TIEGGAD: TARGET-DOMAIN INFORMATION ENHANCEMENT FOR GENERALIST GRAPH ANOMALY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Generalist Graph Anomaly Detection (GGAD) aims to detect anomalies in unseen target graphs using a detector trained on source graphs. However, cross-graph discrepancies may limit the detector's cross-graph generalization ability, leaving anomaly-relevant information in the target domain insufficiently captured. We summarize this bottleneck as Target-domain Anomaly Information Deficiency (TAID). To mitigate TAID, we propose Target-domain Information Enhancement for Generalist Graph Anomaly Detection (TIEGGAD), a multi-scale structural and cross-view hypergraph enhancement framework for zero-shot GGAD. To strengthen structural modeling and complement semantic information, we construct a degree-guided multi-scale structural representation to capture structural deviations across different neighborhood scales. Furthermore, to refine the overall representation using target-domain information, we introduce a training-free cross-view hypergraph module that enables bidirectional information exchange between the semantic and structural views. Finally, parameterized vector-level integration is used to combine semantic, structural, and hypergraph-enhanced representations into a unified representation. The proposed framework requires neither target-domain labels nor retraining. Experiments on multiple real-world graph datasets demonstrate stable and competitive performance for zero-shot cross-domain node anomaly detection, while ablation and mechanism analyses further verify the effectiveness of multi-scale structural modeling and cross-view hypergraph enhancement.

1 INTRODUCTION

Graph Anomaly Detection (GAD) identifies anomalous nodes, edges, subgraphs, or entire graphs whose properties or relational patterns deviate substantially from normal behavior. Node-level GAD focuses on nodes with abnormal attributes or connectivity patterns, with applications spanning financial fraud detection, malicious-account discovery, and cybersecurity (Qiao et al., 2025b; Liu et al., 2022a). Most existing GAD methods follow a "one graph, one model" paradigm, in which a detector is trained or adapted separately for each target graph, which is difficult to scale under scarce labels, rapidly changing scenarios, and strict deployment-cost constraints. Therefore, Generalist Graph Anomaly Detection (GGAD) instead aims to detect anomalies in unseen target graphs using a detector trained on source graphs. Representative recent methods include ARC (Liu et al., 2024), AnomalyGFM (Qiao et al., 2025a), IA-GGAD (Zhang et al., 2025a), and so on. These methods span feature alignment, prototype learning, invariant modeling, and residual representations.

However, cross-graph discrepancies in feature semantics, structural patterns, and anomaly mechanisms may limit the detector's generalization to unseen targets, leaving anomaly-relevant information insufficiently captured. We summarize this bottleneck as Target-domain Anomaly Information Deficiency (TAID). Under zero-shot GGAD, the source-trained detector remains frozen during target inference, while unseen targets may contain anomaly-relevant information not sufficiently represented by the source-trained mapping. Addressing this deficiency therefore requires fuller use of target-graph information without target labels or retraining.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

To address TAID, we propose TIEGGAD, a multi-scale structural and cross-view hypergraph enhancement framework for zero-shot GGAD. To strengthen structural modeling and complement semantic information, TIEGGAD constructs a degree-guided multi-scale structural representation and learns a structural reconstruction branch to capture structural deviations across neighborhood scales. Furthermore, TIEGGAD introduces a training-free cross-view hypergraph module, where hyperedges capture semantic and structural group-level relations beyond pairwise local connections and provide broader context for locally ambiguous deviations. The module enables bidirectional information exchange using only target-observable information, without target labels, retraining, or parameter updates. Finally, parameterized vector-level integration combines semantic, structural, and hypergraph-enhanced representations into a unified representation for anomaly scoring.

Our contributions are summarized as follows:

- We propose TIEGGAD, a multi-scale structural and cross-view hypergraph enhancement framework for zero-shot GGAD, which addresses Target-domain Anomaly Information Deficiency (TAID) without target-domain labels or retraining.
- We design a degree-guided multi-scale structural reconstruction branch to capture structural deviations across neighborhood scales and complement semantic information.
- We introduce a training-free cross-view hypergraph module that captures semantic and structural group-level relations and enables bidirectional information exchange between the two views, thereby enriching target-domain anomaly representations.

2 Related Work

Node-level GAD identifies nodes whose attributes or connectivity depart from normal graph patterns. Representative approaches include AnomalyDAE, CoLA, and TAM, which use reconstruction or self-supervised/structure-aware signals but are generally optimized for individual graphs and do not directly address cross-graph deployment (Qiao et al., 2025b; Liu et al., 2022a; Fan et al., 2020; Liu et al., 2022b; Qiao & Pang, 2023). GGAD instead trains on source graphs and applies a frozen detector to unseen targets. Existing approaches improve transfer through feature or representation alignment, transferable prompts or prototypes, and residual modeling: OWLEYE and ReFi-GAD align cross-graph representations; UNPrompt, AnomalyGFM, and ProMoS construct transferable references; ARC, DR-GGAD, and IA-GGAD model transferable residual deviations; FreeGAD provides a training-free alternative (Zheng et al., 2026; Liu et al., 2026; Niu et al., 2025; Qiao et al., 2025a; Xu et al., 2026; Liu et al., 2024; Fu et al., 2026; Zhang et al., 2025a; Zhao et al., 2025). Despite this progress, frozen target inference still relies mainly on semantic representations, transferable references, or local relations, leaving explicit multi-scale structural statistics and group-level relations comparatively underused.

3 Problem Formulation

3.1 Notation and Preliminaries

Let $\mathcal{G}$ denote an attributed graph with node set $\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$, adjacency matrix $\mathbf{A} \in \{0,1\}^{N\times N}$, and node-attribute matrix $\mathbf{X} \in \mathbb{R}^{N\times d}$; the i-th row of $\mathbf{X}$ is the d-dimensional attribute of v_i. For a labeled graph, $\mathbf{y} \in \{0,1\}^N$ denotes the node labels, where $y_i = 0$ and $y_i = 1$ indicate a normal and an anomalous node, respectively.

3.2 Generalist Graph Anomaly Detection Task

Let $\mathcal{D}_s = \{(\mathcal{G}_s^{(m)}, \mathbf{y}_s^{(m)})\}_{m=1}^{M}$ denote M labeled source graphs, where m indexes the sources and $\mathbf{y}_s^{(m)}$ is the node-label vector of $\mathcal{G}_s^{(m)}$. GGAD trains a detector f_{θ_s}, parameterized by θ_s, on $\mathcal{D}_s$ and keeps it frozen for an unseen target graph $\mathcal{G}_t$ with node set $\mathcal{V}_t$, using no target labels or parameter updates. Given an unlabeled reference set $\mathcal{S}_t \subset \mathcal{V}_t$, the query set is $\mathcal{Q} = \mathcal{V}_t \setminus \mathcal{S}_t$, and the detector outputs $\mathbf{s}_t = f_{\theta_s}(\mathcal{G}_t, \mathcal{S}_t) \in \mathbb{R}^{|\mathcal{Q}|}$, where larger $s_{t,i}$ indicates higher anomaly likelihood for $v_i \in \mathcal{Q}$.

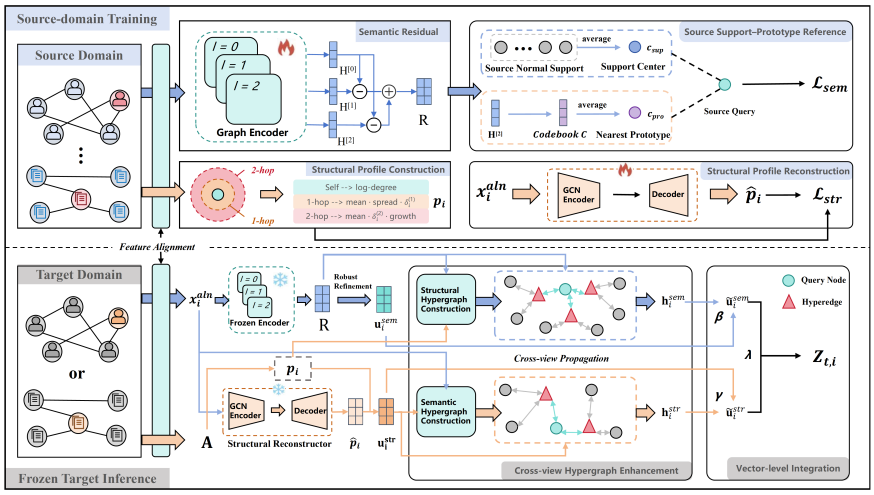


Figure 1: Overview of TIEGGAD. Source-domain training learns the semantic residual and multi-scale structural reconstruction branches. During frozen target inference, the cross-view hypergraph exchanges information between the semantic and structural views, followed by vector-level integration.

3.3 Target-domain Anomaly Information Deficiency

With θ_s learned only from $\mathcal{D}_s$ and frozen during target inference, cross-graph discrepancies can leave anomaly-relevant target information unrepresented by the source-trained mapping. We call this representation-level insufficiency Target-domain Anomaly Information Deficiency (TAID); it describes information loss under frozen target inference rather than the discrepancy itself, following information-theoretic sufficiency and data processing (Cover & Thomas, 2006) and related cross-domain studies (Zhang et al., 2013; Chen et al., 2022).

Complete Target-observable Information and Target Representation. Let P_t denote the target-domain population distribution. Let $V_t^\star$ denote a random query node drawn from $\mathcal{Q}$, and let $Y_t \in \{0,1\}$ denote its anomaly label. Let Obs collect the target information permitted by the GGAD protocol—node attributes, graph structure, reference nodes, and deterministic relations derived from them—and let Φ_{θ_s} be the frozen representation mapping. Then, under P_t, we write

$$(\mathcal{G}_t, \mathcal{S}_t, V_t^\star, Y_t) \sim P_t, \qquad O_t \triangleq \mathrm{Obs}(\mathcal{G}_t, \mathcal{S}_t, V_t^\star), \qquad Z_t \triangleq \Phi_{\theta_s}(O_t), \tag{1}$$

where O_t excludes target labels, training signals, and external information, while Z_t is the complete node-level representation before scalar anomaly scoring. Since $Z_t = \Phi_{\theta_s}(O_t)$, the variables form the *Markov chain*

$$Y_t \to O_t \to Z_t. \tag{2}$$

This notation specifies the Markov relation and does not imply causality.

Anomaly-task Sufficiency and TAID. Under the Markov relation in Equation (2), the following equivalent conditions characterize anomaly-task sufficiency of Z_t with respect to O_t:

$$p_t(Y_t \mid O_t) = p_t(Y_t \mid Z_t) \text{ a.s.} \quad \Longleftrightarrow \quad I_t(Y_t; O_t \mid Z_t) = 0, \tag{3}$$

where $p_t(\cdot \mid \cdot)$ denotes conditional probability under P_t, "a.s." means almost surely, and I_t denotes mutual information under P_t. Related connections to task sufficiency are studied by Rakelly et al. (2021) and Wu et al. (2025). For $I_t(Y_t; O_t) > 0$, define

$$\mathrm{TAID}_t \triangleq \frac{I_t(Y_t; O_t \mid Z_t)}{I_t(Y_t; O_t)} = 1 - \frac{I_t(Y_t; Z_t)}{I_t(Y_t; O_t)}. \tag{4}$$

Equation (2) gives $I_t(Y_t; O_t) = I_t(Y_t; Z_t) + I_t(Y_t; O_t \mid Z_t)$, so TAID_t is the fraction of target-observable anomaly-relevant information not retained in Z_t.

Properties and Multi-domain Aggregation. Non-negativity and data processing yield $0 \leq \mathrm{TAID}_t \leq 1$: 0 indicates anomaly-task sufficiency, while 1 means that O_t contains anomaly-relevant information but Z_t retains none beyond the label prior. If $I_t(Y_t; O_t) = 0$, the ratio is undefined and the target is excluded. For K valid unseen targets indexed by $\{t_1, \ldots, t_K\}$, the macro-average is

$$\overline{\mathrm{TAID}} \triangleq \frac{1}{K} \sum_{k=1}^{K} \mathrm{TAID}_{t_k} . \tag{5}$$

This formulation motivates using complementary structural and group-level relational information from the target graph during frozen inference, without target labels or parameter updates.

