

Learnable Kernel Density Estimation for Graphs and Its Application to Graph-Level Anomaly Detection

Xudong Wang, Ziheng Sun, Chris Ding, Jicong Fan

School of Data Science, The Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen), China
{xudongwang, zihengsun}@link.cuhk.edu.cn {chrisding, fanjicong}@cuhk.edu.cn

Motivation: Density Estimation on Graphs

Goal. Given a collection of normal graphs $\mathcal{G} = \{G_i\}_{i=1}^N$, where $G_i = (V_i, E_i, \mathbf{X}_i)$, learn a density estimator:

$$f: \mathcal{G} \rightarrow \mathbb{R}_+$$

that maps graphs to non-negative density values, capturing the underlying distribution of \mathcal{G} , and its low-density values imply high anomaly scores:

$$s(G) = -\bar{f}(G), \quad \text{flag if } s(G) > \tau$$

Why graph density is Non-trivial

- Graphs are discrete, non-Euclidean, and size-varying.
- The estimator must be permutation-invariant but structurally sensitive.
- Normal graphs may occupy heterogeneous, multi-scale density regions.
- Training is unsupervised: no labeled anomaly graphs are available.

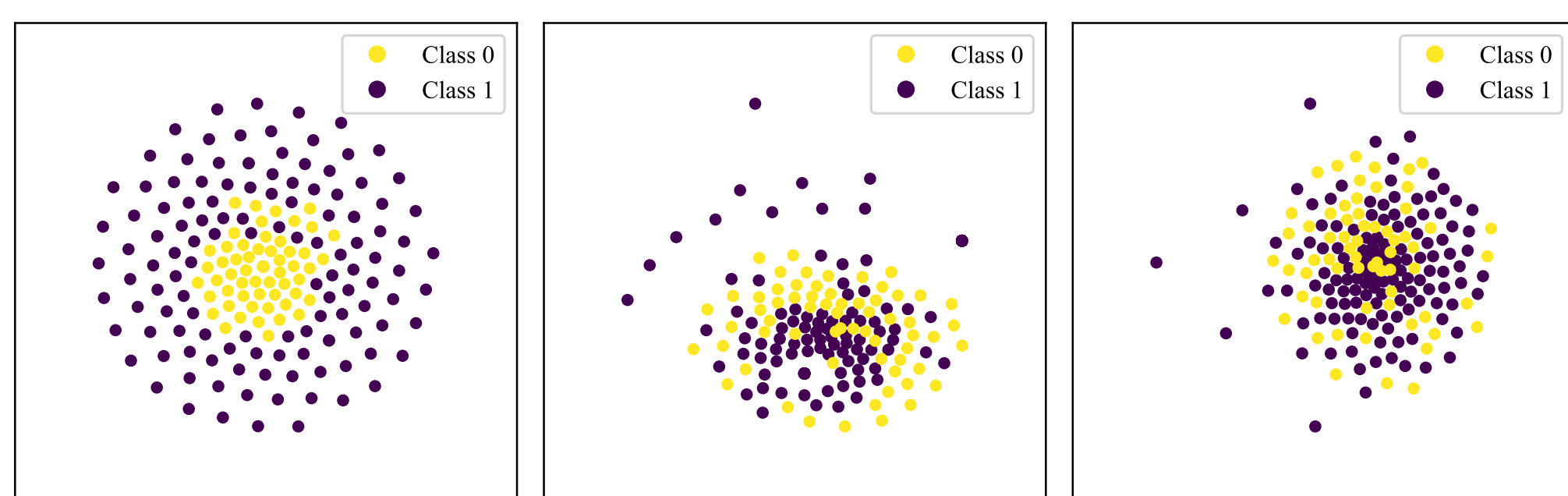


Figure 1: MUTAG kernel t-SNE: learned LGKDE geometry is cleaner than fixed WL/PK kernels.

Limitation of classic graph-kernel KDE: WL/PK kernels are handcrafted and fixed, so the metric and bandwidth cannot adapt to task-relevant graph semantics.

Core Insight and Contributions

Key idea: Learn a graph metric and a multi-scale KDE jointly, using controlled perturbations to expose density boundaries for effective density contrastive learning.

