Fighting against Organized Fraudsters Using Risk Diffusion-based Parallel Graph Neural Network

Jiacheng Ma[1] , **Fan Li**[1] , **Rui Zhang**[1] , **Zhikang Xu**[1] , **Dawei Cheng**[1,2*] , **Yi Ouyang**[3] , **Ruihui Zhao**[3] , **Jianguang Zheng**[3] , **Yefeng Zheng**[3] and **Changjun Jiang**[1,2]

[1]Department of Computer Science and Technology, Tongji University, Shanghai, China
[2]Shanghai Artificial Intelligence Laboratory, Shanghai, China
[3]Tencent Jarvis Laboratory, Shenzhen, China

Abstract

Medical insurance plays a vital role in modern society, yet organized healthcare fraud causes billions of dollars in annual losses, severely harming the sustainability of the social welfare system. Existing works mostly focus on detecting individual fraud entities or claims, ignoring hidden conspiracy patterns. Hence, they face severe challenges in tackling organized fraud. In this paper, we proposed RDPGL, a novel Risk Diffusion-based Parallel Graph Learning approach, to fighting against medical insurance criminal gangs. In particular, we first leverage a heterogeneous graph attention network to encode the *local* context from the beneficiary-provider graph. Then, we devise a community-aware risk diffusion model to infer the *global* context of organized fraud behaviors with the claim-claim relation graph. The *local* and *global* representations are parallel concatenated together and trained simultaneously in an end-to-end manner. Our approach is extensively evaluated on a real-world medical insurance dataset. The experimental results demonstrate the superiority of our proposed approach, which could detect more organized fraud claims with relatively high precision compared with state-of-the-art baselines.

1 Introduction

In recent years, fraudulent insurance claims of medical insurance have become a serious concern in the healthcare and social welfare system [Liang *et al.*, 2019]. According to the report from the U.S. National Health Care Anti-Fraud Association (NHCAA), at least $300 billion, about more than 10% of the nation's $3.6 trillion healthcare expenditure, was lost annually due to healthcare fraud [NHCAA, 2021]. For employers, healthcare fraud inevitably translates into higher premiums and increases the overall cost of doing business. More importantly, financial losses caused by healthcare fraud have led to more out-of-pocket expenses for consumers [Rudman *et al.*, 2009]. There is no doubt that healthcare fraud can have devastating effects on healthy lives, social well-being, as well as sustainable economic development.

Unfortunately, these large-scale healthcare frauds are not just committed by individual dishonest healthcare providers or consumers. An increasing number of criminals are organized like enterprises, which can be far-reaching and move quickly from place to place, to conduct conspiracy frauds to covet the enticing pool of health care money [Timofeyev and Jakovljevic, 2022]. In the year 2018 alone, investigative efforts of the Federal Bureau of Investigation resulted in over 812 operational disruptions of criminal fraud organizations and the dismantlement of the criminal hierarchy of more than 207 healthcare fraud criminal enterprises [Stowell *et al.*, 2018; NHCAA, 2021]. As a result, it is crucial now more than ever to develop a more powerful and flexible approach for combating organized fraudsters.

The healthcare industry has developed anti-fraud approaches for medical insurance since the early 1980s [Hearn Jr, 1989], from the statistical rules [Major and Riedinger, 1992] to classical machine learning methods [Dua and Bais, 2014; Bauder and Khoshgoftaar, 2017]. Later, deep neural networks were introduced to learn the latent fraud patterns [Pandey, 2017; Pumsirirat and Liu, 2018], which uncovers the power of deep architecture in fraud detection. Meanwhile, the cheating methods are also upgraded, becoming too deceptive and concealing for a classical deep model to detect, because the model treats each fraud action as isolated. Recently, graph neural network (GNN) has been employed for fraud detection and achieved remarkable success [Cheng *et al.*, 2020b; Xu *et al.*, 2021; Cheng *et al.*, 2023], as GNN could effectively learn latent features from historical interconnected behavior [Cheng *et al.*, 2020a]. In other words, GNN could more accurately infer the fraud probability by learning the fraud action from relation graphs.

However, existing graph-based fraud detection methods mostly learn one or multiple relation types only from entity graphs [Ma *et al.*, 2018; Dou *et al.*, 2020; Xu *et al.*, 2021; Zhang *et al.*, 2022], which will inevitably lead to suboptimal detection performance. Because entity relation graphs only represent the *local* context of a suspicious claim. In contrast, a wider *global* context can be constructed by action relation graph [Weber *et al.*, 2019], in which a node means an action, like an insurance claim in this paper. Research [Li *et al.*, 2019] has shown that combining the en-

[*] Corresponding Author: dcheng@tongji.edu.cn

tity and action-level relations could improve the model performance in spam review detection. However, we still face two significant challenges in detecting organized healthcare fraud claims: 1) Conspiracy fraudsters are very cunning, can be far-reaching and move quickly from place to place. 2) The patterns of conspiracy fraud are updated frequently.

Therefore, we propose a novel risk diffusion-based parallel graph learning approach, named RDPGL, for organized healthcare fraud detection. In particular, we construct the beneficiary-provider relations as *local* heterogeneous entity graphs (in which a node means a service provider or a beneficiary) and the claim-claim relations as *global* action graphs (in which a node denotes a medical insurance claim). Then, we devise a community-aware risk diffusion graph neural model and graph attentional mechanism to learn from action and entity graphs in parallel to better extract quickly changing behavior patterns of conspiracy fraudsters. The learned representations are then concatenated in the claim level for jointly optimizing by the detection network. Extensive experiments on a real-world medical insurance dataset demonstrate the superior performance of our proposed approach compared with state-of-the-art baselines. In a nutshell, the main contributions of this paper can be summarized:

- To the best of our knowledge, this is the first work that addresses organized healthcare fraud by proposing a novel parallel graph learning approach that could jointly learn from both the local beneficiary-provider relation and the global claim-claim relations.
- We devise a community-aware risk diffusion graph neural model to encode the quickly changing fraudster behavior's *global* representations and graph attentional mechanism to learn *local* entity features. They are jointly optimized in the prediction network to better extract constantly-updated organized conspiracy patterns.
- We validate the effectiveness of the proposed method on a real-world medical insurance dataset, which was manually annotated by healthcare insurance domain experts. Our source codes and the dataset will be available at Github[1].