4 METHOD

4.1 SEMANTIC RESIDUAL

Feature Alignment and Residual Encoding. let $\mathbf{X}^{\mathrm{proj}} \in \mathbb{R}^{N \times d_u}$ be the projected node attributes, where d_u is the aligned dimension, and let $\widetilde{\mathbf{X}}^{\mathrm{proj}}$ be their feature-wise min–max normalization. Let $\mathcal{E}_G = \{(v_i, v_j) \mid A_{ij} = 1\}$ be the edge set, and let $\widetilde{x}_{ik}^{\mathrm{proj}}$ denote the normalized entry of node v_i at coordinate $k \in \{1, \ldots, d_u\}$. We define the coordinate smoothness η_k and let $\pi = (\pi_1, \ldots, \pi_{d_u})$ order the coordinates by nondecreasing smoothness:

$$\eta_k = \frac{1}{|\mathcal{E}_G|} \sum_{(v_i, v_j) \in \mathcal{E}_G} \left(\widetilde{x}_{ik}^{\mathrm{proj}} - \widetilde{x}_{jk}^{\mathrm{proj}} \right)^2, \qquad \mathbf{X}^{\mathrm{aln}} = \mathbf{X}_{:,\pi}^{\mathrm{proj}}, \qquad \eta_{\pi_1} \leq \cdots \leq \eta_{\pi_{d_u}} . \tag{6}$$

Here, $\mathbf{X}^{\mathrm{aln}}$ is the aligned attribute matrix. We use $\widetilde{\mathbf{A}}$ to denote the symmetrically normalized adjacency with self-loops, L to denote the maximum propagation scale, and Enc to denote the semantic encoder. For each $l \in \{0, \ldots, L\}$, we define $\mathbf{X}^{(l)} = \widetilde{\mathbf{A}}^{\,l} \mathbf{X}^{\mathrm{aln}}$ and $\mathbf{H}^{(l)} = \mathrm{Enc}(\mathbf{X}^{(l)})$. Before residual formation, the shared prototype-mixing operation detailed in Appendix B.2 transforms $\mathbf{H}^{(l)}$ into $\overline{\mathbf{H}}^{(l)}$. If $\|$ denotes concatenation, the semantic residual is

$$\mathbf{r}_i = \left[\overline{\mathbf{H}}_{i,:}^{(1)} - \overline{\mathbf{H}}_{i,:}^{(0)} \,\|\, \cdots \,\|\, \overline{\mathbf{H}}_{i,:}^{(L)} - \overline{\mathbf{H}}_{i,:}^{(0)} \right]. \tag{7}$$

Support–Prototype Joint Reference. let $\mathcal{C}$ denote the codebook learned from $\mathbf{H}^{(L)}$, and let $\mathrm{VQ}(\cdot; \mathcal{C})$ denote its vector-quantization operator. With $L = 2$ in our implementation, the codewords are duplicated across the two residual blocks to match the dimensionality of $\mathbf{r}_i$. For the normal source support set $\mathcal{S}_s$, we define the support and prototype centers as

$$\mathbf{c}_{\mathrm{sup}} = \frac{1}{|\mathcal{S}_s|} \sum_{j \in \mathcal{S}_s} \mathbf{r}_j, \qquad \mathbf{c}_{\mathrm{pro}} = \frac{1}{|\mathcal{S}_s|} \sum_{j \in \mathcal{S}_s} \mathrm{VQ}(\mathbf{r}_j; \mathcal{C}). \tag{8}$$

Training uses ranking, compactness, support–prototype consistency, directional separation, and vector-quantization objectives.

Target-Domain Robust Reference. At inference, initial centers $\widehat{\mathbf{c}}_{\mathrm{sup}}$ and $\widehat{\mathbf{c}}_{\mathrm{pro}}$ are estimated from $\mathcal{S}_t$. For $v_j \in \mathcal{S}_t$, define $\mathrm{dist}_j = \|\mathbf{r}_j - \widehat{\mathbf{c}}_{\mathrm{sup}}\|_2 + \|\mathbf{r}_j - \widehat{\mathbf{c}}_{\mathrm{pro}}\|_2$. The lowest $\tau|\mathcal{S}_t|$ scores under retention fraction τ yield purified centers $\widetilde{\mathbf{c}}_{\mathrm{sup}}$ and $\widetilde{\mathbf{c}}_{\mathrm{pro}}$. With blending coefficient $\alpha \in [0, 1]$,the refined centers are

$$\mathbf{c}_{\mathrm{sup}}^{t} = (1 - \alpha)\widehat{\mathbf{c}}_{\mathrm{sup}} + \alpha \widetilde{\mathbf{c}}_{\mathrm{sup}}, \qquad \mathbf{c}_{\mathrm{pro}}^{t} = (1 - \alpha)\widehat{\mathbf{c}}_{\mathrm{pro}} + \alpha \widetilde{\mathbf{c}}_{\mathrm{pro}}.$$

Motivated by robust location estimation (Diakonikolas et al., 2023; Lerasle et al., 2019), we refine the initial support and prototype centers with fixed retention and blending hyperparameters reported in Appendix A.3. For $v_i \in \mathcal{Q}$, let $\mathrm{MinMax}_{\mathcal{Q}}$ denote query-wise scalar min–max normalization and Norm_2 denote vector ℓ_2-normalization. We define the semantic-deviation vector as

$$\mathbf{u}_i^{\mathrm{sem}} = \mathrm{MinMax}_{\mathcal{Q}} \left(\frac{\|\mathbf{r}_i - \mathbf{c}_{\mathrm{sup}}^{t}\|_2 + \|\mathbf{r}_i - \mathbf{c}_{\mathrm{pro}}^{t}\|_2}{2} \right) \mathrm{Norm}_2 \left(2\mathbf{r}_i - \mathbf{c}_{\mathrm{sup}}^{t} - \mathbf{c}_{\mathrm{pro}}^{t} \right). \tag{9}$$

This construction preserves both deviation magnitude and direction.

4.2 MULTI-SCALE STRUCTURAL PROFILE RECONSTRUCTION

To complement semantic residuals, we construct a degree-guided multi-scale structural profile and learn structural deviations by reconstruction.

Structural Profile Construction. For node v_i, let $\mathcal{N}_i$ denote its one-hop neighborhood and let $d_i = |\mathcal{N}_i|$ denote its degree. We define the safeguarded degree as $\bar{d}_i = \max(d_i, 1)$ and the log-degree as $z_i = \log(1 + d_i)$. The structural statistics comprise the first- and second-scale neighborhood means $\mu_i^{(1)}$ and $\mu_i^{(2)}$, the one-hop dispersion $\sigma_i^{(1)}$, the local shift $\delta_i^{(1)}$, the cross-scale shift $\delta_i^{(2)}$, and the neighborhood-growth statistic g_i, defined as follows:

$$\mu_i^{(1)} = \frac{1}{\bar{d}_i} \sum_{v_j \in \mathcal{N}_i} z_j, \qquad \mu_i^{(2)} = \frac{1}{\bar{d}_i} \sum_{v_j \in \mathcal{N}_i} \mu_j^{(1)}, \qquad \sigma_i^{(1)} = \sqrt{\frac{1}{\bar{d}_i} \sum_{v_j \in \mathcal{N}_i} z_j^2 - \left(\mu_i^{(1)}\right)^2},$$

$$\delta_i^{(1)} = |z_i - \mu_i^{(1)}|, \qquad \delta_i^{(2)} = |\mu_i^{(1)} - \mu_i^{(2)}|, \qquad g_i = \log\left(1 + \frac{\sum_{v_j \in \mathcal{N}_i} d_j}{\bar{d}_i}\right).$$

These statistics form the raw profile $\mathbf{p}_i^{\mathrm{raw}} = [z_i, \mu_i^{(1)}, \sigma_i^{(1)}, \mu_i^{(2)}, \delta_i^{(1)}, \delta_i^{(2)}, g_i]^\top \in \mathbb{R}^{d_{\mathrm{str}}}$, where $d_{\mathrm{str}} = 7$. The profile captures connectivity scale, neighborhood statistics, local and cross-scale shifts, and growth. After coordinate-wise standardization by the median and median absolute deviation (MAD), we denote the standardized profile by $\mathbf{p}_i$.

Profile Reconstruction and Structural Deviation. Taking $\mathbf{X}^{\mathrm{aln}}$ and the graph as input, a one-layer GCN encoder (Kipf & Welling, 2017) followed by a two-layer decoder produces $\widehat{\mathbf{p}}_i$, a reconstruction of $\mathbf{p}_i$. Let $\mathbf{w}^{\mathrm{str}} = [w_1^{\mathrm{str}}, \ldots, w_{d_{\mathrm{str}}}^{\mathrm{str}}]^\top$ be reconstruction weights and $\rho \in [0, 1]$ the degree-weight ratio, with $w_1^{\mathrm{str}} = \rho$ and the remaining weight distributed uniformly over the other $d_{\mathrm{str}} - 1$ coordinates, as summarized in Appendix A.3. The reconstruction loss is

$$\mathcal{L}_{\mathrm{rec}} = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{d_{\mathrm{str}}} w_k^{\mathrm{str}} (\widehat{p}_{ik} - p_{ik})^2,$$

where p_{ik} and $\widehat{p}_{ik}$ are the k-th entries of $\mathbf{p}_i$ and $\widehat{\mathbf{p}}_i$, respectively.

Let $\mathrm{MinMax}_{\mathcal{V},\mathrm{dim}}$ and $\mathrm{MinMax}_{\mathcal{V}}$ denote coordinate-wise and scalar min–max normalization over nodes, respectively; let $\odot$ denote element-wise multiplication, let Proj denote the learned projection from weighted structural-error space to semantic-residual space, and let $\varepsilon > 0$ be a stabilizer. The normalized error and structural deviation are

$$\mathbf{e}_i^{\mathrm{str}} = \mathrm{MinMax}_{\mathcal{V},\mathrm{dim}}(|\widehat{\mathbf{p}}_i - \mathbf{p}_i|), \qquad \mathbf{u}_i^{\mathrm{str}} = \mathrm{MinMax}_{\mathcal{V}}((\mathbf{w}^{\mathrm{str}})^\top \mathbf{e}_i^{\mathrm{str}}) \frac{\mathrm{Proj}(\mathbf{w}^{\mathrm{str}} \odot \mathbf{e}_i^{\mathrm{str}})}{\|\mathrm{Proj}(\mathbf{w}^{\mathrm{str}} \odot \mathbf{e}_i^{\mathrm{str}})\|_2 + \varepsilon}.$$

The scalar factor preserves reconstruction-derived magnitude, while the projection determines direction. Training also uses a magnitude-consistency regularizer; its implementation details are reported in Appendix A.3.

4.3 CROSS-VIEW HYPERGRAPH

Hypergraph Construction. During frozen target inference, $\mathcal{H}^{\mathrm{str}}$ and $\mathcal{H}^{\mathrm{sem}}$ contain self, one-hop ego, and relational hyperedges for individual nodes, ego neighborhoods, and view-similar nodes, respectively. Oversized ego-hyperedges retain neighbors by standardized log-degree proximity in the structural view or cosine similarity in the semantic view. To construct relational hyperedges, we define the structural and semantic relation matrices as

$$\mathbf{F}^{\mathrm{str}} = [\mathbf{p}_1^\top; \ldots; \mathbf{p}_N^\top] \in \mathbb{R}^{N \times d_{\mathrm{str}}}, \qquad \mathbf{F}^{\mathrm{sem}} = \mathrm{Std}_{\mathrm{MAD}}((\mathbf{X}^{\mathrm{aln}})_{:,1:d_{\mathrm{rel}}}) \in \mathbb{R}^{N \times d_{\mathrm{rel}}},$$

where $\mathrm{Std}_{\mathrm{MAD}}$ denotes coordinate-wise standardization by the median and MAD, and d_{rel} is the number of semantic-relation dimensions, with $1 \le d_{\mathrm{rel}} \le d_u$. For view $a \in \{\mathrm{str}, \mathrm{sem}\}$ and coordinate k, let B_a be the number of quantile bins and let $q_{k,b}^a$ denote their boundaries, with $q_{k,0}^a = -\infty$ and $q_{k,B_a}^a = +\infty$. The b-th relational hyperedge is

$$\mathcal{B}_{k,b}^a = \{v_i \mid q_{k,b-1}^a \le [\mathbf{F}^a]_{ik} < q_{k,b}^a\}, \qquad b = 1, \ldots, B_a.$$

Small groups are discarded and oversized ones are partitioned; the retained groups are combined with self and ego hyperedges to form $\mathcal{H}^{\mathrm{str}}$ and $\mathcal{H}^{\mathrm{sem}}$.