1. **Deep graph MMD metric:** each graph becomes a node-embedding distribution; graph distances are MMDs.
2. **Learnable multi-scale KDE:** the estimator adapts bandwidth mixture weights instead of fixing one scale.
3. **Structure-aware perturbations:** feature swaps and spectral edge edits generate meaningful density contrasts.
4. **Theory:** consistency, rate, robustness, and generalization for graph density estimation.
5. **Empirics:** recovering underlying graph distribution on synthetic graphs and achieving strong graph-level anomaly detection performance compared with recent advanced baselines.

Training Principle: Density Contrasting

Direct density maximization can collapse representations. LGKDE trains with a local ordering constraint:

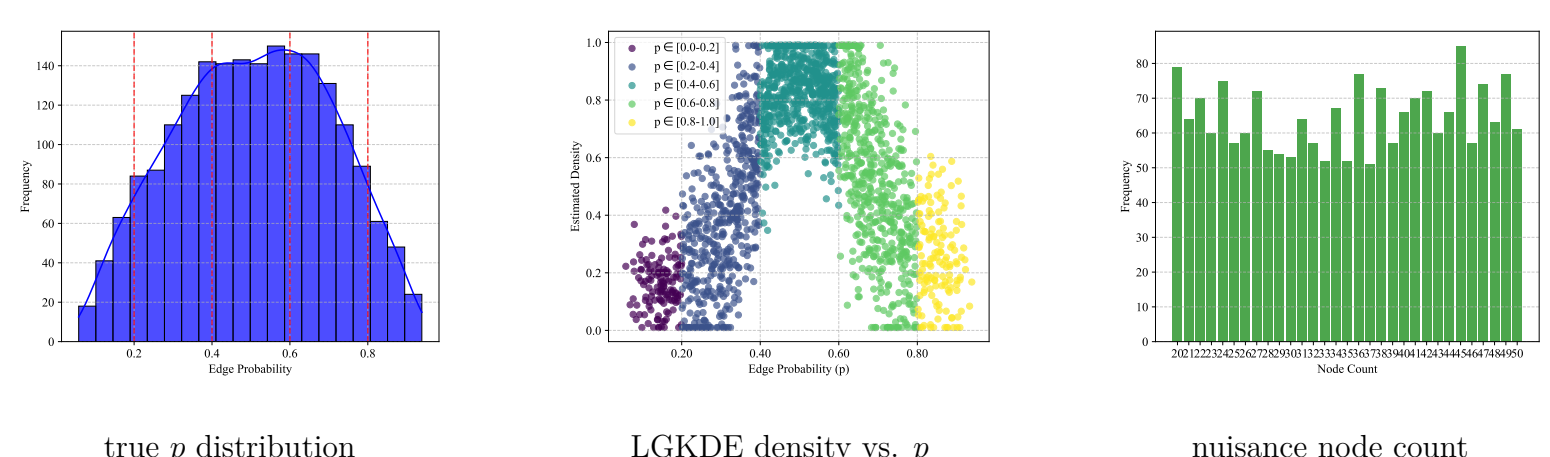
A normal graph should be denser than its controlled perturbations.

$$\max_{\theta, \alpha} \sum_{i=1}^N \sum_{j=1}^{N_{\text{pert}}} \frac{\bar{f}_{\theta, \alpha}(G_i) - \bar{f}_{\theta, \alpha}(G_i^{(j)})}{\bar{f}_{\theta, \alpha}(G_i)}$$

Interpretation

The perturbed graph is *not forced to be anomalous*; it is a locally meaningful contrast that teaches where normal density should decay.

Synthetic Density Recovery



ER graphs are sampled with $n \sim U[20, 50]$, $p \sim \text{Beta}(2, 2)$. The LGKDE learned density peak around $p \approx 0.5$ does not have missance variation in graph size.

