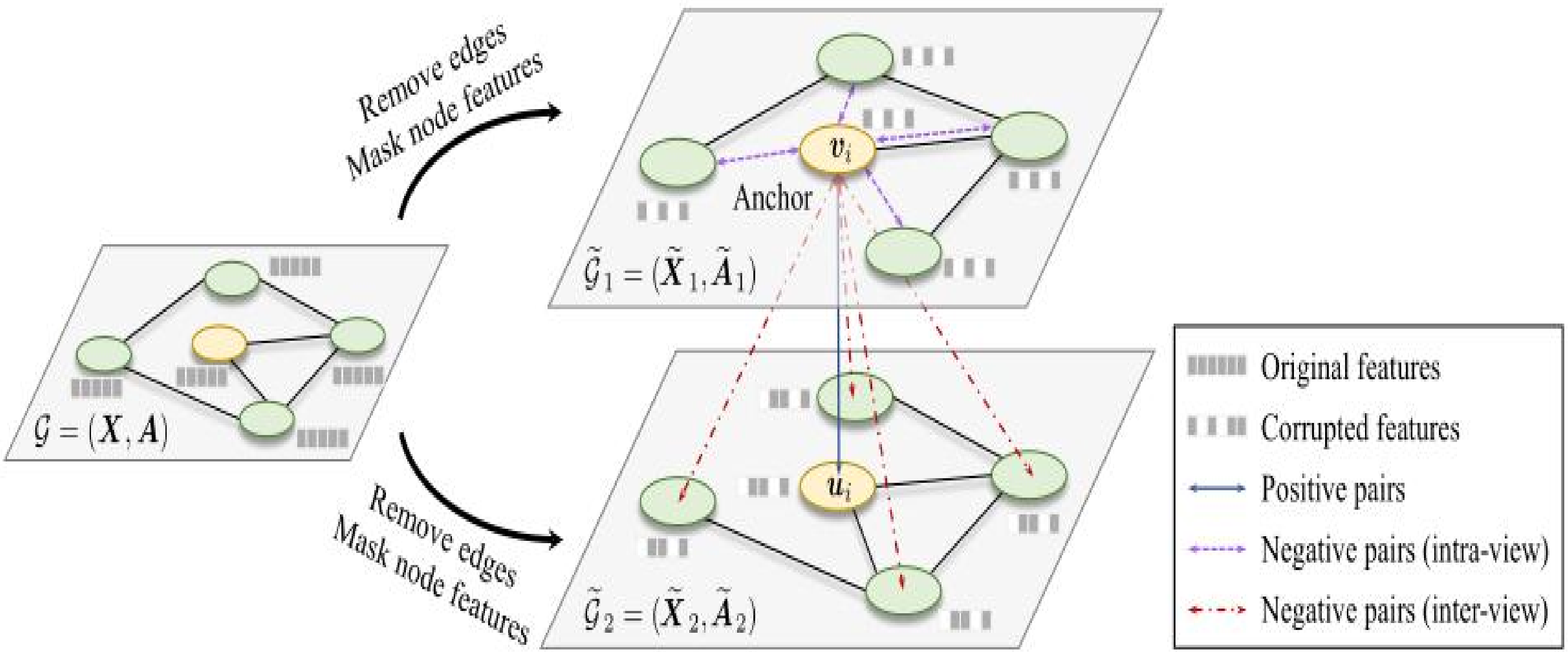
Calibrated Graph Contrastive Learning via Partitioned Similarity and Consistency Discrimination

Background

GCL

General graph contrastive learning (GCL) pipeline:

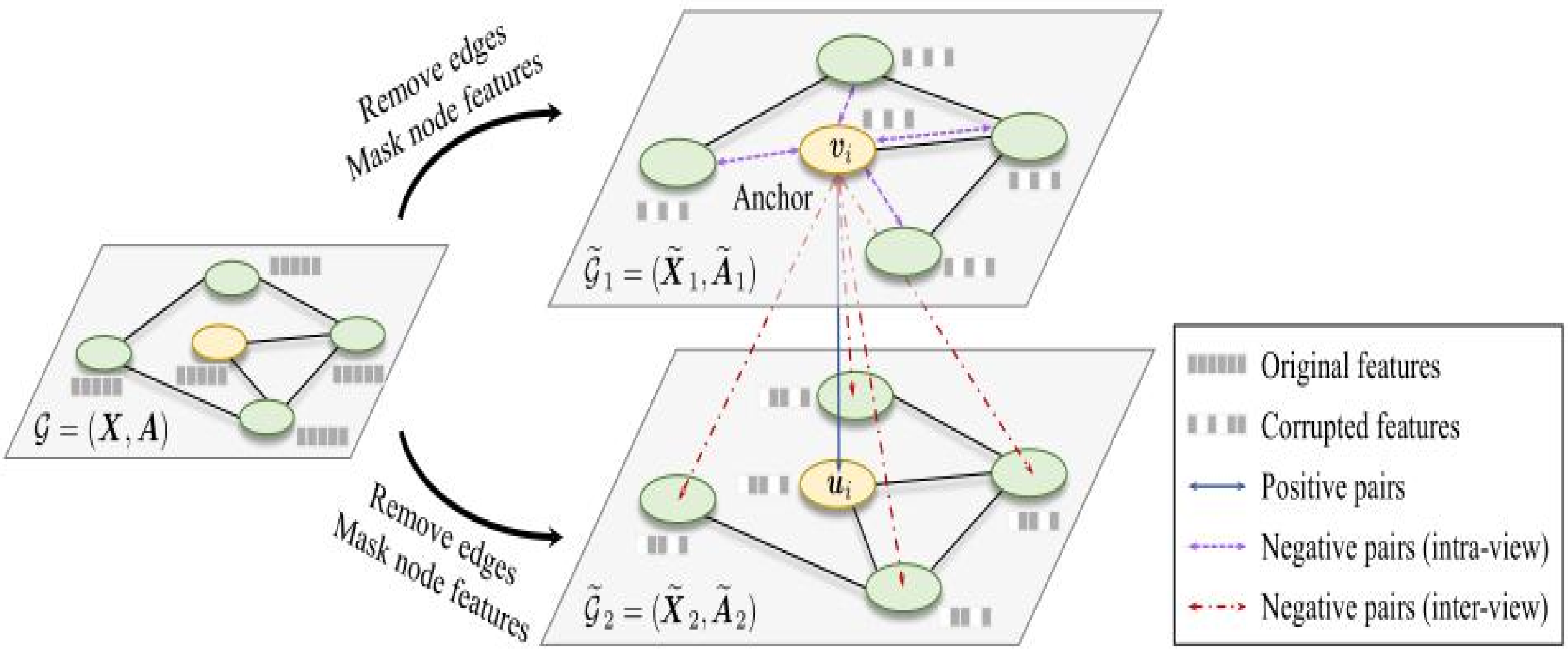
- **Implement augmentations on raw data to generate multiple views;**
- **Corresponding samples across augmented views are positives while others are negatives for contrastive learning.**



Background

Problem

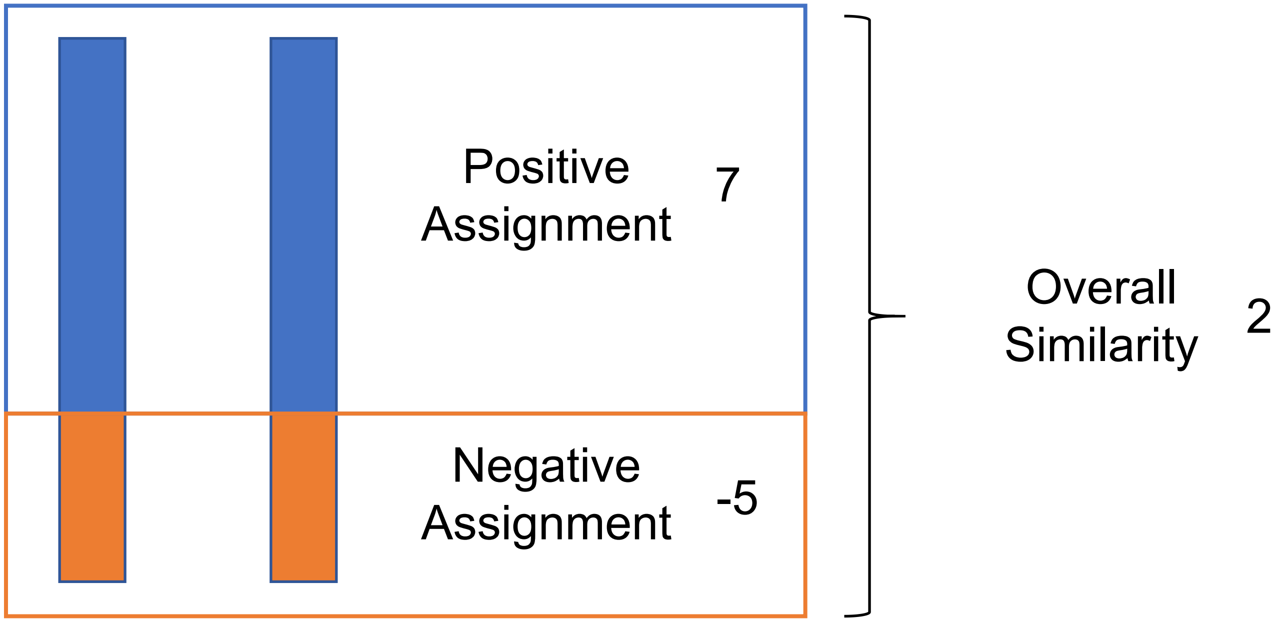
- **Semantic shift bias.** Random augmentations alter the underlying semantics of samples, leading to incorrect positive or negative assignments and injecting noise into training.



