Redundancy-Aware Test-Time Graph Out-of-Distribution Detection

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

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

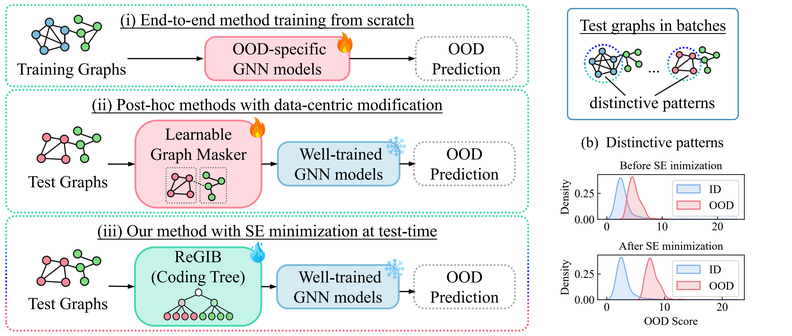
NEURAL INFORMATION PROCESSING SYSTEMS

Paper Code

北京航空航天大學 BEIHANG UNIVERSITY

Introduction

Distributional discrepancy between training and test data can lead models to make inaccurate predictions when encountering out-of-distribution (OOD) samples in real-world applications. Although existing graph OOD detection methods leverage *data-centric* techniques to extract effective representations, their performance remains compromised by structural redundancy that induces semantic shifts.



(a) Comparison between RedOUT and other methods (b) Distinctive patterns (c) Density distributions

Figure 1. (a) A schema comparison. (b) A toy example of distinctive components within test graphs. (c) Scoring distributions before/after structural entropy (*abbr.* SE) minimization.

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

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

Our Proposed Framework: RedOUT

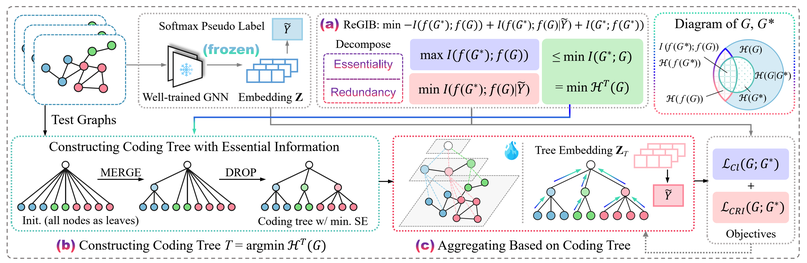


Figure 2. Overview of the overall framework of RedOUT and proposed ReGIB. (a) the ReGIB principle and its bounds. (b) obtained by minimizing structural entropy of graph G, coding tree T serves as an instantiation of the essential view. (c) message passing and aggregating on graph are under the guidance of the coding tree.

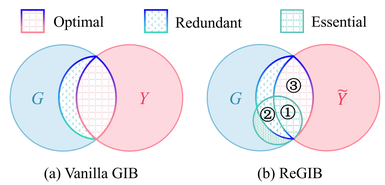
- **Novel Framework:** We propose **RedOUT**, **the first trial** to endow trained models with the ability to capture the essential structural information of graphs during test-time.
- **ReGIB:** Decouples **essential and redundant information** via a coding tree with minimized SE instantiated as essential view. Derive **upper and lower bounds** for optimization.
- **Results:** Extensive experiments validate the effectiveness of RedOUT, demonstrating the **superior performance** over SOTA baselines in unsupervised OOD detection.

Proposition 1 (Lower Bound of $I(f(G^*); \tilde{Y})$).

$$I(f(G^*); \tilde{Y}) = I(f(G^*); f(G)) + I(f(G^*); \tilde{Y}|f(G)) - I(f(G^*); f(G)|\tilde{Y}) \geq \underbrace{I(f(G^*); f(G))}_{\text{Eq. (4) and (8)}} - \underbrace{I(f(G^*); f(G)|\tilde{Y})}_{\text{Eq. (5) and (9)}}. \tag{2}$$

The Principle of ReGIB

Compared to vanilla GIB, the proposed ReGIB accounts for the fact that model trained on ID graphs, leading to the representations containing both the **optimal components** and **irrelevant redundancy**.



(a) Vanilla GIB (b) ReGIB

Figure 3. Comparison between the basic GIB and proposed ReGIB. Note ReGIB balances signals ①, ② with optimal $f(G^*)$ to capture the essential information and discard redundancy.

