PRELIMINARY PREPRINT VERSION – DO NOT CITE
This is not the version of record. The final version will appear in the Proceedings of IJCAI-ECAI 2026.

Geometry-Aware Riemannian Residual Displacement for Cross-Broad-Domain Graph Anomaly Detection

Senyong Wang, **Zihao Yao**, **Chen Zhu** and **Yaying Zhang***
Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai, China
{wangsenyong, yaozihao, zhuchen, yaying.zhang}@tongji.edu.cn

Abstract

Cross-Domain Graph Anomaly Detection supports the transfer of knowledge to unknown targets. Nevertheless, current approaches frequently struggle in the Cross-Broad-Domain paradigm which is characterized by two significant discrepancies: feature heterogeneity and structural disparity. Such diverse geometric topologies result in substantial manifold mismatch, as the traditional dependence on linear residuals in flat Euclidean space induces considerable geometric distortion when accommodating various topological patterns. We propose a novel framework termed Geometry-aware Riemannian Anomaly DEtector (GRADE) to tackle these difficulties. To address the semantic gap, we develop a domain-based contrastive learning technique combined with target-adaptive prompt tuning, which dynamically adjusts source representations to align with the target distribution while maintaining local contexts. To tackle the manifold mismatch, we present the Riemannian Residual Displacement (RRD), a curvature-adaptive metric that maps node representations onto multi-curvature Riemannian manifolds. By quantifying geodesic deviations in curved spaces, RRD corrects geometric distortions and identifies anomaly patterns that remain invariant to topological alterations. Additionally, a geometry-aware prototype alignment approach secures these residuals to data-independent priors for resilient inference. Comprehensive experiments on 11 benchmark datasets across 4 distinct domains demonstrate that GRADE significantly outperforms state-of-the-art baselines in both zero-shot and few-shot settings.

1 Introduction

Graph Anomaly Detection (GAD) is essential for safeguarding critical applications, including financial fraud detection and social network cleansing [Pang *et al.*, 2021; Bian *et al.*, 2020; Chen and Tsourakakis, 2022; Xu *et al.*, 2021]. Although supervised GAD approaches excel with labeled data, they struggle in practical applications where obtaining labels for each new target domain is excessively costly. This constraint has generated interest in Cross-Domain and Generalist models, which seek to transfer knowledge from label-dense source domains to target domains with scarce labels.

Notwithstanding recent advancements [Ding *et al.*, 2021; Wang *et al.*, 2023; Chen *et al.*, 2024a], current methodologies function within Cross-Narrow-Domain settings (e.g., Amazon $\rightarrow$ YelpNYC, both situated within the Co-review domain), capitalizing on inherent topological homogeneity. However, effective implementation demands resilience in Cross-Broad-Domain (CBD) scenarios, where the source and target exhibit significant divergence (e.g., Social $\rightarrow$ Financial, across distinct domains). Such significant distribution shifts pose two challenges: 1) Feature Heterogeneity: distinct semantic spaces that require proper alignment; 2) Structural Disparity: essentially different topological patterns.

Such structural discrepancy induces a geometric constraint referred to as manifold mismatch. This arises from the divergent topological characteristics of the data, where hierarchical structures are naturally suited to Hyperbolic geometry while cycles are best represented in Spherical space. Current approaches, embedding heterogeneous structures into flat Euclidean space, often ignore these curvatures and misalign the geometric metric with the underlying topology. Furthermore, most recent generalist frameworks [Liu *et al.*, 2024; Qiao *et al.*, 2025] rely on Euclidean embeddings. However, rigidly embedding heterogeneous structures into zero-curvature space induces severe geometric distortion. As evidenced in Figure 1, operating in Euclidean space forces residual distributions into a highly concentrated peak, rendering the divergence between normal and anomalous nodes negligible. In contrast, mapping representations onto Riemannian manifolds even with fixed curvatures can unfold these compressed distributions. While generalist models attempt to bridge the gap via massive corpora or prompt tuning, they fail to rectify this intrinsic geometric distortion.

To transcend these geometric limitations, we draw inspiration from geometric graph learning [Dong *et al.*, 2025; Grover *et al.*, 2025] and propose a novel metric building upon traditional residuals in GAD: **Riemannian Residual Displacement (RRD)**. Distinct from Euclidean residuals that suffer from manifold mismatch, RRD is designed to rectify geometric distortion. By dynamically projecting representa-

*Corresponding Author.

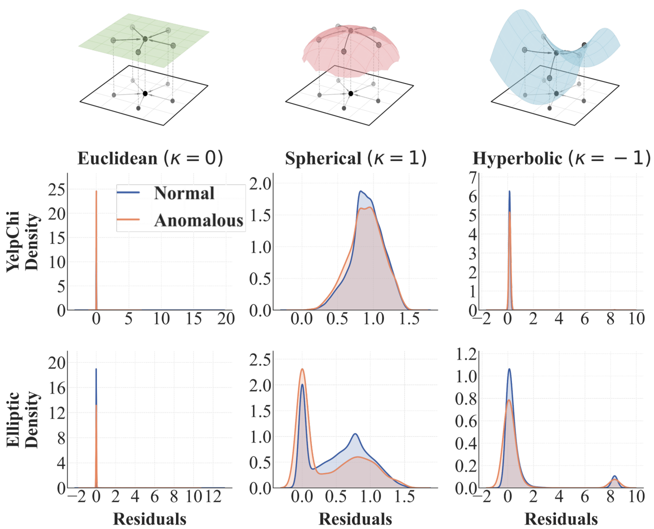

Figure 1: Comparison of residual distributions of YelpChi and Elliptic datasets in Euclidean space and fixed Riemannian manifolds.

tions into Euclidean, Spherical and Hyperbolic spaces with adaptive curvatures, RRD measures deviations on the intrinsic data manifold. This allows for the extraction of geometric anomaly patterns that remain invariant across broad domains, regardless of the underlying topology.

Tailored for the Cross-Broad-Domain GAD, we propose **GRADE** (**G**eometry-aware **R**iemannian **A**nomaly **DE**tector). GRADE addresses the dual challenge via a unified pipeline: it first mitigates feature heterogeneity using domain-based contrastive learning and target-adaptive prompt tuning. It then tackles manifold mismatch by mapping representations into Riemannian manifolds to compute RRD, optimizing an alignment loss against data-independent prototypes. This establishes a structure-agnostic anomaly standard.

The main contributions are summarized as follows:

- We formally define the Cross-Broad-Domain GAD paradigm and identify the dual challenge of feature and structural shifts. To address this, we propose GRADE, a unified pre-training and inference framework adapted for both zero-shot and few-shot settings.
- We propose Riemannian Residual Displacement (RRD) as a novel CD-GAD metric. By addressing the manifold mismatch and rectifying geometric distortion, RRD restores intrinsic structural deviations, establishing a universal standard for anomaly discrimination.
- Extensive experiments on 11 benchmark datasets across 4 distinct domains demonstrate that GRADE significantly outperforms state-of-the-art approaches.

