Proceedings of the Thirty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-26)
Special Track on AI and Social Good

# Deep Reinforcement Learning Enhanced Semi-supervised Graph Neural Network for Credit Card Fraud Detection

**Huilin He**[1] , **Kun Zhu**[1] , **Zewen Hu**[1] , **Jie Wang**[1] and **Dawei Cheng**[1,*]
[1]School of Computer Science and Technology, Tongji University, Shanghai, China
{huilin3, kzhu00, 2251988, wangjie_tongji, dcheng}@tongji.edu.cn

## Abstract

Credit card fraud threatens global payment ecosystems, causing billions in losses and undermining public trust. Efficient fraud detection remains challenging due to surging transaction volumes and evolving tactics. While Graph Neural Networks (GNNs) excel at modeling structural relationships, they struggle in real-world scenarios characterized by label scarcity and often overlook discriminative feature-level signals, leaving rich risk signals underutilized without costly manual engineering. To address this, we propose DRESS, a Deep Reinforcement Learning (DRL) Enhanced Semi-supervised GNN framework. It employs a DRL agent to automatically capture and enhance feature-level risks, fusing them with graph-based structural risks and propagating via a gated temporal attention network for final prediction. To mitigate inefficient exploration of the DRL module, we incorporate a feature self-attention layer to weigh feature contributions to fraud detection and employ self-supervised intrinsic rewards to help optimize the DRL module efficiently. Extensive experiments on real-world datasets demonstrate that DRESS outperforms state-of-the-art methods, especially in low-label scenarios with only 2%–10% labeled samples. By empowering resource-limited institutions to combat fraud and prevent financial loss, DRESS secures the digital trust essential for inclusive growth, contributing to AI for poverty alleviation and economic development.

## 1 Introduction

With the rapid development of financial technologies, transactions have become increasingly frequent and complex, while fraud grows more covert and diverse [AlFalahi and Nobanee, 2024]. As a major financial crime, credit card fraud—involving unauthorized use of funds via theft, skimming, or phishing—has evolved into organized, concealed operations, making detection difficult [PricewaterhouseCoopers, 2024]. Causing billions in annual global losses, these activities severely undermine user confidence and financial security. Consequently, effective fraud detection has become an urgent task for financial institutions.

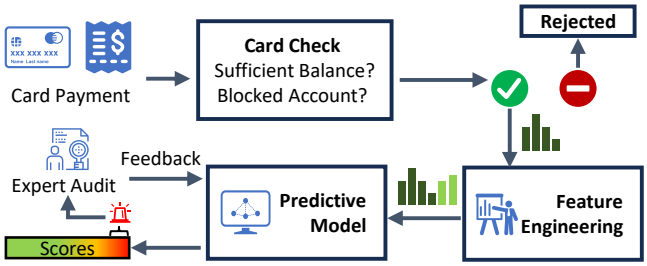


Figure 1: Typical framework of credit card fraud detection.

A typical framework for credit card fraud detection [Cheng *et al.*, 2020a] is shown in Figure 1. A key part of the framework is the predictive model, which has evolved from rule-based systems [K.R. and Zareapoor, 2014] and traditional machine learning [Fu *et al.*, 2016; Hu *et al.*, 2019] to modern deep learning approaches [Liu *et al.*, 2020]. Recently, Graph Neural Networks (GNNs) have emerged as a powerful paradigm for modeling relational data and have achieved promising results in fraud detection [Yu *et al.*, 2025; Liu *et al.*, 2026]. In credit card transaction networks, transactions are represented as nodes and their correlation as edges. GNNs propagate and aggregate information from neighboring nodes to improve node representations.

In real-world settings, labeled data constitute only a small fraction of transactions, while abundant unlabeled data can be used to improve detection [Tian *et al.*, 2023]. To address this, semi-supervised GNNs have been proposed, which leverage unlabeled data by exploiting the similarity and correlations between nodes [Yu *et al.*, 2024]. Meanwhile, as illustrated in Figure 1, feature engineering also plays a critical role in the detection pipeline. It primarily involves exploring risk signals from raw features and generating new features for each transaction (i.e. each node) —a process that is highly valuable yet labor-intensive [Nargesian *et al.*, 2017]. Thus, it would be beneficial for a model to derive additional features that contain risk information from raw features autonomously. However, existing semi-supervised GNNs predominantly focus on propagating risk signals between nodes, partially neglecting intra-node feature fraud patterns. This limits their ability to

∗ Corresponding Author is Dawei Cheng.

detect subtle yet critical fraudulent transactions that do not exhibit strong relational signals.

Recently, Deep Reinforcement Learning (DRL) has gained attention in semi-supervised fraud detection, owing to its ability to learn optimal decision-making policies via interaction with the environment [Bei *et al.*, 2023; Li *et al.*, 2024]. Successful applications, such as CARE-GNN [Dou *et al.*, 2020], utilize RL to dynamically filter structural neighbors for aggregation. To address the aforementioned limitation of underutilized feature-level patterns, we extend this intuition to feature space. Here, DRL is employed to autonomously learn an optimal derivation scheme via trial-and-error, constructing risk-enriched representations from raw attributes to enhance intra-node information for GNN propagation. Existing DRL-based approaches operating in the feature space typically model node features as states and leverage sparse labels to construct reward signals [Zha *et al.*, 2020]. However, these methods face two constraints that hinder their practical effectiveness: (1) they treat all features equally, ignoring their differential contributions to fraud detection [Pang *et al.*, 2021], leading to inefficient action space exploration; (2) their reward designs heavily rely on labels [Zha *et al.*, 2020], resulting in weak reward signals in real-world settings where labeled samples are extremely limited.

To address these challenges, we propose DRESS, Deep Reinforcement Learning Enhanced Semi- supervised Graph Neural Network for Fraud Detection, a novel framework that combines GNNs with deep reinforcement learning to better leverage unlabeled data. We first introduce a feature risk extractor, employing the Deep Q-Network (DQN) [Mnih *et al.*, 2015] framework to automatically extract and amplify fraud signals embedded within node features. Recognizing that different features contribute unevenly to fraud detection, we further incorporate a feature self-attention layer to learn feature-specific contributions, based on which we construct a self-supervised intrinsic reward to enhance the efficiency of DQN in risk information extraction. Meanwhile, inspired by RGTAN [Xiang *et al.*, 2025], we acknowledge the risk-related information inherent in graph structures and introduce a graph-based risk information extractor to quantify node degree and risk neighbor count. Finally, feature-derived risks, graph-structural risks, and node features are fused through a gated temporal attention network (TGAT) [Xiang *et al.*, 2023] for message passing and node classification. Through our proposed framework, the fraud signals in all nodes—particularly the abundant unlabeled ones—are effectively amplified. This enhancement enables more efficient utilization of unlabeled node information for neighbor aggregation and fraud detection tasks.

