CAMERA: Adapting to Semantic Camouflage in Unsupervised Text-Attributed Graph Fraud Detection

Junjun Pan[1] , **Yixin Liu**[1*] , **Yu Zheng**[1*] , **Lianhua Chi**[2] , **Alan Wee-Chung Liew**[1] , **Shirui Pan**[1]

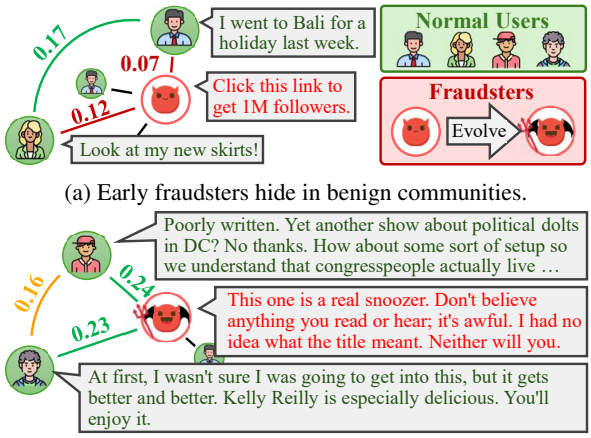
[1]School of Information and Communication Technology, Griffith University, Australia
[2]Department of Computer Science and Information Technology, La Trobe University, Australia
junjun.pan@griffithuni.edu.au, {yixin.liu, yu.zheng, a.liew, s.pan}@griffith.edu.au, l.chi@latrobe.edu.au

Abstract

Text-attributed graph fraud detection (TAGFD) plays a critical role in preventing fraudulent activities on online social and e-commerce platforms. However, to evade detection, fraudsters continuously evolve their camouflaging strategies by deliberately mimicking textual responses of benign users, thereby concealing their malicious purposes. This phenomenon, referred to as **semantic camouflage**, fundamentally undermines commonly relied assumptions on how structural and attribute cues can be exploited to identify fraudsters, and makes it difficult to spot fraudsters with unsupervised TAGFD. To bridge the gaps, we propose a **C**ase-**A**daptive **M**ulti-cue **E**xpert f**RA**mework (CAMERA) for unsupervised TAGFD. CAMERA employs an ego-decoupled mixture-of-experts architecture, where each expert specializes in modeling a distinct type of fraud-indicative cue. A context-informed gating model is introduced to jointly consider the ego node representation and its local neighborhood context for adaptive integration of cues learned by different experts. Furthermore, CAMERA leverages the inherent rarity of fraudsters to support unsupervised one-class learning with expert-level objectives that encourage modeling dominant benign patterns, thereby enabling reliable unsupervised detection of camouflaged fraudsters. Experiments on 4 challenging datasets show that CAMERA consistently outperforms competitors, showing its effectiveness against semantically camouflaged fraudsters. Code available at https://github.com/CampanulaBells/CAMERA

1 Introduction

With the continued advancement of information technology, graph-structured data has become ubiquitous in online services such as e-commerce [Yu *et al.*, 2023] and social media [Hu *et al.*, 2023]. This ubiquity has been accompanied by increasing malicious activities, including fake reviews and spam messages, which underscores the importance of graph fraud detection (GFD) [Pan *et al.*, 2025a; Cai *et al.*, 2025]. Since many real-world platforms are inherently associated with textual content, text-attributed graph fraud detection (TAGFD) has emerged as an essential research direction for identifying suspicious behaviors embedded in text-attributed graphs [Yang *et al.*, 2025a; Qian *et al.*, 2026]. By jointly capturing network structures and attribute information, TAGFD plays a vital role in safeguarding the online platforms [Pan *et al.*, 2026a].



(a) Early fraudsters hide in benign communities.

(b) Evolved fraudsters employ semantic camouflage.

Figure 1: Illustration of fraudster evolution in text-attributed graphs, where the edge weights indicate the affinities between two nodes.

Despite the remarkable progress of detection techniques in recent years, fraudsters also continuously adapt their camouflage strategies to evade detection [Yang *et al.*, 2025b]. As illustrated in Figure 1, earlier fraudsters primarily operated at the structure level, where fraudulent nodes attempted to blend into the benign community by connecting themselves to benign users [Dou *et al.*, 2020]. While existing fraud detection methods could still handle these topology-level camouflages by pruning heterophily edges [Qiao and Pang, 2023], evolved fraudsters also deliberately hide their malicious intent by mimicking normal characteristics, which significantly increases the difficulty of detection. Going beyond structure-level blending, these evolved fraudsters further engage in **semantic camouflage**, i.e., they reshape their language to mimic benign expressions like criticism, while quietly advancing malicious goals like review manipulation

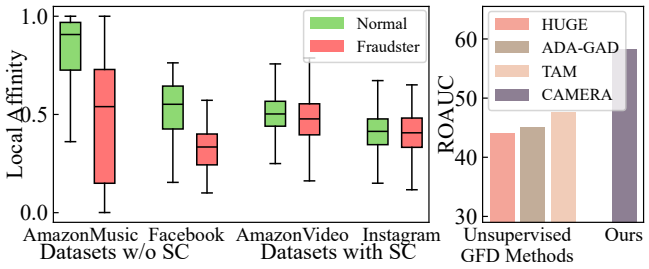


Figure 2: Left: Local affinity of different datasets with and without semantic camouflage (SC). Right: ROAUC on the Instagram dataset.

or the spread of misleading information.

To counter this semantic camouflage, recent TAGFD studies have explored supervised approaches to learn semantic-level anomalies-indicative cues. For instance, FLAG [Yang *et al.*, 2025a] leverages a large language model (LLM) to extract both shared and discriminative contextual signals from collections of fraudulent and normal texts, while DGP [Li *et al.*, 2025] integrates graph structure into LLM fine-tuning. However, these methods typically rely on labels to find the cue from semantics. In real-world applications, however, obtaining reliable annotations for GFD is often costly and impractical. This limitation motivates us to explore a more practical research problem, i.e., **unsupervised TAGFD**, where fraudsters are identified without access to ground-truth labels.

