Breaking the Evidence Trap: Trustworthy Graph Anomaly Detection via Multi-View Evidential Alignment

Songran Bai*†, Xiaolong Zheng*†, Daniel Dajun Zeng*†
* *State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences*
†*School of Artificial Intelligence, University of Chinese Academy of Sciences*
Beijing, China
{baisongran2020, xiaolong.zheng, dajun.zeng}@ia.ac.cn

***Abstract*—Unsupervised graph anomaly detection is fundamentally challenged by the difficulty of distinguishing inherent data noise from genuine out-of-distribution anomalies. Although evidential deep learning offers a computationally efficient framework for uncertainty quantification, its direct adaptation to graph anomaly detection exposes three unresolved pathologies. First, continuous attribute regression suffers from evidence contraction, where suppressed evidence counts induce vanishing gradients that degrade shared encoder representations. Second, topology reconstruction exhibits structural confidence bias, as extreme edge sparsity drives the model toward globally overconfident predictions for edge absence. Third, existing anomaly scoring relies on linear aggregation of reconstruction error and epistemic uncertainty, which neglects their nonlinear interaction and renders the score unreliable under moderate uncertainty. To address these limitations, we propose MVEA, a Multi-View Evidential Alignment framework that integrates multi-view evidential learning with robust anomaly evaluation. We introduce a gradient compensation mechanism via latent evidential clustering to mitigate the risk of regression collapse. For structure learning, we design a cost-sensitive evidential strategy with dual-branch decoding and contrastive ranking to rectify confidence bias. We further propose a neighborhood distribution alignment module to enforce local semantic consistency beyond individual node reconstruction. Finally, we devise an uncertainty-rectified scoring function with a U-shaped gating mechanism to suppress ambiguous noise while amplifying reliable anomaly signals. Extensive experiments on benchmark datasets demonstrate that MVEA achieves superior detection accuracy and robustness compared to state-of-the-art methods.**

***Index Terms*—Unsupervised Graph Anomaly Detection, Deep Evidential Learning, Uncertainty Quantification**

I. Introduction

Graph anomaly detection is pivotal for addressing security issues in graph-structured data [1]. Since anomalies typically lack supervisory signals, unsupervised paradigms based on reconstruction error dominate the field [2]. However, relying solely on reconstruction error is insufficient for complex scenarios, as it fails to distinguish between inherent data noise and genuine out-of-distribution anomalies [3]. Recently, evidential deep learning has emerged as a computationally efficient deterministic alternative to Bayesian methods for uncertainty quantification [4]. By framing reconstruction as a parameter estimation problem for higher-order conjugate prior distributions, evidential deep learning theoretically allows models to jointly output predictions and calibrated confidence estimates in a single forward pass.

However, our theoretical analysis and empirical observations reveal that naively integrating evidential deep learning into graph anomaly detection provokes three fundamental pathologies. First, continuous attribute reconstruction suffers from evidence contraction [5]. In Normal-Inverse-Gamma regression, optimization algorithms tend to suppress the evidence count toward zero when encountering high-error nodes to minimize penalties, inducing vanishing gradients that cause the regression head to stagnate and representation learning to degrade. Second, topology reconstruction exhibits structural confidence bias. Driven by extreme graph sparsity, the model is dominated by negative edge samples, resulting in globally biased high-confidence predictions for edge absence while failing to appropriately express uncertainty on sparse authentic edges. Third, anomaly scoring suffers from a reliability blind spot. Existing linear weighting strategies overlook that reconstruction error is a trustworthy anomaly indicator only when the model is in a decisive state, as moderate error combined with moderate uncertainty tends to misclassify ambiguous noise as genuine anomalies. A schematic illustration of the first two pathologies is provided in Figure 1.

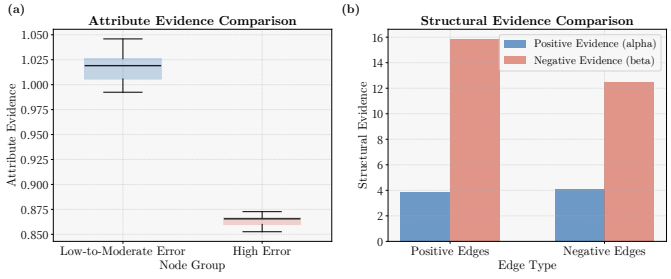

Fig. 1. Schematic illustration of attribute evidence contraction and structural confidence bias in evidential graph anomaly detection.

This work is supported by the Natural Science Foundation of China under Grant Nos. 72225011,72434005 and L242400108.
Corresponding author: Xiaolong Zheng (xiaolong.zheng@ia.ac.cn).

To address these limitations, we propose MVEA, a trustworthy multi-view evidential alignment framework with three targeted contributions. To mitigate evidence contraction, we introduce a gradient compensation mechanism via latent evidential clustering, which imposes a parallel clustering objective on the shared representation to sustain continuous gradient flow even when the regression branch saturates. For structural confidence bias, we propose a cost-sensitive evidential learning strategy that employs a dual-branch decoder to decouple evidence generation for edge existence and absence, coupled with a contrastive ranking loss to optimize the relative ordering of structural beliefs prior to probability fitting. We further incorporate a neighborhood distribution alignment module to constrain the statistical moments of reconstructed neighborhoods, enforcing local semantic consistency beyond individual node reconstruction. Finally, we design an uncertainty-rectified scoring mechanism based on a U-shaped certainty coefficient, which non-linearly modulates the reconstruction error to suppress its contribution under cognitive ambiguity and amplify it under high confidence, thereby robustly separating genuine anomalies from noise. Extensive experiments on five benchmark datasets validate the effectiveness of the proposed method. The source code is publicly available at https://github.com/currywhite30/MVEA-GAD.

