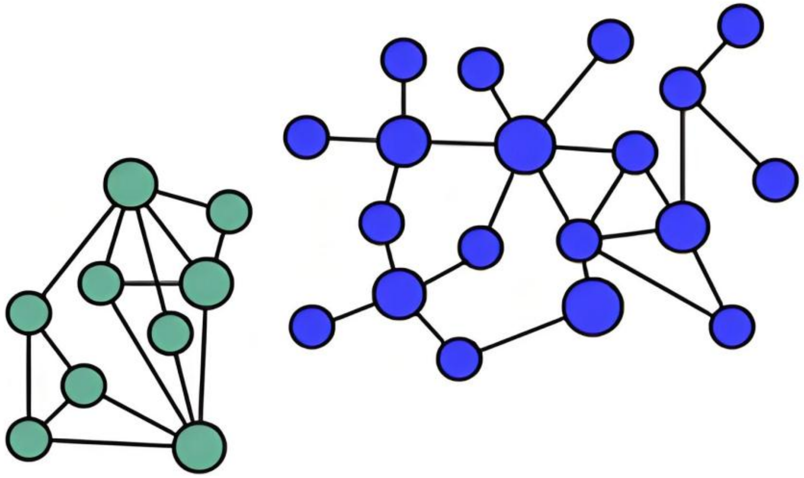
- Data augmentation: Generate two views of a graph through augmentations;
- Contrastive loss: Maximize agreement for augmented views of the same graph (positive pairs) and minimize agreement across different graphs (negative pairs).



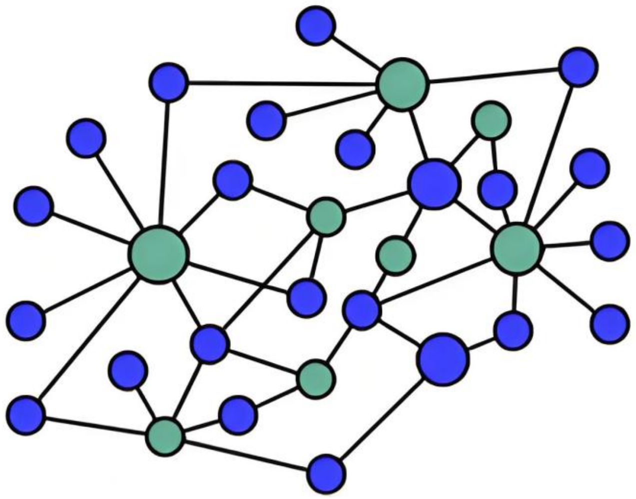
[1] You Y, et al. "Graph contrastive learning with augmentations". *NeurIPS*. 2020.

Heterophilic Graph [2]

- Heterophilic Graph: Connected nodes may have different class labels and dissimilar features.



