DDGAD: DISAGREEMENT-DRIVEN GRAPH ANOMALY DETECTION VIA ADAPT-THEN-COMBINE

Anonymous authors
Paper under double-blind review

ABSTRACT

Graph anomaly detection (GAD) commonly relies on message passing to jointly encode node attributes and neighborhood context. This joint encoding entangles node-specific evidence with contextual influence in a single graph-conditioned representation, making node–context compatibility difficult to observe explicitly. Consequently, the resulting representation alone cannot reveal whether a node agrees with or is contradicted by the local consensus implied by its neighbors. We propose Disagreement-Driven Graph Anomaly Detection (DDGAD), which constructs distinct node-wise and contextual estimates and treats their persistent disagreement as anomaly evidence. Inspired by Adapt-Then-Combine (ATC), DDGAD reverses the conventional role of disagreement in consensus formation. Adapt produces a node-wise estimate without new same-step neighbor aggregation, while Combine forms a neighborhood-dependent contextual estimate from neighboring adapted states. Their pre-consensus disagreement is accumulated across ATC steps for detection. We characterize this signal from graph-spectral and source-response perspectives and derive sufficient conditions for separating anomalous and normal nodes under contextual mixing. Experiments on six benchmarks show the highest average AUROC among the evaluated methods. Controlled context-source interventions and Adapt-operator controls further support persistent node–context disagreement across different Adapt realizations.

Keywords: Graph Anomaly Detection; Adapt-Then-Combine; Node-Context Disagreement; Diffusion Models

1 INTRODUCTION

Many GNN-based graph anomaly detection (GAD) methods use message passing to jointly encode node attributes and neighborhood context (Ding et al., 2019; Fan et al., 2020; Liu et al., 2022). This joint encoding makes anomaly detection inherently context-dependent: the representation of a node reflects not only its own attributes but also evidence aggregated from its neighborhood. When the two are incompatible, however, the resulting representation alone does not reveal whether the node state is supported or contradicted by its local context. This creates a *node–context ambiguity* in anomaly scoring.

Existing GAD methods exploit reconstruction discrepancies (Ding et al., 2019; Fan et al., 2020; Li et al., 2025a) or neighborhood-consistency signals (Liu et al., 2022; Hou et al., 2022; Xu et al., 2022). However, these signals are typically evaluated on graph-conditioned representations in which node and neighborhood information have already interacted. As a result, node–context compatibility remains implicit in the final detection score. A more direct approach is to construct two distinct post-encoding estimates from the same graph-conditioned state, with one updated node-wise and the other formed from neighborhood consensus. Their compatibility can then be measured explicitly.

Our key observation is that compatibility between these two estimates is itself informative. If a node-wise estimate remains compatible with the consensus implied by its neighbors, the two estimates should remain close; persistent disagreement instead indicates a local inconsistency that is difficult to reconcile. We therefore introduce *Disagreement-Driven Graph Anomaly Detection* (DDGAD), which preserves and accumulates node–context disagreement as anomaly evidence.

Inspired by the Adapt-Then-Combine (ATC) estimation paradigm (Sayed, 2014), DDGAD repurposes consensus formation for anomaly detection. To make node–context ambiguity explicit,

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.

DDGAD starts from a shared graph-conditioned state and constructs two post-encoding estimates under different update rules. A diffusion-based Adapt operator (Karras et al., 2022) forms a density-oriented node-wise estimate without new same-step neighbor aggregation, while uniform Combine forms a neighborhood-dependent contextual estimate from neighboring adapted states. Their discrepancy directly reflects node–context compatibility, while persistent disagreement across ATC steps is accumulated for detection. Figure 1 illustrates this mechanism conceptually, whereas Figure 2 presents the full iterative DDGAD framework.

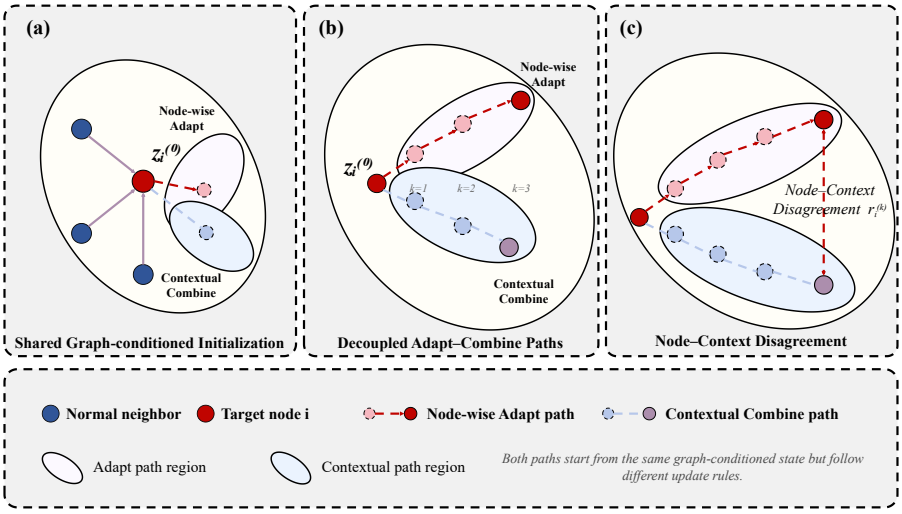

Figure 1: Mechanism illustration of node–context disagreement in DDGAD. Given a shared graph-conditioned state, node-wise Adapt and neighborhood-dependent Combine form estimates under different update rules. Their separation schematically represents node–context incompatibility, with persistent disagreement serving as anomaly evidence.

Our contributions are summarized as follows:

- We introduce a disagreement-driven perspective for GAD that starts from a graph-conditioned initialization and, at each ATC step, contrasts a node-wise Adapt estimate with a neighborhood-dependent contextual Combine estimate. This makes node–context incompatibility directly observable, and persistent pre-consensus disagreement becomes anomaly evidence.
- We characterize the rejected consensus energy spectrally, derive a perturbed population-separation condition from neighborhood composition, and obtain conditional forcing dynamics, a degree-controlled source-response law, and a high-probability persistence bound.
- Experiments on six benchmarks establish strong average AUROC; simple Adapt controls support the generality of disagreement-driven scoring, while matched post-encoder interventions on two Amazon graphs isolate a source-specific response consistent with the theoretical analysis, beyond the effect of added degree.

2 Related Work

Graph anomaly detection spans classical and deep approaches; broader taxonomies, resources, and open challenges are reviewed by Akoglu et al. (2015) and Ma et al. (2023).

2.1 Shallow and Reconstruction-based GAD

Classical non-neural detectors include LOF (Breunig et al., 2000), Isolation Forest (Liu et al., 2008), structural clustering with SCAN (Xu et al., 2007), and residual-based Radar (Li et al., 2017). Deep reconstruction methods score attribute or structural recovery errors, including DOMINANT (Ding et al., 2019), AdONE and DONE (Bandyopadhyay et al., 2020), AnomalyDAE (Fan et al., 2020),

and diffusion-based DiffGAD (Li et al., 2025a). Although their latent models differ, these methods do not explicitly compare a node-wise estimate with a separately constructed neighborhood estimate.

2.2 NEIGHBORHOOD CONSISTENCY-BASED GAD

Contextual detectors exploit neighborhood consistency through contrastive or self-supervised objectives, as in CoLA (Liu et al., 2022), CONAD (Xu et al., 2022), and GraphMAE (Hou et al., 2022). Their compared representations remain neighborhood-dependent, so unreliable context can affect both representation formation and anomaly scoring. DDGAD instead separates post-encoding node-wise and contextual computations before measuring disagreement.

Methods that explicitly model contextual reliability are particularly related. DualGAD captures inconsistency between features and structure through generative and cluster-guided contrastive objectives (Tang et al., 2024), DC-AES suppresses unreliable propagation using abnormal-edge masking (Li et al., 2025b), and GRAM attributes anomaly decisions through gradient attention maps (Yang et al., 2024). DDGAD instead constructs a contextual counter-estimate and retains rejected pre-consensus disagreement as anomaly evidence.

2.3 DIFFUSION AND ADAPT-THEN-COMBINE

Diffusion has been used for counterfactual anomaly generation in CAGAD (Xiao et al., 2024), graph enhancement in DEGAD (Pang et al., 2024), and latent reconstruction in DiffGAD (Li et al., 2025a). DDGAD instead uses EDM as a density-oriented realization of node-wise Adapt, while detection arises from disagreement with a separately computed contextual estimate.

ATC adapts each agent locally and then combines neighboring estimates toward consensus (Sayed, 2014). DDGAD preserves this decomposition but reverses its task semantics: Combine provides a contextual counter-estimate, and the pre-consensus error is retained for detection rather than suppressed. Denoising supplies one realization of node-wise Adapt through the score of a smoothed density (Vincent, 2011; Alain & Bengio, 2014; Stanczuk et al., 2022).

3 METHODOLOGY

The transfer from ATC is semantic rather than modular: the Adapt–Combine decomposition remains recognizable, but the pre-consensus quantity that classical ATC suppresses becomes the detector output. We first formalize the ambiguity created by joint encoding and then define two distinct post-encoding estimates from the same graph-conditioned state.

3.1 PROBLEM FORMULATION

Given an undirected attributed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with adjacency matrix $\mathbf{A} \in \{0,1\}^{N\times N}$ and node features $\mathbf{x}_i \in \mathbb{R}^{d_x}$ stacked into $\mathbf{X} \in \mathbb{R}^{N\times d_x}$, we rank a small anomalous subset $\mathcal{A} \subset \mathcal{V}$ without labels in training or score construction. DDGAD targets contextual anomalies and intrinsic deviations that induce node–context incompatibility; purely intrinsic anomalies that remain context-compatible are outside its separation guarantee. We use d_x for input-feature dimension and d_z for AE latent dimension throughout.

