WHO GETS THE RISK? UNIFIED MULTISCALE OPTIMAL TRANSPORT FOR GRAPH ANOMALY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Label-free graph anomaly detection (GAD) aims to identify anomalous nodes without annotations. Yet anomaly patterns vary across graph domains, structures, and attribute regimes, making node-level evidence difficult to remain consistent across graphs. Moreover, existing methods typically map or aggregate label-free proxies into node scores that jointly determine risk positions and node assignment, leaving the valid scope of evidence implicit. To address these challenges, we propose Mu**l**tiscale **O**ptimal T**r**ansport for Graph Anomaly D**e**tection (LORE), a unified framework designed for label-free graph anomaly detection. LORE reformulates graph anomaly detection as an evidence-guided risk allocation problem by constructing a graph-adaptive risk prior and incorporating label-free node evidence to establish the correspondence between node evidence and risk positions. It further leverages multiscale optimal transport (OT) to reassign risks across different granularities, thereby correcting initial anomaly rankings. We evaluate LORE on over 30 graphs. LORE achieves strong performance on primary benchmarks, scales favorably to large graphs, and mechanism analyses further support its evidence-guided risk-allocation design.

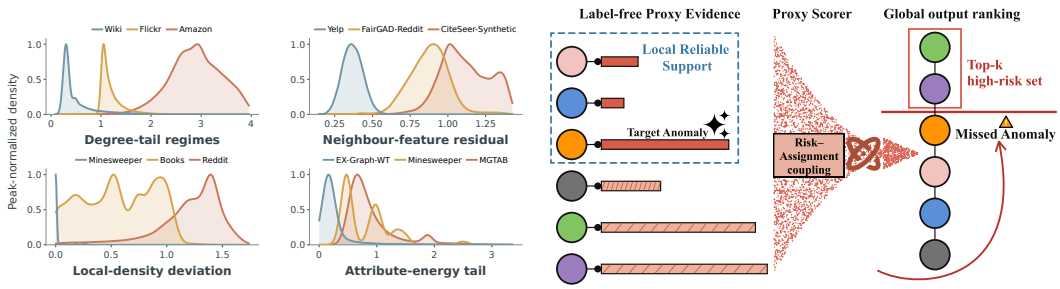
1 INTRODUCTION

Graph anomaly detection (GAD) aims to identify nodes whose behavior departs from the dominant regularities of a graph (Akoglu et al., 2015). It supports relational risk analysis in financial systems (Huang et al., 2022) and the detection of camouflaged malicious behavior on online platforms (Dou et al., 2020). Unlike a fixed semantic class, however, anomalousness is inherently contextual (Chandola et al., 2009): it may appear as an attribute extreme, neighbourhood inconsistency, structural irregularity, or atypical interaction (Qiao et al., 2024), and the same observable pattern may be benign under another graph context (Liu et al., 2022). This ambiguity is especially acute in the label-free setting. Because anomalies are rare and evolve over time, exhaustive annotations are difficult to obtain and cannot fully specify future abnormality (Ruff et al., 2021). A detector therefore cannot learn a task-specific anomaly concept directly; it must infer what is abnormal from the graph itself, using only its topology, node attributes, and relational context.

To operate without labels, mainstream GAD replaces direct supervision with proxies derived from the observed graph. The graph itself provides an internal reference of normality: structural statistics expose unusual connectivity (Akoglu et al., 2010), reconstruction residuals measure departures from dominant structure–attribute patterns (Ding et al., 2019), and self-supervised agreement evaluates whether a node is compatible with its learned neighbourhood context (Liu et al., 2021). Despite their different realizations, these methods follow a common decision path: construct graph-derived evidence, reduce it to a scalar anomaly score, and rank nodes accordingly. Multiple views are typically combined at the score level before ranking (Jin et al., 2021), while recent generalist models broaden the evidence that can be reused across graphs but retain the same proxy-to-score paradigm (Liu et al., 2024). This strategy makes label-free detection operational by replacing unavailable annotations with measurable deviations internal to the graph.

However, this proxy-based decision process leaves two deeper problems unresolved. First, *evidence reliability is graph- and support-dependent* (Figure 1a). Changes in topology, feature regime, anomaly prevalence, and data-generating process can alter the meaning of the same signal; broad benchmarking confirms that no unsupervised detector remains uniformly reliable across graphs and

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.



(a) Evidence varies across datasets.

(b) Coupling distorts global rankings.

Figure 1: Motivation for LORE. Label-free anomaly evidence varies across graphs, and directly coupling evidence with risk assignment may distort node rankings.

anomaly types (Liu et al., 2022). More subtly, anomaly evidence can be meaningful only relative to a particular structural context or network scale (Gutiérrez-Gómez et al., 2020); local-outlier analysis likewise treats the reference neighbourhood as part of the anomaly definition (Schubert et al., 2014). Discrimination within such a support does not justify unrestricted comparison with every node in the graph. The challenge is therefore not merely to identify a strong proxy, but to determine without labels *where its evidence is valid*. Second, *risk–assignment coupling obscures the role of evidence in ranking* (Figure 1b). Direct proxy-to-score mapping lets one scalar jointly construct global risk positions and assign nodes to them, leaving risk structure and node ownership inseparable. Score normalization can align numerical scales (Kriegel et al., 2011), but it does not separate these two roles. Local evidence may consequently induce unsupported movement across distant risk regions. A unified detector must instead distinguish the graph-level structure of available risk positions from the node-level question of *who gets the risk*.

To resolve these problems, we propose Mu**l**tiscale **O**ptimal T**r**ansport for Graph Anomaly D**e**tection (LORE), which recasts label-free GAD as *evidence-guided risk allocation*. Rather than allowing one score to determine both risk positions and their node ownership, LORE explicitly separates the two roles. A graph-adaptive empirical prior first constructs an ordered risk structure without treating it as the final anomaly ranking. Complementary self-supervised evidence then determines which nodes should occupy those positions through transport correspondence. Conditional OT reassigns ownership within prior-defined cells, and Partial OT further resolves the high-risk tail over a nested support. The resulting multiscale allocation preserves the graph-level risk structure, confines evidence to the support on which it is allowed to act, and adapts node ownership to the current graph. We summarize our contributions as follows:

- We identify a shared decision structure across diverse anomaly patterns. Although their manifestations and informative evidence vary across graphs, they can be characterized within a common ordered risk space. This observation provides a unified framework for label-free GAD.
- We reformulate graph anomaly detection as evidence-guided risk allocation and realize it through multiscale optimal transport, decoupling risk structure from node–position assignment. Conditional OT refines ownership within local risk cells, while Partial OT reallocates the high-risk tail over a nested support.
- We conduct extensive empirical analyses across over 30 graphs and diverse label-free baselines to validate the effectiveness and scalability of LORE. Moreover, we systematically investigate evidence correspondence, support restriction, prior transfer, and prior failure cases to reveal when and why risk allocation improves anomaly ranking.

2 Related Work

Label-free Graph Anomaly Detection. Label-free GAD has evolved through increasingly diverse ways of modeling abnormality. Early methods rely on hand-crafted cues, including structural laws (Akoglu et al., 2010), local connectivity patterns (Perozzi and Akoglu, 2016), and attribute residuals (Li et al., 2017). Learning-based methods subsequently model normality and deviation through

structure–attribute reconstruction (Fan et al., 2020), higher-order structural consistency (Yuan et al., 2021), and contrastive objectives (Zheng et al., 2023). More recent approaches explicitly encode mechanisms associated with graph anomalies, including adaptive propagation (Chen et al., 2024), smoothing dynamics (Dong et al., 2025), diffusion reconstruction (Li et al., 2025), heterophily (Pan et al., 2025), multiscale inconsistency (Lian et al., 2026), and uncertainty (Wei et al., 2025). Despite increasingly diverse anomaly modeling, the reliability of proxy signals still varies across graphs (Liu et al., 2022; Tang et al., 2023). LORE therefore shifts the focus from designing another anomaly proxy to establishing a unified label-free decision principle.

Anomaly Score Construction. Regardless of how abnormality is modeled, node-level detection ultimately requires a comparable decision variable. Existing methods therefore transform graph-derived anomaly signals into scalar node scores (Qiao et al., 2024). Reconstruction-based approaches typically use residual magnitude to quantify deviation (Ding et al., 2019), while contrastive methods convert representation disagreement into incompatibility scores (Liu et al., 2021). Density-based approaches measure departure from a reference distribution (Goix et al., 2015), whereas affinity-based approaches quantify deviation from relational patterns (Zhang et al., 2025). When several signals or detector configurations are available, these scalar outputs are further normalized (Kriegel et al., 2011), weighted or fused (Jin et al., 2021), or aggregated at the rank level (Goswami et al., 2023) to obtain a final node ordering. Detector-level methods similarly select candidate rankings (Fung et al., 2024) or combine them without labels (Rayana and Akoglu, 2016). Existing scoring jointly determines risk positions and node assignment. LORE instead decouples them through prior-defined risk and evidence-guided assignment.

Optimal Transport for Anomaly Detection. OT provides a principled way to compare distributions and establish sample correspondence through transport couplings (Peyré and Cuturi, 2019). In anomaly detection, robust OT formulations reduce the influence of contaminated or outlying samples (Mukherjee et al., 2021), while partial transport uses transported mass to identify OOD observations (Ren et al., 2024); unbalanced OT more generally relaxes full mass conservation (Chizat et al., 2018). Other studies use Wasserstein distance (Craig et al., 2024), conditional transport entropy (Feng et al., 2024), or mass-repulsing transport cost (Fernandes Montesuma et al., 2025) as anomaly scores. OT distances have also been used to retrieve suitable detectors from related datasets (Singh and Vanschoren, 2023). More recently, X-GAD uses Wasserstein and Gromov–Wasserstein geometry to model inter-graph anomalies for graph-level detection (Zhang et al., 2026), while GTGD uses the same geometries to align node and structural patterns across graph domains (Huang et al., 2025). LORE instead establishes node–risk-position correspondence within a single graph and progressively refines risk assignment across multiple supports.

3 METHODOLOGY

3.1 PRELIMINARIES

Given an unlabeled attributed graph $G = (V, E, X)$, where V, E, and X denote the node set, edge set, and node attributes, respectively, our goal is to produce an anomaly-risk ranking s_i for each node without access to anomaly labels. Unlike supervised anomaly detection, the absence of labels prevents direct learning of the mapping from evidence to anomalies, requiring anomaly assessment to rely on observable graph statistics, node attributes, and local context.

3.2 RANK CALIBRATION AND GRAPH-DERIVED RISK MEASURE

We consider multiple anomaly views, termed *charts*, each of which produces a node-level anomaly signal. Because their output scales may differ, we map all chart outputs into a common evidence space using the same permutation-invariant empirical mid-rank:

$$e_i^{(j)} \equiv R\Big(h_i^{(j)}\Big) = \frac{r_i^{(j)}}{N}, \text{ where } r_i^{(j)} = \operatorname{rank}\Big(h_i^{(j)} \mid \{h_1^{(j)}, \ldots, h_N^{(j)}\}\Big). \tag{1}$$

Here $h_i^{(j)}$ is the raw score of node i produced by chart j, $r_i^{(j)}$ is its average rank within that chart over all N nodes, and $R(\cdot)$ is the common empirical mid-rank map, so $e_i^{(j)} \in [0, 1]$ is the corresponding

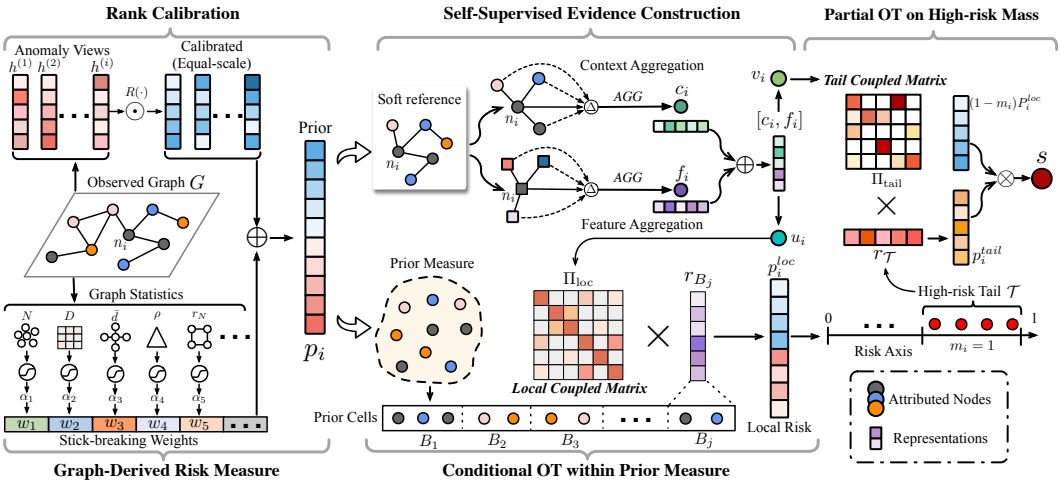


Figure 2: Overview of LORE. LORE reformulates graph anomaly detection as a risk allocation problem, decoupling graph-level risk structure from node ownership and using multiscale optimal transport to assign risk positions to nodes.

rank-normalized evidence. The J anomaly views $\{h_i^{(j)}\}_{j=1}^{J}$ are computed from the observed graph; their complete constructions are provided in Appendix A. All views depend only on graph information and use neither graph identity nor anomaly labels. For combining upper-tail evidence, a rank is converted to a positive unit-mean e-value,

$$E\left(e_i^{(j)}\right) = \frac{-\log\left((N+1-r_i^{(j)})/(N+1)\right)}{N^{-1}\sum_{m=1}^{N} -\log\left((N+1-r_m^{(j)})/(N+1)\right)}. \tag{2}$$

Thus every strictly increasing transformation of a chart produces the same prior ordering, and ties never depend on node storage order.