2 Preliminaries

2.1 κ-Stereographic Model

To encompass the diverse geometric topologies present in graph data, we broaden the framework from Euclidean space to Riemannian manifolds with constant curvatures. In light of recent progress in geometric GAD [Dong *et al.*, 2025; Grover *et al.*, 2025], we employ the κ-stereographic model for integrating various geometries. Let $\mathcal{M}_\kappa^d$ be a d-dimensional Riemannian manifold characterized by constant sectional curvature κ. The manifold is characterized as:

$$\mathcal{M}_\kappa^d = \left\{ \mathbf{x} \in \mathbb{R}^d \mid -\kappa \|\mathbf{x}\|_2^2 < 1 \right\}, \tag{1}$$

where κ indicates the geometry: $\kappa < 0$ corresponds to Hyperbolic space, $\kappa > 0$ to Spherical space, and $\kappa = 0$ returns to Euclidean space. According to the extension proposed by Bachmann [Bachmann *et al.*, 2020], the Möbius addition for $\boldsymbol{x}, \boldsymbol{y} \in M_\kappa^d$ is defined as follows:

$$\mathbf{x} \oplus_\kappa \mathbf{y} = \frac{\left(1 - 2\kappa\langle \mathbf{x}, \mathbf{y}\rangle - \kappa\|\mathbf{y}\|_2^2\right)\mathbf{x} + \left(1 + \kappa\|\mathbf{x}\|_2^2\right)\mathbf{y}}{1 - 2\kappa\langle \mathbf{x}, \mathbf{y}\rangle + \kappa^2\|\mathbf{x}\|_2^2\|\mathbf{y}\|_2^2}. \tag{2}$$

2.2 Tangent Space and Mappings

In contrast to Euclidean space, Riemannian manifolds exhibit curvature and do not accommodate conventional vector operations. Algebraic operations are conducted using the tangent space $T_{\mathbf{x}}\mathcal{M}_\kappa^d$ at a point $\mathbf{x}$, which acts as a local Euclidean approximation of the manifold. The Exponential map $\exp_{\mathbf{x}}^\kappa$ and the Logarithmic map $\log_{\mathbf{x}}^\kappa$ are two crucial bijective mappings that enable the connection between the manifold and its tangent space. $\exp_{\mathbf{x}}^\kappa$ transfers a tangent vector $\mathbf{v} \in T_{\mathbf{x}}\mathcal{M}_\kappa^d$ to a point on the manifold, therefore progressing along the geodesic originating from $\mathbf{x}$ in the direction of $\mathbf{v}$. Conversely, $\log_{\mathbf{x}}^\kappa$ translates a point $\mathbf{y} \in \mathcal{M}_\kappa^d$ back to the tangent space at $\mathbf{x}$, retrieving the vector that signifies the direction and distance from point $\mathbf{x}$ to $\mathbf{y}$.

$$\exp_{\mathbf{x}}^\kappa(\mathbf{v}) = \mathbf{x} \oplus_\kappa \left(\tan_\kappa\left(\frac{\sqrt{|\kappa|}\|\mathbf{v}\|_2}{2}\right) \frac{\mathbf{v}}{\sqrt{|\kappa|}\|\mathbf{v}\|_2} \right), \tag{3}$$

$$\begin{aligned}\log_{\mathbf{x}}^\kappa(\mathbf{y}) = {}& \frac{2}{\sqrt{|\kappa|}} \tan_\kappa^{-1}\left(\sqrt{|\kappa|}\,\|-\mathbf{x} \oplus_\kappa \mathbf{y}\|_2\right) \\ & \times \frac{-\mathbf{x} \oplus_\kappa \mathbf{y}}{\|-\mathbf{x} \oplus_\kappa \mathbf{y}\|_2},\end{aligned} \tag{4}$$

where $\tan_\kappa(\cdot)$ modifies tanh for $\kappa \neq 0$, and $\oplus_\kappa$ indicates the Möbius addition. For any point $\mathbf{x} \in \mathcal{M}_\kappa^d$, the metric tensor $g_{\mathbf{x}}^\kappa = (\lambda_{\mathbf{x}}^\kappa)^2 \mathbf{I}$ provides a conformal factor $\lambda_{\mathbf{x}}^\kappa = \frac{2}{1+\kappa\|\mathbf{x}\|_2^2}$. The distance between two nodes $\mathbf{x}$ and $\mathbf{y}$ on the manifold can simply be calculated using the magnitude of the $\log_{\mathbf{x}}^\kappa$ map:

$$d_{\mathcal{M}_\kappa}(\mathbf{x}, \mathbf{y}) = \frac{2}{\sqrt{|\kappa|}} \tan_\kappa^{-1}\left(\sqrt{|\kappa|}\,\|-\mathbf{x} \oplus_\kappa \mathbf{y}\|_2\right). \tag{5}$$

3 Methodology

3.1 Approach Overview

The proposed GRADE framework adheres to a two-stage pre-training and inference paradigm (Figure 2) to distill domain-invariant geometric patterns. In the pre-training phase, the framework integrates Domain-based Contrastive Learning with Target-Adaptive Prompt Tuning. This combination aligns cross-domain feature distributions while preserving fine-grained local contexts. To address the limitations of Euclidean metrics in modeling complex topologies,

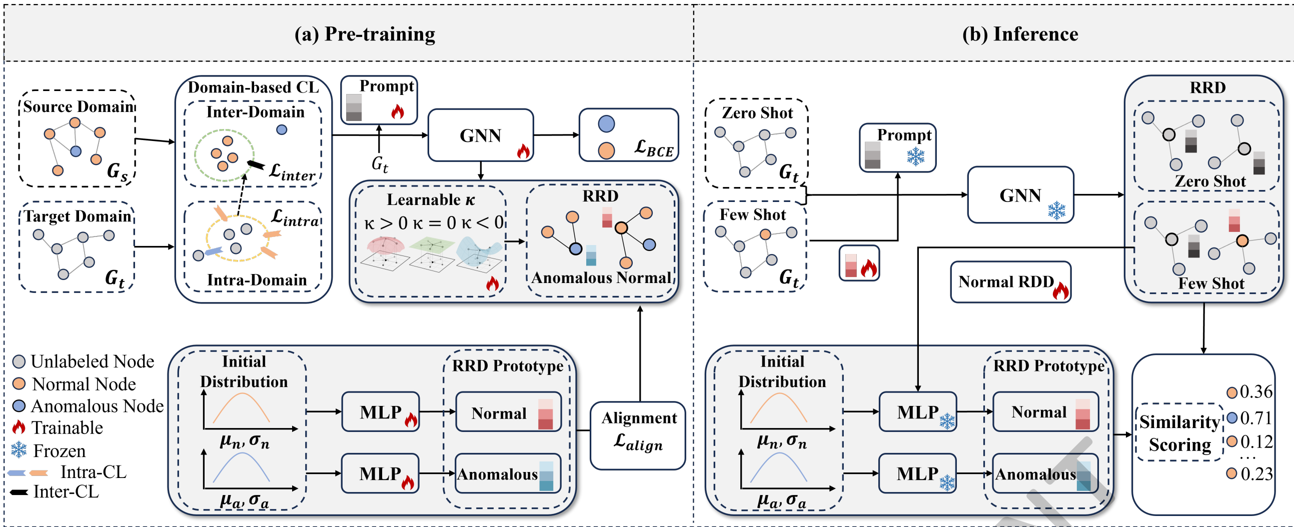

Figure 2: The overall architecture of the proposed framework. (a) The Pre-training phase aligns source and target semantic spaces via contrastive learning and prompt tuning, followed by projecting node representations into Riemannian manifolds to compute RRD for prototype alignment. (b) The Inference phase deploys the pre-trained geometric metric to the target domain, adapting to zero-shot or few-shot settings by measuring deviations from the learned prototypes.