To handle this new research problem, a naive solution is to extend existing unsupervised GFD methods [Qiao and Pang, 2023; He *et al.*, 2024; Pan *et al.*, 2025b] by incorporating a text encoder to extract textual representation as graph features. These methods assume that fraudsters disrupt local affinity [Huang *et al.*, 2023], and hence the structure and attribute information are combined to compute affinity as fraud-indicative cues, either directly detecting fraudsters [Pan *et al.*, 2025b] or revealing camouflage through edge pruning [Qiao and Pang, 2023; He *et al.*, 2024]. However, as shown in Figure 2, fraudsters with semantic camouflage induce only minor changes in local affinity. As a result, simply extending unsupervised GFD methods to TAGFD scenarios with semantic camouflage can lead to degraded performance. In light of this, a core question arises:

How can we perform unsupervised TAGFD under semantic camouflage without access to ground-truth labels?

In answering this question, we identify two critical challenges: ***Challenge 1*** - *Adaptive integration of fraud-indicative cues.* Effective fraud detection relies on capturing diverse fraud-indicative cues from both graph structure and node attributes. However, evolving camouflaging strategies and increasingly complex real-world GFD scenarios continually invalidate the predetermined assumptions regarding the integration of these cues. This motivates the need for approaches that adaptively integrate multiple fraud-indicative cues to detect fraudsters with diverse camouflaging strategies. ***Challenge 2*** - *Capturing fraud-indicative cues under unsupervised settings.* Evolved fraudsters not only hide in benign communities but also actively camouflage in the semantic domain, making their malicious evidence hard to recognize. Such semantic camouflage undermines many commonly adopted heuristic designs, making it particularly challenging to highlight minor deviations from normal patterns without human-annotated labels. As a result, how to leverage the minority nature of fraudsters becomes critical for building an effective TAGFD model.

To fill the gap, we propose a **C**ase-**A**daptive **M**ulti-cue **E**xpert f**RA**mework (CAMERA) for unsupervised TAGFD against evolved fraudsters. To tackle ***Challenge 1***, CAMERA introduces an ego-decoupled Mixture-of-Experts (MoE) architecture that allows adaptive integration of multiple fraud-indicative cues, where each expert specializes in a distinct type of anomaly signal. By explicitly decoupling the specialization of each expert, the experts can extract complementary fraud-indicative cues that highlight diverse fraud patterns. Moreover, we further enhance the gating model with local context to guide the adaptive integration in a finer manner. To address ***Challenge 2***, CAMERA leverages the inherent rarity of fraudsters to support unsupervised training. Through one-class learning combined with expert-specific loss, each expert is encouraged to model dominant benign patterns, making malicious deviations introduced by fraudsters more pronounced. Based on these deviation-aware representations, a lightweight parameter-free fraud detector identifies nodes that deviate from normal patterns, enabling unsupervised detector training. Together, CAMERA avoids reliance on rigid assumptions about fraud behaviors while integrating multiple fraud-indicative cues, providing a solid solution for unsupervised TAGFD in the presence of semantically camouflaged fraudsters. In summary, our contributions are threefold:

Problem. To the best of our knowledge, we are the first to formally address the challenge of unsupervised TAGFD under the semantic camouflage scenarios, where fraudsters can mimic benign behaviors to hide their malicious intent.

Method. We propose CAMERA, a novel unsupervised TAGFD framework that employs a MoE architecture to adaptively integrate multiple fraud-indicative cues to reveal semantically camouflaged fraudsters without supervision.

Experiments. We conduct extensive experiments to demonstrate the superior performance of CAMERA over the state-of-the-art methods on four real-world TAGFD datasets under unsupervised learning scenarios.

2 Related Work

In this section, we provide a summary of two related areas. Detailed reviews are provided in Appendix A.

Fraud Detection on Attributed Graph. Early studies formulate graph fraud detection (GFD) as a supervised class-imbalanced classification problem, relying on labeled fraud instances and sampling-based techniques [Liu *et al.*, 2021a]. However, real-world fraudsters often camouflage themselves by blending into benign communities via heterophilic edges [Gao *et al.*, 2023], which substantially degrades supervised models. To address this, later works explicitly incorporate this finding into model design [Zhao *et al.*, 2025], for example, by pruning heterophilic edges [Gao *et al.*, 2023] or mitigating representation shift caused by

fraudsters [Tang *et al.*, 2022]. However, the cost of obtaining annotations can limit their application in diverse real-world scenarios. To address this, follow-up unsupervised GFD studies incorporate heuristic assumptions to obtain heterophilic edges. For example, TAM [Qiao and Pang, 2023] and ADA-GAD [He *et al.*, 2024] iteratively compute pseudo-labels to prune heterophilic edges, while HUGE [Pan *et al.*, 2025b] establishes a label-free heterophily measure as guidance, thereby training the fraud detector through alignment loss. While recent works further enhance the generalizability by transfer to unseen domains [Pan *et al.*, 2026b; Zhao *et al.*, 2026] or building one-for-all GAD models [Liu *et al.*, 2026], they ignore the rich semantics contained in text attributes, thereby limiting their application.

Fraud Detection on Text-Attributed Graph Many real-world graphs contain rich textual context, making text-attributed GFD (TAGFD) an important research direction, where evolved fraudsters employ semantic camouflage to escape detection. As a representative method, CoLL [Xu *et al.*, 2025a] incorporates LLMs to generate anomaly evidence, but as evolved fraudsters deliberately mimic benign user behaviors, yet when fraudsters deliberately mimic benign behaviors, LLMs are easily misled, resulting in incorrect or misleading evidence. Another line of work integrates structural and textual information into unified graph contrastive learning frameworks [Liu *et al.*, 2025; Xu *et al.*, 2025b]. Nevertheless, semantic camouflage typically induces only subtle ego-neighbor divergence compared to injected structural or feature anomalies [Ding *et al.*, 2019], weakening the effectiveness of contrastive objectives. These limitations motivate us to address semantic camouflage in unsupervised TAGFD.

3 Preliminary

Notations. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{T})$ represents a text-attributed graph, where $\mathcal{V}$ is the set of nodes and $\mathcal{E}$ is the set of edges. We denote the number of nodes and edges as n and m, respectively. $\mathcal{T} = \{t_i \mid v_i \in \mathcal{V}\}$ denotes the set of textual attributes, where $t_i \in \mathcal{D}^{L_i}$ is the text sequence associated with node v_i, $\mathcal{D}$ represents the dictionary of words, and L_i is the length of the text sequence. The graph structure can also be represented as adjacent matrix $\mathbf{A} \in \{0,1\}^{n\times n}$, with $\mathbf{A}_{i,j} = 1$ if there exists an edge between nodes v_i and v_j, and $\mathbf{A}_{i,j} = 0$ otherwise. We denote the set of neighbors of the i th node as $\mathcal{N}(v_i) = \{v_j \mid \mathbf{A}_{i,j} = 1\}$.