Our contributions are summarized as follows:

- We propose DRESS, a deep reinforcement learning enhanced semi-supervised GNN that automatically explores and amplifies intrinsic risk patterns within node features, addressing under-utilization of unlabeled samples in GNNs to improve fraud detection performance.
- We introduce a feature self-attention layer and an intrinsic reward to accelerate RL policy learning under label scarcity. Empirical results demonstrate 37.5%-51% training efficiency gains through this mechanism.
- Extensive experiments on real-world datasets demonstrate DRESS outperforms state-of-the-art baselines, and remains superior performance in extreme semi-supervised scenarios with only 2%–10% labeled data.
- Conducted in collaboration with a leading financial institution, this work incorporates a dataset spanning 10 months of transactions. Validated by the partner's fraud analysts and data scientists for effectiveness and deployability, our approach lowers barriers to advanced AI safeguards, offering practical technical support to safeguard user assets and ensure digital ecosystem stability.

## 2 Related Work

### 2.1 Semi-supervised GNNs for Fraud Detection

Graph Neural Networks (GNNs) have excelled in fraud detection by modeling transaction structures and filtering malicious neighbors [Liu *et al.*, 2021; Zou and Cheng, 2024; Duan *et al.*, 2024; Liu *et al.*, 2025; Wu *et al.*, 2025]. To mitigate real-world label scarcity, semi-supervised variants were developed to facilitate risk propagation across unlabeled data: SemiGNN [Wang *et al.*, 2019] employs multi-view attention to link semantically similar nodes; GTAN [Xiang *et al.*, 2023] utilizes attribute similarity to guide risk flow; and RGTAN [Xiang *et al.*, 2025] further incorporates risk-aware structural attributes. However, these approaches predominantly focus on inter-node structural propagation, often overlooking the discriminative intra-node feature signals embedded within individual transactions, leading to insufficient exploitation of abundant unlabeled data. Although DevNet [Pang *et al.*, 2019] considers feature-level risks, it relies on rigid distributional assumptions. In contrast, our method DRESS addresses these limitations by integrating structural risk propagation with a deep reinforcement learning mechanism that autonomously discovers and amplifies feature-inherent risk patterns. This combined approach enhances semi-supervised GNNs without prior constraints, enabling more comprehensive utilization of unlabeled data.

### 2.2 DRL for Semi-supervised Fraud Detection

While semi-supervised GNNs effectively propagate risk, their aggregation often obscures discriminative intra-node patterns. Deep Reinforcement Learning (DRL), with its label-free policy optimization, circumvents the structural reliance of traditional GNNs, enabling dynamic feature exploration that is particularly suitable for label-scarce scenarios [Peng *et al.*, 2024; Heidari *et al.*, 2024]. Several studies have integrated DRL into semi-supervised anomaly detection through various mechanisms [Zha *et al.*, 2020; Kazari *et al.*, 2023; Li *et al.*, 2024; Zhang *et al.*, 2024]. However, these methods are primarily designed for general anomaly detection and have not been adequately tailored for financial fraud detection. They fail to model the relationships between transactions, but they do show promise in exploring risk signals within individual transactions. A representative example is DPLAN [Pang *et al.*, 2021], a DRL method for semi-supervised anomaly detection that integrates external rewards

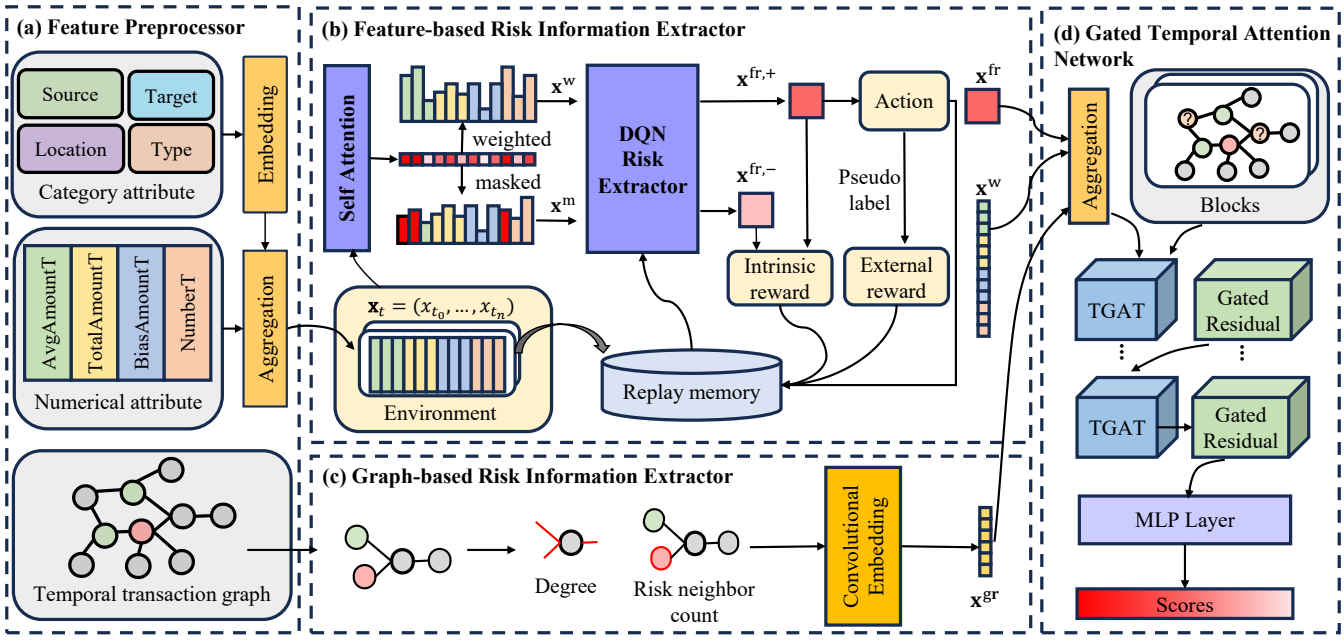


Figure 2: Overview of DRESS. (a) Feature preprocessing. (b) Extract feature risk via DRL, enhanced by feature self-attention and intrinsic rewards to amplify risk-indicative features ($\mathbf{x}^{\mathrm{w}}$) and improve efficiency. (c) Extract graph-based risk (node degree and risk neighbor count). (d) Integrate extracted risks and weighted node features into the gated temporal attention network for propagation and fraud prediction.

with self-supervision from isolation forests applied to features. Yet, it treats all features equally, neglecting their varying importance in indicating fraud. When adapted to fraud detection, existing DRL methods encounter two critical issues: sparse rewards due to limited labels, and ineffective feature exploration that ignores discriminative risk attributes. To overcome these challenges, our framework DRESS introduces a feature self-attention mechanism to adaptively prioritize informative features, alongside feature-aware self-supervised intrinsic rewards. This joint design not only alleviates reward sparsity but also accelerates policy learning by providing denser and more directed learning signals—enabling the DRL agent to focus on meaningful feature combinations and avoid inefficient exploration.