Homophily

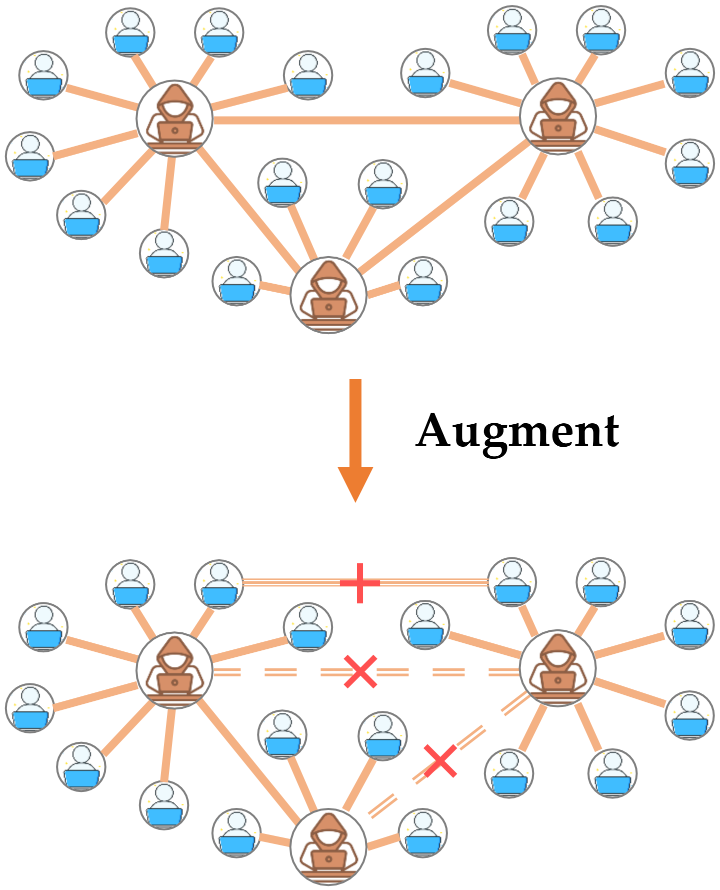


Heterophily

[2] Xiao T, et al. “Simple and Asymmetric Graph Contrastive Learning without Augmentations”. *NeurIPS*. 2023.

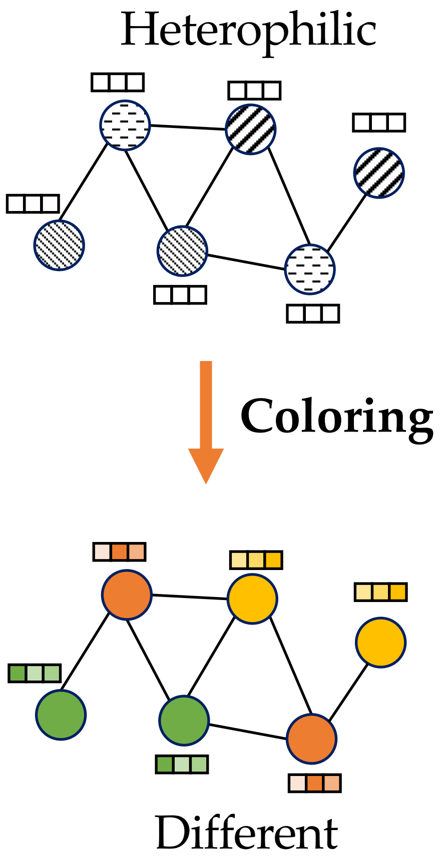
Problem of GCL under Heterophily

- GCL under heterophily depends on the either random or carefully designed augmentation strategies, such as edge dropping. However, in heterophilic graphs, even slight topological perturbations can drastically alter neighborhood semantics;
- Most GCL approaches primarily focus on enforcing local similarity constraints while overlooking global structural consistency.