Problem Definition. In the context of TAGFD, each node $v_i \in \mathcal{V}$ has a label $y_i \in \{0,1\}$, where $y_i = 0$ indicates a benign node and $y_i = 1$ represents a fraudulent node. A commonly accepted assumption is that the number of benign nodes is significantly greater than the number of fraudulent nodes. In unsupervised scenarios, labels are unavailable during the training stage. Given a text-attributed graph $\mathcal{G}$, the goal of unsupervised TAGFD is to learn an scoring function $f : \mathcal{V} \to \mathcal{R}$ to identify whether a node v_i is suspicious by predicting a fraud score $s_i^{\text{fraud}} = f(v_i, \mathcal{G})$, where a higher score indicates a greater likelihood that the node is fraudulent.

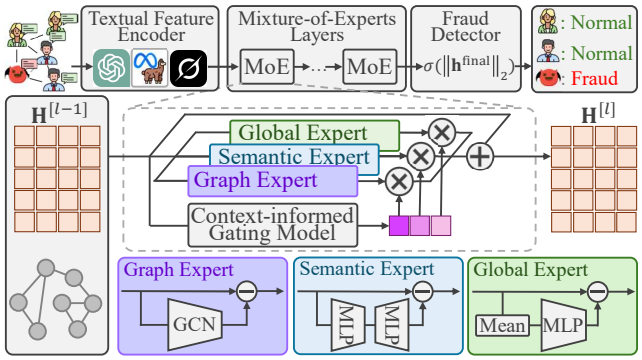


Figure 3: Overall framework of CAMERA.

4 Methodology

In this section, we provide an overview of CAMERA. As shown in Figure 3, we first encode textual node attributes using an LLM to extract high-quality node embeddings. Building on these embeddings, the ego-decoupled Mixture-of-Experts (MoE) layers extract complementary deviation signals, with each expert specializing in a distinct type of fraud-indicative cue. As what constitutes anomalous behavior differs across communities, the contributions of the MoE experts are weighted by a context-informed gating model that leverages both ego and neighborhood embeddings as priors. By learning dominant normal patterns in an unsupervised manner to expose subtle deviations, CAMERA enables adaptive integration of fraud-indicative cues for detecting camouflaged fraudsters without access to ground-truth labels.

4.1 Textual Feature Encoder

Conventional GFD methods often rely on pre-extracted shallow features (e.g., TF-IDF), which are insufficient for capturing the nuanced evidence for malicious purposes required to detect fraud under semantic camouflage. To address these limitations, we employ a pre-trained LLM to encode the textual attributes, providing richer semantic representations for the identification of camouflaged fraudsters. Specifically, given the textual attributes t_i of node v_i, we adopt an LLM-based feature encoder to transform the input text into a dense semantic representation $\mathbf{x}_i$ as the node attribute: $\mathbf{x}_i = \text{LLM}(t_i)$. The representations are subsequently used as the input to the downstream MoE modules, i.e., $\mathbf{H}^{[0]} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{R}^{n\times d}$, where d is the dimension of LLM-generated representations and each row $\mathbf{h}_i^{[0]}$ is the representation vector of node v_i. Compared with small-size text encoders such as SentenceBERT [Reimers and Gurevych, 2019], which were utilized by previous studies of text-attributed graphs, the LLM-based encoder exhibits a stronger semantic compression capability. This is particularly critical under unsupervised TAGFD settings, where subtle cues must be preserved without supervision signals.

4.2 Ego-decoupled MoE Architecture

While the pre-trained LLM-based encoder captures deep semantic information, spotting evolving camouflaged fraudsters

requires the effective and adaptive integration of other fraud-indicative cues, ranging from local structural patterns [Pan *et al.*, 2025b] to global semantic distributions [Jin *et al.*, 2021]. To address this challenge, we adopt a MoE architecture [Li *et al.*, 2026], where each expert is highly specialized and capable of capturing one type of discriminative fraud-indicative cues, providing strong and complementary representations for effective detection against camouflaged fraud [Tan *et al.*, 2025]. To model the interactions among heterogeneous cues, we further employ a multi-layer MoE design, where each MoE layer adaptively routes node representations to different experts and refines the fused representation, which hierarchically integrates the fraud-indicative cues. Moreover, a gating network complements the experts by dynamically learning to weight their contributions for each node. This adaptive mechanism allows the model to flexibly integrate fraud-indicative cues without relying on dataset-specific heuristics or domain knowledge, ensuring the most informative cues are prioritized.

Ego-Decoupled MoE Layer

As the essential building block of CAMERA, we first introduce the proposed MoE layer, which adaptively routes and integrates heterogeneous fraud-indicative cues. Specifically, taking the l-th layer as an example, a standard MoE layer can be written as:

$$\mathbf{H}^{[l]} = \sum_k \mathrm{diag}\big(g_k^{[l]}(\mathbf{H}^{[l-1]}, \mathbf{A})\big)\, e_k^{[l]}(\mathbf{H}^{[l-1]}, \mathbf{A}). \quad (1)$$

where $\mathbf{H}^{[l]} \in \mathbb{R}^{n\times d}$ denotes the output embedding matrix, $e_k^{[l]}(\mathbf{H}^{[l-1]}, \mathbf{A}) \in \mathbb{R}^{n\times d}$ is the embedding matrix produced by the expert k, and $g_k^{[l]}(\mathbf{H}^{[l-1]}, \mathbf{A}) \in \mathbb{R}^n$ denotes the corresponding gating weight vector.

Despite its effectiveness in modeling heterogeneous fraud-indicative cues, this generic formulation does not explicitly enforce functional decoupling among experts. Similar to the over-smoothing effect in graph neural networks [Rusch *et al.*, 2023], redundant expert functions can produce overly homogeneous representations [Liu *et al.*, 2023], which may obscure critical malicious cues. To address this issue, we propose an ego-decoupled MoE layer that separates expert-specific deviation signals from shared ego features:

$$\mathbf{H}^{[l]} = \mathbf{H}^{[l-1]} + \sum_k \mathrm{diag}\big(g_k^{[l]}(\mathbf{H}^{[l-1]}, \mathbf{A})\big) e_k^{[l]}(\mathbf{H}^{[l-1]}, \mathbf{A}). \quad (2)$$

By isolating shared ego embeddings, each expert focuses on complementary fraud signals rather than redundant information. Additionally, the skip connection integrates multi-hop graph features across layers, further enhancing the inter-layer information communication within the MoE model for fraud detection [Dong *et al.*, 2025].