Where the chart weights come from. For each anomaly view, we first determine its weight based on the global statistics of the input graph. Specifically, we compute the graph-level descriptor $g_G = (N, D, \bar{d}, \rho, \nu, r_N)$ directly from the observed graph. These weights are deterministic functions of the descriptor rather than fitted parameters, with implementation details provided in Appendix A.1. For each view, let x denote the evaluated graph statistic and t its reference scale. We quantify the signed deviation of x from t as follows:

$$\ell(x,t) = \log(x/t), \qquad d(x,t) = x - t. \tag{3}$$

Here $\ell(x,t)$ measures relative deviation for positive graph statistics, whereas $d(x,t)$ measures signed absolute deviation for bounded directional statistics. In both cases a positive response places x above its reference t, and a negative response places it below. These responses determine how strongly each node-level chart signal $h_i^{(j)}$ enters the prior. The chart activations are written uniformly as

$$\alpha_j = \mathcal{C}_j\left(S_\tau(z_1), \ldots, S_\tau(z_{m_j})\right), \text{ where } z_k \in \{\ell(x_k,t_k),\ d(x_k,t_k)\}. \tag{4}$$

Here $S_\tau(z) = \sigma(z/\tau)$ is a deterministic soft step with temperature $\tau = 0.1$; reversing the sign of z expresses the opposite direction without introducing a second gate. The function $\mathcal{C}_j(\cdot)$ is the chart-specific deterministic composition of its directional soft contrasts; it may combine them by products, complements, or their fixed compositions. Consequently these shared components produce $\boldsymbol{\alpha}_G = (\alpha_1, \ldots, \alpha_J) \in [0,1]^J$. Their complete graph-statistic wiring, including the auxiliary contrast used only by α_2, is stated in Appendix A.2, Equation A.14. An ordered stick break maps the activations to simplex weights. The last activation is one, so the weights sum exactly to one. The final prior is

$$w_j = \alpha_j \prod_{\ell<j}(1-\alpha_\ell), \qquad p_i = R\left(\sum_{j=1}^{J} w_j E\left(e_i^{(j)}\right)\right). \tag{5}$$

The chart-weight mechanism determines the weight of each view solely from graph-level statistics and fixed reference values, rather than through additional supervision or parameter fitting, thereby providing a deterministic graph-adaptive risk initialization for different graphs. This prior serves only as empirical risk mass for subsequent risk reassignment.

3.3 SELF-SUPERVISED EVIDENCE CONSTRUCTION

To extract node-level evidence beyond the initial prior, we use low-risk nodes to construct a soft normal-reference distribution for the self-supervised views. Specifically, node i is sampled according to $q_i = (1-p_i)^\gamma / \sum_{m=1}^{N}(1-p_m)^\gamma$, with $\gamma = 2$, so that low-risk nodes receive higher sampling probabilities. Guided by this reference distribution, the first view compares each node representation with its neighbourhood context:

$$n_i = \frac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)} z_j, \qquad a_i = \phi_x(z_i), \qquad b_i = \phi_n(n_i). \tag{6}$$

Here, z_i is the standardized feature vector of node i, $\mathcal{N}(i)$ is its leave-self-out symmetric neighbourhood, and ϕ_x and ϕ_n are the attribute and neighbourhood encoders, respectively. Independently, the second view uses an autoencoder ψ_θ to capture feature-level coherence by reconstructing coordinates selected by a binary mask $M_i \in \{0,1\}^D$, where $M_{id} = 1$ indicates that the d-th feature dimension is hidden:

$$\widehat{z}_i = \psi_\theta((\mathbf{1} - M_i) \odot z_i)\,. \tag{7}$$

Together, the two objectives expose complementary departures from the soft reference: neighbourhood mismatch and masked-feature reconstruction error:

$$\mathcal{L}_{\mathrm{ctx}} = -\frac{1}{2}\mathbb{E}_i[\log\sigma(\rho_i^+) + \log\sigma(-\rho_i^-)], \qquad \mathcal{L}_{\mathrm{mask}} = \mathbb{E}_i\left[\frac{\sum_d M_{id}(\widehat{z}_{id} - z_{id})^2}{\sum_d M_{id}}\right]. \tag{8}$$

Here $\rho_i^+ = a_i^\top b_i$ is the matched node–neighbourhood similarity, whereas $\rho_i^- = a_i^\top b_{\pi(i)}$ is its mismatched negative, with π denoting a random permutation.

The two views encode complementary notions of normality. Specifically, context incompatibility captures neighbourhood inconsistency, while feature coherence measures attribute reconstruction consistency. We map both signals to graph-wise ranks, yielding context incompatibility c_i and feature coherence f_i; suspicious nodes are expected to have large c_i or small f_i:

$$f_i = 1 - R\left(\max_d(\widehat{z}_{id} - z_{id})^2\right), \qquad c_i = R\big(-a_i^\top b_i\big)\,. \tag{9}$$

To improve evidence stability, we aggregate multiple self-supervised replicas by node-wise median, $c_i = \mathrm{median}_{r=1}^{3}\, c_i^{(r)}$ and $f_i = \mathrm{median}_{r=1}^{3}\, f_i^{(r)}$. The two signals are then aggregated into the node-level evidence state $\mathbf{e}_i$, which is further projected into two OT transport coordinates, u_i and v_i, for local and tail risk allocation, respectively:

$$\mathbf{e}_i = [c_i, f_i]^\top, \qquad u_i = \kappa_{\mathrm{loc}}(\mathbf{e}_i), \qquad v_i = \kappa_{\mathrm{tail}}(\mathbf{e}_i). \tag{10}$$

Here κ_{loc} and κ_{tail} map the evidence state to scalar transport coordinates. The two OT stages use u_i and v_i to induce distinct transport geometries for local refinement and high-risk-tail allocation. The complete function family is given in Appendix A.3.

3.4 CONDITIONAL OT WITHIN THE PRIOR MEASURE

We partition the continuous prior-risk space into K value-based cells to preserve the coarse-grained risk structure while allowing local risk refinement:

$$B_k = \{i : \min(K-1, \lfloor K p_i \rfloor) = k\}, \qquad k = 0, \ldots, K-1. \tag{11}$$

Within each cell B_k, we define $n_k = |B_k|$ uniformly spaced target risk slots to enable OT-based reassignment while preserving the coarse-grained prior-risk structure:

$$q_{k\ell} = \frac{k + (\ell - \frac{1}{2})/n_k}{K}, \qquad \ell = 1, \ldots, n_k. \tag{12}$$

Within each occupied cell, every node and target risk slot is assigned uniform mass $1/n_k$. Using the local coordinate u_i from Equation 10, Conditional OT solves:

$$\begin{aligned} \pi_k^* = \ & \arg\min_{\pi \geq 0} \sum_{i \in B_k} \sum_{\ell=1}^{n_k} \pi_{i\ell} (u_i - q_{k\ell})^2 \\ \text{subject to} \quad & \sum_{\ell=1}^{n_k} \pi_{i\ell} = \frac{1}{n_k} \quad (i \in B_k), \qquad \sum_{i \in B_k} \pi_{i\ell} = \frac{1}{n_k} \quad (\ell = 1, \ldots, n_k). \end{aligned} \tag{13}$$

Since the OT problem is one-dimensional with quadratic cost, its optimal coupling can be obtained efficiently by monotone matching. The resulting assignments are rank-calibrated to obtain p^{loc}. Restricting transport to prior cells limits the scope over which self-supervised evidence must provide a reliable ordering. Evidence need only resolve ranking ambiguity among nodes with similar prior risk, making its local discriminative value useful even when its reliability varies across the graph.

3.5 Partial OT on high-risk mass

Conditional OT refines positions within prior cells, but ambiguity may remain in the decision-critical high-risk region. Partial OT optionally refines this tail through further evidence-guided reassignment. For a transported mass ratio $m \in (0, 1)$, we select the top $M_0 = \max(2, \lceil mN \rceil)$ nodes under p^{loc}, including all boundary ties. This forms the tail $\mathcal{T}$ with $M = |\mathcal{T}| \geq M_0$ and makes selection independent of node order. The tail target slots are

$$q_\ell^{\mathrm{tail}} = \frac{N - M + \ell - \frac{1}{2}}{N}, \qquad \ell = 1, \ldots, M. \tag{14}$$

Using v_i, Partial OT reassigns high-risk-tail nodes to target risk slots by solving:

$$\begin{aligned} \pi_{\mathrm{tail}}^* = \ & \arg\min_{\pi \geq 0} \sum_{i \in \mathcal{T}} \sum_{\ell=1}^{M} \pi_{i\ell} (v_i - q_\ell^{\mathrm{tail}})^2 \\ \text{subject to} \quad & \sum_{\ell=1}^{M} \pi_{i\ell} = \frac{1}{M} \quad (i \in \mathcal{T}), \qquad \sum_{i \in \mathcal{T}} \pi_{i\ell} = \frac{1}{M} \quad (\ell = 1, \ldots, M). \end{aligned} \tag{15}$$

While Conditional OT resolves local ambiguity within prior cells, Partial OT concentrates evidence-guided correction on the decision-critical high-risk tail. Nodes outside $\mathcal{T}$ retain their incoming risk, restricting tail evidence to the region where it is most reliable. Within $\mathcal{T}$, it reallocates existing tail risk positions according to v_i, refining node ownership while preserving the prior-defined tail structure. Let $\Phi_{\mathrm{loc}}^{K,\kappa_{\mathrm{loc}}}$ and $\Phi_{\mathrm{tail}}^{m,\kappa_{\mathrm{tail}}}$ denote the risk-slot assignment operators defined by Equations 13 and 15, respectively. The final anomaly score is the empirical mid-rank of their composed assignment:

$$s = R\Big[\Big(\Phi_{\mathrm{tail}}^{m,\kappa_{\mathrm{tail}}} \circ \Phi_{\mathrm{loc}}^{K,\kappa_{\mathrm{loc}}}\Big)(p; \mathbf{e})\Big]. \tag{16}$$

The composition of Conditional and Partial OT decouples risk structure from node ownership. The prior defines the global organization of risk positions, while self-supervised evidence determines which nodes occupy these positions through local and tail-level reassignment. Rather than altering the risk measure itself, LORE preserves the prior risk structure and reallocates node ownership through explicit transport couplings.

4 Experiments

4.1 Experimental setup

Datasets and protocols. We evaluate LORE on over 30 graph datasets from diverse domains, with detailed dataset statistics provided in the appendix. Performance is evaluated using both AP and AUROC; representative results are reported in the main paper, with complete results in the appendix. All results are averaged over five random seeds (1–5).