Cross-view Propagation. For view a, let $\mathbf{M}^a$, $\mathbf{W}_e^a$, $\mathbf{D}_v^a$, and $\mathbf{D}_e^a$ denote the node–hyperedge incidence, hyperedge-weight, node-degree, and hyperedge-degree matrices, respectively. Following HGNN (Feng et al., 2019) and HGNN+ (Gao et al., 2023b), the normalized propagation operator is

$$\mathbf{T}^a = (\mathbf{D}_v^a)^{-1/2}\mathbf{M}^a\mathbf{W}_e^a(\mathbf{D}_e^a)^{-1}(\mathbf{M}^a)^\top(\mathbf{D}_v^a)^{-1/2}.$$

Let $\mathbf{R}$ and $\mathbf{U}^{\mathrm{str}}$ stack $\mathbf{r}_i$ and $\mathbf{u}_i^{\mathrm{str}}$ row-wise, with row-normalized forms $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{U}}^{\mathrm{str}}$. Structural relations propagate semantic residuals, while semantic relations propagate structural deviations:

$$\mathbf{\Delta}^{\mathrm{sem}} = \widehat{\mathbf{R}} - \mathbf{T}^{\mathrm{str}}\widehat{\mathbf{R}}, \qquad \mathbf{\Delta}^{\mathrm{str}} = \widehat{\mathbf{U}}^{\mathrm{str}} - \mathbf{T}^{\mathrm{sem}}\widehat{\mathbf{U}}^{\mathrm{str}}.$$

For $a \in \{\mathrm{sem}, \mathrm{str}\}$, let $\mathbf{\Delta}_i^a$ denote the i-th row of $\mathbf{\Delta}^a$. For $v_i \in \mathcal{Q}$, the magnitude-preserving enhancement is

$$\mathbf{h}_i^a = \mathrm{MinMax}_{\mathcal{Q}}(\|\mathbf{\Delta}_i^a\|_2)\frac{\mathbf{\Delta}_i^a}{\|\mathbf{\Delta}_i^a\|_2 + \varepsilon}, \qquad a \in \{\mathrm{sem}, \mathrm{str}\},$$

so a larger propagation discrepancy indicates stronger relational mismatch.

4.4 Vector-level Integration

For query node v_i, let β and γ be semantic and structural enhancement weights computed from unlabeled target statistics. The enhanced representations are

$$\widetilde{\mathbf{u}}_i^{\mathrm{sem}} = \mathbf{u}_i^{\mathrm{sem}} + \beta\mathbf{h}_i^{\mathrm{sem}}, \qquad \widetilde{\mathbf{u}}_i^{\mathrm{str}} = \mathbf{u}_i^{\mathrm{str}} + \gamma\mathbf{h}_i^{\mathrm{str}}. \tag{10}$$

Let $\lambda \in [0, 1]$ be the semantic–structural fusion coefficient, also computed from unlabeled target statistics. The fused target representation and its anomaly score are

$$\mathbf{Z}_{t,i} = (1-\lambda)\widetilde{\mathbf{u}}_i^{\mathrm{sem}} + \lambda\widetilde{\mathbf{u}}_i^{\mathrm{str}}, \qquad s_{t,i} = \|\mathbf{Z}_{t,i}\|_2. \tag{11}$$

Here, $\mathbf{Z}_{t,i}$ is the unified target representation, $s_{t,i}$ is the final anomaly score, and λ balances semantic and structural information. When $V_t^\star = v_i$, the Z_t defined in Section 3.3 corresponds to $\mathbf{Z}_{t,i}$.

5 Experiments

5.1 Experimental Setup

Datasets. We use PubMed, Flickr, Questions, and YelpChi for source training and Cora, CiteSeer, ACM, BlogCatalog, Facebook, Weibo, Reddit, and Amazon as unseen targets. These datasets cover citation, social, review, and e-commerce graphs and include benchmarks introduced by Tang & Liu (2009), Rayana & Akoglu (2015), and Kumar et al. (2019).

Baselines. We compare with GAD baselines, *i.e.*, CoLA (Liu et al., 2022b), TAM (Qiao & Pang, 2023), GHRN (Gao et al., 2023a), Semi-GGAD (Qiao et al., 2024), SmoothGNN (Dong et al., 2025), and GCTAM (Zhang et al., 2025b), and GGAD baselines, *i.e.*, ARC (Liu et al., 2024), AnomalyGFM (Qiao et al., 2025a), UNPrompt (Niu et al., 2025), OWLEYE (Zheng et al., 2026), DR-GGAD (Fu et al., 2026), and IA-GGAD (Zhang et al., 2025a).

Metrics and Protocol. We use the area under the receiver-operating-characteristic curve (AUROC), the area under the precision–recall curve (AUPRC), and average rank over the eight targets as evaluation metrics. AUPRC is informative under class imbalance (Saito & Rehmsmeier, 2015), and lower average rank indicates more consistent cross-dataset performance (Demšar, 2006). All methods are evaluated under frozen target inference without target labels, retraining, or parameter updates. For conventional GAD baselines, we apply the same source-training/frozen-target protocol for a consistent cross-domain comparison.

Implementation. TIEGGAD uses 64-dimensional aligned features, two propagation scales, 40 training epochs, and five trials with seeds 0–4; we report mean $\pm$ standard deviation. Each target

Table 1: Overall performance (%) on eight target datasets. Rank is averaged over the target datasets; lower is better. Best results are boldfaced, second-best results are underlined.

AUROC (%)

Method \ Datasets	Cora	CiteSeer	ACM	BlogCatalog	Facebook	Weibo	Reddit	Amazon	Rank
GAD									
CoLA	63.29±8.88	62.84±9.52	66.85±4.43	50.04±3.25	12.99±11.68	16.27±5.64	52.81±6.69	47.40±7.97	10.38
TAM	62.02±2.39	72.27±0.83	74.43±1.59	49.86±0.73	65.88±6.66	71.54±0.18	55.43±0.33	56.06±2.19	7.63
GHRN	44.30±2.77	44.54±3.58	65.30±1.82	72.20±2.45	63.02±5.42	45.19±1.28	43.09±0.87	47.25±2.86	10.63
Semi-GGAD	39.44±5.41	38.18±4.21	37.47±2.68	50.70±7.34	55.89±8.99	65.73±3.35	55.39±0.44	53.11±4.92	11.00
SmoothGNN	49.23±9.85	49.76±9.28	49.61±6.35	46.99±8.79	49.60±5.36	52.42±24.51	48.27±5.54	48.00±4.05	11.38
GCTAM	58.78±2.17	70.31±1.77	81.21±0.13	67.60±0.77	69.57±1.41	70.61±0.10	59.32±0.73	55.74±0.60	6.25
GGAD									
ARC	87.41±1.04	91.12±0.54	79.91±0.28	74.31±0.18	67.48±1.13	88.88±0.14	59.86±0.82	80.15±1.95	4.19
AnomalyGFM	53.06±1.92	48.47±1.31	55.01±3.02	53.99±1.33	56.05±4.71	82.86±4.20	59.07±5.56	60.15±3.47	8.38
UNPrompt	53.28±1.40	53.18±3.92	69.15±0.81	68.28±0.90	59.10±6.59	41.53±5.80	59.06±1.05	48.62±12.51	8.63
OWLEYE	79.80±0.22	84.19±0.67	78.05±0.25	74.63±0.30	65.82±0.88	88.02±0.21	58.28±0.52	83.69±2.68	5.13
DR-GGAD	92.78±0.72	94.66±0.55	91.19±0.38	74.31±0.44	81.83±1.89	88.94±0.26	57.81±0.32	87.74±1.85	3.31
IA-GGAD	87.99±0.66	92.01±0.27	**93.13±0.27**	74.75±0.22	80.03±1.20	91.53±0.32	59.03±1.07	82.93±1.67	3.00
TIEGGAD (ours)	**96.49±0.43**	**97.74±0.11**	92.27±0.38	**74.79±0.73**	**93.50±1.02**	**94.79±0.87**	**61.78±1.03**	**88.39±0.59**	**1.13**

AUPRC (%)

Method \ Datasets	Cora	CiteSeer	ACM	BlogCatalog	Facebook	Weibo	Reddit	Amazon	Rank
GAD									
CoLA	11.41±3.51	8.33±3.73	7.31±1.45	6.04±0.56	1.90±0.68	7.59±3.26	3.71±0.67	11.06±4.45	10.38
TAM	11.18±0.75	11.55±0.44	23.20±2.36	10.57±1.17	8.40±0.97	16.46±0.09	3.94±0.13	10.75±3.10	8.00
GHRN	5.52±0.85	4.47±1.02	17.48±6.47	26.69±3.01	4.63±2.46	10.30±0.24	2.82±0.16	6.56±0.22	10.50
Semi-GGAD	4.65±0.50	3.72±0.30	9.72±2.18	22.95±6.54	3.44±1.05	52.42±4.13	4.25±0.21	7.37±0.98	10.00
SmoothGNN	6.79±2.75	5.82±3.00	12.44±5.92	15.04±5.48	3.90±2.32	41.67±19.34	3.16±0.44	6.30±0.67	9.75
GCTAM	9.52±0.69	10.29±0.28	48.09±0.28	27.47±0.54	9.61±0.96	16.87±1.31	4.34±0.17	13.71±0.11	6.13
GGAD									
ARC	50.20±1.10	46.55±0.89	40.73±0.06	35.45±0.23	8.47±2.04	64.29±0.46	4.48±0.26	43.29±5.13	4.13
AnomalyGFM	6.73±0.27	4.66±0.12	4.57±0.30	8.28±0.63	3.10±0.56	37.43±3.59	4.90±1.45	8.44±1.10	9.75
UNPrompt	5.71±0.20	4.82±0.60	9.24±0.90	24.40±4.02	3.27±0.80	15.57±5.31	4.98±0.50	7.47±3.65	9.25
OWLEYE	44.80±0.33	43.24±0.76	39.74±0.11	34.69±0.16	5.39±0.52	60.73±0.37	4.10±0.11	55.54±6.84	5.63
DR-GGAD	55.32±2.43	63.06±3.04	51.24±0.83	**37.09±1.07**	13.66±2.07	66.68±0.50	4.87±0.18	62.45±5.33	2.25
IA-GGAD	47.41±1.06	48.78±1.21	**57.03±1.16**	35.36±0.49	5.80±0.41	71.29±0.21	4.43±0.24	44.53±10.23	3.75
TIEGGAD (ours)	**61.17±2.71**	**68.18±2.41**	41.31±0.51	35.71±0.87	**14.31±2.34**	**75.27±2.48**	**5.04±0.31**	**63.93±6.19**	**1.50**

uses 600 reference nodes sampled uniformly without labels. Structural profiles use median/MAD standardization with offset 10^{-6} and clipping to $[-8, 8]$. Hypergraph defaults are ego-cardinality $3:12$, 16 quantile bins, relational-cardinality $8:64$, singleton/ego/relational weights $0.10/1.00/0.45$, and the first eight aligned dimensions for semantic relations.