these aligned representations are projected onto a family of multi-curvature Riemannian manifolds. Within these curved spaces, we compute the RRD, a geometry-aware metric quantifying the geodesic deviation of each node from its neighborhood. To anchor these residuals, the network is optimized via a Prototype Alignment mechanism, forcing geometric residuals to cluster around data-agnostic normal and abnormal prototypes. In the inference phase, GRADE utilizes these learned patterns to detect anomalies in the target domain under both Zero-Shot and Few-Shot settings.

3.2 Domain-Based Contrastive Learning

Intra-Domain Contrastive Learning

Since the target domain lacks supervision, we employ a self-supervised objective to enhance the representation quality of normal nodes (the majority class). We maximize the mutual information between local node patches and global graph summaries within the target domain. Let $\mathbf{H}^t = \{\mathbf{h}_1^t, \dots, \mathbf{h}_{N_t}^t\}$ denote the target node embeddings. A global graph summary $\mathbf{s}^t$ is derived via a readout function $\mathcal{R}(\cdot)$:

$$\mathbf{s}^t = \mathcal{R}(\mathbf{H}^t) = \frac{1}{N_t}\sum_{i=1}^{N_t}\mathbf{h}_i^t. \tag{6}$$

Negative samples $\tilde{\mathbf{H}}^t$ are generated by a corruption function $\mathcal{C}(\cdot)$ that performs row-wise feature shuffling. A discriminator $\mathcal{D}_{intra}$ distinguishes between positive pairs $(\mathbf{h}_i^t, \mathbf{s}^t)$ and negative pairs $(\tilde{\mathbf{h}}_j^t, \mathbf{s}^t)$. The objective is formulated as:

$$\begin{aligned}\mathcal{L}_{intra} &= \mathbb{E}_{(\mathbf{h}_i^t,\mathbf{s}^t)}[\log\sigma(\mathcal{D}_{intra}(\mathbf{h}_i^t,\mathbf{s}^t))] \\ &+ \mathbb{E}_{(\tilde{\mathbf{h}}_j^t,\mathbf{s}^t)}[\log(1-\sigma(\mathcal{D}_{intra}(\tilde{\mathbf{h}}_j^t,\mathbf{s}^t)))],\end{aligned} \tag{7}$$

where $\sigma(\cdot)$ is the sigmoid function. This encourages GRADE to capture intrinsic normality patterns.

Inter-Domain Contrastive Learning

We explicitly correlate the domain representations to transfer anomaly detection knowledge. We consider the aligned global summaries of the target $\mathbf{s}^t$ and the source $\mathbf{s}^s$ as positive examples. We employ a corrupted source summary $\tilde{\mathbf{s}}^s$ for negative samples. An inter-domain discriminator $\mathcal{D}_{inter}$ evaluates the coherence between the views. The alignment loss is defined as follows:

$$\begin{aligned}\mathcal{L}_{inter} &= \mathbb{E}_{(\mathbf{s}^t,\mathbf{s}^s)}[\log\sigma(\mathcal{D}_{inter}(\mathbf{s}^t,\mathbf{s}^s))] \\ &+ \mathbb{E}_{(\mathbf{s}^t,\tilde{\mathbf{s}}^s)}[\log(1-\sigma(\mathcal{D}_{inter}(\mathbf{s}^t,\tilde{\mathbf{s}}^s)))].\end{aligned} \tag{8}$$

The module projects the target graph onto a shared latent space that corresponds with the source by optimizing $\mathcal{L}_{inter}$. The overall contrastive loss is a weighted summation:

$$\mathcal{L}_{contra} = \omega_{intra}\mathcal{L}_{intra} + \omega_{inter}\mathcal{L}_{inter}. \tag{9}$$

3.3 Target-Adaptive Prompt Tuning

Although contrastive learning aligns distributions on a global scale, it may neglect domain-specific attributes. To address this, we implement a prompt learning approach derived from natural language processing. We present a learnable prompt matrix $\mathbf{P} \in \mathbb{R}^{M\times d}$ tailored for the target domain, where M represents the number of prompt tokens and d indicates the dimension. The objective of $\mathbf{P}$ is to modify target representations in accordance with the source distribution while preserving the integrity of the pre-trained GNN backbone. We utilize a query-based attention method to dynamically extract pertinent prompts solely for target nodes. For a target node embedding $\mathbf{h}_i^{t,l}$ at layer l, the attention score $\alpha_{i,j}^l$ for the j-th prompt token is calculated as:

$$\alpha_{i,j}^l = \frac{\exp\left(\mathbf{h}_i^{t,l}\cdot\mathbf{P}_j^\top\right)}{\sum_{k=1}^{M}\exp\left(\mathbf{h}_i^{t,l}\cdot\mathbf{P}_k^\top\right)}. \tag{10}$$

This score reflects the relevance of the prompt to the node's local context. The prompt-enhanced feature is updated via:

$$\mathbf{z}_i^l = \mathbf{h}_i^{t,l} + \sum_{j=1}^{M} \alpha_{i,j}^l \mathbf{P}_j. \tag{11}$$

This mechanism ensures $\mathbf{z}_i^l$ incorporates target-specific knowledge via $\mathbf{P}$ while the frozen GNN parameters preserve the source-derived anomaly detection logic.

3.4 Riemannian Residual Displacement

To accommodate various topologies, we develop a projection method over three separate geometries: Euclidean ($\mathbb{E}$), Hyperbolic ($\mathbb{H}$), and Spherical ($\mathbb{S}$). We consider the curvatures $\mathcal{K} = \{\kappa_{\mathbb{H}}, \kappa_{\mathbb{S}}\}$ as trainable parameters optimized through backpropagation. Tight constraints are established (e.g., $\kappa_{\mathbb{H}} < -\epsilon, \kappa_{\mathbb{S}} > \epsilon$) to guarantee geometric distinctiveness. This enables the model to adaptively map characteristics into the appropriate curvature space for subsequent calculations.

The essence of CD-GAD is to measure the divergence of a node from anticipated patterns. Previous methods predominantly assume a flat Euclidean geometry, modeling discrepancy as the linear residual $\mathbf{r}_i^0 = \mathbf{h}_i - \boldsymbol{\mu}_i$. However, this linear subtraction becomes geometrically ill-posed when applied to graphs with complex topologies embedded in curved manifolds. The vector difference in Euclidean space signifies a chord instead of a geodesic path, leading to structural distortion.