Baselines. We compare against fourteen label-free graph anomaly detection methods, covering classical structural statistics (SCAN (Xu et al., 2007) and OddBall (Akoglu et al., 2010)), reconstruction

Method	Books	Weibo	Reddit	Disney	Enron	Mine.	Tol.	Quest.	Amazon	Yelp
HUGE (AAAI'25)	0.0225	0.1562	0.0500	0.0556	0.0003	0.2890	0.1572	0.0306	OOM	OOM
SCAN (KDD'07)	0.0207	0.0625	0.0283	0.0956	0.0003	0.2005	0.2129	0.0227	0.1759	0.1539
OddBall (PAKDD'10)	0.0245	0.0869	0.0370	0.0704	0.0006	0.1996	0.2968	0.0566	0.1475	0.1510
Radar (IJCAI'17)	0.0269	0.2685	0.0492	0.0547	0.0003	0.2282	0.1795	0.0438	0.3416	0.2363
ANEMONE (CIKM'21)	0.0267	0.0717	0.0337	0.0582	0.0010	0.2240	0.2474	0.0375	0.0585	0.1369
AD-GCL (AAAI'25)	0.0144	0.0967	0.0378	0.0908	0.0003	0.1984	0.2216	0.0300	0.0572	0.1820
DiffGAD (ICLR'25)	0.0218	0.7204	0.0358	0.0574	0.0009	0.2012	0.2681	OOM	0.0606	OOM
SmoothGNN (WWW'25)	0.0190	0.5578	0.0291	0.0671	0.0008	0.2015	0.2578	0.0385	0.1033	0.1377
GAD-NR (WSDM'24)	0.0216	0.2944	0.0382	0.0619	0.0005	0.1900	0.2448	0.0515	0.1548	0.1834
MI-GAD (AAAI'26)	0.0197	0.4046	0.0408	0.0717	0.0003	0.2317	0.1787	0.0247	0.4098	0.2012
GEL (KDD'25)	0.0405	0.1475	0.0432	0.0528	0.0004	0.2276	0.2065	0.0436	0.2503	0.1779
GADAM (ICLR'24)	0.0240	0.0904	0.0483	0.0654	0.0003	0.2020	0.1903	0.0479	0.0542	0.1342
ADA-GAD (AAAI'24)	0.0157	0.7371	0.0376	0.0436	0.0007	0.2096	0.2765	OOM	0.0400	OOM
GCTAM (IJCAI'25)	0.0211	0.2509	0.0360	0.0839	0.0005	0.2173	0.2079	OOM	0.0980	OOM
LORE (Ours)	**0.1526**	**0.8542**	**0.0563**	**0.3443**	**0.1064**	**0.3501**	**0.3878**	**0.0720**	**0.5776**	**0.2594**

Table 1: Average precision on Main-10 benchmark graphs.

and residual modeling (Radar (Li et al., 2017), GAD-NR (Roy et al., 2024), and MI-GAD (Lian et al., 2026)), contrastive or augmentation objectives (ANEMONE (Jin et al., 2021), AD-GCL (Xu et al., 2025), and GCTAM (Zhang et al., 2025)), and recent smoothing, diffusion, and generalist approaches (SmoothGNN (Dong et al., 2025), DiffGAD (Li et al., 2025), HUGE (Pan et al., 2025), GEL (Wei et al., 2025), GADAM (Chen et al., 2024), and ADA-GAD (He et al., 2024)). We use the publicly released implementations and record out-of-memory cases as OOM.

4.2 MAIN RESULTS

Comparison results. Table 1 reports the AP results on Main-10. Complete AP and AUROC results across all datasets are provided in Appendix Tables 9–11. LORE consistently outperforms across graphs with varying scales, structures, and anomaly semantics, suggesting that it is not specialized to a particular anomaly pattern but instead realizes a transferable risk-reassignment mechanism.

Scalability. To evaluate scalability, we conduct experiments on three large-scale datasets. As shown in Table 2, LORE scales reliably to graphs with millions of nodes and achieves the best AP on all three datasets. The result shows that rank-based risk allocation remains practical as graph size grows.

Method	Elliptic	T-Finance	T-Social
FreeGAD	0.1124	0.7882	0.0496
HUGE	0.0709	0.3624	0.0337
GADAM	0.1021	0.1128	0.0372
LORE (current)	**0.6408**	**0.8013**	**0.4508**

Table 2: Performance on large graphs.

Label-free vs. few-shot. We further compare LORE with several few-shot methods. As shown in Table 3, LORE achieves the highest overall macro AP in a fully label-free setting and obtains the best performance on nine of the ten datasets, with SpaceGNN on Amazon as the only exception. These results show that evidence-guided risk allocation can achieve performance comparable to or even stronger than few-shot methods without additional supervision.

Method	Books	Weibo	Reddit	Disney	Enron	Mine.	Tol.	Quest.	Amazon	Yelp
SpaceGNN (5N+5A)	0.0257	0.4581	0.0411	0.0360	0.0001	0.2681	0.3074	0.0329	**0.7667**	0.1753
ARC (5N)	0.0160	0.6598	0.0440	0.0550	0.0004	0.2154	0.2287	0.0476	0.0679	0.1878
AnomalyGFM-N (5N)	0.0196	0.1206	0.0277	0.1428	0.0003	0.2590	0.2034	0.0341	0.2597	0.1520
AnomalyGFM-A (5A)	0.0172	0.1190	0.0286	0.1789	0.0002	0.2562	0.2080	0.0338	0.2507	0.1489
APF (5N+5A)	0.0145	0.7174	0.0431	0.0369	0.0005	0.3152	0.3055	0.0330	0.3990	0.1685
LORE (0)	**0.1526**	**0.8542**	**0.0563**	**0.3443**	**0.1064**	**0.3501**	**0.3878**	**0.0720**	0.5776	**0.2594**

Table 3: Few-shot AP on Main-10 (N/A denote normal/anomalous support).

Variant	Books	Weibo	Reddit	Disney	Enron	Mine.	Tol.	Quest.	Amazon	Yelp	Macro
Risk prior	0.1171	0.8415	0.0503	0.2500	0.0464	0.3014	0.3857	0.0574	0.5450	0.2520	0.2847
Prior + additive evidence	0.1089	0.8364	0.0502	0.2499	0.0719	0.3061	0.3856	0.0572	0.5376	0.2536	0.2857
Shuffled correspondence	0.1286	0.8259	0.0509	0.2566	0.0910	0.3030	0.3840	0.0573	0.5546	0.2378	0.2890
Conditional OT only	**0.1526**	0.8481	0.0532	0.2528	0.0697	0.3366	0.3868	0.0532	0.5709	0.2544	0.2978
Partial OT only	0.1171	0.8444	0.0499	0.3435	**0.1064**	0.3347	0.3860	0.0628	0.5644	0.2533	0.3063
LORE	**0.1526**	**0.8542**	**0.0563**	**0.3443**	**0.1064**	**0.3501**	**0.3878**	**0.0720**	**0.5776**	**0.2594**	**0.3161**

Table 4: Ablations of score correction, evidence correspondence, and LORE's OT components.

4.3 MECHANISM ANALYSIS

Support-dependent evidence. We further investigate why support-restricted risk reassignment is preferable to ordinary score fusion. Using the common weighted fusion of normalized anomaly scores, we construct a minimal operator-replacement control, $s_\alpha = R((1-\alpha)p^{\text{loc}} + \alpha R(v))$. With the same p^{loc} and v, we select α without labels for each graph–seed pair across five random seeds to match LORE's mean ranking displacement (Figure 3(a)). Under this matched condition, Figure 3(b) shows that Partial OT consistently outperforms ordinary fusion across all graph–seed pairs, indicating that the benefit depends on which nodes receive the risk changes rather than on the magnitude of ranking perturbation. Furthermore, Figure 3(c) fixes α and progressively corrupts evidence outside the tail; global fusion degrades as corruption increases, while Partial OT remains stable, showing that evidence effectiveness depends on the support within which it is allowed to alter risk.

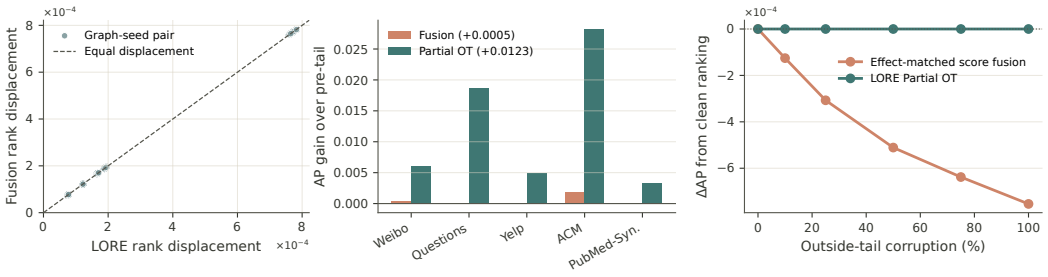


Figure 3: Equal ranking change does not imply equal risk reassignment. (a) Matched ranking change. (b) Risk recipients. (c) Unsupported-evidence sensitivity.

Component ablations. Table 4 shows that directly adding the learned evidence to the prior as a residual yields only a limited gain, whereas the complete LORE produces a substantially larger improvement. The gain is therefore attributable to evidence-driven risk reassignment rather than the direct injection of an additional anomaly score. Performance also drops markedly when node–evidence correspondence is shuffled while the evidence distribution is preserved, demonstrating that LORE relies on the correct correspondence between nodes and evidence rather than arbitrary OT-based reordering. Finally, Conditional OT and Partial OT are each effective in isolation, while their combination performs best, confirming the complementarity of the two risk scales. Appendix Figure 6 provides the corresponding graph-wise decomposition.

Evidence correspondence. To verify whether LORE's gain truly comes from the evidence correspondence rather than arbitrary OT-induced reordering, we progressively disrupt only the correspondence. Figure 4(a) shows that the retained gain over the prior decreases steadily as the correspondence is corrupted, with full shuffling removing 91.4% of the original gain. Figure 4(b) further shows that this degradation trend is observed on most datasets, although the sensitivity to correspondence corruption varies across graphs. Figure 4(c) confirms the same pattern at the graph–seed level: as corruption increases, the number of graph–seed pairs that outperform the prior drops from 50 to 29.

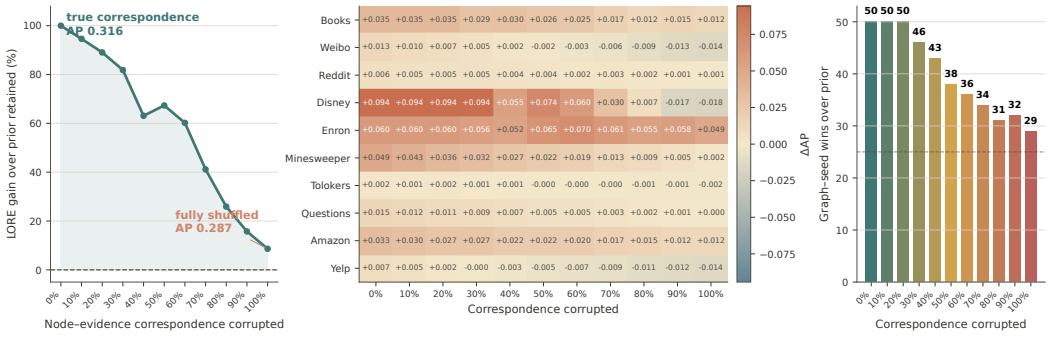


Figure 4: Node–evidence correspondence stress test. (a) Correspondence decay. (b) Gain over the same prior. (c) Pairwise consistency.

Prior transfer and risk recovery. Table 5 compares the frozen risk prior with three classical label-free detectors on five unseen graphs. Its leading macro AP supports cross-graph transfer, although the prior can still fail elsewhere. Table 6 then focuses on graphs where the prior fails. Across all six cases, LORE uses node evidence to reassign misallocated risk positions and recover a more appropriate ranking. This separates transfer of the risk structure from recovery of node ownership, showing that LORE can correct rather than merely refine the prior.

Method	MGTAB	YelpChi	YelpNYC	FairGAD-R	EX-WT	Macro
PCA	0.3301	0.2113	0.1255	0.1954	**0.0122**	0.1749
Isolation Forest	0.2869	**0.2671**	0.1458	0.2022	0.0098	0.1824
HBOS	0.2854	0.2242	**0.1528**	0.2052	0.0061	0.1747
Frozen prior	**0.4700**	0.2520	0.1263	**0.3379**	0.0089	**0.2390**

Table 5: Transfer to additional unseen graphs.

Dataset	Random	PCA	IF	HBOS	Prior	LORE
Cresci-15	0.6321	0.6242	0.6399	0.6119	0.5264	**0.7428**
Bitcoin OTC	0.5696	0.5696	0.5696	0.5696	0.5752	**0.8342**
REV2 Epinions	0.0709	0.0709	0.0709	0.0709	0.0536	**0.1784**
DIAM Ethereum-S	0.4940	0.3184	0.3307	0.3794	0.4516	**0.8086**
Elliptic++	0.0538	0.0970	0.0530	0.0763	0.0437	**0.2566**
DIAM Bitcoin-M	0.1805	0.1346	0.1265	0.1344	0.1322	**0.2963**
Macro	0.3335	0.3024	0.2984	0.3071	0.2971	**0.5195**

Table 6: Recovery from prior failure with OT.

Adaptive risk recovery. Figure 5 groups nodes into ten deciles according to the initial prior risk. The x-axis indicates each node's prior risk region, while the y-axis reports the mean rank displacement $s_i - p_i$ induced by LORE, where positive values indicate movement toward higher-risk positions; anomalous-node displacement on Main-10 (left) is concentrated near the ninth prior decile, whereas on Additional-10 (right), the largest promotion shifts to the fourth through sixth deciles; meanwhile, normal nodes exhibit relatively small overall displacement. This pattern shows that LORE does not correct rankings within a fixed risk region. Instead, the location of risk reassignment adapts to the quality of the prior: when the prior is informative, LORE mainly refines the high-risk tail, whereas under a weaker prior, it recovers missed anomalies from ambiguous middle ranks, shifting from tail refinement to mid-rank recovery.

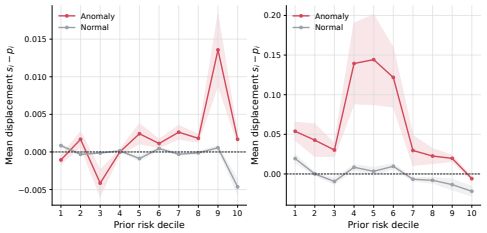


Figure 5: Risk displacement by prior decile.