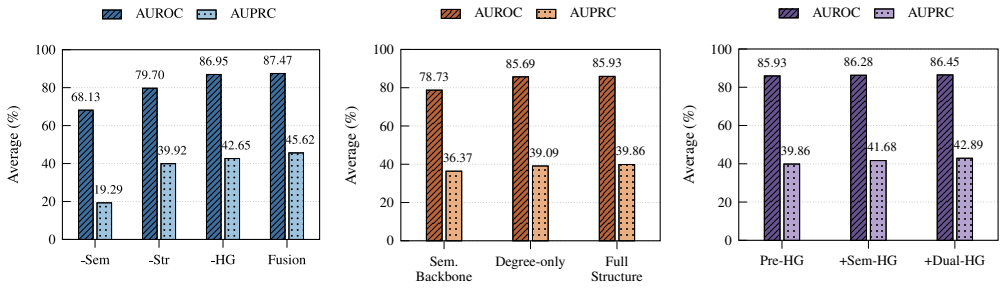
5.2 Overall Performance

Table 1 reports AUROC and AUPRC on the eight unseen target graphs. TIEGGAD achieves the best average rank for both AUROC (1.13) and AUPRC (1.50), with the best results on seven and six targets, respectively. Compared with IA-GGAD, the mean AUROC increases from 82.68 to 87.47 and the mean AUPRC from 39.33 to 45.62, corresponding to gains of 4.79 and 6.29 percentage points. Compared with DR-GGAD, TIEGGAD further improves the mean AUROC from 83.66 to 87.47 and the mean AUPRC from 44.30 to 45.62, yielding gains of 3.81 and 1.32 percentage points. Overall, TIEGGAD shows stronger cross-dataset performance while remaining competitive across heterogeneous targets.

5.3 Ablation Studies

Figure 2 summarizes the main and internal ablations using average AUROC and AUPRC over the eight target graphs. In panel (a), **Fusion** denotes the full model, while $-$**Sem**, $-$**Str**, and $-$**HG** remove the semantic branch, structural branch, and cross-view hypergraph enhancement, respectively.

Main modules. Removing the semantic branch causes the largest degradation, reducing average AUROC/AUPRC from 87.47/45.62 to 68.13/19.29, confirming that transferable semantic residuals provide the primary detection signal. Removing the structural branch reduces the two metrics to 79.70/39.92, corresponding to drops of 7.77 and 5.70 points and demonstrating the complementary value of structural information. Without cross-view hypergraph enhancement, AUROC remains high at 86.95, whereas AUPRC decreases to 42.65. The larger AUPRC change indicates that cross-view enhancement provides a clearer benefit for precision–recall performance while largely preserving overall separability.



(a) Main modules (b) Structural branch (c) Hypergraph module

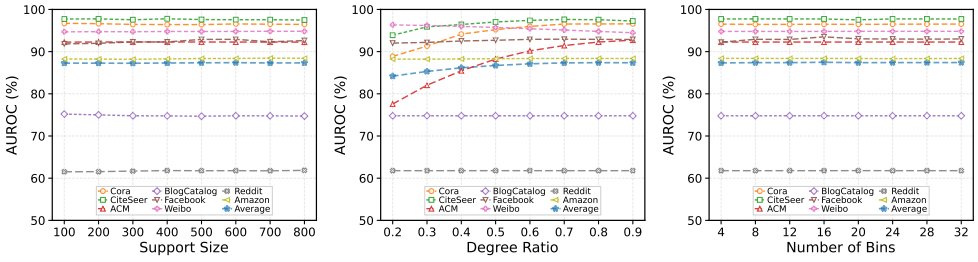
Figure 2: Ablation studies of TIEGGAD using average AUROC and AUPRC over eight target graphs: (a) main modules; (b) structural branch; and (c) hypergraph module.

Structural branch. Figure 2(b) isolates the structural branch with hypergraph enhancement and prototype mixing disabled. Adding the degree coordinate to the semantic backbone raises average AUROC/AUPRC from 78.73/36.37 to 85.69/39.09, while the full multi-scale structural profile further improves them to 85.93/39.86. Overall, the complete structural profile yields gains of 7.20 AUROC and 3.49 AUPRC points over the semantic backbone, showing that degree provides the main structural anchor while the remaining multi-scale statistics provide additional improvements.

Hypergraph module. Figure 2(c) evaluates hypergraph enhancement with prototype mixing disabled while retaining the complete structural branch. Semantic-side enhancement improves average AUROC/AUPRC from 85.93/39.86 to 86.28/41.68, and bidirectional cross-view enhancement further increases them to 86.45/42.89. Thus, the complete cross-view module provides cumulative gains of 0.52 AUROC and 3.03 AUPRC points over the pre-hypergraph representation, with consistent improvements in both metrics. Because these staged ablations use reduced base settings, their values need not match the **Fusion** or $-$**HG** variants in panel (a).

5.4 Parameter Sensitivity

We study sensitivity to the support size, degree-weight ratio, and number of relational-hyperedge bins. Figure 3 reports AUROC across the eight target graphs under different parameter settings, together with the corresponding average performance.



(a) Support size (b) Degree-weight ratio (c) Hyperedge bins

Figure 3: Parameter sensitivity of TIEGGAD in terms of AUROC (%): (a) support size; (b) degree-weight ratio; and (c) number of relational-hyperedge bins.

Support size. Changing the reference-set size causes only minor fluctuations in average AUROC.

Degree-weight ratio. Average AUROC increases more noticeably but smoothly as the degree-weight ratio grows to 0.5 and changes little thereafter, reflecting the central role of degree; the remaining profile statistics still provide complementary multi-scale cues.

Hyperedge bins. The hyperedge-bin curve also varies smoothly but reaches its best average performance at 16 bins. Overall, TIEGGAD is not highly sensitive to these parameters within the tested ranges.

5.5 Experimental Analysis of TAID

Why an Empirical Proxy is Needed. Section 3.3 defines TAID at the target-domain population level as

$$\mathrm{TAID}_t = \frac{I_t(Y_t; O_t \mid Z_t)}{I_t(Y_t; O_t)}.$$

Here, O_t and Z_t follow Section 3.3. Direct evaluation is infeasible because only a single target graph is observed and O_t is high-dimensional, so we use a common post-hoc decoder to quantify how much anomaly-relevant information remains decodable from Z_t.

Evaluation Protocol. All compared methods use the same tree-based HistGradientBoosting probabilistic decoder with five-fold out-of-fold (OOF) evaluation. For $v_i \in \mathcal{Q}$, let $\mathrm{fold}(i) \in \{1, \ldots, 5\}$ be its held-out fold, $\Pr_{-\,\mathrm{fold}(i)}$ the decoder distribution trained on the other folds, and $\mathbf{W}_i = [\mathbf{Z}_{t,i}, s_{t,i}]$ its input; since $s_{t,i}$ is a deterministic readout of $\mathbf{Z}_{t,i}$, it adds no target-side information. Let $\widehat{H}(Y_t)$ denote the empirical label entropy on $\mathcal{Q}$. For targets with $\widehat{H}(Y_t) > 0$, the OOF cross-entropy and empirical proxy are

$$\mathrm{CE}_{\mathrm{OOF}} = -\frac{1}{|\mathcal{Q}|} \sum_{i \in \mathcal{Q}} \log \Pr_{-\,\mathrm{fold}(i)}(Y_{t,i} \mid \mathbf{W}_i), \qquad \text{E-TAID}_t = \frac{\min\{\mathrm{CE}_{\mathrm{OOF}}, \widehat{H}(Y_t)\}}{\widehat{H}(Y_t)}.$$

Here, $Y_{t,i}$ is the held-out label. Hence $0 \leq \text{E-TAID}_t \leq 1$, with smaller values indicating less residual anomaly-task uncertainty. E-TAID is a decoder-dependent empirical proxy rather than a direct estimate of the mutual-information ratio, and target labels are used only for this post-hoc evaluation. Decoder robustness with ExtraTrees and PCA–LR is reported in Appendix E.

Table 2: E-TAID (↓) on eight unseen targets under the shared five-fold OOF decoder. Avg. denotes the macro-average; best and second-best results are boldfaced and underlined.

	E-TAID (↓)								
Method	Cora	CiteSeer	ACM	BlogCatalog	Facebook	Weibo	Reddit	Amazon	Avg.
GHRN	0.99	0.99	0.84	0.83	0.99	0.77	<u>0.97</u>	0.97	0.92
UNPrompt	0.85	0.81	0.75	<u>0.81</u>	0.96	<u>0.26</u>	**0.95**	0.38	0.72
IA-GGAD	<u>0.68</u>	<u>0.65</u>	**0.45**	**0.78**	<u>0.88</u>	0.43	0.99	0.71	<u>0.70</u>
ARC	0.76	0.71	0.74	<u>0.81</u>	0.99	0.28	**0.95**	**0.35**	<u>0.70</u>
AnomalyGFM	0.99	0.99	0.99	0.99	0.99	0.55	**0.95**	0.91	0.92
TIEGGAD (ours)	**0.49**	**0.45**	<u>0.59</u>	<u>0.81</u>	**0.77**	**0.25**	**0.95**	<u>0.36</u>	**0.58**

Results. TIEGGAD obtains the lowest average E-TAID of 0.58, compared with 0.70 for both IA-GGAD and ARC, and is best or tied-best on five of the eight targets. IA-GGAD leads on ACM and BlogCatalog, while ARC leads on Amazon. Across methods, lower average E-TAID broadly coincides with stronger AUROC/AUPRC performance, with TIEGGAD achieving both the lowest average E-TAID and the best average detection ranks. This agreement supports the role of complementary structural and relational information in retaining anomaly-relevant target information, although E-TAID is not derived from AUROC/AUPRC and need not reproduce their dataset-wise ordering.

6 Conclusion

We study zero-shot generalist graph anomaly detection under Target-domain Anomaly Information Deficiency (TAID), where a source-trained frozen detector may fail to preserve anomaly-relevant information on unseen target graphs. TIEGGAD complements semantic residuals with degree-guided multi-scale structural reconstruction and training-free cross-view hypergraph enhancement, then integrates the resulting information at the vector level without target-domain labels or parameter updates. Across eight unseen targets, TIEGGAD achieves the best average ranks in both AUROC and AUPRC; ablation results and the lower average E-TAID further support the complementary value of structural and relational information for mitigating TAID.

AI USE STATEMENT

In this work, we used generative AI tools to assist with manuscript writing and language polishing, assist with the retrieval and discovery of related work, and draft portions of the manuscript. All AI-assisted text and retrieved references were independently reviewed, verified, and revised by the authors. The authors take full responsibility for the final manuscript, including its text, citations, claims, methods, and experimental results.

REPRODUCIBILITY STATEMENT

We provide the source–target split, evaluation protocol, and major implementation settings in Section 5.1. All reported target-domain results follow frozen inference without target labels, retraining, fine-tuning, or parameter updates, and are averaged over five trials with fixed random seeds. Additional implementation details, complete per-dataset ablation results, and complementary AUPRC parameter-sensitivity curves are provided in the Appendix. These details are intended to support faithful reproduction of the experimental setting and reported results.

REFERENCES

Yongqiang Chen, Yonggang Zhang, Yatao Bian, Han Yang, Kaili Ma, Binghui Xie, Tongliang Liu, Bo Han, and James Cheng. Learning causally invariant representations for out-of-distribution generalization on graphs. In *Advances in Neural Information Processing Systems*, volume 35, pp. 22131–22148, 2022. doi: 10.52202/068431-1608.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 2nd edition, 2006. ISBN 9780471241959. doi: 10.1002/047174882X. URL https://onlinelibrary.wiley.com/doi/book/10.1002/047174882X.

Janez Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006.

Ilias Diakonikolas, Daniel M. Kane, Ankit Pensia, and Thanasis Pittas. Near-optimal algorithms for gaussians with Huber contamination: Mean estimation and linear regression. In *Advances in Neural Information Processing Systems*, volume 36, pp. 43384–43422, 2023. doi: 10.52202/075280-1879.

Xiangyu Dong, Xingyi Zhang, Yanni Sun, Lei Chen, Mingxuan Yuan, and Sibo Wang. SmoothGNN: Smoothing-aware GNN for unsupervised node anomaly detection. In *Proceedings of the ACM on Web Conference 2025*, pp. 1225–1236, 2025. doi: 10.1145/3696410.3714615.

Haoyi Fan, Fengbin Zhang, and Zuoyong Li. AnomalyDAE: Dual autoencoder for anomaly detection on attributed networks. In *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5685–5689, 2020. doi: 10.1109/ICASSP40776.2020.9053387.

Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 3558–3565, 2019. doi: 10.1609/aaai.v33i01.33013558.

Changlong Fu, Zhenli He, Xiong Zhang, Cheng Xie, Xin Jin, and Yun Yang. DR-GGAD: Dual residual centering for mitigating anomaly non-discriminativity in generalist graph anomaly detection. In *International Conference on Learning Representations*, 2026.