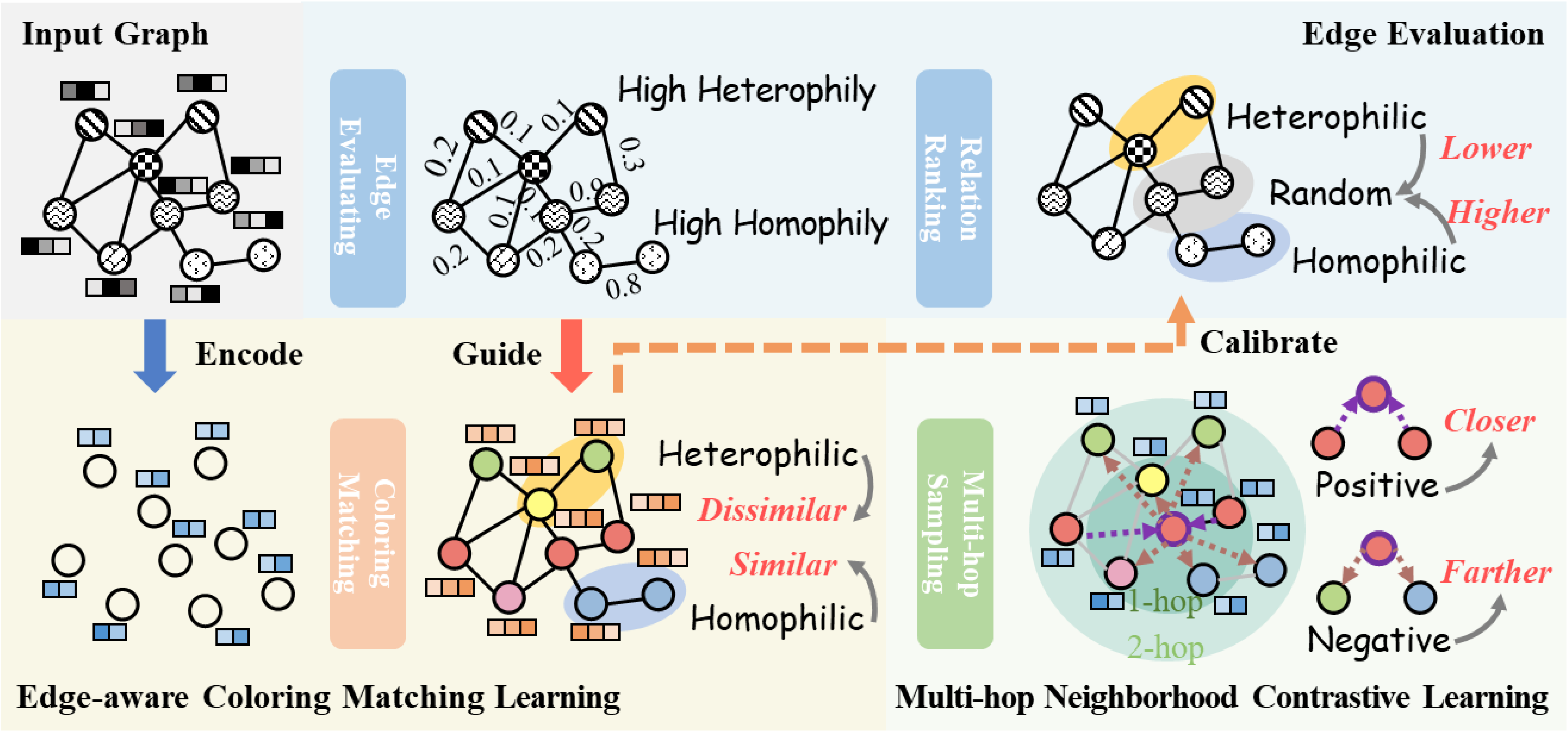


Straightforward Solution

- Graph Coloring: Graph coloring is the process of assigning colors to nodes such that no two adjacent nodes share the same color, while minimizing the total number of colors used;
- Issue 1) Diverse relational patterns: Cannot effectively distinguish heterophilic and homophilic connections;
- Issue 2) Computational complexity: Direct graph coloring is a hard combinatorial problem, difficult to apply in practice.



Framework



Input Graph
Edge Evaluation
Edge Evaluating
High Heterophily
High Homophily
0.2
0.1
0.1
0.1
0.3
0.8
0.2
0.2
0.8
Relation Ranking
Heterophilic
Lower
Random
Higher
Homophilic
Encode
Guide
Calibrate
Coloring Matching
Heterophilic
Dissimilar
Similar
Homophilic
Multi-hop Sampling
1-hop
2-hop
Closer
Positive
Farther
Negative
Edge-aware Coloring Matching Learning
Multi-hop Neighborhood Contrastive Learning