5 Conclusion

Motivated by the heterogeneity of anomaly evidence across graphs, we recast label-free graph anomaly detection as an evidence-guided risk allocation problem. LORE decouples risk structure from node ownership and resolves their evidence-conditioned correspondence through multiscale optimal transport. Extensive experiments show strong overall performance and consistent refinement of the initial risk prior. Mechanism analyses further support the proposed risk-allocation principle, and large-scale evaluations demonstrate good scalability.

REFERENCES

Leman Akoglu, Mary McGlohon, and Christos Faloutsos. Oddball: Spotting anomalies in weighted graphs. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 410–421. Springer, 2010.

Leman Akoglu, Hanghang Tong, and Danai Koutra. Graph based anomaly detection and description: A survey. *Data Mining and Knowledge Discovery*, 29:626–688, 2015. doi: 10.1007/s10618-014-0365-y.

Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM Computing Surveys*, 41(3):1–58, 2009. doi: 10.1145/1541880.1541882.

Jingyan Chen, Guanghui Zhu, Chunfeng Yuan, and Yihua Huang. Boosting graph anomaly detection with adaptive message passing. In *International Conference on Learning Representations*, 2024.

Lénaïc Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. Scaling algorithms for unbalanced optimal transport problems. *Mathematics of Computation*, 87(314):2563–2609, 2018. doi: 10.1090/mcom/3303.

Nathaniel Craig, Jessica N. Howard, and Hancheng Li. Exploring optimal transport for event-level anomaly detection at the large hadron collider. *arXiv preprint arXiv:2401.15542*, 2024.

Kaize Ding, Jundong Li, Rohit Bhanushali, and Huan Liu. Deep anomaly detection on attributed networks. In *Proceedings of the 2019 SIAM International Conference on Data Mining*, pages 594–602, 2019. doi: 10.1137/1.9781611975673.67.

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 Web Conference*, pages 1225–1236, 2025. doi: 10.1145/3696410.3714615.

Yingtong Dou, Zhiwei Liu, Li Sun, Yutong Deng, Hao Peng, and Philip S. Yu. Enhancing graph neural network-based fraud detectors against camouflaged fraudsters. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management*, pages 315–324, 2020. doi: 10.1145/3340531.3411903.

Haoyi Fan, Fengbin Zhang, and Zuoyong Li. AnomalyDAE: Dual autoencoder for anomaly detection on attributed networks. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 5685–5689, 2020. doi: 10.1109/ICASSP40776.2020.9053555.

Chuanwen Feng, Wenlong Chen, Ao Ke, Yilong Ren, Xike Xie, and S. Kevin Zhou. Detecting out-of-distribution samples via conditional distribution entropy with optimal transport. *arXiv preprint arXiv:2401.11726*, 2024.

Eduardo Fernandes Montesuma, Adel El Habazi, and Fred Ngolè Mboula. Unsupervised anomaly detection through mass repulsing optimal transport. *Transactions on Machine Learning Research*, 2025.

Clement Fung, Chen Qiu, Aodong Li, and Maja Rudolph. Model selection of anomaly detectors in the absence of labeled validation data. *arXiv preprint arXiv:2310.10461*, 2024.

Nicolas Goix, Anne Sabourin, and Stéphan Clémençon. On anomaly ranking and excess-mass curves. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, volume 38, pages 287–295, 2015.

Mononito Goswami, Cristian Challu, Laurent Callot, Lenon Minorics, and Andrey Kan. Unsupervised model selection for time-series anomaly detection. In *International Conference on Learning Representations*, 2023.

Leonardo Gutiérrez-Gómez, Alexandre Bovet, and Jean-Charles Delvenne. Multi-scale anomaly detection on attributed networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(1):678–685, 2020. doi: 10.1609/aaai.v34i01.5409.

Junwei He, Qianqian Xu, Yangbangyan Jiang, Zitai Wang, and Qingming Huang. ADA-GAD: Anomaly-denoised autoencoders for graph anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 8481–8489, 2024.

Changqin Huang, Xinxing Shi, Chao Gao, Qian Hu, Xiaodi Huang, Qiang Huang, and Ali Anaissi. Cross-domain graph anomaly detection via graph transfer and graph decouple. *CAAI Transactions on Intelligence Technology*, 10(4):1089–1103, 2025. doi: 10.1049/cit2.70014.

Xuanwen Huang, Yang Yang, Yang Wang, Chunping Wang, Zhisheng Zhang, Jiarong Xu, Lei Chen, and Michalis Vazirgiannis. DGraph: A large-scale financial dataset for graph anomaly detection. In *Advances in Neural Information Processing Systems*, volume 35, 2022. doi: 10.52202/068431-1654.

Ming Jin, Yixin Liu, Yu Zheng, Lianhua Chi, Yuan-Fang Li, and Shirui Pan. ANEMONE: Graph anomaly detection with multi-scale contrastive learning. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management*, pages 3122–3126, 2021. doi: 10.1145/3459637.3482057.

Hans-Peter Kriegel, Peer Kröger, Erich Schubert, and Arthur Zimek. Interpreting and unifying outlier scores. In *Proceedings of the 2011 SIAM International Conference on Data Mining*, pages 13–24, 2011. doi: 10.1137/1.9781611972818.2.

Jinghan Li, Yuan Gao, Jinda Lu, Junfeng Fang, Congcong Wen, Hui Lin, and Xiang Wang. DiffGAD: A diffusion-based unsupervised graph anomaly detector. In *International Conference on Learning Representations*, 2025.

Jundong Li, Harsh Dani, Xia Hu, and Huan Liu. Radar: Residual analysis for anomaly detection in attributed networks. In *International Joint Conference on Artificial Intelligence*, pages 2152–2158, 2017. doi: 10.24963/ijcai.2017/299.

Jie Lian, Zhihao Wu, Jielong Lu, Jiajun Yu, Qianqian Shen, and Haishuai Wang. Beyond local patterns: Multiscale inconsistency learning for graph anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 15216–15224, 2026. doi: 10.1609/aaai.v40i18.38546.

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, pages 27021–27035, 2022.

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*, 2021. 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, 2024.

Debarghya Mukherjee, Aritra Guha, Justin M. Solomon, Yuekai Sun, and Mikhail Yurochkin. Outlier-robust optimal transport. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pages 7850–7860, 2021.

Junjun Pan, Yixin Liu, Xin Zheng, Yizhen Zheng, Alan Wee-Chung Liew, Fuyi Li, and Shirui Pan. A label-free heterophily-guided approach for unsupervised graph fraud detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(12):12443–12451, 2025. doi: 10.1609/aaai.v39i12.33356.

Bryan Perozzi and Leman Akoglu. Scalable anomaly ranking of attributed neighborhoods. In *Proceedings of the 2016 SIAM International Conference on Data Mining*, pages 207–215, 2016. doi: 10.1137/1.9781611974348.24.

Gabriel Peyré and Marco Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5–6):355–607, 2019. doi: 10.1561/2200000073.

Hezhe Qiao, Hanghang Tong, Bo An, Irwin King, Charu Aggarwal, and Guansong Pang. Deep graph anomaly detection: A survey and new perspectives. *arXiv preprint arXiv:2409.09957*, 2024.

Shebuti Rayana and Leman Akoglu. Less is more: Building selective anomaly ensembles. *ACM Transactions on Knowledge Discovery from Data*, 10(4):42:1–42:33, 2016. doi: 10.1145/2890508.

Yilong Ren, Chuanwen Feng, Xike Xie, and S. Kevin Zhou. Partial optimal transport based out-of-distribution detection for open-set semi-supervised learning. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pages 4851–4859, 2024. doi: 10.24963/ijcai.2024/536.

Amit Roy, Juan Shu, Jia Li, Carl Yang, Olivier Elshocht, Jeroen Smeets, and Pan Li. GAD-NR: Graph anomaly detection via neighborhood reconstruction. In *Proceedings of the ACM International Conference on Web Search and Data Mining*, pages 576–585, 2024. doi: 10.1145/3616855.3635767.

Lukas Ruff, Jacob R. Kauffmann, Robert A. Vandermeulen, Grégoire Montavon, Wojciech Samek, Marius Kloft, Thomas G. Dietterich, and Klaus-Robert Müller. A unifying review of deep and shallow anomaly detection. *Proceedings of the IEEE*, 109(5):756–795, 2021. doi: 10.1109/JPROC.2021.3052449.

Erich Schubert, Arthur Zimek, and Hans-Peter Kriegel. Local outlier detection reconsidered: A generalized view on locality with applications to spatial, video, and network outlier detection. *Data Mining and Knowledge Discovery*, 28(1):190–237, 2014. doi: 10.1007/s10618-012-0300-z.

Prabhant Singh and Joaquin Vanschoren. AutoML for outlier detection with optimal transport distances. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 7175–7178, 2023. doi: 10.24963/ijcai.2023/843.

Jianheng Tang, Fengrui Hua, Ziqi Gao, Peilin Zhao, and Jia Li. GADBench: Revisiting and benchmarking supervised graph anomaly detection. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-1289.

Chunyu Wei, Wenji Hu, Xingjia Hao, Yunhai Wang, Yueguo Chen, Bing Bai, and Fei Wang. Graph evidential learning for anomaly detection. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2025. doi: 10.1145/3711896.3736989.

Xiaowei Xu, Nurcan Yuruk, Zhidan Feng, and Thomas A. J. Schweiger. SCAN: A structural clustering algorithm for networks. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 824–833, 2007.

Yiming Xu, Zhen Peng, Bin Shi, Xu Hua, Bo Dong, Song Wang, and Chen Chen. Revisiting graph contrastive learning on anomaly detection: A structural imbalance perspective. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 12972–12980, 2025.

Xu Yuan, Na Zhou, Shuo Yu, Huafei Huang, Zhikui Chen, and Feng Xia. Higher-order structure based anomaly detection on attributed networks. In *2021 IEEE International Conference on Big Data*, pages 2691–2700, 2021.

Xinqiu Zhang, Wei Qin, Tairan Huang, and Jianliang Gao. X-GAD: Explainable graph-level anomaly detection via inter-graph anomalies. *Expert Systems with Applications*, 296:128902, 2026. doi: 10.1016/j.eswa.2025.128902.

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 *International Joint Conference on Artificial Intelligence*, pages 3642–3650, 2025.

Yu Zheng, Ming Jin, Yixin Liu, Lianhua Chi, Khoa T. Phan, and Yi-Ping Phoebe Chen. Generative and contrastive self-supervised learning for graph anomaly detection. *IEEE Transactions on Knowledge and Data Engineering*, 35(12):12220–12233, 2023. doi: 10.1109/TKDE.2021.3119326.

A Complete Construction of the Six Prior Charts

The prior is a graph-adaptive mixture of six interpretable risk charts. Each chart describes a different way in which a node can be unusual—through finite-sample response, canonical attribute energy, exposure-normalised load, edge coherence, diffusion alignment, or scale contrast. The graph-level controller decides how much risk mass each chart contributes; it does not alter the node ordering within a chart. This separation lets the same chart bank cover graphs with very different sizes, feature dimensions, and local geometry. The six charts are evaluated in parallel and appear below in their stick-breaking order. We use R and E for the mid-rank and unit-mean upper-tail maps in Equations 1 and 2, Z for scalar standardisation, and $[x]_+ = \max(x, 0)$.

Chart 1: finite-sample signed response. Let ϕ_{im} be the mth nondegenerate empirical spline atom formed from the graph's complete basic coordinate state, $\bar{\phi}_{im}$ its symmetric neighbour mean, and $\mathcal{S}_m$ its observed edge Dirichlet smoothness. Smooth atoms are more likely to represent stable graph structure, so their reliability is set to $\omega_m = \mathrm{softmax}_m[-Z(\mathcal{S}_m)/\tau_s]$. Let ℓ_i^{resp} be the rank of the three-fold held-out masked joint-pseudolikelihood error. Its role is to refine a structural response, not to define a second global ranking. The binary certificate I_{resp} activates this refinement when the mean held-out error remains nondecreasing through every statistically resolved dyadic upper tail of r. With corr_R denoting rank correlation, the deployed chart is

$$\begin{aligned} A_i &= \sum_m \omega_m [\phi_{im}]_+ [\bar{\phi}_{im}]_+, \\ B_i &= \sum_m \omega_m \sqrt{Z(\phi_{im}\bar{\phi}_{im})^2 + Z((\phi_{im} - \bar{\phi}_{im})^2)^2}, \\ r_i &= R\Big(\tau_r \log\Big[e^{R(A_i)/\tau_r} + e^{R(B_i)/\tau_r}\Big]\Big), \\ \lambda_{\mathrm{resp}} &= I_{\mathrm{resp}}[\mathrm{corr}_R(r, \ell^{\mathrm{resp}})]_+, \\ h_i^{(1)} &= R(r_i + \lambda_{\mathrm{resp}} r_i (1 - r_i) \ell_i^{\mathrm{resp}}). \end{aligned} \tag{A.1}$$

Here A_i measures positive agreement between a node and its neighbourhood, whereas B_i measures both aligned magnitude and local contrast. The log-sum-exp forming r_i acts as a smooth "either mechanism" operator. The factor $r_i(1 - r_i)$ then concentrates the learned correction in the uncertain middle while preserving both ends of the structural ordering. Tail persistence determines whether that correction is informative; when it is absent, the chart naturally reduces to $h^{(1)} = r$.