Background

Problem

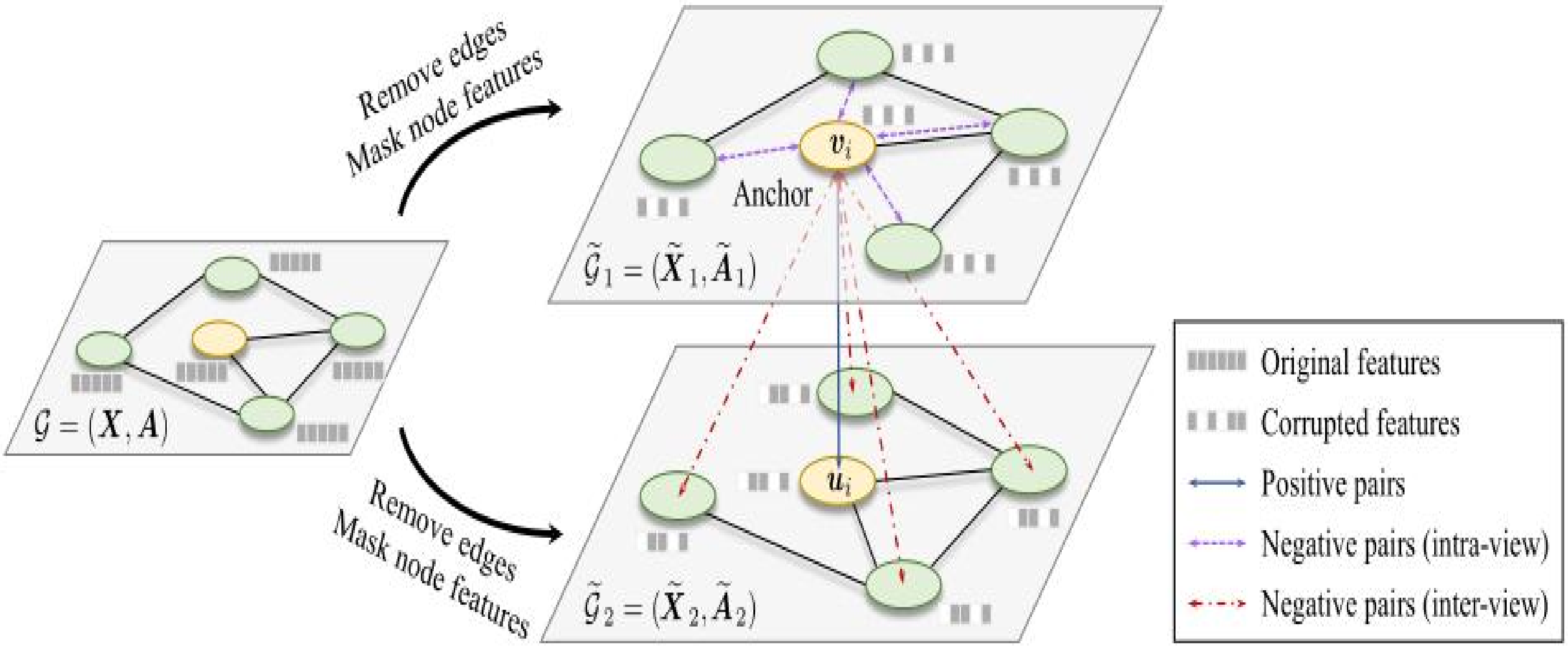
- **Similarity estimation bias.** Feature elements supporting positive pair alignment are suppressed by conflicting components within the representation, causing positive pairs to appear less similar.



Overall similarity of positive pairs are suppressed.

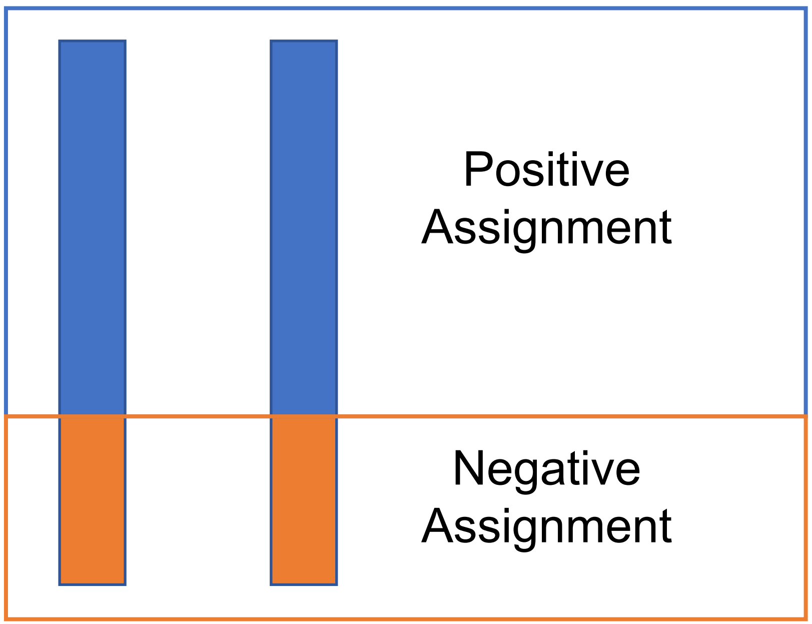
Method

➢ **Semantics-consistency discriminator** corrects the semantic relationships of augmented sample pairs for debiasing semantic shift.



Method

- **Exponential partitioned similarity** restricts the negatively-synergistic features and amplify the positively-synergistic features to alleviate the similarity estimation bias.