2 Preliminary

2.1 Healthcare Fraud Backgrounds

Healthcare fraud is a financial crime in which medical insurance claims are dishonestly filed to profit illegally from the payments received [Villegas-Ortega *et al.*, 2021]. It will not only cause substantial financial losses each year but also raise health insurance premiums, deplete valuable medical resources, and increase business costs. Fraud can be committed by medical service providers, patients, and others who intentionally deceive the healthcare system to receive unlawful benefits or payments. The most common types of fraud by providers are billing for services that were not provided, misrepresenting the service provided, and charging for a more complex or expensive service than was provided [NHCAA, 2021], while dishonest beneficiaries usually claim fake reimbursements or misrepresent service statements [Waghade and Karandikar, 2018]. However, in recent years, conspiracy fraud has appeared in real business, which means organized gangs are involved in the crime, and fraudulent activities may include various deceptive patients and providers. This kind of fraud contains the human brain-armed organized behavioral patterns, causing significant challenges for existing methods to detect. Therefore, this paper addresses this critical task in the healthcare industry, aiming to detect both individual-level and gang-level fraud claims.

2.2 Problem Formulation

In the business procedure of medical insurance, two groups of entities are involved in the commission of healthcare claims. They are (a) *service providers*, including doctors, hospitals, ambulance companies, and laboratories; (b) *insurance* beneficiaries, including patients and patients' employers.

In this paper, we denote the beneficiary-provider relation graph as $\mathcal{G}(\mathcal{B}, \mathcal{P}, \mathcal{E})$ where $\mathcal{B} = \{v_1^b, ..., v_{N_B}^b\}$ is the set of insurance beneficiary nodes, $\mathcal{P} = \{v_1^p, ..., v_{N_P}^p\}$ is the set of medical service provider nodes, and $\mathcal{E} = \{e_1, ..., e_{N_C}\}$ denotes the set of edges in the graph, which represents the claims of medical treatment between the beneficiary and provider. We denote the number of beneficiaries, providers, and claims as N_B, N_P, and N_C, respectively. As for neighbors in the graph, let $\mathcal{N}_v$ be the set of nodes in node v's one-hop neighbors, so $\mathcal{N}_{v_i^b \in \mathcal{B}} \subseteq \mathcal{P}$ and $\mathcal{N}_{v_j^p \in \mathcal{P}} \subseteq \mathcal{B}$ in our network. Each beneficiary node v_i^b is represented by a d^B-dimensional feature vector $\boldsymbol{h_{b,i}^0} \in \mathbb{R}^{d^B}$,while each provider node v_j^p can be described by a d^P-dimensional vector $\boldsymbol{h_{p,j}^0} \in \mathbb{R}^{d^P}$. At the same time, for edge e_k we define $\boldsymbol{h_{e,k}^0} \in \mathbb{R}^{d^C}$ as its attribute vector and let $\mathcal{Y}_c = \{0, 1\}_C^N$ as the set of fraud labels, where 0 represents normal and 1 represents fraud.

Meanwhile, we also construct the claim-claim graph $\mathcal{G}_c(\mathcal{V}_c, \mathcal{E}_c)$, in which a node represents a claim. For two claims $v_i^c, v_j^c \in \mathcal{V}_c$, there will be an edge $e_{i,j}^c$ linked between them if both transactions share the same beneficiary or provider. For each claim record, we aim to infer the possibility of whether it is a fraud event, and our task can be formulated as an edge classification problem in a beneficiary-provider graph, and a node classification problem in the claim-claim graph. In this paper, we study two detection problems: individual-level fraud claims detection and gang-level fraud claims detection. Briefly, it is a pattern discovery and classification problem in graphs. In particular, the model needs to discover the groups of fraudulent entities in medical insurance claims.

3 Methodology

In this section, we present each component of our proposed fraud detection method in medical insurance claims in detail. We first introduce the learning approach in the heterogeneous entity bipartite graph. Then, we present the construction of the claim-claim relation graph and the representation learn-

[1] https://github.com/finint/antifraud

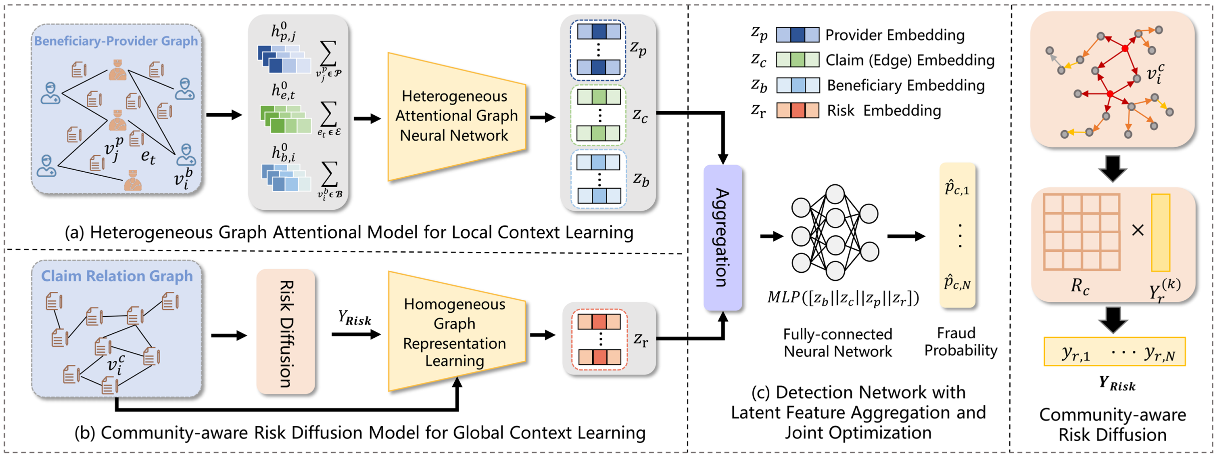

Beneficiary-Provider Graph
v_j^p
e_t
v_i^b
$h_{p,j}^0$
$\sum_{v_j^p \in \mathcal{P}}$
$h_{e,t}^0$
$\sum_{e_t \in \mathcal{E}}$
$h_{b,i}^0$
$\sum_{v_i^b \in \mathcal{B}}$
Heterogeneous Attentional Graph Neural Network
z_p
z_c
z_b
(a) Heterogeneous Graph Attentional Model for Local Context Learning
Claim Relation Graph
v_i^c
Risk Diffusion
Y_{Risk}
Homogeneous Graph Representation Learning
z_r
(b) Community-aware Risk Diffusion Model for Global Context Learning
z_p Provider Embedding
z_c Claim (Edge) Embedding
z_b Beneficiary Embedding
z_r Risk Embedding
Aggregation
$MLP([z_b||z_c||z_p||z_r])$
$\hat{p}_{c,1}$
$\hat{p}_{c,N}$
Fully-connected Neural Network
Fraud Probability
(c) Detection Network with Latent Feature Aggregation and Joint Optimization
v_i^c
R_c
×
$Y_r^{(k)}$
$y_{r,1}$ ⋯ $y_{r,N}$
Y_{Risk}
Community-aware Risk Diffusion

