AAAI-25 / IAAI-25 / EAAI-25
FEBRUARY 25 – MARCH 4, 2025 | PHILADELPHIA, USA

Paper Code

Structural Entropy Guided Unsupervised Graph Out-Of-Distribution Detection

Yue Hou[1,2], He Zhu[1], Ruomei Liu[1], Yingke Su[2], Jinxiang Xia[1], Junran Wu[1★], Ke Xu[1]

[1]State Key Laboratory of Complex & Critical Software Environment, Beihang University [2]Shen Yuan Honors College, Beihang University

北京航空航天大學 BEIHANG UNIVERSITY

Introduction

With the emerging of huge amount of unlabeled data, unsupervised out-of-distribution (OOD) detection is vital for ensuring reliability of GNNs by identifying OOD samples from in-distribution ones during testing, where encountering novel or unknown data is inevitable.

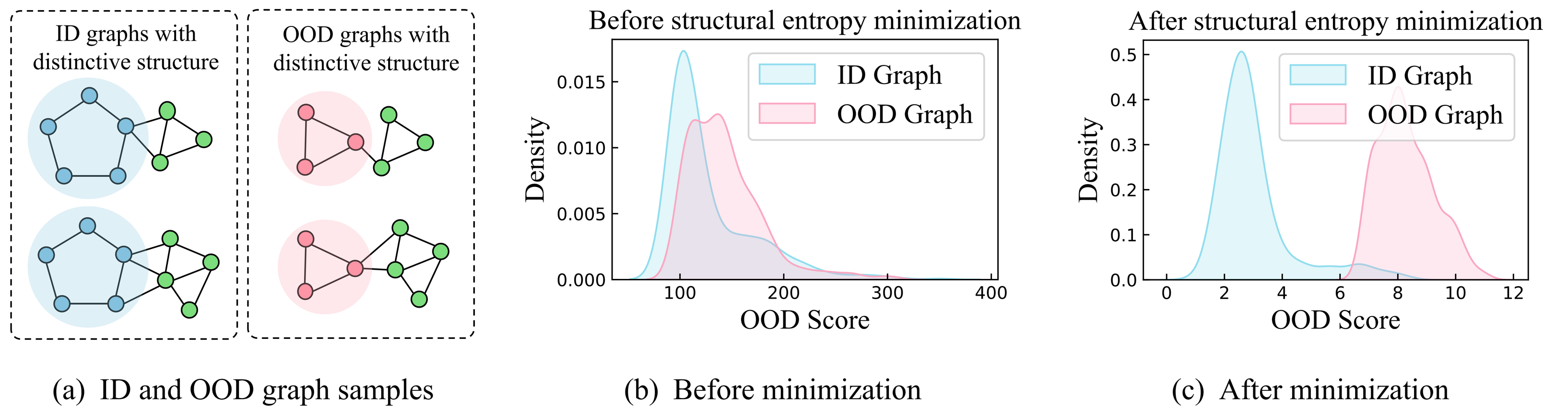

Figure 1. A toy example of ID and OOD graphs and scoring distributions before/after SE minimization.

A challenge still remains less explored: due to the prevalent presence of redundant information in graph structures, current methods struggle to **effectively capture and distinguish the essential structure** between ID and OOD data.

Preliminaries

Structural Entropy. Structural entropy of graph G on its coding tree T provides a hierarchical abstraction to measure the complexity of its structure:

$$\mathcal{H}^T(G) = -\sum_{v_\tau \in T} \frac{g_{v_\tau}}{vol(\mathcal{V})} \log \frac{vol(v_\tau)}{vol(v_\tau^+)}, \quad \mathcal{H}^{(k)}(G) = \min_{\forall T:\mathrm{Height}(T)=k} \{\mathcal{H}^T(G)\}, \tag{1}$$

Graph Contrastive Learning. Typical contrastive loss InfoNCE treats the same graph G_i in different views G_i^α and G_i^β as positive pairs and other nodes as negative pairs.