To rectify this, we propose RRD, a geometry-aware metric defined on the tangent space of a Riemannian manifold. RRD measures the geodesic displacement from the neighborhood centroid $\boldsymbol{\mu}_i^\kappa$ to the node $\mathbf{h}_i$. Utilizing the κ-stereographic model, the closed-form solution for RRD $\mathbf{r}_i^\kappa \in T_{\boldsymbol{\mu}_i^\kappa}\mathcal{M}_\kappa$ is expressed as:

$$\mathbf{r}_i^\kappa = \frac{2\tan_\kappa^{-1}\left(\sqrt{|\kappa|}\|\mathbf{u}_i\|_2\right)}{\sqrt{|\kappa|}\,\lambda_{\boldsymbol{\mu}_i^\kappa}^\kappa\,\|\mathbf{u}_i\|_2} \cdot \mathbf{u}_i, \tag{12}$$

where $\mathbf{u}_i = -\boldsymbol{\mu}_i^\kappa \oplus_\kappa \mathbf{h}_i$ is the Möbius addition difference, and $\lambda_{\boldsymbol{\mu}_i^\kappa}^\kappa = 2/(1+\kappa\|\boldsymbol{\mu}_i^\kappa\|_2^2)$ is the conformal factor. This provides a rigorous geometric quantification of structural anomalies.

3.5 Prototype Alignment

To fully utilize the anomalous discriminability inherent in our proposed RRD, we develop a prototype alignment mechanism. This module associates the acquired RRD with data-independent latent variables, enabling strong transfer generalization. Rather than optimizing static vectors susceptible to overfitting source-specific noise, we suggest instantiating geometric prototypes from a stochastic prior distribution. We propose that the optimal geometric deviations for normal and abnormal instances adhere to separate latent Gaussian distributions. Let $\boldsymbol{\xi}_n$ and $\boldsymbol{\xi}_a$ represent the latent variables drawn from $\mathcal{N}(\mu_n, \sigma_n)$ and $\mathcal{N}(\mu_a, \sigma_a)$, respectively. We employ two independent multilayer perceptron networks to translate the stochastic prior in each scene into the tangent space of the geometric manifold.

$$\mathbf{c}_n^\kappa = \Psi_n^\kappa(\boldsymbol{\xi}_n), \quad \mathbf{c}_a^\kappa = \Psi_a^\kappa(\boldsymbol{\xi}_a). \tag{13}$$

The derived geometric prototypes for the normal and anomalous classes within the curvature space κ are indicated by the symbols $\mathbf{c}_n^\kappa$ and $\mathbf{c}_a^\kappa$. For the subsequent latent alignment, these centroids act as exacting geometric anchors.

Maintaining geometric compactness in the Riemannian tangent space is the main goal. All geometric spaces are covered by the alignment loss $\mathcal{L}_{align}$, which minimizes the geodesic distance between node residuals and the associated generated class prototypes:

$$\mathcal{L}_{align} = \sum_{\kappa\in\mathcal{K}}\sum_{v_i\in\mathcal{V}} \frac{1}{2}\left[(1-y_i)\|\mathbf{r}_i^\kappa - \mathbf{c}_n^\kappa\|_2 + y_i\|\mathbf{r}_i^\kappa - \mathbf{c}_a^\kappa\|_2\right]. \tag{14}$$

Complementary to the geometric alignment, we require the model to learn semantically discriminative representations of the source domain directly from the graph topology before projection. We impose a standard binary classification supervision $\mathcal{L}_{BCE}$ on the output of the GNN backbone:

$$\mathcal{L}_{BCE} = -\frac{1}{|\mathcal{V}|}\sum_{v_i\in\mathcal{V}}\left(y_i\log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\right). \tag{15}$$

Finally, we jointly optimize the prototype alignment, semantic classification, and the contrastive losses defined previously. The total objective function is formulated as:

$$\mathcal{L} = \omega_{bce}\mathcal{L}_{BCE} + \omega_{align}\mathcal{L}_{align} + \mathcal{L}_{contra}, \tag{16}$$

where ω_{bce} and ω_{align} are hyperparameters balancing the contribution of semantic and geometric objectives.

3.6 Model Deployment

Zero-Shot Inference

In the completely unsupervised scenario, we presume the target domain without labels. Utilizing the domain-invariant characteristics of the learnt RRDs, we freeze all model parameters and directly implement the pre-trained model on the target graph. For a given node v_i, we compute its RRD vector $\mathbf{r}_i^\kappa$ within each curvature space. The anomaly score is calculated by examining the aggregate geometric departure from the established normal and abnormal prototypes:

$$S_{zero}(v_i) = \frac{1}{|\mathcal{K}|}\sum_{\kappa\in\mathcal{K}}\left(\|\mathbf{r}_i^\kappa - \mathbf{c}_n^\kappa\|_2 - \|\mathbf{r}_i^\kappa - \mathbf{c}_a^\kappa\|_2\right). \tag{17}$$

Few-Shot Inference

In settings when a limited collection of identified normal nodes $\mathcal{V}_f^t$ is accessible, we execute parameter-efficient tuning to adjust the model. In contrast to conventional fine-tuning that modifies the entire network, we freeze the high-dimensional GNN backbone to avert catastrophic forgetting. GRADE only refines two lightweight elements: the target-specific context prompts $\mathbf{P}$ and the latent normal prototype prior $\boldsymbol{\xi}_n$. The tuning target reduces the discrepancy of support nodes from the standard prototype:

$$\min_{\mathbf{P},\boldsymbol{\xi}_n}\mathcal{L}_{tuning} = \sum_{v_i\in\mathcal{V}_f^t}\sum_{\kappa\in\mathcal{K}}\|\mathbf{r}_i^\kappa(\mathbf{P}) - \Psi_n^\kappa(\boldsymbol{\xi}_n)\|_2^2. \tag{18}$$