Figure 1: The model architecture of our proposed risk diffusion-based parallel graph learning (RDPGL) method. It contains three components: (a) the heterogeneous graph attentional network for *local* context learning on the beneficiary-provider graph; (b) the community-aware risk diffusion graph model for the learning of the quickly changing fraudster behavior's *global* context on the claim-claim graph; (c) The detection network with latent feature aggregation and joint optimization in parallel to better extract constantly-updated organized fraud patterns.

ing process. Lastly, we introduce the detection network and optimization strategy of the proposed methods.

3.1 Local Entity Graph Learning

The interconnectedness of medical insurance transactions is significant, and the goal of the beneficiary-provider bipartite graph is to mine knowledge hidden in the interaction network topology among distinct insurance entities. In consideration that the beneficiary nodes and provider nodes share a similar information aggregation phase, we choose the beneficiary entities here to report the operation procedure.

We first utilize the attention mechanism to indicate the importance of local neighboring node v_j^p's features to a node v_i^b where $v_j^p, v_k^p \in \mathcal{N}_{v_i^b}$. For the hidden state of nodes in the l-th layer, the attention coefficients can be computed as:

$$\alpha_{ij}^l = \frac{\exp\left(\sigma(\mathbf{u}^T\left[\mathbf{W_b}\boldsymbol{h}_{\boldsymbol{b,i}}^{\boldsymbol{l-1}}\|\mathbf{W_p}\boldsymbol{h}_{\boldsymbol{p,j}}^{\boldsymbol{l-1}}\right])\right)}{\sum_{v_k^p}\exp\left(\sigma(\mathbf{u}^T\left[\mathbf{W_b}\boldsymbol{h}_{\boldsymbol{b,i}}^{\boldsymbol{l-1}}\|\mathbf{W_p}\boldsymbol{h}_{\boldsymbol{p,k}}^{\boldsymbol{l-1}}\right])\right)}, \quad (1)$$

where $\mathbf{W_b} \in \mathbb{R}^{d^B \times d^B}$, $\mathbf{W_p} \in \mathbb{R}^{d^P \times d^P}$ are weight matrices of the beneficiary and provider entities, respectively, and $\boldsymbol{u} \in \mathbb{R}^{d^B + d^P}$ is the weight vector. We choose σ as the $LeakyReLU$ nonlinearity activation function and let $||$ be the concatenation operation. Then, we generate the message representations passed from neighboring nodes and edges. We assume that the edges (claims) between the two nodes contain the same amount of information and their sum of messages will be weighted by node attention coefficients computed above. We denote $\mathcal{N}_{ij}(e)$ as the set of edges connected between beneficiary node i and provider node j, and N_{ij}^e refers to the number of edges connecting them. The neighborhood message construction phase can be formulated as:

$$\boldsymbol{m}_{\boldsymbol{v_i^b \leftarrow v_i^b}}^{\boldsymbol{l}} = \sigma(\mathbf{W_1}\boldsymbol{h}_{\boldsymbol{b,i}}^{\boldsymbol{l-1}}),$$

$$\boldsymbol{m}_{\boldsymbol{v_i^b \leftarrow \mathcal{N}(e_i^b)}}^{\boldsymbol{l}} = \sigma\Big(\sum_{v_k^p \in \mathcal{N}_{v_i^b}} \alpha_{ik}^l \big(\frac{1}{N_{ik}^e}\sum_{e_s \in \mathcal{N}_{ik}(e)} \mathbf{W_2}\boldsymbol{h}_{\boldsymbol{e,s}}^{\boldsymbol{l-1}}\big)\Big),$$

$$\boldsymbol{m}_{\boldsymbol{v_i^b \leftarrow \mathcal{N}(v_i^b)}}^{\boldsymbol{l}} = \sigma\Big(\sum_{v_k^p \in \mathcal{N}_{v_i^b}} \alpha_{ik}^l \mathbf{W_3}\boldsymbol{h}_{\boldsymbol{p,k}}^{\boldsymbol{l-1}}\Big), \quad (2)$$

where $\boldsymbol{m}_{\boldsymbol{v_i^b \leftarrow \mathcal{N}(v_i^b)}}^{\boldsymbol{l}}$, $\boldsymbol{m}_{\boldsymbol{v_i^b \leftarrow \mathcal{N}(e_i^b)}}^{\boldsymbol{l}}$, $\boldsymbol{m}_{\boldsymbol{v_i^b \leftarrow v_i^b}}^{\boldsymbol{l}}$ represent the aggregation message from adjacent nodes, edges of v_i^b, and the hidden state of the node in the last layer, respectively, and $\mathbf{W_1} \in \mathbb{R}^{d^B \times d^B}$, $\mathbf{W_2} \in \mathbb{R}^{d^B \times d^C}$, $\mathbf{W_3} \in \mathbb{R}^{d^B \times d^P}$ are transformation matrices. We incorporate the propagated message above and formulate the updating paradigm as:

$$\boldsymbol{h}_{\boldsymbol{b,i}}^{\boldsymbol{l}} = \boldsymbol{m}_{\boldsymbol{v_{b,i} \leftarrow v_{b,i}}}^{\boldsymbol{l}} + \boldsymbol{m}_{\boldsymbol{v_{b,i} \leftarrow \mathcal{N}(v_{b,i})}}^{\boldsymbol{l}} + \boldsymbol{m}_{\boldsymbol{v_{b,i} \leftarrow \mathcal{N}(e_{b,i})}}^{\boldsymbol{l}}. \quad (3)$$

The generation of hidden vector $\boldsymbol{h}_{\boldsymbol{p}}^{\boldsymbol{l}}$ for provider nodes shares similar operation illustrated above. With aggregation sublayers, the model has the sufficient expressive power to map input features of node entities into high-level representations and capture deep structural relationships among them.