## 3 Method

### 3.1 Overview

**Preliminary.** Based on the temporal aggregation of fraud patterns [Cheng *et al.*, 2020b], we construct a temporal transaction graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$. Here, $\mathcal{V} = \{v_1, ..., v_N\}$ represents transactions, and $\mathbf{X} \in \mathbb{R}^{N \times d}$ denotes the attribute matrix. Edges in $\mathcal{E}$ connect transactions sharing properties (e.g., same cardholder) within a temporal window to capture potentially coordinated activities. The label set is $Y = \{y_1, \ldots, y_N\}$, where $y_i \in \{0, 1, 2\}$ corresponds to normal, fraudulent, and unlabeled nodes, respectively. Our objective is to maximize the probability nodes being classified correctly given graph $\mathcal{G}$, i.e. $\max_\theta P_\theta(Y|\mathcal{G}, \mathbf{X})$.

**Overview of Framework.** As illustrated in Figure 2 , our framework consists of four components. First, we preprocess the features by embedding categorical attributes and aggregating them with numerical attributes. Then, to automatically explore and enhance risk signals within node features, in the feature-based risk information extractor, we employ a Deep Q-Network (DQN) framework for learning risk information from node features. Besides, recognizing the varying importance of different features for fraud detection, we introduce a feature self-attention layer to learn feature contribution weights, based on which we generate a feature pair—weighted features and masked features, and construct a self-supervised intrinsic reward to accelerate DQN's convergence toward effective risk extraction strategies. Subsequently, we extract structural risk information from the graph topology through node degree and risky neighbor counting. Finally, we concatenate the feature-derived risk signals, structural risk information, and weighted node features for graph propagation and final prediction.

### 3.2 DRL-based Feature Risk Extractor

Effectively capturing subtle yet critical fraud signals from raw transaction features is beneficial for identifying concealed fraudulent activities in label-scarce settings. To autonomously discover and amplify these intrinsic risk patterns, we introduce a DQN-based feature risk extractor (Figure 2(b)). Within this deep reinforcement learning setup, the environment consists of preprocessed transaction data $\mathcal{D} = (\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_n), \mathbf{x}_i = (x_{i_1}, ..., x_{i_d})$, where $x_{i_j}$ is the $j$-th feature of the transaction $\mathbf{x}_i$. The state $\mathbf{s}$ is a batch of transaction data sampled from $\mathcal{D}$, and the action $a$ involves selecting pseudo-labels based on risk information derived from transaction features via the risk extractor. Consequently, the action space is $\{0, 1\}$, i.e., $a \in \{0, 1\}$. Rewards include both extrinsic rewards $r^e$ and intrinsic rewards $r^i$.

The framework works as follows: (1) **State Sampling**. At each time step $t$, a batch of transaction data are sampled from $\mathcal{D}$ to form the current state $\mathbf{s}_t = (\mathbf{x}_{t_1}, ..., \mathbf{x}_{t_m})$. (2) **Policy Generation**. The feature self-attention layer (detailed in Sec. 3.3) first generates weighted features $\mathbf{x}^{\text{w}}$, which are then processed by the Q-network. This network comprises two distinct MLPs: the $Q_{agent}$ network for adaptive exploration, and the $Q_{target}$ network for stable prediction. The output of Q-network is the extracted feature risk information $\mathbf{x}^{\text{fr}} \in \mathbb{R}^{m\times 2}$, which encodes both benign and risk scores.
(3) **Action Taking**. Actions $a_{t_i}$ are determined by risk scores $p$ (from the feature risk information $\mathbf{x}^{\text{fr}}$), and a F1-optimized dynamic threshold $\epsilon$, which can be formulated as follows:

$$a_{t_i} = \begin{cases} 1, & \text{p}>\epsilon \\ 0, & \text{p}\leq \epsilon \end{cases} \tag{1}$$

(4) **Reward Calculating**. The reward $r = r^e + r^i$ combines intrinsic reward $r^i$ which will be detailed later and extrinsic reward $r^e$ which is calculated as follows:

$$r^e = \frac{1}{2}\left(\frac{\text{TP}}{\text{TP}+\text{FP}} + \frac{\text{TP}}{\text{TP}+\text{FN}}\right) \tag{2}$$

where $\text{TP} = \sum_{i=1}^{m} I(a_{t_i} = 1 \wedge y_{t_i} = 1)$, $\text{FN} = \sum_{i=1}^{m} I(a_{t_i} = 0 \wedge y_{t_i} = 1)$, $\text{FP} = \sum_{i=1}^{m} I(a_{t_i} = 1 \wedge (y_{t_i} = 0 \vee y_{t_i} = 2))$. The extrinsic reward is computed as the average of precision and recall, guiding the model to balance its ability to identify positive and negative samples. In calculating FP, we adopt a Positive-Unlabeled (PU) learning strategy [Elkan and Noto, 2008]. Given the extreme rarity of fraud, we treat the unlabeled pool as the background distribution. Consequently, predicting fraud within this set reduces immediate extrinsic reward, acting as a regularizer. This imposes a strict selectivity constraint, encouraging the agent to distinguish definitive fraud patterns from the background noise rather than blindly expanding predictions.
(5) **Experience Replay and Policy Update**. The next state $\mathbf{s}_{t+1}$ is sampled from $\mathcal{D}$, and the transition tuple $(\mathbf{s}_t, \mathbf{a}_t, r, \mathbf{s}_{t+1})$ is stored in the experience replay buffer. To update the policy, we randomly sample an experience $(\mathbf{s}_t, \mathbf{a}_t, r, \mathbf{s}_{t+1})$ from the buffer. The state $\mathbf{s}_t$ in it is fed into $Q_{\text{agent}}$ to obtain risk information $\mathbf{x}_t^{\text{fr}}$, which encodes the Q-values for each action. From it, the column corresponding to action $\mathbf{a}_t$ is used as the state-action value, i.e. $V_t = Q_{\text{agent}}(\mathbf{s}_t, \mathbf{a}_t) = \mathbf{x}_t^{\text{fr},\mathbf{a}_t}$. For $\mathbf{s}_{t+1}$, the next state's action value $V_{t+1} = \max_{\mathbf{a}'} Q_{\text{target}}(\mathbf{s}_{t+1}, \mathbf{a}')$. The loss $\mathcal{L}_{DRL}$ is then computed to update $Q_{\text{agent}}$, and the parameters of $Q_{\text{agent}}$ are copied to $Q_{\text{target}}$ every $z$ steps. $\mathcal{L}_{DRL}$ is formulated as:

$$\mathcal{L}_{\text{DRL}} = \frac{1}{m}\sum_{i=1}^{m}(r + \gamma V_{(t+1)_i} - V_{t_i})^2 \tag{3}$$

Here, $r+\gamma V_{t+1}$ represents the target value based on the maximum expected future return of the next state, while $V_t$ is the current estimated value. The difference between them reflects the gap between the current policy and the optimal policy. Minimizing this MSE makes the Q-network approximate the Bellman-optimal value function and incorporating future reward expectations is crucial for detecting temporally aggregated fraud patterns, where the discount factor $\gamma$ regulates the weight given to future transaction rewards.