Exponential partitioned similarity:

$$\mathcal{S}_{part}(\boldsymbol{h}_{v_i}, \boldsymbol{h}_{v_j}) = \frac{1}{K}\sum_{k=1}^{K} \exp\left(\frac{K \cdot \boldsymbol{h}_{v_i}^{(k)\top} \boldsymbol{h}_{v_j}^{(k)}}{\tau}\right),$$

Example:

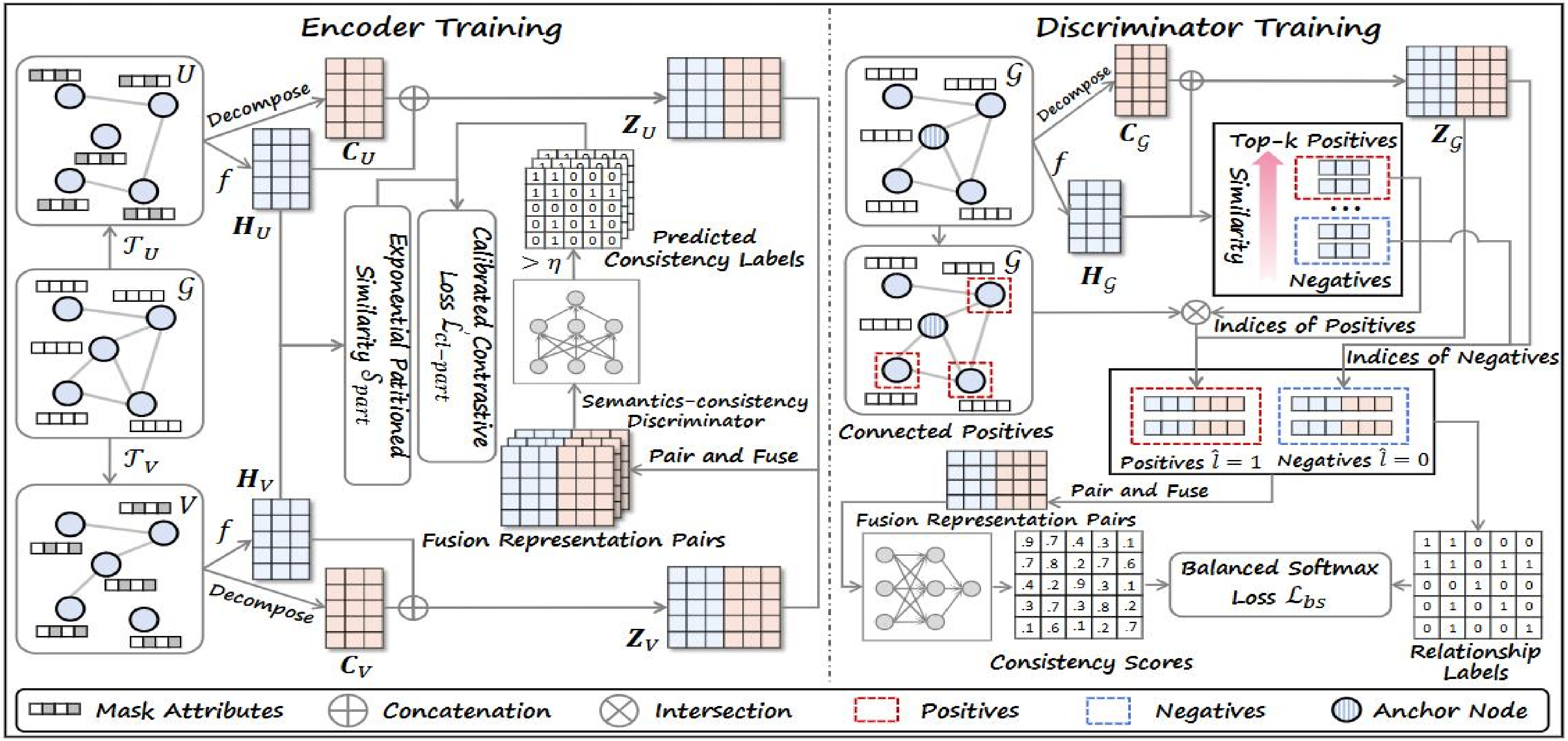
$$\boldsymbol{h}_{v_i}^{\top}\boldsymbol{h}_{v_j} = \boldsymbol{h}_{v_i}^{(1)\top}\boldsymbol{h}_{v_j}^{(1)} + \boldsymbol{h}_{v_i}^{(2)\top}\boldsymbol{h}_{v_j}^{(2)} \text{ and } \boldsymbol{h}_{v_i}^{(2)\top}\boldsymbol{h}_{v_j}^{(2)} > \boldsymbol{h}_{v_i}^{\top}\boldsymbol{h}_{v_j}$$

$$\exp\left(\frac{2 \cdot \boldsymbol{h}_{v_i}^{(1)\top}\boldsymbol{h}_{v_j}^{(1)}}{\tau}\right) \rightarrow 0, \ and \ \exp\left(\frac{2 \cdot \boldsymbol{h}_{v_i}^{(2)\top}\boldsymbol{h}_{v_j}^{(2)}}{\tau}\right) \geq \exp\left(\frac{\boldsymbol{h}_{v_i}^{\top}\boldsymbol{h}_{v_j}}{\tau}\right)$$