The Proposed LGKDE Framework

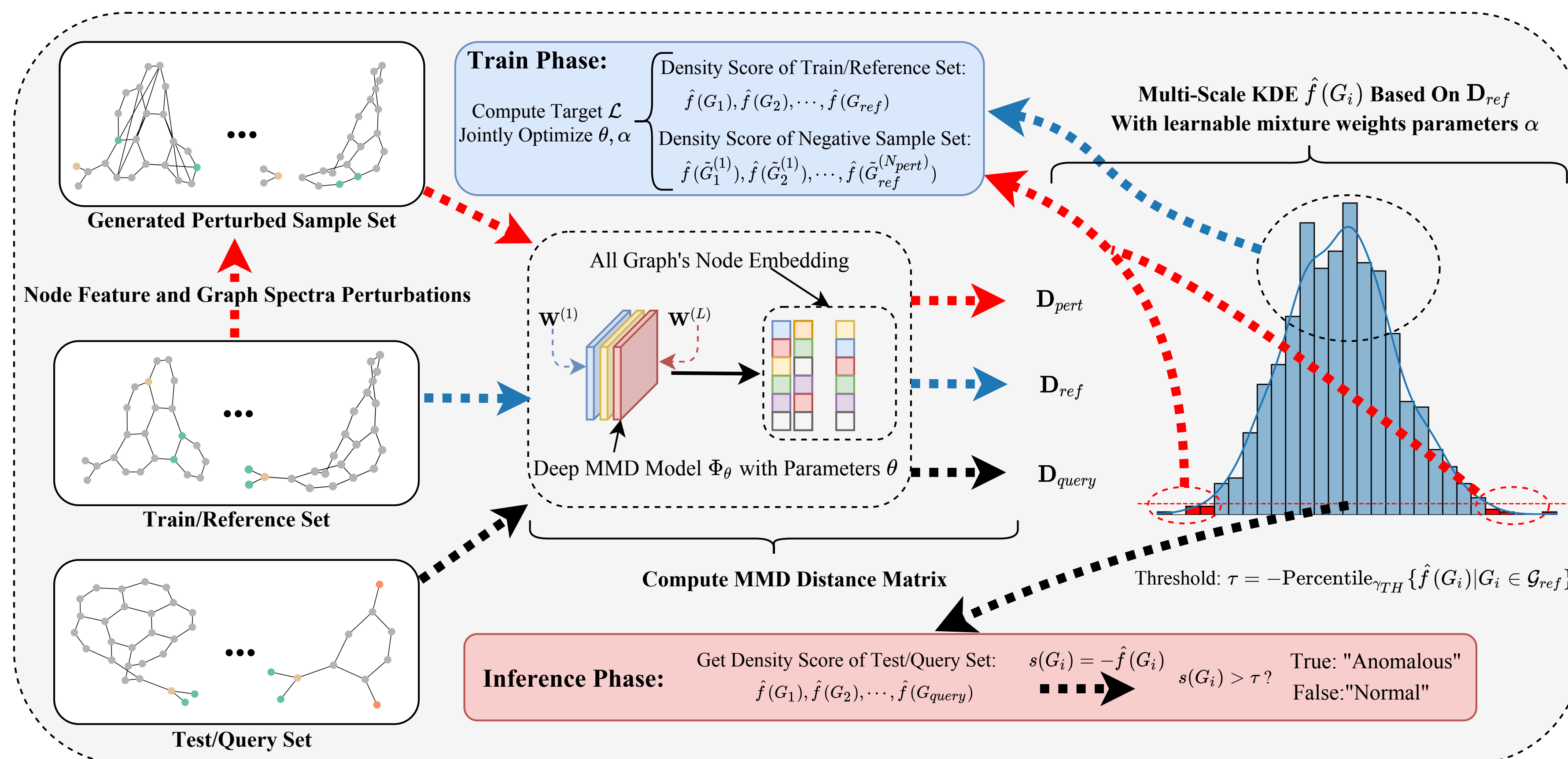


Figure 2: LGKDE jointly learns graph representations, an MMD distance matrix, and a multi-scale KDE. Training contrasts reference graphs against structure-aware perturbations; inference scores low-density query graphs as anomalous.

End-to-end pipeline: train/reference graphs \rightarrow GNN node embeddings \rightarrow graph MMD distances \rightarrow learnable multi-scale KDE \rightarrow anomaly score $s(G) = -\bar{f}_{\text{KDE}}(G)$.