II. Related Work

A. Unsupervised Graph Anomaly Detection

Unsupervised graph anomaly detection has become the predominant approach due to the scarcity of labeled anomalies [6]. The field is largely built upon graph autoencoders, exemplified by DOMINANT [7], which detects anomalies by jointly minimizing node attribute and graph structure reconstruction errors. Subsequent research has increasingly adopted contrastive learning frameworks to capture richer topological semantics. Methods such as CONAD [8] enforce representation invariance between original and artificially perturbed graphs, while MuSE [9] contrasts multi-scale topological views to identify anomalies at varying granularities. Recent innovations move beyond simple reconstruction, with GAD-NR [10] leveraging neighborhood reconstruction under homophily assumptions and G3AD [11] introducing guarding mechanisms to resist anomalous information propagation during message passing. Despite their effectiveness, these deterministic models remain susceptible to overconfident errors when handling complex anomalies, as they lack principled mechanisms to quantify predictive uncertainty.

B. Uncertainty Quantification on Graphs

Uncertainty quantification enhances the reliability of graph learning systems by estimating predictive confidence [12]. Traditional Bayesian methods such as variational inference and Monte Carlo dropout provide probabilistic guarantees but often incur substantial computational costs that limit scalability to large graphs [13], [14]. As a computationally efficient alternative, evidential deep learning directly models predictive distributions as higher-order conjugate priors, with methods such as GPN [15] employing Dirichlet distributions to identify out-of-distribution nodes in a single forward pass. Conformal prediction has also been adapted to graphs to deliver statistically rigorous coverage guarantees [16], [17]. While these techniques are well-established for node classification tasks, their principled extension to unsupervised anomaly detection remains an open and underexplored area.

C. Uncertainty-Aware Graph Anomaly Detection

Integrating uncertainty quantification into graph anomaly detection is an emerging direction aimed at improving interpretability and detection robustness [3], [18]. OODGAT [19] explicitly models out-of-distribution nodes during graph propagation to better separate known-class deviations from unknown anomalies. MITIGATE [20] employs multi-task active learning to quantify node informativeness and guide sample selection in low-label settings. Within purely unsupervised paradigms, recent frameworks model reconstruction likelihoods using evidential distributions, quantifying both structural and attribute uncertainty through higher-order evidence accumulation [3]. This evidential modeling reduces the sensitivity of traditional autoencoders to noise contamination and provides a principled confidence measure for anomaly scoring.

III. Preliminary

A. Problem Definition

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$ denote an attributed graph with node set $\mathcal{V}$ of size N, edge set $\mathcal{E}$, adjacency matrix $\mathbf{A} \in 0, 1^{N\times N}$, and node feature matrix $\mathbf{X} \in \mathbb{R}^{N\times D}$. The objective of Unsupervised Graph Anomaly Detection is to learn a scoring function $f : \mathcal{V} \to \mathbb{R}$ that assigns anomaly scores based on deviations in structural or attribute patterns, using only $\mathcal{G}$ during training [2].

B. Evidential Deep Learning Theory

Evidential Deep Learning quantifies uncertainty by treating model outputs as parameters of conjugate prior distributions. For continuous attributes, a Normal-Inverse-Gamma prior is placed over the Gaussian likelihood parameters (μ, σ^2):

$$p(\mu, \sigma^2|\gamma, \nu, \alpha, \beta) = \frac{\sqrt{\nu}\beta^\alpha}{\sqrt{2\pi}\Gamma(\alpha)} \left(\frac{1}{\sigma^2}\right)^{\alpha+\frac{3}{2}} \times \exp\left(-\frac{2\beta + \nu(\mu-\gamma)^2}{2\sigma^2}\right) \tag{1}$$

with aleatoric uncertainty $\mathbb{E}[\sigma^2] = \frac{\beta}{\alpha-1}$ and epistemic uncertainty $\text{Var}[\mu] = \frac{\beta}{\nu(\alpha-1)}$ [21]. For binary structures, a Beta prior models the Bernoulli likelihood:

$$p(p|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \tag{2}$$

where the total evidence $S = \alpha + \beta$ yields uncertainty $u = 2/S$ [3]. Categorical uncertainty is modeled via a Dirichlet prior with concentration parameters $\alpha = [\alpha_1, \ldots, \alpha_K]$, giving $u_{cluster} = K/\sum_{k=1}^{K} \alpha_k$ [22].

IV. METHODOLOGY

MVEA adopts an encoder-decoder architecture with multi-view evidential alignment at the decoding stage. Given an attributed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$, a shared graph neural network encoder f_{enc} maps node attributes and graph structure into a low-dimensional latent space:

$$\mathbf{Z} = f_{\text{enc}}(\mathbf{X}, \mathbf{A}; \boldsymbol{\Theta}_{\text{enc}}) \tag{3}$$

The latent representation $\mathbf{Z}$ is simultaneously fed into three parallel branches: an attribute evidential decoder, a structure evidential decoder, and a latent clustering head. These branches are jointly optimized end-to-end to exploit complementary information across views. The overall framework is illustrated in Figure 2.