3.2 MOTIVATION: AMBIGUITY AFTER JOINT ENCODING

Message passing jointly encodes node attributes and neighborhood context:

$$\mathbf{z}_i = f_\theta(\mathbf{x}_i, \{\mathbf{x}_j : j \in \mathcal{N}(i)\}), \tag{1}$$

so that a graph-conditioned embedding contains both node and contextual contributions. In a linearized L-layer GCN with normalized adjacency $\widetilde{\mathbf{A}}$ and composed feature map $\mathbf{W}_{1:L}$, a perturbation $\Delta\mathbf{X}$ on anomalous nodes gives

$$\Delta\mathbf{H}^{(L)} = \widetilde{\mathbf{A}}^L \Delta\mathbf{X}\mathbf{W}_{1:L}, \qquad \Delta\mathbf{h}_i^{(L)} = \sum_{j\in\mathcal{A}} (\widetilde{\mathbf{A}}^L)_{ij} \Delta\mathbf{x}_j \mathbf{W}_{1:L}. \tag{2}$$

Hence contextual influence can propagate along length-L walks and alter a node's graph-conditioned state. A single mixed nonlinear embedding does not explicitly expose whether a node-wise estimate remains compatible with the consensus implied by its neighborhood. DDGAD therefore con-

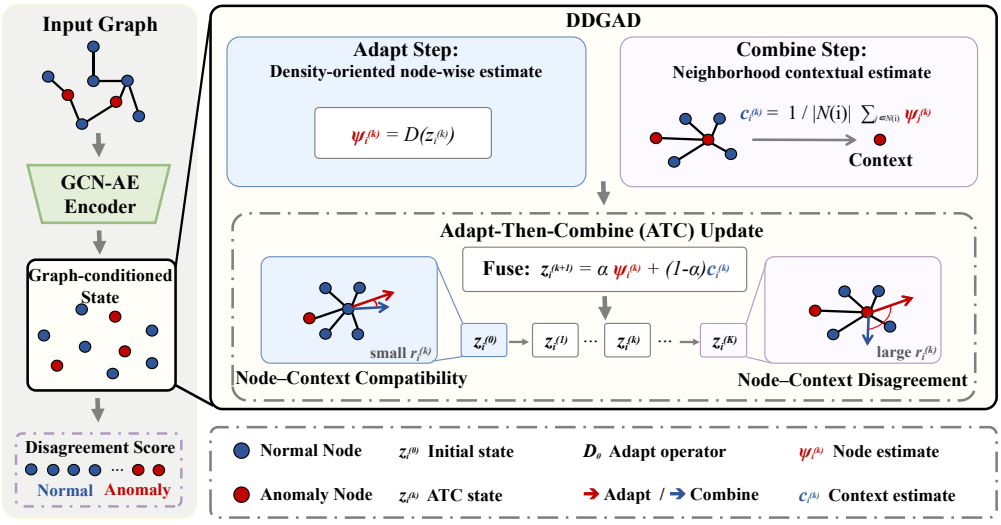

Figure 2: Overall framework of DDGAD. A GCN autoencoder provides the graph-conditioned initialization $\mathbf{z}_i^{(0)}$. At each ATC step, node-wise Adapt produces $\psi_i^{(k)}$, uniform Combine constructs the contextual estimate $c_i^{(k)}$, and Fuse updates $\mathbf{z}_i^{(k+1)}$. Their pre-consensus disagreement is accumulated over K steps to obtain the anomaly score.

structs two distinct post-encoding estimates at each ATC step: a node-wise Adapt estimate and a neighborhood-dependent contextual Combine estimate. Their disagreement directly exposes node–context incompatibility; Appendix A gives the propagation argument.

3.3 Consensus-to-Detection through Inverted ATC

The transfer from ATC is semantic rather than modular: DDGAD preserves the Adapt–Combine decomposition but repurposes it for detection. Starting from the graph-conditioned initialization $\mathbf{z}_i^{(0)}$, DDGAD iterates a shared ATC state $\mathbf{z}_i^{(k)}$. At step k, Adapt produces the node-wise estimate $\psi_i^{(k)}$ without new same-step neighbor aggregation, while Combine forms the neighborhood-dependent contextual estimate $c_i^{(k)}$ from neighboring adapted states. Rather than suppressing their discrepancy before Fuse, DDGAD retains and accumulates it as anomaly evidence.

Adapt. Adapt constructs a node-wise reference

$$\psi_i^{(k)} = \mathcal{D}_\theta(\mathbf{z}_i^{(k)}; \sigma_k, \boldsymbol{\xi}_i^{(k)}), \tag{3}$$

using a shared MLP-based EDM denoiser, noise level σ_k, and sampling noise $\boldsymbol{\xi}_i^{(k)}$. Conditioned on the current states and noises, Adapt reads neither graph topology nor $\mathbf{z}_j^{(k)}$ for $j \neq i$, so it performs no new same-step neighbor aggregation. For $k \geq 1$, however, $\mathbf{z}_i^{(k)}$ may already contain cross-node information introduced by the preceding Fuse. Appendix A gives the conditional Jacobian statement and the mixture-aware density interpretation.

Combine. Combine uses the uniform mean of adapted neighbors:

$$c_i^{(k)} = \begin{cases} \dfrac{1}{|\mathcal{N}(i)|} \displaystyle\sum_{j \in \mathcal{N}(i)} \psi_j^{(k)}, & |\mathcal{N}(i)| > 0, \\ \psi_i^{(k)}, & |\mathcal{N}(i)| = 0. \end{cases} \tag{4}$$

This parameter-free operator deliberately preserves the ordinary neighborhood context. Its difference from Adapt provides the node–context incompatibility signal used for detection.

Fuse. With self-confidence $\alpha \in [0, 1]$, Fuse updates

$$\mathbf{z}_i^{(k+1)} = \alpha\, \psi_i^{(k)} + (1 - \alpha)\, c_i^{(k)}. \tag{5}$$

This convex update retains ATC's distributed-estimation interpretation; its local objective is given in Appendix A. Uniform Combine preserves the observed neighborhood context as the contextual estimate against which Adapt is compared.

3.4 From Consensus Error to Detection Evidence

Classical ATC regards $(\mathbf{I}-\mathbf{W})\mathbf{\Psi}^{(k)}$ as a transient error to be dissipated. DDGAD exposes the same quantity before Fuse and accumulates its node-wise magnitude. Here $\mathbf{W}$ is the fixed row-normalized neighborhood matrix (with an identity row for an isolated node), and $\mathbf{\Psi}^{(k)}$ stacks adapted states. The defining principle is the reinterpretation of rejected pre-consensus energy as anomaly evidence, with EDM serving as one density-oriented realization of Adapt.

3.5 Training and Implementation Details

DDGAD has three stages: a graph convolutional autoencoder (AE) supplies initial embeddings $\mathbf{z}^{(0)}$, an EDM denoiser implements node-wise Adapt, and ATC performs Combine and Fuse. The AE and EDM are trained offline without anomaly labels and remain fixed during inference; ATC introduces no additional trainable parameters. The AE reconstructs node attributes and graph structure; because all nodes participate in this transductive objective, $\mathbf{z}^{(0)}$ remains graph-conditioned and retains contextual influence introduced by graph encoding. Its loss and implementation details are reported in Appendix C.

Diffusion denoiser training. The Adapt operator is trained on frozen AE embeddings with the EDM objective (Karras et al., 2022)

$$\mathcal{L}_{\mathrm{EDM}} = \mathbb{E}_{\mathbf{z}^{(0)},\sigma,\epsilon}\left[\lambda(\sigma)\left\|\mathcal{D}_\theta(\mathbf{z}^{(0)}+\sigma\epsilon;\sigma)-\mathbf{z}^{(0)}\right\|_2^2\right], \tag{6}$$

where $\epsilon \sim \mathcal{N}(0,\mathbf{I})$. Because the target is the observed latent population rather than a label-filtered normal subset, EDM provides a shared density-oriented adaptation rule whose node-wise application performs no new same-step neighbor aggregation. Complete architecture, preconditioning, and sampling details are in Appendix B.

3.6 Disagreement-based Anomaly Scoring

At each ATC step, DDGAD measures the pre-consensus disagreement between the node-wise Adapt estimate and the neighborhood-dependent contextual Combine estimate:

$$r_i^{(k)} \;=\; \|\psi_i^{(k)} - c_i^{(k)}\|_2. \tag{7}$$

Let $\mathbf{C}^{(k)}$ and $\mathbf{R}^{(k)}$ stack the Combine estimates and Adapt–Combine residuals, respectively. Then

$$\mathbf{C}^{(k)} = \mathbf{W}\mathbf{\Psi}^{(k)}, \qquad \mathbf{R}^{(k)} = \mathbf{\Psi}^{(k)} - \mathbf{C}^{(k)} = (\mathbf{I}-\mathbf{W})\mathbf{\Psi}^{(k)}. \tag{8}$$

Here $\mathbf{W}\mathbf{\Psi}^{(k)}$ is the local consensus formed by uniform neighborhood aggregation, so $(\mathbf{I}-\mathbf{W})\mathbf{\Psi}^{(k)}$ is the part of the node-wise reference left unexplained by neighborhood consensus. The Fuse update further gives

$$\|\mathbf{z}_i^{(k+1)} - \psi_i^{(k)}\|_2 = (1-\alpha)r_i^{(k)}. \tag{9}$$

Thus, at iteration k, the distance moved from the node-wise Adapt reference toward neighborhood consensus is exactly proportional to $r_i^{(k)}$. A small residual indicates compatibility between the node-wise reference and its local neighborhood, whereas a large residual indicates a substantial conflict between them. Because the denoiser is stochastic and a single iteration may be transient, DDGAD averages disagreement over the full trajectory:

$$\mathrm{score}_i \;=\; \frac{1}{K}\sum_{k=0}^{K-1} r_i^{(k)}. \tag{10}$$

Persistent averaging retains repeated node–context incompatibility while reducing sensitivity to any one sampling step. The main detector uses only Eq. (10); Table 2 additionally examines its quadratic counterpart.