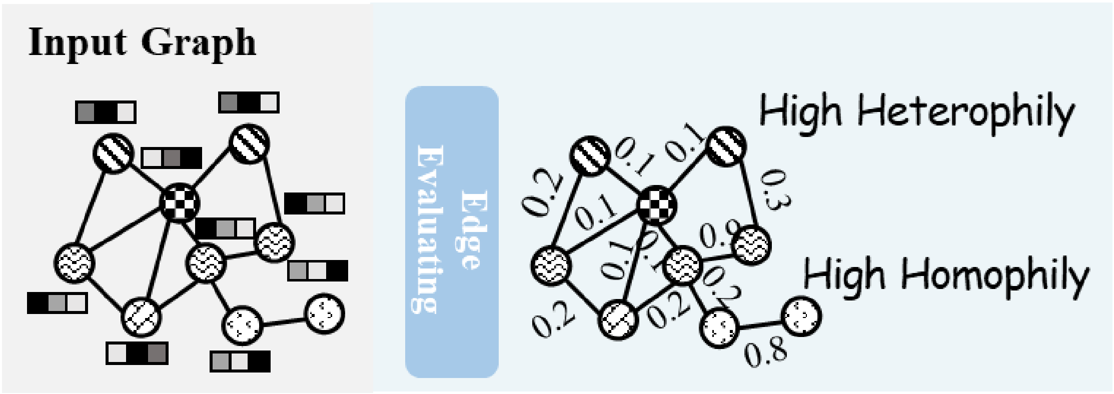
Edge Evaluation

To address the issue of diverse relational patterns, we introduce an edge evaluator that integrates feature and structural information inside the graph, to identify the relationships between node pairs.

$$\mathrm{x}'_i = \phi_1([\mathrm{x}_i \parallel \mathrm{p}_i]), \mathrm{x}'_j = \phi_1([\mathrm{x}_j \parallel \mathrm{p}_j]),$$

$$\omega_{i,j} = \left(\phi_2([\mathrm{x}'_i \parallel \mathrm{x}'_j]) + \phi_2([\mathrm{x}'_j \parallel \mathrm{x}'_i])\right)/2$$

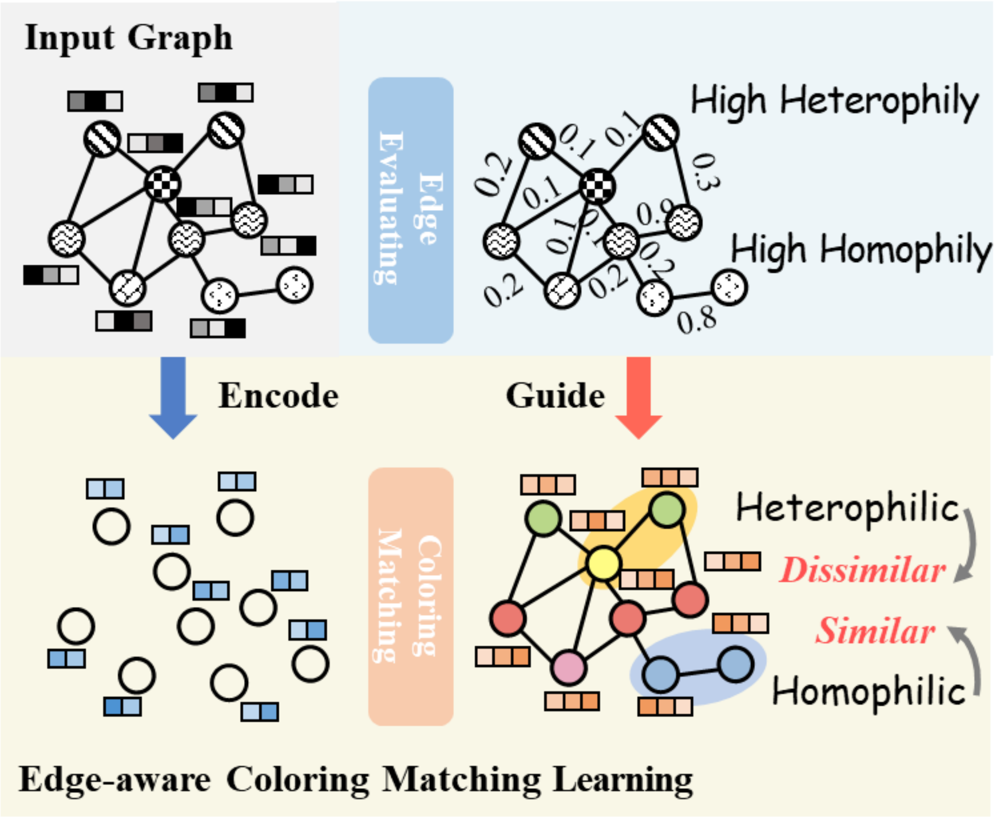
$$\hat{\omega}_{i,j} = \mathrm{Sigmoid}\left((\omega_{i,j} + \log \varpi - \log(1-\varpi))/\tau_m\right)$$