$$\ell(\mathbf{z}_i^\alpha, \mathbf{z}_i^\beta) = -\log \frac{e^{sim(\mathbf{z}_i^\alpha, \mathbf{z}_i^\beta)/\tau}}{\sum_{j=1, j\neq i}^N e^{sim(\mathbf{z}_i^\alpha, \mathbf{z}_j^\alpha)/\tau} + e^{sim(\mathbf{z}_i^\alpha, \mathbf{z}_j^\beta)/\tau}}, \quad \mathcal{L} = \frac{1}{2N}\sum_{i=1}^N \left[\ell(\mathbf{z}_i^\alpha, \mathbf{z}_i^\beta) + \ell(\mathbf{z}_i^\beta, \mathbf{z}_i^\alpha)\right]. \tag{2}$$

Our Proposed Framework: SEGO

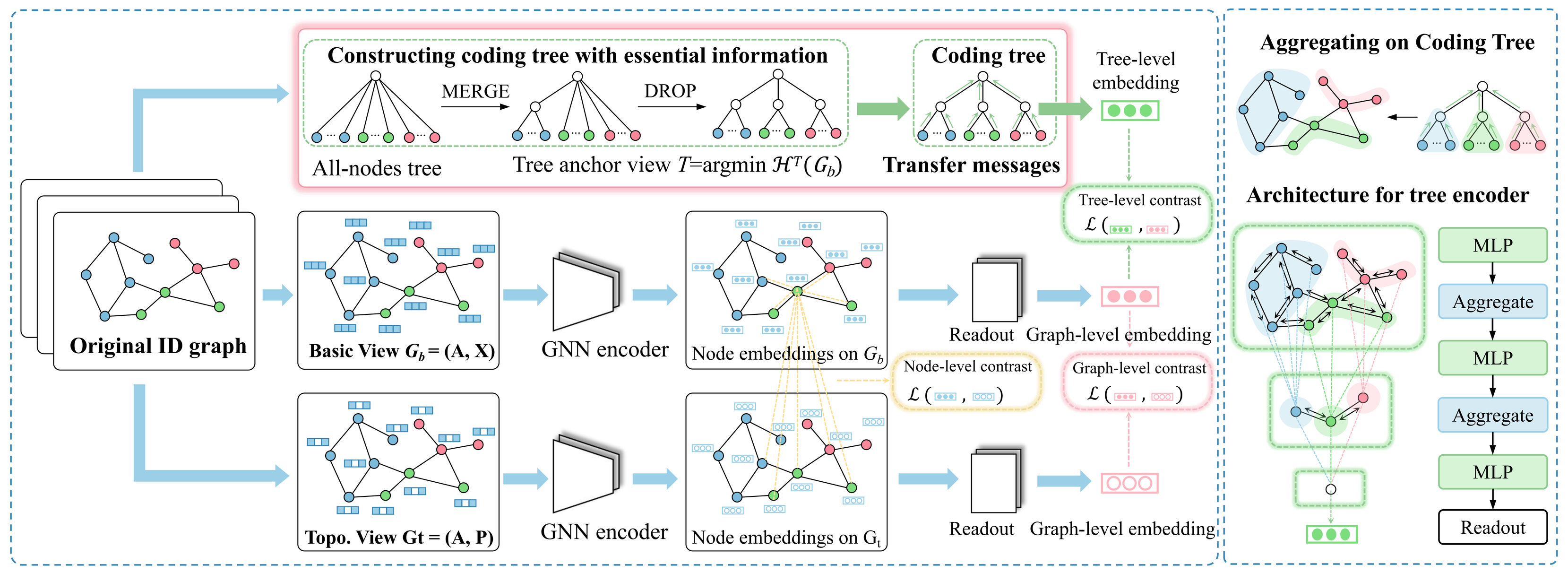

Figure 2. Overview of our proposed SEGO, which employs multi-grained contrast using triplet views.

- Guided by structural entropy theory, we propose a novel framework for unsupervised graph OOD detection, termed **SEGO**, which can remove redundant information and capture the essential structure of graphs, significantly improving model performance.
- To mitigate the information gap between node and graph embeddings, we employ a multi-grained contrastive learning scheme using triplet views, which includes coding tree as an anchor view and operates at local, global, and tree levels.
- Extensive experiments validate the effectiveness of SEGO, demonstrating superior performance over SOTA baselines in OOD detection.