A. Attribute Evidential Reconstruction

1) Theoretical Analysis of Evidence Contraction: In the Normal-Inverse-Gamma regression framework, the model minimizes the negative log-likelihood loss $\mathcal{L}_{\text{NLL}}$ to estimate the parameters $\boldsymbol{\theta} = \{\gamma, \nu, \alpha, \beta\}$. The gradient driving the update of the predicted mean γ is given by [5]

$$\frac{\partial \mathcal{L}_{\text{NLL}}}{\partial \gamma} = -(2\alpha + 1)\frac{(\mathbf{x} - \gamma)\nu}{(\mathbf{x} - \gamma)^2\nu + 2\beta(1 + \nu)} \tag{4}$$

For anomalous nodes with large reconstruction errors, the optimizer tends to suppress the evidence count ν toward zero to reduce penalties rather than adjust network parameters to fit the observations, as the former incurs lower update costs [5]. Since ν is parameterized via $\nu = \text{Softplus}(o_\nu)$, driving $\nu \to 0$ forces the logit $o_\nu \to -\infty$. Substituting $\nu \to 0$ into Equation 4 reveals a vanishing gradient:

$$\lim_{\nu \to 0} \frac{\partial \mathcal{L}_{\text{NLL}}}{\partial \gamma} = 0 \tag{5}$$

Furthermore, the gradient with respect to the network weights also vanishes through the activation saturation:

$$\frac{\partial \mathcal{L}_{\text{NLL}}}{\partial o_\nu} = \frac{\partial \mathcal{L}_{\text{NLL}}}{\partial \nu} \cdot \sigma'(o_\nu) \xrightarrow{o_\nu \to -\infty} 0 \tag{6}$$

Consequently, for the most informative hard samples such as anomalous nodes, the regression head stops providing meaningful gradients for representation learning, causing encoder updates on these nodes to stagnate.

2) Latent Evidential Clustering with Gradient Compensation: To mitigate evidence contraction, we introduce a gradient compensation mechanism via latent evidential clustering. A parallel clustering head is imposed on the shared latent representation, parameterized by a Dirichlet distribution. The total objective is $\mathcal{J} = \mathcal{L}_{\text{NLL}} + \lambda_c \mathcal{L}_{\text{clus}}$, where $\mathcal{L}_{\text{clus}}$ is the KL divergence between the soft assignment distribution Q and the sharpened target distribution P [23]. The soft assignment probability q_{ij} between node embedding $\mathbf{h}_i$ and centroid $\mathbf{c}_j$ is computed via the Student-t kernel, and the target distribution is obtained by squaring and renormalizing Q to enforce compact cluster assignments.

The total gradient with respect to the shared embedding $\mathbf{h}$ is

$$\nabla_{\mathbf{h}} \mathcal{J} = \underbrace{\left(\frac{\partial \mathcal{L}_{\text{NLL}}}{\partial \nu}\sigma'(o_\nu)\right)\mathbf{w}_\nu}_{\text{Regression Term } (\to \mathbf{0})} + \lambda_c \underbrace{\nabla_{\mathbf{h}} \mathcal{L}_{\text{clus}}}_{\text{Clustering Term}} \tag{7}$$

When evidence contraction occurs and $\sigma'(o_\nu) \approx 0$, the regression term vanishes. The clustering gradient with respect to $\mathbf{h}_i$ remains analytically non-zero:

$$\nabla_{\mathbf{h}_i} \mathcal{L}_{\text{clus}} = \sum_{j=1}^{K} \left(1 + \|\mathbf{h}_i - \mathbf{c}_j\|^2\right)^{-1} (p_{ij} - q_{ij})(\mathbf{h}_i - \mathbf{c}_j) \tag{8}$$

Anomalous node embeddings typically reside in low-density regions far from cluster centroids or near cluster boundaries with high assignment entropy [24]. This structural mismatch keeps the assignment discrepancy $(p_{ij} - q_{ij})$ and displacement $(\mathbf{h}_i - \mathbf{c}_j)$ large, sustaining a non-zero clustering gradient that continues to update the shared encoder. These updates gradually push the logit o_ν out of the saturation region, partially restoring gradient flow through the regression branch. The attribute view loss is

$$\mathcal{L}_{\text{attr}} = \mathcal{L}_{\text{NLL}} + \mathcal{L}_{\text{reg}} + \mathcal{L}_{\text{clus}} \tag{9}$$

where $\mathcal{L}_{\text{reg}}$ is a KL regularization term between the predicted Normal-Inverse-Gamma distribution and an uninformative prior, preventing overconfident predictions in sparse evidence regions.

B. Structure Evidential Reconstruction

1) Analysis of Structural Confidence Bias: Graph topology reconstruction faces a severe class imbalance problem, where negative edge samples vastly outnumber positive edges. In the evidential learning context, this imbalance causes the aggregate gradient to be dominated by the negative class. The model is driven to globally maximize the absence evidence β, so that for true edges that are difficult to reconstruct, it inherits this global bias and outputs a high β rather than the appropriate low-evidence high-uncertainty response. The model thereby produces overconfident yet erroneous predictions for edge absence rather than expressing the epistemic uncertainty that structural anomalies warrant.

2) Cost-Sensitive Evidential Learning: To rectify structural confidence bias, we propose a multi-stage cost-sensitive objective that decouples evidence generation, aligns relative confidence rankings, and calibrates the distribution with targeted regularization.