Edge-aware Coloring Matching Learning

To address the issue of computational complexity, we take graph coloring as an inspiration. By assigning the same colors to homophilic node pairs and different colors to heterophilic node pairs, we enable self-supervised representation learning.

$$\mathcal{L}_m = \frac{1}{n} \sum_{v_i \in \mathcal{V}} \frac{1}{|\mathcal{N}(v_i)|} \sum_{v_j \in \mathcal{N}(v_i)} (1 - \hat{\omega}_{i,j}) \cdot \Delta(\pi_i, \pi_j)$$



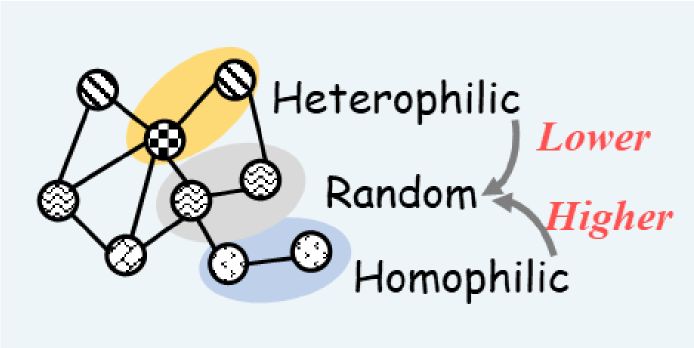
Edge-aware Coloring Matching Learning

A sparse-induced color redundancy constraint is introduced to suppress redundant colors and enhance intra-class compactness;

$$\mathcal{L}_d = \sum_{j \in [\![\chi_{\mathcal{G}}]\!]} \Phi_{max}\left(\{\mathscr{C}_{ij}^{col}\}_{i \in [\![n]\!]}\right), \; \mathscr{C}_i^{col} = \frac{\exp\left((\log(\pi_i) + \varrho_i)/\tau_o\right)}{\sum_{j \in [\![\chi_{\mathcal{G}}]\!]} \exp\left((\log(\pi_j) + \varrho_j)/\tau_o\right)}$$

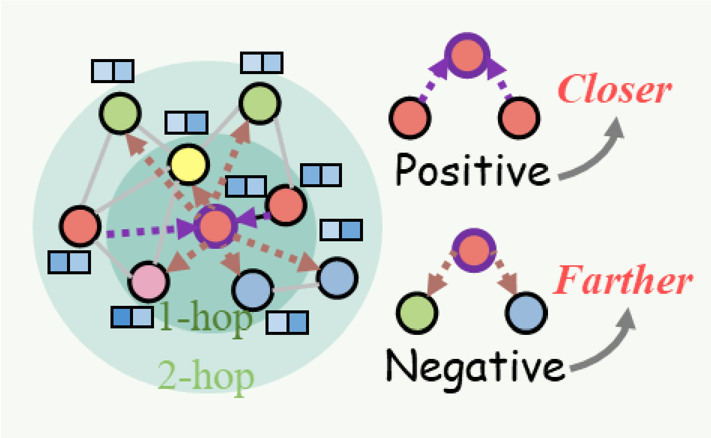
To dynamically calibrate the edge evaluator, we control the ranking of homophilic, heterophilic, and randomly sampled node pairs.

$$\mathcal{L}_r = -\sum_{e_{i,j} \in \mathcal{E}} (1 - \hat{\omega}_{i,j}) \log\left(\sigma\left(\Delta_{p,q}^{rnd} - \Delta_{i,j}^{mat}\right)\right) + \hat{\omega}_{i,j} \log\left(1 - \sigma\left(\Delta_{p,q}^{rnd} - \Delta_{i,j}^{mat}\right)\right)$$



Multi-hop Neighborhood Contrastive Learning

Due to the inherent heterophily of graphs, semantically similar nodes are often located in distant, non-adjacent regions. Local neighbors alone cannot capture global semantic consistency. To address this, we construct positive samples from multi-hop same-color nodes and pull them closer to enforce global consistency.