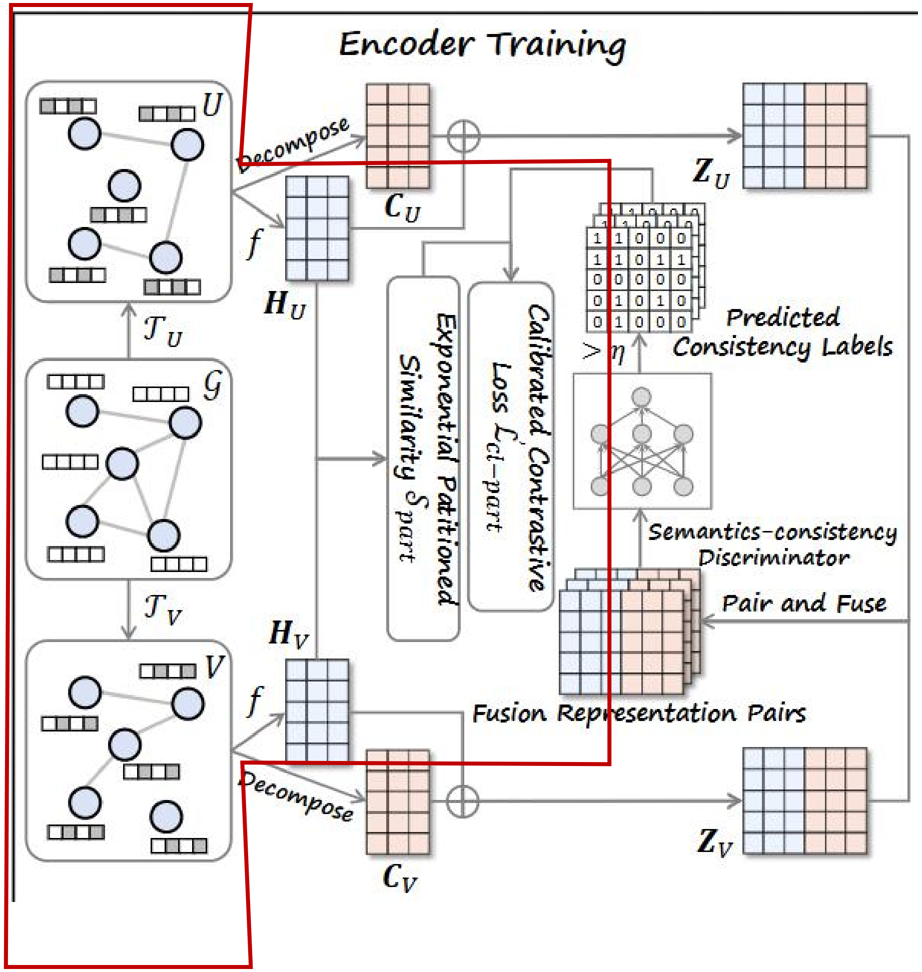
Method

Overview

- The discriminator calibrates the contrastive loss by correcting false positive or negative sample pair assignments caused by augmentation-induced semantic flips;
- The calibrated encoder generates more reliable representations to further enhance the discriminator in distinguishing the semantics consistency relationships.



Pre-train GNN Encoder



Implement augmentations T_U and T_V to obtain two augmented views U and V, which are processed by f to generate representations H_U and H_V.

$$H_U = f(U)$$
$$H_V = f(V)$$

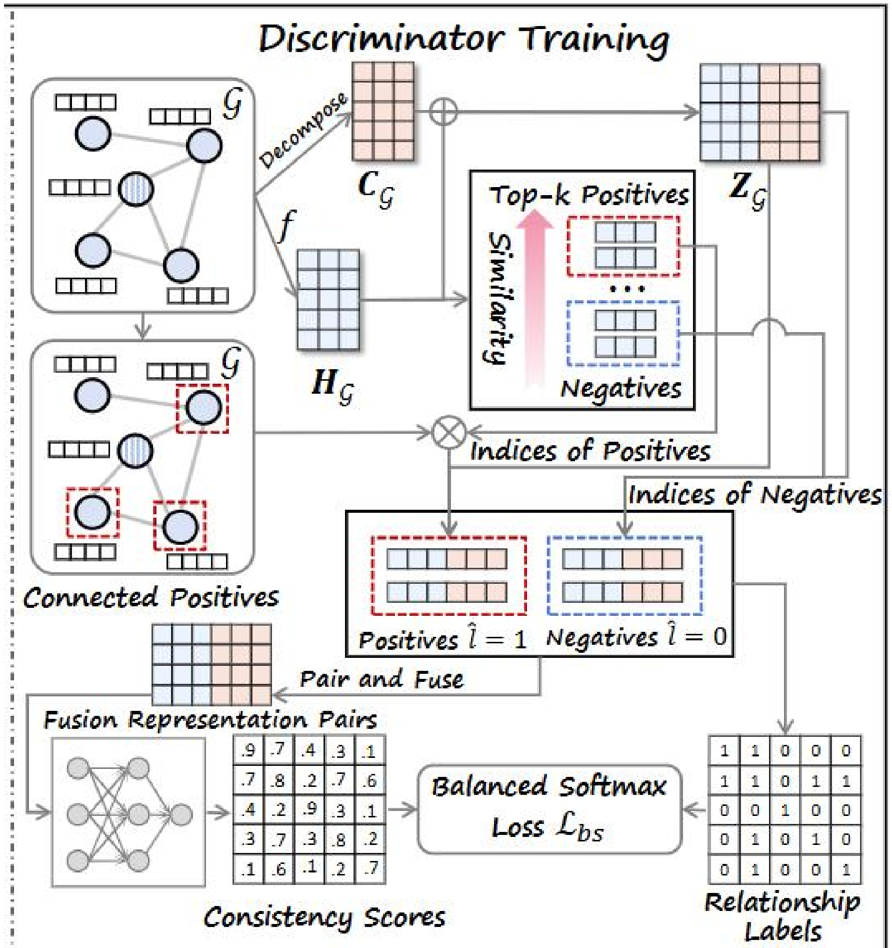
Exponential partitioned similarity:

$$\mathcal{S}_{part}(\boldsymbol{h}_{v_i}, \boldsymbol{h}_{v_j}) = \frac{1}{K}\sum_{k=1}^{K} \exp\left(\frac{K \cdot \boldsymbol{h}_{v_i}^{(k)\top} \boldsymbol{h}_{v_j}^{(k)}}{\tau}\right),$$

Partitioned contrastive loss:

$$\ell_{cl-part}(u_i) = -\log \frac{\mathcal{S}_{part}(\boldsymbol{h}_{u_i}, \boldsymbol{h}_{v_i})}{\sum_j \mathcal{S}_{part}(\boldsymbol{h}_{u_i}, \boldsymbol{h}_{v_j}) + \sum_{j \neq i} \mathcal{S}_{part}(\boldsymbol{h}_{u_i}, \boldsymbol{h}_{u_j})},$$

Pre-train Discriminator



Decompose the graph Laplacian matrix to obtain the eigenvectors as structural information:

$$\tilde{A} = U \Lambda U^{\top},$$

Feature encoding $\left(h_{v_i}, h_{v_j}\right)$ and structural encoding $\left(c_{v_i}, c_{v_j}\right)$ are concatenated to form a fusion representation pair

$$\left(z_{v_i} = h_{v_i} \parallel c_{v_i}, z_{v_j} = h_{v_j} \parallel c_{v_j}\right)$$

p_{v_i,v_j} is used to judge the semantic relationship between v_i and v_j

$$p_{v_i,v_j} = D(z_{v_i} \odot z_{v_j}),$$

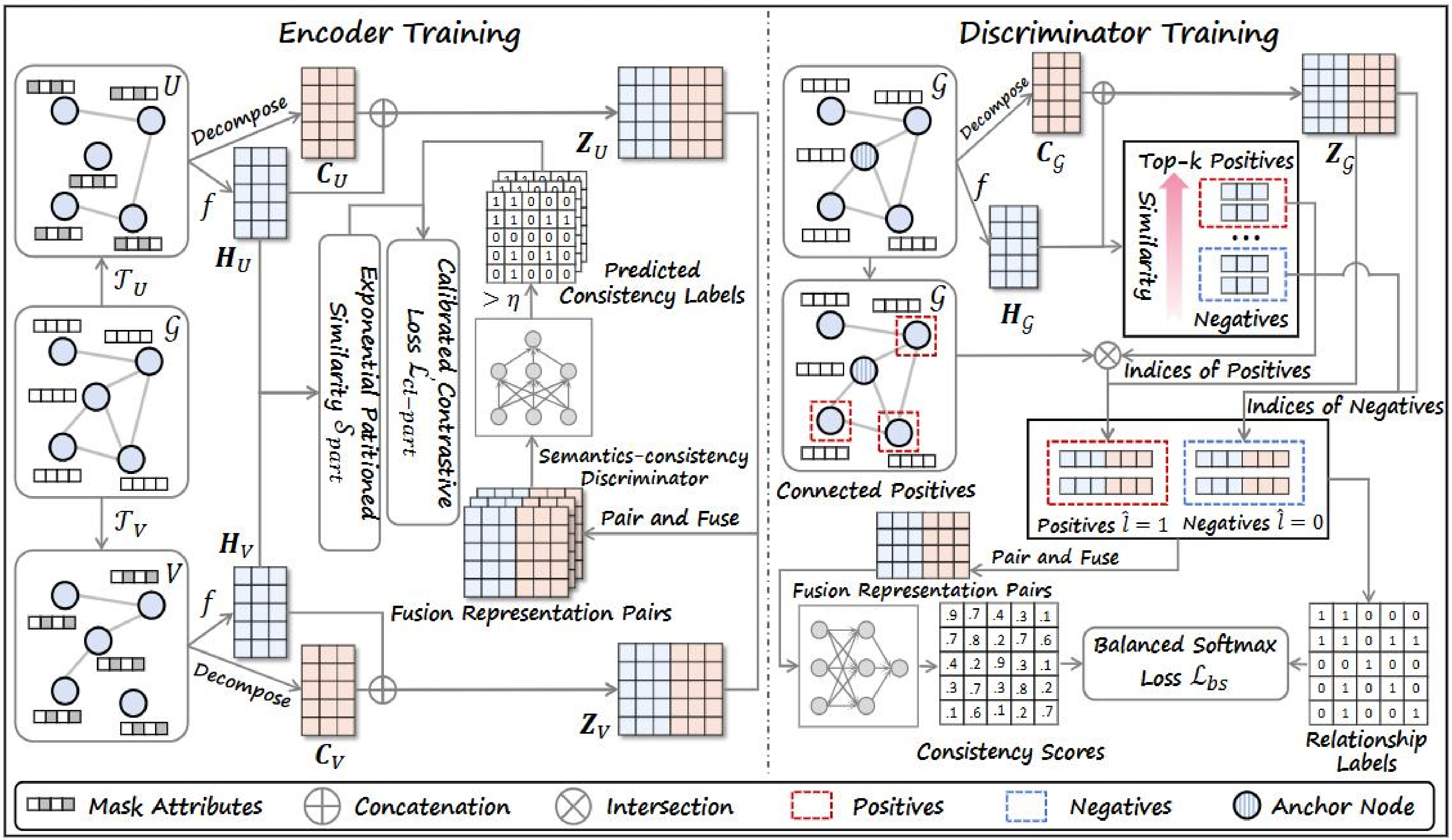
$$\tilde{l}_{v_i,v_j} = \begin{cases} 1, & \text{if } p_{v_i,v_j} \geq \eta \\ 0, & \text{if } p_{v_i,v_j} < \eta \end{cases}$$