3.7 ALGORITHM

Algorithm 1 summarizes ATC inference after the AE and EDM have been fixed.

Algorithm 1: DDGAD inference

Require: Graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with adjacency $\mathbf{A}$, node features $\mathbf{X}$, iterations K, α
Ensure: Anomaly scores for all $v_i \in \mathcal{V}$
1: $\mathbf{z}^{(0)} = \mathrm{GCN}(\mathbf{X}, \mathbf{A})$; initialize $e_i = 0$
2: **for** $k = 0$ to $K - 1$ **do**
3: **Adapt:** $\psi_i^{(k)} = \mathcal{D}_\theta(\mathbf{z}_i^{(k)}; \sigma_k, \boldsymbol{\xi}_i^{(k)})$
4: **Combine:** compute $c_i^{(k)}$ by Eq. (4)
5: **Fuse:** $\mathbf{z}_i^{(k+1)} = \alpha\psi_i^{(k)} + (1-\alpha)c_i^{(k)}$
6: **Score:** $e_i \leftarrow e_i + \|\psi_i^{(k)} - c_i^{(k)}\|_2$
7: **end for**
8: **Output:** $\mathrm{score}_i = e_i/K$

3.8 THEORY: REJECTED CONSENSUS ENERGY AND STRUCTURAL SEPARATION

For an undirected graph, let $\mathbf{\Pi}$ be the degree-stationary measure of $\mathbf{W}$ and write its symmetric conjugate as $\mathbf{S} = \mathbf{\Pi}^{1/2}\mathbf{W}\mathbf{\Pi}^{-1/2} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$. For eigenvector $\mathbf{u}_\ell$ and eigenvalue λ_ℓ, define the graph-frequency coefficient

$$\widehat{\mathbf{\Psi}}_\ell^{(k)} = \mathbf{u}_\ell^\top \mathbf{\Pi}^{1/2}\mathbf{\Psi}^{(k)}, \tag{11}$$

and the weighted norm $\|\boldsymbol{Y}\|_{\mathbf{\Pi},F}^2 = \mathrm{tr}(\boldsymbol{Y}^\top \mathbf{\Pi}\boldsymbol{Y})$.

Theorem 1 (Pre-consensus spectral energy)**.** *The pre-consensus residual satisfies*

$$\|\mathbf{R}^{(k)}\|_{\mathbf{\Pi},F}^2 = \sum_{\ell=1}^{N}(1-\lambda_\ell)^2\|\widehat{\mathbf{\Psi}}_\ell^{(k)}\|_2^2. \tag{12}$$

Every component-wise consensus mode has $\lambda_\ell = 1$ and contributes zero; modes rejected most strongly by neighborhood averaging receive the largest weight. At the graph level, Eq. (12) identifies the $\mathbf{\Pi}$-weighted aggregate residual energy rejected by consensus; it does not equate each node score $r_i^{(k)}$ with a localized spectral energy.

Theorem 2 (Population separation under contextual mixing)**.** *At step k, let $\mu_N^{(k)}$ and $\mu_A^{(k)}$ be normal and anomalous Adapt-space prototypes, with $D_k = \|\mu_A^{(k)} - \mu_N^{(k)}\|_2$. Suppose normal and detectable-anomaly neighborhoods contain at most anomaly proportions $\rho_N^{(k)}$ and $\rho_A^{(k)}$, respectively. If node-side estimates deviate from their class prototype by at most $\epsilon_{M,y}^{(k)}$ and contexts deviate from the corresponding two-population mixture by at most $\epsilon_{C,y}^{(k)}$, then any non-isolated detectable anomaly a and non-isolated compatible normal node n satisfy*

$$r_a^{(k)} - r_n^{(k)} \geq (1 - \rho_A^{(k)} - \rho_N^{(k)})D_k - \epsilon_{M,A}^{(k)} - \epsilon_{M,N}^{(k)} - \epsilon_{C,A}^{(k)} - \epsilon_{C,N}^{(k)}. \tag{13}$$

Thus $\rho_A + \rho_N < 1$ defines a sufficient detectable regime rather than a universal graph property. Averaging Eq. (13) over K steps gives a positive score margin whenever its right-hand side has positive mean. Appendix A proves the result and makes the context error degree-dependent under a sub-Gaussian sampling model.

The matched-source experiment also admits an exact population law. For a victim with neighborhood size $m > 0$ and $\beta_r = r/(m+r)$, Appendix A proves for squared disagreement that

$$\Delta_q(r) = A\beta_r + B\beta_r^2. \tag{14}$$

Under the coefficient conditions stated in the appendix, anomalous sources yield a positive response that increases with dose while added degree is exactly matched. The appendix also gives a con-

Table 1: AUROC (%) on six benchmarks. Every cell is reported as mean±sample standard deviation. The highest value is **bold** and the second-highest is underlined. "Avg" is the unweighted mean of the six dataset means.

	Books	Disney	Reddit	Amazon-5KN	Amazon-Bal	Tolokers	Avg
Traditional / shallow non-neural							
LOF (Breunig et al., 2000)	37.96±8.22	51.36±13.12	56.84±2.29	53.33±1.00	54.52±1.32	47.66±0.38	50.28
IF (Liu et al., 2008)	57.58±5.26	42.43±14.22	55.57±1.32	**80.06±0.44**	**80.01±1.48**	39.68±0.48	59.22
SCAN (Xu et al., 2007)	45.42±5.95	51.65±10.86	49.58±0.16	50.00±0.00	50.00±0.00	53.81±0.31	50.08
Radar (Li et al., 2017)	58.17±11.69	70.33±12.41	44.82±1.87	78.43±0.30	78.17±0.74	49.49±0.22	63.24
GNN reconstruction / generative							
GCN-AE (Kipf & Welling, 2017)	51.62±5.61	39.13±19.97	51.18±1.66	26.40±1.02	26.14±1.21	45.84±1.24	40.05
DOMINANT (Ding et al., 2019)	38.51±6.76	49.53±4.27	55.65±0.99	56.78±7.47	56.10±6.82	54.35±0.61	51.82
AnomalyDAE (Fan et al., 2020)	42.95±7.25	49.29±21.68	48.97±0.54	52.51±1.82	45.98±10.08	49.97±5.51	48.28
AdONE (Bandyopadhyay et al., 2020)	43.25±1.56	33.10±9.91	47.85±2.17	36.36±2.92	35.99±2.59	46.24±1.95	40.47
DONE (Bandyopadhyay et al., 2020)	34.33±2.84	38.89±11.36	56.90±1.77	35.94±0.27	32.01±3.33	48.96±1.45	41.17
GAD-NR (Roy et al., 2024)	38.78±5.96	28.49±9.11	55.51±1.38	36.09±2.51	35.19±1.76	55.74±0.89	41.63
GAAN (Chen et al., 2020)	46.99±6.84	48.94±22.06	49.51±2.80	53.49±0.85	53.25±0.67	51.72±0.63	50.65
Contrastive / self-supervised							
CoLA (Liu et al., 2022)	43.17±12.04	50.24±7.44	51.33±1.30	38.17±0.47	37.80±1.30	51.57±0.34	45.38
GraphMAE (Hou et al., 2022)	61.85±4.45	76.60±12.77	48.70±5.15	73.70±0.64	74.59±0.61	43.08±0.40	63.09
CONAD (Xu et al., 2022)	38.39±6.81	47.52±7.58	56.11±0.96	60.09±11.02	56.05±6.80	54.55±0.80	52.12
OCGNN (Wang et al., 2021)	53.10±8.10	74.23±1.68	49.66±7.66	51.94±1.13	52.97±2.22	44.19±0.65	54.35
PREM (Pan et al., 2023)	33.68±7.02	24.00±10.46	57.02±1.42	41.56±7.59	41.16±6.19	47.58±2.05	40.83
GADAM (Chen et al., 2024)	49.91±7.23	28.61±19.12	58.03±1.76	30.42±4.03	31.21±4.49	42.89±1.19	40.18
TAM (Qiao & Pang, 2023)	45.11±8.26	40.54±11.09	**58.20±2.56**	61.33±1.20	62.91±0.31	40.30±0.57	51.40
Diffusion-based							
DiffGAD (Li et al., 2025a)	**69.10±6.42**	52.96±20.06	56.49±0.63	49.69±2.36	48.11±3.06	54.39±2.17	55.12
Ours							
DDGAD (ours)	62.18±8.11	**80.85±9.76**	57.29±2.14	69.98±3.60	67.02±0.74	**62.53±0.24**	**66.64**

Table 2: Signal ablation (AUROC, %). Here r_i is Eq. (10) and $q_i = K^{-1}\sum_{k=0}^{K-1}\|\psi_i^{(k)} - c_i^{(k)}\|_2^2$.

Variant	Score used	Books	Disney
AE only	reconstruction error	51.25	39.97
AE + EDM	denoising deviation	55.88	35.31
Persistent disagreement	r_i	64.04	71.75
Quadratic disagreement	q_i	64.17	72.18

ditional input-to-state bound and concentration results for neighborhood estimation and stochastic averaging.