Metric	Method	Reddit	Amz-all	Amazon	Yelp-all	Tolokers	Question	Elliptic	Disney	T-Finance	T-Social
AUROC	CoLA	0.4627	0.4095	0.4578	0.4883	0.4509	0.4951	0.5568	0.4689	0.4893	OOM
	TAM	0.5721	0.7538	0.4724	0.4222	0.5348	0.5113	0.3289	0.4768	0.2985	OOM
	GADAM	0.4532	0.5959	0.6646	0.4829	0.4832	<u>0.5594</u>	0.3922	0.4288	0.1382	0.5218
	BWGNN	0.5374	0.5036	0.5146	0.5739	<u>0.5766</u>	0.5194	0.6729	0.5395	0.6694	0.4924
	GHRN	0.5498	0.7617	0.5079	0.5645	<u>0.5584</u>	0.4823	<u>0.6986</u>	0.5932	0.6663	0.4934
	PMP	0.5651	0.6054	0.6261	0.5325	0.4474	0.4474	**0.8001**	0.5000	0.3720	0.5000
	ACT	0.4319	0.5454	0.5305	0.5741	0.4266	0.4815	0.3135	0.4901	0.5570	OOM
	UNPrompt	0.5483	0.8052	0.7667	0.5784	0.4541	0.4596	0.5901	0.6511	0.2073	0.3957
	AnomalyGFM	<u>0.5909</u>	<u>0.9094</u>	<u>0.8486</u>	<u>0.5855</u>	0.5269	0.5395	0.6259	<u>0.6992</u>	<u>0.7889</u>	<u>0.5452</u>
	Ours	**0.6293**	**0.9388**	**0.8827**	**0.6004**	**0.6141**	**0.5851**	0.6648	**0.7952**	**0.8055**	**0.6890**
AUPRC	CoLA	0.0388	0.0865	0.0664	0.1462	0.0842	0.0298	0.0991	0.0708	0.0437	OOM
	TAM	0.0418	0.1743	0.0681	0.1240	0.0975	0.0312	0.0692	0.0633	0.0337	OOM
	GADAM	0.0293	0.1595	0.4179	0.1371	0.1001	<u>0.0395</u>	0.0733	0.0651	0.0261	0.0293
	BWGNN	0.0426	0.0712	0.0707	<u>0.1869</u>	<u>0.2766</u>	0.0309	0.1355	0.0531	0.0667	0.0298
	GHRN	<u>0.0428</u>	0.1487	0.0663	<u>0.1723</u>	<u>0.2564</u>	0.0273	<u>0.1496</u>	0.0617	0.0632	0.0303
	PMP	<u>0.0385</u>	0.1108	0.1416	0.1599	0.0325	0.0325	**0.2217**	0.0484	0.0373	0.0301
	ACT	0.0267	0.0272	0.0758	0.1685	0.1776	0.0307	0.0654	0.0737	0.0464	OOM
	UNPrompt	0.0376	0.2599	0.1651	0.0705	0.2049	0.0297	0.1278	0.1197	0.0267	0.0232
	AnomalyGFM	0.0386	<u>0.6997</u>	<u>0.5874</u>	0.1829	0.2466	0.0353	0.0747	<u>0.1288</u>	<u>0.1082</u>	<u>0.0325</u>
	Ours	**0.0603**	**0.7743**	**0.6578**	**0.1988**	**0.2771**	**0.0421**	0.1495	**0.3027**	**0.2892**	**0.0435**

Table 1: Performance comparison on zero-shot CBD-GAD with Facebook as the source domain. OOM indicates that the model is Out Of Memory. The best results are highlighted in **bold** and the second best results are <u>underlined</u>.

4 Experiments

4.1 Experimental Setup

Datasets

We conduct experiments on 11 benchmark datasets across 4 distinct domains: Social Networks, Finance, Co-review, and Collaborative Networks. The Social Networks include Facebook [Xu *et al.*, 2022], Reddit [Kumar *et al.*, 2019], Question [Platonov *et al.*, 2023], and T-Social [Tang *et al.*, 2022]. The Finance domain comprises T-Finance [Tang *et al.*, 2022] and Elliptic [Weber *et al.*, 2019]. The Co-review domain consists of Amazon, Amazon-all (Amz-all), and YelpChi-all (Yelp-all) [Dou *et al.*, 2020]. Finally, the collaborative networks include Disney [Sánchez *et al.*, 2013] and Tolokers [Platonov *et al.*, 2023]. The detailed statistics of datasets are summarized in Appendix C. We select Facebook as the source domain for zero-shot and few-shot settings.

Competing Methods

We evaluate GRADE against state-of-the-art approaches, categorized as follows: (1) **Unsupervised methods:** CoLA [Liu *et al.*, 2021], TAM [Qiao and Pang, 2023], and GADAM [Chen *et al.*, 2024b]. (2) **Supervised methods:** BWGNN [Tang *et al.*, 2022], GHRN [Gao *et al.*, 2023], and PMP [Zhuo *et al.*, 2024]. (3) **Cross-Domain and Generalist methods:** ACT [Wang *et al.*, 2023], UNPrompt [Niu *et al.*, 2025], and AnomalyGFM [Qiao *et al.*, 2025]. For the few-shot setting, we compare against generalist methods: ARC [Liu *et al.*, 2024], and AnomalyGFM.

4.2 Cross-Broad-Domain Effectiveness

Zero-Shot Performance Analysis

Table 1 presents the zero-shot inference results where the model is pre-trained on the source domain and transferred to diverse target datasets. GRADE establishes a new performance frontier, consistently outperforming state-of-the-art baselines. Remarkably, our model outperforms the competitive AnomalyGFM baseline by significant margins, achieving AUROC improvements of **16.55%** and **13.73%** on the Tolokers and Disney datasets, respectively. This result is particularly insightful given the Cross-Broad-Domain setting, where the source domain (Social Networks) and target domains (Collaborative Networks) diverge macroscopically in both disjoint feature spaces and heterogeneous geometric topologies. Existing Euclidean-based methods struggle to adapt to such structural disparity, precipitating the manifold mismatch problem. In contrast, GRADE effectively bridges this gap by projecting representations into multi-curvature Riemannian manifolds, thereby recovering intrinsic anomaly patterns that remain invariant across broad domains.

PMP exhibits improved performance on the Elliptic dataset, potentially stemming from partitioned message transmission inadvertently identifying anomalous patterns that resemble those in both the source and target domains. However, this performance exhibits insufficient robustness to perturbations in the source domain. GRADE consistently outperforms all baselines when the source domain is changed to Amazon-all, as demonstrated empirically in Appendix G. This validates that GRADE achieves its superior universality by acquiring geometry-aware invariant representations instead of succumbing to overfitting specific source-target correlations.

Few-Shot Performance Analysis

We further evaluate GRADE under low-resource settings (1/5/10-shot) against generalist baselines, as summarized in Table 2. GRADE exhibits exceptional sample efficiency, consistently surpassing techniques even in severely constrained 1-shot situations. The swift adaptability arises from the geometric compactness of the acquired Riemannian manifold. In contrast to Euclidean-based approaches that necessitate large samples to accurately define decision boundaries in distorted spaces, our RRD metric guarantees geometric distinction be-