Chart 2: canonical empirical measure. This chart provides a stable default measure when the attributes support several complementary summaries. Let $b_i^{(0)}$ be the all-resolution base score. The coordinate a_i combines attribute magnitude, neighbourhood residual, and rooted third-return energy; d_i measures departure along the graph's positive third-cumulant direction. If μ_0 is the mean coordinatewise central-atom mass, the component capitals are $C_0 = 1$, $C_a = \mu_0(1 - \mu_0)^{-1} \mathrm{odds}([\mathrm{corr}_R(b^{(0)}, a)]_+)$, and $C_d = (1 - \mu_0)\Lambda(b^{(0)}, d)$, where Λ is the least-favourable product of positive alignment odds over the resolved dyadic chains. These capitals favour an auxiliary coordinate only when it agrees with the base ordering over resolved portions of the graph. Let ℓ_i^{can} be the three-fold held-out denoising-energy rank over the four canonical operator spectra and let I_{can} be its tail-persistence certificate. The complete chart is

$$\begin{aligned} a_i &= R\left(\left\| \Big[Z(\|x_i\|_2),\ Z(\|x_i - \bar{x}_i\|_2),\ Z(t_i^{(3)}) \Big]_+ \right\|_2\right), \\ g_3 &= \frac{[N^{-1}\sum_i z_i^{\odot 3}]_+}{\|[N^{-1}\sum_i z_i^{\odot 3}]_+\|_1}, \qquad d_i = R(-z_i^\top g_3), \\ b_i &= R\left(\frac{C_0 b_i^{(0)} + C_a a_i + C_d d_i}{C_0 + C_a + C_d}\right), \\ x_i^{\times} &= R(\log E(b_i) + \log E(\ell_i^{\mathrm{can}})), \\ \lambda_{\mathrm{can}} &= I_{\mathrm{can}}[\mathrm{corr}_R(b, \ell^{\mathrm{can}})]_+, \\ h_i^{(2)} &= R\big((1 - \lambda_{\mathrm{can}}) b_i + \lambda_{\mathrm{can}} x_i^{\times}\big). \end{aligned} \tag{A.2}$$

Here $t_i^{(3)}$ is the rooted third-return field and $\bar{x}_i$ is the symmetric neighbour mean. The positive cone in a_i highlights nodes that are energetic in one or more canonical views, while d_i captures directional skew that a norm alone would miss. The e-value product in $x_i^{\times}$ rewards agreement in the upper tails of the canonical and denoising orderings rather than a large value in only one of them. The final mixture therefore reads as "canonical risk, sharpened by reproducible spectral evidence"; without such agreement it reduces to b.

Charts 3–6: explicit relational coordinates. The remaining charts are deliberately simpler relational coordinates. They become useful when the graph geometry makes a single structural contrast more informative than the composite charts above. Let z_i be coordinatewise standardised features, d_i^{out} the directed out-degree, d_i^{sym} the symmetric degree, and $\bar{z}_i^{-}$ the symmetric leave-self-out neighbour mean. Let S contain the lowest 20% of feature coordinates by observed edge quadratic variation, A_d be the degree-squared damped symmetric step, and $\widetilde{A}$ the symmetric mean step with self contribution. Exposure-normalised load is

$$h_i^{(3)} = Z(-d_i^{\mathrm{out}}) + Z(\|z_i\|_1). \tag{A.3}$$

It contrasts attribute mass with structural exposure: a feature-heavy node supported by few outgoing relations receives a stronger response than an equally large node embedded in many relations. Positive edge coherence on the graph-smooth coordinates is

$$h_i^{(4)} = \sum_{d \in S} [z_{id}]_+ [\bar{z}_{id}^{-}]_+. \tag{A.4}$$

Restricting the product to empirically smooth dimensions suppresses noisy attributes and retains coherent positive excursions shared with the local neighbourhood. Damped two-step diffusion alignment is

$$h_i^{(5)} = Z(d_i^{\mathrm{sym}}) - 1.5Z(\|x_i - A_d^2 x_i\|_1) + 3Z(\cos(x_i, A_d^2 x_i)). \tag{A.5}$$

This chart favours high-exposure nodes whose attributes remain aligned after two-step propagation; the residual and cosine terms distinguish genuine diffusion support from degree alone. The 3:1.5 weighting gives angular agreement twice the magnitude of the residual penalty, with degree acting as a lower-order exposure term. Finally, conditional-versus-two-hop scale contrast is

$$h_i^{(6)} = 2Z(\|z_i - \widetilde{A} z_i\|_2) - Z(\|x_i - \widetilde{A}^2 x_i\|_2). \tag{A.6}$$

The 2:1 contrast emphasises deviations visible against the immediate neighbourhood but attenuates deviations already explained at two hops. It is therefore a scale-comparison chart rather than another raw residual score. After construction, the common $E(\cdot)$ map makes all six upper tails comparable and Equation 5 combines them into the coarse risk measure. Conditional and Partial OT subsequently decide which nodes occupy positions in that measure.

A.1 Graph geometry seen by the weight controller

The controller needs only a coarse description of the regime in which a chart will operate. We use six global coordinates: sample size N, feature dimension D, mean stored degree $\bar{d}$, feature occupancy ρ, relative feature-norm dispersion ν, and the finite-sample resolution r_N. Together they distinguish small from asymptotic graphs, sparse from dense relations, narrow from wide attributes, and homogeneous from heterogeneous node magnitudes. After replacing nonfinite feature entries by zero, write the resulting matrix as $X^0 \in \mathbb{R}^{N \times D}$, let M be the number of stored directed edge entries, and set $u_i = \|x_i^0\|_2$. With $\varepsilon = 10^{-6}$, the descriptor is

$$\begin{aligned}
&N = \mathrm{rows}(X^0), && D = \mathrm{cols}(X^0), && \bar{d} = \frac{M}{\max(1, N)}, \\
&\rho = \frac{1}{ND} \sum_{i=1}^{N} \sum_{d=1}^{D} \mathbf{1}[x_{id}^0 \neq 0], && \bar{u} = \frac{1}{N} \sum_{i=1}^{N} u_i, && s_u = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (u_i - \bar{u})^2}, \\
&\nu = \frac{s_u}{\max(|\bar{u}|, \varepsilon)}, && r_N = \max(2, N)^{-1/2}, && g_G = (N, D, \bar{d}, \rho, \nu, r_N).
\end{aligned} \tag{A.7}$$

The population standard deviation in ν makes norm dispersion comparable across graph sizes, while $r_N = \max(2, N)^{-1/2}$ supplies the natural shrinking scale for finite-sample contrasts. The use of

stored degree keeps $\bar{d}$ consistent with the message-passing representation; a symmetrically stored edge contributes in both directions. Feature occupancy ρ complements D: two graphs may have equally wide attributes but very different amounts of observed information.

A.2 Complete graph-level activation controller

The activations translate graph geometry into a soft preference over the six charts. Most regimes are described by simple one- or two-coordinate contrasts. Chart 2 additionally covers a dense low-dimensional regime in which high structural exposure, heterogeneous feature norms, and high feature occupancy jointly make the canonical measure informative:

$$q_{\mathrm{dense}} = S_\tau[-\ell(D,16)]S_\tau[\ell(\bar{d},512)]S_\tau[\ell(\nu,1)] \times S_\tau\left[d\left(\rho, 1-\frac{1}{\max(2,D)}\right)\right]. \tag{A.8}$$

In chart order, the complete deterministic activations are

$$\alpha_1 = S_\tau[-\ell(N,512)], \tag{A.9}$$
$$\alpha_2 = 1 - \{1 - S_\tau[\ell(D,128)]\}(1-q_{\mathrm{dense}}), \tag{A.10}$$
$$\alpha_3 = S_\tau[\ell(\bar{d},512)]S_\tau[\ell(\nu,1)], \tag{A.11}$$
$$\alpha_4 = S_\tau[-\ell(\bar{d},4)]S_\tau[-\ell(\nu,r_N)]S_\tau[d(\rho,0.5)], \tag{A.12}$$
$$\alpha_5 = S_\tau[-\ell(D,16)]S_\tau[\ell(\bar{d},16)], \tag{A.13}$$
$$\alpha_6 = 1. \tag{A.14}$$

The equations have a direct regime interpretation. The finite-sample chart α_1 is favoured when N is below its reference scale. The canonical chart α_2 responds either to wide attributes or to the dense low-dimensional regime q_{dense}. The load chart α_3 targets graphs with both high relational exposure and heterogeneous feature norms. The coherence chart α_4 is most relevant for sparse, low-dispersion graphs with sufficiently occupied attributes, where a smooth local coordinate is less likely to be washed out by hubs. The diffusion chart α_5 covers dense graphs with narrow attributes. Finally, $\alpha_6 = 1$ closes the stick and assigns any remaining mass to the scale-contrast chart.

The reference values serve as interpretable regime landmarks: $N = 512$ marks the finite-sample transition; $D = 16$ and $D = 128$ distinguish narrow and wide attributes; $\bar{d} = 4, 16, 512$ separate sparse, moderately connected, and high-exposure graphs; and $\nu = r_N$ versus $\nu = 1$ distinguishes sampling-scale variation from order-one norm heterogeneity. The occupancy contrast at $\rho = 0.5$ asks whether at least half of the feature matrix is active, while the stricter $1 - 1/D$ threshold in q_{dense} identifies the nearly fully occupied low-dimensional case. These are soft landmarks rather than hard graph categories.

Under Equation 5, α_j is best read as a conditional gate rather than a standalone weight: w_j receives the fraction α_j of the mass left by earlier charts. The ordering therefore moves from specialised, sharply identified regimes to a broadly applicable fallback. Soft steps make this transition continuous near each reference scale, avoiding abrupt changes in the prior for nearly identical graphs.

A.3 Admissible evidence-cost coordinates

For $\mathbf{e} = (c, f)^\top \in [0,1]^2$, the OT solver uses the following low-complexity scalar coordinates:

$$\kappa_{\mathrm{lin}}(\mathbf{e}) = \lambda_c c + \lambda_f f, \qquad \kappa_{\min}(\mathbf{e}) = \min(1-c, f), \tag{A.15}$$
$$\kappa_{\mathrm{opp}}(\mathbf{e}) = (1-c)f, \qquad \kappa_{\mathrm{joint}}(\mathbf{e}) = cf, \tag{A.16}$$
$$\kappa_{\max}(\mathbf{e}) = \max(c, f), \qquad \kappa_{\mathrm{rad}}(\mathbf{e}) = (c-\tfrac{1}{2})^2 + (f-\tfrac{1}{2})^2. \tag{A.17}$$

These coordinates encode different relations between the same two evidence views. The linear form selects a direction in evidence-rank space; $\min(1-c, f)$ and $(1-c)f$ express increasingly strict agreement between context compatibility and feature coherence; cf captures their cross-view interaction; $\max(c, f)$ lets the dominant coordinate determine position; and κ_{rad} measures directional departure from the joint centre. This small family changes the geometry used for matching, not the evidence available to the matcher. It is expressive enough to represent dominance, agreement, interaction, and radial departure while remaining interpretable.

A.4 DETAILED DATASET STATISTICS

Table 7 highlights the breadth of the complete evaluation suite. The 34 graphs range from 124 to 5.78 million nodes, from 1 to 10,000 feature dimensions, and from 0.04% to 63.21% anomaly prevalence, spanning markedly different graph structures and anomaly constructions. Such breadth makes it difficult for any fixed structural scale, feature scale, or predefined anomaly type to characterize every graph. Consequently, if the same risk-reallocation mechanism remains stable across graphs with such pronounced differences, this provides stronger evidence that it captures a principle of risk organization shared across graphs rather than a pattern specific to a single graph family.