Balanced Softmax loss is used for training the discriminator:

$$\mathcal{L}_{bs} = -\sum_{i,j} \log \frac{\mathbb{I}(\hat{l}_{v_i,v_j} = 1)\exp(p_{v_i,v_j} + \log N_{pos}) + \mathbb{I}(\hat{l}_{v_i,v_j} = 0)\exp(p_{v_i,v_j} + \log N_{neg})}{\exp(p_{v_i,v_j} + \log N_{pos}) + \exp(p_{v_i,v_j} + \log N_{neg})},$$

Fine-tuning Process

An alternating training strategy between GNN and discriminator.



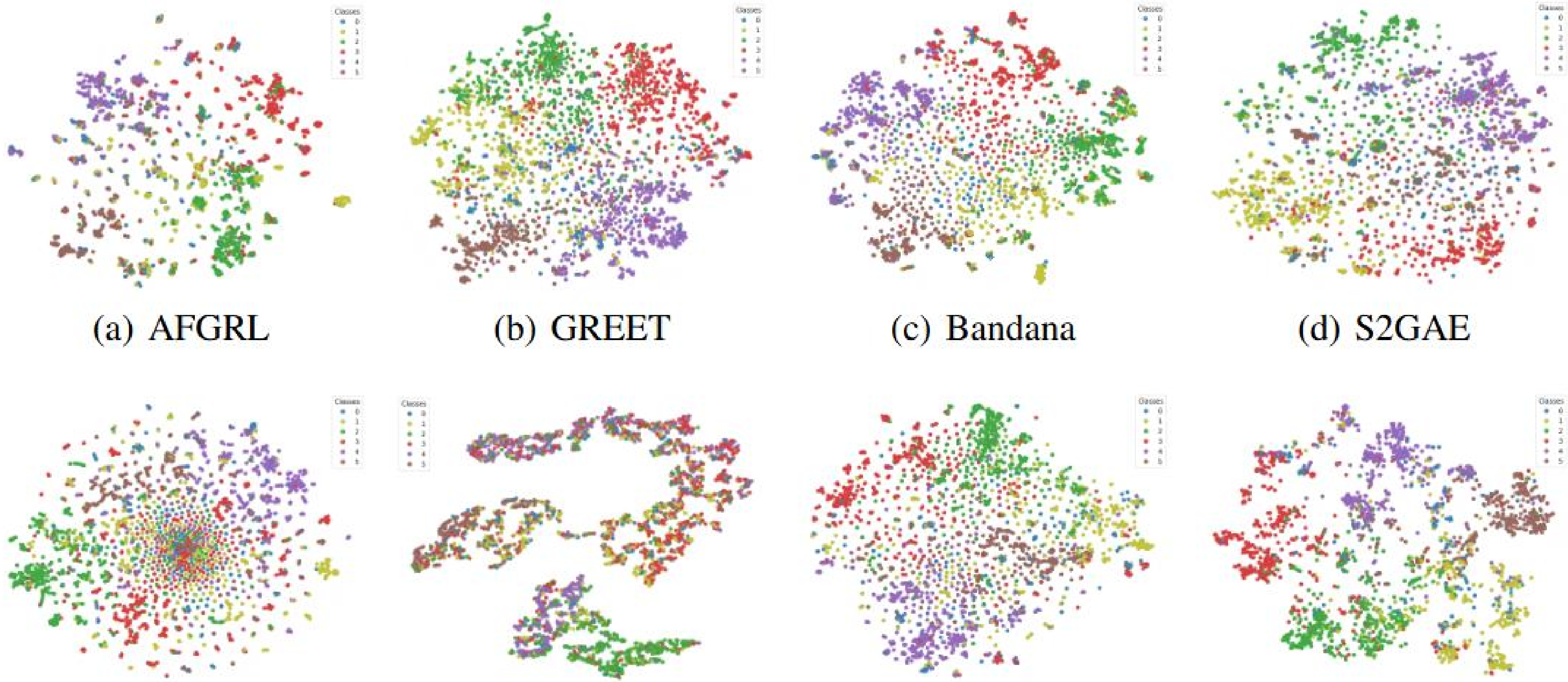
Experiments

Dataset	Cora	Citeseer	Pubmed	DBLP	Photo	Computers
AFGRL	83.34 ± 0.87	71.49 ± 0.78	85.21 ± 0.22	83.08 ± 0.14	93.22 ± 0.28	89.88 ± 0.33
SUGRL	85.37 ± 0.55	73.49 ± 0.66	86.68 ± 0.21	83.03 ± 0.23	93.20 ± 0.40	88.90 ± 0.20
NCLA	85.32 ± 0.44	73.24 ± 0.55	85.47 ± 0.40	83.97 ± 0.17	93.48 ± 0.22	89.14 ± 0.35
GREET	<u>85.70 ± 0.46</u>	73.26 ± 0.60	<u>86.95 ± 0.31</u>	83.81 ± 0.13	92.85 ± 0.31	87.94 ± 0.35
HomoGCL	85.40 ± 0.46	72.34 ± 0.40	86.31 ± 0.18	<u>84.39 ± 0.16</u>	93.25 ± 0.26	<u>90.09 ± 0.32</u>
S2GAE	85.46 ± 0.28	73.37 ± 0.47	86.46 ± 0.09	84.08 ± 0.00	93.50 ± 0.16	89.99 ± 0.14
iGCL	84.31 ± 0.43	72.85 ± 0.85	86.17 ± 0.25	83.74 ± 0.23	93.10 ± 0.26	90.06 ± 0.41
PiGCL	84.63 ± 0.78	<u>73.51 ± 0.64</u>	86.30 ± 0.20	84.30 ± 0.28	93.14 ± 0.30	89.25 ± 0.27
Bandana	85.48 ± 0.84	73.04 ± 0.75	86.35 ± 0.25	83.93 ± 0.02	93.44 ± 0.11	89.62 ± 0.09
SGRL	85.36 ± 0.23	73.20 ± 0.22	86.17 ± 0.01	84.12 ± 0.01	**93.83 ± 0.04**	90.02 ± 0.02