$$\mathcal{P}_{\mathcal{G}}(v_i) = \left\{ v_j \mid v_j \in \mathcal{N}^{(\varkappa)}(v_i),\ \mathscr{C}_j^{col} = \mathscr{C}_i^{col} \right\}$$

$$\mathcal{L}_c = -\frac{1}{n} \sum_{v_i \in \mathcal{V}} \frac{1}{|\mathcal{P}_{\mathcal{G}}(v_i)|} \sum_{v_j \in \mathcal{P}_{\mathcal{G}}(v_i)} \log \frac{\exp\left(\mathrm{sim}(\mathbf{z}_i, \mathbf{z}_j)/\tau_c\right)}{\sum_{v_k \in \mathcal{V} \setminus \mathcal{P}_{\mathcal{G}}(v_i)} \exp\left(\mathrm{sim}(\mathbf{z}_i, \mathbf{z}_k)/\tau_c\right)}$$

Comparison with the State-of-the-art

Table 1: Results in terms of classification accuracies (in percent ± standard deviation) on homophilic benchmarks. The best and second-best performance under each dataset are marked with **boldface** and <u>underline</u>, respectively. OOM indicates Out-Of-Memory.

Methods	Cora	CiteSeer	PubMed	Wiki-CS	Computers	Photo	CS	Physics
GCN	81.50±1.30	70.30±0.28	78.80±2.90	76.89±0.37	86.34±0.48	92.35±0.25	93.10±0.17	95.54±0.19
GAT	82.80±1.30	71.50±0.49	78.50±0.27	77.42±0.19	87.06±0.35	92.64±0.42	92.41±0.27	95.45±0.17
MLP	56.11±0.34	56.91±0.42	71.35±0.05	72.02±0.21	73.88±0.10	78.54±0.05	90.42±0.08	93.54±0.05
H2GCN	80.23±0.20	69.97±0.66	78.79±0.30	79.73±0.13	84.32±0.52	91.86±0.27	91.18±0.58	93.56±0.48
FAGCN	77.80±0.66	69.81±0.80	76.74±0.66	74.34±0.53	83.51±1.04	92.72±0.22	93.81±0.24	<u>96.16±0.15</u>
PC-Conv	82.47±0.56	69.92±1.33	79.57±1.23	<u>79.94±0.52</u>	87.89±0.26	**93.89±0.14**	94.24±0.12	95.99±0.14
DeepWalk	69.47±0.55	58.82±0.61	69.87±1.25	74.35±0.06	85.68±0.06	89.44±0.11	84.61±0.22	91.77±0.15
node2vec	71.24±0.89	47.64±0.77	66.47±1.00	71.79±0.05	84.39±0.08	89.67±0.12	85.08±0.03	91.19±0.04
GAE	71.07±0.39	65.22±0.43	71.73±0.92	70.15±0.01	85.27±0.19	91.62±0.13	90.01±0.71	94.92±0.07
VGAE	79.81±0.87	66.75±0.37	77.16±0.31	75.63±0.19	86.37±0.21	92.20±0.11	92.11±0.09	94.52±0.00
DGI	82.29±0.56	71.49±0.14	77.43±0.84	75.73±0.13	84.09±0.39	91.49±0.25	91.95±0.40	94.57±0.38
GMI	82.51±1.47	71.56±0.56	79.83±0.90	75.06±0.13	81.76±0.52	90.72±0.33	OOM	OOM
MVGRL	83.03±0.27	<u>72.75±0.46</u>	79.63±0.38	77.97±0.18	87.09±0.27	92.01±0.13	91.97±0.19	95.53±0.10
GRACE	80.08±0.53	71.41±0.38	80.15±0.34	79.16±0.36	87.21±0.44	92.65±0.32	92.78±0.23	95.39±0.32
GCA	80.39±0.42	71.21±0.24	80.37±0.75	79.35±0.12	87.84±0.27	92.78±0.17	93.32±0.12	95.87±0.15
BGRL	81.08±0.17	71.59±0.42	79.97±0.36	78.74±0.22	<u>88.92±0.33</u>	93.24±0.29	93.26±0.36	95.76±0.38
HGRL	80.66±0.43	68.56±1.10	80.35±0.58	76.68±0.17	84.30±0.47	93.53±0.22	93.99±0.15	OOM
GREET	<u>83.32±0.49</u>	72.20±1.01	80.50±0.66	79.87±0.49	87.55±0.37	92.99±0.38	<u>94.68±0.21</u>	95.91±0.14
HeteGCL	81.55±0.65	70.63±1.16	<u>82.50±0.57</u>	79.12±0.25	85.76±0.21	93.82±0.32	**94.79±0.06**	OOM
CoRep	**85.04±0.34**	**73.67±0.40**	**83.50±0.47**	**82.20±0.51**	**89.17±3.81**	<u>93.84±1.89</u>	94.39±0.31	**96.21±0.11**