Dataset	Nodes	Directed edges	Features	Labeled	Anomalies	Rate (%)
Main (10): primary accuracy and external-baseline comparison						
Books	1,418	3,695	21	1,418	28	1.97
Weibo	8,405	407,963	400	8,405	347	4.13
Reddit	10,984	168,016	64	10,984	366	3.33
Disney	124	335	28	124	6	4.84
Enron	13,533	176,987	18	13,533	5	0.04
Minesweeper	10,000	39,402	7	10,000	2,000	20.00
Tolokers	11,758	519,000	10	11,758	2,566	21.82
Questions	48,921	153,540	301	48,921	1,460	2.98
Amazon	11,944	8,796,784	25	11,944	821	6.87
Yelp	45,954	7,693,958	32	45,954	6,677	14.53
Additional (10): breadth and robustness beyond Main-10						
ACM	16,484	164,350	8,337	16,484	597	3.62
BlogCatalog	5,196	345,566	8,189	5,196	300	5.77
BITalpha	3,219	24,530	128	3,219	300	9.32
Wiki	9,227	18,257	64	9,227	217	2.35
Cora-SAMCL	2,708	11,606	1,433	2,708	150	5.54
CiteSeer-SAMCL	3,327	10,275	3,703	3,327	150	4.51
PubMed-SAMCL	19,717	92,849	500	19,717	600	3.04
Cora-Synthetic	2,708	11,096	1,433	2,708	135	4.99
CiteSeer-Synthetic	3,327	9,824	3,703	3,327	166	4.99
PubMed-Synthetic	19,717	93,058	500	19,717	986	5.00
Large (3): scalability and bounded-memory evaluation						
Elliptic	203,769	234,355	165	46,564	4,545	9.76
T-Finance	39,357	42,484,443	10	39,357	1,803	4.58
T-Social	5,781,065	146,211,016	10	5,781,065	174,280	3.01
Mechanism check (11): strict-unseen transfer and recovery diagnosis						
MGTAB	10,199	1,700,108	788	10,199	2,748	26.94
YelpChi	45,954	7,693,958	32	45,954	6,677	14.53
YelpNYC	21,040	3,316,274	10,000	21,040	1,000	4.75
FairGAD-Reddit	9,892	1,211,748	385	9,892	1,353	13.68
EX-Graph-WT	452,930	2,333,911	8	452,930	939	0.21
Cresci-15	5,301	14,220	1,542	5,301	3,351	63.21
Bitcoin OTC	5,881	35,592	1	316	180	56.96
REV2 Epinions	253,071	27,336,208	1	9,427	668	7.09
DIAM Ethereum-S	1,329,729	6,794,521	48	3,360	1,660	49.40
Elliptic++	822,942	2,868,964	55	265,354	14,266	5.38
DIAM Bitcoin-M	2,527,649	19,187,400	56	259,956	46,930	18.05

Table 7: Dataset groups, graph statistics, and evaluation roles.

B ADDITIONAL FEW-SHOT COMPARISONS

Table 8 extends the Main-10 few-shot support comparison to Additional-10 and further examines cross-graph coverage across different graph types. Although several comparison methods can use a small number of labeled nodes as support, LORE uses no additional labeled support nodes and still achieves the highest AP on all 10 graphs, with pronounced advantages on BITalpha, the SAMCL family, and the synthetic variants. Under the current evaluation protocol, these results indicate that LORE's gains do not depend on adaptation to the provided few-shot labels, but arise primarily from modeling and reallocating graph-scale risk structure through node-level evidence correspondence.

Method	ACM	Blog	BITalpha	Wiki	Cora-S	Cite-S	Pub-S	Cora-Syn	Cite-Syn	Pub-Syn
SpaceGNN (5N+5A)	0.0500	0.0758	0.1059	0.0268	0.1241	0.1452	0.0390	0.0733	0.0946	0.0555
ARC (5N)	0.4069	0.3736	0.1966	0.0183	0.4841	0.4078	0.3088	0.4398	0.4018	0.3710
AnomalyGFM-N (5N)	0.0389	0.0617	0.0894	0.0251	0.0593	0.0546	0.0368	0.0703	0.0510	0.0555
AnomalyGFM-A (5A)	0.0400	0.0608	0.0882	0.0245	0.0588	0.0575	0.0369	0.0677	0.0515	0.0561
APF (5N+5A)	0.0389	0.0746	0.1132	0.0257	0.0622	0.0564	0.0624	0.0565	0.0652	0.0861
LORE (0)	**0.8560**	**0.3932**	**0.7481**	**0.0312**	**0.8108**	**0.8949**	**0.7830**	**0.6742**	**0.8141**	**0.7956**

Table 8: Few-shot AP on Additional-10 under the Table 3 support protocol.

C ADDITIONAL MECHANISM DIAGNOSTICS

C.1 ADDITIONAL-10 RESULTS

Method	ACM	Blog	BITalpha	Wiki	Cora-S	Cite-S	Pub-S	Cora-Syn	Cite-Syn	Pub-Syn
HUGE	OOM	OOM	0.1243	0.0194	0.1407	0.0946	0.2038	0.1251	0.0811	0.2350
SCAN	0.0246	0.0412	0.0594	0.0310	0.0369	0.0297	0.0196	0.0352	0.0344	0.0324
OddBall	0.4005	0.0559	0.5419	0.0209	0.5155	0.4914	0.5288	0.4479	0.4376	0.5209
Radar	0.2287	0.3057	0.1407	0.0197	0.1375	0.1909	0.2251	0.1127	0.2347	0.2733
ANEMONE	0.1723	0.0718	0.2385	0.0256	0.1378	0.0759	0.0947	0.0987	0.1376	0.1242
AD-GCL	0.0439	0.0583	0.0888	0.0225	0.0813	0.0854	0.0379	0.0621	0.0737	0.0581
DiffGAD	OOM	OOM	0.1133	0.0207	0.3802	0.2596	0.1269	0.2830	0.2762	0.1535
SmoothGNN	0.0754	0.2874	0.0932	0.0221	0.0869	0.0528	0.0320	0.0787	0.0582	0.0522
GAD-NR	0.1694	OOM	0.1063	0.0194	0.0707	0.0595	0.1320	0.0669	0.0795	0.1648
MI-GAD	0.2496	0.0356	0.1775	0.0162	0.3632	0.4910	0.5611	0.3393	0.5165	0.5966
GEL	0.1598	0.0393	0.2676	0.0185	0.1471	0.1525	0.2958	0.1023	0.1244	0.3503
GADAM	0.4119	0.3075	0.2031	0.0210	0.6713	0.4039	0.6166	0.5451	0.3928	0.6315
ADA-GAD	OOM	OOM	0.0867	0.0192	0.3094	0.3369	0.1012	0.1688	0.2273	0.1194
GCTAM	OOM	OOM	0.0879	0.0198	0.0358	0.0314	0.0232	0.0305	0.0344	0.0397
LORE (ours)	**0.8560**	**0.3932**	**0.7481**	**0.0312**	**0.8108**	**0.8949**	**0.7830**	**0.6742**	**0.8141**	**0.7956**

Table 9: Average precision on Additional-10 graphs.

Table 9 further shows that the improvement is not confined to a single graph family, but appears simultaneously on dense social graphs, sparse citation graphs, high-dimensional attributed graphs, and graphs with different anomaly-generation mechanisms. In particular, the consistent advantage on the SAMCL and synthetic variants indicates that the method does not depend on a fixed feature dimension or anomaly pattern. This consistency across graph families and anomaly constructions further supports our mechanistic interpretation: the gain primarily arises from resolving the coupling between graph-level risk structure and node evidence. Specifically, the Prior provides the graph-scale risk structure, while evidence determines which nodes should occupy the corresponding risk positions.

C.2 MULTISCALE MECHANISM FINGERPRINTS

Figure 6 presents the multiscale OT mechanism fingerprints across the Main-10 graphs. The left panel exactly decomposes the AP gain of the full model over the Prior into the Conditional OT main effect, the Partial OT main effect, and their interaction:

$$\begin{aligned}\Delta_{\text{full}} = &\underbrace{(\text{AP}_{\text{cond}} - \text{AP}_{\text{prior}})}_{\Delta_{\text{cond}}} + \underbrace{(\text{AP}_{\text{partial}} - \text{AP}_{\text{prior}})}_{\Delta_{\text{partial}}} \\ &+ \underbrace{(\text{AP}_{\text{full}} - \text{AP}_{\text{cond}} - \text{AP}_{\text{partial}} + \text{AP}_{\text{prior}})}_{\Delta_{\text{int}}}.\end{aligned} \tag{C.1}$$

A positive interaction indicates superadditive cooperation between the two stages, whereas a negative interaction indicates subadditive cancellation because the stages correct partly overlapping nodes; it does not necessarily imply that either stage is harmful. The right panel divides each component by the corresponding graph's total AP gain, so the three normalized contributions sum to one up to display rounding.

The graphs exhibit clearly different mechanism compositions. The gain on Books is driven primarily by Conditional OT, whereas Disney and Enron depend more strongly on Partial OT. Questions, Reddit, and Yelp show relatively strong positive interactions, indicating that much of their final improvement comes from joint correction at the two scales. Amazon and Minesweeper benefit from both Conditional and Partial main effects, while their negative interactions show that the stages capture partly overlapping correction signals. Overall, no single OT scale dominates uniformly across graphs. This graph-dependent composition supports retaining both Conditional and Partial transport so that the model can express different risk-reallocation structures.

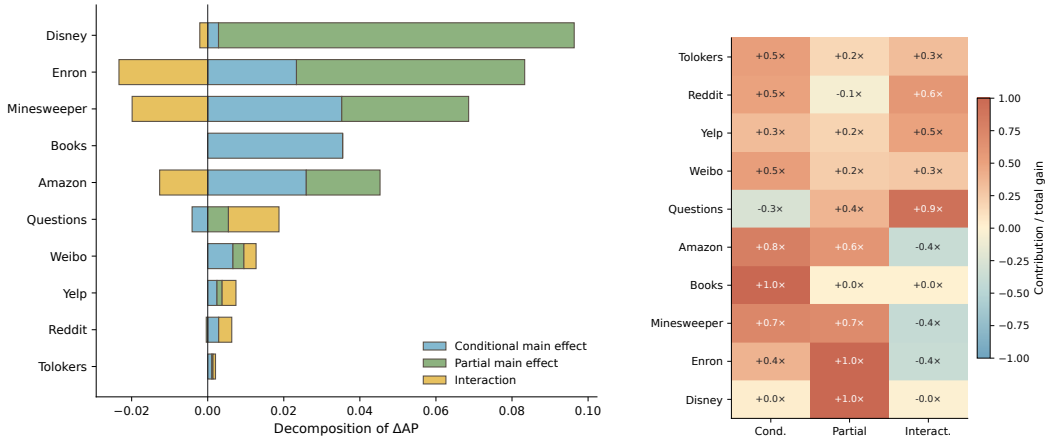


Figure 6: Multiscale OT mechanism fingerprints on Main-10. Left: decomposition of AP gain into Conditional, Partial, and interaction effects. Right: normalized contribution of each effect to the total gain.

C.3 Cross-Graph Proxy Reliability

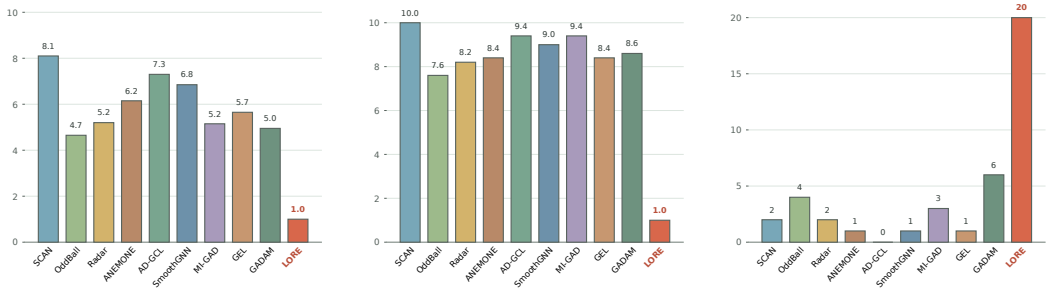


Figure 7: Cross-graph ranking robustness. From left to right: mean rank, worst-five rank, and Top-2 count; lower ranks and higher Top-2 counts are better.

Figure 7 evaluates the stability of different detectors from the perspective of cross-graph ranking. Mean rank summarizes overall competitiveness across the full suite, worst-five rank characterizes the performance floor on unfavorable graphs, and Top-2 count measures how often a method remains among the leading detectors. Although existing methods can perform well on particular graphs, their advantages are fragmented across the suite: the complete baselines have mean ranks between 4.65 and 8.10 and worst-five ranks between 7.60 and 10.00, and none enters the top two on more than six of the 20 graphs. The reliability of a single proxy is therefore strongly graph-dependent; different proxies fit different graph structures and anomaly patterns, making it difficult for any one of them to provide a uniformly reliable ranking. In contrast, under the recorded graph-wise configuration protocol, LORE attains a mean rank and worst-five rank of 1.00 and enters the top two on all 20 graphs. Its advantage is thus not driven by a small subset of particularly favorable datasets, but

reflects broader and more stable cross-graph coverage. This result supports treating multiple proxies as graph-dependent candidate risk coordinates and node evidence, and then constructing the final ranking through risk reallocation, rather than relying on any fixed proxy to impose a uniform ordering across all graphs.

C.4 Support-Dependent Evidence Reliability

Figure 8 examines whether the reliability of tail evidence depends on its scope of application. The gold bars evaluate the configured evidence over the full graph, whereas the green bars keep the evidence unchanged but evaluate it only within the fixed high-risk support identified by LORE. The coral bars preserve both this support and the multiset of evidence values, while randomly permuting the correspondence between evidence and nodes. Across these representative datasets, the same evidence exhibits substantially varying graph-wide discriminative ability, whereas its pairwise anomaly–normal ordering becomes markedly stronger within the high-risk support. Once the node–evidence correspondence inside that support is destroyed, AUC returns approximately to the random baseline. The reliability of evidence is therefore not an intrinsic global property; it depends jointly on the decision region in which the evidence is applied and on its node-level correspondence. This result also explains why Partial OT uses tail evidence only within the high-risk tail instead of allowing it to impose an unsupported uniform ordering over the entire graph.