4 Experiments

4.1 Datasets

We evaluate on six attributed-graph benchmarks: Books, Disney, Reddit, two Amazon co-purchasing variants, and Tolokers. Training is transductive and label-free, and scores are produced for all nodes before a fixed evaluation mask is applied. Dataset statistics, partitions, and implementation details are in Appendix C.

4.2 Main results

We compare with nineteen non-neural, reconstruction/generative, contrastive/self-supervised, and diffusion-based baselines on the same graph instances and evaluation partitions; implementation and score definitions are given in Appendix C.

Over five seeds, DDGAD achieves the highest average AUROC of 66.64% across six benchmarks, exceeding the next-highest average, obtained by Radar, by 3.40 percentage points. It ranks first on Disney and Tolokers and second on Books. Architecture and reporting details are in Appendix C.

4.3 Ablation Study

Table 2 isolates label-free inference signals using one shared AE/EDM checkpoint per dataset.

EDM deviation does not consistently improve AE scoring, whereas both disagreement summaries are substantially stronger. The quadratic variant is only marginally higher than persistent disagreement, by 0.13 AUROC on Books and 0.43 on Disney. We retain the linear persistent score in Eq. (10) as the default because it preserves the scale of each per-iteration Adapt–Combine discrepancy and is less sensitive to isolated large deviations. Their comparable performance identifies Adapt–Combine disagreement, rather than a particular trajectory aggregation form, as the core signal.

4.4 Hyperparameter Sensitivity

Figure 3 traces AUROC across ATC iteration counts K for each dataset. The dataset-dependent response shows that persistence is an active component of the detector rather than an inert repetition count.

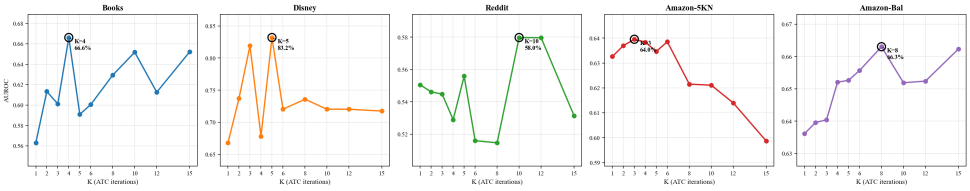

Figure 3: AUROC versus ATC iteration count K on each dataset.

4.5 Controlled Context-Source Intervention

We test whether disagreement responds to the *source* of contextual evidence rather than merely to adding edges. For each Amazon graph, we compute the clean encoder output $\mathbf{z}^{(0)}$ once and freeze both it and all EDM reference snapshots from the clean trajectory. We then change only the adjacency consumed by a uniform Combine operator. In each of 30 randomized trials, 500 normal victims receive $r \in \{1, 2, 4, 8, 16\}$ distinct incoming sources sampled either from anomalous nodes or from normal nodes. The paired endpoint is

$$\Delta_{\text{src}}(r) = s_{\text{anomaly-add}}(r) - s_{\text{normal-add}}(r), \tag{15}$$

so both counterfactuals add the same number of edges. Labels are used only to construct this post-hoc intervention, never for training or score construction. We restrict victims to nodes with non-empty pre-intervention neighborhoods: adding the first neighbor to an isolated node changes the existence of the contextual reference, rather than only its source composition.

Figure 4 reports the complete dose response. Identity already yields a positive source effect, showing that the principle extends beyond diffusion, while persistent EDM is strongest throughout the tested range. The response increases monotonically on both graphs, and all 30 trials have positive dose slopes for every displayed method. Table 3 reports the $r = 16$ matched-source effects and the uncertainty and consistency statistics for persistent EDM.

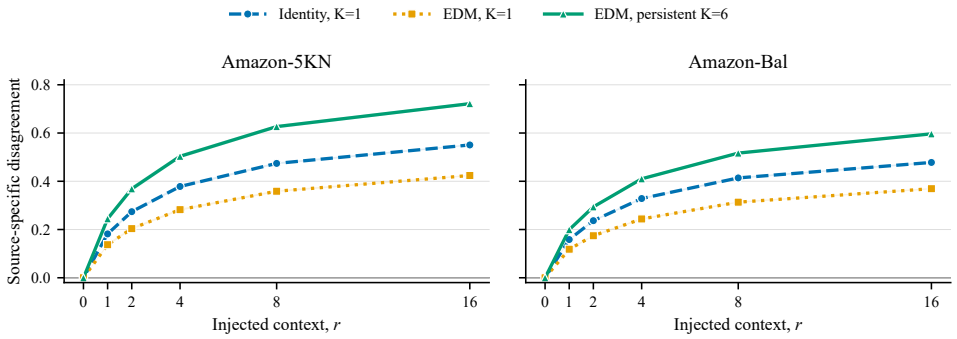

Figure 4: Matched post-encoder source intervention validating source-specific disagreement response. Curves show mean $\Delta_{\text{src}}(r)$ over 30 randomized trials with 500 paired victims per trial; shaded regions are 95% trial-level paired-bootstrap confidence intervals. The clean encoder output and EDM reference trajectory are frozen, and only uniform Combine receives the intervention edges.

Table 3: Matched context-source intervention at $r = 16$. The three method columns report $\Delta_{\mathrm{src}}(16)$ from Eq. (15), not accuracy. Brackets give the persistent-EDM 95% trial-level paired-bootstrap interval from 50,000 resamples. The remaining columns report the standardized persistent-EDM effect, positive-victim fraction, and exact two-sided Wilcoxon signed-rank p-value over 30 trials; both datasets have 30/30 positive trial-level effects.

	Matched-source effect $\Delta_{\mathrm{src}}(16)$			**Persistent-EDM statistics**		
Dataset	**Identity** $K = 1$	**EDM** $K = 1$	**EDM** $K = 6$ **[95% CI]**	$\Delta_{\mathrm{src}}/\sigma_{\mathrm{clean,N}}$	**Victim** $+$	p
Amazon-5KN	0.5505	0.4240	**0.7217 [0.7155, 0.7280]**	0.448	98.0%	1.86×10^{-9}
Amazon-Bal	0.4781	0.3693	**0.5966 [0.5921, 0.6012]**	0.343	98.3%	1.86×10^{-9}

With node-side information fixed and added degree matched, anomalous sources induce consistently larger Adapt–Combine disagreement than normal sources. The stronger persistent-EDM effect further supports the source-sensitivity of disagreement beyond additional connectivity.

Is diffusion necessary for Adapt? To test whether effective node–context disagreement relies on diffusion, we replace EDM with Identity, Linear, and MLP Adapt operators while keeping the frozen AE initialization, uniform Combine, and disagreement score fixed.

Table 4: Trade-off between detection performance and efficiency for Adapt controls on Disney. AUROC/AP are percentages. Time is median complete-ATC latency on an RTX 4070; parameter counts exclude the shared frozen AE. All rows share one frozen AE checkpoint and uniform Combine.

Adapt Operator	**AUROC**	**AP**	**Time (ms)**	**Adapt Params**
Identity	74.82	7.74	1.31	0
Linear	79.43	9.15	1.23	272
MLP	79.08	8.93	1.59	2,128
EDM (Karras et al., 2022)	**85.11**	**14.39**	134.65	471,056

All non-diffusion Adapt controls produce effective detection signals, indicating that Adapt–Combine disagreement is not unique to EDM. Identity already achieves 74.82% AUROC, while Linear and MLP both approach 79% AUROC. EDM performs best with 85.11% AUROC and 14.39% AP, improving over the strongest non-diffusion control at substantially higher computational cost. Thus, diffusion is not necessary for informative disagreement, but provides a stronger, albeit more expensive, Adapt realization.

4.6 Efficiency and Limitations

With latent width d_z, denoiser width h_D, depth L_D, and S_D solver steps, one EDM-based ATC iteration costs $O(S_D N L_D h_D^2 + |\mathcal{E}| d_z)$ up to input/output projections; uniform Combine is the edge-dependent term. On Disney, the complete ATC loop takes 134.65 ms with EDM and 1.59 ms with MLP.

DDGAD exposes rather than removes upstream contamination: coherent anomalous neighborhoods can yield little disagreement as $\rho_A \to 1$, and degree-zero nodes have zero contextual residual.

5 Conclusion

DDGAD reinterprets pre-consensus node–context disagreement as anomaly evidence. Theory characterizes its spectral, structural, and source-response properties, while benchmarks, matched-source interventions, and Adapt controls support the mechanism. The signal is not diffusion-specific, although EDM provides the strongest tested realization. Overall, rejected consensus error offers an informative and interpretable basis for GAD.

AI use statement. Generative AI tools assisted with LaTeX migration, language editing, and compilation troubleshooting. The authors reviewed the resulting manuscript and take responsibility for its content, claims, and artifacts.

REFERENCES

Leman Akoglu, Hanghang Tong, and Danai Koutra. Graph based anomaly detection and description: A survey. *Data Mining and Knowledge Discovery*, 29(3):626–688, 2015.

Guillaume Alain and Yoshua Bengio. What regularized auto-encoders learn from the data-generating distribution. *Journal of Machine Learning Research*, 15(110):3743–3773, 2014.

Sambaran Bandyopadhyay, Lokesh N, Saley Vishal Vivek, and M. Narasimha Murty. Outlier resistant unsupervised deep architectures for attributed network embedding. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, pp. 25–33, 2020. doi: 10.1145/3336191.3371788.

Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. LOF: Identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, pp. 93–104, 2000. doi: 10.1145/342009.335388.

Jingyan Chen, Guanghui Zhu, Chunfeng Yuan, and Yihua Huang. Boosting graph anomaly detection with adaptive message passing. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=CanomFZssu.