Essential View with Redundancy-eliminated Information

Redundancy-eliminated Essential Information. According to graph information bottleneck theory (GIB), the objective is to generate an essential view that retains sufficient information while reducing uncertainty (i.e., redundant information) as much as possible.

$$\text{GIB: } \max I(f(G); y) - \beta I(G; f(G)) \Rightarrow \min I(G; f(G)), \tag{3}$$

Definition 1. *The anchor view with redundancy-eliminated essential information is supposed to be a distinctive substructure of the given graph. Let G^* be target anchor view of graph G:*

$$I(G^*; G) = \mathcal{H}(G^*) - \mathcal{H}(G^*|G). \tag{4}$$

Theorem 1. *The information in G^* is a subset of information in G (i.e., $\mathcal{H}(G^*) \subseteq \mathcal{H}(G)$); thus, we have $\mathcal{H}(G^*|G) = 0$, where $\mathcal{H}(G^*)$ is the structural entropy of G^*.*

Here, the mutual information between G and G^* can be rewritten as $I(G^*; G) = \mathcal{H}(G^*)$. Accordingly, to acquire the anchor view with essential information, we need to optimize:

$$\min I(G; f(G)) \Rightarrow \min \mathcal{H}(G^*). \tag{5}$$

Thus, we argue that the view obtained by minimizing structural entropy of a given graph represents the redundancy-eliminated information that retains distinctive substructure.

Maximum Effective Mutual Information. With the essential view eliminating redundant information, we also theoretically prove that our SEGO effectively captures the maximum mutual information between the representations obtained from the anchor view and labels.

Theorem 2. *Optimizing the contrastive loss is equivalent to maximizing $I(f(G^*); f(G))$, leading to the maximization of $I(f(G^*); y)$.*

Redundancy-aware Multi-grained Triplet Contrastive Learning (1)

Instantiation of Triplet Views. We first treat a given graph $G_b = (\mathbf{A}, \mathbf{X})$ as a basic view to directly learn from the original input of the ID data. From this basic view, we construct an anchor view of the graph (▷Eq. (5)). The total process can be divided into two steps: 1) construction of the full-height binary coding tree and 2) compression of the binary coding tree to height k. We design two efficient operators, MERGE and DROP, to construct minimal structural entropy coding tree T with fixed height k.