Yuan Gao, Xiang Wang, Xiangnan He, Zhenguang Liu, Huamin Feng, and Yongdong Zhang. Addressing heterophily in graph anomaly detection: A perspective of graph spectrum. In *Proceedings of the ACM Web Conference 2023*, pp. 1528–1538, 2023a. doi: 10.1145/3543507.3583268.

Yue Gao, Yifan Feng, Shuyi Ji, and Rongrong Ji. HGNN+: General hypergraph neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3181–3199, 2023b. doi: 10.1109/TPAMI.2022.3182052.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=SJU4ayYgl.

Srijan Kumar, Xikun Zhang, and Jure Leskovec. Predicting dynamic embedding trajectory in temporal interaction networks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1269–1278, 2019. doi: 10.1145/3292500.3330895.

Matthieu Lerasle, Zoltan Szabo, Timothée Mathieu, and Guillaume Lecue. MONK outlier-robust mean embedding estimation by median-of-means. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 3782–3793. PMLR, 2019. URL https://proceedings.mlr.press/v97/lerasle19a.html.

Kay Liu, Yingtong Dou, Yue Zhao, Xueying Ding, Xiyang Hu, Ruitong Zhang, Kaize Ding, Canyu Chen, Hao Peng, Kai Shu, Lichao Sun, Jundong Li, George H. Chen, Zhihao Jia, and Philip S. Yu. BOND: Benchmarking unsupervised outlier node detection on static attributed graphs. In *Advances in Neural Information Processing Systems*, volume 35, pp. 27021–27035, 2022a. doi: 10.52202/068431-1959.

Yixin Liu, Zhao Li, Shirui Pan, Chen Gong, Chuan Zhou, and George Karypis. Anomaly detection on attributed networks via contrastive self-supervised learning. *IEEE Transactions on Neural Networks and Learning Systems*, 33(6):2378–2392, 2022b. doi: 10.1109/TNNLS.2021.3068344.

Yixin Liu, Shiyuan Li, Yu Zheng, Qingfeng Chen, Chengqi Zhang, and Shirui Pan. ARC: A generalist graph anomaly detector with in-context learning. In *Advances in Neural Information Processing Systems*, volume 37, pp. 50772–50804, 2024. doi: 10.52202/079017-1606.

Yujing Liu, Yixin Liu, Yu Zheng, Alan Wee-Chung Liew, Xiaofeng Cao, and Shirui Pan. Rethinking feature alignment in generalist graph anomaly detection: A relational fingerprint-based approach. In *Proceedings of the 43rd International Conference on Machine Learning*, volume 306 of *Proceedings of Machine Learning Research*. PMLR, 2026.

Chaoxi Niu, Hezhe Qiao, Changlu Chen, Ling Chen, and Guansong Pang. Zero-shot generalist graph anomaly detection with unified neighborhood prompts. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pp. 3226–3234, 2025. doi: 10.24963/ijcai.2025/359. URL https://www.ijcai.org/proceedings/2025/359.

Hezhe Qiao and Guansong Pang. Truncated affinity maximization: One-class homophily modeling for graph anomaly detection. In *Advances in Neural Information Processing Systems*, volume 36, pp. 49490–49512, 2023. doi: 10.52202/075280-2154.

Hezhe Qiao, Qingsong Wen, Xiaoli Li, Ee-Peng Lim, and Guansong Pang. Generative semi-supervised graph anomaly detection. In *Advances in Neural Information Processing Systems*, volume 37, pp. 4660–4688, 2024. doi: 10.52202/079017-0152.

Hezhe Qiao, Chaoxi Niu, Ling Chen, and Guansong Pang. AnomalyGFM: Graph foundation model for zero/few-shot anomaly detection. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, pp. 2326–2337, 2025a. doi: 10.1145/3711896.3736843.

Hezhe Qiao, Hanghang Tong, Bo An, Irwin King, Charu Aggarwal, and Guansong Pang. Deep graph anomaly detection: A survey and new perspectives. *IEEE Transactions on Knowledge and Data Engineering*, 37(9):5106–5126, 2025b. doi: 10.1109/TKDE.2025.3581578.

Kate Rakelly, Abhishek Gupta, Carlos Florensa, and Sergey Levine. Which mutual-information representation learning objectives are sufficient for control? In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Shebuti Rayana and Leman Akoglu. Collective opinion spam detection: Bridging review networks and metadata. In *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 985–994, 2015. doi: 10.1145/2783258.2783370.

Takaya Saito and Marc Rehmsmeier. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*, 10(3):e0118432, 2015. doi: 10.1371/journal.pone.0118432.

Lei Tang and Huan Liu. Relational learning via latent social dimensions. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 817–826, 2009. doi: 10.1145/1557019.1557109.

Qilong Wu, Yiyang Shao, Jun Wang, and Xiaobo Sun. Learning optimal multimodal information bottleneck representations. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 67584–67622. PMLR, 2025. URL https://proceedings.mlr.press/v267/wu25x.html.

Yiming Xu, Zihan Chen, Zhen Peng, Song Wang, Bin Shi, Bo Dong, and Chao Shen. Generalist graph anomaly detection via prototype-based distillation. In *Proceedings of the 43rd International Conference on Machine Learning*, volume 306 of *Proceedings of Machine Learning Research*. PMLR, 2026.

Kun Zhang, Bernhard Schölkopf, Krikamol Muandet, and Zhikun Wang. Domain adaptation under target and conditional shift. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pp. 819–827. PMLR, 2013. URL https://proceedings.mlr.press/v28/zhang13d.html.

Xiong Zhang, Zhenli He, Changlong Fu, and Cheng Xie. IA-GGAD: Zero-shot generalist graph anomaly detection via invariant and affinity learning. In *Advances in Neural Information Processing Systems*, volume 38, pp. 93324–93356, 2025a. doi: 10.52202/085713-2807.

Xiong Zhang, Hong Peng, Zhenli He, Cheng Xie, Xin Jin, and Hua Jiang. GCTAM: Global and contextual truncated affinity combined maximization model for unsupervised graph anomaly detection. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pp. 3642–3650, 2025b. doi: 10.24963/ijcai.2025/405.

Yunfeng Zhao, Yixin Liu, Shiyuan Li, Qingfeng Chen, Yu Zheng, and Shirui Pan. FreeGAD: A training-free yet effective approach for graph anomaly detection. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*, pp. 4379–4389, 2025. doi: 10.1145/3746252.3761125.

Lecheng Zheng, Dongqi Fu, Zihao Li, and Jingrui He. OwlEye: Zero-shot learner for cross-domain graph data anomaly detection. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=ugGIpVfGE6.

A Detailed Reproducibility Protocol

This appendix records the experimental protocol, implementation details, complete per-dataset ablation results, and complementary AUPRC parameter-sensitivity curves underlying Sections 5.3 and 5.4. Target labels are accessed only after inference for evaluation.

A.1 Experimental Scope and Trial Protocol

The source graphs are PubMed, Flickr, Questions, and YelpChi; the eight reported targets are Cora, CiteSeer, ACM, BlogCatalog, Facebook, Weibo, Reddit, and Amazon. A single detector is optimized sequentially over the four sources and then frozen for target inference.

We run five trials with seeds $0, 1, 2, 3, 4$. Each target uses 600 uniformly sampled unlabeled reference nodes to form $\mathcal{S}_t$, with the remaining nodes forming the query set $\mathcal{Q}$. AUROC and AUPRC are evaluated only on $\mathcal{Q}$, and Table 1 reports mean $\pm$ standard deviation over the five trials. Deterministic CUDA execution is enabled and cuDNN benchmarking is disabled.

Dataset Descriptions

Following the benchmark construction used by ARC and IA-GGAD (Liu et al., 2024; Zhang et al., 2025a), the twelve graphs span citation, social, and co-review networks, with both injected and real anomalies.

PubMed (source). PubMed is a citation network in which nodes are scientific publications, edges are citation links, and node attributes are bag-of-words features. The benchmark uses injected structural and attribute anomalies.

Flickr (source). Flickr is a social network whose nodes represent users and whose edges represent social connections; user-generated textual and tag information forms the attributes. The benchmark uses injected anomalies.

Questions (source). Questions is derived from the Yandex Q question-answering platform, with users as nodes and question–answer interactions as edges. User-description embeddings form the attributes, and the benchmark provides real anomaly labels.

YelpChi (source). YelpChi is a co-review network for deceptive-review detection. We use the RUR relation, where reviews posted by the same user are connected, and the benchmark contains real review anomalies.

Cora (target). Cora is a citation network of scientific publications with citation edges and bag-of-words node attributes. The benchmark evaluates injected structural and attribute anomalies.

CiteSeer (target). CiteSeer is a citation network with papers as nodes, citations as edges, and text-derived node attributes. Its anomaly labels are generated by the same benchmark injection protocol as the other citation targets.

ACM (target). ACM is a larger citation graph whose nodes are publications and whose edges encode citation relations, with text-based publication features. The benchmark uses injected anomalies.

BlogCatalog (target). BlogCatalog is a social-network benchmark with users as nodes, social links as edges, and user-generated textual content as attributes. Its anomalies are injected under the shared benchmark protocol.

Facebook (target). Facebook is a social graph in which nodes correspond to users and edges represent friendship relations. It belongs to the real-anomaly group in the benchmark.

Weibo (target). Weibo is a user social-media graph with hashtag and posting information; location and bag-of-words signals are included in the node features. Repeated suspicious posting behavior provides real anomaly labels.

Reddit (target). Reddit is a forum-based interaction network with user nodes and text-derived attributes. Banned users are treated as real anomalies in the benchmark.

Amazon (target). Amazon is a co-review graph designed for review-fraud analysis. We use the UPU relation, which connects users who reviewed at least one common product, and the benchmark provides real anomaly labels.

BASELINE DESCRIPTIONS

The baselines in Section 5.1 cover both conventional graph anomaly detection (GAD) and generalist graph anomaly detection (GGAD) settings.

CoLA. CoLA (Liu et al., 2022b) is a contrastive self-supervised GAD method that contrasts each node with sampled local substructures and derives anomaly scores from node–context agreement.

TAM. TAM (Qiao & Pang, 2023) models one-class homophily through local node affinity and iteratively truncates likely non-homophilous connections so that normal-node affinities are strengthened for anomaly scoring.

GHRN. GHRN (Gao et al., 2023a) addresses graph heterophily from a spectral perspective, using high-frequency structural signals to identify and prune likely inter-class edges before anomaly detection.

Semi-GGAD. Semi-GGAD (Qiao et al., 2024) exploits available normal nodes to generate pseudo anomalies from local-affinity and egocentric-closeness priors, providing negative samples for a discriminative one-class detector.

SmoothGNN. SmoothGNN (Dong et al., 2025) exploits differences in individual and neighborhood smoothing patterns between normal and anomalous nodes and combines smoothing-aware representation learning with anomaly-oriented scoring.

GCTAM. GCTAM (Zhang et al., 2025b) extends truncated affinity maximization with contextual and global affinity modeling, enabling more adaptive truncation of suspicious nodes than a fixed local threshold.

ARC. ARC (Liu et al., 2024) combines smoothness-based feature alignment, ego-neighbor residual encoding, and in-context anomaly scoring with normal reference nodes to generalize across unseen graphs.

AnomalyGFM. AnomalyGFM (Qiao et al., 2025a) pre-trains graph-agnostic normal and abnormal prototypes aligned with node residual representations, supporting zero-shot inference and optional few-shot prompt tuning on new graphs.

UNPrompt. UNPrompt (Niu et al., 2025) learns unified neighborhood prompts that capture transferable normal patterns, allowing a one-for-all detector to perform zero-shot anomaly detection on unseen graphs without target-side retraining.

OWLEYE. OWLEYE (Zheng et al., 2026) combines cross-domain feature alignment, multi-domain dictionary learning, and truncated-attention reconstruction to transfer normal patterns and detect anomalies in unseen graph domains.

DR-GGAD. DR-GGAD (Fu et al., 2026) uses dual residual centering: a multi-scale residual center captures node-to-center deviations, while an affinity-residual module promotes local directional consistency for cross-domain separability.

IA-GGAD. IA-GGAD (Zhang et al., 2025a) addresses feature-space and graph-structure shifts through invariant representation learning and structure-insensitive affinity learning, producing anomaly predictions on unseen graphs without target fine-tuning.