### 3.3 Feature Self-attention and Intrinsic Reward

To enhance the efficiency of DRL-based feature risk extraction, we address the varying contributions of features by introducing a feature self-attention layer, which prioritizes critical fraud indicators to guide the agent's focus. This layer, implemented as an MLP, adaptively learns feature attention weights $\mathbf{w} = (w_1, ..., w_d)$, where $d$ is the number of features. Higher attention weights indicate greater contributions to fraud detection. Based on these weights, we design a self-supervised contrastive task using a feature pair: weighted features $\mathbf{x}_{t_i}^{\text{w}} = \mathbf{x}_{t_i}\mathbf{w}$ (signal amplified) and masked features $\mathbf{x}_{t_i}^{\text{m}}$ (signal suppressed). The masked features $\mathbf{x}_{t_i}^{\text{m}}$ are generated by replacing the top-$k$ highest-attention features with their mean values. To ensure differentiability, we employ a soft masking mechanism, which is formalized as:

$$\mathbf{m} = \sigma\left(\frac{\mathbf{w} - \theta_k}{\tau}\right) \tag{4}$$

$$\mathbf{x}_{t_i}^{\text{m}} = \mathbf{x}_{t_i} * (1 - \mathbf{m}) + \mathbf{x}_{t_i}^{\text{mean}} * \mathbf{m} \tag{5}$$

where $\mathbf{m}$ is the soft mask vector, $\theta_k$ the the top-k proportion attention threshold, $\tau$ a temperature coefficient controlling the steepness of masking, and $\mathbf{x}_{t_i}^{\text{mean}}$ the feature-wise mean vector. After this masking, high-attention (fraud-indicative) features in $\mathbf{x}_{t_i}^{\text{m}}$ are smoothed, while low-attention features remain largely unchanged. This results in a suppressed risk score $\mathbf{x}_{t_i}^{\text{fr},-}$. Conversely, weighted features $\mathbf{x}_{t_i}^{\text{w}}$ amplify risk signals, creating an enhanced risk score $\mathbf{x}_{t_i}^{\text{fr},+}$. To help the DQN extractor distinguish benign and risky patterns, we maximize the divergence between $\mathbf{x}_{t_i}^{\text{fr},-}$ and $\mathbf{x}_{t_i}^{\text{fr},+}$ using a pairwise loss [Ouyang *et al.*, 2022] as the intrinsic reward:

$$r^i = -\frac{1}{m}\sum_{i=1}^{m}\log(1 + \exp(\mathbf{x}_{t_i}^{\text{fr},-} - \mathbf{x}_{t_i}^{\text{fr},+})) \tag{6}$$

We only use fraudulent samples to compute intrinsic rewards. First, for normal samples, amplifying the divergence between enhanced and masked risk scores is unnecessary as they reflect legitimate patterns. Second, this focus prevents reward dilution by normal-sample noise, thereby enhancing the discriminative power of limited fraudulent samples' features. Ultimately, the intrinsic rewards guide DQN to find effective risk extraction strategies faster, improving training efficiency. In parallel, the feature self-attention layer works globally to highlight risk signals and suppress noise across all samples, thereby exposing latent fraud patterns for GNN propagation.

### 3.4 Risk Propagation via Graph

This module extracts structural patterns and propagates risk signals to detect coordinated fraud activities within complex credit card transaction networks.

**Graph-based Risk Extractor.** Graph structures inherently encapsulate fraud signals [Xiang *et al.*, 2025]. Compared to legitimate transactions, fraudulent transactions exhibit a higher proportion of malicious neighbors. For each transaction $l_i$, we extract two structural features: degree feature $f_{\text{degree},g}^i$ and risk neighbor count $f_{\text{risk},g}^i$, where $g$ denotes the

$g$-hop neighborhood. These features are formulated as:

$$f^i_{\text{degree},g} = \sum_{u \in \mathcal{N}^i_g} \deg(u) \tag{7}$$

$$f^i_{\text{risk},g} = \sum_{u \in \mathcal{N}^i_g, y_u = 1} y_u \tag{8}$$

where $\mathcal{N}^i_g$ represents the $g$-hop neighbors of node $l_i$, and $\deg(u)$ denotes the degree of node $u$. These features are processed via a convolutional embedding layer [Xiang *et al.*, 2025] to derive the graph-structural risk representation $\mathbf{x}^{\text{gr}}$.

**Gated Temporal Graph Attention Mechanism.** We perform message passing and node representation updates on the constructed temporal transaction graph using a gated temporal attention mechanism (TGAT) [Xiang *et al.*, 2023]. The input $\mathbf{x}_{t_i}$ integrates weighted features from the self-attention layer ($\mathbf{x}^{\text{w}}_{t_i}$), feature and structural risk representations ($\mathbf{x}^{\text{fr}}_{t_i}, \mathbf{x}^{\text{gr}}_{t_i}$), and label embeddings $\mathbf{x}^{\text{L}}_{t_i} = \tilde{\mathbf{y}}\mathbf{W}_{\text{L}}$ [Shi *et al.*, 2021] (where $\tilde{\mathbf{y}}$ masks the central label to prevent leakage): $\mathbf{x}_{t_i} = \text{Concat}(\mathbf{x}^{\text{w}}_{t_i}, \mathbf{x}^{\text{L}}_{t_i}, \mathbf{x}^{\text{fr}}_{t_i}, \mathbf{x}^{\text{gr}}_{t_i})$. TGAT aggregates embeddings using multi-head attention:

$$\alpha_{\mathbf{x}_t,\mathbf{x}_i} = \frac{\exp(\text{LeakyReLU}(\mathbf{a}^T[\mathbf{x}_t \| \mathbf{x}_i]))}{\sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_t)} \exp(\text{LeakyReLU}(\mathbf{a}^T[\mathbf{x}_t \| \mathbf{x}_j]))} \tag{9}$$

$$\text{Head} = \sum_{\mathbf{x}_i \in \mathcal{X}} \sigma\Big( \sum_{\mathbf{x}_t \in \mathcal{N}(x_i)} \alpha_{\mathbf{x}_t,\mathbf{x}_i} \mathbf{x}_t \Big) \tag{10}$$

$$\mathbf{H} = \text{Concat}(\text{Head}_1, ..., \text{Head}_{h_{att}})\mathbf{W}_o \tag{11}$$

where $\alpha_{\mathbf{x}_t,\mathbf{x}_i}$ represents the attention weight for edge $(\mathbf{x}_t, \mathbf{x}_i)$, $[\mathbf{x}_t || \mathbf{x}_i]$ denotes vector concatenation, $\mathbf{a} \in \mathbb{R}^{2d}$ is the attention weight vector of each head, $h_{att}$ is the number of attention heads, $\mathbf{W}_o \in \mathbb{R}^{d \times d}$ is a learnable projection matrix, and $\mathbf{H}$ denotes the aggregated embedding. To mitigate gradient vanishing, we apply a gated residual connection fusing raw attributes $\mathbf{x}_{t_i}$ and aggregated embeddings $\mathbf{h}_{t_i}$:

$$\text{gate}_{t_i} = \sigma([\mathbf{x}_{t_i} \| \mathbf{h}_{t_i}]\beta_{t_i}) \tag{12}$$

$$\mathbf{z}_{t_i} = \text{gate}_{t_i} \cdot \mathbf{h}_{t_i} + (1 - \text{gate}_{t_i}) \cdot \mathbf{x}_{t_i} \tag{13}$$

where $\beta_{t_i}$ is a learnable gating vector, and $gate_{t_i}$ determines the relative importance of aggregated embeddings $\mathbf{h}_{t_i}$ versus raw attributes $\mathbf{x}_{t_i}$. The output $\mathbf{z}_{t_i}$ serves as input to the next TGAT layer, iteratively refining representations.