Zhenxing Chen, Bo Liu, Meiqing Wang, Peng Dai, Jun Lv, and Liefeng Bo. Generative adversarial attributed network anomaly detection. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pp. 1989–1992, 2020.

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*, pp. 594–602. SIAM, 2019.

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.

Zhenyu Hou, Xiao Liu, Yukuo Cen, Yuxiao Dong, Hongxia Yang, Chunjie Wang, and Jie Tang. GraphMAE: Self-supervised masked graph autoencoders. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 594–604, 2022. doi: 10.1145/3534678.3539321.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 26565–26577, 2022.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.

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*, 2025a.

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

Mark Junjie Li, Shiyang He, Jun Li, Jinren Li, Gen Zhao, Jiang Liu, and Sunjie Huang. Dual-space contrastive learning with abnormal edge suppression for graph anomaly detection. In *ECAI 2025*, pp. 2586–2593, 2025b. doi: 10.3233/FAIA251109.

Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *IEEE International Conference on Data Mining*, pp. 413–422. IEEE, 2008.

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, 2022. doi: 10.1109/TNNLS.2021.3068344.

Xiaoxiao Ma, Jia Wu, Shan Xue, Jian Yang, Chuan Zhou, Quan Z. Sheng, Hui Xiong, and Leman Akoglu. A comprehensive survey on graph anomaly detection with deep learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(12):12012–12038, 2023. doi: 10.1109/TKDE.2021.3118815.

Junjun Pan, Yixin Liu, Yizhen Zheng, and Shirui Pan. PREM: A simple yet effective approach for node-level graph anomaly detection. In *2023 IEEE International Conference on Data Mining (ICDM)*, pp. 1253–1258. IEEE, 2023.

Shikang Pang, Chunjing Xiao, Wenxin Tai, Zhangtao Cheng, and Fan Zhou. Graph anomaly detection with diffusion model-based graph enhancement (student abstract). In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 23610–23612, 2024. doi: 10.1609/aaai.v38i21.30494.

Hezhe Qiao and Guansong Pang. Truncated affinity maximization: One-class homophily modeling for graph anomaly detection. In *Advances in Neural Information Processing Systems*, 2023.

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 17th ACM International Conference on Web Search and Data Mining*, 2024.

Ali H Sayed. Adaptation, learning, and optimization over networks. *Foundations and Trends in Machine Learning*, 7(4-5):311–801, 2014.

Jan Stanczuk, Georgios Batzolis, Teo Deveney, and Carola-Bibiane Schönlieb. Your diffusion model secretly knows the dimension of the data manifold. *arXiv preprint arXiv:2212.12611*, 2022.

Hui Tang, Xun Liang, Jun Wang, and Sensen Zhang. Dualgad: Dual-bootstrapped self-supervised learning for graph anomaly detection. *Information Sciences*, 668:120520, 2024. doi: 10.1016/j.ins.2024.120520.

Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural Computation*, 23(7):1661–1674, 2011.

Xuhong Wang, Baihong Jin, Ying Du, Ping Cui, Yingshui Tan, and Yupu Yang. One-class graph neural networks for anomaly detection in attributed networks. *Neural Computing and Applications*, 33(18):12073–12085, 2021. doi: 10.1007/s00521-021-05924-9.

Chunjing Xiao, Shikang Pang, Xovee Xu, Xuan Li, Goce Trajcevski, and Fan Zhou. Counterfactual data augmentation with denoising diffusion for graph anomaly detection. *IEEE Transactions on Computational Social Systems*, 11(6):7555–7567, 2024. doi: 10.1109/TCSS.2024.3403503.

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*, pp. 824–833, 2007. doi: 10.1145/1281192.1281280.

Zhiming Xu, Xiao Huang, Yue Zhao, Yushun Dong, and Jundong Li. Contrastive attributed network anomaly detection with data augmentation. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2022.

Yifei Yang, Peng Wang, Xiaofan He, and Dongmian Zou. Gram: An interpretable approach for graph anomaly detection using gradient attention maps. *Neural Networks*, 178:106463, 2024. doi: 10.1016/j.neunet.2024.106463.

A THEORETICAL ANALYSIS OF ATC DYNAMICS

We analyze the object implemented by DDGAD: a stochastic node-wise diffusion Adapt operator, a uniform neighborhood Combine operator, and their persistent pre-consensus residual. The results establish structural isolation, a mixture-aware manifold interpretation, an exact spectral characterization of rejected consensus energy, conditional dynamics under forcing, population separation under contextual mixing, a matched-source dose law, and concentration of neighborhood estimation and persistent averaging.

A.1 PROPAGATION BEFORE ATC AND ISOLATION INSIDE ADAPT

Proposition 1 (Propagation support in a linearized GCN)**.** *Let an L-layer linearized GCN be $\mathbf{H}^{(L)} = \widetilde{\mathbf{A}}^L \mathbf{X} \mathbf{W}_{1:L}$, where $\widetilde{\mathbf{A}}$ is the normalized adjacency with self-loops and $\mathbf{W}_{1:L}$ is the composed feature map. For a perturbation $\Delta\mathbf{X}$ supported on node set $\mathcal{A}$,*

$$\Delta\mathbf{H}^{(L)} = \widetilde{\mathbf{A}}^L \Delta\mathbf{X} \mathbf{W}_{1:L}, \qquad \Delta\mathbf{h}_i^{(L)} = \sum_{j \in \mathcal{A}} (\widetilde{\mathbf{A}}^L)_{ij} \Delta\mathbf{x}_j \mathbf{W}_{1:L}. \tag{16}$$

Therefore node i is unaffected whenever no walk of length at most L in the original graph connects it to the perturbed support; equivalently, no length-L walk connects them in the self-loop-augmented graph.

Proof. Substituting $\mathbf{X} + \Delta\mathbf{X}$ into the linearized encoder and subtracting the unperturbed output gives the first equality. Expanding its ith row and using the support of $\Delta\mathbf{X}$ gives the second. The result characterizes where contamination can propagate; it does not imply monotonic growth with depth.□

Proposition 2 (Conditional cross-node isolation of Adapt)**.** *Condition on the EDM sampling noises $\{\boldsymbol{\xi}_i^{(k)}\}_{i=1}^N$. Because the shared Adapt map is applied node by node,*

$$\frac{\partial \psi_i^{(k)}}{\partial \mathbf{z}_j^{(k)}} = \mathbf{0}_{d_z \times d_z}, \qquad j \neq i. \tag{17}$$

If two latent states differ only on a node set S, their adapted outputs, under coupled equal noises, differ only on S.

Proof. The input to $\psi_i^{(k)} = \mathcal{D}_\theta(\mathbf{z}_i^{(k)}; \sigma_k, \boldsymbol{\xi}_i^{(k)})$ contains no $\mathbf{z}_j^{(k)}$ for $j \neq i$; hence Eq. (17) holds. Equal node-wise inputs and coupled noises outside S then produce equal outputs. This is a same-step conditional-isolation statement: Adapt performs no new cross-node aggregation, but for $k \geq 1$ its input $\mathbf{z}_i^{(k)}$ may already contain neighbor information introduced by the preceding Fuse.□

A.2 MIXTURE-AWARE MANIFOLD INTERPRETATION OF DIFFUSION ADAPT

The identity between denoising and score estimation used below follows the standard connection between denoising objectives and the score of a smoothed data density (Vincent, 2011; Alain & Bengio, 2014). Let the frozen AE latent distribution be the mixture

$$p(\mathbf{z}) = (1 - \pi_A) p_N(\mathbf{z}) + \pi_A p_A(\mathbf{z}), \tag{18}$$

where the components denote normal and anomalous latent populations only for analysis; these labels are not used by training. Let $p_{N,\sigma}$, $p_{A,\sigma}$, and p_σ be their Gaussian-smoothed densities.

Proposition 3 (Diffusion score and latent-mixture bias)**.** *For $\mathbf{y} = \mathbf{z} + \sigma\boldsymbol{\epsilon}$ with $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$, the population-optimal MSE denoiser is*

$$\mathcal{D}_\sigma^*(\mathbf{y}) = \mathbb{E}[\mathbf{z} \mid \mathbf{y}] = \mathbf{y} + \sigma^2 \nabla_{\mathbf{y}} \log p_\sigma(\mathbf{y}). \tag{19}$$

Define the local posterior anomaly mass and component scores by

$$\omega_A(\mathbf{y}) = \frac{\pi_A p_{A,\sigma}(\mathbf{y})}{p_\sigma(\mathbf{y})}, \qquad s_q(\mathbf{y}) = \nabla_{\mathbf{y}} \log p_{q,\sigma}(\mathbf{y}), \quad q \in \{N, A\}. \tag{20}$$

Then

$$\nabla \log p_\sigma(\mathbf{y}) = (1 - \omega_A(\mathbf{y})) s_N(\mathbf{y}) + \omega_A(\mathbf{y}) s_A(\mathbf{y}). \tag{21}$$

If the learned denoiser has pointwise error at most $\epsilon_\theta(\mathbf{y}, \sigma)$ from $\mathcal{D}^*_\sigma$, its deviation from the normal-component denoiser obeys

$$\|\mathcal{D}_\theta(\mathbf{y};\sigma) - \mathcal{D}^*_{N,\sigma}(\mathbf{y})\|_2 \le \epsilon_\theta(\mathbf{y},\sigma) + \sigma^2\omega_A(\mathbf{y})\|s_A(\mathbf{y}) - s_N(\mathbf{y})\|_2. \tag{22}$$