Comparison with the State-of-the-art

Table 2: Results in terms of classification accuracies (in percent ± standard deviation) on heterophilic benchmarks. The best and second-best performance under each dataset are marked with **boldface** and underline, respectively. OOM indicates Out-Of-Memory.

Methods	Chameleon	Squirrel	Actor	Cornell	Texas	Wisconsin
GCN	59.63±2.32	36.28±1.52	30.83±0.77	57.03±3.30	60.00±4.80	56.47±6.55
GAT	56.38±2.19	32.09±3.27	28.06±1.48	59.46±3.63	61.62±3.78	54.71±6.87
MLP	46.91±2.15	29.28±1.33	35.66±0.94	81.08±7.93	81.62±5.51	84.31±3.40
H2GCN	59.39±1.98	37.90±2.02	35.86±1.03	<u>82.16±4.80</u>	84.86±6.77	86.67±4.69
FAGCN	<u>63.44±2.05</u>	<u>41.17±1.94</u>	36.81±0.26	81.35±5.05	84.32±6.02	83.33±2.01
PC-Conv	53.20±1.60	35.79±0.62	36.07±0.61	78.65±2.70	<u>85.68±2.97</u>	**88.63±2.94**
DeepWalk	47.74±2.05	32.93±1.58	22.78±0.64	39.18±5.57	46.49±6.49	33.53±4.92
node2vec	41.93±3.29	22.84±0.72	28.28±1.27	42.94±7.46	41.92±7.76	37.45±7.09
GAE	33.84±2.77	28.03±1.61	28.03±1.18	58.85±3.21	58.64±4.53	52.55±3.80
VGAE	35.22±2.71	29.48±1.48	26.99±1.56	59.19±4.09	59.20±4.26	56.67±5.51
DGI	39.95±1.75	31.80±0.77	29.82±0.69	63.35±4.61	60.59±7.56	55.41±5.96
GMI	46.97±3.43	30.11±1.92	27.82±0.90	54.76±5.06	50.49±2.21	45.98±2.76
MVGRL	51.07±2.68	35.47±1.29	30.02±0.70	64.30±5.43	62.38±5.61	62.37±4.32
GRACE	48.05±1.81	31.33±1.22	29.01±0.78	54.86±6.95	57.57±5.68	50.00±5.83
GCA	49.80±1.81	35.50±0.91	29.65±1.47	55.41±4.56	59.46±6.16	50.78±4.06
BGRL	47.46±2.74	32.64±0.78	29.86±0.75	57.30±5.51	59.19±5.85	52.35±4.12
HGRL	48.29±1.64	35.79±0.89	36.97±0.98	79.46±4.45	82.16±6.00	86.28±3.58
GREET	63.09±2.18	40.86±1.93	35.75±1.08	73.78±3.64	85.41±3.67	84.12±4.76
HeteGCL	48.77±1.55	34.27±1.58	**37.59±1.22**	81.32±6.26	82.37±5.83	80.39±5.23
CoRep	**65.64±1.39**	**46.88±1.56**	<u>37.32±1.13</u>	**82.70±4.55**	**88.65±3.97**	<u>86.86±3.17</u>

Thanks!