A.2 DATA LOADING, ALIGNMENT, AND GRAPH PREPROCESSING

The loader accepts cached `.npz` or raw `.mat` graphs. Sparse Amazon and YelpChi attributes are row-normalized before alignment. When the original feature dimension is below 64, a fixed Gaussian random projection with seed 0 expands it to 256 dimensions, followed by principal component analysis (PCA) with seed 0 to obtain $\mathbf{X}^{\mathrm{proj}} \in \mathbb{R}^{N\times 64}$.

Coordinates are then ordered by the edge-smoothness criterion in Equation (6) to obtain $\mathbf{X}^{\mathrm{aln}}$. Semantic propagation uses the symmetrically normalized adjacency; the structural graph is canonicalized separately to contain exactly one self-loop per node.

The seven-dimensional structural profile is computed from the unweighted adjacency after removing its diagonal. Table 3 records only the implementation order and interpretation because the exact formulas are already given in Section 4.2. Each coordinate is standardized independently by the median and median absolute deviation (MAD), using an offset of 10^{-6}, replacing non-finite values with zero, and clipping the result to $[-8, 8]$.

Table 3: Coordinate order and interpretation of the seven-dimensional structural profile used by TIEGGAD.

Coordinate	Statistic	Interpretation
1	Log-degree	Direct node-connectivity scale
2	One-hop mean	Average log-degree in the one-hop neighborhood
3	One-hop spread	Dispersion of one-hop log-degrees
4	Two-hop mean	Neighborhood mean propagated to the second scale
5	Local shift	Difference between the node and its one-hop context
6	Cross-scale shift	Difference between one- and two-hop neighborhood statistics
7	Neighborhood growth	Local expansion reflected by neighboring degrees

This order matches $\mathbf{p}_i$ in Section 4.2 and separates direct connectivity, neighborhood statistics, and cross-scale deviations.

A.3 Reported Configuration

Table 4 summarizes the reported eight-target configuration. Bold rows are group labels spanning both columns, not settings with missing values. When `params/GCN_600.json` is available, its optimization settings take precedence over the parser fallbacks; all other entries correspond to the reported runs.

The launcher fixes the unlabeled reference size, five evaluation trials, and the same relational-bin count in both hypergraph directions; the evaluation list is restricted to the eight targets in Section A.1. These settings are shared across targets rather than selected from target labels or dataset-specific validation. The internal depth flag and normal-prompt count are implementation-level launcher controls reported verbatim for exact reproduction and do not introduce additional paper-level symbols. The vector-loss weight corresponds to η_{mag} in Equation (13); the top-score fraction is used by the overlap statistic in Section B.5, and the λ grid step is the quantization increment of the deterministic fusion rule. For the E-TAID analysis in Section 5.5, the same fixed HistGradientBoosting decoder configuration is used for every method and target, without dataset-specific decoder tuning.

B Optimization and Inference Details

B.1 Semantic Training Objective

The semantic encoder produces residual vectors $\mathbf{r}_i$ as defined in Section 4.1. For a D-dimensional residual, we use the root-mean-square distance

$$\mathrm{RMSDist}(\mathbf{x}, \mathbf{c}) = \sqrt{D^{-1}\|\mathbf{x} - \mathbf{c}\|_2^2 + 10^{-12}},$$

and write $\mathrm{CosSim}(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle / (\|\mathbf{x}\|_2 \|\mathbf{y}\|_2)$ for cosine similarity. Let $\mathbf{c}_{\mathrm{sup}}$ and $\mathbf{c}_{\mathrm{pro}}$ be the source support and prototype centers from Equation (8), and let $\mathbf{c}_{\Sigma} = \mathbf{c}_{\mathrm{sup}} + \mathbf{c}_{\mathrm{pro}}$. The training-

Table 4: Reported TIEGGAD configuration.

Setting	Reported value
Protocol and representation	
Source graphs / target graphs	$4/8$
Trials / seeds	$5/\{0,1,2,3,4\}$
Epochs / target reference size	$40/600$
Aligned dimension / propagation scales	$64/2$
Semantic hidden width / activation / dropout	1024 / ELU / 0.30
Internal depth flag / normal prompts	$4/10$
Prototype codewords per source / top-k	$2048/15$
Optimization and reference construction	
Optimizer / learning rate / weight decay (fallback)	AdamW / 10^{-5} / 5×10^{-5}
Structural profile dimension / hidden width	$7/32$
Degree-weight ratio / vector-loss weight	$0.80/0.10$
Reference retention τ / blend α	$0.90/0.25$
Hypergraph and integration	
Ego / relational cardinality limits	3:12 / 8:64
Relational quantile bins	16
Self / ego / relational edge weights	$0.10/1.00/0.45$
Semantic relation dimensions	8
Top-score fraction / λ grid step	$0.05/0.05$
Maximum β / maximum γ	$0.55/0.45$

time semantic score is

$$d_{\mathrm{sup}}(\mathbf{x}) = \mathrm{RMSDist}(\mathbf{x}, \mathbf{c}_{\mathrm{sup}}), \quad d_{\mathrm{pro}}(\mathbf{x}) = \mathrm{RMSDist}(\mathbf{x}, \mathbf{c}_{\mathrm{pro}}), \quad a(\mathbf{x}) = 1 - \mathrm{CosSim}(\mathbf{x}, \mathbf{c}_{\Sigma}),$$
$$\widetilde{a}(\mathbf{x}) = a(\mathbf{x}) - 0.35[\max(a(\mathbf{x}) - 1, 0)]^2, \qquad s_{\mathrm{tr}}(\mathbf{x}) = \tfrac{1}{2}[d_{\mathrm{sup}}(\mathbf{x}) + d_{\mathrm{pro}}(\mathbf{x})] + \sqrt{d_{\mathrm{sup}}(\mathbf{x}) d_{\mathrm{pro}}(\mathbf{x})}\, \widetilde{a}(\mathbf{x}).$$

Up to ten source-normal nodes are used as supports. An equal number of anomalous nodes and non-support normal nodes are sampled as queries. Hard normals are the normal queries with the largest $s_{\mathrm{tr}}(\cdot)$. Let $\mathbf{x}_n$, $\mathbf{x}_a$, and $\mathbf{x}_h$ denote a random normal query, an anomalous query, and a hard-normal query, respectively. With ranking margin $m_{\mathrm{rank}} = 0.25$, the two ranking losses are

$$\mathcal{L}_{\mathrm{rand}} = \mathbb{E}[\mathrm{softplus}(s_{\mathrm{tr}}(\mathbf{x}_n) - s_{\mathrm{tr}}(\mathbf{x}_a) + m_{\mathrm{rank}})]\,,$$
$$\mathcal{L}_{\mathrm{hard}} = \mathbb{E}[\mathrm{softplus}(s_{\mathrm{tr}}(\mathbf{x}_h) - s_{\mathrm{tr}}(\mathbf{x}_a) + m_{\mathrm{rank}})]\,.$$

The compactness loss is the mean support score. The support–prototype consistency loss is implemented as the mean RMS distance from each support to its nearest detached codeword, and the directional loss averages cosine-embedding terms that pull random normals toward $\mathbf{c}_{\mathrm{sup}}$ and push anomalies away from it. We denote these terms by $\mathcal{L}_{\mathrm{compact}}$, $\mathcal{L}_{\mathrm{code}}$, and $\mathcal{L}_{\mathrm{dir}}$, respectively, and let $\mathcal{L}_{\mathrm{VQ}}$ denote the vector-quantization objective. The implemented semantic objective is

$$\mathcal{L}_{\mathrm{sem}} = 0.65\mathcal{L}_{\mathrm{rand}} + 0.35\mathcal{L}_{\mathrm{hard}} + 0.08\mathcal{L}_{\mathrm{compact}} + 0.05\mathcal{L}_{\mathrm{code}} + 0.10\mathcal{L}_{\mathrm{dir}} + \mathcal{L}_{\mathrm{VQ}}. \tag{12}$$

The vector-quantization objective contains reconstruction of the deepest semantic embedding from its quantized version together with the commitment loss, whose coefficient is 0.25. The exponential-moving-average codebook decay is 0.8. Semantic and structural parameters use separate AdamW optimizers.

B.2 Source Codebooks and Target Reference Refinement

At the final source-training epoch, a 2048-entry codebook is saved for each source graph. Because the semantic residual concatenates two propagation scales, each 1024-dimensional codeword is duplicated along the feature dimension. During target inference, the four source codebooks are concatenated along the codeword axis to form the joint codebook used by the prototype reference.

Target reference refinement follows the robust-reference procedure in Section 4.1. Initial support and prototype centers are computed from all reference nodes. Reference nodes are ranked by the

sum of their distances to these centers, and a fixed fraction is retained before the initial and retained centers are blended with a fixed interpolation coefficient. Target labels are not used for either ranking or refinement.

The implementation also applies shared prototype mixing within the semantic branch before residual formation: each propagated embedding is averaged with the mean of its top-k highest-dot-product codewords, using the same k reported in Table 4 for every target. This operation does not depend on target labels or dataset identity. The staged internal ablations in Section 5.3 disable prototype mixing and use the same reduced semantic configuration across variants, so their base values need not match the full-model ablation.

B.3 Structural Reconstruction Objective

The structural branch takes $\mathbf{X}^{\mathrm{aln}}$ and the graph as input. A linear layer of width 32 with ELU activation is followed by one graph-convolution layer and a second ELU; a two-layer decoder reconstructs the seven-dimensional profile $\mathbf{p}_i$.

For degree-weight ratio ρ, the coordinate weights are

$$w_1^{\mathrm{str}} = \rho, \qquad w_k^{\mathrm{str}} = \frac{1-\rho}{d_{\mathrm{str}}-1}, \quad k = 2, \ldots, d_{\mathrm{str}}, \qquad d_{\mathrm{str}} = 7.$$

As in Section 4.2, define

$$\mathbf{e}_i^{\mathrm{str}} = \mathrm{MinMax}_{\mathcal{V},\mathrm{dim}}(|\widehat{\mathbf{p}}_i - \mathbf{p}_i|), \qquad a_i^{\mathrm{str}} = \mathrm{MinMax}_{\mathcal{V}}((\mathbf{w}^{\mathrm{str}})^\top \mathbf{e}_i^{\mathrm{str}}),$$

where $\mathbf{w}^{\mathrm{str}} = [w_1^{\mathrm{str}}, \ldots, w_{d_{\mathrm{str}}}^{\mathrm{str}}]^\top$. The projection Proj maps the weighted error vector to the semantic-residual space. With a small stabilizer $\varepsilon > 0$, the structural deviation is

$$\mathbf{u}_i^{\mathrm{str}} = a_i^{\mathrm{str}} \frac{\mathrm{Proj}(\mathbf{w}^{\mathrm{str}} \odot \mathbf{e}_i^{\mathrm{str}})}{\|\mathrm{Proj}(\mathbf{w}^{\mathrm{str}} \odot \mathbf{e}_i^{\mathrm{str}})\|_2 + \varepsilon}.$$

The implemented structural objective augments weighted reconstruction with a magnitude-consistency term:

$$\mathcal{L}_{\mathrm{struct}} = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{d_{\mathrm{str}}} w_k^{\mathrm{str}}(\widehat{p}_{ik} - p_{ik})^2 + \eta_{\mathrm{mag}}\ \mathrm{MSE}\big(\mathrm{MinMax}_{\mathcal{V}}(\|\mathrm{Proj}(\mathbf{w}^{\mathrm{str}} \odot \mathbf{e}_i^{\mathrm{str}})\|_2), a_i^{\mathrm{str}}\big). \tag{13}$$

Here MSE denotes mean squared error evaluated over the node index i, and η_{mag} is the vector-loss weight reported in Table 4. The scalar target a_i^{str} is detached in the second term. This term aligns the projected-vector magnitude with the reconstruction-derived magnitude used in the structural deviation.

B.4 Exact Hyperedge Construction and Propagation

Each target hypergraph contains three hyperedge families.