### 3.5 Loss Function and Model Optimization

For the aggregated embedding $\mathbf{H}$, we apply a MLP to predict fraud risk. The prediction layer is formulated as:

$$\hat{y} = \sigma(\text{PReLU}(\mathbf{H}\mathbf{W}_0 + \mathbf{b}_0)\mathbf{W}_1 + \mathbf{b}_1) \tag{14}$$

where $\hat{y} \in \mathbb{R}^{n \times 1}$ represents the predicted risk outcomes for all transactions, and $\mathbf{W}$ and $\mathbf{b}$ denote learnable parameters in the MLP. The final loss function combines binary cross-entropy and the DRL-based feature risk extractor loss $\mathcal{L}_{DRL}$:

$$\mathcal{L}_{\text{CE}} = -\frac{1}{|V|} \sum_{i=0}^{|V|} [y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)] \tag{15}$$

$$\mathcal{L} = \mathcal{L}_{\text{CE}} + \alpha \mathcal{L}_{\text{DRL}} \tag{16}$$

where $|V|$ is the number of central nodes, and the hyperparameter $\alpha$ controls the weight of $\mathcal{L}_{DRL}$ in the total loss. The optimization objective of our proposed model is to identify effective extraction strategies for risk-indicative information within feature while ensuring the learning of maximally discriminative node representations for fraud detection.

## 4 Experiment

### 4.1 Experiment Setup

**Datasets.** We conduct experiments on three real-world datasets: two public benchmarks, Amazon [McAuley and Leskovec, 2013] and YelpChi [Rayana and Akoglu, 2015], and a proprietary credit card fraud dataset, FFSD. Collected from a leading financial institution, FFSD comprises transaction records characterized by attributes such as amount and type, with ground-truth labels manually annotated by domain experts. Notably, FFSD mirrors real-world label scarcity: out of 927,555 transactions, only 140,576 (15%) are labeled, leaving 85% unlabeled. This high sparsity underscores the critical need for semi-supervised methods to leverage abundant unlabeled data in resource-constrained financial settings.

**Compared Methods.** To verify the effectiveness of DRESS, we compare it with 12 state-of-the-art GNN-based methods. As there are few GNN-based semi-supervised fraud detectors, we adopt nine supervised baselines: GraphSAGE [Hamilton *et al.*, 2017], GEM [Liu *et al.*, 2018], GraphConsis [Liu *et al.*, 2020], PC-GNN [Liu *et al.*, 2021], CARE-GNN [Dou *et al.*, 2020], GAGA [Wang *et al.*, 2023], FRAUDRE [Zhang *et al.*, 2021], HOGRL [Zou and Cheng, 2024], DGA-GNN [Duan *et al.*, 2024]; and three semi-supervised ones: SemiGNN [Wang *et al.*, 2019], GTAN [Xiang *et al.*, 2023], RGTAN [Xiang *et al.*, 2025].

**Evaluation Metrics and Parameter Settings.** We evaluate performance using AUC, Macro-F1, and AP. To simulate semi-supervised settings, we adopt a 10/10/80% train/val/test split. The hidden dimension is set to 256. Key hyperparameters include an initial learning rate of $1e^{-3}$, $\tau = 0.1$, $k = 0.5$, $\gamma = 0.99$, and $\alpha = 0.6$ (see Sec. 4.6). The TGAT layer depth is set to 2 for YelpChi and FFSD and 1 for Amazon.

### 4.2 Fraud Detection Experiment

To evaluate the performance of our proposed model, we compare DRESS against various baselines on three datasets (Table 1). The results indicate that DRESS outperforms baselines on most metrics across all datasets, demonstrating its strong fraud detection capability. On the Amazon and YelpChi datasets, supervised GNN models achieve relatively strong performance. This can be attributed to their specialized architectures tailored for supervised learning, along with the fact that semi-supervised models are reduced to supervised learning on these two fully labeled datasets. Among supervised approaches, DGA-GNN performs the best, yet DRESS still matches or exceeds its results on most metrics. On Amazon, DRESS achieves performance comparable to DGA-GNN, which is due to the dataset's small scale and susceptibility to overfitting. In contrast, on the real-world financial dataset FFSD, semi-supervised models exhibit advan-

| Datasets | | Amazon | | | YelpChi | | | FFSD | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Metrics | | AUC | F1 | AP | AUC | F1 | AP | AUC | F1 | AP |
| Supervised | GraphSAGE | 0.7381 | 0.4971 | 0.2058 | 0.5364 | 0.4606 | 0.1664 | 0.6717 | 0.4514 | 0.4563 |
| | GEM | 0.8006 | 0.5120 | 0.2714 | 0.5890 | 0.4657 | 0.2331 | 0.6913 | 0.4514 | 0.3070 |
| | GraphConsis | 0.8705 | 0.6383 | 0.8572 | 0.6844 | 0.4609 | 0.5895 | 0.7347 | 0.6749 | 0.5528 |
| | PC-GNN | 0.9341 | 0.8857 | 0.8064 | 0.7416 | 0.4608 | 0.3497 | 0.6718 | 0.4514 | 0.4036 |
| | CARE-GNN | 0.9303 | 0.9009 | 0.8387 | 0.7547 | 0.4606 | 0.3701 | 0.6811 | 0.4654 | 0.4571 |
| | FRAUDRE | 0.9331 | 0.7452 | 0.8436 | 0.7551 | 0.4624 | 0.3690 | 0.6814 | 0.4609 | 0.4589 |
| | GAGA | 0.9108 | 0.8835 | 0.7701 | 0.8974 | 0.7567 | 0.6430 | 0.7467 | 0.7009 | 0.5575 |
| | HOGRL | 0.9251 | 0.8788 | 0.8179 | 0.7833 | 0.6200 | 0.4440 | 0.6587 | 0.6160 | 0.4860 |
| | DGA-GNN | **0.9667** | 0.9016 | 0.8747 | 0.9127 | 0.7790 | 0.7044 | 0.7748 | **0.7559** | 0.5845 |
| Semi-supervised | SemiGNN | 0.6214 | 0.3432 | 0.0907 | 0.5463 | 0.3911 | 0.1421 | 0.5105 | 0.4237 | 0.1796 |
| | GTAN | 0.8755 | 0.8952 | 0.8140 | 0.8030 | 0.6349 | 0.4305 | 0.7526 | 0.7059 | 0.5090 |
| | RGTAN | 0.9159 | 0.9084 | 0.7789 | 0.8940 | 0.7716 | 0.6537 | 0.7792 | 0.7177 | 0.6073 |
| Ours | DRESS | 0.9608 | **0.9138*** | **0.8785** | **0.9527*** | **0.8496*** | **0.8090*** | **0.8049*** | 0.7389 | **0.6583*** |

Table 1: Fraud detection performance of various methods on three datasets. The boldface refers to the highest score of the models. "*" indicates the statistically significant improvements (i.e., paired t-test with $p < 0.05$) over the best baseline.