Proof. Tweedie's identity for additive Gaussian corruption gives Eq. (19). Differentiating the log-density of the smoothed mixture yields Eq. (21). Finally, subtracting $\mathcal{D}^*_{N,\sigma}(\mathbf{y}) = \mathbf{y} + \sigma^2 s_N(\mathbf{y})$ from the mixture denoiser and applying the triangle inequality with the learned approximation error yields Eq. (22).□

Corollary 1 (Conditional approach to a dominant manifold)**.** *Let $\mathcal{M}_N$ be a dominant normal latent manifold and suppose the ideal normal-component denoiser satisfies*

$$\operatorname{dist}(\mathcal{D}^*_{N,\sigma}(\mathbf{y}), \mathcal{M}_N) \le q_\sigma \operatorname{dist}(\mathbf{y}, \mathcal{M}_N) + \epsilon_M(\sigma), \qquad 0 \le q_\sigma < 1. \tag{23}$$

Then

$$\operatorname{dist}(\mathcal{D}_\theta(\mathbf{y};\sigma), \mathcal{M}_N) \le q_\sigma \operatorname{dist}(\mathbf{y}, \mathcal{M}_N) + \epsilon_M(\sigma) + \epsilon_\theta(\mathbf{y},\sigma) + \sigma^2\omega_A(\mathbf{y})\|s_A(\mathbf{y}) - s_N(\mathbf{y})\|_2. \tag{24}$$

Proof. Distance to a closed set is 1-Lipschitz. Applying this fact to Eq. (22) and then substituting Eq. (23) proves the claim.□

The corollary formalizes the limited sense in which Adapt is manifold-oriented. A high global anomaly ratio is not by itself decisive; the bias is governed by the local posterior mass $\omega_A(\mathbf{y})$. Conversely, a coherent anomalous mode can make this term large, so diffusion is not assumed to be a clean-manifold projector. Multi-step EDM sampling composes such denoising fields; its network and numerical errors are collected in the Adapt approximation term used below.

A.3 Inverted Consensus and Pre-consensus Spectral Energy

Let $\mathbf{W}$ be the fixed row-normalized neighborhood matrix,

$$W_{ij} = \begin{cases} 1/|\mathcal{N}(i)|, & j \in \mathcal{N}(i),\ |\mathcal{N}(i)| > 0, \\ 1, & j = i,\ |\mathcal{N}(i)| = 0, \\ 0, & \text{otherwise.} \end{cases} \tag{25}$$

Let $\mathbf{\Psi}^{(k)}$ stack the adapted states. Then

$$\mathbf{C}^{(k)} = \mathbf{W}\mathbf{\Psi}^{(k)}, \qquad \mathbf{R}^{(k)} = (\mathbf{I} - \mathbf{W})\mathbf{\Psi}^{(k)}. \tag{26}$$

Proposition 4 (Local estimation and correction identity)**.** *For a non-isolated node, the DDGAD update is the unique minimizer*

$$\mathbf{z}_i^{(k+1)} = \arg\min_{\mathbf{u}\in\mathbb{R}^{d_z}} \Big[\alpha\|\mathbf{u} - \psi_i^{(k)}\|_2^2 + \frac{1-\alpha}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)} \|\mathbf{u} - \psi_j^{(k)}\|_2^2\Big]. \tag{27}$$

Furthermore,

$$\|\mathbf{z}_i^{(k+1)} - \psi_i^{(k)}\|_2 = (1-\alpha)\|c_i^{(k)} - \psi_i^{(k)}\|_2 = (1-\alpha)r_i^{(k)}. \tag{28}$$

Proof. The first-order condition for the strictly convex objective in Eq. (27) gives the unique minimizer $\alpha\psi_i^{(k)} + (1-\alpha)c_i^{(k)}$. Subtracting $\psi_i^{(k)}$ from this minimizer yields Eq. (28). For an isolated node, Eq. (25) gives $c_i^{(k)} = \psi_i^{(k)}$, so the residual is zero.□

Because $c_i^{(k)}$ is a convex combination of adapted neighbors and $\mathbf{z}_i^{(k+1)}$ is a convex combination of $\psi_i^{(k)}$ and $c_i^{(k)}$, bounded Adapt outputs imply bounded ATC states. We therefore use this exact boundedness statement instead of claiming geometric convergence of the stochastic, stateful map.

For the undirected graph, set $m_i^+ = \max\{|\mathcal{N}(i)|, 1\}$, $\varpi_i = m_i^+ / \sum_j m_j^+$, and $\mathbf{\Pi} = \operatorname{diag}(\boldsymbol{\varpi})$. The isolated-node convention in Eq. (25) makes $\mathbf{W}$ reversible:

$$\mathbf{\Pi}\mathbf{W} = \mathbf{W}^\top\mathbf{\Pi}. \tag{29}$$

Hence $\mathbf{S} = \mathbf{\Pi}^{1/2}\mathbf{W}\mathbf{\Pi}^{-1/2}$ is symmetric. Write $\mathbf{S} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$, with eigenvalues $\lambda_\ell \in [-1, 1]$, and define the row-vector graph-frequency coefficient

$$\widehat{\mathbf{\Psi}}_\ell^{(k)} = \mathbf{u}_\ell^\top \mathbf{\Pi}^{1/2}\mathbf{\Psi}^{(k)}. \tag{30}$$

Theorem 3 (Pre-consensus spectral characterization)**.** *For* $\|\boldsymbol{Y}\|_{\mathbf{\Pi},F}^2 = \mathrm{tr}(\boldsymbol{Y}^\top \mathbf{\Pi}\boldsymbol{Y})$*,*

$$\|\mathbf{R}^{(k)}\|_{\mathbf{\Pi},F}^2 = \sum_{\ell=1}^{N}(1-\lambda_\ell)^2 \|\widehat{\mathbf{\Psi}}_\ell^{(k)}\|_2^2. \tag{31}$$

Proof. Using reversibility and Eq. (26),

$$\mathbf{\Pi}^{1/2}\mathbf{R}^{(k)} = (\mathbf{I}-\mathbf{S})\mathbf{\Pi}^{1/2}\mathbf{\Psi}^{(k)} = \mathbf{U}(\mathbf{I}-\mathbf{\Lambda})\mathbf{U}^\top \mathbf{\Pi}^{1/2}\mathbf{\Psi}^{(k)}. \tag{32}$$

Taking squared Frobenius norms and using the orthogonality of $\mathbf{U}$ yields Eq. (31).□

Every eigenvector with $\lambda_\ell = 1$ is a component-wise consensus mode and contributes zero. Components most strongly rejected by the averaging operator receive weight $(1-\lambda_\ell)^2$. Thus, in the $\mathbf{\Pi}$-weighted graph-level aggregate, the residual energy is exactly the pre-consensus spectral energy in Eq. (31); the theorem does not assign this identity separately to each node score.

A.4 Persistent ATC Dynamics under Forcing

Consider one connected component, renormalize $\boldsymbol{\varpi}$ on that component, and let $\mathbf{J} = \mathbf{1}\boldsymbol{\varpi}^\top$ be its consensus projector. Define

$$\mathbf{B} = \alpha\mathbf{I} + (1-\alpha)\mathbf{W}, \qquad \rho_\perp(\mathbf{B}) = \max_{\ell \geq 2} |\alpha + (1-\alpha)\lambda_\ell|. \tag{33}$$

Write the stochastic Adapt output as

$$\mathbf{\Psi}^{(k)} = \mathcal{F}_k(\mathbf{Z}^{(k)}) + \mathbf{\Upsilon}^{(k)}, \tag{34}$$

where the shared row-wise map $\mathcal{F}_k$ is L_k-Lipschitz in $\|\cdot\|_{\mathbf{\Pi},F}$ and $\mathbf{\Upsilon}^{(k)}$ collects sampling noise, approximation error, and node- or source-specific forcing.

Theorem 4 (Conditional input-to-state bound)**.** *Let* $\boldsymbol{\delta}^{(k)} = (\mathbf{I}-\mathbf{J})\mathbf{Z}^{(k)}$*. The Fuse dynamics satisfy*

$$\|\boldsymbol{\delta}^{(k+1)}\|_{\mathbf{\Pi},F} \leq q_k\|\boldsymbol{\delta}^{(k)}\|_{\mathbf{\Pi},F} + u_k, \quad q_k = L_k\rho_\perp(\mathbf{B}), \quad u_k = \rho_\perp(\mathbf{B})\|(\mathbf{I}-\mathbf{J})\mathbf{\Upsilon}^{(k)}\|_{\mathbf{\Pi},F}. \tag{35}$$

If $q_k \leq q < 1$ *and* $u_k \leq \bar{u}$*, then*

$$\|\boldsymbol{\delta}^{(k)}\|_{\mathbf{\Pi},F} \leq q^k\|\boldsymbol{\delta}^{(0)}\|_{\mathbf{\Pi},F} + \left(\frac{1-q^k}{1-q}\right)\bar{u}. \tag{36}$$

Proof. Because $\mathbf{B}$ commutes with $\mathbf{J}$ and has weighted operator norm $\rho_\perp(\mathbf{B})$ on the disagreement subspace,

$$\|(\mathbf{I}-\mathbf{J})\mathbf{B}\mathbf{X}\|_{\mathbf{\Pi},F} \leq \rho_\perp(\mathbf{B})\|(\mathbf{I}-\mathbf{J})\mathbf{X}\|_{\mathbf{\Pi},F}. \tag{37}$$

Moreover, $\mathcal{F}_k(\mathbf{J}\mathbf{Z})$ is a consensus matrix. Projection optimality and Lipschitzness therefore give $\|(\mathbf{I}-\mathbf{J})\mathcal{F}_k(\mathbf{Z})\|_{\mathbf{\Pi},F} \leq L_k\|(\mathbf{I}-\mathbf{J})\mathbf{Z}\|_{\mathbf{\Pi},F}$. Substitution into $\mathbf{Z}^{(k+1)} = \mathbf{B}\mathbf{\Psi}^{(k)}$ yields Eq. (35), and induction yields Eq. (36).□

This conditional stability result characterizes contraction under compatible dynamics; stochastic or source-specific inputs determine the residual floor, and persistent anomaly evidence requires non-canceling forcing across iterations.