Afterward, different from the traditional graph neural network only processing nodes, we also need an update function for edge attributes to generate claim embedding, which contains high-order information from node entities. For $\forall e \in \mathcal{E}$, let $\boldsymbol{h}_{\boldsymbol{e}}^{\boldsymbol{l-1}}, \boldsymbol{h}_{\boldsymbol{B(e)}}^{\boldsymbol{l-1}}, \boldsymbol{h}_{\boldsymbol{P(e)}}^{\boldsymbol{l-1}}$ be the hidden states of claim and its corresponding beneficiary and provider output from layer $l-1$, respectively. We define the aggregation function as:

$$\boldsymbol{h}_{\boldsymbol{e}}^{\boldsymbol{l}} = \sigma\Big(\mathbf{W_e}\left[\boldsymbol{h}_{\boldsymbol{e}}^{\boldsymbol{l-1}}||\boldsymbol{h}_{\boldsymbol{B(e)}}^{\boldsymbol{l-1}}||\boldsymbol{h}_{\boldsymbol{P(e)}}^{\boldsymbol{l-1}}\right]\Big), \quad (4)$$

where $\mathbf{W_e}$ is a learnable matrix for edge update. Here, we use a concatenation operation to aggregate messages. By stacking multiple layers, the network can learn more complex features from input spaces and better encode the local structural information. The final output embeddings for the beneficiary, provider, and claim are denoted as $\boldsymbol{z_b}$, $\boldsymbol{z_p}$ and $\boldsymbol{z_c}$, respectively. They will further be used in the downstream classification task.

3.2 Global Claim Graph Learning

To inject claim-wise inter-dependent knowledge into our detection model and measure the global gang-level risk for each claim behavior, we construct a claim-claim graph $\mathcal{G}_c(\mathcal{V}_c, \mathcal{E}_c)$. In particular, for claim event nodes $v_i^c, v_j^c \in \mathcal{V}_c$, there will be an edge $e_{i,j}^c$ linked between them if both transactions share the same beneficiary or provider. We devote our efforts to utilizing this relation graph to model a behavior-level network that is far more large and complex than the entity bipartite graph to better learn quickly changing organized conspiracy fraud patterns.

Community-aware Risk Diffusion

To effectively learn the far-reaching and quickly changing conspiracy fraud patterns, we propose a risk diffusion graph neural model in the *global* claim-claim graph. It is natural to employ the fraud classification labels of known nodes as the risk rating targets and use a label propagation algorithm to calculate the passed risks. However, this method has two limitations to our task. On the one hand, the traditional label propagation (LP) algorithm [Raghavan *et al.*, 2007] is based on an adjacency matrix that only considers the 1-hop neighborhood of the starting node and it is infeasible for risk information to reach remote nodes in a large criminal group. On the other hand, the traditional LP algorithm only manages the distance between two nodes in the graph and ignores the importance of node attributes. To better represent the risk rating of each node by combing the knowledge of a larger behavior community, we propose an improved risk diffusion algorithm based on a well-designed risk matrix (RM) $\mathbf{R_c}$, which is extended from the original adjacency matrix $\mathbf{A_c}$.

In a graph structure, the effect of multiple indirect diffusion is the same as direct information diffusion [Dong *et al.*, 2021]. This means if we have access to the multi-hop neighborhood of node v and approximate their direct risk diffusion degree, the diffusion range can be scaled during the iteration. Thus, in the implementation, we sample a subset of known fraud claim nodes in the training set as sources of risk infection S_c [Niu *et al.*, 2020]. The nodes chosen will be treated as starting nodes to build contagion links. Formally, given source nodes, we utilize a biased random walk [Grover and Leskovec, 2016] to explore their neighborhoods in depth-first traversal, taking into account the mixture of two notions of equivalence in a real-world network. Furthermore, we build direct links between source nodes and their multi-hop neighborhoods to extend the adjacency matrix as $\mathbf{A_c^{'}}$ and record their original hop distance. After that, for $\forall\, v_i^c, v_j^c \in \mathcal{V}_c$ those have a link in $\mathbf{A_c^{'}}$, we measure the degree of risk transmission from v_j^c to v_i^c as below:

$$Risk_{ij} = \frac{sim(v_i^c, v_j^c)}{\gamma \cdot hop_{ij}}, \tag{5}$$

where $sim(u, v)$ denotes the cosine similarity of nodes u and v's attributes, hop_{ij} refers to the hop number between them, and γ is a hyper-parameter that controls the relative importance of two criteria. We replace the link record in $\mathbf{A_c^{'}}$ with the risk diffusion degree computed above and normalized it by row to form our $\mathbf{R_c}$. Finally, we introduce the technique in personalized PageRank [Brin, 1998] to propagate the risk. We denote initial risk labels as $\mathbf{Y_r^{(0)}}$, which assigns 1 to the known fraud nodes in the training set and a small constant $\epsilon > 0$ to other nodes since the normal or unknown claims may be the camouflage of criminals and still have potential risk. The diffusion process can be expressed as:

$$\mathbf{Y_r^{(k)}} = (1-\alpha)\mathbf{R_c}\mathbf{Y_r^{(k-1)}} + \alpha\mathbf{Y_r^{(0)}}, \tag{6}$$

$$\mathbf{Y_{Risk}} = \mathbf{Y_r^{(K)}}, \tag{7}$$

where $\mathbf{Y_{Risk}}$ is the risk rating label matrix which incorporates community-aware risk information, K is the iteration number, and $\alpha \in (0, 1)$ represents a propagation hyper-parameter. The labels will be used in the pre-training task.

Pre-training for Risk Representation Learning

As mentioned before, we aim to incorporate gang-level risks into the healthcare fraud detection task. Inspired by recent advances in multi-task learning [Zhang and Yang, 2021], we devise an auxiliary pre-training prediction task to better learn the global pattern in a claim-level relation graph. Let $\mathbf{Y_{Risk}}$ be the soft labels of the regression task, and we leverage a vanilla GCN [Kipf and Welling, 2016] to learn community-aware risk representation for each claim node. We denote $\mathbf{H_c^{(l)}}$ as the input for the layer $\mathbf{l}$ in the network, and the layer-wise aggregation rule can be formulated as:

$$\mathbf{H_c^{(l+1)}} = \sigma\left(\tilde{\mathbf{D_c}}^{-\frac{1}{2}}\tilde{\mathbf{A_c}}\tilde{\mathbf{D_c}}^{-\frac{1}{2}}\mathbf{H_c^{(l)}}\mathbf{W_c^{(l)}}\right), \tag{8}$$

where $\tilde{\mathbf{A_c}} = \mathbf{A_c} + \mathbf{I_c}$ is the adjacency matrix with self-loops and $\tilde{\mathbf{D_c}}$ is the diagonal matrix of $\tilde{\mathbf{A_c}}$ representing the node degrees. The final result $\mathbf{H_c}$ through multi-layer transformation can be denoted as the risk representation $\boldsymbol{z_r}$. Then, a fully-connected neural layer is applied to predict the risk labels and we use back-propagation of mean squared error (MSE) loss to update the network in the pre-training stage.