Dual-Branch Evidential Generation. We decouple the generation of positive and negative evidence by projecting node representations into two distinct latent spaces $\mathbf{H}_\alpha$ and $\mathbf{H}_\beta$ via independent projection heads. The evidential parameters for each node pair (i, j) are computed as:

$$\mathbf{E}^\alpha = \text{Softplus}(\mathbf{H}_\alpha \mathbf{H}_\alpha^\top) + 1, \quad \mathbf{E}^\beta = \text{Softplus}(\mathbf{H}_\beta \mathbf{H}_\beta^\top) + 1 \tag{10}$$

The offset of 1 enforces $\alpha_{ij}, \beta_{ij} \geq 1$, ensuring the Beta distribution remains unimodal and capable of expressing intermediate uncertainty rather than collapsing to bimodal extremes.

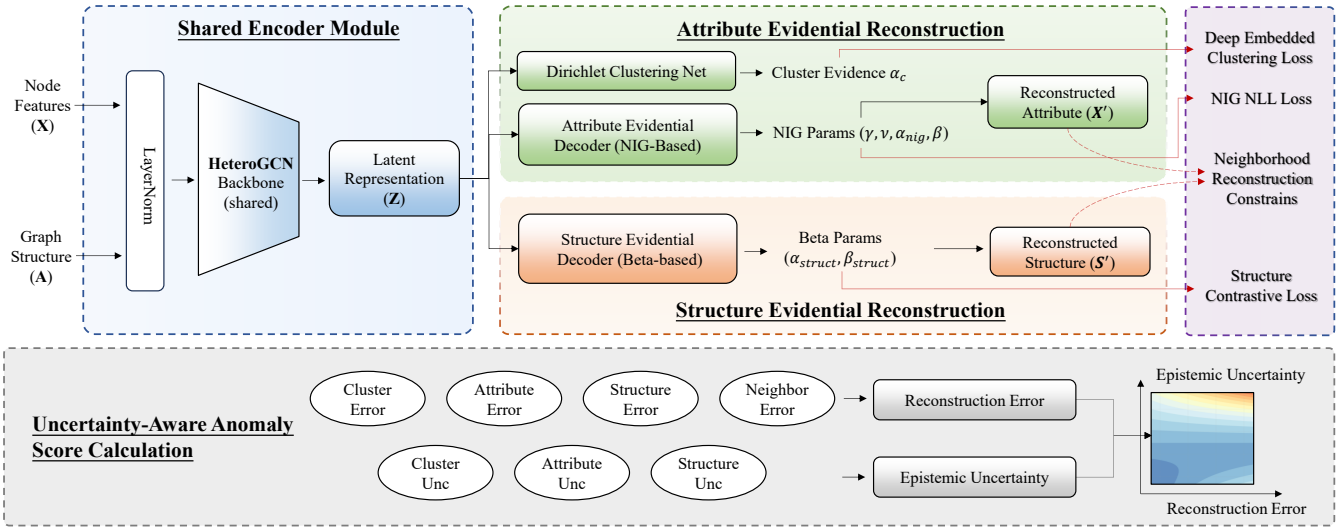

Fig. 2. The overall framework of the proposed MVEA.

Contrastive Ranking on Evidential Logits. We define the evidential logit s_{ij} as the log-odds of the expected edge probability:

$$s_{ij} = \log(\alpha_{ij}) - \log(\beta_{ij}) \tag{11}$$

Rather than supervising absolute probability values directly, we optimize the relative ordering of structural beliefs via an InfoNCE-style ranking loss [25]:

$$\mathcal{L}_{\text{rank}} = \sum_{i \in \mathcal{V}} \left(\log \sum_{k \in \mathcal{V}} \exp(s_{ik}/\tau) - \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} s_{ij}/\tau \right) \tag{12}$$

where τ is a temperature parameter. This loss lifts the evidential logits of true neighbors above the aggregate background, mitigating the sparsity-induced bias without relying on absolute probability thresholds.

Error-Aware Uncertainty Regularization. To calibrate the absolute evidence magnitude, we introduce an error-aware KL divergence that penalizes confident incorrect predictions. For each node pair (i, j), the regularization is weighted by the prediction error [3]:

$$\ell_{\text{KL}}^{(ij)} = |A_{ij} - p_{ij}| \cdot \text{KL}[\text{Beta}(\alpha_{ij}, \beta_{ij}) \parallel \text{Beta}(1, 1)] \tag{13}$$

This dynamic constraint forces evidence parameters toward 1 when predictions deviate from ground truth, converting overconfident errors into appropriately uncertain predictions. The loss is aggregated separately for positive and negative edges to handle sparsity:

$$\mathcal{L}_{\text{KL}} = \frac{\lambda_{\text{pos}}}{|\mathcal{E}|} \sum_{(i,j) \in \mathcal{E}} \ell_{\text{KL}}^{(ij)} + \frac{\lambda_{\text{neg}}}{|\bar{\mathcal{E}}|} \sum_{(i,j) \notin \mathcal{E}} \ell_{\text{KL}}^{(ij)} \tag{14}$$

Misleading Evidence Penalty. To further suppress overconfident erroneous predictions, we directly penalize evidence parameters that contradict the ground truth:

$$\mathcal{L}_{\text{mis}} = \frac{\lambda_{\text{pos}}}{|\mathcal{E}|} \sum_{(i,j) \in \mathcal{E}} (\beta_{ij} - 1) + \frac{\lambda_{\text{neg}}}{|\bar{\mathcal{E}}|} \sum_{(i,j) \notin \mathcal{E}} (\alpha_{ij} - 1) \tag{15}$$

The total structural loss is $\mathcal{L}_{\text{struct}} = \mathcal{L}_{\text{rank}} + \mathcal{L}_{\text{KL}} + \mathcal{L}_{\text{mis}}$.