Learned Graph Metric + Multi-Scale KDE

1. **Graph as node-embedding distribution**

$$\mathbf{Z}_i = \text{GNN}_{\theta}(\mathbf{A}_i, \mathbf{X}_i), \quad \mathbf{z}_p^{(i)} \in \mathbb{R}^d.$$

2. **Deep graph MMD distance**

$$d_{\text{MMD}}(G_i, G_j) = \sup_{k_{\gamma} \in \mathcal{K}} \|\mu_{\gamma}(G_i) - \mu_{\gamma}(G_j)\|_{\mathcal{H}}.$$

For Gaussian kernels, $k_{\gamma}(\mathbf{u}, \mathbf{v}) = \exp(-\gamma \|\mathbf{u} - \mathbf{v}\|^2)$; multiple γ values capture multiple structural scales.

3. **Learnable multi-scale KDE**

$$\bar{f}_{\text{KDE}}(G) = \sum_{k=1}^M \pi_k(\alpha) \phi_k(G), \quad \pi_k = \text{softmax}(\alpha)_k.$$

$$\phi_k(G) = \frac{1}{N} \sum_{i=1}^N K_{\text{KDE}}(d_{\text{MMD}}(G, G_i), h_k).$$

Inference:

$$s(G) = -\bar{f}_{\text{KDE}}(G), \quad \tau = -\text{Percentile}_{\gamma_{\text{TH}}} \{\bar{f}_{\text{KDE}}(G_i)\}_{i=1}^N.$$

Structure-Aware Perturbations

Node feature perturbation. Randomly swap features on a subset $\mathcal{V}_{\text{swap}}$:

$$\mathbf{X}'_v = \begin{cases} \mathbf{X}_{\text{perm}(v)}, & v \in \mathcal{V}_{\text{swap}}, \\ \mathbf{X}_v, & \text{otherwise.} \end{cases}$$

Energy-based spectral perturbation. Given an adjacency matrix \mathbf{A} , we perform SVD decomposition $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, and let $\{\sigma_i\}_{i=1}^n$ be the singular values in descending order. The cumulative energy ratio up to index k as:

$$E(k) = \frac{\sum_{i=1}^k \sigma_i^2}{\sum_{j=1}^n \sigma_j^2},$$

Energy thresholds τ_1, τ_2 and $\tau_1 < \tau_2$. Default $(\tau_1, \tau_2) = (0.50, 0.75)$.