A.5 Population Separation under Contextual Mixing

At step k, let $\mu_N^{(k)}, \mu_A^{(k)} \in \mathbb{R}^{d_z}$ be normal and anomalous Adapt-space prototypes and define

$$D_k = \|\mu_A^{(k)} - \mu_N^{(k)}\|_2. \tag{38}$$

For node i, let $\rho_i^{(k)}$ denote the anomalous population mass in its neighborhood and let

$$\bar{c}_i^{(k)} = (1-\rho_i^{(k)})\mu_N^{(k)} + \rho_i^{(k)}\mu_A^{(k)} \tag{39}$$

be the corresponding two-population context. This is a reference model, not an assumption that every embedding equals its class center: node heterogeneity, finite-neighborhood error, and mismatch from two-population mixing appear explicitly below.

Theorem 5 (Perturbed population separation)**.** *Let a be a non-isolated anomaly in the detectable regime and n a non-isolated compatible normal node. Suppose*

$$\rho_a^{(k)} \le \rho_A^{(k)}, \qquad \rho_n^{(k)} \le \rho_N^{(k)}, \tag{40}$$

$$\|\psi_i^{(k)} - \mu_{y_i}^{(k)}\|_2 \le \epsilon_{M,y_i}^{(k)}, \qquad \|c_i^{(k)} - \bar{c}_i^{(k)}\|_2 \le \epsilon_{C,y_i}^{(k)} \tag{41}$$

for $i \in \{a,n\}$ and $y_i \in \{A,N\}$. Then

$$r_a^{(k)} - r_n^{(k)} \ge \Gamma_k, \quad \Gamma_k = (1-\rho_A^{(k)}-\rho_N^{(k)})D_k - \epsilon_{M,A}^{(k)} - \epsilon_{M,N}^{(k)} - \epsilon_{C,A}^{(k)} - \epsilon_{C,N}^{(k)}. \tag{42}$$

Consequently,

$$\mathrm{score}_a - \mathrm{score}_n \ge \frac{1}{K}\sum_{k=0}^{K-1}\Gamma_k, \tag{43}$$

and the two nodes are strictly ordered whenever the right-hand side is positive.

Proof. For the anomaly, the reverse triangle inequality and Eqs. (39)–(41) give

$$r_a^{(k)} \ge (1-\rho_a^{(k)})D_k - \epsilon_{M,A}^{(k)} - \epsilon_{C,A}^{(k)} \ge (1-\rho_A^{(k)})D_k - \epsilon_{M,A}^{(k)} - \epsilon_{C,A}^{(k)}. \tag{44}$$

For the normal node, the triangle inequality gives

$$r_n^{(k)} \le \rho_n^{(k)} D_k + \epsilon_{M,N}^{(k)} + \epsilon_{C,N}^{(k)} \le \rho_N^{(k)} D_k + \epsilon_{M,N}^{(k)} + \epsilon_{C,N}^{(k)}. \tag{45}$$

Subtracting the two bounds yields Eq. (42). Averaging over k then yields Eq. (43).□

The condition $\rho_A^{(k)} + \rho_N^{(k)} < 1$ is sufficient, not necessary, and defines a structurally detectable regime for non-isolated nodes. Isolated nodes are excluded because Eq. (25) sets $c_i^{(k)} = \psi_i^{(k)}$ and hence $r_i^{(k)} = 0$. Moreover, as a coherent anomalous community drives $\rho_A^{(k)} \to 1$, its guaranteed margin vanishes. These are expected failure modes of a detector based only on node–context disagreement.

Corollary 2 (Degree-dependent Combine error)**.** *Assume all nodes are non-isolated and write $m_i = |\mathcal{N}(i)|$. Conditional on neighbor types, suppose the centered neighbor Adapt outputs are independent and each coordinate of $\psi_j^{(k)} - \mu_{y_j}^{(k)}$ is σ_C-sub-Gaussian. Then for any $\delta_C \in (0,1)$, with probability at least $1-\delta_C$, simultaneously for all N nodes and K steps,*

$$\|c_i^{(k)} - \bar{c}_i^{(k)}\|_2 \le \sigma_C \sqrt{\frac{2d_z \log(2NKd_z/\delta_C)}{m_i}}. \tag{46}$$

Any additional deviation from the two-population reference adds as a structural-mismatch term to the right-hand side.

Proof. For a fixed node, step, and coordinate, the average of m_i conditionally independent σ_C-sub-Gaussian variables has tail probability at most $2\exp(-m_i t^2/(2\sigma_C^2))$. Applying a union bound over the NKd_z events and using $\|x\|_2 \le \sqrt{d_z}\|x\|_\infty$ yields Eq. (46).□

Corollary 2 makes the role of degree explicit while retaining the unavoidable d_z dependence of a high-dimensional Euclidean-norm bound. It is an idealized finite-sample statement: correlated neighbors or heterogeneous class-conditional means require an effective sample size or an added mismatch term.

A.6 Matched-Source Dose Response

Consider a non-isolated victim with original neighborhood size m, base context c_0, and fixed node-side reference ψ. Add r sources with normal population mean μ_N or anomalous mean $\mu_A = \mu_N + \Delta\mu$. With $\beta_r = r/(m+r)$,

$$c_N(r) = (1-\beta_r)c_0 + \beta_r\mu_N, \qquad c_A(r) = c_N(r) + \beta_r\Delta\mu. \tag{47}$$

Theorem 6 (Degree-controlled source-response law)**.** *For $q_N(r) = \|\psi - c_N(r)\|_2^2$ and $q_A(r) = \|\psi - c_A(r)\|_2^2$,*

$$\Delta_q(r) := q_A(r) - q_N(r) = A\beta_r + B\beta_r^2, \tag{48}$$

where

$$A = -2\langle \psi - c_0, \Delta\mu\rangle, \qquad B = \|\Delta\mu\|_2^2 + 2\langle \mu_N - c_0, \Delta\mu\rangle. \tag{49}$$

If $A \geq 0$, $B \geq 0$, and $(A,B) \neq (0,0)$, then $\Delta_q(r) > 0$ and is strictly increasing for $r > 0$.

Proof. Substituting the matched contexts from Eq. (47) gives

$$\Delta_q(r) = \beta_r^2\|\Delta\mu\|_2^2 - 2\beta_r\langle \psi - c_N(r), \Delta\mu\rangle \tag{50}$$

$$= -2\beta_r\langle \psi - c_0, \Delta\mu\rangle + \beta_r^2\left(\|\Delta\mu\|_2^2 + 2\langle \mu_N - c_0, \Delta\mu\rangle\right). \tag{51}$$

The resulting expression is Eq. (48). Under the stated conditions, β_r is strictly increasing in r and $A + 2B\beta_r > 0$; hence the response is positive and strictly increasing.□

The theorem explains why equal added degree is essential: both counterfactuals share the same β_r, leaving source composition as the difference. For a persistent quadratic score, summing Eq. (48) over steps preserves positivity when the conditions hold step-wise. A positive squared response also orders the corresponding per-step norms, but monotonicity of the difference between raw norms requires additional scale conditions; Figure 4 is therefore empirical evidence consistent with the law rather than a direct proof of its assumptions.

A.7 High-Probability Persistent Separation

Theorem 5 derives a deterministic margin from Adapt-space geometry and contextual composition. We now quantify the reduction of stochastic EDM variation with K. Let $\mathcal{H}_k$ denote the trajectory filtration and define

$$\eta_i^{(k)} = r_i^{(k)} - \mathbb{E}[r_i^{(k)} \mid \mathcal{H}_{k-1}], \qquad \bar{m}_i = \frac{1}{K}\sum_{k=0}^{K-1}\mathbb{E}[r_i^{(k)} \mid \mathcal{H}_{k-1}]. \tag{52}$$

Theorem 7 (Persistent-score concentration)**.** *Suppose $\eta_i^{(k)}$ is conditionally σ_i-sub-Gaussian, i.e., for all $t \in \mathbb{R}$,*

$$\mathbb{E}[\exp(t\eta_i^{(k)}) \mid \mathcal{H}_{k-1}] \leq \exp(t^2\sigma_i^2/2). \tag{53}$$

Then

$$\Pr(|\text{score}_i - \bar{m}_i| \geq t) \leq 2\exp\left(-\frac{Kt^2}{2\sigma_i^2}\right). \tag{54}$$

Proof. Iterating the conditional moment-generating-function bound produces an exponential supermartingale for $\sum_k \eta_i^{(k)}$. Chernoff's method applied to both tails, followed by division by K, yields Eq. (54).□

Consequently, for a fixed pair consisting of an anomaly and a normal node, with conditional-mean gap $\bar{m}_A - \bar{m}_N$, both scores deviate by at most

$$t_\delta = \sigma\sqrt{\frac{2\log(4/\delta)}{K}}, \tag{55}$$

with probability at least $1-\delta$ when $\sigma_A, \sigma_N \leq \sigma$. The ordering is therefore preserved whenever $\bar{m}_A - \bar{m}_N > 2t_\delta$. This formalizes the $K^{-1/2}$ variance reduction of persistent averaging under the stated martingale condition. Thus K reduces stochastic variation, while its effect on the underlying mean gap remains governed by the dynamics.