Expert Specialization

MoE architectures typically achieve expert functional divergence through large-scale unsupervised pretraining [Fedus *et al.*, 2022] or domain partitioning [Gururangan *et al.*, 2022] with supervision. However, in unsupervised TAGFD, neither human annotations nor sufficiently large curated datasets are available, which hinders the operationally identical experts from learning diverse and discriminative fraud-indicative cues. Therefore, to ensure functionality divergence, we explicitly design the operation of each expert $e_k^{[l]}$ to specialize in a distinct fraud-indicative cue, ensuring functional differentiation and capturing complementary cues critical for detecting camouflaged fraudsters. Concretely, three specialized experts, i.e., the graph expert $e_{\text{graph}}^{[l]}$, the semantic expert, and the global expert, are instantiated in CAMERA, which are defined as follows.

Graph Expert. The graph expert focuses on capturing *structural deviation* signals, thereby spotting any unusual structural patterns that cannot be encoded with text attributes alone. We first encode local structural patterns using a GCN layer and compute the residual between ego representations and aggregated neighborhood features to expose anomalous structural discrepancies. Specifically, for each node v_i with representation $\mathbf{h}_i^{[l-1]}$, the operation of the graph expert can be written as:

$$e_{\text{graph}}^{[l]}(\mathbf{h}_i^{[l-1]}, \mathbf{A}) = \mathbf{h}_i^{[l-1]} - \mathrm{GCN}\Big(\mathbf{h}_i^{[l-1]}, \{\mathbf{h}_j^{[l-1]}\}_{v_j\in\mathcal{N}(v_i)}\Big). \quad (3)$$

Semantic Expert. To precisely spot any malicious semantic cues, the semantic expert employs an MLP-based autoencoder to learn a compact representation of benign semantics, and uses the difference between that and the input embedding to reveal *semantic deviations*:

$$e_{\text{semantic}}^{l}(\mathbf{h}_i^{[l-1]}) = \mathbf{h}_i^{[l-1]} - \mathrm{Decoder}(\mathrm{Encoder}(\mathbf{h}_i^{[l-1]})). \quad (4)$$

Hence, the semantic deviation encodes fine-grained malicious cues of the text, which is critical for detecting semantically camouflaged fraudsters that mimic normal behaviors.

Global Expert. While the graph and semantic experts focus on local deviations, fraudsters are inherently rare instances that deviate from the overall data distribution, which can be exposed from a global perspective. In light of this, we utilized the global expert to measure each node's discrepancy from the dominant benign prototype:

$$e_{\text{global}}^{[l]}(\mathbf{h}_i^{[l-1]}) = \mathbf{h}_i^{[l-1]} - \mathrm{MLP}(\mathbf{h}_{\text{global}}^{[l-1]}), \quad (5)$$

where $\mathbf{h}_{\text{global}}^{[l-1]} = \frac{1}{n}\sum_{v_j\in\mathcal{V}} \mathbf{h}_j^{[l-1]}$. The estimated *global deviation* characterizes how much a node diverges from the majority, complementing the local structural and semantic deviations captured by other experts.

Collectively, the graph, semantic, and global experts maintain functional differentiation while capturing complementary fraud residuals across structural, textual, and distributional fraud-indicative cues, which ensures that the MoE layer learns diverse and complementary fraud-indicative cues.

Context-informed Gating Model

After embedding extraction by specialized experts, a gating model is responsible for dynamically weighting their contributions for every node, enabling adaptive integration of fraud-indicative cues without relying on prior assumptions or domain-specific knowledge. In standard MoE architectures,

gating is typically implemented using input-dependent mechanisms such as self-attention [Lewis *et al.*, 2021]. However, in TAGFD, the importance of different fraud-indicative cues is not solely determined by a node's attributes, but is highly dependent on its local context. For example, in topic-focused communities, abnormal connection patterns often provide stronger evidence of fraud, whereas in more diverse communities, semantic irregularities such as misleading or toxic content become more informative. Relying only on ego features may therefore lead to suboptimal expert selection. Motivated by this observation, we propose a context-informed gating model that jointly considers the ego node representation and its local neighborhood context when determining expert contributions to allow for more fine-grained adaptation.

Specifically, to obtain the local neighborhood context $\mathbf{c}_i^{[l-1]}$ of node v_i, we aggregate the embeddings of its neighbors as:

$$\mathbf{c}_i^{[l-1]} = \frac{1}{\deg(v_i)} \sum_{v_j \in \mathcal{N}(v_i)} \mathbf{h}_j^{[l-1]}, \tag{6}$$

where $\mathbf{h}_j^{[l-1]}$ denotes the input embedding of node v_j at the $l-1$-th layer, $\mathcal{N}(v_i)$ is the neighbor set of v_i, and $\deg(v_i)$ is its degree. The gating network then leverages both the ego representation and the local context to compute expert weights:

$$\begin{aligned} \mathbf{g}_i^{[l]} &= \text{Softmax}\big(\text{Linear}([\mathbf{h}_i^{[l-1]} \,||\, \mathbf{c}_i^{[l-1]}])\big) \in \mathbb{R}^3, \\ g_k^{[l]}(\mathbf{H}^{[l-1]}, \mathbf{A}) &= \{\mathbf{g}_{i,k}^{[l]} \mid v_i \in \mathcal{V}\}, \end{aligned} \tag{7}$$

where $\mathbf{g}_{i,k}^{[l]}$ represents the k-th entry of $\mathbf{g}_i^{[l]}$ with $k = \{1, 2, 3\}$ corresponds to graph, semantic, and global experts, respectively, $||$ denotes concatenation and $\text{Linear}(\cdot)$ is a learnable linear transformation. By incorporating both ego features and neighborhood context, the proposed gating model enables more informed selection and aggregation of the malicious cues captured by different experts. Together with the unsupervised training objectives, our context-aware weighting allows the MoE layer to adaptively integrate fraud-indicative cues, thereby providing the flexibility needed to effectively detect camouflaged fraudsters under the diverse and evolving conditions of TAGFD.