3.3 Detection Network and Optimization

In the downstream detection task, after the concatenation of learned embeddings of beneficiary $\boldsymbol{z_b}$, provider $\boldsymbol{z_p}$, claim $\boldsymbol{z_c}$ and group-level risk attributes $\boldsymbol{z_r}$, we utilize a multi-layer perceptron (MLP) as the detection network to infer the fraud probability of a claim as:

$$\hat{p}_c = \boldsymbol{MLP}\big([\boldsymbol{z_b}||\boldsymbol{z_c}||\boldsymbol{z_p}||\boldsymbol{z_r}]\big). \tag{9}$$

For the edge classification task, we adopt the cross-entropy loss function for optimization, which can be formulated as:

$$\mathcal{L} = -\sum_i \left[y_{c,i}\log\left(\hat{p}_{c,i}\right) + (1-y_{c,i})\log\left(1-\hat{p}_{c,i}\right)\right], \tag{10}$$

where $y_{c,i} \in \mathcal{Y}_c$ is the ground-truth fraud label of the i-th claim edge and $\hat{p}_{c,i}$ denotes its predicted fraud probability.

Dataset	Type	Claim	Beneficiary	Provider
Training (Jan-Sep)	Outpatient	399,508	122,797	4,956
Test (Oct-Dec)	Outpatient	118,229	70,080	4,512
Training (Jan-Sep)	Inpatient	31,329	24,843	2,038
Test (Oct-Dec)	Inpatient	9,145	8,484	1,570

Table 1: The statistics of the dataset.

The proposed method can be optimized through the standard stochastic gradient descent-based algorithms. In this paper, we used the Adam optimizer [Kingma and Ba, 2014] to learn the parameters. We set the initial learning rate to 10^{-3} and the weight decay to 10^{-5} by default.

4 Experiments

4.1 Experimental Settings

Datasets

The dataset was collected from real-world medicare claims [Gupta, 2020], which include inpatient claims, outpatient claims, and beneficiary details. It contains more than 0.55 million claims (38.1% are labeled as fraudulent, according to the ground truth reported by the healthcare system) and over 0.2 million beneficiaries. After a thorough analysis of the dataset with our collaborating domain experts of the National Healthcare Security Administration (NHSA), we observed that many fraudulent activities involve multiple organized parties, known as conspiracy fraud. Then, the organized fraud labels are elaborately annotated by the anti-fraud experts of NHSA. They judge whether a claim is involved in organized fraud by their domain expertise in combating fraudsters over multiple years. According to the criterion of organized fraud in medical insurance procedures and existing fraud labels in the dataset, each record is annotated with two fraud labels indicating whether it is individual or organized fraud. In the experiment, we train the model on the dataset of the first nine months (from January to September) and detect on the rest three months for testing (from October to December). Table 1 reports detailed statistics of the dataset, including both the inpatient and outpatient claims. We train and evaluate our proposed model by individual-level and gang-level label settings for organized fraud detection tasks. We will contribute the annotated labels of the dataset for the research community to inspire more work in the future.

Baseline Methods

We compare our proposed techniques with the baselines as shown below:

- **RF** [Bauder and Khoshgoftaar, 2018]: A widely-used random forest-based fraud detection method.
- **DNN** [Kazemi and Zarrabi, 2017]: A typical deep neural network that includes three hidden layers.
- **GCN** [Kipf and Welling, 2016]: A well-known graph neural network that performs similar operations as CNN to learn the features by inspecting neighboring nodes.
- **GAT** [Liu *et al.*, 2021]: A powerful model with the attention mechanism for graph learning. We set the attention head to 4 and the number of stacked layers to 2.
- **DCI** [Wang *et al.*, 2021]: A simple yet effective graph self-supervised learning scheme for node representation learning, which captures the intrinsic graph properties in more concentrated feature spaces by clustering the entire graph into multiple parts.
- **AMNet** [Chai *et al.*, 2022]: A GNN-based model aiming to capture both low-frequency and high-frequency signals, and adaptively combining signals of different frequencies to detect fraud samples that are dissimilar to their neighboring normal nodes.
- **BWGNN** [Tang *et al.*, 2022]: A popular GNN approach for mining structural data with spectral and spatial localized band-pass filters to better detect fraud patterns.

Parameter Settings and Evaluation Metrics

In our implementation, we first pre-train a two-layer GCN for generating risk embedding, which has a dimension of 32. The number of attentional layers in the entity graph is set as 2 and the dimensions of output representation z_b, z_c, z_p are 16, 32, 16, respectively. In the parallel training phase, the maximum number of epochs is set to 100. We adopt a dropout mechanism with the rate of 0.6. Our method is implemented using PyTorch 1.12.1 with CUDA 11.3 and Python 3.7 as the backend. The model is trained on a server with two 32GB NVIDIA Tesla V100 GPUs. In the experiment, we leverage three commonly-used metrics: area under curve (AUC), recall, and F1 score to evaluate the effectiveness of our model comprehensively. For all three metrics, the higher score indicates the higher performance of the methods.

4.2 Overall Performance Comparison

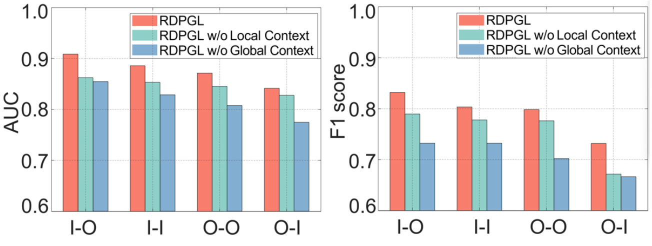
We evaluate our method and all compared baselines in two healthcare fraud detection tasks: individual-level and gang-level (organized) fraud. Table 2 shows the model performance in terms of three metrics. It can be seen that our proposed RDPGL consistently obtains better results across different fraud types, which proves the effectiveness of the proposed method.