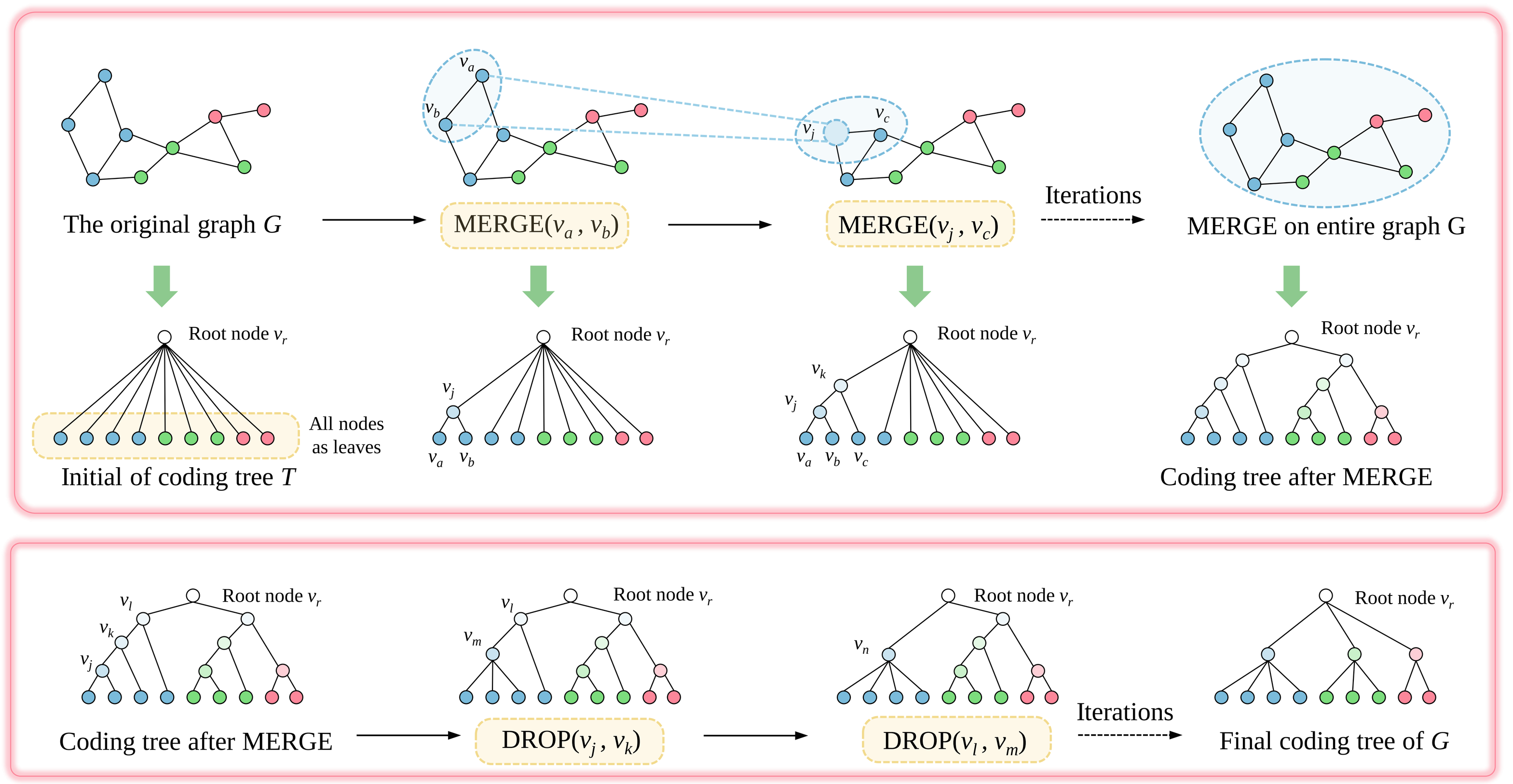

Figure 3. Overview of operators MERGE and DROP.

Definition 2. *Assuming v_a and v_b as two child nodes of root node v_r, $\mathrm{MERGE}(v_a, v_b)$ is defined as adding a new node v_j as the child of v_r and the parent of v_a and v_b:*

$$v_j.children = \{v_a, v_b\}, \quad v_r.children = \{v_j\} \cup v_r.children, \tag{6}$$

$$(v_a, v_b) = argmax\{\mathcal{H}^T(G) - \mathcal{H}^{T_{ab}}(G) | v_a, v_b \in v_r.children\}. \tag{7}$$

Redundancy-aware Multi-grained Triplet Contrastive Learning (2)

Definition 3. *Given node v_m and its parent node v_m^+ in T, the operator $\mathrm{DROP}(v_m)$ is defined as adding the children of v_m and itself to the child set of v_m^+:*

$$v_m^+.children = v_m^+.children \cup v_m.children, \tag{8}$$

$$v_m = argmin\{\mathcal{H}^{T_m}(G) - \mathcal{H}^T(G) | v_m \in T, v_m \neq v_r, v_m \notin \mathcal{V}\}. \tag{9}$$

This tree anchor view $T = \arg\min \mathcal{H}^T(G_b)$ effectively removes redundant information from graphs while preserving distinctive essential structural information.

Multi-grained Contrastive Learning Objectives. We employ a multi-grained contrastive learning scheme that extracts features at three distinct levels: the **local-level** for fine-grained feature extraction, the **global-level** for coarse-grained feature extraction, and the **tree-level** for capturing essential information of the entire graph (▷Figure 2).

During the training phase, we introduce the standard deviation of prediction errors to adaptively adjust the balance of local and global information. The overall loss is calculated by:

$$\mathcal{L} = \mathcal{L}_{tree} + \sigma_l^\theta \mathcal{L}_{local} + \sigma_g^\theta \mathcal{L}_{global}, \tag{10}$$

where σ_l and σ_g are the standard deviations of predicted errors of the node and graph levels, respectively. During the inference phase, we employ z-score normalization based on the mean values and standard deviations of the predicted errors of training samples: $s_{G_i} = \frac{s_l - \mu_l}{\sigma_l} + \frac{s_g - \mu_g}{\sigma_g}$, where μ_l and μ_g represent the mean values of predicted errors for training samples.