Metric	Setting	Method	Reddit	Amz-all	Amazon	Yelp-all	Tolokers	Question	Elliptic	Disney
AUROC	1-shot	ARC	0.5904	0.6244	0.5729	0.5083	0.5231	0.5705	0.2663	0.4561
		AnomalyGFM	0.5922	0.8972	0.8531	0.5872	0.5200	0.5303	0.6199	0.6649
		Ours	**0.6300**	**0.9351**	**0.9130**	**0.5881**	**0.6040**	**0.5747**	**0.6638**	**0.7613**
	5-shot	ARC	0.5956	0.6602	0.6229	0.4981	0.5671	0.5757	0.2729	0.4569
		AnomalyGFM	0.6023	0.9011	0.8600	0.5951	0.5353	0.5426	0.6119	0.6613
		Ours	**0.6302**	**0.9350**	**0.8865**	**0.5970**	**0.6034**	**0.5807**	**0.6664**	**0.7627**
	10-shot	ARC	0.5988	0.6655	0.6442	0.4953	0.5488	**0.5901**	0.3419	0.4771
		AnomalyGFM	0.6252	0.9215	0.8649	**0.6064**	0.5356	0.5611	0.6303	0.6649
		Ours	**0.6313**	**0.9346**	**0.8913**	0.5967	**0.6030**	0.5816	**0.6667**	**0.7642**
AUPRC	1-shot	ARC	0.0417	0.0926	0.0729	0.0497	0.2316	0.0383	0.0677	0.0521
		AnomalyGFM	0.0398	0.6921	0.5801	0.1852	0.2407	0.0332	0.1401	0.1223
		Ours	**0.0594**	**0.7882**	**0.7226**	**0.1985**	**0.2701**	**0.0408**	**0.1486**	**0.3387**
	5-shot	ARC	0.0429	0.1075	0.0973	0.0521	0.2489	0.0386	0.0683	0.0588
		AnomalyGFM	0.0401	0.6985	0.5831	0.1918	0.2482	0.0336	0.1437	0.1257
		Ours	**0.0575**	**0.7882**	**0.6679**	**0.2051**	**0.2699**	**0.0417**	**0.1506**	**0.3476**
	10-shot	ARC	0.0434	0.1107	0.1058	0.0526	0.2432	**0.0461**	0.0724	0.0681
		AnomalyGFM	0.0444	0.7124	0.5895	0.1990	0.2481	0.0346	**0.1570**	0.1399
		Ours	**0.0574**	**0.7872**	**0.6763**	**0.2043**	**0.2698**	0.0434	0.1508	**0.3473**

Table 2: Performance comparison on few-shot CBD-GAD. The best results are highlighted in **bold** and the second best results are underlined.

tween normal and aberrant patterns. The stability of the geometric metric space against data scarcity is thus validated, and the model can successfully calibrate the prototype to the target distribution against data scarcity.

4.3 Scalability and Efficiency

To evaluate realistic implementation potential, we expand our assessment to extensive graph datasets. Recognizing resource restrictions in processing such large-scale graphs, we employ a subgraph-based inference technique following protocols provided in the recent framework [Qiao *et al.*, 2025]. By standardizing the batch size to 1024 and utilizing a neighbor sampling technique with sizes of 5 and 7 for T-Finance and T-Social correspondingly, we mitigate memory constraints while preserving local structural integrity. The comprehensive performance comparisons for detection on these extensive benchmarks are included in the Table 1. Figure 3 analyzes the computational efficiency. Although GRADE results in a modest rise in training expenses due to the manifold projections, it preserves a competitive standing in terms of inference latency and peak memory consumption. Since GRADE freezes most parameters after pre-training, the inference phase uses lightweight geometric metric fine-tuning instead of expensive dynamic parameter adjustments. This attribute makes GRADE exceptionally appropriate for real-time applications, providing an advantageous balance between advanced detection efficiency and computing resource utilization.

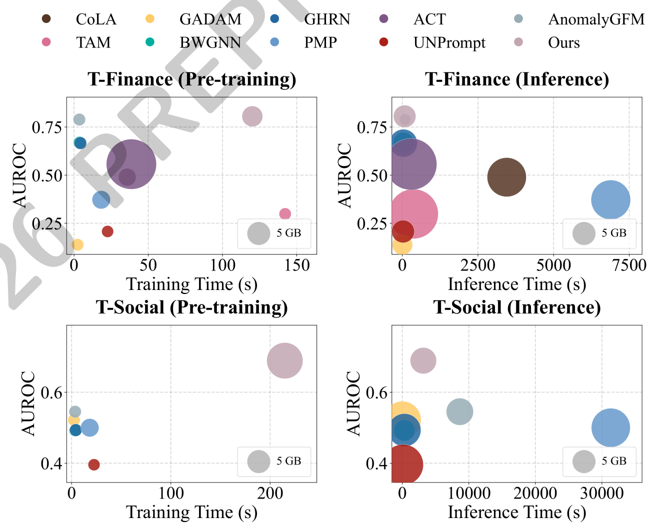

Figure 3: Comparison of running time, AUROC, and peak memory usage on T-Finance and T-Social datasets.

4.4 Ablation Study

We verify the contribution of each component by evaluating three variants: (1) w/o CL (without domain-based CL), (2) w/o Prompt (without prompt tuning), and (3) w/o RRD (Euclidean residuals). As shown in Table 3, GRADE consistently outperforms all variants. Specifically, the performance drop in w/o CL and w/o Prompt indicates that relying solely on geometric projection is insufficient without semantic alignment and target-aware adaptation. Most notably, the significant degradation in w/o RRD confirms that rigid Euclidean metrics fail to capture complex manifold patterns, distinguishing RRD as the core driver for resolving structural disparity. This is corroborated by Figure 4, where w/o CL exhibits distinct domain shifts, while GRADE effectively aligns heterogeneous feature spaces. Furthermore, Figure 5 illustrates that while Euclidean and Spherical spaces show overlaps on this specific dataset, the Hyperbolic space yields a clear separation boundary, validating the necessity of our multi-curvature design to adaptively capture diverse topological patterns.

4.5 Hyperparameter Sensitivity

We investigate robustness regarding two key hyperparameters: Prompt Tokens (M) and Loss Weights. As shown in Figure 6, performance improves as M increases from 1 to 5, indicating that sufficient prompt is essential to bridge se-

Metric	Method	Reddit	Amz-all	Amazon	Yelp-all	Tolokers	Question	Elliptic	Disney
AUROC	w/o CL	0.5849	0.8636	0.6078	<u>0.5996</u>	0.5492	0.5547	<u>0.6428</u>	<u>0.7867</u>
	w/o Prompt	<u>0.6279</u>	<u>0.9355</u>	<u>0.8647</u>	0.5707	<u>0.6091</u>	<u>0.5768</u>	0.6010	0.7641
	w/o RRD	0.5275	0.5567	0.4971	0.5060	0.5117	0.5119	0.5947	0.7599
	Ours	**0.6293**	**0.9388**	**0.8827**	**0.6004**	**0.6141**	**0.5851**	**0.6648**	**0.7952**
AUPRC	w/o CL	0.0523	0.4433	0.1097	<u>0.1982</u>	0.2335	0.0381	<u>0.1347</u>	<u>0.3004</u>
	w/o Prompt	<u>0.0597</u>	<u>0.7689</u>	<u>0.4782</u>	0.1813	<u>0.2657</u>	<u>0.0409</u>	0.1287	0.2834
	w/o RRD	0.0363	0.0754	0.0644	0.1525	0.2279	0.0342	0.1143	0.2639
	Ours	**0.0603**	**0.7743**	**0.6578**	**0.1988**	**0.2771**	**0.0421**	**0.1495**	**0.3027**

Table 3: Ablation study of different components. The best results are highlighted in **bold** and the second best results are <u>underlined</u>.

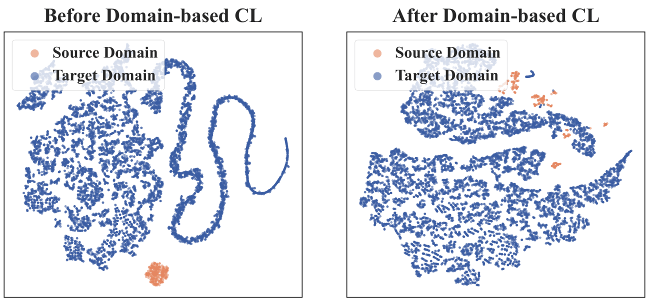

Figure 4: t-SNE visualization of node feature distributions in Facebook (source domian) and Amazon-all (target domain) datasets before and after Domain-based CL.