To encourage context-dependent expert usage and sparse gating weights, we apply an entropy-based regularization loss on the gating weights $\mathbf{g}_i$ to train the gating layer:

$$\mathcal{L}_{\text{gating}} = \sum_{l} \sum_{k \in \{\text{graph, semantic, global}\}} -\frac{1}{N} \sum_{i=1}^{N} g_{i,k}^{[l]} \log(g_{i,k}^{[l]} + \epsilon), \tag{8}$$

where ϵ is a small constant for numerical stability. Notably, during training, we block the gradient from propagating to earlier layers to ensure that the gating loss only updates the gating network, preventing it from directly updating weights of experts to ensure their specialization. By minimizing $\mathcal{L}_{\text{gating}}$, we encourage the gating distribution to have a sharper distribution, thereby encouraging each node to strategically utilize the relevant experts given its local context.

4.3 Rarity-driven Unsupervised Fraud Detection

While the proposed MoE layers are designed to adaptively integrate fraud-indicative cues, training them for TAGFD in a fully unsupervised setting remains challenging. As evolved fraudsters deliberately camouflage their malicious purpose by mimicking benign semantics, the assumptions underlying traditional detection methods (e.g., affinity assumption [Qiao and Pang, 2023]) no longer hold, thereby rendering them ineffective in handling evolved fraudsters.

To seek alternative supervision signals, the intrinsic rarity of fraudsters in real-world datasets is a stable and persistent property that can serve as a reliable training prior. Concretely, the "rarity assumption" is that fraudulent behaviors can be characterized as deviations from dominant benign patterns, and hence fraudsters can be identified as low-density outliers in the learned normality space. Motivated by this insight, we draw inspiration from one-class (OC) classification to model the dominant patterns of benign nodes with a parameter-free OC fraud detector, thereby exposing the deviations introduced by minority fraudulent nodes. Meanwhile, we design an expert loss that encourages each expert to capture normal patterns, thereby highlighting any cues of fraudsters in the residual deviations.

Parameter-free One-class Fraud Detector

Building on the discriminative embeddings produced by the MoE experts, we employ a parameter-free OC classifier that models normal patterns and identifies deviations. Specifically, the ℓ_2 norm of each embedding serves as a normality measure, which is converted into a bounded fraud score via a sigmoid function $\sigma(\cdot)$, i.e.,

$$s_i = \sigma(\|\mathbf{h}_i^{\text{final}}\|_2), \tag{9}$$

where $\mathbf{h}_i^{\text{final}}$ is the representation of v_i at the final MoE layer. This design provides a lightweight yet effective mechanism to fully leverage the fraud-indicative cues captured by the MoE layers, while also enabling unsupervised detection of subtle fraudulent behaviors with an OC detector.

Under the assumption that fraudsters are a minority, we encourage most fraud scores to approach zero by minimizing

$$\mathcal{L}_{\text{OC}} = \frac{1}{N} \sum_{i=1}^{N} \text{BCE}(s_i, 0), \tag{10}$$

where $\text{BCE}(\cdot, \cdot)$ denotes binary cross-entropy. Optimizing $\mathcal{L}_{\text{OC}}$ pushes the embeddings of normal nodes into a compact hypersphere around the origin while naturally allowing minority fraudsters to stand out due to their larger norms [Wang *et al.*, 2021]. Leveraging the rarity assumption, the OC loss suppresses fraud scores for the majority of nodes while keeping the detector lightweight and easy to optimize. More discussion on the property of OC loss is given in Appendix B.

Expert Loss

Although the OC loss can supervise the whole network at a global level, the functionality of each expert could be unconstrained, which leads to expert redundancy and weak specialization. To further regularize the specialization of experts, we introduce an expert loss to enforce functional usefulness at a fine-grained expert level. As we explicitly design each expert to specialize in a distinct type of fraud-indicative cue, minimizing their corresponding deviation signals can naturally serve as the optimization objective. Since the majority

Type	Method	AUROC (%)				AUPRC (%)			
		Reddit	Instagram	AmazonVideo	YelpChi	Reddit	Instagram	AmazonVideo	YelpChi
GAD	DOMINANT (SDM'19)	62.89±0.09	45.35±0.08	OOM	OOM	20.60±0.04	8.92±0.01	OOM	OOM
	CoLA (TNNLS'21)	58.63±1.81	45.05±0.20	51.06±1.95	48.62±0.16	13.44±0.60	8.95±0.06	12.83±0.56	14.08±0.08
	PREM (ICDM'23)	54.85±4.99	54.38±3.59	58.08±1.91	OOM	12.16±2.31	11.51±1.16	15.06±0.68	OOM
GFD	TAM (NeurIPS'23)	63.07±0.10	47.51±0.06	OOM	OOM	14.76±0.11	10.40±0.07	OOM	OOM
	ADA-GAD (AAAI'24)	64.09±0.02	44.98±0.01	OOM	OOM	**21.78±0.02**	8.87±0.01	OOM	OOM
	HUGE (AAAI'25)	60.64±0.78	44.07±0.17	51.77±0.55	OOM	12.80±0.57	9.00±0.15	12.49±0.17	OOM
TAGAD	TAGAD (arXiv'25)	56.65±0.12	44.83±0.06	51.55±0.12	42.21±0.11	11.64±0.04	9.04±0.02	12.14±0.04	10.78±0.02
	CMUCL (ECAI'25)	60.72±0.15	44.41±0.30	53.99±0.22	OOM	13.52±0.14	8.81±0.11	13.80±0.19	OOM
	CoLL (MM'25)	59.26±0.06	44.02±0.72	50.27±1.26	47.25±0.32	13.22±0.18	8.88±0.20	11.99±0.41	12.50±0.09
TAGFD	CAMERA (Ours)	**65.09±0.04**	**58.21±0.29**	**63.05±0.60**	**61.74±0.04**	18.99±0.04	**14.23±0.54**	**17.37±0.83**	**18.47±0.02**

Table 1: Performance comparison in terms of AUROC and AUPRC, with best results in **bold**. OOM means out-of-memory on a 24GB GPU.

of nodes are normal, this encourages each expert to model the patterns of benign nodes, causing the residual deviations of minority fraudulent nodes to become more pronounced. To maintain expert disentanglement, each loss is applied only to the parameters of the corresponding expert, and gradients are blocked from propagating to earlier layers or the gating network. Formally, the expert loss is computed as the average squared deviation across nodes, experts, and layers:

$$\mathcal{L}_{\text{expert}} = \sum_{l} \sum_{k \in \{\text{graph, semantic, global}\}} \frac{1}{N} \sum_{i=1}^{N} \| e_k^{[l]}(\mathbf{h}_i^{[l-1]}, \mathbf{A}) \|_2^2, \tag{11}$$

where N is the number of nodes, and $e_k^{[l]}(\cdot)$ denotes the residual deviation captured by the k-th expert at layer l. By modeling benign patterns, the experts naturally amplify the subtle malicious signals caused by camouflaged fraudsters, thereby producing discriminative embeddings $\mathbf{h}_i^{\text{final}}$ for the subsequent detector.