C. Neighborhood Distribution Alignment

Individual node reconstruction often fails to capture collective anomaly patterns at the subgraph level [10]. We propose a neighborhood distribution alignment module that constrains the statistical moments of reconstructed neighborhoods to match their ground-truth counterparts. The ground-truth neighborhood mean and variance for node v_i are computed under a diagonal Gaussian approximation:

$$\boldsymbol{\mu} = \mathbf{D}^{-1}\mathbf{A}\mathbf{X}, \quad \boldsymbol{\sigma}^2 = \mathbf{D}^{-1}\mathbf{A}(\mathbf{X} \odot \mathbf{X}) - \boldsymbol{\mu}^2 \tag{16}$$

On the reconstructed side, the soft adjacency matrix $\hat{\mathbf{A}}$ is derived from the posterior expectation of the Beta distribution:

$$\hat{A}_{ij} = \mathbb{E}[p_{ij}] = \frac{\alpha_{ij}}{\alpha_{ij} + \beta_{ij}} \tag{17}$$

The reconstructed neighborhood statistics are then computed using $\hat{\mathbf{A}}$ and the predicted attribute matrix $\hat{\mathbf{X}} = \boldsymbol{\gamma}$:

$$\hat{\boldsymbol{\mu}} = \hat{\mathbf{D}}^{-1}\hat{\mathbf{A}}\hat{\mathbf{X}}, \quad \hat{\boldsymbol{\sigma}}^2 = \hat{\mathbf{D}}^{-1}\hat{\mathbf{A}}(\hat{\mathbf{X}} \odot \hat{\mathbf{X}}) - \hat{\boldsymbol{\mu}}^2 \tag{18}$$

The alignment loss minimizes the KL divergence between the ground-truth and reconstructed neighborhood distributions:

$$\mathcal{L}_{\text{dist}} = \sum_{i \in \mathcal{V}} \text{KL}\big(\mathcal{N}(\boldsymbol{\mu}_i, \text{diag}(\boldsymbol{\sigma}_i^2)) \parallel \mathcal{N}(\hat{\boldsymbol{\mu}}_i, \text{diag}(\hat{\boldsymbol{\sigma}}_i^2))\big) \tag{19}$$

This module couples the attribute and structure views at the subgraph level, enabling the detection of collective anomalies that evade node-wise reconstruction.

D. Overall Objective

The total training objective combines the three complementary losses:

$$\mathcal{L}_{\text{total}} = \gamma_1 \mathcal{L}_{\text{attr}} + \gamma_2 \mathcal{L}_{\text{struct}} + \gamma_3 \mathcal{L}_{\text{dist}} \tag{20}$$

where $\gamma_1, \gamma_2, \gamma_3$ balance the contribution of each view. At inference, a single forward pass yields all evidential parameters, enabling efficient anomaly scoring without sampling or post-processing.

E. Uncertainty-Rectified Anomaly Scoring

1) Limitations of Linear Aggregation: Existing evidential graph anomaly detection methods derive anomaly scores as a linear combination $S = (1-\lambda)\,\text{Rec} + \lambda\,\text{Unc}$. This formulation implicitly treats reconstruction error and epistemic uncertainty as independent indicators, overlooking their non-linear interaction. Genuine anomalies manifest in two distinct patterns. Structural anomalies exhibit high reconstruction error coupled with low uncertainty, indicating that the model is confident yet incorrect. Out-of-distribution anomalies exhibit high uncertainty, reflecting the model's insufficient knowledge of these samples. Ambiguous noise, by contrast, produces moderate error and moderate uncertainty simultaneously. The linear aggregation fails to separate this third category, as the summation of moderate values can exceed the score of genuine structural anomalies, producing false positives.

2) Rectified Scoring via U-Shaped Gating: We propose a non-linear scoring mechanism grounded in the observation that reconstruction error is a reliable anomaly indicator only when the model is in a decisive state, either with high confidence or complete ignorance, and becomes unreliable under cognitive ambiguity. Let $\text{Rec}_{\text{total}} \in [0,1]$ and $\text{Unc}_{\text{total}} \in [0,1]$ denote the min-max normalized total reconstruction error and epistemic uncertainty, respectively. We define a U-shaped certainty coefficient:

$$\kappa(\text{Unc}_{\text{total}}) = (2 \cdot \text{Unc}_{\text{total}} - 1)^2 \tag{21}$$

This parabolic function attains its minimum of zero at $\text{Unc} = 0.5$, completely suppressing the error contribution under maximal ambiguity, and recovers to one at both extremes, fully activating the error signal under decisive model states. The final anomaly score for node v_i is:

$$\text{Score}(v_i) = (1-\lambda) \cdot \text{Rec}_{\text{total},i} \cdot \kappa_i + \lambda \cdot \text{Unc}_{\text{total},i} \tag{22}$$

where λ balances the contribution of the gated reconstruction error and the uncertainty term. This design amplifies the penalty for high-confidence errors while suppressing ambiguous noise, yielding a more precise anomaly ranking.