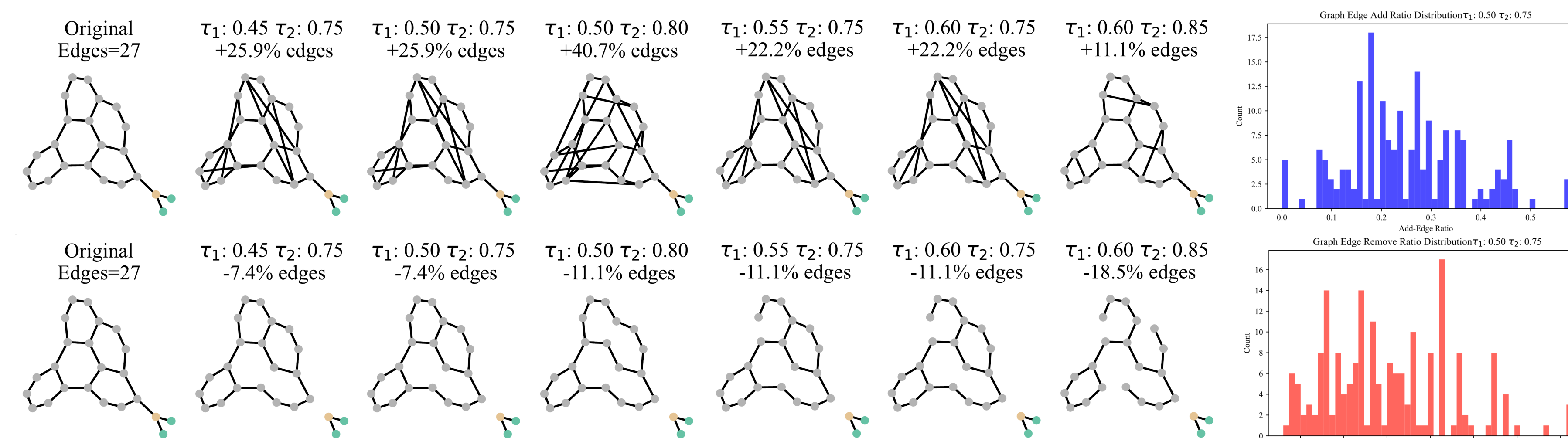


Figure 3: Spectral perturbations visualization of the MUTAG 1st sample graph. The default $(\tau_1, \tau_2) = (0.50, 0.75)$ produces controlled, topology-preserving add/remove variants instead of random noise. The right two histograms show the edge change ratio of spectral perturbation across the whole MUTAG datasets with $(\tau_1, \tau_2) = (0.50, 0.75)$.

Theoretical Guarantees

Assumptions. Bounded GNN weights; f^* is bounded and smooth; bandwidths satisfy $h_k \rightarrow 0$ and $Nh_k^{d_{\text{int}}} \rightarrow \infty$.

1. **Consistency.** The learned graph KDE converges to the target density:

$$\|\bar{f}_{\text{KDE}} - f^*\|_1 \xrightarrow{P} 0.$$

2. **Convergence Rate.** With $h^* \asymp N^{-1/(4+d_{\text{int}})}$,

$\text{MISE} = O(N^{-\frac{4}{4+d_{\text{int}}}})$, $d_{\text{int}} = 1 \Rightarrow O(N^{-0.8})$. Bounded node/spectral edits therefore induce bounded density-score changes.

4. **Generalization.** With N normal reference graphs and bounded GNN weight norms,

$$\mathbb{E}_{\mathcal{G}} |\bar{f}_{\text{KDE}}(G) - f^*(G)| \leq \bar{\Delta}_{\mathcal{G}} + O\left(\frac{(1+\alpha) \log N}{N}\right) + O\left(\sqrt{\frac{\log(1/\delta)}{N}}\right).$$

Implication: LGKDE is a learnable graph-density estimator: more references improve accuracy, while controlled perturbations give stable anomaly scores.

Main Results: Graph-Level Anomaly Detection

Avg. AUROC (% \uparrow) Due to space constraints, we present representative results in this poster. Full comparisons with all baselines and metrics are in the paper.

Method	MUT	PROT	DD	ENZ	DHFR	BZR	COX2	AIDS	IMDB	NCII	COLL	RED-B	Avg. Rank
WL-iF	65.7	65.8	70.5	51.0	51.6	51.7	49.6	61.4	51.8	50.4	51.4	49.8	55.90
OCCGN	79.6	76.5	79.1	62.4	61.1	69.1	57.8	96.9	61.5	69.5	60.6	82.1	71.34
GLocalKD	86.3	77.3	80.8	61.8	61.8	68.6	58.9	96.9	53.3	65.3	51.9	80.3	70.25
SIGNET	88.8	75.9	74.5	63.1	72.9	80.8	72.4	97.6	70.1	74.3	72.5	85.2	77.34
CVTGAD	86.6	76.5	78.8	68.6	63.2	77.7	64.4	99.2	69.8	69.1	71.0	87.4	76.03
MUSE	85.9	76.9	79.2	67.8	71.5	78.3	65.9	99.0	67.8	74.5	67.5	85.3	76.63
UniFORM	88.5	77.2	78.6	69.3	69.8	79.9	65.1	98.5	68.4	72.9	66.2	84.7	76.58
LGKDE	91.6	79.0	79.8	71.0	82.6	81.1	66.7	99.1	68.8	76.7	67.9	88.1	79.37

Avg. AUPRC (% \uparrow).