Definition 1 (Redundancy-aware Graph Information Bottleneck). Given a test graph G, and the pseudo-label $\tilde{Y}$ predicted by pre-trained f, ReGIB aims to capture the distinctive structure G^* with essential information and further learn the optimal representation $f(G^*)$ that satisfies:

$$\min \text{ReGIB} \triangleq -I(f(G^*); f(G)) + I(f(G^*); f(G)|\tilde{Y}) + \beta I(G^*; f(G^*)). \tag{3}$$

Proposition 2 (Lower Bound of $I(f(G^*); f(G))$).

$$I(f(G^*); f(G)) \geq -\mathcal{L}_{CI}(G^*, G) + \log(N). \tag{4}$$

Proposition 3 (Upper Bound of $I(f(G^*); f(G)|\tilde{Y})$). For any $\mathbb{Q}(f(G)|\tilde{Y})$, the variational upper bound of conditional mutual information $I(f(G^*); f(G)|\tilde{Y})$ is:

$$I(f(G^*); f(G)|\tilde{Y}) \leq \mathbb{E}_{\mathbb{P}}\left[\log \frac{\mathbb{P}(f(G), f(G^*)|\tilde{Y})}{\mathbb{P}(f(G^*)|\tilde{Y})\mathbb{Q}(f(G)|\tilde{Y})}\right]. \tag{5}$$

Redundancy-eliminated Essential Information. Suppose the optimal view G^* can capture the essential information while eliminating redundancy of graph G. The mutual information between G and G^* can be formulated as $I(G^*; G) = \mathcal{H}(G^*) - \mathcal{H}(G^*|G)$.

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: $\mathcal{H}(G^*|G) = 0$. Here, the mutual information between G and G^* can be rewritten as:

$$I(G^*; G) = \mathcal{H}(G^*). \tag{6}$$

Proposition 4 (Upper Bound of $I(G^*; f(G^*))$) Given the tuple $(G, G^*, f(G^*))$, the learning procedure follows the Markov Chain $< G \to G^* \to f(G^*) >$. Accordingly, to acquire the essential view with essential information, we need to optimize:

$$\min I(G^*; f(G^*)) \leq \min I(G^*; G) = \min \mathcal{H}(G^*). \tag{7}$$

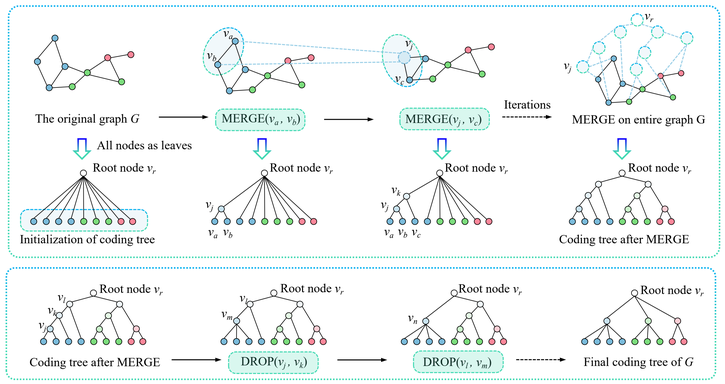


Figure 4. Overview of operators **MERGE** and **DROP** for coding tree construction.

Coding Tree Construction (Instantiation for G^*). The coding tree, as a hierarchical abstraction of the original graph structure, can be considered a contrastive view $T = \arg\min \mathcal{H}^T(G)$, which serves as an instantiation of the essential information in graph G.

Correspondingly, we design two efficient operators, **MERGE** and **DROP**, to construct coding tree T with minimum structural entropy (in Figure 4).

ReGIB Principle Instantiation

Instantiation for $I(f(G^*); f(G))$. Based on the lower bound derived in Eq. (4), the mutual information between the essential view G and basic view G is estimated by:

$$I(f(G^*); f(G)) \doteq -\mathcal{L}_{CI}(G^*, G). \tag{8}$$

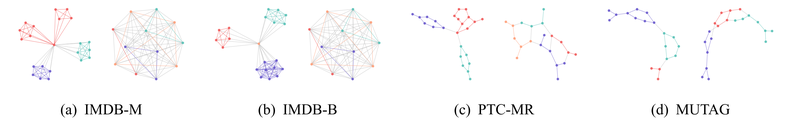
Instantiation for $I(f(G^*); f(G)|\tilde{Y})$. To specify the variational upper bound, we treat the similarity between $\mathbf{Z}_T$ and $\mathbf{Z}$ given the predicted $\tilde{Y} = \arg\max(\text{softmax}(\mathbf{Z}))$ as an approximation of the log-likelihood $\log \mathbb{P}(f(G)|f(G^*), \tilde{Y})$, and set $\mathbb{Q}(f(G)|\tilde{Y}) = \mathbb{E}_{\mathbf{Z}^- \sim \mathbb{P}(f(G)|\tilde{Y})} e^{\text{sim}(\mathbf{Z}^-, \mathbf{Z}_T)}$. Thus $I(f(G^*); f(G) \mid \tilde{Y})$ can be approximately instantiated as $I(f(G^*); f(G)|\tilde{Y}) \doteq \mathcal{L}_{CRI}(G, G^*)$, where $\mathcal{L}_{CRI}$ is conditional redundancy-eliminated loss, *i.e.*,