Compared with traditional approaches, GNN-based methods achieve better performance due to the modeling of high-order interactions among different behaviors. We observe that GCN and GAT perform inadequately w.r.t. RDPGL, which confirms the necessity of learning representations from the heterogeneous beneficiary-provider relation graph. In addition, different from state-of-the-art models proposed in recent years, our RDPGL could also incorporate risk diffusion in the action-level claim-claim graph and generates more effective fraud risk features based on global claim interactions, leading to more accurate prediction performance.

As we can see, the proposed RDPGL obtains significant improvements in organized fraud detection. It outperforms DCI and BWGNN, which are two of the strongest baselines, by 3%-5% in terms of AUC in both inpatient and outpatient scenarios. The model achieves the highest F1 score and also the highest recall, which means that RDPGL can significantly

Model	Individual Healthcare Fraud						Organized Healthcare Fraud					
	I-Outpatient			I-Inpatient			O-Outpatient			O-Inpatient		
	AUC	Recall	F1	AUC	Recall	F1	AUC	Recall	F1	AUC	Recall	F1
RF (2018)	0.7337	0.5719	0.5852	0.7022	0.5669	0.5847	0.7327	0.5256	0.4924	0.6921	0.5349	0.4881
DNN (2017)	0.7425	0.5913	0.6076	0.7085	0.5819	0.5523	0.7181	0.5311	0.4858	0.6813	0.5157	0.4896
GCN (2021)	0.7743	0.5986	0.6417	0.7557	0.6036	0.6369	0.7598	0.5472	0.6189	0.7245	0.5584	0.6332
GAT (2021)	0.7822	0.6886	0.7126	0.7437	0.6348	0.6581	0.7719	0.5912	0.6542	0.7415	0.6018	0.6356
AMNet (2022)	0.7819	0.7532	0.7953	0.7577	0.7236	0.6847	0.7583	0.6456	0.6132	0.7129	0.6253	0.6339
BWGNN (2022)	0.8512	0.8471	0.7582	0.8129	0.8176	0.7312	0.8258	0.8098	0.6375	0.7983	0.7349	0.6353
DCI (2021)	0.8679	0.8536	**0.8472**	0.8361	0.8083	0.7754	0.8353	0.7994	0.7486	0.8195	0.7719	0.6658
RDPGL	**0.9086**	**0.8553**	0.8319	**0.8858**	**0.8311**	**0.8032**	**0.8713**	**0.8302**	**0.7982**	**0.8415**	**0.8022**	**0.7317**

Table 2: The performance comparison of our proposed method with seven state-of-the-art baselines in individual("I-") and organized("O-") healthcare fraud detection tasks. The result proves that our method significantly outperforms recent baselines in most metrics.



(a) AUC

(b) F1 Score

Figure 2: Ablation studies for different RDPGL framework variants, in terms of AUC and F1.

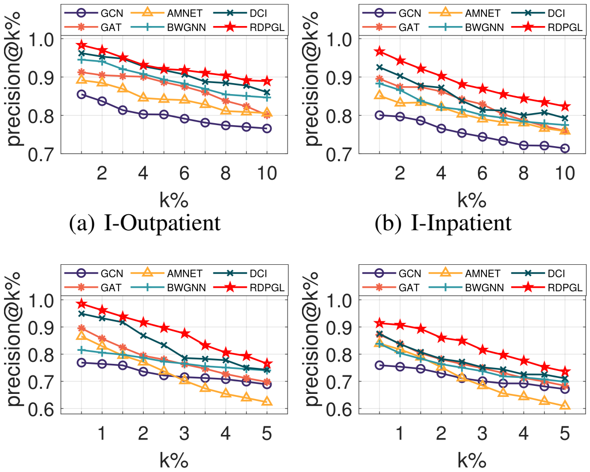
reduce the false-negative rate and accurately detect real fraud activities, demonstrating its superiority in extracting complicated behavior patterns from gang fraud. Additionally, we see that the model performance on inpatient data is generally worse than that on outpatient data, possibly due to the occurrence of data imbalance and the distinct hidden patterns between inpatient and outpatient records. In particular, DCI is the only baseline that can achieve over 0.7 F1 score in the inpatient data and it is significantly superior to other baseline models, while our RDPGL can achieve near 0.8 F1 score. Besides, only RDPGL reaches over 80% recall and over 0.7 F1 score in terms of organized fraud detection on inpatient data, proving its effectiveness in capturing gang-level fraud claim behaviors. In summary, our method retains its effectiveness for two different detection tasks in both inpatient and outpatient settings, exhibiting its robustness trait.

4.3 Ablation Study

We further investigate two variants of RDPGL to explore the effectiveness of each component of our proposed approach:

- **RDPGL w/o Local Context**: Removing the heterogeneous attentional graph neural network during parallel learning which only employs the risk embedding learned from the claim relation network.
- **RDPGL w/o Global Context**: Removing the original community-aware risk diffusion model and replacing it with the traditional label propagation algorithm for the pre-training task.

We report the comparisons of detection performance in terms of AUC and F1 score for the above-described model configurations in Figure 2. RDPGL outperforms both counterparts on different fraud detection tasks. Specifically, the RDPGL w/o Global Context performs worst on different datasets, indicating that *local* and *global* context in risk embedding learned by the parallel pre-trained graph learning module plays an important role in detecting both individual-level and gang-level fraud. The RFPGL performs better than the RDPGL w/o Local Context, especially in organized fraud detection. The reason is that the variant fails to propagate community-level risk messages and incorporate node attributes in the diffusion process as RDPGL does, thus generating a less representative risk embedding vector and resulting in suboptimal performances in both metrics. This experiment validates the essential and effectiveness of each component of our proposed method in healthcare fraud detection.



(a) I-Outpatient

(b) I-Inpatient

(c) O-Outpatient

(d) O-Inpatient

Figure 3: The precision@k for individual and organized fraud detection in both inpatient and outpatient settings.