1. **Self (singleton) hyperedges.** One self hyperedge is created for every node with the fixed self-edge weight reported in Appendix A.3.
2. **Ego hyperedges.** A center and its one-hop neighbors form an ego hyperedge if its cardinality is at least 3. Cardinality is capped at 12, including the center. For a structural hypergraph, excess neighbors are retained by proximity of the standardized log-degree coordinate. For a semantic hypergraph, they are retained by cosine similarity in the first eight median/MAD-standardized aligned feature dimensions. Ego hyperedges use the fixed edge weight reported in Appendix A.3.
3. **Relational hyperedges.** For every relation coordinate, nodes are assigned to 16 empirical-quantile bins after duplicate boundaries are collapsed. Groups smaller than 8 are discarded. Larger groups are sorted by node index and partitioned into chunks of at most 64; a final chunk is retained only if it contains at least 8 nodes. Relational hyperedges use the fixed edge weight reported in Appendix A.3.

Incidence matrices are cached for each graph and hypergraph configuration. Propagation uses the symmetric operator in Section 4.3 and introduces no learnable target-side parameters.

The structural hypergraph propagates row-normalized semantic residuals, while the semantic hypergraph propagates row-normalized structural vectors. In each direction, the propagated context is subtracted from the original representation and the query-wise min–max-normalized discrepancy magnitude is restored after row normalization. Hypergraph construction and propagation are performed only after source training is complete.

B.5 Unsupervised Target Characterization and Adaptive Integration

The adaptive integration policy uses no target labels and does not assign special rules to named datasets. Instead, each unseen graph is characterized by the relative strength and agreement of the semantic, structural, and hypergraph-derived anomaly evidence on the query nodes. Let $\mathbf{q}_s$ and $\mathbf{q}_r$ denote the normalized semantic-vector magnitude and structural reconstruction score. We summarize their behavior through a robust tail-separation statistic, Pearson correlation, and top-score overlap:

$$R_s = \mathrm{Rel}(\mathbf{q}_s), \qquad R_r = \mathrm{Rel}(\mathbf{q}_r), \qquad C_{sr} = \mathrm{Corr}(\mathbf{q}_s, \mathbf{q}_r), \qquad O_{sr} = \mathrm{Overlap}_{\mathrm{top}}(\mathbf{q}_s, \mathbf{q}_r).$$

Here $\mathrm{Rel}(\cdot)$ measures how clearly the upper tail separates from the robust central spread, while C_{sr} and O_{sr} measure cross-view agreement. The same types of statistics are computed between each base view, its cross-view hypergraph discrepancy, and the auxiliary signal from the opposite view. These quantities provide a target-side description without using labels.

The policy can be interpreted through four broad evidence regimes.

Semantic-dominant targets. When the semantic signal has a clearly stronger tail than the structural signal and the two views show little complementary agreement, the structural fusion weight λ is kept small. In this regime, the source-trained semantic residual already provides the clearer anomaly evidence, so a large structural contribution could dilute a strong transferable signal. Hypergraph gates are also conservative when their discrepancies add little evidence beyond the base semantic representation.

Structure-helpful targets. When structural reconstruction produces evidence that is comparable to or clearer than the semantic signal, especially when the semantic tail is weak, λ is increased so that the structural branch contributes more strongly. Positive cross-view overlap or correlation further supports this increase because it indicates that the two branches identify compatible abnormal regions while contributing different cues.

Balanced or complementary targets. When semantic and structural evidence are both informative and neither view clearly dominates, the policy retains a moderate mixture rather than collapsing to one branch. This is the intended operating regime of vector-level integration: semantic residuals provide transferable attribute evidence, while the structural profile contributes target-observable topology that is not guaranteed to be preserved by the source mapping.

Weak-agreement or uncertain targets. When a candidate auxiliary signal is weak, negatively correlated with the base evidence, or has little overlap among high-scoring nodes, its contribution is reduced or suppressed. The policy is also more conservative on very large targets when the two views show weak overlap, preventing graph size or a diffuse discrepancy distribution from producing an excessively large structural weight.

The hypergraph weights β and γ follow the same evidence principle. The semantic-side gate β is increased only when structural relations produce a reliable discrepancy that is supported by the structural reconstruction cue; the structural-side gate γ is increased when semantic relations provide similarly consistent evidence for the structural representation. If the base view is already strong while the propagated discrepancy is weak or inconsistent, the corresponding gate is reduced, and it can be set to zero. Thus the adaptive policy favors complementary relational evidence rather than increasing hypergraph contribution by default.

All mappings from these unlabeled statistics to β, γ, and λ are deterministic, bounded, and shared across datasets. The implementation uses fixed clipping, thresholding, and quantization rules; the top-score fraction and λ quantization step are reported in Table 4. The remaining fixed thresholds are implementation constants shared across all targets and are specified in the accompanying implementation rather than selected from target performance. Neither target labels nor dataset identity

enter the computation. This preserves the frozen zero-shot protocol while allowing the relative contribution of semantic, structural, and relational evidence to respond to the observed target graph.

B.6 Frozen Target-Inference Sequence

Target inference proceeds in the following order:

1. sample the unlabeled reference set and define its complement as the query set;
2. generate aligned multi-scale embeddings, apply the shared prototype mixing, and form semantic residuals with the frozen semantic encoder;
3. form robust support and prototype centers, then obtain the query semantic-deviation vectors;
4. reconstruct the seven-dimensional profiles with the frozen structural branch and obtain the structural vectors;
5. construct the two target hypergraphs and compute cross-view discrepancy vectors;
6. compute β, γ, and λ from unlabeled target statistics and integrate vectors as in Equation (11);
7. take the ℓ_2-norm of the integrated vector as the anomaly score, then access query labels only to compute AUROC and AUPRC.

No optimizer, gradient step, fine-tuning, or target-label-dependent selection occurs in Steps 1–6.

C Complete Ablation Results

This section reports the per-dataset values underlying Figure 2. All entries are percentages; "Average" is the unweighted macro-average over the eight reported target graphs. Values are shown to the two-decimal precision of the experiment records.

C.1 Main-Module Ablation

Table 5 compares the complete system with three removal variants. $-$**Sem** removes the semantic branch and semantic-side hypergraph enhancement, $-$**Str** removes the structural branch and structural-side hypergraph enhancement, $-$**HG** keeps both base branches but removes cross-view hypergraph enhancement, and **Fusion** denotes the complete model.

Table 5: Complete main-module ablation results (%): AUROC.

AUROC (%)

Dataset	$-$Sem	$-$Str	$-$HG	Fusion
Cora	72.96	86.22	95.75	**96.49**
CiteSeer	75.59	90.89	96.33	**97.74**
ACM	70.34	79.75	**92.27**	**92.27**
BlogCatalog	44.93	**74.79**	**74.79**	**74.79**
Facebook	93.41	68.22	92.04	**93.50**
Weibo	89.52	90.84	**94.81**	94.79
Reddit	44.11	58.37	**61.78**	**61.78**
Amazon	54.15	**88.52**	87.81	88.39
Average	68.13	79.70	86.95	**87.47**

The AUROC results show that the semantic branch provides the largest contribution: removing it lowers the macro-average from 87.47 to 68.13. Removing the structural branch produces a smaller but still substantial drop of 7.77 points, confirming that the structural profile supplies complementary target-side evidence. Removing cross-view hypergraph enhancement changes the average by only

0.52 points, indicating that its AUROC contribution is more incremental. Across individual targets, Fusion remains best or nearly best, with only very small deviations on Weibo and Amazon.

Table 6: Complete main-module ablation results (%): AUPRC.

AUPRC (%)

Dataset	−Sem	−Str	−HG	Fusion
Cora	31.42	48.07	52.26	**61.17**
CiteSeer	35.93	50.17	54.94	**68.18**
ACM	8.36	40.25	**41.31**	**41.31**
BlogCatalog	5.83	**35.71**	**35.71**	**35.71**
Facebook	13.95	6.36	12.83	**14.31**
Weibo	47.13	70.44	**75.30**	75.27
Reddit	2.73	4.27	**5.04**	**5.04**
Amazon	8.93	**64.05**	63.78	63.93
Average	19.29	39.92	42.65	**45.62**

AUPRC makes the complementary effects more visible. Relative to Fusion, removing the semantic, structural, and hypergraph components reduces the macro-average by 26.33, 5.70, and 2.97 points, respectively. The larger hypergraph gap in AUPRC than in AUROC indicates that relational enhancement is particularly useful for ranking scarce anomalous nodes near the top of the score distribution. Fusion achieves the highest macro-average, while the few per-target exceptions are small and do not change the overall trend.

C.2 Structural-Branch Ablation

The structural study uses the same reduced semantic configuration across all three variants and removes hypergraph enhancement. **Sem. Backbone** uses the reduced semantic backbone, **Degree-only** adds standardized log-degree as the sole structural coordinate, and **Full Structure** uses all seven coordinates in Table 3.

Table 7: Complete internal ablation of the structural branch (%): AUROC.

AUROC (%)

Dataset	Sem. Backbone	Degree-only	Full Structure
Cora	84.67	96.09	**96.23**
CiteSeer	90.33	95.42	**96.20**
ACM	79.77	**93.27**	92.44
BlogCatalog	**74.62**	**74.62**	**74.62**
Facebook	68.99	91.33	**92.04**
Weibo	90.81	94.20	**94.86**
Reddit	**60.15**	59.81	**60.15**
Amazon	80.48	80.79	**80.92**
Average	78.73	85.69	**85.93**

The AUROC comparison shows that degree supplies most of the structural improvement, increasing the macro-average from 78.73 to 85.69. Using the complete seven-dimensional profile further raises the average to 85.93, with additional gains on Cora, CiteSeer, Facebook, Weibo, and Amazon. BlogCatalog is unchanged across the three settings and Reddit remains essentially stable, suggesting that the structural branch does not force gains uniformly across all targets. Overall, the full profile preserves the strong degree anchor while adding modest cross-scale information where it is useful.

Table 8: Complete internal ablation of the structural branch (%): AUPRC.

AUPRC (%)

Dataset	Sem. Backbone	Degree-only	Full Structure
Cora	46.15	51.44	**52.61**
CiteSeer	47.51	50.33	**54.38**
ACM	40.31	**44.37**	43.59
BlogCatalog	**35.76**	**35.76**	**35.76**
Facebook	6.61	12.56	**12.83**
Weibo	70.61	74.41	**75.55**
Reddit	**4.54**	4.46	**4.54**
Amazon	39.48	39.37	**39.61**
Average	36.37	39.09	**39.86**

The AUPRC results reinforce the same pattern: Degree-only improves the reduced semantic backbone by 2.72 points on average, and the remaining six coordinates add another 0.77 points. The full profile gives clear additional gains on CiteSeer and Weibo and smaller improvements on several other targets, while ACM shows a slight decrease relative to Degree-only. This variation is consistent with the profile being complementary rather than uniformly dominant. The macro-average nevertheless favors Full Structure, supporting retention of the multi-scale coordinates alongside degree.

C.3 Cross-view Hypergraph Ablation

The hypergraph study fixes the same semantic–structural base across all three variants and varies only cross-view propagation. **Pre-HG** uses no hypergraph enhancement; **+Sem-HG** propagates semantic residuals over structural relations; and **+Dual-HG** additionally propagates structural vectors over semantic relations. This shared base isolates the contribution of one-way and bidirectional cross-view propagation.

Table 9: Complete internal ablation of cross-view hypergraph enhancement (%): AUROC.

AUROC (%)

Dataset	Pre-HG	+Sem-HG	+Dual-HG
Cora	96.23	96.93	**97.11**
CiteSeer	96.20	97.53	**97.86**
ACM	**92.44**	**92.44**	**92.44**
BlogCatalog	**74.62**	**74.62**	**74.62**
Facebook	92.04	92.04	**92.91**
Weibo	**94.86**	**94.86**	94.84
Reddit	**60.15**	**60.15**	**60.15**
Amazon	80.92	**81.65**	81.64
Average	85.93	86.28	**86.45**

AUROC increases from 85.93 before hypergraph enhancement to 86.28 with semantic-side propagation and 86.45 with the bidirectional variant. The gains are concentrated on Cora, CiteSeer, Facebook, and Amazon, whereas ACM, BlogCatalog, and Reddit remain unchanged and Weibo varies only marginally. This pattern is expected because the base semantic–structural representation is already strong on several targets. The consistent macro-average increase nevertheless shows that cross-view propagation adds useful relational evidence without broadly destabilizing the base detector.