Experiments (Partial)

ID dataset	BZR	PTC-MR	AIDS	ENZYMES	IMDB-M	Tox21	FreeSolv	BBBP	ClinTox	Esol	A.A.	A.R.
OOD dataset	COX2	MUTAG	DHFR	PROTEIN	IMDB-B	SIDER	ToxCast	BACE	LIPO	MUV		
PK-LOF	42.22±8.39	51.04±6.04	50.15±3.29	50.47±2.87	48.03±2.53	51.33±1.81	49.16±3.70	53.10±2.07	50.00±2.17	50.82±1.48	49.63	12.9
PK-OCSVM	42.55±8.26	49.71±6.58	50.17±3.30	50.46±2.78	48.07±2.41	51.33±1.81	48.82±3.29	53.05±2.10	50.06±2.19	51.00±1.33	49.52	12.8
PK-iF	51.46±1.62	54.29±4.33	51.10±1.43	51.67±2.69	50.67±2.47	49.87±0.82	52.28±1.87	51.47±1.33	50.81±1.10	50.85±3.51	51.45	11.1
WL-LOF	48.99±6.20	53.31±8.98	50.77±2.87	52.66±2.47	52.28±4.50	51.92±1.58	51.47±4.23	52.80±1.91	51.29±3.40	51.26±1.31	51.68	10.4
WL-OCSVM	49.16±4.51	53.31±7.57	50.98±2.71	51.77±2.21	51.38±2.39	51.08±1.46	50.38±3.81	52.85±2.00	50.77±3.69	50.97±1.65	51.27	11.1
WL-iF	50.24±2.49	51.43±2.02	50.10±0.44	51.17±2.01	51.07±2.25	50.25±0.96	52.60±2.38	50.78±0.75	50.41±2.17	50.61±1.96	50.87	12.4
OCGIN	76.66±4.17	80.38±6.84	86.01±6.59	57.65±2.96	67.93±3.86	46.09±1.66	59.60±4.78	61.21±8.12	49.13±4.13	54.04±5.50	63.87	7.9
GLocalKD	75.75±5.99	70.63±3.54	93.67±1.24	57.18±2.03	78.25±4.35	66.28±0.98	64.82±3.31	73.15±1.26	55.71±3.81	86.83±2.35	72.23	5.1
InfoGraph-iF	63.17±9.74	51.43±5.19	93.10±1.35	60.00±1.83	58.73±1.96	56.28±0.81	56.92±1.69	53.68±2.90	48.51±1.87	54.16±5.14	59.60	8.5
InfoGraph-MD	86.14±6.77	50.79±8.49	69.02±11.67	55.25±3.51	**81.38±1.14**	59.97±2.06	58.05±5.46	70.49±4.63	48.12±5.72	77.57±1.69	65.68	7.4
GraphCL-iF	60.00±3.81	50.86±4.30	92.90±1.21	61.33±2.27	59.67±1.65	56.81±0.97	55.55±2.71	59.41±3.58	47.84±0.92	62.12±4.01	60.65	8.7
GraphCL-MD	83.64±6.00	73.03±2.38	93.75±2.13	52.87±6.11	79.09±2.73	58.30±1.52	60.31±5.24	75.72±1.54	51.58±3.64	78.73±1.40	70.70	5.3
GOOD-D$_{simp}$	93.00±3.20	78.43±2.67	98.91±0.41	61.89±2.51	79.71±1.19	65.30±1.27	70.48±2.75	81.56±1.97	66.13±2.98	91.39±0.46	78.68	3.2
GOOD-D	94.99±2.25	81.21±2.65	99.07±0.40	61.84±1.94	79.94±1.09	66.50±1.35	80.13±3.43	82.91±2.58	69.18±3.61	91.52±0.70	80.73	2.2
SEGO	**96.66±0.91**	**85.02±0.94**	**99.48±0.11**	**64.42±4.95**	80.27±0.92	**66.67±0.82**	**90.95±1.93**	**87.55±0.13**	**78.99±2.81**	**94.59±0.94**	**84.46**	**1.1**

Table 1. OOD detection results in terms of AUC (%, mean ± std).