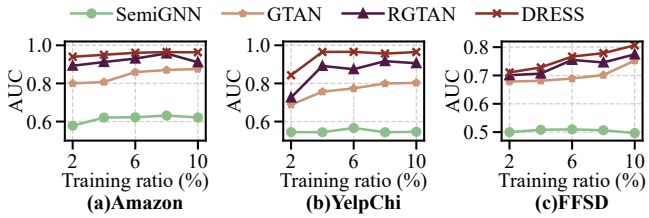


Figure 3: Semi-supervised experiment results under different ratios of labeled training data.

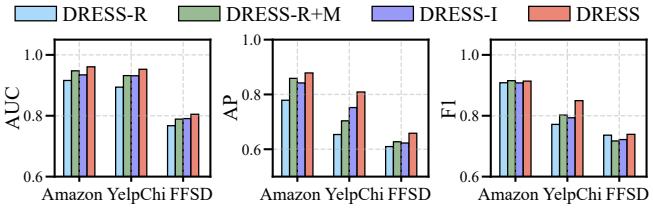


Figure 4: Ablation study comparing DRESS with its three ablated variants across three datasets.

tages by leveraging unlabeled samples. Notably, DRESS consistently surpasses RGTAN—the strongest semi-supervised baseline—across all evaluation metrics. This confirms the effectiveness of our method in realistic, label-scarce scenarios. The performance gain can be largely attributed to DRESS's DRL-driven feature risk extraction, and the effective fusion of feature and structural risks. These findings underscore the value of DRESS in advancing fraud detection systems, offering more reliable identification of fraudulent transactions.

### 4.3 Semi-supervised Experiment

We evaluate DRESS against three semi-supervised baselines (SemiGNN, GTAN, RGTAN) across Amazon, YelpChi, and FFSD datasets, varying the labeled training ratio from 2% to 10% (Figure 3). DRESS consistently outperforms all baselines across all ratios. Notably, in extreme low-label scenarios (2% ratio), DRESS achieves substantial AUC gains over the best competitor RGTAN, specifically 11.69% on YelpChi and 0.98% on FFSD, while maintaining consistently high performance on Amazon. While extreme sparsity inevitably causes performance degradation on YelpChi across all methods, DRESS maintains a significant absolute performance lead over the baselines. As label availability increases, DRESS sustains this superiority. This improvement probably stems from the DRL mechanism, which autonomously mines feature-level risks within individual nodes. By propagating these enhanced signals—especially from abundant unlabeled nodes—DRESS leverages unlabeled data more effectively than baselines, confirming its robustness and practical value for reliable transaction monitoring with minimal labels.

### 4.4 Ablation Study

To validate each component, we compare DRESS against three variants: DRESS-R (w/o RL extractor), DRESS-R+M (replacing RL with a machine learning method), and DRESS-I (w/o intrinsic rewards). The machine learning(ML) method used in DRESS-R+M is DevNet [Pang *et al.*, 2019], which employs a deviation loss to enforce a statistically significant separation of anomaly scores from normal data distributions.

As shown in Figure 4, DRESS consistently outperforms all ablated versions. First, compared to DRESS-R, DRESS achieves significant gains across all datasets (e.g., AP improves by 12.82%, 10.33%, and 4.87% on Amazon, YelpChi, and FFSD), highlighting the DRL extractor's critical role in capturing fraud patterns. Second, DRESS surpasses DRESS-R+M. Unlike traditional ML methods (e.g., DevNet) constrained by rigid distributional assumptions like Gaussianity, our DRL approach enables dynamic, distribution-free feature enhancement. This supports the necessity and superiority of adopting reinforcement learning. Third, DRESS outperforms DRESS-I, particularly in AP (gains of 3.65%, 9.75%, and 3.06%). Given AP's sensitivity to minority class detection,

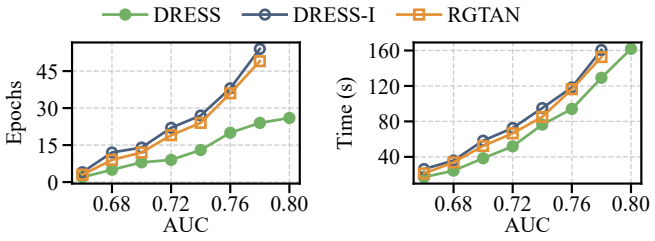


Figure 5: Training epochs (left) and time (right) required to achieve different AUC levels for DRESS, its ablated variant DRESS-I (w/o intrinsic rewards), and the RGTAN baseline.

the results indicate that intrinsic rewards effectively guide the model to focus on key fraud-indicative features, boosting performance in imbalanced classification. Collectively, these ablation results validate the design of DRESS and the synergistic effect of its DRL extractor and self-supervised intrinsic reward mechanism, which together enhance real-world fraud detection under label scarcity.

### 4.5 Efficiency Improvement Study

To further evaluate the effectiveness of the proposed intrinsic reward mechanism in improving training efficiency within the DRL-enhanced framework, we record the epochs and time required to reach specific AUC thresholds (FFSD results in Figure 5). Comparing DRESS with DRESS-I (w/o intrinsic rewards) reveals that intrinsic rewards significantly accelerate convergence. For instance, achieving 0.76 AUC on FFSD takes 36 epochs (116.35s) for DRESS-I, compared to just 20 epochs (94.12s) for DRESS. This efficiency improvement probably stems from its ability to prioritize critical feature attention, which mitigates unproductive exploration in high-dimensional spaces with sparse extrinsic rewards, thereby accelerating agent policy optimization. Furthermore, DRESS achieves target AUC levels faster than RGTAN, demonstrating superior efficiency over state-of-the-art semi-supervised baselines. Crucially, DRESS achieves this efficiency while introducing negligible computational overhead, as evidenced by its smooth operation on our financial partner's existing infrastructure. This combination of high accuracy and low training cost enhances the feasibility of deploying DRESS in real-world financial systems.

### 4.6 Parameter Sensitivity

In this section, we analyze DRESS's sensitivity to four hyperparameters ($\tau, k, \gamma, \alpha$), as shown in Figure 6.