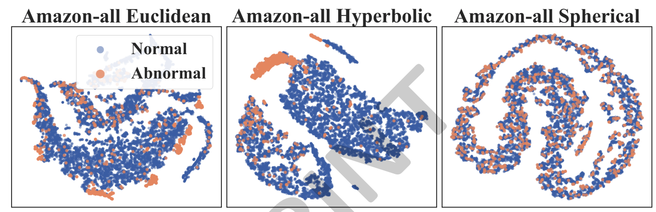

Figure 5: t-SNE visualization of RRD distributions across Euclidean, Hyperbolic, and Spherical spaces on Amazon-all.

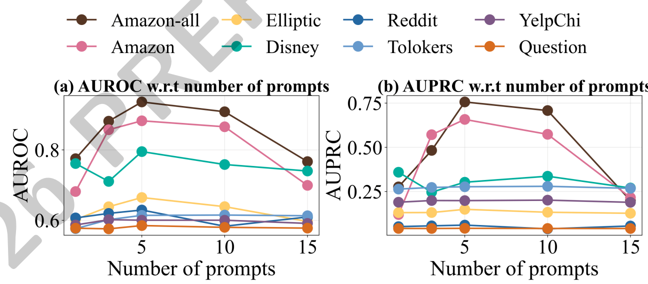

Figure 6: Impact of the number of prompt tokens on anomaly detection performance.

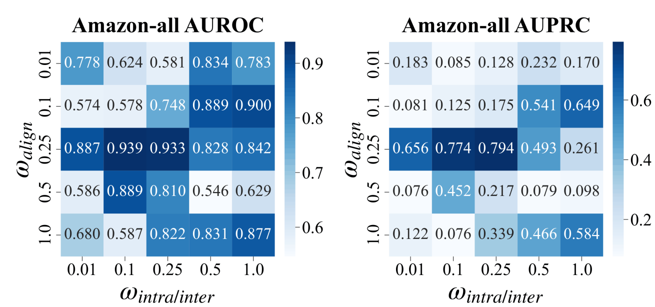

Figure 7: Hyperparameter sensitivity analysis of the balancing coefficients ω_{align} and $\omega_{intra/inter}$ on Amazon-all.

mantic gaps. However, larger M leads to diminishing returns, likely due to introducing adaptation noise that disrupts the pre-trained source knowledge. Regarding the loss coefficients, Figure 7 presents sensitivity heatmaps varying ω_{align} and $\omega_{intra/inter}$. We observe a robust joint region where $\omega_{align} \in [0.25, 0.5]$ and $\omega_{intra/inter} \in [0.1, 0.25]$. This suggests a critical synergy between tasks: moderate geometric regularization effectively prevents manifold collapse without dominating the semantic classification objective, whereas extreme weights either fail to enforce structure or over-constrain the representation space.

4.6 Discussion

Given the difficulty of determining the optimal curvature space for a target domain within the CBD-GAD framework, GRADE is pre-trained to perceive and compute geometric properties across three distinct spaces. Although this approach incurs substantial computational costs during pre-training, it enables GRADE to resolve the critical manifold mismatch challenge by aligning RRD prototypes. In addition, we employ an auxiliary subgraph-based method to efficiently manage time and memory consumption when processing large-scale datasets.

5 Conclusion

In this work, we formalized the Cross-Broad-Domain GAD paradigm and pinpointed the manifold mismatch as a fundamental challenge. To overcome this, we proposed GRADE, a geometry-aware framework that bridges macroscopic domain discrepancies. GRADE incorporates domain-based contrastive learning with prompt tuning to mitigate feature heterogeneity, while leveraging multi-curvature Riemannian manifolds to resolve structural disparity. At its core, the Riemannian Residual Displacement metric rectifies geometric distortion, enabling the recovery of intrinsic anomaly patterns that remain invariant across distinct domains. Extensive evaluations on 11 benchmark datasets confirm that GRADE establishes a new state-of-the-art in both zero-shot and few-shot settings. Future work will explore extending this geometric perspective to dynamic GAD and investigating the potential of learnable product manifolds.

Acknowledgments

This work was partly supported by the National Key Research and Development Program of China under Grant 2023YFB3002201, the National Natural Science Foundation of China under Grant 72342026, and Fundamental Research Funds for the Central Universities under Grant 2024-6-ZD-02.

References

[Bachmann *et al.*, 2020] Gregor Bachmann, Gary Bécigneul, and Octavian Ganea. Constant curvature graph convolutional networks. In *International conference on machine learning*, pages 486–496. PMLR, 2020.

[Bandyopadhyay *et al.*, 2020] 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*, pages 25–33, 2020.

[Bian *et al.*, 2020] Tian Bian, Xi Xiao, Tingyang Xu, Peilin Zhao, Wenbing Huang, Yu Rong, and Junzhou Huang. Rumor detection on social media with bi-directional graph convolutional networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 549–556, 2020.

[Chen and Tsourakakis, 2022] Tianyi Chen and Charalampos Tsourakakis. Antibenford subgraphs: Unsupervised anomaly detection in financial networks. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2762–2770, 2022.

[Chen *et al.*, 2024a] Jiazhen Chen, Sichao Fu, Zhibin Zhang, Zheng Ma, Mingbin Feng, Tony S Wirjanto, and Qinmu Peng. Towards cross-domain few-shot graph anomaly detection. In *2024 IEEE International Conference on Data Mining (ICDM)*, pages 51–60. IEEE, 2024.

[Chen *et al.*, 2024b] 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.

[Ding *et al.*, 2019] 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. SIAM, 2019.

[Ding *et al.*, 2021] Kaize Ding, Kai Shu, Xuan Shan, Jundong Li, and Huan Liu. Cross-domain graph anomaly detection. *IEEE Transactions on Neural Networks and Learning Systems*, 33(6):2406–2415, 2021.

[Dong *et al.*, 2025] Xiangyu Dong, Xingyi Zhang, Lei Chen, Mingxuan Yuan, and Sibo Wang. SpaceGNN: Multi-space graph neural network for node anomaly detection with extremely limited labels. *arXiv preprint arXiv:2502.03201*, 2025.

[Dou *et al.*, 2020] 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 & knowledge management*, pages 315–324, 2020.

[Duan *et al.*, 2025] Mingjiang Duan, Da He, Tongya Zheng, Lingxiang Jia, Mingli Song, Xinyu Wang, and Zunlei Feng. Global attribute-association pattern aggregation for graph fraud detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 11616–11624, 2025.

[Fan *et al.*, 2020] Haoyi Fan, Fengbin Zhang, and Zuoyong Li. AnomalyDAE: Dual autoencoder for anomaly detection on attributed networks. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5685–5689. IEEE, 2020.