$\mathcal{L}_{tree}$	$\mathcal{L}_{global}$	$\mathcal{L}_{local}$	BZR	PTC-MR	AIDS	ENZYMES	IMDB-M	Tox21	FreeSolv	BBBP	ClinTox	Esol
			COX2	MUTAG	DHFR	PROTEIN	IMDB-B	SIDER	ToxCast	BACE	LIPO	MUV
✓	-	-	54.79±4.08	58.20±3.87	43.68±7.36	49.26±1.11	49.56±5.76	49.26±5.10	49.89±2.95	50.53±0.63	51.97±4.58	54.49±3.57
-	✓	-	87.44±4.66	77.84±3.71	97.60±1.05	56.74±1.96	75.22±1.91	65.07±1.32	78.40±6.44	77.66±2.29	70.11±2.44	89.57±2.80
-	-	✓	83.51±4.14	72.48±3.77	96.84±0.58	60.85±2.95	79.34±1.81	62.58±0.67	59.48±2.20	69.53±2.29	53.29±4.32	86.49±1.20
✓	✓	-	87.27±8.21	**87.71±1.35**	97.97±0.04	54.82±2.74	74.51±1.52	64.84±0.29	89.34±0.06	**88.34±1.64**	**79.21±4.55**	94.13±1.32
✓	-	✓	79.36±8.69	55.08±1.29	90.66±3.40	63.38±4.18	72.96±3.73	55.68±2.67	61.01±5.29	70.13±0.26	52.14±2.58	77.78±1.01
-	✓	✓	86.29±1.09	77.53±4.03	98.23±0.19	61.55±1.47	75.27±0.54	65.44±1.14	88.04±1.15	80.43±2.58	65.89±4.58	90.94±1.17
✓	✓	✓	**96.66±0.91**	85.02±0.94	**99.48±0.11**	**64.42±4.95**	**80.27±0.92**	**66.67±0.82**	**90.95±1.93**	87.55±0.13	78.99±2.81	**94.59±0.94**

Table 2. Ablation study results of SEGOand its variants in terms of AUC (%, mean ± std).

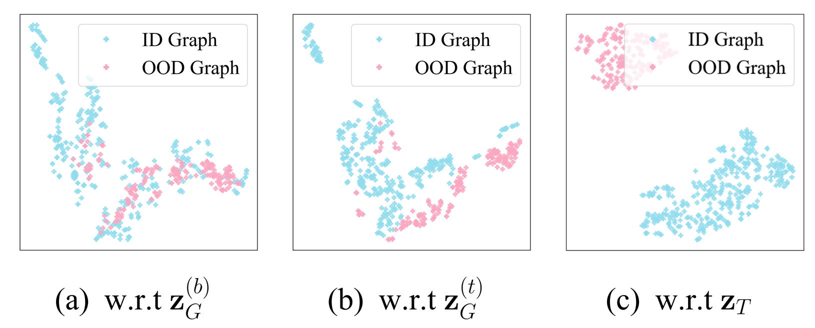

Figure 4. T-SNE visualization of embeddings.

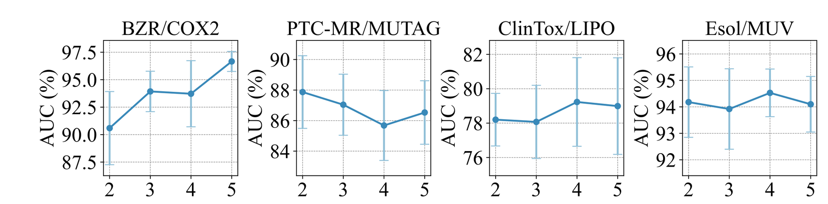

Figure 5. Natural hierarchy of graph.

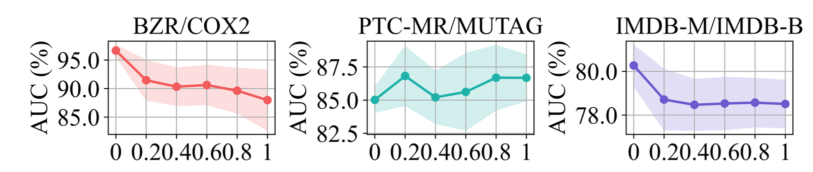

Figure 6. Sensitivity of self-adaptiveness strength θ.