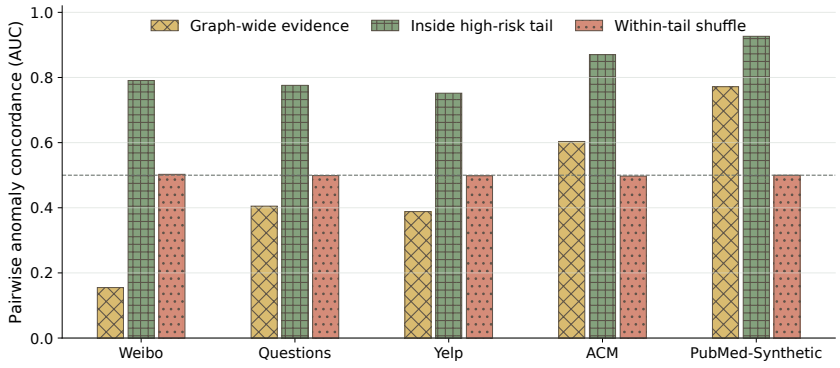


Figure 8: Support-dependent evidence reliability.

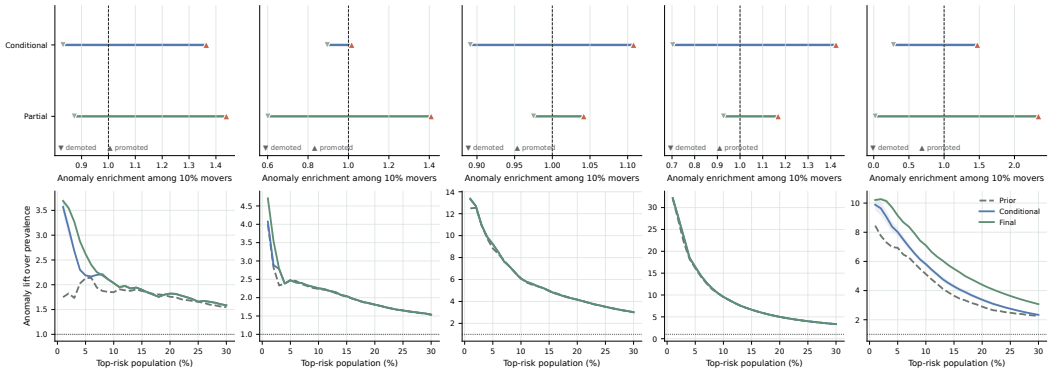
C.5 Anomaly-Aligned Risk Migration

Figure 9 examines at the node level whether LORE's risk migration aligns with ground-truth anomaly status. The upper row reports anomaly enrichment among nodes promoted and demoted by Conditional OT and Partial OT. Overall, promoted nodes tend to exhibit greater anomaly enrichment, whereas demoted nodes are closer to or below the overall anomaly prevalence. Thus, on these representative graphs, LORE does not reassign risk arbitrarily; it tends to allocate higher risk to nodes that are more likely to be anomalous while relatively reducing the risk assigned to normal nodes. The varying migration magnitudes across graphs and transport stages further indicate that effective risk adjustment is graph-dependent.

The lower row further examines whether this anomaly alignment is confined to a small number of the most extreme movers. As the evaluation range expands from the highest-risk nodes to a larger top-risk population, anomaly lift generally remains above the prevalence baseline. Anomaly concentration is therefore not restricted to a few extremely displaced nodes, but extends over a broader high-risk region. On graphs where the Prior, Conditional, and Final curves are close, LORE mainly refines the ordering within the high-risk region while largely preserving its aggregate anomaly concentration. Where the curves separate visibly, risk reallocation additionally improves anomaly concentration across that region.

Taken together, Conditional OT and Partial OT exhibit different mover profiles, showing that they do not simply repeat the same correction. The former primarily resolves node-ordering ambiguity within local risk cells, whereas the latter further adjusts node ownership in the high-risk tail. Fig-

ure 9 therefore supports LORE's multi-scale risk-reallocation mechanism from the complementary perspectives of migration direction and effective range.



(a) Minesweeper (b) Yelp (c) Amazon (d) PubMed-SAMCL (e) BITalpha

Figure 9: Anomaly-Aligned Risk Migration.

C.6 Marginal risk versus node-level assignment

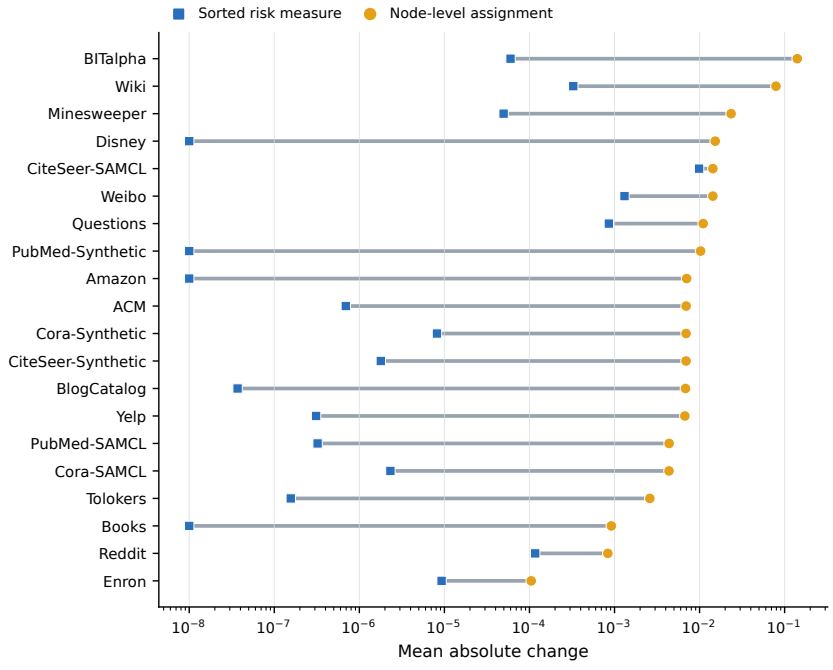


Figure 10: Risk-measure preservation versus node-level reassignment.

Figure 10 examines what LORE changes during risk reallocation. We compare the risk scores before and after OT from two complementary perspectives. The sorted risk measure first orders the scores and therefore ignores specific node identities, measuring whether the overall distribution of risk values changes. Node-level assignment instead aligns scores by the same node identity before and after transport, measuring the extent to which risk is reassigned across nodes. Across most datasets, the sorted risk measure changes only minimally, whereas node-level assignment changes substantially more. LORE therefore generally does not reshape the overall risk distribution; rather, it changes which nodes occupy the existing risk positions. In other words, OT markedly changes the risk received by individual nodes while leaving the global collection of risk values nearly stable. This

shows that LORE's core role is not to generate a new risk structure, but to redistribute node ownership within the existing one, separating "risk structure" from "which nodes occupy its positions."

C.7 View-Specific Evidence Correspondence

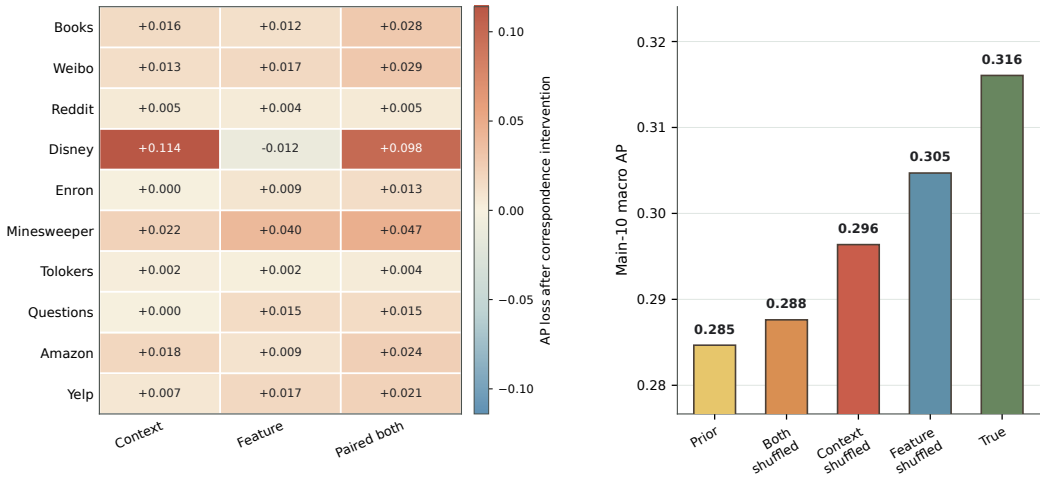
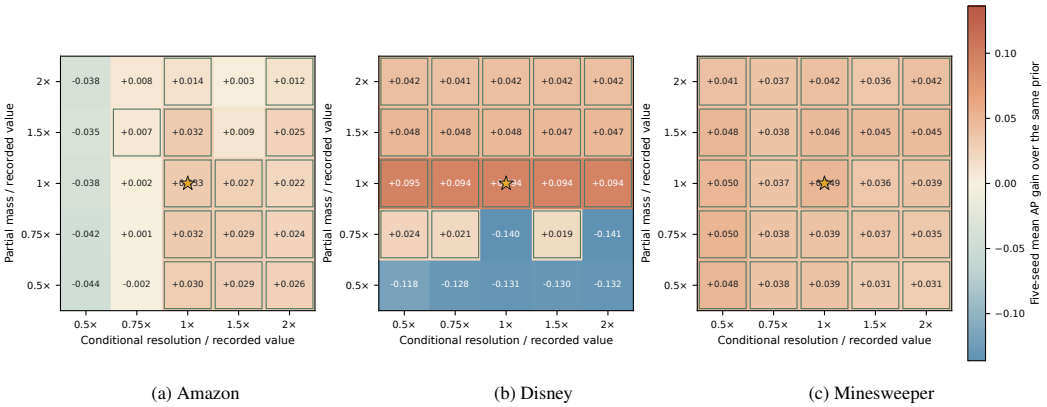


Figure 11: View-specific correspondence interventions on Main-10. (a) Correspondence loss across datasets. (b) Aggregate intervention.

Figure 11 examines whether the two evidence views contribute through stable node-level correspondence rather than through their marginal distributions alone. The intervention results reveal a clear graph-dependent pattern: context and feature evidence are complementary, but their relative importance varies across graphs. This indicates that no single view is uniformly reliable, and that useful evidence must be interpreted with respect to the current graph structure. More importantly, jointly disrupting the node correspondence of both views removes most of the gain over the prior, showing that the benefit of the learned evidence lies primarily in identifying which nodes should receive the available risk, rather than merely supplying favorable global score distributions. These results further support LORE's design choice of treating evidence as a graph-dependent assignment signal instead of relying on a fixed fusion rule.

C.8 Resolution and Transported-Mass Sensitivity

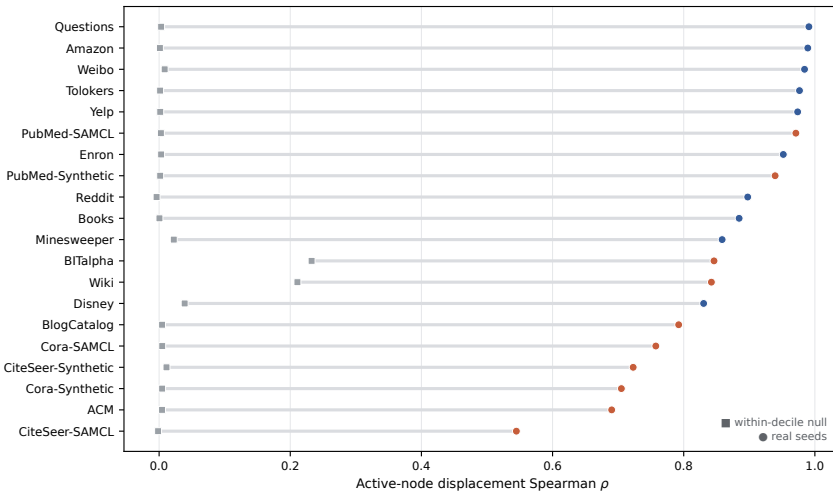


(a) Amazon (b) Disney (c) Minesweeper

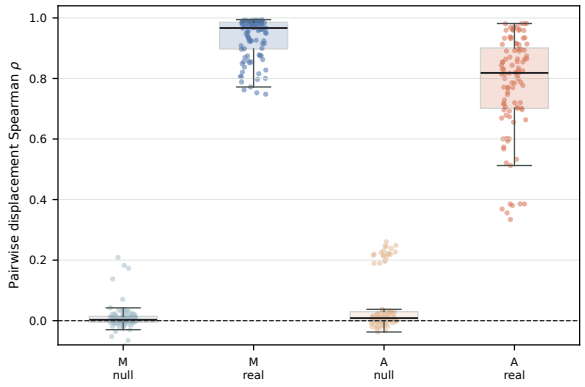
Figure 12: Sensitivity of LORE to Conditional resolution and Partial transported mass.