Method	MUT	PROT	DD	ENZ	DHFR	BZR	COX2	AIDS	IMDB	NCII	COLL	RED-B	Avg. Rank
WL-iF	55.2	48.0	46.6	24.1	34.4	76.3	81.1	29.7	59.3	51.6	52.2	74.6	52.75
OCCGN	74.1	79.8	84.2	32.7	50.8	88.3	83.0	93.4	60.3	54.2	57.0	85.6	70.28
GLocalKD	75.5	77.0	87.5	29.3	47.3	84.2	76.6	95.3	54.5	40.2	47.9	81.5	66.38
SIGNET	77.4	75.2	75.4	22.2	58.6	92.0	86.4	93.6	67.1	66.5	69.2	85.5	69.92
CVTGAD	81.3	83.9	72.9	30.2	52.4	90.1	85.2	96.4	68.9	65.1	70.6	88.2	73.77
MUSE	82.7	82.3	85.2	31.9	61.3	87.2	65.9	94.2	58.7	65.5	63.9	87.9	71.93
UniFORM	80.5	81.6	83.8	34.6	58.9	87.7	61.5	94.1	56.3	64.8	62.3	86.5	71.03
LGKDE	96.8	85.1	89.3	52.8	78.6	88.8	85.0	95.3	87.2	92.2	68.5	88.6	84.01

Ablation Studies

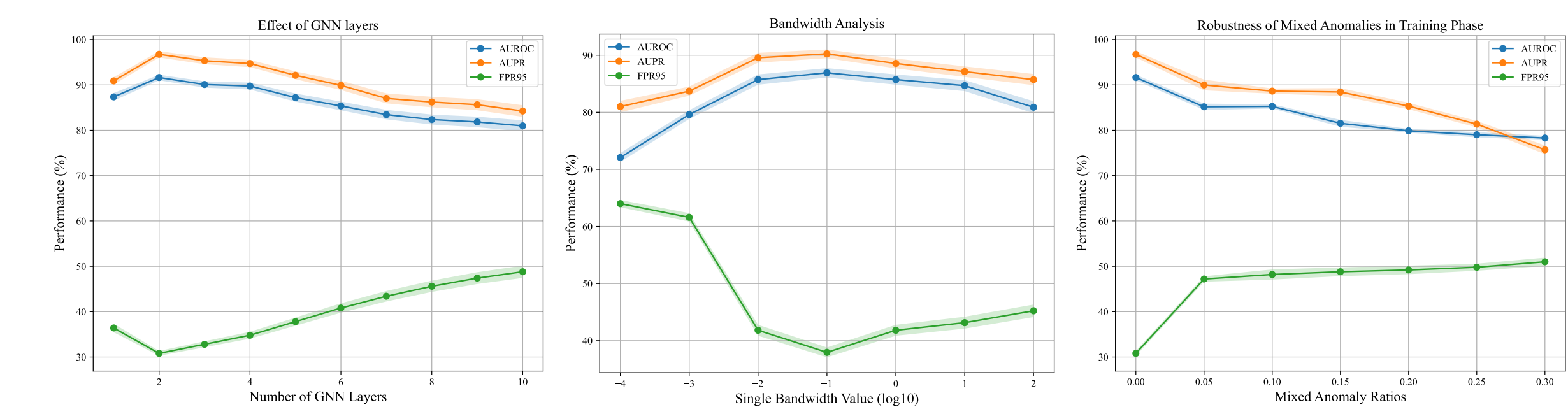
Due to space constraints, we only present representative results in this poster. See full results in our paper.

Component ablation on MUTAG

Variant	AUROC	AUPRC	FPR95
Single $h = 10^{-1}$	86.92	90.24	37.95
Readout avg.	83.89	86.92	44.73
Fixed weights	88.92	89.64	36.97
Full LGKDE	91.63	96.75	30.80

Perturbation Ablation

Variant	MUTAG AUROC	MUTAG AUPRC	PROT AUROC	PROT AUPRC
Feature-only	84.23	89.67	72.34	78.56
Spectral-only	87.89	92.45	75.67	81.23
Random edges	81.45	86.78	69.12	75.89
Full	91.63	96.75	78.97	85.08



Insight: gains come from joint metric learning, adaptive multi-scale KDE, and structure-aware perturbations rather than simply increasing GNN capacity.

Takeaway

LGKDE = learned graph metric + adaptive multi-scale KDE + novel graph perturbations for effective density contrastive learning; graph-level anomaly detection \Leftrightarrow principled graph density estimation.