4.4 Ranking Performance for Detection

Due to the humongous quantity of medical insurance claim data, it is almost impossible to annotate the whole dataset in a real industrial scenario. The accuracy of a fraud detection algorithm is usually evaluated with the precision of the top returned fraud detection results, in which experts only need to

Metric	F1-Score				
ϵ α	0.05	0.1	0.15	0.2	0.25
0.1	0.7398	0.7458	0.7510	0.7426	0.7305
0.2	0.7533	0.7620	**0.7649**	0.7593	0.7421
0.3	0.6879	0.7012	0.7245	0.7119	0.7023

Table 3: Impact of the propagation coefficientα and the initial potential riskϵ on organized fraud detection.

check those high-risk candidates (a tiny portion of the whole dataset). In the following experiment, we utilize the precision of the predicted top k percentage of fraud claims with the highest confidence to evaluate the detection performance. In particular, we select the top 1%–10% of most confident fraud claims predicted by RDPGL and other GNN-based approaches for individual-level fraud, while for organized fraud, we choose the range of 0.5%–5% because the positive samples of this fraud type are relatively small. As shown in Figure 3, the precision of baselines gradually decreases with the increase of k for the reason that the more samples predicted, the less reliable the result is. We observe that the precision of GCN is not comparable with other methods and the latest deep graph methods, such as BWGNN and DCI, achieve the best performance among these baselines. Our RDPGL performs better than other models in all settings of k on all datasets. It achieves an average of over 95% precision for the top 1% of predictions on both detection tasks, which is better than the performances of compared baselines. Specifically, for organized fraud detection in an inpatient scenario, the improvements vary from 1% to 9%. The improvements are more remarkable in the top 3% predictions. These results strongly prove the superiority of RDPGL in real industrial scenarios.

4.5 Parameter Sensitivity

We further investigate the model generalization performance on hyper-parameters of initial potential risk ϵ and propagation parameter α. Specifically, we evaluate their influence on risk diffusion on organized fraud detection tasks and present averaged F1 score of performance on inpatient and outpatient test data in Table 3. We vary the initial potential risk ϵ for the normal node from 0.05 to 0.25 with an incremental step of 0.05 and the propagation coefficient α is searched from the set of $\{0.1, 0.2, 0.3\}$. We can observe that our RDPGL performs better when increasing ϵ from 0.05 to 0.15, and the average F1 score reaches the peak when $\epsilon = 0.15$ and $\alpha = 0.2$. The performance degrades if we keep increasing the ϵ value. We suspect that the overestimation of risk diffused by legal behavior may lead to the over-sensitivity of the model on discriminating fraud claims. As the α increases from 0.2 to 0.3, we can find the performance decreases even quicker. This is probably due to the higher α restricting the diffusion breadth and later deteriorating the learning capacity of community-level risk representation.

4.6 Case Studies

In RDPGL, the learned representation is utilized to indicate the risk information diffused across the community and help improve the performance of organized fraud detection.

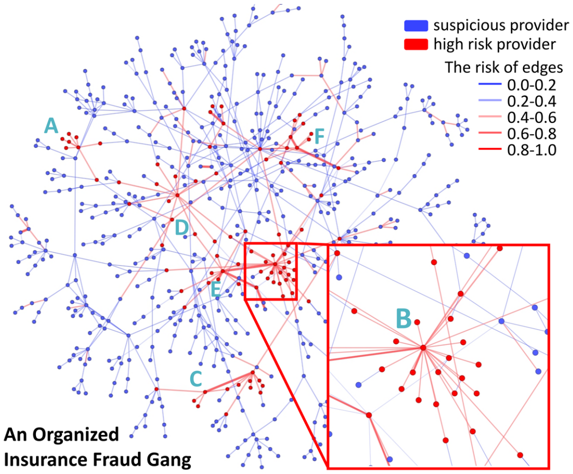


Figure 4: The layout of a typical medical service provider graph. There are six criminal groups marked as A, B, C, D, E and F. Our method successfully detects most of these organized fraud claims where the color of the edges denotes the detected fraud probability.

Figure 4 presents a typical case of a medical service provider network with 663 providers. Edges between two providers denote that they have claims with the same beneficiaries and the color and thickness show the fraud probability, which is produced by our proposed method, of these claims. We denote the provider with claims of higher fraud probability as a high-risk provider and others as low-risk providers. There are six organized fraud gangs annotated by domain experts, which are marked as A, B, C, D, E and F. It is apparent that the fraud gangs have a cluster-like pattern and our method successfully detects most claims in these criminal groups. To further analyze the organized fraud pattern, we zoom in on group B with red squares and it shows that most neighboring nodes are involved in organized fraud and form a cluster structure, which reveals that our well-designed community-aware risk diffusion method is suitable for detecting this fraud pattern. From high-risk providers and their neighboring suspicious nodes, we can observe that the claims' fraud probability is relevant to its behavior entities, and the integration of information from the heterogeneous entity is essential.

5 Conclusion

In this paper, we proposed a novel risk diffusion-based parallel graph learning system for organized fraud detection in medical insurance claims. To better extract constantly-updated organized fraud patterns, we devised the heterogeneous graph attentional network for *local* context learning on the beneficiary-provider graph and the community-aware risk diffusion graph model for quickly changing fraudster behavior's *global* context learning on the claim-claim graph, which is then jointly optimized in the downstream detection network. Extensive experiments on the real-world dataset demonstrated the advantages of our method over existing state-of-the-art baselines. The superior performance of RDPGL also exposes the vulnerabilities of organized fraudsters and safeguards our healthcare insurance system.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant no. 62102287) and the Shanghai Science and Technology Innovation Action Plan Project (Grant no. 22YS1400600 and 22511100700)

References

[Bauder and Khoshgoftaar, 2017] Richard A Bauder and Taghi M Khoshgoftaar. Medicare fraud detection using machine learning methods. In *16th IEEE international conference on machine learning and applications (ICMLA)*, pages 858–865. IEEE, 2017.

[Bauder and Khoshgoftaar, 2018] Richard Bauder and Taghi Khoshgoftaar. Medicare fraud detection using random forest with class imbalanced big data. In *2018 IEEE international conference on information reuse and integration (IRI)*, pages 80–87. IEEE, 2018.

[Brin, 1998] Sergey Brin. The pagerank citation ranking: bringing order to the web. *Proceedings of ASIS, 1998*, 98:161–172, 1998.

[Chai *et al.*, 2022] Ziwei Chai, Siqi You, Yang Yang, Shiliang Pu, Jiarong Xu, Haoyang Cai, and Weihao Jiang. Can abnormality be detected by graph neural networks? In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI), Vienna, Austria*, pages 23–29, 2022.