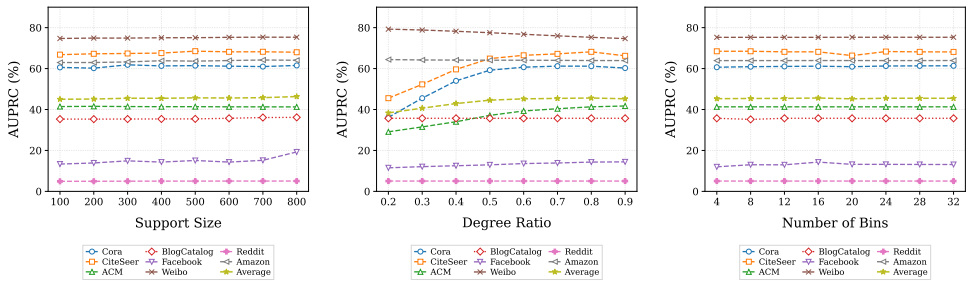
Table 10: Complete internal ablation of cross-view hypergraph enhancement (%): AUPRC.

Dataset	Pre-HG	+Sem-HG	+Dual-HG
	AUPRC (%)		
Cora	52.61	58.12	**62.18**
CiteSeer	54.38	63.62	**69.16**
ACM	**43.59**	**43.59**	**43.59**
BlogCatalog	**35.76**	**35.76**	**35.76**
Facebook	12.83	12.83	**12.99**
Weibo	**75.55**	**75.55**	75.52
Reddit	**4.54**	**4.54**	**4.54**
Amazon	**39.61**	39.39	39.39
Average	39.86	41.68	**42.89**

The AUPRC improvement is larger than the AUROC change: +Sem-HG raises the macro-average by 1.82 points and +Dual-HG by a further 1.21 points, for a total gain of 3.03 over Pre-HG. Cora and CiteSeer account for the clearest improvements, while most remaining targets are stable or change only slightly. The result suggests that cross-view propagation is especially helpful for precision–recall ranking even when the corresponding AUROC changes are comparatively small. Bidirectional enhancement therefore provides the strongest aggregate result without requiring target labels or target-side parameter updates.

D AUPRC Parameter-Sensitivity Results

Section 5.4 reports the AUROC sensitivity curves for support size, degree-weight ratio, and the number of relational-hyperedge bins. Figure 4 provides the corresponding AUPRC results under the same settings. Each sweep varies one parameter while keeping the remaining reported configuration fixed; the "Average" curve is the macro-average across the eight target graphs.



(a) Support size (b) Degree-weight ratio (c) Hyperedge bins

Figure 4: Parameter sensitivity of TIEGGAD in terms of AUPRC (%): (a) support size; (b) degree-weight ratio; and (c) number of relational-hyperedge bins.

D.1 Support Size

The support-size parameter is the number of unlabeled target reference nodes used to estimate the robust semantic reference. We vary it from 100 to 800 while keeping the sampling and inference protocol unchanged. As shown in Figure 4(a), average AUPRC changes moderately across the sweep, ranging from 45.01 to 46.33. The increase at 800 is driven mainly by Facebook, while most targets vary gradually. Together with the AUROC curve in Figure 3(a), this indicates that the reported setting of 600 provides a robust shared choice without target-specific selection.

D.2 Degree-Weight Ratio

The degree-weight ratio $\rho = w_1^{\mathrm{str}}$ controls how strongly the structural profile emphasizes node degree while keeping the profile dimension and branch architecture fixed. The remaining weight is distributed over the six complementary coordinates, so this sweep tests the balance between the degree anchor and the multi-scale statistics. Figure 4(b) shows that average AUPRC rises from 38.34 at $\rho = 0.2$ to its maximum of 45.62 at the reported $\rho = 0.8$, then changes only slightly at 0.9. The opposing per-target trends, especially for Weibo, support retaining nonzero weight for the complementary coordinates rather than reducing the structural profile to degree alone. Combined with the AUROC results in Figure 3(b), the selected ratio lies in a broad high-performing region rather than at an isolated target-specific optimum.

D.3 Number of Relational-Hyperedge Bins

The number of quantile bins controls the resolution of the relational hyperedges. We vary the shared bin count from 4 to 32 in both hypergraph directions while keeping the remaining hypergraph construction fixed. Figure 4(c) shows that average AUPRC varies by only 0.40 points and reaches its maximum at the reported setting of 16 bins. Most target curves are nearly flat, with the main visible variation coming from Facebook. Together with the AUROC sweep in Figure 3(c), this supports stable quantile-based grouping under one shared bin count rather than a narrowly tuned discretization.

Overall, the AUROC curves in Section 5.4 and the complementary AUPRC curves above support one shared configuration across heterogeneous targets: the reported settings achieve the best or near-best macro-average behavior while remaining in stable neighborhoods.

Rounding. Tables 5–10 use the two-decimal experiment records; average rows are unweighted macro-averages over the eight displayed targets.

E Decoder Robustness of E-TAID

Evidence states and common evaluation protocol. For E-TAID, each method retains its native evidence dimension; no cross-model dimensionality matching is imposed. We use the complete node-level evidence immediately before each method's original scalar scoring map: GHRN uses its two final stage-2 logits; UNPrompt uses the concatenation of its ℓ_2-normalized projected MLP and neighborhood embeddings; IA-GGAD uses its semantic- and affinity-deviation scores; ARC uses the query residual-difference vector $\mathbf{H}_q - \widetilde{\mathbf{H}}_q$; AnomalyGFM uses the normal- and abnormal-prompt similarity pair; and TIEGGAD uses the fused vector $\mathbf{Z}_{t,i}$ in Equation (11). The accompanying scalar $s(Z)$ is the method's unchanged original anomaly score, hence is a deterministic readout of the same evidence. The six methods are included because these complete pre-score evidence states can be extracted without altering their original scoring paths. Within each target and trial, the same post-hoc non-support evaluation mask is used for all six methods.

OOF split and fallback. The decoder uses stratified shuffled OOF splitting with the trial seed as the split seed and $n_{\mathrm{fold}} = \min(5, n_{\mathrm{min}})$, where n_{min} is the minority-class count. Decoder random states are $100\,s + f$ for trial seed s and fold index f. If the empirical label entropy is zero, E-TAID is undefined for that target; if both classes are present but $n_{\mathrm{min}} < 2$, the label-prior predictor is used, giving E-TAID $= 1$. Otherwise, the prior fallback in Section 5.5 uses $\min\{\mathrm{CE}_{\mathrm{OOF}}, \widehat{H}(Y_t)\}$.

Decoder configurations. The primary HGB decoder uses log loss, learning rate 0.05, 150 iterations, at most 15 leaf nodes, maximum depth 3, $\max(20, \lceil 0.01 n_{\mathrm{train}} \rceil)$ samples per leaf, L_2 regularization 1.0, and early stopping with validation fraction 0.15, 12 no-improvement iterations, and tolerance 10^{-7}. ExtraTrees uses 200 trees, maximum depth 8, minimum leaf size 10, square-root feature sampling, no class weighting, and no bootstrap. For PCA–LR, Z is standardized within each OOF training fold and reduced only when needed to $d_{\mathrm{PCA}} = \min(64, d_Z, n_{\mathrm{train}} - 1)$ components using randomized PCA; the scalar score is standardized separately and appended afterward. Logistic regression uses L_2 regularization with $C = 0.5$, the `lbfgs` solver, and at most 3000 iterations.

Section 5.5 reports E-TAID using HGB as the shared post-hoc decoder. Since E-TAID is decoder-dependent, we further examine whether the comparison is sensitive to decoder choice using ExtraTrees and PCA followed by logistic regression (PCA–LR). The evidence state, scalar readout,

evaluation nodes, OOF split rule, and entropy-based prior fallback remain unchanged; only the post-hoc decoder is replaced. Tables 11 and 12 report the resulting six-method comparison.

Table 11: E-TAID (↓) on eight unseen targets under the shared five-fold OOF ExtraTrees decoder. Avg. denotes the macro-average; best and second-best results are boldfaced and underlined.

E-TAID (↓, ExtraTrees)									
Method	Cora	CiteSeer	ACM	BlogCatalog	Facebook	Weibo	Reddit	Amazon	Avg.
GHRN	0.99	0.99	0.92	0.89	0.98	0.84	0.98	0.98	0.95
UNPrompt	0.90	0.87	0.85	<u>0.80</u>	0.94	**0.29**	<u>0.95</u>	0.54	0.77
IA-GGAD	<u>0.75</u>	<u>0.71</u>	0.79	0.81	<u>0.89</u>	0.64	0.99	0.77	0.79
ARC	0.81	0.76	<u>0.75</u>	**0.79**	0.97	0.44	**0.93**	**0.38**	<u>0.73</u>
AnomalyGFM	0.99	0.98	0.98	0.99	0.98	0.62	0.97	0.94	0.93
TIEGGAD (ours)	**0.61**	**0.58**	**0.60**	**0.79**	**0.77**	<u>0.32</u>	0.96	<u>0.39</u>	**0.63**

Table 12: E-TAID (↓) on eight unseen targets under the shared five-fold OOF PCA–LR decoder. Avg. denotes the macro-average; best and second-best results are boldfaced and underlined.

E-TAID (↓, PCA–LR)									
Method	Cora	CiteSeer	ACM	BlogCatalog	Facebook	Weibo	Reddit	Amazon	Avg.
GHRN	0.99	0.99	0.93	0.87	0.98	0.92	<u>0.99</u>	0.98	0.96
UNPrompt	0.97	0.92	0.83	<u>0.85</u>	0.99	**0.29**	**0.97**	0.48	0.79
IA-GGAD	<u>0.68</u>	<u>0.63</u>	**0.78**	**0.81**	**0.93**	0.53	<u>0.99</u>	0.74	<u>0.76</u>
ARC	0.82	0.77	0.83	1.00	1.00	0.41	**0.97**	**0.35**	0.77
AnomalyGFM	0.99	0.99	0.98	0.99	<u>0.97</u>	0.71	**0.97**	0.92	0.94
TIEGGAD (ours)	**0.58**	**0.51**	<u>0.79</u>	0.99	0.99	<u>0.34</u>	<u>0.99</u>	<u>0.36</u>	**0.69**

ExtraTrees. Under the ExtraTrees decoder, TIEGGAD achieves the lowest macro-average E-TAID of 0.63, compared with ARC (0.73), UNPrompt (0.77), IA-GGAD (0.79), AnomalyGFM (0.93), and GHRN (0.95). At the dataset level, TIEGGAD obtains the best or tied-best result on Cora, CiteSeer, ACM, BlogCatalog, and Facebook, while remaining second-best on Weibo and Amazon. Although the relative ordering varies on individual targets, particularly Reddit and Weibo, the overall results show that the lower E-TAID of TIEGGAD is preserved when the primary HGB decoder is replaced by a different tree-based ensemble.

PCA–LR. With PCA–LR, TIEGGAD again achieves the lowest macro-average E-TAID, reaching 0.69 compared with 0.76 for IA-GGAD, 0.77 for ARC, 0.79 for UNPrompt, 0.94 for AnomalyGFM, and 0.96 for GHRN. TIEGGAD performs best on Cora and CiteSeer and remains second-best on ACM, Weibo, and Amazon, while tying for second-best on Reddit. Larger dataset-wise variations are observed on BlogCatalog and Facebook, where IA-GGAD obtains the lowest E-TAID and AnomalyGFM becomes second-best on Facebook. Nevertheless, TIEGGAD retains the lowest macro-average result under this substantially different linear decoder.

Taken together with the HGB results in Table 2, the three decoder evaluations provide a consistent aggregate picture. TIEGGAD achieves the lowest macro-average E-TAID under HGB (0.58), ExtraTrees (0.63), and PCA–LR (0.69), despite noticeable changes in dataset-wise rankings across decoders. This consistency suggests that the lower residual anomaly-task uncertainty of TIEGGAD is not tied to a particular post-hoc decoder. Combined with the stronger average AUROC and AUPRC performance reported in Section 5.2, the results further support that the structural and cross-view information introduced by TIEGGAD helps retain anomaly-relevant target information under frozen cross-domain inference.