To examine LORE's sensitivity to Conditional resolution and Partial transported mass, we scale both parameters relative to their default settings by $\{0.5, 0.75, 1, 1.5, 2\}$ on Amazon, Disney, and Minesweeper, while keeping the evidence coordinates and cost functions fixed. As shown in Figure 12, 14 of 25 parameter combinations on Amazon, 18 of 25 on Disney, and all 25 on Minesweeper improve over the corresponding same-run prior for every random seed. These connected regions of positive gain show that LORE is not tied to a single hyperparameter setting but remains stable across a broad portion of the tested grid. The dataset-specific boundaries further reveal the distinct roles of the two OT scales. Amazon's gain declines when Conditional cells become too broad, indicating greater reliance on Conditional OT; Disney deteriorates when too little tail mass is transported, indicating greater reliance on Partial OT; and Minesweeper retains positive gains under every parameter combination, demonstrating robustness to both scales.

C.9 Node-Level Migration Stability



(a) Real–null agreement



(b) Pairwise agreement

Figure 13: Node-level migration stability across five evidence seeds.

Similar final AP across evidence seeds does not guarantee stable node-level risk allocation, because evidence learned from different random initializations may act on different nodes yet yield similar final rankings. We therefore further test whether LORE's risk redistribution is consistent at the node level across seeds. Specifically, we compare all ten pairwise combinations from five fixed evidence-seed initializations, characterize risk migration stability at three levels—node displacement magnitude, the identity of decision-relevant nodes, and migration direction—and use matched permutation controls

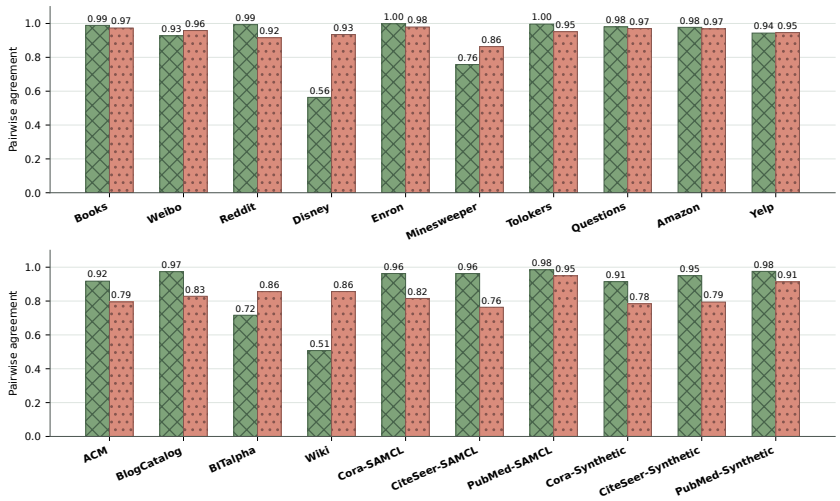


Figure 14: Membership and direction agreement. Cross-hatched green bars and dotted coral bars denote Top-5% membership Jaccard and migration-sign agreement, respectively.

to exclude apparent agreement arising only from the prior-risk structure. Figure 13(a) first examines whether node displacement magnitude remains consistent across random initializations. For any two runs a and b, we compute the risk displacement of node i as $\Delta_i^{(r)} = s_i^{(r)} - p_i$ and use Spearman correlation to compare the relative ordering of node displacements across the two runs. A large correlation means that the runs agree on which nodes move more and which move less. The circles (real seeds) show correlations obtained while retaining the true node correspondence, whereas the squares (within-decile null) show the control obtained after randomly permuting node identities within the same prior-risk decile. This control preserves the coarse prior-risk region of each node but breaks the correspondence between specific nodes, allowing us to distinguish stability arising from genuine node-level decisions from superficial agreement induced by similar prior-risk regions. Across almost all datasets, the real correlations are high whereas the permutation controls concentrate near zero. Once the true node correspondence is broken, the agreement largely disappears. Figure 13(a) therefore shows that LORE's cross-seed stability has a clear node-level basis rather than reflecting only superficial agreement at the level of prior-risk regions. Figure 13(b) summarizes the correlation distributions over all seed pairs. The real correlations remain consistently well above the permutation nulls; Main-10 is more concentrated, whereas Additional-10 exhibits greater variation. This indicates that node migration is stable overall but is more sensitive to random initialization on the more challenging datasets. However, high displacement correlation shows only that different initializations produce similar relative orderings of node displacement magnitudes; it does not guarantee that the same nodes ultimately determine the decision, or that those nodes consistently move in the same direction. Figure 14 therefore further evaluates node-level stability from the complementary perspectives of decision-relevant membership and migration direction. Top-5% membership Jaccard measures whether different initializations select the same high-risk nodes, while migration-sign agreement measures whether the same node is consistently promoted or demoted across runs. Both dataset groups maintain high Top-5% membership Jaccard, and migration-sign agreement is also high overall. Thus, although random initialization introduces some variation in the exact displacement magnitudes, LORE remains consistent in which nodes enter the decision-critical high-risk set and whether those nodes are promoted or demoted.

D AUROC Comparisons

Tables 10–14 report the AUROC counterparts of the main-text AP results.

These results examine whether LORE's improvements are confined to the leading high-risk nodes or extend to the ranking more broadly. AP emphasizes the identification quality of anomalous nodes near the top of the ranking, whereas AUROC reflects positive–negative pairwise ordering over the entire

Method	Books	Weibo	Reddit	Disney	Enron	Mine.	Tol.	Quest.	Amazon	Yelp
HUGE	0.5196	0.5074	**0.5930**	0.3980	0.3193	**0.6473**	0.3460	0.4980	OOM	OOM
SCAN	0.5126	0.2260	0.4439	0.4718	0.4206	0.5016	0.4730	0.3719	0.6570	0.4851
OddBall	0.5691	0.4476	0.5562	0.3821	0.5471	0.4987	0.5909	0.6142	0.6243	0.5245
Radar	0.5042	0.6757	0.5919	0.5113	0.3161	0.5425	0.4036	0.6103	0.8768	0.6043
ANEMONE	0.5648	0.3055	0.4990	0.3791	0.6612	0.5394	0.5662	0.5506	0.4319	0.4810
AD-GCL	0.3364	0.3831	0.5590	0.3774	0.3155	0.4958	0.5136	0.5003	0.3877	0.5938
DiffGAD	0.5259	0.9065	0.5407	0.4932	0.5786	0.5022	0.5243	OOM	0.4845	OOM
SmoothGNN	0.4944	0.7595	0.4530	0.4997	0.4846	0.5041	0.5290	0.5231	0.4488	0.4700
GAD-NR	0.4993	0.7928	0.5588	0.4718	0.4645	0.4590	0.5342	**0.6428**	0.7201	0.5876
MI-GAD	0.5000	0.5908	0.5387	0.4165	0.3370	0.5269	0.4081	0.4525	0.8311	0.5903
GEL	0.5748	0.5556	0.5720	0.4209	0.4727	0.5367	0.4389	0.6243	0.8705	0.5729
GADAM	0.5608	0.4274	0.5781	0.3220	0.2600	0.5063	0.4259	0.6079	0.4173	0.4762
ADA-GAD	0.3925	0.9340	0.5627	0.3805	0.5617	0.5016	0.5743	OOM	0.2084	OOM
GCTAM	0.5143	0.5728	0.5104	**0.6297**	0.4818	0.5205	0.4958	OOM	0.5752	OOM
LORE	**0.6993**	**0.9710**	0.5628	0.6186	**0.7656**	0.6365	**0.6804**	0.5828	**0.9078**	**0.6230**

Table 10: AUROC on Main-10.

Method	ACM	Blog	BITalpha	Wiki	Cora-S	Cite-S	Pub-S	Cora-Syn	Cite-Syn	Pub-Syn
HUGE	OOM	OOM	0.5738	0.4368	0.8162	0.7593	0.8860	0.8197	0.7075	0.8618
SCAN	0.2993	0.3350	0.2476	**0.6195**	0.2790	0.2685	0.2597	0.3203	0.2978	0.2614
OddBall	0.7498	0.4888	0.7450	0.3806	0.7728	0.7176	0.7378	0.7144	0.7150	0.7376
Radar	0.7758	0.7427	0.4808	0.4470	0.7188	0.8299	0.8437	0.7504	0.8067	0.8356
ANEMONE	0.6547	0.4375	0.5700	0.5069	0.6444	0.5715	0.5867	0.6414	0.6221	0.5726
AD-GCL	0.5821	0.5011	0.4972	0.4896	0.6494	0.6974	0.5800	0.5955	0.6449	0.5639
DiffGAD	OOM	OOM	0.4546	0.4634	0.8838	0.8625	0.8797	0.8549	0.8454	0.8549
SmoothGNN	0.4517	0.6548	0.4367	0.4741	0.6283	0.4831	0.4581	0.6198	0.4846	0.4560
GAD-NR	0.7427	OOM	0.4491	0.4432	0.6004	0.5902	0.7167	0.6396	0.6262	0.7087
MI-GAD	0.7172	0.2595	0.6946	0.3337	0.8591	0.9041	0.9124	0.8587	0.8879	0.9020
GEL	0.7360	0.3088	0.7792	0.4171	0.6880	0.6824	0.8363	0.6637	0.6586	0.8208
GADAM	0.8384	**0.7900**	0.7121	0.4775	0.9360	0.8939	0.9511	0.9387	0.8763	0.9302
ADA-GAD	OOM	OOM	0.4764	0.4377	0.8434	0.8747	0.8101	0.8106	0.8495	0.7860
GCTAM	OOM	OOM	0.4905	0.4445	0.2721	0.3086	0.3755	0.2450	0.3110	0.3876
LORE	**0.9904**	0.7686	**0.9333**	0.5303	**0.9771**	**0.9911**	**0.9867**	**0.9502**	**0.9833**	**0.9830**

Table 11: AUROC on Additional-10.

Method	Elliptic	T-Finance	T-Social
FreeGAD	0.4870	0.9296	0.6791
HUGE	0.3715	0.8782	0.5836
GADAM	0.4721	0.6977	0.6325
LORE (current)	**0.8885**	**0.9454**	**0.8581**

Table 12: AUROC on the large graphs in Table 2.

Method	MGTAB	YelpChi	YelpNYC	FairGAD-R	EX-WT	Macro
PCA	0.5183	0.5896	**0.6897**	0.5755	0.7337	0.6214
Isolation Forest	0.4887	**0.6399**	0.6672	0.5853	**0.7683**	0.6299
HBOS	0.4847	0.5766	0.6835	0.5706	0.5805	0.5792
Frozen prior	**0.7727**	0.6224	0.5939	**0.7926**	0.7227	**0.7009**

Table 13: AUROC on the additional unseen graphs in Table 5.

Dataset	Random	PCA	IF	HBOS	Prior	LORE
Cresci-15	**0.5000**	0.4033	0.4351	0.3896	0.3606	0.4880
Bitcoin OTC	0.5000	0.5000	0.5000	0.5000	0.4101	**0.7691**
REV2 Epinions	0.5000	0.5000	0.5000	0.5000	0.3975	**0.6236**
DIAM Ethereum-S	0.5000	0.1199	0.1682	0.3059	0.5065	**0.7419**
Elliptic++	0.5000	**0.7265**	0.5063	0.6728	0.4452	0.6454
DIAM Bitcoin-M	0.5000	0.3603	0.3246	0.3716	0.3571	**0.5505**
Macro	0.5000	0.4350	0.4057	0.4567	0.4128	**0.6364**

Table 14: AUROC counterparts of the prior-failure recovery results in Table 6.

evaluation set. Overall, the two metrics exhibit broadly consistent trends: LORE remains strongly competitive on Main-10, Additional-10, and the three large graphs, indicating that its gains are generally not restricted to the high-risk tail but extend to a wider portion of the ranking. At the same time, the remaining gaps on a small number of datasets, including Reddit and Questions, show that improved identification of high-risk nodes does not necessarily translate into an equal improvement in ordering across the full distribution. Viewed globally, the AUROC results complement the main-text AP analysis and further show that LORE typically improves both the leading ranks and the broader risk-ordering structure.

E Computational Complexity

The apparent transport problem is inexpensive because both source evidence and target risk positions are one-dimensional with uniform mass. Their exact quadratic OT solution is monotone matching, so a dense coupling matrix is unnecessary. Conditional OT sorts scalar evidence within disjoint prior cells, and Partial OT sorts only the selected tail. The resulting worst-case cost is $O(N \log N)$ time and $O(N)$ memory after fitting the two self-supervised views. This reduction from pairwise transport to ordered assignment is what makes the same allocation principle practical on graphs with millions of nodes.

AI use statement

Generative AI tools were used to assist with code refinement, writing improvement, and language polishing during the research process. The authors retained full responsibility for all research decisions, analyses, interpretations, and the final content of the manuscript.