[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, Xiaoyang Wang, Ying Zhang, and Liqing Zhang. Risk guarantee prediction in networked-loans. In *IJCAI International Joint Conference on Artificial Intelligence*, pages 1–7, 2020.

[Cheng *et al.*, 2023] Dawei Cheng, Zhibin Niu, Jie Li, and Changjun Jiang. Regulating systemic crises: Stemming the contagion risk in networked-loans through deep graph learning. *IEEE Transactions on Knowledge and Data Engineering*, 35:6278–6289, 2023.

[Dong *et al.*, 2021] Hande Dong, Jiawei Chen, Fuli Feng, Xiangnan He, Shuxian Bi, Zhaolin Ding, and Peng Cui. On the equivalence of decoupled graph convolution network and label propagation. In *Proceedings of the Web Conference 2021*, pages 3651–3662, 2021.

[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.

[Dua and Bais, 2014] Prerna Dua and Sonali Bais. Supervised learning methods for fraud detection in healthcare insurance. pages 261–285, 2014.

[Grover and Leskovec, 2016] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864, 2016.

[Gupta, 2020] Rohit Anand Gupta. Healthcare provider fraud detection analysis. https://www.kaggle.com/datasets/rohitrox/healthcare-provider-fraud-detection-analysis, 2020. Accessed: 202-01-25.

[Hearn Jr, 1989] Francis J Hearn Jr. Curing the health care industry: Government response to medicare fraud and abuse. *J. Contemp. Health L. & Pol'y*, 5:175, 1989.

[Kazemi and Zarrabi, 2017] Zahra Kazemi and Houman Zarrabi. Using deep networks for fraud detection in the credit card transactions. In *2017 IEEE 4th International conference on knowledge-based engineering and innovation (KBEI)*, pages 0630–0633. IEEE, 2017.

[Kingma and Ba, 2014] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.

[Kipf and Welling, 2016] Thomas Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *ArXiv*, abs/1609.02907, 2016.

[Li *et al.*, 2019] Ao Li, Zhou Qin, Runshi Liu, Yiqun Yang, and Dong Li. Spam review detection with graph convolutional networks. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 2703–2711, 2019.

[Liang *et al.*, 2019] Chen Liang, Ziqi Liu, Bin Liu, Jun Zhou, Xiaolong Li, Shuang Yang, and Yuan Qi. Uncovering insurance fraud conspiracy with network learning. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1181–1184, 2019.

[Liu *et al.*, 2021] Can Liu, Li Sun, Xiang Ao, Jinghua Feng, Qing He, and Hao Yang. Intention-aware heterogeneous graph attention networks for fraud transactions detection. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 3280–3288, 2021.

[Ma *et al.*, 2018] Jun Ma, Danqing Zhang, Yun Wang, Yan Zhang, and Alexey Pozdnoukhov. Graphrad: a graph-based risky account detection system. In *Proceedings of ACM SIGKDD conference, London, UK*, volume 9, 2018.

[Major and Riedinger, 1992] John A Major and Dan R Riedinger. Efd: A hybrid knowledge/statistical-based system for the detection of fraud. *International Journal of Intelligent Systems*, 7(7):687–703, 1992.

[NHCAA, 2021] NHCAA. The challenge of health care fraud. https://www.nhcaa.org/tools-insights/about-health-care-fraud/the-challenge-of-health-care-fraud, 2021. Accessed: 2023-01-19.

[Niu *et al.*, 2020] Zhibin Niu, Runlin Li, Junqi Wu, Dawei Cheng, and Jiawan Zhang. iconviz: Interactive visual exploration of the default contagion risk of networked-guarantee loans. In *2020 IEEE conference on visual analytics science and technology (VAST)*, pages 84–94. IEEE, 2020.

[Pandey, 2017] Yamini Pandey. Credit card fraud detection using deep learning. *International Journal of Advanced Research in Computer Science*, 8(5), 2017.

[Pumsirirat and Liu, 2018] Apapan Pumsirirat and Yan Liu. Credit card fraud detection using deep learning based on auto-encoder and restricted boltzmann machine. *International Journal of advanced computer science and applications*, 9(1), 2018.

[Raghavan *et al.*, 2007] Usha Nandini Raghavan, Réka Albert, and Soundar Kumara. Near linear time algorithm to detect community structures in large-scale networks. *Physical review E*, 76(3):036106, 2007.

[Rudman *et al.*, 2009] William J Rudman, John S Eberhardt, William Pierce, and Susan Hart-Hester. Healthcare fraud and abuse. *Perspectives in Health Information Management/AHIMA, American Health Information Management Association*, 6(Fall), 2009.

[Stowell *et al.*, 2018] Nicole F Stowell, Martina Schmidt, and Nathan Wadlinger. Healthcare fraud under the microscope: improving its prevention. *Journal of Financial Crime*, 25(4):1039–1061, 2018.

[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.

[Timofeyev and Jakovljevic, 2022] Yuriy Timofeyev and Mihajlo Jakovljevic. Fraud and corruption in healthcare. *Frontiers in Public Health*, 10, 2022.

[Villegas-Ortega *et al.*, 2021] José Villegas-Ortega, Luciana Bellido-Boza, and David Mauricio. Fourteen years of manifestations and factors of health insurance fraud, 2006–2020: a scoping review. *Health & justice*, 9(1):1–23, 2021.

[Waghade and Karandikar, 2018] Shivani S Waghade and Aarti M Karandikar. A comprehensive study of healthcare fraud detection based on machine learning. *International Journal of Applied Engineering Research*, 13(6):4175–4178, 2018.

[Wang *et al.*, 2021] Yanling Wang, Jing Zhang, Shasha Guo, Hongzhi Yin, Cuiping Li, and Hong Chen. Decoupling representation learning and classification for gnn-based anomaly detection. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, pages 1239–1248, 2021.

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

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

[Zhang and Yang, 2021] Yu Zhang and Qiang Yang. A survey on multi-task learning. *IEEE Transactions on Knowledge and Data Engineering*, 34(12):5586–5609, 2021.

[Zhang *et al.*, 2022] Ge Zhang, Zhao Li, Jiaming Huang, Jia Wu, Chuan Zhou, Jian Yang, and Jianliang Gao. efraudcom: An e-commerce fraud detection system via competitive graph neural networks. *ACM Transactions on Information Systems (TOIS)*, 40(3):1–29, 2022.