[Gao *et al.*, 2023] Yuan Gao, Xiang Wang, Xiangnan He, Zhenguang Liu, Huamin Feng, and Yongdong Zhang. Addressing heterophily in graph anomaly detection: A perspective of graph spectrum. In *Proceedings of the ACM web conference 2023*, pages 1528–1538, 2023.

[Grover *et al.*, 2025] Karish Grover, Geoffrey J Gordon, and Christos Faloutsos. CurvGAD: Leveraging curvature for enhanced graph anomaly detection. *arXiv preprint arXiv:2502.08605*, 2025.

[Huang *et al.*, 2025] Changqin Huang, Xinxing Shi, Chengling Gao, Qintai Hu, Xiaodi Huang, Qionghao Huang, and Ali Anaissi. Cross-domain graph anomaly detection via graph transfer and graph decouple. *CAAI Transactions on Intelligence Technology*, 2025.

[Kumar *et al.*, 2019] Srijan Kumar, Xikun Zhang, and Jure Leskovec. Predicting dynamic embedding trajectory in temporal interaction networks. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1269–1278, 2019.

[Li *et al.*, 2017] Jundong Li, Harsh Dani, Xia Hu, and Huan Liu. Radar: Residual analysis for anomaly detection in attributed networks. In *IJCAI*, volume 17, pages 2152–2158, 2017.

[Liu *et al.*, 2021] 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, 2021.

[Liu *et al.*, 2023] Zemin Liu, Xingtong Yu, Yuan Fang, and Xinming Zhang. Graphprompt: Unifying pre-training and downstream tasks for graph neural networks. In *Proceedings of the ACM web conference 2023*, pages 417–428, 2023.

[Liu *et al.*, 2024] Yixin Liu, Shiyuan Li, Yu Zheng, Qingfeng Chen, Chengqi Zhang, and Shirui Pan. ARC: A generalist graph anomaly detector with in-context learning. *Advances in Neural Information Processing Systems*, 37:50772–50804, 2024.

[Ma *et al.*, 2024] Xiaoxiao Ma, Fanzhen Liu, Jia Wu, Jian Yang, Shan Xue, and Quan Z Sheng. Rethinking unsupervised graph anomaly detection with deep learning: Residuals and objectives. *IEEE Transactions on Knowledge and Data Engineering*, 2024.

[Niu *et al.*, 2025] Chaoxi Niu, Hezhe Qiao, Changlu Chen, Ling Chen, and Guansong Pang. Zero-shot generalist graph anomaly detection with unified neighborhood prompts. *arXiv preprint arXiv:2410.14886*, 2025.

[Pang *et al.*, 2021] Guansong Pang, Chunhua Shen, Longbing Cao, and Anton Van Den Hengel. Deep learning for anomaly detection: A review. *ACM computing surveys (CSUR)*, 54(2):1–38, 2021.

[Platonov *et al.*, 2023] Oleg Platonov, Denis Kuznedelev, Michael Diskin, Artem Babenko, and Liudmila Prokhorenkova. A critical look at the evaluation of gnns under heterophily: Are we really making progress? *arXiv preprint arXiv:2302.11640*, 2023.

[Qiao and Pang, 2023] Hezhe Qiao and Guansong Pang. Truncated affinity maximization: One-class homophily modeling for graph anomaly detection. *Advances in Neural Information Processing Systems*, 36:49490–49512, 2023.

[Qiao *et al.*, 2025] Hezhe Qiao, Chaoxi Niu, Ling Chen, and Guansong Pang. Anomalygfm: Graph foundation model for zero/few-shot anomaly detection. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pages 2326–2337, 2025.

[Sánchez *et al.*, 2013] Patricia Iglesias Sánchez, Emmanuel Müller, Fabian Laforet, Fabian Keller, and Klemens Böhm. Statistical selection of congruent subspaces for mining attributed graphs. In *2013 IEEE 13th international conference on data mining*, pages 647–656. IEEE, 2013.

[Song *et al.*, 2025] Yudan Song, Yuecen Wei, Yuhang Lu, Qingyun Sun, Minglai Shao, Li-e Wang, Chunming Hu, Xianxian Li, and Xingcheng Fu. Mitigating message imbalance in fraud detection with dual-view graph representation learning. *arXiv preprint arXiv:2507.06469*, 2025.

[Sun *et al.*, 2022] Mingchen Sun, Kaixiong Zhou, Xin He, Ying Wang, and Xin Wang. Gppt: Graph pre-training and prompt tuning to generalize graph neural networks. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1717–1727, 2022.

[Tang *et al.*, 2022] Jianheng Tang, Jiajin Li, Ziqi Gao, and Jia Li. Rethinking graph neural networks for anomaly detection. In *International conference on machine learning*, pages 21076–21089. PMLR, 2022.

[Tang *et al.*, 2023] Jianheng Tang, Fengrui Hua, Ziqi Gao, Peilin Zhao, and Jia Li. GADBench: Revisiting and benchmarking supervised graph anomaly detection. *Advances in Neural Information Processing Systems*, 36:29628–29653, 2023.

[Wang *et al.*, 2021] 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.

[Wang *et al.*, 2023] Qizhou Wang, Guansong Pang, Mahsa Salehi, Wray Buntine, and Christopher Leckie. Cross-domain graph anomaly detection via anomaly-aware contrastive alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 4676–4684, 2023.

[Weber *et al.*, 2019] Mark Weber, Giacomo Domeniconi, Jie Chen, Daniel Karl I Weidele, Claudio Bellei, Tom Robinson, and Charles E Leiserson. Anti-money laundering in bitcoin: Experimenting with graph convolutional networks for financial forensics. *arXiv preprint arXiv:1908.02591*, 2019.

[Wo *et al.*, 2025] Zengyi Wo, Minglai Shao, Shiyu Zhang, and Ruijie Wang. Local homophily-aware graph neural network with adaptive polynomial filters for scalable graph anomaly detection. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pages 3180–3191, 2025.

[Xu *et al.*, 2021] Bingbing Xu, Huawei Shen, Bingjie Sun, Rong An, Qi Cao, and Xueqi Cheng. Towards consumer loan fraud detection: Graph neural networks with role-constrained conditional random field. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 4537–4545, 2021.

[Xu *et al.*, 2022] 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*, pages 444–457. Springer, 2022.

[Xu *et al.*, 2025] 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.

[Yan *et al.*, 2025] Junyi Yan, Enguang Zuo, Ke Liang, Meng Liu, Miaomiao Li, Xinwang Liu, Xiaoyi Lv, and Kai Lu. Address anomalies at critical crossroads for graph anomaly detection. *IEEE Transactions on Knowledge and Data Engineering*, 2025.

[Zhang *et al.*, 2025] Xiong Zhang, Hong Peng, Zhenli He, Cheng Xie, Xin Jin, and Hua Jiang. Gctam: Global and contextual truncated affinity combined maximization model for unsupervised graph anomaly detection. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, pages 3642–3650, 2025.

[Zhuo *et al.*, 2024] Wei Zhuo, Zemin Liu, Bryan Hooi, Bingsheng He, Guang Tan, Rizal Fathony, and Jia Chen. Partitioning message passing for graph fraud detection. *arXiv preprint arXiv:2412.00020*, 2024.