V. Experiments

A. Experimental Setup

1) Datasets: We evaluate **MVEA** on five heterogeneous benchmarks spanning social (**Weibo**, **Reddit**), communication (**Enron**), and e-commerce (**Disney**, **Books**) domains [26], [27]. These datasets exhibit diverse topological sparsity and attribute dimensionality, providing a robust testbed where anomalies correspond to semantic outliers such as spammers, banned users, and fraudulent actors.

2) Baselines and Metrics: We benchmark **MVEA** against six state-of-the-art methods: classic reconstruction models (**DOMINANT** [7], **AnomalyDAE** [28]); augmented frameworks incorporating contrastive or neighborhood constraints (**CONAD** [8], **GAD-NR** [10], **G3AD** [11]); and the uncertainty-aware evidential model **GEL** [3]. Evaluation is performed using **AUC** for overall detection accuracy and **Recall@K** [2] to assess the capability of retrieving top-ranked anomalies.

B. Main Results

1) Performance Analysis: As presented in Table I, **MVEA** achieves state-of-the-art performance on three benchmarks including Reddit, Disney, and Books while securing competitive rankings on Weibo, thereby demonstrating robust improvements over representative baselines. Specifically, our method outperforms traditional reconstruction-based methods such as AnomalyDAE and consistently surpasses auxiliary-enhanced models like GAD-NR on complex datasets. Compared to the strongest evidential competitor GEL, **MVEA** exhibits distinct advantages in differentiating anomalies. While GEL remains competitive on Enron, our model achieves substantial gains elsewhere, peaking with an AUC improvement of approximately 2.3% on Disney. Notably, **MVEA** demonstrates a remarkable leap in Recall@K, a metric critical for practical high-confidence detection. For instance, on the Reddit dataset, it improves Recall@K by over 260% compared to GEL and achieves a 53% relative improvement on Disney. This performance leap is attributed to two fundamental advancements. First, unlike GEL which is prone to evidence contraction, our high-fidelity evidence construction integrates latent clustering and structure-aware constraints to compel the learning of semantically meaningful evidential parameters. Second, in contrast to the simplistic linear aggregation employed by baselines, our non-linear uncertainty rectification models the interaction between reconstruction error and epistemic uncertainty. By utilizing the certainty coefficient as a dynamic gate, this mechanism effectively amplifies the penalty for high-confidence errors while suppressing the contribution of uncertain noisy samples to yield a significantly more precise anomaly ranking.

2) Ablation Study: To rigorously verify the contribution of each component within **MVEA**, we conduct a comprehensive ablation study on the Reddit dataset, incrementally evaluating Latent Evidential Clustering, the Structure-aware Objective, the Neighborhood Constraint, and the Uncertainty-Rectified Scoring mechanism as summarized in Table II. First, regarding attribute reconstruction, solely minimizing the negative log-likelihood in Variant A proves insufficient due to potential evidence contraction. The integration of Latent Evidential Clustering in Variant B boosts the AUC to 59.30%, confirming that clustering constraints act as critical regularizers to enforce compact semantic prototypes. Second, for topology modeling, the ranking-based objective in Variant D surpasses the standard

TABLE I

PERFORMANCE COMPARISON (MEAN ± STD%) WITH STATE-OF-THE-ART BASELINES ON FIVE BENCHMARK DATASETS. THE BEST RESULTS ARE HIGHLIGHTED IN **BOLD**, AND THE SECOND-BEST ARE UNDERLINED. METRICS INCLUDE AUC AND RECALL@K (REC@K).

Method	Weibo		Reddit		Enron		Disney		Books	
	AUC	Rec@K	AUC	Rec@K	AUC	Rec@K	AUC	Rec@K	AUC	Rec@K
DOMINANT	85.0±14.6	19.7±13.8	56.0±0.2	0.9±0.4	73.1±8.9	0.0±0.0	47.1±4.5	3.3±6.7	50.1±5.0	1.6±3.1
AnomalyDAE	91.5±1.2	42.2±23.7	55.7±0.4	0.9±0.5	54.3±11.2	0.0±0.0	48.8±2.2	0.0±0.0	62.2±8.1	2.7±2.2
GAD-NR	87.71±5.39	28.5±5.4	57.99±1.67	1.6±0.5	80.87±2.95	0.0±0.0	76.76±2.75	12.4±2.8	65.71±4.98	3.9±1.5
CONAD	85.4±14.3	20.3±13.3	56.1±0.1	1.3±1.6	71.9±4.9	0.0±0.0	48.0±3.5	0.8±3.6	52.2±6.9	1.7±2.9
G3AD	**95.1±1.35**	**52.1±4.2**	62.1±0.22	2.4±0.8	72.39±2.9	0.0±0.0	65.3±1.7	6.8±1.5	54.0±4.2	1.8±0.9
GEL	89.32±3.36	35.6±3.3	62.76±2.42	2.9±0.6	**82.34±2.93**	0.0±0.0	78.21±4.94	15.2±3.4	70.79±3.16	5.5±1.2
MVEA (Ours)	91.91±4.0	35.7±2.1	**63.38±2.2**	**10.66±2.5**	80.46±2.24	**0.0±0.0**	**80.5±0.7**	**23.3±5.4**	**71.6.0±1.2**	**6.43±3.7**

TABLE II

ABLATION STUDY OF **MVEA** ON THE REDDIT DATASET. WE INCREMENTALLY EVALUATE THE EFFECTIVENESS OF LATENT EVIDENTIAL CLUSTERING (LEC), COST-SENSITIVE EVIDENTIAL LEARNING (INFONCE), NEIGHBORHOOD CONSTRAINT (NEIGH.), AND UNCERTAINTY-RECTIFIED SCORING (RECTIFIED). "LINEAR" DENOTES LINEAR SUMMATION OF ERROR AND UNCERTAINTY.