The overall training objective is defined by combining $\mathcal{L}_{\text{OC}}$ with $\mathcal{L}_{\text{expert}}$ and $\mathcal{L}_{\text{gating}}$, which is written as:

$$\mathcal{L} = \mathcal{L}_{\text{expert}} + \alpha \mathcal{L}_{\text{gating}} + \beta \mathcal{L}_{\text{OC}}, \tag{12}$$

where α and β are trade-off hyperparameters. The overall algorithm and complexity analysis of CAMERA is given in Appendices C and D, respectively.

5 Experiments

5.1 Experimental Setup

Datasets. We conduct experiments on four public GAD datasets spanning diverse application domains, including Reddit [Li *et al.*, 2024], Instagram [Li *et al.*, 2024], AmazonVideo [McAuley and Leskovec, 2013], and YelpChi [Rayana and Akoglu, 2015]. Following the previous work [Li *et al.*, 2025; Yang *et al.*, 2025a], we construct the text-attributed graph for fraud detection purposes, and convert all graphs to homogeneous undirected graphs for our experiments. Detailed dataset statistics are summarized in Appendix E.

Baseline and Evaluation Metrics. We compare our approach against 9 state-of-the-art methods. CoLA [Liu *et al.*, 2021b], DOMINANT [Ding *et al.*, 2019], and PERM [Pan *et al.*, 2023] represent contrastive learning, reconstruction-based, and affinity-based paradigms for unsupervised graph anomaly detection (GAD). TAM [Qiao and Pang, 2023], ADA-GAD [He *et al.*, 2024], and HUGE [Pan *et al.*, 2025b] focus on unsupervised GFD by pruning heterophily edges to mitigate the impact of camouflaged fraudsters. We also include recent studies that incorporate text attributes (TAGAD), namely TAGAD [Liu *et al.*, 2025], CMUCL [Xu *et al.*, 2025b], and CoLL [Xu *et al.*, 2025a]. To ensure a fair comparison, we use OpenAI's text-embedding-3-small model to encode textual attributes for CAMERA and baselines. We use gpt-4o-mini for TAGAD methods that incorporate LLM. The area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPRC) are used as the evaluation metrics. Average results with standard deviations are reported over five runs with different random seeds. The implementation detailsand more experiments are presented in Appendices E and F, respectively.

Method	Reddit	Instagram	AmazonVideo	YelpChi
OpenAI (Ours)	**65.09±0.04**	**58.21±0.29**	**63.05±0.60**	**61.74±0.04**
SentenceBERT	61.45±0.10	53.23±1.67	53.95±3.82	53.79±0.16
BoW	48.88±0.04	51.52±0.08	51.58±0.05	39.93±0.04

Table 2: Ablation study on text encoder. Result in AUROC (%).

5.2 Experimental Results

Performance Comparison

The comparison results of CAMERA is illustrated in Table 1. From the table, we make the following key observations: ❶ CAMERA outperforms SOTA methods in most datasets, with the only exception being AUPRC on Reddit. These results demonstrate the effectiveness of CAMERA in diverse real-world scenarios. ❷ Among the baselines, methods that address structure camouflage by pruning heterophilic edges (TAM, ADA-GAD, HUGE) achieve strong performance on Reddit but perform poorly on other datasets such as Instagram, revealing the limitation of static assumptions that fail to capture evolving fraudster characteristics. In contrast, the adaptive fraud-indicative cues integration strategy of CAMERA allows it to automatically learn and exploit the most informative malicious signal, thereby effectively addressing evolving camouflaging strategies. ❸ The scalability of existing unsupervised GFD methods is limited by out-of-memory (OOM) issues when applied to large-scale real-world datasets such as AmazonVideo and YelpChi . In contrast, CAMERA demonstrates better scalability, underscoring its effec-

Method	Reddit	Instagram	AmazonVideo	YelpChi
Number of Experts				
1 Expert	56.60	56.19	58.66	59.39
2 Experts	58.72	56.91	59.34	60.15
3 Experts (Ours)	**65.09**	**58.21**	**63.05**	**61.74**
Gating Mechanism				
Uniform weighting	63.08	57.43	59.40	61.19
Ego only (w/o context)	64.89	57.13	59.65	61.44
Context-informed (Ours)	**65.09**	**58.21**	**63.05**	**61.74**

Table 3: Ablation studies on #experts and the gating mechanism.

Experts			AUROC (%)			
Graph	Global	Semantic	Reddit	Instagram	AmazonVideo	YelpChi
✓	✓	✓	**65.09±0.04**	**58.21±0.29**	**63.05±0.60**	**61.74±0.04**
–	✓	✓	55.74±0.51	56.55±1.55	58.97±3.37	60.32±0.17
✓	–	✓	59.96±1.50	56.52±0.57	57.69±5.59	60.24±0.18
✓	✓	–	60.46±1.71	57.65±0.34	61.35±1.38	59.90±0.07
✓	–	–	61.40±0.34	57.95±0.47	61.35±1.56	60.29±0.06
–	✓	–	54.00±0.33	57.06±0.35	60.00±0.53	58.14±0.35
–	–	✓	54.41±0.29	53.55±1.98	54.64±7.80	59.74±0.22

Table 4: Ablation study on expert participation.

tiveness in safeguarding large-scale real-world graph applications like e-commerce and social networks. ❹ Although existing TAGAD incorporate large language models to enhance feature alignment [Xu *et al.*, 2025b] or to generate auxiliary evidence [Xu *et al.*, 2025a], they achieve limited performance on the TAGFD task. This is because malicious intent is often concealed by semantic camouflage employed by fraudsters, making it substantially harder to identify than anomalies in constructed benchmark datasets [Ma *et al.*, 2021].

Ablation Study

To examine the contribution of key design in CAMERA, we conduct ablation studies on critical model components.