For an all-node statement, if $\sigma_i \le \sigma_T$ for every node, a union bound gives

$$\max_i |\text{score}_i - \bar{m}_i| \le t_T := \sigma_T \sqrt{\frac{2\log(2N/\delta_T)}{K}} \tag{56}$$

with probability at least $1 - \delta_T$. Hence every detectable anomaly is ranked above every compatible normal node whenever their conditional-mean margin is uniformly larger than $2t_T$. The structural term in Eq. (42), the degree-dependent context error in Eq. (46), and the temporal term in Eq. (56) respectively expose the roles of class separation, neighborhood sample size, and persistence. Their joint application rests on the stated prototype, conditional-independence, and martingale assumptions.

B Additional Method and Training Details

B.1 Autoencoder objective

The graph convolutional autoencoder reconstructs node attributes and graph structure with

$$\mathcal{L}_{\text{AE}} = \frac{1}{N}\sum_{i=1}^{N} \lambda_x \|\mathbf{x}_i - \widehat{\mathbf{x}}_i\|_2 + \frac{1}{N}\sum_{i=1}^{N} \lambda_A \|\mathbf{A}_{i,:} - \widehat{\mathbf{A}}_{i,:}\|_2 + \lambda_e \mathcal{L}_{\text{edge}}. \tag{57}$$

For large graphs, the dense structure term is omitted to control memory. All nodes participate in this transductive objective, so the resulting initialization remains graph-conditioned.

This section documents the node-wise EDM implementation (Karras et al., 2022) used by DDGAD. Training and inference operate on frozen AE latent embeddings. The MLP denoiser never receives graph adjacency; topology is introduced only by Combine.

B.2 Forward perturbation.

The forward process adds Gaussian noise with level σ to an observed AE latent vector $\mathbf{x}_0$:

$$\mathbf{x}_\sigma = \mathbf{x}_0 + \sigma\mathbf{e}, \qquad \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{58}$$

The noise level σ is sampled from a log-normal distribution with mean -1.2 and standard deviation 1.2. We adopt the EDM parameterization as a generic continuous denoising framework and use fixed dataset-level noise ranges rather than borrowing the image-domain schedule directly.

B.3 Training loss.

The denoising network $D_\theta(\mathbf{x}, \sigma)$ predicts the original observed latent $\mathbf{x}_0$ from $\mathbf{x}_\sigma$ by minimizing

$$\mathcal{L}_{\text{EDM}} = \mathbb{E}_{\sigma,\mathbf{x}_0,\mathbf{e}}\Big[w(\sigma)\,\big\|D_\theta(\mathbf{x}_0 + \sigma\mathbf{e}, \sigma) - \mathbf{x}_0\big\|_2^2\Big], \tag{59}$$

with the EDM weighting function

$$w(\sigma) = \frac{\sigma^2 + \sigma_{\text{data}}^2}{(\sigma \cdot \sigma_{\text{data}})^2}, \qquad \sigma_{\text{data}} = 0.5. \tag{60}$$

This is the EDM preconditioning weight used by the implementation.

B.4 Architecture and training.

The denoiser is a node-wise MLP with a sinusoidal noise embedding, input projection, LayerNorm, SiLU activations, and an output projection back to the AE latent dimension. Its internal width is at least 128. Training is full-graph, uses Adam, gradient clipping, a step learning-rate schedule, and early stopping on the unsupervised EDM training loss.

B.5 Sampling.

At each ATC iteration, DDGAD perturbs the current latent with fresh Gaussian noise and applies a power-law EDM schedule followed by Heun correction. Writing the numerical sampler as $\mathcal{S}_\theta$,

$$\psi_i^{(k)} = \mathcal{S}_\theta\Big(\mathbf{z}_i^{(k)} + \sigma_{\max}\boldsymbol{\xi}_i^{(k)}; \sigma_{\min}, \sigma_{\max}\Big), \qquad \boldsymbol{\xi}_i^{(k)} \sim \mathcal{N}(0, \mathbf{I}). \tag{61}$$

No edge index is passed to $\mathcal{S}_\theta$.

B.6 Interpretation and boundary.

EDM is trained on the full latent population rather than a label-filtered normal subset, so it provides a density-oriented node-wise reference rather than a guaranteed clean-manifold projector. At each step, conditional on the current graph-conditioned inputs $\mathbf{z}^{(k)}$, Adapt introduces no new same-step cross-node aggregation; earlier Combine/Fuse updates may already have mixed neighboring information. Proposition 3 quantifies how local anomaly posterior mass biases the ideal denoising direction. Anomaly evidence arises only when this Adapt reference and the neighborhood context exhibit a persistent gap.

C Additional Experimental Details

C.1 Dataset statistics

Table 5: Dataset statistics.

Dataset	Nodes	Edges	Features	Anomaly ratio
Books	1,418	3,695	21	1.97%
Disney	124	335	28	4.84%
Reddit	10,984	168,016	64	3.33%
Amazon-5KN	5,000	10,000	25	16.42%
Amazon-Bal	5,000	10,000	25	16.42%
Tolokers	11,758	530,758	10	21.82%

C.2 Implementation and reporting protocol

The autoencoder contains a two-layer GCN encoder and a two-layer GCN attribute decoder with a dot-product structure decoder, residual input projection, and layer normalization. We use Adam for both stages. The denoiser uses EDM preconditioning with $P_{\mathrm{mean}} = -1.2$, $P_{\mathrm{std}} = 1.2$, $\sigma_{\mathrm{data}} = 0.5$, and $\sigma_{\mathrm{min}} = 0.002$. Early stopping monitors only the unsupervised EDM training loss.

For all six benchmarks and five seeds reported in Table 1, the Fuse weight is $\alpha = 0.7$ and the ATC horizon is $K = 6$.

Transductive evaluation partitions. The detector is trained transductively on the full graph: features and edges of every node, including nodes later assigned to the test partition, are available to the AE and EDM, whereas anomaly labels are never provided to either model. Labels are used only outside the detector to construct deterministic class-stratified evaluation partitions and to compute metrics. For each seed, 50% of anomalous nodes and 20% of normal nodes are first assigned to a train/validation pool; 20% of each class-specific pool is then assigned to validation, and the remainder of that pool is assigned to train. All nodes not selected into the pool form the test evaluation partition. Up to integer rounding, the resulting train/validation/test proportions are therefore 40%/10%/50% for anomalous nodes and 16%/4%/80% for normal nodes. The names "train" and "validation" are retained for partition bookkeeping and compatibility across experiments: DDGAD does not use these masks for parameter optimization, early stopping, checkpoint selection, inference, or score construction. It produces one anomaly score for every node before the test mask is applied, and the reported benchmark AUROC is computed only on the fixed test partition.

Main-table values are the mean and sample standard deviation over five independently run seeds. Every method uses the same AUROC computation on its deterministic seed-specific test partition. Within each controlled ablation table, all rows share one checkpoint and evaluation protocol; independently trained ablation suites are interpreted through their within-table relative differences.

Main-table reruns. We rerun all main-table methods on the six stored graph instances with five seeds and report the resulting mean and sample standard deviation. GCN-AE is our label-free reconstruction control: a two-layer GCN is trained for 100 epochs with learning rate 0.01 to reconstruct node attributes under mean-squared error, and the per-node squared reconstruction error is used as its anomaly score. GraphMAE is originally a self-supervised representation learner rather than a dedicated anomaly detector. We use the PyG implementation with 200 pretraining epochs, row-normalized features, a 0.5 mask rate, and the scaled-cosine-error objective with exponent 2; its

node score is the mean masked-node reconstruction error over 20 independently sampled masks. SCAN is evaluated as a non-neural structural clustering baseline. The remaining baselines use their corresponding implementations and the seed-specific test partitions described above.

C.3 CONTEXT-SOURCE INTERVENTION PROTOCOL

The intervention in Section 4.5 is a post-hoc mechanism analysis, not part of detector training or benchmark evaluation. For each Amazon graph, we load one pre-existing checkpoint, compute $\mathbf{z}^{(0)}$ on the original graph once, and generate five stochastic realizations of the clean EDM trajectory. These node-reference snapshots are frozen before any intervention. Intervention edges are visible only to a uniform Combine operator; neither the GCN encoder nor EDM receives the modified adjacency.

Amazon-5KN and Amazon-Bal each contain 5,000 nodes, including 821 labeled anomalies. Among their 4,179 normal nodes, 1,454 and 1,395 nodes, respectively, are isolated under the adjacency used by Combine. We exclude these degree-zero nodes from the matched intervention because adding their first neighbor changes the existence of the contextual reference itself. This leaves 2,725 and 2,784 eligible normal victims. Each of 30 trials samples 500 victims and, for every victim, nested sets of $r \in \{1, 2, 4, 8, 16\}$ distinct non-neighbor sources. The anomaly-add and normal-add counterfactuals use the same victim and edge count; sources differ only in post-hoc label-defined type.

Uncertainty is computed over the 30 trial-level paired mean effects. Figure 4 uses 50,000-resample percentile bootstrap confidence intervals at each dose. Table 3 reports the same interval for persistent EDM at $r = 16$ and an exact two-sided Wilcoxon signed-rank test of the 30 trial-level paired effects against zero. The standardized effect divides the mean paired source effect by the clean-score standard deviation among normal nodes. "Victim $+$" averages, over trials, the within-trial fraction of victims satisfying $s_{\text{anomaly-add}} > s_{\text{normal-add}}$.

Including isolated victims changes the estimand by simultaneously introducing the first contextual reference and can reverse the aggregate response; we therefore restrict the source-controlled comparison to non-isolated victims and the equal-degree normal-source counterfactual.