Variant	Attribute Recon.		Structure Recon.		Neigh.	Scoring Function		AUC (%)
	NLL+Reg	LEC	BCE	Ranking		Linear	Rectified	
A	✓	✗	✗	✗	✗	✓	✗	58.52
B	✓	✓	✗	✗	✗	✓	✗	59.30
C	✗	✗	✓	✗	✗	✓	✗	57.49
D	✗	✗	✗	✓	✗	✓	✗	58.18
E	✓	✓	✗	✓	✗	✓	✗	61.28
F	✓	✓	✗	✓	✓	✓	✗	62.24
MVEA	✓	✓	✗	✓	✓	✗	✓	**63.38**

Binary Cross Entropy used in Variant C. This indicates that prioritizing relative edge ordering yields more discriminative structural evidence in sparse graph scenarios compared to classification-based losses which are prone to class imbalance. Further building upon the integrated baseline of Variant E, the inclusion of the Neighborhood Constraint in Variant F elevates the AUC to 62.24%. This demonstrates that explicitly enforcing distribution consistency between reconstructed and ground-truth neighborhoods captures subtle local anomalies often overlooked by node-wise reconstruction alone. Finally, replacing the linear aggregation with our Uncertainty-Rectified Scoring achieves the optimal performance of 63.38%. This significant improvement validates that non-linearly modulating reconstruction error via the certainty coefficient effectively decouples aleatoric noise from epistemic anomalies, thereby resolving the reliability blind spot inherent in simplistic linear combinations.

3) Robustness Analysis: Real-world graphs are often susceptible to corruption. To evaluate the resilience of **MVEA** in non-ideal scenarios, we conduct robustness tests on the Weibo dataset by simulating two types of perturbations: Gaussian attribute noise injection and structural adversarial attacks (DICE). Table III reports the performance comparison against the strongest baseline, GEL, across perturbation rates ranging from 0% to 30%. As observed, while both methods experience performance degradation as corruption intensifies, **MVEA** exhibits significantly superior stability. Specifically, under attribute noise, **MVEA** maintains an AUC of over 86% even at a 20% perturbation rate, whereas GEL suffers a sharper decline. This resilience stems from our Latent Evidential Clustering (LEC) module, which effectively filters feature jitter as aleatoric uncertainty. Similarly, in the face of structural attacks that disrupt homophily, our Contrastive Ranking mechanism prioritizes the relative ordering of reliable neighbors, allowing **MVEA** to outperform GEL by a margin of 8.6% at the 30% attack rate. These results confirm that **MVEA** effectively rectifies uncertainty to mitigate the impact of data noise and adversarial anomalies.

TABLE III

ROBUSTNESS COMPARISON (AUC %) ON WEIBO DATASET UNDER ATTRIBUTE PERTURBATION AND STRUCTURAL ATTACK (DICE). UGAD DEMONSTRATES SLOWER PERFORMANCE DECAY COMPARED TO GEL.

Rate	Attribute Perturbation		Structural Attack	
	GEL	**UGAD**	GEL	**UGAD**
0%	89.32	**91.91**	89.32	**91.91**
10%	85.15	**89.40**	84.76	**88.65**
20%	79.80	**86.25**	78.92	**85.10**
30%	72.45	**82.33**	74.50	**83.10**

VI. Conclusion

We have introduced **MVEA**, a robust graph anomaly detection framework that effectively bridges the gap between reconstruction-based learning and uncertainty quantification. Through Latent Evidential Clustering and a novel Uncertainty-Rectified Scoring mechanism, **MVEA** successfully overcomes the evidence contraction and linear aggregation flaws inherent in prior works. Empirical results on five datasets confirm that our method not only establishes new state-of-the-art results in terms of AUC and Recall@K but also maintains exceptional stability under adversarial noise. These findings underscore the necessity of non-linear uncertainty modulation in high-stakes detection tasks. In the future, we plan to adapt this uncertainty-guided principle to dynamic and large-scale graph environments.

References

[1] X. Ma, J. Wu, S. Xue, J. Yang, C. Zhou, Q. Z. Sheng, H. Xiong, and L. Akoglu, "A comprehensive survey on graph anomaly detection with deep learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 12, pp. 12 012–12 038, 2023.

[2] K. Liu, Y. Dou, Y. Zhao, X. Ding, X. Hu, R. Zhang, K. Ding, C. Chen, H. Peng, K. Shu, L. Sun, J. Li, G. H. Chen, Z. Jia, and P. S. Yu, "Bond: Benchmarking unsupervised outlier node detection on static attributed graphs," in *Advances in Neural Information Processing Systems*, vol. 35, 2022, pp. 27 021–27 035.

[3] C. Wei, W. Hu, X. Hao, Y. Wang, Y. Chen, B. Bai, and F. Wang, "Graph evidential learning for anomaly detection," in *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, 2025, p. 3122–3133.

[4] J. Gao, M. Chen, L. Xiang, and C. Xu, "A comprehensive survey on evidential deep learning and its applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–20, 2025.

[5] Y. Wu, B. Shi, B. Dong, Q. Zheng, and H. Wei, "The evidence contraction issue in deep evidential regression: Discussion and solution," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 19, pp. 21 726–21 734, 2024.