Textual Feature Encoder. We replace the LLM-based feature encoder with bag-of-words (BoW) and SentenceBERT representations [Reimers and Gurevych, 2019] to assess the impact of semantic representation quality. As shown in Table 2, the LLM-based encoder consistently achieves the highest detection performance. In contrast, although bag-of-words representations are the most commonly adopted choice in GAD [Pan *et al.*, 2023], they fail to capture the critical semantic information required for TAGFD. These findings confirm that rich semantic embeddings are critical for detecting camouflaged fraudsters, aligning with our method's emphasis on preserving subtle textual cues in an unsupervised setting.

Specialized Experts. To quantify the contribution of each expert, we compare the full model with variants using only one or two experts. Table 4 shows that the full MoE achieves the best overall performance. Our additional experiments in Table 3 further demonstrate that adding more experts consistently improves the results. Together, these results demonstrate that successful TAGFD demands diverse fraud-indicative cues. Among the experts, the graph expert contributes most significantly, highlighting the importance of local structural deviations.

Context-informed Gating. We evaluate the gating model by replacing it with two alternatives: uniform weighting and ego-only gating. As shown in Table 3, incorporating local neighborhood context consistently improves performance across all datasets, highlighting the importance of incorporating community-specific information for weighting experts. In contrast, uniform weighting leads to overall inferior results. Together, the experiment findings confirm that adaptively integrating fraud-indicative cues is critical for identifying evolved, camouflaged fraudsters.

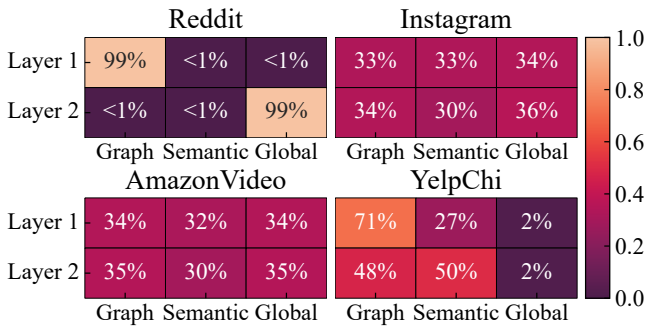


Figure 4: Visualization of expert weight.

Visualization of Expert Allocation

To investigate how CAMERA adaptively integrates fraud-indicative cues across different scenarios, we visualize the gating weights at the dataset and case levels.

Dataset-level visualization is shown in Figure 4. On Instagram and AmazonVideo, the three experts receive relatively balanced weights. However, as illustrated in Table 3, replacing them with uniform weights results in a significant performance drop, which shows that the gating model also provides sample-specific fraud-indicative cues integration. On YelpChi and Reddit, the model demonstrates sparse activation. Specifically, YelpChi primarily relies on graph and semantic cues, while Reddit emphasizes the graph expert at layer 1 and the global expert at layer 2. This visualization result also shows that CAMERA performs hierarchical cue integration through expert specialization at different layers. An additional case study on YelpChi can be found in Appendix F, confirming that CAMERA can achieve adaptive integration of fraud-indicative cues.

6 Conclusion

In this paper, we propose a novel unsupervised TAGFD framework, CAMERA, that enables detection against evolved fraudsters that engage in semantic camouflage to mimic benign behaviors and evade detection. By utilizing ego-decoupled MoE architecture together with a one-class unsupervised training objective, CAMERA can adaptively integrate different fraud-indicative cues to address evolving camouflaging strategies. Comprehensive empirical results show the effectiveness and robustness of CAMERA in unsupervised TAGFD, validating CAMERA as a practical solution for detecting evolved fraudsters in real-world scenarios.

Acknowledgements

The work of S. Pan was partially supported by the Australian Research Council (ARC) under Grant Nos. DP240101547

and FT210100097. The work of Y. Liu was partially supported by the ARC under Grant No. DE260101172.

References

[Cai *et al.*, 2025] Tingyi Cai, Yunliang Jiang, Yixin Liu, Ming Li, Changqin Huang, and Shirui Pan. Out-of-distribution detection on graphs: A survey. *arXiv preprint arXiv:2502.08105*, 2025.

[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.

[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. In *The Thirteenth International Conference on Learning Representations*, 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.

[Fedus *et al.*, 2022] William Fedus, Barret Zoph, and Noam Shazeer. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. *Journal of Machine Learning Research*, 23(120):1–39, 2022.

[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.

[Gururangan *et al.*, 2022] Suchin Gururangan, Mike Lewis, Ari Holtzman, Noah A Smith, and Luke Zettlemoyer. Demix layers: Disentangling domains for modular language modeling. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5557–5576, 2022.

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

[Hu *et al.*, 2023] Xinxin Hu, Haotian Chen, Hongchang Chen, Shuxin Liu, Xing Li, Shibo Zhang, Yahui Wang, and Xiangyang Xue. Cost-sensitive gnn-based imbalanced learning for mobile social network fraud detection. *IEEE Transactions on Computational Social Systems*, 11(2):2675–2690, 2023.

[Huang *et al.*, 2023] Yihong Huang, Liping Wang, Fan Zhang, and Xuemin Lin. Unsupervised graph outlier detection: Problem revisit, new insight, and superior method. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, pages 2565–2578. IEEE, 2023.

[Jin *et al.*, 2021] Ming Jin, Yixin Liu, Yu Zheng, Lianhua Chi, Yuan-Fang Li, and Shirui Pan. Anemone: Graph anomaly detection with multi-scale contrastive learning. In *Proceedings of the 30th ACM international conference on information & knowledge management*, pages 3122–3126, 2021.

[Lewis *et al.*, 2021] Mike Lewis, Shruti Bhosale, Tim Dettmers, Naman Goyal, and Luke Zettlemoyer. Base layers: Simplifying training of large, sparse models. In *International Conference on Machine Learning*, pages 6265–6274. PMLR, 2021.

[Li *et al.*, 2024] Yuhan Li, Peisong Wang, Xiao Zhu, Aochuan Chen, Haiyun Jiang, Deng Cai, Victor W Chan, and Jia Li. Glbench: A comprehensive benchmark for graph with large language models. *Advances in Neural Information Processing Systems*, 37:42349–42368, 2024.

[Li *et al.*, 2025] Yuan Li, Jun Hu, Bryan Hooi, Bingsheng He, and Cheng Chen. Dgp: A dual-granularity prompting framework for fraud detection with graph-enhanced llms. *arXiv preprint arXiv:2507.21653*, 2025.

[Li *et al.*, 2026] Shiyuan Li, Yixin Liu, Yu Zheng, Xiaofeng Cao, Shirui Pan, and Heng Tao Shen. Towards one-for-all anomaly detection for tabular data. In *International Conference on Machine Learning (ICML)*, 2026.