$$\mathcal{L}_{CRI}(G, G^*) \triangleq \mathbb{E}_{\tilde{y} \sim \mathbb{P}(\tilde{Y})} \mathbb{E}_{\mathbf{Z}, \mathbf{Z}_T \sim \mathbb{P}(f(G), f(G^*)|\tilde{y})} \left[sim(\mathbf{Z}, \mathbf{Z}_T) - \log \mathbb{E}_{\mathbf{Z}^- \sim \mathbb{P}(f(G)|\tilde{y})} e^{sim(\mathbf{Z}^-, \mathbf{Z}_T)} \right]. \tag{9}$$

Experiment

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

ID dataset / OOD dataset	BZR / COX2	PTC-MR / MUTAG	AIDS / DHFR	ENZYMES / PROTEIN	IMDB-M / IMDB-B	Tox21 / SIDER	FreeSolv / ToxCast	BBBP / BACE	ClinTox / LIPO	Esol / MUV	A.A.	A.R.
Non-deep Two-step Methods												
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	14.9
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	14.8
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	12.9
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	12.1
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	13.0
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	14.2
Deep Two-step Methods												
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	9.9
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	8.5
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	10.2
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	6.3
End-to-end Training Methods												
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	9.3
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	6.1
GOOD-$D_{\textit{vimp}}$	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	3.4
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	2.4
Test-time and Data-centric Methods												
GTrans	55.17±5.04	62.38±2.36	60.12±1.98	49.94±5.67	51.55±2.90	61.67±0.73	50.81±3.03	64.02±2.10	58.54±2.38	76.31±3.85	59.05	10.0
AAGOD$_S$+	76.75	–	–	66.22	59.00	64.26	–	67.80	–	–	–	–
AAGOD$_L$+	76.00	–	–	65.89	62.70	57.59	–	57.13	–	–	–	–
GOODAT	82.16±0.15	81.84±0.57	96.43±0.25	66.29±1.54	79.03±0.03	68.92±0.01	68.83±0.02	77.07±0.03	62.46±0.54	85.91±0.27	76.89	3.9
RedOUT	**95.06±0.54**	**94.45±1.66**	**99.98±0.16**	**66.75±2.02**	79.54±0.72	**71.67±0.50**	**92.97±0.84**	**92.60±0.23**	**86.56±0.76**	**95.00±0.54**	**87.46**	**1.3**
Improve	△+0.07	△+12.61	△+0.91	△+0.46	▽	△+2.75	△+12.84	△+9.69	△+17.38	△+3.48	△+6.73	△



(a) IMDB-M (b) IMDB-B (c) PTC-MR (d) MUTAG

Figure 5. Case study with visualization of the essential structure based on the coding tree preserved by RedOUT.

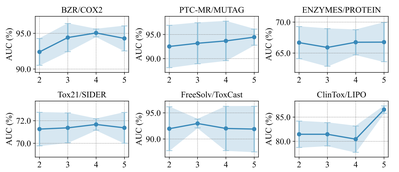


Figure 6. The natural hierarchy of graph.

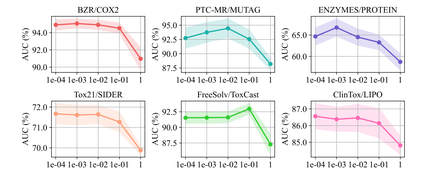
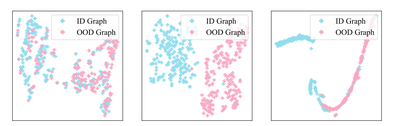


Figure 7. The sensitivity of hyperparameter λ.



(a) w.r.t $\mathbf{Z}$ (b) w.r.t $\mathbf{Z}_T$ (c) w.r.t GOODAT

Figure 8. T-SNE visualization of embeddings.

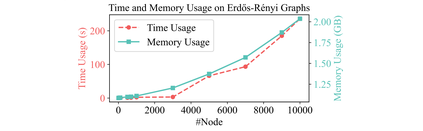


Figure 9. Scalability of coding tree construction on Erdős-Rényi graphs with $|\mathcal{E}| = 2|\mathcal{V}|$.