[6] X. Ma, F. Liu, J. Wu, J. Yang, S. Xue, and Q. Z. Sheng, "Rethinking unsupervised graph anomaly detection with deep learning: Residuals and objectives," *IEEE Transactions on Knowledge and Data Engineering*, vol. 37, no. 2, pp. 881–895, 2025.

[7] K. Ding, J. Li, R. Bhanushali, and H. Liu, *Deep Anomaly Detection on Attributed Networks*, pp. 594–602.

[8] Z. Xu, X. Huang, Y. Zhao, Y. Dong, and J. Li, "Contrastive attributed network anomaly detection with data augmentation," in *Advances in Knowledge Discovery and Data Mining*, 2022, pp. 444–457.

[9] S. Kim, S. Y. Lee, F. Bu, S. Kang, K. Kim, J. Yoo, and K. Shin, "Rethinking reconstruction-based graph-level anomaly detection: Limitations and a simple remedy," in *Advances in Neural Information Processing Systems*, vol. 37, 2024, pp. 95 931–95 962.

[10] A. Roy, J. Shu, J. Li, C. Yang, O. Elshocht, J. Smeets, and P. Li, "Gad-nr: Graph anomaly detection via neighborhood reconstruction," in *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, ser. WSDM '24, 2024, p. 576–585.

[11] Y. Bei, S. Zhou, J. Shi, Y. Ma, H. Wang, and J. Bu, "Guarding graph neural networks for unsupervised graph anomaly detection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 9, pp. 16 840–16 853, 2025.

[12] F. Wang, Y. Liu, K. Liu, Y. Wang, S. Medya, and P. S. Yu, "Uncertainty in graph neural networks: A survey," *Transactions on Machine Learning Research*, 2024. [Online]. Available: https://openreview.net/forum?id=0e1Kn76HM1

[13] D. Wu, L. Gao, M. Chinazzi, X. Xiong, A. Vespignani, Y.-A. Ma, and R. Yu, "Quantifying uncertainty in deep spatiotemporal forecasting," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, ser. KDD '21, 2021, p. 1841–1851.

[14] Y. Zhang, S. Pal, M. Coates, and D. Ustebay, "Bayesian graph convolutional neural networks for semi-supervised classification," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 5829–5836, 2019.

[15] M. Stadler, B. Charpentier, S. Geisler, D. Zügner, and S. Günnemann, "Graph posterior network: Bayesian predictive uncertainty for node classification," in *Advances in Neural Information Processing Systems*, vol. 34, 2021, pp. 18 033–18 048.

[16] K. Huang, Y. Jin, E. Candes, and J. Leskovec, "Uncertainty quantification over graph with conformalized graph neural networks," in *Advances in Neural Information Processing Systems*, vol. 36, 2023, pp. 26 699–26 721.

[17] S. Bai, X. Zheng, and D. D. Zeng, "Crc-sgad: Conformal risk control for supervised graph anomaly detection," in *2025 IEEE International Conference on Intelligence and Security Informatics (ISI)*, 2025, pp. 151–156.

[18] S. Bai, Y. Ji, Y. Liu, X. Zhang, X. Zheng, and D. D. Zeng, "Alleviating performance disparity in adversarial spatiotemporal graph learning under zero-inflated distribution," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 11, pp. 11 436–11 444, Apr. 2025.

[19] L. Wang, D. He, H. Zhang, Y. Liu, W. Wang, S. Pan, D. Jin, and T.-S. Chua, "Goodat: Towards test-time graph out-of-distribution detection," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 14, pp. 15 537–15 545, Mar. 2024.

[20] W. Chang, K. Liu, K. Ding, P. S. Yu, and J. Yu, "Multitask active learning for graph anomaly detection," 2024. [Online]. Available: https://arxiv.org/abs/2401.13210

[21] A. Amini, W. Schwarting, A. Soleimany, and D. Rus, "Deep evidential regression," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 14 927–14 937.

[22] M. Sensoy, L. Kaplan, and M. Kandemir, "Evidential deep learning to quantify classification uncertainty," in *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[23] J. Xie, R. Girshick, and A. Farhadi, "Unsupervised deep embedding for clustering analysis," in *Proceedings of The 33rd International Conference on Machine Learning*, vol. 48, 2016, pp. 478–487.

[24] L. Zheng, J. Birge, H. Wu, Y. Zhang, and J. He, "Cluster aware graph anomaly detection," in *Proceedings of the ACM on Web Conference 2025*, ser. WWW '25, 2025, p. 1771–1782.

[25] Z. Wang, B. Xu, Y. Yuan, H. Shen, and X. Cheng, "Infonce is a free lunch for semantically guided graph contrastive learning," in *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '25, 2025, p. 719–728.

[26] J. He, Q. Xu, Y. Jiang, Z. Wang, and Q. Huang, "Ada-gad: Anomaly-denoised autoencoders for graph anomaly detection," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 8, pp. 8481–8489, 2024.

[27] Y. Bei, S. Zhou, Q. Tan, H. Xu, H. Chen, Z. Li, and J. Bu, "Reinforcement neighborhood selection for unsupervised graph anomaly detection," in *2023 IEEE International Conference on Data Mining (ICDM)*, 2023, pp. 11–20.

[28] H. Fan, F. Zhang, and Z. Li, "Anomalydae: Dual autoencoder for anomaly detection on attributed networks," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 5685–5689.