[Liu *et al.*, 2021a] Yang Liu, Xiang Ao, Zidi Qin, Jianfeng Chi, Jinghua Feng, Hao Yang, and Qing He. Pick and choose: a gnn-based imbalanced learning approach for fraud detection. In *Proceedings of the web conference 2021*, pages 3168–3177, 2021.

[Liu *et al.*, 2021b] 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] Boan Liu, Liang Ding, Li Shen, Keqin Peng, Yu Cao, Dazhao Cheng, and Dacheng Tao. Diversifying the mixture-of-experts representation for language models with orthogonal optimizer. In *Proceedings of the European Conference on Artificial Intelligence (ECAI)*. IOS Press, 2023.

[Liu *et al.*, 2025] Xudong Liu, Yanan Ren, Hengtong Zhang, Run-An Wang, Shenghe Zheng, and Zhaonian Zou. Towards anomaly detection on text-attributed graphs. https://openreview.net/forum?id=LMKYd9JHgU, 2025. OpenReview preprint.

[Liu *et al.*, 2026] Yixin Liu, Shiyuan Li, Yu Zheng, Qingfeng Chen, Chengqi Zhang, Philip S Yu, and Shirui Pan. From few-shot to zero-shot: Towards generalist graph anomaly detection. *arXiv preprint arXiv:2602.18793*, 2026.

[Ma *et al.*, 2021] Xiaoxiao Ma, Jia Wu, Shan Xue, Jian Yang, Chuan Zhou, Quan Z Sheng, Hui Xiong, and Leman Akoglu. A comprehensive survey on graph anomaly detection with deep learning. *IEEE transactions on knowledge and data engineering*, 35(12):12012–12038, 2021.

[McAuley and Leskovec, 2013] Julian John McAuley and Jure Leskovec. From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews. In *Proceedings of the 22nd international conference on World Wide Web*, pages 897–908, 2013.

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

[Pan *et al.*, 2025a] J. Pan, Y. Zheng, Y. Tan, and Y. Liu. A survey of generalization of graph anomaly detection: From transfer learning to foundation models. In *2025 IEEE International Conference on Knowledge Graph (ICKG)*, pages 316–323. IEEE, 2025.

[Pan *et al.*, 2025b] Junjun Pan, Yixin Liu, Xin Zheng, Yizhen Zheng, Alan Wee-Chung Liew, Fuyi Li, and Shirui Pan. A label-free heterophily-guided approach for unsupervised graph fraud detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 12443–12451, 2025.

[Pan *et al.*, 2026a] Junjun Pan, Yixin Liu, Rui Miao, Kaize Ding, Yu Zheng, Quoc Viet Hung Nguyen, Alan Wee-Chung Liew, and Shirui Pan. Explainable and fine-grained safeguarding of llm multi-agent systems via bi-level graph anomaly detection. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics*, 2026.

[Pan *et al.*, 2026b] Junjun Pan, Yixin Liu, Chuan Zhou, Fei Xiong, Alan Wee-Chung Liew, and Shirui Pan. Correcting false alarms from unseen: Adapting graph anomaly detectors at test time. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2026.

[Qian *et al.*, 2026] Yanyu Qian, Yue Tan, Yixin Liu, Wang Yu, and Shirui Pan. Dynhd: Hallucination detection for diffusion large language models via denoising dynamics deviation learning. *arXiv preprint arXiv:2603.16459*, 2026.

[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.

[Rayana and Akoglu, 2015] Shebuti Rayana and Leman Akoglu. Collective opinion spam detection: Bridging review networks and metadata. In *Proceedings of the 21th acm sigkdd international conference on knowledge discovery and data mining*, pages 985–994, 2015.

[Reimers and Gurevych, 2019] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*, 2019.

[Rusch *et al.*, 2023] T Konstantin Rusch, Michael M Bronstein, and Siddhartha Mishra. A survey on oversmoothing in graph neural networks. *arXiv preprint arXiv:2303.10993*, 2023.

[Tan *et al.*, 2025] Yue Tan, Xiaoqian Hu, Hao Xue, Celso De Melo, and Flora Salim. Bisecle: Binding and separation in continual learning for video language understanding. *Advances in Neural Information Processing Systems*, 38:33752–33782, 2025.

[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.

[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.

[Xu *et al.*, 2025a] Yiming Xu, Jiarun Chen, Zhen Peng, Zihan Chen, Qika Lin, Lan Ma, Bin Shi, and Bo Dong. Court of llms: Evidence-augmented generation via multi-llm collaboration for text-attributed graph anomaly detection. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pages 2437–2446, 2025.

[Xu *et al.*, 2025b] Yiming Xu, Xu Hua, Zhen Peng, Bin Shi, Jiarun Chen, Xingbo Fu, Song Wang, and Bo Dong. Text-attributed graph anomaly detection via multi-scale cross-and uni-modal contrastive learning. *arXiv preprint arXiv:2508.00513*, 2025.

[Yang *et al.*, 2025a] Chengdong Yang, Hongrui Liu, Daixin Wang, Zhiqiang Zhang, Cheng Yang, and Chuan Shi. Flag: Fraud detection with llm-enhanced graph neural network. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pages 5150–5160, 2025.

[Yang *et al.*, 2025b] Jie Yang, Rui Zhang, Ziyang Cheng, Dawei Cheng, Guang Yang, and Bo Wang. Grad: Guided relation diffusion generation for graph augmentation in graph fraud detection. In *Proceedings of the ACM on Web Conference 2025*, pages 5308–5319, 2025.

[Yu *et al.*, 2023] Jianke Yu, Hanchen Wang, Xiaoyang Wang, Zhao Li, Lu Qin, Wenjie Zhang, Jian Liao, and Ying Zhang. Group-based fraud detection network on e-commerce platforms. In *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining*, pages 5463–5475, 2023.

[Zhao *et al.*, 2025] Yunfeng Zhao, Yixin Liu, Shiyuan Li, Qingfeng Chen, Yu Zheng, and Shirui Pan. Freegad: A training-free yet effective approach for graph anomaly detection. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*, pages 4379–4389, 2025.

[Zhao *et al.*, 2026] Yunfeng Zhao, Yixin Liu, Qingfeng Chen, Shiyuan Li, Yue Tan, and Shirui Pan. Fedcigar: A personalized reconstruction approach for federated graph-level anomaly detection. In *International Joint Conference on Artificial Intelligence*, 2026.