**Impact of the temperature coefficient $\tau$.** $\tau$ controls the sharpness of soft masking (Eq. 4). Results confirm that a smaller $\tau$ yields superior performance. A lower $\tau$ induces a harder mask, effectively suppressing high-contribution features. This maximizes the discriminative divergence between $\mathbf{x}^{\mathrm{fr},+}$ and $\mathbf{x}^{\mathrm{fr},-}$, thereby enhancing the model's ability to distinguish fraudulent patterns.

**Impact of the masking ratio $k$.** The hyperparameter $k$ controls the proportion of features being masked. As shown in the figure, model performance peaks when $k = 0.5$. A lower ratio ($k < 0.5$) fails to generate sufficient divergence for contrastive learning. Conversely, a higher ratio ($k > 0.5$) masks too many informative features, introducing noise into the intrinsic reward calculation and degrading optimization.

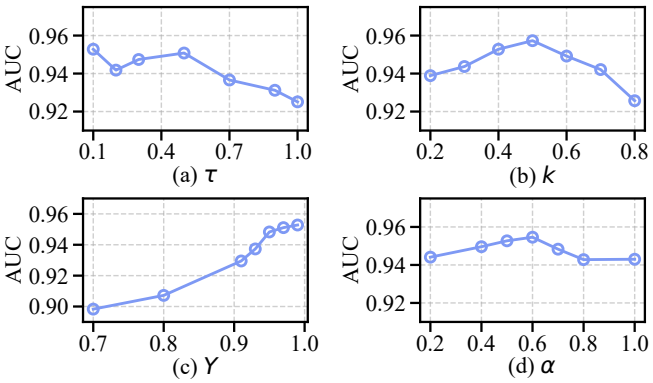


Figure 6: Parameter Sensitivity Analysis with respect to (a) masking temperature coefficient $\tau$ in the feature self-attention layer, (b) feature masking ratio $k$, (c) the discount factor $\gamma$, and (d) the weight $\alpha$ of the DRL-based risk information extractor in the total loss.

**Impact of the discount factor $\gamma$.** $\gamma$ regulates the weight assigned to future rewards in the DRL-based feature risk extractor. It balances the influence of immediate and future states when estimating the cumulative return. As illustrated in the figure, increasing $\gamma$ generally improves model performance. This may be attributed to the temporal aggregation of fraudulent activities: successive states are closely related in time, and a higher $\gamma$ helps the model better leverage such temporal correlations, thereby improving fraud detection accuracy.

**Impact of the loss weight $\alpha$.** The hyperparameter $\alpha$ controls the contribution of the reinforcement learning loss $\mathcal{L}_{\mathrm{DRL}}$ to the overall loss. The model shows stable behavior with respect to $\alpha$, and achieves the best balance between feature risk extraction and graph-based propagation when $\alpha = 0.6$. This indicates a proper trade-off between feature-level risk extraction and graph-based information propagation, both of which are essential for robust semi-supervised fraud detection.

## 5 Conclusion

In this paper, we propose a Deep Reinforcement Learning enhanced semi-supervised graph neural network, which efficiently detects transaction fraud under real-world label scarcity. To uncover latent fraud within node features, we devise a DRL-driven feature risk extractor with a self-supervised intrinsic reward, mitigating DRL's low efficiency in semi-supervised settings. By adaptively generating new risk features, the model enhances risk signals within abundant unlabeled nodes, improving credit card fraud detection performance. Extensive experiments on real-world datasets demonstrate DRESS's superiority over baselines. Semi-supervised evaluations further confirm its optimal effectiveness even with limited labeled samples. These results highlight its capability to uncover subtle fraud patterns, providing a practical and powerful tool for financial institutions. The effectiveness of DRESS will also provide new perspectives to enhance financial security and justice, contributing to sustainable economic growth and poverty alleviation by safeguarding vulnerable populations and fostering a fair ecosystem.

## Acknowledgments

The work is supported by the National Science Fundation of China (62472317, 62522213) and the Fundamental Research Funds for the Central Universities.

## References

[AlFalahi and Nobanee, 2024] Latifa AlFalahi and Haitham Nobanee. Conceptual building of sustainable economic growth and corporate bankruptcy. *Social Science Electronic Publishing*, 2024. [Online; accessed 2024-12-12].

[Bei *et al.*, 2023] Yuanchen Bei, Sheng Zhou, Qiaoyu Tan, Hao Xu, Hao Chen, Zhao Li, and Jiajun Bu. Reinforcement neighborhood selection for unsupervised graph anomaly detection. In *2023 IEEE International Conference on Data Mining (ICDM)*, pages 11–20. IEEE, 2023.

[Cheng *et al.*, 2020a] Dawei Cheng, Xiaoyang Wang, Ying Zhang, and Liqing Zhang. Graph neural network for fraud detection via spatial-temporal attention. *IEEE Transactions on Knowledge and Data Engineering*, 34(8):3800–3813, 2020.

[Cheng *et al.*, 2020b] Dawei Cheng, Sheng Xiang, Chencheng Shang, Yiyi Zhang, Fangzhou Yang, and Liqing Zhang. Spatio-temporal attention-based neural network for credit card fraud detection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 362–369, 2020.

[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.*, 2024] Mingjiang Duan, Tongya Zheng, Yang Gao, Gang Wang, Zunlei Feng, and Xinyu Wang. Dga-gnn: Dynamic grouping aggregation gnn for fraud detection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pages 11820–11828, 2024.

[Elkan and Noto, 2008] Charles Elkan and Keith Noto. Learning classifiers from only positive and unlabeled data. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 213–220, 2008.

[Fu *et al.*, 2016] Kang Fu, Dawei Cheng, Yi Tu, and Liqing Zhang. Credit card fraud detection using convolutional neural networks. In *International conference on neural information processing, ICONIP 2016*, pages 483–490, 2016.

[Hamilton *et al.*, 2017] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems*, 30, 2017.

[Heidari *et al.*, 2024] Marzi Heidari, Hanping Zhang, and Yuhong Guo. Reinforcement learning guided semi-supervised learning. *Advances in Neural Information Processing Systems*, 37:136990–137009, 2024.

[Hu *et al.*, 2019] Binbin Hu, Zhiqiang Zhang, Chuan Shi, Jun Zhou, Xiaolong Li, and Yuan Qi. Cash-out user detection based on attributed heterogeneous information network with a hierarchical attention mechanism. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 946–953, 2019.

[Kazari *et al.*, 2023] Kiarash Kazari, Ezzeldin Shereen, and György Dán. Decentralized anomaly detection in cooperative multi-agent reinforcement learning. In *IJCAI*, pages 162–170, 2023.

[K.R. and Zareapoor, 2014] Seeja K.R. and Masoumeh Zareapoor. Fraudminer: A novel credit card fraud detection model based on frequent itemset mining. *The Scientific World Journal*, 2014(1):252797, 2014.