CaliGCL	**85.87 ± 0.62**	**74.13 ± 0.54**	**87.16 ± 0.19**	**85.32 ± 0.13**	<u>93.72 ± 0.24</u>	**90.79 ± 0.28**

Experiments

Datasets	Biochemical Molecules			Social Networks			
	NCI1	DD	MUTAG	COLLAB	RDT-B	RDT-M5K	IMDB-B
InfoGCL	76.2 ± 1.1	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	77.9 ± 0.4	78.6 ± 0.4	86.8 ± 1.4	71.4 ± 1.2	89.5 ± 0.8	56.0 ± 0.3	71.1 ± 0.4
JOAO	78.1 ± 0.5	77.3 ± 0.5	87.4 ± 1.0	69.5 ± 0.4	85.3 ± 1.4	55.7 ± 0.6	70.2 ± 3.1
JOAOv2	78.4 ± 0.5	77.4 ± 1.2	87.7 ± 0.8	69.3 ± 0.3	86.4 ± 1.5	56.0 ± 0.2	70.8 ± 0.3
AD-GCL	69.7 ± 0.5	74.5 ± 0.5	88.6 ± 1.3	73.3 ± 0.6	85.5 ± 0.8	53.0 ± 0.8	71.6 ± 1.0
SimGRACE	79.1 ± 0.4	77.4 ± 1.1	89.0 ± 1.3	71.7 ± 0.8	89.5 ± 0.9	55.9 ± 0.3	71.3 ± 0.8
SPAN	71.4 ± 0.5	75.8 ± 0.5	89.1 ± 0.8	75.0 ± 0.5	83.6 ± 0.6	54.1 ± 0.5	73.7 ± 0.7
GPA	<u>80.4 ± 0.4</u>	<u>79.9 ± 0.4</u>	<u>89.7 ± 0.8</u>	<u>76.2 ± 0.1</u>	89.3 ± 0.4	53.7 ± 0.2	<u>74.6 ± 0.4</u>
DRGCL	79.7 ± 0.4	78.4 ± 0.7	89.5 ± 0.6	70.6 ± 0.8	**90.8 ± 0.3**	<u>56.3 ± 0.2</u>	72.0 ± 0.5
CaliGCL	**80.8 ± 2.0**	**80.4 ± 2.5**	**90.4 ± 5.8**	**77.2 ± 1.7**	<u>90.3 ± 2.7</u>	**56.5 ± 1.5**	**76.9 ± 5.0**

Experiments



(a) AFGRL (b) GREET (c) Bandana (d) S2GAE

(e) iGCL (f) PiGCL (g) SGRL (h) CaliGCL

Figure 7: t-SNE visualization of CaliGCL and the comparative models on Citeseer dataset.