[Li *et al.*, 2024] Fan Li, Zhiyu Xu, Dawei Cheng, and Xiaoyang Wang. Adarisk: risk-adaptive deep reinforcement learning for vulnerable nodes detection. *IEEE Transactions on Knowledge and Data Engineering*, 36(11):5576–5590, 2024.

[Liu *et al.*, 2018] Ziqi Liu, Chaochao Chen, Xinxing Yang, Jun Zhou, Xiaolong Li, and Le Song. Heterogeneous graph neural networks for malicious account detection. In *Proceedings of the 27th ACM international conference on information and knowledge management*, pages 2077–2085, 2018.

[Liu *et al.*, 2020] Zhiwei Liu, Yingtong Dou, Philip S Yu, Yutong Deng, and Hao Peng. Alleviating the inconsistency problem of applying graph neural network to fraud detection. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*, pages 1569–1572, 2020.

[Liu *et al.*, 2021] 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.*, 2025] Xin Liu, Haojun Rui, Dawei Cheng, Li Han, Zhongyun Zhou, and Guoping Zhao. Multi-granularity augmented graph learning for spoofing transaction detection. In *Proceedings of the ACM on Web Conference 2025*, pages 5151–5160, 2025.

[Liu *et al.*, 2026] Xin Liu, Yuanhang Yu, Peng Zhu, Dawei Cheng, and Changjun Jiang. Role perceptual augmented temporal graph network for related-party transaction detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 38943–38951, 2026.

[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.

[Mnih *et al.*, 2015] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through

deep reinforcement learning. *nature*, 518(7540):529–533, 2015.

[Nargesian *et al.*, 2017] Fatemeh Nargesian, Horst Samulowitz, Udayan Khurana, Elias B Khalil, and Deepak S Turaga. Learning feature engineering for classification. In *IJCAI*, volume 17, pages 2529–2535, 2017.

[Ouyang *et al.*, 2022] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744, 2022.

[Pang *et al.*, 2019] Guansong Pang, Chunhua Shen, and Anton Van Den Hengel. Deep anomaly detection with deviation networks. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 353–362, 2019.

[Pang *et al.*, 2021] Guansong Pang, Anton van den Hengel, Chunhua Shen, and Longbing Cao. Toward deep supervised anomaly detection: Reinforcement learning from partially labeled anomaly data. In *Proceedings of the 27th ACM SIGKDD International conference on knowledge discovery and data mining*, pages 1298–1308, 2021.

[Peng *et al.*, 2024] Tianhao Peng, Wenjun Wu, Haitao Yuan, Zhifeng Bao, Zhao Pengru, Xin Yu, Xuetao Lin, Yu Liang, and Yanjun Pu. Graphrare: Reinforcement learning enhanced graph neural network with relative entropy. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, pages 2489–2502. IEEE, 2024.

[PricewaterhouseCoopers, 2024] PricewaterhouseCoopers. Global economic crime survey 2024. https://www.pwc.com/gx/en/services/forensics/economic-crime-survey.html, 2024. Accessed: 2025-11-30.

[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.

[Shi *et al.*, 2021] Yunsheng Shi, Zhengjie Huang, Shikun Feng, Hui Zhong, Wenjing Wang, and Yu Sun. Masked label prediction: Unified message passing model for semi-supervised classification. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 1548–1554. International Joint Conferences on Artificial Intelligence Organization, 8 2021.

[Tian *et al.*, 2023] Sheng Tian, Jihai Dong, Jintang Li, Wenlong Zhao, Xiaolong Xu, Baokun Wang, Bowen Song, Changhua Meng, Tianyi Zhang, and Liang Chen. Sad: semi-supervised anomaly detection on dynamic graphs. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 2306–2314, 2023.

[Wang *et al.*, 2019] Daixin Wang, Jianbin Lin, Peng Cui, Quanhui Jia, Zhen Wang, Yanming Fang, Quan Yu, Jun Zhou, Shuang Yang, and Yuan Qi. A semi-supervised graph attentive network for financial fraud detection. In *2019 IEEE international conference on data mining (ICDM)*, pages 598–607. IEEE, 2019.

[Wang *et al.*, 2023] Yuchen Wang, Jinghui Zhang, Zhengjie Huang, Weibin Li, Shikun Feng, Ziheng Ma, Yu Sun, Dianhai Yu, Fang Dong, Jiahui Jin, et al. Label information enhanced fraud detection against low homophily in graphs. In *Proceedings of the ACM Web Conference 2023*, pages 406–416, 2023.

[Wu *et al.*, 2025] Jiasheng Wu, Xincheng Wang, Jie Yang, Dawei Cheng, Guang Yang, and Bo Wang. Defending attacks on anti-fraud model with generative graph representations. *IEEE Transactions on Knowledge and Data Engineering*, 38(2):969–982, 2025.

[Xiang *et al.*, 2023] Sheng Xiang, Mingzhi Zhu, Dawei Cheng, Enxia Li, Ruihui Zhao, Yi Ouyang, Ling Chen, and Yefeng Zheng. Semi-supervised credit card fraud detection via attribute-driven graph representation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 14557–14565, 2023.

[Xiang *et al.*, 2025] Sheng Xiang, Guibin Zhang, Dawei Cheng, and Ying Zhang. Enhancing attribute-driven fraud detection with risk-aware graph representation. *IEEE Transactions on Knowledge and Data Engineering*, 2025.

[Yu *et al.*, 2024] Hang Yu, Zhengyang Liu, and Xiangfeng Luo. Barely supervised learning for graph-based fraud detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 16548–16557, 2024.

[Yu *et al.*, 2025] Zhizhi Yu, Chundong Liang, Xinglong Chang, Dongxiao He, Di Jin, and Jianguo Wei. Dynamic neighborhood modeling via node-subgraph contrastive learning for graph-based fraud detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 13115–13123, 2025.

[Zha *et al.*, 2020] Daochen Zha, Kwei-Herng Lai, Mingyang Wan, and Xia Hu. Meta-aad: Active anomaly detection with deep reinforcement learning. In *2020 IEEE International Conference on Data Mining (ICDM)*, pages 771–780. IEEE, 2020.

[Zhang *et al.*, 2021] Ge Zhang, Jia Wu, Jian Yang, Amin Beheshti, Shan Xue, Chuan Zhou, and Quan Z Sheng. Fraudre: Fraud detection dual-resistant to graph inconsistency and imbalance. In *2021 IEEE international conference on data mining (ICDM)*, pages 867–876. IEEE, 2021.

[Zhang *et al.*, 2024] Jinghui Zhang, Zhengjia Xu, Dingyang Lv, Zhan Shi, Dian Shen, Jiahui Jin, and Fang Dong. Digin-gnn: Discriminative feature guided gnn-based fraud detector against inconsistencies in multi-relation fraud graph. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pages 9323–9331, 2024.

[Zou and Cheng, 2024] Yao Zou and Dawei Cheng. Effective high-order graph representation learning for credit card fraud detection. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, pages 7581–7589. International Joint Conferences on Artificial Intelligence Organization, 8 2024.